

Využití metod dálkového průzkumu Země pro monitoring stavu koupacích míst

VÁCLAVA MAŤAŠOVSKÁ, FILIP KOTHAN, ONDŘEJ LEDVINKA, PETR PUMANN,
TOMÁŠ FOJTÍK, MARCELA MAKOVCOVÁ, LENKA BENDAKOVSKÁ

Klíčová slova: dálkový průzkum – monitoring stavu koupacích vod – spektrální indexy – regrese

SOUHRN

Ministerstvo zdravotnictví ve spolupráci s dalšími dotčenými orgány sestavuje každoročně Seznam přírodních koupališť na povrchových vodách, ve kterých nabízí službu koupání provozovatel, a dalších povrchových vod ke koupání. Ten je sestavován na základě dlouhodobého monitoringu prováděného krajskými hygienickými stanicemi, a to výhradně *in situ* na bázi kalendářních let. Vzhledem k tomu, že je tento způsob monitoringu nejen časově, ale i finančně náročný, vzniká tedy přirozeně potřeba tyto nároky minimalizovat.

V současnosti probíhající projekt *Využití metod dálkového průzkumu Země pro monitorování stavu a kvality koupacích míst v České republice* nabízí možnost některé ukazatele potřebné pro tato hodnocení monitorovat distančně. Jeho cílem je za pomoci moderních statistických technik v kombinaci s využitím GIS nástrojů nalézt a popsat vzájemný vztah mezi daty získanými při terénních šetřeních a satelitními daty. Prostřednictvím svých výstupů tak může nabídnout krajským hygienickým stanicím nástroj nejen pro stanovení běžně hodnocených ukazatelů, ale i pro identifikaci možných nových koupacích míst. Díky vytvořeným časovým řadám může zároveň sloužit jako podklad pro hodnocení koupací sezony.

V předloženém příspěvku jsou prezentovány zejména postupy použité pro dosažení stanovených cílů. Na čtrnácti modelových lokalitách proběhly dvě etapy terénních šetření s navazujícími laboratorními pracemi. Zároveň byly hledány optimální způsoby zpracování satelitních dat vysokého prostorového rozlišení (Sentinel-2). Po základním zpracování multispektrálního obrazu bylo otestováno 105 spektrálních indexů, přičemž byl potvrzen počáteční předpoklad silné korelace některých indexů s hodnotami naměřených ukazatelů. Pro predikci hodnot sledovaných ukazatelů byly využity dvě modelovací techniky, a to *Random Forests* a *Partial Least Squares regression*, kterým byla satelitní data předkládána jako vysvětlující proměnné.

ÚVOD

Koupací vody

Za koupací vody lze považovat veškeré vody, které jsou využívány ke koupání větším počtem obyvatel. Lze je podrobněji rozčlenit na přírodní a umělá koupaliště. Pro potřeby řešeného projektu jsou v úvahu brána pouze koupaliště přírodní. Těmi se podle § 6 odst. 1 zákona č. 258/2000 Sb., o ochraně veřejného zdraví, rozumí *stavba povolená k účelu koupání nebo nádrží ke koupání, v nichž je voda ke koupání obměňována řízeným přítokem a odtokem pitné vody nebo trvalým*

přítokem a odtokem chemicky neupravené podzemní nebo povrchové vody, nebo stavba povolená k účelu koupání vybavená systémem přírodního způsobu čištění vody ke koupání, nebo povrchová voda, ve které nabízí službu koupání provozovatel.

Problematika koupacích vod je legislativně ukotvena zejména ve výše uvedeném zákoně, který reflektuje nařízení a požadavky směrnice Evropského parlamentu a Rady 2006/7/ES ze dne 15. února 2006, o řízení jakosti vod ke koupání. Jsou zde mj. upravena práva a povinnosti fyzických a právnických osob v oblasti ochrany a podpory veřejného zdraví. Dále je zde definována soustava orgánů ochrany veřejného zdraví, včetně jejich pravomocí a působnosti [1, 2]. Dalším významným předpisem je vyhláška č. 238/2011 Sb., o stanovení hygienických požadavků na koupaliště, sauny a hygienické limity písku v pískovištích venkovních hracích ploch. Ta stanovuje základní pravidla pro monitoring a posouzení jakosti vody v přírodních koupalištích a její klasifikaci, rovněž určuje rozsah informování veřejnosti o jakosti povrchových vod ke koupání.

Pokud zvláštní prováděcí předpis nestanoví jinak, je při odběru vzorků pro zjišťování hodnot ukazatelů jakosti vody (*tabulka 1*) postupováno podle následujících technických norem:

- ČSN EN ISO 5667-1 (75 7051) Jakost vod – Odběr vzorků, část 1: Návod a návrh programu odběru vzorků a pro způsoby odběru vzorků,
- ČSN ISO 5667-4,6 (75 7051) Jakost vod – Odběr vzorků, část 4: Pokyny pro odběr vzorků z vodních nádrží,
- ČSN ISO 5667-4,6 (75 7051) Jakost vod – Odběr vzorků, část 6: Návod pro odběr vzorků z řek a potoků,
- ČSN EN ISO 19458 (75 7801) Jakost vod – Odběr vzorků pro mikrobiologickou analýzu,
- ČSN 75 7717 Jakost vod – Stanovení planktonních sinic [3, 4].

Četnost odběrů vzorků a jejich rozložení po dobu koupací sezony, včetně míst odběrů v případě dalších povrchových vod ke koupání, je dána monitorovacím kalendářem. Ten je vydáván pravidelně před zahájením koupací sezony 1. května krajskou hygienickou stanicí (KHS) jako opatření obecné povahy. Koupací sezona je pak zpravidla vymezena obdobím od 30. května do 1. září, nebo obdobím, během něhož lze očekávat velký počet koupajících se osob [1].

Vyhláška rovněž stanovuje pravidla pro posuzování a postupy při klasifikaci jakosti koupacích vod [3, 4]. Přírodní koupaliště a povrchové vody využívané ke koupání, u nichž je základním předpokladem splnění hygienických limitů, jsou zařazeny do Seznamu přírodních koupališť na povrchových vodách, ve kterých nabízí službu koupání provozovatel, a dalších povrchových vod ke koupání (dále jen seznam). Ten je každoročně sestavován do 31. března a po dobu 10 kalendářních dnů zpřístupněn na úředních deskách Ministerstva zdravotnictví (MZd) a KHS, včetně jejich územních pracovišť, veřejnosti k připomínkám. Po vyhodnocení všech uplatněných připomínek je upravený seznam ještě před

Tabulka 1. Sledované ukazatele pro hodnocení jakosti koupacích vod
Table 1. Monitored indicators for the assessment of bathing waters quality

Mikrobiologické ukazatele	příloha vyhlášky č. 1, sloupec A	intestinální enterokoky [KTJ/100 ml]
		<i>Escherichia coli</i> [KTJ/100 ml]
		průhlednost [m]
		vodní květ [stupeň]
Sledování výskytu sinic	příloha vyhlášky č. 4, tab. č. 1 a 2	sinice – buněčný objem [buňky/ml], [mm ³ /l]
		chlorofyl-a [µg/l]
		vodní květ [stupeň]
		mikroskopický obraz (slovní popis)
Vizuální kontrola	příloha vyhlášky č. 5	znečištění odpady [stupeň]
		přírodní znečištění [stupeň]

zahájením koupací sezony znovu uveřejněn na úředních deskách a na portálu veřejné správy [5]. Zároveň je seznam s uvedením důvodů jeho změn oproti předchozímu roku, pokud k nim došlo, předložen Ministerstvem životního prostředí (MŽP) Evropské komisi. Do 31. prosince kalendářního roku po uplynutí koupací sezony MŽP ve spolupráci s MZd zpracuje a předloží zprávu o výsledcích monitorování a posouzení jakosti povrchových vod uvedených v seznamu s popisem významných opatření, která byla přijata příslušnými orgány k řízení jakosti vody ke koupání [1, 2].

