VŠB – Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra informatiky

# Modelování neurčitostí ve srážko-odtokových modelech

# Rainfall-Runoff Uncertainity Modelling

2014 Martin Golasowski

**Tuto stránku nahradíte v tištěné verzi práce oficiálním zadáním Vaší diplomové či bakalářské práce.**

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

V Ostravě 7. května 2014 .....

Rád bych na tomto místě poděkoval svému vedoucímu Ing. Štěpánovi Kuchařovi za ochotu, pomoc, cenné rady a připomínky v průběhu tvorby této práce. Dále děkuji Ing. Martině Litschmannové, PhD. za její cenná doporučení a konzultace při zpracovávání statistické analýzy, která je součástí této práce. V neposlední řadě děkuji své rodině a přátelům za trpělivost a podporu a všem kolegům, kteří do této práce přispěli svými vědomostmi.

## Abstrakt

Cílem této práce je prozkoumat srážko-odtokové modely s ohledem na neurčitosti jejich vstupních parametrů, provést analýzy těchto vstupních parametrů na základě historických dat projektu Floreon+ a navrhnout a implementovat obecný mechanismus modelování neurčitostí srážko-odtokových modelů, který bude přizpůsobený ke spouštění na HPC.

**Klíčová slova:** srážko-odtokové modely, neurčitosti, modelování, simulace, jádrové odhady hustoty, MPI, paralelní optimalizace, HPC

## Abstract

The aim of this work is to investigate the rainfall-runoff models with regard to the uncertainty of their input parameters, perform the analysis of the input parameters based on historical data provided by project Floreon+ and design and implement a general mechanism for modeling the uncertainty of rainfall-runoff models, which will be adapted to run on HPC.

**Keywords:** rainfall-runoff models, uncertainity, modelling, simulation, kernel density estimate, MPI, paralel optimization, HPC

## Seznam použitých zkratek a symbolů

| | | |
|---|---|---|
| ALADIN | – | Aire Limitée, Adaption Dynamique, Development International |
| ADO.NET | – | ActiveX Data Objects |
| ČHMÚ | – | Český Hydrometeorologický Ústav |
| SCS | – | Soil Conservation Service |
| GIS | – | Geoinformační systémy |
| R | – | Statistický a vizualizační software R |
| T-SQL | – | Procedurální rozšíření jazyka SQL pro databázi Microsoft SQL Server |
| HPC | – | High performance computing |

# Obsah

# Seznam tabulek

# Seznam obrázků

# Seznam výpisů zdrojového kódu

# 1 Úvod

Motivace k předpovídání stavu počasí pramení od nepaměti z jeho význačného vlivu na náš každodenní život. Starověké civilizace byly na přízni počasí přímo existenčně závislé, neboť zemědělství představovalo téměř výhradní zdroj potravy. Podle své geografické polohy byly tyto rané civilizace odkázané na do jisté míry nepředvídatelné rozmary přírodních živlů. Pokud očekávané období dešťů přišlo například o měsíc později navzdory očekávání, došlo k ohrožení výnosů z úrody potraviny nutné pro přežití. Podobně s příchodem neočekávané bouře či prudkého deště došlo k rozvodnění řek, které mohly ohrozit obydlená místa. Postupem času se lidstvo s příchodem technologického pokroku stalo méně závislé na proměnlivosti počasí i díky schopnosti vývoj počasí předpovídat.

Ty největší výkyvy počasí však dokážou náš každodenní život ovlivnit a často i narušit. Příkladem extrémních výkyvů jsou například povodně v červenci roku 1997, kdy na stanici Lysá hora činil měsíční úhrn dešťových srážek pouze za měsíc červenec 811,9 mm, což byla více než polovina průměrného ročního úhrnu srážek, který se pohybuje okolo hodnoty 1500 mm. Následkem těchto srážek byly povodně, které zasáhly téměř celou Moravu a Slezsko [1]. Předpověď počasí je v případě extrémních výkyvů nezbytným nástrojem, který může v rámci své přesnosti poskytovat informace, na základě kterých lze varovat obyvatele před následky těchto výkyvů, případně učinit bezpečnostní opatření.

Tato diplomová práce se zabývá srážko-odtokovými modely, které slouží ke krátkodobé simulaci hydrologické situace na daném povodí podle množství spadlých dešťových srážek. Cílem této práce je prozkoumání problematiky neurčitostí parametrů takových modelů a aplikovat tyto poznatky na model Math1D.

V kapitole 2 je nastíněna problematika srážko-odtokového procesu a jeho modelování. Je zde uvedena klasifikace modelů podle jejich přístupu k parametrizaci modelovaného povodí. Dále je zde popsán model Math1D, jeho klasifikace a metody, které využívá pro simulaci a jeho vstupní parametry, s přihlédnutím k jejich možným neurčitostem. Jsou zde vybrány parametry, jejichž chybovost je vhodná pro simulaci stochastickými metodami a bude předmětem statistické analýzy.

Kapitola 3 se zabývá statistickou analýzou chyby předpovědi srážek numerického modelu Aladin, jehož výsledky jsou hlavním vstupním parametrem modelu Math1D. Je zde popsán postup analýzy, způsob získávání dat a očekávané výsledky. V první části je provedena explorační analýza dat, na kterou navazuje analýza předpokládaných závislostí a její výsledky. Dále je zde zahrnut a proveden způsob modelování chyby s přihlédnutím k jeho použití při stochastickém modelování.

Kapitola 4 popisuje metodu Monte Carlo. Je zde popsán princip jejího fungování a navrhnut způsob jejího použití pro simulaci chybovosti vstupních parametrů modelu Math1D. Dále je zde popsán způsob implementace této metody do modelu, její parametrizace a způsob její paralelizace pro spuštění v prostředí výkonných výpočetních clusterů.

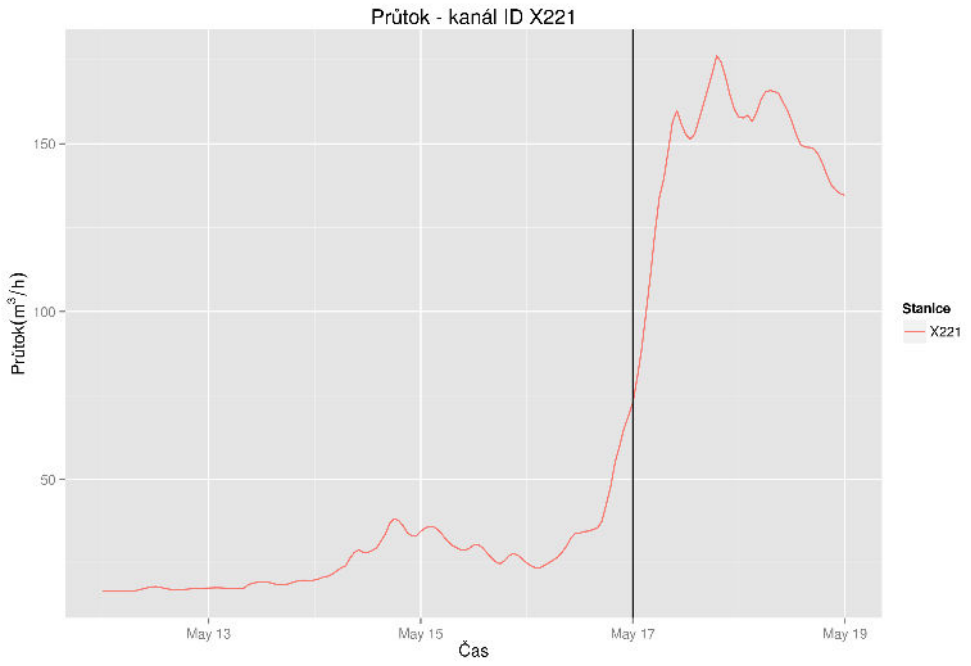
Kapitola 5 obsahuje prezentaci výsledků experimentů s různými parametry a zrychlení dosažené paralelním spouštěním.

# 2 Srážko-odtokové modely

Účelem srážko-odtokových modelů je simulace vlivu dešťových srážek na celkový odtok z povodí. Vstupem modelu jsou obyčejně srážky pro zvolený časový úsek a schematizace povodí. Výstupem modelu je pak vývoj výsledného odtoku z povodí, do kterého je zahrnut objem srážek, které po dopadu stekly do koryta řeky, podzemní odtok a příspěvky jednotlivých přítoků, příp. pramenů řeky.

Výsledný odtok je prezentován ve formě hydrogramu, což je graf, který zobrazuje vývoj průtoku vody v daném místě jako funkci času. Na obrázku 1 je vykreslen hydrogram simulovaný modelem Math1D pro stanici Sviadnov, která se nachází na spodní části toku řeky Ostravice. Hydrogram zobrazuje výsledek krátkodobé simulace pro období 12. - 19. května 2010, kdy se vyskytly prudké deště a došlo k rozvodnění řek na území Moravskoslezského kraje. Svislá čára na časové ose rozděluje simulaci podle použitých vstupních dat, prvních pět dní jsou použity měřené úhrny srážek, na další dva dny jsou použity srážky z předpovědi meteorologického modelu Aladin.

Objem srážek, které se přemění na výsledný odtok z povodí je snížen o ztráty způsobené vsakováním, vypařováním a fyzikálními vlastnostmi povodí. V praxi se využívá mnoha přístupů k modelování jednotlivých elementů srážko-odtokového procesu, přičemž specifikace srážek a vstupních parametrů povodí závisí na konkrétní implementaci. Tato práce si klade za cíl provést analýzu možných neurčitostí vstupních parametrů modelu Math1D, proto se následující sekce obsáhleji věnuje pouze modelům a přístupům využívaným v tomto modelu, další informace lze nalézt v [5], [6] a [9].




Obrázek 1: Simulovaný hydrogram pro stanici Sviadnov v období povodní v květnu 2010

## 2.1 Proces odtoku dešťových srážek

Odtok dešťových srážek ze zemského povrchu je ovlivněn mnoha faktory. V případě, kdy srážky přímo dopadají na hladinu vody v korytě je situace jednoduchá, objem srážek se přičte ke stávajícímu objemu vody v korytě. V reálné situaci však srážky dopadají na mnoho různých povrchů, různě vzdálených od cílového koryta přičemž část objemu srážek se do koryta nemusí dostat. Celkový odtok z povodí se dělí na základní tok *Base Flow* a přímý odtok *Direct runoff*. Při výskytu srážek se nejdříve zvýší přímý odtok a základní odtok na srážky reaguje se zpožděním. V praxi se odtok dělí na povrchový a podpovrchový.

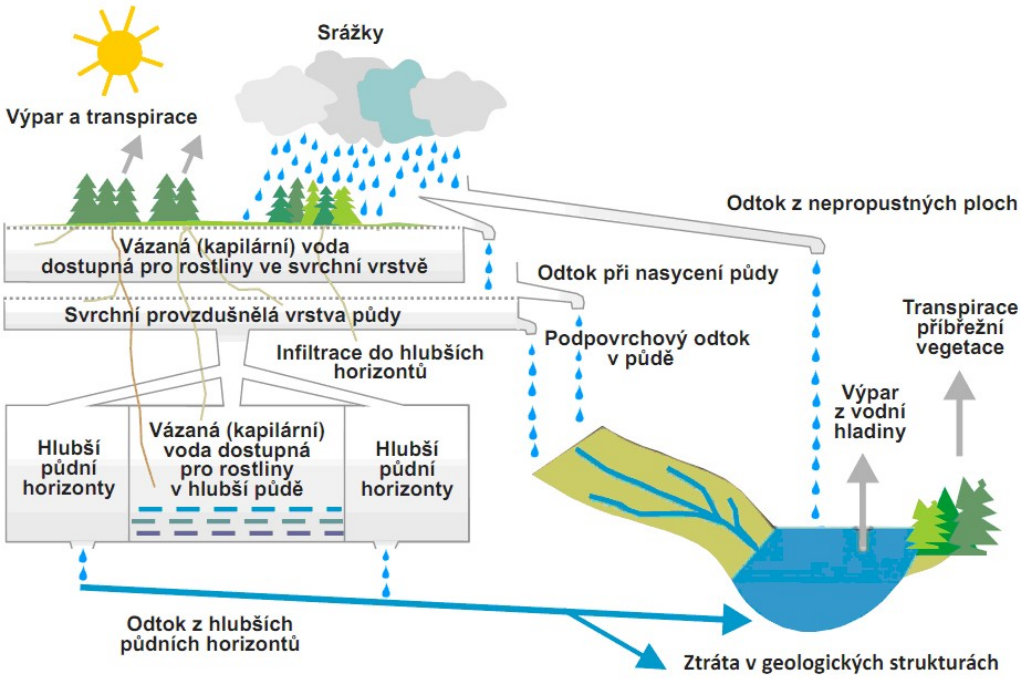
Povrchový odtok nastane v případě, kdy je půda nasycená a dochází ke stékání vody po povrchu půdy do koryta. Rychlý podpovrchový odtok je ta část vody, kterou nestačí ještě nenasycená půda vsakovat. Společně s povrchovým odtokem pak tvoří přímý odtok. Pomalý podpovrchový odtok probíhá vsakováním srážek do půdy - infiltrací a dalším prosakováním srážek do spodních vod, kde tvoří tzv. základní tok.

Kromě procesu vsakování při podpovrchovém odtoku se také vyskytuje jev opačný, a to vzlínání vody kapilárami zpět na povrch. Stejně pak při povrchovém odtoku dochází k odpařování. Je také nutno brát v potaz geologické složení půdy a prostorové proporce povodí, jako je např. velikost plochy, podélný a příčný sklon, možný výskyt jiných rezervoárů, prohlubní, kde se můžou srážky při odtoku hromadit.

Roli také hraje vegetace a využití povrchu (zástavba, parkoviště, louka...), tyto faktory společně ovlivňují jeho mechanickou drsnost. Voda může stékat po listech a větvích nebo může propadnout přímo na úroveň půdy. Je nutno brát v potaz její plochu a hustotu, ze které vyplývá míra zpětného vypařování srážek. V případě, kdy voda dopadne na listy, dochází částečně k odpařování a částečně ke stékání vody, tzv. intercepci. Přízemní vegetace, jako například tráva zvyšuje drsnost půdy a přispívá k rychlejšímu vsakování vody, dochází ke zpomalování odtoku.

Po bouři lze vidět na některými typy lesů bílá oblaka páry, které představují určitou část srážek vypařených zpět do vzduchu. Při modelování těchto jevů je nutné také vzít v potaz počáteční stav povodí. Velká část povodní vzniká díky přesycení povodí a postupnému zrychlování odtoku vody do koryt přičemž intenzita srážek zůstává přibližně stejná po delší dobu. Před výskytem dešťových srážek mohla být půda již do určité míry nasáklá, v korytě se již vyskytovalo určité množství vody, teplota a vlhkost vzduchu mohla ovlivnit míru vypařování srážek.

Součástí procesu je také odezva povodí, tedy způsob, jakým se projeví srážky spadlé na horní části povodí na průtok na nižších částech. Svou roli hraje doba, za kterou se zvýšený průtok projeví na zbytku povodí a s jakou intenzitou [10][6][5].




Obrázek 2: Schéma srážko-odtokového procesu [3]

### 2.1.1 Klasifikace modelů

Tato práce se zabývá matematickými srážko-odotokovými modely. Existují však i historické modely pracující např. se zmenšeným fyzickým modelem povodí[6] nebo modely, využívající k modelování toku elektrického proudu[6], tyto však v praxi svou přesností daleko zaostávají za matematickými modely, které se díky technologickému rozvoji stávají stále přesnějšími [6]. Obecně lze srážko-odtokové matematické modely rozdělit na *stochastické* a *deterministické*. Pokud jsou vstupní parametry a procesy v modelu předem známy a nezahrnují jakoukoli náhodnost, jedná se o model *deterministický*. Stejný model pro stejné vstupní parametry tak vždy podá ekvivalentní výsledek. Pokud se však model snaží o modelování náhodného jevu a některé vstupní parametry nejsou předem známy, jedná se o model *stochastický* [6] [10].

Podle způsobu specifikace rozdílů mezi jednotlivými částmi povodí se modely dělí na *koncentrované*, *semi-distribuované* a *distribuované*. *Koncentrované* modely obecně zanedbávají nebo ignorují prostorové náležitosti povodí a využívají pro celé povodí společných aproximovaných parametrů.*Distribuované* oproti tomu počítají do jisté míry s odlišnostmi mezi jednotlivými součástmi povodí, detailní specifikace parametrů povodí pak závisí na kvalitě dostupných dat v GIS. Přílišná podrobnost modelu však může ovlivňovat rychlost výpočtu takového modelu, na druhou stranu aproximace většího množství parametrů může ovlivňovat přesnost reprezentace fyzikálních procesů na povodí. Míra detailní specifikace parametrů závisí na mnoha faktorech (dostupnost dat o průtoku na jednotlivých

kanálech, dostupnost dat o srážkách, členitost povodí, apod.). *Semi-distribuované* modely jsou pak takové modely, které kombinují oba výše zmíněné přístupy a které pracují s povodím rozčleněným na jednotlivé odtokové plochy (subpovodí), které představují oblast s homogenními fyzikálními parametry. Jejich výhodou je zmenšení výpočetní náročnosti aproximací prostorové členitosti povodí.

Dále se srážko-odtokové modely dělí na *empirické* a *konceptuální*, podle způsobu, jakým byl určen proces, který má model simulovat. *Konceptuální* model (např. kinematická vlna) modeluje fyzikální, chemické nebo biologické procesy, působící na vstupní parametry. *Empirické* modely (např. Snyderův jednotkový hydrogram) využívají pozorovaných změn vstupních dat, bez hlubší znalosti procesu, který tyto změny způsobuje - jsou založeny na pozorování skutečnosti.

Poslední kategorií jsou modely, jejichž parametry můžeme získat měřením, například odebráním vzorku půdy nebo měřením teploty vzduchu. Pokud parametry modelu nelze změřit, je označen anglickým termínem *fitted-parameter* a hodnota parametru je určena expertním odhadem nebo kalibrací. Podle způsobu použití jsou srážko-odtokové modely děleny dále na *krátkodobé* a *dlouhodobé*. *Krátkodobé* modely jsou používány pro předpovědi v horizontu několika dní, zejména při výskytu silných srážek, *dlouhodobé* modely jsou pak využívány pro předpověď stavů vody bez ohledu na výskyt srážek v řádu měsíců [6], [5].

Srážko-odtokové modely využívají velké množství metod. V následujících kapitolách budou zmíněny jen ty, které využívá model Math1D, kterým se tato práce zabývá. Popisem modelu se zabývá podkapitola 2.2

### 2.1.2 Jednotkový hydrogram

Metoda jednotkového hydrogramu je jednou z metod využívaných pro stanovení povrchového a rychlého odtoku v nenasycené půdě. Tato metoda, kterou v roce 1932 představil hydrolog Sherman, spočívá v aplikaci proporcionálního tvaru nárustu objemu vody na již existující odtok [6] [10].

Jednotkový hydrogram představuje základní odtok z povodí generovaný jednotkovým objemem srážek, dopadlým rovnoměrně a s rovnoměrnou rychlostí na danou plochu za daný časový úsek. Hodnoty jednotkového diagramu určují tvar odezvy povodí na srážkovou událost.

Příchozí odtok se konvolucí skládá s jednotkovým hydrogramem, který tvaruje výsledný hydrogram. Metoda tak efektivně modeluje vliv intezity srážek na odtok z povodí v čase. Způsoby modelování pomocí jednotkového diagramu se dělí na *konceptuální* a *empirické*. Mezi *konceptuální* metody patří např. metoda lineární kaskády, Nashova kaskáda apod. Mezi *empirické* metody, využívající zdroje GIS patří Snyderova metoda a SCS-UH metoda [6].

### 2.1.3 Kinematická vlna

Alternativní metoda pro stanovení přímého odtoku je metoda kinematické vlny. Oproti jednotkovému diagramu, který je metodou empirickou, tato metoda je čistě konceptuální, založená na modelování fyzikálních vlastností povrchového odtoku vody. Základní tok je

roven objemu povrchového odtoku. Řešení soustavy diferenciálních rovnic pak simuluje nerovnoměrný odtok srážek z povrchu do daného kanálu. Výsledný hydrogram pak představuje vývoj toku v daném kanálu, ovlivněný fyzikálními vlastnostmi odtokové plochy. V praxi se tato metoda používá kromě modelování odtoku z ploch, příslušícím jednotlivým korytům, tak i vlivu jednotlivých přítoků, které přispívají k odtoku z daného povodí.

