Vedúci práce: doc. Ing. Viera Rozinajová, PhD.

máj 2016

Slovenská technická univerzita v Bratislave  
Fakulta informatiky a informačných technológií

FIIT-5212-72257

Oliver Moravčík

PREDIKCIA VÝROBY ELEKTRINY Z OBNOVITEĽNÝCH ZDROJOV SO ZOHĽADNENÍM EXTERNÝCH FAKTOROV

Bakalárska práca

Študijný program: Informatika

Študijný odbor: Informatika

Miesto vypracovania: Ústav informatiky a softvérového inžinierstva, FIIT STU, Bratislava

Vedúci práce: doc. Ing. Viera Rozinajová, PhD.

máj 2016

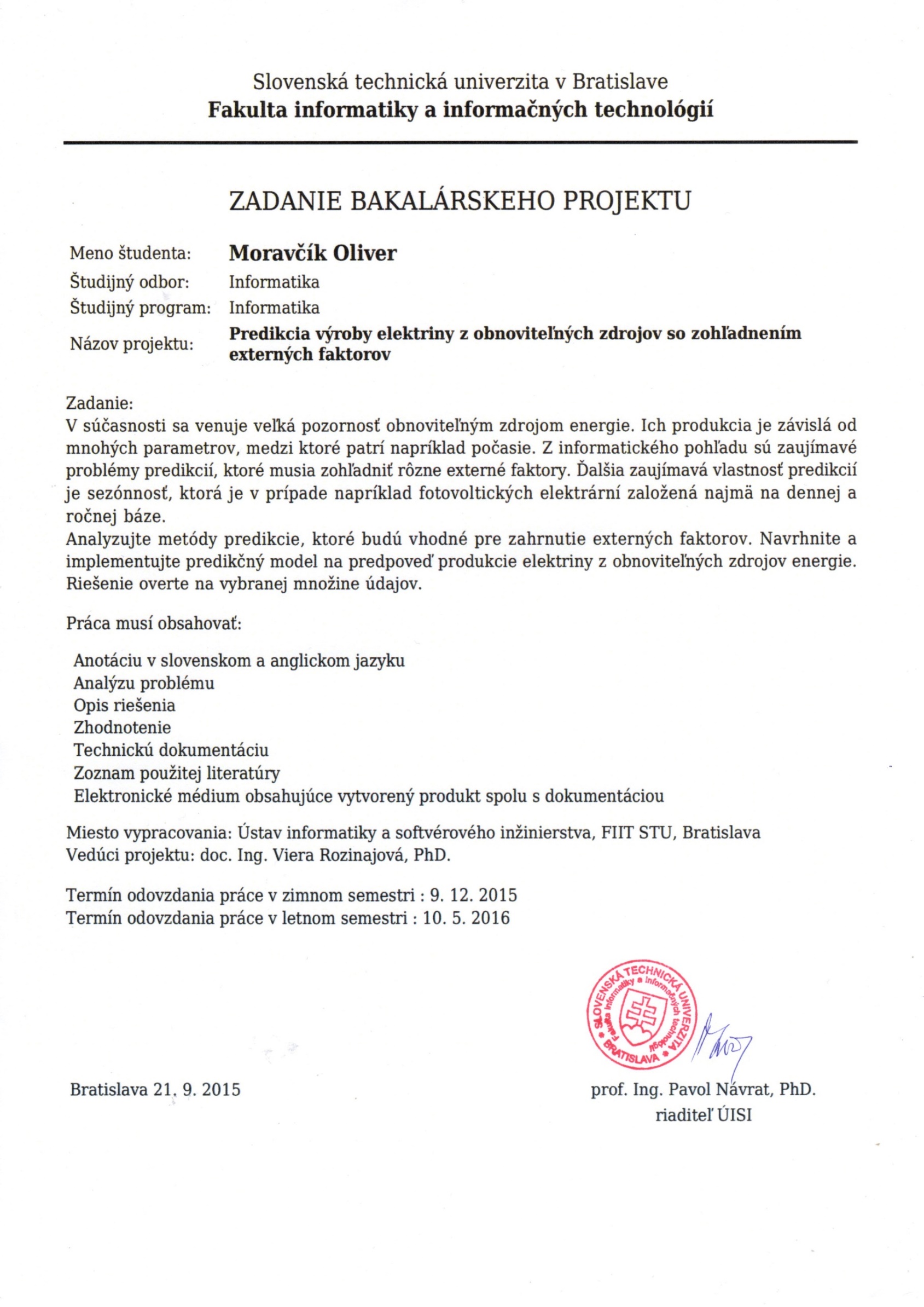
Oliver Moravčík

PREDIKCIA VÝROBY ELEKTRINY Z OBNOVITEĽNÝCH ZDROJOV SO ZOHĽADNENÍM EXTERNÝCH FAKTOROV

Bakalárska práca

Slovenská technická univerzita v Bratislave  
Fakulta informatiky a informačných technológií

FIIT-5212-72257



Slovenská technická univerzita v Bratislave  
**FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLÓGIÍ**

PREDIKCIA VÝROBY ELEKTRINY Z OBNOVITEĽNÝCH ZDROJOV SO ZOHĽADNENÍM EXTERNÝCH FAKTOROV

Študijný program: Informatika

Autor: Oliver Moravčík

Vedúci bakalárskej práce: doc. Ing. Viera Rozinajová, PhD.

máj 2016

Cieľom tejto bakalárskej práce je zanalyzovať predikciu výroby elektriny zo solárnych fotovoltaických elektrární, zanalyzovať metódy používané pri tejto predikcii a vytvoriť predikčný model, ktorý bude predpovedať množstvo vyrobenej elektrickej energie na deň dopredu. V teoretickej časti je zanalyzovaná doména predikcie výroby elektriny z fotovoltaických elektrární, externé faktory vplývajúce na výrobu elektriny a predikčné metódy používané na riešenie daného problému. V praktickej časti je opísané vlastné riešenie problému predikcie výroby elektriny, implementácia predikčného modelu a dosiahnuté výsledky.

Slovak University of Technology Bratislava  
**FACULTY OF INFORMATICS AND INFORMATION TECHNOLOGIES**

FORECASTING ELECTRICITY PRODUCTION FROM RENEWABLE RESOURCES WITH EXTERNAL FACTORS TAKEN INTO ACCOUNT

Degree Course: Informatics

Author: Oliver Moravčík

Supervisor: doc. Ing. Viera Rozinajová, PhD.

May 2016

The aim of this bachelor thesis is analysis of forecasting electricity production from solar photovoltaic power plants, analysis of methods used for forecasting this prediction and to create forecasting model, which will predict electricity production a day ahead. In theoretical part the domain of forecasting electricity production from photovoltaic power plants, external factors affecting the electricity production and forecasting methods used to solve this problem are analyzed. In practical part the own solution of forecasting electricity production problem, prediction model implementation and achieved results are described.Používané skratky

|  |  |
| --- | --- |
| FVE | Fotovoltaické elektrárne |
| GHI | Globálne horizontálne žiarenie (z ang. global horizontal irradiance) |
| NPP | Numerická predpoveď počasia |
| UNS | Umelá neurónová sieť |
| WNS | Waveletová neurónová sieť |
| SVM | Support vector machine |
| MOS | Model output statistics |
| NLRS | Náhodný les regresných stromov |

Obsah

[1 Úvod 1](#_Toc449626871)

[1.1 Motivácia 1](#_Toc449626872)

[1.2 Úvod do predikcie produkcie FVE 3](#_Toc449626873)

[2 Metódy predikcie 5](#_Toc449626874)

[2.1 Metódy založené na analýze časových radov 5](#_Toc449626875)

[2.2 Metódy strojového učenia 9](#_Toc449626876)

[2.3 Fyzikálne metódy 15](#_Toc449626877)

[2.4 Hybridné metódy 16](#_Toc449626878)

[3 Metriky presnosti predikcie 18](#_Toc449626879)

[4 Predikcia výroby elektriny fotovoltaickými elektrárňami 20](#_Toc449626880)

[4.1 Predikcia globálneho horizontálneho žiarenia 21](#_Toc449626881)

[5 Návrh vlastného riešenia 23](#_Toc449626882)

[6 Implementácia vlastného riešenia 25](#_Toc449626883)

[6.1 Dáta 25](#_Toc449626884)

[6.2 Prvé predikcie 28](#_Toc449626885)

[6.3 Výber trénovacej množiny 29](#_Toc449626886)

[6.4 Nastavenia predikčného modelu a faktorov podobnosti 33](#_Toc449626887)

[6.5 Experimenty 33](#_Toc449626888)

[6.6 Najpresnejšie výsledky 33](#_Toc449626889)

[6.7 Zrýchlenie predikcie v R 34](#_Toc449626890)

[6.8 Pokračovanie 35](#_Toc449626891)

[7 Zhodnotenie 37](#_Toc449626892)

[8 Technická dokumentácia 38](#_Toc449626893)

[8.1 Import dát do databázy 38](#_Toc449626894)

[9 Zdroje 46](#_Toc449626895)

# Úvod

Našou úlohou pri riešení bakalárskej práce je predpovedať výrobu elektrickej energie fotovoltaickou elektrárňou podľa predpovede počasia. Pre riešenie daného problému boli použité viaceré metódy. Tieto metódy máme za úlohu analyzovať a jednu metódu vybrať a použiť pre implementáciu vlastného riešenia tohto problému. Predikcia má predpovedať produkciu fotovoltaickej elektrárne na deň dopredu.

Na predikciu výroby elektrickej energie fotovoltaickou elektrárňou (FVE) sa používa viacero metód, ktoré sa bežne používajú na riešenie predikčných problémov. Sú nimi metódy založené na analýze časových radov ako regresné procesy ARMA a ARIMA, metódy strojového učenia ako napríklad umelá neurónová sieť ale aj metódy navrhnuté na základe fyzikálnych faktorov a vzťahov vplývajúcich na produkciu FVE. Opisu a charakteristike týchto metód sa venujeme v kapitole 2 Metódy predikcie.

Dosiahnuté výsledky predikcie je potrebné štatisticky spracovať. Na to slúžia viaceré metriky presnosti, ktorých vymenovanie a stručná charakteristika sú obsahom kapitoly 3 Metriky presnosti predikcie.

Vo štvrtej kapitole sme objasnili problém predikcie výroby elektrickej energie fotovoltaickými elektrárňami.