Dálkový průzkum Země

Dálkový průzkum Země (DPZ) je relativně moderní metoda získávání informací o objektech a jevech na zemském povrchu, přičemž musí být splněna základní podmínka, a to že přenos informací je zajištěn pomocí elektromagnetického záření. To je členěno podle vlnové délky do tzv. elektromagnetického spektra, které je spojitě a velmi rozsáhlé, avšak pro potřeby DPZ je využita pouze jeho část (pásma viditelného záření, infračerveného záření a pásma mikrovlonná). Za základní členění metod DPZ by se dalo považovat rozdělení na konvenční a nekonvenční. Zatímco v případě konvenčních metod jsou pořizovány a zpracovávány fotografické snímky zachycené centrální projekcí na fotografický materiál, nekonvenční metody využívají registrace elektromagnetického záření pomocí snímačů. Pro účely tohoto výzkumu byly využívány výhradně produkty metod nekonvenčních, a to snímání zemského povrchu pomocí družicových systémů.

Existuje několik důvodů, proč metody DPZ využít, ať už v přírodovědných, či technických oborech. První nespornou výhodou je aktuálnost získaných informací, která výrazně předčí informace pořízené metodami klasického pozemního šetření. S tím je úzce spjata další vysoce ceněná vlastnost, opakovatelnost pořízení dat. To se týká především družicových systémů, které mají pevně stanovenou dráhu oběhu. Díky tomu je umožněno zachycení časové dynamiky sledovaných jevů. Nelze opomenout ani rozsah zachyceného území v jednom okamžiku (řádově několik desítek až stovek km²). Po provedení geometrických a radiometrických korekcí, jejichž účelem je odstranění zkreslení dat, je zpravidla získán geometricky přesný a standardizovaný obraz. Ten umožňuje nejen zjištění přesné polohy objektu zájmu, ale i stanovení a porovnání získaných hodnot registrovaného elektromagnetického záření [6]. Relativně novou výzvou pro oblast DPZ jsou volně dostupné produkty Evropské kosmické agentury (ESA). Ta v současné době zajišťuje provoz šesti typů družicových systémů (mise

Sentinel-6 zahájena 21. listopadu 2020) se zaměřením na různé složky životního prostředí, díky kterým jsou naplňovány cíle programu Copernicus. Jeho hlavním cílem je poskytovat informační služby založené na DPZ, a umožnit tak sledování nejrůznějších komponent životního prostředí. Pro účely projektu byla využita data družicového systému Sentinel-2. Hlavním úkolem této mise je především monitoring stavu lesů a zemědělství, avšak díky parametrům výsledných dat (prostorové rozlišení, rozsah snímaného elektromagnetického spektra a perioda snímání) jsou současně snímky družic Sentinel-2 vysoce atraktivní i pro jiná odvětví monitoringu životního prostředí [7].

Výzkumný projekt Zéta II TJ02000091 – *Využití metod dálkového průzkumu Země pro monitoring stavu a kvality koupacích míst v České republice* je zaměřen na nalezení optimálních metod pro hodnocení ukazatelů sledovaných na koupacích vodách s využitím satelitních dat. Cílem řešeného projektu je pomocí kombinace těchto dat s daty z pozemního šetření identifikovat rozsah a intenzitu znečištění vodních ploch (především zelenou organickou složkou) využívaných pro koupání (např. chlorofyl-a, množství sinic a řas nebo zákal). Základní motivací pro využití informací získaných metodami DPZ pro řešení projektu je skutečnost, že již v minulosti byla prokázána závislost mezi projevy odrazivosti a množstvím pevných částic ve vodním prostředí. V případě fytoplanktonu (mikroskopické sinice a řasy volně se vznášející ve vodě) byly popsány změny v projevech odrazivosti při využití vlnových délek 0,5–0,75 µm, tedy v červeném a infračerveném pásmu elektromagnetického spektra [8, 9]. Hlavním očekávaným výsledným produktem je nástroj ve formě webové mapové aplikace pro zvýšení efektivity zejména KHS při činnostech souvisejících s monitoringem stavu koupacích vod a hodnocením uplynulé koupací sezony. Přidanou hodnotou by pak měla být možnost získání některých parametrů jakosti koupacích vod bez nutnosti přímého kontaktu s nimi, a to i na lokalitách, které nepodléhají současnému monitoringu, a ucelený přehled o heterogenitě vodních ploch a jejím vývoji v čase.

METODIKA ŘEŠENÍ

Terénní šetření

První etapa terénního šetření proběhla v období měsíců června až října 2019, kdy byl pracovníky Státního zdravotního ústavu (SZÚ) prováděn pravidelný odběr vzorků vody na čtrnácti modelových lokalitách v Praze, Středočeském,

Pardubickém a Královéhradeckém kraji (*tabulka 2*). Při výběru modelových lokalit bylo dbáno na rozmanitost vodních ploch z hlediska jejich velikosti, charakteru a oživení fytoplanktonem – v tomto případě byla brána v úvahu nejen kvantita fytoplanktonu, ale i jeho předpokládané složení. Předběžný časový harmonogram terénních šetření byl vytvořen tak, aby korespondoval s dobou přeletu družice. Letový plán družic Sentinel-2 je dostupný online [10] od listopadu 2015 ve formátu KML.

Pro každou modelovou lokalitu bylo zvoleno jedno hlavní odběrové místo (H). U oficiálně sledovaných vod bylo většinou totožné s monitorovaným místem odběru provozovatele či KHS. Na hlavním odběrovém místě byly z horizontu 0–30 cm (standardní odběr pro koupací vody) Andělovou tyčí odebírány vzorky pro stanovení chlorofylu-a, zákalu, fluorimetrické stanovení fytoplanktonu a mikroskopický rozbor. V souladu s postupem ČSN 75 7717 byl vytvořen směsný vzorek z několika dílčích vzorků (zpravidla z pěti) z okruhu několika metrů, kde hloubka dosahovala minimálně 1 m. Pro vytvoření představy o přítomnosti a složení větších zástupců fytoplanktonu (především sinic) byl odebírán vzorek s použitím planktonní sítě s oky o průměru 20 µm. Odběry vzorků z horizontu 0–10 cm (těsně pod hladinou) byly využity pro fluorimetrické stanovení fytoplanktonu a zákalu. Současně byla stanovena průhlednost vody, přítomnost vodního květu a další přímo pozorovatelné ukazatele kvality koupacích vod a změřena teplota v obou horizontech odběru. Kromě odběrových míst (H) byl vytipován i různý počet vedlejších odběrových míst (O) – většinou 2–4 podle plochy modelové lokality. Na těch byly vzorky odebírány pouze z horizontu těsně pod hladinou. Odběry byly prováděny vstupem do vody, z mola, v některých případech z lodi.

Laboratorní práce

Fluorescenční analýzy, filtrace a extrakce vzorků pro stanovení chlorofylu-a byly vždy prováděny po návratu do laboratoře v den odběru. Ukazatele spjaté se sinicemi a řasami byly mikroskopicky stanoveny metodou podle ČSN 75 7717 a podle ČSN 75 7712. Vzorky pro kvantifikaci byly konzervovány v Lugolově roztoku. Chlorofyl-a byl stanoven pomocí standardní metody ČSN ISO 10260. Fluorescence fytoplanktonu byla měřena fluorometrem AquaPen AP100, který měří odezvu vzorků excitovaných při dvou vlnových délkách (450 a 620 nm). Pro odebrané vzorky byla stanovena hodnota okamžité fluorescence chlorofylu (Ft) po excitaci při obou vlnových délkách a OJIP křivky (450 a 620 nm), ze kterých, kromě množství fytoplanktonu, lze získat informace i o jeho fyziologickém stavu. Pro stanovení zákalu byl využíván turbidimetr HACH 2100P [4].

