



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

ZABEZPEČENÁ WEBOVÁ APLIKACE PRO PLOŠNÝ MONITORING TEPLoty ZALOŽENA NA METODÁCH STROJOVÉHO UČENÍ

SECURE WEB APPLICATION FOR SURFACE TEMPERATURE MONITORING BASED ON MACHINE
LEARNING METHODS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Jan Hejna

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. et Ing. Petr Musil

BRNO 2023



Bakalářská práce

bakalářský studijní program **Informační bezpečnost**

Ústav telekomunikací

Student: Jan Hejna

ID: 221545

Ročník: 3

Akademický rok: 2022/23

NÁZEV TÉMATU:

Zabezpečená webová aplikace pro plošný monitoring teploty založena na metodách strojového učení

POKYNY PRO VYPRACOVÁNÍ:

V rámci teoretické části práce prozkoumejte vliv slunečního záření na provoz venkovních telekomunikačních zařízení a možnosti kvantifikace tohoto vlivu při využívání hodnot teplot oportunistických senzorů.

Součástí práce je také korelační analýza získaných hodnot oportunistických senzorů s hodnotami naměřenými pomocí zvolených meteorologických stanic.

V praktické části navrhnete algoritmus, který bude schopen vytvářet 2D teplotní mapy a také kvantifikovat tepelný přírůstek venkovních telekomunikačních zařízení způsobený slunečním zářením na základě metod strojového učení. Klíčovým požadavkem tohoto návrhu je zajištění naprostého skrytí pozic koncových telekomunikačních zařízení (oportunistických senzorů) koncovému uživateli a zabezpečení celé webové aplikace, aby nemohlo dojít k žádnému úniku citlivých informací.

Výstupem bakalářské práce je návrh webové aplikace, která bude umožňovat zobrazení 2D teplotní mapy s posuvnou časovou osou. Výstup bude sloužit jako zabezpečený demonstrátor výzev v oblasti provozu a měření oportunistickými metodami využitím telekomunikačních zařízení a metod strojového učení.

DOPORUČENÁ LITERATURA:

Podle pokynů vedoucího práce

Termín zadání: 6.2.2023

Termín odevzdání: 26.5.2023

Vedoucí práce: Ing. et Ing. Petr Musil

doc. Ing. Jan Hajný, Ph.D.
předseda rady studijního programu

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Bakalářská práce je zaměřena na prozkoumání vlivu slunečního záření na naměřenou teplotu teplotními senzory v mikrovlnných zařízeních a na možnosti kvantifikace tohoto vlivu. Práce zkoumá možnost využití těchto zařízení pro měření teploty okolí pomocí zvolených metod strojového učení. Cílem bylo vytvoření zabezpečené webové aplikace, která bude sloužit pro plošný monitoring teploty okolí a rozšířit tak stávající sítě meteorologických měřících bodů.

Teoretická část práce se zabývá technologií mikrovlnných spojů a okolními vlivy, které působí na naměřenou teplotu těchto zařízení.

V praktické části práce byla provedena korelační analýza mezi naměřenými teplotami z různých technologií telekomunikačních zařízení a teplotami okolních meteorologických stanic. Analýza si kladla za cíl výběr vhodné technologie pro predikci okolní teploty. Další část praktické práce zahrnovala proces trénování modelu strojového učení. Poslední částí zahrnuje vývoj webové aplikace zobrazující teplotní mapy České republiky.

KLÍČOVÁ SLOVA

mikrovlnné zařízení, sluneční záření, teplota oportunistických senzorů, teplotní přírůstek, měření teploty okolí, plošný monitoring, strojové učení, Lineární regrese, webová aplikace, Python.

ABSTRACT

The bachelor thesis is focused on the exploration of the influence of solar radiation on the measured temperature by temperature sensors in microwave devices and on the possibilities of quantifying this influence. The thesis explores the possibility of using these devices to measure ambient temperature using selected machine learning methods. The aim was to create a secure web application that will be used for area-based ambient temperature monitoring, thus expanding existing networks of meteorological measurement points.

The theoretical part of the thesis focuses on the technology of microwave links and the environmental influences that affect the measured temperature of these devices.

In the practical part of the thesis, a correlation analysis was performed between the measured temperatures from different telecommunication equipment technologies and the temperatures of the surrounding meteorological stations. The analysis aimed at selecting an appropriate technology for ambient temperature prediction. The next part of the practical work involved the process of training a machine learning model. The last part involves the development of a web application that displays temperature maps of the Czech Republic.

KEYWORDS

microwave devices, solar radiation, opportunistic sensor temperature, temperature increment, ambient temperature monitoring, area monitoring, machine learning, linear regression, web application, Python.

HEJNA, Jan. *Zabezpečená webová aplikace pro plošný monitoring teploty založena na metodách strojového učení*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2023, 93 s. Bakalářská práce. Vedoucí práce: Ing. et. Ing. Petr Musil

Prohlášení autora o původnosti díla

Jméno a příjmení autora:	Jan Hejna
VUT ID autora:	221545
Typ práce:	Bakalářská práce
Akademický rok:	2022/23
Téma závěrečné práce:	Zabezpečená webová aplikace pro plošný monitoring teploty založena na metodách strojového učení

Prohlašuji, že svou závěrečnou práci jsem vypracoval samostatně pod vedením vedoucí/ho závěrečné práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....

podpis autora*

*Autor podepisuje pouze v tištěné verzi.

PODĚKOVÁNÍ

Děkuji vedoucímu mé bakalářské práce panu Ing. et Ing. Petru Musilovi za odborné vedení, konzultace, trpělivost a cenné rady a návrhy při zpracování bakalářské práce. Zvláště pak panu Ing. Milanu Bubniakovi za ochotnou spolupráci. Rád bych chtěl také poděkovat své rodině za podporu v průběhu studia.

Obsah

Úvod	12
1 Mikrovlnné spoje	13
1.1 Mikrovlnná zařízení	14
1.1.1 Mikrovlnná anténa	14
1.1.2 Vnější jednotka(ODU)	14
1.1.3 Vnitřní jednotka (IDU)	15
2 Faktory ovlivňující naměřenou hodnotu senzorů	16
2.1 Sluneční záření	17
2.2 Jouleovo teplo	19
2.3 Výskyt srážek	20
2.4 Umístění mikrovlnné jednotky	21
3 Statistické metody analýzy korelace a pravděpodobnostního rozdělení	24
3.1 Korelační analýza	24
3.1.1 Pearsonův korelační koeficient	24
3.1.2 Spearmanův koeficient pořadové korelace	25
3.1.3 Kendallův koeficient pořadové korelace	27
3.2 Pravděpodobnostní rozdělení souborů dat	27
3.3 Gaussovo normální rozdělení	27
3.3.1 Šikmost a špičatost	28
4 Nástroje pro analýzu a zkoumání vlivů	31
4.1 Nástroj pro výpočet Offsetu	31
4.2 Kvantifikace tepelného přírůstku vyprodukovaný slunečním zářením	36
5 Statistická analýza	39
5.1 Testování normální distribuce naměřených teplot	39
5.2 Korelační analýza	44
6 Vytvoření modelu založeného na strojovém učení	55
6.1 Shromažďování a příprava dat	55
6.2 Výběr vhodného modelu strojového učení	56
6.3 Natrénování modelu strojového učení	58

7	Vývoj zabezpečené webové aplikace	62
7.1	Návrh webové aplikace	62
7.2	Použité technologie	62
7.2.1	Databáze	62
7.2.2	Hlavní knihovny	63
7.3	Architektura aplikace	65
7.3.1	Backendová část	66
7.3.2	Frontendová část	70
7.4	Definování tras a komunikace s backendovou částí	72
7.4.1	Zabezpečení citlivých informací	73
7.5	Testování predikovaných teplot	74
7.5.1	Ověření přesnosti predikce	75
7.6	Kvantifikace teplotní přírůstku	79
7.6.1	Výsledky kvantifikace teplotní přírůstku	80
	Závěr	82
	Literatura	84
	Seznam symbolů a zkratk	88
	Seznam příloh	90
	A Ukázka webové aplikace	91
	B Obsah elektronické přílohy	93

Seznam obrázků

1.1	Rozdělení mikrovlnného spektra.	13
1.2	Blokové schéma vysílače.	15
1.3	Blokové schéma přijímače.	15
2.1	Faktory působící na senzor měření teplot mikrovlnné jednotky	17
2.2	Graf průběhu intenzity slunečního záření	18
2.3	Vliv výskytu srážek na změnu naměřené teploty	20
2.4	Rozložení mikrovlnných jednotek na střeše panelové budovy	21
2.5	Výškový rozdíl senzorů měření teplot	22
2.6	Výškový rozdíl senzorů měření teplot	23
3.1	Ukázkové grafy Pearsonové korelace	25
3.2	Ukázkové grafy Spearmanové pořadové korelace	26
3.3	Grafické vyjádření Gaussova normálního rozdělení pro náhodnou veličinu X	28
3.4	Ilustrační obrázek pozitivní a negativní šikmosti	29
3.5	Ilustrační obrázek různých typů špičatosti křivek	29
4.1	Zobrazení offsetu mezi mikrovlnnou jednotkou a ČHMÚ jednotkou	32
4.2	Průběh teplotního offsetu	33
4.3	Průběh teplotního offsetu v noci	34
4.4	Teplotní offset vyprodukovaný Jouleovým teplem	35
4.5	Kružnice popisující situaci	36
4.6	Průběhy hodnot teplotního offsetu mikrovlnného spoje na obou stranách spoje	37
5.1	Průběh grafu pro hodnoty naměřené s časovým intervalem 10 minut	43
5.2	Průběh grafu po zprůměrování hodnot na 2 hodiny	43
6.1	Histogram reziduí	60
6.2	Vztah mezi skutečnými a predikovanými teplotami	61
7.1	Výběrový seznam konkrétních hodin	72
7.2	Interaktivní teplotní mapa	73
7.3	Referenční teplotní mapa - 9.5.2023 15:00	75
7.4	Interpolovaná a okresová teplotní mapa	76
7.5	Referenční teplotní mapa	77
7.6	Interpolovaná a okresová teplotní mapa	78
7.7	Průběhy teplotních přírůstků vyprodukované slunečním zářením	81
A.1	Interaktivní teplotní mapa okresů	91
A.2	Výběrový seznam pro vykreslení interpolované teplotní mapy	91
A.3	Interpolovaná teplotní mapa	92

Seznam tabulek

5.1	Výsledky testování normality pro hodnoty s časovým intervalem 10 minut	40
5.2	Výsledky testování normality pro hodnoty s časovým intervalem 30 minut	41
5.3	Výsledky testování normality pro hodnoty s časovým intervalem 1 hodiny	42
5.4	Výsledky testování normality pro hodnoty s časovým intervalem 2 hodin	42
6.1	První záznamy parametrů MW zařízení	56
6.2	Koeficienty metrik pro jednotlivé modely strojového učení	57
6.3	Výsledky metrik modelu	59

Seznam výpisů

4.1	Ukázka vynesení hodnot z jednotek	31
4.2	Odečtení teplotních hodnot	32
6.1	Ukázka trénování a ukládání modelu v Pythonu	59
7.1	Navázání spojení s databázemi	66
7.2	Funkce pro určení vlivu slunečního záření	67
7.3	Predikce teploty na základě vstupních parametrů	68
7.4	Funkce pro přiřazení okresu jednotlivým záznamům	68
7.5	Funkce pro přiřazení okresu jednotlivým záznamům	69
7.6	Callback funkce pro funkcionalitu časové osy	70

Úvod

Venkovní telekomunikační zařízení hrají klíčovou roli v přenosu informací, avšak jsou vystavena různým vnějším meteorologickým faktorům, které mohou ovlivnit jejich provoz. Jedním z těchto faktorů je sluneční záření, jehož vliv na naměřenou teplotu těchto zařízení je relativně neprobádaným tématem.

Bakalářská práce se zabývá prozkoumáním vlivu slunečního záření na naměřené teploty mikrovlnných zařízení. Zvláštní pozornost je věnována možnostem kvantifikace tohoto vlivu pomocí oportunistických senzorů, které teplotu měří. Cílem této bakalářské práce je nejen porozumět, jak sluneční záření ovlivňuje naměřené teploty, ale také otestovat, zdali je možné pomoci těmto telekomunikačním zařízením měřit teplotu okolí oportunistickou metodou. Výstupem práce je zabezpečená webová aplikace, zobrazující 2D teplotní mapu České republiky, která vykresluje průběh predikovaných teplot s použitím natrénovaného modelu strojového učení. Při vývoji webové aplikace byl kladen důraz na naprosté skrytí geografických pozic telekomunikačních zařízení před koncovým uživatelem a zabezpečení celé webové aplikace tak, aby nemohlo dojít k žádnému úniku citlivých informací.

Webová aplikace by mohla poskytnout významný přínos prostřednictvím potenciálního rozšíření stávajících meteorologických sítí měřících bodů pro monitorování teploty okolí, což by následně mohlo vést ke zvýšení přesnosti a spolehlivosti meteorologických dat.

Součástí práce je také korelační analýza, která porovnává naměřené hodnoty teplot komunikačními zařízeními s hodnotami naměřenými pomocí vybraných meteorologických stanic. Tato analýza poskytuje podklad pro výběr vhodné technologie pro předpovídání okolní teploty.

Práce je rozdělena do 7 kapitol. První kapitola poskytuje stručný přehled o mikrovlnných spojích. Druhá kapitola se zabývá konkrétními faktory ovlivňující naměřenou teplotu mikrovlnných jednotek. Třetí kapitola poskytuje teoretický přehled různých typů korelačních koeficientů, které byly použity pro korelační analýzu naměřených teplot. Čtvrtá kapitola popisuje vývoj nástrojů pro analýzu a kvantifikaci vlivů, které působí na naměřenou teplotu. Nástroje byly vytvořeny v databázi ukládající veškeré hodnoty naměřené mikrovlnným zařízením a vybranými meteorologickými stanicemi. Pátá kapitola se zabývá statistickou analýzou naměřených teplot, konkrétně zkoumá pravděpodobnostní rozdělení naměřených teplot a také korelační koeficienty jednotlivých technologií. Šestá kapitola popisuje samotný proces vytvoření modelu strojového učení, od shromažďování a přípravy dat až po výběr vhodného modelu a jeho natrénování. V poslední, sedmé kapitole je podrobně popsán vývoj zabezpečené webové aplikace, včetně návrhu aplikace, použitých technologií, architektury aplikace a testování predikovaných teplot.

1 Mikrovlnné spoje

Název mikrovlnné spoje reprezentuje rádiové bezdrátové spojení mezi dvěma body v síťové topologii označované termínem point-to-point. Hlavním důvodem použití dvoubodové topologie je malá vlnová délka vysílaných paprsků, která dosahuje rozsahu 1 mm - 1 m. Elektromagnetické vlny jsou tedy nasměrovány v úzkých paprscích přímo na přijímací anténu, díky čemuž jsme schopni šetřit šířku pásma rádiového spektra, jelikož nedochází k vzájemnému rušení mikrovlnných zařízení. Mikrovlny jsou elektromagnetické vlny o kmitočtu mezi 0,3 – 300 GHz, jejichž frekvence nepřímo úměrně koreluje s jejich vlnovou délkou, což můžeme vidět na obrázku níže. [1]:

Frekvence	3 30 300 kHz 3 30 300 MHz 3 30 300 GHz
Vlnová délka	1 10 km 1 100 m 10 1 100 mm 10 1
Označení	VLB LF MF HF VHF UHF SHF EHF

Obr. 1.1: Rozdělení mikrovlnného spektra. Převzato z [1]

Mikrovlnné spoje můžeme rozdělit do dvou kategorií:

- Analogové spoje
- Digitální spoje [2]

V současné době se však analogové spoje téměř nevyužívají právě kvůli jednoduchosti, spolehlivosti a menším nákladům na zavedení digitálního přenosu. Postupný přechod z analogových spojů na digitální se označuje jako konvergence. [2] Pro zdařilé zavedení mikrovlnných spojů je třeba, aby byla zajištěna přímá viditelnost mezi oběma stanicemi spoje, jelikož oproti rádiovým vlnám s nižší frekvencí mikrovlnné paprsky nedokáží překonávat překážky, které jim stojí v přímé cestě. Vzdálenost mezi spolupracujícími stanicemi obvykle nepřekračuje desítky kilometrů. Při větších vzdálenostech je nutné do spoje vkládat mezistanice (retranslační stanice), aby podmínka přímé viditelnosti byla splněna. Takový spoj se pak nazývá mikrovlnný reléový spoj. Mezistanice přijme vysílaný signál, který zesílí a následně ho vysílá další stanici v topologii. [1]

1.1 Mikrovlnná zařízení

Typická architektura mikrovlnného zařízení se skládá z mikrovlnné antény, vnější a vnitřní jednotky. V následující kapitole si tyto části popíšeme.

1.1.1 Mikrovlnná anténa

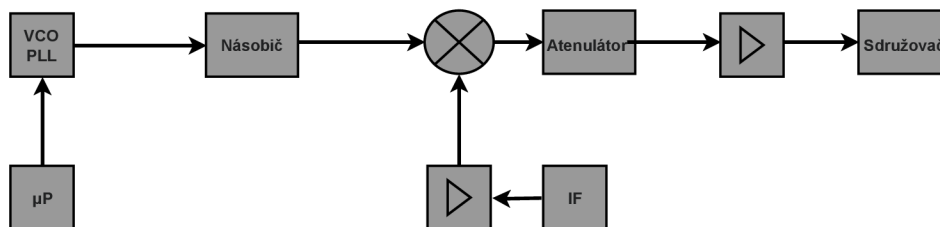
V případě mikrovlnných spojů se zpravidla používají směrové antény, které jsou charakteristické svým malým vyzařovacím úhlem. Primárním úkolem antén je vysílání a předávání signálů generovaného z okolní stanice. Existuje několik druhů směrových antén, ve většině případů se však používají antény parabolické. Dalším typem, se kterým se můžeme setkat, jsou trychtýřové antény. Ty se však používají na méně exponovaných místech. Vlastnosti antény závisí jednak na průměru, ale také na použitém kmitočtu. Nejčastěji se používají antény průměru 0,35/0,65/0,90/1,20/1,80/2,40 m, ale ve zvláštních případech lze použít průměry větší. V systému radioreleové komunikace se vyžaduje, aby anténa měla co největší účinnost, vyzařovací diagram s malými postranními laloky, velké potlačení křížové polarizace, co nejširší pracovní kmitočtové pásmo a v něm nízký poměr stojatých vln. [1, 3]

1.1.2 Vnější jednotka(ODU)

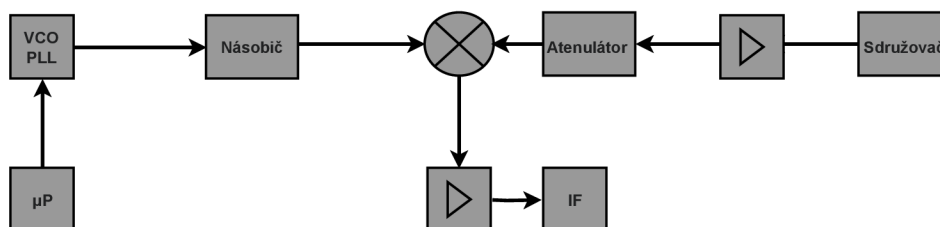
Vnější jednotka se stará o přeměnu mezifrekvenčního kmitočtu na kmitočet rádiový, úpravu rádiového kmitočtu a jeho následné zesílení. Parametry ODU vychází z použitého kmitočtu a přenosové kapacity. Jedna jednotka obvykle nedokáže pokrýt celé kmitočtové pásmo, rozděluje kmitočtové pásmo nejčastěji do podpásem A, B, ve speciálních případech na ještě vícero podpásem. Jednotka se skládá ze dvou hlavních částí – přijímače a vysílače. [1] Vysílač se rovněž skládá z několika funkcí, kterými musí signál projít, než jej bude možné přenést anténou k další stanici. V první řadě se mezifrekvenční kmitočet na vstupu zesílí na požadovanou signální úroveň. Následně je tento signál převeden do směšovače, kde je převeden na požadovaný signál v RF pásmu. Poté je signál ze směšovače přiveden do vhodného filtru, který zadržuje jedno postranní pásmo. Tyto dva kroky signál velmi oslabí, proto je nutné zesílit jeho úroveň v mikrovlnném zesilovači. Signál je převeden do antény a již nic nebrání úspěšnému šíření k další stanici. [1]

Jak už z názvu vypovídá, přijímač se stará o přijímání signálu zachyceného přijímací anténou. Na vstupu přijímače je pásmový filtr, který odstraní všechny zrcadlové kmitočty. Přijatý signál následně pomocí nízkosumových zesilovačů zesiluje signál, přemění jej na IF signál a poté předá demodulátoru. Lokální oscilátor přijímače generuje kmitočet, který s přijímaným kmitočtem RF vytvoří ve směšovači mezifrekvenční kmitočet. Poté signál putuje do mezifrekvenčních zesilovačů, jejichž

úkolem je zachování stálé hodnoty mezifrekvence v důsledku nutnosti přivádět do demodulátoru stálý signál. Pro nejvhodnější úroveň kvality IF signálu je mikrovlnná část přijímače zdvojnásobena a doplněna o výběrový blok, který vyselektuje ten nejvhodnější IF signál. Bloková schémata vysílače i přijímače jsou vyobrazena na obrázcích níže [1]:



Obr. 1.2: Blokové schéma vysílače. Převzato z [1]



Obr. 1.3: Blokové schéma přijímače. Převzato z [1]

1.1.3 Vnitřní jednotka (IDU)

IDU je hlavní část mikrovlnného přenosového systému. Vnitřní jednotka se stará o vnitřní procesy probíhající ve vnější jednotce. Zajišťuje procesy modulace, demodulace, multiplexování a servisní dohled. Dalo by se tedy říct, že se jedná o samotný „mozek“ celého systému. [1]

2 Faktory ovlivňující naměřenou hodnotu senzorů

Kvantifikování faktorů podílejících se na naměřené hodnotě teplot mikrovlnných senzorů je velmi složitou problematikou. To je zapříčiněno proměnlivostí mnoha faktorů, které působí na samotné senzory a mají vliv na výslednou teplotu. Obecně lze teplotu T , naměřenou mikrovlnnými senzory reprezentovat jako součet teplot vyprodukovaných několika faktory:

$$T = t_j + t_s + t_{vz} + t_r \quad (2.1)$$

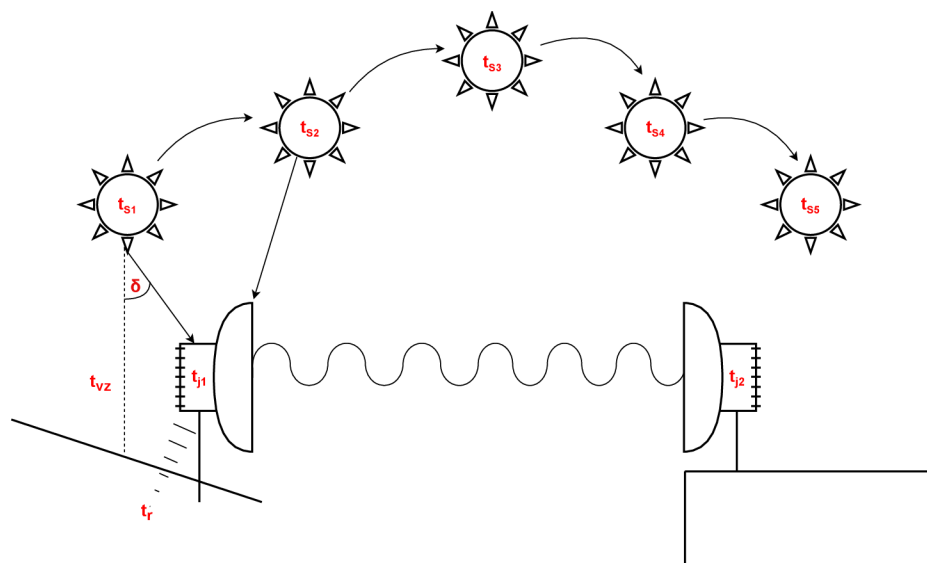
- t_j označuje samotnou teplotu jednotky
- t_s značí teplotu vyprodukovanou slunečním zářením
- t_{vz} reprezentuje teplotu produkovanou okolím
- t_r představuje teplotu reflektovanou okolním prostředím

Jedním z hlavních vlivů, které působí na samotné senzory teplot je vystavení senzorů na přímý sluneční svit. Tento faktor je označený v obecném vztahu jako t_s . Teplotní senzory se následně nahřejí a naměřená hodnota bude vyšší než reálná teplota okolí. Kvantifikování samotného vlivu slunečního záření je velmi složitou problematikou. Je nutné brát na vědomí samotné umístění senzoru - například to, jak velkému slunečnímu záření je senzor vystaven, (paprskům slunečního záření mohou překážet různé budovy, střechy, nebo také mraky, které brání slunečním paprskům proniknout přímo na senzor). Zároveň také musíme brát v potaz, že Slunce vychází a zapadá, mění tedy svoji polohu na obloze vůči Zemi.

Teplota reflektovaná okolním prostředím (označována jako t_r) se také významně podílí na výsledné teplotě. Jedná se o faktor, který souvisí se samotným umístěním mikrovlnných zařízení. Ty se ve většině případech umísťují na střechy budov, které se slunečním zářením mohou nahřát a následně reflektovat teplo přímo na jednotku.

Samotná teplota jednotky t_j hraje rovněž velkou roli ve výsledné teplotě. Jedná se o elektrické zařízení, které se díky svému provozu zahřívá (Jouleovo teplo) a ovlivňuje naměřenou hodnotu jednotky. Teplota okolí (značená t_{vz}) ovlivňuje naměřenou hodnotu senzorů spíše opačným vlivem, jelikož teplota vzduchu dosahuje razantně menších hodnot, než právě teplota mikrovlnné jednotky. Můžeme tedy říct, že slouží spíše jako chlazení.

Situaci, která zobrazuje všechny okolní vlivy působící na mikrovlnnou jednotku, můžeme vidět na obrázku níže:



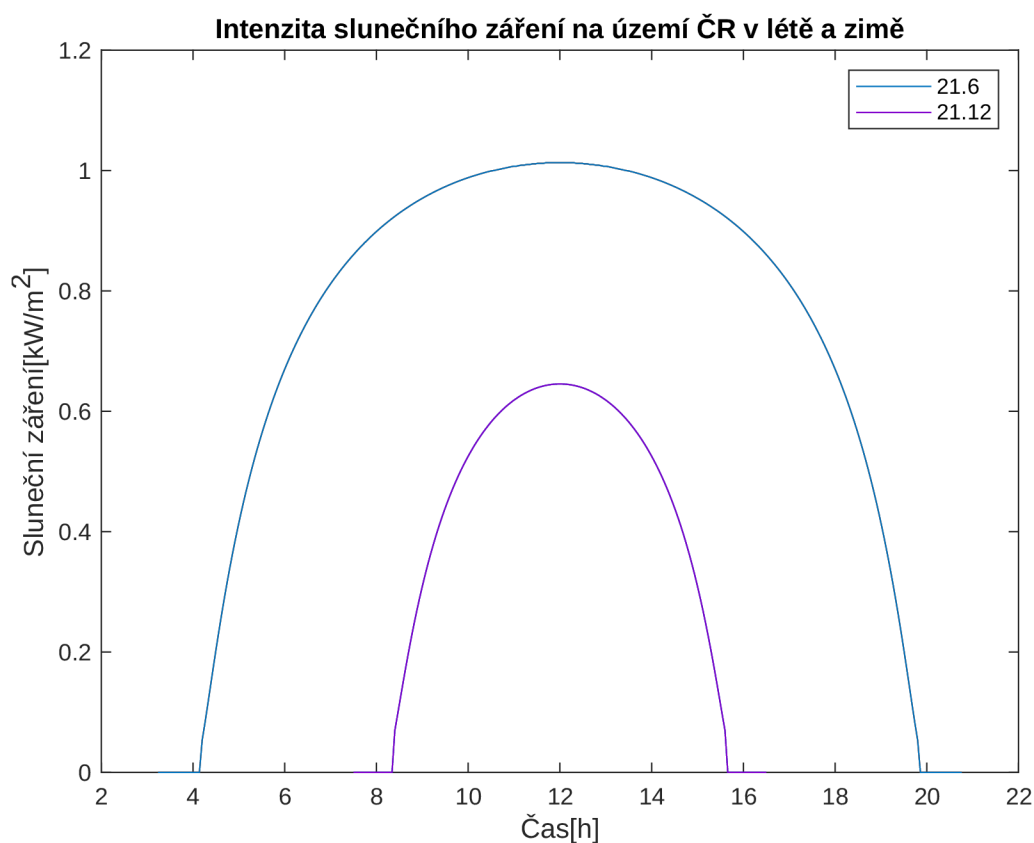
Obr. 2.1: Faktory působící na senzor měření teplot mikrovlnné jednotky

2.1 Sluneční záření

Určování intenzity slunečního záření je velmi komplexní problematikou - je nutné brát v úvahu několik faktorů, které se na následné intenzitě podílí. Jedná se o proměnnou vzdálenost mezi Zemí a Sluncem, znečištění atmosféry, zeměpisná poloha (zejména šířka) daného místa [4, 5]. Slunce vyzařuje energii o výkonu $3,89 \cdot 10^{26}$ W, avšak z tohoto množství dopadá na okraj stratosféry průměrných 1350 W/m^2 . Skutečný energetický tok dopadající na povrch Země je následně ještě menší než uvedených 1350 W/m^2 . Konkrétně na území ČR dopadá sluneční záření o intenzitě, která v celoročním průměru dosahuje hodnot 620 W/m^2 . Toto množství však není během roku rovnoměrně rozděleno. Doba slunečního svitu na území ČR se v průměrném roce pohybuje mezi 1750 až 1900 hodinami za rok a na 1 m^2 vodorovné plochy dopadá za rok celkové množství energie kolem $1200 \text{ kW} \cdot \text{h}$. Takové množství však není během roku rozděleno rovnoměrně - v chladnější polovině roku, a to tedy od začátku října až do konce března, dopadá jen 25% z tohoto množství, tedy přibližně $300 \text{ kW} \cdot \text{h/m}^2$. V teplejší polovině roku - od dubna do konce září dopadá zbylých 75%, což je přibližně $900 \text{ kW} \cdot \text{h/m}^2$. [4]

Průběh intenzity slunečního záření se v průběhu roku velmi liší na základě střídání ročních období. Největší intenzity slunečního záření dosahuje v létě, a to konkrétně 21. června, tedy v den jarní rovnodennosti. Maximální sluneční intenzita v tento den dosahuje hodnot kolem 1 kW/m^2 [6], jedná se tak o nejdelší den v roce. Intenzita slunečního záření v ČR pro tento den začíná působit krátce po 4. hodině ranní a přestává kolem 20. hodiny.

