



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA DOPRAVNÍ

Bc. Jiřina Lucia Varon Izová

**Analýza dat plovoucích vozidel ve vztahu
k meteorologickým podmínkám**

Diplomová práce

Studijní program: Technika a technologie v dopravě a spojích

Studijní obor: Inteligentní dopravní systémy

Hlavní vedoucí práce: Ing. Zuzana Purkrábková

Vedoucí práce: doc. Ing. Pavel Hruběš, Ph.D.

2023



K620 **Ústav dopravní telematiky**

ZADÁNÍ DIPLOMOVÉ PRÁCE
(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení studenta (včetně titulů):

Bc. Jiřina Lucia Varon Izová

Studijní program (obor/specializace) studenta:

navazující magisterský – IS – Inteligentní dopravní systémy

Název tématu (česky): **Analýza dat plovoucích vozidel ve vztahu k meteorologickým podmínkám**

Název tématu (anglicky): Data analysis of floating vehicles in relation to meteorological conditions

Zásady pro vypracování

Při zpracování diplomové práce se řiďte následujícími pokyny:

- Seznamte se s problematikou a vypracujte rešerši k tématu zadání z dostupné odborné literatury.
- Vyberte vhodné lokality, časové úseky a vypracujte statistické analýzy vztahu dostupných meteorologických dat a dat plovoucích vozidel.
- Vyhodnoťte a diskutujte vliv počasí na rychlost dopravního proudu.
- Vytvořte lokální prediktivní model rychlosti dopravního proudu vzhledem k meteorologickým podmínkám.
- Zhodnoťte dosažené výsledky, navrhňte další postup výzkumu.



- Rozsah grafických prací: dle potřeby
- Rozsah průvodní zprávy: minimálně 55 stran textu (včetně obrázků, grafů a tabulek, které jsou součástí průvodní zprávy)
- Seznam odborné literatury: Statistics, Crawley, Michael J. John Wiley & Sons Inc · 2014; Statistical Modeling and Computation, Rubinstein, Reuven Y.; Kroese, Dirk P. Springer-Verlag New York Inc. · 2013; GIS for Transportation, Miller, Shaw, Oxford Uni. Press, 2001; Spatial Databases, Rigaux et all., Morgan Kaufmann P., 2002

Vedoucí diplomové práce: **Ing. Zuzana Purkrábková**
doc. Ing. Pavel Hrubeš, Ph.D.


Datum zadání diplomové práce: **15. června 2022**
(datum prvního zadání této práce, které musí být nejpozději 10 měsíců před datem prvního předpokládaného odevzdání této práce vyplývajícího ze standardní doby studia)

Datum odevzdání diplomové práce: **15. května 2023**
a) datum prvního předpokládaného odevzdání práce vyplývající ze standardní doby studia a z doporučeného časového plánu studia
b) v případě odkladu odevzdání práce následující datum odevzdání práce vyplývající z doporučeného časového plánu studia


Ing. Zuzana Bělinová, Ph.D.
vedoucí
Ústavu dopravní telematiky


prof. Ing. Ondřej Příbyl, Ph.D.
děkan fakulty

Potvrzuji převzetí zadání diplomové práce.


Jiřina Lucia Varon Izová
jméno a podpis studenta


V Praze dne 15. června 2022

Prohlášení

Prohlašuji, že jsem tuto práci vypracovala samostatně a využila k tomu pouze literaturu uvedenou v seznamu zdrojů na konci práce.

Jsem obeznámena se zákonem č.121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon). Nemám závažný důvod proti užití tohoto školního díla ve smyslu § 60 zmíněného zákona.

V Praze dne: 15. května 2023

Podpis: 

Poděkování

Ráda bych touto cestou poděkovala vedoucím diplomové práce paní Ing. Zuzaně Purkrábkové a panu doc. Ing. Pavlu Hrubešovi, Ph.D. za poskytnutí námětu pro diplomovou práci a pomoc během jejího zpracování. Dále bych také chtěla poděkovat panu Ing. Filipovi Týcovi z ŘSD za poskytnutí dat z meteorologických stanic. V neposlední řadě děkuji svým nejbližším, kteří mi během celého mého studia byli velkou oporou.

Abstrakt

Diplomová práce se zabývá vztahem meteorologických podmínek a rychlostí dopravního proudu. Na základě dat ze silničních meteorologických stanic a FCD jsou zpracovány regresní modely, které popisují dopad meteorologických podmínek na rychlost dopravního proudu. Pomocí poskytnutých dat je také vytvořen prediktivní model pomocí vybraných algoritmů.

Klíčová slova

FCD, plovoucí vozidla, silniční meteorologie, silniční meteorologická stanice, regresní model, prediktivní model

Abstract

This diploma thesis deals with the relationship between meteorological conditions and traffic flow speed. Based on data from road weather stations and FCD, regression models are developed to describe the impact of meteorological conditions on traffic flow speed. Using the provided data, a predictive model is also developed using selected algorithms.

Key words

FCD, floating cars, road meteorology, road weather station, regression model, predictive model

OBSAH

Seznam použitých zkratek.....	8
Úvod.....	9
1. Rešerše.....	11
1.1. Související práce.....	11
1.2. Popis vybrané oblasti.....	13
1.2.1. Jihočeský kraj.....	13
1.2.2. Středočeský kraj.....	14
1.3. Popis meteorologických jevů, prvků a situací.....	15
1.3.1. Teplota vzduchu.....	15
1.3.2. Rosný bod.....	16
1.3.3. Teplota mraznutí.....	16
1.3.4. Atmosférické srážky.....	17
1.3.5. Vlhkost vzduchu.....	18
1.3.6. Tlak vzduchu.....	19
1.3.7. Rychlost a směr větru.....	19
1.3.8. Radiace.....	19
1.3.9. Dohlednost.....	19
1.4. Vliv počasí na dopravní situaci a vozovku.....	22
1.5. Nehodovost.....	23
2. Analýza datových zdrojů.....	25
2.1. Silniční meteorologické stanice.....	25
2.1.1. Popis dat z meteorologických stanic.....	27
2.2. Flotila plovoucích vozidel.....	29
2.2.1. Využití FCD dat.....	29
2.2.2. TMC segmenty.....	31
2.2.3. Popis FCD dat.....	32
3. Teoretický popis statistických metod.....	34

3.1.	Vícerozměrná regresní analýza.....	34
3.2.	Korelační analýza	37
4.	Analýza vybraných úseků.....	38
4.1.	Kubova Huť	38
4.1.1.	Příprava dat.....	39
4.1.2.	Výsledky metod.....	42
4.2.	Dálnice D11	53
4.2.1.	Příprava dat.....	54
4.2.2.	Výsledky metod.....	57
5.	Lokální prediktivní model.....	65
5.1.	Výsledky prediktivního modelu.....	66
6.	Závěr.....	70
7.	Seznam obrázků	72
8.	Seznam tabulek.....	73
9.	Seznam grafů	74
10.	Seznam příloh.....	74
11.	Citace.....	75

Seznam použitých zkratek

FCD	Data plovoucích vozidel (Floating car data)
DIS	Dálniční informační systém
PDZ	Proměnné dopravní značení
PDZ-M	Proměnné dopravní značení – meteo
ŘSD ČR	Ředitelství silnic a dálnic ČR
WAN ŘSD	Privátní celorepubliková datová síť ŘSD
CDT	Místní centrum dopravní telematiky
ZPI-T	Zařízení pro provozní informace – teploměr
DSMS	Doplňkové silniční meteorologické stanice
GPS	Globální polohový systém (Global Positioning System)
m n. m.	Metry nad mořem
WMS	World Map Service
ČHMÚ	Český hydrometeorologický ústav
SIVS	Systém integrované výstražné služby
NDIC	Národní dopravní informační centrum
IZS	Integrované záchranné složky
CSV	Čárkou oddělené údaje (Comma Separated Values)
TMC	Kanál dopravních zpráv (Traffic Message Channel)
NA	Nedostupné hodnoty (Not Available)
SQL	Strukturovaný dotazovací jazyk (Structured Query Language)
RMSE	Střední kvadratická chyba (Root Mean Square Error)
MAE	Střední absolutní chyba (Mean Absolute Error)
K-NN	Algoritmus k nejbližších sousedů (K-Nearest Neighbor)
C-ITS	Kooperativní inteligentní dopravní systémy (Cooperative-Intelligent Transport Systems)
ICT	Informační a komunikační technologie (Information and Communication Technologies)

Úvod

Sběr dat je pro dopravu zásadní téma, které zlepšuje mnoho jejích aspektů. Zdrojem velkých dat jsou data z plovoucích vozidel (Floating car data, FCD), kdy se kontinuálně zaznamenává GPS záznam flotily vozidel na vybraných úsecích pozemních komunikací. Vozidlo je zdrojem i jiných údajů, jako je časové razítko, kategorie vozidla či rychlost a směr vozidla. Ze surových dat vznikne množina parametrů, která charakterizuje skupinu vozidel na určitém segmentu v reálném čase. FCD data se obvykle používají k analytickým úlohám zkoumající dopravní proud, především pak jeho plynulost. Typické jsou také navigační úlohy.

Silniční meteorologie je odvětví aplikované meteorologie, které se zabývá meteorologickými prvky a jevy z hlediska jejich vlivu na silniční dopravu. Zahrnuje měření na silničních meteorologických stanicích a věnuje se speciální předpovědi povětrnostních podmínek, kdy se nejčastěji předpovídá stav a teplota povrchu vozovky. [1] V oboru silniční meteorologie patří mezi nejpodstatnější měřené parametry – teplota vzduchu a vozovky, teplota namrzání vozovky, množství srážek a informace o stavu vozovky (suchá, vlhká, mokrá, namrzlá).

Analýza dopravní situace a meteorologických jevů spojuje nevýhoda v nejednoznačné interpretovatelnosti, jelikož podmínky jsou v obou případech neopakovatelné a nekontrolovatelné, proto je komplexní posouzení značně ztíženo, čemuž napomáhá i dynamická povaha. Přínosy analýzy spočívají v pochopení vztahu mezi meteorologickými podmínkami a dopravní situací.

Cílem této diplomové práce je zpracovat analýzu vztahu mezi meteorologickými podmínkami a rychlostí dopravního proudu na základě dat z plovoucích vozidel a dat z meteorologických stanic. Dalším cílem je vytvoření prediktivního modelu na základě poskytnutých dat.

V úvodní části se práce věnuje důkladné rešerši studií, které se zaměřují na meteorologické a klimatické podmínky s vlivem na dopravu. Dále navazuje popis lokalit, stručné vysvětlení základních pojmů týkající se meteorologických jevů, prvků a situací, vliv počasí na dopravní situaci a vozovku a prezentace nehodových dat ve vztahu k meteorologickým podmínkám.

Následuje analytická část, kde jsou popisována data určená k analýze vztahu, jejich využití a zdroje těchto dat. Data pochází ze silničních meteorologických stanic a plovoucích vozidel.

Další část se zaměřuje na teoretický popis statistických metod. Jedná se o vícerozměrnou regresní analýzu a korelační analýzu, které slouží k popisu statistických závislostí. Cílem regresní analýzy bude najít souvislost mezi meteorologickými podmínkami a chováním dopravy, přesněji rychlostí dopravního proudu.

Pro statistickou analýzu jsou vybrány dva úseky, které jsou v následujících kapitolách označeny jako Kubova Huť a Dálnice D11. Většina dat poskytnutých od ŘSD se nachází na dálniční síti. Kubova Huť

umístěna u komunikace I. třídy je tedy vybrána z důvodu lokace. Jedná se o meteostanici umístěnou na pohoří Šumava. Předpokládá se, že zde budou rozdílné hodnoty oproti meteostanicím umístěných na dálniční síti. Před analýzou daného úseku jsou data upravena tak, aby s nimi šlo dále pracovat. Následně je pro každý úsek zpracována korelační analýza a regresní analýza.

Jako poslední jsou pak vytvořeny lokální prediktivní modely, kdy se pomocí dat z dálnice D11 a vybraných algoritmů vytváří predikce a pomocí RMSE se výsledky vybraných přístupů porovnávají. V rámci této práce jsou zmíněny algoritmy náhodného lesa, K-NN algoritmus a lineární regrese. Nakonec přichází zhodnocení a závěr.

1. Rešerše

Tato část začíná rešerší prací, které souvisí s tématem meteorologických podmínek a dopravy. Rešerše byla provedena na studii, které se tematicky zaměřují na meteorologické a klimatické podmínky s vlivem na dopravu. Dále následuje popis oblastí zdroje meteorologických dat a popis meteorologických jevů, prvků a situací. Závěrem přichází shrnutí vlivu počasí na dopravní situaci a vozovku a nehodová data ve spojitosti s povětrnostními podmínkami.

1.1. Související práce

Zimní podmínky, zejména pak náledí na povrchu komunikace, mají významné důsledky pro bezpečnost provozu na dálnicích. Predikcí kritických meteorologických podmínek na mostních komunikacích se zabývali ve studii s názvem *Using multiple linear regression and BP neural network to predict critical meteorological conditions of expressway bridge pavement icing*, kdy byla využita vícenásobná lineární regrese a algoritmus back propagation neuronových sítí na základě dat z meteorologické stanice na mostě v provincii Shandong v Číně. Přesnost výsledků korelační analýzy byly porovnány mezi lineárním regresním modelem a modelem neuronové sítě. [2]

Tématem bezpečnosti se věnovali i ve studii *Assessment of Influence of Meteorological Parameters on the Risk of Accidents on Roads Outside Settlements*, která potvrzuje důležitost vlivu meteorologických podmínek na nehodovost. Výzkum zkoumá vliv pěti hlavních meteorologických parametrů na nehodovost, kdy pomocí výpočtů byly zjištěny nelineární závislosti počasí a rizika dopravních nehod a nalezeny vzorce popisující vztah mezi hodnotami meteorologických prvků a variací rizika dopravní nehody. [2]

Pohled na vztah silničního provozu a počasí poskytl i článek *A review of approaches to the study of weather's effect on road traffic parameters*, který zkoumal různé kombinace dopravních charakteristik a meteorologických podmínek. Autoři nejprve identifikovali optimální povětrnostní podmínky pro dopravní provoz a vnímání řidiče pomocí rozličných metodologií, zdrojů dat a pozadí. Výsledky poukázaly na skutečnost, že velmi záleží na oblasti (zvláště kvůli klimatickému pásmu), lidském chování, právním rámci (týká se např. omezení rychlosti), specifických attributech pozemní komunikace a místnímu pohledu. [3]

Existují i studie, které se věnují tématu meteorologických jevů a dopravy vzhledem k využití veřejné hromadné dopravy (VHD). Ve studii *Impacts of weather on public transport ridership: Results from mining data from different sources* autoři zohledňují změny počasí a zkoumají vliv na chování uživatelů veřejné hromadné dopravy na základě dat z čínského Šen-čenu ze září 2014. K analýze byl použit vícerozměrný model, jehož výsledky poukázaly, že některé aspekty povětrnostních podmínek ovlivňují veřejnou dopravu více než jiné. [4]

Případová studie *Influence of Weather on Transport Demand* se věnuje vlivu počasí na poptávku městské hromadné dopravy ve Vídni. Podle počasí v jednotlivých dnech by mohlo být odlišně nastavené řízení dopravních systémů. Pomocí statistického modelování byl dokázán vliv počasí na volbu mobility většiny populace, zároveň závisí na způsobu a účelu cesty a vlastnostech cestujícího. [5]

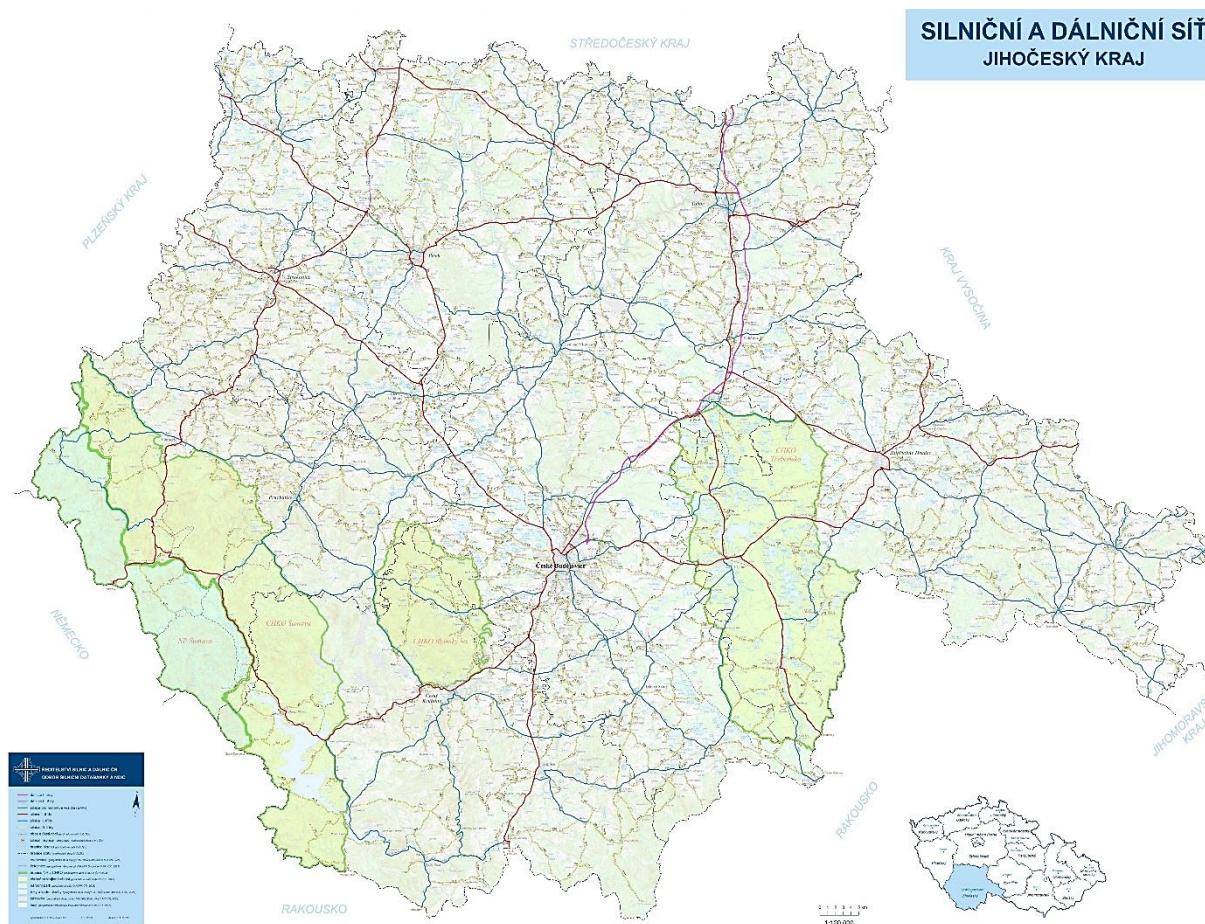
Vliv počasí na dopravní situaci se zvláště diskutuje ve vztahu k tématu klimatické změny. Dopad povětrnostních podmínek na silniční síť zkoumá studie *Weather and climate risks to road transport*, která identifikuje a posuzuje rizika změny klimatu ve Spojeném království. Poskytuje také shrnutí hlavních dopadů počasí a klimatu na silniční síť a otevírá otázku potřeby zvýšení adaptačních opatření, jako jsou inženýrská řešení a nové technologie, aby silniční doprava fungovala efektivně bez větších ohledů na stav počasí. Nezanedbává ani zohlednění socioekonomických faktorů a popisuje vzájemné závislosti systémů, zároveň apeluje na dostatek kvalitních dat pro pochopení problému. [6]

1.2. Popis vybrané oblasti

Data z meteorologických stanic byla poskytnuta z Jihočeského a Středočeského kraje, které jsou popsány v této kapitole z pohledu širších dopravních vztahů.

1.2.1. Jihočeský kraj

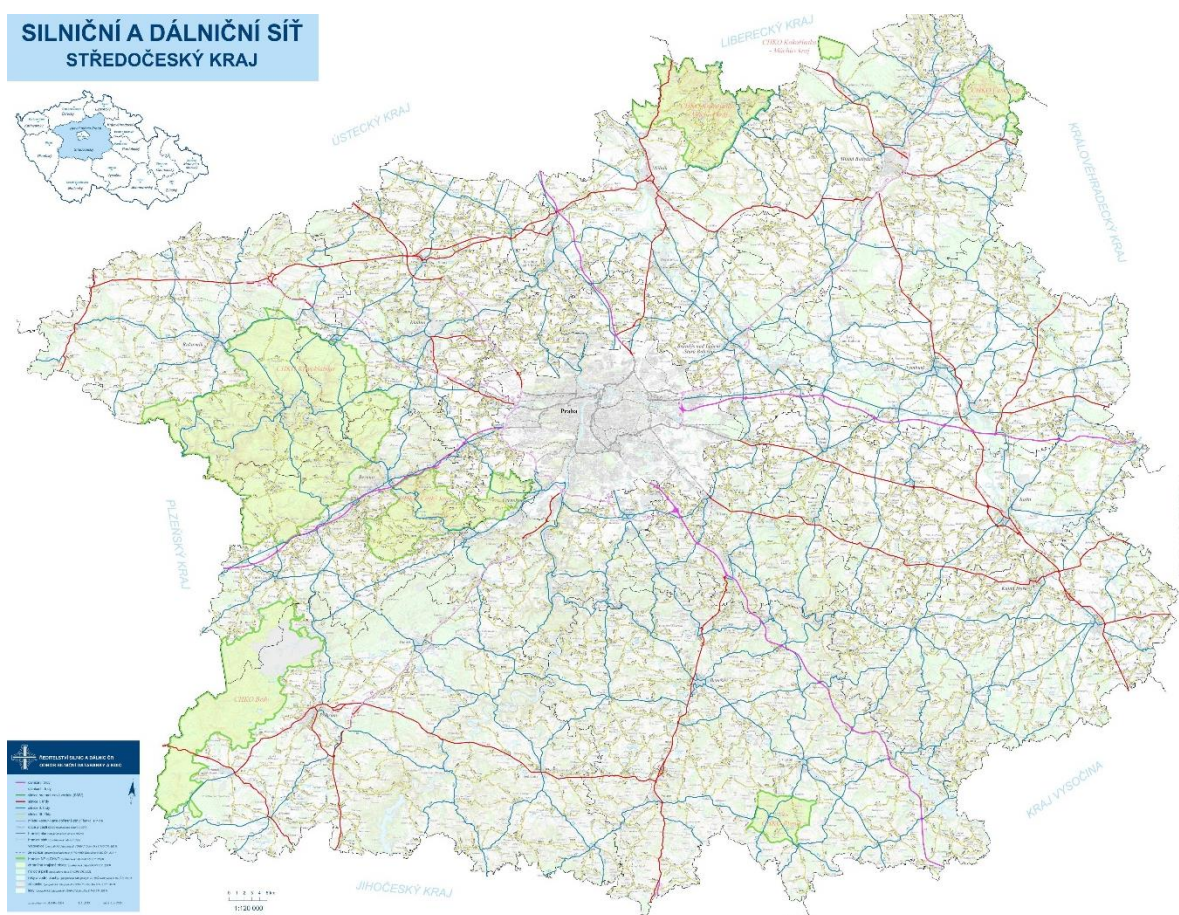
V Jihočeském kraji se k 1. 1. 2022 nacházelo 74 km dálnic a 646 km silnic I. třídy. [7] Nyní je v realizaci 54,8 km dálnic, z toho 28,4 km na dálnici D3 a 26,4 km dálnici D4. [8] Právě dálnice D3 patří mezi nejvytíženější pozemní komunikace v Jihočeském kraji. Dálnice D3 v budoucnu bude sloužit jako průtah Praha – Rakousko. Pro Jihočeský kraj a město České Budějovice je tato dálnice bezesporu nejdůležitější dopravní stavbou. Její dokončení v Jihočeském kraji se plánuje na rok 2025 zprovozněním úseku Nažidla – Dolní Dvořiště. [9] Po jejím dokončení se uleví dopravě v kraji a zbytná doprava I. stupně již nebude nevhodně zatěžovat místní dopravní infrastrukturu. Dálnice D4 je převážně ve výstavbě a v budoucnu by měla složit jako spojení Písek – Příbram – Praha. Další významná komunikace je silnice I/20 České Budějovice – Písek, která je vedena severozápadně směr Písek a pak dále navazuje na D4. Na obrázku 1 můžeme vidět silniční a dálniční síť Jihočeského kraje.