Zpracování multispektrálního obrazu

Zpracovávaná data pocházela z mise Sentinel-2, která byla zahájena 23. června 2015 vypuštěním družice Sentinel-2A do oběhu. Je vybavena senzorem MultiSpectral Imager, který snímá data ve 13 pásmech ve vysokém prostorovém rozlišení (10–60 m). Za nejdůležitější je považováno pásmo viditelné (B2, B3, B4) a infračervené části (B5, B6, B7, B8 a B8a) elektromagnetického spektra. Ostatní pásma jsou určena především pro radiometrické korekce obrazu. Doba oběhu této družice je 10 dní, ve spolupráci se sesterskou družicí Sentinel-2B (vypuštěna do oběhu 7. března 2017) se však perioda snímání stejné lokality snižuje na polovinu, v případě rovníkových oblastí dokonce až na tři dny. V rámci jedné pořízené scény je zachyceno 290 × 290 km zemského povrchu [11].

Pro základní předzpracování obrazu bylo testováno několik softwarových prostředí. Prvním byl SNAP – ver. 6.0 vyvíjený společností Brockmann Consult, Array Systems Computing and C-S pro účely programu Copernicus.

Tabulka 2. Modelové lokality 2019
Table 2. Model sites 2019

Název lokality	Typ lokality	Výměra vodní plochy [km ²]	Uvedeno v seznamu [5]	Složení fytoplanktonu
Očko	písník	0,11	ne	tenké vláknité sinice, anorganický zákal
Ovčáry	písník	0,17	ne	koncem sezony vláknité i kokální sinice, vodní květ
Proboštská jezera	písník	0,27	ano	obrněnky rodu <i>Ceratium</i>
Opatovický písník	písník	0,31	ano	velmi nízké oživení fytoplanktonem
Mělice	písník	0,43	ano	fekální kontaminace, nízké oživení fytoplanktonem
Mělice (malý písník)	písník		ne	velmi nízké oživení fytoplanktonem
Eliška	nádrž	0,02	ne	střední oživení fytoplanktonem (spíše řasy)
Marvánek	nádrž	0,02	ne	sinicové vodní květy <i>Microcystis</i> , <i>Aphanizomenon</i>
Vyžlovka	nádrž	0,20	ano	koncem sezony vláknité sinice (<i>Planktothrix</i>)
Hostivař	nádrž	0,35	ano	vláknité sinice (<i>Planktothrix</i> a další)
Seč	nádrž	1,62	ano	sinice (vodní květy), fekální kontaminace
Rozkoš (horní)	nádrž	8,95	do 2018	silné vodní květy sinic (<i>Microcystis</i>)
Rozkoš (dolní)	nádrž		ano	střední oživení fytoplanktonem (sinice i řasy)
Labe (Stará Boleslav)	řeka		ne	silné oživení fytoplanktonem (řasy), anorganický zákal

Tento software je distribuován pod licencí GPL (General Public Licence), je tedy volně k dispozici. Jeho nevýhodou však je, že se stále vyvíjí, jeho moduly jsou vyvíjeny samostatně a velmi často se vyskytují problémy s aplikací v různých operačních systémech [12]. Dále byly jednotlivé kroky předzpracování testovány a prováděny v komerčním softwaru Geomatica (PCI). Prvním krokem bylo převzorkování jednotlivých pásem na stejné prostorové rozlišení 10 m. Tento krok byl nezbytný pro další předzpracování obrazu. Následně byly provedeny pro jednotlivé scény atmosférické korekce. Jedná se o nejkomplicovanější formy opravy obrazu. Vlivem rozptylu, pohlcování a také emisivity atmosféry dochází k modifikaci hodnot pixelu. Všeobecně je pro úpravu hodnot pixelů využíváno přírodního modelu a jeho aproximací, s cílem zajistit, aby tyto hodnoty co nejvíce odpovídaly skutečným odrazovým nebo zářivým vlastnostem sledovaného objektu. Zároveň zde byly vygenerovány tzv. masky mraků, tedy pixely oblačnosti, které byly z dalších analýz vyloučeny.

Vzhledem k tomu, že bylo získáno velké množství dat, vznikla potřeba (z důvodů časové a výpočetní náročnosti) velikost získaných dat a čas nutný k jejich zpracování snížit. Prvním krokem bylo tzv. mozaikování dlaždic. Jedná se o proces, při kterém v případě, kdy sledované území je pokryto více snímky, je dosaženo bezešvé a barevně vyrovnané mozaiky (nebudou zde patrné přechody mezi jednotlivými snímky). Kvalita výsledné mozaiky závisí především na geometrické přesnosti snímků, velikosti jejich překryvů a rozdílech ve způsobu jejich pořízení. K mozaikování byl využit modul softwaru Geomatica, Mosaic Tool, ve kterém bylo na základě uživatelských recenzí využito manuální metody [13]. Díky tomuto procesu byl následně snížen počet opakování u dalších kroků digitálního zpracování snímku. Ke snížení celkového objemu dat bylo nutné vybrat z mozaiky pouze vodní plochy. Ty byly získány výběrem prvků z vrstvy vodních ploch ZABAGED*. Takto získané datové sady očištěné o „nepotřebná data“ vstupovaly do dalších analýz.

Je patrné, že zdrojová satelitní data a manipulace s nimi představují poměrně velkou zátěž pro paměť pracovní stanice a vysoké časové nároky na získání a zpracování obrazových dat. Poměrně zajímavou možností automatizace nabízí R balíček „sen2r“. Při využití některých jeho funkcí lze podle zadání požadovaných parametrů získat více snímků najednou, provést na nich atmosférickou korekci (za využití nástroje Sen2Cor), vytvořit jejich mozaiky (po dnech snímání), maskování oblačnosti, ořez na požadované vodní plochy a výpočet spektrálních indexů v uživatelsky přívětivém prostředí [14].

TVORBA MODELŮ

Vlastní modelování hodnot ukazatelů jakosti vody prostřednictvím satelitních dat proběhlo pomocí dvou modelovacích technik v prostředí statistického softwaru R [15], který je rovněž distribuován pod licencí GPL.

Random Forests (RF) je jednou z technik využívajících např. klasifikačních a regresních stromů (CART). Jedná se o ansámblovou metodu učení pro klasifikaci, predikci, měření významnosti proměnných, měření efektu proměnných na predikci, shlukování a detekci odlehklých hodnot. Spočívá především v konstrukci velkého počtu rozhodovacích stromů, přičemž výsledná regresní funkce je vážený průměr regresních funkcí všech stromů. Při konstrukci RF jsou využívány binární stromy, při jejichž tvorbě se vstupní data dělí na testovací a trénovací soubor. Souborem vstupních dat, se kterým se zde pracuje, je myšlen ansámbl bootstrapových výběrů. Jedná se tedy o náhodné výběry s opakováním vzniklé z původního vzorku pozorovaných dat. Tím lze zajistit, že i velmi malé soubory dat mohou být rozděleny na velký počet testovacích a trénovacích souborů. Trénovací soubory jsou použity pro konstrukci stromů.