Pro zobrazení opačného extrému, a to tedy, kdy intenzita slunečního záření dosahuje nejmenších hodnot, byl pro změnu vybrán den zimního slunovratu - 21. prosince. Intenzita slunečního záření v tento den dosahuje maximální hodnoty $0,7 \text{ kW/m}^2$ [6]. Jedná se o nejkratší den v roce a intenzita slunečního záření začíná narůstat krátce po 8. hodině a klesne na 0 krátce před 16. hodinou. Průběh intenzity slunečního záření obou případů je zobrazen na grafu níže:



Obr. 2.2: Graf průběhu intenzity slunečního záření. Převzato z [6]

V obou případech se jedná o ideální stav, kdy nepůsobí žádné okolní faktory (například oblačnost), které by slunečnímu záření bránily v průniku na povrch Země. [6] Průběh a velikost intenzity záření se v průběhu roku mění každým dnem, kdy se sluneční svit se prodlužuje, nebo zkracuje. Země při oběhu kolem Slunce mění svoji vzdálenost vůči Slunci a tím pádem se mění také intenzita slunečního záření.

2.2 Jouleovo teplo

Dalším vlivem, který dokáže ovlivnit naměřené hodnoty tepla je ohřev samotného elektrického zařízení. Tento jev se označuje termínem Jouleovo teplo. Je to proces, při kterém se průchodem elektrického proudu vodičem produkuje teplo. Hodnota teploty vyprodukovaná tímto jevem se označuje písmenem Q a lze ji poměrně jednoduše určit na základě následujících vztahů [7]:

$$Q = U \cdot I \cdot t \quad (2.2)$$

Kde:

- Q značí velikost Jouleova tepla
- U je napětí
- I je elektrický proud
- t je čas

V případě, kdy není známo přesné napětí vodiče, ale je dán jeho odpor, lze ze vzorce pro výpočet napětí odvodit tento vzorec na níže uvedený vztah:

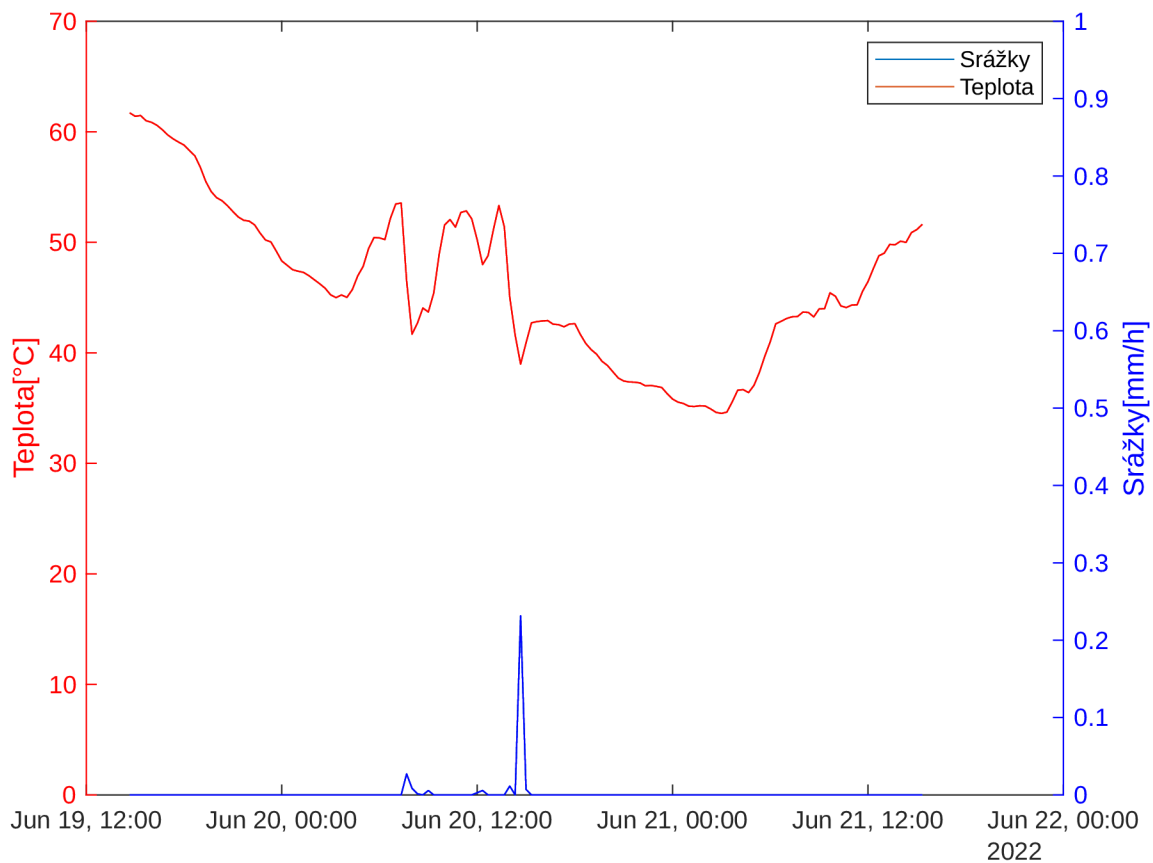
$$Q = R \cdot I^2 \cdot t \quad (2.3)$$

Kde:

- Q značí velikost Jouleova tepla
- R je odpor
- I je elektrický proud
- t je čas

2.3 Výskyt srážek

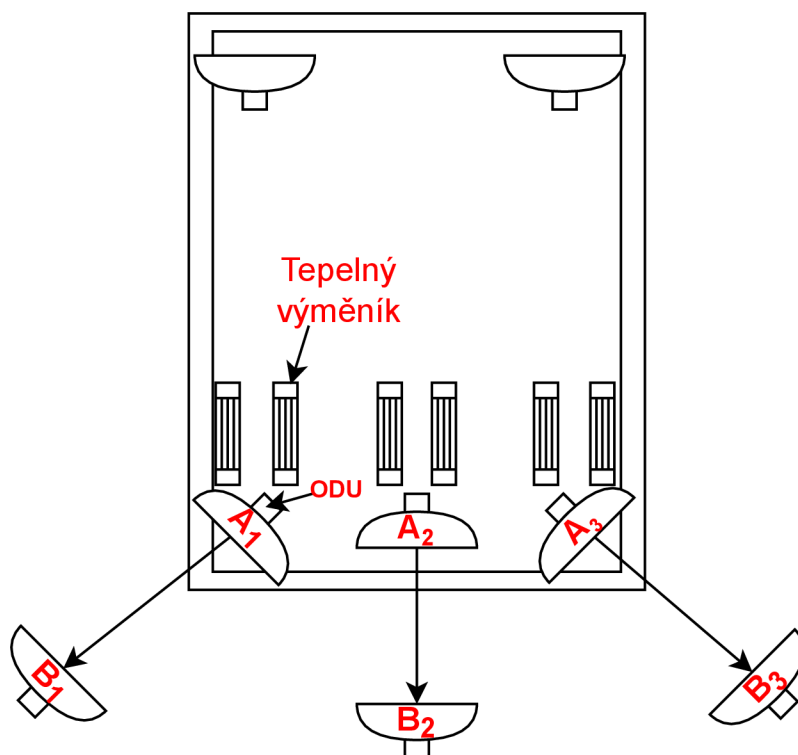
Dalším případem, kdy by mohlo dojít ke zkreslení teplot, je výskyt deště v měřené oblasti. Jednak dokáže ovlivnit samotnou úroveň přijímaného signálu, ale také velmi dobře působí jako chladící kapalina a pro tento stav je tedy charakteristický prudký pokles naměřené teploty. Tuto situaci můžeme vidět na následujícím grafu, kdy dochází k okamžitému poklesu naměřené hodnoty, který je zapříčiněn výskytem srážek v okolí mikrovlnné jednotky. Graf je zobrazený níže:



Obr. 2.3: Vliv výskytu srážek na změnu naměřené teploty

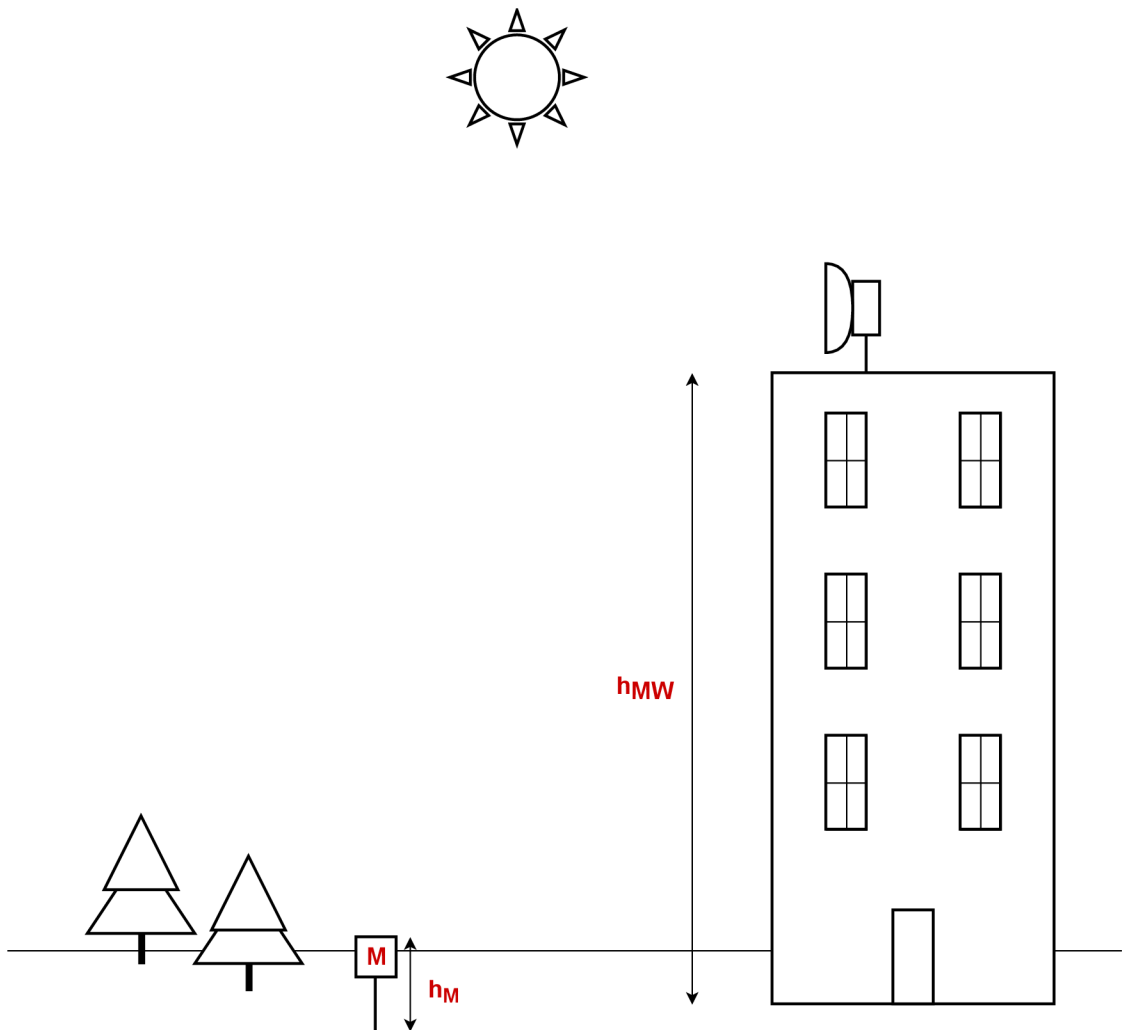
2.4 Umístění mikrovlnné jednotky

Mikrovlnné jednotky se nejčastěji instalují do takové výšky, aby na sebe jednotky měly přímý dohled a nic vysílaným paprskům nebránilo v cestě. Z tohoto důvodu se mikrovlnné jednotky umísťují nejčastěji na střechy panelových domů, kde jsou jednotky vystaveny extrémní intenzitě slunečního záření prakticky po celý den, jelikož na střeše neexistují překážky, které by s pohybem slunce tvořily na jednotky stín. Dalším faktorem, který by se mohl podepsat na výsledné teplotě naměřené senzorem mikrovlnné jednotky je, že právě na střechy panelových domů jsou také v některých případech vyvedeny například tepelné výměníky, který se na budovu umísťují tehdy, kdy je nutné z místnosti odvádět teplo. Nejčastější případy jsou například serverovny (nebo místnosti, odkud je nutné odvádět teplo). Pokud je tedy tepelný výměník nainstalován na stejné střeše právě jako mikrovlnná jednotka, odváděné teplo působí na vnější jednotku (ODU), ve které se nachází senzor na měření teploty a následně se vyprodukované teplo tepelným výměníkem reflektuje na naměřené teplotě. Situace rozložení mikrovlnných jednotek můžeme vidět na následujícím obrázku, kde je zobrazen jednak tepelný výměník, ale také rozvržení a orientace mikrovlnných jednotek s jejich protistranami:



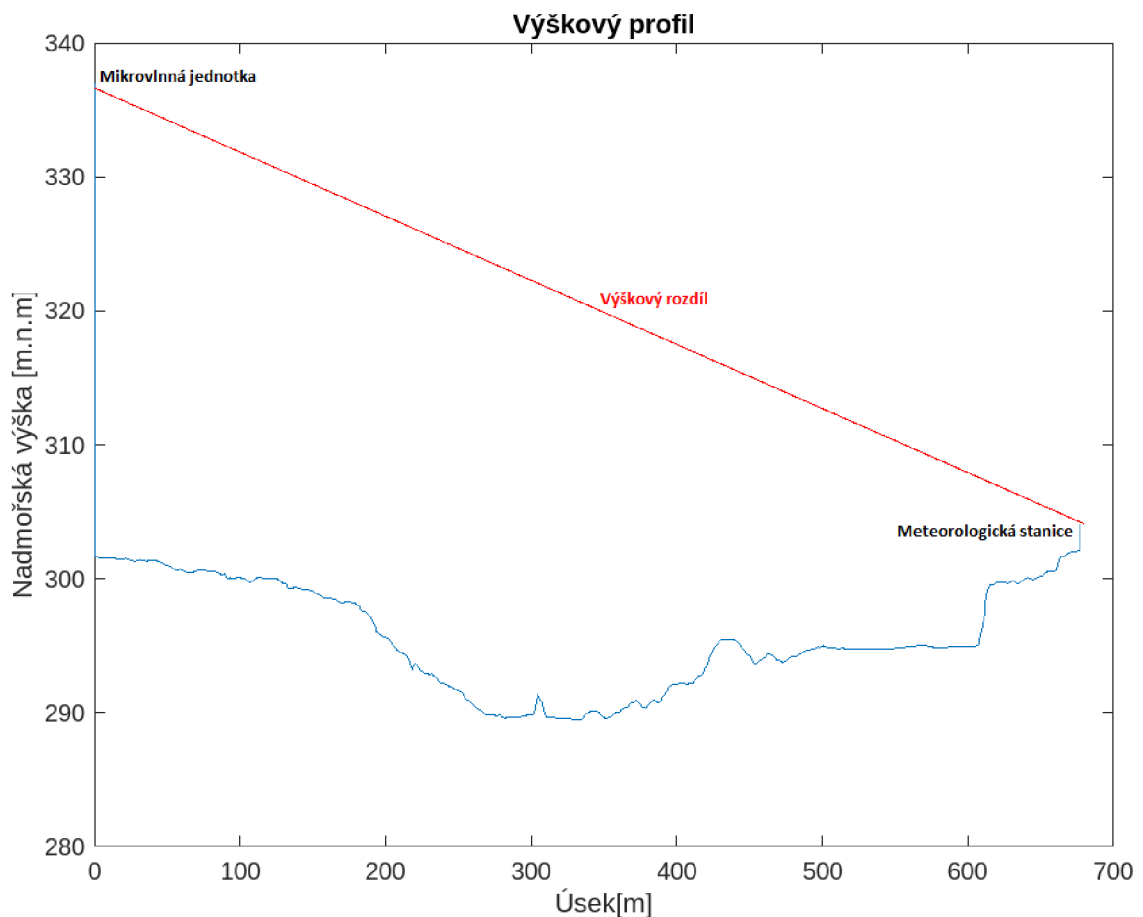
Obr. 2.4: Rozložení mikrovlnných jednotek na střeše panelové budovy

Dalším možným faktorem, který v rozdílech naměřených hodnot hraje roli je výška umístění senzoru měření teplot. ČHMÚ měří hodnoty teplot zpravidla ve výšce dvou metrů, zatímco mikrovlnné jednotky, které jsou umístěny na samém vrcholu budov dosahují výšky klidně i 12 metrů a výše, což zapříčiní působení větší intenzity slunečního záření na mikrovlnnou jednotku. [8] Situaci můžeme vidět na obrázku níže, kde h_M je výška jednotky ČHMÚ a h_{MW} je výška mikrovlnného senzoru:



Obr. 2.5: Výškový rozdíl senzorů měření teplot

Celkový výškový rozdíl může být ještě větší v důsledku umístění jednotek v jiné nadmořské výšce. Situaci si můžeme zobrazit na průběhu grafu výškového profilu, který zobrazuje průběh rozdílu nadmořské výšky mezi dvěma body na mapě. Pro referenční příklad byl vybrán panelový dům, který na své střeše nese řadu mikrovlnných jednotek, jedná se o panelový dům v Praze na ulici Pšenčíkova. V tomto panelovém domě se nachází celkem 13 pater, z čehož můžeme jednoduše vypočítat téměř přesnou výšku mikrovlnného zařízení, jelikož výška patra v panelovém domě činí dle norem 2800 mm. Pro následné porovnání bylo nutné vybrat stanici ČHMÚ, která bude v těsné blízkosti právě mikrovlnné jednotky, jedná se o jednotku, která se nachází v Praha-Libuš. Mikrovlnná jednotka se nachází v celkové nadmořské výšce zhruba 337 metrů, zatímco stanice ČHMÚ snímá hodnoty teplot v nadmořské výšce 304 metrů. Celkový výškový rozdíl tedy činí 33 metrů. Na následujícím grafu můžeme vidět průběh grafu výškového profilu:



Obr. 2.6: Výškový rozdíl senzorů měření teplot

3 Statistické metody analýzy korelace a pravděpodobnostního rozdělení

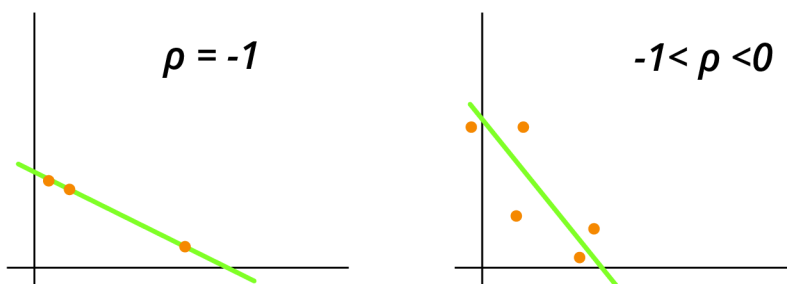
V následující kapitole si popíšeme jednotlivé statistické metody, které byly použity pro analýzu zkoumaných dat a porozumění vzájemných vztahů mezi hodnotami. Konkrétně se zaměříme na korelační analýzu a teorii pravděpodobnostního rozdělení.

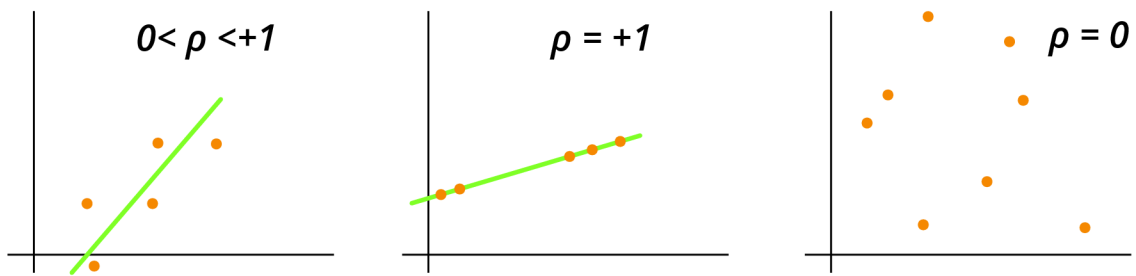
3.1 Korelační analýza

Důležitou problematikou určení faktorů, které se podílejí na ovlivnění výsledné teploty, je analýza vztahů mezi jednotlivými hodnotami a následné určení jejich konkrétních vlivů na výsledné hodnoty. Ve statistice se k určení tohoto jevu využívá několik korelačních analýz, jako je například Pearsonův korelační koeficient nebo Spearmanův koeficient pořadové korelace.

3.1.1 Pearsonův korelační koeficient

Jedná se o druh lineární korelace mezi dvěma soubory dat. Koeficient představuje poměr mezi kovariancí dvou proměnných a součinem jejich směrodatné odchylky. Výsledek se označuje hodnotou ρ a pohybuje se v rozsahu od -1 do 1. V ideálním případě, kdy se koeficient rovná 1, se dvě veličiny ovlivňují v přímé úměrnosti. V opačném případě, tedy když je koeficient roven -1, se veličiny ovlivňují protichůdně. V situaci, kdy ρ je rovno nule, spolu veličiny vůbec nekorelují, neexistuje mezi nimi závislost. Všechny možné situace, které mohou nastat, můžeme vidět na vyobrazených grafech níže [9]:





Obr. 3.1: Ukázkové grafy Pearsonové korelace. Převzaty z [9]

Pro přesné určení Pearsonova korelačního koeficientu se používá následující vztah:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (3.1)$$

Kde:

- cov značí kovarianci mezi dvěma hodnotami
- σ označuje směrodatnou odchylku jednotlivých hodnot [9]

3.1.2 Spearmanův koeficient pořadové korelace

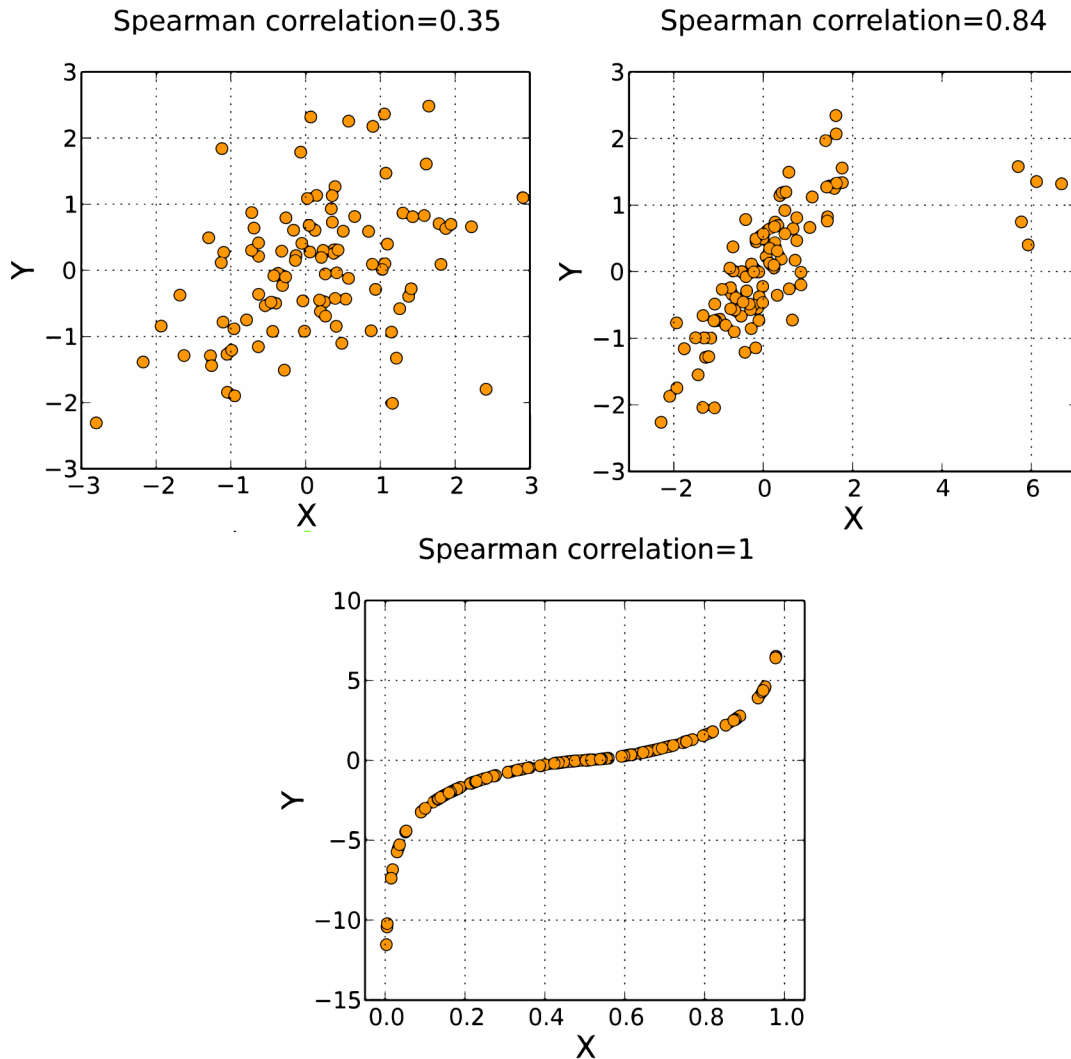
Další druh korelační analýzy se nazývá Spearmanův koeficient pořadové korelace. Svůj název dostal po Charlesi Spearmanovi, psychologovi, který se proslavil svojí prací ve statistice. Koeficient tohoto vztahu také dosahuje stejného rozsahu jako u Pearsonova koeficientu, a to tedy od -1 do 1. Na rozdíl od Pearsonova koeficientu korelace, který popisuje lineární vztah veličin X a Y , Spearmanův koeficient korelace popisuje, jak dobře vztah veličin X a Y odpovídá monotónní funkci, která může být nelineární. Dalším rozdílem mezi těmito dvěma druhy korelací je, že Spearmanova analýza nepracuje s hodnotami, nýbrž s jejichž pořadím [10]. Pro výpočet koeficientu Spearmanové pořadové korelace se využívá následujících vztahů [11, 10]:

$$r_s = \rho_{R(X), R(Y)} = \frac{\text{cov}(R(X), R(Y))}{\sigma_{R(X)} \sigma_{R(Y)}} \quad (3.2)$$

Kde:

- ρ označuje obvyklý Pearsonův korelační koeficient aplikovaný na proměnné pořadí
- cov značí kovarianci proměnných pořadí
- σ označuje směrodatnou odchylku jednotlivých hodnot [11]

