

Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra geoinformatiky

Romana KUDĚLKOVÁ

**PRAVDĚPODOBNOST POŠKOZENÍ POROSTU –
VYUŽITÍ METOD NELINEÁRNÍ REGRESE**

Bakalářská práce

Vedoucí práce: Mgr. Pavel Tuček, Ph.D.

Olomouc 2014

Čestné prohlášení

Prohlašuji, že jsem bakalářskou práci bakalářského studia oboru Geoinformatika a geografie vypracovala samostatně pod vedením Mgr. Pavla Tučka, Ph.D..

Všechny použité materiály a zdroje jsou citovány s ohledem na vědeckou etiku, autorská práva a zákony na ochranu duševního vlastnictví.

Všechna poskytnutá i vytvořená digitální data nebudu bez souhlasu školy poskytovat.

V Olomouci 15. dubna 2014

podpis

Bakalářská práce vznikla rovněž za podpory projektu Vnitřní grantové agentury Univerzity Palackého v Olomouci s názvem Statistické inference nad daty ze senzorů a senzorových sítí - předzpracování, modelování, verifikace, interpretace a vizualizace. Registrační číslo projektu PrF_2013_024.

Děkuji vedoucímu práce Mgr. Pavlu Tučkovi, Ph.D. za podněty a připomínky při vypracování práce. Dále děkuji konzultantům Ing. Pavlu Samcovi a Doc. RNDr. Pavlu Cudlínovi, CSc. za poskytnuté rady k řešené problematice.

Za poskytnutá data děkuji Centru výzkumu globální změny AV ČR, v. v. i..

Za poskytnuté rady a materiály děkuji Ústavu pro hospodářskou úpravu lesa Brandýs nad Labem.

Vložený originál **zadání** bakalářské práce (s podpisy vedoucího katedry, vedoucího práce a razítkem katedry). Ve druhém výtisku práce je vevázána fotokopie zadání.

OBSAH

ÚVOD	6
1 CÍLE PRÁCE	7
2 POUŽITÉ METODY A POSTUPY ZPRACOVÁNÍ	8
2.1 Použitá data	8
2.2 Postup práce	10
2.3 Použité programy	11
3 TEORETICKÁ ČÁST	12
3.1 Metodika hodnocení zdravotního stavu lesa	12
3.1.1 Hodnocení zdravotního stavu lesa pozemním šetřením	13
3.2 Využití regresních modelů v lesnictví.....	14
4 LOKALITA JABLUNKOV	15
5 ZÁKLADNÍ OPERACE S DATY	16
5.1 Úprava dat	16
5.2 Základní statistická analýza	16
5.2.1 Kategoriaální proměnné	17
5.2.2 Spojité proměnné	18
6 PRAVDĚPODOBNOST POŠKOZENÍ POROSTU	21
6.1 Regresní model.....	21
6.2 Predikce a validace.....	23
6.3 Shluková analýza	25
7 VÝSLEDKY	27
7.1 Výsledky základní statistické analýzy	27
7.2 Sestavení regresního modelu.....	28
7.3 Predikce a validace.....	28
7.4 Statistické vyhodnocení výsledků	29
8 DISKUZE	30
9 ZÁVĚR	31
POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE	
SUMMARY	
PŘÍLOHY	

ÚVOD

Poškození lesního ekosystému je v současné době problémem, který je potřeba řešit. V období technologického věku, ve kterém se lidstvo nachází, je více než nutné využít tyto technologie právě k ochraně životního prostředí. V posledních letech je snaha odborníků spojit své síly napříč širokým spektrem jiných oborů a řešit problém společně.

Lesní ekosystém je jedním z nejvíce rozšířených ekosystémů na území České republiky. Je přírodním bohatstvím, domovem mnoha druhů živočichů, kulturním lákadlem, ale také místem, které je velmi citlivé na jakékoliv změny životního prostředí. Tato skutečnost je patrná, zejména jedná-li se o změny dlouhodobějšího charakteru. Tento fakt je mnohdy opomíjen, a proto často převládá přesvědčení, že je les naprosto v pořádku. To však není zcela pravda a například působení imisní zátěže z dob nedávno minulých, je stále akutním problémem.

Další skutečností je, že v posledních letech stále více vzrůstá poptávka po dřevě, jelikož je to jeden z ekonomicky lukrativních způsobů, jak v době ekonomické krize získat finanční prostředky, je ho využíváno o to více. Nejde však o samotnou těžbu dřeva, jde především o nepochopení základních principů lesního ekosystému. Dřevo nestačí jen těžít, ale je potřeba tento ekosystém postupně znovuobnovovat. Slovo postupně je jedno z nejdůležitějších v pochopení tohoto problému. Je nutné soustavně přemýšlet i o budoucnosti tohoto ekosystému. Správné pochopení této problematiky nepřinese pouze spokojenost ochráncům lesního porostu, ale také možnost stále ekonomicky využívat tento ekosystém, a především, což je nejdůležitější, dědictví budoucí generacím, ať už v podobě možnosti poznat jedinečný koloběh přírody, tak i v podobě zkušeností a poučení se, že člověk je malým pánem a měl by se daleko lépe starat o bohatství, jež nám planeta Země dennodenně nabízí a poskytuje.

Hodnocení stavu poškození porostu je jedním ze základních stavebních kamenů v problematice hospodaření s lesními porosty. Ačkoliv byly informace o poškození porostu dlouhou dobu získávány pouze pozemním šetřením, které prováděl kvalifikovaný pracovník, je nemožné, aby subjektivní vnímání do tohoto procesu nezasáhlo. Proto zde byl požadavek ohodnotit tuto charakteristiku matematicko-statistickými metodami a pokusit se eliminovat tento subjektivní vliv.

Práce má za úkol určit poškození porostu vybranými metodami a následně vyhodnotit správnost hodnot získaných pozemním šetřením. Pomocí vhodných vizualizačních metod a grafických výstupů podává spolu s textovou částí jedno z řešení daného problému objektivního hodnocení pravděpodobnosti poškození porostu.

1 CÍLE PRÁCE

Cílem bakalářské práce je zpracovat data o desetiletém pozorování lesního porostu v okrese Frýdek-Místek a vyhodnotit na základě doporučených matematicko-statistických metod pravděpodobnost poškození porostu. Data budou poskytnuta výzkumným ústavem CzechGlobe. Součástí práce bude také vytvoření teoretické rešerše, která se bude zabývat expertním hodnocením poškození porostu zejména metodikou a pozemním šetřením prováděným odborníky a využitím logistické regrese a nelineárních regresních modelů v lesnictví.

Na datech bude provedena základní statistická analýza, pomocí níž budou data statisticky vyhodnocena a výsledky interpretovány v podobě tabelárních a grafických výstupů. Bude také sestaven vhodný regresní model pro modelování pravděpodobnosti poškození porostu. V tomto kroku dojde také ke stanovení statisticky významných faktorů ovlivňujících poškození porostu a k určení parametrů modelu. Následně proběhne statistické vyhodnocení výsledků metodou shlukové analýzy. Výsledky této analýzy budou zobrazeny pomocí vhodné vizualizace.

Výsledkem práce bude jednak validace expertního hodnocení, ale také predikce této charakteristiky v místech, kde není expertně ohodnocena. Tyto výsledky budou zobrazeny pomocí vhodné vizualizace a datových náhledů obsahujících výsledky a predikce. K bakalářské práci budou rovněž vytvořeny internetové stránky.

2 POUŽITÉ METODY A POSTUPY ZPRACOVÁNÍ

2.1 Použitá data

Data byla poskytnuta výzkumným ústavem CzechGlobe v podobě .xls dokumentu. Jednalo se o desetiletá pozorování vybraných charakteristik lesů v okrese Frýdek-Místek. Sledovaným druhem stromů byl smrk. Tabulka obsahovala 33 sloupců a 7361 záznamů.

Jako identifikátory jednotlivých oblastí byly použity kódy skládající se ze tří jednotek prostorového rozdělení lesa. Jak uvádí Sequens (2007), první část kódu lze označit jako oddělení, což je nejvyšší jednotka prostorového rozdělení lesa, jeho význam je především orientační a je označen arabskými číslicemi. Druhá část je dílec, který bývá označován velkými písmeny a je vytvářen na základě podobnosti přírodních podmínek s cílem postupného dosažení jednotného způsobu hospodaření. Třetí částí je porostní skupina vylišující se jako část porostu, u něhož se v důsledku vývoje bude měnit hranice, je označena číslem příslušného věkového stupně vždy s intervalem deseti let.

K takto definovaným oblastem byly doplněny další údaje:

- rok – časový údaj, ke kterému se vztahují naměřené a určené charakteristiky,
- plocha porostní skupiny – plocha území, kterou zaujímá porost příslušného věkového stupně dané oblasti, udává se v hektarech,
- zastoupení dřeviny – poměr redukováných ploch, plocha, kterou pokrývá právě jeden druh, udává se v procentech,
- plocha dřeviny – zastoupení dřeviny převedené na plošnou jednotku, udává se v hektarech,
- věk 51–80 – věková kategorie porostu 51–80 let, udává, zda porost v příslušné oblasti do této kategorie spadá či nikoliv, je označen hodnotami 1 a 0 tzv. pravda a nepravda,
- věk ≥ 80 – věková kategorie porostu 80 a více let, udává, zda porost v příslušné oblasti do této kategorie spadá či nikoliv, je označen hodnotami 1 a 0 tzv. pravda a nepravda,
- zakmenění – veličina sloužící k určení stupně obsazení plochy lesním porostem, lze ji vypočítat jako poměr redukované plochy jednotlivé dřeviny k celkové ploše porostní skupiny, častěji však bývá odhadnuta a označena dle standardů určitým stupněm zakmenění,
- lesní typ – kategorická proměnná, nemění se v čase pouze v prostoru, udává podmínky pro růst (úživnost), skládá se ze tří částí, první částí je vertikální členění neboli vegetační stupeň označován čísly 1–9 (v případě čísla 0 se jedná o ostrovní výskyt, vzniklý po době ledové), druhou částí je horizontální škála, tzv. edafické kategorie zahrnující hodnotu půdy, dostatek živin a vody, označeny jsou velkými písmeny např. B – bohatá, třetí částí je indikovaná bylina, označována čísly, ale od této charakteristiky se postupně upouští,

- SLT – soubor lesních typů, skládá se z první a druhé části lesního typu,
- hospodářský soubor – jedná se o generalizaci souboru lesních typů, je dán vyhláškou Ministerstva zemědělství č. 83/1996 Sb., o zpracování oblastních plánů rozvoje lesů a o vymezení hospodářských souborů (ÚHÚL, 1996),
- nahodilá těžba – udává se v m³, jde o těžbu způsobenou vlivem škodlivých faktorů – kůrovec, ostatní hmyz, tracheomykóza, živelné pohromy, ostatní,
- celkový součet nahodilé těžby – udává se v m³, součet všech nahodilých těžeb bez důrazu na jejich vznik,
- zásoba daného roku – dva údaje, první je zásoba daného roku po odečtení nahodilé těžby z předchozích let, druhá je zásoba daného roku po odečtení nahodilé těžby z předchozích let a aktuálního roku,
- procentuální vyjádření nahodilé těžby – udává se v procentech, jedná se o procento nahodilé těžby ze zásoby postupné a počáteční,
- kumulativní procentuální vyjádření nahodilé těžby – udává se v procentech, získává se kumulativním sčítáním procenta nahodilé těžby ze zásoby počáteční,
- orientace – orientace vůči světovým stranám, udává se ve stupních,
- sklon – sklon svahu, udává se ve stupních,
- nadmořská výška – výškový rozdíl mezi sledovanou lokalitou a hladinou moře, udává se v metrech nad mořem,
- ozáření – intenzita záření dopadající na povrch, je vyjádřena kategoricky,
- lesní vegetační stupeň – vertikální členění lesních porostů podle charakteristických typů dřevin pro určité změny klimatu a teploty, vyjádřen kategoricky čísly 0–9,
- trofická úroveň – úživnost, půdní podmínky daného porostu, k vyjádření těchto podmínek se používají trofické řady, zde je však použita číselná klasifikace,
- hydrická úroveň – vlhkostní režim půd, k vyjádření se používají hydrické řady, zde je však použita rozšířená klasifikace,
- přesný věk – skutečný, přesný věk porostu,
- vertikální struktura – rozrůzněnost, nemění se v čase jen v prostoru,
- počáteční poškození – poškození na začátku plánu vlivem biotických nebo abiotických činitelů,
- imise – imisní zátěž, určena kategoricky,
- zdravotní stav – zdravotní stav porostu, určen kategoricky
- procentuální poškození stromů – udává se v procentech, zjišťuje se pozemním šetřením, odborník sleduje zejména stav a olistění koruny, stav kmene a kořenového systému.