Obrázek 1: Silniční a dálniční síť Jihočeský kraj [10]

1.2.2. Středočeský kraj

Středočeský kraj má hned po Praze nejhustší dopravní síť, která je zobrazena na obrázku 2. Hlavním důvodem je, že geograficky obklopuje Prahu a je tedy přes tento kraj vedeno nejvíce dálnic. [11] Stěžejní komunikace vedoucí přes Středočeský kraj je dálnice D1 Praha – Brno – Ostrava – polské hranice, dálnice D11 Praha – Hradec Králové – Jaroměř, dálnice D5 Praha – Plzeň – hranice s Německem. Dále zde najdeme dálnici D6 Praha – Karlovy Vary, dálnici D7 Praha – Chomutov, dálnici D8 Praha – Ústí nad Labem, dálnici D10 Praha – Liberec, výhledově dálnici D3 Praha – České Budějovice a dálnici D4 Praha – Písek. Částečně se ve Středočeském kraji nachází i Pražský okruh D0. Ačkoli má Středočeský kraj relativně dobře vybudovanou dopravní síť komunikací, tak dochází pravidelně k přetížení silnic a dálnic, a to zejména na hlavních průtazích. Nejproblematičtější úsek je zmiňovaná dálnice D1 a Pražský okruh D0.



Obrázek 2: Silniční a dálniční síť Středočeského kraje [10]

1.3. Popis meteorologických jevů, prvků a situací

Počasi vystihuje stav atmosféry, který je charakterizován souhrnem hodnot všech meteorologických prvků atmosférickými jevy v určitém místě a čase. Předpověď počasí vystihuje budoucí stav povětrnostních podmínek na základě aplikací meteorologických poznatků, mezi které patří přízemní pozorování a měření na meteorologických stanicích, aerologická sondážní měření ve vyšších vrstvách atmosféry, družicová a radarová měření, výstupy z regionálních a globálních numerických předpovědních modelů, statistických modelů, koncepčních modelů apod. Pro specifické požadavky některých uživatelů se využívají speciální předpovědi, jedná se například o zimní údržbu pozemních komunikací, leteckou dopravu, energetiku, zemědělství, mobilní operátory či o státní orgány a instituce.

Kombinace či nabývání konkrétních hodnot jednotlivých hydrometeorologických jevů, které mají za následek materiální škody, újmy na zdraví a na životě, se řadí pod pojem nebezpečné jevy. Pro nebezpečné jevy jsou stanovena kritéria pro jejich monitorování a varování veřejnosti a pro každý jev je dán i stupeň nebezpečí, který představuje. Řadí se sem například vysoké teploty, déletrvající sucha, silné (vydatné, dlouhotrvající, přívalové) srážky vedoucí k povodním, krupobití, silný nárazový vítr, nebezpečí požárů, silné mrazy, ledovka, náledí, silná námraza, vysoká sněhová pokrývka, sněhové jazyky a závěje, rychlé tání sněhu, silné mlhy, mrznoucí mlhy, ale i smogové situace a lavinové nebezpečí. Před nebezpečnými jevy se vydává výstraha ČHMÚ v rámci Systému integrované výstražné služby (SIVS).

Meteorologické jevy souvisí i s nadmořskou výškou. V rámci ČR existuje následující dělení poloh dle nadmořské výšky:

- Nižší polohy do 400 m n. m.
- Střední polohy od 400 do 600 m n. m.
- Vyšší polohy od 600 do 800 m n. m.
- Horské polohy od 800 m n. m.

Meteorologické prvky a jevy označují jevy v atmosféře nebo na zemském povrchu, které jsou pozorovány na meteorologických stanicích a v jejich okolí. [12] Následující rozdělení a popis vybraných jevů vychází z měřených parametrů na silniční meteorologické stanici.

1.3.1. Teplota vzduchu

Teplota vzduchu, určující tepelný stav ovzduší, je většinou udávána ve dvou metrech nad povrchem země v jednotkách Celsiovy teplotní stupnice s přesností na desetiny °C. Pro teplotu jsou typické pojmy nejnižší noční teplota, což je tří až čtyř stupňové rozmezí předpokládaných minimálních nočních teplot v období od 21:00 do 7:00 hodin místního času, a nejvyšší denní (tj. maximální) teplota, což je rozmezí předpokládaných maximálních denních teplot od 7:00 do 21:00 hodin místního času. Zpravidla se nejnižší hodnoty vyskytují ráno kolem východu Slunce, nejvyšší hodnoty pak mezi 13. a 16. hodinou.

Teplota se měří teploměrem, většinou elektrickým, který se nachází v dobrém tepelném kontaktu se vzduchem, zároveň je dokonale uchráněn před přímým slunečním zářením. [8] [1]

Teplota vzduchu se proměňuje celoročně, závisí hlavně na cirkulaci vzduchu, na stavu terénu, ročním období (tj. na poloze Slunce), na přítomnosti oblačnosti a srážek, větru a na nadmořské výšce. V rámci místních podmínek lze vyzorovat specifický denní a roční chod teploty vzduchu, přičemž průběh této veličiny se může někdy i značně vymykat. Během roku má teplota typický průběh jako v mírných klimatických oblastech. Nejnižší teploty nastávají v zimním období, obvykle v lednu, naopak nejvyšší teploty jsou v letním období, obvykle v červenci. Týká se to jak průměrných denních či měsíčních teplot, tak i absolutního ročního minima. Stabilní průběh teploty vzduchu je charakteristický pro mrazivé zimní období, kdy existují stabilní tlakové výše, podobně pro období letní, kdy jde jen o vysoké teploty s opačnou cirkulací. Největší rozdíly teplot v jednotlivých dnech jsou registrovány během jarního období. [13]

1.3.2. Rosný bod

Rosný bod spolu s měřenou teplotou vzduchu patří k základním charakteristikám vlhkosti vzduchu. Teplota rosného bodu vyjadřuje stav, kdy se vlhký vzduch o dané hodnotě směšovacího poměru vodní páry stane nasyceným vzhledem k vodě následkem izobarického ochlazování. Při poklesu teploty pod hodnotu teploty rosného bodu nastává kondenzace vodní páry obsažené ve vzduchu a vzniká tak rosa či mlha. Pokud je relativní vlhkost vzduchu nižší než 100 %, teplota rosného bodu bude vždy menší než teplota vzduchu. Deficit teploty rosného bodu narůstá s o to menší relativní vlhkostí vzduchu při určité teplotě vzduchu. Na meteorologických stanicích v rámci České republiky se teplota přízemního rosného bodu vypočítává ze staničního tlaku vzduchu, teploty vzduchu a relativní vlhkosti, které jsou měřeny pomocí teplotně–vlhkostních senzorů, případně výpočtem z údajů psychrometru, eventuálně lze teplotu rosného bodu také určit z psychrometrických tabulek. Přibližnou hodnotu teploty rosného bodu lze též měřit přímo kondenzačním vlhkoměrem nebo termohygroskopem. [1]

1.3.3. Teplota mrznutí

Teplota mrznutí, neboli bod mrznutí, označuje bod tuhnutí čisté vody při daném atmosférickém tlaku vzduchu. Přízemní minimální teplota určuje nejnižší teplotu vzduchu měřenou ve výšce 5 cm nad povrchem země za určitý časový interval, obvykle za noc a ráno. Přízemní mrazíky se pak vyskytují, pokud přízemní teplota vzduchu klesne pod bod mrazu. V zimním období se pak v souvislosti s teplotou namrzání mohou objevovat následující jevy:

- Ledovka je charakterizována jako homogenní průhledná ledová vrstva s hladkým povrchem, která se tvoří na povrchu země, vozovkách, chodnících, na stožárech aj. Příčinou je mrznoucí mrholení nebo mrznoucí déšť, kdy buď zmrznou přechlazené vodní kapky, které jsou v kapalném stavu i při teplotě pod bodem mrazu, při dopadu na zemský povrch nebo na

předměty, jejichž teplota je pod nulou či těsně kolem nuly, nebo zmrznou nepřechlazené vodní kapky okamžitě po dopadu na zemský povrch nebo na předměty, jejichž teplota je výrazně pod bodem mrazu.

- Ledová vrstva pokrývající zem může být také náledí, které vzniká třemi způsoby. Vznik může nastat v případě, že nepřechlazené kapky deště či mrholení později na zemi zmrznou, nebo jestliže voda z úplně či částečně roztátého sněhu na zemi opět pomrzne. Třetí příčinou může být sníh, který kvůli provozu vozidel na komunikacích zledovatí. Poslední dvě příčiny se pak nazývají zmrazky.
- Námraza označuje usazeninu, která se vytváří rychlým zmrznutím přechlazených kapek mlhy nebo oblaků, usazujících se na předmětech při teplotách v rozmezí od -1 do -10 °C.
- Na vozovce se také může formovat jíní, což je druh usazených tuhých srážek, který nastává při jasných nocích kondenzací vodní páry z okolního vzduchu a jejím mrznutím na povrchu. Jde o stejný princip jako v případě rosy, ale při záporných teplotách. [8] [1]

1.3.4. Atmosférické srážky

Atmosférické srážky značí soustavu vodních částic, které vznikají kondenzací či sublimací vodní páry v následujících stavech – v kapalném stavu, kam řadíme déšť, mrznoucí déšť, mrholení, mrznoucí mrholení, dále existuje stav tuhý v podobě sněžení, sněhových zrn, sněhových krupek, krupek, zmrzlého deště, krup a ledových jehliček. Lze rozlišovat i stav padající, vznášející se atmosféře, zdvižené větrem z povrchu, usazené na předmětech na zemi nebo ve volné atmosféře, což je například rosa, jíní, námraza a ledovka.

Jednotlivé jevy disponují různými charakteristikami. Déšť se definuje jako vodní srážky, které vypadávají z oblaků ve tvaru kapek větších než 0,5 mm, případně i menším, jestliže jsou velmi rozptýlené. Objevují se hlavně z vrstevnaté oblačnosti.

Přeháňky lze popsat jako srážky s náhlým začátkem i konce, rychlým kolísáním intenzity a krátkým trváním. Doprovází je často rychlé střídání velké oblačnosti s krátkým vyjasněním. Přeháňky vypadávají z kupovité (konvektivní) oblačnosti, kdy mohou vypadávat i sněhové krupky.

Jako husté kapalně srážky se označuje mrholení, které je složené z drobných kapiček o průměru menším než 0,5 mm a vypadává z nízké oblačnosti typu stratus nebo z mlhy.

Sněžení jsou pak srážky ze sněhových vloček či z ledových krystalků vypadávající z vrstevnaté oblačnosti. Sněhové přeháňky vypadávají pak z oblačnosti kupovité (konvektivní).

Padající srážky můžeme rozdělovat podle délky trvání do následujících skupin:

- **Trvalé srážky** – vyznačují se delší dobu výskytu a vcelku neměnicí se intenzitou velkoplošných srážek (např. déšť, sníh)

- **Občasné srážky** – mezi jednotlivými srážkovými jevy, které se opakují, se objevují přestávky trvající řádově hodiny. Může se jednat jak o velkoplošné, tak i o konvektivní srážky.
- **Přeháňky** – fáze padajících srážek trvá relativně krátkou dobu, řádově minuty, případně desítky minut. Intenzita srážek a množství oblačnosti je poměrně kolísavé, mnohdy se mezi jednotlivými přeháňkami i vyjasní.
- **Četné srážky** – vztahují se k přeháňkám, které se opakují v celkem krátkých intervalech, řádově desítky minut.
- **Bez specifikace** – není časově určená doba trvání, většinou se tato charakteristika používá při přechodu fronty.

Charakteristikou srážek je také jejich intenzita. Množství atmosférických srážek spadlých za jednotku času, neboli intenzitu, lze podle výšky vrstvy vody nebo výškou sněhu rozlišit do následujících úrovní (tabulka č. 1). Průměrná intenzita srážek se vyhodnocuje ze srážkoměrů, okamžitá intenzita se pak měří váhovým srážkoměrem. [8] [1]

Tabulka 1: Kategorizace intenzity srážek

Intenzita	Děšť v mm/h	Výška sněhové pokrývky v cm/h
Velmi slabá	Neměřitelné množství	Jednotlivé vločky nepokrývají celý povrch země
Slabá	0,1 – 2,5	Méně než 0,5
Mírná	2,6 – 8,0	0,6 – 4,0
Silná	8,0 – 40,0	Více než 4,0
Velmi silná	Více než 40	Krátkodobé intenzivní sněhové přeháňky

1.3.5. Vlhkost vzduchu

Vlhkost vzduchu popisuje množství vodní páry ve vzduchu. Hodnotí se obsah vodní páry v určitém místě atmosféry za určité teploty, obvykle se měří relativní vlhkost vzduchu nebo tlak vodní páry. Relativní vlhkost vzduchu se určuje v procentech, tlak vodní páry pak v hPa. Pro měření se využívá vlhkoměr, na meteorologických stanicích je instalováno vlhkostní číslo umístěné v radiačním krytu. Pomocí údajů z meteorologických družic lze také vyhodnotit vertikální profil vlhkosti vzduchu, v důsledku pohlcování odraženého nebo vlastního záření zemského povrchu v absorpčních pásech vodní páry. Vertikální profil vlhkosti vzduchu rozděluje obsah vodní páry v atmosféře s výškou, kdy v mezní vrstvě atmosféry se odvíjí od výparu na zemském povrchu, na vertikálním promíchávání vzduchu a na teplotě, zatímco ve volné atmosféře pak závisí hlavně na advekci a vertikálních pohybech vzduchu.

V letním anticyklonálním období je vlhkost obvykle nízká, ojediněle i méně než 30% relativní nasycení vzduchu. V podzimním mlhavém či zimním období může nasycení dosáhnout až nad 95 %. Vlhkost vzduchu ovlivňují různé faktory, například srážky, dodaná energie ze slunečního svitu nebo vítr. [1][14]

1.3.6. Tlak vzduchu

Tlak vzduchu určuje sílu, která působí v určitém místě atmosféry kolmo na libovolně orientovanou jednotkovou plochu a která je vyvolaná tíhou vzduchového sloupce rozprostírajícího se od úrovně, ve které se tlak zjišťuje, až k horní hranici atmosféry. Je standardně přepočtený na střední výšku hladiny moře, jednotkou jsou pascaly, hlavně pak jeho násobky (hektopascaly, hPa). Na meteostanicích se měří staničními tlakoměry s přesností na desetiny hPa. Normální tlak vzduchu na hladině moře ve standardní atmosféře je 1 013,25 hPa. Na základě změn tlaku, náhlosti a intenzity těchto změn lze pozorovat a s určitou pravděpodobností určovat, jak se počasí bude vyvíjet v nejbližším časovém horizontu. [14]

1.3.7. Rychlost a směr větru

Vítr, neboli pohyb vzduchu vůči zemskému povrchu, je charakterizován rychlostí a směrem horizontální složky pohybu (vertikální složka většinou není brána v potaz). Vítr vzniká zejména působením horizontální složky síly tlakového gradientu a uplatňuje se i Coriolisova síla. Slouží jako prostředek přenosu vody v atmosféře, přenosu energie, hybnosti a jiných fyzikálních vlastností ve vzduchových hmotách. Přímé měření rychlosti větru, ale i směr větru, eviduje anemometr. Rychlost větru je zaznamenává v m/s, případně v km/h nebo slovním vyjádřením, tj. klidno, slabý vítr, mírný, čerstvý, silný, velmi silný. Náraz větru pak vyjadřuje krátkodobé zvýšení rychlosti větru, což musí být oproti průměru alespoň o 5 m/s po dobu 1 až 20 s. Náraz větru je stanoven od 12 m/s. [8] [1]

1.3.8. Radiace

Radiace, neboli záření, se v meteorologii vztahuje k šíření elektromagnetických vln nebo toků hmotných částic (korpuskulárního záření) atmosférou. Velikost záření vystihuje především intenzita toku energie, jejíž jednotkou jsou W/m^2 . [1]

1.3.9. Dohlednost

Dohlednost je definována jako největší vzdálenost, při které lze vidět a spolehlivě rozpoznat černý předmět vhodných rozměrů umístěný u země, pokud je pozorován za denního světla proti obloze horizontu či na pozadí mlhy, nebo který lze vidět a rozpoznat v noci, jestliže existuje umělé osvětlení na úrovni normálního denního světla. Dohlednost závisí na stavu zraku pozorovatele. Pro účely přístrojového měření je definován práh kontrastové citlivosti na hodnotu 0,025, v noci se určuje prahová hodnota osvětlení např. za občanského soumraku (fáze soumraku, která nastává po západu nebo před východem Slunce) 10^6 luxů a za tmavé noci při svitu hvězd $10^{7,5}$ luxů. Tímto způsobem je pak možné

porovnávat výsledky vizuálního a přístrojového pozorování. Dohlednost se měří většinou měřičem průzračnosti nebo transmisometrem, případně měřičem dohlednosti v metrech či v jeho násobcích.

Dohlednost závisí mimo jiné na rozptylových podmínkách pro rozptyl znečišťujících látek v ovzduší. Stav znečištění ovzduší se odvíjí od proudění vzduchu v horizontálním i vertikálním směru: rozptylové podmínky lze rozlišit do následujících skupin:

- **Dobré** – dobré rozptylové podmínky značí, že ve výšce 1 000 až 1 500 m nad terénem se nevyskytuje zádržná vrstva, jako je inverze a izotermie, která by omezovala rozptyl škodlivin. V případě výškové zádržné vrstvy může záviset i na rychlosti větru pod spodní hranicí zádržné vrstvy.
- **Mírně nepříznivé** – objevuje se výskyt zádržné vrstvy, kdy záleží na rychlosti větru, do jaké míry umožní rozptylu škodlivin. Zádržná vrstva nesplňuje kritéria dobrých ani nepříznivých rozptylových podmínek.
- **Nepříznivé** – ve výšce do 1 000 m se nachází mohutná zádržná vrstva nad terénem v kombinaci se slabým nebo žádným prouděním. Vzniká situace, při které je rozptyl příměsí v atmosféře téměř znemožněn a předpokládá se déletrvající významné překročení imisních limitů.

Rozptylové podmínky ovlivňují míru promíchání a ředění znečišťujících látek v ovzduší, čímž určuje jejich koncentraci. Nicméně se ale odlišují od samotné kvality ovzduší a jeho znečištění. Lze to pozorovat na případu nepříznivých podmínek, které neznamenají nutně vysoké koncentrace škodlivin (jako tomu může být v letních měsících při nízkých emisích znečišťujících látek). Zároveň vysoké koncentrace většinou nastávají při nepříznivých rozptylových podmínkách a při působení jiných faktorů (např. nízká teplota vzduchu). Rozptylové podmínky jsou indikovány ventilačním indexem, který se určuje na základě součin výšky mezní vrstvy a průměrné rychlosti větru v mezní vrstvě.

Při překročení informativní prahové hodnoty znečišťující látky v ovzduší (např. suspendovaných částic PM10, přízemního ozonu, oxidu siřičitého, oxidu dusičitého) se vyhláší smogová situace. Při překročení regulační prahové hodnoty znečišťující látky se vyhláší regulace, kdy je nutné regulovat množství emisí pro dané zdroje. Pokud se překročí varovné prahové hodnoty přízemního ozonu, vyhláší se varování.

Dohlednost může být ovlivněna inverzní situací, což je synoptický stav, kdy teplota vzduchu vzrůstá s vyšší nadmořskou výškou. Inverzní vrstva zamezuje v atmosféře promíchávání vzduchu, což vede k znečištění vzduchu pod vrstvou inverze a k hromadění vodní páry, v důsledku toho k tvoření mlh a nízké inverzní oblačnosti.

Dohlednost se může vztahovat i k intenzitě sněžení. V tabulce č. 2 je podle množství spadáných sněhových vloček za hodinu uveden vliv na dohlednost.

Tabulka 2: Klasifikace dohlednosti dle množství sněhu

Množství sněhu v cm/h	Dohlednost
Méně než 0,5	Neovlivňuje
0,6 – 4,0	Mírně zhoršená
Více než 4,0	Zhoršená již na 500 m
Krátkodobé intenzivní sněhové přeháňky	Zhoršená pod 500 m

Vodorovnou dohlednost může ovlivňovat také mlha, která se popisuje jako soustava velmi malých vodních kapek, případně nepatrných ledových krystalů, rozptýlených ve vzduchu. Relativní vlhkost vzduchu může dosahovat až 100 %, je v mlze vysoká. Dohlednost je snížena pod 1 km, dokonce i jen v jednom pruhu. [8] [12] [1]

1.4. Vliv počasí na dopravní situaci a vozovku

Následující tabulka obsahuje základní přehled jednotlivých meteorologických parametrů a jejich vliv na dopravní proud, vozovku a dopady na provoz.

Tabulka 3: Vliv počasí na dopravu [15]

Proměnné	Vliv na dopravní proud	Vliv na vozovku	Dopady na provoz
Teplota a vlhkost vzduchu	-	-	Strategie údržby komunikací (např. v případě sněhu, ledu...) Plánování oprav (např. oprava krytu vozovek, obnova vodorovného značení...)
Srážky	Kapacita vozovky Rychlost provozu Riziko nehody Cestovní zpoždění	Viditelná vzdálenost Překážka v jízdním pruhu Tření vozovky	Výkon vozidla (např. trakce) Strategie údržby komunikací Řízení rychlostních limitů Řízení signálního plánu Řídící schopnosti/chování řidiče Vliv na rozhodování o evakuaci
Rychlost větru	Rychlost provozu Cestovní zpoždění Riziko nehody	Viditelná vzdálenost * Překážka v jízdním pruhu**	Výkon vozidla (např. stabilita) Řízení vjezdu vozidel (např. omezení typu vozidla, uzavření komunikace...) Vliv na rozhodování o evakuaci
Mlha	Rychlost provozu Cestovní zpoždění Riziko nehody Rozptyl rychlosti	Viditelná vzdálenost	Strategie údržby komunikací Řídící schopnosti/chování řidiče Řízení rychlostních limitů Řízení vjezdu vozidel
Teplota vozovky	-	Poškození infrastruktury	Strategie údržby komunikací
Stav vozovky	Kapacita vozovky Rychlost provozu Cestovní zpoždění Riziko nehody	Poškození infrastruktury Tření vozovky	Strategie údržby komunikací Výkon vozidla Řídící schopnosti/chování řidiče (např. výběr trasy) Řízení signálního plánu Řízení rychlostních limitů Řízení vjezdu vozidel
Hladina vody	Rychlost provozu Cestovní zpoždění Riziko nehody	Zatopení jízdního pruhu.	Řízení vjezdu vozidel Vliv na rozhodování o evakuaci Koordinaci v rámci území

*kvůli navátému sněhu, prachu, písku apod.

**kvůli navátému sněhu, fyzickým překážkám, úlomkům apod.

1.5. Nehodovost

Povětrnostní podmínky zvyšují riziko nehodovosti. Z tabulky č. 4 a 5 lze pozorovat, v kolika případech byly zaznamenány ztížené povětrnostní podmínky, což mohlo přispět ke vzniku nehody či zvýšit její závažnost. Údaje se týkají Jihočeského a Středočeského kraje za rok 2022 na dálnicích a silnicích I. třídy.

Tabulka 4: Statistika nehodovosti v Jihočeském kraji [16]

Jihočeský kraj			
Povětrnostní podmínky v době nehody	Druh pozemní komunikace	Počet nehod	Počet nehod celkem
Na počátku deště, slabý déšť, mrholení apod.	Dálnice	3	36
	Silnice I. třídy	33	
Déšť	Dálnice	10	77
	Silnice I. třídy	67	
Mlha	Dálnice	1	15
	Silnice I. třídy	14	
Sněžení	Dálnice	1	17
	Silnice I. třídy	16	
Námraza, náledí	Dálnice	2	18
	Silnice I. třídy	16	
Nárazový vítr (boční, vichřice apod.)	Dálnice	2	8
	Silnice I. třídy	6	
Jiné ztížené	Dálnice	1	2
	Silnice I. třídy	1	

Tabulka 5: Statistika nehodovosti ve Středočeském kraji [16]

Středočeský kraj			
Povětrnostní podmínky v době nehody	Druh pozemní komunikace	Počet nehod	Počet nehod celkem
Na počátku deště, slabý déšť, mrholení apod.	Dálnice	85	142
	Silnice I. třídy	57	
Déšť	Dálnice	228	321
	Silnice I. třídy	93	
Mlha	Dálnice	34	77
	Silnice I. třídy	43	
Sněžení	Dálnice	60	83
	Silnice I. třídy	23	
Námraza, náledí	Dálnice	14	26
	Silnice I. třídy	12	
Nárazový vítr (boční, vichřice apod.)	Dálnice	7	16
	Silnice I. třídy	9	
Jiné ztížené	Dálnice	18	26
	Silnice I. třídy	8	

Z údajů lze vyčíst, že z meteorologických jevů má na nehodách největší podíl déšť a jeho mírnější podoby, jako je mrholení a začínající déšť, dále pak sněžení či mlha. Povětrnostní podmínky nemusí být nutně příčinou vzniku nehod, ale jednoznačně zvyšují riziko jejího vzniku.