Kladná hodnota koeficientu označuje situaci, kdy hodnota na ose Y má tendenci se zvyšovat v případě, že dojde ke zvýšení hodnoty na ose X. Může také nastat situace, kdy je koeficient roven nule. Tato hodnota značí, že obě hodnoty se navzájem neovlivňují. V případě záporného koeficientu na sebe hodnoty působí v nepřímé úměrnosti, tedy tak, že pokud klesne hodnota Y, tak hodnota X se zvyšuje. Situace, které mohou potencionálně nastat můžeme vidět na grafech zobrazených níže:



Obr. 3.2: Ukázkové grafy Spearmanové pořadové korelace. Převzaty z [11]

3.1.3 Kendallův koeficient pořadové korelace

Kendallův koeficient pořadové korelace je další metodou korelační analýzy. Obdobně jako u předchozího Spearmanového pořadového koeficientu tato metoda zkoumá monotonní vztahy obou hodnot a tím tedy měří pravděpodobnost, že se dvě proměnné budou pohybovat stejným směrem, ale ne nutně konstantní rychlostí. Stejně jako u předchozích dvou metod korelačních analýz koeficient této korelace nabývá hodnot v rozsahu od -1 do 1.[12] Tato metoda korelace pracuje se souhlasnými a nesouhlasnými páry hodnot. Souhlasnými páry jsou myšleny takové páry, u kterých se zvyšuje nebo klesá hodnota u obou proměnných, zatímco u nesouhlasných párů se jedna hodnota zvyšuje a druhá klesá, nebo naopak. Korelační koeficient Kendallové korelace se značí řeckým písmenem τ . Pro výpočet tohoto koeficientu je využito následujícího vztahu[13] :

$$\tau = \frac{(\text{Počet souhlasných párů}) - (\text{Počet nesouhlasných párů})}{(\text{Celkový počet párů})} \quad (3.3)$$

Kendall korelace je stejně jako předchozí Spearmanové korelace neparametrickým statistickým testem, což znamená, že není nutné aby obě dvě sady hodnot spadaly do zvonovité křivky distribuce dat.[12] Hodnoty koeficientů pro Kendallovu korelaci obvykle vychází jako menší čísla oproti Spearmanové korelaci, jelikož nejsou citlivé na chyby. Hodnoty korelačního koeficientu jsou přesnější při menších velikostech datových vzorků. [14]

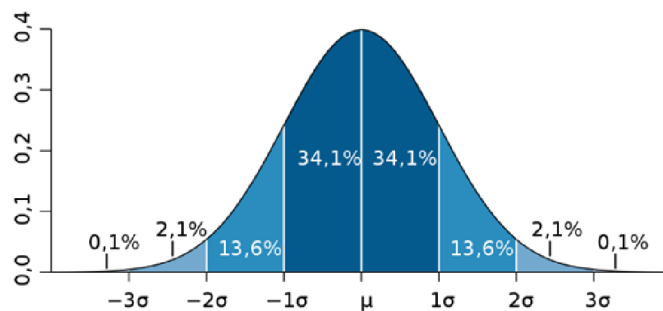
3.2 Pravděpodobnostní rozdělení souborů dat

Teorie pravděpodobnostního rozdělení je zapotřebí tehdy, pokud chceme navrhovat pravděpodobnostní modely vztahu proměnných a zdůvodnit procedury pro statistické usuzování v korelační a regresivní analýze[15]. Jednoduše tedy můžeme říct, že z důvodu zkoumání korelací naměřených hodnot je nutné popsat pravděpodobnostní rozdělení naměřených proměnných a následně díky tomu určit, zdali je pro tento případ vhodnější využití parametrických (Pearsonův korelační koeficient), nebo neparametrických testů (Spearmanův a Kendallův pořadový korelační koeficient). V našem konkrétním případě bylo potřeba zjistit, zdali jsou hodnoty teplot distribuovány v Gaussovo normálním rozdělení.[14]

3.3 Gaussovo normální rozdělení

Normální rozdělení (Gaussovo rozdělení) patří mezi nejdůležitější rozdělení pravděpodobnosti spojité náhodné veličiny. Většina jevů, vyskytujících se v přírodě nebo

společnosti lze modelovat právě Gaussovým rozdělením[16]. Nejčastější výsledky se pohybují kolem průměru, výsledky vzdálenější od průměru se objevují řidčeji a přitom se přibližně stejně často objevují hodnoty nadprůměrné a podprůměrné[17]. Matematická definice normálního rozdělení je založena na normální hustotě pravděpodobnosti a má několik vlastností, jako jsou střední hodnota, rozptyl, šikmost a špičatost[18]. Grafickým vyjádřením Gaussova normálního rozdělení je zvonovitá křivka, symetrická kolem střední hodnoty μ („parametr polohy“ – udává polohu křivky na ose x). Šířku křivky v tzv. inflexním bodě (bod obratu křivky) udává směrodatná odchylka σ („parametr rozptýlení“)[19]. Grafické vyjádření Gaussova normálního rozdělení pro náhodnou veličinu X můžeme vidět na následujícím obrázku:



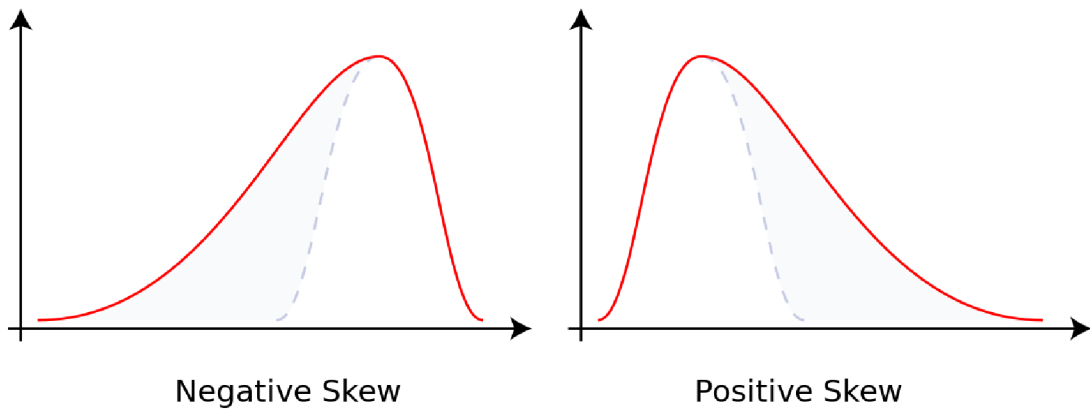
Obr. 3.3: Grafické vyjádření Gaussova normálního rozdělení pro náhodnou veličinu X . Převzato z [17]

3.3.1 Šikmost a špičatost

Oba tyto statistické pojmy slouží k vystižení dalších charakteristických vlastností rozložení hodnot souboru. Šikmost zjišťuje, zda jsou hodnoty rozloženy okolo průměru symetricky, špičatost pak porovnává koncentraci hodnot blízko průměru a dále od něho, špičatost tedy vyjadřuje, jak často se vyskytují odlehlé hodnoty.[17]

Šikmost(Skewness)

Na dalším obrázku je vyobrazeno několik křivek každá s různou šikmostí, kde můžeme vidět přerušovanou čáru, která značí, jak se díky šikmosti křivka liší od symetrické křivky.

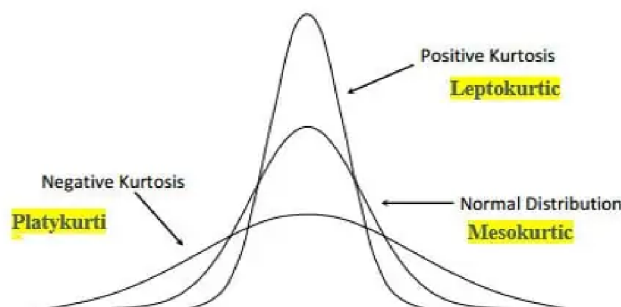


Obr. 3.4: Ilustrační obrázek pozitivní a negativní šikmosti. Převzato z [17]

Pokud je tedy distribuce dat pozitivně zkreslená, tedy data mají pozitivní šikmost, znamená to, že mají vyšší počet datových bodů s nízkými hodnotami. Jednoduše řečeno, menší hodnoty se v datech vyskytují s vyšší frekvencí než ty vyšší. V opačném případě, tedy datové soubory, které mají negativní šikmost rozložení dat, se vyskytují datové body s většími hodnotami s větší frekvencí, než ty nižší. Šikmost nám také říká o směru odlehlých hodnot, to znamená, že datové soubory s pozitivní šikmostí mají většinu odlehlých hodnot na pravé straně distribuce[20].

Špičatost(Kurtosis)

Pomocí špičatosti můžeme zjistit, zda a do jaké míry jsou hodnoty koncentrovány poblíž průměru[17]. Normální distribuce má špičatost rovnu 3, tudíž pokud chceme spočítat nadměrnou špičatost musíme tuto hodnotu odečíst od celkové špičatosti[21]. Na následujícím obrázku můžeme vidět několik křivek vyobrazující rozdíl mezi špičatostí daných křivek:



Obr. 3.5: Ilustrační obrázek různých typů špičatosti křivek. Převzato z [22]

Z obrázku je zřetelné, že pozitivní hodnoty špičatosti, jejichž tvar křivky nese název Leptokurtic, má tendenci mít výrazný vrchol v blízkosti průměru. Opačný extrém, tedy datové sady, které mají data distribuována s nízkou špičatostí, jejichž tvar křivky je označován termínem Platykurtic. Distribuce dat s nízkou špičatostí je charakteristická menším počtem extrémních odlehlých kladných či záporných hodnot. Mesokurtic je tvar křivky, který nesou data distribuovaná normálním rozdělením.[23, 21]

4 Nástroje pro analýzu a zkoumání vlivů

Následující kapitola se věnuje tvorbě nástrojů, které slouží k analýze a kvantifikaci vlivů, které působí na mikrovlnné zařízení a ovlivňují tak jejich naměřenou teplotu. Nástroje byly vytvořeny v databázi InfluxDB, která v časových řadách ukládá parametry z jednotlivých mikrovlnných zařízení (Teplota, Přijímaná úroveň signálu, atd.) umístěných po celé ČR, a také dostupné data z meteorologických stanic v okolí Prahy.

4.1 Nástroj pro výpočet Offsetu

Prvním krokem evaluace vlivu slunečního záření bylo vynesení nástroje, který bude zobrazovat teplotní rozdíl mezi jednotkou ČHMÚ a mikrovlnným senzorem v časové řadě. Teplotní rozdíl mezi dvěma jednotkami je označován termínem offset. Díky samotnému offsetu můžeme pozorovat právě takové vlivy, které se podepíší na výsledné naměřené hodnotě.

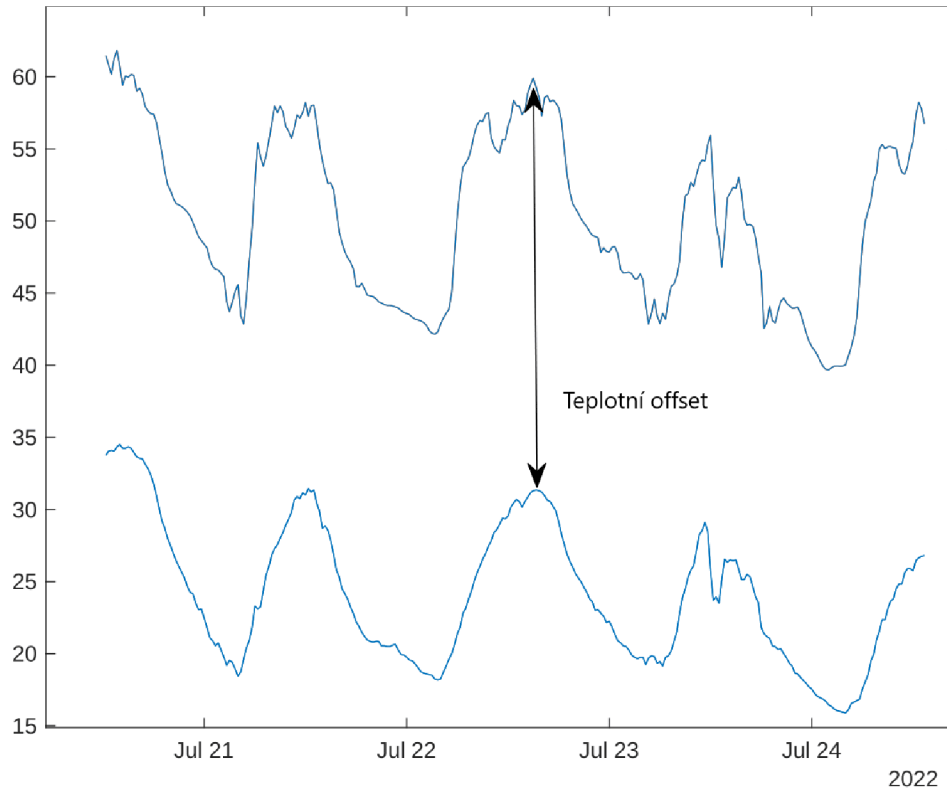
Pro vynesení samotného rozdílu obou hodnot je nutné prvně do nástroje vytáhnout z databáze hodnoty naměřené oběma jednotkami. Ukázka zdrojového kódu pro vynesení obou hodnot můžeme vidět na obrázku níže:

Výpis 4.1: Ukázka vynesení hodnot z jednotek

```
temp_gage = from(bucket: "gauges")
  |> range(start: v.timeRangeStart, stop: v.timeRangeStop)
  |> filter(fn: (r) => r["_field"] == "temperature")
  |> filter(fn: (r) => r["id"] == v.Gauge_ID)
  |> aggregateWindow(every: duration(v: v.Sample_Interval),
    fn: mean, createEmpty: false)

temp_cml = from(bucket: "mws")
  |> range(start: v.timeRangeStart, stop: v.timeRangeStop)
  |> filter(fn: (r) => r["_field"] == "temperature")
  |> filter(fn: (r) => r["ip"] == v.IP_A)
  |> aggregateWindow(every: duration(v: v.Sample_Interval),
    fn: mean, createEmpty: false)
```

Po vynesení těchto hodnot si můžeme zobrazit graf časové řady, která zobrazuje teploty postupem časového intervalu. Na grafu můžeme vidět teplotní offset, který nám zobrazuje právě okolní vlivy, které působí na mikrovlnnou jednotku. Grafy obou hodnot vynesené do jedné časové řady můžeme vidět na obrázku níže:



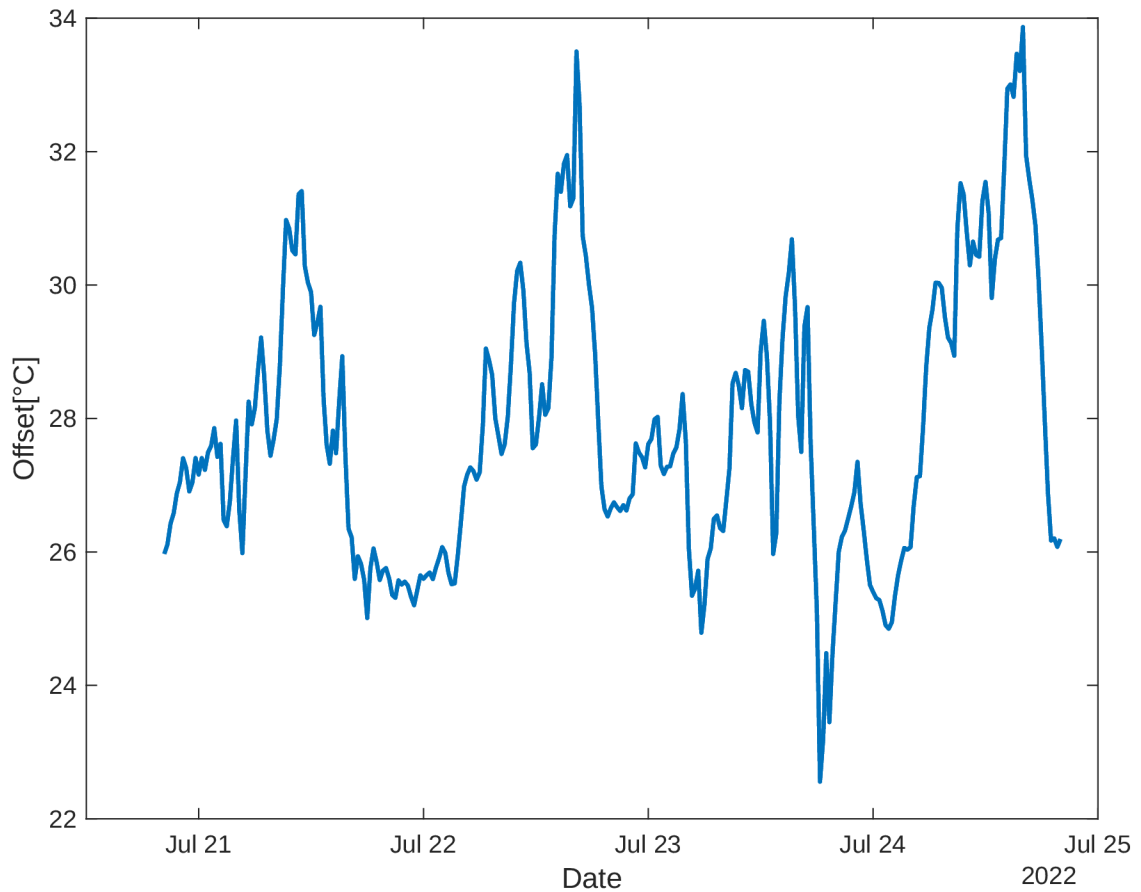
Obr. 4.1: Zobrazení offsetu mezi mikrovlnnou jednotkou a ČHMÚ jednotkou

Pro přímé vynesení offsetu je nutné tyto dvě hodnoty od sebe odečíst. Problémem však je, že se jedná o dva druhy datových sad, proto je nutné tyto dvě sady spojit do jedné, aby bylo možné je od sebe odečíst. Pro takové případy je v databázi funkce *join*, která obě datové sady spojí. Následně nám nic nebrání v odečtení hodnot. Pro odečtení byla použita následující funkce:

Výpis 4.2: Odečtení teplotních hodnot

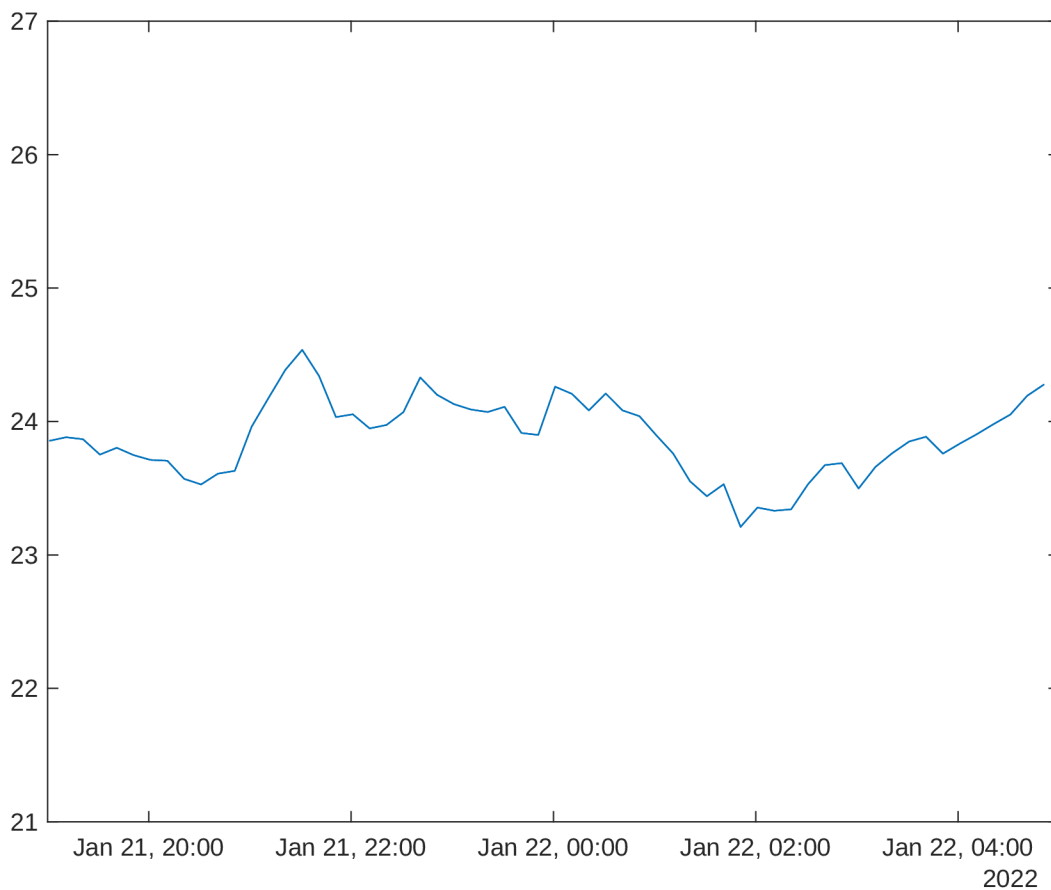
```
|> map(fn: (r) => ({r with _value:
(r["_value_temp_cm1"] - r["_value_temp_gage"]}))
```


Poté si v časové řadě vyneseme průběh teplotního offsetu, na kterém si můžeme všimnout velkého nárůstu, ke kterému dochází v situaci, kdy vychází slunce. Z tohoto faktu můžeme určit, že vliv slunečního záření na mikrovlnnou jednotku má velký dopad na výslednou teplotu jednotky. Graf průběhu teplotního offsetu pro nejjasnější dny v roce 2022 je zobrazen níže:



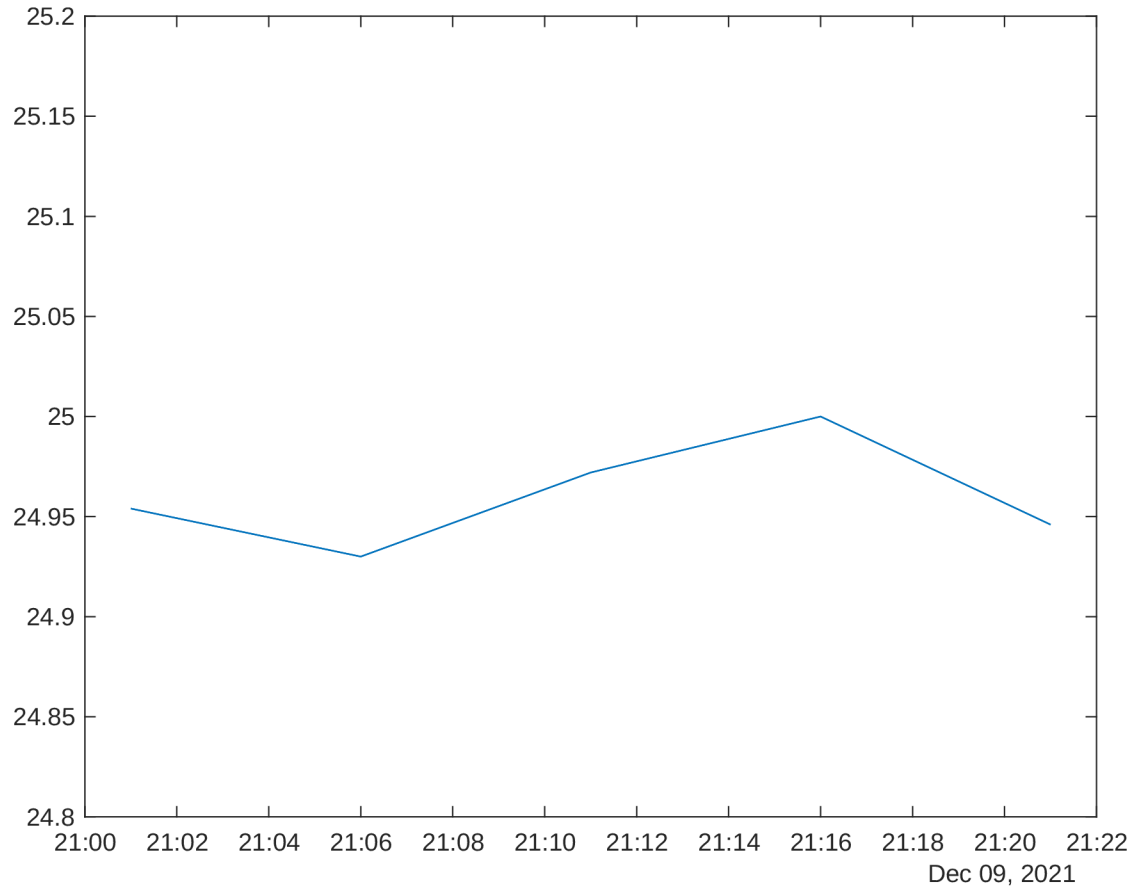
Obr. 4.2: Průběh teplotního offsetu

Z nástroje pro vynesení teplotního offsetu také můžeme kvantifikovat i jiné vlivy, které se podílejí na celkové teplotě jednotky a to tím způsobem, že si jako časový interval zvolíme dobu, kdy slunce nesvítí. V tomto případě dosahuje offset viditelně menších hodnot a průběh grafu je o dost stabilnější, než je tomu tak v případě nástupu slunečního záření. Tato situace je zobrazena na grafu zobrazeném níže:



Obr. 4.3: Průběh teplotního offsetu v noci

V situaci, kdy vybereme takové období, kdy se teplota okolí v noci pohybuje okolo 0°C , můžeme kvantifikovat vliv tepla vyprodukovaného samotnou jednotkou, a to konkrétně Jouleovým teplem, na základě odstranění ostatních faktorů které nemají na mikrovlnnou jednotku v noci vliv. Graf průběhu teplotního offsetu, kdy se hodnota okolí pohybuje okolo 0°C je zobrazen níže:

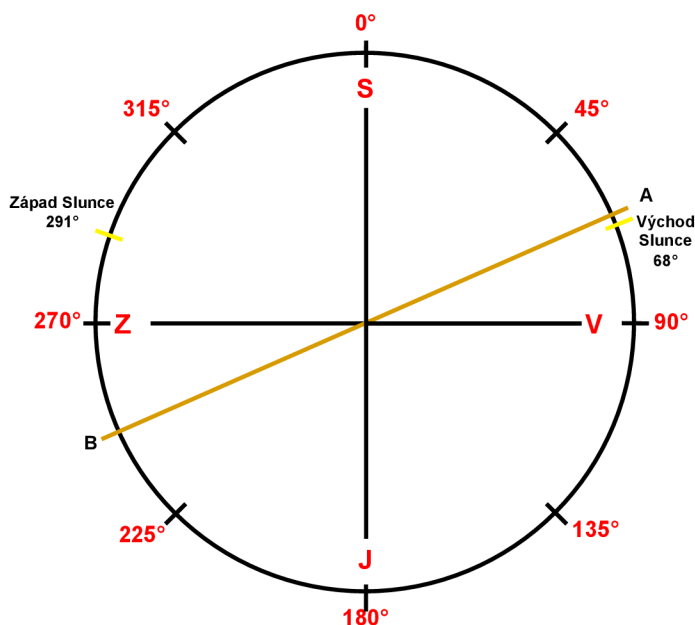


Obr. 4.4: Teplotní offset vyprodukovaný Jouleovým teplem

Z grafu můžeme vyčíst, že teplota vyprodukovaná Jouleovým teplem se pohybuje okolo 25°C . Tato hodnota se však může lišit od jednotlivého zařízení, jelikož vztah, ze kterého je Jouleovo teplo odvozeno, obsahuje veličiny, které se mohou lehce lišit na jednotlivých zařízeních.