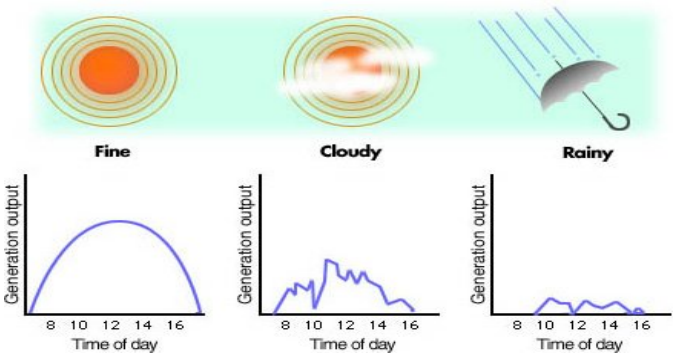
## Motivácia

Slnečná energia je z pohľadu človeka nevyčerpateľný zdroj energie. Slnečná energia dopadá na zemský povrch neustále a na rozdiel od fosílnych palív je zadarmo a nikdy ju nevyčerpáme. Za jednu hodinu dopadne na zemský povrch viac energie, než zemská civilizácia spotrebuje za celý rok. Slnko poháňa aj vetry a vlny. Veterné elektrárne navyše na rozdiel od solárnych zaberajú veľmi málo zemského povrchu, a dokonca žiadny, ak sú postavené v mori, kde sú vetry najsilnejšie. Ak by sa ľudstvu podarilo zužitkovať iba miniatúrny zlomok prístupnej slnečnej a veternej energie, mohlo by produkovať elektrickú energiu pre svoje energetické potreby navždy a bez vypúšťania akéhokoľvek uhlíku do atmosféry.

O nevyužitom potenciály slnečnej energie píše aj docentka Morvová vo svojej knihe Princípy metód a využitie obnoviteľných zdrojov energie(1): „Každý rok dopadne zo Slnka na Zem asi 10 tisíckrát viac energie, ako ľudstvo za toto obdobie spotrebuje. Množstvo dopadajúcej slnečnej energie na územie Slovenska je asi 200-násobne väčšie, ako je súčasná spotreba primárnych energetických zdrojov u nás. Je to obrovský, doposiaľ takmer úplne nevyužitý potenciál. Využívanie slnečnej energie je dnes najčistejším spôsobom využívania energie vôbec a na rozdiel od iných zdrojov (aj obnoviteľných) sú dopady na okolité životné prostredie zanedbateľné.“

„V mnohých krajinách by stačilo pokryť menej ako 1 % územia (napr. strechy budov, nevyužité plochy) slnečnými technológiami, aby bol zabezpečený dostatok energie pre celú krajinu.“

„Podstatné je, že aj v našich klimatických podmienkach je potenciál slnečnej energie obrovský, veď len energia dopadajúca na strechu budovy vo väčšine prípadov presahuje spotrebu energie v nej. Intenzita slnečného žiarenia u nás predstavuje asi 1100 kWh/m2 za rok, kým priemerná spotreba v obytných domoch je len asi 150 kWh/m2 na vykurovanie a 25-50 kWh/m2 na chod elektrospotrebičov a na varenie. Z uvedeného vyplýva, že množstvo dopadajúcej slnečnej energie je až 5-krát väčšie alebo vyjadrené inak je postačujúce na pokrytie spotreby až 5-poschodovej obytnej budovy (merané v hodnotách na m2 horizontálneho povrchu).“

Kvôli miznúcim zásobám fosílnych palív sú solárne elektrárne stále žiadanejšie a buduje sa ich stále viac. Ich produkcia je ale veľmi závislá na počasí ako je vidieť na nasledujúcom obrázku z knihy docentky Morvovej (1) (viď Obrázok 1).

Obrázok 1: Závislosť produkcie elektriny fotovoltaikou od počasia (1)

Pre zapojenie solárnych elektrární do elektrickej siete, pre efektívnu produkciu elektrickej energie a pre marketing na trhu s elektrickou energiou je potrebné predpovedať vo viacerých časových horizontoch produkciu solárnej elektrárne. Predikcia produkcie solárnych elektrární je potrebná aj z dôvodu, že elektrická energia sa nedá efektívne skladovať a solárne elektrárne na rozdiel od konvenčných elektrární nevedia prispôsobiť produkciu očakávanej spotrebe. Efektívna predikcia pomáha operátorom elektrickej siete lepšie manažovať rovnováhu medzi množstvom požadovanej a produkovanej elektrickej energie.

## Úvod do predikcie produkcie FVE

Predpoveď produkcie elektrickej energie solárnymi fotovoltaickými elektrárňami je blízko spojená s predpoveďou počasia. Predpoveď produkcie FVE sa v skutočnosti delí na dve časti. Prvou je predpoveď meteorologických premenných, ktoré majú vplyv na produkciu FVE, a druhou je predpoveď množstva vyrobenej elektrickej energie z FVE na základe predpovedaných meteorologických premenných a charakteristiky FVE. Podľa prečítaných zdrojov sa vždy ako hlavná meteorologická premenná používa *globálne horizontálne žiarenie* (GHI). Je súčtom priameho kolmého žiarenia a rozptýleného (difúzneho) horizontálneho žiarenia. (2) Keďže fotovoltaika znamená priamu premenu slnečnej energie na elektrinu (1), tak pri fotovoltaickej technológii sú dôležité obe zložky. Pri solárnych termálnych elektrárňach, ktoré používajú zrkadlá na nasmerovanie slnečného žiarenia, ktoré ohrieva vodu a vytvára tak paru na výrobu elektriny v parnej turbíne, je dôležité len kolmé žiarenie.

GHI môžeme brať ako numerický ukazovateľ množstva slnečnej energie dopadajúcej na jednotku plochy. GHI sa predpovedá vždy pre konkrétnu oblasť, pretože premenné vstupujúce do výpočtu sú relevantné vždy len pre jednu konkrétnu oblasť. Docentka Morvová vo svojej knihe píše o GHI: „Množstvo dopadajúceho žiarenia na konkrétnom mieste však závisí od viacerých faktorov ako sú napr.:

• zemepisná poloha

• miestna klíma

• ročné obdobie

• sklon povrchu k dopadajúcemu žiareniu.“ (1)

Výpočet GHI závisí od slnečného svitu, a teda aj od dĺžky slnečného svitu a uhla, pod ktorým slnečné lúče dopadajú na povrch Zeme v danej zemepisnej šírke. Najväčšou premennou pri výpočte GHI je oblačnosť, ktorá významne blokuje slnečné lúče a jej premenlivosť hlavnou príčinou chýb vo výpočte predpovede GHI.

Predpoveď GHI je pre predpovedanie výslednej produkcie FVE veľmi dôležitá, pretože presnosť výpočtu GHI má približne 90% vplyv na presnosť predpovede produkcie FVE. Vplyv teploty okolia je na úrovni 10% a vplyv vetra asi 1%. (3)

Prepojenie predpovede GHI s ostatnými časťami predikcie produkcie FVE je znázornené v na nasledujúcom obrázku (viď. Obrázok 2).



Obrázok 2: Prepojenie častí predikcie produkcie FVE

# Metódy predikcie

Autori publikujúci v téme predikcie výroby elektriny z FVE, ako napríklad Diagne a kolektív (4), predikčné metódy rozdeľujú na štatistické a fyzikálne, kde štatistické sú založené na analýze a spracovaní historických dát a fyzikálne využívajú výpočty rovníc popisujúcich vzťahy platné medzi vstupnými premennými. Medzi štatistické metódy tak zaraďujú metódy založené na analýze časových radov aj metódy strojového učenia. Metódy strojového učenia však môžu byť použité aj na analýzu časového radu a predikciu ďalšej postupnosti časového radu, ako aj výpočet predikcie nezávisle od akejkoľvek postupnosti.

Letendre a kolektív (2) taktiež píšu, že predikčné metódy sú všeobecne charakterizované ako fyzikálne alebo štatistické, ale v praxi je hranica medzi týmito prístupmi nejednoznačná. Fyzikálne metódy explicitne modelujú fyzikálne atmosférické javy pri predikcii GHI použitím numerickej predpovede počasia (NPP) alebo snímok oblohy. Štatistické metódy predikujú GHI pomocou trénovania a štatisticky odvodených hodnôt. Ako príklad uvádzajú, že fyzikálny prístup k predpovedi môže použiť vektorovo založenú predikciu rozvoja oblakov použitím interpolácie nedávnych, po sebe nasledujúcich snímok oblohy a štatistický prístup môže použiť súčasné a historické výstupné hodnoty elektrárne k predikcii budúcich výstupných hodnôt.

## Metódy založené na analýze časových radov

Do tejto kategórie patria hlavne autoregresné modely. Regresné metódy sa úspešne používajú v predikcii časových radov už dlhší čas. Použitím tohto prístupu zistíme vzťahy medzi prediktormi, premennými použitými na vstupe a premennými, ktoré máme predikovať. (4)

### Perzistentný (stály) model

Perzistentný model nie je autoregresný ale je jednoduchý predikčný model, ktorý je nutné spomenúť. Tento model predpokladá, že predpovedaná hodnota v čase *t* je rovnaká ako v čase *t – 1* :

Tento model je naivný, pretože predpokladá nemennosť predpovedanej hodnoty. V praxi sa však používa na predikciu globálneho horizontálneho žiarenia s krátkym časovým horizontom, napríklad 15 minút a malou úpravou predikcie podľa pozície Slnka na oblohe. (2) Využívajú ho operátori FVE pri riadení procesu výroby elektrickej energie. Model je ale nepresný pre predpoveď na dlhší časoví horizont než 1 hodina. Jeho presnosť sa výrazne znižuje, ak sa mení oblačnosť. Perzistentný model je ale užitočný ako základ porovnania predikčných modelov. Je užitočné porovnávať výsledky predikčného modelu s jednoduchým predikčným modelom, ako je tento. Implementovať komplexný predikčný nástroj je úplne zbytočné, pokiaľ nepreukáže jasne lepšie vlastnosti ako jednoduchý predikčný model. (4)

### AR proces

AR (autoregressive) model, alebo proces znamená autoregresný proces. Je to stacionárny stochastický proces. Stochastický proces je nekonečnou postupnosťou náhodných veličín usporiadaných v čase. Pre modelovanie stochastického procesu je potrebné porozumieť povahe jeho náhodnosti. Túto náhodnú zložku je ľahšie definovať pri radoch, ktorých rozdelenie pravdepodobnosti sa v čase nemení. Takéto rady nazývame stacionárnymi. Časové rady môžu byť silne stacionárne a slabo stacionárne. Podrobnejšie vysvetlenie týchto vlastností nebudeme potrebovať. Vlastnosť stacionarity opisuje závislosť rozdelenia bieleho šumu.