2. Analýza datových zdrojů

Dopravní informace zaštiťují komplexní a rozsáhlou síť informací, které umožňují v reálném čase reagovat na aktuální dopravní situace či analyzovat historická data pro statistické vyhodnocování dopravního proudu a tvorbu modelů. Zdroje dat můžeme dělit na dvě skupiny dle způsobu pořizování: data stacionární a data mobilní. Stacionární zdroje dat poskytují informace z fixní pozice, mobilní zdroje poskytují celoplošné pokrytí a pracují s určitou mírou pravděpodobnosti a tolerance. Mezi stacionární zdroje dat řadíme různé senzory, jako jsou indukční smyčky, kamery a jiná čidla, které poskytují aktuální dopravní data z nejvytíženějších komunikací. Řešením pro sběr dat z méně frekventovaných úseků či objízdných tras v případě nehod mohou být data pocházející z mobilních zdrojů umístěných ve vozidlech. [17]

Mezi zdroje dat pro účely diplomové práce byly využity silniční meteorologické stanice a flotila plovoucích vozidel, které patří mezi významné zdroje pro řízení dopravy a jiné dopravní systémy.

2.1. Silniční meteorologické stanice

Silniční meteorologická stanice je zařízení vybavené senzory pro celoroční měření teploty vzduchu, vlhkosti vzduchu, srážek, rychlosti a směru větru, atmosférického tlaku a případně jiných parametrů. Měřicí stanoviště jsou vybavena přehledovými kamerami a senzory, mezi něž patří senzory teploty a vlhkosti, barometrický senzor, vozovkové senzory, srážkoměry, dohledoměry, senzory směru a rychlosti větru.

Stanice tvoří bezdrátovou senzorskou síť, jejíž prvky předávají informace místnímu dispečinku, kde na technologickém serveru proběhne validace a následně transformace do centrálních systémů do WAN ŘSD (privátní celorepubliková datová síť ŘSD) ve formátu RWIS_RSD.xml, který operuje se Silničním meteorologickým informačním systémem (SMIS). Pro integraci stanic do dálničního informačního systému (DIS) je potřeba napojení na datakoncentrátor, který periodicky stahuje údaje z meteostanice a ukládá je do centrální databáze DIS serveru v CDT. Data musejí být v databázi přístupná pro místní vizualizaci DIS pomocí standardního rozhraní, tj. pro nadstavbové systémy v rámci WAN ŘSD, ať se jedná o SMIS či jiné technologické servery smluvních servisních organizací.

Vybrané meteorologické stanice disponují také funkcí řídit proměnné dopravní značení, které zobrazují meteorologické informace – teplotu vozovky a vzduchu (ZPI-T) nebo varování před smykem a náledím (PDZ-M). Pro komunikace se v tomto případě používá komunikační napojení a řízení pomocí datakoncentrátorů DIS, pokud jsou v dané lokalitě k dispozici. Pro přímé řízení z meteostanice je nutné, aby byla řídicí jednotka stanice schopná zpracovat stavové a provozní informace z ovládaného zařízení a zároveň musí být schopná předávat tyto informace do dohledových systémů.

Komunikační rozhraní mezi meteostanicí a přenosovým prvkem je tvořeno galvanicky oddělenými sériovými porty RS232/RS485/RS422, popřípadě ethernetovým rozhraním. Stanice jsou také obvykle

vybaveny kamerovým bodem snímajícím stav vozovky, který dovytváří komplexní přehled o stavu počasí zprostředkováním obrazové informace. Obrazový snímek se přenáší v periodě do 5 minut s minimální kvalitou 640 x 480 px.

Poloha meteostanic se určuje dle meteorologického průzkumu ČHMÚ, termálního mapování nebo na základě požadavků správce komunikace. Lokalita je omezena i možnostmi technologického napojení na napájecí síť a přenosové datové trasy. Doporučená vzdálenost od krajů vozovky je 10 m, která zajišťuje ochranu čidel před nepříznivými vlivy vozovky a bezpečnou vzdálenost od vozovky. Zamýšlená instalace (a s tím související umístění, konfigurace apod.) schvaluje oddělení inteligentních dopravních systémů ŘSD.

Podle charakteristiky umístění lze silniční meteorologické stanice rozdělit do tří skupin:

Referenční meteostanice

Referenční meteostanice jsou umístěny v širší oblasti nebo úseku komunikace, nejsou vybudovány na potencionálně nebezpečném místě se zhoršenou sjízdností nebo zvýšenou nehodovostí. Využívají se také pro tvorbu regionální nebo liniové specializované předpovědi počasí k údržbě komunikací. Referenční meteostanice se na dálnicích instalují každých 8 - 10 km.

Lokální meteostanice

Lokální meteostanice se nachází na kratších úsecích (cca do délky 3 km), v lokalitě, která byla analýzou silniční sítě vyhodnocena jako riziková z pohledu sjízdnosti nebo nehodovosti. Typicky se umísťují na mostech, vrcholech stoupání, na lesních úsecích apod. V některých případech se referenční a lokální význam lokality překrývá.

Doplňkové meteostanice (DSMS)

Doplňkové meteostanice jsou situovány na místech, kde je potřeba potvrdit hypotézu změny stavu povrchu vozovky. Nedisponují rozhodovací pozicí, ale slouží jen jako doplněk k dané lokalitě. Typicky se jedná například o podružné stanice, které jsou umístěny několik stovek metrů od hlavní řídicí stanice, ke které jsou datově připojeny. [18]

2.1.1. Popis dat z meteorologických stanic

První druh byla historická data ze sítě meteostanic v Jihočeském a Středočeském kraji z období 1.1.2019 – 31.12.2021, která byla poskytnuta ŘSD v CSV formátu.

Data ze sítě meteostanic byla uspořádána v souborech pro jednotlivou meteostanici a jednotlivý rok. Jednotlivé stanice jsou označeny čtyřmístným kódem (Cxxx Jihočeský kraj, Sxxx Středočeský kraj). Meteorologickou stanici lze lokalizovat GPS souřadnicemi a údaji o poloze m n. m., umístění na dané komunikaci, kilometrem a názvem.

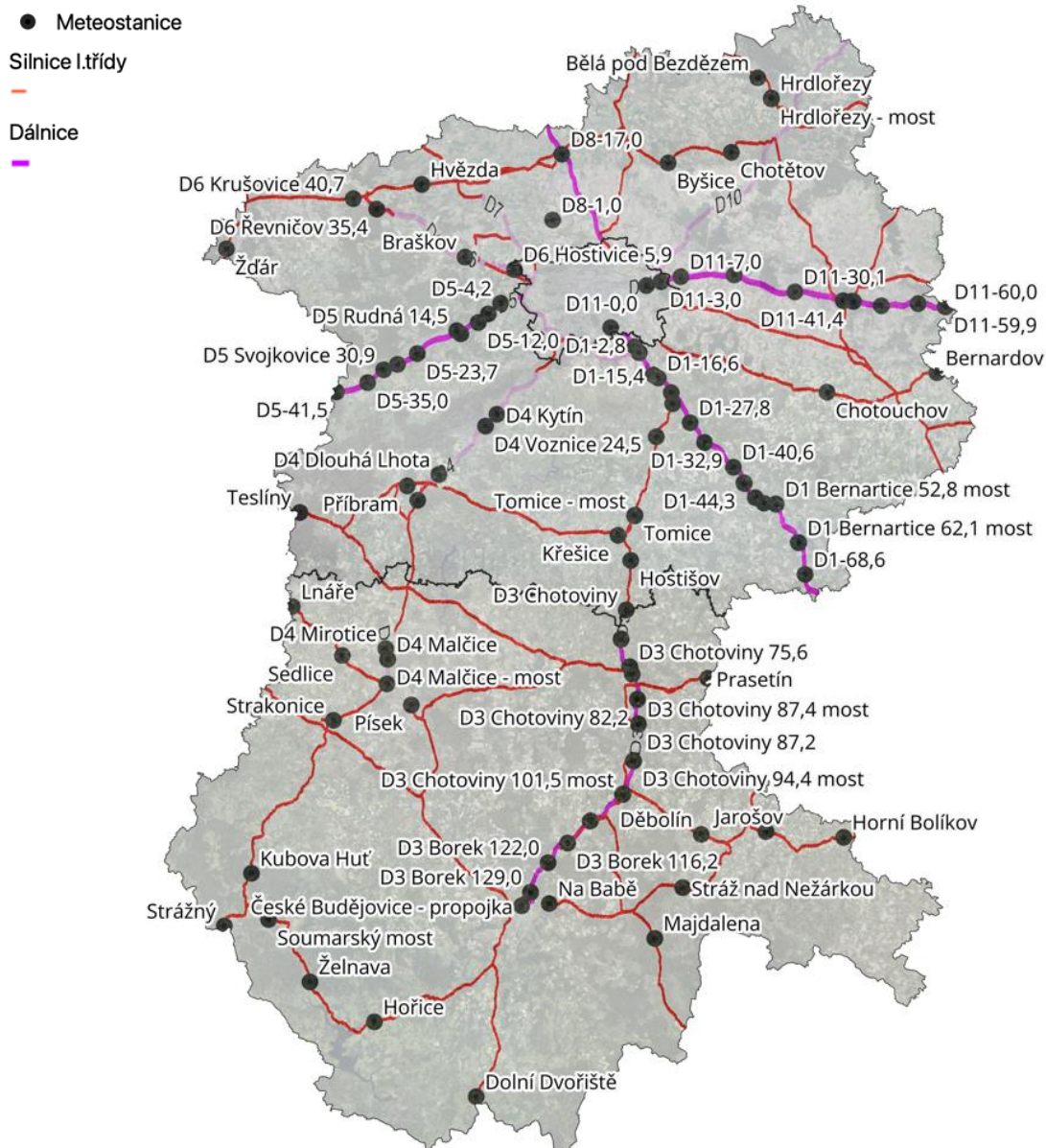
Jednotlivý soubor obsahuje identifikaci v čase, údaje o teplotě vzduchu, rosném bodu, teplotě vozovky, teplotě vozovky v hloubce 5 cm a 30 cm, teplotě namrzání, dále pak údaje o stavu, srážkách, vlhkosti, tlaku, rychlosti větru, směru větru, speciální radiaci a dohlednosti. V tabulce níže jsou jednotlivé sloupce rozepsány i s příslušnými jednotkami. Hodnoty jsou zaznamenávány každých 10 s.

Tabulka 6: Data z meteostanice

Údaj	Jednotky
Teplota vzduchu	°C
Rosný bod	°C
Teplota vozovky	°C
Teplota vozovky v hloubce 5 cm	°C
Teplota vozovky v hloubce 30 cm	°C
Teplota namrzání	°C
Stav	Mokro, Mokro nasoleno, Možnost namrzání, Sníh/led, Sucho, Vlhko, Zbytková sůl
Srážky	mm
Vlhkost	%
Tlak	hPa
Rychlost větru	m/s
Směr větru	°
Specifická radiace	W/m ²
Dohlednost	m

Mimo zmíněné základní meteorologické prvky jsou měřeny i jiné parametry. Teplota povrchu vozovky je stěžejní veličinou v silniční meteorologii. Závisí hlavně na radiační bilanci povrchu vozovky, která je podstatně ovlivněna zastíněním povrchu před přímým slunečním zářením. Teplota povrchu je také podmíněna jinými faktory, jako je teplota vzduchu, padající a usazené srážky, rychlost větru, podpovrchová teplota, chemické ošetření povrchu vozovky apod. Stav vozovky udává informace o stavu vozovky, která může být charakterizována pojmy mokro, mokro nasoleno, možnost namrzání, výskyt sněhu/ledu, sucho, vlhko a výskyt zbytkové soli.

Na dálnicích (D1, D3, D4, D5, D6, D8, D11) je umístěno 75 meteostanic, na silnicích I. třídy pak 40 jednotek, celkem tedy v oblastech Jihočeského a Středočeského kraje 115 stanic. Na mapě níže jsou polohy jednotlivých meteostanic vyobrazeny. Jak lze pozorovat, majorita se nachází na dálnicích. Silniční a dálniční síť ČR byla získána z WMS služby. [19] Zobrazovaná data jsou v majetku a správě Ředitelství silnic a dálnic ČR – odboru silniční databanky a NDIC.



Obrázek 3: Poloha meteorologických stanic.

2.2. Flotila plovoucích vozidel

Floating Car Data, neboli FCD, jsou dopravní informace z plovoucích vozidel. V ČR se nachází minimálně 150 000 plovoucích vozidel, z nichž 75 % musí být osobní vozidla. Systém FCD má několik prvků – vozidla, systém FCD, Národní dopravní informační centrum (NDIC) a uživatele, díky kterým se z dat stanou informace. Vozidla průběžně odesílají data o poloze, směru a rychlosti, musí být proto vybavena GPS jednotkami. Z nich se získávají strukturovaná data, nad kterými se provádí operace jako je agregace, anonymizace, filtrace a verifikace dat. V rámci systému FCD se také data převádí na strukturovaný výstup nad TMC segmenty. Data z FCD a jiné výstupy jsou zpracovány v NDIC, kde se zvyšuje kvalita informací na výstupu, aby mohly být použity pro další informování, detekci kolon, detekci dopravních nehod, omezení provozu apod. Dopravní informace pak jsou k dispozici pro uživatele, kdy se FCD data využívají pro zkvalitnění informací o dopravní situaci, čímž se může zlepšit kvalita řízení dopravy. [20]

FCD lze získat z Národního dopravního informačního centra (NDIC) ŘSD. K již zpracovaným datům mají přístup všichni odběratelé z veřejného i soukromého sektoru, jako jsou města, kraje, Ministerstvo dopravy, složky IZS, akademická obec, výrobci vozidel, soukromé poskytovatelé služeb, navigační služby (například Waze, Google Maps), ICT sektor a jiné. NDIC má tímto k dispozici spolehlivý a garantovaný zdroj dopravních dat, díky kterým mohou v reálném čase sledovat stav dopravního proudu na vybraných dopravně významných komunikacích. Informace pak mohou sloužit k sledování dopadu událostí, ke kterým patří například nehody, uzavírky, modernizace úseku komunikace, a k rychlé identifikaci problémových míst s výrazným zhoršením plynulosti dopravního proudu či snížením rychlosti. [21]

2.2.1. Využití FCD dat

Data mají v rámci dopravy mnoho využití. Mohou zlepšovat bezpečnost a plynulost dopravy, její kvalitnější organizování, jelikož FCD lze využít pro řízení silničního provozu. Konkrétními kroky je například hodnocení bezpečnosti pozemních komunikací, posuzování dopadu uzavírek, analýza chování účastníků provozu při různých dopravních situacích, výpočet ekonomických ztrát z kongescí, analýza dopadu na životní prostředí nebo pozorování důsledků výstavby nových staveb a dopravních řešení (např. výstavba obchvatu). Na rozdíl od komunitních systémů čerpají vývojáři a následní koncoví uživatelé větší množství autorizovaných informací, například o nehodách či uzavírkách. Díky spojení všech informací není viděn jen následek, ale lze určit i příčinu vzniklé situace.

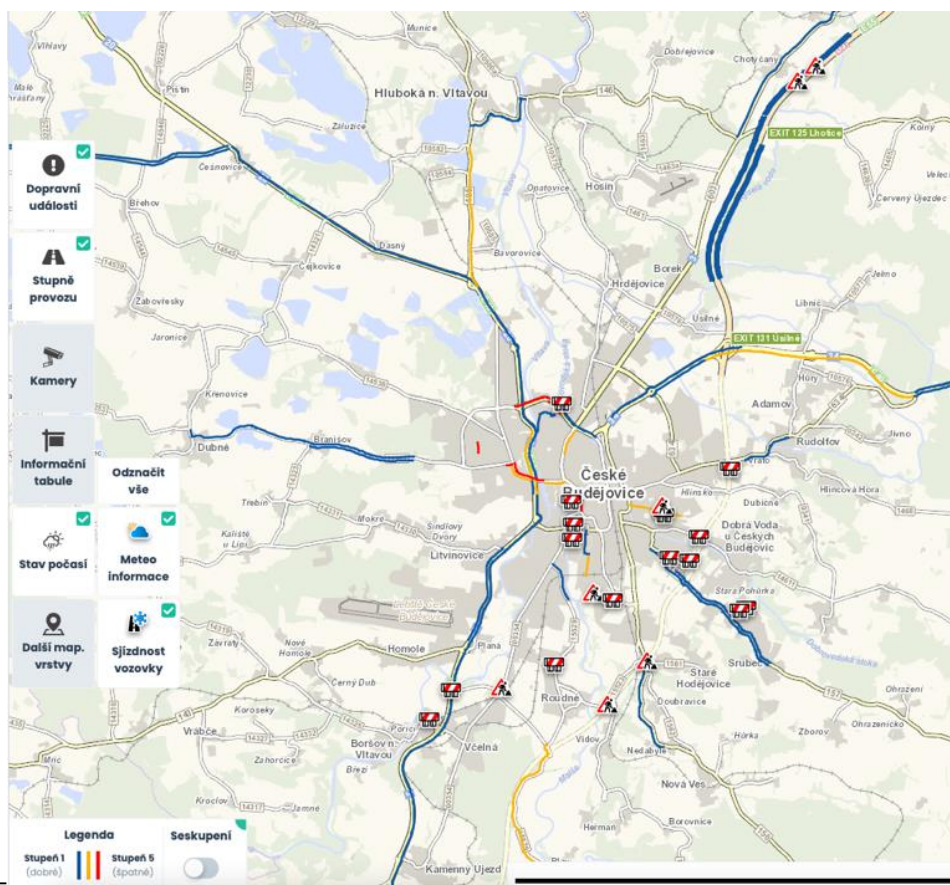
Využití FCD dat lze shrnout do tří skupin:

- I. Informace z plovoucích vozidel slouží jako prostředek k organizaci dopravy z místa A do místa B, k orientaci pro poskytovatele flotilových řešení, jako pomocník subjektů

nabízející služby (například ubytování) v blízkosti komunikací k plánování. Jedná se o základní desku, nad kterou lze postavit business k službám veřejnosti.

- II. K dispozici jsou také data historická, které slouží dopravním expertům a jiným odborným subjektům pro různé analýzy týkající se dopravního proudu a pro vyhodnocování a plánování na základě dopravních toků.
- III. Data z plovoucích vozidel začlenit do pojmu „big data“, jelikož vozidla generují vysoké objemy dat. Big data nacházejí využití v dopravě v podobě predikcí a modelování vývoje dopravy. Díky tomu lze například efektivněji řídit dopravu i na úrovni širšího okolí velkých aglomerací (týká se například Prahy, Brna, Ostravy či Plzně). [21]

Využití FCD dat je možné sledovat na webovém portále dopravniinfo.cz, který poskytuje vizualizaci rychlosti dopravního proudu v reálném čase. Ukázka rozhraní je na obrázku č. 4. Postupně jsou přidávány funkce, například výpočet zdržení na definovaném úseku či výpočet dojezdové doby. [21]



Obrázek 4: Zobrazení aktuální dopravní situace na stránkách dopravniinfo.cz [22]

2.2.2. TMC segmenty

Síť FCD pokrývá dálnice, silnice I. a II. třídy, některé úseky komunikací III. třídy a hlavní průjezdní tahy větších měst. Jedná se celkem o necelých 30 tisíc úseků, které se spadají pod geografický referenční systém používaný v systému TMS (Traffic Message Channel). Funkčnost georeferenčního systému je založena na lokalizačních tabulkách, což je databázové prostředí pro ukládání dat. Poskytovaná data se tedy nevztahují k jednotlivým vozidlům, ale vypovídají o skupině vozidel na daných úsecích, které jsou označeny identifikačním číslem. [21]

Na mapě níže lze vidět uspořádání segmentů v Jihočeském, Středočeském kraji a v Praze.



Obrázek 5: TMC segmenty na území Jihočeského, Středočeského kraje a hlavního města Prahy

2.2.3. Popis FCD dat

Data z flotily plovoucích vozidel byla poskytnuta pro účely práce ŘSD. Výpočty dopravních informací na výstupu se periodicky opakují každou minutu, jsou pravidelně aktualizované a anonymizované. [20]

Vstupní data z vozidla jsou anonymizované ID vozidla, časové razítko pozorování z palubní jednotky, zeměpisná šířka, zeměpisná délka, rychlost vozidla, směr pohybu vozidla, kategorie vozidla a čas zpracování dat pro anonymizaci. [23]

Výstupy z plovoucích vozidel jsou vypsány v tabulce níže. Jedná se o údaje, které obsahují poskytnuté datové soubory.

Tabulka 7: FCD data

Údaj	Jednotky
Aktuální počet plovoucích vozidel osobních	-
Aktuální počet plovoucích vozidel nákladních	-
Aktuální vypočtená rychlost dopravního proudu	km/h
Aktuální čas průjezdu definovaného TMC segmentu	s
Aktuální zpoždění na definovaném segmentu	s
Typická průměrná rychlost při volném průjezdu segmentu	km/h
Doba volného průjezdu	s
Lokalizace kolony na segmentu	-
Délka kolony	m
Vzdálenost čela kolony od počátku segmentu	m
Míra spolehlivosti	-
Reakční doby systému	s
Parametr Level of Service	-

Rychlost dopravního proudu je udávána v km/h za aktuálních podmínek a při volném provozu (1. stupeň provozu). Dojezdové doby jsou měřeny také při stávajících podmínkách a za volného provozu.

Aktuální zpoždění na definovaném segmentu je vyjádřený rozdíl aktuální dojezdové doby oproti době při volném průjezdu. Výpočtem rozdílu doby dostupnosti vstupních dat a doby výpočtu stavové informace se zase dojde k parametru reakční doba systému.

Mezi stavový parametr patří Level of Service vyjadřující stupeň dopravy v rozmezí 1 až 5, kdy 1. stupeň značí volný dopravní proud, zatímco stupeň 5 dopravní kolaps. Stavovými informacemi jsou také data týkající se kolon – lokalizace kolony na segmentu, odhad její délky vzdálenost čela kolony od počátku segmentu.

V tabulce č. 8 jsou jednotlivé stupně dopravy popsány. Interval značí rozmezí, v kterém se pohybuje rozdíl aktuální rychlosti dopravního proudu a rychlosti dopravního proudu na volném segmentu.

Tabulka 8: Stupně dopravy (LOS)

Stupeň dopravy	Popis stavu na komunikaci	Interval
1	Výskyt jednotlivých vozidel, jízda je plynulá	<85 %, 100 %)
2	Výskyt malých skupin vozidel, jízda je plynulá. Na křižovatkách není problém s odbavováním.	<75 %, 85 %)
3	Tvorba proudů vozidel, přesto je provoz plynulý se sníženou průměrnou rychlostí oproti maximální povolené.	<50 %, 75 %)
4	Tvorba kolon vozidel, provoz již není plynulý. Průměrná rychlost je výrazně snížena, průjezd křižovatkami je omezen.	<20 %, 50 %)
5	Dopravní kolaps, stojící či jen pomalu popojíždějící kolony a provoz téměř zastaven s minimální průměrnou rychlostí.	<0 %, 20 %)

Kvalitativně lze datový soubor ohodnotit pomocí míry spolehlivosti a počtu plovoucích vozidel v daném segmentu. Míra spolehlivosti udává míru, se kterou je možné datům důvěřovat. Pohybuje se v intervalu od 0 do 1, kde 1 značí 100% spolehlivost dat. [24]

3. Teoretický popis statistických metod

Cílem statistické analýzy bylo vytvoření analýzy vztahu mezi chováním dopravního proudu a povětrnostními podmínkami jako je teplota, srážky, rychlost větru apod. K analýze chování dopravního proudu vzhledem k meteorologickým podmínkám byla použita vícerozměrná regresní analýza a korelační analýza, které slouží k popisu statistických závislostí. Jejich úkolem je hledání a hodnocení závislostí mezi dvěma statistickými veličinami.

3.1. Vícerozměrná regresní analýza

Vícerozměrná regrese uvažuje n vysvětlujících proměnných a jednu vysvětlovanou proměnnou. Vysvětlující proměnné nabývají hodnot x_{i1} až x_{in} , kde $i = 1, \dots, k$. Mezi proměnnými a vysvětlovanou proměnnou existuje následující vztah.