Metoda bere v potaz především náklon (podélný, příčný), rozměry a drsnost odtokových ploch a koryt samotných. Podle druhu modelu se této metody mimo jiné používá i k modelování příspěvků z částí městské kanalizační sítě, která může při výskytu srážek významným způsobem ovlivnit vývoj hydrologické situace. Pro popis drsnosti povrchu odtokových ploch se využívá tzv. *Manningův koeficient*, jehož hodnota je pro různé typy povrchů určena hydrologickými tabulkami [6].

### 2.1.4 Efektivní srážky (SCS - CN)

Tato metoda byla vyvinuta americkou agenturou pro ochranu půdy (SCS - Soil Conservation Service). Oproti metodám předešlým, které řeší časové rozložení vlivu srážek a přítoků na odtok z povodí, je tato metoda využívána v modelu Math1D pro výpočet tzv. efektivních srážek, tedy objemu srážek sníženého o ztráty způsobené intercepcí, evapotranspirací, povrchovou a podpovrchovou retencí.

Metoda při výpočtu zahrnuje mimo samotný srážkový úhrn také úroveň počátečního nasycení [1] a parametr označovaný jako číslo CN křivky.

Hodnota počátečního nasycení udává srážkový úhrn, který může dopadnout na povrch půdy a vsákne se, aniž by generoval odtok.

Hodnota CN se stanovuje jako funkce geologických vlastnosti půdy, na kterou srážky dopadají, jejího pokrytí vegetací a způsobu jejího využití. Obecně se určuje z hydrologických tabulek, které jsou vydávány konkrétně pro každou modelovanou oblast, příslušnou autoritativní institucí.

V praxi se pro stanovení objemu efektivních srážek mimo SCS - CN používají i další metody, mezi které patří například *Green and Ampt metoda* nebo *Initial and constant rate loss* [6].

## 2.2 Model Math1D

Matematický model Math1D je srážko-odtokový model pro krátkodobé předpovědi vyvíjený na katedře aplikované matematiky Fakulty elektrotechniky a informatiky na VŠB - TU Ostrava v rámci projektu *Floreon+* [11]. Simuluje vývoj hydrologické situace v čase na kanálech definovaných schematizací zvoleného povodí. Pro výpočet efektivních srážek využívá výše zmíněné SCS-CN metody. Dále využívá metody kinematické vlny pro výpočet vlivu příspěvku z jednotlivých přítoků v čase a jednotkového hydrogramu pro rozložení vlivu efektivních srážek na odtok z povodí v rámci zadaného časového intervalu.

[1] v angl. literatuře označováno jako *Initial Abstraction* příp. *Initial Loss*

| Název | Popis |
|---|---|
| **N** | Drsnost (Manningův koeficient) |
| **Length** | Délka kanálu |
| **Slope** | Podélný slon koryta |
| **Bank slope** | Sklon břehů lichoběžníkového profilu koryta |
| **Width** | Šířka koryta |

Tabulka 1: Parametry kanálů ve schematizaci

### 2.2.1 Implementace

Původní implementace byla vyvinuta pomocí nástroje Matlab, v průběhu zpracovávání této diplomové práce byla provedena konverze stávajícího kódu do jazyka C++. Konverze původního kódu byla předmětem diplomové práce diplomanta Bc. Radima Vavříka, s nímž jsem na konverzi pracoval a výsledný kód pak využíval pro implementaci neurčitostních simulací [8].

### 2.2.2 Rozhraní

Model je implementován jako konzolová aplikace. Pomocí argumentů lze ovlivnit spuštění kalibrace modelu a spuštění neurčitostní simulace. Aplikace čte nastavení z XML souboru, který obsahuje cestu k XML souborům obsahujícím schematizaci povodí, měřené průtoky na vybraných stanicích, srážky a cestu pro ukládání výsledných hydrogramů.

### 2.2.3 Schematizace povodí

Schematizace je XML soubor popisující prostorovou organizaci daného povodí. Každé povodí má své prameny a přítoky, na jeho konci se nachází závěrný profil, do něhož vtéká veškerá voda z příslušného povodí.

Povodí je rozděleno na jednotlivé kanály, které jsou obyčejně ohraničeny jednotlivými soutoky a prameny, každému kanálu také přísluší počáteční a koncová stanice, přičemž některé z nich mohou odpovídat reálným hlásným profilům, což jsou hydrologické měřící stanice, které zaznamenávají reálný průtok. Nejčastěji je měřící stanice namapována na závěrný profil. Každý kanál má parametry popisující jeho fyzikální vlastnosti, jejich výčet je uveden v tabulce 1.

Každému kanálu přísluší tzv. subpovodí (subbasin), které představuje plochu, na kterou dopadá déšť stékající do příslušného kanálu. Ke každému subpovodí přísluší meteorologická stanice poskytující data o aktuální meteorologické situaci, v tomto případě data o hodinových srážkových úhrnech. Tyto meteorologické stanice nemusí nutně odpovídat fyzickým stanicím, důležité je mapování jejich polohy na stanice, pro které jsou vydávány předpovědi modelu Aladin. Parametry popisující vlastnosti subpovodí jsou uvedeny v tabulce 2.

| Název | Popis |
|---|---|
| **Slope** | Sklon subpovodí |
| **Base flow** | Základní tok |
| **Area** | Plocha subpovodí |
| **Cn** | CN křivka |
| **N** | Drsnost (Manningův koeficient) |
| **InitialAbstraction** | Počáteční nasycení |

Tabulka 2: Parametry subpovodí ve schematizaci

## 2.3 Parametry schematizace povodí

Vstupními parametry modelu Math1D jsou srážky, měřené průtoky na vybraných stanicích a schematizace povodí. Nepřesnost každého z nich může negativním způsobem ovlivňovat vypovídací hodnotu výsledku simulace a může mít více příčin. Navržení způsobu jejího modelování spočívá v analýze jejího původu a navržení způsobu vyčíslení. Následující kapitoly se zabývají analýzou jednotlivých skupin parametrů podle druhu jejich neurčitosti.

### 2.3.1 Prostorové parametry

Parametry jako podélný a příčný sklon a další rozměry jsou získány měřením nebo jsou stanoveny expertním odhadem. Neurčitost těchto parametrů může být způsobena chybou měřicích přístrojů, nesprávnou metodikou nebo chybou při zpracování dat. Vyčíslit takovou chybu lze pouze pokud jsou k dispozici data, která považujeme za správná a se kterými můžeme zkoumaná data porovnat. Za správná, přesnější data lze považovat např. údaje získané přesnějšími přístroji, autoritativní institucí nebo přesnější metodikou.

V tomto případě se jedná o geografická data popisující prostorové náležitosti jednotlivých kanálů povodí. Tato data jsou získána měřením, které provádí příslušná instituce (Povodí Odry), která je pak poskytuje pro použití v GIS. Je nepravděpodobné, že by mapování prostorových náležitostí povodí provádělo více nezávislých subjektů, které by se významně lišily přesností získaných dat. Neexistují tedy data, se kterými by se dala původní geografická data porovnat, vyčíslení chyby takovým způsobem je zřejmě nereálné.

Další možností jak vyčíslit chybu a odhadnout tak neurčitost měřených dat je získání údajů o přesnosti použité instrumentace, jako například kalibrované korekční křivky, údaje o rozlišení apod. Tyto údaje však nemusí být vždy k dispozici a vzhledem k množství používaných přístrojů, kdy každý přístroj má svou vlastní korekční křivku, by modelování neurčitosti bylo značně neefektivní a zdlouhavé. Prostorové parametry povodí mají oproti jiným parametrům srážko-odtokového modelu zanedbatelný vliv na výslednou chybu předpovědi, předpokládá se tedy, že naměřená data jsou správná, korigována o možné chyby. Neurčitost prostorových parametrů povodí tedy lze zanedbat.

### 2.3.2 Fyzikální parametry

Parametrem popisujícím vlastnosti subpovodí je číslo CN křivky. Hodnota parametru se určuje z tabulek sestavených organizací SCS a vydaných v dokumentu *Technical Release 55, TR-55* [6] [9]. V současné době se však používají jiné metody pro stanovení CN křivek, zejména proto, že původní tabulky nezahrnují všechny možné typy půdy a jejího využití. Obecně se CN křivky určují podle typu povrchu, vegetace, její hustoty a typu. Neurčitost tohoto parametru je zatížena jednak metodikou určování samotné CN křivky a také možnou chybou měření při určování parametrů pro výpočet CN křivky. Metoda SCS-CN se používá desítky let s poměrně velkou přesností s původními parametry, míra neurčitosti tohoto parametru je proto relativně malá. Modelování jeho neurčitosti je proveditelné porovnáváním výsledků založených na původních parametrech z *TR-55* a některých nových metod založených na automatickém zpracování satelitních snímků pomocí GIS. Rozbor jednotlivých metod je nad rámec této práce, více informací v [9].

Manningův koeficient drsnosti N je využíván v modelu kinematické vlny v manningově rovnici pro výpočet průměrné rychlosti odtoku kapaliny z otevřeného kanálu. Je určen hydrologem, empiricky, analýzou tvaru koryta a materiálu, ze kterého je tvořeno. Jeho hodnota vyjadřuje odpor, který je kladen proudění vody korytem samotným, kameny či jinými objekty a vegetací. Jeho hodnota se dá určit z hydrologických tabulek, které stanovují hodnoty manningova koeficientu pro různé typy povrchů, viz [5]. Kromě tabulkových hodnot také existují i empirické metody stanovení hodnoty koeficientu, které by se daly využít pro analýzu neurčitosti tohoto parametru. Opět se zde vyskytuje problém, kdy není známa správná referenční hodnota parametru, ze které by se dala odvodit chyba plynoucí z metodiky stanovení použité hodnoty parametru. Určení správné hodnoty parametru N je tedy otázkou kalibrace modelu. Jeho hodnotu je nutno přizpůsobit podmínkám v daném povodí případně provést kalibraci pro získání nejvhodnější možné hodnoty. Pokud není známo pravděpodobnostní rozdělení chyby, se kterou je hodnota parametru stanovena, není tento parametr vhodný pro neurčitostní simulace.

## 2.4 Srážky

Srážky jsou kapky vody, vzniklé kondenzací vodní páry v atmosféře a dopadnuvší na zem v pevné nebo v kapalné podobě. Tato práce se zabývá srážkami dešťovými, pokud nebude řečeno jinak. Jednotka, ve které se množství srážek měří, je výška vodního sloupce v milimetrech za hodinu *mm/h* a představuje tzv. srážkový úhrn. Oficiální definice jednotky srážkového úhrnu zní:

> Množství srážek se udává v milimetrech (s přesností na desetiny milimetru). Je to výška, do které by na povrchu země sahaly spadlé (usazené) srážky ve formě vody nebo voda vzniklá rozpuštěním tuhých srážek, kdyby se nevsákla do půdy, neodtekla ani neodpařila. Výšce srážek 1 mm odpovídá množství vody 1 litr na 1 $m^2$ vodorovné plochy [2].

### 2.4.1 Měřené srážky

Srážky jsou modelu předávány jako vektory diskrétních hodnot úhrnu srážek za dané časové období pro každou stanici. Každému subpovodí přísluší jedna stanice. Nemusí se nutně jednat o fyzickou meteorologickou stanici, jedná se zejména o geografickou polohu místa, na které je vydávána předpověď srážkového úhrnu.

Nejčastější využití modelu Math1D je v režimu 5+2 dny, na krátkodobou předpověď. Simulace je provedena na 7 dní, z toho srážky za prvních 5 dní jsou měřené a zbývající 2 dny jsou predikované. Měřené srážkové úhrny dodává Povodí Odry z měřicích stanic.

Neurčitost v měřených srážkách může být opět způsobena chybou měření a zkreslením dat při jejich přenosu a ukládání. Modelování neurčitosti způsobené chybou měření by mohlo zahrnovat například aplikaci korekční křivky každého použitého měřicího přístroje na každou jím dodanou hodnotu. Lze předpokládat, že data dodávána Povodí Odry jsou již korigována o chybu měření a vykazují tedy nejlepší možnou přesnost, a že nedochází k významným odchylkám při přenosu dat do databáze projektu Floreon+. Neurčitost takových dat lze považovat za irelevantní.

### 2.4.2 Predikované srážky

Predikované srážkové úhrny jsou dodávány numerickým modelem Aladin. Model Aladin slouží pro generování krátkodobých předpovědí počasí, obyčejně na 54 hodin dopředu. Model je vyvíjen v rámci stejnojmenného konsorcia ALADIN založeného v roce 1990 ve Francii ve spolupráci s mnoha vědeckými a meteorologickými pracovišti ze střední a východní Evropy. Provoz modelu Aladin ve specifikaci pro Českou Republiku zajišťuje Český hydrometeorologický úřad.

Predikce je založena na řešení systému diferenciálních rovnic řešených spektrální metodou na omezené oblasti. Vstupem modelu jsou data z meteorologických stanic a data z globálního modelu počasí. Výstupem jedné predikce jsou údaje o oblačnosti, teplotě, směru a rychlosti větru a relativní vlhkosti ve 2 metrech. Jedna predikce se vydává na 54 hodin dopředu s hodnotami po 6hodinových krocích. Česká specifikace modelu využívá síť pokrývající plochu ČR čtverci o straně 9 km [7].

Přesnost předpovězených srážkových úhrnů na jednotlivých stanicích lze vyčíslit jejich porovnáním s naměřenými srážkovými úhrny. Rozdíl naměřené a předpovězené hodnoty je absolutní chyba, jejíž pravděpodobnostní rozdělení lze při dostatečně velkém počtu dat relativně přesně odhadnout.

## 2.5 Neurčitost v simulaci

Účelem meteorologického modelu je simulace meteorologických procesů s pokud možno co největší přesností. Tu je možné vyčíslit jako odchylku simulované hodnoty od měřené hodnoty dané veličiny. Můžou ji ovlivňovat jak použité metody k modelování dílčích procesů, tak neurčitost parametrizace těchto metod. Běžný výstup modelu může být zatížen chybou, avšak bez bližší znalosti povahy této chyby není možné posoudit přesnost konkrétní simulace. Pokud by však byl výstup modelu doplněn o informaci o intervalu,

v jakém se může chyba pohybovat a s jakou pravděpodobností, je možno uživateli modelu podat informaci, která mu umožní interpretovat výsledky simulace s větší přesností.

Meteorologické modely modelují data na základě vstupní parametrizace modelované domény a vstupních dat. Parametrizace a vstupní data jsou buď určena empiricky nebo jsou výsledkem jiného modelování. Provedením citlivostní analýzy lze identifikovat vliv jednotlivých vstupů modelu na jeho výsledky a jejich neurčitost. Modelování neurčitosti lze provést, pokud je možno určit chybu vybraného vstupního parametru. Statistickou analýzou chyby pak lze určit její pravděpodobnostní rozdělení a toto použít pro stochastické modelování neurčitosti výsledku simulace. Pokud lze určit i interval, ze kterého chyba parametru pochází, je možno pro stochastické modelování použít metodu Monte Carlo.

Tato práce je zaměřena na srážko-odtokové modely, jejichž vstupem jsou data o dešťových srážkách. Pokud je model využíván pro predikování vývoje hydrologické situace, data o srážkách taktéž pocházejí z modelování meteorologických jevů v atmosféře a mohou být zatížena určitou chybou. Pravděpodobnostní rozdělení chyby meteorologického modelu lze určit poměrně dobře, meteorologická data se na mnoha místech zaznamenávají již mnoho desítek let a není problém výsledky modelu porovnávat s měřenými daty. Analýzou chyby předpovědi srážek se zabývá následující kapitola, aplikací jejích výsledků do modelu Math1D se zabývá kapitola 4.

# 3 Statistická analýza chyby předpovědi srážek

Náplní této kapitoly je statistická analýza chyby, jejíž hodnota je stanovená jako rozdíl **měřené** a **predikované hodnoty srážek**. Cílem analýzy je potvrdit nebo vyvrátit závislost chyby na stanici, pro kterou byla předpověď vydána, dále na délce předpovědi a na vybraných intervalech predikované intezity srážek.

Nejprve je provedena explorační analýza, kde jsou určeny hodnoty výběrových charakteristik, je provedeno ověření normality dat a je určen intervalový odhad mediánu chyby.

Následně je pomocí analýzy rozptylu stanovena statistická významnost výše zmíněných druhů závislostí. Pokud se prokáže, že existuje statisticky významná závislost chyby na některém z výše zmíněných kriterií, bude provedena explorační analýza chyby rozdělené podle daného kritéria, které se následně bude brát v potaz při jejím modelování.

Dále je nutné stanovit způsob modelování závislosti chyby na intenzitě predikovaných srážek za účelem neurčitostní simulace. Jedním ze způsobů modelování závislosti daných veličin je pomocí lineární regrese. Je zde proveden bodový odhad parametrů lineárního regresního modelu a následně je provedena jeho verifikace.

Dalším způsobem modelování závislosti chyby na intenzitě predikovaných srážek je neparametrický odhad hustoty pravděpodobnosti. Ten je zde proveden pomocí metody jádrových odhadů.

Na závěr je provedeno srovnání výsledků výše zmíněných přístupů k modelování a zvolen je ten, jehož výsledky budou použity pro implementaci neurčitostních simulací.

## 3.1 Dostupná data

Databáze projektu Floreon+ obsahuje jak předpovídané srážkové úhrny tak měřené srážkové úhrny pro vybrané meteorologické stanice na území Moravskoslezského kraje. Pro model Math1D byly v době tvorby této práce dostupné schematizace pro povodí vybraných řek, jmenovitě Ostravice, Olše, Odry a Opavy. Pro analýzu byla použita data ze stanic náležících povodím těchto řek, jejichž povodí jsou popsané v jednotlivých schematizacích.

Měřené srážkové úhrny jsou měřeny s rozlišením 0,1 mm/h, v potaz jsou proto brány pouze chyby předpovědi, pro které je naměřená hodnota stejná nebo větší než 0,1 mm/h.

Předpovědi modelu Aladin jsou v databázi dostupné v datových sadách generovaných každých 12 hodin. Každá datová sada obsahuje pro jednu stanici 8 hodnot, pro každý 6hodinový úsek na 2 dny dopředu. Předpovídané srážkové úhrny byly v databázi projektu Floreon+ rovnoměrně rozděleny na dílčí hodinové úhrny, aby bylo zachováno stejné časové rozlišení s měřenými srážkovými úhrny. Analýza proto počítá s časovým rozlišením jedné hodiny, pokud není řečeno jinak.

## 3.2 Explorační analýza

Tato podkapitola se zabývá explorační analýzou chyby předpovědi srážek. Je zde popsán způsob, jakým byla analyzovaná data získána a zpracována. Jsou zde určeny hodnoty

vybraných výběrových charakteristik. Je zde také ověřena normalita dat, na základě které jsou pak voleny další statistické testy a je proveden intervalový odhad mediánu chyby.

### 3.2.1 Získávání dat

Data o srážkových úhrnech jsou dostupná v SQL databázi projektu Floreon+ ve společné tabulce pro měřené a předpovídané hodnoty. Chyby předpovědi pro všechny časy, sady předpovědí a stanice bylo nutno získat ve formátu CSV, vhodném pro zpracovávání statistickým software R. Každé hodnotě chyby přísluší čas, pro který byla předpověď vydána (offset), čas vydání předpovědi, původní měřená hodnota a stanice, pro kterou byla předpověď vydána.

Pro oddělení kódu samotných SQL dotazů a způsobu ukládání získaných dat byl proces rozdělen na dílčí části - uloženou proceduru, která pracuje s tabulkami databáze a konzolovou aplikaci, která ukládá získaná data do CSV souborů na disk. Uložená procedura *CompareDatasets* využívá prostředků T-SQL, poskytovaných databázovým strojem. Získává z databáze chyby pro zadanou délku offsetu (délku předpovědi označovanou jako $T + n$, kde $n$ je počet hodin od času vydání předpovědi), pro stanici, poskytovatele dat a zadaný rozsah naměřeného srážkového úhrnu. Procedura nejprve pomocí kurzoru vytáhne předpovídané hodnoty pro zadané parametry. Předpovězené hodnoty jsou procházeny v cyklu, ve kterém je ke každé korespondující hodnotě získána měřená hodnota srážek. Měřené hodnoty srážek mohou být v databázi uloženy vícekrát pro jeden čas a jednu stanici. Procedura bere hodnotu, která byla přidána nejpozději z toho důvodu, že poskytovatel dat měření pro jeden čas v průběhu několika hodin postupně zpřesní a vydá novou hodnotu. Jako nejpřesnější hodnota se proto bere ta, která byla do databáze uložena jako poslední. Hodnoty jsou dále v cyklu porovnávány, vyřazeny jsou takové předpovězené hodnoty, které nemají měřenou hodnotu. Dále jsou vyřazeny všechny případy, kdy je předpovězená i naměřená hodnota nulová, z toho důvodu, že analýza se primárně nezabývá četností hodnot vzhledem k zadaným parametrům. Dále je ověřeno, zda získaná naměřená hodnota vyhovuje zadanému limitu a nakonec je vypočtena chyba předpovědi, dle vztahu uvedeného v definici 3.1

**Definice 3.1** *Vztah pro výpočet chyby předpovědi srážek modelu Aladin*

$$S_{delta} = S_m - S_p$$

kde, $S_{delta}$ je chyba předpovědi srážkového úhrnu v mm/h, $S_m$ je měřený a $S_p$ je predikovaný srážkový úhrn v mm/h.