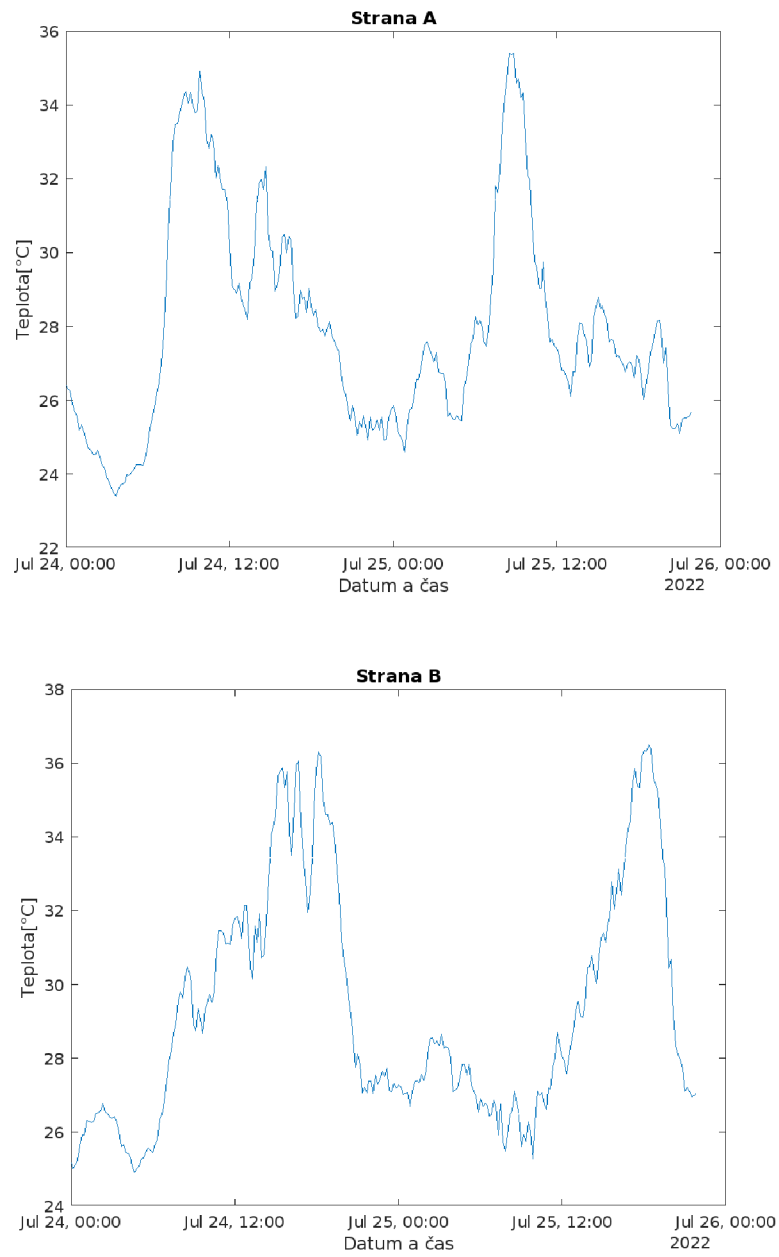
4.2 Kvantifikace tepelného přírůstku vyprodukovaný slunečním zářením

Pro kvantifikování tepelného nárůstku způsobeném slunečním zářením je nutné nastavit nástroje v InfluxDB tak, aby mikrovlnné jednotky a jednotky ČHMÚ byly geografickou polohou co nejbližší, jelikož teplota a různé meteorologické podmínky mohou být velmi proměnlivé pro každé vzdálenější místo. Pro ukázkový příklad byl vybrán mikrovlnný spoj jehož strana A se nachází v Praze na ulici Pšenčíkova a její protistrana v obci Kosoř. Následně bylo nutné k mikrovlnnému spoji vybrat geograficky nejbližší meteorologickou stanici, jedná se o meteorologickou stanici Praha-Libuš. Dalším faktorem, pro co nejpřesnější určení tepelného přírůstku slunečním zářením je vybrat období, kdy není taková oblačnost, jelikož oblačnost odráží sluneční záření. To poté nemusí na jednotku působit skoro vůbec. Doba, kdy začíná sluneční záření působit na jednotku, je velmi proměnlivá na základě orientace dané jednotky. Pro zobrazení tepelného přírůstku byl vybrán časový interval ve dne, kdy se vyskytuje minimální oblačnost - 24-25/7/2022. Mikrovlnný spoj, na kterém bude zobrazen průběh teplot se skládá ze strany A, která je orientovaná na západojihozápad, což odpovídá azimutu $247,5^\circ$ a strany B, která je orientovaná na východoseverovýchod, kde se azimut rovná $67,5^\circ$ [24]. Celou situaci si můžeme detailně ukázat na následující kružnici, která zobrazuje jednak přesnou orientaci daného mikrovlnného spoje, ale také přesnou deklinaci východu a západu Slunce pro daný den. Kružnice je vyobrazena na následujícím obrázku:



Obr. 4.5: Kružnice popisující situaci

Průběhy teplotních offsetů obou stran jsou zobrazeny níže:



Obr. 4.6: Průběhy hodnot teplotního offsetu mikrovlnného spoje na obou stranách spoje

Z grafů můžeme vyčíst, že sluneční záření působící na mikrovlnné jednotky má na obou stranách nástup v jiný čas, který závisí na orientaci dané jednotky, dokazuje to právě periodičnost průběhů grafů pro oba dva dny, kdy největších hodnot teplotního offsetu dosahují obě strany téměř ve stejný čas. Jednotka na straně A dosahuje nejvyšších hodnot vyprodukovaných sluncem zhruba o půl 9 ráno a její protistrana kolem 18 hodiny. Z toho se dá usoudit, že největší tepelný přírůstek slunce vyprodukuje na jednotku, kdy jeho paprsky dopadají na zadní stranu jednotky. Je také nutné od výše naměřených hodnot odečíst teplotu vyprodukovanou teplotou okolí a Jouleovým teplem, která se pro stranu A v tento den pohybuje okolo $24,5^{\circ}\text{C}$. Na protistraně se tato hodnota pohybuje průměrně kolem 27°C . Pokud tedy tyto hodnoty odečteme od naměřených hodnot teplotního offsetu, zjistíme tak, že sluneční záření na straně A vyprodukovalo maximální tepelný přírůstek $10,4^{\circ}\text{C}$. Na straně B po odečtu tato hodnota vychází $10,8^{\circ}\text{C}$.

5 Statistická analýza

Tato kapitola se zabývá podrobným prozkoumáním teplotních dat získaných z různých technologií mikrovlnných zařízení, ale také teplot naměřených meteorologickými stanicemi. Konkrétně bude ve statistické analýze provedeno testování normální distribuce naměřených teplot, pomocí široké škály vybraných metod testování, které byly zvoleny s ohledem na maximalizaci přesnosti výsledků testování. Výsledkem testování bude sada tabulek zobrazující p hodnoty pro jednotlivé měsíce. Ačkoliv bylo provedeno mnoho testů normality, do této kapitoly byla vložena pouze část z nich. Hlavním důvodem, proč není nutné prezentovat všechny testovací tabulky, je skutečnost, že byla vypořazována přibližně stejná míra potencionálního normálního rozložení napříč různými technologiemi. Výsledky následně poskytnou podklad pro rozhodnutí, zda pro korelační analýzu bude vhodnější využití parametrických, nebo neparametrických metod.

Po dokončení testování normality přichází na řadu korelační analýza, která bude zkoumat vzájemné vztahy mezi naměřenými teplotami. Korelační analýza byla opět zaznamenaná do tabulek, tentokrát však pro jednotlivé technologie. Výsledné tabulky zobrazují korelační koeficienty pro jednotlivé měsíce a budou pomáhat při výběru vhodné technologie pro predikci teploty okolí.

5.1 Testování normální distribuce naměřených teplot

Pro jednotlivé testování normální distribuce dat byl vytvořen nástroj pro zautomatizování provedených testů. Testování je totiž nutné provést na obrovském množství dat, tudíž pro urychlení a zjednodušení práce bylo vytvoření skriptu žádoucí. Vytvořený skript přebírá vstupní data, která následně upraví do takového tvaru, se kterým následně může pracovat. Vstupní data jsou prvně rozdělena po jednotlivých měsících a nástroj následně testuje normální distribuci dat pro jednotlivé měsíce zvlášť. Pro testování normální distribuce bylo zvoleno několik metod testování z důvodu větší výpovědní hodnoty a maximalizaci přesnosti provedené analýzy. Testování normality se obvykle neprovádí ručně z důvodu velké výpočetní náročnosti, tudíž implementace tohoto testování byla provedena pomocí Python knihoven[25]. Existuje mnoho metod, které lze použít pro testování normality, ve vytvořeném nástroji byly konkrétně implementovány tyto metody – Shapirův-Wilkův, Lillieforsův, Anderson-Darlingův a D’Agostino-Pearsonův test. Nástroj také zkoumá šikmost a špičatost distribuce dat a vypisuje jejich konkrétní koeficienty pro dané měsíce. Tyto koeficienty lze využít také pro popis naměřených teplot pro jednotlivé měsíce v roce. Testování samotných datasetů bylo provedeno na naměřených teploty s proměnlivou časovou periodou. Surová data v datasetech mají časovou periodu 5 minut,

pro odstranění některých krátkodobých výkyvu byly tyto hodnoty také agregovány na 30m/1h/2h a testy normality provedeny i na data s takovou časovou periodou. Data byla agregována tak, že se sečetly všechny hodnoty z tohoto časového intervalu a následně byl vypočítán jejich průměr. Výsledky provedených testů můžeme vidět v následujících tabulkách, kde jsou vypsány p hodnoty jednotlivých měsíců v roce. Jedná se o teploty naměřené meteostanicemi a k nim přilehlými mikrovlnnými zařízeními:

PIVIN -10min												
year	2021	2021	2021	2022	2022	2022	2022	2022	2022	2022	2022	2022
month	10	11	12	1	2	3	4	5	6	7	8	9
p_shapiro	1,69E-13	1,04E-20	1,8E-23	4,96E-35	6,21E-11	8,45E-25	8,76E-19	2,19E-15	3,02E-27	2,52E-27	1,51E-33	1,99E-17
p_lilliefors	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
p_ad	5,31E-11	0	0	0	7,03E-10	0	1,03E-21	9,96E-18	0	0	0	9,37E-25
p_normal	7,3E-12	1,04E-40	1,6E-15	2,32E-81	0,004716	1,74E-44	9,55E-26	1,73E-15	1,29E-52	5,63E-41	8,3E-64	5,23E-24
skewness	0,199815	-0,10471	0,306769	0,720116	-0,11985	0,349315	0,190726	0,301337	0,608733	0,507052	0,685947	0,209536
kurtosis	2,704794	2,351541	3,069612	3,733507	2,918562	2,44165	2,483975	2,876266	3,256852	2,757143	2,984017	2,512959

10.0.3.157 -10min												
year	2021	2021	2021	2022	2022	2022	2022	2022	2022	2022	2022	2022
month	10	11	12	1	2	3	4	5	6	7	8	9
p_shapiro	6,82E-14	2,71E-18	1,22E-15	2,69E-26	2,17E-16	2,1E-23	1,45E-21	5,31E-12	2,31E-19	9,39E-23	3,11E-20	1,76E-18
p_lilliefors	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
p_ad	4,23E-21	0	6,24E-30	0	0	0	0	1,76E-20	0	0	0	2,02E-24
p_normal	1,24E-11	2,25E-08	3,1E-10	1,7E-47	1,41E-06	1,03E-39	4,18E-31	5,77E-14	4,17E-52	5,81E-54	3,36E-26	2,7E-32
skewness	0,157333	-0,2009	0,150405	0,549126	0,172347	0,311596	0,298075	0,07834	0,150432	0,242865	0,303873	0,16939
kurtosis	2,656429	2,821991	2,678608	3,356467	2,809451	2,450412	2,507896	2,570007	2,303622	2,329951	2,585258	2,419797

10.0.54.101 -10min												
year	2021	2021	2021	2022	2022	2022	2022	2022	2022	2022	2022	2022
month	10	11	12	1	2	3	4	5	6	7	8	9
p_shapiro	1,03E-29	8,67E-28	2,83E-18	3,51E-33	3,13E-27	1,18E-32	1,04E-31	1,64E-29	1,76E-30	2,7E-32	1,09E-34	1,34E-28
p_lilliefors	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
p_ad	0	0	0	0	0	0	0	0	0	0	0	0
p_normal	5,98E-53	1,19E-50	2,93E-12	1,06E-81	1,11E-77	1,12E-54	2,46E-62	1,92E-48	5,1E-55	4,67E-56	2,07E-62	2,25E-47
skewness	0,613702	0,500879	0,257363	0,71278	0,705812	0,584636	0,6837	0,511204	0,498571	0,542842	0,61288	0,586212
kurtosis	2,9881	3,849914	3,177232	3,806582	3,995055	2,673831	3,158386	2,60291	2,484404	2,556598	2,607801	2,964099

Tab. 5.1: Výsledky testování normality pro hodnoty s časovým intervalem 10 minut

V předchozí tabulce byla normální distribuce testována na hodnotách s periodou 10 minut. Můžeme vidět, že p hodnoty provedených testů dosahují opravdu malých hodnot blízkých se k 0. Tyto nízké hodnoty mohou být způsobeny větším výskytem krátkodobě odlehých hodnot a šumem v datech. Z tohoto důvodu došlo ke zprůměrování hodnot na větší časový interval, čím více hodnot bude zprůměrováno, tím dojde k co největší eliminaci těchto dvou jevů. Následující tabulka zobrazuje p hodnoty pro agregované hodnoty po 30 minutách:

PIVIN -30min												
year	2021	2021	2021	2022	2022	2022	2022	2022	2022	2022	2022	2022
month	10	11	12	1	2	3	4	5	6	7	8	9
p_shapiro	3,08E-06	5,92E-11	9,47E-13	4,75E-21	0,000108	9,98E-14	1,24E-09	1,92E-07	1,87E-15	1,75E-15	6,22E-20	1,05E-08
p_lilliefors	0,018043	0,001	0,001	0,001	0,024019	0,001	0,001	0,001	0,001	0,001	0,001	0,001
p_ad	0,000965	8,02E-15	2,75E-24	0	0,002366	6,73E-16	1,67E-07	4,64E-06	2,02E-22	1,23E-20	0	1,84E-08
p_normal	0,00022	5,02E-14	1,05E-05	5,82E-28	0,177658	2,91E-15	5,68E-09	1,23E-05	2,91E-18	3,54E-14	7,6E-22	2,38E-08
skewness	0,198859	-0,10663	0,306264	0,720924	-0,11762	0,348851	0,192021	0,300502	0,60963	0,506678	0,684437	0,206606
kurtosis	2,703511	2,348283	3,074205	3,727229	2,913025	2,439473	2,483588	2,874623	3,257087	2,75625	2,982495	2,511698

10.0.3.157 -30min												
year	2021	2021	2021	2022	2022	2022	2022	2022	2022	2022	2022	2022
month	10	11	12	1	2	3	4	5	6	7	8	9
p_shapiro	1,35E-05	1,29E-07	6,95E-06	9,22E-14	4,12E-05	1,65E-12	3,41E-11	6,91E-05	8,25E-10	3,81E-12	3,07E-10	4,85E-09
p_lilliefors	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
p_ad	2,86E-05	9,36E-10	2,27E-06	1,14E-20	2,32E-05	4,25E-16	5,49E-11	6,38E-06	4,57E-11	9,57E-14	1,01E-12	4,15E-07
p_normal	0,000235	0,001866	0,000706	8,67E-17	0,011178	9,13E-14	7,79E-11	3,9E-05	3,86E-18	1,55E-18	3,06E-09	3,78E-11
skewness	0,158721	-0,20813	0,151263	0,554392	0,171764	0,313064	0,300484	0,079576	0,151233	0,243598	0,304835	0,170334
kurtosis	2,654077	2,813368	2,676465	3,361634	2,802924	2,447306	2,507467	2,566375	2,296	2,32557	2,582577	2,418449

10.0.54.101 -30min												
year	2021	2021	2021	2022	2022	2022	2022	2022	2022	2022	2022	2022
month	10	11	12	1	2	3	4	5	6	7	8	9
p_shapiro	4,26E-17	9,95E-16	2,94E-09	1,13E-19	2,62E-15	2,91E-19	1,41E-18	3,25E-17	1,05E-17	5,21E-19	1,07E-20	2,65E-16
p_lilliefors	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
p_ad	9,32E-27	8,81E-24	3,24E-15	0	1,76E-15	0	1,05E-29	0	0	0	0	3,46E-22
p_normal	4,01E-18	3,42E-17	0,000135	5,6E-28	1,67E-26	9,77E-19	1,92E-21	5,63E-17	8,84E-19	3,88E-19	2,7E-21	2,94E-16
skewness	0,610699	0,495074	0,256018	0,711355	0,700702	0,583346	0,683159	0,512898	0,49805	0,541823	0,609909	0,583217
kurtosis	2,982731	3,823115	3,176462	3,80051	3,987815	2,671475	3,158378	2,590934	2,48371	2,555919	2,602619	2,960787

Tab. 5.2: Výsledky testování normality pro hodnoty s časovým intervalem 30 minut

Z tabulek zřetelné, že p hodnoty po zprůměrování na větší časový interval dosahují znatelně větších hodnot, hodnoty Lillieforsova testu dosahují konstantních hodnot 0,001 z důvodu toho, že tato hodnota je nejmenší hodnota, kterou je daná knihovna schopna vypsat. Konkrétně u meteostanice, tedy tabulka označená P1VIN, v únoru nelze zamítnout nulovou hypotézu o normálním rozložení dat, p hodnota totiž dosahuje větší hodnoty než 0,05.

Na následujících tabulkách, které zobrazují zprůměrované hodnoty po 1 a 2 hodinách, si můžeme všimnout, že zprůměrování hodnot má poměrně velký vliv na výsledky testů normality, p hodnoty se díky zprůměrování přibližují k hodnotě, kdy nelze zamítnout nulovou hypotézu o normálním rozložení dat. Na statistické ukazatele šikmosti a špičatosti agregace hodnot velký vliv nemá, jejich koeficienty se razantním způsobem nemění. Pokud porovnáme tabulky s časovou periodou 10 minut a 2 hodin, můžeme vidět, že dochází k razantním změnám v p hodnotách a dochází k většímu počtu výskytu buněk, u kterých nelze zamítnout nulovou hypotézu normálního rozdělení. Tabulky zobrazující výsledky pro zprůměrované hodnoty:

P1VIN -1h												
year	2021	2021	2021	2022	2022	2022	2022	2022	2022	2022	2022	2022
month	10	11	12	1	2	3	4	5	6	7	8	9
p_shapiro	0,001913	8,17E-07	5,35E-08	1,9E-14	0,020876	8,71E-09	8,7E-06	0,000278	4,47E-10	4,2E-10	1,76E-13	4,41E-05
p_lilliefors	0,365714	0,001	0,001	0,001	0,247217	0,00148	0,021641	0,014279	0,001	0,001	0,001	0,007367
p_ad	0,058777	1,12E-07	2,06E-12	1,8E-26	0,088016	3,42E-08	0,000596	0,003081	1,48E-11	1,31E-10	1,32E-18	0,000245
p_normal	0,016313	1,93E-07	0,003035	1,33E-14	0,442267	5,76E-08	8,89E-05	0,003232	1,06E-09	1,7E-07	2,13E-11	0,000195
skewness	0,197731	-0,1105	0,305037	0,722239	-0,11443	0,348864	0,194325	0,301951	0,61181	0,507494	0,68389	0,202862
kurtosis	2,702166	2,340965	3,08071	3,715821	2,908276	2,436451	2,483519	2,870944	3,259585	2,757807	2,985412	2,510332

10.0.3.157 -1h													
year	2021	2021	2021	2022	2022	2022	2022	2022	2022	2022	2022	2022	2022
month	10	11	12	1	2	3	4	5	6	7	8	9	
p_shapiro	0,006039	0,000236	0,005217	1,2E-08	0,018082	7,03E-08	6,38E-07	0,011431	3,12E-06	9,59E-08	3,71E-06	1,92E-05	
p_lilliefors	0,027412	0,001	0,01252	0,001	0,02443	0,001	0,002357	0,017226	0,001	0,001	0,001	0,060928	
p_ad	0,01644	8,63E-05	0,00536	4,09E-10	0,029145	2,89E-08	1,23E-05	0,00514	3,35E-06	3,05E-07	1,82E-06	0,000947	
p_normal	0,016523	0,034869	0,028898	5,29E-09	0,109888	2,98E-07	9,86E-06	0,007025	9,93E-10	1,28E-09	5,49E-05	7,32E-06	
skewness	0,158838	-0,21571	0,152464	0,5591	0,173733	0,31647	0,301906	0,086944	0,156445	0,24915	0,307776	0,17069	
kurtosis	2,652008	2,804588	2,676516	3,357409	2,807162	2,44496	2,507267	2,567053	2,283387	2,322407	2,582032	2,416236	

10.0.54.101 -1h													
year	2021	2021	2021	2022	2022	2022	2022	2022	2022	2022	2022	2022	2022
month	10	11	12	1	2	3	4	5	6	7	8	9	
p_shapiro	3E-11	3,54E-10	1,73E-05	2,69E-13	6,57E-10	5,56E-13	1,93E-12	1,66E-11	8,17E-12	8,22E-13	4,51E-14	1,19E-10	
p_lilliefors	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	
p_ad	1,81E-13	2,55E-12	5,99E-08	2,82E-22	6,53E-08	7,33E-18	2,21E-15	6,19E-19	4,31E-17	3,69E-18	3,92E-25	2,57E-11	
p_normal	2,27E-09	1,29E-08	0,011312	1,65E-14	9,2E-14	9,3E-10	3,46E-11	5,15E-09	1,02E-09	6,12E-10	5,05E-11	1,86E-08	
skewness	0,604424	0,479724	0,25292	0,70848	0,694989	0,581663	0,681483	0,514365	0,498723	0,543522	0,608082	0,577789	
kurtosis	2,973438	3,769665	3,177653	3,789383	3,980782	2,667726	3,153054	2,579284	2,484369	2,558351	2,600144	2,94838	

Tab. 5.3: Výsledky testování normality pro hodnoty s časovým intervalem 1 hodiny

P1VIN01 -2h													
year	2021	2021	2021	2022	2022	2022	2022	2022	2022	2022	2022	2022	2022
month	10	11	12	1	2	3	4	5	6	7	8	9	
p_shapiro	0,066293	0,000455	0,000114	1,72E-09	0,261097	2,83E-05	0,002959	0,020764	2,81E-06	3,08E-06	9,78E-09	0,007577	
p_lilliefors	0,67876	0,008761	0,001	0,001	0,609745	0,038018	0,179286	0,085989	0,001	0,014082	0,001	0,089363	
p_ad	0,342849	0,000345	2,2E-06	6,11E-14	0,455527	0,000302	0,038196	0,060811	3,82E-06	1,86E-05	2,24E-09	0,019823	
p_normal	0,128642	0,000382	0,04882	5,98E-08	0,653318	0,000268	0,011373	0,047102	1,67E-05	0,000365	3,58E-06	0,015924	
skewness	0,200186	-0,1114	0,305828	0,72403	-0,11858	0,346968	0,198771	0,310522	0,621194	0,509881	0,682518	0,196936	
kurtosis	2,696855	2,325723	3,085232	3,712627	2,905856	2,43059	2,485389	2,864186	3,276365	2,762318	2,988801	2,503055	

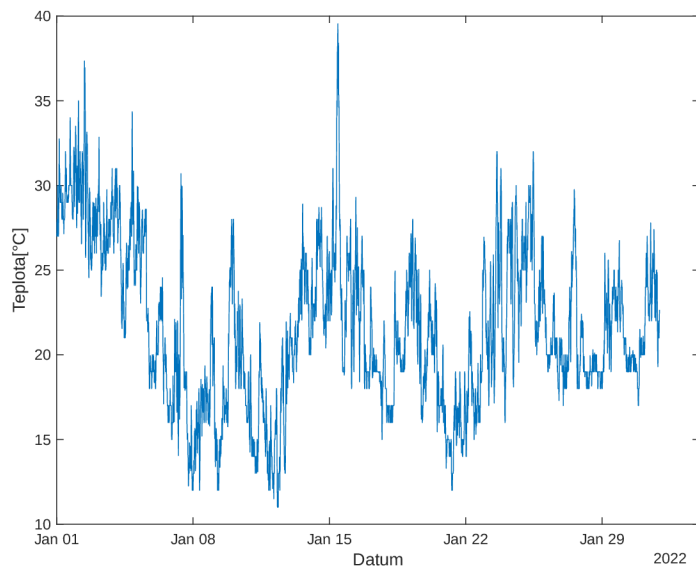
10.0.3.157 -2h													
year	2021	2021	2021	2022	2022	2022	2022	2022	2022	2022	2022	2022	2022
month	10	11	12	1	2	3	4	5	6	7	8	9	
p_shapiro	0,071126	0,005566	0,132621	2,2E-05	0,329864	0,000111	0,000483	0,111622	0,002477	6,96E-05	0,00139	0,004915	
p_lilliefors	0,076511	0,022548	0,229617	0,001	0,91793	0,00298	0,083039	0,024234	0,051241	0,011848	0,001648	0,43408	
p_ad	0,113935	0,006242	0,109473	1,68E-05	0,628979	0,000187	0,004287	0,042669	0,006168	0,000231	0,002864	0,054639	
p_normal	0,097949	0,119418	0,178966	3,25E-05	0,322334	0,000861	0,003878	0,078283	0,000149	4,68E-05	0,008476	0,002648	
skewness	0,172818	-0,24098	0,154113	0,574468	0,181473	0,325401	0,304992	0,115642	0,168608	0,27363	0,313323	0,172669	
kurtosis	2,627625	2,768721	2,672671	3,364374	2,810982	2,464663	2,51395	2,563576	2,312173	2,329037	2,593054	2,405187	

10.0.54.101 -2h													
year	2021	2021	2021	2022	2022	2022	2022	2022	2022	2022	2022	2022	2022
month	10	11	12	1	2	3	4	5	6	7	8	9	
p_shapiro	8,84E-07	8,66E-06	0,004399	1,01E-08	1,01E-05	2,23E-08	7,98E-08	2,14E-07	2,3E-07	2,86E-08	6,41E-09	1,89E-06	
p_lilliefors	0,009178	0,001	0,010975	0,001	0,070129	0,001	0,001	0,001	0,001	0,001	0,001	0,003234	
p_ad	1,47E-06	2,36E-06	0,000202	5,75E-12	0,00066	2,6E-09	5,62E-08	2,09E-10	1,19E-08	1,61E-09	3,37E-12	7,64E-06	
p_normal	7,4E-05	0,000581	0,107701	5,26E-08	3,67E-07	3E-05	5,03E-06	6,44E-05	5,12E-05	2,36E-05	8,97E-06	0,000153	
skewness	0,580436	0,433004	0,24832	0,711813	0,666627	0,581317	0,675656	0,524625	0,504961	0,551435	0,599107	0,566076	
kurtosis	2,906572	3,5856	3,156636	3,811996	3,962555	2,677563	3,145076	2,59309	2,516431	2,57166	2,604242	2,924266	

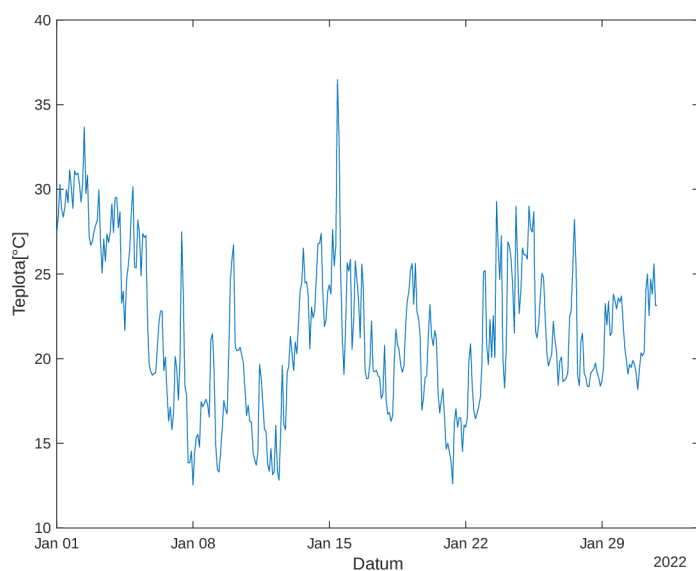
Tab. 5.4: Výsledky testování normality pro hodnoty s časovým intervalem 2 hodin

Zobrazení vlivu agregace hodnot

Pro zobrazení vlivu zprůměrování hodnot si můžeme zobrazit dva průběhy grafů, kde si zobrazíme průběh grafu naměřených teplot s časovou periodou 10 minut a následně hned druhý graf, který má tuto periodu nastavenou po 2 hodinách:



Obr. 5.1: Průběh grafu pro hodnoty naměřené s časovým intervalem 10 minut



Obr. 5.2: Průběh grafu po zprůměrování hodnot na 2 hodiny

Díky zprůměrování naměřených teplot na časový interval po 2 hodinách je krásně vidět, že průběh grafů je vyhlazenější a díky agregaci hodnot byl také odstraněn šum v datech. Hodnoty nemají takové výkyvy jako je tomu při hodnotách s periodou 5 minut.

Jelikož se při testování normální distribuce dat jednoznačně neprokázalo, zda naměřené teploty mají nebo nemají normální distribuci, bude vhodné při korelační analýze využít jak parametrické (Pearsonův korelační koeficient), tak neparametrické (Spearmanův koeficient pořadové korelace, Kendallův koeficient pořadové korelace) metody testování.