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_nx_{in} + e_i, \quad i = 1, \dots, k$$

y	závislá proměnná, neboli k – rozměrný sloupcový vektor hodnot vysvětlované proměnné
b_0	konstanta nebo odhadovaný průměr y , když jsou všechny nezávislé proměnné rovny nule
b_1, b_2, b_n ...	regresní koeficienty, které představují změnu v y_i v reakci na jednotkovou změnu v každé nezávislé proměnné, neboli vektor parametrů $b = (b_0, \dots, b_n)^T$
x_1, x_2, x_n ...	hodnoty nezávislých proměnných, neboli $(k+1)$ – rozměrný sloupcový vektor neznámých parametrů
ε	náhodná chyba nebo nevysvětlená variabilita, jež nelze vysvětlit pomocí nezávislých proměnných

Pomocí metody nejmenších čtverců se odhaduje vektor parametrů $b = (b_0, \dots, b_n)^T$ minimalizací funkce. [25][26][27]

Rovnice byla upravena pro potřeby vytvořeného modelu takto:

$$\Delta Y = b_0 + b_t \Delta T + b_{tv} T_v + b_s S + b_v V + b_d D + b_h H + b_p P + b_c C$$

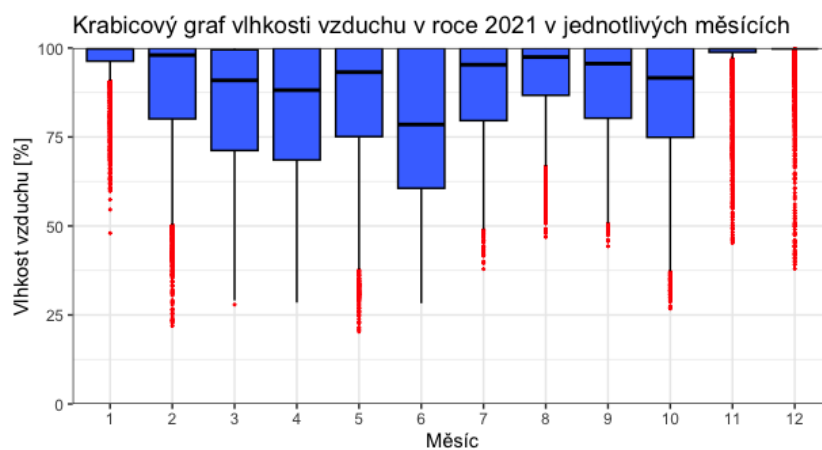
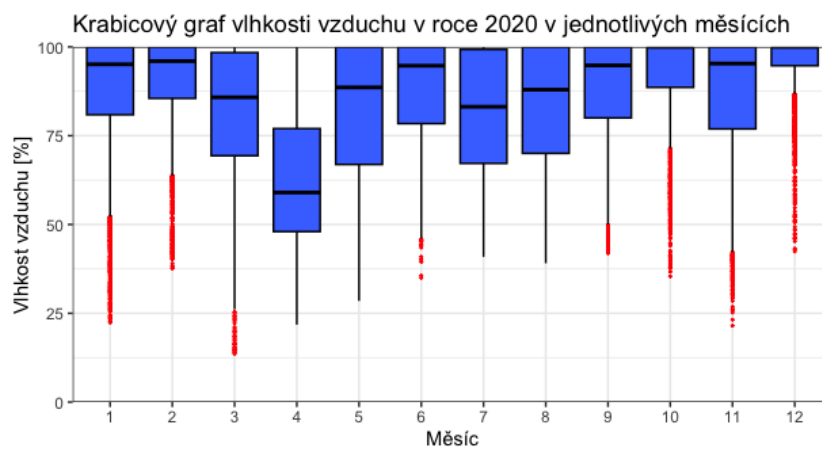
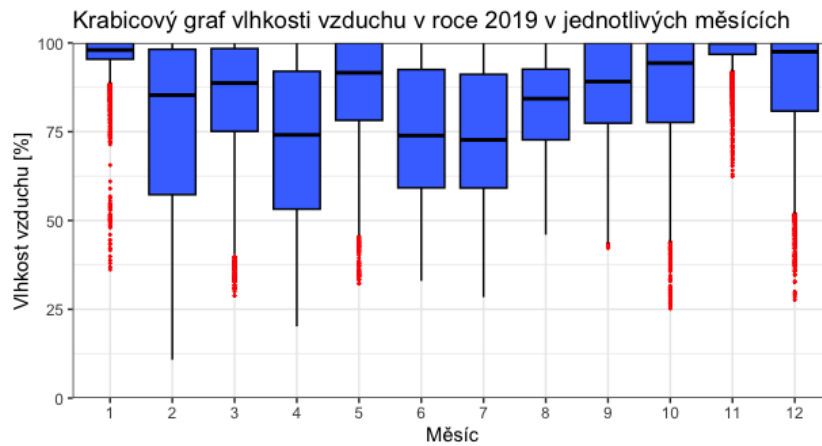
ΔY značí odchylku rychlosti dopravního proudu od průměru, která se vypočítá rozdílem aktuální vypočtené rychlosti dopravního proudu od typické průměrné rychlosti při volném průjezdu segmentu. T je teplota vzduchu, T_v teplota vozovky, S jsou srážky, V vyjadřuje rychlost větru, D dohlednost, H vlhkost vzduchu, P tlak vzduchu a C je stav vozovky.

Vzhledem k podnebí v České republice, které je charakterizováno sezónností, byla proměnná teplota vzduchu brána jako odchylka od měsíčního průměru. Tím se zahrne vliv ročního období a redukuje se variabilita dat, čímž se může zpřesnit predikce a odhad regresního koeficientu. Samozřejmě to s sebou nese i riziko ztráty informací. Původně bylo zamýšleno zachovat se stejně i v případě vlhkosti vzduchu, která jistou charakteristiku odpovídající ročnímu období má. Ale po zobrazení distribuce hodnot vlhkosti v krabicovém grafu nelze jistě říci, že existuje nějaký vzorec chování hodnot vlhkosti vzhledem

k ročnímu období. Krabicové grafy jsou zobrazeny níže (graf č. 1), kde můžeme pozorovat, že výkyvy jsou spíše náhodné a není zde výrazný vliv sezónnosti. Červené hodnoty vyznačují odlehlé hodnoty v jednotlivém měsíci.

Pro teplotu byla odchylka od průměru vypočítána z následujícího vzorce, kde T_{ij} je daná hodnota v určitém i -tou hodinu, j je den v měsíci, n pak počet dní v měsíci.

$$\Delta T_{ij} = T_{ij} - \frac{\sum_{j=1}^n T_{ij}}{n}$$



Graf 1: Krabicové grafy vlhkosti vzduchu (Kubova Huť)

Nezávislými proměnnými v této analýze je teplota vzduchu, teplota a stav vozovky, srážky, rychlost větru, dohlednost, vlhkost, tlak, které byly vybrány z dostupných dat jako významné proměnné. Závislou proměnnou je pak rychlost dopravního proudu. Analýza hodnotí, jakou mírou každá z nezávislých proměnných ovlivňuje rychlost dopravního proudu.

Testování probíhá pomocí testování hypotéz, kdy testujeme nulovou hypotézu H_0 : Mezi proměnnou X a Y neexistuje žádný vzájemný vztah. Alternativní hypotézou pak je H_A : Mezi proměnnou X a Y existuje nějaký vzájemný vztah. K tomu slouží statistika t , která vyjadřuje pravděpodobnost náhodného výskytu.

Na začátku byla také otestována normalita data kvůli odhadu parametrů. Jednotlivé proměnné byly zobrazeny pomocí histogramu, kdy se vizuálně pozorovalo, zda data mají normální rozdělení. Následně byl ještě aplikován tzv. kvantil-kvantilový graf (q-q plot) pro otestování normality. Ten umožňuje graficky zhodnotit, z jakého rozložení data pocházejí, a to tím způsobem, že vykresluje uspořádané vzorky z rozložení poskytnutých dat oproti podobnému počtu uspořádaných kvantilů odebraných z normálního rozložení.

Z podstaty lineárních regresních modelů vyplývá několik předpokladů, které je nutné splňovat pro správný výsledek modelu. Nezávislé proměnné by měly mít normální rozdělení, avšak mnohem důležitější je normální distribuce reziduí modelu s nulovou střední hodnotou a konstantním rozptylem. Vysvětlující proměnné musí splňovat podmínku nezávislosti.

Výstupními údaji z regresní analýzy jsou odhady regresních koeficientů, odhady směrodatných odchylek odhadů, statistika t_j pro testování hypotéz, že se parametr rovná nule, a p -hodnoty. P -hodnota udává nejmenší hladinu významnosti, na které by bylo možné hypotézu zamítnout.

Výstupem z modelu regresní analýzy je také koeficient determinace (R^2), což je statistická míra, která značí, jak velký podíl z celkové variability zkoumaných hodnot lze vysvětlit nezávislými proměnnými. Koeficient nabývá hodnot v intervalu $\langle 0, 1 \rangle$. Čím vyšší hodnota je, tím je model považován za spolehlivější. Koeficient determinace lze vyjádřit i v procentech, kdy vyjadřuje, z kolika procent je závislá proměnná Y ovlivněna nezávislou proměnnou X .

Pro posouzení, zda existuje statisticky významný vztah mezi závislou a nezávislými proměnnými v modelu, byl proveden test významnosti regresního modelu. Provádí se testováním nulové hypotézy, která předpokládá, že se nevyskytuje žádný statisticky významný vztah mezi závislou a nezávislými proměnnými. Provádí se nejčastěji pomocí f -testu.

Vícerozměrná regresní analýza byla provedena v programovacím jazyce R v prostředí RStudio. Výstupy modelu regresní analýzy v RStudiosu jsou následující:

- Odhad regresního koeficientu b_n – vystihuje průměrný vliv jednotkového navýšení proměnné x na proměnnou y , jestliže se ostatní proměnné nemění.

- Odhad reziduální směrodatné odchylky odhadu
- Statistika t – jedná se o hodnotu testovací statistiky z t -testu, která udává poměr mezi odhadnutou hodnotou regresního koeficientu a standardní chybou odhadu. Čím větší hodnota t , tím menší je pravděpodobnost, že výsledky modelu jsou náhodné.
- P -hodnota – údaj zobrazuje, s jak velkou pravděpodobností by k vypočtené statistice t došlo náhodou, pokud by byla pravdivá nulová hypotéza o nulovém účinku parametru. Pokud je p -hodnota nízká, resp. nižší než hladina významnosti (obvykle 0,05), lze vyloučit nulovou hypotézu.
- F -statistika modelu – používá se k testování hypotéz, které říkají, že všechny regresní koeficienty modelu jsou nulové a model tedy nevysvětluje proměnnou y . Jestliže je f -statistika větší, než kritická hodnota, zároveň p -hodnota je menší než hladina významnosti, tak lze zamítnout nulovou hypotézu a přijmout hypotézu alternativní. F -test se používá hlavně v regresních analýzách s parametry. [25][26][27][28]

3.2. Korelační analýza

Jak je zmíněno výše, lineární regresní model předpokládá, že hodnoty jednotlivých proměnných jsou nezávislé od hodnot ostatních proměnných.

Korelační analýza zkoumá vliv mezi dvěma proměnnými. Předpokládá se, že některé meteorologické prvky mohou vykazovat mezi sebou kolinearitu. Příkladem může být například vysoká vlhkost, která může značit výskyt srážek. Byla proto provedena korelační analýza, která zkoumala, jak spolu jednotlivé parametry korelují. Následně byly vyhodnoceny výsledky, kdy jako významná byla posouzena hodnota nad 0,5 a v případě, že byl výsledek v korelační matici vyšší než 0,6, byl jeden z parametrů z modelu vyřazen. [25][29]

4. Analýza vybraných úseků

V rámci analýzy byly vybrány dvě lokality, pro které byly vytvořeny modely. Lokality byly vybrány záměrně tak, aby se od sebe odlišovaly svou charakteristikou a mohl být případně pozorován rozdíl mezi jízdou na dálniční síti a jízdou v odlehlejší oblasti. Datová analýza byla provedena v programovacím jazyku R v prostředí RStudio. Práce s geografickými daty se odehrávala v softwaru QGIS.

4.1. Kubova Huť

Pro analýzu oblasti byla vybrána oblast nacházející se na komunikaci I/4 na 137. km v Jihočeském kraji.

Na obrázku č. 6 a 7 můžeme vidět pozici meteorologické stanice a TMC segmentu. Meteostanice se nachází 990 m n. m., tím je zařazena do horské oblasti a jedná se tedy o nejvýše položenou meteorologickou stanici. Roční průměrná denní intenzita dopravy činila v roce 2020 při Celostátním sčítání dopravy 3 333 vozidel za den v obou směrech. [30]



Obrázek 6: Pozice meteorologické stanice v Kubově Huti



Obrázek 7: Pozice meteorologické stanice v Kubově Hutí z pohledu ČR

4.1.1. Příprava dat

Data z meteostanice byla uložena v souborech s názvem C010, což je identifikační číslo meteostanice, rozdělená podle roků (2019, 2020, 2021). Na začátku proběhlo v příkazovém řádku překódování znaků na obecný formát UTF-8. Data byla následně načtena v programu RStudio, kde proběhlo jejich předzpracování – čištění, filtrování a jejich úprava na správný datový typ. Nulové hodnoty byly buď upraveny průměrem, nebo byly vymazány. Poté byla vytvořena analýza, kdy byla zhotovena deskriptivní statistika a dále proběhlo testování normality dat, korelační analýza a regresní analýza. Odlehlé hodnoty byly ponechány, aby nebyla výsledná statistika zkreslená. Po úpravě se pracovalo s následujícími proměnnými: teplota vzduchu, teplota vozovky, srážky, vlhkost, tlak a rychlost větru. Ostatní proměnné buď nebyly vůbec zaznamenány (dohlednost), nebo nebyly shledány významnými pro danou analýzu. To se týká například rosného bodu, který byl vzhledem k tomu, že úzce souvisí s teplotou vzduchu, vlhkostí a tlakem vzduchu (viz. kap. 1.3.2.), z modelu vyloučen. Pro účely modelu také nebyly shledány relevantní údaje o teplotě vozovky v hloubce 5 a 30 cm, směr větru, radiace a teplota namrzání, kterou lépe vystihuje stav vozovky. Z teploty vzduchu byl vypočítán parametr rozdíl teploty, který určuje, jak moc se teplota odlišuje od průměrné měsíční teploty.

Data z plovoucích vozidel byla k meteostanici přiřazena pomocí TMC segmentu, který se u meteorologické stanice nachází. Stanici C010 odpovídal TMC segment TS23363T26307. Ten byl vyfiltrován v RStudio napojením na SQL databázi FCD dat. Byla tak získána data odpovídající časovému období dat z meteostanice – tj. roky 2019, 2020 a 2021. Data byla také upravena a vyčištěna podle potřeby a podobně bylo nakládáno s nulovými hodnotami, které byly vymazány. Z datové sady nás zajímaly především údaje o aktuální vypočtené rychlosti dopravního proudu a typické průměrné rychlosti při volném průjezdu segmentu. Z rozdílu těchto dvou parametrů byla vytvořena nová proměnná, rozdíl rychlosti, která určuje změnu rychlosti. Z ostatních parametrů byl extrahován také

sloupec určující míru spolehlivosti a stupeň dopravy, které však sloužily spíše jen pro orientaci v daných hodnotách.

Datová sada z meteorologické stanice pak byla napojena na data z plovoucích vozidel pomocí časového razítka, vytvořil se tak nový datový rámec s 28 426 řádky. Hodnoty proměnných byly převážně spojité veličiny, kategoriální byl pouze stav vozovky, který nabýval hodnot mokro, mokro nasoleno, možnost namrzání, sníh/led, sucho, vlhko a zbytková sůl. Aby se s ním mohlo pracovat v rámci modelu, byl upraven do matice, kdy se z jednotlivých stavů vytvořily sloupce a ty pomocí 0 a 1 určovaly, jaký stav vozovky v daný čas nastává.

Z dat byly provedeny vybrané deskriptivní charakteristiky. V následujících tabulkách č. 9, 10 a 11 jsou hodnoty statistických veličin, jako je průměr, minima, maxima či četnosti.

Tabulka 9: Statistické veličiny meteorologických dat (Kubova Huť)

Parametr	Rok	Minimum	Maximum	Průměr	Medián
Teplota [°C]	2019	-14,4	29,5	6,14	5,4
	2020	-11,2	27,6	6,99	6,5
	2021	-14,9	28	5,75	5,2
Rozdíl teplot [°C]	2019	-13,6	15,47	0	-0,12
	2020	-15,02	13,41	0	-0,11
	2021	-14,6	18,52	0	-0,46
Teplota vozovky [°C]	2019	-12,9	49,4	8,47	6,3
	2020	-8,3	45,4	9,75	8,1
	2021	-10,2	48	10,11	8,3
Srážky [mm]	2019	0	19,3	0,01	0
	2020	0	18,5	0,01	0
	2021	0	12,2	0,01	0
Vlhkost [%]	2019	10,8	100	84,03	91,5
	2020	13,6	100	84,94	92,8
	2021	20,3	100	88,39	97,4
Tlak [hPa]	2019	855,1	918,6	887,11	887,8
	2020	853	912,5	888,37	889,7
	2021	862,8	906,1	887,78	889,1
Rychlost větru [m/s]	2019	0	8,1	1,86	1,6
	2020	0	8,7	1,93	1,7
	2021	0	8,9	1,82	1,6

Tabulka 10: Statistické veličiny z dat plovoucích vozidel (Kubova Huť)

Rok	Průměrná rychlost vypočtená	Průměrná rychlost typická	Průměrná míra spolehlivosti	Průměrný stupeň dopravy
2019	65,148	82,374	0,827	2,046
2020	65,396	85,687	0,84	2,261
2021	65,127	85,846	0,814	2,284

Stav vozovky byl vzhledem ke kategoriím popsán pomocí četností (tabulka č. 11).

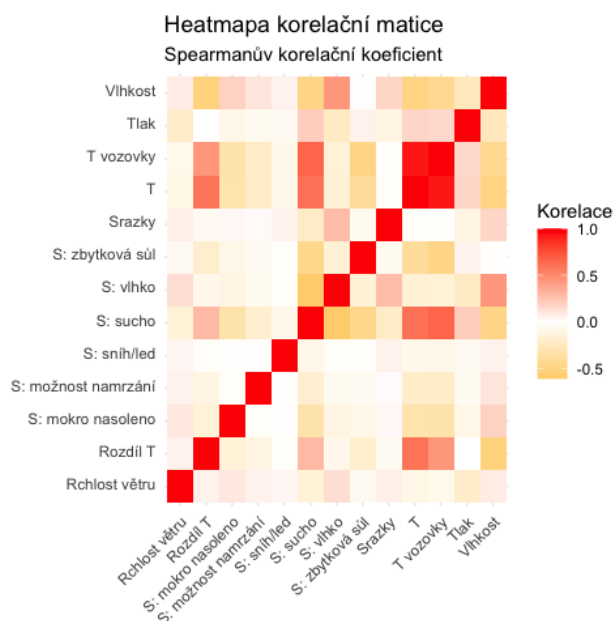
Tabulka 11: Statistické vyhodnocení stavu vozovky (Kubova Huť)

Rok	Stav vozovky	Četnost	Relativní četnost [%]
2019	mokro	35	0,07
	mokro nasoleno	3663	6,95
	možnost namrzání	1657	3,14
	sníh/led	657	1,25
	sucho	31469	59,72
	vlhko	8679	16,47
	zbytková sůl	6534	12,4
	2020	mokro	53
mokro nasoleno		3796	7,24
možnost namrzání		1707	3,26
sníh/led		165	0,31
sucho		25813	49,24
vlhko		12668	24,17
zbytková sůl		8216	15,67
2021		mokro	633
	mokro nasoleno	3766	7,25
	možnost namrzání	1514	2,91
	sníh/led	164	0,32
	sucho	34982	67,33
	vlhko	8507	16,37
	zbytková sůl	2390	4,6

4.1.2. Výsledky metod

4.1.2.1. Korelační analýza

Korelační analýza byla provedena pomocí Spearmanova testu nezávislosti, který byl vybrán vzhledem k přítomnosti odlehlých hodnot, ale pro porovnání byla provedena i parametrická korelace pomocí Pearsonova korelačního koeficientu. Byla vytvořena korelační matice, jejíž hodnoty pak byly vyhodnoceny. Korelační koeficient může nabývat hodnot od $\langle -1,1 \rangle$, kdy hraniční hodnoty znamenají velmi silnou závislost, 0 pak absolutní nezávislost. Pokud byla hodnota vyšší než 0,5, tak byla považována jako významná. Níže jsou výsledky z korelační matice se Spearmanovým testem nezávislosti – heatmapa (obrázek č. 8), která graficky zobrazuje korelaci jednotlivých veličin, a tabulkové hodnoty korelací s hodnotou nad 0,5 (tabulka č. 12).



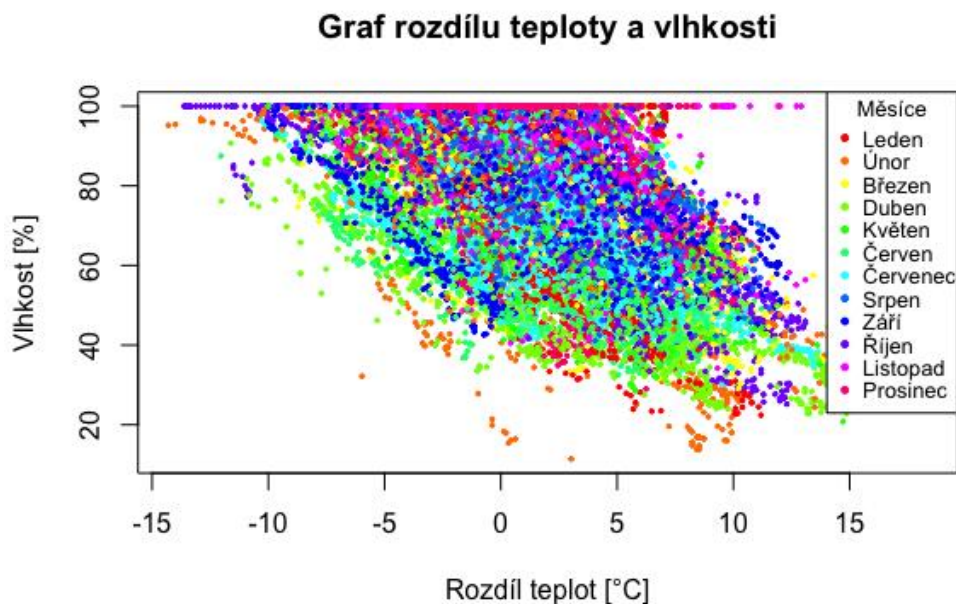
Obrázek 8: Heatmapa korelační matice (Kubova Huť)

Tabulka 12: Výsledky korelační matice se Spearmanovým testem nezávislosti (Kubova Huť)

	Teplota vozovky	Vlhkost	Stav vozovky suchá	Rozdíl teplot			
Teplota	0,93	-0,50	0,60	0,60	Stav vozovky suchá	0,70	-0,61
					Rozdíl teplot		-0,53

Přestože s parametrem teplota nebylo dále počítáno, byl pro zajímavost do matice začleněn. Jak můžeme vidět, suchý stav vozovky silně koreluje s teplotou vozovky a vlhkým stavem vozovky. Proto v modelu se tyto dvě proměnné se suchým stavem vozovky nepotkávají.

Rozdíl teplot také významně koreluje s vlhkostí, přesto nebyl ani jeden z parametrů v modelu odebrán, jelikož byla na začátku zvolena hraniční mez 0,6 pro vyloučení z modelu. Graficky je znázorněn vztah rozdílu teplot a vlhkosti, který byl významný a můžeme zde pozorovat i jak se vztah vyvíjí v jednotlivých měsících (graf č. 2).



Graf 2: Graf parametrů rozdíl teplot a vlhkost (Kubova Huť)

4.1.2.2. Normalita dat

Vzhledem k lineární regresi je vhodné se také zabývat normalitou dat. Data byla vyobrazena pomocí histogramů, které jsou proloženy křivkou normálního rozložení, která je vypočítána z poskytnutých dat. Jak můžeme vidět na vytvořených histogramech (graf č. 3), jednotlivé parametry mohou odpovídat normálnímu rozdělení s určitou křivostí. Tvar histogramu srážek je daný odlehlými hodnotami, které neodpovídají průměru, což může být dáno tím, že některé hodnoty se srážkami byly kvůli NA hodnotám promazány. Histogram stavu vozovky je klasický histogram s četnostmi na ose y, jelikož stav je z kategorického rozdělení.