Pre pochopenie AR procesu definujem najskôr autoregresný proces rádu 1. Autoregresný proces rádu 1 = AR(1) je definovaný nasledovne:

kde *Xt* je hodnota v čase *t* a *at* je takzvaný *biely šum*. Je to náhodná zložka. Biely šum je postupnosť nezávislých náhodných premenných (hodnôt) z normálneho rozdelenia. (5) *Φ1*je váha (koeficient) zložky. Využitím operátora spätného posunu *B*, pre ktorý platí , môžeme vzťah definujúci AR(1) proces prepísať do tvaru:

Po pochopení jednoduchosti AR(1) môžeme ľahšie pochopiť autoregresný proces vyššieho rádu. Autoregresný proces rádu *p* = AR(*p*) je taký proces, kedy je aktuálna hodnota časového radu v čase *t* lineárnou kombináciou predchádzajúcich hodnôt tohto radu. Ak je len konečný počet váh π nenulový, a pre , tak výsledný proces je autoregresným procesom (modelom) rádu *p*. (5)

AR(*p*) je definovaný vzťahom:

alebo v tvare s operátorom spätného posunu:

pričom

Pretože , AR proces je vždy invertovateľný. Aby bol AR proces stacionárny, korene musia ležať mimo jednotkového kruhu.

AR procesy sú užitočné pri opise situácií, v ktorých súčasná hodnota časového radu záleží na jeho predchádzajúcich hodnotách a bielom šume (náhodnej zložke). (5)

### MA proces

Skratka MA znamená „moving average“ a je to proces kĺzavých priemerov. V tomto procese je opisovaná hodnota časového radu lineárnou kombináciou súčasných a minulých hodnôt bieleho šumu (náhodnej veličiny). To znamená, že sa v časovom rade prejaví s rôznou intenzitou vplyv hodnôt z minulosti. MA je tiež stacionárnym procesom. Opäť pre pochopenie definujem najskôr MA(1) – proces kĺzavých priemerov prvého rádu:

respektíve:

Ak je len konečný počet váh *ψ* nenulový, a pre , tak výsledný proces je procesom pohyblivých priemerov rádu *q*. (5)

MA(*p*) je definovaný vzťahom:

A v tvare s použitím operátora spätného posunu:

pričom

a *θt* je predstavuje doznievajúcu náhodnú hodnotu *t* krokov dozadu.

Pretože , konečný MA proces je vždy stacionárny. MA proces je invertovateľný ak korene ležia mimo jednotkového kruhu. MA procesy sú užitočné pri opisovaní javu, v ktorom udalosti produkujú okamžitý efekt, ktorý trvá len krátke časové obdobie. (5)

### ARMA proces

Ako je evidentné už zo skratky názvu tohto procesu, ARMA proces (autoregressive moving average = autoregresný proces kĺzavých priemerov) je spojením predchádzajúcich dvoch procesov, teda ARMA proces má aj autoregresnú zložku aj zložku kĺzavých priemerov.

Stacionárne procesy môžu byť reprezentované v AR forme alebo MA forme. Problém v týchto reprezentáciách ale je, že obsahujú až priveľa parametrov ako pre konečný MA rad, tak aj pre konečný AR rad, pretože pre dobrú aproximáciu je potrebný model vysokého rádu. Vo všeobecnosti vysoký počet parametrov znižuje efektivitu odhadu. Preto pri budovaní modelu môže byť nevyhnutné zahrnúť obe zložky, AR aj MA, čo vedie k ARMA modelu. (5)

Takýto model je užitočný pri opise rôznych časových radov. Spojenie AR a MA procesu často presnejšie popisuje skutočné dáta. Je to zapríčinené nevyhovujúcimi podmienkami použitia AR alebo MA procesov.

Bolo dokázané, že model je vhodný na predikciu, ak je v časovom rade lineárna závislosť. Hlavnou požiadavkou na ARMA model je, že časový rad musí byť stacionárny, (4) čo znamená, že stredná hodnota a rozptyl bieleho šumu sa nemenia s časom.

ARMA(1,1) obsahuje AR(1) aj MA(1):

alebo:

Analogicky s predchádzajúcimi modelmi môžeme definovať ARMA(*p, q*) ako:

alebo:

kde

V ARMA(*p, q*) procese *p* a *q* reprezentujú rády príslušných AR a MA polynómov. Preto ARMA(*p,* 0) je ekvivalentný procesu AR(*p*) a opačne aj pre MA proces.

### ARIMA proces

ARIMA je integrovaný autoregresný proces kĺzavých priemerov. Oproti ARMA procesu má ešte o jednu zložku naviac. ARIMA(*p, d, q*) slúži na spracovanie nestacionárnych časových radov a práve táto pridaná zložka *d* opisuje, koľko krát musel byť časový rad diferencovaný, kým z neho bol stacionárny rad, respektíve po diferencii koľkého rádu bol z pôvodného nestacionárneho časového radu rad stacionárny. Stacionárny proces, ktorý je výsledkom diferencie nestacionárneho radu nie je nevyhnutne biely šum. Všeobecne diferencovaný rad pripomína ARMA(*p, q*) proces. Preto môže byť ARIMA(*p, d, q*) zapísaný nasledovne:

kde stacionárny AR operátor

a invertovateľný MA operátor

nezdieľajú žiadny spoločný faktor. Parameter hrá veľmi rozdielnu rolu pre a . Ak je *d* rovné nule, pôvodný proces je stacionárny a je viazaný na priemer radu. Ak je *d* väčšie ako nula, je nazývané deterministickým trendom a je často vynechávané z modelu, pokiaľ nie je naozaj potrebné. (5)

Daigne (4) píše, že Reikard aplikoval ARIMA metódu na predikciu GHI. Porovnal ARIMA s inými metódami na časovom horizonte 24 hodín, a tvrdí, že ARIMA model zachytáva presné prechody v žiarení spojené s denným cyklom presnejšie než ostatné metódy.

## Metódy strojového učenia

Metódy strojového učenia patria do oblasti umelej inteligencie. O výskum umelej inteligencie je stále väčší záujem. Pribúdajú nové techniky, známe techniky sa zlepšujú. Techniky umelej inteligencie sú použiteľné vo veľa oblastiach, nie len v predikcii, ale aj pre široké spektrum aplikácií, kompresii dát, optimalizácii, rozoznávaní vzorov a klasifikácii. Techniky umelej inteligencie boli brané aj pri predikcii za alternatívne riešenia, ale stali sa vďaka svojim výsledkom plnohodnotným prostriedkom na riešenie problému predikcie a sú častokrát po vhodnej implementácií presnejšie a spoľahlivejšie ako klasické riešenia. Je to kvôli tomu, že na rozdiel od klasických metód analýzy časových radov, kedy predikované hodnoty časového radu závisia od predchádzajúcich hodnôt, metódy strojového učenia môžu brať tieto hodnoty ako samostatné, neuvažujú pôvodnú štruktúru dát a dokážu tak riešiť aj nelineárne problémy, pri ktorých klasické metódy zlyhávajú alebo dodávajú neuspokojivé výsledky.

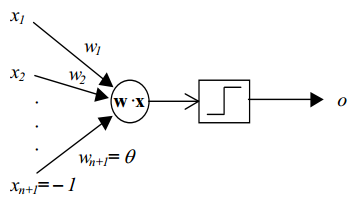
Tieto metódy väčšinou fungujú ako čierna skrinka, čo znamená, že nevieme ako model vypočítal výstupnú hodnotu. Pri metódach strojového učenia sa spolieha na ich schopnosť naučiť sa rozoznať vzory vo vstupných dátach. Metódy strojového učenia je možné natrénovať na rôznych dátach, čo z nich robí veľmi univerzálny nástroj.

### Umelá neurónová sieť

Umelá neurónová sieť (UNS) sa snaží napodobiť štruktúru a funkcionalitu prirodzených neurónových sietí, teda nervových systémov živých organizmov. Tak ako sa prirodzená neurónová sieť skladá z množstva neurónov, ktoré sú medzi sebou poprepájané synapsiami, UNS sa skladá z formálnych neurónov. Sinčák a Andrejková (6) definujú neurónovú sieť ako masívne paralelný procesor, ktorý má sklon k uchovávaniu experimentálnych znalostí a ich ďalšieho využívania. Napodobňuje ľudský mozog v dvoch aspektoch:

* Poznatky sú zbierané v neurónovej sieti počas učenia
* Medzineurónové spojenia (synaptické váhy) sú využívané na ukladanie znalostí.

Jednou z veľmi významných vlastností neurónových sietí je, že svojim spôsobom je tzv. univerzálnym aproximátorom funkcií. Môže sa nám stať, že máme systém, ktorého popis je mimoriadne náročný alebo je systém natoľko zložitý, že jeho popis je skoro nemožný. Máme však dáta, ktoré do systému vstupujú, a k nim odpovedajúce výstupy. V takejto situácii, môžeme použiť vhodnú UNS a pokúsiť sa ju naučiť chovať sa ako sledovaný systém pomocou trénovacích údajov. (6) Presne tento prístup k modelovaniu výpočtu sa najčastejšie využíva pri predikcii výroby elektrickej energie z FVE.

Základným prvkom UNS je formálny neurón. Špeciálnym typom formálneho neurónu je perceptrón. Perceptrón má na vstupných prepojeniach váhy (viď Obrázok 3). Váha na danom prepojení patrí k hodnote, ktorú neurón prijíma daným prepojením. Vo váhach perceptrónu sa uchovávajú znalosti UNS.