Jednotlivé výsledky se ukládají do dočasné tabulky, jejíž obsah je po skončení procedury vypsán na výstup. Procedura vrací výsledky pro všechny časy predikce v databázi pro které je dostupná predikovaná i měřená hodnota.

Konzolová aplikace spouští uloženou proceduru pro zadanou množinu časů, limitů a stanic. Pomocí knihovny *ADO.NET* poté vyčítá výsledky po řádcích, jejichž jednotlivé položky jsou převedeny na znakový řetězec oddělený středníkem. Aplikace ukládá výsledky do složek podle poskytovatele dat a limitů měřených srážek. Výsledky jsou uloženy

| Výběrová charakteristika | Hodnota |
|---|---|
| Velikost výběrového souboru | 70 804 |
| Průměr (mm/h) | 0,5 |
| Dolní kvartil (mm/h) | -0,1 |
| Medián (mm/h) | 0,1 |
| Horní kvartil (mm/h) | 0,6 |
| Minimum (mm/h) | -9,2 |
| Maximum (mm/h) | 38,2 |
| Šikmost | 6,2 |
| Špičatost | 71,3 |
| Směrodatná odchylka (mm/h) | 1,6 |

Tabulka 3: Výběrové charakteristiky chyby predikce srážek

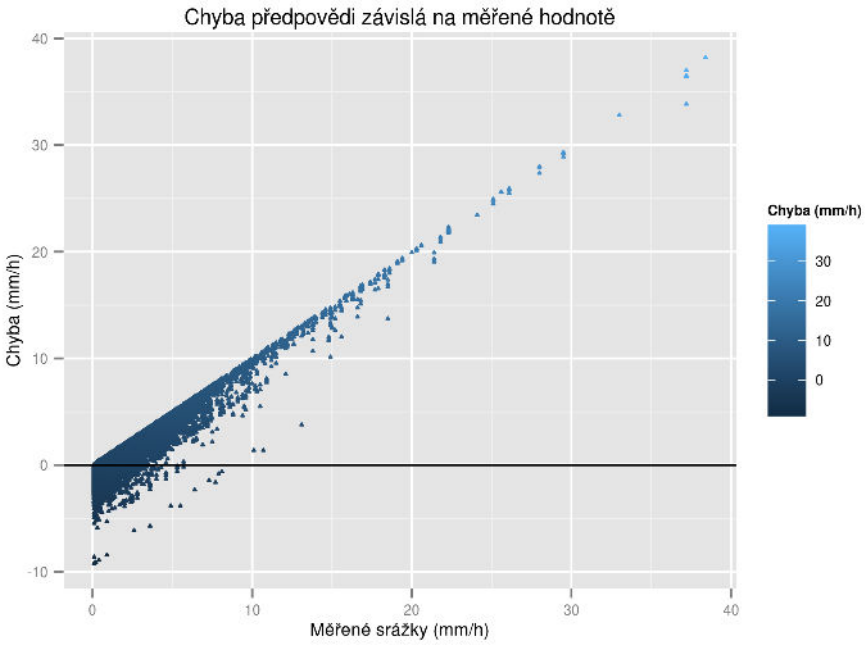
do souborů podle jednotlivých offsetů. Každý soubor obsahuje chyby pro všechny zadané stanice a všechny zadané časy pro každý offset.

### 3.2.2 Výběrové charakteristiky chyby

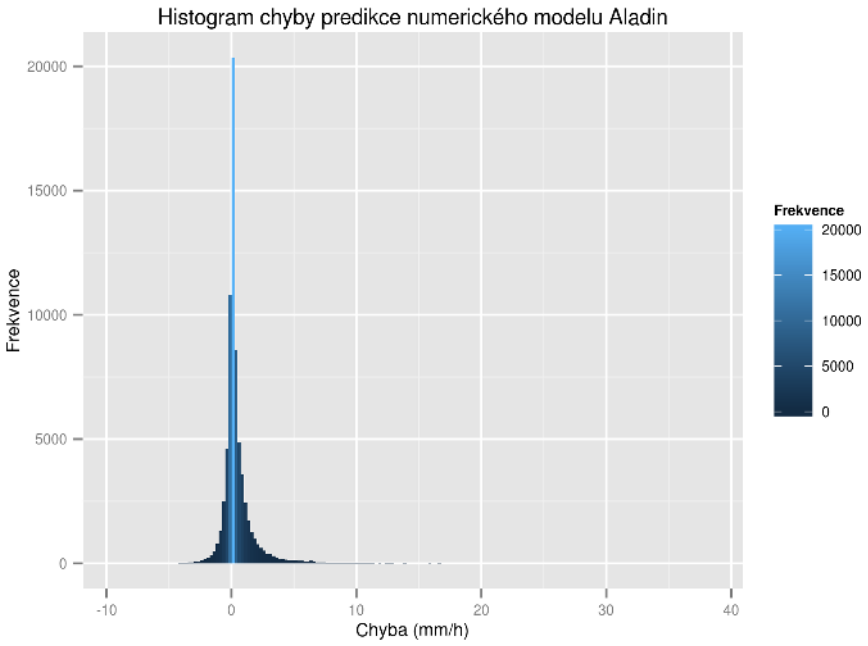
Prvním krokem statistické analýzy je explorační analýza dat. V tabulce 3 jsou uvedeny hodnoty některých výběrových charakteristik chyby. Na obrázku 3 se nachází bodový graf, zobrazující závislost velikosti chyby na skutečných měřených srážkách.

Hodnota mediánu výběrového souboru je kladná, což poukazuje na větší počet výskytů kladných chyb než záporných, jinými slovy model má tendenci podhodnocovat, což jde dobře vypozorovat z grafu na obrázku 3, kde lze vidět, že model nadhodnocuje zhruba do intenzity 5 mm/h, ve výjimečných případech do 7 mm/h. Hodnoty jsou také rozmístěné relativně blízko okolo svého průměru, což dokazuje vysoká špičatost, což může značit, že velké množství chyb je relativně malých, avšak vyskytují se zde i velké odchylky, zejména se zvětšující se intenzitou skutečných srážek. Výskyt velkého množství kladných odlehlých pozorování také potvrzuje kladná odchylka průměru od mediánu.

Přes vysoké nahuštění hodnot okolo jejich průměru zde lze také pozorovat poměrně velké hodnoty odlehlých pozorování. Velkým výchylkám také odpovídá poměrně velká hodnota výběrové směrodatné odchylky (1,6 mm/h), která je přibližně trojnásobkem průměru chyby. Minimální hodnota chyby byla $-9,1$ mm/h, v tomto případě model predikoval hodnotu $9,1$ mm/h a ve skutečnosti nepršelo vůbec. Oproti tomu maximální hodnota chyby $38,2$ mm/h, kdy pršelo velice intenzivně a model predikoval nulový srážkový úhrn. Tyto výchylky chyby se objevují zjevně zřídkakdy, avšak při neurčitostní simulaci je nutné brát jejich hodnoty v potaz.




Obrázek 3: Chyba předpovědi závislá na měřené intenzitě srážek




Obrázek 4: Histogram chyby předpovědi pro výběrový soubor o velikosti $n = 70804$

### 3.2.3 Ověření normality dat

Statistické testy s větší vypovídací hodnotou většinou mají normalitu testovaných dat jako jeden z předpokladů. Z toho důvodu je potřeba před pokračováním ověřit, zda testovaná data pocházejí z normálního rozdělení. Toto lze ověřit vizuální analýzou histogramu a posouzením jejich šikmosti a špičatosti, resp. exaktně s využitím statistických testů [12].

Tvar histogramu na obrázku 4 zjevně neodpovídá tvaru hustoty pravděpodobnosti normálního rozdělení. Hodnota špičatosti dat se nenachází v intervalu $\langle -2; 2 \rangle$, což je obecné pravidlo pro posouzení špičatosti dat. Z toho lze usuzovat, že data nepochází z normálního rozdělení, což je nutné brát v potaz při výběru a používání statistických testů.

### 3.2.4 Intervalový odhad mediánu chyby

Jednou ze základních metod statistické indukce je intervalový odhad daného parametru populace. Pomocí ní lze z výběrového souboru odhadnout interval, ve kterém se daný parametr pohybuje, a lze tak odhadnout hodnotu parametru pro celou populaci sledované proměnné. Pokud hodnoty ve výběrovém souboru dat pocházejí z normálního rozdělení, lze určit například intervalový odhad střední hodnoty sledované veličiny. Pokud data nepocházejí z normálního rozdělení nebo obsahují velké množství odlehlých pozorování, které nelze vyloučit, je možné použít tzv. robustní odhady, mezi které patří například intervalový odhad mediánu [12].

V tomto případě je k dispozici výběrový soubor chyby predikce srážkového úhrnu, na základě kterého lze provést robustní odhad mediánu populace, která je reprezentována chybami všech predikcí srážek numerického modelu Aladin. Odhad dolní a horní meze intervalu mediánu s 95 % spolehlivostí je dán níže uvedeným vztahem, kde $\hat{x}_p$ jsou dané $100p\%$ výběrové kvantily a $n$ je velikost výběrového souboru. Dosazením hodnot za všechny proměnné a provedením nezbytných úprav je obdržen výsledný intervalový odhad.

Vztah pro výpočet intervalového odhadu mediánu [12]

$$\left\langle x_{0,5} - 1,57\frac{(\hat{x}_{0,75} - \hat{x}_{0,25})}{\sqrt{n}}; x_{0,5} + 1,57\frac{(\hat{x}_{0,75} - \hat{x}_{0,25})}{\sqrt{n}} \right\rangle$$

Po dosazení:

$$\left\langle 0,1 - 1,57\frac{(0,6 - (-0,1))}{\sqrt{70804}}; 0,1 + 1,57\frac{(0,6 - (-0,1))}{\sqrt{70804}} \right\rangle \sim \left\langle 0,095; 0,104 \right\rangle$$

Medián chyby predikce srážek se tedy bude s 95 % spolehlivostí pohybovat v intervalu $\langle 0,095; 0,104 \rangle$ (mm/h), což může značit poměrně velkou přesnost modelu. Hodnota mediánu chyby pohybující se okolo hodnoty 0,1 mm/h ukazuje na mírné podhodnocení skutečných srážkových úhrnů.

### 3.2.5 Vliv odlehlých pozorování na statistické testy

Je obecně známo, že velmi rozsáhlé výběry dat obsahující množství odlehlých pozorování zkreslují výsledky statistických testů, které tak mohou podávat špatné výsledky. Jedna z možností jak eliminovat vliv odlehlých pozorování je jejich odstranění z výběru dat. V případě analyzovaného výběrového souboru dat chyby predikce srážek nastává problém s určením hranice odlehlých pozorování. V případě, že by byla hranice nastavena příliš nízko, mohlo by dojít ke zkreslení dat. Oproti tomu, kdyby byla hranice příliš vysoká, vliv odlehlých pozorování by opět mohl zkreslovat výsledky statistických testů.

Řešením problému odlehlých pozorování spočívá v provedení rovnoměrně náhodného výběru dat o přiměřené velikosti. Počet odlehlých pozorování se tak vzhledem k původnímu souboru proporčně zmenší, při zachování přibližně stejných hodnot výběrových charakteristik původního souboru.

## 3.3 Analýza rozptylu chyby

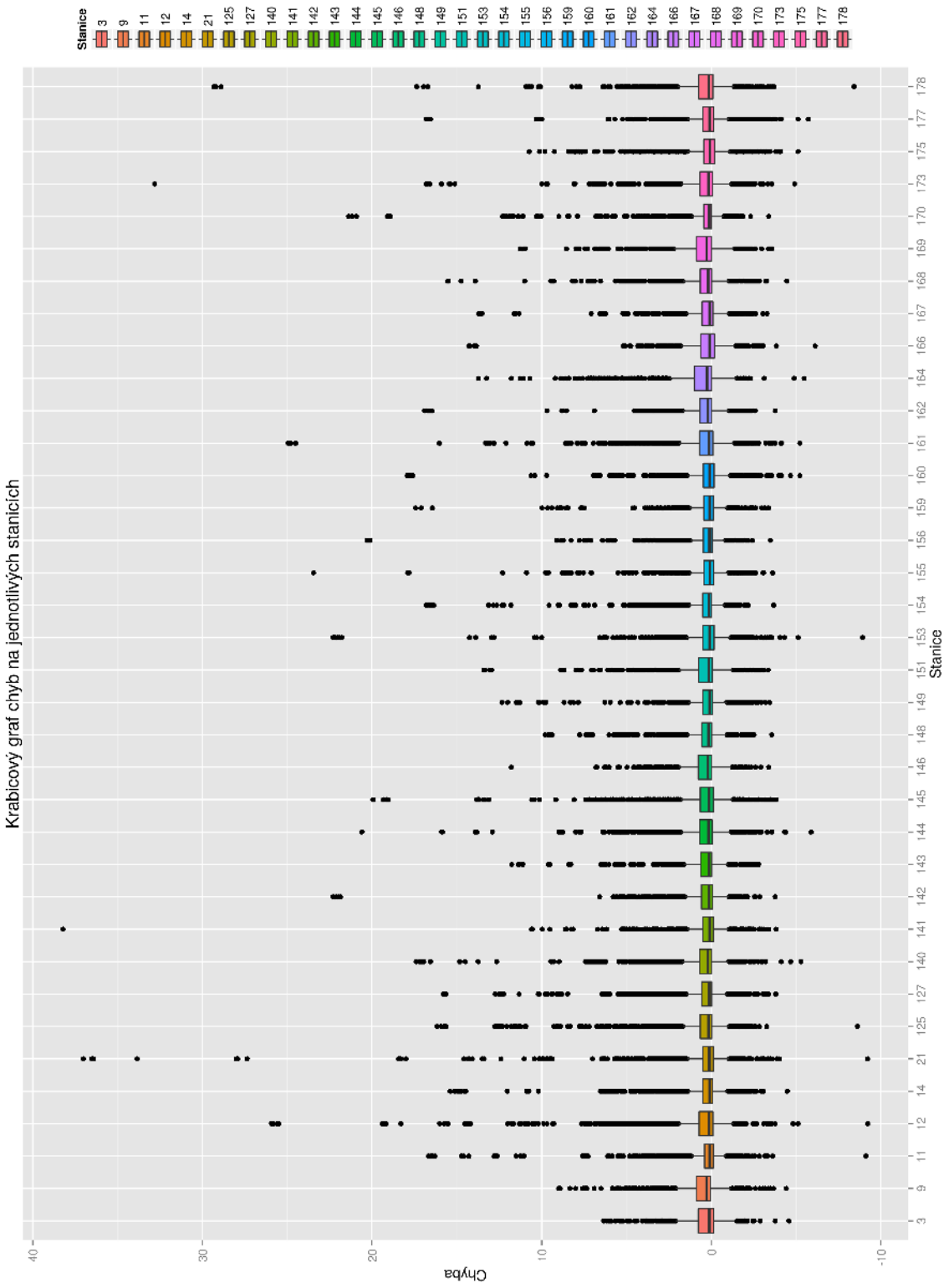
Náplní této podkapitoly je analýza možné závislosti chyby na stanici, době předpovědi a na intenzitě srážek. Pomocí statistických testů je postupně ověřena statistická významnost výše zmíněných závislostí. Statisticky významné závislosti pak budou využity pro modelování chyby při implementaci neurčitostní simulace.

### 3.3.1 Závislost chyby na měřicích stanicích

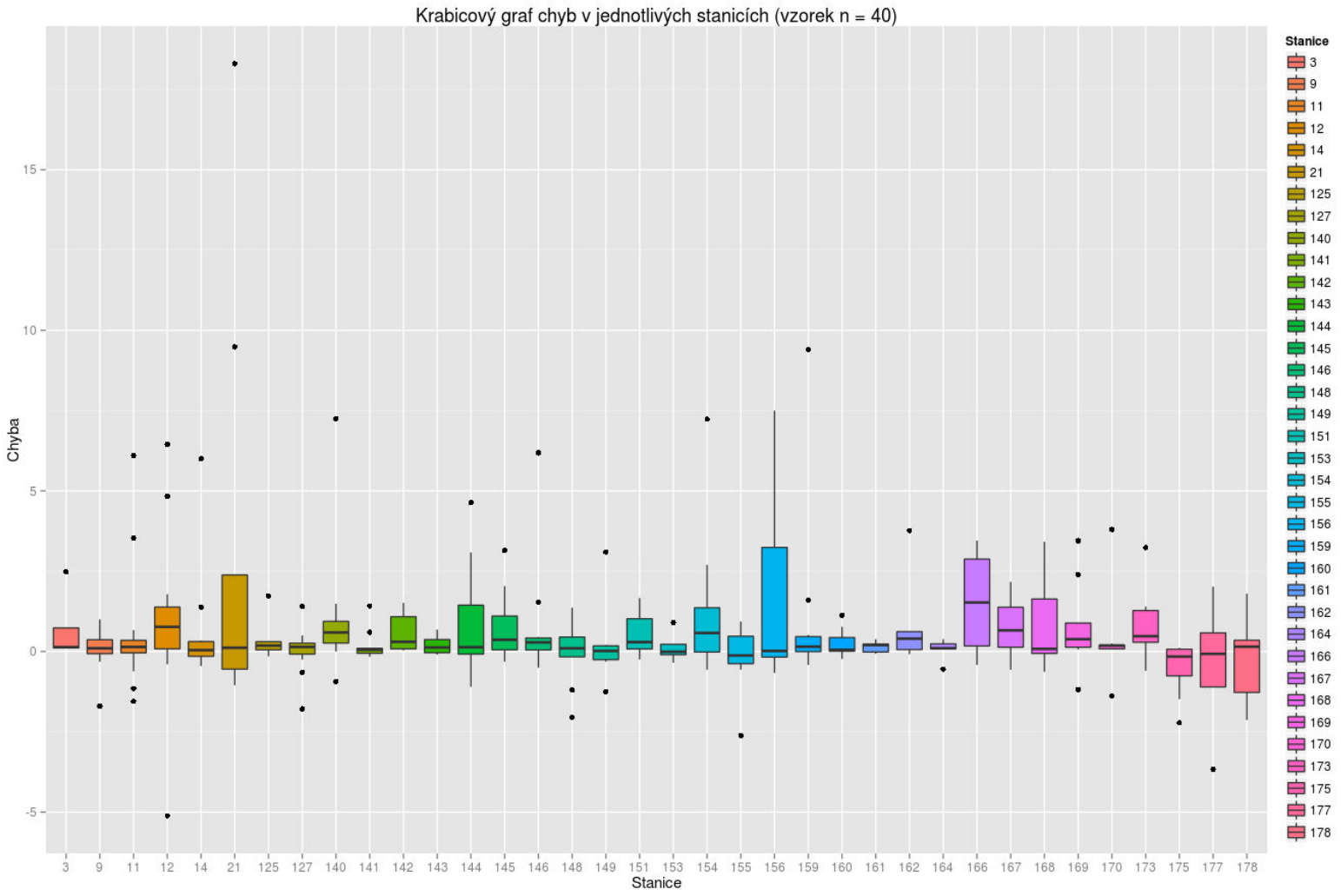
Další část analýzy se zabývala možnou závislostí velikosti absolutní chyby předpovědi na stanici, pro kterou byla předpověď vydána. V případě, že by existovala závislost mezi jednotlivými stanicemi, předpověď modelu Aladin by byla nekonzistentní, přesnost předpovědi by byla ovlivněna umístěním stanice. Vhodným statistickým testem pro analýzu závislosti chyby na měřené stanici je parametrická analýza rozptylu (ANOVA) [12]. Tento test je možno použít pouze v případě, že jsou splněny jeho předpoklady. Parametrická analýza rozptylu vyžaduje, aby data pocházela z normálního rozdělení, jednotlivé výběry byly nezávislé a měly srovnatelné rozptyly. Shoda s normálním rozdělením však nebyla prokázána, je proto nutné pro ověření závislosti chyby na měřící stanici použít neparametrický Kruskal-Wallisův test [12].

> Předmětem testu bude **nulová hypotéza**, která předpokládá, že mediány chyby předpovědi pro jednotlivé stanice se mezi sebou statisticky významně neliší, oproti **alternativní hypotéze**, která předpokládá, že mezi chybovostí alespoň dvou stanic existuje statisticky významný rozdíl.