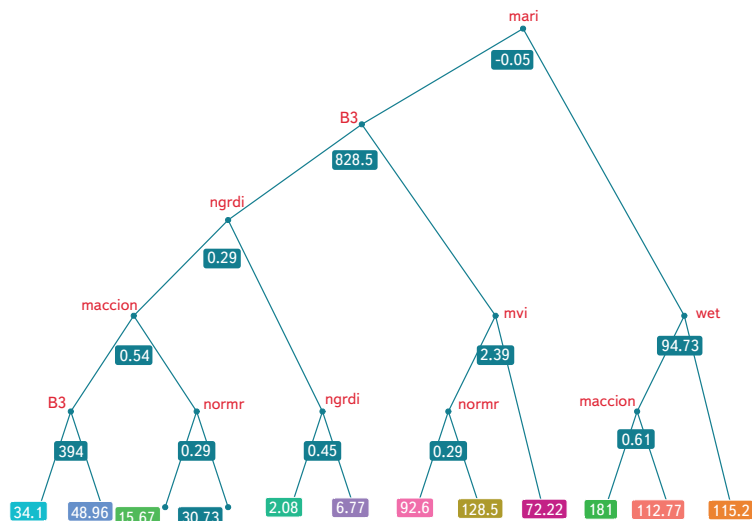
Tabulka 3. Ukazatele sledované v terénu
Table 3. Indicators monitored in terrain

Zkratka ukazatele	Slovní popis	Jednotka
ALG.30	řasy pro standardní odběr pro koupací vody (0–30 cm)	jedinci/ml
CHA.30	chlorofyl-a pro standardní odběr pro koupací vody (0–30 cm)	µg/l
CYANBACCELL.30	sinice (buňky) pro standardní odběr pro koupací vody (0–30 cm)	buňky/ml
CYANBACVOL.30	sinice (objem) pro standardní odběr pro koupací vody (0–30 cm)	mm ³ /l
FEOPIG.30	feopigmenty (buňky) pro standardní odběr pro koupací vody (0–30 cm)	µg/l
OJIP450.10	fluorescence (křivka OJIP ve vlnové délce 450 nm) pro odběr těsně pod hladinou (0–10 cm)	bezrozměrné
OJIP450.30	fluorescence (křivka OJIP ve vlnové délce 450 nm) pro standardní odběr pro koupací vody (0–30 cm)	bezrozměrné
OJIP620.10	fluorescence (křivka OJIP ve vlnové délce 620 nm) pro odběr těsně pod hladinou (0–10 cm)	bezrozměrné
OJIP620.30	fluorescence (křivka OJIP ve vlnové délce 620 nm) pro standardní odběr pro koupací vody (0–30 cm)	bezrozměrné
OJIPSUM.10	fluorescence (součet křivek OJIP ve vlnových délkách 450 nm a 620 nm) pro odběr těsně pod hladinou (0–10 cm)	bezrozměrné
OJIPSUM.30	fluorescence (součet křivek OJIP ve vlnových délkách 450 nm a 620 nm) pro standardní odběr pro koupací vody (0–30 cm)	bezrozměrné
TEMP.10	teplota pro odběr těsně pod hladinou (0–10 cm)	°C
TEMP.30	teplota pro standardní odběr pro koupací vody (0–30 cm)	°C
TURB.10	zákal pro odběr těsně pod hladinou (0–10 cm)	NTU
TURB.30	zákal pro standardní odběr pro koupací vody (0–30 cm)	NTU

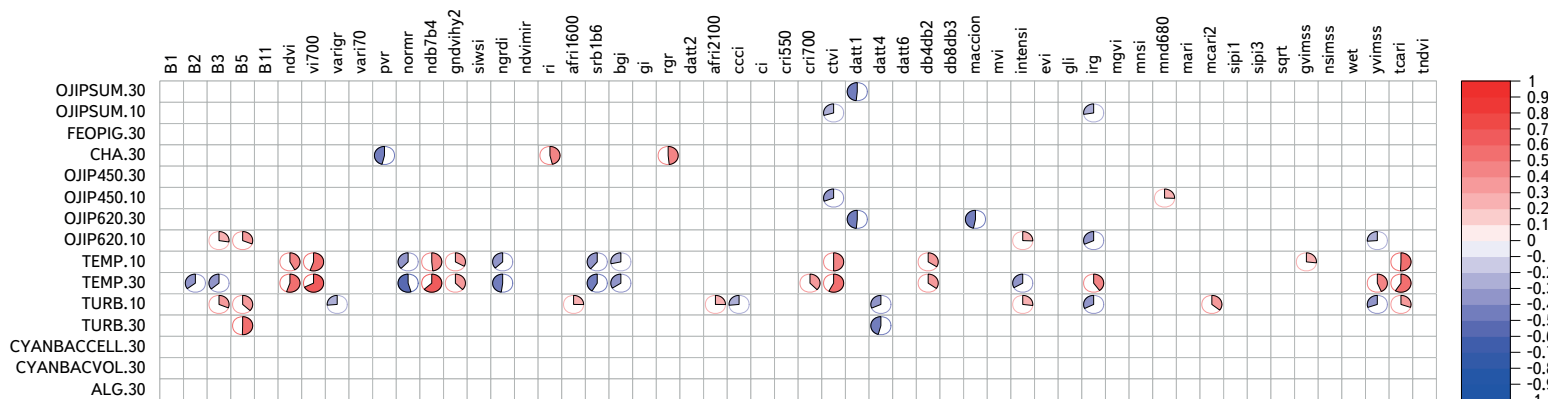
Testovací soubory, tedy pozorování, která se do výběru pro trénování nedostala, jsou využity k odhadu chyby. Základním algoritmem tvorby RF je:

1. vytvoření bootstrapového podsouboru o velikosti N (trénovací soubor),
2. výběr náhodné podmnožiny prediktorů s předem zvoleným počtem těchto prediktorů,
3. vytvoření stromu na bootstrapovém souboru pouze pomocí této podmnožiny prediktorů,
4. zařazení hodnot z testovacího souboru vytvořeným stromem a určení predikce,
5. opakování kroků 1–4 až do vytvoření lesa s požadovaným počtem stromů [16].

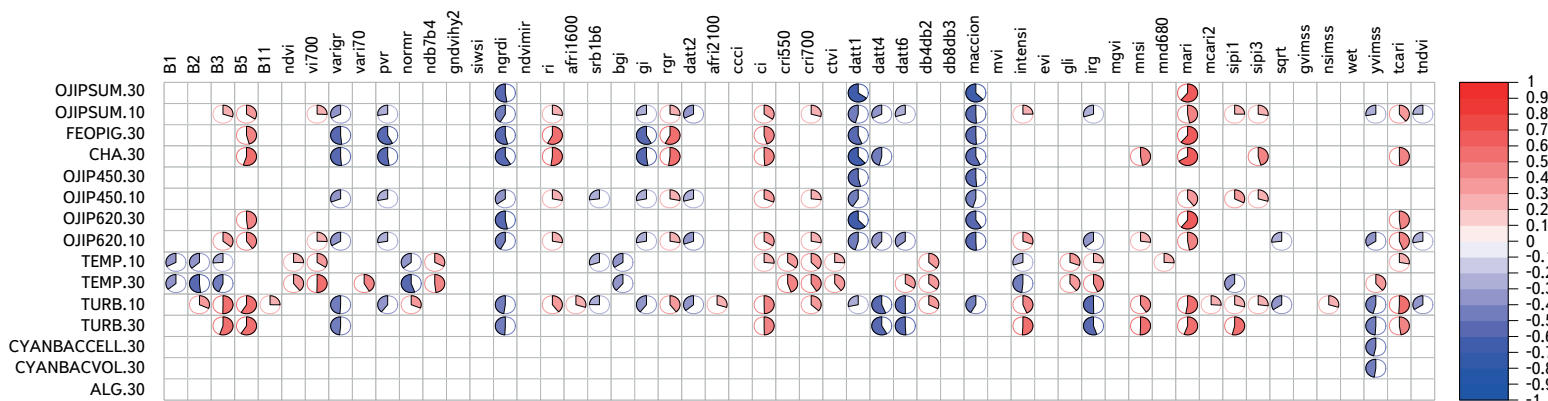
Volba optimálního počtu prediktorů ve druhém kroku algoritmu byla zajištěna postupem implementovaným v R balíčce „CAST“, kde bylo využito funkcí dalších balíčků „caret“ a „randomForest“ [17–22]. Ostatní parametry byly ponechány ve výchozím nastavení. Příklad zkonstruovaného regresního stromu s nejmenším počtem uzlů pro ukazatel CHA.30 je znázorněn na obr. 1.