2.2 Postup práce

Nejdříve bylo potřeba získat data, s nimiž mělo být pracováno. S tímto krokem nebyl problém, poněvadž výzkumný ústav CzechGlobe již dříve projevil zájem o práci na téma poškození porostu z dat, která má tato firma k dispozici. Po poskytnutí těchto dat proběhlo seznámení s určitými aspekty, které bylo nutné prokonzultovat s odborníky. V listopadu roku 2013 proběhly tyto konzultace, které byly nezbytné ke správnému pochopení a interpretaci lesnických charakteristik a odbornému vysvětlení zkoumané problematiky. Těmito odborníky byli Ing. Pavel Samec a Doc. RNDr. Pavel Cudlín, CSc., kteří díky svým zkušenostem a dlouhodobé práci v oboru lesnictví poskytli požadované informace na toto téma.

Následovalo vypracování rešeršní části, která byla věnována samotnému poškození porostu a metodám jeho zjišťování především s důrazem na pozemní šetření. Dále se tato část zabývá již řešenými pracemi v oboru lesnictví, při nichž bylo využito regresních modelů.

Dalším krokem byla úprava dat, jelikož v těchto datech byly nalezeny četné chyby, a jejich předzpracování do požadovaného formátu, v němž následně mělo dojít ke zpracování základní statistické analýzy těchto dat. Tato úprava byla časově náročná, neboť bylo mnohokrát téměř nemožné zjistit, jakým způsobem byly vytvořeny stupnice pro zařazení do určitých kategorií zkoumaných veličin.

Dále byl proveden výběr vhodného druhu regrese v závislosti na zkoumaném jevu, tzn. pravděpodobnosti poškození porostu. Jelikož se nabízela velká řada možností, byla zvolena konzultace s odborníky z katedry matematické analýzy Univerzity Palackého. Po poradě s nimi byla zvolena beta regrese, jejíž vlastnosti nejlépe odpovídaly požadavkům. Dle zvoleného druhu regrese byl zhotoven model pro výpočet pravděpodobnosti poškození porostu, pomocí kterého došlo k validaci a predikci této charakteristiky. Modelů bylo vytvořeno více, avšak podle určitých kritérií byl zvolen právě jeden.

Výsledky z toho modelu byly statisticky vyhodnoceny pomocí základních statistických ukazatelů a následně interpretovány. Byla také provedena shluková analýza pro statisticky významné faktory na pět, sedm a deset shluků podle určitých let. Jako výstup byla vyhotovena heat mapa, z níž díky dobré barevné názornosti a odlišnosti je možné vyčíst, které oblasti jsou si podle významných faktorů ovlivňujících poškození porostu v určitých letech podobné, a zda se během let tato příslušnost do určitého shluku mění. Shluky byly rovněž statisticky popsány a charakterizovány.

V posledním bodě byla provedena vhodná vizualizace. Požadavkem výzkumného ústavu CzechGlobe bylo vytvořit rozdílové mapy, z nichž je možné lépe sledovat změnu poškození porostu mezi rokem 2008 a ostatními roky. Těchto map bylo zhotoveno deset. Pro názornost byl však z jedné z nich vytvořen poster. Do něho byly zakomponovány i výsledky shlukové analýzy, které nesou zajímavé informace, a proto je vhodné, aby zde byly vloženy.

2.3 Použité programy

Pro práci bylo použito několik programů. Pro práci s daty byl použit Microsoft Office Excel 2007, ve kterém došlo k jejich úpravě a náhledu. Následovala základní statistická analýza dat, jež byla provedena také pomocí programu Microsoft Office Excel 2007.

Složitější úkony statistické analýzy byly řešeny v programu RStudio, kde také probíhala konstrukce vhodných regresních modelů. Ve stejném prostředí byla zpracována shluková analýza a výsledné vyhodnocení výsledků. Statické vyhodnocení výsledků proběhlo v Microsoft Office Excel 2007 a programu RStudio. Vizualizace poškození porostu byla řešena pomocí ArcGIS 10.1 a výsledné grafické úpravy ve volně dostupném programu Inkscape.

3 TEORETICKÁ ČÁST

3.1 Metodika hodnocení zdravotního stavu lesa

Hodnocení příčin poškození zdravotního stavu lesa je komplikováno skutečností, že charakter a množství škodlivých faktorů podléhají stále rychleji změnám a lesní ekosystémy reagují na tyto změny s určitým zpožděním. Monitorování zdravotního stavu lesa je při jeho současném stavu nezbytnou součástí nástrojů potřebných pro kvalifikovaná a včasná rozhodnutí v oblasti ochrany lesních ekosystémů jako důležitého prvku životního prostředí. V současné době se klade stále větší důraz na zajištění vysoké a trvalé produkce kvalitní dřevní hmoty a současně zachování ochranných a společenských funkcí lesa, tzn. princip trvale udržitelného rozvoje (Mze & VULHM, 2004).

Vzhledem k této skutečnosti bylo potřeba sjednotit tyto metodiky a vlivem tohoto rozhodnutí se začal monitoring stavu lesa provádět v rámci celoevropského mechanismu získávání dat o stavu lesních ekosystémů a následné možnosti jejich porovnávání. Tato snaha vytvořit standard hodnocení lesních porostů vyústila ve vznik celoevropské sítě monitorovacích ploch Mezinárodního kooperativního programu sledování a vyhodnocování vlivu znečištění ovzduší na lesy, zkráceně označované jako ICP Forests. Tento program představuje jeden z nejdůležitějších systémů, využitelných pro posílení snah kontrolovat lesní ekosystémy (Mze & VULHM, 2004).

Česká republika vstoupila do programu ICP Forests v roce 1986 s 61 monitorovacími plochami v síti 16x16 km. Počet ploch se v následujícím roce zvýšil na 106. Pro tuto nadnárodní síť byla přijata dosavadní koncepce uplatňovaná v České republice již od padesátých let, kdy byly založeny první monitorovací plochy na sledování vlivu SO₂ na lesní porosty. V roce 1991 byla založena síť 8x8 km s dalšími 334 plochami. Kromě toho byly v rámci regionálních studií založeny plochy v síti 1x1 km v lesních oblastech Šumava, Brdy a Krkonoše. Informace z těchto monitorovacích ploch umožnily získat podrobný a reprezentativní obraz o situaci v daných lesních oblastech a porovnat zdravotní stav porostů se stavem půd a výživou porostů. V letech 1997–1998 byla provedena celková rekonstrukce ploch národní i nadnárodní sítě s cílem optimalizovat druhovou a věkovou skladbu na monitorovacích plochách tak, aby lépe charakterizovaly skutečnou skladbu v lesích České republiky. V současné době je pravidelné hodnocení I. úrovně prováděno na plochách základní sítě 16x16 km a vybraných plochách ze sítě 8x8 km v celkovém počtu 306 ploch, které jsou rozmístěny rovnoměrně podle lesnatosti po celém území České republiky (VULHM, 2014)

Na základě programu ICP Forests probíhala celoplošná reprezentativní šetření (stav koruny, porostní charakteristiky, půdní poměry atd.). Na každé monitorovací ploše byly zjišťovány základní stanovištní a porostní charakteristiky (souřadnice, výška n. m., orientace, věk, zastoupení dřevin, dostupnost vody aj.). V pravidelných intervalech (1–5 let) se prováděla tato odborná šetření: hodnocení stavu koruny (defoliace, barevné změny, mechanické poškození aj.), zjišťování sociálního postavení, měření dendrometrických

parametrů a fytoocenologické snímkování. V nepravidelných intervalech se jako doplňující šetření prováděly letokruhové analýzy a jednorázově byly provedeny listové a půdní analýzy (Mze & VULHM, 2004).

Jak udává VULHM (2014), směrnice ustavující program ICP Forests vypršela 31. 12. 2002. Program ICP Forests zůstal součástí CLRTAP, mezinárodní financování monitoringu zdravotního stavu lesů bylo od roku 2003 zajištěno v rámci programu Evropské komise Forest Focus, ustanoveného směrnicí ES 2152/2003. Cílem projektu FutMon bylo pokračovat v monitoringu zdravotního stavu lesů v systematické síti ploch a harmonizovat metody sledování s programy národních inventarizací lesů v jednotlivých evropských zemích. Na plochách intenzivního monitoringu bylo cílem zkvalitnit již probíhající měření a zároveň doplnit některé sledované parametry, aby bylo hodnocení skutečně komplexní a vyhovovalo některým používaným modelům. V současné době je monitoring zdravotního stavu lesa financován pouze z národních zdrojů, v České republice poskytuje finanční prostředky Ministerstvo zemědělství.

Jedním z nejvíce poškozených lesních porostů jsou porosty smrkové. Dle Samce a kol. (2012) se rychlé plošné hynutí těchto porostů v několika středoevropských zemích stává stále závažnějším problémem, proti němuž jsou vyvíjeny strukturované systémy lesnických opatření. Současná vlna plošného hynutí začala po roce 1995 navzdory souběžnému odsiřování velkých průmyslových zdrojů znečištění životního prostředí. Nejdříve postihla průmyslovou Slezskou nížinu v Polsku, odkud se přesouvala na jihovýchod do Vnějších Západních Karpat a Nízkého Jeseníku. Plošné chřadnutí smrkových porostů v Beskydech začalo přibližně v roce 2002. Mezi hlavní příčiny tak rychlého chřadnutí v souvislé části střední Evropy patří vedle zbytkových projevů okyselování ekosystémů i celková ztráta autoregulace hospodářsky změněných geobiocenóz. Průběh kalamity nastolil otázky ke způsobům dostatečně rychlé a efektivní přestavby lesů tak, aby nebyly trvale narušeny jejich přirozené, ani hospodářsky významné funkce.

3.1.1 Hodnocení zdravotního stavu lesa pozemním šetřením

Zdravotní stav lesních dřevin je charakterizován při pozemním šetření především stupněm defoliace, jež je definován jako relativní ztráta asimilačního aparátu v koruně stromu v porovnání se zdravým stromem rostoucím ve stejných porostních a stanovištních podmínkách. Je to ztráta, která je způsobena především vlivem nepříznivých změn prostředí lesních ekosystémů jako důsledek dlouhodobého a nadměrného znečištění ovzduší různými škodlivými látkami (Mze & VULHM, 2004).

Hodnocení defoliace se provádí na monitorovacích plochách v období srpen – září. Na každé ploše musí být zhodnoceno v každém roce minimálně 25 stromů, přitom jejich sociální postavení musí mít vyšší hodnotu než 4 (podle Krafta). Při hodnocení defoliace se do ní nezahrnuje ztráta asimilačního aparátu, včetně celých větví nebo částí koruny, která je způsobena mechanickým vlivem různých škodlivých faktorů prostředí (vítr, sníh, námraza, nešetná těžba). Na částečné zkreslení skutečného stavu a větší subjektivitu při hodnocení má vliv i skutečnost, kdy v důsledku působení stresových faktorů v lesním

prostředí dochází ke zkracování letorostů a tím také k zahuštění koruny a zvýšení zápoje. Hodnocení defoliace by nemělo být ovlivněno transparentností koruny, přitom je nutné vzít v úvahu také různé typy větvení koruny (především u smrku). Defoliace se vyjadřuje procenty v intervalech po 5 %. Hodnotí se vizuálně, a je proto zatížena určitou chybou vyplývající ze subjektivního vlivu hodnotitele (Mze & VULHM, 2004).