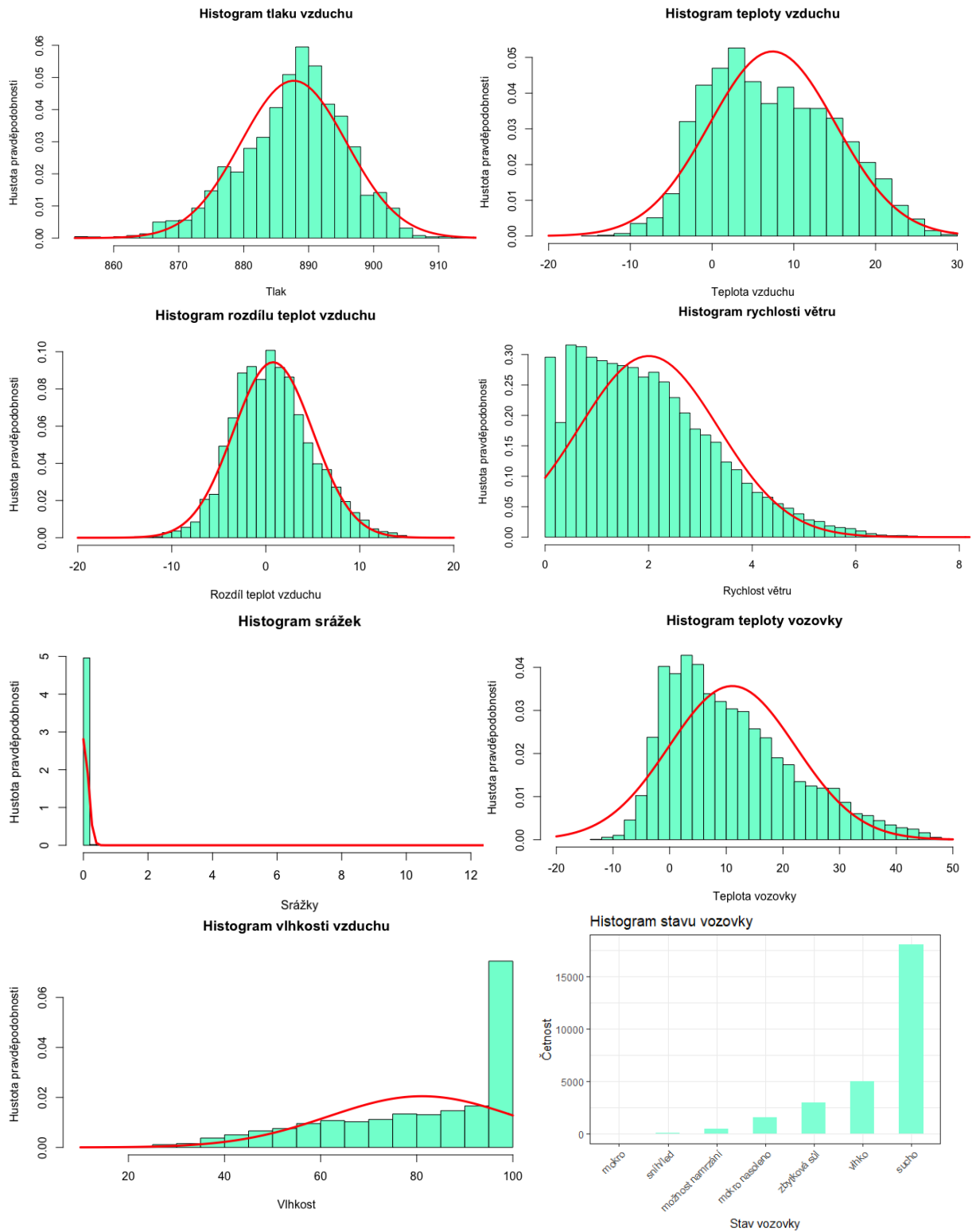
V tabulce č. 13 jsou vypočteny koeficienty šikmosti a špičatosti. Koeficient šikmosti značí míru asymetrie distribuce. Pokud je rozložení dokonale symetrické, tak je koeficient roven nule. Kladné zešikmení značí, že odlehlé hodnoty se nacházejí více napravo od průměru než vlevo. U záporného zešikmení je tomu naopak – odlehlé hodnoty leží od průměru vlevo.

Koeficient špičatosti udává rozdíl rozložení hodnot oproti normálnímu rozdělení. Koeficient špičatosti normálního rozdělení se rovná třem. Pokud je menší než tři, jako je tomu u teploty a vlhkosti, znamená to, že data obsahují méně extrémních odlehlých hodnot než u normálního rozdělení. Pokud je koeficient větší než tři, znamená to, že data obsahují více odlehlých hodnot, než je běžné u normálního rozdělení.

V našem případě nejsou hodnoty koeficientů šikmosti a špičatosti příliš vzdálené hodnotám normálního rozdělení a odchylky jsou pravděpodobně dané odlehlými hodnotami. Výrazně jiné hodnoty jsou u srážek. Pozitivní koeficient šikmosti značí, že je rozdělení zkreslené směrem doprava, což odpovídá hodnotám v datovém souboru, kde je větší počet nižších údajů a několik ojedinělých vysokých hodnot. Koeficient šikmosti tedy neodpovídá normálnímu rozdělení. Koeficient špičatosti u srážek byl vypočten na 3 626,83, což představuje vyšší koncentraci hodnot kolem střední hodnoty a vyšší pravděpodobnost odlehlých hodnot. Celkově jsou srážková data zešikmená doprava s vysokou špičatostí. Je to dané tím, že v datech existuje mnoho hodin bez zaznamenaných srážek a případy, kdy jsou zaznamenané, se stávají odlehlými hodnotami. [28][31]

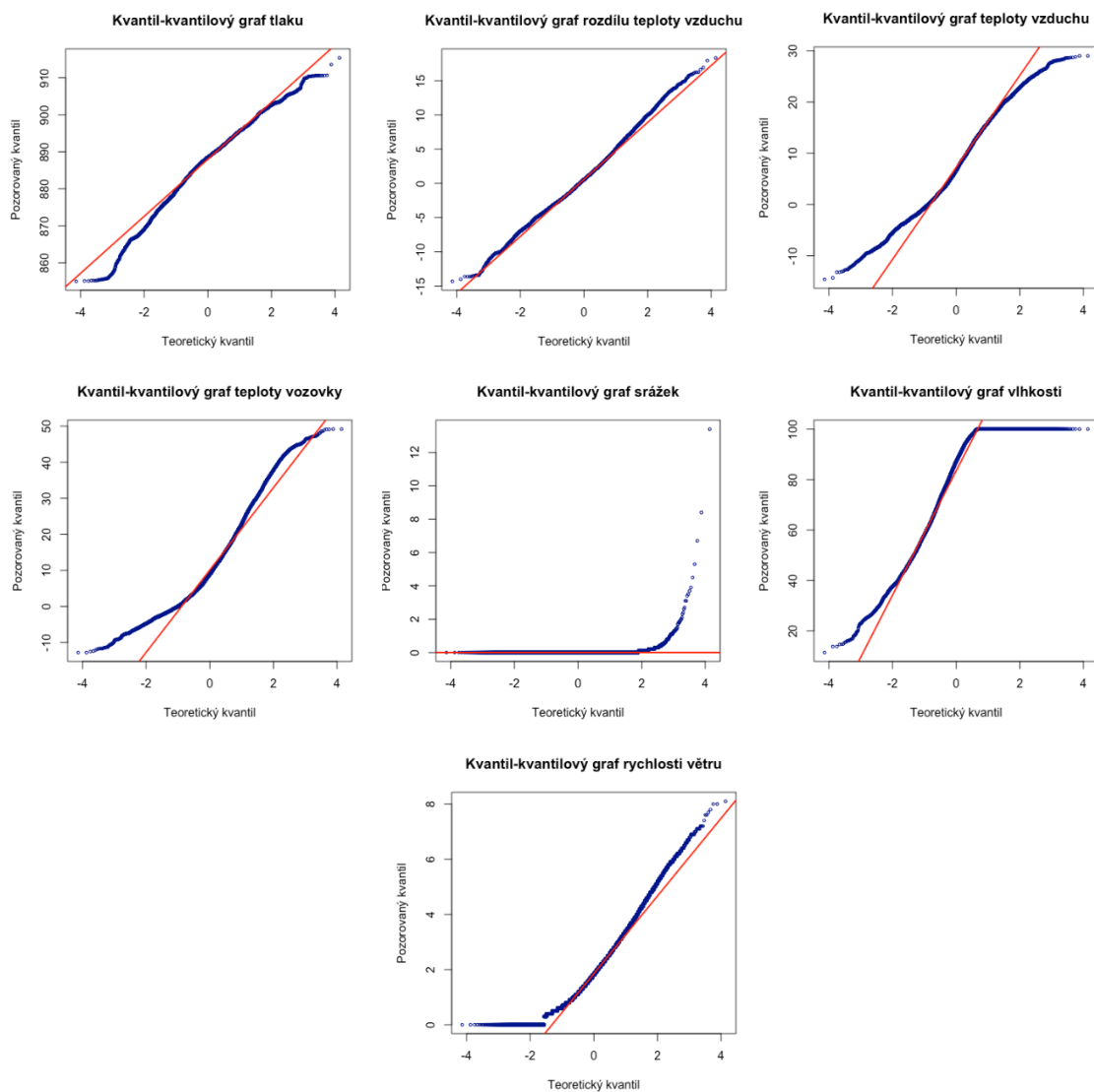
Tabulka 13: Koeficienty šikmosti a špičatosti (Kubova Huť)

Koeficient	Tlak	Teplota	Rozdíl teploty	Rychlost větru	Srážky	Teplota vozovky	Vlhkost
Šikmosti	-0,44	0,25	0,29	0,71	49,32	0,77	-0,81
Špičatosti	3,32	2,3	3,18	3,21	3 626,83	3,03	2,59



Graf 3: Histogramy (Kubova Huť)

Pro lepší porovnání normality dat byly parametry zobrazeny ještě pomocí kvantil-kvantilového grafu (také q-q plot, viz. graf č. 4). Na ose x jsou zobrazeny kvantily teoretického rozdělení (v tomto případě normální), zatímco na osy y jsou kvantily pozorovaného rozdělení. Pokud jsou data normálně rozdělená, kopírují přímku (na grafu červeně). Nejblíže normálnímu rozdělení je parametr rozdílu teplot, který i z histogramu odpovídá normální distribuci. Rovné ocásky, které můžeme pozorovat u grafu vlhkosti, rychlosti větru či srážek, jsou dané celými, často opakujícími se hodnotami, které působí diskrétně, zároveň také odlehlými hodnotami. Z grafů tedy můžeme potvrdit, že některé parametry se podobají normálnímu rozdělení, některé méně. Přesto to není překážkou pro lineární regresní model, který předpokládá normalitu reziduí. [28]



Graf 4: Kvantil–kvantilové grafy (Kubova Huť)

4.1.2.3. Regresní analýza

Regresní analýza byla provedena pomocí dvou modelů. Jelikož stav vozovky je kategoričtý, po převedení pak nabývají jednotlivé stavy binárních hodnot, bylo nutné pro hodnoty stavu vozovky vytvořit jiný model, který reflektuje binární data. Celkově analýza probíhala na vzorku 28 426 dat.

Model 1

V RStudiosu byl vytvořen lineární model, kdy původní zamýšlená regresní rovnice byla upravena pro tento model následovně:

$$\Delta Y = b_0 + b_t \Delta T + b_{tv} T_v + b_s S + b_v V + b_h H + b_p P$$

V modelu tedy je 6 proměnných – rozdíl teploty vzduchu, teplota vozovky, srážky, rychlost větru, vlhkost a tlak. Výsledkem modelu je následující tabulka (tabulka č. 14).

Tabulka 14: Výsledky regresního modelu 1 (Kubova Huť)

	Odhad parametru	Odhad směrodatné odchylky	Statistika t	P hodnota
Intercept	-93,29473	6,918455	-13,485	$< 2 \cdot 10^{-16}$
Rozdíl teplot	0,150829	0,017973	8,392	$< 2 \cdot 10^{-16}$
Teplota vozovky	-0,047293	0,006253	-7,563	$4,07 \cdot 10^{-14}$
Srážky	-1,105072	0,41247	-2,679	0,00739
Vlhkost	-0,044823	0,004018	-11,154	$< 2 \cdot 10^{-16}$
Tlak	0,089142	0,007687	11,597	$< 2 \cdot 10^{-16}$
Rychlost větru	0,075455	0,044838	1,683	0,09242

Výsledek modelu udává mimo jiné hodnota F -statistiky, v tomto případě 117,5 na 4 a 28 421 volných stupňů volnosti. Kritická hodnota f -testu se vypočítala pro tyto stupně volnosti a porovnála se s výsledkem, který byl mnohem vyšším, tím pádem lze říci, že model disponuje vysvětlovací schopností s pomocí minimálně jedné vysvětlovací proměnné a alespoň částečně objasňuje stav hodnot rozdílu aktuální rychlosti dopravního proudu oproti průměrné rychlosti. Kritická mez pro f -test je v tomto modelu 0,1776767.

P -hodnota modelu je zároveň se svojí hodnotou menší než $2 \cdot 10^{-16}$ nižší než určená hladina významnosti (0,05), lze tedy uznat, že alespoň jeden regresní koeficient je statisticky významně odlišný od nuly.

Z výsledků modelu můžeme vyhodnotit následující:

Pro proměnnou Rozdíl teploty

- Odhad regresního koeficientu pro proměnnou rozdíl teploty vzduchu je 0,150829 s odhadem směrodatné odchylky 0,017973, kdy 95% interval spolehlivosti pro parametr se nachází v rozmezí 0,1156 až 0,18606, což značí, že s 95% pravděpodobností regresní koeficient rozdílu teplot náleží tomuto intervalu. Hodnota regresního koeficientu 0,150829 znamená, že pro každé jednotkové zvýšení v hodnotách rozdílu teplot se očekává průměrné zvýšení rozdílu

rychlosti dopravního proudu o 0,150829, za předpokladu, že jsou všechny ostatní vysvětlující hodnoty konstantní. Statistika t s hodnotou 8,392 udává, že odhadnutý regresní koeficient je 8,392krát větší než standardní chyba tohoto odhadu a napovídá, že pravděpodobnost náhodného výskytu je na nízké úrovni. Z p -hodnoty můžeme vyhodnotit na hladině významnosti zamítnutí nulové hypotézy, že regresní koeficient je nulový, jelikož p -hodnota je menší než $2 \cdot 10^{-16}$. Tento regresní koeficient je tedy statisticky významný a lze říci, že existuje vztah mezi vysvětlující a vysvětlovanou proměnnou.

Pro proměnnou Teplota vozovky

- Odhad regresního koeficientu pro proměnnou teplota vozovky je -0,047293 s odhadem směrodatné odchylky 0,006253, kdy 95% interval spolehlivosti pro parametr se nachází v rozmezí -0,05955 až -0,03504. Hodnota regresního koeficientu -0,047293 znamená, že pro každé jednotkové snížení v hodnotách teplot vozovky se očekává průměrné zvýšení rozdílu rychlosti dopravního proudu o 0,047293, za předpokladu, že jsou všechny ostatní vysvětlující hodnoty konstantní. Statistika t s hodnotou -7,563 udává, že odhadnutý regresní koeficient je -7,563krát menší než standardní chyba tohoto odhadu. Tato hodnota je relativně daleko od nuly, zároveň je p -hodnota menší než hladina významnosti. Parametr teplota vozovky je proto shledána statisticky významný, i když vliv na vysvětlovanou proměnnou je vzhledem k odhadu parametru nepatrný.

Pro proměnnou Srážky

- Odhad regresního koeficientu pro proměnnou srážky je -1,105072 s odhadem směrodatné odchylky 0,412470, kdy 95% interval spolehlivosti pro parametr se nachází v rozmezí -1,91353 až -0,29661. Hodnota regresního koeficientu -1,105072 znamená, že pro každé jednotkové snížení v hodnotách srážek se očekává průměrné zvýšení rozdílu rychlosti dopravního proudu o 1,105072, za předpokladu, že jsou všechny ostatní vysvětlující hodnoty konstantní. Statistika t s hodnotou -2,679 udává, že odhadnutý regresní koeficient je -2,679krát menší než standardní chyba tohoto odhadu. P -hodnota 0,00739 není v tomto případě tolik nízká, jako v předešlých případech, přesto je nižší než hladina významnosti. Odhad regresního koeficientu je v tomto modelu nejvyšší, bohužel výsledný t -hodnota může naznačovat poněkud vyšší pravděpodobnost náhodného výskytu. Přesto můžeme tvrdit, že existuje důkaz pro zamítnutí nulové hypotézy o nevýznamnosti regresního koeficientu.

Pro proměnnou Vlhkost

- Odhad regresního koeficientu pro proměnnou vlhkost je -0,044823 s odhadem směrodatné odchylky 0,004018, kdy 95% interval spolehlivosti pro parametr se nachází v rozmezí -0,0527 až -0,03695. Hodnota regresního koeficientu -0,044823 značí, že pro každé jednotkové snížení hodnoty vlhkosti se očekává průměrné zvýšení rozdílu rychlosti dopravního

proudu o 0,004018, za předpokladu, že jsou všechny ostatní vysvětlující hodnoty konstantní. Statistika t s hodnotou -11,154 sice předvídá menší pravděpodobnost náhodného výskytu a spolu s p -hodnotou, která je menší než $2 \cdot 10^{-16}$, naznačují, že je vlhkost statisticky významná, přesto kvůli nízkému odhadu regresního koeficientu nelze proměnné přisuzovat velký vliv na změnu rychlosti.

Pro proměnnou Tlak

- Odhad regresního koeficientu pro proměnnou tlak je 0,089142 s odhadem směrodatné odchylky 0,007687, kdy 95% interval spolehlivosti pro parametr se nachází v rozmezí 0,07408 až 0,10421. Hodnota regresního koeficientu 0,089142 znamená, že pro každé jednotkové zvýšení hodnoty tlaku se průměrně sníží rozdíl rychlosti dopravního proudu o 0,089142, za předpokladu, že jsou všechny ostatní vysvětlující hodnoty konstantní. Statistika t s hodnotou 11,597 udává, že odhadnutý regresní koeficient je 11,597krát větší než standardní chyba tohoto odhadu. Z p -hodnoty můžeme vyhodnotit na hladině významnosti zamítnutí nulové hypotézy, že regresní koeficient je nulový, jelikož p -hodnota je menší než $2 \cdot 10^{-16}$. Tento regresní koeficient je tedy statisticky významný a může vysvětlovat vysvětlovanou proměnnou, přesto vzhledem k velikosti odhadu parametru nelze proměnnou považovat za relevantní vysvětlující proměnnou.

Pro proměnnou Rychlost větru

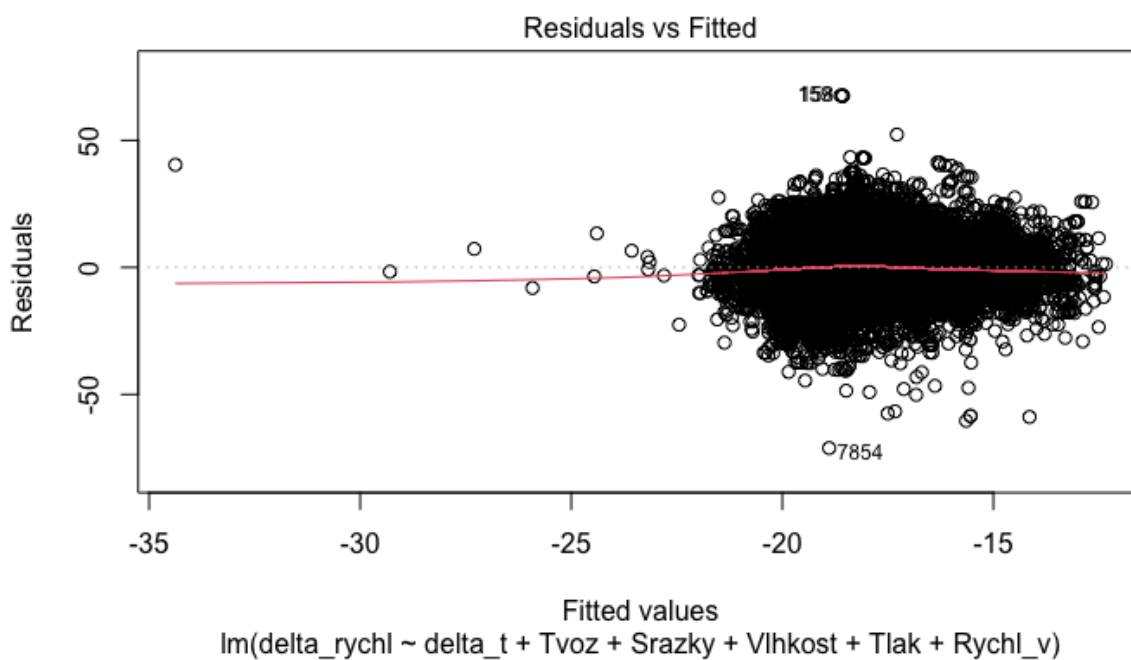
- Regresní koeficienty u proměnné rychlost větru je 0,075455 s odhadem směrodatné odchylky 0,044838, kdy 95% interval spolehlivosti pro parametr se nachází v intervalu od -0,01243 do 0,16334. Statistika t s hodnotou 1,683 již napovídá o trchu větší pravděpodobnost náhodného výskytu. P -hodnota je v tomto případě s hodnotou 0,09242 větší než hranice hladiny významnosti 0,05. Vzhledem k p -hodnotě v kombinaci s nízkým regresním koeficientem a nízkou hodnotou t nelze s jistotou vyloučit nulovou hypotézu.

Celkově tedy lze říci, že na hladině významnosti 0,05 na vysvětlující proměnnou nemá vliv rychlost větru. Nepatrný vliv má tlak vzduchu, kdy hodnota regresního koeficientu tlaku (0,089142) je oproti teplotě vzduchu (0,150829) téměř poloviční, a dále také teplota vozovky spolu s vlhkostí vzduchu. Největší vliv z posuzovaných proměnných na rozdíl od rychlosti větru mají dle modelu srážky, na druhou stranu u nich je statistika t nejbližší k nule, což zvyšuje pravděpodobnost náhodného výskytu. Teplota vzduchu je hned po srážkách proměnná s nejvyšším odhadem regresního koeficientu a i t -hodnota napovídá, že to pravděpodobně nebude náhodné.

Koeficient determinace R^2 se rovná 0,02095, přičemž může nabývat hodnot od 0 do 1 a čím blíže je 1, tím lepší vysvětlovací schopnost model má. V našem případě z toho vyplývá, že vybrané proměnné v modelu vysvětlují jen malou část variability hodnot závislé proměnné. Upravená hodnota koeficientu, která zohledňuje počet regresních proměnných a velikost vzorku dat, vyšla na 0,02074. Hodnota je velmi

podobná jako u normálního R^2 , což může znamenat, že v modelu se nenachází žádné zbytečné regresní proměnné.

S výsledkem modelu bylo možné také zobrazit, jestli byl dodržen předpoklad normálního rozdělení reziduí. Jak můžeme na grafu č. 5 vidět, jsou tam zobrazeny rezidua proti predikovaným hodnotám. Rezidua jsou náhodně rozptýlena kolem nuly, nelze tam pozorovat nějaký systematický vzorec.



Graf 5: Výsledek z modelu – grafické znázornění reziduí (Kubova Huť)

Po vyřazení proměnných, které byly statisticky nevýznamné či jejich odhad regresního koeficientu byl nepatrný, byl vytvořen upravený model s následujícími výsledky vypsány v tabulce č. 15.

$$\Delta Y = b_0 + b_t \Delta T + b_s S$$

Tabulka 15: Výsledky upraveného regresního modelu (Kubova Huť)

	Odhad parametru	Odhad směrodatné odchylky	Statistika t	P hodnota
Intercept	-18,21775	0,05972	-305,06	$< 2 \cdot 10^{-16}$
Rozdíl teplot	0,21683	0,01386	15,65	$< 2 \cdot 10^{-16}$
Srážky	-1,66245	0,41350	-4,02	$5,83 \cdot 10^{-5}$

Z výsledků lze pozorovat, že obě hodnoty mají nízkou p -hodnotu, zároveň se mírně zvýšily odhady parametrů. Odhady obou proměnných se zvýšily zhruba o polovinu, u srážek i rozdílu teploty se pak významněji zvýšila hodnota statistiky t . Z toho lze usoudit, že se zmenšila pravděpodobnost, že výsledky modelu jsou náhodné. Na druhou stranu došlo k snížení koeficientu determinace na hodnotu 0,009158, takže model vysvětluje velmi malou část variability proměnné.

Model 2

Byl vytvořen také model zohledňující stav vozovky, jehož cílem bylo zjistit, jak jednotlivé proměnné ovlivňují pravděpodobnost snížení rychlosti. Změna rychlosti byla upravena tak, aby odpovídala alternativnímu rozdělení, což je speciální druh binomického rozdělení, které nabývá jen dvou hodnot. Vzhledem k povaze stavů vozovky – mokro, mokro nasoleno, možnost namrzání, sníh/led, sucho, vlhko a zbytková sůl – nás spíše zajímalo, kdy se rychlost sníží. Byla určena hraniční hodnota změny rychlosti o 20 km/h oproti typické průměrné rychlosti při volném průjezdu.

Pro tento případ byl použit zobecněný lineární model, který umožňuje modelovat data s binomickým rozdělení. V modelu byly zadány proměnné – stav mokro, stav mokro nasoleno, stav možnost namrzání, stav sníh/led, stav vlhko a stav zbytková sůl. Stav sucho byl kvůli vysoké korelační hodnotě se stavem vlhko z modelu odstraněn.

Upravená rovnice modelu vypadá následovně:

$$\Delta Y = b_0 + b_m C_m + b_{mn} C_{mn} + b_s C_s + b_v C_v + b_z C_z$$

Výsledné hodnoty nalezneme v následující tabulce č. 16.