Na obrázku 5 jsou vykresleny krabicové grafy pro všechny stanice. Lze zde pozorovat velký počet odlehlých pozorování, navzdory podobným proporcím chyby mezi jednotlivými stanicemi. Na základě výsledku Kruskal-Wallisova testu ($p\text{-}hodnota < 0,001$) lze na hladině významnosti $0,05$ zamítnout nulovou hypotézu ve prospěch alternativní. Tento výsledek však neodpovídá zmíněnému krabicovému grafu.




Obrázek 5: Krabicový graf chyb pro všechny stanice

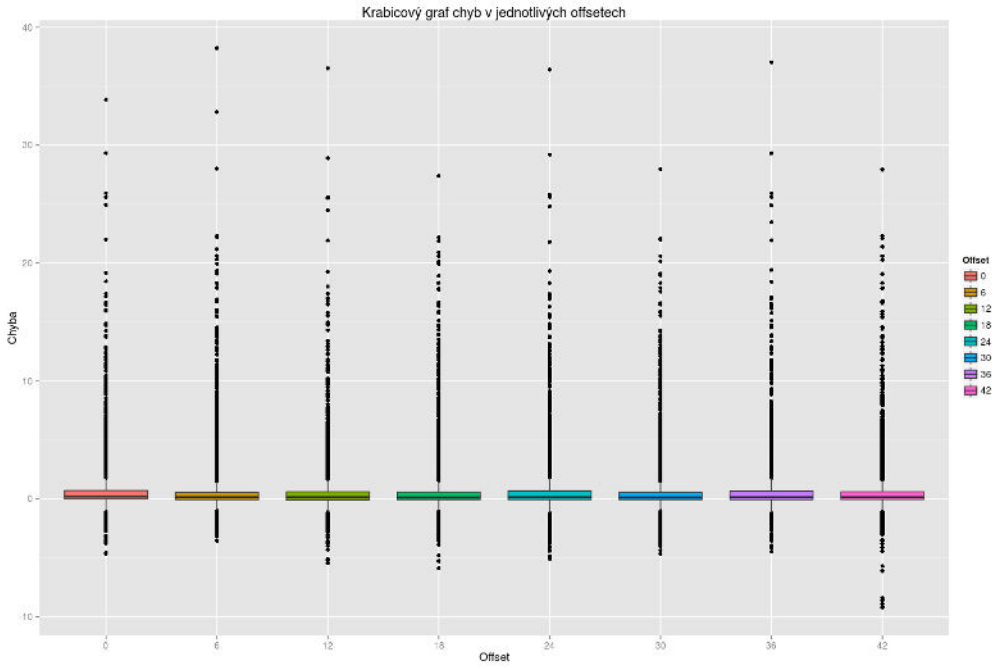



Obrázek 6: Krabicový graf chyb pro všechny stanice, náhodný výběr o velikosti n = 40

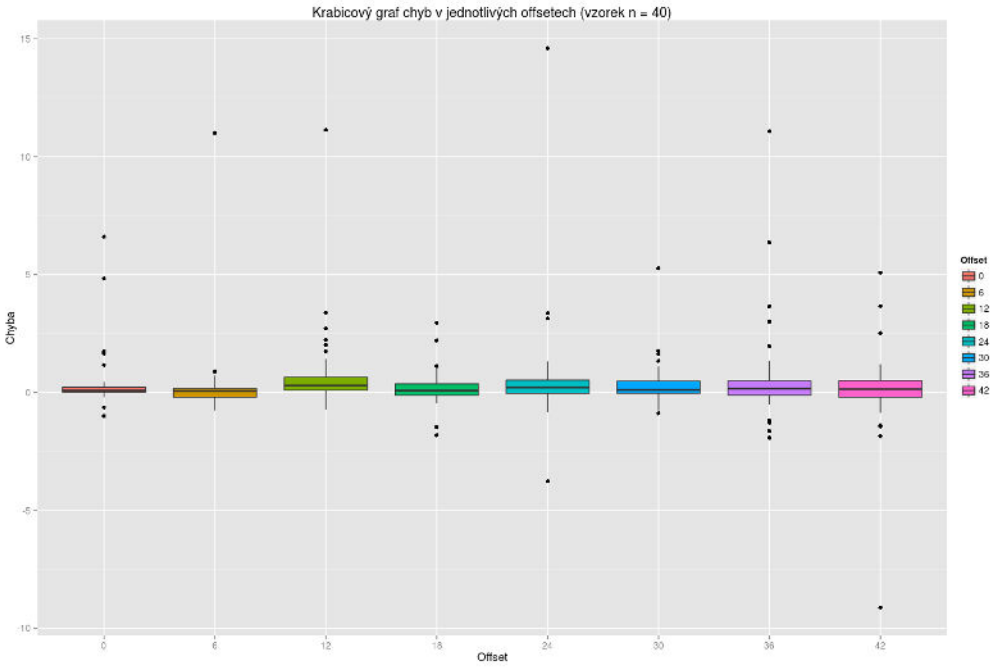
Provedením náhodného výběru o velikosti $n = 40$ měření z původního výběrového souboru lze eliminovat vliv odlehlých pozorování. Na obrázku 6 je vykreslen krabicový graf pro náhodný výběr z původních dat. Oproti grafu pro celý datový soubor zde ubylo odlehlých pozorování, naopak proporce pravděpodobnostního rozdělení zůstaly přibližně stejné. Náhodný výběr o dostatečné velikosti eliminuje vliv odlehlých pozorování a zároveň zachovává hodnoty statistických znaků. Provedením Kruskal-Wallisova testu nad tímto vzorkem lze získat průkaznější výsledky. Na základě výsledku testu (*p-hodnota* $0,560$) nelze na hladině významnosti $0,05$ zamítnout nulovou hypotézu a lze říci, že mezi jednotlivými stanicemi neexistuje statisticky významný rozdíl.

### 3.3.2 Závislost chyby na době předpovědi

Dalším kritériem, které může ovlivňovat velikost chyby předpovědi je délka času, na který byla předpověď vydána (offset). Je logické se domnívat, že přesnost předpovědi se bude zhoršovat s její délkou, avšak v případě krátkodobé předpovědi rozdíl nemusí být natolik významný, aby ovlivňoval její celkovou chybovost. Pozorováním krabicových grafů na obrázcích 7 a 8 lze vidět obdobně jako u analýzy závislosti stanic, že velkou roli zde hrají odlehlá pozorování. Stejným způsobem byl proto proveden náhodný výběr o velikosti $n = 40$ prvků z původního datového souboru a pomocí Kruskal-Wallisova testu ověřena existence statisticky významného rozdílu chybovosti mezi jednotlivými offsety.




Obrázek 7: Krabicový graf chyby předpovědi pro jednotlivé offsety




Obrázek 8: Krabicový graf chyby předpovědi pro jednotlivé offsety, náhodný výběr o velikosti n = 40

> **Nulová hypotéza** v tomto případě předpokládá, že neexistuje statisticky významný rozdíl v chybovosti předpovědi mezi jednotlivými offsety. **Alternativní hypotéza** předpokládá, že statisticky významný rozdíl mezi nimi existuje.

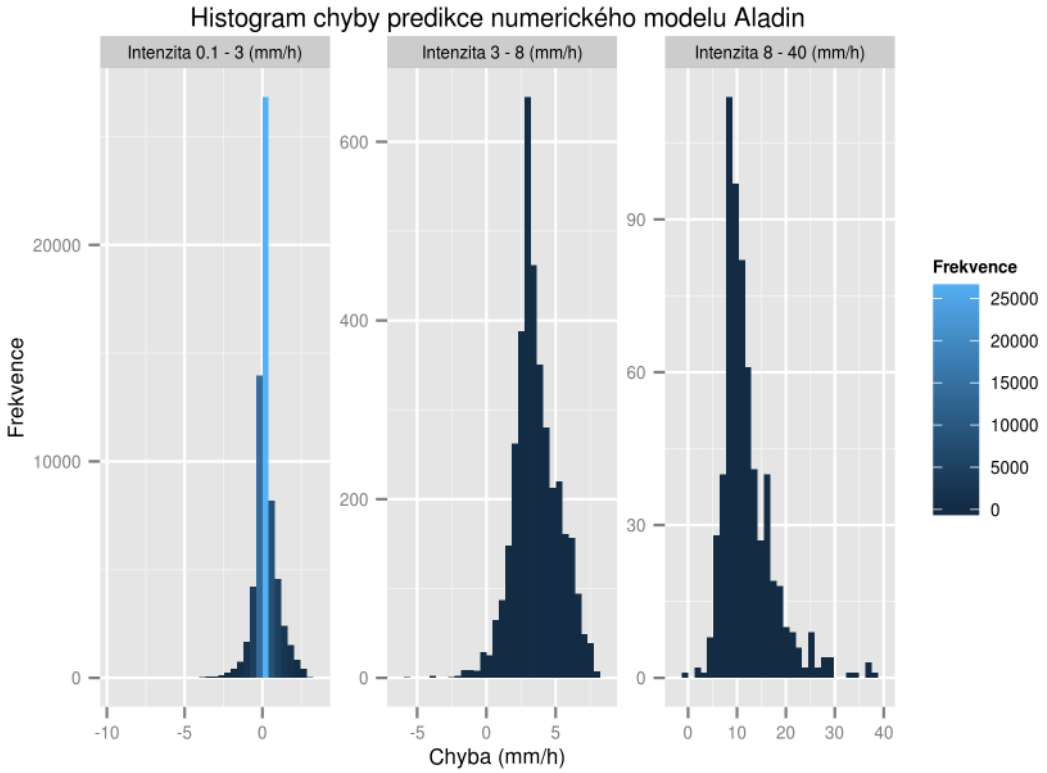
Podle výsledků Kruskal-Wallisova testu (*p-hodnota* $0,465$) pro náhodný výběr ze souboru nelze na hladině významnosti $0,05$ zamítnout nulovou hypotézu, chybovost modelu není závislá na délce předpovědi.

### 3.3.3 Závislost chyby na skutečné intenzitě srážek

Podle výstupu explorační analýzy, lze na základě grafu na obrázku 3 předpokládat jistou závislost chyby předpovědi na intenzitě předpovídaných srážek. Rozdělením intenzity srážek do kategorií a modelováním chyby pro každou kategorii zvlášť tak lze dále zpřesnit modelování neurčitosti předpovědi. Identifikace vhodných intervalů by spočívala v analýze četností výskytu jednotlivých intenzit na území ČR, což by bylo zdlouhavé a nemuselo by významným způsobem přispět ke zpřesnění neurčitostní simulace. ČHMÚ rozděluje srážky do kategorií uvedených v tabulce 4. Uvedené kategorie byly stanoveny experty na meteorologii a jsou využívány ČHMÚ jakožto autoritativní institucí.

| Kategorie | mm/h |
|---|---|
| Velmi slabé | 0 - 0,1 |
| Slabé | 0,1 - 3 |
| Mírné | 3 - 8 |
| Silné | 8 - 40 |
| Velmi silné | nad 40 |

Tabulka 4: Kategorie intenzity dešťových srážek podle ČHMÚ [2]




Obrázek 9: Histogramy chyb pro všechny intervaly intenzity, proporčně normalizované

V tabulce 5 se nacházejí výběrové charakteristiky chyby pro jednotlivé kategorie intenzity srážek. Jsou zde zřejmé značné rozdíly ve velikostech jednotlivých výběrových souborů, což může značit příliš malý počet kategorií intenzity srážek. Průměr chyby se v rámci jednotlivých kategorií značně liší, je zde viditelná závislost chyby na intenzitě srážek. Podhodnocování lze pozorovat na maximálních hodnotách, které se pohybují těsně u horní hranice intervalu. Odstupy horního a dolního kvartilu od mediánu jsou pro každou z kategorií relativně stejné, což spolu s relativně malými směrodatnými odchyl-

kami značí rovnoměrné rozložení chyb okolo jejich průměru, model má tedy tendenci chybovat stejnoměrně, odchylka předpovědi je stabilní.

| **Výběrová charakteristika** | **Intenzita (mm/h)** | | |
|---|---|---|---|
| | **0,1 - 3** | **3 - 8** | **8 - 40** |
| Velikost výběrového souboru | 66454 | 3719 | 631 |
| Průměr (mm/h) | 0,2 | 3,6 | 12,1 |
| Dolní kvartil (mm/h) | -0,1 | 2,6 | 8,7 |
| Medián (mm/h) | 0,1 | 3,4 | 10,8 |
| Horní kvartil (mm/h) | 0,5 | 4,6 | 14,1 |
| Minimum (mm/h) | -9,2 | -5,7 | -0,6 |
| Maximum (mm/h) | 3 | 8 | 38,2 |
| Šikmost | -0,2 | 0,1 | 1,65 |
| Špičatost | 6,2 | 0,7 | 4,0 |
| Směrodatná odchylka (mm/h) | 0,7 | 1,6 | 5,25 |

Tabulka 5: Výběrové charakteristiky chyby pro jednotlivé kategorie intenzity

## 3.4 Odhad parametrů regresního modelu chyby a jeho verifikace

Určení vhodného lineárního regresního modelu je jedním ze způsobů jak popsat stochastickou závislost mezi dvěma veličinami. Regresorem je v tomto případě intenzita srážky a závislou proměnnou chyba předpovědi. Při vhodně zvoleném modelu by se pak dala pro danou intenzitu srážky odhadnout chyba, s jakou může být hodnota předpovězena.

Lineární regrese je nejjednodušším z regresních modelů (definice 3.2), popisujícím chování závislé veličiny přímkou.

**Definice 3.2** *Obecný předpis lineárního regresního modelu je určen jako*

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i$$

,

kde $\beta_1$ a $\beta_2$ jsou koeficienty rovnice přímky a $\epsilon_i$ je náhodná složka chyby, představující vliv neznámých a nevysvětlených regresorů.

Bodovým odhadem hodnot parametrů modelu lze získat odhad regresní funkce (definice 3.3). Běžně používanou metodou pro stanovení bodového odhadu hodnot je metoda nejmenších čtverců. Tato metoda spočívá v minimalizaci součtu druhých mocnin chyb zvoleného modelu [12]. Minimalizační úlohy jsou obecně pro svou náročnost vhodné pro strojové zpracování, proto bude pro určení hodnot koeficientů použit statistický software.

**Definice 3.3** *Odhad regresní funkce*

$$\hat{Y} = b_1 + b_2 x$$

,

| Parametr | Bodový odhad | Rozptyl náhodné složky | t-test (p-hodnota) |
|---|---|---|---|
| $b_1$ | -0,290 | 0,686 | $< 0,001$ |
| $b_2$ | 0,801 | -0,242 | $< 0,001$ |
| Adjustovaný index determinace | | | 0,404 |
| **Výběrová reziduální směrodatná odchylka** | | | 0,594 (8325 st. volnosti) |
| **Celkový F-test (p-hodnota)** | | | $< 0,001$ (8325 st. volnosti) |

Tabulka 6: Výsledky regresní analýzy pro interval **0,1 - 3 mm/h**

| Parametr | Bodový odhad | Rozptyl náhodné složky | t-test (p-hodnota) |
|---|---|---|---|
| $b_1$ | -0,242 | 0,242 | 0,317 |
| $b_2$ | 0,838 | -0,051 | $< 0,001$ |
| Adjustovaný index determinace | | | 0,385 |
| **Výběrová reziduální směrodatná odchylka** | | | 1,422 (428 st. volnosti) |
| **Celkový F-test (p-hodnota)** | | | $< 0,001$ (428 st. volnosti) |

Tabulka 7: Výsledky regresní analýzy pro interval **3 - 8 mm/h**

### 3.4.1 Bodový odhad parametrů modelu

Hodnoty parametrů přímkového modelu byly odhadnuty pomocí funkce `lm` nacházející se v balíku `stats`, který je součástí statistického software R [15]. Funkce využívá pro nalezení hodnoty parametrů metody nejmenších čtverců [15]. Na obrázku 10 jsou vykresleny grafy obsahující původní data proložená přímkou s odhadnutými parametry. Mezery mezi hodnotami na vodorovných osách jsou určeny rozlišením měřených srážek 0,1 mm/h [4].

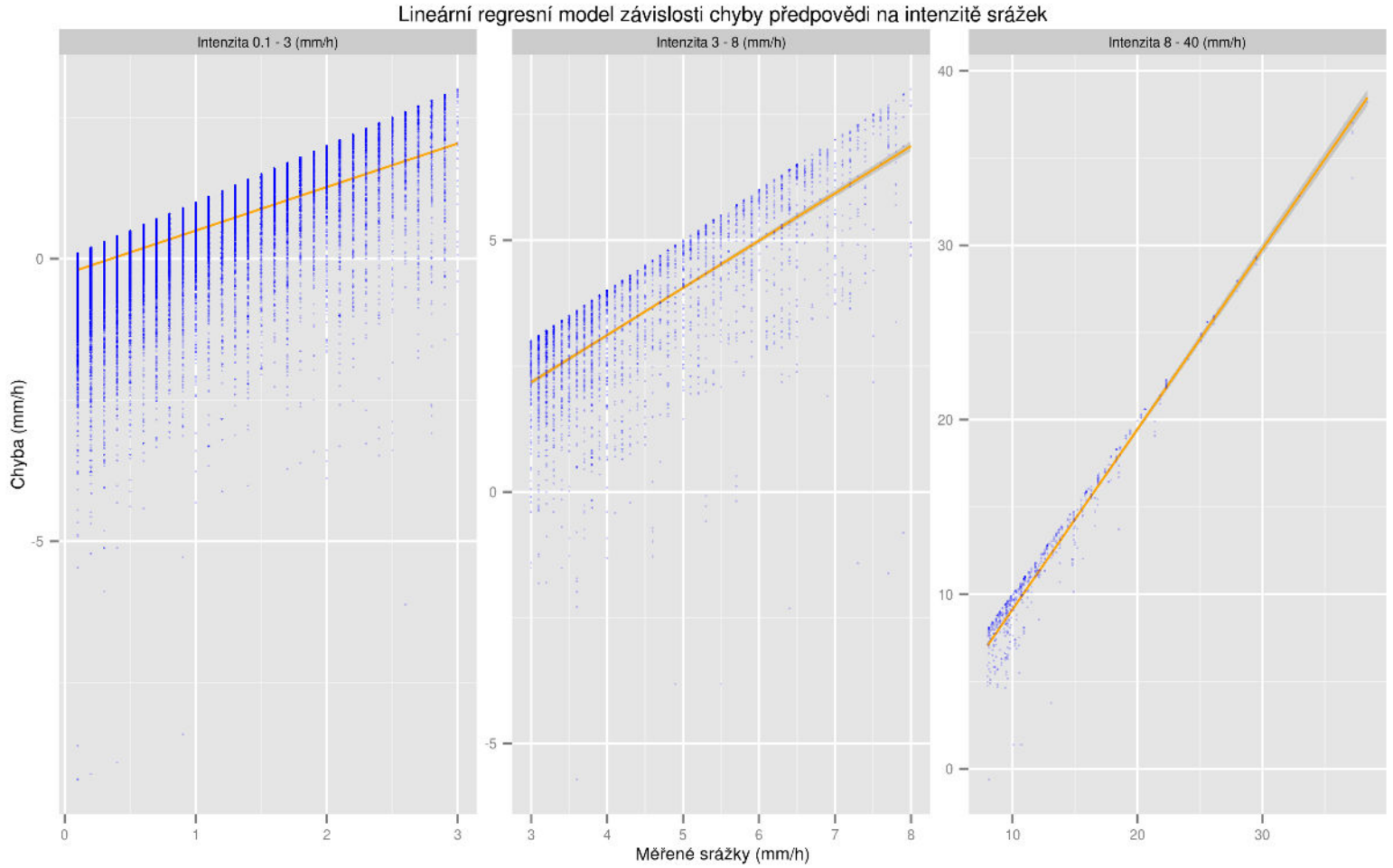
### 3.4.2 Verifikace modelu

Po provedení odhadu parametrů modelu je nutné přistoupit k jeho verifikaci. Pokud model vyhoví, je možno prohlásit, že závislost chyby na intenzitě srážek lze modelovat přímkovým modelem a že mezi zkoumanými veličinami existuje lineární závislost.