Podle Mze & VULHM (2004) probíhá hodnocení stavu koruny na celé monitorovací ploše, což je důležité z hlediska následného vyhodnocení a statistického zpracování dat. Toto sledování se provádí i u stromů se sociálním postavením 4, údaje o těchto stromech však nejsou zahrnuty do standardního statistického zpracování. Evidovány jsou také mrtvé stromy a to do doby, než dojde k jejich vytěžení. Tato informace je využívána např. pro vyhodnocování mortality porostu nebo pro srovnání výsledků pozemního šetření a dálkového průzkumu zdravotního stavu lesa.

3.2 Využití regresních modelů v lesnictví

Jednou z možností využití regresních modelů v lesnictví je hodnocení vlivu různých faktorů na zdravotní stav a poškození porostu. Nejčastěji se jich využívá při přírodních katastrofách postihující lesní porost. V těchto případech se především využívá regrese nelineární. Nelineární regrese byla využita Cruzem a kol. (2005) pro určení rychlosti šíření požáru na jemně zvlněném terénu v lesních porostech v závislosti na typu těchto porostů a jejím zápoji. Tato analýza byla založena na experimentálním souboru dat týkající se především boreálních druhů lesních paliv. Tento model má několik nedostatků a oblastí, které by se daly zlepšit, i tak ale může být využit na podporu požárního managementu a další studie.

Logistický regresní model a Coxův regresní model byl aplikován Kamimuro a kol. (2013) k zjištění poškození lesů způsobené tropickou bouří v Japonsku. Vše bylo aplikováno na lesní porosty především Kryptomérie japonské. Byl zde zkoumán vliv různých vlastností lesního porostu a meteorologických vlastností tropické bouře na poškození těchto porostů. Výsledkem této analýzy bylo vyhodnocení stability porostů vůči tropickým bouřím.

Další možností je vytvoření růstových modelů. Lekwadi a kol. (2012) využili Chapman-Richardsonovu nelineární růstovou funkci a pomocí NLQR (Nonlinear quantile regression) byl sestaven růstový model pro smrk Sitka v lesních porostech Irsko. Společně s určitou klasifikací je to důležitý nástroj pro rozhodování týkající se hospodaření v těchto oblastech.

U dalšího růstového modelu byla využita logistická regrese. Gillner a kol (2013) použil tuto regresi k vytvoření růstového modelu buku lesního ve východním Německu. Na tomto území byly vytvořeny modely růstu z letokruhové analýzy, ty však nezohledňovaly určité změny v růstu. Proto byl sestaven tento model, aby tak došlo k doplnění informací z původně používaného modelu a k lepší schopnosti predikovat úmrtnost stromů.

4 LOKALITA JABLUNKOV

Poškozené lesy, které jsou předmětem zájmu této bakalářské práce, se nachází v Moravskoslezském kraji, v okrese Frýdek-Místek, v Lesní správě Jablunkov. Jedním z klasifikačních celků jsou revíry, jichž se na tomto území nachází hned několik. V případě zkoumaných dat jde konkrétně o pět revírů, z nichž bylo možné tato data získat.

Celá tato oblast Lesní správy Jablunkov se nachází u hranic České republiky s několika státy. Na východě je to hranice s Polskem a na jihu se Slovenskem. V severní části sousedí s Lesní správou Ostrava a na západě s Lesní správou Frýdek-Místek. Jediným městem této LS (lesní správa) je Jablunkov, který je také nejvýchodnějším městem České republiky, a je tak významnou křižovatkou a centrem pro mnoho okolních obcí.

Z hlediska obecně zeměpisné charakteristiky je toto území lesní správy Jablunkov rozděleno řekou Olší a spojnici Jablunkov-Jablunkovský průsmyk na tři části. Na východní část Slezské Beskydy, kde dominuje Čantoryje (954 m n. m.) a Stožek (978 m n. m.). Na jihovýchodní část Jablunkovskou vrchovinu s Gírovou (840 m n. m.) a na západní, největší část Moravskoslezské Beskydy s Velkým Polomem (1067 m n. m.), Malým Polomem (1061 m n. m.), Kozubovou (982 m n. m.), Ostrým (1044 m n. m.) a Javorovým (1032 m n. m.). Nadmořská výška od 350 m n. m. do 1067 m n. m. (Lesy ČR, 2012).

Podle Samce a kol. (2012) patří Podbeskydský a Beskydský bioregion k územím, kde došlo k plošnému chřadnutí kulturních smrkových porostů. V Podbeskydském bioregionu teoreticky převládají silně rizikové podmínky pro růst smrkových porostů, které tvoří 76% území a zahrnují 70% rozlohy zdejších lesů. Tyto změny jsou viditelné i na fotografiích z této oblasti (Obr. 1).



Obr. 1 Fragmentované lesy (autor: Ing. Pavel Samec).

5 ZÁKLADNÍ OPERACE S DATY

5.1 Úprava dat

Jak už bylo zmíněno (viz. kapitola 2.1 „Použitá data“), data použitá k práci obsahovala několik tisíc záznamů o stavu vybraných lesních oblastí z okresu Frýdek-Místek s největším zastoupením smrku. Tato data byla získána v průběhu 11 let v období od roku 1999 až 2009. Při prvním seznámení s těmito daty vyšlo najevo, že nejsou zcela správná. Bylo tedy nutné je opravit. Hlavním problémem byly záznamy z roku 2009. Především se to týkalo chybějících charakteristik datovaných k tomuto roku. Ty nebyly vyplněny minimálně u jedné čtvrtiny zkoumaných oblastí.

Po kontaktování odborníků, kteří tato data spravovali, nebyl důvod, proč tyto informace nejsou kompletní v plném rozsahu, vysvětlen. Dále se jednalo o sloupce s informacemi týkající se zásoby dřeva, ve kterých bylo potřeba vysvětlit, zda hodnota 0 znamená to stejné jako záznam „VYTĚŽENO“, či nikoliv. Jelikož se ani odborníci nedokázali shodnout a dopátrat se k řešení, byl rok 2009 na návrh výzkumného ústavu CzechGlobe z následující operací vypuštěn. Tento problém se nyní pokoušejí v tomto výzkumném ústavu vyřešit, ale vzhledem k časovému harmonogramu bakalářské práce se v budoucnu aktualizovaná data nepodařilo zanést do výpočtů této bakalářské práce, proto zde byla zpracována data pouze z období let 1999 – 2008.

Z tohoto důvodu musela být data upravena do požadovaného formátu. To znamenalo smazat záznamy z roku 2009, přejmenovat sloupce s charakteristikami a odstranit veškeré formátování. Dále byla přezkoumána struktura dat. V této struktuře došlo především k rozdělení dat do pěti revírů, které se v oblasti nacházejí a jejichž číslo bylo určeno z první cifry identifikačního kódu.

Pro další práci s daty bylo nutné detailně se zaměřit na typ dat jednotlivých charakteristik. V souboru byly nalezeny všechny typy dat: nominální, ordinální, poměrová i intervalová. Vzhledem k tomu muselo být později toto zjištění bráno v úvahu v každé následující operaci s tímto datovým souborem. Jednou z možností bylo převést data do jednotného typu, jelikož ale nebylo možné zvolit jednotný typ, byla tato možnost zamítnuta a dále tak bylo pracováno s daty v původní podobě.

5.2 Základní statistická analýza

Před samotným tvořením modelu poškození porostu byla provedena základní statistická analýza nad daty. Tato analýza podává informace o souboru z hlediska vybraných skupin dat stejného typu. Dále byl proveden statistický rozbor souboru dle rozdělení do pěti revírů a také dle každého roku desetiletého pozorování. Z dat byla odstraněna pole:

- plocha dřeviny,
- věk 51–80,
- věk ≥ 81 ,

- lesní typ,
- SLT,
- hospodářský soubor,
- zásoby daného roku.

Tato pole nebyla podrobena základní statistické analýze, protože byla zastoupena jinými poli vyjadřující stejnou charakteristiku nebo se jednalo o veličiny určující strukturu oblasti z hlediska zařazení do typických vlastností výskytu oblasti. Následně byly proměnné rozděleny na kategoriální a spojité. Toto rozdělení bylo zvoleno s ohledem na další zpracování dat.

5.2.1 Kategoriální proměnné

Hodnotami kategoriálních proměnných jsou celá čísla také označovaná jako kategorie. Tyto proměnné byly v rámci základní statistické analýzy hodnoceny pomocí histogramu. Histogram je graf rozdělení četností sledovaného jevu. Na ose x byly zobrazeny příslušné kategorie a na ose y byly vyneseny četnosti těchto kategorií. K tvorbě histogramů byla použita data vždy jen za jeden rok, jelikož se v průběhu deseti let tyto kategorie nemění. Histogramy byly vytvořeny v programu Microsoft Excel 2007 a k vizualizaci byl použit sloupcový graf (Obr. 2).



Obr. 2 Ukázka histogramu pro kategoriální proměnnou „počáteční poškození“.

Tento graf podává uživateli informaci o kategoriích, které jsou zastoupeny v datech nejvíce a naopak nejméně. V případě proměnné počátečního poškození je nejčetnější hodnotou vyskytující se v datech hodnota jedna. Do této kategorie spadá přibližně 500 oblastí, jejichž počáteční poškození není nulové, ale dá se charakterizovat jako minimální. Podobných histogramů bylo vytvořeno devět právě podle počtu kategoriálních proměnných. Těmito proměnnými jsou zakmenění, ozáření, lesní vegetační stupeň, trofická a hydrická úroveň, vertikální struktura, počáteční poškození, imise a zdravotní stav.

5.2.2 Spojité proměnné

Hodnotami spojitých proměnných mohou být jakékoli hodnoty z určitého intervalu. S ohledem na tento typ byly k statistickému popisu souboru zvoleny charakteristiky polohy, variability, šikmosti a špičatosti.

Charakteristiky polohy

Podle R. Brázdila a kol. (1995) se jedná o tzv. střední hodnoty, které jednoduše reprezentují jednotlivé hodnoty uvažovaného statistického znaku. Udávají polohu rozdělení četností a charakterizují obecnou velikost zkoumaného jevu v daném souboru. Jsou často užívány, protože lze jejich prostřednictvím snáze porovnávat dva nebo více souborů. Z těchto charakteristik byl vybrán aritmetický průměr a medián.

Aritmetický průměr je definován jako úhrn hodnot kvantitativního statistického znaku, dělený rozsahem souboru (Brázdil a kol., 1995).

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

Medián \tilde{x} (prostřední hodnota) je takový prvek řady, uspořádané v neklesajícím pořadí, který ji rozděluje tak, že polovina prvků této řady má menší hodnotu znaku a polovina větší, než je hodnota mediánu. Jelikož není tato hodnota ovlivněna extrémně nízkými nebo vysokými hodnotami, v některých případech lépe popisuje úroveň než průměr (Brázdil a kol., 1995).

Charakteristiky variability

Důležitým doplňkem informací středních hodnot jsou charakteristiky variability. Jsou to čísla, která charakterizují stupeň proměnlivosti statistického znaku v daném souboru. Poskytují tím také představu o rozmístění variant znaku na ose úseček (Brázdil a kol., 1995). Z těchto charakteristik byl vybrán rozsah (variační rozpětí), rozptyl a směrodatná odchylka.

Rozsah (variační rozpětí) R je dle R. Brázdila a kol. (1995) rozdílem nejvyšší a nejnižší hodnoty sledovaného znaku.

$$R = x_{max} - x_{min} \quad (2)$$

Rozptyl s^2 z n hodnot znaku x_i je průměr ze čtverců odchylek jednotlivých hodnot znaku od jejich aritmetického průměru \bar{x} (Brázdil a kol., 1995).