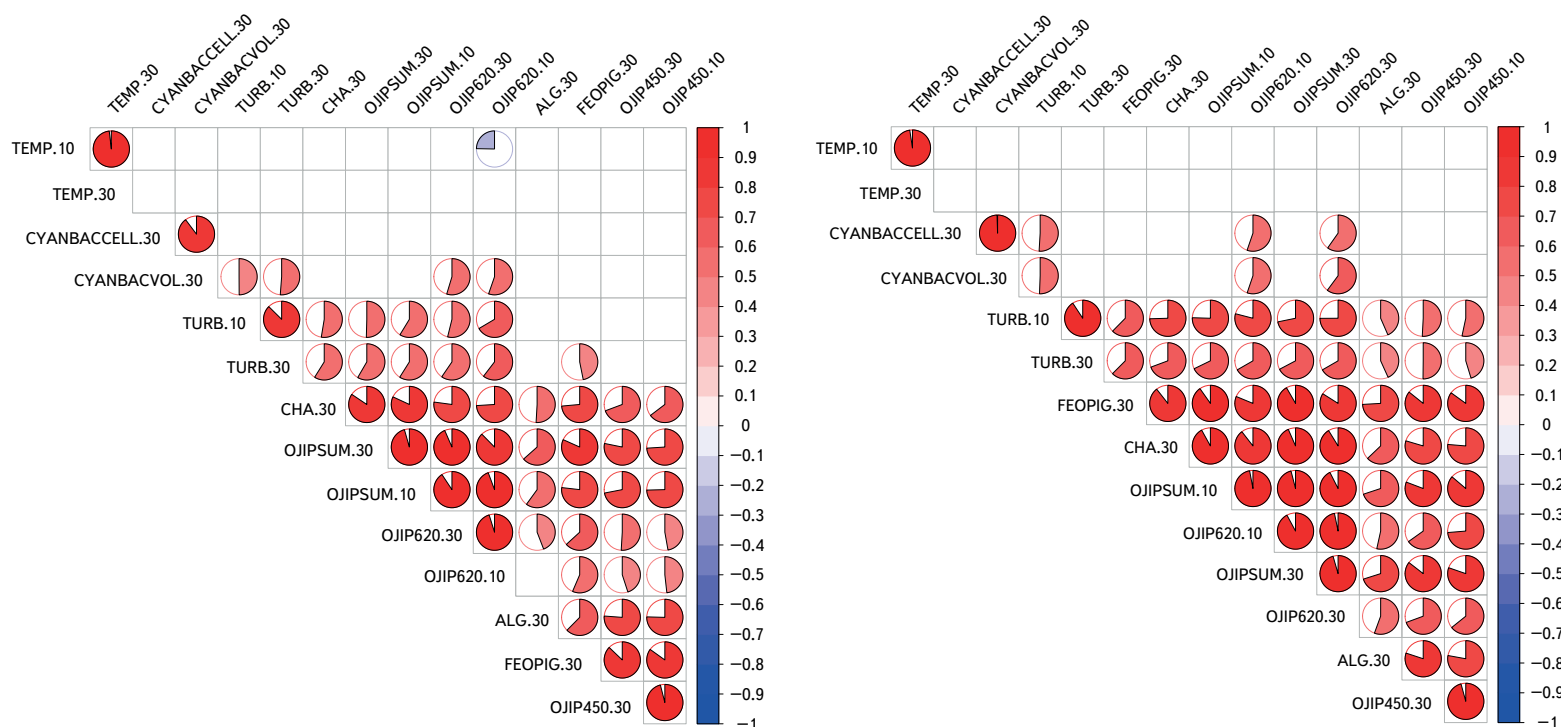
Obr. 1. Příklad regresního stromu s nejmenším počtem uzlů pro ukazatel CHA.30
 Fig. 1. An example of the chlorophyll-a indicator (30 cm depth) regression tree with the smallest number of nodes



Obr. 2. Korelační graf znázorňující vypočítané hodnoty Pearsonova korelačního koeficientu a jejich statistické významnosti na hladině 0,001 pro měřené ukazatele a prediktory vstupující do jakéhokoliv RF modelu
 Fig. 2. Correlation plot showing computed values of the Pearson correlation coefficient and their statistical significance at the 0.001 level for measured indicators and predictors entering any of the RF models



Obr. 3. Korelační graf znázorňující vypočítané hodnoty Spearmanova korelačního koeficientu a jejich statistické významnosti na hladině 0,001 pro měřené ukazatele a prediktory vstupující do jakéhokoliv RF modelu
 Fig. 3. Correlation plot showing computed values of the Spearman correlation coefficient and their statistical significance at the 0.001 level for measured indicators and predictors entering any of the RF model



Obr. 4. Korelační grafy znázorňující vypočítané hodnoty Pearsonova (vlevo) a Spearmanova (vpravo) korelačního koeficientu a jejich statistické významnosti na hladině 0,001 pro měřené terénní ukazatele samotné

Fig. 4. Correlation plots showing computed values of the Pearson (left) and Spearman (right) correlation coefficients and their statistical significance at the 0.001 level for measured terrain indicators themselves

Partial Least Square regression (PLSR) technika pro vytvoření modelu a následnou predikci je v současné době hojně využívána nejen v přírodních vědách, ale například v ekonomii či marketingu. Původně byla vyvinuta pro účely zpracování chemických dat a poprvé byla popsána statistikem H. Woldem [23]. Svě uplatnění nalézá především při posuzování velkého množství možných korelací a vysvětlujících proměnných. Promítnutím mnoha proměnných do výrazně nižšího počtu ortogonálních latentních proměnných redukuje jejich dimenzi. Jinými slovy, informační obsah prediktorů přenesení do uměle vytvořených vzájemně nekorelovaných proměnných, při zachování maximálního informačního rozsahu. Tyto proměnné jsou následně optimalizovány pro aplikaci vícerozměrné lineární regrese. PLSR se snaží pomocí latentních proměnných v prostoru matice prediktorů popsat směr rozptylu v prostoru matice vysvětlovaných proměnných. Před samotným modelováním je část vstupních dat skryta. Kalibrace tedy probíhá nasazením modelu na data známá, zatímco validace modelu pomocí skryté části dat. Při výpočtech byla využita kombinace R balíčků „caret“ a „pls“ [24]. Pro individuální vysvětlované proměnné byly konstruovány modely s maximálním počtem latentních proměnných získávaných z původních proměnných vysvětlujících (tj. spektrálních indexů či pásem).

Výkonnost finálních modelů byla posouzena běžnými „performance“ statistikami, jako jsou např. koeficient determinace (R^2), symetrická střední absolutní procentní chyba (SMAPE), střední čtvercová chyba (MSE) a odmocninová střední čtvercová chyba (RMSE).

VÝSLEDKY A DISKUSE

Na čtrnácti modelových lokalitách (51 odběrových míst) bylo v rámci první etapy terénního šetření (v roce 2019) odebráno 218 vzorků koupacích vod. Pro každé odběrové místo byly stanoveny hodnoty 15 ukazatelů (tabulka 3).