V tabulkách 6, 7 a 8 jsou uvedeny výsledky regresní analýzy vygenerované statistickým software R. Celkový F-test u všech třech získaných modelů nevyvrátil vhodnost zvolené regresní funkce, která je tedy vhodná pro modelování dané závislosti. Podle dílčích t-

| Parametr | Bodový odhad | Rozptyl náhodné složky | t-test (p-hodnota) |
|---|---|---|---|
| $b_1$ | -2,049 | 0,686 | 0,004 |
| $b_2$ | 1,090 | 0,052 | $< 0,001$ |
| Adjustovaný index determinace | | | 0,839 |
| **Výběrová reziduální směrodatná odchylka** | | | 1,967 (83 st. volnosti) |
| **Celkový F-test (p-hodnota)** | | | $< 0,001$ (83 st. volnosti) |

Tabulka 8: Výsledky regresní analýzy pro interval **8 - 40 mm/h**




Obrázek 10: Chyba závislá na intenzitě srážek predikovaná přímkovou regresí pro různé kategorie srážek s konfidenčními intervaly

| **Intenzita (mm/h)** | **p-hodnota** | | |
|---|---|---|---|
| | Kolmogorov-Smirnovův test | t-test | Durbin-Watsonův test |
| 0,1 - 3 | $< 0,001$ | $> 0,999$ | $< 0,001$ |
| 3 - 8 | $< 0,001$ | $> 0,999$ | $0,738$ |
| 8 - 40 | $2,571 \cdot 10^{-7}$ | $> 0,999$ | $0,462$ |

Tabulka 9: Výsledek statistických testů vlastností reziduí

testů je možno potvrdit statistickou významnost všech odhadovaných parametrů, kromě parametru $b_1$ v intervalu 3 - 8 mm/h, který lze na základě výsledku t-testu zanedbat, což by znamenalo změnu modelu, který by neobsahoval parametr $b_1$.

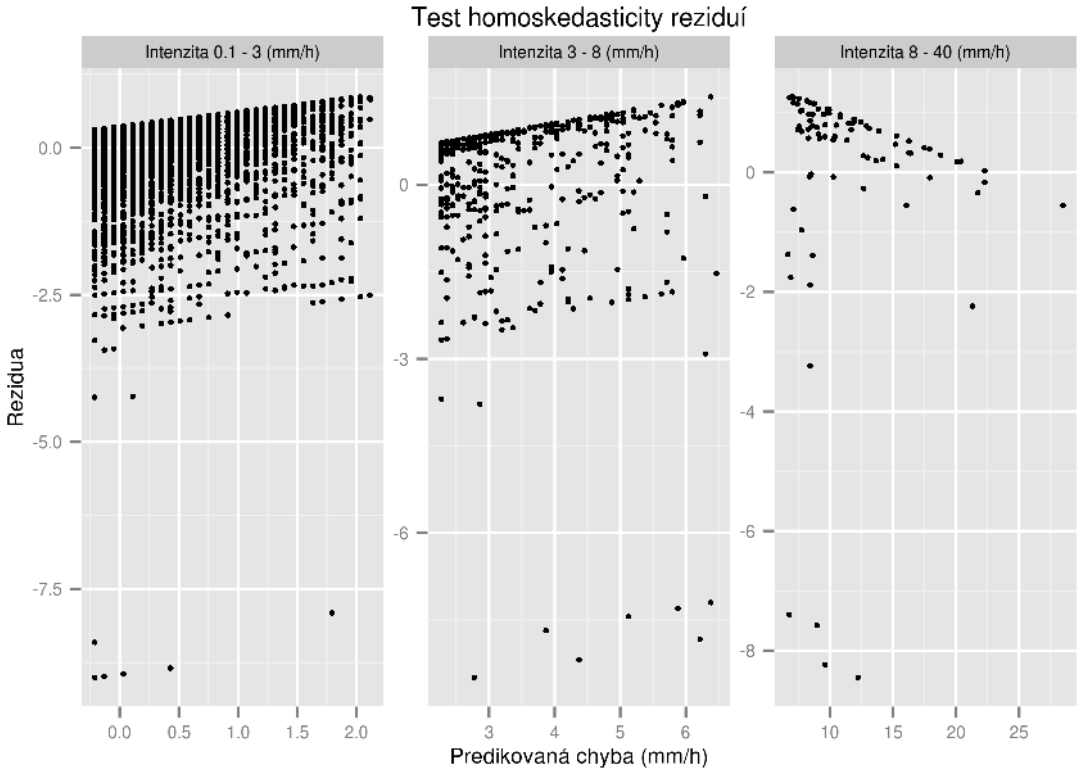
Nízké hodnoty adjustovaného indexu determinace ($R^2_{adj}$) v prvních dvou případech značí, že velikost srážek vysvětluje chybu měření pouze zhruba ze 40 % pro všechny kategorie intenzity. Zbylých 60 % hodnot je důsledkem vlivu neznámých regresorů, které nejsou v modelu zahrnuty.

Dalším krokem je ověření vlastností reziduí. Aby byla potvrzena vhodnost modelu, musejí pocházet z normálního rozdělení, musí mít konstantní rozptyl, nulovou střední hodnotu a nesmí být autokorelována.

Pro ověření normality reziduí lze použít Kolomogorov-Smirnovův test dobré shody, pro ověření nulovosti střední hodnoty standardní t-test. Konstatní rozptyl se dá přibližně ověřit konstrukcí bodového grafu závislosti hodnoty reziduí na hodnotách generovaných ověřovaným modelem. Autokorelaci reziduí lze testovat pomocí Durbinova-Watsonova testu [12].

Dle výsledků v tabulce 9 lze říci, že rezidua nevyhovují ani požadavku na nulovost střední hodnoty ani požadavku na normalitu. Stejně tak v grafu na obrázku 11 jde vypozorovat zjevnou závislost reziduí na predikovaných hodnotách, rezidua pravděpodobně nemají stejné rozptyly. Model pro první kategorii srážek má autokorelovaná rezidua, což značí pro tuto kategorii nevhodně zvolený model. Autokorelace reziduí nebyla prokázána pro dva zbývající modely.

Výsledky analýzy reziduí, které poukazují na nesplnění požadavků pro správné použití lineární regrese spolu s výsledkem ověření stability modelu značí nevěrohodnost modelu, závislost chyby na předpovědi tedy není možno modelovat přímkovým regresním modelem.




Obrázek 11: Graf závislosti reziduí na predikovaných hodnotách pro všechny intervaly intenzity

## 3.5 Jádrový odhad hustoty pravděpodobnosti chyby

Pokud není nutné znát předpis pravděpodobnostní funkce náhodné veličiny, lze provést její neparametrický odhad. Jedna z možností provedení neparametrického odhadu je konstrukce histogramu. Zkoumané hodnoty se zařadí do přihrádek o šířce $h$. Výška každé přihrádky je funkcí počtu hodnot v přihrádce. Odhad pravděpodobnosti pro konkrétní hodnotu se dá provést prostřednictvím vztahu v definici 3.4. Příklad výsledného histogramu se nachází na obrázku 9.

**Definice 3.4** *Vztah pro odhad hustoty pravděpodobnosti v bodě $x$ pomocí histogramu:*

$$\hat{p}(x) = \frac{n_p}{h \cdot n}$$

*kde:*

- $n$ *je celkový počet bodů*

- $n_p$ *je počet bodů v jedné přihrádce*
- $h$ *je šířka přihrádky*

Nevýhodou odhadu hustoty pravděpodobnosti pomocí histogramu je jeho nespojitost a relativní nepřesnost odhadu plynoucí ze špatně určené šířky přihrádky $h$. Pro zvýšení přesnosti odhadu je možno hodnotu $h$ snížit, potom však může dojít k situaci, kdy není dostatek dat pro naplnění všech přihrádek. Konstrukce histogramu je tedy vhodná pouze pro získání obecného náhledu na možný tvar hustoty pravděpodobnosti.[18]

Mezi další neparametrické metody odhadu patří jádrové odhady, prostřednictvím kterých lze vzorkováním přibližně určit tvar hustoty pravděpodobnosti. Při dostatečně velkém počtu vzorků pak lze použít výsledek pro generování hodnot se zkoumaným pravděpodobnostním rozdělením. V tomto případě bude proveden jádrový odhad hustoty pravděpodobnosti chyby numerického modelu Aladin z dostupných vzorků dat.

Metoda využívá vyhlazovací funkci, tzv. jádro, jejíž vlastnosti jsou totožné s vlastnostmi hustoty pravděpodobnosti. Princip této metody spočívá v odhadu hustoty pravděpodobnosti pro daný počet hodnot rovnoměrně rozmístěných v intervalu určeném hodnotami realizací zkoumané veličiny. Odhad hustoty pravděpodobnosti v daném bodě je proveden pomocí vztahu uvedeného v definici 3.5. [17], [18]. Jednotlivé body jsou proloženy jádrovými funkcemi, odhad hustoty v daném bodě se pak rovná součtu hodnot jádrových funkcí, který daný bod překrývají.

**Definice 3.5** *Vztah pro odhad hustoty pravděpodobnosti v bodě $x$ pomocí jádrového odhadu:*

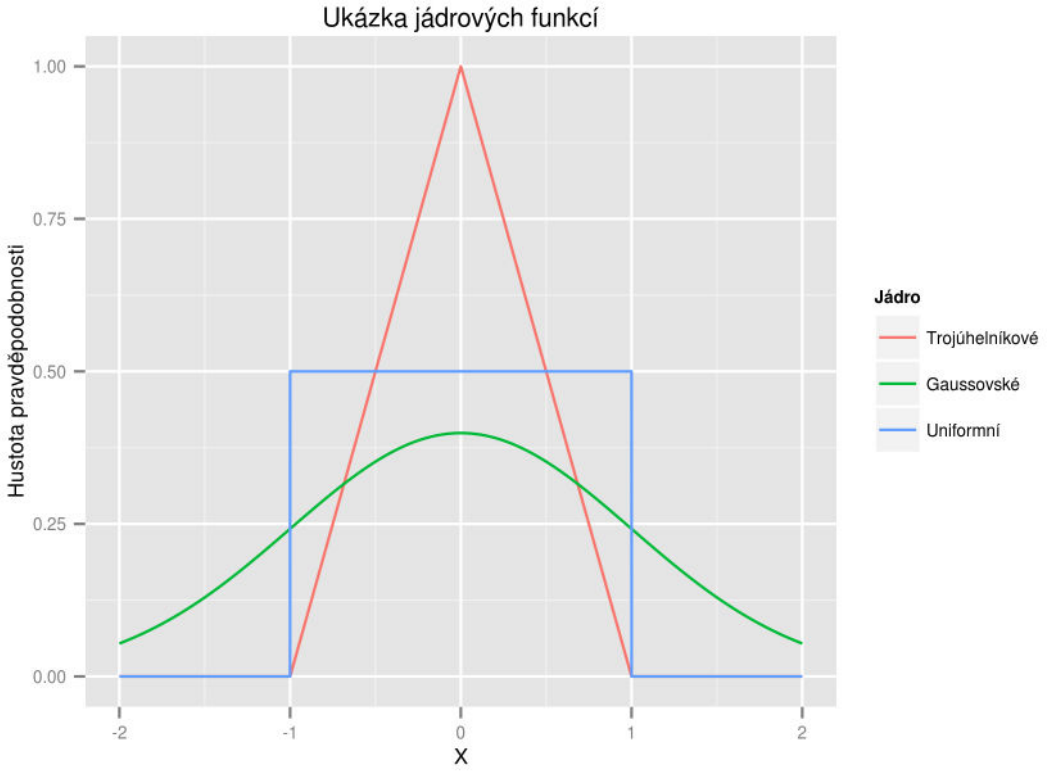
$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} w(x - x_i, h)$$

*kde:*

- $w$ *je zvolená jádrová funkce*
- $h$ *je šířka vyhlazovacího okna*
- $n$ *počet vzorků*

Vyhlazovací funkce $w$ se nazývá jádrem. Volba vhodného typu funkce závisí na tvaru histogramu zkoumaného náhodného výběru. Na obrázku 12 se nachází grafy vybraných jádrových funkcí. Jejich parametrem je kromě pozorované hodnoty také vyhlazovací parametr $h$, nazývaný také šířka pásma [2], který má významný vliv na kvalitu odhadu. Špatně zvolená šířka pásma má za následek buďto přehlazený odhad, který nemusí vystihovat všechny malé odchylky zkoumané hustoty pravděpodobnosti nebo naopak příliš malá šířka pásma může vnášet do výsledku nežádoucí výchylky a rušení [17].

[2] v angl. literatuře *bandwidth*




Obrázek 12: Grafy vybraných jádrových funkcí

### 3.5.1 Určení šířky pásma

Metod pro určování hodnoty šířky pásma existuje velké množství, jejich rozbor však není náplní této práce. Budou zde proto zmíněny pouze ty metody, které jsou poskytované statistickým software R a které byly použity při analýze.

V prostředí software R jsou jádrové odhady poskytovány prostřednictvím funkce `density`. Pro stanovení vhodné hodnoty šířky pásma lze použít *Silvermannovo pravidlo*, které je vhodné pro odhady využívající gaussovské jádro, dále je zde dostupná metoda *vychýleného* a *nevychýleného křížového ověřování* a *Sheather-Jonesova* metoda [15],[17],[18].

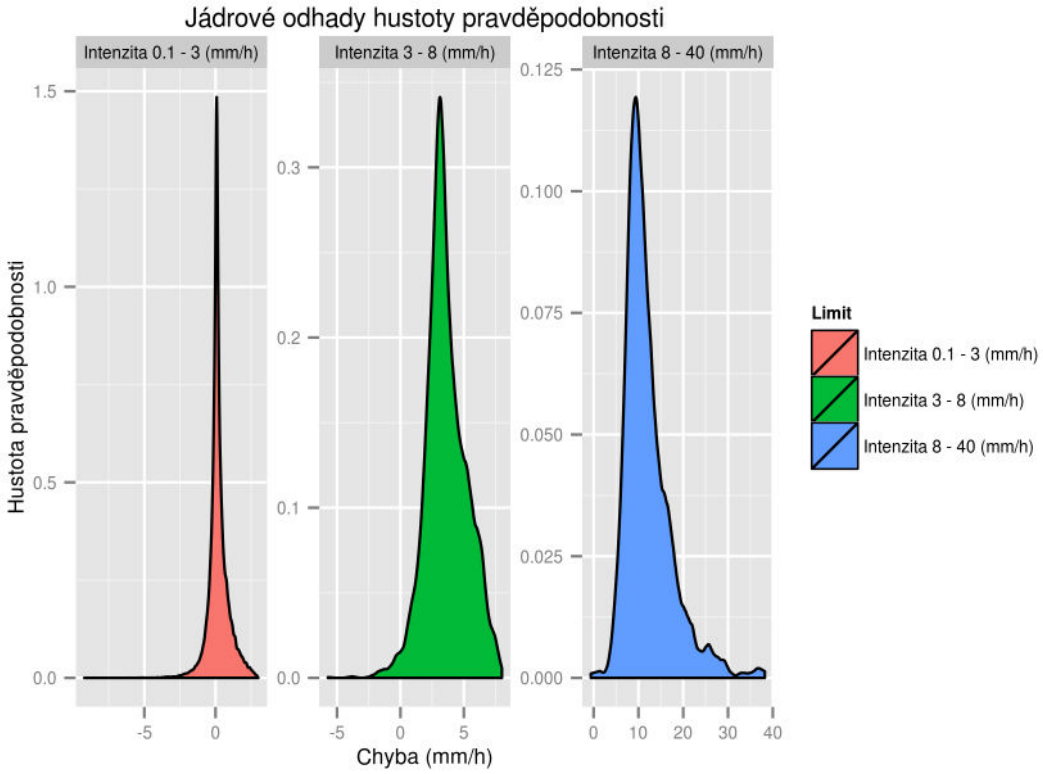
Při analýze byla použita metoda *nevychýleného křížového ověřování*, která spočívá v minimalizaci ukazatele MISE[3], který svou hodnotou vyjadřuje kvalitu odhadu [17],[18].

[3] střední integrální kvadratická chyba, angl. Mean integrated squared error

### 3.5.2 Odhad hustoty pravděpodobnosti chyby

Odhad hustoty pravděpodobnosti byl proveden pomocí funkce `density`, která provádí interpolaci získaného odhadu pro zadaný počet bodů, které jsou rovnoměrně rozmístěné v intervalu ohraničeném minimální a maximální hodnotou zkoumaných dat. Výsledkem je vektor souřadnic, jejichž vynesením do grafu pak lze získat odhadovanou křivku hustoty pravděpodobnosti.

Vzhledem ke tvaru histogramů pro jednotlivé intervaly bylo zvoleno trojúhelníkové jádro. Porovnáním výsledků pro různé metody určení hodnoty vyhlazovacího parametru byla zvolena metoda nevychýleného křížového ověřování. Odhady provedené pomocí vygenerované hodnoty vyhlazovacího parametru nejvíce odpovídaly tvaru histogramu zkoumaných dat. Odhadnuté křivky hustoty pravděpodobnosti se nachází na obrázku 13 byly navzorkovány pro 70 000 hodnot, což zhruba odpovídá velikosti výběrového souboru.



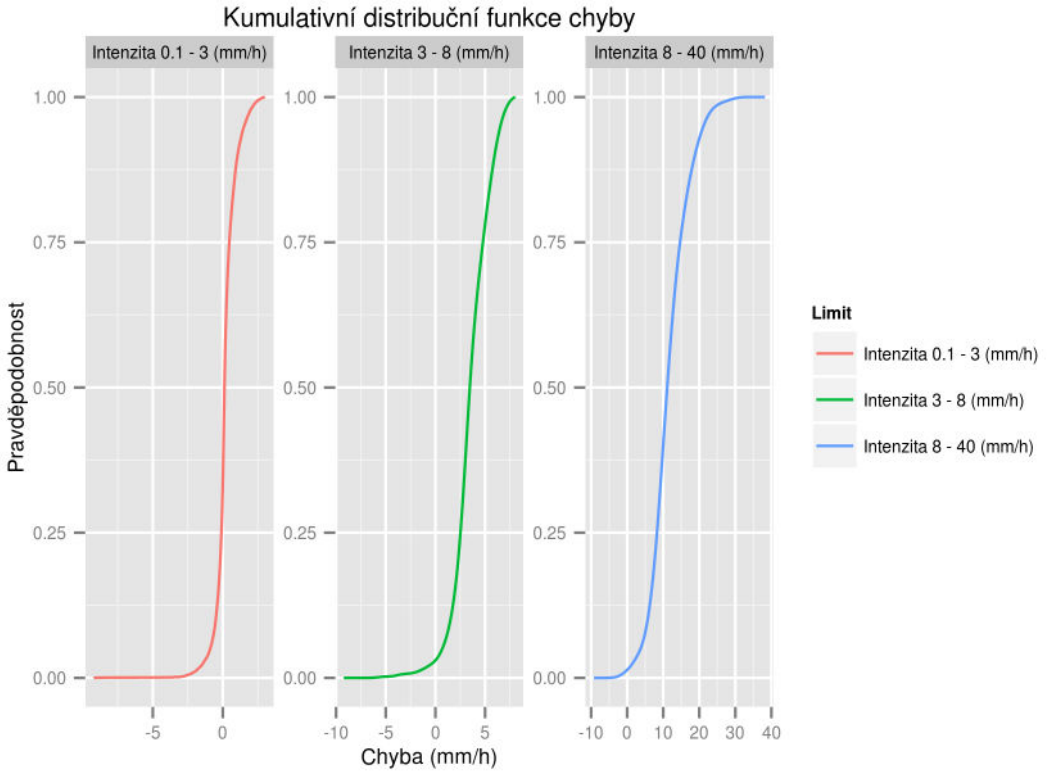

Obrázek 13: Jádrový odhad hustoty pravděpodobnosti pro jednotlivé limity

Ze získaných vzorků hustoty pravděpodobnosti chyby lze pomocí vztahu v definici 3.6 získat křivku její distribuční funkce, která je následně využita pro generování hodnot

chyby pocházejících z odhadnutého pravděpodobnostního rozdělení. Výsledné křivky distribuční funkce se nacházejí na obrázku 14.

**Definice 3.6** *Distribuční funkce vyjádřená pomocí pravděpodobnostní funkce*

$$F(x) = \sum_{x_i<x} P(x_i)$$




Obrázek 14: Grafy vygenerovaných kumulativních distribučních funkcí chyby

## 3.6 Srovnání modelů

Výsledky verifikace lineárního regresního modelu značí, že závislost chyby na intenzitě srážek jím nelze věrohodně modelovat. Oproti tomu jádrové odhady hustoty pravděpodobnosti chyby poskytují výsledky vhodné pro stochastické modelování metodou Monte Carlo, kterou se zabývá následující kapitola.

# 4 Modelování neurčitosti

Tato kapitola se zabývá stochastickým modelováním neurčitostí pomocí metody Monte Carlo a její implementací do modelu Math1D s využitím výsledků předcházející statistické analýzy.