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (3)$$

V praxi se však více využívá druhé odmocniny rozptylu – směrodatné odchylky, která je mírou proměnlivosti čili rozptylu hodnot x_i náhodné veličiny kolem průměru (Brázdil a kol., 1995).

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (4)$$

Charakteristiky šikmosti a špičatosti

Míry šikmosti jsou čísla, která charakterizují nesouměrnost rozdělení četností. Poskytují představu o tvaru rozdělení četností co do nesouměrnosti. Míra šikmosti souměrného rozdělení je nulová, v ostatních případech je kladná nebo záporná. Nejdokonalější z těchto měř je koeficient asymetrie (Brázdil a kol., 1995).

Koeficient asymetrie α podle R. Brázdila a kol. (1995) je aritmetický průměr z třetích mocnin odchylek jednotlivých hodnot znaku od aritmetického průměru vyjádřených v jednotkách směrodatné odchylky.

$$\alpha = \frac{\sum_{i=1}^k n_i \cdot (x_s - \bar{x})^3}{n s^3} \quad (5)$$

Rozdělení četností: $\alpha = 0$	souměrné
$\alpha > 0$	zešikmeno doleva (kladná špičatost)
$\alpha < 0$	zešikmeno doprava (záporná špičatost)

Míry špičatosti jsou čísla popisující koncentraci prvků souboru v blízkosti určité hodnoty znaku. Poskytují představu o tvaru rozdělení četností co do špičatosti nebo plochosti. S vyšší koncentrací prvků kolem dané hodnoty nabudou vyšší číselné hodnoty než v případě plochého tvaru rozdělení četností. Nejdokonalější z těchto měř je koeficient špičatosti (Brázdil a kol., 1995).

Koeficient špičatosti ε je podle R. Brázdila a kol. (1995) průměrná hodnota součtu čtvrtých odmocnin odchylek x_i od průměru \bar{x} měřených v jednotkách směrodatné odchylky s .

$$\varepsilon = \frac{\sum_{i=1}^k (x_s - \bar{x})^4 n_i}{n s^4} - 3 \quad (6)$$

Rozdělení četností: $\varepsilon = 0$	normálně zašpičatělé
$\varepsilon > 0$	špičaté (kladně zašpičatělé)
$\varepsilon < 0$	ploché (záporně zašpičatělé)

Tyto charakteristiky byly na daném souboru dat provedeny v programu RStudio pomocí příkazu *Describe*. Následně byly vytvořeny tabulky v programu Microsoft Excel 2007. V první řadě byly statistické ukazatele spojitých proměnných dopočítány pro pět revírů dané oblasti a následně pro jednotlivé roky. Nejvíce informací nám tato analýza poskytla právě u proměnné procentuálního poškození porostu.

Tab. 1 Základní statistická analýza procentuálního poškození porostu pro pět revírů

Revír	průměr	směr. odchylka	medián	min.	max.	rozsah	šikmost	špičatost
%poškození 1	0,43	0,14	0,41	0,14	1	0,86	0,64	1,76
%poškození 2	0,36	0,13	0,34	0,13	1	0,87	1,23	3,02
%poškození 3	0	0	0	0	0	0	NA	NA
%poškození 4	0	0	0	0	0	0	NA	NA
%poškození 5	0	0	0	0	0	0	NA	NA

Z tabulky je jasně patrné, že v případě revírů tři, čtyři a pět nebylo procentuální poškození porostu vůbec expertně vyhodnoceno. Porovnání je tak možné pouze u prvních dvou revírů. Revír jedna je podle této analýzy průměrně poškozen více než revír dva. Také je zde patrný rozdíl v šikmosti a špičatosti. Oba revíry mají rozdělení četností zešikmené doleva, ale první revír o polovinu méně než druhý. U špičatosti je to podobné, oba mají špičaté rozdělení četností, ale druhý o téměř polovinu více než první.

Tab. 2 Základní statistická analýza procentuálního poškození porostu pro jednotlivé roky

Rok	průměr	směr. odchylka	medián	min.	max.	rozsah	šikmost	špičatost
%poskození 99	0,14	0,20	0	0	1	1	1,18	0,35
%poskození 00	0,14	0,20	0	0	1	1	1,18	0,35
%poskození 01	0,14	0,20	0	0	1	1	1,18	0,35
%poskození 02	0,14	0,20	0	0	1	1	1,18	0,35
%poskození 03	0,14	0,20	0	0	1	1	1,18	0,35
%poskození 04	0,14	0,20	0	0	1	1	1,18	0,35
%poskození 05	0,14	0,20	0	0	1	1	1,18	0,35
%poskození 06	0,14	0,20	0	0	1	1	1,18	0,35
%poskození 07	0,14	0,20	0	0	1	1	1,18	0,35
%poskození 08	0,12	0,20	0	0	1	1	1,18	0,35

Z této tabulky číslo dva bylo zjištěno, že se expertně určené procentuální poškození porostu v průběhu let nijak neměnilo. Údaje z prvního roku tzn. 1999, byly po celých deset let pouze opisovány a nebyly nijak aktualizovány. Vzhledem ke změnám ostatních proměnných je velmi nepravděpodobné, že by se toto procentuální poškození porostu v letech nijak neměnilo. I z těchto důvodů byla snaha určit toto poškození nezávisle na expertním ohodnocení této veličiny. Z důvodu chybějících údajů této veličiny pro tři revíry je minimum každého roku hodnota nula. Maximum s hodnotou jedna (100%) znamená ohodnocení poškození porostu jako porost zcela poškozený.

6 PRAVDĚPODOBNOST POŠKOZENÍ POROSTU

Tato kapitola je věnována určení pravděpodobnosti poškození porostu s využitím metod nelineární regrese. Bylo nutné zvolit a sestavit regresní model odpovídající charakteru dané veličiny. Po konzultacích s odborníky z katedry matematické analýzy Univerzity Palackého byla zvolena beta regrese. Úkolem bylo sestavit model pro rok 2008 a pomocí něho predikovat poškození porostu v daných oblastech. Tyto úkony byly prováděny v programu RStudio (Obr. 3).

```
#nacteni dat#
`2008_vybrane` <- read.delim("C:/Users/Romanka/Desktop/bk/data/model/2008_vybrane.txt")
data2008 <- `2008_vybrane`
view(data2008)

#model2008_12 = FINALNI MODEL(p < 0.05)#
model2008_final <- betareg(data2008$poskozeni ~ data2008$areaporost + data2008$areadreviny
+ data2008$hospsoubor + data2008$zasobaroku + data2008$vek + data2008$imise)
summary(model2008_final)

#zjisteni vzorce k dopoctu %poskozeni, kde poskozeni={0;1}#
predict2008_final=predict(model2008_final)
write.table(predict2008_final, "predict2008_final.txt", sep = "\t")
class(predict2008_final)
length(predict2008_final)
#spravne hodnoty#
predict2008_final
#stejne jako hodnoty z excelu#
predict(model2008_final,type="link")
#prepocet hodnot z excelu na spravne hodnoty#
#ex = 1,16-0,26*areaporost+0,39*areadreviny+0,00012*hospsoubor-0,00033*zasobaroku+0,0057*vek+0,084*imise#
ex = predict(model2008_final,type="link")
1/(1+exp(-ex))
```

Obr. 3 Náhled na programový kód ve vývojovém prostředí programu RStudio.

6.1 Regresní model

V mnoha případech regresních vztahů nelze vyjádřit regresní závislost přímkou nýbrž křivkou. Nejčastěji je to nelineární závislost – nelineární regrese (Brázdil a kol., 1995). Použita byla beta regrese.

Beta regrese je speciální typ regrese vhodný zejména pro analýzy procent, poměrů a koncentračních indexů. Je vhodná pro datové sady, kde hodnoty zkoumaného jevu náleží do otevřeného intervalu (0; 1). Tento regresní model navrhli Ferrari a Cribari-Neto (2004) a nazvali jej jako beta regresní model. Metoda je založena na předpokladu, že značně nezávislá proměnná má beta rozdělení pravděpodobností. Rozdělení beta má následující hustotu z hlediska proměnné střední hodnoty a přesnosti parametru

$$f(y; \alpha, \beta) = \frac{y^{\alpha-1}(y-1)^{\beta-1}}{B(\alpha, \beta)} \quad (7)$$

pro $y \in (0; 1)$, kde α a β jsou kladné reálné parametry a $B(\alpha, \beta)$ označuje tuto funkci beta.

$$B(\alpha, \beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1} dy \quad (8)$$

Dle Ferrari a Cribari-Neto (2004) byla navržena parametrizace nastavením $\mu = \frac{\alpha}{\alpha+\beta}$ a $\phi = \alpha + \beta$. Hustota y může být zapsána jako:

$$f(y; \mu, \phi) = \frac{y^{\mu\phi-1}(y-1)^{\phi-\mu\phi-1}}{B(\mu\phi, \phi-\mu\phi)} \quad (9)$$

s $0 < \mu < 1$ a $\phi > 0$. Proto platí, že $y \sim B(\mu, \phi)$ a potom $E(y) = \mu$ a $VAR(y) = \frac{\mu(1-\mu)}{1+\phi}$.

Pro stanovené μ se rozptyl y zvyšuje tak, jako přesný parametr ϕ klesá.

Tato beta regrese je vhodná pro model poškození porostu, protože hodnoty z intervalu (0; 1) jsou obvykle heteroskedastické (více kolísají kolem průměru a méně na horní a dolní hranici intervalu). Jejich distribuce je asymetrická, proto může být aproximace založená na Gaussovské bázi nepřesná. To platí zejména pro malé vzorky. Z těchto důvodů byla beta regrese použita k nalezení a sestavení nejvhodnějšího modelu poškození porostu (Ferrari a Cribati-Neto, 2004).

V programu RStudio byla použita pro práci s touto regresí knihovna *Betareg*. Tato knihovna obsahuje funkce, pomocí nichž bylo sestaveno několik modelů pro rok 2008. Ty byly hodnoceny dle Waldových statistik, kritéria AIC a Likelihood. Dle těchto ukazatelů byl vybrán jeden optimální model pro rok 2008, do něhož byla následně dosazena data z ostatních let.

Důležitou částí bylo zjištění parametrů pro regresní model. Nejdříve byl sestaven saturovaný model, což znamená, že do něj byly dosazeny veškeré proměnné. Dle již zmíněných kritérií byly tyto proměnné následně z modelu odebírány. Konkrétně byla u každé proměnné řešena hodnota P (Obr. 4, poslední sloupec), pokud tato hodnota splňovala podmínku, že $P < 0,05$, pak byla proměnná v modelu ponechána. V opačném případě došlo k odebrání proměnné z modelu a sestavení dalšího modelu bez této proměnné. Tyto změny pozitivně ovlivňovaly ukazatele kritérií AIC a Likelihood. Tímto postupem byl nakonec určen jeden výsledný model pro rok 2008, který obsahoval proměnné, jež dle těchto kritérií ovlivňují poškození porostu. Parametry tohoto modelu byla čísla ze sloupce Estimate (Obr. 4).

Coefficients (mean model with logit link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.162e+00	1.732e-01	-6.712	1.92e-11 ***
data2008\$areaporost	-2.631e-01	4.568e-02	-5.761	8.37e-09 ***
data2008\$areadreviny	3.936e-01	6.103e-02	6.450	1.12e-10 ***
data2008\$hospsoubor	1.195e-04	5.189e-05	2.303	0.021261 *
data2008\$zasobaroku	-3.341e-04	6.982e-05	-4.785	1.71e-06 ***
data2008\$vek	5.717e-03	1.530e-03	3.737	0.000186 ***
data2008\$imise	8.445e-02	4.134e-02	2.043	0.041059 *

Obr. 3 Tabulka výsledného modelu v programu RStudio.