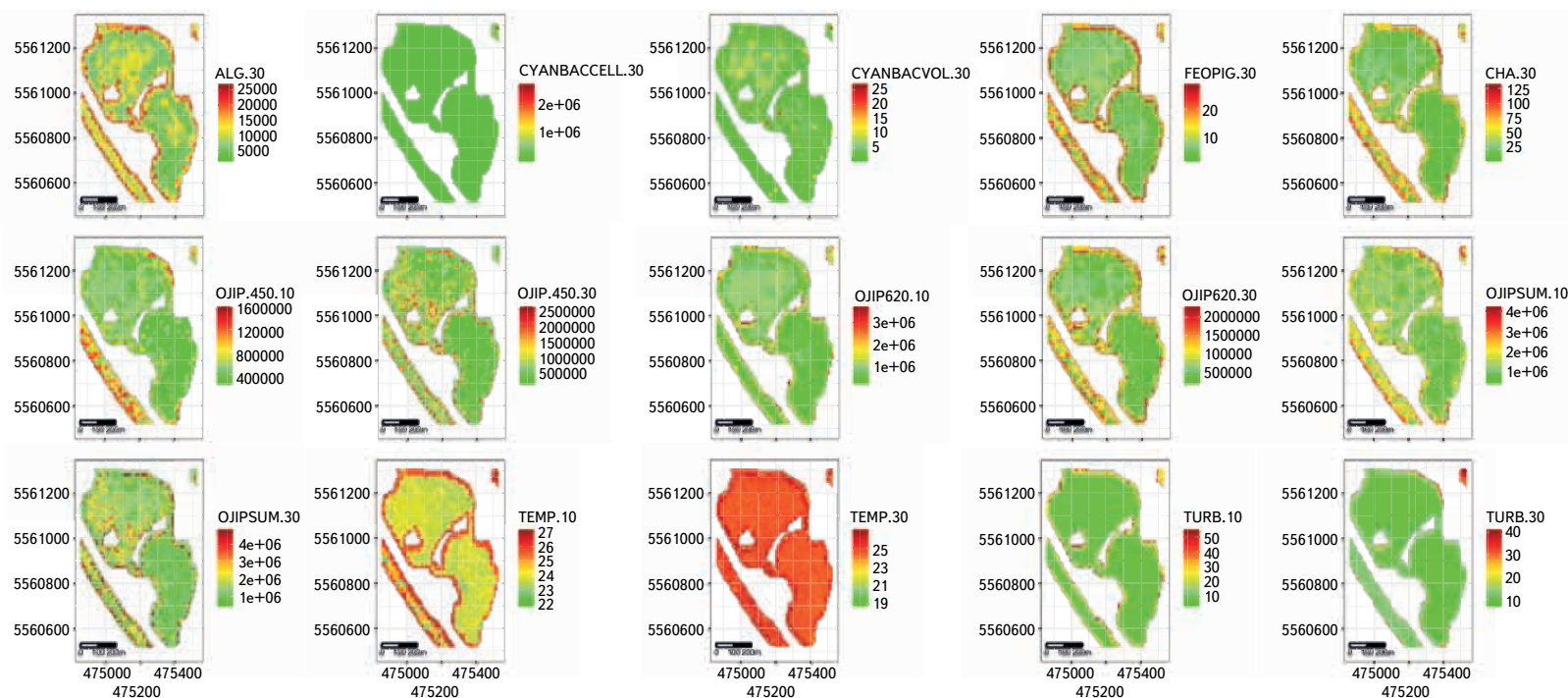
Jedním z významných omezení při využití optických dat DPZ pro monitoring je oblačnost. Je třeba poznamenat, že i když četnost terénních šetření byla výrazně vyšší, nízká kvalita satelitních dat z důvodu nadměrné oblačnosti v několika dnech způsobila, že pro další analýzy bylo relevantních pouze 19 odběrových dní v rámci koupací sezony 2019.

Po skončení první etapy byla terénními pracovníky předána data nejen s hodnotami naměřených ukazatelů, ale i s lokalizací odběrových míst v souřadnicovém systému WGS 84. Tato data byla ještě před vstupem do procesu modelování při standardizaci (např. převod z dlouhého formátu tabulky na formát široký) zbavena zjevných chyb, které mohly být způsobeny častou konverzí datových typů v rámci různých uživatelských prostředí. Pro pozice se známými souřadnicemi bylo po transformaci na společný souřadnicový systém z družicových dat (WGS 84/UTM zone 33N) pro budoucí modely vypočteno 105 spektrálních indexů a určeny hodnoty odrazivosti 11 samostatných spektrálních pásem družic Sentinel-2, tedy celkem 116 prediktorů. Algoritmy pro výpočet vhodných spektrálních indexů byly sestaveny na základě formulací uvedených v databázi indexů využitelných v DPZ (IDB) dostupné online na <https://www.indexdatabase.de/>. Při výběru zdrojových kombinací spektrálních pásem byl především brán ohled na uvedenou primární aplikaci konkrétního spektrálního indexu a na spektrální pásma obsažena v kombinaci. Bylo zjištěno, že ve zdrojové databázi IDB jsou uvedeny duplicitní kombinace spektrálních pásem uvedené pod jiným názvem nebo chybně uvedeny rovnice. Duplicitní prediktory a kombinace s nestandardními či chybějícími hodnotami byly odstraněny, tedy byl redukován počet prediktorů vstupujících do modelu na 104 (93 spektrálních indexů a 11 samostatných spektrálních pásem).

V počáteční fázi prediktivního modelování byly zkoumány vzájemné vztahy mezi veličinami získanými v terénu a prediktory reprezentovanými výstupy ze satelitních dat. Pro tento účel byly vypočteny hodnoty korelací prostřednictvím Pearsonova koeficientu pro lineární vztahy a Spearmanova koeficientu

Tabulka 4. Celková výkonnost modelů RF a PLSR měřená vybranými statistikami
 Table 4. Total performance of the RF and PLSR models measured by selected statistics

	R^2		SMAPE		RMSE	
	RF	PLSR	RF	PLSR	RF	PLSR
ALG.30	0,2	0,05	0,82	1,03	10 243,99	11 147,99
CHA.30	0,59	0,13	0,54	0,74	33,36	48,47
CYANBACCELL.30	0,26	0,07	1,32	1,46	697 731,24	784 006,77
CYANBACVOL.30	0,39	0,19	1,27	1,32	9,09	10,48
FEOPIG.30	0,43	0,04	0,68	0,84	10,06	13,03
OJIP450.10	0,61	0,12	0,45	0,71	484 544,69	723 564,01
OJIP450.30	0,4	0,16	0,58	0,62	541 691,78	641 248,93
OJIP620.10	0,65	0,13	0,49	0,82	865 358,57	1 355 970,77
OJIP620.30	0,49	0,05	0,49	0,75	689 958,74	943 232,68
OJIPSUM.10	0,62	0,09	0,47	0,75	1 203 672,5	1 864 324,12
OJIPSUM.30	0,63	0,03	0,41	0,69	913 420,07	1 471 812,86
TEMP.10	0,69	0,26	0,07	0,11	2,09	3,23
TEMP.30	0,64	0,4	0,07	0,11	2,27	2,91
TURB.10	0,51	0,19	0,43	0,62	15,8	20,35
TURB.30	0,57	0,37	0,34	0,41	7,69	9,32



Obr. 5. Vizualizace výsledných rastrů odhadů sledovaných ukazatelů pro den 3. 6. 2019 vzniklých pomocí RF modelu za využití satelitních dat (lokality Proboštská jezera)
 Fig. 5. Visualization of resulting raster layers of monitored indicators estimates for 3 June 2019 produced using the RF model fed by satellite data (Proboštská lakes site)

pro zahrnutí vztahů nelineárních. Jejich statistická významnost byla sledována na hladinách 0,05, 0,01 a 0,001. U některých vysvětlujících proměnných byla zaznamenána poměrně silná korelace (s absolutními hodnotami koeficientů $> 0,7$) s ukazateli měřenými v terénu. Pouze vysoké hodnoty korelačních koeficientů však nejsou zárukou pro dosažení relevantních výsledků modelování. Základním předpokladem je znalost kauzálních vztahů. Proto byly průběžně výsledky konzultovány v rámci řešitelského týmu s odborníky na jakost koupacích vod. Na základě korelací, s přihlédnutím k jejich statistické významnosti na zvolené hladině, lze tedy usuzovat o vztazích lineárních i nelineárních, na nichž je možné stavět další analýzy.