Obrázok 3: Perceptrón (7)

Beňušková (7) definovala perceptrón nasledovne: Perceptrón je model neurónu, ktorý prijíma vstupné signály cez synaptické váhy tvoriace váhový vektor . Vstupný vektor sa nazýva *vzor* alebo obrazec (angl. *pattern*). Zložky vstupného vektora môžu nadobúdať reálne alebo binárne hodnoty. Zložky váhového vektora sú reálne čísla. Výstup perceptrónu o je daný vzťahom:

kde premenná *net* označuje váhovanú sumu vstupov, t.j. skalárny (zložkový) súčin váhového a vstupného vektora. Funkcia *f* sa volá *aktivačná funkcia* perceptrónu. V tejto notácii predpokladáme, že perceptrón má *n+1* vstupov. Hodnota *(n+1)-*ého vstupu je vždy –1 a , čo je hodnota prahu excitácie perceptrónu. (7)

Spôsobov opisu štruktúry neurónovej siete je viacero. Neuróny môžeme označiť za vrcholy, spojenia neurónov za hrany a celú neurónovú sieť môžeme opísať ako orientovaný graf. Analyzovať takto štruktúrovanú sieť je ťažké, preto bývajú častejšie neurónové siete organizované do jednoduchších pravidelných štruktúr ako je napríklad viacvrstvová štruktúra. Vo viacvrstvovej štruktúre rozlišujeme nasledujúce vrstvy:

* Vstupná vrstva – neuróny tejto vrstvy prijímajú vstupné informácie z prostredia mimo neurónovej siete. Informačný tok (výstup týchto neurónov) pokračuje k neurónom ďalšej vrstvy.
* Skrytá vrstva – neuróny tejto vrstvy prijímajú vstupné informácie zo vstupnej vrstvy, inej skrytej vrstvy, alebo aj z prostredia mimo neurónovej siete. Informačný tok pokračuje k neurónom ďalšej vrstvy.
* Výstupná vrstva – neuróny tejto vrstvy prijímajú vstupné informácie z predchádzajúcej vrstvy ale výstup neurónov tejto vrstvy vyúsťuje do prostredia mimo neurónovej siete.

Neurónová sieť má jednu vstupnú a jednu výstupnú vrstvu. Skrytých vrstiev môže byť niekoľko. Neuróny dvoch vrstiev sú prepojené spôsobom „každý s každým“, čiže ak je vrstva M s počtom neurónov *m* prepojená s vrstvou N s počtom neurónov *n*, počet prepojení medzi týmito dvomi vrstvami je .

Vo všeobecnosti sa štruktúra neurónových sietí rozdeľuje do dvoch kategórií:

* Dopredné neurónové siete – pri týchto sieťach sa signál šíri po orientovaných synaptických prepojeniach len jedným smerom a to dopredu. (6)
* Rekurentné neurónové siete - pri rekurentných sieťach je dosť ťažké rozdelenie vrstiev a neurónov na vstupné, resp. výstupné. Niekedy neuróny v rekurentných sieťach predstavujú vstupné ale aj výstupné typy neurónov a tým aj vrstiev. Špeciálnym prípadom sú tzv. čiastočne rekurentné UNS, v ktorých je stanovená určitá požiadavka na štruktúru a na prepojenia. Napríklad vrstvové čiastočne rekurentné siete pripúšťajú šírenie signálu oboma smermi. (6)

UNS sa trénujú na trénovacej množine dát. V tejto množine je vzorka historických vstupných dát aj s požadovaným výstupom, ktorý má UNS predikovať. Existuje viacero trénovacích algoritmov, ktoré tu nie je nutné opisovať. Všeobecne sa pri trénovaní menia váhy perceptrónov tak, aby sa celkový výsledok priblížil k požadovanej hodnote. Pri trénovaní UNS hrozí takzvané *pretrénovanie*, kedy to vyzerá, že sa UNS naučí naspamäť predikovať hodnoty z trénovacej množiny, ale pri predikciách zo vstupných údajov mimo trénovacej množiny je veľmi nepresná.

Viacvrstvové UNS, ktoré sa trénujú metódou spätného šírenia chýb, sú schopné riešiť aj nelineárne problémy. (7) Preto sa využívajú pri predikcii tak, že sa natrénujú na historických dátach a potom vedia generovať výstup podľa vstupných dát, ktoré sú predikovanými vstupnými údajmi pre čas, pre ktorý má neurónová sieť predikovať výstupné údaje. Rekurentné UNS dokážu generovať časový rad podľa vzoru vstupnej vzorky časového radu. Vďaka svojej rekurentnosti si UNS uloží sebou generovaný člen časového radu, zahrnie ho do vstupnej vzorky a generuje nový člen časového radu.

Daigne (4) píše, že použitím trénovacích dát UNS zredukovali odmocninu zo strednej kvadratickej chyby (RMSE, vysvetlené v kapitole 3) priemerného denného GHI až o 15% v porovnaní s 12 až 18 hodinovými predikciami NPP. Reikardove výsledky ukázali, že v rozlišnosti 60, 30 a 15 minút dávajú ARIMA modely presnejšie výsledky. Podľa výsledkov Sfetsosa a Coonicka je najvhodnejšia UNS. (4)

### Waveletová neurónová sieť

K predikcii sa používajú aj neurónové siete s použitím waveletovej transformácie. *Wavelet* je funkcia „malej vlny“ zvyčajne označovaná ako . Malá vlna rastie a klesá v konečnej perióde na rozdiel od „veľkej vlny“, ako je sínusoida, ktorá rastie a klesá opakovane v nekonečnej perióde. Wavelety sa používajú pri spracovaní signálov a analýze časových radov. Wavelety sú základom *waveletovej transformácie*, ktorá závisí od dvoch premenných – od frekvencie a času. Poznáme dva základné typu waveletovej transformácie – spojitú a diskrétnu. Spojitá sa používa na funckie definované na celej množine reálnych čísel a diskrétna na funkcie definované na množine celých čísel. Zvyčajne , kde je počet hodnôt v časovom rade. Aby funkcia mohla byť zatriedená ako wavelet, musí mať tieto tri vlastnosti (8):

1. Integrál funkcie je rovný nule.
2. Integrál druhej mocniny funkcie je rovný jednotke.
3. Funkcia musí spĺňať podmienku prípustnosti.

Prvá vlastnosť hovorí o tom, že každý rast funkcie nad nulu musí byť vyrovnaný klesaním pod nulu. K tomu by stačilo splnenie rovnosti , ale druhá vlastnosť hovorí, že funkcia musí nadobudnúť konečný počet stúpnutí nad nulu. Ak je splnená podmienka prípustnosti, signál opísaný funkciou môže byť zrekonštruovaný z jeho súvislej waveletovej transformácie (8).

Waveletové neurónové siete (WNS) sú kombináciou waveletov a neurónových sietí. Vo všeobecnosti WNS pozostávajú z doprednej neurónovej siete s jedným alebo viacerými vstupmi, s jednou skrytou vrstvou neurónov, ktorých aktivačné funkcie sú waveletové funkcie, a s výstupnou vrstvou s jedným alebo viacerými lineárnymi kombinátormi (sumármi).

WNS sú použiteľné na odhad funkcií, kedy môžu byť natrénované na sérii hodnôt funkcie, k výpočtu očakávaných hodnôt funkcie pre daný vstup. WNS sú veľmi dobré pre aproximáciu neznámych funkcií a sú odolné voči bielemu šumu.

### Support vector machine

Názov support vector machine (SVM) môžeme preložiť ako stroj podporného vektora, ale ani v slovenskej odbornej literatúre sa preklady tohto názvu nepoužívajú a táto metóda je známa pod svojím anglickým názvom.

SVM je klasifikátor odvodený od strojového učenia, ktorý mapuje vektor prediktorov (vstupné parametre predikcie) do viacrozmerného priestoru cez lineárne alebo nelineárne jadrové (z anglického „kernel“) funkcie. V probléme binárnej klasifikácie sú dve skupiny (-1 a +1) oddelené vo viacrozmernej nadrovine podľa princípu minimalizácie rizika. Zámerom je nájsť rozdeľujúcu nadrovinu skonštruovanú podľa vektora prediktorov namapovaného do viacrozmerného priestoru nelineárnou funkciou a vektor váh s počiatkom odchýlky, ktorá klasifikuje všetky hodnoty do jednej z dvoch skupín. Keďže v probléme binárnej klasifikácie je týchto rozdeľujúcich nadrovín nekonečný počet, cieľom je nájsť optimálnu rovinu, ktorá je od oboch skupín najviac vzdialená. (9)

Zjednodušene by sme mohli povedať, že pri binárnej klasifikácii SVM rozdeľuje body v priestore do dvoch skupín (priestorov/nadrovín), ktoré môže oddeliť rovinou. Najlepším rozdelením je také, ktoré tieto dve skupiny od seba rozdeľuje s čo najväčšou medzerou medzi nimi.

Metóda SVM bola pôvodne používaná na rozoznávanie vzorov a binárnu klasifikáciu, ale princípy jej fungovania môžu byť ľahko rozšírené pre regresiu a predikciu časových radov. SVM dosahuje dobré výsledky pre nelineárne problémy, pričom nepotrebuje poznať štruktúru dát, čo je výhodou pri predikcii nelineárnych časových radov. SVM sa používa iba zriedka, hoci majú veľa teoretických výhod pre klasifikačné aj regresné úlohy. (10)

Thissen a kolektív (10) píšu, že pri zložitejšej trénovacej množine určenej na porovnanie výsledkov SVM model dosiahol lepšie výsledky pri predikcii časových radov ako ARMA model a vo väčšine prípadov aj lepšie výsledky ako Elmanova neurónová sieť. Pri použití menšej trénovacej množiny, ktorá obsahovala len desatinu dát boli výsledky SVM a ARMA modelov rovnako dobré ale Elmanovu neurónovú sieť nebolo možné použiť pre predikciu časového radu.

### Náhodné lesy klasifikačných alebo regresných stromov

Náhodné lesy sú kombináciou stromových prediktorov takých, že každý strom je založený na náhodnej vzorke z vektora hodnôt a s normálnym rozdelením pre všetky stromy v lese. Tento princíp môže byť aplikovaný aj na regresiu. (11)

Teóriu náhodných lesov predstavil Breiman (11) aj s metódou *bagging*-u klasifikačných stromov, kedy stromy nezávisia od predchádzajúcich stromov a každý je nezávisle skonštruovaný zo vzorky z množiny dát. Breiman navrhol náhodné lesy, ktoré pridávajú ďalšiu vrstvu náhodnosti do metódy bagging-u. Okrem konštrukcie každého stromu z inej vzorky dát náhodné lesy menia spôsob ako sú klasifikačné alebo regresné stromy konštruované. V klasických stromoch je každý uzol rozdelený najlepším možným rozdelením medzi všetkými premennými. V náhodných lesoch je každý uzol rozdelený najlepším rozdelením medzi podmnožinou prediktorov, ktorá je náhodne vybraná pri danom uzle. Táto kontra-intuitívna stratégia dosahuje veľmi dobré výsledky v porovnaní s mnohými inými metódami klasifikácie ako SVM a UNS, a je odolná voči pretrénovaniu. Výsledná predikovaná hodnota je vybraná agregáciou predikcií jednotlivých stromov. Pri klasifikácii je vybraná väčšinová (najpočetnejšia) výstupná hodnota a pri regresii je vybraný priemer z hodnôt. (12)

V Španielsku použili kvantilové regresné lesy, ktoré sú založené na Breimanových náhodných lesoch. Kvantily dávajú viac informácií o rozložení výstupných hodnôt ako o funkcii prediktorov než o samotnom priemere. S použitím tejto metódy implementovali model, ktorý predikoval dennú produkciu so strednou chybou menšou než 1,3% a s celkovou priemernou absolútnou chybou menšou než 9,5%. (13)

### Meranie presnosti metód strojového učenia

Pri meraní presnosti modelu, ktorý je založený na metóde strojového učenia, je potrebné vytvoriť dve množiny údajov:

1. Trénovaciu množinu – obsahuje údaje, na ktorých sa model natrénuje (nastavia sa hodnoty váh perceptrónov UNS).
2. Testovaciu množinu – obsahuje údaje, na ktorých sa overí schopnosť modelu odhadnúť predikovanú hodnotu.