5.2 Korelační analýza

Pro co nejpřesnější výsledky korelační analýzy bylo nutné porovnávat teploty u obou zařízení, které jsou v těsné blízkosti sebe. Z tohoto důvodu byly na základě geografických pozic ze seznamu mikrovlnných zařízení vyfiltrována zařízení v okolí meteorologických stanic. Po vyfiltrování zařízení již nic nebránilo v započatí korelační analýzy. Pro tuto analýzu byl také pomocí Pythonu vytvořen nástroj, kde na vstupu jsou dva datasety, jeden dataset pro teploty naměřené meteorologickou stanicí a druhý pro hodnoty naměřené mikrovlnným zařízením. Datasety jsou následně na základě časového razítka spojeny do jedné tabulky, čímž bylo docíleno stavu, kdy bude zkoumána korelace pro hodnoty naměřené vždy ve stejný čas. Následně se tabulka rozdělí po jednotlivých měsících a korelační koeficienty jsou vypisovány právě pro jednotlivé měsíce. Vytvořený nástroj také ošetřuje stavy, kdy se v datasetech vyskytují jednotlivé mezery v datech, které jsou způsobeny například výpadkem daného zařízení, nebo nesmyslné hodnoty, tyto stavy bylo nutné ošetřit, aby bylo docíleno, co nejpřesnějších výsledků

Výsledky korelační analýzy pro jednotlivé spoje jsou zobrazeny v následujících tabulkách. Hodnoty teplot naměřených mikrovlnnými zařízeními jsou porovnávány s hodnotami naměřené meteorologickou stanicí. K této stanici bylo vyfiltrováno několik typů mikrovlnných zařízení pro porovnání, která technologie mikrovlnných spojů nejvíce koreluje s teplotou okolí. Tabulky s výsledky korelační analýzy provedené pro technologii *O* jsou zobrazeny na následující straně:

Technologie	IP	Vzdálenost		Orientace
O	10.0.21.137	0,386957334		176,6595674
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
June 2020	0,9522	0,8231	0,9587	1327
July 2020	0,9600	0,8243	0,9608	8864
August 2020	0,9625	0,8291	0,9602	8908
September 2020	0,9561	0,8166	0,9554	8061
October 2020	0,9190	0,7500	0,9165	8489
November 2020	0,9490	0,7840	0,9344	8427
December 2020	0,8738	0,6979	0,8697	8928
January 2021	0,9236	0,7614	0,9240	8906
February 2021	0,9126	0,7256	0,8654	651
April 2021	0,9666	0,8538	0,9702	7062
May 2021	0,9514	0,8112	0,9503	8914
June 2021	0,9622	0,8282	0,9618	8636
July 2021	0,9475	0,8033	0,9511	5414
December 2021	0,9430	0,7766	0,9198	1574
January 2022	0,9115	0,7328	0,9002	8860
February 2022	0,8949	0,7305	0,9003	8063
March 2022	0,9577	0,8138	0,9504	8848
April 2022	0,9631	0,8366	0,9636	8535
May 2022	0,9472	0,7995	0,9494	8922
June 2022	0,9577	0,8240	0,9600	8564
July 2022	0,9596	0,8197	0,9557	8922
August 2022	0,9605	0,8211	0,9533	8908
September 2022	0,9299	0,7541	0,9157	3922
October 2022	0,8665	0,6542	0,8370	3889
November 2022	0,8277	0,6395	0,8303	4627

Technologie	IP	Vzdálenost		Orientace
O	10.0.21.138	3,0549185		356,6612127
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
June 2020	0,9601	0,8499	0,9682	1327
July 2020	0,9690	0,8568	0,9726	8856
August 2020	0,9744	0,8661	0,9765	8908
September 2020	0,9707	0,8731	0,9767	8058
October 2020	0,9475	0,8259	0,9592	8489
November 2020	0,9384	0,8142	0,9385	8187
December 2020	0,8830	0,7456	0,9049	8928
January 2021	0,9130	0,7680	0,9185	8906
February 2021	0,8025	0,5593	0,7187	651
April 2021	0,9768	0,8911	0,9833	7062
May 2021	0,9669	0,8403	0,9623	8912
June 2021	0,9633	0,8438	0,9642	8560
July 2021	0,9649	0,8490	0,9685	5414
December 2021	0,9587	0,8259	0,9511	1574
January 2022	0,8877	0,7149	0,8753	8860
February 2022	0,8802	0,7232	0,8962	8063
March 2022	0,9694	0,8639	0,9726	8848
April 2022	0,9696	0,8616	0,9722	8535
May 2022	0,9680	0,8556	0,9721	8922
June 2022	0,9726	0,8641	0,9749	8564
July 2022	0,9799	0,8878	0,9831	8922
August 2022	0,9756	0,8826	0,9803	8908
September 2022	0,9597	0,8322	0,9616	3922
October 2022	0,9089	0,7595	0,9093	3889
November 2022	0,8867	0,7436	0,9026	4627

Předchozí dvě tabulky zobrazují korelační koeficienty pro mikrovlnná zařízení technologie *O*. Tato technologie dosahuje poměrně velkých korelačních koeficientů a tyto koeficienty se zdají být až na pár měsíců poměrně konzistentní, některé měsíce dosahují oproti ostatním lehce zkreslených hodnot z důvodu malého množství naměřených teplot, proto do tabulek byl přidán také sloupec, který zobrazuje konkrétní počet zkoumaných hodnot. Největších průměrných hodnot korelačních koeficientů dosahují tyto technologie pro slunečnější měsíce – jarní a letní. Naopak nejmenších hodnot tyto koeficienty dosahují pro podzimní a zimní měsíce, zde koeficienty v porovnání s předchozími měsíci dosahují znatelně menších hodnot. Z toho můžeme usoudit, že jednotky tohoto typu jsou více náchylné na změnu teploty při větším výskytu srážek, nebo větru, čímž je toto období v roce charakteristické, než při působení paprsků slunečního záření s větší intenzitou.

Následují tabulky korelací pro technologii *C*, konkrétně se jedná o *C1*, tato technologie se dělí do podkategorií *C1*, *C2* a *C3*. Korelační koeficienty byly zkoumány pro jednotlivé technologie zvlášť. Pro demonstraci byla vybrána dvě referenční zařízení. Tabulky korelačních koeficientů pro technologii *C1*:

Technologie	IP	Vzdálenost		Orientace
C1	10.0.21.174	0,386957334		211,8930192
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
June 2020	0,7079	0,4862	0,6668	1595
July 2020	0,7775	0,5967	0,7901	8888
August 2020	0,7880	0,6152	0,8020	8908
September 2020	0,8024	0,6382	0,8278	8638
October 2020	0,7332	0,5525	0,7262	8776
November 2020	0,7764	0,5608	0,7388	8427
December 2020	0,5919	0,4851	0,6350	8928
January 2021	0,6544	0,4733	0,6437	8906
February 2021	0,7403	0,4173	0,5669	651
August 2021	0,8509	0,5969	0,7762	5347
September 2021	0,8832	0,6733	0,8465	1250
December 2021	0,7866	0,6169	0,7971	1574
January 2022	0,6234	0,4181	0,5738	8899
February 2022	0,5570	0,4271	0,5831	8063
March 2022	0,8718	0,6841	0,8629	8848
April 2022	0,8572	0,6826	0,8626	8535
May 2022	0,7948	0,5910	0,7890	8922
June 2022	0,8272	0,6270	0,8159	8563
July 2022	0,8652	0,6644	0,8504	8922
August 2022	0,8870	0,7058	0,8712	8908
September 2022	0,7406	0,5390	0,7242	3922
October 2022	0,6016	0,3977	0,5549	3889
November 2022	0,5386	0,3571	0,5063	4627

Technologie	IP	Vzdálenost		Orientace
C1	10.0.21.175	1,61557858		31,88526777
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
June 2020	0,7785	0,5488	0,7403	1595
July 2020	0,8187	0,6260	0,8148	8885
August 2020	0,8274	0,6407	0,8254	8908
September 2020	0,8147	0,6264	0,8161	8635
October 2020	0,7817	0,5888	0,7804	8768
November 2020	0,7879	0,5399	0,7220	8427
December 2020	0,6210	0,4723	0,6371	8926
January 2021	0,7080	0,5250	0,7235	8906
February 2021	0,8317	0,6429	0,8208	651
August 2021	0,9034	0,6558	0,8241	5346
September 2021	0,9225	0,7407	0,9005	1250
December 2021	0,7498	0,6376	0,8196	1574
January 2022	0,6628	0,4258	0,5924	8899
February 2022	0,5956	0,4279	0,5999	8063
March 2022	0,9192	0,7455	0,9068	8848
April 2022	0,8816	0,7048	0,8848	8529
May 2022	0,8704	0,6811	0,8633	8922
June 2022	0,8726	0,6907	0,8668	8544
July 2022	0,9282	0,7605	0,9224	8922
August 2022	0,9164	0,7281	0,8942	8908
September 2022	0,7927	0,5873	0,7813	3918
October 2022	0,7535	0,5341	0,7064	3889
November 2022	0,6320	0,4301	0,5725	4627

Z předchozích tabulek je zřejmé, že korelační koeficienty pro technologii *C1* dosahují menších hodnot. Koeficienty mají celkově větší výkyvy oproti technologii *O*. Tato technologie obdobně jako předchozí dosahuje největších hodnot koeficientů pro slunečné měsíce, pro ostatní měsíce se však vyskytují poměrně velké výkyvy. Statistické ukazatele rozptyl a směrodatná odchylka u koeficientů pro technologii *C1* dosahují řádově větších hodnot, než je tomu tak u technologie označené *O*. Celkový průměrný rozptyl pro tuto technologii je se pohybuje nad 0,1 pro všechny 3 korelační koeficienty.

Následují korelační koeficienty také z technologie *C*, tentokrát se však jedná o *C2*. Pro zobrazení korelačních koeficientů byly opět vybrány dvě zařízení, které jsou viditelné na následující straně:

Technologie	IP	Vzdálenost		Orientace
C2	10.0.21.246	0,386957334		100,4386277
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
April 2021	0,8479	0,6731	0,8480	7062
May 2021	0,7124	0,5159	0,6893	8914
June 2021	0,8610	0,6801	0,8704	8636
July 2021	0,7835	0,5900	0,7807	5414
August 2021	0,8134	0,5675	0,7550	5348
September 2021	0,8725	0,6587	0,8361	1250
December 2021	0,7266	0,5796	0,7692	1574
January 2022	0,5554	0,3679	0,5107	8899
February 2022	0,4700	0,3684	0,5115	8063
March 2022	0,8565	0,6643	0,8466	8848
April 2022	0,8545	0,6847	0,8556	8534
May 2022	0,8712	0,6880	0,8680	8922
June 2022	0,8914	0,7210	0,8909	8563
July 2022	0,9129	0,7372	0,9021	8922
August 2022	0,9215	0,7545	0,9056	8908
September 2022	0,8217	0,6201	0,8015	3922
October 2022	0,6971	0,4752	0,6439	3889
November 2022	0,6469	0,4571	0,6293	4627

Technologie	IP	Vzdálenost		Orientace
C2	10.0.21.247	1,696393874		280,4530485
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
April 2021	0,8022	0,6116	0,8120	7062
May 2021	0,6892	0,4795	0,6572	8914
June 2021	0,8415	0,6510	0,8440	8636
July 2021	0,7832	0,5891	0,7891	5414
August 2021	0,8476	0,5681	0,7470	5348
September 2021	0,8193	0,6553	0,8305	1250
December 2021	0,7322	0,6124	0,7999	1574
January 2022	0,5806	0,3608	0,5109	8899
February 2022	0,4245	0,3062	0,4352	8063
March 2022	0,8791	0,6746	0,8588	8848
April 2022	0,8372	0,6498	0,8419	8530
May 2022	0,8986	0,7310	0,8976	8922
June 2022	0,9015	0,7354	0,9010	8563
July 2022	0,9319	0,7741	0,9275	8922
August 2022	0,9225	0,7499	0,9078	8908
September 2022	0,8796	0,6956	0,8695	3922
October 2022	0,8069	0,6008	0,7717	3889
November 2022	0,7820	0,5747	0,7441	4627

Z korelačních koeficientů je zřejmé, že technologie C2 se od předchozí C1 razantním způsobem neliší a hodnoty koeficientů vychází obdobně. Jednotlivé rozdíly mohou způsobit umístěním daných zařízení, nebo jejich azimutem.

Poslední ze série technologií C je $C3$. Její korelační koeficienty jsou zobrazeny v následujících tabulkách:

Technologie	IP	Vzdálenost		Orientace
C3	172.19.9.75	0,344300565		221,7104288
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
May 2020	0,9203	0,7538	0,9065	7038
June 2020	0,9147	0,7387	0,8955	8534
July 2020	0,9134	0,7543	0,9134	8863
August 2020	0,9309	0,7857	0,9319	8897
September 2020	0,9139	0,7618	0,9170	8627
October 2020	0,8558	0,6816	0,8465	8829
November 2020	0,8974	0,7121	0,8759	8426
December 2020	0,7916	0,6402	0,8002	8924
January 2021	0,8323	0,6629	0,8313	8906
February 2021	0,8796	0,6088	0,7465	651
April 2021	0,9456	0,8188	0,9482	7061
May 2021	0,9216	0,7597	0,9059	8910
June 2021	0,9426	0,8015	0,9440	8633
July 2021	0,9198	0,7672	0,9197	5414
August 2021	0,9355	0,7503	0,9029	5347
September 2021	0,9395	0,7803	0,9205	1250
December 2021	0,9014	0,7125	0,8629	1574
January 2022	0,8609	0,6743	0,8299	8897
February 2022	0,8352	0,6798	0,8422	8061
March 2022	0,9313	0,7757	0,9213	8832
April 2022	0,9389	0,8066	0,9428	8471
May 2022	0,9135	0,7488	0,9110	8905
June 2022	0,9288	0,7799	0,9284	8552
July 2022	0,9330	0,7760	0,9232	8903
August 2022	0,9343	0,7818	0,9231	8898
September 2022	0,8733	0,6949	0,8606	3917
October 2022	0,7876	0,5652	0,7413	3881
November 2022	0,7521	0,5750	0,7550	4621

Technologie	IP	Vzdálenost		Orientace
C3	172.19.9.76	2,718719724		41,69233928
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
May 2020	0,9654	0,8440	0,9612	7038
June 2020	0,9515	0,7944	0,9325	8533
July 2020	0,9642	0,8469	0,9634	8856
August 2020	0,9626	0,8471	0,9640	8881
September 2020	0,9594	0,8356	0,9585	8627
October 2020	0,9174	0,7675	0,9189	8828
November 2020	0,9239	0,7521	0,9011	8426
December 2020	0,7863	0,6221	0,8014	8924
January 2021	0,8788	0,7085	0,8733	8906
February 2021	0,8338	0,5425	0,6925	651
April 2021	0,9695	0,8723	0,9739	7061
May 2021	0,9523	0,8162	0,9466	8910
June 2021	0,9678	0,8575	0,9694	8633
July 2021	0,9555	0,8418	0,9643	5414
August 2021	0,9614	0,8164	0,9471	5295
September 2021	0,9634	0,8397	0,9591	1250
December 2021	0,8927	0,7329	0,8786	1574
January 2022	0,9003	0,7154	0,8744	8897
February 2022	0,9195	0,7805	0,9235	8061
March 2022	0,9685	0,8554	0,9670	8832
April 2022	0,9694	0,8627	0,9691	8471
May 2022	0,9582	0,8373	0,9589	8905
June 2022	0,9623	0,8510	0,9666	8544
July 2022	0,9751	0,8750	0,9755	8903
August 2022	0,9623	0,8395	0,9593	8880
September 2022	0,9428	0,7879	0,9293	3917
October 2022	0,8829	0,6829	0,8474	3881
November 2022	0,8486	0,6770	0,8488	4621

Můžeme vidět, že zařízení *C3*, poslední z rodiny *C*, oproti ostatním silně koreluje s teplotou okolí. Hodnoty korelačních koeficientů lze přirovnat k předchozí technologii *O*, avšak tato technologie pořád dosahuje průměrně dvakrát tak větších hodnot rozptylů a směrodatné odchylky, tudíž není tak konzistentní oproti *O* technologii.

Poslední technologií, kde byla zkoumána korelace s okolní teplotou je technologie *S*. Konkrétně se jedná o 3 typy zařízení této technologie – *S1*, *S2*, *S3*. Jedním z hlavních problémů u této technologie je malý vzorek naměřených teplot, teploty naměřené teplotními čidly v těchto zařízeních jsou snímány a ukládány zhruba od poloviny Listopadu minulého roku, přesto je do korelační analýzy zahrneme také:

Technologie	IP	Vzdálenost		Orientace
S1	172.19.9.18	0,386928424		138,6308989
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
November 2022	0,8953	0,7207	0,8633	4149
December 2022	0,9391	0,8313	0,9451	8770
January 2023	0,9372	0,8209	0,9389	8918
February 2023	0,9733	0,8802	0,9737	6829
March 2023	0,9543	0,8120	0,9277	1863

Technologie	IP	Vzdálenost		Orientace
S1	172.19.9.19	2,748726519		318,6746599
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
November 2022	0,9414	0,8382	0,9420	4149
December 2022	0,9505	0,8626	0,9540	8770
January 2023	0,9559	0,8716	0,9642	8918
February 2023	0,9712	0,8770	0,9651	6829
March 2023	0,9096	0,8209	0,9322	1868

Technologie	IP	Vzdálenost		Orientace
S1	172.19.9.64	0,344300565		186,6630667
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
November 2022	0,8953	0,7248	0,8661	4149
December 2022	0,9391	0,8339	0,9456	8770
January 2023	0,9347	0,8204	0,9367	8918
February 2023	0,9761	0,8893	0,9769	6829
March 2023	0,9480	0,8065	0,9271	1881

Technologie	IP	Vzdálenost		Orientace
S1	172.19.9.65	2,717531022		6,660109272
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
November 2022	0,9137	0,8018	0,9172	4149
December 2022	0,9409	0,8592	0,9540	8770
January 2023	0,9404	0,8564	0,9595	8918
February 2023	0,9602	0,8572	0,9549	6829
March 2023	0,9467	0,8456	0,9582	1883

Z dosavadních naměřených hodnot je zřejmé, že naměřené teploty zařízením technologie *S1* silně korelují s okolní teplotou. Tyto koeficienty dosahují obrovských hodnot oproti předchozím technologiím také v chladnějším období roku. Průměrně dosahují hodnoty koeficientů ještě větších hodnot než všechny ostatní technologie, je však nutné tyto výsledky brát s rezervou, a to z důvodu malého zkoumaného vzorku dat. Dalším z rodiny S je *S2*, korelační koeficienty pro tuto technologii jsou zobrazeny v následujících tabulkách:

Technologie	IP	Vzdálenost		Orientace
S2	172.19.9.42	0,344300565		86,98235539
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
November 2022	0,7658	0,5696	0,7337	4149
December 2022	0,8823	0,7406	0,8876	8770
January 2023	0,8686	0,7243	0,8782	8918
February 2023	0,9342	0,7929	0,9311	6829
March 2023	0,8980	0,6793	0,8267	1884

Technologie	IP	Vzdálenost		Orientace
S2	172.19.9.43	2,569363437		267,0070251
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
November 2022	0,8867	0,7224	0,8653	4149
December 2022	0,9190	0,7905	0,9230	8770
January 2023	0,9256	0,7996	0,9307	8918
February 2023	0,9548	0,8318	0,9539	6829
March 2023	0,9489	0,7954	0,9217	1884

Koeficienty pro tuto technologii dosahují menších hodnot oproti *S1*, ale pořád se jedná o poměrně vysoké hodnoty. Až na jeden měsíc na jedné straně spoje se zdají být korelační koeficienty poměrně konzistentní. Pro poslední z rodiny *S* byly zkoumány korelační koeficienty pouze pro jedno zařízení, jelikož se žádné další v okolí meteorologické stanice nenachází a korelační koeficienty by nebyly dostatečně vypovídající. Tabulka s korelačními koeficienty:

Technologie	IP	Vzdálenost		Orientace
S3	172.19.9.71	0,344300565		203,1532251
Měsíc a rok	Pearson koeficient	Kendall koeficient	Spearman koeficient	Počet vzorků
November 2022	0,9186	0,7596	0,8901	4149
December 2022	0,9459	0,8503	0,9531	8770
January 2023	0,9465	0,8415	0,9475	8918
February 2023	0,9814	0,9039	0,9815	6829
March 2023	0,9579	0,8361	0,9428	1885

Zařízení technologie *S3* také poměrně silně koreluje s okolní teplotou, problémem však je, že bohužel nejsou k dispozici také historická data těchto zařízení. Zařízení se zatím jeví méně náchylné na změnu naměřené teploty při působení vlhkosti, srážek a větru, což je pro chladnější období v roce charakteristické.

Technologie *O*, *S1* a *S3* dosahují největších koeficientů korelací. Tento fakt zobrazuje následující tabulka, která zobrazuje průměrné korelační koeficienty pro jednotlivé technologie:

Průměrné korelační koeficienty pro jednotlivé technologie			
Technologie	Pearson	Kendall	Spearman
O	0,93807	0,79818	0,93555
C1	0,75733	0,56829	0,73843
C2	0,85301	0,68431	0,83880
C3	0,79101	0,60620	0,77945
S1	0,94131	0,82940	0,93863
S2	0,89840	0,74463	0,88520
S3	0,95007	0,83827	0,94298

Tyto technologie jsou také poměrně konzistentní v korelačních koeficientech, což si můžeme zobrazit v následující tabulce, která zobrazuje průměrné odchylky pro jednotlivé technologie:

Průměrné odchylky pro jednotlivé technologie			
Technologie	Pearson	Kendall	Spearman
O	0,03133	0,05123	0,03532
C1	0,10203	0,10405	0,11275
C2	0,06979	0,08014	0,07982
C3	0,09404	0,09370	0,09730
S1	0,01595	0,03019	0,02110
S2	0,03098	0,04385	0,03924
S3	0,01567	0,03234	0,02125

Na základě provedené korelační analýzy došlo k závěru, že nejvhodnějšími kandidáty pro predikci teploty okolí jsou mikrovlnné zařízení technologie *O*, *S1*, *S3*. To znamená, že tyto technologie by mohly poskytnout nejpřesnější predikci na základě jejich naměřených teplot. Technologie rodiny *S* dosahují největších korelačních koeficientů a také mají průměrně nejmenší odchylky, problémem však je jejich malý vzorek historických dat. Proto pro predikci teploty okolí byly zvoleny mikrovlnné zařízení technologie *O*, ke kterým máme k dispozici velké množství historických dat.

6 Vytvoření modelu založeného na strojovém učení

Pro vytvoření co nejpřesnějšího modelu založeného na strojovém učení je nutné před samotným vytvářením a trénováním modelu provést několik klíčových kroků. Tento proces zahrnuje přípravu a shromažďování dat, výběr vhodného modelu strojového učení, trénování modelu a následná validace a testování výsledků, které daný model predikuje na základě vstupních dat. [26] Vytvořený model strojového učení bude predikovat teplotu okolí na základě několika vstupních parametrů z jednotlivých zařízení umístěných na území celé ČR. Samotné řešení vývoje modelu strojového učení si nyní detailně popíšeme v následujících podkapitolách.

6.1 Shromažďování a příprava dat

Pro shromažďování a přípravu dat byl vytvořen Python skript pro tvorbu trénovacích dat, který zpracovává několik datových souborů - data z jednotlivých MW zařízení, naměřená data z meteostanic a pracuje s databází ukládající geografické souřadnice jednotlivých zařízení. Surová data z mikrovlnných zařízení obsahují několik parametrů, z nichž některé budou využity pro predikci okolní teploty. Shromažďované parametry zahrnují časové razítko, IP adresu zařízení, přijímaný signál a teplotu zařízení měřenou teplotním čidlem v mikrovlnném zařízení.

Vytvořený skript na základě IP adresy přiřazuje geografické souřadnice k jednotlivým zařízením, umožňuje přiřadit jejich azimut a vypočítat vzdálenost mezi geografickými souřadnicemi pomocí knihovny `geopy`. Díky těmto informacím lze určit nejbližší meteorologickou stanici, jejíž naměřené teploty okolí poslouží jako referenční hodnoty pro trénování algoritmu a testování modelu.

Pro správné sladění dat z různých zdrojů bylo nutné spojit teploty z meteorologických stanic a parametry z MW zařízení na základě časového razítka, což bylo provedeno pomocí funkce `pd.merge()`.

Dalším parametrem pro trénování modelu je sloupec `Sun` ve finálním datovém rámci, který určuje, zda má sluneční záření vliv na jednotku v daném záznamu (řádku). Pokud slunce svítí, má hodnotu 1, v opačném případě 0. K implementaci tohoto parametru byla použita knihovna `Astral`, sloužící k výpočtu časů různých aspektů slunce [27]. V našem případě bylo třeba zjistit časy východu a západu Slunce pro ČR v průběhu roku, takže záznamy s časovým razítkem mezi východem a západem Slunce mají hodnotu 1, ostatní 0.

Jako poslední krok bylo časové razítko ve sloupci `Timestamp` převedeno na unixový formát pro jednodušší reprezentaci času. Výstupem Python skriptu je CSV sou-

bor obsahující následující parametry: *Temperature_Meteo*, *Timestamp*, *Unix*, *Temperature_MW*, *Signal*, *IP*, *Sun* a *Azimut*. Výstupní soubory reprezentují parametry z jednotlivých MW zařízení a k nim přilehlých meteorologických stanic, jedná se o zaznamenaná data za celý jeden rok s časovou periodou 10 minut.

Prvních pár řádků výstupního souboru pro jedno MW zařízení si zobrazíme v následující tabulce:

Tab. 6.1: První záznamy parametrů MW zařízení

Teplota_Meteo	Časové razítko	Unix	Teplota_MW	Signál	IP	Sun	Azimut
4,9	01.10.2021 0:00	1633046400	31,524	-61,233	10.0.65.103	0	252,7229
4,6	02.10.2021 0:10	1633046400	31,949	-61,352	10.0.65.103	0	252,7229
5,0	03.10.2021 0:20	1633047600	31,127	-61,597	10.0.65.103	0	252,7229
4,4	04.10.2021 0:30	1633047600	31,774	-61,428	10.0.65.103	0	252,7229

6.2 Výběr vhodného modelu strojového učení

Pro výběr vhodného modelu testování bylo nutné pomocí vytvořeného skriptu pro přípravu dat vytvořit několik datových souborů, které byly využity pro natrénování modelů a následné testování jejich přesnosti, která bude určena na základě zvolených koeficientů metrik. Pro případ predikce teploty okolí na základě parametrů naměřené mikrovlnným zařízením byly zvoleny regresní modely strojového učení.

Regresní modely strojového učení jsou jedním z hlavních nástrojů pro predikci časových řad, jako je teplota, a jsou zvláště vhodné pro tento účel, protože mohou zpracovávat velké množství dat a zohledňovat různé vstupní proměnné [28]. Tyto modely se právě často používají s dobrými výsledky v oblasti předpovědi počasí.[29] V rámci výběru konkrétního regresního modelu strojového učení bylo provedeno natrénování a následné testování jednotlivých modelů strojového učení. Pro náš konkrétní případ byly vybrány tyto konkrétní modely strojového učení – Linear Regression, Lasso, Elastic Net, SVR(Support Vector Regression), Decision Tree, Random Forest.

Vstupní data byla rozdělena na trénovací a testovací pomocí python knihovny scikit-learn a tohoto řádku kódu:

```
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.3, random_state=42, shuffle=False)
```

Pomocí tohoto kódu jsme schopni natrénovat a následně ověřit výkonnost modelu a získat tak realističtější odhad jeho přesnosti. Parametr `test_size` udává, v jakém poměru budou data rozdělena na trénovací a testovací, konkrétně parametr tedy určuje velikost testovací sady. Další parametr `random_state` slouží ke specifikaci

semínka pro generování náhodných čísel, tímto parametrem lze zajistit reprodukovatelnost rozdělení dat. Poslední boolean parametr `shuffle`, určuje, zda mají být data náhodně zamíchána před rozdělením. V případě predikce průběhu teploty byl tento parametr nastaven na hodnotu `False`, čímž bylo zachováno pořadí dat.