Výsledky statistického hodnocení vzájemných vztahů mezi proměnnými jsou zobrazeny pro hladinu 0,001 ve formě korelačních grafů znázorněných na obr. 2 a 3 (z důvodu stručnosti jen pro RF modely a prediktory do nich vstupující). Popisují nejen vzájemné vztahy, ale jsou zároveň i výčtem spektrálních pásem a jejich kombinací, které vstoupily do vlastního procesu modelování (v případě PLSR modelů tvořily latentní proměnné všechny prediktory). Zde je patrné, že v modelech jsou zahrnuty i vysvětlující proměnné, které statisticky významnou korelaci nevykazovaly, nebo byl jejich vstup do modelu založen na nelineárním vztahu. Ty měly při náhodném výběru, prováděném balíčkem „CAST“, zásadní vliv na vysvětlení měřených ukazatelů. Skutečnost, že korelace vždy nemusí indikovat kauzalitu, potvrzuje i fakt, že nejvyšší četnost vstupu do modelů měly prediktory B1, ndb7b4, cri700, datt4, maccion, mnd680.

Z výše zobrazených korelačních grafů rovněž vyplývá, že ukazatele chlorofyl-a a zákal významně korelují s pásmy o vlnových délkách 0,69–0,71 μm a jejich kombinacemi. Pásma o vlnových délkách 0,5–0,6 μm , na která byla zaměřena studie provedená Jihočeskou univerzitou v Českých Budějovicích [8, 9], v tomto případě nevykazovala významnou závislost. To si lze vysvětlit zejména vyšší heterogenitou námi zvolených modelových lokalit.

Pro prediktivní modelování bylo využito dvou modelovacích technik, jejichž základní principy jsou popsány výše. Celkem bylo zkonstruováno 30 modelů (tj. pro 15 ukazatelů měřených v rámci terénního šetření a dvě modelovací techniky). Pro modelové lokality vzniklo celkem 570 rastrových vrstev s hodnotami transformovanými do hodnot ukazatelů sledovaných v terénu (tj. pro všech 19 možných dnů). Z hodnocení jejich výkonnosti v tabulce 4 vyplývá, že výrazně kvalitnějších výsledků dosahují modely konstruované pomocí RF, kde u většiny ukazatelů nabývá R^2 hodnot nad 0,5. Nejlepších výsledků dosahuje ukazatel teploty vody v obou měřených horizontech, rovněž hodnoty křivek OJIP (v obou vlnových délkách) a v odběrovém horizontu těsně pod hladinou. Podle očekávání byly vysoko hodnoceny i ukazatele chlorofyl-a a zákal, což koresponduje s výsledky studií [8, 9]. Modely pro predikci sinic nedosahují bohužel uspokojivých výsledků. To lze přiknout nedostatečnému zohlednění sezonnosti a různých vývojových stádií sinic.

Odhady hodnot ukazatelů, pro které se neosvědčil ani jeden z konstruovaných modelů využitých k transferu informace ze satelitních dat do výsledných rastrů, mohou rovněž vycházet ze vztahů mezi ukazateli samotnými, které naznačují korelační grafy na obr. 4. Podmínkou však je, že ukazatel, ze kterého bude potenciálně možné informaci přenést, bude dobře vystižen výše uvedenými modely a zároveň bude vykazovat dobrou korelaci s ukazatelem, kterému bude třeba tuto informaci dodat. Při jejím hodnocení je nezbytně nutné uvědomit si příčinné souvislosti. Z hodnot korelačních koeficientů je též možno posoudit vhodnost využití lineárního, potažmo nelineárního regresního modelu, příp. nutnost veličiny transformovat tak, aby těmto modelům coby vstupy vyhovovaly.

Na obr. 5 je zobrazeno všech 15 ukazatelů pro jeden odběrový den (konkrétně 3. 6. 2019) na modelové lokalitě Proboštská jezera v Brandýse nad Labem. Již na první pohled je patrné, že v mnoha případech zde vznikají na okrajích vodní plochy nadhodnocené predikce. Jedná se převážně o zahrnutí okolní vegetace do pixelů, které zasahují do vodních ploch. Tato systematická chyba byla potlačena ořezem vstupní polygonové vrstvy s vodními plochami a posunem prvků v bodové vrstvě s odběrnými místy.

ZÁVĚR

Monitoring a hodnocení stavu a kvality koupacích vod jsou významné nejen pro pravidelný reporting příslušným institucím, ale i pro informování veřejnosti o stavu životního prostředí. Doposud získané výsledky v rámci řešeného projektu TJ02000091 – *Využití metod dálkového průzkumu Země pro monitoring stavu a kvality koupacích míst v České republice* potvrzují počáteční hypotézu o tom, že data získaná pomocí metod dálkového průzkumu Země mohou představovat významnou podporu a zefektivnění těchto činností. V předloženém článku byla nastíněna nejen základní problematika koupacích vod a monitoringu jejich stavu, ale především zde byly popsány dílčí činnosti postupu, který lze použít pro odhad požadovaných ukazatelů pomocí satelitních snímků, včetně jeho předností a nedostatků. Vzhledem k tomu, že v současné době projekt stále probíhá, budou v průběhu příštího roku jeho kompletní výsledky promítnuty do výstupů projektu v podobě tištěného Atlasu koupacích míst a webové mapové aplikace provozované Výzkumným ústavem vodohospodářským T. G. Masaryka, v. v. i. Rovněž budou závěrečné výsledky i s postupy pro jejich dosažení představeny na workshopu určeném především pracovníkům krajských hygienických stanic a dotčených ministerstev.

Poděkování

Autoři děkují za finanční podporu Technologické agentuře České republiky. Příspěvek byl zpracován s podporou projektu 2. veřejné soutěže Programu aplikovaného výzkumu ZÉTA TJ02000091 – Využití metod dálkového průzkumu Země pro monitoring stavu a kvality koupacích míst v České republice (2019–2021).