Tato čísla byla do tohoto modelu dosazena a vznikl tak model finální. Finální model pro rok 2008 s již určenými parametry proměnných vypadá následovně:

Poškození porostu

$$= 1,16 - 0,26 \cdot \text{plocha porostní skupiny} + 0,36 \cdot \text{plocha dřeviny} \\ + 0,00012 \cdot \text{hospodářský soubor} - 0,00033 \cdot \text{zásoba roku} + 0,0057 \\ \cdot \text{věk} + 0,084 \cdot \text{imise}$$

6.2 Predikce a validace

Po zhotovení modelu pro rok 2008 do něj byla dosazena data i z předešlých devíti let. Tento krok byl nejdříve řešen v programu Microsoft Excel 2007, ve kterém však vyšly tyto hodnoty záporně a zcela špatně. Ke správnému postupu bylo nutné seznámit se s terminologií a využít tyto znalosti k úpravě vzorce pro predikci pravděpodobnosti poškození porostu.

Problémem však nebyl zmíněný program, ale funkce, která byla použita k predikování hodnot z výsledného modelu. V programu RStudio s použitím funkce *predict* byly výsledky správné, a tak bylo zjištěno, že k tomuto účelu byla použita linková funkce. Bylo tedy nutné vytvořit vzorec pro výpočet pravděpodobnosti s ohledem na použitou funkci. Pravděpodobnost je šance, že sledovaný jev A nastane k šanci, že nenastane. Hodnoty pravděpodobnosti náležejí otevřenému intervalu (0; 1).

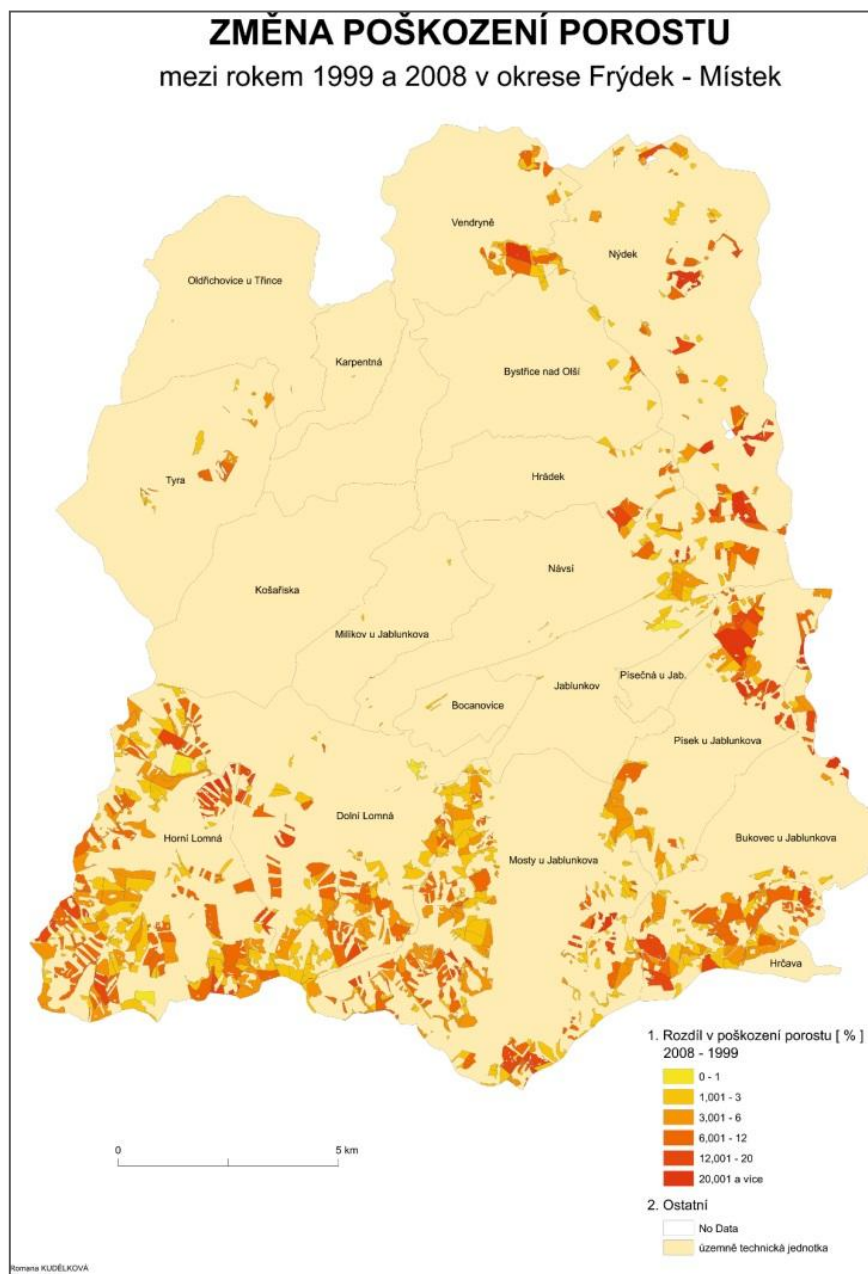
$$P(A) = \frac{P(A)}{1-P(A)} \quad (10)$$

Znamenalo to tedy vzorec pro výpočet pravděpodobnosti zlogaritmovat. Je-li pravděpodobnost jako π a finální model pro rok 2008 jako η , pak platí:

$$\begin{aligned} \text{logit}(\pi) &= \log\left(\frac{\pi}{1-\pi}\right) \\ \log\frac{\pi}{1-\pi} &= \eta \\ \frac{\pi}{1-\pi} &= e^\eta \\ \pi &= e^\eta - \pi e^\eta \\ \pi(1 + e^\eta) &= e^\eta \\ \pi &= \frac{1}{1+e^{-\eta}} \end{aligned} \quad (16)$$

Je-li do vzorce (č. 16) za η dosazen finální model, výsledky jsou v programu Microsoft Excel 2007 stejné jako v programu RStudio určené pomocí funkce *predict*. Výsledky pravděpodobnosti poškození porostu byly přidány k použitým datům, a tak bylo možné s nimi dále pracovat a porovnávat s hodnotami expertně určeného procentuálního poškození porostu.

V místech, kde nebyla tato charakteristika expertně určena, došlo k její predikci a u již ohodnocených oblastí došlo k validaci těchto výsledků. Z modelových hodnot bylo vytvořeno deset rozdílových map, které zobrazují rozdíly procentuálního poškození porostu mezi rokem 2008 a zbylými devíti lety (Obr. 5). Tyto mapy byly vytvořeny v programu ArcGIS 10.1 a jsou k dispozici pouze v elektronické podobě. Z jedné z nich byl vytvořen poster, který obsahuje nejen tuto mapu, ale také výsledky shlukové analýzy. Tento poster je v elektronické i papírové podobě.



Obr. 5 Rozdílová mapa pro roky 1999 a 2008.

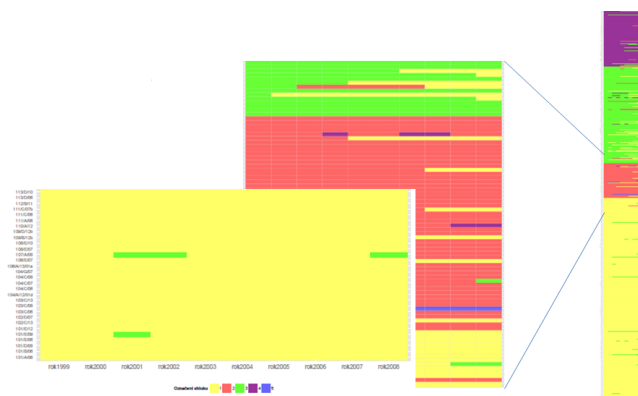
6.3 Shluková analýza

Jak uvádí Kučera (2008), shluková analýza se zabývá algoritmy a metodami, pomocí nichž sdružuje data s podobnými vlastnostmi do shluku. Snaží se uspořádat data do smysluplných struktur. Je to nástroj datové analýzy, jenž třídí různé objekty do shluků tak, že podobnost dvou objektů náležících do jedné skupiny je maximální, zatímco podobnost s objekty mimo tento shluk je naopak minimální. Algoritmy shlukové analýzy je možné rozdělit na hierarchické a nehierarchické.

Tato analýza byla provedena na statisticky významné faktory finálního modelu podle let 1999 – 2008. Těchto faktorů bylo šest a jsou to již uvedené proměnné modelu. K tomuto účelu bylo vybráno shlukování nehierarchické, které se upravuje tak, aby se optimalizovala vzájemná vzdálenost a odlišnost shluků a zároveň aby objekty v nich byly rovnoměrně rozloženy (Kučera, 2008).

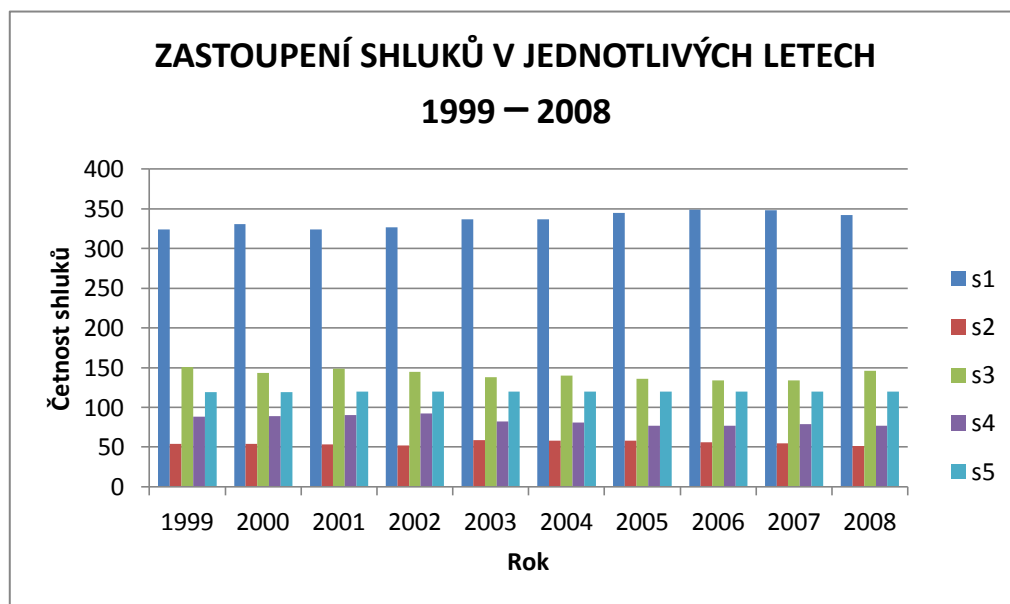
V první řadě bylo nutné stanovit počet shluků. Shlukování bylo provedeno do pěti, sedmi a deseti shluků. Všechny operace probíhaly v programu RStudio. Nejdříve pomocí funkce *Cluster.Sim* byla vygenerována html stránka, na které byly zobrazeny všechny možné metody nehierarchického shlukování s ohodnocením vhodnosti použití na shlukovaných datech. Z této stránky byla vybrána metoda shlukování i metoda, na jejímž základě byly vytvořeny matice vzdáleností. Metodou pro vytvoření této matice byla metoda GDM1 (General Distance Measure 1), která je vhodná pro smíšená data. Po vytvoření těchto matic došlo k samotnému shlukování metodou PAM (Partition Around Medoid). Jak uvádí Kučera (2008) je základem metody PAM vytvoření jednoho reprezentativního objektu, medoidu pro každý shluk. Medoidy byly vybrány tak, aby měl každý objekt k medoidu uvnitř svého shluku minimální vzdálenost.

Výsledky shlukování byly převedeny do tabelární podoby. Vizualizace této analýzy proběhla pomocí funkce *ggplot*. Byla vytvořena heatmapa (teplotní mapa), která graficky znázorňuje shluky a oblasti, které podle vybraných proměnných do těchto shluků v jednotlivých deseti letech náleží. Heatmapa pro pět shluků (Obr. 6) byla přidána i na poster rozdílové mapy. Lze díky ní snadno porovnat, které oblasti jsou si podobné z hlediska faktorů ovlivňující poškození porostu.