Tabulka 16: Výsledky regresního modelu (Kubova Huť)

	Odhad parametru	Odhad směrodatné odchylky	Statistika z	P hodnota
Intercept	-0,38461	0,01512	25,429	$< 2 \cdot 10^{-16}$
Mokro nasoleno	0,36857	0,05193	7,098	$1,27 \cdot 10^{-12}$
Možnost namrzání	0,72044	0,08898	8,096	$5,67 \cdot 10^{-16}$
Sníh/led	1,52767	0,24531	6,228	$4,74 \cdot 10^{-10}$
Vlhko	0,56040	0,03218	17,414	$< 2 \cdot 10^{-16}$
Zbytková sůl	-0,11240	0,04033	2,787	0,00532

Intercept v tabulce znamená regresní koeficient b_0 , hodnota 0,18232 nastává, pokud jsou všechny nezávislé proměnné rovny nulové. V tomto modelu představuje logaritmus poměru pravděpodobností výskytu daného jevu, pokud jsou všechny ostatní parametry v modelu rovné nule.

Oproti předchozímu modelu se zde nachází místo statistiky t statistika z pro test hypotézy, že odhad regresního koeficientu je nulový. Vypočítává se jako podíl odhadnutého koeficientu a standardní chyby odhadu koeficientu.

Odhady udávají logaritmus poměru pravděpodobnosti, když se hodnota závislé proměnné změní o jednotku. Všechny hodnoty jsou udávány na hladině významnosti 0,05.

Pro jednotlivé nezávislé proměnné bylo vyhodnoceno:

Pro stav vozovky Mokro nasoleno

- Odhadnutý regresní koeficient je 0,36857 a podle p -hodnoty je statisticky významný. Z hodnota je 7,098, což udává, o kolik je odhadnutý koeficient vzdálený od nulového koeficientu, který říká, že proměnná nemá vliv na závislou proměnnou. V tomto případě nám kombinace vyšší hodnoty z a naopak nízké p -hodnoty dovoluje zamítnout nulovou hypotézu a přijmout alternativní hypotézu.

Pro stav vozovky Možnost namrzání

- Odhadnutý regresní koeficient má hodnotu 0,72044, což je v modelu druhý nejvyšší odhad. Modelem byla vypočtena také vyšší hodnota statistiky z a nízká p -hodnota, i zde proto můžeme vyhodnotit, že parametr má vliv na výsledek modelu a je statisticky významný.

Pro stav vozovky Sníh/led

- Nejvyšší odhadnutý regresní koeficient v tomto modelu připadá stavu sníh/led s hodnotou 1,52767. P -hodnota nižší než hladina významnosti a vyšší výsledek z -testu nám umožňuje odmítnout nulovou hypotézu.

Pro stav vozovky Vlhko

- Odhadnutý regresní koeficient je 0,56040 podle p -hodnoty také statisticky významný s odhadem směrodatné odchylky 0,03218, tím pádem má z modelu nejvyšší hodnotu z , která naznačuje, že odhad regresního koeficientu je s nízkou chybou.

Pro stav vozovky Zbytková sůl

- Odhadnutý regresní koeficient je -0,11240 a statistika z je z modelu nejbližší nule. V tomto případě se p -hodnota nejvíce blíží hladině významnosti, a přestože i u tohoto parametru můžeme vyloučit nulovou hypotézu, je stav vozovky se zbytkovou solí nejslabším článkem modelu vzhledem k vlivu na snížení rychlosti o více než 20 km/h oproti typické.

U všech proměnných můžeme vyloučit nulovou hypotézu a říci, že jednotlivé parametry mají vliv na snížení rychlosti o více než 20 km/h oproti typické průměrné rychlosti při volném průjezdu. Největší regresní koeficient byl odhadnut u stavu vozovky sníh/led, dále pak u namrzání a u vlhko. Nižší odhad oproti vlhké vozovce je u stavu mokro nasoleno. Nejnižší odhad regresního koeficientu s nejnižší p -hodnotou je u stavu zbytková sůl.

Výsledky modelu udávají nulovou a reziduální odchylku. Nulová odchylka vystihuje odchylku mezi modelem jen s interceptem a s daty, zatímco reziduální odchylka udává odchylku mezi modelem s prediktory a daty. Rozdílem těchto hodnot dostaneme vyjádření schopnosti modelu vysvětlit variabilitu dat. V navrženém modelu vychází rozdíl na 456, což vykazuje, že je model schopný vysvětlit 456 jednotek variability dat. Podílem této hodnoty a nulové odchylky zjistíme koeficient determinace s hodnotou 0,011709. Můžeme tedy říci, že tento model vysvětluje 1,1709 % variability dat.

4.2. Dálnice D11

Pro další analýzu byla vybrána oblast vyobrazená na obrázcích č. 9 a 10, která se nachází na 17,5 km dálnice D11 ve Středočeském kraji. Dálnice D11 spojuje Prahu na východ s Hradcem Králové, dále vede do Jaroměře, odkud má vést k hraničnímu přejezdu Královec/Lubawka v Polsku. Délka dálnice je 114,483 km.

Na obrázku č. 9 můžeme vidět pozici meteorologické stanice a TMC segmentu. Meteostanice se nachází ve Středočeském kraji v 196 m n. m. Roční průměr denních intenzit dopravy bylo v roce 2020 při Celostátním sčítání dopravy 45 499 vozidel za den v obou směrech. [30]



Obrázek 9: Pozice meteorologické stanice na dálnici D11



Obrázek 10: Pozice meteorologické stanice na dálnici D11 z pohledu ČR

4.2.1. Příprava dat

Data z meteostanice byla uložena v souborech s názvem S008, což je identifikační číslo meteostanice, rozdělená podle roků (2019, 2020, 2021). Nejprve proběhla úprava a čištění dat, nulové hodnoty byly odstraněny. Původně bylo zamýšleno hodnoty nahradit pomocí interpolace, ale vzhledem k velkému množství za sebou chybějících dat nakonec nebyl algoritmus na výsledný dataset použit a řádky s NA hodnotami byly odstraněny.

Odlehle hodnoty byly ponechány, aby nebyla výsledná statistika zkreslená. Po úpravě se pracovalo s následujícími proměnnými: teplota vzduchu, teplota vozovky, srážky, vlhkost, dohlednost a rychlost větru. Ostatní proměnné buď nebyly vůbec meteorologickou stanicí zaznamenány, nebo nebyly shledány významnými pro danou analýzu. Pro účely modelu tak nebyly shledány relevantní údaje o teplotě vozovky v hloubce 5 a 30 cm, směr větru, radiace a teplota namrzání, kterou lépe vystihuje stav vozovky. Vzhledem k výsledkům předchozího modelu nebyla v tomto případě teplota vzduchu upravena na rozdíl od průměru, ale bylo počítáno s její absolutní hodnotou.

Data z plovoucích vozidel byla k meteostanici přiřazena pomocí TMC segmentu, který se u meteorologické stanice nachází. Z programu QGIS bylo zjištěno identifikační číslo příslušného TMC segmentu k meteostanici (TS25617T01349). Ten byl pak vyfiltrován v RStudiosu napojením na SQL databázi FCD dat. Byla tak získána data odpovídající časovému období dat z meteostanice – tj. roky 2019, 2020 a 2021. Data byla také upravena a nulové hodnoty byly odstraněny. Z datové sady nás zajímaly především údaje o aktuální vypočtené rychlosti dopravního proudu a typické průměrné rychlosti při volném průjezdu segmentu, dále jsme stejně jako v případě Kubovy Hutě pracovali s rozdílem hodnot těchto dvou parametrů, který vyjadřuje změnu rychlosti. Z ostatních parametrů byla extrahována také časová značka a sloupec určující míru spolehlivosti a stupeň dopravy, které však sloužily spíše jen pro orientaci v daných hodnotách.

Datová sada z meteorologické stanice pak byla napojena na data z plovoucích vozidel pomocí časového razítka, tím vznikl nový datový rámec o 283 684 řádcích. Hodnoty proměnných byly převážně spojité veličiny, kategoriální byl opět pouze stav vozovky, který nabýval hodnot mokro, mokro nasoleno, možnost namrzání, sucho, vlhko a zbytková sůl. Aby se s ním mohlo pracovat v rámci modelu, byl upraven do matice, kdy se z jednotlivých stavů vytvořily sloupce a ty pomocí nul a jedniček určovaly, jaký stav vozovky v daný časový okamžik byl zaregistrován.

Z dat byly provedeny vybrané deskriptivní charakteristiky. V následujících tabulkách jsou hodnoty statistických veličin, jako je průměr, minima, maxima či četnosti. Jak můžeme vidět v tabulce č. 17, u srážek byly po úpravě dat, kdy se mimo jiné promazávaly řádky obsahující NA hodnoty, vypočteny velmi nízké hodnoty. Zvláště pak u maximálních hodnot zbyla například v roce 2021 v datovém setu po úpravě hodnota jen 0,9 mm, což neodpovídá maximální naměřené hodnotě, ale bohužel měly NA hodnoty v jiných sloupcích dopad na srážková data.

Tabulka 17: Statistické veličiny meteorologických dat (dálnice D11)

Parametr	Rok	Minimum	Maximum	Průměr	Medián
Teplota [°C]	2019	-5,4	15,6	4,03	3,9
	2020	-6,9	5,7	-2,08	-3,6
	2021	-9,9	29,1	7,21	5,8
Rosný bod [°C]	2019	-8,7	10,7	1,44	1,5
	2020	-7,4	0,7	-3,51	-3,9
	2021	-13,7	19,7	3,26	2,4
Teplota vozovky [°C]	2019	-4,8	21,1	4,62	4,2
	2020	-5,1	8,6	-0,34	-1,2
	2021	-9	40,4	9,79	7,6
Srážky [mm]	2019	0	15,2	0,04	0
	2020	0	0,9	0	0
	2021	0	23,2	0,04	0
Vlhkost [%]	2019	43,2	99,4	83,93	84,9
	2020	58,6	99,1	90,72	97
	2021	19,4	100	78,68	83,4
Dohlednost [m]	2019	149	2000	1999,42	2000
	2020	2000	2000	2000	2000
	2021	20	2000	1993,95	2000
Rozdíl teplot [°C]	2019	-9,93	10,12	0	0,02
	2020	-4,82	7,78	0	-1,52
	2021	-13	20,4	0	-0,39

Proto byl nakonec k analýze vybrán jen rok 2021. Zároveň byla poté provedena korelační analýza, po jejímž vyhodnocení byl upraven datový set jen s navzájem nekorelovanými proměnnými z toho důvodu, aby promazané NA hodnoty v jednotlivých sloupcích neovlivňovaly přílišným způsobem naměřené hodnoty. V tabulce č. 18 pak můžeme sledovat statistické veličiny u proměnných, které vstupují do modelu.

Tabulka 18: Statistické veličiny meteorologických dat – proměnné v modelu (dálnice D11)

Parametr	Rok	Minimum	Maximum	Průměr	Medián
Teplota [°C]	2021	-9,9	35,5	11,53	11,4
Srážky [mm]	2021	0	201,7	0,07	0
Dohlednost [m]	2021	20	2000	1996,46	2000

Dále byl vytvořen popis stavu vozovky pomocí četností (tabulka č. 19).

Tabulka 19: Statistické vyhodnocení stavu vozovky (dálnice D11)

Rok	Stav	Četnost	Relativní četnost [%]
2021	možnost namrzání	150	0,04
	zbytková sůl	4415	1,10
	vlhko	18730	4,68
	mokro nasoleno	21171	5,29
	mokro	33559	8,38
	sucho	322276	80,51

Výsledky statistických veličin dat z plovoucích vozidel jsou v tabulce č. 20.

Tabulka 20: Statistické veličiny dat z plovoucích vozidel (dálnice D11)

Rok	Průměrná rychlost vypočtená	Průměrná rychlost typická	Průměrná míra spolehlivosti	Průměrný stupeň dopravy
2019	109,515	128,795	0,936	1,634
2020	114,017	130	0,878	1,556
2021	111,497	130	0,847	1,671

4.2.2. Výsledky metod

4.2.2.1. Korelační analýza

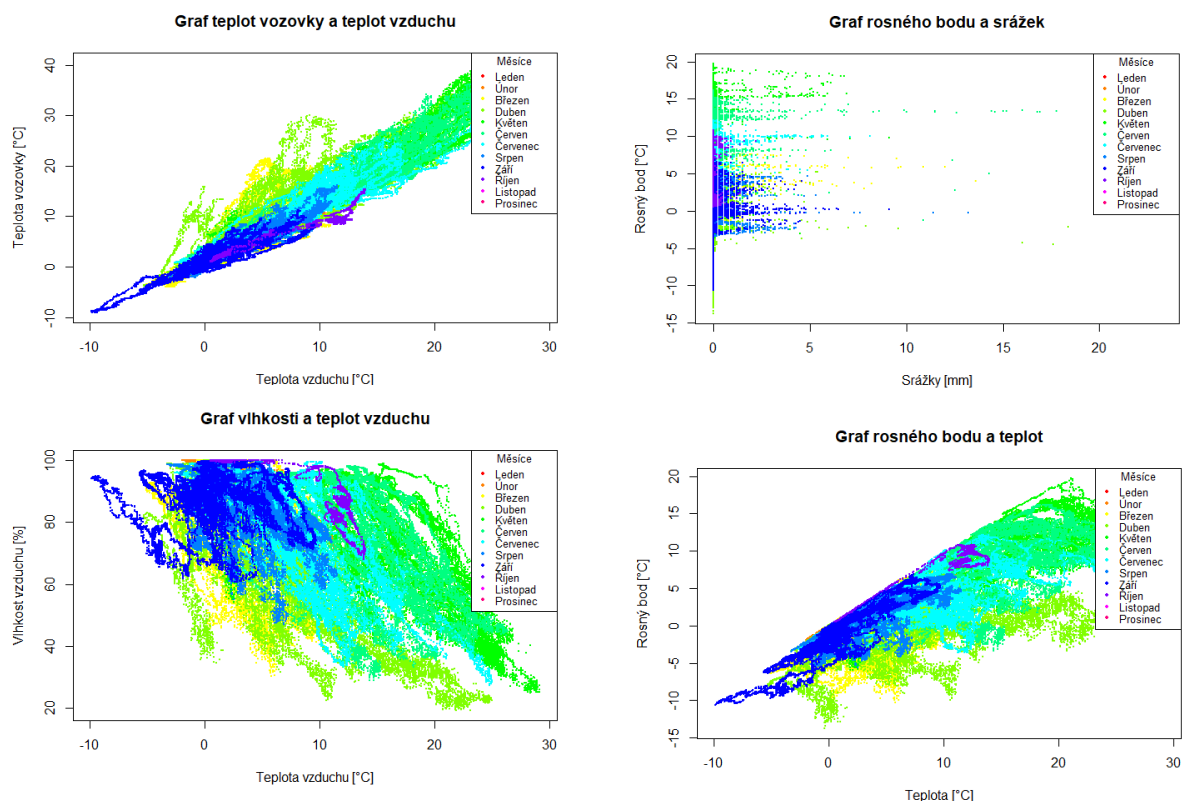
Korelační matice byla i v tomto případě vytvořena pomocí Spearmanova testu nezávislosti, který byl vybrán vzhledem k přítomnosti odlehklých hodnot ve srážkových datech a datech o dohlednosti. Z výsledků korelační matice můžeme pozorovat, že parametr teploty silně koreluje s rosným bodem, teplotou vozovky a vlhkostí.

Parametry s hodnotou korelačního vztahu nad 0,6 značí silnou závislost, proto vybraný prvek z tohoto silně korelovaného vztahu není do modelu zařazen. V tomto případě do modelu rovnice regresní analýzy nevstupuje rosný bod, teplota vozovky a vlhkost. Hodnoty korelačního koeficientu jsou v tabulce č. 21.

Tabulka 21: Výsledky korelační matice se Spearmanovým testem nezávislosti (dálnice D11)

	Srážky	Rosný bod	Teplota vozovky	Vlhkost	
Rosný bod	0,72	Teplota	0,83	0,96	0,63

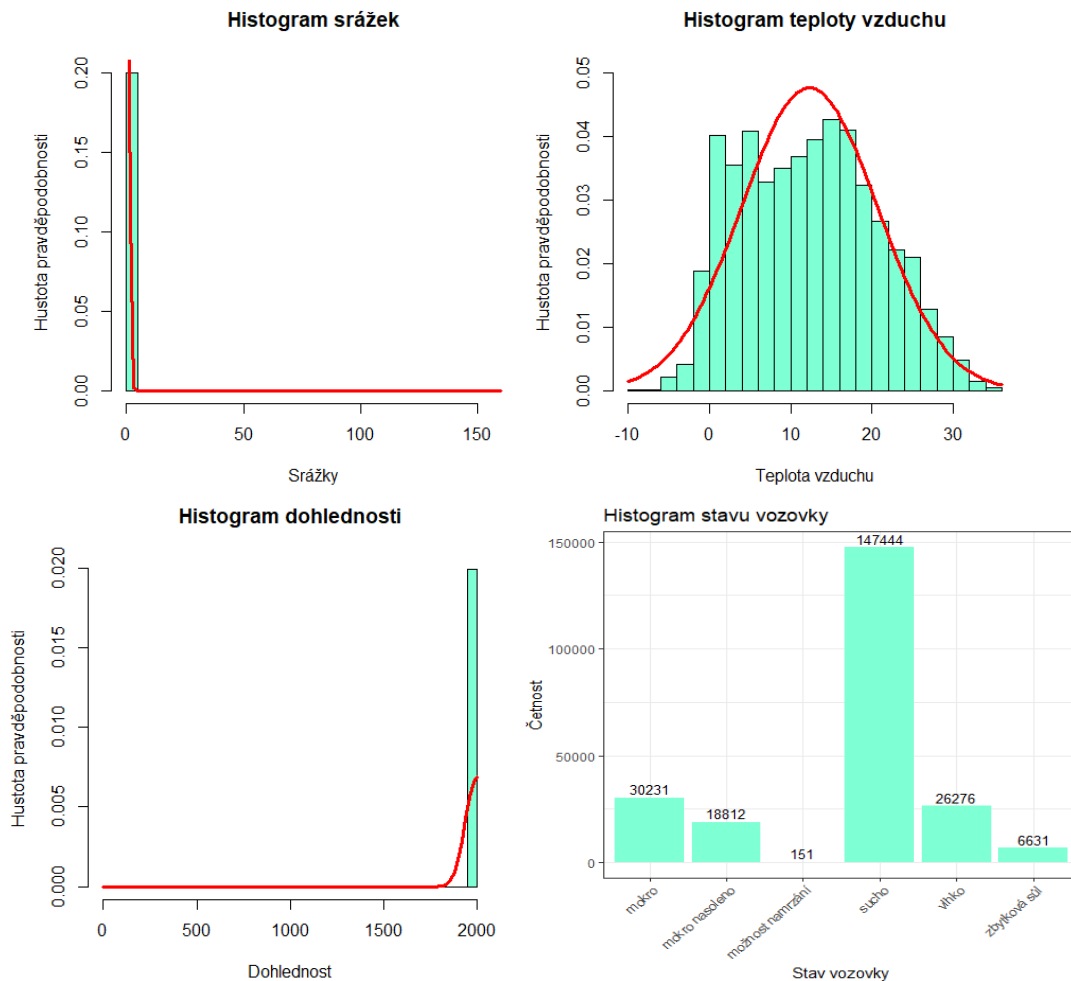
V souhrnném grafu č. 6 jsou silně korelované vztahy vyobrazeny. Pomocí barevného zobrazení hodnot můžeme pozorovat souvislost s měsícem, kdy byly údaje zaznamenány.



Graf 6: Grafy závislosti (dálnice D11)

4.2.2.2. Normalita dat

Proměnné, které vstupují do modelu, byly pomocí histogramů vyobrazeny tak, aby z nich bylo možné posoudit, zda pochází z normálního rozdělení. Histogramy jsou červeně proloženy křivkou normálního rozložení a je k nim přidán i histogram reprezentující četnosti stavu vozovky (graf č. 7). Tvar histogramu srážek a dohlednosti je velmi silně zešikmený doprava, resp. doleva, s nulou jako nejčastější naměřenou hodnotou u srážek, resp. s hodnotou 2 000 u dohlednosti. Data byla nakonec ponechána v tomto stavu, jelikož máme dostatečně velký vzorek dat a kdyby proběhly úpravy hodnot, výsledek modelu by byl zkreslený. Podobně se vyjadřuje centrální limitní věta, která říká, že pokud máme dostatečně velké množství dat, tak se data normálnímu rozdělení začnou přibližovat. Zároveň lineární model nemá za podmínku normální distribuci proměnných, ale spíše reziduí. [32][29]



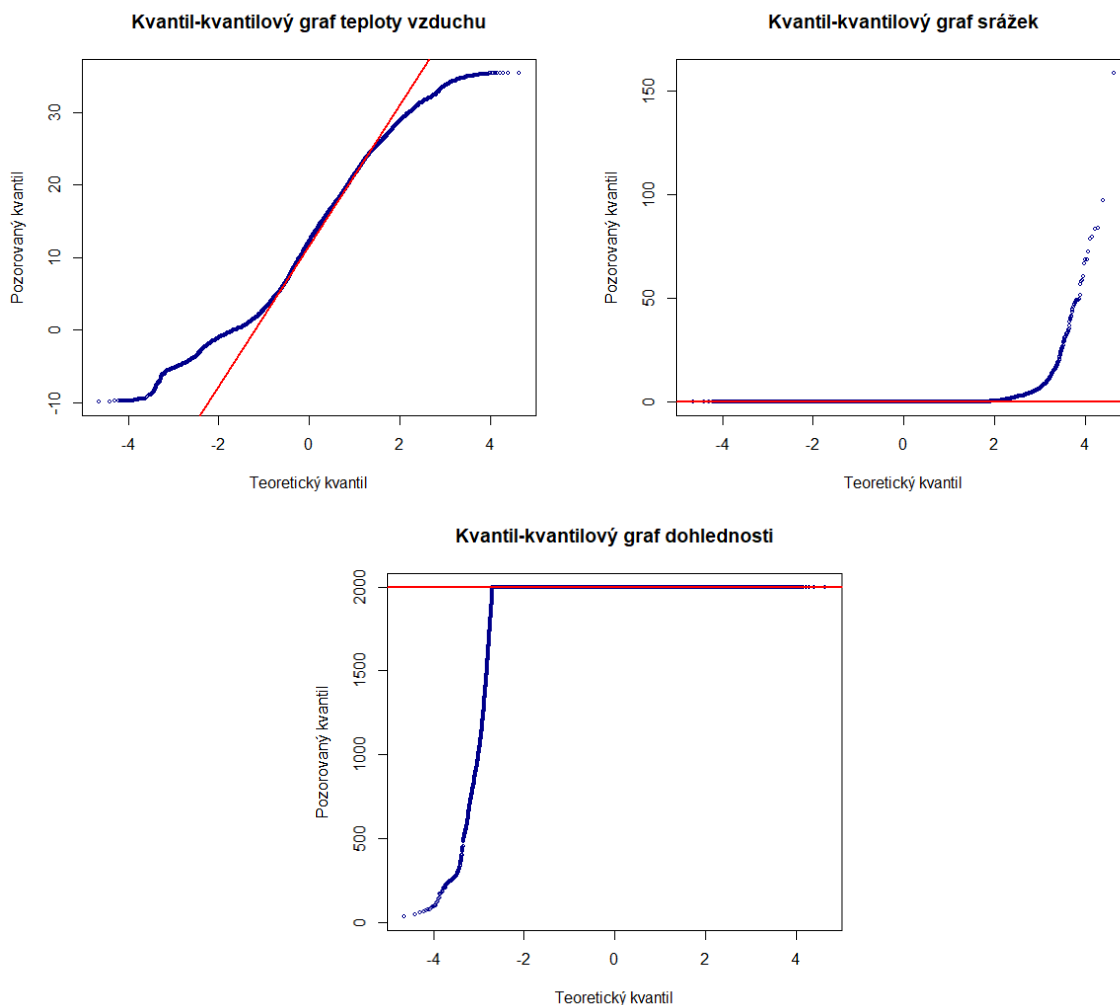
Graf 7: Histogramy (dálnice D11)

Vypočtené koeficienty šikmosti a špičatosti jsou zapsány v tabulce č. 22. Jak můžeme pozorovat, situace u srážek je velmi podobná jako u modelu Kubovy Hutě – data mají velkou špičatost a jsou zešikmená doprava. U dohlednosti je tomu podobně, avšak odlehle hodnoty se nacházejí od průměru nalevo a mají o něco menší špičatost. Situace je opět pravděpodobně daná odlehlými hodnotami. Jedině teplota se v tomto případě přibližuje dokonale normálnímu rozdělení.

Tabulka 22: Koeficienty šikmosti a špičatosti (dálnice D11)

Koeficient	Srážky	Teplota	Dohlednost
Šikmosti	57,23	0,19	-22,87
Špičatosti	5 550,45	2,2	570,8

Proměnné byly zobrazeny (viz. graf č. 8) ještě pomocí kvantil-kvantilového grafu, kdy jsou na ose x zobrazeny kvantily teoretického rozdělení (v tomto případě normální), zatímco na osy y jsou kvantily pozorovaného rozdělení. Normálně rozdělená data kopírují červeně vyznačenou linii. V tomto případě můžeme pozorovat jisté odchýlení, např. ve tvaru S v případě kvantilového grafu teploty vzduchu, jak již naznačovaly histogramy.