Pre zachovanie správnych podmienok pre testovanie musí platiť pravidlo, že prienikom týchto dvoch množín musí byť prázdna množina, čo znamená, že žiadna jednotka vstupných údajov z trénovacej množiny nesmie byť v testovacej množine. Zabezpečí sa tak nezávislosť testovacej množiny. Správny výber trénovacej a testovacej množiny je kritickým pre hodnotenie výkonu predikčného modelu.

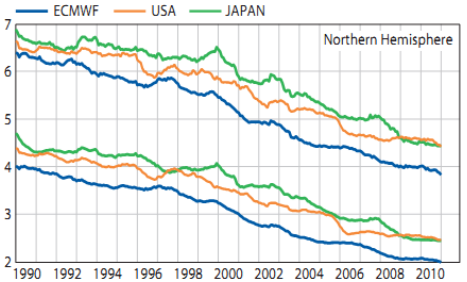
## Fyzikálne metódy

Pri fyzikálnych metódach predikčné modely predikujú výstupné hodnoty počítaním rovníc opisujúcich fyzikálne zákony a vzťahy medzi vstupnými parametrami. Fyzikálne predikčné modely na predikciu produkcie FVE sa implementujú na mieru pre konkrétnu FVE podľa charakteristiky fotovoltaických panelov elektrárne a vstupných hodnôt ako predpovedaná hodnota GHI a teplota vzduchu alebo teplota zadnej strany fotovoltaických panelov, respektíve teplota buniek fotovoltaických panelov pre daný časový interval, pre ktorý má model predikovať výslednú hodnotu.

Fyzikálne modely sú veľmi presné po správnej implementácii na mieru danej FVE ale sú náchylné na chyby v predikcii vstupných parametrov ako GHI. Z tohto dôvodu sa pri fyzikálnych modeloch často využíva dodatočné spracovanie výstupu nazývané *postprocessing*.

Fyzikálnym modelom pre predikciu celkového stavu atmosféry a teda aj potrebných vstupných údajov pre predikciu produkcie FVE ako teplota, oblačnosť alebo aj GHI je model numerickej predpovede počasia (NPP).

### Modely numerickej predpovede počasia

Modely NPP sa všeobecne používajú na predikciu stavu atmosféry až na 15 dní dopredu. Časový vývoj stavu atmosféry je modelovaný základnými diferenčnými rovnicami, ktoré popisujú fyzikálne zákony vládnuce počasiu. Začiatočné podmienky sú získavané z celosvetového pozorovania a v prvom kroku je budúci stav atmosféry vypočítaný globálnym NPP modelom. Globálne NPP v súčasnosti fungujú v asi pätnástich spoločnostiach poskytujúcich predpoveď počasia. Vlastné globálne modely NPP majú napríklad US National Oceanic and Atmospheric Administration (NOAA) a European Centre for Medium-Range Weather Forecasts (ECMWF). Vlastný globálny NPP majú aj v Japonsku. ECMWF v súčasnosti poskytuje najkvalitnejšie strednodobé a dlhodobé meteorologické predpovede na svete, ako je aj vidieť na obrázku (viď Obrázok 4).

Obrázok 4: Porovnanie odmocniny zo strednej kvadratickej chyby (RMSE) pre trojdňovú (nižšie) a päťdňovú (vyššie) predpoveď (3)

Od ECMWF má dáta aj Slovenský Hydrometeorologický ústav (SHMU), ktorý vytvoril vlastný model NPP s názvom Aladin, ktorý však nie je globálny, ale regionálny a poskytuje najpresnejšiu predpoveď počasia pre územie Slovenska. (3)

Globálne modely majú zvyčajne hrubé rozlíšenie a nedovoľujú detailné mapovanie. Síce sa rozlíšenie za posledné roky zvýšilo, dnes je to stále v rozmedzí 16 až 50 kilometrov (priestorové rozlíšenie). Regionálne modely pokrývajú menšiu časť Zeme, preto môžu operovať nad vyšším rozlíšením. Na NPP modely sa môže aplikovať dodatočné spracovanie výstupu. Postprocessing môže byť aplikovaný za účelom modelovania špecifických lokálnych efektov a vlastností konkrétnych oblastí ako nížiny a pohoria, kde je rozlíšenie na niekoľko kilometrov skresľujúce, alebo môže byť postprocessing aplikovaný za účelom zníženia chýb modelu NPP.

## Hybridné metódy

Hybridné modely boli navrhnuté pre prekonanie nedostatkov individuálnych modelov ako sú napríklad fyzikálne modely. Hybridné metódy spájajú rozdielne metódy pre zvýšenie presnosti predpovede. Tieto modely dosahujú vyššiu presnosť predpovede kombináciou modelov s rôznym prístupom. Často spájajú metódy založené na analýze časových radov pre spracovanie lineárneho vstupu a využívajú umelú inteligenciu na rozoznávanie nelineárnych vzorov. Veľmi často používanou kombináciou je predikcia meteorologických parametrov využitím metód založených na analýze časových radov alebo metódou strojového učenia a následné zadanie týchto predikovaných hodnôt ako vstupu do fyzikálneho modelu, ktorý vypočíta množstvo vyprodukovanej elektrickej energie za daných vstupných meteorologických podmienok.

Letendre (2) tvrdí, že predpovedať produkciu FVE použitím iba štatistických metód nie je vo všeobecnosti súčasťou moderných predikčných systémov solárnej energie, ale hybridný prístup využíva pokročilé štatistické techniky ku korekcii známych nedostatkov spojených s rozdielnymi predikčnými metódami, úpravami stredných chýb alebo metódami strojového učenia.

Pri fyzikálnych modeloch sa často využíva postprocessing so štatistickým prístupom. Pri takomto prístupe sa spracovávajú historické dáta pre korekciu predikovaných hodnôt. Metóda štatistickej korekcie výstupu sa nazýva *model output statistics* (MOS).

Model output statistics (MOS) používa štatistické vzťahy medzi pozorovanými elementmi počasia a meteorologickými dátami, satelitnými údajmi alebo modelovanými parametrami k získaniu štatistickej úpravy výstupu. (2) Nevýhodou MOS je ako pri všetkých štatistických prístupoch potreba presných historických dát.

Hoci sa na MOS používajú prevažne metódy založené na analýze časových radov, akákoľvek štatistická metóda patrí do konceptu MOS. Daigne (4) píše o brazílskom NPP modeli, ktorého predpovede GHI boli veľmi nadhodnotené a aplikovaním UNS bolo dosiahnuté značné zlepšenie výsledkov.

# Metriky presnosti predikcie

Existuje viacero metrík pre určenie presnosti predikcie výroby elektrickej energie z FVE. Rôzne metriky sa hodia pre rôzne účely. Operátori FVE potrebujú metriky, ktoré presne odrážajú ceny predikčných chýb (straty). Výskumníci vyžadujú indikátory na porovnanie výsledkov viacerých predikčných modelov alebo jedného modelu pri rozdielnych podmienkach. (14)

Pri všetkých metrikách sa počíta s chybou *ei*, ktorá je rozdielom predpovedanej hodnoty *yp* a nameranej hodnoty *yn,*, dolný index *i* označuje *i*-te poradie premenných v intervale od 1 po *N*, kde *N* je počet všetkých párov predpovedaná-nameraná hodnota v testovacej množine:

Štandardnými používanými metrikami presnosti predikcie sú nasledovné štatistické ukazovatele:

* Stredná chyba = mean bias error - MBE:
* Stredná kvadratická chyba = mean square error = MSE
* Odmocnina zo strednej kvadratickej chyby = root mean square error - RMSE:
* Priemerná absolútna chyba = mean absolute error = MAE:
* Maximálna absolútna chyba = maximal absolute error = MAXAE:
* Štandardná odchýlka = standard deviation = SDE:

Charakteristika MBE (označuje sa aj „bias“) je priemernou predikčnou chybou a zapuzdruje systémovú tendenciu modelu k nadhodnoteniu alebo podhodnoteniu predpovedanej hodnoty (systematická chyba). Charakteristika MAE je priemerným rozsahom (rozptylom) predikčnej chyby. Charakteristika RMSE je MAE veľmi podobná, ale dáva viac váhy väčším chybám. Pellandová (14) spomína Madsena a kolektív, ktorý argumentuje, že väčšie chyby majú disproporčnú cenu, a preto RMSE lepšie odráža cenu predikčných chýb pre systémových operátorov než MAE. Preto je RMSE najpoužívanejšou metrikou presnosti predikcie výroby elektriny. Charakteristika MAXAE vyjadruje maximálnu chybu. Tento ukazovateľ je vhodný pre výskumníkov, ktorý môžu na základe tohto ukazovateľa porovnať maximálnu chybovosť porovnávaných modelov. Vzťahy medzi MSE, SDE a MBE sú nasledovné:

Inými slovami, SDE zachytáva časť RSME, ktorá nie je spôsobená systematickou chybou a poskytuje náznak hodnoty RMSE, ktorá môže byť docielená elimináciou MBE. (14)

Charakteristiky RMSE, MAE a MAXAE majú vlastnosť metriky, takže merajú vzdialenosť predpovede od nameraných údajov. Pre dokonalú predpoveď sú tieto charakteristiky rovné nule. Používajú sa aj relatívne resp. percentuálne hodnoty rRMSE, rMAE a rMBE pričom normalizácia je vzhľadom na namerané hodnoty. (3) Takéto ukazovatele sú vhodné pre celkové hodnotenie presnosti predikcie, pretože sú normalizované. Odchýlka od presného výsledku o 10 jednotiek je rovnaká aj pri očakávanej hodnote 100 jednotiek, aj pri očakávanej hodnote 1000 jednotiek. Po normalizácii vzhľadom na očakávané (namerané) hodnoty dostaneme odchýlku 10% pri prvej hodnote a 1% pri druhej hodnote, čo je desaťnásobný rozdiel. Preto je vhodné porovnávať výsledky v normalizovanom tvare.