Obr. 6 Náhled na heatmapu pro pět shluků.

Dále byly v programu Microsoft Excel 2007 vypočítány četnosti jednotlivých kategorií v letech 1999 – 2008. Výsledky těchto výpočtů byly převedeny do grafu četností těchto shluků. Pro lepší názornost byl vytvořen jeden graf, který obsahuje informace o četnosti shluků celého zkoumaného období (Obr. 7).



Obr. 7 Graf četností pěti shluků v jednotlivých letech.

Z grafu je patrné, že nejvíce oblastí spadá do shluku jedna. Je také zřejmé, že počet oblastí v daných shlucích se nijak výrazně během let nemění.

7 VÝSLEDKY

Prvním dílčím výsledkem bakalářské práce byla interpretace dat. Vzhledem k tomu, že tato data byla pořízena v průběhu deseti let, bylo obtížné dohledat tvůrce dat pro jednotlivé roky. To se nakonec podařilo, a tak mohla být data dále zpracovávána.

7.1 Výsledky základní statistické analýzy

Základní statistická analýza byla provedena dle rozdělení podle typu dat. Pro data kategoriální byly vytvořeny histogramy, z nichž bylo určeno nejčastější kategoriální zařazení zkoumaných oblastí vzhledem ke zvolené charakteristice. Jednou z podstatných charakteristik pro tuto bakalářskou práci bylo „počáteční poškození porostu“. Z grafu je patrné, že nejčastější kategorií je kategorie jedna, což znamená, že oblast nebyla na počátku šetření poškozena nijak výrazně (Obr. 8).



Obr. 8 Histogram počátečního poškození porostu.

Spojité proměnné byly vyhodnoceny charakteristikami polohy, variability, šikmosti a špičatosti. Nejdůležitější z těchto proměnných bylo „procentuální poškození porostu“, protože právě tato charakteristika byla následně predikována a validována. Proto jsou zde uvedeny výsledky právě této proměnné. Nejprve byly vyhodnoceny z hlediska pěti revírů (Tab. 3) a následně ze všech deseti let (Tab. 4).

Ze statistického hodnocení pěti revírů bylo zjištěno, že ve třech revírech nebylo procentuální poškození určeno vůbec. U těchto tří revírů došlo tedy k predikci této charakteristiky, u zbylých dvou k validaci. Ze statistického hodnocení charakteristiky procentuálního poškození porostu z hlediska jednotlivých let se potvrdila domněnka, že hodnoty tohoto poškození se vůbec neměnily. Existují tedy dvě vysvětlení, buď byly tyto hodnoty pouze opisovány z předchozích let, nebo hodnotitel na místě neshledal žádný rozdíl oproti předchozímu roku.

Tab. 3 Základní statistická analýza procentuálního poškození porostu pro pět revírů

Revír	průměr	směr. odchylka	medián	min.	max.	rozsah	šikmost	špičatost
%poškození 1	0,43	0,14	0,41	0,14	1	0,86	0,64	1,76
%poškození 2	0,36	0,13	0,34	0,13	1	0,87	1,23	3,02
%poškození 3	0	0	0	0	0	0	NA	NA
%poškození 4	0	0	0	0	0	0	NA	NA
%poškození 5	0	0	0	0	0	0	NA	NA

Tab. 4 Základní statistická analýza procentuálního poškození porostu pro jednotlivé roky

Rok	průměr	směr. odchylka	medián	min.	max.	rozsah	šikmost	špičatost
%poskození 99	0,13	0,20	0	0	1	1	1,18	0,35
%poskození 00	0,13	0,20	0	0	1	1	1,18	0,35
%poskození 01	0,13	0,20	0	0	1	1	1,18	0,35
%poskození 02	0,13	0,20	0	0	1	1	1,18	0,35
%poskození 03	0,13	0,20	0	0	1	1	1,18	0,35
%poskození 04	0,13	0,20	0	0	1	1	1,18	0,35
%poskození 05	0,13	0,20	0	0	1	1	1,18	0,35
%poskození 06	0,13	0,20	0	0	1	1	1,18	0,35
%poskození 07	0,13	0,20	0	0	1	1	1,18	0,35
%poskození 08	0,13	0,20	0	0	1	1	1,18	0,35

7.2 Sestavení regresního modelu

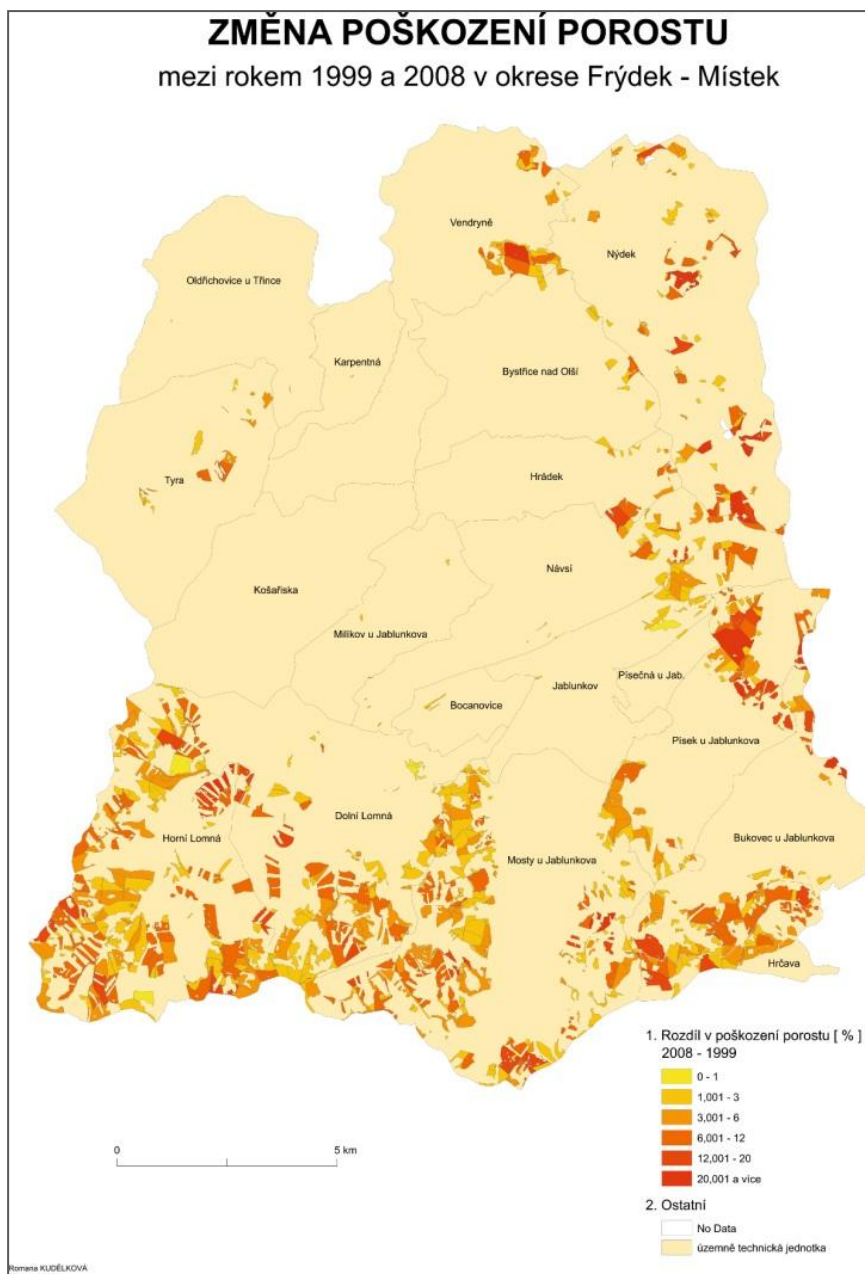
K sestavení tohoto modelu byla zvolena beta regrese. Na základě určených kritérií byly zvoleny faktory ovlivňující procentuální poškození porostu. Dále byly určeny parametry tohoto modelu a vytvořen model výsledný. Tento model byl sestaven z dat pro rok 2008 a následně do něj byly dosazeny všechny ostatní roky.

Poskození porostu

$$\begin{aligned}
 &= 1,16 - 0,26 \cdot \text{plocha porostní skupiny} + 0,36 \\
 &\cdot \text{plocha dřeviny} + 0,00012 \cdot \text{hospodarsky soubor} - 0,00033 \\
 &\cdot \text{zasoba roku} + 0,0057 \cdot \text{vek} + 0,084 \cdot \text{imise}
 \end{aligned}$$

7.3 Predikce a validace

K predikování výsledných hodnot byla použita linková funkce. Došlo k přepočtu podle sestaveného vzorce. Formou vizualizace výsledných predikovaných a validovaných hodnot byly rozdílové mapy. Těchto map bylo zhotoveno deset, jsou k dispozici pouze v elektronické podobě (Obr. 9) a vystihují změnu poškození porostu mezi roky 1999–2008. Dle předpokladů se tyto rozdíly zvětšují společně s rozstupem jednotlivých let. Z jedné mapy byl vytvořen poster, který je v analogové podobě přiložen jako příloha.



Obr. 9 Rozdílová mapa pro rok 1999 a 2008.

7.4 Statistické vyhodnocení výsledků

K tomuto vyhodnocení byla použita shluková analýza, pomocí které byly vytvořeny shluky podobných oblastí podle významných faktorů modelu. Tato shluková analýza byla provedena pro pět, sedm a deset shluků. Použita byla však pouze ta první. Toto zařazení do shluků bylo srovnáno podle jednotlivých let. K vizualizaci byla použita heatmapa, díky níž lze snadno porovnat, zda oblasti během jednotlivých let mění své zařazení do shluků. Tato heatmapa je součástí analogového posteru v přílohách. Dále bylo prostřednictvím grafu četností jednotlivých shluků v období deseti let vyhodnoceno, že nejvíce oblastí spadá do shluku jedna a počet objektů se v jednotlivých shlucích nijak výrazně nemění.

8 DISKUZE

Hlavním problémem celé této práce byla už od samého začátku data. Tato data obsahovala spoustu informací, nikdo však nebyl schopen vysvětlit význam všech charakteristik datového souboru. To je ovšem jen část problému. Druhá část spočívá vtom, že u mnohých charakteristik, jejichž význam byl známý, se vyskytovaly hodnoty, které podle standardů určení vůbec neexistují. Bylo tedy jasné, že data někdo upravoval, nepodařil se ale najít konkrétní člověk, který by vše potřebné vysvětlil.

Dále bylo mnohokrát s odborníky diskutováno, zda by nebylo vhodné tvořit model zároveň také s meteorologickými daty. To však nebylo předmětem zájmu práce, poněvadž výzkumný ústav CzechGlobe projevil zájem o sestavení model z jim dostupných dat. I kdyby tu však tento požadavek nebyl, bylo by velmi složité standardizovat data do podoby potřebné k tomuto účelu. Především vzhledem k časovému rozpětí původních dat a také k relativně malým oblastem, jež byly předmětem zájmu. Dále je nutné říci, že i když by bylo jistě velmi přínosné a užitečné mít k daným oblastem nějaká více rozšířená a komplexní data, jsou to v dané chvíli ta nejkomplexnější data, která jsou k dispozici. V dalších letech budou zcela určitě tato data rozšiřována a vzniknou tak další měřitelné charakteristiky, jež nepochybně budou hrát jednu z hlavních rolí i v poškození lesního porostu.

Co se týče použité regrese, je zde určitě ještě stále mnoho nevyužitých možností, se kterými by se dalo pracovat. Na toto téma by tak mohly vzniknout další práce, v nichž by byl způsob vytvoření regresního modelu proveden zcela odlišně. Beta regrese však byla zvolena s ohledem na poskytnutá data a jevila se tak v danou chvíli jako nejlepší řešení.