Graf 8: Kvantil–kvantilové grafy (dálnice D11)

4.2.2.3. Regresní analýza

Stejně jako tomu bylo u Kubovy Hutě, i zde byla provedena regresní analýza pomocí dvou modelů, kdy jeden model je pro binární data vyjadřující stav vozovky.

Model 1

Vícerozměrný lineární model byl vytvořen v RStudiosu podle následující upravené rovnice z kapitoly 4.1.

$$\Delta Y = b_0 + b_t T + b_s S + b_d D$$

Do modelu vstupují jako nezávislé proměnné teplota vzduchu, srážky a dohlednost. Ostatní proměnné byly buď vyřazeny kvůli korelaci, nebo nebyly shledány významnými. Výsledkem modelu je následující tabulka č.23.

Tabulka 23: Výsledky regresního modelu 1 (dálnice D11)

	Odhad parametru	Odhad směrodatné odchylky	Statistika t	P hodnota
Intercept	-33,73	1,091	-30,91	$<2 \cdot 10^{-16}$
Teplota vzduchu	0,272	$3,633 \cdot 10^{-3}$	74,86	$<2 \cdot 10^{-16}$
Srážky	-0,5232	$3,402 \cdot 10^{-2}$	-15,38	$<2 \cdot 10^{-16}$
Dohlednost	0,006209	$5,468 \cdot 10^{-4}$	11,36	$<2 \cdot 10^{-16}$

Oproti předchozímu lineárnímu modelu se model zabývá i dohledností. Změna se také týká teploty vzduchu, která je v tomto případě v modelu v absolutních číslech. Model pracoval se vzorkem o velikosti 283 684 dat.

Výsledná statistika f modelu byla spočtena na hodnotu 2 079 na 3 a 283 680 volných stupňů volnosti. Kritická hodnota f -testu se vypočítala s hladinou významnosti 0,05 pro tyto stupně volnosti a porovnála se s výsledkem, který byl mnohem vyšší, tím pádem můžeme tvrdit, že model disponuje vysvětlovací schopností s pomocí minimálně jedné vysvětlovací proměnné a alespoň částečně objasňuje stav hodnot rozdílu aktuální rychlosti dopravního proudu oproti průměrné rychlosti. Kritická mez pro f -test činí v tomto modelu 0,117282.

P -hodnota modelu je zároveň se svojí hodnotou menší než $2 \cdot 10^{-16}$ nižší než určená hladina významnosti (0,05), lze tedy uznat, že minimálně jeden regresní koeficient je statisticky významně odlišný od nuly.

Z výsledků modelu můžeme vyhodnotit následující:

Pro parametr Teplota vzduchu

- Odhad regresního koeficientu pro proměnnou teplota vzduchu je 0,272 s odhadem směrodatné odchylky 0,003633, kdy 95% interval spolehlivosti pro parametr se nachází v rozmezí od 0,26486 do 0,27911. Hodnota regresního koeficientu 0,272 znamená, že pro každé jednotkové zvýšení v hodnotách teplot se očekává průměrné zvýšení rozdílu rychlosti dopravního proudu o 0,272, za předpokladu, že jsou všechny ostatní vysvětlující hodnoty

konstantní. Statistika t s hodnotou 74,86 udává, že odhadnutý regresní koeficient je téměř 75krát menší než standardní chyba tohoto odhadu. Tato hodnota je v rámci modelu nejvyšší a značí, že je velmi malá pravděpodobnost, že jde o náhodný výskyt. P -hodnota je menší než hladina významnosti, proto u teploty můžeme říct, že je statisticky významná.

Pro parametr Srážky

- Odhad regresního koeficientu pro proměnnou srážky je -0,5232 s odhadem směrodatné odchylky $3,402 \cdot 10^{-2}$, kdy 95% interval spolehlivosti pro parametr se nachází v rozmezí -0,58991 až -0,45654. Hodnota regresního koeficientu -0,5232 značí, že pro každé jednotkové snížení hodnoty srážek se očekává průměrné zvýšení rozdílu rychlosti dopravního proudu o 0,5232, jestliže jsou všechny ostatní vysvětlující hodnoty konstantní. Statistika t s hodnotou -15,38 je daleko od nuly, proto lze dostatečně průkazně odmítnout nulovou hypotézu a nabídnout alternativní hypotézu, za přispění p -hodnoty menší než hladina významnosti.

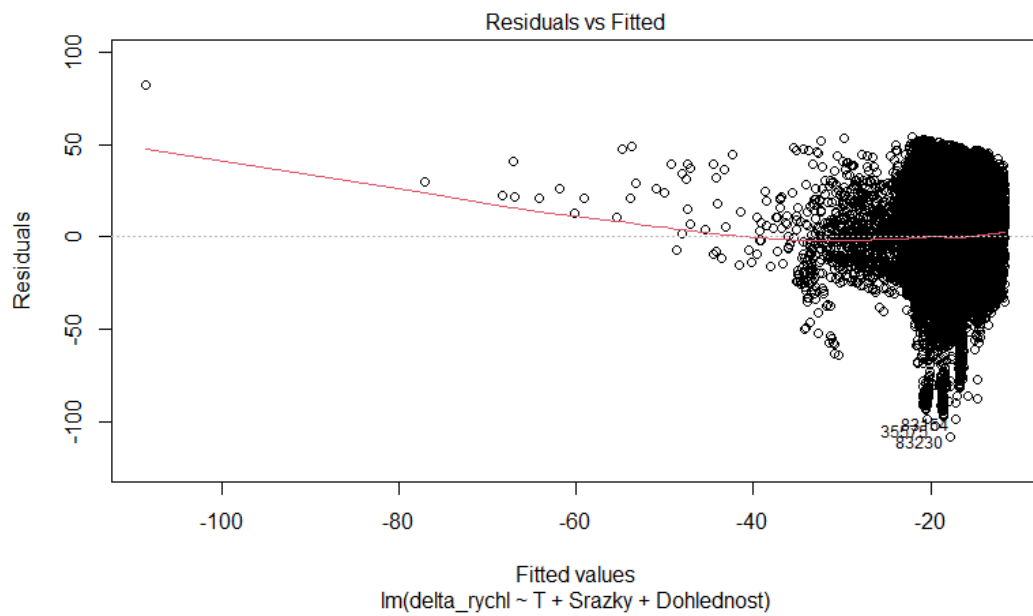
Pro parametr Dohlednost

- Odhad regresního koeficientu pro proměnnou dohlednost je 0,006209 s odhadem směrodatné odchylky 0,0005468, kdy 95% interval spolehlivosti pro parametr se nachází v intervalu od 0,00514 do 0,00728. Hodnota regresního koeficientu 0,006209 znamená, že pro každé jednotkové zvýšení hodnoty dohlednosti se průměrně zvýší rozdíl rychlosti dopravního proudu o 0,006209, v případě že jsou všechny ostatní vysvětlující hodnoty konstantní. Statistika t s hodnotou 11,36 udává, že odhadnutý regresní koeficient je 11,36krát větší než standardní chyba tohoto odhadu a je dostatečně velká na to, aby se jednalo o náhodný výskyt. Z p -hodnoty můžeme vyhodnotit na hladině významnosti zamítnutí nulové hypotézy, že regresní koeficient je nulový, jelikož p -hodnota je menší než $2 \cdot 10^{-16}$. Tento regresní koeficient je tedy statisticky významný a může vysvětlovat vysvětlovanou proměnnou, přestože jeho hodnota je velmi nízká.

Celkově předkládá model lepší výsledky než lineární model z Kubovy Hutě, jelikož p -hodnoty a výsledek t -testu u všech proměnných poukázaly na jejich statistickou významnost. Největší vliv na změnu rychlosti dopravního proudu mají srážky, s téměř polovičním odhadem koeficientu je v závěsu teplota vzduchu. Nepatrný vliv, přesto ne náhodný, má dohlednost.

Koeficient determinace R^2 je 0,02152 na intervalu od $\langle 0,1 \rangle$, přičemž jeho blízkost nuly značí nízkou vysvětlovací schopnost modelu. Z toho vyplývá, že vybrané proměnné v modelu vysvětlují jen malou část variability hodnot závislé proměnné. Upravená hodnota koeficientu, která zohledňuje počet regresních proměnných a velikost vzorku dat, vyšla na 0,02151. Hodnota je velmi podobná jako u normálního R^2 , což může znamenat, že v modelu se nenachází žádné zbytečné regresní proměnné. V porovnání s modelem Kubovy Hutě se vysvětlovací schopnost modelu neznatelně zvýšila.

Na grafu č. 9 pak můžeme pozorovat závislost mezi predikovanými hodnotami a chybou, neboli rezidui. V grafu by neměl být nalezen žádný vzorec a rezidua by měla být rozptýlena kolem nuly, což náš model splňuje a znamená to, že byl nastaven správně. Zároveň je tím splněna podmínka lineárního modelu, že rezidua mají mít normální distribuci. [28]



Graf 9: Výsledek z modelu – grafické znázornění reziduí (dálnice D11)

Model 2

Pro stav vozovky byl vytvořen zobecněný lineární model s binomickým rozdělením, který má za cíl zjistit, jak jednotlivé proměnné ovlivňují pravděpodobnost snížení rychlosti. Změna rychlosti byla upravena tak, aby odpovídala alternativnímu rozdělení. Vzhledem k povaze stavů vozovky – mokro, mokro nasoleno, sníh/led, sucho, vlhko a zbytková sůl – nás spíše zajímalo, kdy se rychlost sníží. Byla určena hraniční hodnota změny rychlosti o 20 km/h oproti typické průměrné rychlosti při volném průjezdu. Při srovnání s předchozím zobecněným modelem nebyl v datech zaznamenán stav sníh/led a korelační matice v tomto případě neshledala žádné silné korelace.

V modelu se pracovalo se vzorkem o velikosti 400 301 dat.

V tomto případě nebylo v případě stavů vozovky zjištěna korelace nad hodnotu 0,6, proto se rovnice modelu trochu mění. Zároveň se v datech neobjevil stav sníh/led.

Upravená rovnice modelu vypadá následovně:

$$\Delta Y = b_0 + b_m C_m + b_{mn} C_{mn} + b_{su} C_{su} + b_v C_v + b_z C_z$$

Výsledné hodnoty nalezneme v následující tabulce č. 24.

Tabulka 24: Výsledky zobecněného regresního modelu (dálnice D11)

	Odhad parametru	Odhad směrodatné odchylky	Statistika z	P hodnota
Intercept	0,235243	0,013247	17,759	$<2 \cdot 10^{-16}$
Mokro nasoleno	0,003447	0,021907	0,157	0,875
Možnost namrzání	1,522615	0,279886	5,44	$5,32 \cdot 10^{-8}$
Sucho	-0,5273	0,013894	-37,953	$<2 \cdot 10^{-16}$
Vlhko	-0,14313	0,022457	-6,373	$1,85 \cdot 10^{-10}$
Zbytková sůl	0,039954	0,045093	0,886	0,376

Hodnoty parametrů jsou udávány na hladině významnosti 0,05. Pro jednotlivé nezávislé proměnné bylo vyhodnoceno:

Pro parametr Mokro nasoleno

- Odhadnutý regresní koeficient je 0,003447 a podle p -hodnoty není statisticky významný, jelikož je nad hladinou signifikantnosti. Z hodnota je 0,157, což také značí, že proměnná nemá vliv na závislou proměnnou. V tomto případě nám kombinace nízké hodnoty z a naopak vysoké p -hodnoty potvrzuje přijetí nulové hypotézy.

Pro parametr Možnost namrzání

- Odhadnutý regresní koeficient má hodnotu 1,522615, hodnotu z -testu o 5,44 více vzdálenou od nuly a nízkou p -hodnotu. V tomto případě můžeme vyhodnotit, že parametr má vliv na výsledek modelu a je statisticky významný.

Pro parametr Sucho

- Odhadnutý regresní koeficient je -0,5273 a nízká p -hodnota spolu s velmi vysokou statistikou z znamenají, že je parametr také statisticky významný s odhadem směrodatné odchylky 0,013894, tím pádem i vyšší hodnoty z . Můžeme z toho vyvodit, že i tento parametr ovlivňuje snížení rychlosti o více než 20 km/h.

Pro parametr Vlhko

- Odhadnutý regresní koeficient je pro stav vlhko -0,14313, podle p -hodnoty také statisticky významný s odhadem směrodatné odchylky 0,022457, tím pádem i vyšší hodnoty z . Můžeme z toho vyvodit, že i tento parametr ovlivňuje snížení rychlosti o více než 20 km/h.

Pro parametr Zbytková sůl

- Pro stav zbytková sůl je odhadnutý regresní koeficient 0,039954. V tomto případě přesahuje p -hodnota hladinu významnosti, proto u tohoto parametru nemůžeme vyloučit nulovou hypotézu.

Model vyhodnotil, že stav mokro nasoleno a stav zbytková sůl nemají vliv na snížení rychlosti o více jak 20 km/h. Největší odhad regresního koeficientu měl stav možnost namrzání, dále byl s velmi malým odhadem chyby významný stav sucho. Stav vlhko pak může mít velmi malý vliv.

Pomocí nulové a reziduální odchylky můžeme říci, že je navržený model schopný vysvětlit 2 667 jednotek variability dat. Podílem této hodnoty a nulové odchylky zjistíme koeficient determinace s hodnotou 0,00683. Můžeme tedy říci, že tento model vysvětluje 0,68 % variability dat.

Pokud z modelu tedy odstraníme statisticky nevýznamné proměnné, dobereme se následujících výsledků (tabulka č. 25). Všechny proměnné byly shledány jako statisticky významné.

Tabulka 25: Výsledky upraveného regresního modelu (dálnice D11)

	Odhad parametru	Odhad směrodatné odchylky	Statistika z	P hodnota
Intercept	0,2387	0,01025	23,3	$< 2 \cdot 10^{-16}$
Možnost namrzání	1,51916	0,27976	5,43	$5,63 \cdot 10^{-8}$
Sucho	-0,53076	0,01107	-47,9	$< 2 \cdot 10^{-16}$
Vlhko	-0,14658	0,02083	-7,037	$1,96 \cdot 10^{-12}$

Z tabulky můžeme vyčíst, že stav vozovky možnost namrzání ovlivňuje snížení rychlosti nejvíce. V závěsu je stav sucha, jehož odhad parametru říká, že jestliže je vozovka suchá, změna rychlosti je o to menší. Stav vlhko je v porovnání s ostatními odhady nejmenší, jeho vliv tedy bude nepatrný.

5. Lokální prediktivní model

Výběr modelu pro vytvoření predikce je náročné, protože model musí být dostatečně robustní, aby vytvářel přesné predikce, zároveň musí mít nízké nároky na zdroje. Po nastudování vybraných prediktivních metod byly vybrány tři algoritmy, jejichž výsledky pak byly porovnány. Jedná se o algoritmus náhodného lesa, metoda nejbližších sousedů a prediktivní lineární regrese.

Pro tento model byla použita data z roku 2021 z dálnice D11, která byla zpracována v kapitole 4.2. Zdroje dat tedy pocházejí z meteorologické stanice S008 a segmentu TS25617T01349.

Kritériem pro porovnání jednotlivých modelů je především střední kvadratická chyba, neboli root mean square error (RMSE). Dále pak byla vypočtena i střední absolutní chyba (mean absolute error, MAE) a relativní chyba. Relativní chyba je poměr absolutní chyby k absolutní hodnotě.

Náhodný les – Random Forest

Algoritmus náhodného lesa je ensemble učící algoritmus, který se skládá z většího určeného množství rozhodovacích stromů pro predikci výsledné proměnné. Každý strom je vytvořen pomocí náhodného výběru pozorování a proměnných z trénovacího datasetu. Tímto způsobem model získává odolnost vůči přeučení, které je typické pro jednotlivé rozhodovací stromy, a zároveň zachovává jejich schopnost zachytit složité nelineární vztahy mezi proměnnými. Náhodný les vypočítává predikci tím způsobem, že výstupy jednotlivých stromů určují výsledný odhad. [33]

Algoritmus k nejbližších sousedů – K-Nearest Neighbor (K-NN)

Tento algoritmus, založený na instanci, funguje na principu hledání k nejbližších sousedů v trénovacích datech pro nové, neznámé pozorování a poté vypočítává predikci na základě hodnot těchto sousedních instancí. Výsledná hodnota predikce je většinou průměrem cílových hodnot těchto k sousedů. Co se týče distribuce dat, K-NN nevyžaduje žádné předpoklady o distribuci dat, je tedy vhodný v případě, že jsou vztahy mezi proměnnými nelineární a komplikované. Na začátku je potřeba nastavit vhodný počet nejbližších sousedů. [34]

Lineární regrese

Lineární regrese disponuje jednoduchým přístupem k učení s učitelem, kdy se v kontextu predikce používá k odhadu cílové proměnné na základě jedné nebo více nezávislých proměnných. Lineární regresi se zabývá předcházející kapitola. Zde byla zahrnuta z toho důvodu, aby bylo možné ji porovnat se zbývajícími algoritmy.

5.1. Výsledky prediktivního modelu

Nejprve byla data načtena do RStudia, kde byl ze souboru vybrán náhodně vzorek 20 000 dat. Data pak byla náhodně rozdělena na trénovací a testovací data v poměru 80 % trénovací a z 20 % testovací data, jde tedy o učení s učitelem. Byla také provedena kontrola křížové validace. Při křížové validaci se datový soubor rozdělí na n podmnožin, kdy se vždy jedna podmnožina použije na testovací fázi a zbylých $n-1$ podmnožin na fázi trénovací. Postup se opakuje tím způsobem, že každá podmnožina se použije právě jednou. [32]

Pro každý algoritmus byl vytvořen jeden model. Model byl spuštěn na trénovacím vzorku, kdy se vytvořily předpovědi pro testovací vzorek. Následně se předpovědi porovnaly se skutečnými hodnotami v testovacím vzorku. Vzhledem k velkému datovému vzorku, který obsahoval přes 400 000 údajů, bylo z datového souboru náhodně vybráno 20 000 řádků.

Do modelů vstupují proměnné teplota vzduchu, rosný bod, teplota vozovky, stav vozovky, srážky, vlhkost a dohlednost. Prediktivní proměnnou je aktuální rychlost. Kategorická proměnná stav byla upravena do datového typu faktor, aby s ní bylo možné pracovat. Datový soubor byl ještě rozdělen na dva různé datové sady, aby se zjistilo, jak se modely chovají za různých podmínek.

Jednotlivé modely jsou mezi sebou porovnávány pomocí střední čtvercové chyby, která udává míru nepřesnosti aproximace.

Algoritmus regresního náhodného lesa počítá u všech případů s 500 stromy. Byly vyzkoušeny i jiné počty stromů, ale při počtu s 500 stromy docházelo k nejlepším výsledkům.

Algoritmus K-NN byl vyzkoušen s různými počty k . Ve všech případech ale nejlépe vycházelo 10 nejbližších sousedů.

Datový soubor se všemi proměnnými

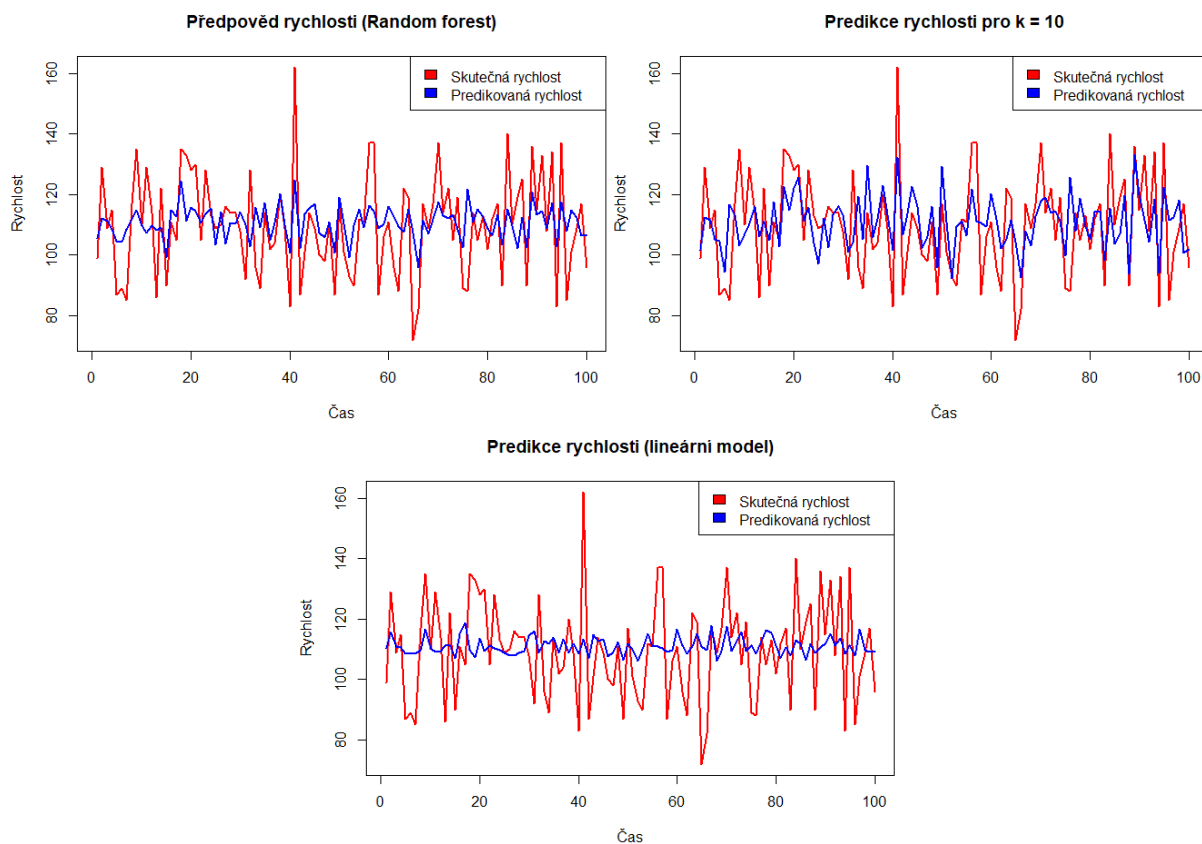
Nejprve byl namodelován soubor se všemi zmíněnými proměnnými, nehledě na jejich vzájemnou korelaci či výsledky regresní analýzy. V tabulce č. 26 můžeme vidět jednotlivé RMSE pro model, pokud byly zahrnuty všechny proměnné. Mimo RMSE byla vypočtena ještě střední absolutní chyba (MAE) a relativní chyba.

Tabulka 26: Porovnání modelů v případě všech proměnných

Typ predikce	RMSE	MAE	Relativní chyba
Random Forest	14,868	11,848	0,134
KNN	15,194	11,876	0,136
Lineární	16,024	12,785	0,144

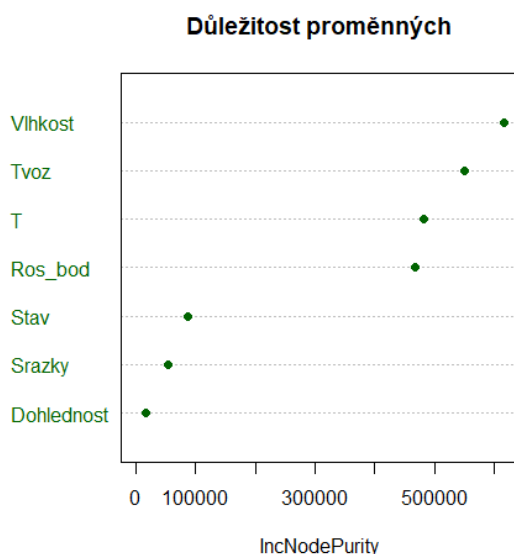
Jak můžeme z hodnot RMSE vidět, nejlépe v tomto případě byl schopný predikovat model s algoritmem náhodných lesů. Dále pak měl lepší výsledek algoritmus k nejbližších sousedů a nejhůře dopadla lineární predikce. Rozdíl hodnot ale není příliš vysoký.

Na následujících grafech č. 10 je vyobrazena předpověď rychlosti vůči jejím skutečným hodnotám na testovacích datech.



Graf 10: Předpověď rychlosti v případě všech proměnných

Pomocí algoritmu Random forest byl také vyjádřen podíl vysvětlených proměnných, který byl spočten na 14,65 %. Zároveň byl zjištěna důležitost proměnných, která určuje, jak moc proměnné pomáhají predikcím. Graficky je důležitost zobrazena v grafu č 11.