# Predikcia výroby elektriny fotovoltaickými elektrárňami

Pri predikovaní produkcie FVE sa používajú metódy opísané v kapitole 2 Metódy predikcie. Predikčné metódy sme v tejto kapitole rozdelili do dvoch skupín: fyzikálne a štatistické. Fyzikálny prístup využíva vstupné dáta a využíva vzťahy medzi premennými pre výpočet rovníc popisujúcich produkciu elektrickej energie fotovoltaickými panelmi. Štatistický prístup sa spolieha na vstupné dáta a historické dáta, podľa ktorých sa môže model trénovať. Fyzikálny model funguje ako biela skrinka, kedy vidíme ako model vypočíta výstupné hodnoty a štatistický model väčšinou funguje ako čierna skrinka, kde nevidíme ako model dospel k výsledkom.

Pre fyzikálny model sú potrebné presné meteorologické vstupné údaje ako GHI a teplota vzduchu alebo teplota zadnej steny panelov alebo teplota fotovoltaických článkov, a presné údaje o fotovoltaických paneloch, ktoré charakterizujú produkciu elektrickej energie. Preto sú fyzikálne modely navrhnuté presne na mieru pre danú špecifikáciu FVE a charakteristiku používaných fotovoltaických panelov.

Pri štatistickom prístupe sú dôležité historické dáta ako predpovedané alebo namerané hodnoty GHI, ukazovatele počasia (teplota, rýchlosť vetra, oblačnosť), údaje sledovania oblačnosti, a k nim nameraný výstup FVE, teda množstvo vyprodukovanej elektrickej energie za daných meteorologických podmienok. Takéto historické dáta sa použijú ako časový rad alebo ako trénovacia množina dát, ktorá sa použije na natrénovanie predikčného modelu ako napríklad umelá neurónová sieť. Pre štatistické metódy je lepšie, ak sa trénujú na historických dátach predpovedí, než na skutočných nameraných dátach, pretože predikcia sa vždy vypočítava na dátach predpovede meteorologických premenných a tie majú rovnakú chybu predikcie ako majú historické dáta predpovede meteorologických premenných. Rozdiel medzi predpovedanými a skutočnými hodnotami má v takomto prípade menší negatívny vplyv na presnosť výslednej predikcie výroby elektriny.

Presnosť predikcie výroby elektrickej energie závisí na presnosti predpovedi vstupných meteorologických premenných. Keďže modely využívajúce fyzikálny prístup berú do úvahy vstupné dáta pre daný jeden deň, ich presnosť býva nižšia v porovnaní so štatistickým prístupom, kvôli chybám v predpovedi hodnôt vstupných premenných. Hoci štatistický prístup nemodeluje priamo produkciu FVE, jeho výhodou je, že je menej náchylný na chyby v predpovedi vstupných údajov (GHI) v jednotlivých dňoch, pretože trénovacia množina obsahuje viacero dní, podľa ktorých sa model učí predikovať produkciu FVE.

Fyzikálny prístup použili Lorenz a kolektív pre predikciu produkcie FVE pre dve oblasti v Nemecku. Počas ročného testovacieho obdobia sa im podarilo dosiahnuť výsledky s odmocninou zo strednej kvadratickej chyby rovnej 3,9% pre predpoveď v rámci hodiny a 4.6% pre predpoveď na deň dopredu. (14) Huang a kolektív a aj Kudo a kolektív porovnávali fyzikálny a štatistický prístup k predikcii. Štatistický prístup dosahoval o málo lepšie výsledky než fyzikálny.

## Predikcia globálneho horizontálneho žiarenia

Štatistické metódy pre predikciu GHI sa používajú rovnaké ako štatistické metódy pre predikciu produkcie FVE. Rozdiel je vo vstupných a výstupných parametroch a štruktúre dát v trénovacích množinách. Fyzikálnymi modelmi pre predikciu GHI sú modely NPP (opísané v kapitole 2.3.1).

V štatistických modeloch sa používajú historické dáta, čo je problémom pre nové FVE v oblasti, z ktorej historické dáta zatiaľ neexistujú. To v praxi znamená, že FVE si musí dáta zbierať a až po čase sa presnosť predpovede spresní do prijateľnej miery. Väčšinou sa meracie prístroje v oblasti inštalujú dopredu, aby boli prístupné historické dáta už pri dokončení výstavby FVE.

Výber správnej metódy alebo modelu závisí od časového horizontu, pre ktorý chceme predikovať. Štatistické modely sú presnejšie pre krátkodobé predpovede a ich výpočet predikcie je na počítači veľmi rýchly, takže sa môžu často opakovať, čo ich priamo predurčuje pre krátkodobé predpovede. NPP modely sú presnejšie pri predikcii na vzdialenejší časový horizont a aj ich výpočet trvá dlhšie. Výpočet sa vykonáva dvakrát až štyrikrát za deň a beží na superpočítačoch, ktoré zvládnu viacero operácii naraz a majú väčšiu pamäť pre spracovávanie veľkého množstva dát.

Ako som spomínal vyššie, najväčší vplyv na hodnotu GHI a hlavne na zmenu tejto hodnoty má oblačnosť. Preto je *sledovanie oblačnosti* veľmi dôležitou súčasťou predikcie GHI. Niektoré predikčné modely sú založené z väčšej časti na sledovaní oblačnosti a tvoria tak samostatnú kategóriu. Dáta zo sledovania oblačnosti využívajú ale aj oba vyššie spomínané typy predikčných modelov.

### Sledovanie oblačnosti

Najväčší vplyv na úroveň GHI má stav oblohy, konkrétne oblačnosť. Oblačnosť je veľmi premenlivá s časom aj priestorom, pretože sa oblaky v atmosfére hýbu a menia. Preto je sledovanie oblohy a stavu oblačnosti základnou úlohou pre predpoveď GHI.

V súčasnosti sa pre predikciu GHI používajú snímky oblačnosti zo satelitov aj snímky zo zeme. Oba spôsoby sledovania oblačnosti poskytujú pomerne veľké časové aj priestorové rozlíšenie. Základom pre využitie týchto snímok je sledovanie štruktúry a pohybu oblakov v zaznamenaných časových krokoch. Pokročilejšie modely sledujú aj tiene oblakov. Na základe sledovaných snímok je možné predpovedať ako sa oblaky zmenia, kam smerujú a aké z toho vyplývajú následky pre úroveň GHI na povrchu v danej sledovanej oblasti.

Chyby v predikciách založených na sledovaní oblačnosti prudko vzrastajú pri nízkej výške Slnka na oblohe (nízkom uhle dopadu slnečných lúčov), vysokých priestorových nepravidelnostiach povrchu (pohoria, doliny), a nízkych hodnotách GHI.

Spracovaním satelitných aj pozemných snímok, oblaky môžu byť detegované, charakterizované a využité k predpovedi GHI pomerne presne až na 6 hodín dopredu. Modely časových radov založené na dátach zo sledovania oblačnosti detegujú pohyb oblakových štruktúr používaním vektorových polí. (4)

Modely spracovávajúce satelitné snímky dokážu zachytiť väčšiu oblasť a môžu preto sledovať pohyb oblačnosti na väčšom priestore po dlhší čas. Modely spracovávajúce snímky oblohy získané z povrchu Zeme nedokážu sledovať tak veľký priestor. Prakticky nevidia oblačnosť vzdialenú viac než 30 minút (závisí od rýchlosti pohybu oblačnosti). Majú ale vyššie priestorové a časové rozlíšenie a majú schopnosť zaznamenať náhle zmeny. Preto sú presnejšie pre predikcie do časového horizontu 30 minút. Satelitné snímky sú najpresnejšie pre predikcie pre časový horizont 30 minút až 6 hodín. Pre predpovede na obdobie vzdialenejšie ako 6 hodín sú už potom presnejšie NPP modely.

# Návrh vlastného riešenia

Úlohou nášho riešenia je predpovedať produkciu fotovoltaickej elektrárne na deň dopredu podľa predpovedaného počasia. Pre vlastné riešenie sme si vybrali použitie umelej neurónovej siete. UNS je podľa nás najvhodnejšou predikčnou metódou vzhľadom na dostupné dáta, horizont predikcie a faktory ovplyvňujúcich produkciu FVE ako je vysvetlené nižšie.

Meteorologické máme k dispozícii od Slovenského hydrometeorologického ústavu (SHMU). Dáta od SHMU sú z NPP modelu Aladin. Predikovať budeme produkciu FVE na nasledujúci deň, čo je časové obdobie za horizontom šiestich hodín, takže použitie dát z NPP modelu je najlepšou možnosťou, pretože v takomto časovom horizonte dosahujú NPP modely najpresnejšie výsledky v porovnaní s ostatnými predikčnými modelmi. Model Aladin navyše dosahuje najpresnejšie predpovede počasia na území Slovenska (3), čo tiež prispeje k presnosti nami implementovaného predikčného modelu. Dáta obsahhujú potrebné meteorologické premenné na predpoveď výroby elektriny ako sú globálne horizontálne žiarenie (GHI), teplota vzduchu, rýchlosť vetra a iné.

K dispozícii máme aj namerané dáta z prevádzky fotovoltaických elektrární, ktorých produkciu máme za úlohu predikovať. Meteorologické dáta sú lokalizované na umiestnenie týchto fotovoltaických elektrární. Môžeme tak vytvoriť dvojicu – množinu premenných opisujúcich meteorologické podmienky v danom okamihu konkrétneho dňa s periódou jednej hodiny a množstvo vyrobenej elektrickej energie v prislúchajúcom časovom intervale.

Pri predikcii množstva vyprodukovanej elektriny za celý deň nebudeme potrebovať predikovať časový rad, keďže predikovaná hodnota nezávisí od postupnosti hodnôt produkcie predchádzajúcich dní, ale od predpovedaných meteorologických a iných premenných. Pri predikcií v rámci dňa po hodinových časových intervaloch predikovanie časového radu prichádza do úvahy, pretože za jasného počasia hodnoty produkcie elektrickej energie do poludnia stúpajú a po poludní klesajú. Výsledné hodnoty sú ale ovplyvnené zmenou meteorologických podmienok viac ako postupnosťou časového radu, preto nebudeme predikovať časový rad ani v rámci dňa.