Jistou alternativou by bylo převést data na kategorické proměnné a získat tak možnost využít dalších regresních modelů.

9 ZÁVĚR

Úkolem práce bylo zpracovat a vyhodnotit data o desetiletém pozorování smrkových porostů v okrese Frýdek-Místek. Vytvořit z těchto dat pomocí matematicko-statistických metod model pro rok 2008, do něhož by byla dosazena data z let předchozích. Výsledkem měla být predikce a validace expertního ohodnocení této charakteristiky.

Práce je tvořena dvěma částmi. První z nich je část teoretická, která se zabývá metodikou určení poškození porostu především pozemním šetřením. Dále jsou zde zmíněny možnosti využití metod nelineární a logistické regrese v lesnictví.

Druhou část tvoří praktické řešení, jehož základ spočíval ve zpracování a úpravě dat k dalšímu použití. Tato část byla časově velmi náročná z důvodů již několikrát zmíněných problému s interpretací části těchto dat.

Dalším krokem k úspěšnému splnění stanovených cílů bylo statistické vyhodnocení upravených dat. Tato data byla podrobena základní statistické analýze, při níž musel být brán ohled na jejich typ a podobu. Proto byla rozdělena na dvě části, z nichž každá byla vyhodnocena jinými statistickými metodami. Těmito metodami byly: histogram, charakteristiky polohy, variability, šikmosti a špičatosti. Důraz byl kladen především na proměnné týkající se poškození porostu.

Postup pro sestavení regresního modelu se skládal z několika jednotlivých kroků. Nejdříve byla zvolena odpovídající regrese. Touto regresí byla beta regrese. Následně byly dle Waldových statistik, Likelihood a AIC kritéria vybrány statisticky významné faktory ovlivňující poškození porostu. Těmito faktory jsou: plocha porostní skupiny, plocha dřeviny, hospodářský soubor, zásoba roku, věk a imise. K těmto faktorům byly odvozeny parametry a poté byl sestaven výsledný model pro rok 2008.

Následovala predikce a validace procentuálního poškození porostu. Do upraveného modelu pro rok 2008 byla dosazena data z ostatních let, čímž byly získány hodnoty procentuálního poškození v místech, kde expertní ohodnocení této charakteristiky chybělo a v místech, kde již určeno bylo, byly získány hodnoty nové, které bylo možné porovnat s těmi již určenými.

Vyhodnocení těchto výsledků bylo provedeno formou rozdílových map, při jejichž porovnání lze snadno zjistit, že rozdíl hodnot vždy jednoho roku s rokem 2008 postupně narůstá se zvyšujícím se časovým rozestupem mezi zkoumanými roky. V tom případě byla potvrzena domněnka, že pravděpodobnost poškození porostu se mění a není stále stejná, jako tomu bylo u expertního ohodnocení této charakteristiky.

Pro statistické zhodnocení výsledků byla provedena shluková analýza, pomocí které bylo vytvořeno pět shluků oblastí podle podobnosti vzhledem k statisticky významným faktorům modelu. Z vytvořené heatmapy je patrné, jak jsou si oblasti podobné v jednotlivých letech. Nejvíce zastoupen je shluk jedna a počet objektů ve shlucích se v deseti letech nijak výrazně nemění.

Výsledky této práce budou dále využity výzkumným ústavem CzechGlobe k dalšímu řešení problémů s hodnocením zdravotního stavu lesa.

POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

Knižní zdroje

BRÁZDIL, Rudolf et al. *Statistické metody v geografii: cvičení*. 3. vyd. Brno: Masarykova univerzita, 1995, 177 s. ISBN 80-210-1260-9.

CRUZ, Miguel G, Martin E ALEXANDER a Ronald H WAKIMOTO. Development and testing of models for predicting crown fire rate of spread in conifer forest stands. *Canadian Journal of Forest Research*. 2005, vol. 35, issue 7, s. 1626-1639. DOI: 10.1139/X05-085. Dostupné z: <http://www.nrcresearchpress.com/doi/abs/10.1139/x05-085>

GILLNER, Sten, Nadja RÜGER, Andreas ROLOFF a Uta BERGER. Low relative growth rates predict future mortality of common beech (*Fagus sylvatica* L.). *Forest Ecology and Management*. 2013, vol. 302, s. 372-378. DOI: 10.1016/j.foreco.2013.03.032. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0378112713001795>

KAMIMURA, K., S. SAITO, H. KINOSHITA, K. KITAGAWA, T. UCHIDA a H. MIZUNAGA. Analysis of wind damage caused by multiple tropical storm events in Japanese *Cryptomeria japonica* forests. *Forestry*. 2013-08-06, vol. 86, issue 4, s. 411-420. DOI: 10.1093/forestry/cpt011. Dostupné z: <http://forestry.oxfordjournals.org/cgi/doi/10.1093/forestry/cpt011>

LEKWADI, S.O., A. NEMESOVA, T. LYNCH, H. PHILLIPS, A. HUNTER a M. MAC SIÚRTÁIN. Site classification and growth models for Sitka spruce plantations in Ireland. *Forest Ecology and Management*. 2012, vol. 283, s. 56-65. DOI: 10.1016/j.foreco.2012.07.003. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0378112712003994>

SAMEC Pavel, RYCHTECKÁ Petra, HORÁČEK Matěj, TUREK Kamil, VLČKOVÁ Veronika 2012. Geoinformační model rizikových růstových podmínek pro smrkové porosty v Podbeskydském a Beskydském bioregionu (Česká republika). In *Acta Mus. Beskid.*, 4: 1–22. ISSN 1803-960X

Mze, VULHM. *Monitoring stavu lesa v České republice 1984-2003: Forest condition monitoring in the Czech Republic 1984-2003*. Praha: Výzkumný ústav lesního hospodářství a myslivosti, 2004, 431 s. ISBN 80-864-6123-8.

Internetové zdroje:

CRIBARI - NETO, Francisco a Silva L. P. FERRARI. *Beta regression for modelling rates and propotions*. 2004 [cit. 2014-05-19]. Dostupné z:
<http://www.ime.usp.br/~sferrari/beta.pdf>

KUČERA, Jiří. *Shluková analýza* [online]. 2011 [cit. 2014-05-19]. Dostupné z:
[z:http://is.muni.cz/th/172767/fi_b/5739129/web/web/main.html](http://is.muni.cz/th/172767/fi_b/5739129/web/web/main.html)

LESY ČR, s. p. *Lesy ČR* [online]. 2012 [cit. 2014-05-19]. Dostupné z:
www.lesycr.cz/ls111/charakteristika-uzemi/Stranky/default.aspx

SEQUENS, Josef. *Hospodářská úprava lesů: Souhrn*. Praha, 2007[cit. 2014-05-19].
Dostupné z: <http://dl.webcore.czu.cz/file/T0Nqb084d05IR0E9>

VULHM. *Vulhm* [online]. 2009 [cit. 2014-05-19]. Dostupné z:
www.vulhm.cz/?p=icp_forests&site=default

ÚHÚL, *Eagri* [online]. 2009-2013 [cit. 2014-05-19]. Dostupné z: <http://eagri.cz/public/web/uhul/legislativa/vyhlaska-1996-83.html>

SUMMARY

The aim of this thesis is processing of data on decade of observation of forest and evaluation, based on best mathematical and statistical methods, of likelihood of forest damage. The thesis contains theoretical research, which deals with expert evaluation of forest damage and using of logistic and nonlinear regression in forestry. The data originate from district Frýdek-Místek. A problem with modelling of the forest damage is a global problem and it is important solve it, because forests are indispensable part of Earth's surface.

Before the evaluation the data had to be corrected, because some input values were incorrect. Then a basic statistical analysis was performed. Histograms were created for categorical variables and tables were created for continuous variables. These graphs and tables were evaluated. The most important variable was an initial damage of forest and expert forest damage. Results of the evaluation were that the initial damage was low in the most of areas and the expert forest damage was badly determined. Values of the expert forest damage were copied from previous years.

A model of likelihood of forest damage was created by beta regression in RStudio. Significant variables were determined by using this regression and then parameters of the model were calculated. This model was created according the data of year 2008. The data of other years were substituted to the model. Values of likelihood of forest damage were calculated and tabulated. The likelihood of forest damage was also determined for areas where the characteristic has not been determined. Then a cluster analysis was made in RStudio and results of the analysis were shown in heat map. The significant variables were clustered into five clusters. And then it was possible to evaluate the similarity of areas according to the significant variables.

Prediction and validation of expert evaluation of forest damage was done for each other year. Then maps were created by using these results. The ten maps are difference maps that show differences of forest damage between various years. The maps were made in ArcGIS 10.1 and the maps are available in PDFs. A poster was created from one selected map, result of cluster analysis, graph and tables.

This approach of determination of likelihood of forest damage used in this thesis is one of the many ways how the likelihood of forest damage can be determined. Many kinds of regressions can be used but they must respect a form of using data. The beta regression was the best for data used in the thesis.

Results of the thesis were used by Research Institute CzechGlobe for other work with data of conditions of forests in district Frýdek-Místek. Also there is a possibility to address this issue from the perspective of effects of meteorological phenomena. It could be a topic of other thesis.

PŘÍLOHY

SEZNAM PŘÍLOH

Vázané přílohy:

- Příloha 1 Vyhodnocení základní statistické analýzy
- Příloha 2 Zdrojový kód finálního modelu a predikce
- Příloha 3 Zdrojový kód shlukové analýzy
- Příloha 4 Fotografie poškozených lesů

Volné přílohy

- Příloha 5 Poster změny poškození porostu mezi rokem 1999 a 2008
v okrese Frýdek-Místek
- Příloha 6 DVD

Příloha1 Výsledky základní statistické analýzy



Obr. 1 Histogram proměnné „počáteční poškození porostu“.

Tab. 1 Základní statistická analýza procentuálního poškození porostu pro pět revírů

Revír	průměr	směr. Odchylka	medián	min.	max.	rozsah	šíkmost	špičatost
%poškození 1	0,43	0,14	0,41	0,14	1	0,86	0,64	1,76
%poškození 2	0,36	0,13	0,34	0,13	1	0,87	1,23	3,02
%poškození 3	0	0	0	0	0	0	NA	NA
%poškození 4	0	0	0	0	0	0	NA	NA
%poškození 5	0	0	0	0	0	0	NA	NA

Tab. 2 Základní statistická analýza procentuálního poškození porostu pro jednotlivé roky

Rok	průměr	směr. Odchylka	medián	min.	max.	rozsah	šíkmost	špičatost
%poškození 99	0,13	0,20	0	0	1	1	1,18	0,35
%poškození 00	0,13	0,20	0	0	1	1	1,18	0,35
%poškození 01	0,13	0,20	0	0	1	1	1,18	0,35
%poškození 02	0,13	0,20	0	0	1	1	1,18	0,35
%poškození 03	0,13	0,20	0	0	1	1	1,18	0,35
%poškození 04	0,13	0,20	0	0	1	1	1,18	0,35
%poškození 05	0,13	0,20	0	0	1	1	1,18	0,35
%poškození 06	0,13	0,20	0	0	1	1	1,18	0,35
%poškození 07	0,13	0,20	0	0	1	1	1,18	0,35
%poškození 08	0,13	0,20	0	0	1	1	1,18	0,35

Příloha 2 Zdrojový kód finálního modelu a predikce

```
#Načtení dat#
`2008_vybrane` <-
read.delim("C:/Users/Romanka/Desktop/bk/data/model/2008_vybr
ane.txt")
data2008 <- `2008_vybrane`
View(data2008)

#model2008_12 = FINÁLNÍ MODEL (p < 0.05)#
model2008_final <- betareg(data2008$poskozeni ~
data2008$areaporost + data2008$areadreviny +
data2008$hospsoubor + data2008$zasobaroku + data2008$vek +
data2008$imise)
summary(model2008_final)

#Zjištění vzorce k dopočtu %poškození, kde poškození = {0;1}#
predict2008_final=predict(model2008_final)
write.table(predict2008_final, "predict2008_final.txt",
sep = "\t")
class(predict2008_final)
length(predict2008_final)

#Správné hodnoty#
predict2008_final

#Stejně jako hodnoty z excelu (nepřepočítané)#
predict(model2008_final,type="link")

#Přepočet hodnot z excelu na správné hodnoty#
#ex = 1,16-0,26*areaporost+0,39*areadreviny+0,00012*hospsoubor-
0,00033*zasobaroku+0,0057*vek+0,084*imise#
ex = predict(model2008_final,type="link")
1/(1+exp(-ex))
```