Graf 11: Důležitost proměnných v případě všech proměnných

Datový soubor s vybranými proměnnými

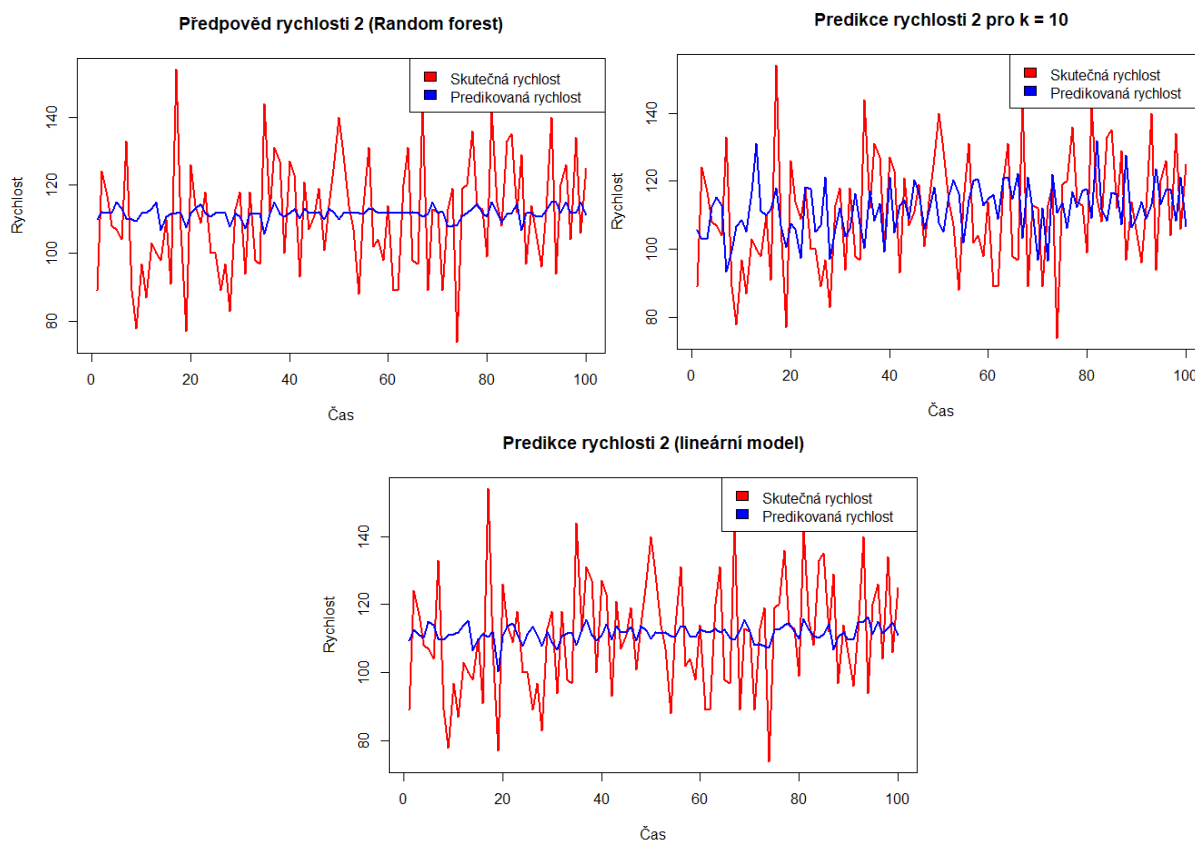
Pro porovnání byly vytvořeny modely, kde byly z proměnných odstraněny korelované veličiny, takže se pracovalo s teplotou vzduchu, stavem, srážkami a dohledností. Jedná se o stejný datový soubor z regresní analýzy dálnice D11.

Výsledné hodnoty RMSE v tabulce č. 27 ukazují, že střední kvadratická chyba je u modelů náhodného lesa a lineární regrese velmi podobné, na druhé straně K-NN model vykazuje největší hodnotu RMSE.

Tabulka 27: Porovnání modelů v případě vybraných proměnných

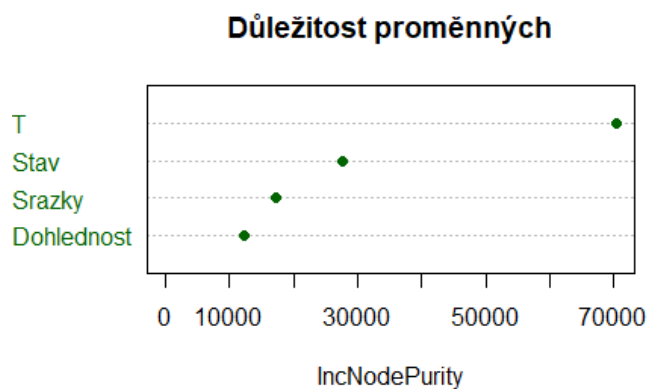
Typ	RMSE	MAE	Relativní chyba
Random Forest	15,846	12,548	0,142
KNN	17,254	13,775	0,154
Lineární	15,865	12,573	0,142

Níže na grafu č. 12 můžeme pozorovat, jak se predikce liší od skutečných hodnot na prvních 100 datech testovacích dat.



Graf 12: Předpověď rychlosti v případě vybraných proměnných

Pokud si zobrazíme důležitost proměnných z modelu náhodného lesa, dostaneme se k následujícímu grafu č. 13. Podíl variability hodnot závislé proměnné byl spočten na 2,54 %.



Graf 13: Důležitost proměnných v případě vybraných proměnných

6. Závěr

Práce se věnuje analýze vztahu mezi meteorologickými podmínkami a rychlostí dopravního proudu na základě dat z plovoucích vozidel a dat z meteorologických stanic. Pro práci s daty byly využity softwarové nástroje RStudio a pgAdmin. Pro zpracování map a mapových podkladů byl použit software QGIS.

V rámci regresního modelu z dat v Kubově Huti byly využity proměnné rozdíl teploty, teplota vozovky, srážky, vlhkost a tlak. Byl zjištěn nepatrný vliv tlaku vzduchu a teploty vozovky, která spolu s vlhkostí může snižovat rozdíl rychlosti. Z toho by se mohlo usuzovat, že vyšší vlhkost a teplota vozovky jsou spíše v létě, proto rozdíl rychlosti od typické při volném průjezdu nebude takový, ale bohužel vzhledem k velmi nízkým odhadům parametru nelze s jistotou něco takového usuzovat. Největší regresní koeficient byl u srážek, s o něco vyšší pravděpodobností náhodného výskytu. Jistější výsledek vlivu na rychlost má z tohoto modelu teplota vzduchu. Upravený model s rozdílem teploty vzduchu a srážkami potvrdil větší regresní koeficient u srážek.

Kvůli stavu vozovky, který byl v datech vyjádřen kategoricky, byl vytvořen ještě jeden model pracující s binomickým rozdělením. Rychlost dopravního proudu jako závislá proměnná byla upravena na kategorie vyjadřující, jestli je rozdíl aktuální rychlosti o 20 km/h vyšší nebo nižší oproti typické průměrné rychlosti při volném průjezdu. Model v tomto případě nepřijmul ani v případě jednoho stavu nulovou hypotézu. Největší regresní koeficient byl odhadnut u stavu vozovky sníh/led, u namrzání a u stavu vlhko.

Po zhodnocení modelu z Kubovy Hutě byl vytvořen podobný model i u dálnice D11, ovšem s jinou skladbou proměnných a jen z roku 2021. Do modelu vstupovaly proměnné teplota vzduchu, srážky a dohlednost a všechny byly shledány statisticky významné. Největším vlivem podle hodnoty regresního koeficientu měly srážky, pak teplota vzduchu. Nepatrný vliv byl zjištěn u dohlednosti. Upravený model jen s teplotou a srážkami potvrdil silnější odhad parametru u srážek.

Druhý model se stavy vozovky vyhodnotil největší odhad koeficientu u stavu možnost namrzání, jako statisticky nevýznamné určil stav vozovky mokro nasoleno a stav zbytková sůl. Do upraveného modelu pak vstupovaly proměnné možnost namrzání, stav sucho a vlhko, z nichž byl opět s největším vlivem shledán stav možnost namrzání.

Lineární regresní model byl vybrán z toho důvodu, že se snadno interpretuje a porovnává. Koeficient determinace je u všech modelů malý, přesto vyjadřuje, že nějaký vliv proměnné na vysvětlovanou proměnnou rychlost dopravního proudu mají. Nízké koeficienty a koeficient determinace mohou být dány velkou variabilitou dat, proto odhady a předpovědi nejsou tak přesné.

Dalším bodem byla tvorba lokálního prediktivního modelu, který predikuje rychlost dopravního proudu vzhledem k meteorologickým podmínkám. Model byl vytvořen na základě dat z předchozí analýzy a pro

porovnání byly použity algoritmy náhodného lesa, K-NN a lineární regrese. Výsledky byly porovnány podle střední kvadratické chyby, kterou měly modely velmi podobnou. Nejlépe byl vyhodnocen algoritmus náhodného lesa v případě, že predikoval na základě všech dostupných parametrů.

Vzhledem k výsledkům modelů by se práce mohla dále věnovat dopadu meteorologických podmínek na dopravní nehody, které bývají zapříčiněny i povětrnostními vlivy (např. zhoršenou viditelností, námrazou), případně k vyššímu riziku nehodovosti přispívají. Takový model by mohl dokázat, zda mají povětrnostní podmínky prokazatelně větší vliv na nehodovost než na rychlost dopravního proudu.

Práce by se dále mohla rozšířit o integraci dalších proměnných, jako je například roční období, denní doba či události na silnici. Poté by mohl vzniknout velmi komplexní model, který by lépe predikoval rychlost dopravy. Prostor je také určitě pro optimalizaci prediktivního modelu a jeho validaci. Model by se pak mohl implementovat i v praxi, kdy by se mohla na základě zadaných meteorologických dat predikovat rychlost dopravního proudu, případně i intenzita. Výsledky by se mohly využít jako pomocný nástroj k řízení dopravy, bylo by možné například omezovat rychlost na základě aktuálních meteorologických podmínek a snížit tak pravděpodobnost vzniku nehody. Model by mohl také být základem pro vytvoření dlouhodobých trendů vztahu mezi meteorologickými podmínkami a vlivem na charakteristiku dopravního proudu.

S tím mohou souviset i C-ITS služby, kdy by vylepšená a přesnější predikce rychlosti dopravního proudu mohla sloužit pro včasná upozornění řidičů na očekávané snížení rychlosti. S tím by mohla být spojená i spolupráce s ostatními poskytovanými službami C-ITS, což by umožnilo efektivnější využití a koordinaci opatření pro zmenšení vlivu nepříznivého počasí v provozu.

Využití datové analýzy meteorologických dat ve spojitosti s dopravou má obecně v dopravě velký potenciál, zvláště pak pro její využití při řízení dopravy a jiných procesech, které by se mohly pomocí výstupů z datové analýzy zlepšit. Následkem může být zvýšení bezpečnosti dopravy, optimalizace provozu, minimalizace zbytných opatření a zlepšení plynulosti dopravy.

7. Seznam obrázků

Obrázek 1: Silniční a dálniční síť Jihočeský kraj [10]	13
Obrázek 2: Silniční a dálniční síť Středočeského kraje [10]	14
Obrázek 3: Poloha meteorologických stanic	28
Obrázek 4: Zobrazení aktuální dopravní situace na stránkách dopravniinfo.cz [22]	30
Obrázek 5: TMC segmenty na území Jihočeského, Středočeského kraje a hlavního města Prahy	31
Obrázek 6: Pozice meteorologické stanice v Kubově Huti	38
Obrázek 7: Pozice meteorologické stanice v Kubově Huti z pohledu ČR	39
Obrázek 8: Heatmapa korelační matice (Kubova Huť)	42
Obrázek 9: Pozice meteorologické stanice na dálnici D11	53
Obrázek 10: Pozice meteorologické stanice na dálnici D11 z pohledu ČR	53

8. Seznam tabulek

Tabulka 1: Kategorizace intenzity srážek	18
Tabulka 2: Klasifikace dohlednosti dle množství sněhu	21
Tabulka 3: Vliv počasí na dopravu [15].....	22
Tabulka 4: Statistika nehodovosti v Jihočeském kraji [16].....	23
Tabulka 5: Statistika nehodovosti ve Středočeském kraji [16]	24
Tabulka 6: Data z meteostanice	27
Tabulka 7: FCD data	32
Tabulka 8: Stupně dopravy (LOS)	33
Tabulka 9: Statistické veličiny meteorologických dat (Kubova Huť).....	40
Tabulka 10: Statistické veličiny z dat plovoucích vozidel (Kubova Huť)	40
Tabulka 11: Statistické vyhodnocení stavu vozovky (Kubova Huť)	41
Tabulka 12: Výsledky korelační matice se Spearmanovým testem nezávislosti (Kubova Huť)	42
Tabulka 13: Koeficienty šikmosti a špičatosti (Kubova Huť).....	44
Tabulka 14: Výsledky regresního modelu 1 (Kubova Huť).....	47
Tabulka 15: Výsledky upraveného regresního modelu (Kubova Huť)	50
Tabulka 16: Výsledky regresního modelu (Kubova Huť).....	51
Tabulka 17: Statistické veličiny meteorologických dat (dálnice D11)	55
Tabulka 18: Statistické veličiny meteorologických dat – proměnné v modelu (dálnice D11).....	55
Tabulka 19: Statistické vyhodnocení stavu vozovky (dálnice D11)	56
Tabulka 20: Statistické veličiny dat z plovoucích vozidel (dálnice D11)	56
Tabulka 21: Výsledky korelační matice se Spearmanovým testem nezávislosti (dálnice D11)	57
Tabulka 22: Koeficienty šikmosti a špičatosti (dálnice D11)	59
Tabulka 23: Výsledky regresního modelu 1 (dálnice D11).....	60
Tabulka 24: Výsledky zobecněného regresního modelu (dálnice D11).....	63
Tabulka 25: Výsledky upraveného regresního modelu (dálnice D11).....	64
Tabulka 26: Porovnání modelů v případě všech proměnných	66
Tabulka 27: Porovnání modelů v případě vybraných proměnných.....	68

9. Seznam grafů

Graf 1: Krabicové grafy vlhkosti vzduchu (Kubova Huť)	35
Graf 2: Graf parametrů rozdíl teplot a vlhkost (Kubova Huť)	43
Graf 3: Histogramy (Kubova Huť).....	45
Graf 4: Kvantil–kvantilové grafy (Kubova Huť)	46
Graf 5: Výsledek z modelu – grafické znázornění reziduí (Kubova Huť).....	50
Graf 6: Grafy závislosti (dálnice D11).....	57
Graf 7: Histogramy (dálnice D11)	58
Graf 8: Kvantil–kvantilové grafy (dálnice D11)	59
Graf 9: Výsledek z modelu – grafické znázornění reziduí (dálnice D11).....	62
Graf 10: Předpověď rychlosti v případě všech proměnných.....	67
Graf 11: Důležitost proměnných v případě všech proměnných.....	67
Graf 12: Předpověď rychlosti v případě vybraných proměnných.....	68
Graf 13: Důležitost proměnných v případě vybraných proměnných	69

10. Seznam příloh

Příloha 1: R skripty

11. Citace

- [1] *Elektronický meteorologický slovník (eMS)* [online]. Česká meteorologická společnost, c2017 [cit. 2023-03-02]. Dostupné z: <http://slovník.cmes.cz/>
- [2] HAN, Shuo, Jinliang XU, Menghua YAN, Zhaoxin LIU a Feng CHEN. Using multiple linear regression and BP neural network to predict critical meteorological conditions of expressway bridge pavement icing. *PLOS ONE* [online]. 2022, **17**(2) [cit. 2023-03-29]. ISSN 1932-6203. Dostupné z: doi:10.1371/journal.pone.0263539
- [3] JARMUŽ, Dominika a Jerzy CHMIEL. A REVIEW OF APPROACHES TO THE STUDY OF WEATHER'S EFFECT ON ROAD TRAFFIC PARAMETERS. *Transport Problems* [online]. 2020, **15**(4, Part 2), 241-251 [cit. 2023-03-04]. ISSN 2300-861X. Dostupné z: doi:10.21307/tp-2020-063
- [4] ZHOU, Meng, Donggen WANG, Qingquan LI, Yang YUE, Wei TU a Rui CAO. Impacts of weather on public transport ridership: Results from mining data from different sources. *Transportation Research Part C: Emerging Technologies* [online]. 2017, **75**, 17-29 [cit. 2023-02-28]. ISSN 0968090X. Dostupné z: doi:10.1016/j.trc.2016.12.001
- [5] RUDLOFF, Christian, Maximilian LEODOLTER, Dietmar BAUER, Roland AUER, Werner BRÖG a Knud KEHNSCHERPER. Influence of Weather on Transport Demand. *Transportation Research Record: Journal of the Transportation Research Board* [online]. 2015, **2482**(1), 110-116 [cit. 2023-04-01]. ISSN 0361-1981. Dostupné z: doi:10.3141/2482-14
- [6] CHAPMAN, Lee. Weather and climate risks to road transport. *Infrastructure Asset Management* [online]. 2015, **2**(2), 58-68 [cit. 2023-03-04]. ISSN 2053-0242. Dostupné z: doi:10.1680/iasma.14.00032
- [7] *Dopravní infrastruktura v kraji k 1. 1. 2022* [online]. České Budějovice: Český statistický úřad, 2022 [cit. 2023-04-16]. Dostupné z: <https://www.czso.cz/csu/xc/infrastruktura-silnicni-dopravy-v-kraji-k-1-1-2022>
- [8] *Interactive map: stavby v realizaci* [online]. Praha: Ministerstvo dopravy ČR, 2023 [cit. 2023-04-16]. Dostupné z: [https://www.mdcr.cz/Mapa#/stavby?komunikace=Dalnice&kraj=JH%C4%8C&filters\[\]=StavbyRealizace&page=0](https://www.mdcr.cz/Mapa#/stavby?komunikace=Dalnice&kraj=JH%C4%8C&filters[]=StavbyRealizace&page=0)
- [9] *Úseky dálnice D3* [online]. Praha: Ředitelství silnic a dálnic ČR, c2016 [cit. 2023-04-16]. Dostupné z: <https://www.dalniced3.cz/#useky-dalnice-d3>
- [10] *Mapy ke stažení: náhledy map silniční a dálniční sítě ČR po krajích (platné k 1. 1. 2023)* [online]. Praha: Ředitelství silnic a dálnic ČR, 2023 [cit. 2023-04-16]. Dostupné z: <https://www.rsd.cz/mapy#/mapy?>
- [11] *Ekonomický vývoj: 5.6 Doprava* [online]. Praha: Český statistický úřad, 2014 [cit. 2023-04-16]. Dostupné z: https://www.czso.cz/csu/czso/13-2105-05-v_letech_2000_az_2004-5_6_doprava

- [12] Základní meteorologická terminologie. *Český hydrometeorologický ústav* [online]. Praha: ČHMÚ, c2022 [cit. 2023-03-23]. Dostupné z: <https://www.chmi.cz/predpovedi/predpovedi-pocasi/ceska-republika/meteorologicka-terminologie>
- [13] Teplota vzduchu. *Meteo aktuality: Seriózní hydrometeorologický servis* [online]. Praha: Meteo aktuality, c2023 [cit. 2023-03-23]. Dostupné z: <https://www.pocasimeteoaktuality.cz/meteorologie/meteorologicke-prvky/teplota-vzduchu/2/>
- [14] Meteorologické prvky. *Meteo aktuality: Seriózní hydrometeorologický servis* [online]. Praha: Meteo aktuality, c2023 [cit. 2023-03-23]. Dostupné z: <https://www.pocasimeteoaktuality.cz/meteorologie/meteorologicke-prvky/>
- [15] How Do Weather Events Impact Roads?. *U.S. Department of Transportation Federal Highway Administration Office of Operations* [online]. Washington, DC: U.S. Department of Transportation, 2023 [cit. 2023-04-01]. Dostupné z: https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm
- [16] *Statistika nehodovosti* [online]. Praha: Policie ČR, c2023 [cit. 2023-04-01]. Dostupné z: <https://www.policie.cz/clanek/statistika-nehodovosti-900835.aspx>
- [17] HRUBEŠ, Pavel. *Data plovoucích vozidel v aplikacích monitorování dopravy: Floating car data for traffic management applications*. V Praze: České vysoké učení technické, 2011. ISBN 978-80-01-04826-9.
- [18] *PPK - ITS: Požadavky na provedení a kvalitu inteligentních dopravních systémů na dálnicích a silnicích ve správě ŘSD ČR*. 08/2018. Praha: Ředitelství silnic a dálnic ČR, 2018. Dostupné také z: <https://www.rsd.cz/web/guest/technicke-dokumenty/ppk-a-dopravni-znaceni#zalozka-ppk>
- [19] *Geoportál silniční a dálniční sítě ČR* [online]. Praha: ŘSD ČR, 2023 [cit. 2023-03-11]. Dostupné z: https://geoportal.rsd.cz/arctgis/services/WMS_ULS/MapServer/WMSServer?
- [20] TÝC, Filip. *Fórum pro dynamické dopravní informace: Informace a data jako základ mobility*. Praha, 2020. Dostupné také z: http://www.top-expo.cz/domain/top-expo/files/smart-city/smart-city-2020/cm/prezentace/06_tyc_filip.pdf
- [21] Vývojáři dopravních aplikací získají kvalitní data pro svoji práci. *Ministerstvo dopravy* [online]. Praha: Ministerstvo dopravy ČR, c2023, 12. 12. 2019 [cit. 2023-03-23]. Dostupné z: <https://www.mdcz.cz/Media/Media-a-tiskove-zpravy/Vyvojari-dopravnich-aplikaci-ziskaji-kvalitni-data?lang=cs-CZ>
- [22] *Přehledná mapa: Aktuální dopravní informace, nehody, uzavírky a omezení* [online]. Praha: Ředitelství silnic a dálnic ČR, 2008 [cit. 2023-04-15]. Dostupné z: <https://dopravniinfo.cz/>
- [23] KOVALJOV, Michal a Pavel HRUBEŠ. *Prezentace nezávislé autority sledující kvalitu dat ŘSD FCD*. FD ČVUT. Dostupné také z: http://www.telematika.cz/download/doc/04_Prezentace_nezavisle_autority_sleduj%C3%ADci_kvalitu_dat_RSD_FCD_20191106.pdf

- [24] Dokumentace pro NDIC - DATEX II Elaborated Data Publication: Koncepty. *Národní registr dopravních informací České republiky* [online]. TamTam Research, c2022 [cit. 2023-04-02]. Dostupné z: https://registr.dopravniinfo.cz/docs/x-format/cz-ndic_d2-fcd-v1.0-cs-html/concepts.html
- [25] KLADIVO, Petr. *Základy statistiky* [online]. Olomouc: Univerzita Palackého v Olomouci, 2013 [cit. 2023-02-28]. ISBN 978-80-244-3842-9. Dostupné z: <https://geography.upol.cz/soubory/studium/e-ucebnice/978-80-244-3842-9.pdf>
- [26] NAVARA, Mirko. *Pravděpodobnost a matematická statistika*. Praha: Nakladatelství ČVUT, 2007. ISBN isbn978-80-01-03795-9.
- [27] JARUŠKOVÁ, Daniela. *Pravděpodobnost a matematická statistika*. 3. vyd. V Praze: České vysoké učení technické, 2011. ISBN isbn978-80-01-04829-0.
- [28] CRAWLEY, Michael J. *Statistics: an introduction using R*. Second edition. Chichester: John Wiley & Sons, 2014. ISBN 9781118941119.
- [29] BEVANS, Rebecca. Multiple Linear Regression. *Scribbr* [online]. Amsterdam, 2022 [cit. 2023-02-28]. Dostupné z: <https://www.scribbr.com/statistics/multiple-linear-regression/>
- [30] Celostátní sčítání dopravy 2020: Přehledová mapa. *Celostátní sčítání dopravy* [online]. ŘSD ČR, c2022 [cit. 2023-03-08]. Dostupné z: https://scitani.rsd.cz/CSD_2020/pages/map/default.aspx
- [31] How to Calculate Skewness & Kurtosis in R. *Statology* [online]. Statology Study, c2023 [cit. 2023-04-13]. Dostupné z: <https://www.statology.org/skewness-kurtosis-in-r/>
- [32] HOLČÍK, Jiří a Martin KOMENDA. *Matematická biologie: e-learningová učebnice* [online]. Brno: Masarykova univerzita, 2015 [cit. 2023-04-12]. ISBN 978-80-210-8095-9. Dostupné z: <https://portal.matematickabiologie.cz/>
- [33] *Random forest Algorithm in Machine learning: An Overview* [online]. Bangalore: Great Lakes E-Learning Services Pvt., c2013–2023 [cit. 2023-04-20]. Dostupné z: <https://www.mygreatlearning.com/blog/random-forest-algorithm/?fbclid=IwAR27O4BZfBJ40Z8V081QOJR8R2QfmurVNST88OaxMI9q2JGq0IUVMwioKmM>
- [34] *A Quick Introduction to KNN Algorithm* [online]. Bangalore: Great Lakes E-Learning Services Pvt., c2013–2023 [cit. 2023-04-20]. Dostupné z: https://www.mygreatlearning.com/blog/knn-algorithm-introduction/?fbclid=IwAR15zO-yLulK_JSG4NbQUBkb00q3oqoMz0FpflS5q_djEnuWYalaLUWI67E