Všechny vybrané modely byly natrénovány na předzpracovaných trénovacích datech a následně byl proveden test jednotlivých modelů s cílem zvolit ten nejvhodnější a nejvýkonnější model pro predikci teploty. Pro vyhodnocení kvality modelů byly použity následující metriky: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) a R-Squared (R2). Výsledky koeficientů pro jednotlivé metriky jsou zobrazeny v následující tabulce:

Tab. 6.2: Koeficienty metrik pro jednotlivé modely strojového učení

Model	MSE	MAE	RMSE	R2
Linear Regression	8.539309	2.044402	2.922210	0.884947
Lasso	9.710007	2.369239	3.116088	0.869173
Elastic Net	15.924210	3.215541	3.990515	0.785447
SVR	20.860839	3.547667	4.567367	0.718934
Decision Tree	15.536850	2.748843	3.941681	0.790666
Random Forest	10.222069	2.210721	3.197197	0.862274

Mean Squared Error(MSE) značí průměrnou kvadratickou chybu mezi skutečnými a předpovídanými hodnotami. Je to metrika citlivější na výrazné odchylky než MAE. [30]

Mean Absolute Error(MAE) je metrika, která značí průměr rozdílu mezi původními hodnotami a předpovídanými hodnotami. Udává nám míru toho, jak daleko byly předpovědi od skutečného výsledku. [30]

Root Mean Squared Error(RMSE) je metrika, která dává představu o velikosti chyby modelu. Ukazuje průměrnou kvadratickou chybu v jednotkách závislé proměnné. [31]

A jako poslední R-squared(R2) koeficient určuje, kolik procent variability dat je model schopný vysvětlit variabilitou regresní přímky. [31] Po pečlivém zhodnocení výsledků metrik byla zvolena Lineární regrese jako nejvhodnější model pro predikci teplot.

Důvodem pro zvolení Lineární regrese byl samotný výkon modelu, jelikož metriky pro hodnocení výkonnosti modelu dosahovali nejlepších výsledků. Další výhodou lineární regrese je rychlost trénování a následné predikce hodnot a to z toho důvodu, že natrénovaný model bude pracovat s obrovským množstvím dat, tudíž je rychlost zpracování dalším klíčovým parametrem při výběru modelu strojového učení.[32]

6.3 Natrénování modelu strojového učení

V této podkapitole se podrobněji zaměříme na proces vytváření modelu strojového učení a optimalizaci tréninkového postupu s cílem dosáhnout co nejpřesnějších predikcí teplot. Pro účinný trénink modelu bylo nutné zajistit větší množství vstupních trénovacích dat, než bylo původně vytvořeno pro výběr vhodného modelu strojového učení. Jako zdroj dat bylo vybráno několik mikrovlnných zařízení technologie *O* umístěných v blízkosti Prahy, jelikož dostupná referenční data z meteorologických stanic se také nachází v tomto regionu. Následně po přípravě trénovacích dat se můžeme zaměřit na trénování modelu strojového učení založeného na Lineární regresi.

Zdrojový kód pro vytváření, natrénování a následné testování výkonu modelu používá tyto hlavní knihovny:

- pandas: manipulace a analýzu dat.
- sklearn: strojové učení obsahující různé algoritmy a nástroje.
- matplotlib: vizualizace dat pomocí grafů.
- numpy: numerické výpočty.
- joblib: ukládání a načítání modelů.

Po načtení těchto knihoven bylo nutné načíst připravené trénovací datové sady pomocí knihovny Pandas, následně pomocí metody `pd.concat` došlo ke sloučení datových souborů do jednoho datového rámce. Dále jsou data rozdělena na vstupní (X) a výstupní hodnoty (Y). Výstupní hodnoty jsou referenční teploty okolí naměřené okolní meteorologickou stanicí, které konkrétně reprezentuje sloupec s názvem *Temperature_Meteo*. Na druhé straně vstupní hodnoty obsahují všechny zbylé sloupce, kromě dvou sloupců *Timestamp* a *IP*, které mají v datových sadách pouze informativní účely. Abychom získali lepší odhad výkonnosti modelu a maximalizovali využití dostupných dat, byla použita technika křížové validace místo dělení dat na trénovací a testovací množiny. Snížením velikosti trénovacích dat totiž riskujeme ztrátu důležitých vzorců/trendů v sadě dat. Křížová validace rozdělí datové sady do několika částí (tzv. "foldů") a následně se v každém kroku křížové validace jeden z foldů použije jako testovací množina, zatímco ostatní foldy slouží jako trénovací. Tento proces je opakován pro každý fold, což znamená, že každý fold je použit jak pro trénování, tak pro testování modelu. [33]

Po přípravě dat bylo nutné definovat, natrénovat a uložit samotný model strojového učení. Pro tento účel slouží následující řádky kódu:

Výpis 6.1: Ukázka trénování a ukládání modelu v Pythonu

```
# Definice modelu
model = LinearRegression()

# Natrénování modelu
model.fit(X, Y)

# Uložení natrénovaného modelu
joblib.dump(model, "Linear_model.joblib")
```

Tato část kódu v první řadě vytváří instanci lineárního regresního modelu z knihovny Sklearn a přiřazuje ji proměnné `model`. Pomocí metody `fit`, vytvořenou instanci natrénuje na datových souborech `X` a `Y`. Pro pozdější použití natrénovaného modelu je nutné jej uložit, tento proces zprostředkovává funkce `dump`, která ukládá natrénovaný model do souboru s názvem `Linear_model.joblib`.

Pro vyhodnocení výkonnosti zvoleného modelu strojového učení byly vybrány shodné metriky, které byly použity při selekci nejvhodnějšího modelu. K těmto ukazatelům byly přidány další metriky, které byly získány prostřednictvím křížové validace - průměrná přesnost, směrodatná odchylka. Současně bylo také vytvořeno několik diagramů, které mají za úkol poskytnout vizuální prezentaci a usnadnit interpretaci získaných výsledků predikovaných modelem. Tabulka zobrazující výsledky jednotlivých metrik po natrénování modelu:

Tab. 6.3: Výsledky metrik modelu

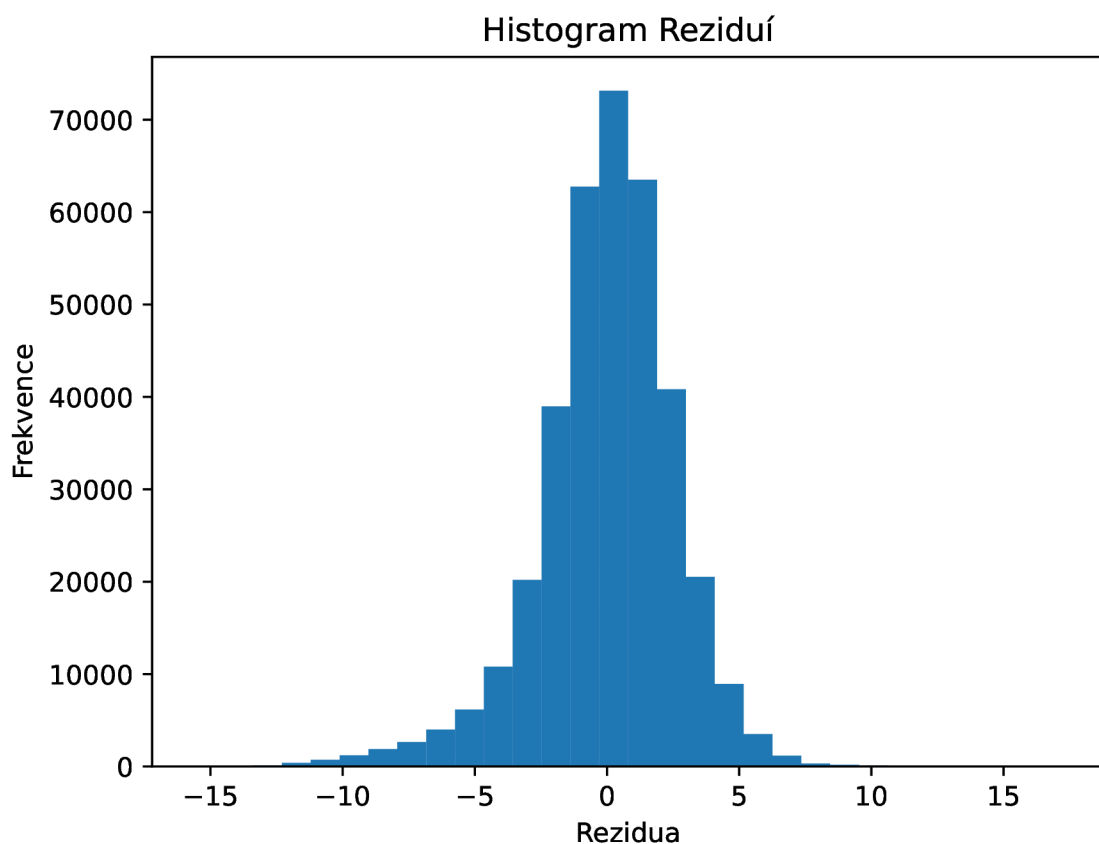
Metrika	Hodnota
Average Score	0.89
Score Deviation	0.02
Mean Squared Error	6.72
Mean Absolute Error	1.92
Root Mean Squared Error	2.59
Average Score	0.89

Celkově lze konstatovat, že natrénování modelu proběhlo úspěšně, přidané metriky křížové validace udávají průměrnou přesnost 89% s nízkou odchylkou 0,2 značí, že natrénovaný model je konzistentní přes různé podmnožiny dat. R-squared značí, že model dokáže vysvětlit 90% variability dat, Mean Absolute Error ukazuje, že průměrná chyba předpovědi modelu je 1,92 jednotek. Na základě dostupných parametrů k předpovědi lze říct, že model je natrénován dobře a předpovídá relativně přesné hodnoty.

Nyní se zaměříme na vizuální interpretaci výsledků pomocí grafů a diagramů. Tyto vizualizace nám umožní lépe porozumět výkonu modelu strojového učení a jeho schopnosti předpovědět správné hodnoty. Analýzou těchto vizualizací získáme komplexnější pohled na výkonnost a kvalitu natrénovaného modelu.

Prvním grafem pro vizuální analýzu natrénování modelu je histogram reziduí. Tento histogram zobrazuje frekvenci reziduí, v ideálním případě by měly být hodnoty nejvíce frekventované v oblasti nuly a histogram by měl mít tvar zvonové křivky (normální rozdělení). Histogram reziduí můžeme vidět na následujícím obrázku:

Obr. 6.1: Histogram reziduí

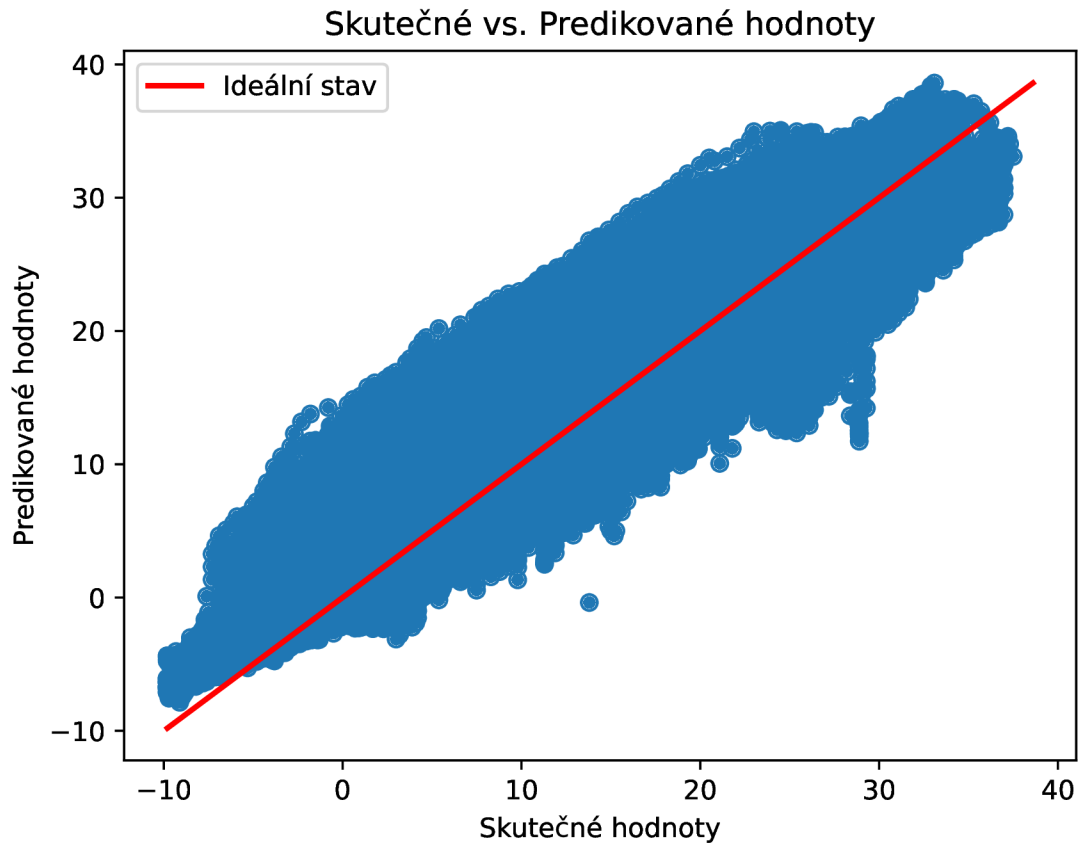


Z histogramu je možné vyčíst, že nejvíce reziduí je frekventovaných kolem nuly, jak by tomu mělo být v ideálním případě. Frekvence hodnot reziduí by se dalo říct, že je relativně symetrická kolem nuly, mírná asymetrie se vyskytuje v odlehlých hodnotách a frekvence výskytu těchto odlehlých hodnot je v celkovém počtu hodnot zanedbatelná. [34]

Dalším prvkem vizuální analýzy je graf skutečných vs. predikovaných hodnot, tím je zobrazen vztah mezi skutečnými (osa x) a predikovanými (osa y) hodnotami. V ideálním případě by většina bodů měla ležet na červené diagonální čáře. V praxi

je však ideálního stavu dosaženo však velmi vzácně. Graf zobrazující vztah mezi skutečnými a predikovanými hodnotami je zobrazen níže:

Obr. 6.2: Vztah mezi skutečnými a predikovanými teplotami



Z grafu je zřejmé, že vztah mezi skutečnými hodnotami a predikovanými hodnotami modelem splňuje tvar lineární křivky. Je ale také vidět, že v některých případech dochází k odchylce predikované teploty od té skutečné. Tento stav může být zapříčiněn několika faktory jako jsou například nepravidelnost, nebo výskyt vztahů, které nemohou být plně vysvětleny pouze použitím dostupných vstupních proměnných.

Vyvinutý model strojového učení byl otestován a podroben zkoumání z hlediska výkonnosti a přesnosti předpovědí. Vizualní analýza prostřednictvím vybraných grafů potvrdila, že tento model je schopen předpovídat teploty s dostatečnou přesností. Ačkoliv byly zaznamenány určité odchylky mezi predikovanými a skutečnými hodnotami, tyto odchylky byly vzhledem k dostupným parametrům očekávány a ve srovnání s celkovým výkonem modelu jsou zanedbatelné. S ohledem na dosažené výsledky lze model využít pro predikci teploty okolí na základě parametrů jednotlivých mikrovlnných spojů.

7 Vývoj zabezpečené webové aplikace

Tato kapitola popisuje proces vývoje samotné webové aplikace, která byla vyvinuta za účelem plošného monitoringu teploty okolí na území ČR, teploty budou určeny na základě vstupních parametrů z jednotlivých mikrovlnných zařízení technologie *O*. K monitoringu teplot jsou využívána data z telekomunikačních zařízení, které pokrývají rozsáhlé území České republiky. Výstupem aplikace je vykreslení 2D teplotní mapy s posuvnou časovou osou, díky které bude možné zobrazit průběh teplot v jednotlivých oblastech. Webová aplikace se skládá z frontendové a backendové části, kde frontendová část se týká konkrétně koncových uživatelů, co uživatelé vidí a s čím interagují přímo v jejich webovém prohlížeči (design, uživatelské rozhraní, funkčnost). Na druhé straně backendová část se zaměřuje na tu část webové aplikace, kterou koncový uživatel nevidí, provádí logiku aplikace, komunikuje s databází a dalšími službami. Obě tyto části společně tvoří komplexní webovou aplikaci. [35]

Backendová část aplikace byla naprogramována v jazyce Python (CPython 3.10.3), a to z důvodu výskytu velkého množství potřebných knihoven, které pomohly v implementaci webové aplikace. Frontendovou část webové aplikace zajišťují základní technologie používané při vytváření webových stránek a aplikací, konkrétně se jedná o programovací jazyky HTML (struktura), CSS (vzhled) a Javascript (interaktivita). V následujících kapitolách se podrobněji zaměříme na technologie a postupy použité při vývoji této aplikace.

7.1 Návrh webové aplikace

Jak bylo dříve uvedeno, webová aplikace bude sloužit k zobrazení teplotní mapy České republiky založené na vstupních datech a predikcích modelu strojového učení. Aplikace bude zpracovávat data z vybraných telekomunikačních zařízení, která budou získávána z databáze InfluxDB prostřednictvím API. V rámci této aplikace byl kladen velký důraz na ochranu soukromí, zejména na zajištění, že geografické polohy jednotlivých zařízení budou zcela skryty před koncovými uživateli.

7.2 Použité technologie

7.2.1 Databáze

InfluxDB

Jedná se o časově orientovanou databázi, která je optimalizovaná pro rychlé zpracování časových řad. Databázový systém je založen na verzi 2.6.1 a používá vlastní

dotazovací jazyk s názvem Flux. Konkrétně pro webovou aplikaci databáze uchovává periodicky měřené parametry mikrovlnných spojů a společně s časovým razítkem měření je ukládá, parametry jsou do databáze zapisovány v reálném čase. Komunikace aplikace s databází probíhá prostřednictvím API, na základě vygenerovaného tokenu dokáže webová aplikace z databáze číst. [36]

PostgreSQL

Jedná se o open-source objektově relační databázový systém, který využívá dotazovací jazyk SQL. Klíčové prvky pro zvolení této databáze byl její výkon a velké množství možností zabezpečení celého systému před únikem uložených dat. Databáze ukládá data ve formě tabulky jednotlivé metadata mikrovlnných spojů (IP adresa, technologie, orientace jednotky, zeměpisná šířka a délka zařízení). Webová aplikace s je s databází propojena a na základě IP adresy a konkrétního SQL dotazu přiřazuje k jednotlivým záznamům z předchozí InfluxDB databáze orientaci a geografické souřadnice zařízení. [37]

7.2.2 Hlavní knihovny

Webová aplikace pro svůj chod používá několik knihoven, z nichž nejdůležitější knihovny jsou:

Flask

Mikroframework postavený na programovacím jazyce Python, který umožňuje vytváření webových aplikací. Tento framework si zakládá na co nejjednodušším jádru aplikace, tak, aby bylo rozšiřitelné. Flask je založený na knihovnách Werkzeug a Jinja2 (šablonovací engine), díky kterým umožňuje snadný vývoj webových aplikací s minimálním množstvím kódu. Flask je vhodný pro webové aplikace, které nepotřebují robustní architekturu, vyznačuje se jednoduchostí, flexibilitou a škálovatelností. [38]

V rámci webové aplikace má Flask následující funkci:

1. Inicializace a spuštění webové aplikace: Aplikace je založena na instanci Flasku, která je vytvořena při importu této knihovny. Instance představuje jádro webové aplikace a je použita při definování různých funkcí a tras (endpointů).
2. Definování tras(endpointů): Flask umožňuje pomocí dekorátoru `@app.route()` definovat trasu(endpoint). Dekorátor je připojen ke konkrétní funkci a vytváří pro ni konkrétní URL. Když uživatel zadá konkrétní URL adresu do webového prohlížeče, Flask automaticky vyvolá přiřazenou funkci, která obsahuje logiku pro zpracování této konkrétní trasy.

3. Renderování šablon: Flask používá šablonovací engine Jinja2 pro generování HTML stránek. Konkrétně jsou šablony načítány z adresáře templates a renderovány pomocí funkce `render_template()`, do které jsou předány proměnné, které mají být vykresleny.
4. Zpracování HTTP požadavků: Další funkcí, kterou Flask framework ve webové aplikaci zprostředkovává je zpracování HTTP požadavků. Ve vytvořené webové aplikaci jsou konkrétně zpracovávány GET a POST požadavky.

Bokeh

Tato knihovna v programovacím jazyce Python slouží k vizualizaci dat. Jednou z hlavních funkcí, které tato knihovna poskytuje je tvorba interaktivních grafů, a také grafů, které cílí na moderní webové prohlížeče. [39]

Knihovna Bokeh se ve webové aplikaci používá pro následující účely:

1. Vytvoření teplotní mapy: Knihovna vykresluje geografické data ve formátu GeoJSON (konkrétně hranice jednotlivých okresů ČR) a vyznačuje teplotu jednotlivých okresů pomocí barevné škály.
2. Interaktivita mapy: Teplotní mapa nabízí interaktivní prvky, jako je například zobrazení informací o jednotlivém okrese při najetí kurzorem na daný okres (název okresu, teplota). Kromě toho mapa poskytuje interaktivní časovou osu, která umožňuje uživatelům prohlížet průběh teplot v různých časových okamžicích.
3. Zobrazení teplotní mapy na webovém prohlížeči: Knihovna obsahuje možnost zobrazení interaktivního grafu ve webovém prohlížeči, kde použitím CustomJS callbacků zajišťuje aktualizaci teplot na základě změn v posuvníku časové osy.

SQLAlchemy

SQLAlchemy je knihovna pro Python, která obsahuje sadu nástrojů zjednodušujících práci s relačními databázemi. Zajišťuje komunikaci Python aplikace s databází, přičemž komunikace probíhá prostřednictvím ORM (objektově-relačního mapování). [40] V rámci webové aplikace je knihovna použita pro následující účely:

1. Komunikace webové aplikace s databází: Pomocí této knihovny bylo zajištěno spojení aplikace s lokální PostgreSQL databází.
2. Získání geografických souřadnic a azimutů: V databázi jsou na základě IP adresy mikrovlnných zařízení hledány konkrétní zeměpisné souřadnice a jejich azimuty. Tímto způsobem jsou predikované teplotě určené přesné geografické souřadnice. Výpis SQL dotazu, který se stará o vyhledání jednotlivých metadat pro konkrétní IP adresy je následující:

```
SELECT latitude, longitude, azimuth FROM mw_list WHERE ip=:ip
```


InfluxDBClient

Jak název napovídá, knihovna InfluxDBClient se stará o komunikaci s databází InfluxDB. Po úspěšném připojení webové aplikace k InfluxDB serveru, je tato knihovna schopná z databáze číst, nebo zapisovat data. Pro zahájení komunikace je nutné předat několik parametrů – API klíč, URL adresa serveru. Ve webové aplikaci se tato knihovna používá pro získání potřebných parametrů pro predikci teploty okolí (Teplota, Kvalita signálu, IP adresa zařízení). [41]

Pandas a Geopandas

Pandas

Pandas je Python knihovna používaná pro práci s datovými sadami. Používá primárně dvě datové struktury – Series (1-dimenzionální) a DataFrame (2-dimenzionální). Knihovna obsahuje několik nástrojů pro analýzu a manipulaci s tabulkovými daty. Ve webové aplikaci se tato knihovna používá pro veškeré zpracování dat. [42]

Geopandas

Geopandas je rozšíření knihovny Pandas, které přidává do Pandas objektů podporu pro geografická data. Ve webové aplikaci slouží k načítání geografických dat České republiky. Dále se toto rozšíření používá pro nalezení příslušného okresu, ve kterých se nachází měřící zařízení na základě jejich geografických souřadnic. [43]

7.3 Architektura aplikace

Tato část se zabývá samotnou strukturou webové aplikace, a to jak na úrovni backendu, tak i frontendu. Struktura zahrnuje jednotlivé komponenty aplikace, funkce ale také způsoby, jakými jednotlivé komponenty mezi sebou komunikují. Architektura webové aplikace byla navržena tak, aby poskytovala koncovým uživatelům aktuální údaje o naměřené teplotě na území České republiky. Serverová část (backend) aplikace je zodpovědná za zpracování a přípravu dat pro zobrazení, komunikaci s databází. Zatímco klientská část (frontend) aplikace má na starosti vykreslení předaných dat z backendové části a poskytnout tak koncovému uživateli přehlednou interaktivní mapu. V této části si detailněji popíšeme logiku celé aplikace a jednotlivé části, které stojí za strukturou celou webové aplikace.

7.3.1 Backendová část

Backendová část webové aplikace je tvořena několika Python soubory, které jsou do hlavního souboru `main.py` volány. Rozdělení částí webové aplikace do několika souborů zvyšuje přehlednost a zajišťuje potencionálně jednodušší údržbu a aktualizaci aplikace. Mezi hlavní funkce a komponenty serverové části aplikace patří:

Komunikace s databází

Serverová část aplikace vytváří spojení s databázemi. Funkce `create_postgres_connection` vytváří spojení webové aplikace s PostgreSQL databází. Aplikace rovněž komunikuje s databází InfluxDB, tato komunikace je realizována prostřednictvím API. Zahájení komunikace pro obě databáze si zobrazíme v následujícím výpisu kódu:

Výpis 7.1: Navázání spojení s databázemi

```
#PostgreSQL
engine = create_postgres_connection(database="MW_list",
user=database_user, password=database_password,
host="localhost", port="5432")

#InfluxDB
client = InfluxDBClient(url=url, token=token)
```

Pro úspěšné zahájení komunikace je nutné funkcím předat potřebné parametry, u PostgreSQL databáze jsou předávány parametry – název, uživatelské jméno, heslo, hostitelská IP adresa, naslouchací port. Z důvodu zabezpečení nejsou přístupové údaje definovány přímo ve zdrojovém kódu. Při komunikaci s InfluxDB databází, která probíhá prostřednictvím API, jsou předávány tyto dva parametry – URL adresa, na které je server databáze spuštěn a API token.

Získání a zpracování vstupních dat

Aplikace z InfluxDB databáze získává naměřené parametry z mikrovlnných zařízení, konkrétně z databáze přebírá naměřené teploty a přijímanou úroveň signálu jednotlivých mikrovlnných zařízení, která jsou identifikována pomocí IP adresy. Tento proces zajišťuje externí Python soubor s názvem `data_extraction.py`, který obsahuje funkci `get_data10m()`. Výstupem této funkce je datový rámec obsahující parametry jednotlivých zařízení. Tyto parametry jsou následně zpracovány a připraveny tak pro predikci modelem strojového učení. Při zpracování datových rámců byl kladen důraz na zajištění stejné struktury dat, na které byl natrénován model strojového učení.

K jednotlivým záznamům ve zpracovaném datovém rámci jsou přidány geografické souřadnice z databáze PostgreSQL. Databáze na základě IP adresy v daném záznamu vyhledá v databázi příslušné souřadnice. Funkce `get_coordinates_and_azimuth()` zajišťuje tuto operaci. Dalším parametrem v datovém rámci je sloupec s názvem *Sun*, který označuje, zda má v daný moment na jednotku sluneční záření vliv. Tento parametr je určován pomocí knihovny *astral*, která pro daný den a oblast dokáže určit čas východu a západu Slunce. Záznamy v datovém rámci, které mají časové razítko v rozsahu mezi východem a západem Slunce dostávají označení 1, v opačném případě 0. Výpis funkce `is_daylight()`:

Výpis 7.2: Funkce pro určení vlivu slunečního záření

```
location = LocationInfo(latitude=lat, longitude=lng)
s = sun(location.observer, date=time.date())
sunrise = s['sunrise'].astimezone(pytz.utc)
sunset = s['sunset'].astimezone(pytz.utc)

if sunrise <= time <= sunset:
    return 1
else:
    return 0
```

Tímto způsobem bylo zajištěno, že vstupní data jsou kompatibilní s natrénovaným modelem a připravena pro následnou predikci. Výsledný datový rámec, označený proměnnou `df`, obsahuje tyto parametry – časové razítko i v Unixovém formátu, teplota teplotního čidla, přijímaná úroveň signálu, IP adresa, označení působení slunečního záření, azimut jednotky, zeměpisné souřadnice.