Literatura

- [1] Zákon č.258/2000 Sb., o ochraně veřejného zdraví a o změně některých souvisejících zákonů.
- [2] Legislativa vztahující se ke koupalištím. Státní zdravotní ústav [online]. SZÚ: ©2020 [cit. 2020-11-27]. Dostupné z: <http://www.szu.cz/tema/zivotni-prostredi/koupani-ve-volne-prirode/legislativa-vztahujici-se-ke-koupalitim>
- [3] Vyhláška č. 238/2011 Sb., o stanovení hygienických požadavků na koupaliště, sauny a hygienické limity písku v pískovištích venkovních hracích ploch.
- [4] PUMANN, P., BAUDIŠOVÁ, D., KOŽÍŠEK, F., ŠAŠEK, J. a MYŠÁKOVÁ, M. *Metodický návod na vzorkování, terénní a laboratorní vyšetřování a hodnocení jakosti vody v přírodních koupalištích a povrchových vodách ke koupání* [online]. Praha: Státní zdravotní ústav [certifikovaná metodika], 2014. Dostupné z: <http://www.szu.cz/tema/zivotni-prostredi/koupaliste-metody?lang=1>
- [5] Ministerstvo zdravotnictví. *Seznam přírodních koupališť na povrchových vodách* [online]. [cit. 2020-12-01], 2020. Dostupné z: <https://www.mzcr.cz/category/ochrana-verejneho-zdravi/koupani-v-prirode>
- [6] DOBROVOLNÝ, P. *Dálkový průzkum Země: digitální zpracování obrazu*. Brno: Masarykova univerzita, 1998. ISBN 80-210-1812-7.
- [7] European Space Agency. [online]. *European Space Agency* [cit. 2020-12-02], 2020. Dostupné z: <https://sentinel.esa.int>
- [8] BROM, J., NEDBAL, V. a DURAS, J. Rybníky pohledem z výšky – hodnocení kvality vody a eutrofizace pomocí dálkového průzkumu Země. In: *Rybníky 2019* [online]. Praha: Česká společnost krajinných inženýrů, 2019, s. 98–106. ISBN 978-80-01-06595-2. Dostupné z: <http://www.cski-cr.cz/uskutecnene-udalosti/13-14-6-rybniky-2019-konference/>
- [9] NEDBAL, V. a BROM, J. Hodnocení znečištění povrchových vod pomocí satelitních snímků. *Vodní hospodářství*, 2019, roč. 69, č. 5, s. 6–8. ISSN 1211-0760.
- [10] European Space Agency: Sentinel-2 Acquisition Plans [online]. [cit. 2020-12-02]. Dostupné z: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/acquisition-plans>
- [11] European Space Agency: Sentinel-2 – Missions [online]. [cit. 2020-12-02]. Dostupné z: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>
- [12] SNAP. ESA Step: Science Toolbox Exploitation [online]. 2020 [cit. 2020-12-01]. Dostupné z: <http://step.esa.int/main/toolboxes/snap/>
- [13] PCI Geomatica (2018). Geomatica Help.
- [14] RANGHETTI, L., BOSCHETTI, M., NUTINI, F., and BUSETTO, L. "sen2r": An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. *Computers & Geosciences* [online], 2020, Vol. 139, 104473. ISSN 0098-3004. Dostupné z: DOI: 10.1016/j.cageo.2020.104473.
- [15] R CORE TEAM. *R: A language and environment for statistical computing* [online]. Vienna, Austria: R Foundation for Statistical Computing, 2020. Dostupné z: <https://www.R-project.org/>

[16] BREIMAN, L. Random forests. *Machine Learning* [online], 2001, Vol. 45, No. 1, p. 5–32. ISSN 0885-6125. Dostupné z: doi:10.1023/A:1010933404324.

[17] MEYER, H., REUDENBACH, CH., HENGL, T., KATURJI, M., and NAUSS, T. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software* [online], 2018, Vol. 101, p. 1–9. ISSN 1364-8152. Dostupné z: doi:10.1016/j.envsoft.2017.12.001.

[18] MEYER, H., REUDENBACH, CH., WÖLLAUER, S., and NAUSS, T. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling* [online], 2019, Vol. 411, 108815. ISSN 0304-3800. Dostupné z: doi:10.1016/j.ecolmodel.2019.108815.

[19] MEYER, H. CAST: „caret“ Applications for Spatial-Temporal Models [online], 2020. Dostupné z: https://CRAN.R-project.org/package=CAST

[20] KUHN, M. *caret: Classification and Regression Training*. [online], 2020. Dostupné z: https://CRAN.R-project.org/package=caret

[21] KUHN, M. and JOHNSON, K. *Applied Predictive Modeling*. New York, NY: Springer New York, 2013. ISBN 978-1-4614-6848-6.

[22] LIAW, A. and WIENER, M. Classification and Regression by randomForest. *R News*, 2002, Vol. 2, No. 3, p. 18–22. ISSN 1609-3631.

[23] WOLD, H. Estimation of principal components and related models by iterative least squares. In: KRISHNAIAH, P.R. ed. *Multivariate Analysis. Proceedings of the International Symposium held in Dayton, Ohio, June 14-19, 1965*. New York: Academic Press, 1966, p. 391–420.

[24] MEVIK, B.-H., WEHRENS, R., and LILAND, K.H. *pls: Partial Least Squares and Principal Component Regression*. [online], 2020. Dostupné z: https://CRAN.R-project.org/package=pls

Autoři

Ing. Bc. Václava Maťašovská¹

✉ vaclava.matasovska@vuv.cz
ORCID: 0000-0001-9229-463X

Mgr. Filip Kothan²

✉ filip.kothan@szu.cz
ORCID: 0000-0003-3210-5782

Mgr. Ondřej Ledvinka, Ph.D.³

✉ ondrej.ledvinka@chmi.cz
ORCID: 0000-0002-0203-7064

Mgr. Petr Pumann²

✉ petr.pumann@szu.cz
ORCID: 0000-0002-8199-1883

Ing. Tomáš Fojtík¹

✉ tomas.fojtik@vuv.cz
ORCID: 0000-0001-6480-3900

Ing. Marcela Makovcová¹

✉ marcela.makovcova@vuv.cz
ORCID: 0000-0002-1060-4188

Ing. Lenka Bendakovská, Ph.D.²

✉ lenka.bendakovska@szu.cz
ORCID: 0000-0001-8795-8008

¹Výzkumný ústav vodohospodářský T. G. Masaryka, v. v. i.

²Státní zdravotní ústav

³Český hydrometeorologický ústav

Příspěvek prošel lektorským řízením.

DOI: 10.46555/VTEI.2020.12.003

USE OF EARTH REMOTE SENSING METHODS TO MONITOR THE CONDITION OF BATHING SITES

MATASOVSKA, V.¹; KOTHAN, F.²; LEDVINKA, O.³; PUMANN, P.²; FOJTIK, T.¹; MAKOVCOVA, M.¹; BENDAKOVSKA, L.²

¹T. G. Masaryk Water Research Institute, p.r.i.

²National Institute of Public Health

³Czech Hydrometeorological Institute

Keywords: remote sensing – monitoring of bathing waters status – spectral indices – regression

The Ministry of Health in cooperation with other related bodies compiles, on an annual basis, the *List of outdoor bathing sites on surface waters where the bathing service is offered by the operator, and other surface waters used for bathing*. It is compiled based on the long-term monitoring, performed, exclusively as *in-situ* measurements, by the regional hygiene stations. With respect to the fact that this way of monitoring is not only time-consuming but also financially demanding, the need for minimizing these expenses naturally arises.

The currently running project *Use of remote sensing methods for monitoring the status and quality of bathing waters in the Czech Republic* offers possibilities of distant monitoring of some of these indicators necessary for such assessment. The objective is to find and describe the relationships between data obtained in terrain and remote sensing data, by means of up-to-date statistical techniques combined with the use of suitable GIS tools. The project outcomes may then provide the regional hygiene stations with a tool beneficial not only from the perspective of conducting the routine assessment of the parameters but also from the perspective of identification of new bathing sites. In addition, thanks to its time-series character, the produced material can serve as a basis for the assessment of the bathing season.

In the present paper, the procedure leading to the fulfillment of the desired goals is mainly presented. Two stages of field analysis were carried out in 14 selected model sites, followed by laboratory works. In the meantime, optimal ways of processing the satellite data with a huge spatial resolution (Sentinel-2) were sought. After basic multispectral image processing, 105 spectral indices were tested, while the initial assumption was confirmed regarding the strong correlation between some of the indices and the values of measured parameters. For the prediction of these values using the satellite data, two machine-learning techniques were employed, namely *Random Forests* and *Partial Least Squares regression*.