## 4.1 Metoda Monte Carlo

Historie metody Monte Carlo ve své současné podobě sahá zpět až do 40. let minulého století, kdy vědci Stanislaw Ulam a John von Neumann pracovali v rámci projektu Manhattan na vývoji atomové bomby. První formulace metody je však známá už od 16. století, kdy tuto metodu použil francouzský matematik Buffon pro řešení úlohy tzv. Buffonovy jehly [13].

Obecně se tato metoda používá pro stochastické modelování náhodné veličiny nebo náhodného děje, který není možné modelovat deterministicky. Pomocí této metody se obecně určuje střední hodnota náhodné veličiny, v případě modelu Math1D se tedy pomocí metody Monte Carlo bude určovat střední hodnota chyby předpovědi výsledného odtoku. Implementace metody spočívá především v určení způsobu modelování náhodné veličiny tak, aby výsledný model pokud možno co nejpřesněji simuloval pravděpodobnostní charakteristiky zkoumané náhodné veličiny. Pomocí pseudonáhodného generátoru čísel se poté provádí vzorkování z daného pravděpodobnostního rozdělení. Je zřejmé, že čím větší počet vzorků je vygenerován, tím víc se jeji pravděpodobnostní rozdělení blíží modelovanému. Vzhledem k vysokému počtu realizací, které je nutno generovat je metoda vhodná pro zpracování počítačem, je však nutno brát v potaz omezení plynoucí z determinismu generátorů pseudonáhodných čísel, jež jsou popsány níže. Výsledky takové simulace jsou pak podrobeny statistické analýze (výběr kvantilů, výpočet číselných charakteristik).

V případě neurčitostní simulace modelu Math1D je zkoumanou náhodnou veličinou chybovost předpovědi srážek, které slouží jako vstup pro simulaci jejich odtoku z povodí. Jak již bylo řečeno v podkapitole číslo 2.5, chybovost nelze dobře modelovat deterministicky, lze však statistickou analýzou přibližně určit její pravděpodobnostní rozdělení. Vzorkováním chyb ze zjištěného rozdělení lze generovat sady srážek, zatížené chybou odpovídající pravděpodobnostnímu rozdělení chyby předpovědi. Pokud je pro každou sadu srážek provedena simulace srážko-odtokového modelu, lze ze všech získaných výsledků statistickými metodami zjistit, s jakou pravděpodobností se bude chyba simulace modelu pohybovat v konkrétních mezích, tedy jakou neurčitostí je simulace zatížena.

Jednou z metod, jak analyzovat výslednou neurčitost je výběr kvantilů z hodnot pro všechny provedené iterace pro daný časový úsek a daný kanál. $100p\%$ kvantil je taková hodnota, kde právě $100p\%$ hodnot datového souboru je menších než hodnota tohoto kvantilu [12]. Vybráním stejnoměrně rozmístěných kvantilů lze spolehlivě zjistit variabilitu zkoumaného výběru, porovnáváním jejich odstupů.

### 4.1.1 Pseudonáhodné generátory čísel

Získání absolutně náhodného čísla ve smyslu takovém, že nelze deterministicky určit následující náhodné číslo, není pouze pomocí počítače možné. Pro získání takového náhodného čísla je většinou nutné přídavné hardwarové zařízení, které generuje náhodné čísla například podle generátoru šumu nebo podle Geigerova počítače detekujícího rozpad radioaktivního materiálu. Takové zařízení mohou být velmi nákladná a nejsou běžnou součástí běžných počítačů [13].

Pro potřeby simulace Monte Carlo však postačuje tzv. pseudonáhodný generátor čísel. Takový druh náhodného generátoru generuje pomocí matematických transformací sekvenci čísel v závislosti na svém počátečním stavu (tzv. *seed*).

Důležitá vlastnost pseudonáhodného generátoru čísel je tzv. *periodicita*, jejíž hodnota udává počet vzorků, které je možné vygenerovat, aniž by se generovaná sekvence začala opakovat. Periodicita je však pouze jedním z důležitých kvalitativních požadavků na generátory náhodných čísel. Dalším významným požadavkem je požadavek na nezávislost a uniformitu generovaných vzorků, kdy generované vzorky jsou uniformně rozdělené v daném intervalu a případné vzorkování jiných typů pravděpodobnostních rozdělení závisí na konkrétní implementaci.

Dalšími požadavky na generátory pseudonáhodných čísel jsou především opakovatelnost, na stejný počáteční seed generátor vygeneruje pokaždé stejnou sekvenci čísel, dostatečná rychlost, možnost souběžného využití ve více vláknech a v neposlední řadě také ověřitelnost používaných matematických transformací [14].

## 4.2 Implementace v Math1D

Neurčitostní simulace byly implementovány jako rozšíření stávajícího kódu modelu vyvinutého v C++ [8]. Byly zde využívány prostředky standardu C++11, speciálně kolekce typu `vector`, sdílené ukazatele a další. Byla zde také zohledněna potřeba spouštění výsledného kódu v prostředí výpočetních clusterů použitím rozšíření OpenMP pro usnadnění správy vláken. Rozložení výpočtu mezi více procesů běžících na různých výpočetních uzlech bylo realizováno pomocí knihovny MPI.

Pro generování náhodných vzorků byl použit generátor pseudonáhodných čísel typu *Mersenne twister (M19937)* jehož výhodou je velmi dlouhá perioda $(2^{19937} - 1)$ a dobré statistické vlastnosti [14].

### 4.2.1 Simulace neurčitostí

Neurčitostní simulace implementuje třída `Uncertainity`. Při inicializaci třídy se zvolí dle konfigurace parametry, jejichž neurčitost bude simulována, dále je třídě předán ukazatel na inicializovanou třídu `MatData` jakožto modelu, jehož parametry se budou v rámci simulace měnit.

Parametry modelu, které je možno v rámci simulace mění reprezentují následující třídy:

- `PrecipitationUncertainity` - předpovězené srážky

- `CnUncertainity` - číslo CN křivky (subpovodí)
- `ManningUncertainity` - Manningův koeficient drsnosti (kanál, subpovodí)

Tyto třídy implementují změnu konkrétního parametru modelu pro jednotlivé iterace a jsou potomky abstraktní třídy `AbstractParam`, která obsahuje metody a atributy společné všem třídám reprezentujícím měněný parametr.

Při inicializaci třídy `Uncertainity` je inicializován generátor náhodných čísel a každému parametru je dle konfigurace přiřazena instance třídy implementující vlastní generování hodnot chyby předpovědi podle daného pravděpodobnostního rozdělení. Generování hodnot zajišťují třídy `KernelDensity`, `Regression a UniformRandom`, které jsou rovněž potomky abstraktní třídy `AbstractRandom`, která poskytuje pro každý způsob generování generátor náhodných čísel a ostatní metody a atributy společné těmto třídám. Díky této architektuře lze jednoduše přidávat další parametry a způsoby generování jejich náhodných hodnot. Vytvořením třídy, která je potomkem abstraktní třídy `AbstractRandom` lze jednoduše implementovat další pravděpodobnostní rozdělení pro generování náhodných hodnot. Třídní diagram zobrazující tyto třídy, jejich vybrané atributy a metody se nachází na obrázku 15.

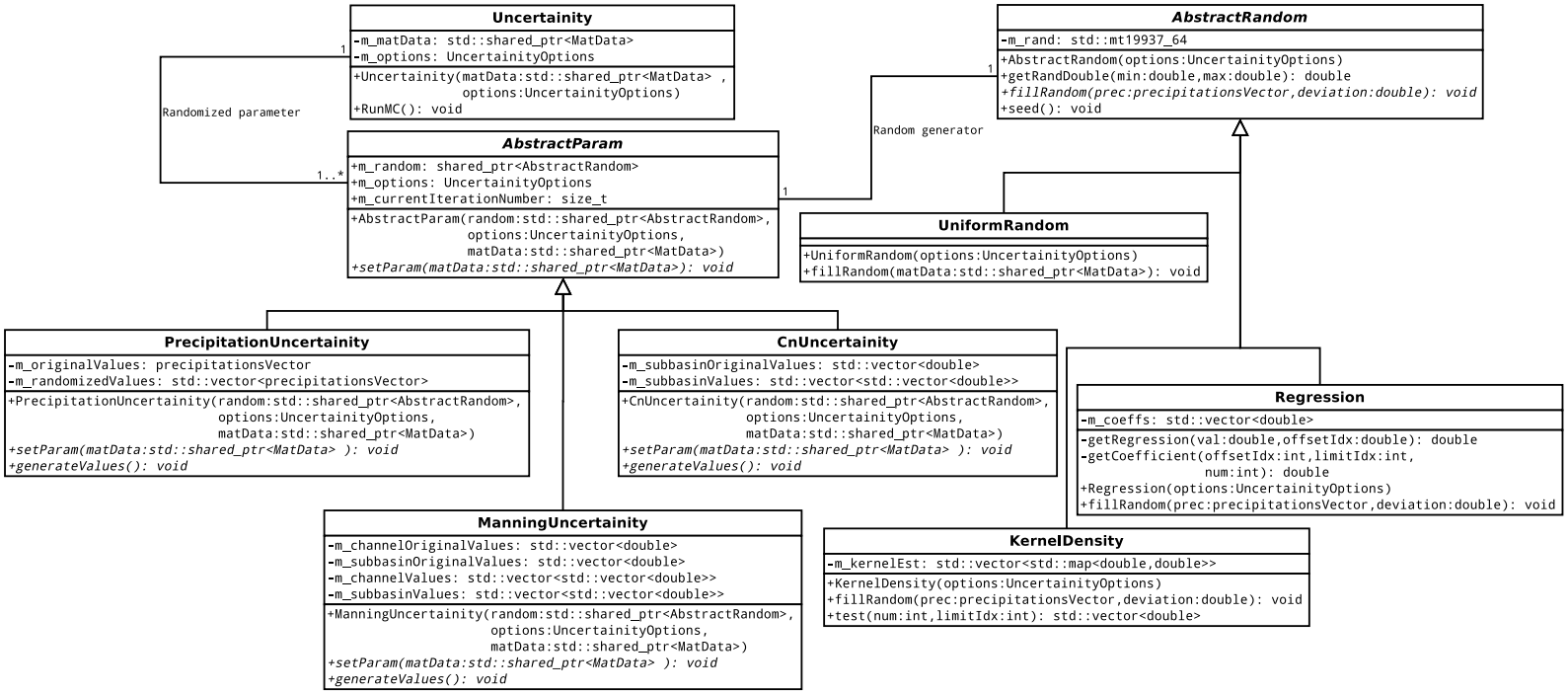
### 4.2.2 Konfigurace

V rámci simulace neurčitostí lze měnit srážky předpovězené numerickým modelem Aladin, čísla CN křivek pro subpovodí a hodnoty Manningova koeficientu pro subpovodí a pro kanály. Chybovost srážek lze modelovat pomocí regresního modelu, nebo pomocí distribuční funkce získané jádrovými odhady. Je zde také možné určit jednotlivé kategorie srážek, kterým mohou odpovídat odlišné způsoby generování hodnot.

Vzhledem k tomu, že není známo pravděpodobnostní rozdělení chyby hodnot CN křivky ani Manningova koeficientu, lze provádět pouze experimenty, kdy se hodnoty generují z určitého známého rozdělení, jehož parametrizace však nemusí odpovídat reálným fyzikálním vlastnostem, které tyto dva parametry popisují. Z toho důvodu je pro tyto dva parametry nutné v konfiguraci kromě volby samotného pravděpodobnostního rozdělení a jeho parametrů také nastavit interval, ve kterém se hodnoty mohou pohybovat, přičemž pokud generovaná hodnota interval přesáhne, je automaticky nahrazena jeho spodní resp. horní hranicí. Interval, ve kterém jsou hodnoty fyzikálně korektní je určen hydrologem.

Dále se v konfiguraci nastavuje počet dní, ve kterých byly modelu předány měřené srážky a počet následujících dní, kdy jsou modelu předány srážky předpovídané. V simulaci je simulována pouze neurčitost předpovídaných srážek, pravděpodobnostní rozdělení chyby měřených srážek nebylo určeno. V konfiguraci je dále nutné určit kvantily, které se budou vybírat z výsledků všech iterací a cestu k adresáři, do kterého jsou výsledky uloženy.

Nejdůležitějším parametrem je počet iterací simulace, který je určen hodnotou uvedenou za argumentem `-mc` předaným aplikaci z příkazové řádky. Výpis 1 obsahuje příklad konfigurace simulace neurčitosti.



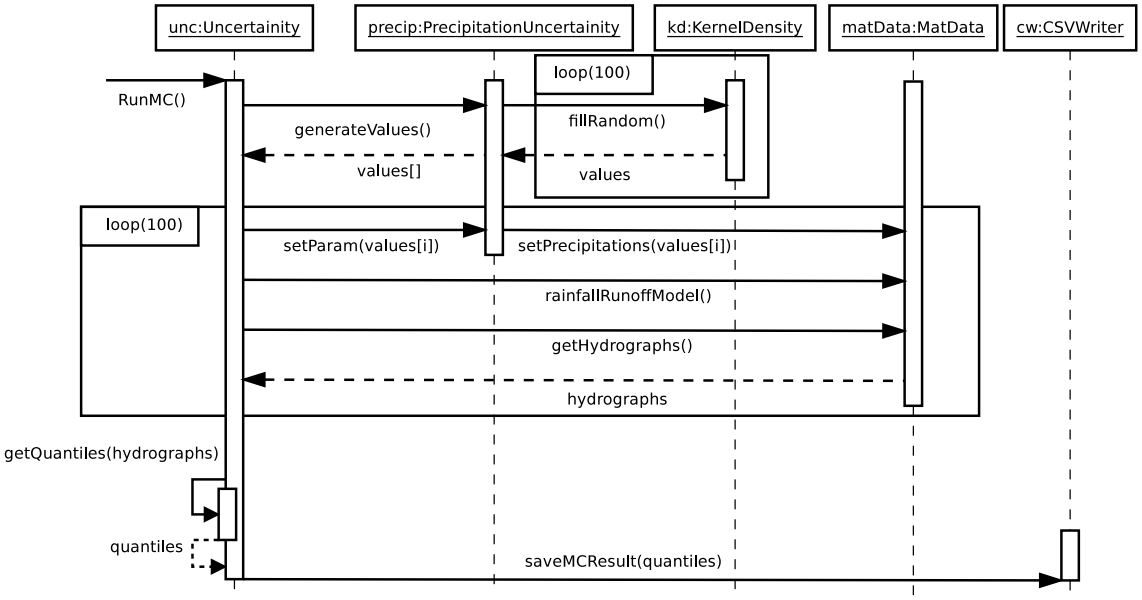

Obrázek 15: Třídní diagram implementace simulace neurčitosti

```
math1d_cl::UncertainityOptions options;
options.simulationCount = mcCount;
options.quantiles = {5.0, 25.0, 75.0, 95.0};
options.precipRandType = math1d_cl::RandomType::KERNEL_DENSITY;
options.daysMeasured = 0;
options.daysPredicted = 7;
options.valuesPerDay = 24;
options.resultsDir = "test_data/results";
options.precipDeviation = 0.0;

options.limits = {
    {3.0, "./cdf/Kernel_estimate_CDF_limit_2.csv"},
    {8.0, "./cdf/Kernel_estimate_CDF_limit_3.csv"},
    {40.0, "./cdf/Kernel_estimate_CDF_limit_4.csv"}
};

options.simParameter.precipitations = false;

// Manning's N
options.simParameter.manning = true;
options.nDeviation = 1.5;
options.nLimits.lower = 0.005;
options.nLimits.upper = 0.1;

// CN Curve number
options.simParameter.cn = true;
options.cnDeviation = 1.5;
options.cnLimits.lower = 30;
options.cnLimits.upper = 100;
```

Výpis 1: Příklad konfigurace simulace neurčitosti

## 4.3 Sekvenční průběh simulace

Inicializace simulace probíhá načtením vstupních souborů a konfigurace modelu. Po inicializaci je spuštěna hlavní část simulace. Po zavolání metody `RunMC` třídy `Uncertainity` se postupně generují sady náhodných hodnot pro zvolené parametry. Hodnoty parametrů v každé sadě jsou předány modelu, který provede vlastní simulaci. Výsledné hydrogramy dílčích simulací jsou uchovávány a po provedení všech iterací je proveden výběr příslušných kvantilů. Hydrogramy příslušící jednotlivým kvantilům jsou následně uloženy do CSV souborů na disk. Na obrázku 16 se nachází přehledový sekvenční diagram popisující jednotlivá volání mezi třídami v průběhu simulace neurčitosti.




Obrázek 16: Sekvenční diagram simulace neurčitosti srážek pro 100 iterací

## 4.4 Generování náhodných hodnot neurčitých parametrů

Třída představující konkrétní neurčitý parametr dědí z abstraktní třídy `AbstractParam` a implementuje mimo jiné metodu `generateValues()`. Tato metoda má za úkol generovat náhodné hodnoty parametru ze zvoleného pravděpodobnostního rozdělení a je volána v cyklu pro každý parametr zahrnutý do simulace neurčitosti. Třída implementující konkrétní rozdělení pak dědí z abstraktní třídy `AbstractRandom` a implementuje metodu `fillRandom()`, která generuje náhodné hodnoty daného parametru pro jednu iteraci.

### 4.4.1 Generování hodnot pomocí jádrových odhadů hustoty

Generování chyby z odhadnutého pravděpodobnostního rozdělení získaného jádrovými odhady zajišťuje třída `KernelDensity`. Při inicializaci této třídy jsou načteny navzorkované křivky kumulativních distribučních funkcí pro každou kategorii intenzity srážek.

Jednotlivé body křivek jsou uloženy do asociativní datové struktury `std:map` jako dvojice souřadnic $x$ a $y$. Souřadnice $y$ v případě kumulativní pravděpodobnostní funkce označuje pravděpodobnost $p \in (0, 1)$ výskytu příslušné hodnoty chyby, reprezentované hodnotou souřadnice $x$. Výhodou kolekce `std:map` je, že data v ní uložené jsou (v závislosti na konkrétní implementaci) reprezentovány ve formě vyváženého binárního vyhledávacího stromu (nejčastěji se jedná o *red-black* strom) [19]. Výhodou takového stromu je složitost vyhledávání klíče, která je v nejhorším případě $O(\log n)$ [19]. Metoda

`fillRandom()` poté generuje jednotlivé náhodné chyby předpovědi, které jsou přičítány k prvkům vektoru srážek, předaného parametrem.

Pro náhodný výběr chyby se nejprve určí kategorie intenzity, do které předpovídaná hodnota srážky spadá. Poté se vygeneruje hodnota kanonické uniformní pravděpodobnosti $p \in (0, 1)$ a v příslušné mapě obsahující navzorkovanou kumulativní distribuční funkci je podle $p$ nalezena nejbližší nižší a nejbližší vyšší hodnota pravděpodobnosti. Nalezeným pravděpodobnostem odpovídají příslušné hodnoty chyby předpovědi, které spolu s hodnotami pravděpodobnosti udávají souřadnice v grafu kumulativní distribuční funkce. Mezi těmito dvěma body je pak dle vztahu 4.1 provedena lineární interpolace vzhledem k hodnotě $p$, jejíž výsledkem je pak konkrétní hodnota chyby předpovědi, která je přičtena k původní hodnotě srážky.

**Definice 4.1** *Vztah pro lineární interpolaci dvou bodů vzhledem k $y$:*

$$x = x_0 + \frac{(y - y_0) \cdot (x_1 - x_0)}{y_1 - y_0}$$

### 4.4.2 Generování hodnot pomocí lineární regrese

Modelování chyby předpovědi pomocí lineární regrese je implementováno ve třídě `Regression`. Při inicializaci jsou načteny odhadnuté hodnoty lineárního regresního modelu pro jednotlivé kategorie intenzity. Hodnoty jsou ukládány do pole, přičemž index odpovídajícího vektoru hodnot regresního modelu je určen z hodnoty vstupního srážkového úhrnu.

Nejprve je vygenerována náhodná hodnota srážkového úhrnu z intervalu určeného původní vstupní hodnotou srážkového úhrnu. Hodnota je generována z uniformního pravděpodobnostního rozdělení, kde hranice jeho horního a spodního intervalu je určena procentuální odchylkou od původního srážkového úhrnu uvedenou v konfiguraci simulace neurčitosti.

Dosazením vygenerované hodnoty do vzorce odhadhu lineárního regresního modelu podle definice 3.3 je poté aplikována možná chybovost modelu Aladin. Hodnoty parametrů regresního modelu jsou určeny výše zmíněným indexem.