Umelú neurónovú sieť nebudeme samy implementovať. V jazyku R, ktorý pre vlastné riešenie použijeme, je balík *neuralnet*, ktorý umožňuje jednoduché použitie UNS. V jazyku R taktiež využijeme dostupné funkcie na meranie presnosti predikcie, spracovanie štatistiky a vyobrazenie štatistických údajov do grafov.

Ťažiskom našej implementácie bude výber trénovacej množiny. Výberom správnej trénovacej množiny sa zvyšuje presnosť predikcie neurónovej siete. UNS chceme pre predikciu trénovať na množine obsahujúcej údaje o dňoch podobných s dňom, pre ktorý má UNS predikovať produkciu elektriny.

Podobnosť meteorologických podmienok budeme porovnávať podľa úrovne GHI, miery oblačnosti a uhla dopadajúceho priameho slnečného žiarenia, pretože tieto údaje súvisia so slnečnou energiou dopadajúcou na fotovoltaické panely. V jazyku R je možné vypočítať uhol priameho slnečného žiarenia ale ekvivalentom by mohla byť aj časová vzdialenosť daného dňa od zimného slnovratu. Ďalej budeme porovnávať podobnosť meteorologických podmienok podľa teploty vzduchu a rýchlosti vetra, pretože tie súvisia s teplotou fotovoltaických panelov.

Dni v dostupných historických dátach plánujeme podľa podobnosti s dňom predpovede číselne ohodnotiť, podľa tejto hodnoty usporiadať a vybrať potrebný počet dní do trénovacej množiny.

Veľkosť trénovacej množiny je dôležitý a zaujímavý faktor. Pri väčšej trénovacej množine trénovanie UNS trvá dlhšie, ale vzorka údajov je pestrejšia. Keďže ale budeme predpovedať na deň dopredu, rýchlosť trénovania neurónovej siete ani výpočtu predikcie nie je prioritou. Prioritou je presnosť, preto môže byť trénovacia množina väčšia a aj neurónovú sieť môžeme trénovať pred každou predikciou s vhodnou trénovacou množinou.

Pri veľmi veľkom počte podobných dní v historických dátach predpokladáme, že by nebolo najlepšie vybrať najpodobnejšie dni, pretože by v nich bol pravdepodobne veľmi malý rozdiel. Takáto trénovacia množina by nemusela mať dostatočnú pestrosť údajov na správne určenie váh perceptrónov. Pre zabezpečenie dostatočnej pestrosti dát v trénovacej množine by obsah trénovacej množiny mohol byť vyberaný náhodne z dní dostatočne podobných dňu, pre ktorý bude UNS predikovať. Do úvahy prichádza aj úprava pravdepodobnosti náhodného výberu dát podľa hodnoty podobnosti (podobnejšie dni by mali väčšiu pravdepodobnosť výberu), alebo možnosť vybrania aj menej podobných dát do trénovacej množiny.

Dôležité pri výbere dát do trénovacej množiny je, že raz vybrané dáta sa nesmú do trénovacej množiny vybrať druhý krát. UNS by sa tak na týchto dátach trénovala dvakrát viacej ako na ostatných a hrozilo by takzvané pretrénovanie, kedy by to vyzeralo, ako keby sa sieť naučila dané predpovede naspamäť ale predpovede pre iné vstupné dáta by boli nepresné. Možným znížením rizika pretrénovania je aj väčšia a teda pestrejšia trénovacia množina a menší počet opakovaní cyklov pri trénovaní neurónovej siete.

Najvhodnejšiu veľkosť trénovacej množiny a najvhodnejšiu možnosť výberu trénovacej množiny bude možné určiť porovnaním presnosti výsledkov predikcie pri použití rôzne veľkých trénovacích množín a rôznych možností výberu trénovacej množiny.

# Implementácia vlastného riešenia

## Dáta

Pre našu prácu máme k dispozícií dáta meteorologických predpovedí z modelu Aladin lokalizované na umiestnenie fotovoltaických elektrární, z ktorých máme záznamy o ich produkcii za rovnaké obdobie. K týmto dátam sme doplnili hodnoty slnečných koordinátov a dĺžku dňa (slnečného svitu), ktoré sme vypočítali v jazyku R.

Model Aladin je model numerickej predpovede počasia, a dáta z tohto modelu sú najlepšími možnými dátami, aké sme mohli pre svoju prácu získať, pretože pre predpoveď počasia s časovým horizontom 24 hodín dopredu sú predpovede z modelov NPP najpresnejšie a model Aladin je regionálny model NPP a na území Slovenska dosahuje najpresnejšie výsledky predpovede, ako sme aj napísali už v kapitole 2.3.1 Modely numerickej predpovede počasia.

Každý záznam predpovede počasia z modelu Aladin je predpoveď na 24 hodín dopredu, takže každý záznam má rovnakú chybu predpovede, čo nám dovoľuje pristupovať ku každému záznamu rovnako a každý náš výsledok predikcie závisí od rovnakej chyby predpovede počasia. Znamená to aj, že pri hypotetickom použití v reálnom procese produkcie elektrickej energie by boli tieto dáta dostupné 24 hodín dopredu, čo zabezpečuje dostatočné množstvo času na ich spracovanie a použitie pri predikcií výroby elektrickej energie.

Záznamy predpovede počasia sú bodové záznamy s hodinovým krokom, čo znamená že predpovedajú pre konkrétny časový okamih s periódou jednej hodiny. Časové údaje sú uložené v UTC (Coordinated Universal Time), teda v koordinovanom svetovom čase, ktorý sa často používa v technických kruhoch, pretože je akýmsi základom, na ktorý sa odkazujú všetky časové pásma a nepoužíva letný čas.

Predpoveď počasia obsahuje údaje o teplote vzduchu v dvoch metroch nad povrchom, rýchlosti a smere vetra v desiatich metroch nad povrchom, celkovej oblačnosti vyjadrenej percentuálne (100% = úplne zamračené), relatívnej vlhkosti v dvoch metroch nad povrchom vyjadrenej percentuálne, atmosférickom tlaku redukovanom na hladinu mora a globálnom horizontálnom žiarení (W/m2). Z týchto údajov považujeme za relevantné pre predpoveď hodnoty globálneho horizontálneho žiarenia, teploty vzduchu, rýchlosti vetra, celkovej oblačnosti a relatívnej vlhkosti.

Dáta o produkcii FVE majú niekoľko rozdielností oproti dátam z modelu Aladin. Tieto dáta sme preto museli upraviť tak, aby sme dvojicu dát: predpoveď počasia – produkcia FVE. Prvým rozdielom je, že záznamy z FVE sú uložené s časovým záznamom v CET (Central European Time) a v CEST (Central European Summer Time), teda v centrálnom európskom čase s posunom na letný čas. V praxi to znamená, že sme časové údaje museli posunúť o jednu, respektíve dve hodiny, aby boli v UTC, rovnako ako záznamy z modelu Aladin.

Druhým rozdielom je, že záznamy z FVE sú intervalové záznamy s pätnásť minútovou periódou, teda záznam z 12:00 vyjadruje produkciu FVE za časové obdobie od 11:45 do 12:00. Preto sme tieto záznamy zoskupili do hodinových intervalov a hodnoty o produkcii FVE sme sčítali. Pre najlepšie vytvorenie dvojíc predpoveď počasia – produkcia FVE sme spolu zoskupili záznamy z 11:45, 12:00, 12:15 a 12:30 do záznamu s časom 12:00. Takto môžeme zdvojiť predpoveď počasia s produkciou FVE, kde je predpoveď počasia platná pre okamih v polovici časového intervalu produkcie FVE.

V záznamoch o produkcii FVE je veľa hodnôt, no pre nás sú zaujímavé len dve hodnoty: výkon a energia (práca). Spočiatku sme z dát vyberali hodnoty výkonu, no po rade nášho kolegu Simona Sudoru sme vybrali zo záznamov hodnoty vyprodukovanej energie. Ako nám upresnil Ing. Peter Janiga, PhD. z Fakulty elektrotechniky a informatiky, hodnoty výkonu nie sú merané hodnoty, ale sú vypočítané z hodnôt vyprodukovanej energie a vyjadrujú priemerný výkon za daný časový interval. V dôsledku tak pracujeme priamo s nameranými hodnotami vyprodukovanej energie v kWh, podobne ako sa pri predikcii spotreby energie pracuje s hodnotami spotrebovanej energie. Hodnoty vyprodukovanej energie boli v záznamoch uložené aditívne (každým záznamom sa hodnota zvýšila o množstvo vyprodukovanej energie za daný časový interval), preto sme pri spracovávaní záznamov vypočítali rozdiel v hodnote vyprodukovanej energie oproti predchádzajúcemu záznamu.

Slnečnými koordinátmi nazývame údaje opisujúce polohu slnka na oblohe. Sú nimi azimut a elevácia. Azimut je hodnota veľkosti uhlu medzi priamkou smerujúcou k severnému magnetickému pólu a priemetom priamky slnečných lúčov do roviny zemského povrchu v danom miesto. Tento uhol je orientovaný v smere hodinových ručičiek. Elevácia zácisí od výšky Slnka na oblohe a opisuje uhol medzi rovinou zemského povrchu a priamkou vodorovnou s dopadajúcimi slnečnými lúčmi na túto rovinu. Elevácia teda vyjadruje uhol dopadajúcich slnečných lúčov, čo je dôležitý faktor pri FVE. Pri východe a západe Slnka má elevácia nulovú hodnotu. Dĺžka dňa je pre nás dĺžka slnečného svitu, teda počet hodín počas jedného dňa, kedy je Slnko nad obzorom.

Elevácia je priamo úmerná s hodnotou GHI počas jasného dňa a dĺžka dňa je priamo úmerná so vzdialenosťou dňa od zimného slnovratu a teda so zmenou maximálnej hodnoty elevácie v rámci dňa aj priemernou teplotou vzduchu. Preto sú obe tieto hodnoty relevantné pri predikcii produkcie FVE. Azimut ale nemá pre nás podobnú výhodu a bolo to vidieť aj na výsledkoch experimentu, kedy sme ho pri predikcii použili. Z tohto dôvodu sme ho ďalej pri predikcii nepoužívali.

### Import dát do databázy

Rozhodli sme sa vložiť dáta do relačnej databázy pre uľahčenie výberu dát pre trénovaciu množinu. Konkrétnou databázou je PostgreSQL, ktorá je najpoužívanejšou „open source“ databázou.