Predikce a příprava dat pro vizualizaci

Jednotlivé parametry mají v rámci aplikace různé účely. Geografické souřadnice slouží k vizualizaci predikovaných dat, IP adresa slouží pro identifikaci jednotky a časové razítko slouží k určení hodiny predikovaných teplot. Časové razítko v Unixovém formátu, teplota teplotního čidla, sluneční indikátor a přijímaná úroveň signálu jsou klíčové hodnoty, které jsou přímo vstupem do predikčního modelu. Webová aplikace k predikci teploty používá natrénovaný model strojového učení, který je pomocí knihovny *joblib*, do aplikace zavolán. Tento model slouží pro predikci teplot pro každý záznam v datech `X`, což je proměnná, která reprezentuje pouze datový rámec potřebných parametrů pro predikci. Predikovaná teplota je následně přidána do původního datového rámce:

Výpis 7.3: Predikce teploty na základě vstupních parametrů

```
X = df.drop(['Timestamp', 'Latitude', 'Longitude', 'IP'],
axis=1)
model = joblib.load("Linear_model_final1.joblib")
predicted_temperatures = model.predict(X)
df["Predicted_Temperature"] = predicted_temperatures
```

Pro vizualizaci predikovaných teplot je přidána funkce `find_region`, která na základě zeměpisných souřadnic v jednotlivých záznamech dokáže určit okres, ve kterém se zařízení nachází. Tato funkce pro každý záznam v datovém rámci označí ID jednotlivých okresů:

Výpis 7.4: Funkce pro přiřazení okresu jednotlivým záznamům

```
def find_region(gdf, latitude, longitude):
    point = Point(longitude, latitude)
    for index, row in gdf.iterrows():
        if row['geometry'].contains(point):
            return row['district']
    return None
df['district_id'] = df.apply(lambda row: find_region(gdf,
row['Latitude'], row['Longitude']), axis=1)
```

Pro sledování průběhu teploty byl vypočítán medián predikovaných teplot pro každé zařízení a jednotlivou hodinu. Pro každou hodinu obsahoval datový rámec pro každé zařízení 6 naměřených záznamů, ze kterých byl pro každou hodinu počítán medián predikovaných teplot. Datový rámec ukládající agregované teploty reprezentuje proměnná `hourly_median_df`, ze kterého je vytvořený finální datový rámec pro vizualizaci predikovaných teplot, označený `median_df`. Tento datový rámec je tvořen seskupením jednotlivých hodin a okresů, přičemž v okresech, ve kterých se nachází vícero záznamů, jsou jejich teploty průměrovány. Teploty jednotlivých okresů jsou následně předány do `GeoDataFrame` datového rámce, díky kterému je provedeno kompletní vykreslení okresové teplotní mapy.

Interpolovaný teplotní graf

Webová aplikace mimo jiné také vykresluje teplotní mapu interpolovaných teplot, tento proces má na starosti funkce `draw_plot()` ze souboru `int_graph.py`. Interpolovaná teplotní mapa je vykreslena na základě předané konkrétní hodiny jako parametr při volání této funkce.

Klíčovou částí tohoto procesu je interpolace predikovaných teplot, pro kterou je použita metoda vážené inverzní vzdálenosti (IDW). Tato metoda se používá k

odhadu teplot v bodech, pro které nejsou k dispozici přímá měření. Principem odhadování teplot je, že teplota bližších bodů má na odhadovanou teplotu větší vliv než teplota vzdálenějších bodů. Jako váhy jednotlivých teplot jsou v našem případě použity jednotlivé vzdálenosti bodů od interpolovaného místa.[44] Implementace interpolační metody byla zprostředkována pomocí souboru pro počítání interpolačních dat v prostoru, tento soubor byl získán z repozitáře pycomlink na platformě Github [45]. Původní zaměření tohoto repozitáře je interpolace srážkových dat.

Následující výpis kódu demonstruje, jak jsou teploty v rámci funkce `draw_plot()` interpolovány, aby bylo pokryté celé území ČR:

Výpis 7.5: Funkce pro přiřazení okresu jednotlivým záznamům

```
# Získání hranic České republiky
czech_bounds = czech_rep.bounds

# Vytvoření mřížky pro interpolaci podle hranic ČR
grid_x, grid_y = np.mgrid[czech_bounds.minx.min():
czech_bounds.maxx.max():500j, czech_bounds.miny.min():
czech_bounds.maxy.max():500j]

#Funkce pro vytvoření masky
def create_mask(czech_rep, grid_x, grid_y):
    mask = np.zeros_like(grid_x, dtype=bool)
    for i in range(grid_x.shape[0]):
        for j in range(grid_x.shape[1]):
            point = Point(grid_x[i, j], grid_y[i, j])
            mask[i, j] = czech_rep.contains(point).any()
    return mask

#Vytvoření masky a aplikace masky
mask = create_mask(czech_rep, grid_x, grid_y)
grid_x = grid_x[mask]
grid_y = grid_y[mask]

#Vytvoření instance interpolátoru
idw_interpolator = IdwKdtreeInterpolator(nnear=20, p=2,
exclude_nan=True, max_distance=1)

#Interpolace teplot na místech, kde je maska True
grid_z_masked = idw_interpolator(df['Longitude'].values,
df['Latitude'].values, df['Predicted_Temperature'].values,
xgrid=grid_x, ygrid=grid_y)
```

```
#Vytvoření výsledné mřížky 'grid_z' zobrazující teploty  
grid_z = np.full(mask.shape, np.nan)  
grid_z[mask] = grid_z_masked
```

Tento výpis zahrnuje celý proces interpolace teplot a následné upravení interpolovaných dat, aby byla vykreslována pouze na území označeném hranicemi ČR. Interpolace dat je provedena pomocí funkce `IdwKdtreeInterpolator()`, která obsahuje několik parametrů, které nastavují vlastnosti interpolace. Parametry `nnear=20` a `p=2` definují, kolik bude interpolace brát v úvahu nejbližších okolních bodů a jak moc bude váha těchto bodů se vzdáleností klesat. Parametr `max_distance` určuje maximální vzdálenost bodů, které budou zahrnuty do interpolace.

7.3.2 Frontendová část

Frontendová část webové aplikace je zodpovědná za interakci s uživatelem prostřednictvím webového prohlížeče. Webová aplikace poskytuje uživateli interaktivní mapu s posuvnou časovou osou, díky které lze sledovat průběh jednotlivých okresů za posledních 24 hodin. Interpolovaná teplotní mapa je vykreslena při výběru konkrétní hodiny z nabídky, pomocí které dojde k vykreslení mapy interpolovaných teplot. Frontendová část webové aplikace má tyto funkce:

Interaktivní mapa

Pomocí knihovny Bokeh je vykreslena teplotní mapa, která umožňuje vizualizaci předpřipravených dat v GeoJSON formátu. Tato data v sobě zahrnují také predikované teploty. Pomocí této knihovny byly do teplotní mapy přidány některé widgety a nástroje, které slouží k interaktivitě mapy. Widget Slider, slouží k přidání časové osy, Hovertools je nástroj, který zobrazuje informace o daném okresu (název, teplota) při převedení kurzoru na daný okres. Posledním interaktivním prvkem mapy je tlačítko Button s názvem Start/Stop, které při spuštění zobrazuje animaci teplotního průběhu. Pro zajištění funkčnosti těchto prvků je každý z nich napojen na vlastní JavaScriptový kód, který jim dodá potřebnou interaktivitu. Tento proces je realizovaný pomocí CustomJS v knihovně Bokeh, pomocí něhož je možné prvkům přidat tzn. callback funkce. Callback funkce umožňují vytvářet interaktivní vizualizace, které reagují na akce uživatele. Výpis callback funkce, která byla implementována pro časovou osu:

Výpis 7.6: Callback funkce pro funkcionalitu časové osy

```
var date = new Date(cb_obj.value);
```

```

var hour = date.getUTCHours();
date.setUTCMinutes(0);
date.setUTCSeconds(0);
date.setUTCMilliseconds(0);
var median_df = median_df_cds.data;
var gdf_data = gdf_cds.data;

// Aktualizace teplot v gdf_data podle median_df
for (var i = 0; i < median_df['Hour'].length; i++) {
    var row_date = new Date(median_df['Hour'][i]);

    if (row_date.getUTCHours() === hour &&
        row_date.getUTCDate() === date.getUTCDate() &&
        row_date.getUTCMonth() === date.getUTCMonth() &&
        row_date.getUTCFullYear() === date.getUTCFullYear
        ()) {

        var district = median_df['district_id'][i];
        var temperature = median_df['
            Predicted_Temperature'][i];

        for (var j = 0; j < gdf_data['xs'].length; j++) {
            if (gdf_data['district'][j] == district) {
                gdf_cds.data['temperature'][j] =
                    temperature;
            }
        }
    }
}

gdf_cds.change.emit();

// Formátování času
var formatted_date = date.toISOString().slice(0, 10) + "
    " + ("0" + date.getUTCHours()).slice(-2) + ":00";
plot.title.text = "Teplotní mapa ČR, Datum a čas: " +
    formatted_date;

```

Tato část kódu aktualizuje teploty na mapě na základě hodnoty časové osy, která určuje aktuální čas. Hodnota časové osy se převede na časový objekt `Date` a poté jsou procházena teplotní data, aby došlo k nalezení teploty k odpovídajícímu času.

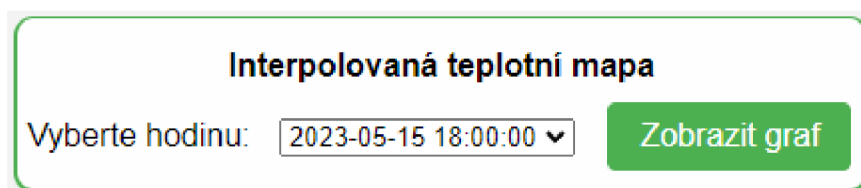
7.4 Definování tras a komunikace s backendovou částí

Pomocí knihovny Flask jsou definovány trasy(endpointy) a také zajištěna komunikace mezi frontendovou a backendovou částí aplikace. Aplikace obsahuje dvě hlavní routy:

- `index`: slouží jako domovská stránka, která zobrazuje interaktivní mapu a také zobrazuje výběrový seznam jednotlivých hodin, pomocí kterých je vykreslena interpolovaná teplotní mapa
- `interpolation_graph`: tato routa zobrazuje interpolovanou teplotní mapu na základě vybrané hodiny

Komunikace probíhá pomocí HTTP POST dotazů, pomocí kterých je zasílána do backendové části webové aplikace uživatelem vybraná hodina. Pro tuto konkrétní hodinu je následně na endpointu `interpolation_graph` vykreslena interpolovaná teplotní mapa. Výběrový seznam, který slouží pro zasílání dotazů pro vykreslení interpolované mapy:

Obr. 7.1: Výběrový seznam konkrétních hodin



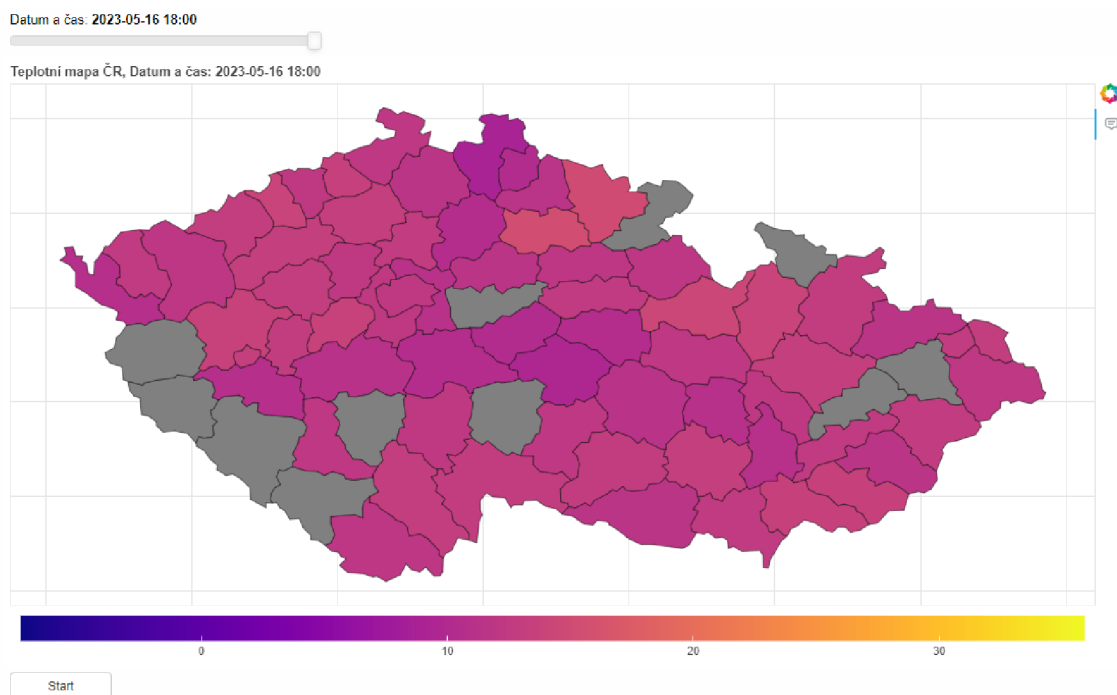
Interpolovaná teplotní mapa

Vyberte hodinu:

Po stisknutí tlačítka `Zobrazit graf` je zasílán HTTP POST dotaz s konkrétní vybranou hodinou a následně je pro tuto hodinu vykreslena interpolovaná teplotní mapa.

Interaktivní mapa vykreslovaná frontendovou částí aplikace vypadá následovně:

Obr. 7.2: Interaktivní teplotní mapa



Okresy, které jsou označeny šedou barvou nemají v GeoJSON souboru přiřazeny žádné teploty. Na území tohoto okresu se totiž nevyskytuje žádné zařízení, pomocí kterého by byla teplota predikována.

7.4.1 Zabezpečení citlivých informací

Při vytváření webové aplikace byl také kladen důraz na zabezpečení citlivých informací, konkrétně geografických souřadnic jednotlivých zařízení. Tyto citlivé informace jsou uloženy v PostgreSQL databázi, se kterou webová aplikace komunikuje. Základem každého zabezpečení databáze je pravidelné aktualizování na nejnovější verze, které mimo jiné také řeší naskytnuté bezpečnostní chyby a zranitelnosti, kterých by mohl potenciální útočník využít pro získání dat nelegitimním způsobem.

Prvním, základním zabezpečením databáze je nastavení přístupových údajů. Tím bylo zamezeno neoprávněnému přístupu do databáze, podmínkou je nastavení dostatečně silného hesla. Během iniciace komunikace s databází ve webové aplikaci je rovněž nezbytné tyto přístupové údaje zadat přímo ve zdrojovém kódu. Tento bezpečnostní problém je vyřešen vytvořením proměnného prostředí na serveru, na kterém je aplikace spuštěna. Proměnné prostředí tyto přístupové údaje ukládá v registrech OS a následně je s použitím vhodných knihoven a názvu proměnného

prostředí možné tyto údaje zavolat do zdrojového kódu. Důležitým aspektem přístupového účtu jsou také jeho jednotlivé oprávnění, které určují, jaké akce může účet v rámci aplikace provádět. Vytvořený účet pro webovou aplikaci potřebuje pouze oprávnění pro čtení z databáze, toho bylo docíleno tímto příkazem:

```
GRANT SELECT ON ALL TABLES IN SCHEMA public TO nazev_uctu;
```

Select je v SQL terminologii typ oprávnění, který umožní přidělenému účtu z databáze pouze číst.

Co se týče konkrétního zabezpečení webové aplikace před únikem geografických pozic. Ve webové aplikaci je zajištěno, že citlivá data, tedy geografické souřadnice jednotlivých zařízení, nejsou zasílána na klientovu stranu. Klientovi jsou pouze zasílány agregované teplotní hodnoty pro jednotlivé okresy a všechny operace s geografickými souřadnicemi probíhají pouze na straně serveru v backendové části aplikace.

Získání citlivých dat z databáze probíhá prostřednictvím SQL Alchemy, které používá parametrizované SQL dotazy, tím je zamezeno provedení SQL dotazů na databázi, který nebyl původně zamýšlen, takový typ útoku je nazýván SQL Injection.[46] Konkrétně se SQL dotaz volá ve funkci `get_coordinates_and_azimuth`.

Dalším možným zabezpečovacím prvkem je zabezpečení komunikace webové aplikace s databází ukládající citlivé informace. Takové zabezpečení se provádí pomocí SSL/TSL protokolů, které komunikaci šifrují. V našem konkrétním případě však není nutná implementace těchto protokolů. Důvodem je, že webová aplikace a PostgreSQL databáze jsou obě hostovány lokálně, což znamená, že komunikace neprobíhá přes síťové rozhraní a tím pádem není náchylná k útokům typu man in the middle. Implementace SSL/TSL protokolů by představovala zbytečné náklady na výkon.

Nicméně, v kontextu reálného nasazení, kde by komunikace mezi serverem a databází probíhala přes síťové rozhraní, by bylo nutné tuto komunikaci zabezpečit. Neimplementace těchto protokolů v reálném nasazení by představovala závažné bezpečnostní riziko.

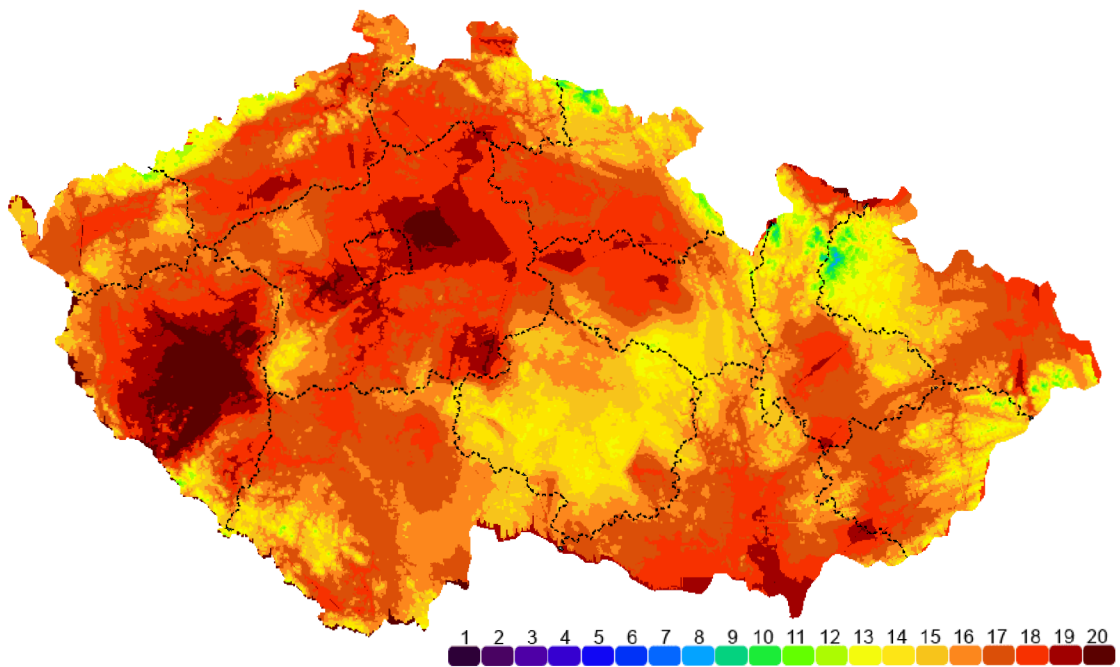
7.5 Testování predikovaných teplot

V následující části se zaměříme na testování predikovaných výsledků modelem strojového učení. Výsledky porovnáme s referenčními hodnotami získanými na serveru `in-pocasi.cz`, tím bude ověřena přesnost predikovaného modelu pro jednotlivé oblasti na mapě. Pro zajištění co největší přesnosti testování je důležité se zaměřit na široké spektrum klimatických podmínek.

7.5.1 Ověření přesnosti predikce

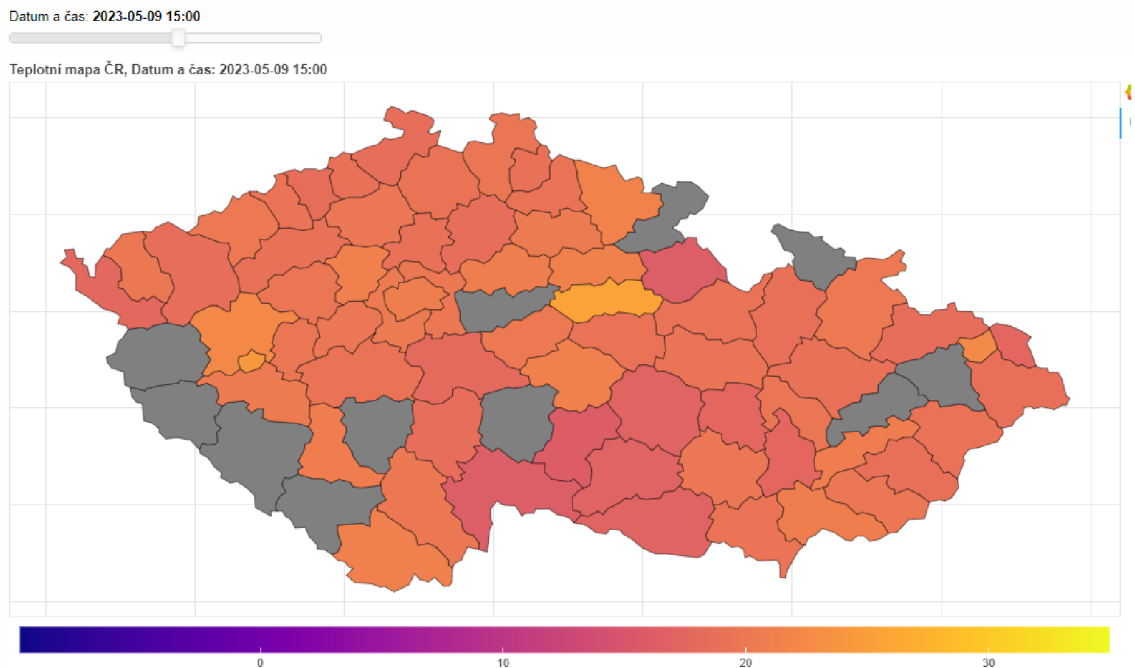
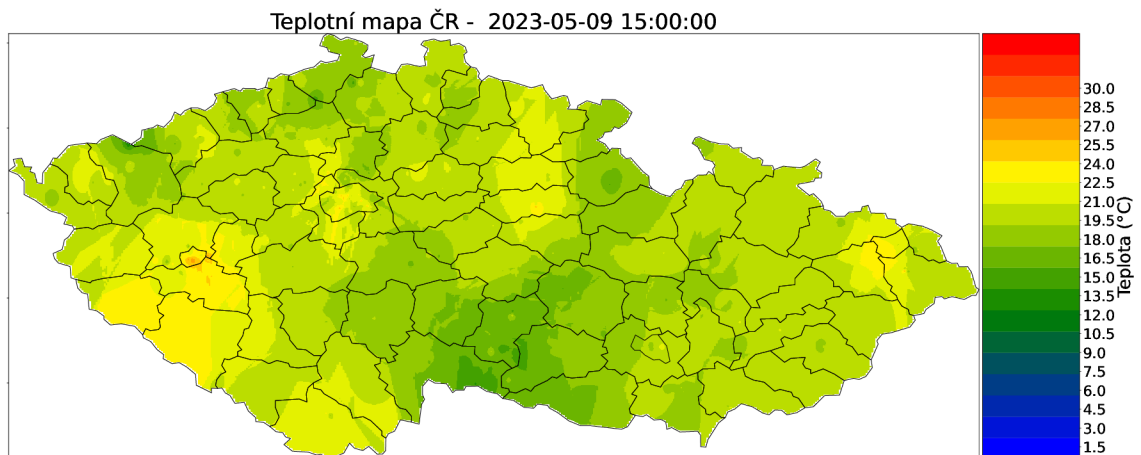
Jasno/polojasno

Pro ověření přesnosti predikce za jasného/polojasného počasí bylo vybráno datum 9.5, kdy je na celém území ČR jasno, místy polojasno. Pro toto počasí je charakteristické působení slunečního záření, které jednotku v průběhu dne nahřívá. Referenční teplotní mapa pro tento den, konkrétně pro čas 15:00:



Obr. 7.3: Referenční teplotní mapa - 9.5.2023 15:00

Predikované výsledky natrénovaným modelem pro tento den:



Obr. 7.4: Interpolovaná a okresová teplotní mapa

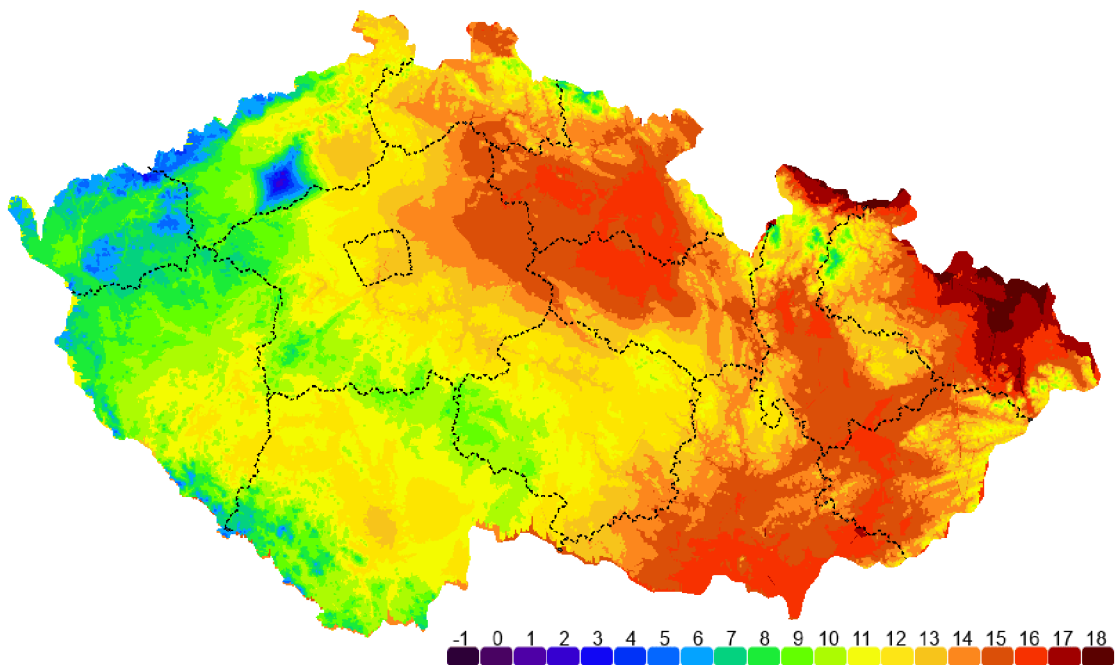
Při porovnání obou teplotních map můžeme vidět, že predikované teploty jsou velmi podobné referenčním. Model strojového učení dosahuje poměrně přesných výsledků i při intenzivnějším působení slunečního záření. Při detailním prozkoumání referenční teplotní mapy je vidět, že oblast Vysočina se vyznačuje mírně sníženými teplotami, což je odlišnost, kterou úspěšně zachycují i vizualizace generované modelem strojového učení.

I přesto, že jsou teplotní predikce obecně poměrně přesné, můžeme si všimnout

také chybně predikovaných hodnot, například okresu Pardubice bylo predikováno 26 °C, to může být zapříčiněno několika důvody – jednotka v tomto okrese může být umístěna atypicky, může být ovlivněna jiným teplotním zdrojem, může být v technické závadě nebo se může jednat o chybu v modelu, který tuto predikci provádí.