### 4.4.3 Generování hodnot z uniformního rozdělení

Hodnoty uniformního rozdělení je možné generovat pomocí třídy `UniformRandom`. Voláním metody `getRandDouble()` je vytvořena instance třídy `std::uniform_real_distribution`, která v C++11 implementuje generování hodnot z uniformního rozdělení. Při vytváření instance této třídy jsou konstruktoru předány příslušné parametry, v tomto případě dolní a horní hranice intervalu pro generování náhodných hodnot. Třídě je pak následně předána instance inicializovaného pseudonáhodného generátoru čísel, která následně vygeneruje náhodné číslo z uniformního rozdělení ze zadaného intervalu.

## 4.5 Předání generovaných hodnot modelu a jeho spouštění

Náhodné hodnoty parametrů jsou modelu předávány voláním metody `setParam()`, kterou implementují všechny třídy, které jsou potomkem třídy `AbstractParam`. Volání metody nastaví jednu sadu parametrů odpovídající jedné iteraci simulace. Tuto metodu v simulaci obyčejně volá třída `Uncertainity`, pokud je do simulace zahrnuto více parametrů je rovněž volána v cyklu pro každý z nich.

Po nastavení jedné vygenerované sady parametrů je spuštěna vlastní simulace srážko-odtokového procesu, voláním metody `rainfallRunoffModel()`. Tento proces nastavení hodnot parametrů a spouštění simulace se opakuje v cyklu pro zadaný počet iterací.

## 4.6 Výběr kvantilů

Po provedení všech dílčích simulací je proveden výběr kvantilů. Pro sestavení hydrogramu pro jeden konkrétní kanál je nutné vybrat daný kvantil z vektoru příslušícímu danému časovému úseku a obsahujícímu simulované hodnoty průtoku pro všechny provedené iterace.

Výběr kvantilu probíhá tak, že vektor hodnot průtoku je vzestupně seřazen a podle jeho velikosti je vypočten index odpovídající hledanému $n$-procentnímu kvantilu. Pokud vektor obsahuje sudý počet hodnot, je kvantil vypočítán jako aritmetický průměr dvou sousedících hodnot nejbližších vypočtenému indexu. V případě lichého počtu hodnot je kvantil určen hodnotou elementu s vypočteným indexem.

Pro seřazení daného vektoru hodnot je využita funkce `std:sort()`, která implementuje algoritmus *Quicksort*. Jeho složitost je v nejhorším případě $O(n^2)$ [19]. Nevýhodou tohoto přístupu k výběru kvantilů je jeho sekvenční implementace. Ačkoli výběr kvantilů představuje malou část času běhu celé simulace, v případě, že bude prováděno velké množství simulací (např. $10^6$), omezení plynoucí ze sekvenční implementace výběru kvantilu bude znatelnější. Vhodnější způsob výběru kvantilů spočívá v použití proprietární knihovny třetí strany. V tomto případě by bylo vhodné použití knihovny *Intel ® Math Kernel Library - MKL*, resp. funkcí poskytovaných její částí - *Vector Statistical Library (VSL)* [20].

## 4.7 Paralelní implementace

Paralelizace výpočtu pro více procesů je realizována pomocí knihovny MPI. Po startu aplikace je inicializován zvolený počet procesů. Jednotlivé procesy z disku načtou schematizaci, srážkové úhrny a konfiguraci své instance modelu. Hlavní proces (rank 0) poté vytváří podle počtu spuštěných procesů a počtu iterací sady náhodných hodnot daných parametrů, které postupně posílá ostatním procesům. Generování hodnot jednotlivých parametrů je prováděno ve více vláknech najednou užitím direktiv OpenMP. První vygenerovaná sada připadne hlavnímu procesu, ostatní sady jsou odeslány zbylým procesům.

Po vygenerování všech sad hodnot jednotlivé procesy postupně nastavují jednotlivé sady měněných parametrů a pro každou z nich spustí srážko-odtokovou simulaci. Samotná srážko-odtoková simulace probíhá paralelně ve více vláknech s využitím knihovny

OpenMP [8]. Výsledné hydrogramy dílčích simulací jsou pak po skončení dílčích simulací seskupeny do výsledného pole pomocí kolektivního volání `MPI_Gather`. Z výsledného pole hodnot jsou poté pro každý kanál a každý časový úsek vybrány hodnoty příslušných kvantilů. Výsledky experimentů se spouštěním simulace neurčitosti pro různý počet procesů a různý počet iterací se nacházejí v podkapitole 5.5.

# 5 Výsledky simulací

V této kapitole se nachází výsledky simulací s využitím reálných dat. Jsou zde uvedeny hydrogramy doplněné o neurčitostní křivky vybraných kvantilů, ve kterých je postupně zahrnuta jak chybovost predikce srážek, tak případná zvolená odchylka čísla CN křivky a hodnoty koeficientu drsnosti. Je zde také provedeno ověření funkčnosti generování náhodné chyby pomocí jádrových odhadů.

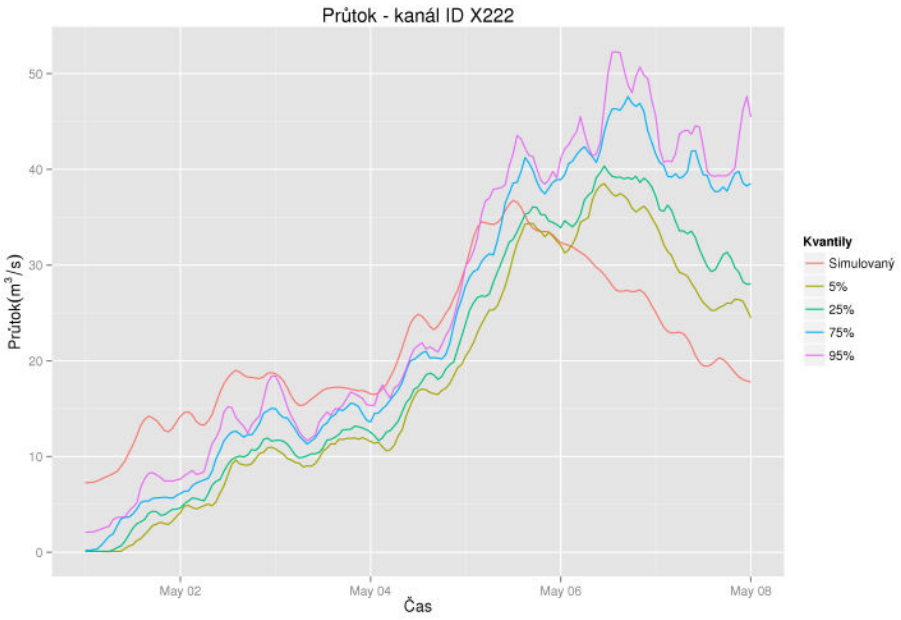
Křivky jednotlivých kvantilů reprezentují pravděpodobnost, s jakou se může hodnota průtoku na daném kanálu odchýlit od simulované hodnoty určené ze srážkového úhrnu predikovaného numerickým modelem Aladin.

Simulace byly prováděny s daty dostupnými v databázi projektu Floreon+ na schematizaci povodí řeky Ostravice pro období 1. - 8. května 2010. Pokud není řečeno jinak, simulace je provedena pro období 7 dní predikovaných srážek. Zobrazeny jsou hydrogramy pro závěrný profil nacházející se před soutokem s Odrou v Ostravě. Z výsledků neurčitostních simulací jsou vybírány kvantily $5\%, 25\%, 75\%$ a $95\%$.

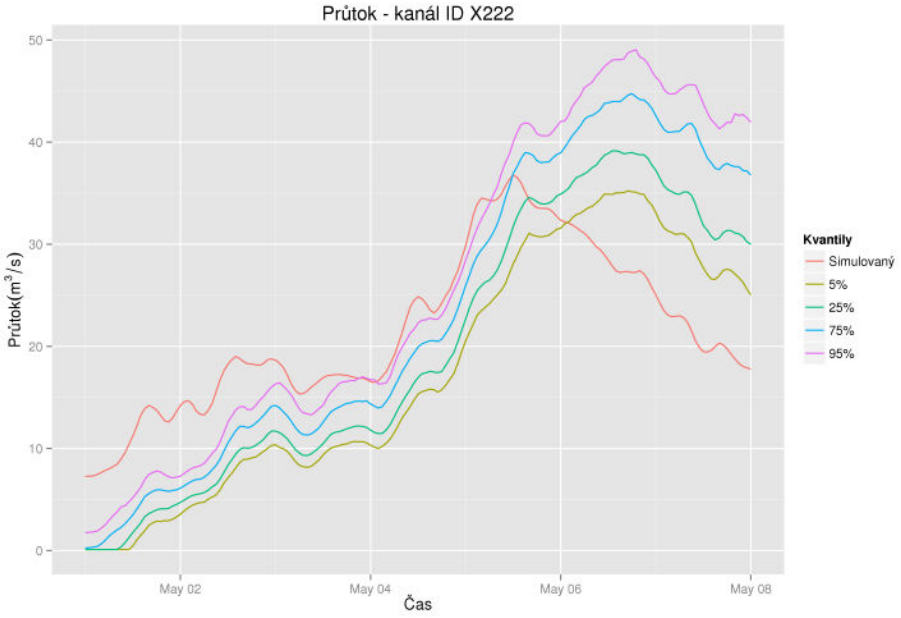
## 5.1 Srážky

Na obrázcích 17 - 19 lze pozorovat vliv počtu iterací na kolísání grafů jednotlivých kvantilů. Se zvyšujícím se počtem iterací se hodnoty jednotlivých kvantilů pohybují blíže své ustálené hodnotě. Kvantily mají tendenci posouvat se nahoru v případě výskytu vysokých srážkových úhrnů $>$ 3mm/h. Původní simulovaná hodnota by měla odpovídat mediánu, kterému se však kvantily se zvyšující se intenzitou srážek vzdalují. Toto je způsobeno příliš širokými intervaly intenzity srážek, kdy např. srážkový úhrn o hodnotě 3,1 mm/h již spadá do intervalu, ve kterém je průměrná chyba 3,6 mm/h (viz. tabulka 5). Tak velká hodnota chyby odpovídá více než 100% odchylce, srážky, což se nutně projevuje ve výsledných hydrogramech.

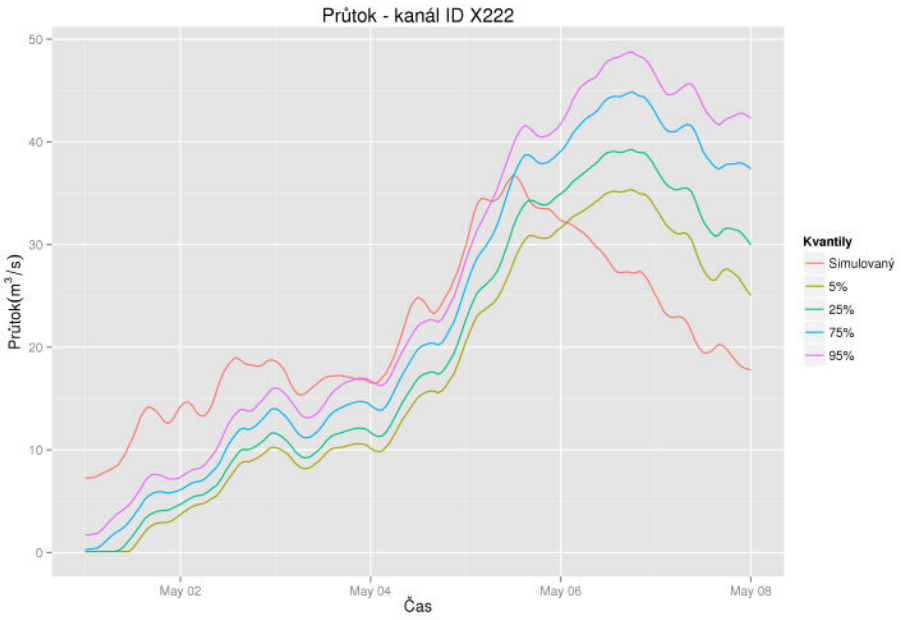
Na obrázku 20 se nachází simulace pro stejný časový úsek s tím rozdílem, že zde jsou pro prvních 5 dní použity měřené srážkové úrhny a pro zbývající 2 dny úhrny predikované. Simulace neurčitosti zde tak má vliv pouze na predikované srážky.



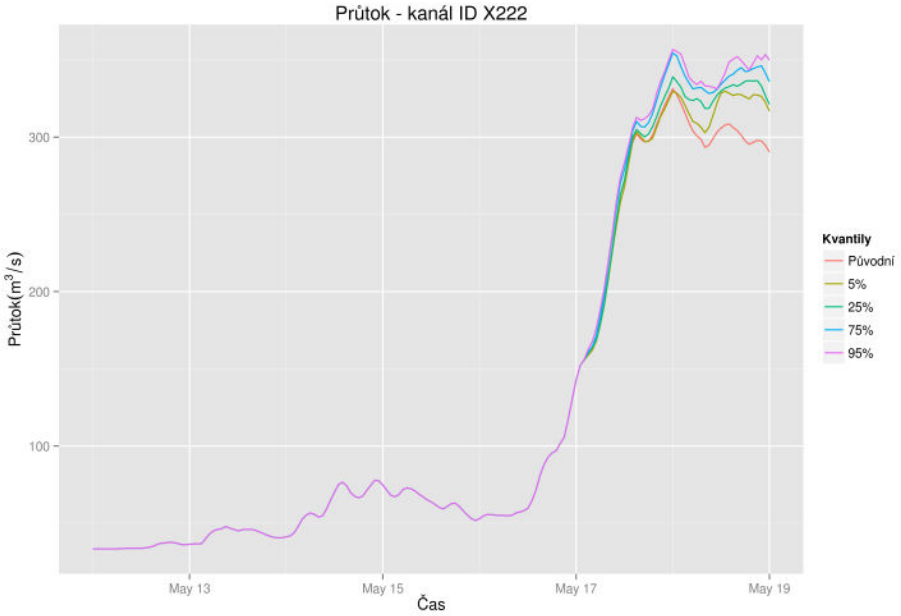

Obrázek 17: Simulace neurčitosti srážek, 10 iterací




Obrázek 18: Simulace neurčitosti srážek, 1000 iterací



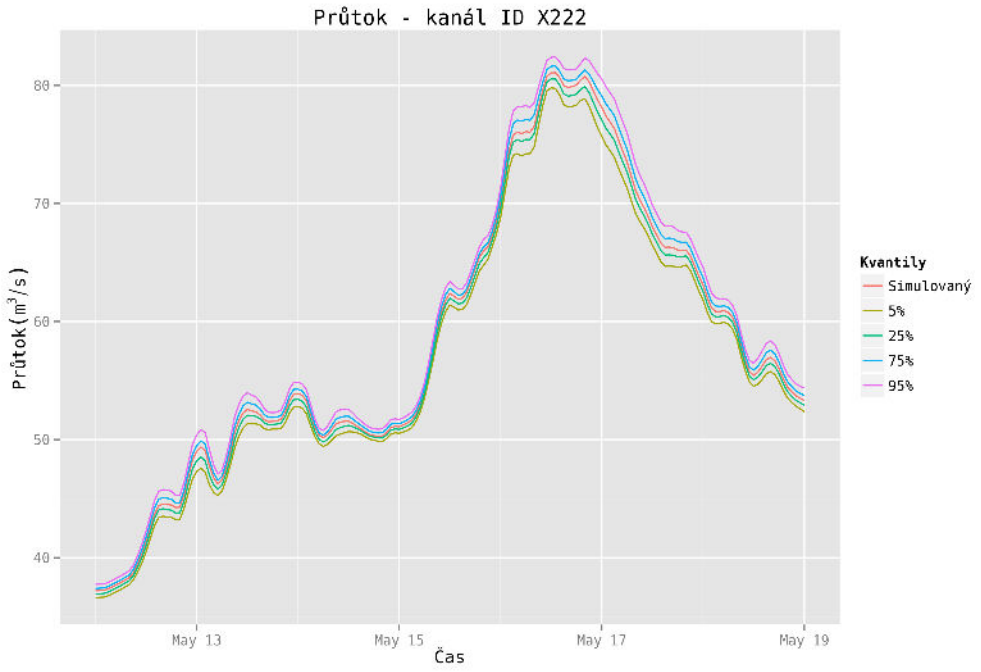

Obrázek 19: Simulace neurčitosti srážek, 20000 iterací



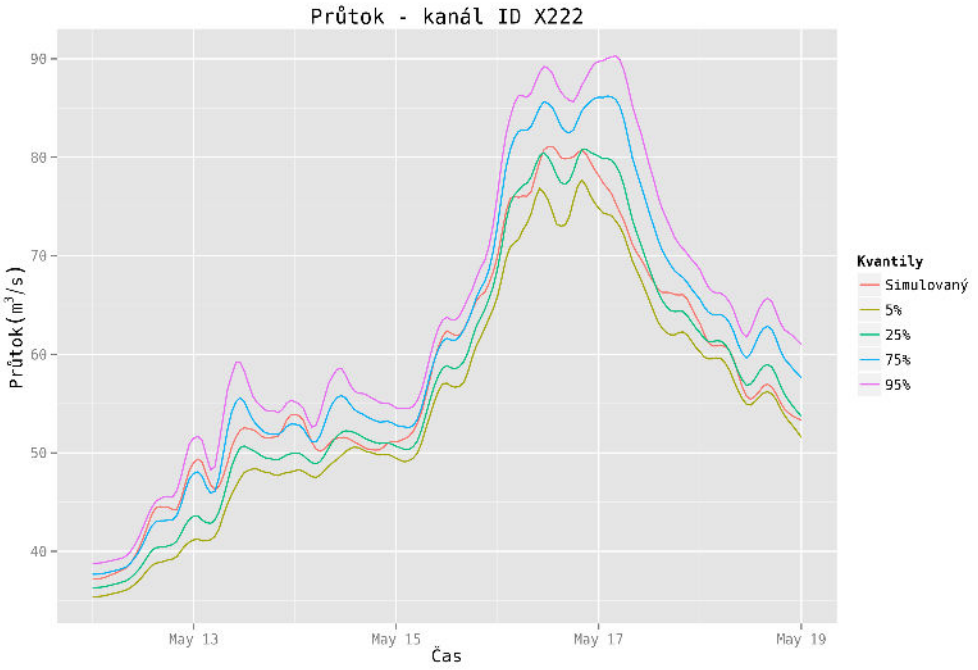

Obrázek 20: Simulace neurčitosti srážek, 100 iterací, 5 dní měřené, 2 dny predikce

## 5.2 CN křivka

Výsledky simulace neurčitosti čísla CN křivky se nacházejí na obrázcích 21 a 22. Simulace jsou prováděny pro 10000 iterací pro danou odchylku. Náhodné hodnoty čísla CN křivky jsou generovány z uniformního rozdělení, neboť pravděpodobnostní rozdělení tohoto parametru není známo.