Příloha 3 Zdrojový kód shlukové analýzy

```
#nahrani knihoven#
```

```
library(clusterSim)
library(cluster)
library (amap)
library (reshape2)
library (ggplot2)
```

```
#nahrani dat#
```

```
data1999 <-
read.delim("C:/Users/Romanka/Desktop/bk/data/shlukovka/shluk
ova_analyza_kudelkova/data_txt/data1999.txt")
data2000 <-
read.delim("C:/Users/Romanka/Desktop/bk/data/shlukovka/shluk
ova_analyza_kudelkova/data_txt/data2000.txt")
data2001 <-
read.delim("C:/Users/Romanka/Desktop/bk/data/shlukovka/shluk
ova_analyza_kudelkova/data_txt/data2001.txt")
data2002 <-
read.delim("C:/Users/Romanka/Desktop/bk/data/shlukovka/shluk
ova_analyza_kudelkova/data_txt/data2002.txt")
data2003 <-
read.delim("C:/Users/Romanka/Desktop/bk/data/shlukovka/shluk
ova_analyza_kudelkova/data_txt/data2003.txt")
data2004 <-
read.delim("C:/Users/Romanka/Desktop/bk/data/shlukovka/shluk
ova_analyza_kudelkova/data_txt/data2004.txt")
data2005 <-
read.delim("C:/Users/Romanka/Desktop/bk/data/shlukovka/shluk
ova_analyza_kudelkova/data_txt/data2005.txt")
data2006 <-
read.delim("C:/Users/Romanka/Desktop/bk/data/shlukovka/shluk
ova_analyza_kudelkova/data_txt/data2006.txt")
data2007 <-
read.delim("C:/Users/Romanka/Desktop/bk/data/shlukovka/shluk
ova_analyza_kudelkova/data_txt/data2007.txt")
data2008 <-
read.delim("C:/Users/Romanka/Desktop/bk/data/shlukovka/shluk
ova_analyza_kudelkova/data_txt/data2008.TXT")
```

```

#Nadefinování prvního sloupečku = id dat#
rownames(data1999) <- data1999[,1]
rownames(data2000) <- data2000[,1]
rownames(data2001) <- data2001[,1]
rownames(data2002) <- data2002[,1]
rownames(data2003) <- data2003[,1]
rownames(data2004) <- data2004[,1]
rownames(data2005) <- data2005[,1]
rownames(data2006) <- data2006[,1]
rownames(data2007) <- data2007[,1]
rownames(data2008) <- data2008[,1]

#Vybrání dat pro shlukovou analýzu#
data1999s <- data1999[,c(4,6,8,9,17,20)]
data2000s <- data2000[,c(4,6,8,9,17,20)]
data2001s <- data2001[,c(4,6,8,9,17,20)]
data2002s <- data2002[,c(4,6,8,9,17,20)]
data2003s <- data2003[,c(4,6,8,9,17,20)]
data2004s <- data2004[,c(4,6,8,9,17,20)]
data2005s <- data2005[,c(4,6,8,9,17,20)]
data2006s <- data2006[,c(4,6,8,9,17,20)]
data2007s <- data2007[,c(4,6,8,9,17,20)]
data2008s <- data2008[,c(4,6,8,9,17,20)]

#Stanovení optimálního typu shlukování - cluster.sim#
#Interval or mixed data without normalization#
data1999s_sim <- cluster.Sim(data1999s, 7, 5, 10, icq = "S",
outputHtml = "data1999s_sim", normalizations = c("n0"),
distances = c("d1", "d2", "d3", "d4", "d5"), methods =
c("m1", "m2", "m3", "m4", "m5", "m6", "m7", "m8"))
data2000s_sim <- cluster.Sim(data2000s, 7, 5, 10, icq = "S",
outputHtml = "data2000s_sim", normalizations = c("n0"),
distances = c("d1", "d2", "d3", "d4", "d5"), methods =
c("m1", "m2", "m3", "m4", "m5", "m6", "m7", "m8"))
data2001s_sim <- cluster.Sim(data2001s, 7, 5, 10, icq = "S",
outputHtml = "data2001s_sim", normalizations = c("n0"),
distances = c("d1", "d2", "d3", "d4", "d5"), methods =
c("m1", "m2", "m3", "m4", "m5", "m6", "m7", "m8"))
data2002s_sim <- cluster.Sim(data2002s, 7, 5, 10, icq = "S",
outputHtml = "data2002s_sim", normalizations = c("n0"),
distances = c("d1", "d2", "d3", "d4", "d5"), methods =
c("m1", "m2", "m3", "m4", "m5", "m6", "m7", "m8"))
data2003s_sim <- cluster.Sim(data2003s, 7, 5, 10, icq = "S",
outputHtml = "data2003s_sim", normalizations = c("n0"),

```



```

distances = c("d1", "d2", "d3", "d4", "d5"), methods =
c("m1", "m2", "m3", "m4", "m5", "m6", "m7", "m8"))
data2004s_sim <- cluster.Sim(data2004s, 7, 5, 10, icq = "S",
outputHtml = "data2004s_sim", normalizations = c("n0"),
distances = c("d1", "d2", "d3", "d4", "d5"), methods =
c("m1", "m2", "m3", "m4", "m5", "m6", "m7", "m8"))
data2005s_sim <- cluster.Sim(data2005s, 7, 5, 10, icq = "S",
outputHtml = "data2005s_sim", normalizations = c("n0"),
distances = c("d1", "d2", "d3", "d4", "d5"), methods =
c("m1", "m2", "m3", "m4", "m5", "m6", "m7", "m8"))
data2006s_sim <- cluster.Sim(data2006s, 7, 5, 10, icq = "S",
outputHtml = "data2006s_sim", normalizations = c("n0"),
distances = c("d1", "d2", "d3", "d4", "d5"), methods =
c("m1", "m2", "m3", "m4", "m5", "m6", "m7", "m8"))
data2007s_sim <- cluster.Sim(data2007s, 7, 5, 10, icq = "S",
outputHtml = "data2007s_sim", normalizations = c("n0"),
distances = c("d1", "d2", "d3", "d4", "d5"), methods =
c("m1", "m2", "m3", "m4", "m5", "m6", "m7", "m8"))
data2008s_sim <- cluster.Sim(data2008s, 7, 5, 10, icq = "S",
outputHtml = "data2008s_sim", normalizations = c("n0"),
distances = c("d1", "d2", "d3", "d4", "d5"), methods =
c("m1", "m2", "m3", "m4", "m5", "m6", "m7", "m8"))
# Matice vzdálenosti pro smíšená data (klasická čísla)#
data1999s_dist <- GDM1(data1999s, weightsType = "equal")
data2000s_dist <- GDM1(data2000s, weightsType = "equal")
data2001s_dist <- GDM1(data2001s, weightsType = "equal")
data2002s_dist <- GDM1(data2002s, weightsType = "equal")
data2003s_dist <- GDM1(data2003s, weightsType = "equal")
data2004s_dist <- GDM1(data2004s, weightsType = "equal")
data2005s_dist <- GDM1(data2005s, weightsType = "equal")
data2006s_dist <- GDM1(data2006s, weightsType = "equal")
data2007s_dist <- GDM1(data2007s, weightsType = "equal")
data2008s_dist <- GDM1(data2008s, weightsType = "equal")
#Shlukování smíšených dat funkcí pam#
#Počet shluků - 5#
data1999s_pam5 <- pam(data1999s_dist, 5, diss = TRUE)
data2000s_pam5 <- pam(data2000s_dist, 5, diss = TRUE)
data2001s_pam5 <- pam(data2001s_dist, 5, diss = TRUE)
data2002s_pam5 <- pam(data2002s_dist, 5, diss = TRUE)
data2003s_pam5 <- pam(data2003s_dist, 5, diss = TRUE)
data2004s_pam5 <- pam(data2004s_dist, 5, diss = TRUE)
data2005s_pam5 <- pam(data2005s_dist, 5, diss = TRUE)
data2006s_pam5 <- pam(data2006s_dist, 5, diss = TRUE)

```

```

data2007s_pam5 <- pam(data2007s_dist, 5, diss = TRUE)
data2008s_pam5 <- pam(data2008s_dist, 5, diss = TRUE)
#Vytvoření dat se skupinami - 5 clusterů#
data_skupiny_5 <-
as.data.frame(cbind(data1999s_pam5$clustering,
data2000s_pam5$clustering, data2001s_pam5$clustering,
data2002s_pam5$clustering, data2003s_pam5$clustering,
data2004s_pam5$clustering, data2005s_pam5$clustering,
data2006s_pam5$clustering, data2007s_pam5$clustering,
data2008s_pam5$clustering), file = "data_skupiny_5.txt",
col.names = c("1999_5cl", "2000_5cl", "2001_5cl",
"2002_5cl", "2003_5cl", "2004_5cl", "2005_5cl", "2006_5cl",
"2007_5cl", "2008_5cl"), row.names = rownames(data1999s))
colnames(data_skupiny_5) <- c("1999_5cl", "2000_5cl",
"2001_5cl", "2002_5cl", "2003_5cl", "2004_5cl", "2005_5cl",
"2006_5cl", "2007_5cl", "2008_5cl")
write.table(data_skupiny_5, "data_skupiny_5.txt",
sep = "\t")
#Heatmapy#
#Sort, melt - 5 clusterů#
data_skupiny_sort5 <- cbind(rownames(data_skupiny_5),
data_skupiny_5)
colnames(data_skupiny_sort5) <- c("JPRL",
"rok1999", "rok2000", "rok2001", "rok2002", "rok2003", "rok2004",
"rok2005", "rok2006", "rok2007", "rok2008")
data_skupiny_sort5$JPRL <- with(data_skupiny_sort5,
reorder(JPRL, rok1999))
data_skupiny_melt5 <- melt(data_skupiny_sort5)
paleta5 <- c( "#FF4040", "#CAFF70", "#008B8B", "#FFFF00 ",
"#FF82AB")
#Stanovení parametrů#
p5 <- ggplot(data_skupiny_melt5, aes(variable, JPRL, fill =
factor(value))) + geom_tile()
pp5 <- p5 + labs(title = "SROVNÁNÍ SHLUKOVÁNÍ PODLE
VÝZNAMNÝCH FAKTORŮ BĚHEM LET 1999-2008") +
scale_fill_manual(name = "Označení shluku", values = paleta5)
ppp5 <- pp5 + theme_grey(base_size = 9) + labs(x = "", y =
"") + theme(legend.position = "bottom", axis.ticks =
element_blank(), axis.text.x = element_text(size = 10))
ppp5
#Export do png 5 cl#
png(file = "heatmap5.png", width = 700, height = 5800)
ppp5
dev.off()

```

```
#Export do pdf 5 cl#
  cairo_pdf(filename = "heatmap_cluster5.pdf", width = 10,
height = 100, pointsize = 10, onefile = FALSE, family =
"sans", bg = "white", antialias = c("default"))
ppp5
dev.off()
```

Příloha 4 Fotografie poškozených lesů



Obr. 2 Poškozený les (autor: Ing. Pavel Samec).



Obr. 3 Fragmentované lesy (autor: Ing. Pavel Samec).



Obr. 4 Poškozený strom (autor: Ing. Pavel Samec).



Obr. 5 Nespecifické žloutnutí smrku (autor: Ing. Pavel Samec).