Proměnlivé počasí

Abychom lépe ověřili přesnost modelu, byl zvolen den s proměnlivým počasím v různých oblastech České republiky. Konkrétně se jedná o 28. dubna, kdy na západě republiky převládalo zatažené počasí a občasný déšť, zatímco na východě bylo oblačno, místy polojasno, což umožnilo pronikání mírného slunečního záření na mikrovlnné jednotky. Následuje referenční teplotní mapa:



Obr. 7.5: Referenční teplotní mapa

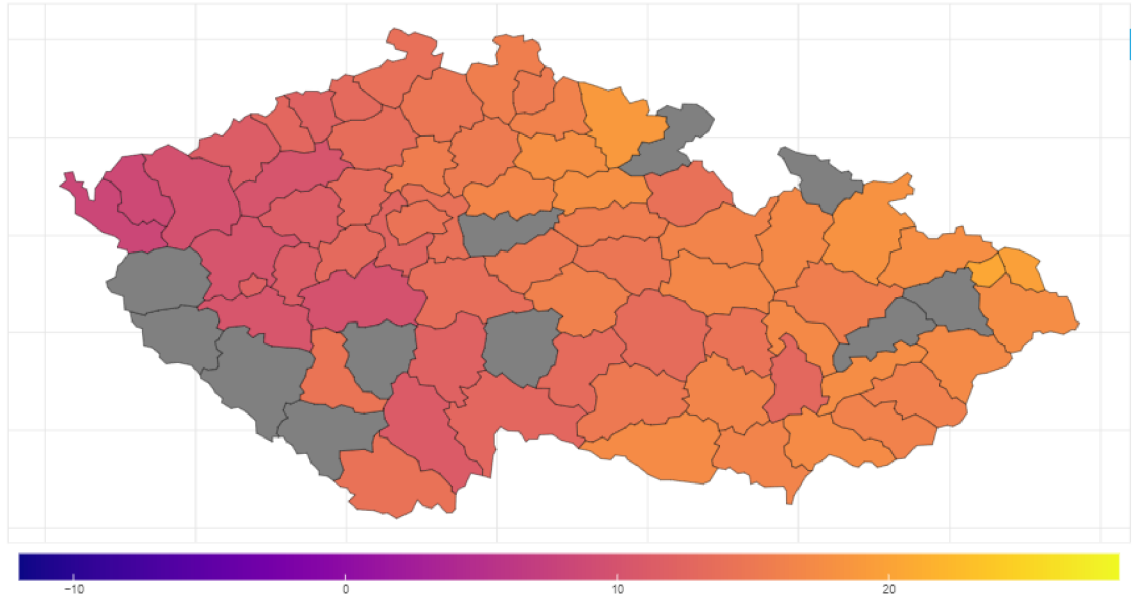
Referenční teplotní mapa

Predikované teplotní mapy pro tento den:

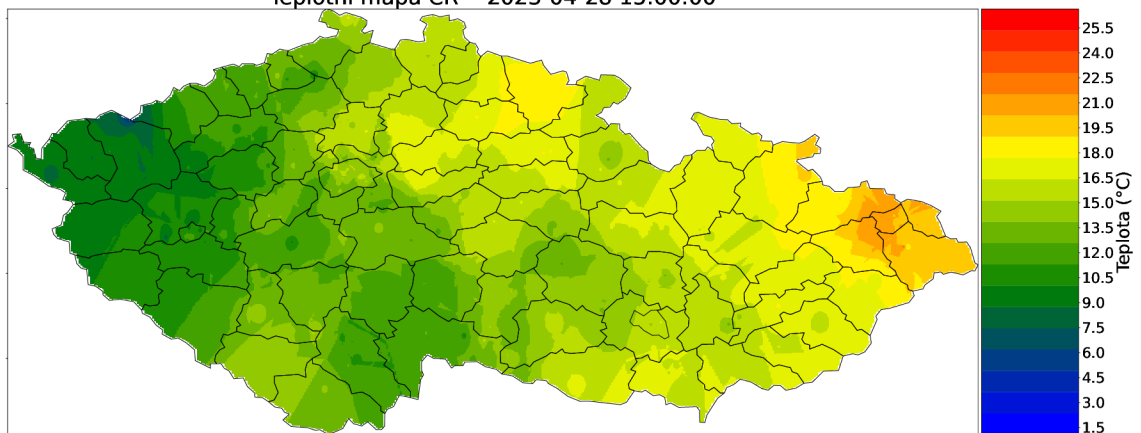
Datum a čas: 2023-04-28 15:00



Teplotní mapa ČR, Datum a čas: 2023-04-28 15:00



Teplotní mapa ČR - 2023-04-28 15:00:00



Obr. 7.6: Interpolovaná a okresová teplotní mapa

V predikovaných teplotních mapách pro dané počasí lze pozorovat poměrně vysokou přesnost. Při porovnání teplotních map, je krásně vidět, že největších teplot dosahují východní oblasti, přičemž nejvyšší oblasti se vyskytují v Moravskoslezském kraji. Na západním území hodnoty dosahují poměrně nižších hodnot, a to z důvodu zataženého počasí. Nejnižší teploty jsou predikovány pro Karlovarský kraj. Zajímavým zjištěním je také fakt, že přechod teplot je interpolován téměř identicky napříč územím, s tím, že severnější oblasti vyšších teplot se rozprostírají dále na západ než právě jižní oblasti.

Na interpolované teplotní mapě je také možné identifikovat některé rozdíly v predikovaných teplotách. Například v oblasti Jeseníků si můžeme všimnout, že teplota by měla být razantně nižší, to může být zapříčiněno například tím, že model strojového učení byl natrénován pouze na dostupných datech z urbanizovaných oblastí. Tyto oblasti jsou charakteristické jiným klimatickým chováním, než je tomu v oblastech pohoří. Kromě toho by mohla být variabilita teplot mezi různými oblastmi ovlivněna také nadmořskými výškami těchto lokalit.

7.6 Kvantifikace teplotní přírůstku

S využitím natrénovaného modelu strojového učení a jeho předpovědí teplot lze také následně kvantifikovat přibližný teplotní přírůstek vyprodukovaný slunečním zářením. Kvantifikace probíhá tím způsobem, že kontroluje záznamy ve vytvořeném datovém rámci, který obsahuje veškeré hodnoty včetně parametrů teploty zařízení, predikované teploty a označení zdali pro daný moment působí sluneční záření, které budou použity pro výpočet teplotního přírůstku.

Pro záznamy, které jsou označeny působením slunečního záření (hodnota ve sloupci *Sun* je rovna 1) je počítán teplotní přírůstek tím způsobem, že od naměřené teploty mikrovlnné jednotky je odečtena predikovaná teplota okolí a také ještě průměrná teplotní konstanta, která označuje teplotu, kterou v daném záření vyprodukuje fyzikální jev zvaný Jouleovo teplo. Byla využita průměrná konstanta pro všechny zařízení, která byla vypočítána pro danou technologii O na základě velkého vzorku sledovaných zařízení, konstanta reprezentuje teplotu 24°C. Jelikož se pro teplo vyprodukované samotnou jednotkou může lehce lišit pro dané zařízení, jedná se o přibližný teplotní přírůstek. Vzorec pro výpočet kvantifikace teplotního přírůstku tedy vypadá následovně:

$$T_P = T_{MW} - T_O - c \quad (7.1)$$

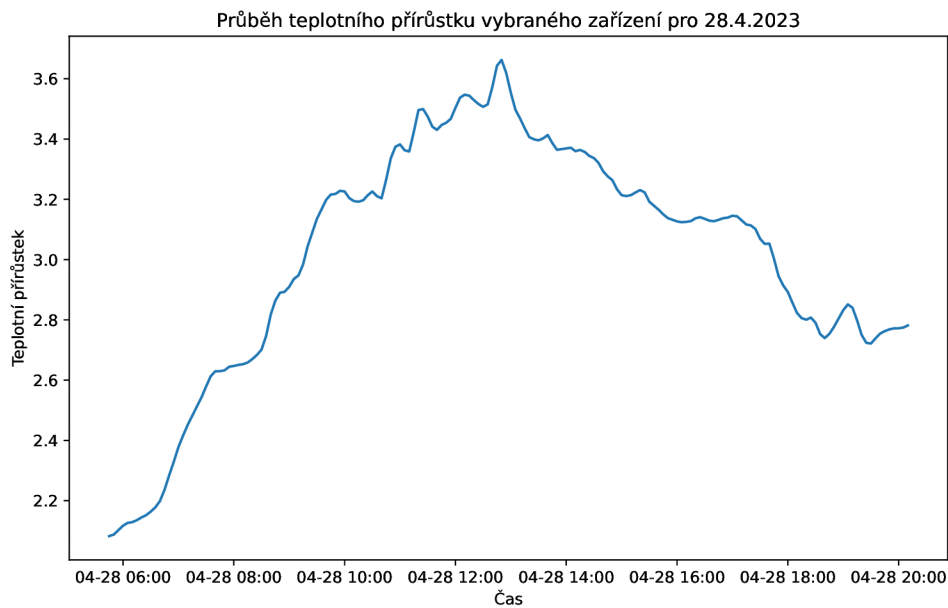
Kde:

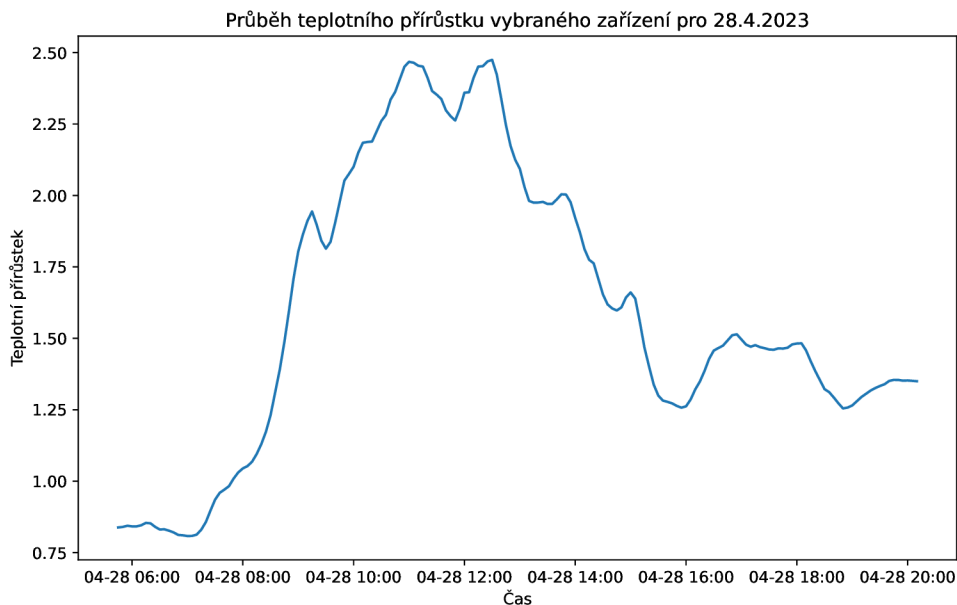
- T_P označuje teplotní přírůstek vyprodukovaný slunečním zářením

- T_{MW} značí teplotu mikrovlnného zařízení
- T_O reprezentuje predikovanou teplotu
- c je průměrná konstanta teplot vyprodukovaných Jouleovým teplem

7.6.1 Výsledky kvantifikace teplotní přírůstku

Nyní si zobrazíme teplotní přírůstky vyprodukované v průběhu dne pro vybrané zařízení. Výsledky budou zobrazeny v grafu, který bude zobrazovat průběh teplotního přírůstku za jeden den. Pro zobrazení vlivu oblačnosti na kvantifikaci teplotního přírůstku byl vybrán den 28.4, kdy se východě vyskytuje působení většího množství slunečního záření z důvodu oblačnosti na západní části republiky. Teplotní přírůstky budou zobrazeny pro dvě zařízení, první zařízení se nachází v okolí Brna, druhé pak v okolí Prahy. Grafy teplotních přírůstků:





Obr. 7.7: Průběhy teplotních přírůstků vyprodukované slunečním zářením

Můžeme si všimnout, že první graf, který zobrazuje teplotní přírůstek pro Brněnské zařízení dosahuje větších hodnot, než je tomu tak pro druhé zařízení v Praze. To je zapříčiněno větší oblačností, která se v Praze pro daný den vyskytuje. Průběh teplotního přírůstku je také ovlivněn azimutem konkrétní jednotky.

Existuje řada možných důvodů, proč teplotní přírůstky nezačínají od nuly. Jedním z nich je fakt, že teplota vygenerovaná samotnou jednotkou byla určena jako konstantní, ale přesto se pro jednotlivé mohou vyskytovat odchylky od této konstanty. Dalším možným důvodem může být, že mikrovlonné zařízení se obvykle instalují na střechy panelových domů, které mohou dosahovat výšek i přes 30 metrů. Teplotní přírůstek byl nicméně počítán pouze pro ty záznamy v datovém rámci, které byly označeny ve sloupci *Sun* hodnotou 1. Tato hodnota indikuje, že se daný záznam nachází v časovém intervalu mezi východem a západem Slunce. Výpočet východu a západu Slunce je mimo jiné ovlivněn konkrétní nadmořskou výškou daného místa. V našem konkrétním případě byl východ a západ Slunce počítán pro obecnou nadmořskou výšku místa na mapě a to pro střed České republiky. Zjednodušeně lze konstatovat, že pokud je zařízení umístěno na střechě panelového domu, východ Slunce nastane o několik minut dříve a západ Slunce bude o několik minut posunut. To znamená, že sluneční záření na jednotku bude mít vliv dříve, než bude ve sloupci *Sun* tento vliv označen. [47]

Závěr

Práce se zabývala technologií mikrovlnných spojů a vlivy, které se podílí na naměřené hodnotě mikrovlnné jednotky. Práce byla zejména zaměřena na vliv slunečního záření na naměřenou hodnotu teplotními senzory v mikrovlnných jednotkách. Pro zkoumání tohoto vlivu bylo nutné také zanalyzovat působení jiných vlivů, které naměřenou teplotu ovlivňují.

V databázi ukládající naměřené hodnoty byl vytvořen dashboard, který obsahuje několik zajímavých nástrojů, které slouží k analýze a díky kterým je možné sledovat a kvantifikovat průběhy různých vlivů, které naměřenou hodnotu ovlivňují.

Na základě provedené korelační analýzy, která se zaměřila na prozkoumání vztahů mezi teplotou telekomunikačních zařízení různých technologií a teplotou zaznamenanou okolními meteorologickými stanicemi, byla zvolena nejvhodnější technologie pro predikci teploty okolí. Tato predikce je založena na natrénovaném modelu strojového učení, konkrétně Lineární regresi, který na základě naměřených hodnot jednotlivých zařízení předpovídá teplotu okolí.

S využitím natrénovaného modelu byla vytvořena webová aplikace zobrazující interaktivní teplotní mapu České republiky, která vykresluje teploty jednotlivých okresů na území ČR. Součástí interaktivní teplotní mapy je posuvná časová osa, pomocí které je možné zobrazit průběh teplot za posledních 24 hodin. Další funkcí webové aplikace je možnost vykreslení grafu interpolované teplotní mapy pro konkrétní vybranou hodinu. Interpolovaná teplotní mapa poskytuje detailnější přehled o teplotě napříč územím České republiky a pomocí zvolené interpolační metody bylo možné předpovědět teploty v oblastech, ve kterých se nenachází žádné telekomunikační zařízení. Vstupem do webové aplikace jsou hodnoty jednotlivých zařízení naměřené v reálném čase, které ukládá databáze InfluxDB a následně je pomocí API předává webové aplikaci. Aplikace komunikuje také s lokálně vytvořenou PostgreSQL databází, ze které získává geografické souřadnice a orientaci konkrétních zařízení.

Při testování predikovaných teplot byla prokázána vysoká úroveň přesnosti předpovědí modelem strojového učení. Testování bylo provedeno na širokém spektru klimatických podmínek, které se v některé dny vyskytovaly. Model prokázal svou schopnost úspěšně predikovat teploty jak za jasného, tak i proměnlivého počasí. Detailnější analýza modelu odhalila některé specifické oblasti, kde by se přesnost modelu dala potenciálně zlepšit, to je zapříčiněno z velké části dostupnými trénovacími daty.

Webová aplikace byla následně zabezpečena proti úniku konkrétních geografických souřadnic. Hlavním prvkem zabezpečení bylo zajištění, aby se na klientskou stranu nezasílaly konkrétní geografické souřadnice jednotlivých zařízení. Dále byl

zdrojový kód aplikace zabezpečen proti různým typům bezpečnostních hrozeb, jako je například SQL Injection a navrhuje další možné zabezpečení pomocí šifrování komunikace mezi webovou aplikací a PostgreSQL databází proti známým útokům Man-in-the-middle s využitím protokolu SSL.

Pomocí predikovaných teplot byla provedena přibližná kvantifikace teplotního přírůstku vyprodukovaného slunečním zářením. Kvantifikace tohoto vlivu představuje potenciální využití v indikaci teplotních hrozeb, které by mohly ovlivnit výkon telekomunikačního zařízení.

Webová aplikace představuje obrovský potenciál v rozšíření stávajících sítí meteorologických měřících bodů. Existuje zde však stále prostor pro další vývoj. Zvláště se jedná o využití slibných technologií rodiny S , které dosahovali vysokých korelačních koeficientů. Ačkoliv pro tyto technologie není k dispozici dostatek historických dat, data jsou již nyní zaznamenávána a ukládána pro jejich budoucí využití. Dalším možným aspektem, který by mohl vést ke zpřesnění predikovaných teplot, by mohlo být zajištění dalších parametrů, například konkrétní nadmořské výšky, ve kterých se telekomunikační zařízení nachází.

Literatura

- [1] Květoslava Miloslava Matějková. Principy digitální mikrovlnné komunikace. Technical report, ALCOMA, October 2019.
- [2] Miroslav KASAL. Směrové a družicové spoje. *Skriptum*. Brno: FEKT VUT v Brně, 2003.
- [3] Miroslav Pavelec. Vliv vnějšího prostředí na parametry mikrovlnných antén. 2015.
- [4] Miroslav CENEK. Obnovitelné zdroje energie. 2. upr. a dopl. vyd. Praha: Fcc public, 2001, 208 s.
- [5] Lenka DOSOUDILOVÁ. *Simulace slunečního záření*. PhD thesis, Masarykova univerzita, Přírodovědecká fakulta, 2009.
- [6] Calculation of solar insolation. URL: <https://www.pveducation.org/pvc/drom/properties-of-sunlight/calculation-of-solar-insolation>.
- [7] Jouleovo teplo, Feb 2022. URL: https://cs.wikipedia.org/wiki/Jouleovo_teplo.
- [8] Blanka Vítková. *Měření slunečního záření*. PhD thesis, Masarykova univerzita, Pedagogická fakulta, 2012.
- [9] Pearson correlation coefficient, Nov 2022. URL: https://en.wikipedia.org/wiki/Pearson_correlation_coefficient.
- [10] Tomáš Pavlík et al. Biostatistika. 2012.
- [11] Spearman's rank correlation coefficient, Dec 2022. URL: https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient.
- [12] A refresher on statistics for data scientists, May 2023. URL: <https://www.phdata.io/blog/data-science-stats-review/>.
- [13] Kendall rank correlation coefficient, May 2023. URL: https://en.wikipedia.org/wiki/Kendall_rank_correlation_coefficient.
- [14] Kendall rank correlation explained, May 2023. URL: <https://towardsdatascience.com/kendall-rank-correlation-explained-dee01d99c535>.
- [15] Jan Hendl. Korelace, 2020. Sken z knihy. URL: https://is.muni.cz/el/fss/podzim2020/POL593/um/Hendl_-_Korelace.pdf?kod=BZ402Zk;lang=en.

- [16] Normální rozdělení, March 2023. URL: https://www.wikiskripta.eu/w/Norm%C3%A1ln%C3%AD_rozd%C4%9Blen%C3%AD.
- [17] Statistické pojmy, May 2023. URL: <https://www.scio.cz/o-vzdelavani/teorie-a-metodika-testu/statisticke-pojmy/>.
- [18] Normal distribution, March 2023. URL: <https://www.investopedia.com/terms/n/normaldistribution.asp>.
- [19] Normální (gaussovo) rozdělení, May 2023. URL: <https://cit.vfu.cz/statpotr/POTR/Teorie/Predn2/rozdelZS.htm>.
- [20] Shape of data: Skewness and kurtosis, March 2023. URL: <https://www.analyticsvidhya.com/blog/2020/07/what-is-skewness-statistics/>.
- [21] Shape of data: Skewness and kurtosis, March 2023. URL: <https://www.analyticsvidhya.com/blog/2021/05/shape-of-data-skewness-and-kurtosis>.
- [22] Kurtosis, March 2023. URL: <https://quantra.quantinsti.com/glossary/Kurtosis>.
- [23] Platykurtic, March 2023. URL: <https://www.investopedia.com/terms/p/platykurtic.asp>.
- [24] Světové strany a určování azimutu. URL: <https://clanky.rvp.cz/clanek/r/ZFADA/2628/SVETOVE-STRANY-A-URCOVANI-AZIMUTU.html>.
- [25] Testy normality, March 2023. URL: https://www.wikiskripta.eu/w/Testy_normality.
- [26] Strojové učení: Metody klasifikace, May 2023. URL: <https://www.rascasone.com/cs/blog/strojove-uceni-ml-metody-klasifikace>.
- [27] Astral documentation, May 2023. URL: <https://astral.readthedocs.io/en/latest/>.
- [28] Regression analysis in machine learning, May 2023. URL: <https://www.javatpoint.com/regression-analysis-in-machine-learning>.
- [29] Ashan Lakmal Thilakarathne. Implementing a machine learning model for weather prediction. 2023. Accessed: 2023-05-22. URL: <https://medium.com/swlh/creating-a-model-for-weather-forecasting-using-linear-regression-b18c1590e8d>

- [30] Metrics to evaluate your machine learning algorithm, May 2023. URL: <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>.
- [31] Performance metrics in machine learning: A complete guide, May 2023. URL: <https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide>.
- [32] OpenGenus. Advantages and disadvantages of linear regression, 2023. [Online; accessed 22-May-2023]. URL: <https://iq.opengenus.org/advantages-and-disadvantages-of-linear-regression/>.
- [33] Cross-validation in machine learning, May 2023. URL: <https://towardsdatascience.com/cross-validation-in-machine-learning-72924a69872f>.
- [34] Residual plot analysis, May 2023. URL: <https://www.originlab.com/doc/origin-help/residual-plot-analysis>.
- [35] Front-end vs back-end: Jaký je mezi nimi rozdíl?, May 2023. URL: <https://www.apitree.cz/blog/front-end-vs-back-end-jaky-je-mezi-nimi-rozdil>.
- [36] Influxdb: Purpose-built open source time series database, May 2023. URL: <https://www.influxdata.com/products/influxdb/>.
- [37] About postgresql, May 2023. URL: <https://www.postgresql.org/about/>.
- [38] Úvod do frameworku flask a webových aplikací v pythonu, May 2023. URL: <https://www.itnetwork.cz/python/flask/uvod-do-frameworku-flask-a-webovych-aplikaci-v-pythonu>.
- [39] Guide for data visualization with bokeh python library, May 2023. URL: <https://www.analyticsvidhya.com/blog/2021/09/guide-for-data-visualization-with-bokeh-python-library/>.
- [40] Sqlalchemy orm tutorial for python developers, May 2023. URL: <https://auth0.com/blog/sqlalchemy-orm-tutorial-for-python-developers/>.
- [41] Influxdb python client, May 2023. URL: <https://github.com/influxdata/influxdb-client-python>.
- [42] Overview — pandas 1.3.4 documentation, May 2023. URL: https://pandas.pydata.org/docs/getting_started/overview.html.

- [43] About geopandas, May 2023. URL: <https://geopandas.org/en/stable/about.html>.
- [44] Prostorová analýza: Interpolace - qgis, May 2023. URL: https://docs.qgis.org/2.18/cs/docs/gentle_gis_introduction/spatial_analysis_interpolation.html.
- [45] Pycomlink: Python for communication link analysis, May 2023. URL: <https://github.com/pycomlink/pycomlink>.
- [46] Sql injection, May 2023. URL: <https://portswigger.net/web-security/sql-injection>.
- [47] Tom Skilling. The article title, 2023. Accessed: 2023-05-20. URL: <https://www.chicagotribune.com/weather/ct-wea-0928-asktom-20160927-column.html>.

Seznam symbolů a zkratek

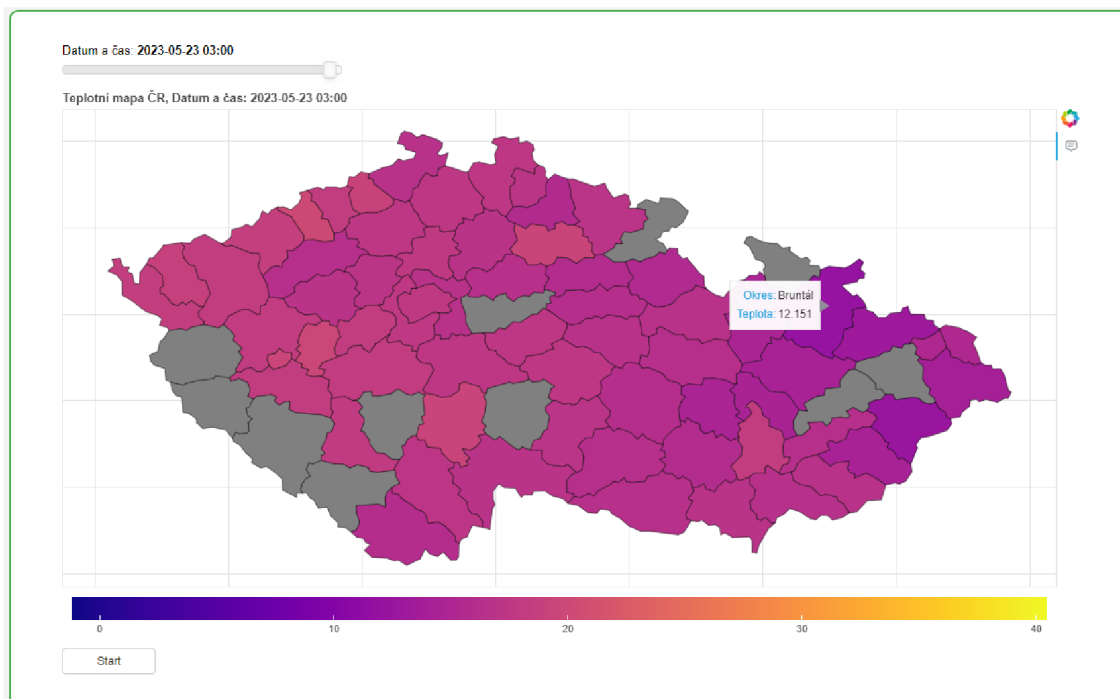
API	Application Programming Interface
ČHMÚ	Český hydrometeorologický ústav
ČR	Česká republika
GHz	Gigahertz
HTTP	Hypertext Transfer Protocol
I	Elektrický proud
ID	Identifikátor
IDU	Vnitřní jednotka
IDW	Vážené inverzní vzdálenosti
IF	Mezifrekvenční
MAE	Mean Absolute Error
m	Metr
mm	Milimetr
MSE	Mean Squared Error
ODU	Vnější jednotka
ORM	Objektově-relační mapování
OS	Operační systém
Q	Jouleovo teplo
R	Odpor
R²	R-Squared
RF	Vysokofrekvenční
RMSE	Root Mean Squared Error
ρ	Pearsonův korelační koeficient
°C	Stupeň Celsia

SQL	Structured Query Language
SSL	Security Socket Layer
r_s	Spearmanův korelační koeficient
SVR	Support Vector Regression
σ	Směrodatná odchylka
t	Čas
τ	Kendallův korelační koeficient
TEM	Transverzálně elektromagnetické vlny
TLS	Transport Layer Security
U	Napětí
URL	Uniform Resource Locator
μ	Střední hodnota
<i>cov</i>	Kovariance

Seznam příloh

A Ukázka webové aplikace	91
B Obsah elektronické přílohy	93

A Ukázka webové aplikace

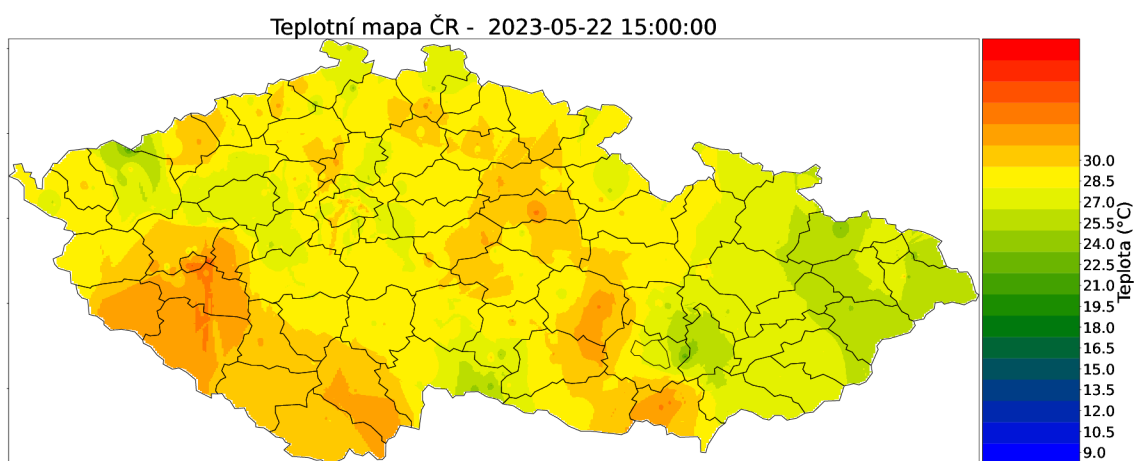


Obr. A.1: Interaktivní teplotní mapa okresů

Interpolovaná teplotní mapa

Vyberte hodinu:

Obr. A.2: Výběrový seznam pro vykreslení interpolované teplotní mapy



Obr. A.3: Interpolovaná teplotní mapa

B Obsah elektronické přílohy

```
/.....kořenový adresář přiloženého archivu
├── country_data adresář s JSON daty pro vykreslení teplotní mapy České republiky
│   └── czech_republic_.json
├── templates.....adresář obsahující HTML šablony
│   ├── index.html
│   └── interpolation_graph.html
├── data_extraction.py.....Python soubor pro získání vyčítání dat z InfluxDB
├── main.py.....Python soubor pro spuštění webové aplikace s hlavními funkcemi
├── int_graph.py.....Python soubor pro vykreslení interpolované teplotní mapy
├── idw.py.....Převzatý soubor pro interpolační metodu
├── temporal.py.....Soubor zprostředkovávající interpolační metody
├── interpolator.py.....Převzatý soubor pro interpolační metodu
├── Linear_Regression_model1.joblib.....natrénovaný model strojového učení
└── requirements.txt....seznam potřebných knihoven použitých ve webové aplikaci
```