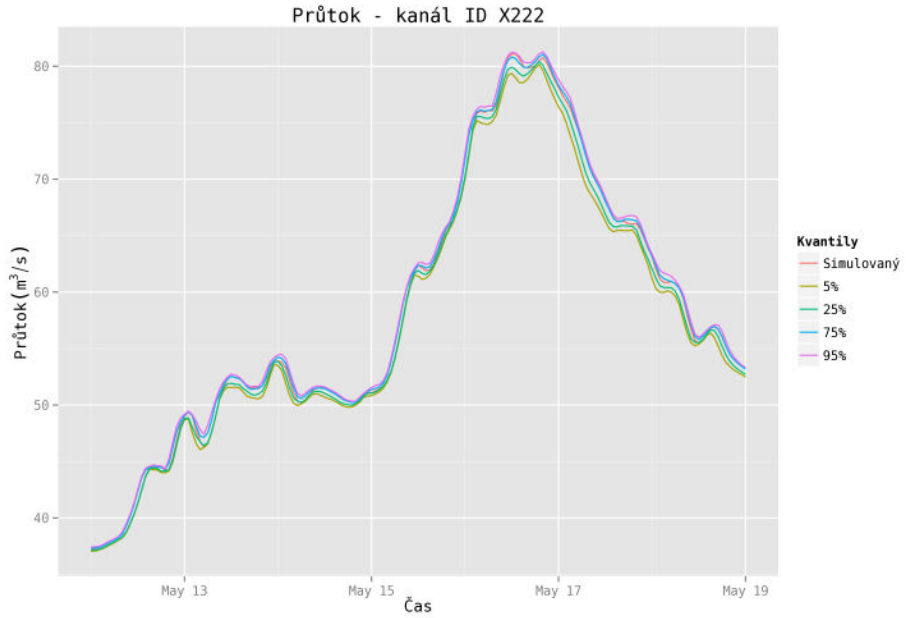

Obrázek 21: Simulace neurčitosti čísla CN křivky, max. odchylka 10 %, 10000 iterací, uniformní rozdělení



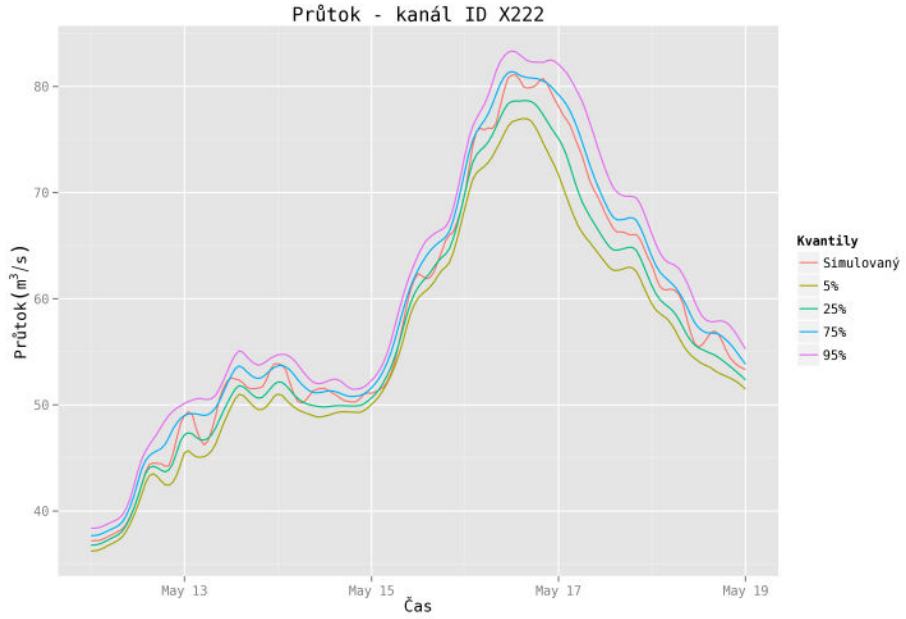

Obrázek 22: Simulace neurčitosti čísla CN křivky, max. odchylka 100 %, 10000 iterací, uniformní rozdělení

## 5.3 Manningův koeficient

Simulaci neurčitosti lze také provádět pro Manningův koeficient drsnosti. Výsledky simulací se nacházejí na obrázcích 23 a 24. Vzorky jsou obdobně generovány z uniformního pravděpodobnostního rozdělení, protože není známo skutečné pravděpodobnostní rozdělení tohoto parametru. Je zde možno vidět rozdíl v citlivosti modelu na změnu tohoto parametru, která je menší oproti parametru CN.




Obrázek 23: Simulace neurčitosti Manningova koeficientu drsnosti, max. odchylka 10 %, 10000 iterací, uniformní rozdělení

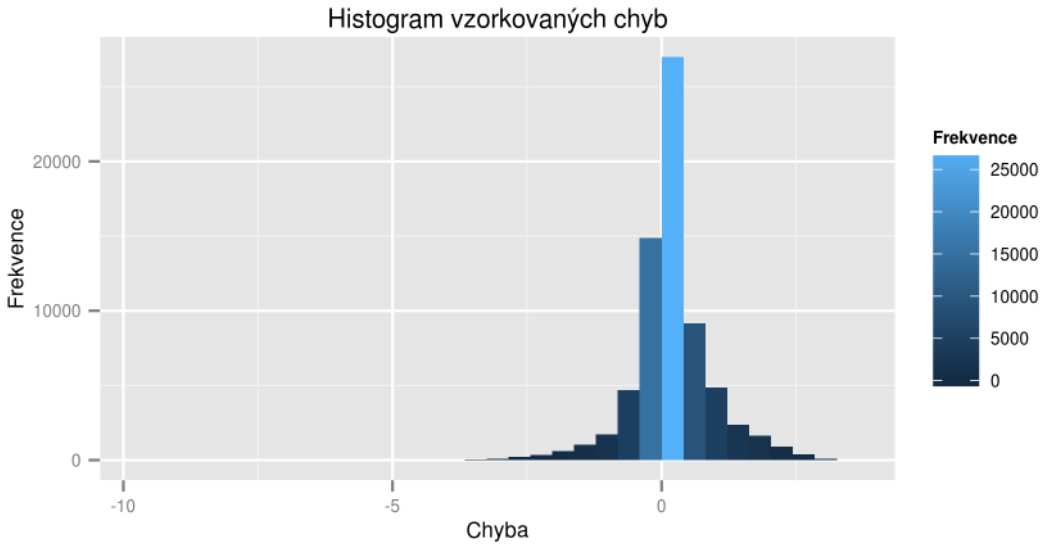



Obrázek 24: Simulace neurčitosti Manningova koeficientu drsnosti, max. odchylka 100 %, 10000 iterací, uniformní rozdělení

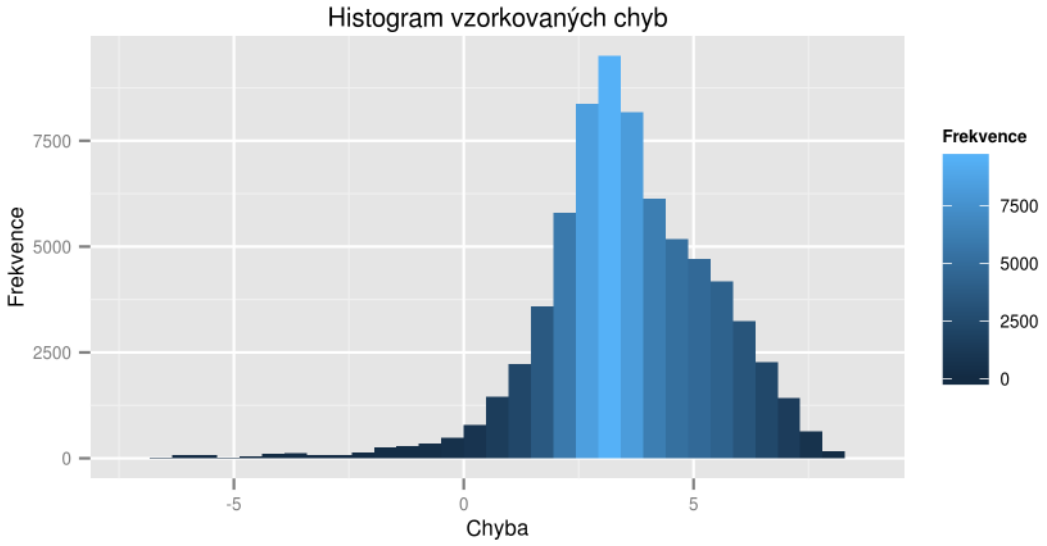
## 5.4 Test hodnot chyby generovaných pomocí jádrových odhadů

Porovnáním histogramu generovaných chyb s původním histogramem chyb lze ověřit, zda je zvolený způsob použití jádrových odhadů korektní. Porovnáním tvaru histogramů generovaných hodnot na obrázcích 25 - 27 s histogramy původních zjištěných hodnot na obrázku 9 lze potvrdit správnost generovaných hodnot, jejichž pravděpodobnostní rozdělení pro jednotlivé kategorie odpovídá původním měřeným rozdělením.

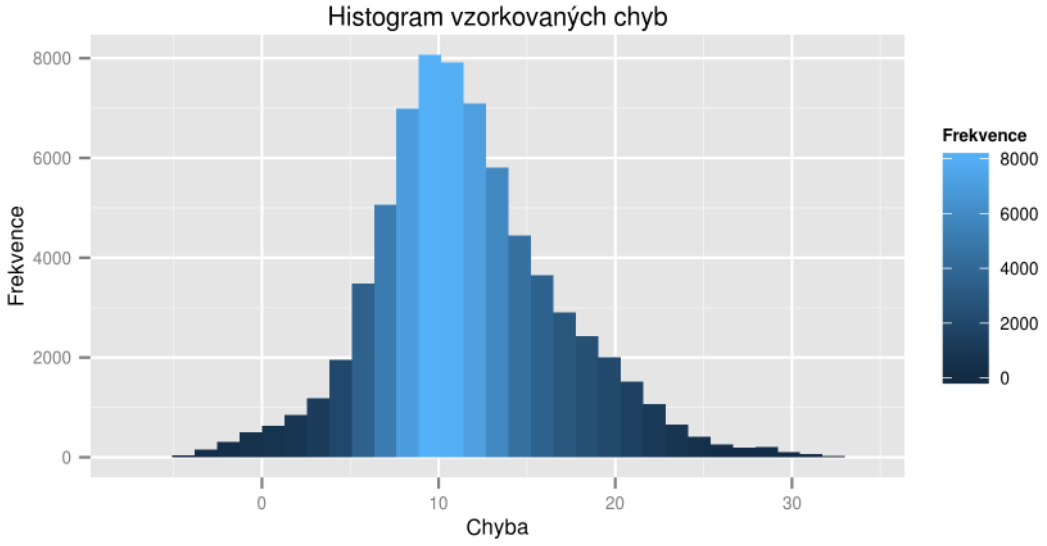
Porovnáním střední hodnoty generovaných a původních hodnot pro různé počty iterací lze odhadnout, jak rychle se v závislosti na počtu vzorků přibližuje střední hodnota chyby generované pomocí jádrových odhadů střední hodnotě chyby získané analýzou původních dat v kapitole 3. Analýzou získaného grafu lze pak také zhruba odhadnout, kolik simulací je nutno provést, dokud se hodnoty chyby neustálí. Na obrázku 28 se nachází graf odchylek střední hodnoty vzhledem k počtu iterací. První dvě kategorie se hodnota odchylky pohybuje okolo nuly, zhruba po vygenerování 20 000 vzorků. U poslední kategorie se hodnota odchylky pohybuje okolo 0,4 při stejném počtu vzorků, což může naznačovat chybně zvolenou vyhlazovací funkci.



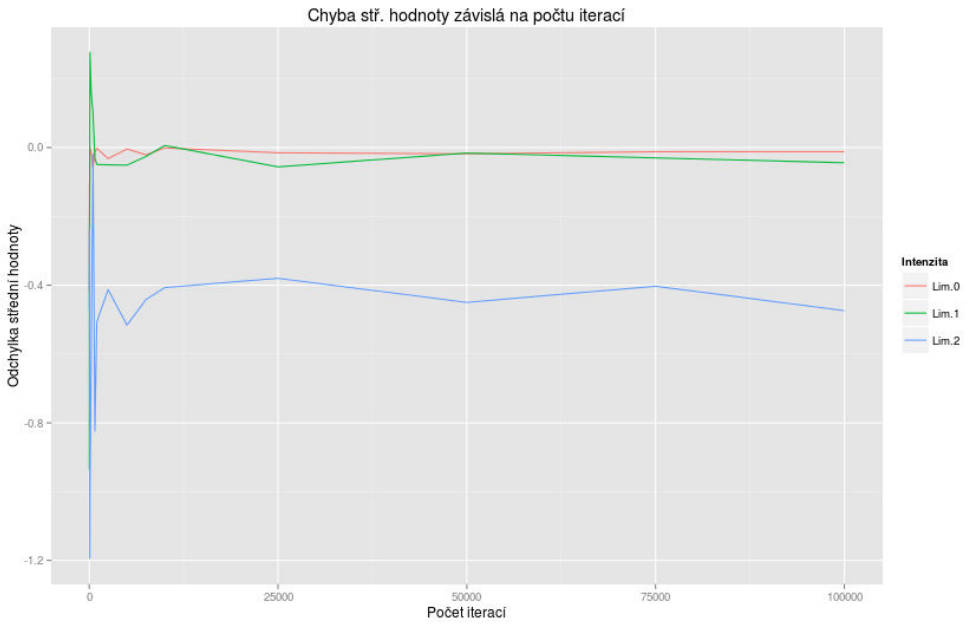

Obrázek 25: Histogram generovaných hodnot chyby pro interval 0,1 - 3,0 mm/h




Obrázek 26: Histogram generovaných hodnot chyby pro interval 3,0 - 8,0 mm/h




Obrázek 27: Histogram generovaných hodnot chyby pro interval 8,0 - 40,0 mm/h




Obrázek 28: Závislost odchylky střední hodnoty na počtu iterací pro všechny intervaly srážek

## 5.5 Rychlost paralelního spouštění na HPC

Podle výsledků v tabulce 10 lze usoudit, že zvolený způsob paralelizace simulace neurčitostí významným způsobem přispěl ke zvýšení rychlosti simulace. Při zdvojnásobení počtu uzlů pro daný počet iterací se čas běhu simulace zkrátí přibližně o polovinu. Při malém počtu iterací je rozdíl méně patrný, neboť je nutno brát v potaz náročnost MPI komunikace. Počet volání MPI je proto omezen, balík změn daného parametru, pro jeden proces je odeslán najednou v jednom volání. Sběr výsledných hydrogramů pro výběr kvantilů je realizován jediným kolektivním voláním.

| Počet iterací | Počet uzlů | Čas (s) |
|---|---|---|
| 100 | 2 | 48 |
| | 4 | 30 |
| | 8 | 12 |
| 1000 | 2 | 340 |
| | 4 | 165 |
| | 8 | 90 |
| 2500 | 2 | 701 |
| | 4 | 351 |
| | 8 | 194 |

Tabulka 10: Časy běhu simulace pro různé počty uzlů výpočetního clusteru

# 6 Závěr

Tato práce se zabývala neurčitostí vstupních parametrů srážko-odtokových modelů. Poznatky získané v průběhu tvorby práce byly použity pro rozšíření funkcionality srážko-odtokového modelu Math1D a přispěly tak k obohacení jeho výstupů o informaci o možné neurčitosti objemu simulovaných průtoků.

V kapitole 2 jsem nastínil problematiku modelování srážko-odtokového procesu. Jsou zde popsány jeho jednotlivé elementy, přístupy k jejich modelování a jejich parametrizace. Popsal jsem zde také model Math1D a metody, které využívá. V této kapitole jsem rovněž popsal možnou neurčitost jeho vstupních parametrů modelu a na základě této analýzy jsem vybral vhodné parametry pro zahrnutí do simulace neurčitosti.

V dalším kroku jsem provedl statistickou analýzu chybovosti predikce srážkových úhrnů generovaných numerickým modelem Aladin, která se nachází v kapitole 3. V této kapitole jsem nejprve navrhl způsob získávání dat z databáze projektu *Floreon+*. Získaná data jsem pomocí statistického software podrobil analýze. Nejprve jsem provedl úvodní explorační analýzu získaného datového souboru a na základě zjištění, že data nepocházejí z normálního rozdělení jsem provedl oboustranný intervalový odhad mediánu chyby. Dále zde byla pomocí statistických testů vyvrácena závislost chyby predikce srážek na stanici, pro kterou byla předpověď vydána, rovněž byla vyvrácena závislost chyby předpovědi na délce časového úseku, na který byla vydána. Oproti tomu se potvrdila závislost chyby na skutečné intenzitě srážek. Byly proto stanoveny kategorie intenzity, které byly následně zahrnuty do simulace neurčitosti. V další části statistické analýzy jsem postupně využil dva přístupy k modelování chyby. Prvním z nich byl bodový odhad parametrů lineárního regresního modelu, kde jsem po jeho provedení přistoupil k verifikaci získaného modelu. Verifikace poukázala na nevěrohodnost získaného modelu, kdy závislost chyby na intenzitě nelze modelovat lineárním regresním modelem. Dalším přístupem, který jsem využil při modelování byly jádrové odhady hustoty pravděpodobnosti. Po prozkoumání této metody jsem pomocí statistického software provedl jádrový odhad hustoty pravděpodobnosti chyby pro každou z výše zmíněných kategorií intenzity. Výsledné navzorkované křivky hustoty pravděpodobnosti jsem po jejich převodu na křivky kumulativní distribuční funkce mohl jednoduše využít pro generování vzorků chyby se zjištěným tvarem pravděpodobnostní funkce.

Princip metody Monte Carlo využité pro simulaci neurčitosti parametrů modelu a její implementaci jsem popsal v kapitole 4. Implementoval jsem obecnou strukturu tříd pro simulaci neurčitosti a její konfiguraci pro predikované srážky, hodnotu CN křivky a Manningova koeficientu drsnosti. Pro simulaci neurčitosti predikce srážek jsem implementoval generování vzorků chyby s využitím výsledků jádrového odhadu hustoty pravděpodobnosti, pro ostatní parametry bylo implementováno generování vzorků jejich hodnot z uniformního rozdělení. Implementaci simulace neurčitosti jsem přizpůsobil pomocí knihoven MPI a OpenMP ke spouštění na výkonných výpočetních clusterech. Zrychlení dosažené touto optimalizací je spolu s výsledky simulací prezentováno v kapitole 5.

Předmětem dalšího vývoje by mohlo být určení vhodnějších kategorií intenzity srážek. Statistická analýza četnosti výskytu srážek o dané intenzitě by mohla odhalit preciznější rozdělení jejich intenzity do kategorií, mohla by se tak zvýšit přesnost výsledků simulace

neurčitosti. Dalším rozšířením této práce by mohlo být navržení dynamického způsobu paralelizace simulace, který by byl vhodný pro běh v heterogenních výpočetních prostředích a využití prostředků dostupných *Intel Math Kernel Library* ®.

# 7 Reference


[1] ČHMÚ - Webové stránky *Vyhodnocení povodňové situace v červenci 1997,* [cit. 13.2.2014], Dostupné z: http://voda.chmi.cz/pov97/obsah.html

[2] *Návod pro pozorovatele meteorologických stanic,* ČHMÚ Ostrava, 2003. 90 s.

[3] ČHMÚ - Webové stránky *Průvodce informacemi Hlásné a předpovědní povodňové služby ČHMÚ pro povodňové orgány,* [cit. 20.3.2014], Dostupné z: http://www.chmi.cz/files/portal/docs/poboc/CB/pruvodce/hydrologicke_predpovedi_v_cr.htm

[4] *Srážkoměr MR3 a MR3H,* Meteoservis v.o.s, 2008. 2 s.

[5] VIEUX, BAXTER E. *Distributed hydrologic modelling using GIS*, Second edition, Dodrecht: Kluwer Academic Publishers, 2004. 312 s. ISBN 1-4020-2460-6.

[6] *Hydrologic Modeling System HEC - HMS Technical Reference Manual*, U.S. Army Corps of Engineers, 2000. 158 s.

[7] VÁCHOVÁ, J. *Matematické Modelování - Model ALADIN*, Západočeská Univerzita v Plzni, 2009. 8 s. Semestrální práce z předmětu matematické modelování.

[8] VAVŘÍK, R. *Vývoj srážko-odtokového modelu Math1D pro spouštění na HPC*, Ostrava, 2014. Diplomová práce na Fakultě elektrotechniky a informatiky VŠB - TU Ostrava. Vedoucí diplomové práce Ing. Štěpán Kuchař.

[9] *Urban hydrology for small watersheds, Technical Release 55*, SCS, USA, 1986. 164 s.

[10] JENÍČEK, Michal. Rainfall-runoff modelling in small and middle-large catchments–an overview. *Geografie–Sbornik ČGS*, 2007, 111: 305-313.

[11] MARTINOVIČ, J., ŠTOLFA, S., KOŽUSZNIK, J., UNUCKA, J., VONDRÁK, I. *Floreon–the system for an emergent flood prediction.*, ECEC-FUBUTEC-EUROMEDIA, Porto, 2008.

[12] LITSCHMANNOVÁ, M. *Úvod do statistiky*, VŠB-TU Ostrava, Fakulta elektrotechniky a informatiky, 2011. 380 s.

[13] SHONKWILER,Ronald W. - MENDIVIL, Franklin *Explorations in Monte Carlo Methods*, Springer Science + Business Media, 2009. 248 s. ISBN 978-0-387-87836-2

[14] KROESE, Dirk. P. - TAIMRE, Thomas - BOTEV, Zdravko I. *Handbook of Monte Carlo Methods*, Wiley, 2011. 775 s. ISBN 978-0-470-17793-8

[15] R Development Core Team *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. 1706 s. ISBN 3-900051-07-0

[16] WICKAM, H. *ggplot2: elegant graphics for data analysis*, Springer New York, 2009. 195 s. ISBN 978-0-387-98140-6

[17] NOVÁKOVÁ, J. *Jádrové odhady distribuční funkce*, Brno, 2009. 66 s. Diplomová práce na Přírodovědecké fakultě Masarykovy univerzity. Vedoucí diplomové práce Mgr. Jan Koláček, PhD.

[18] ZUCCHINI, Walter. *Applied Smoothing Techniques, Part 1: Kernel Density Estimation.*, Temple University, Philadephia, PA, 2003. 20 s.

[19] SOLTER, Nicholas A.; KLEPER, Scott J. *Professional C++*.John Wiley & Sons, 2005. 867 s.

[20] *Intel® Math Kernel Library Reference Manual*.Intel, 2011. 3388 s.