Pri spracovaní dát sme narazili na niekoľko problémov v dátach z FVE, ktoré sme museli vyriešiť. Prvým problémom bolo formátovanie *csv* súborov, v ktorých sú dáta uložené. Súbory mali za hlavičkou dva nadbytočné riadky a na oddelenie desatinných miest je použitá čiarka, pričom pre kompatibilitu s databázou preferujeme použitie bodky. V týchto súboroch je aj veľký počet stĺpcov dát, ktoré nepotrebujeme. Pre tieto dôvody sme sa rozhodli dáta najskôr predspracovať v programovacom jazyku R a nie priamou funkcionalitou databázy ako v prípade spracovania dát predpovede počasia z modelu Aladin. Dôsledkom takéhoto spracovania dát je dlhšia doba, ktorú celkové importovanie dát trvá, ktorá sa približuje k štyrom minútam.

Druhým problémom boli rozdielnosti v dátach opísané vyššie, teda časové záznamy v CET a CEST na rozdiel od UTC, intervalové záznamy na rozdiel od bodových a 15-minútová perióda záznamov na rozdiel od hodinovej periódy.

Tretím problémom bola absencia niektorých záznamov, respektíve absencia dát v záznamoch. Niektoré záznamy chýbali úplne a vytvorili v dátach „dieru“. Chýba v nich niekoľko dní, resp. hodín. Tento problém sme sa rozhodli vyriešiť odstránením dát za celé tieto dni, pretože neexistuje spôsob, ako by sme mohli dané dáta nahradiť a ich odstránením prídeme len o malé množstvo dát.

K dátam sme pridali údaj o počte dní od, respektíve do najbližšieho zimného slnovratu, ktorým sme substituovali uhol dopadajúceho slnečného žiarenia. Neskôr sme ale túto hodnotu nahradili skutočnými hodnotami slnečných koordinátov a dĺžku dňa, ktoré sme opísali vyššie. Tieto hodnoty sme vypočítali v R použitím funkcií balíku *insol*.

Z databázy sme následne odstránili aj dáta, ktoré sme neodstránili v dôsledku problémov v dátach ale za účelom zmenšenia celkového objemu dát o záznamy, ktoré nepotrebujeme. Boli nimi záznamy, ktoré nemali v dátach svoju dvojicu *predpoveď počasia – produkcia FVE* a záznamy, kde bolo v tejto dvojici predpovedané GHI aj výsledný výkon FVE, resp. výsledná práce, nulové, teda dáta z času, kedy na fotovoltaické panely nedopadalo slnečné žiarenie a panely neprodukovali žiadnu elektrickú energiu, respektíve po doplnení dát slnečných koordinátov sme vymazali záznamy, kde bola hodnota elevácie menšia ako nula, čo sú záznamy z času, kedy slnko nebolo nad obzorom a teda žiadne slnečné žiarenie na fotovoltaické panely nedopadalo.

V databáze sme pripravili aj tabuľku na ukladanie dát o experimentoch, kde môžeme uložiť údaje o nastavení predikčného modelu, spôsobu výberu trénovacej množiny a výsledkoch predikcií vo forme štatistických metrík presnosti predikcie za daných podmienok.

V technickej dokumentácii sú zdrojové kódy, na ktorých je vidieť štruktúra dát uložených v databáze v jednotlivých tabuľkách a ich prepojenie cudzími kľúčmi, funkcie pre spracovanie údajov a funkcie pre vymazanie nepotrebných údajov v jazyku SQL a skript v jazyku R, ktorý dané funkcie volal a vykonáva celkové importovanie dát do databázy.

## Prvé predikcie

Po tom, ako sme importovali dáta do databázy, mohli sme začať implementovať skripty pre predikciu v jazyku R. V skripte sa dotiahnu potrebné dáta z databázy, vykoná sa predikcia a štatistické výsledky presnosti predikcie sa uložia do databázy aj s údajmi o nastavení predikčného modelu aj o výbere záznamov do trénovacej množiny. Pre prácu s databázou sme použili balík *RPostgreSQL* a pre výpočet štatistiky presnosti hlavne funkciu *modeval* z balíku *sirad*. Výhodou práce v jazyku R je, že podporuje „open source“ vývoj balíkov a tak mnoho používateľov R vytvorilo vlastné balíky, ktoré sprístupnili pre používanie. Vďaka tomu sme nemuseli mnohé výpočty sami implementovať a mohli sme použiť dostupnú funkcionalitu balíkov, čo zjednodušilo v mnohých smeroch našu implementáciu na volanie funkcií z týchto balíkov.

Pre predikciu sme použili umelú neurónovú sieť z balíku *neuralnet* ako sme plánovali, a keďže nás pri analýze zaujala aj metóda náhodného lesa regresných stromov, rozhodli sme sa ju použiť pre porovnanie. Použitie oboch predikčných metód po naštudovaní dokumentácie a dostupných návodov na použitie bolo v jazyku R jednoduché. Pre náhodný les regresných stromov sme použili balík *randomForest*.

Pre testovanie presnosti nastavení predikčných modelov sme vytvorili skripty, ktoré predikovali produkciu FVE pre všetky dni vybranej jednej elektrárne s rôznymi kombináciami parametrvo nastavení predikčných modelov. Pre neurónovú sieť sme nastavovali veľkosť trénovacej množiny, kombinácie počtu skrytých vrstiev a počtu neurónov na jednotlivých vrstvách a algoritmus trénovania. Pre NLRS sme nastavovali veľkosť trénovacej množiny, počet vytvorených stromov, počet uzlov stromu a počet náhodne vybraných prediktorov pre vytvorenie uzla stromu. Pri obidvoch predikčných metódach sme testovali aj rozdielnosť výsledkov pre rôzne kombinácie prediktorov. Napríklad prvé skripty predikovali len podľa hodnôt GHI, teploty vzduchu a rýchlosti vetra.

Už pri prvých skriptoch bolo vidieť, že trénovanie neurónovej siete trvá dlhšie než vytvorenie náhodného lesa a že štatistické výsledky presnosti predikcie ukazujú na presnejšiu predikciu použitím NLRS oproti UNS. Pri UNS sme navyše narazili na chyby, ktoré sme nespôsobili sami, ale po analýze chybových hlášok a prečítaní viacerých diskusií na webových fórach sme usúdili, že chyba je v implementácií balíka *neuralnet*. Otestovali sme aj implementáciu UNS v balíku *nnet*, no pri tejto implementácií sme mohli vytvoriť neurónovú sieť s len jednou skrytou vrstvou perceptrónov, takže výsledky tiež neboli presnejšie než výsledky predikcie s použitím NLRS.

Po neúspechu s UNS sme sa rozhodli zostať pri používaní náhodného lesa regresných stromov, čo je aj veľmi logické, pretože podľa štatistických ukazovateľov dosahoval tento predikčný model menšiu chybu predikcie a predikcia trvala kratší čas. S odstupom času sme prišli aj na to, že použitie NLRS je pre nás vhodnejšie, v skutku asi najvhodnejšou predikčnou metódou vzhľadom na nami navrhnutý výber dát do trénovacej množiny, pretože môžeme vyberať najpodobnejšie záznamy do trénovacej množiny, ktorej veľkosť môže byť malá (len 30 záznamov) a nemusíme sa obávať takzvaného *pretrénovania* ako u UNS, pretože niečo podobné u NLRS nie je možné.

## Výber trénovacej množiny

Pre zvýšenie presnosti predikcie sme v našom riešení navrhli výber záznamov do trénovacej množiny podľa ich podobnosti so záznamom, pre ktorý predikujeme produkciu elektrickej energie. Produkcia FVE vo veľkej miere závisí od meteorologických podmienok a hodnoty vyprodukovanej energie sa výrazne líšia aj pri malých rozdieloch v hodnotách premenných opisujúcich meteorologické podmienky. Preto veríme, že trénovanie predikčných modelov na záznamoch, ktoré opisujú výrazne odlišné meteorologické podmienky má viac negatívny vplyv na presnosť predikcie než pozitívny. Tak ako je evidentný rozdiel v produkcii FVE pri jasnom a zamračenom stave oblačnosti, pri nízkej (ráno alebo večer) a pri vysokej (poludnie) elevácii Slnka na oblohe, alebo pri porovnaní produkcie v lete a v zime, tak výrazná musí byť aj chyba predikcie pri trénovaní predikčného modelu na záznamoch odlišných od záznamu, pre ktorý má predikovať výslednú hodnotu vyprodukovanej energie.

Naopak, ak natrénujeme predikčný model na záznamoch podobných so záznamom, pre ktorý má predikovať, zúžime tak rozsah intervalu možných výstupných hodnôt predikcie a pri viac homogénnej trénovacej množine, môže trénovanie predikčného modelu byť cielenejšie a „detailnejšie“ alebo „jemnejšie“.

Predpokladáme, že takáto trénovacia množina umožní UNS presnejšie simulovať lineárne a nelineárne súvislosti medzi vstupnými prediktormi a výstupnou hodnotou, a pri NLRS umožní vytvoriť regresné stromy, ktorých priemerná výstupná hodnota bude bližšia skutočnej hodnote, ktorú sa snažíme predikovať a rozdiel medi skutočnou a predikovanou hodnotou nebude nikdy tak veľký ako môže byť veľký bez použitia výberu najpodobnejších záznamov do trénovacej množiny.

Klasickým postupom práce pri predikcii metódami strojového učenia je rozdelenie dát na trénovaci a testovaciu množinu, ako sme vysvetlili v kapitole 2.2.5 Meranie presnosti metód strojového učenia. Na záznamov v trénovacej množine sa natrénuje predikčný model a jeho presnosť sa testuje na záznamoch v testovacej množine. Keďže chceme predikovať pre každý záznam čo najpresnejšie, v našom riešení vyberieme novú trénovaciu množinu a natrénujeme nový predikčný model pre každý záznam, pre ktorý predikujeme. Rozdeľujeme tak dáta na množinu, ktorá obsahuje záznamy z dňa kedy bol vytvorený záznam, pre ktorý predikujeme, a na množinu obsahujúcu všetky ostatné záznamy, ktoré nie sú v prvej množine. Z tejto druhej množiny záznamov potom vyberáme potrebný počet záznamov do trénovacej množiny.

Produkcia FVE nezávisí od produkcie v predchádzajúcich hodinách alebo dňoch, ale závisí od meteorologických podmienok, ktoré samotné majú túto závislosť. Hodnoty meteorologických premenných však pre nás predpovedal model Aladin a my môžeme obmedziť predikciu iba na závislosť od aktuálnych hodnôt meteorologických premenných. Keďže v našom riešení nezávisí predikcia od predchádzajúcich hodnôt a postupnosti záznamov, nepracujeme s časovým radom, ako to býva pri iných predikciách (predikcia spotreby elektrickej energie). Pretože nepracujeme s časovým radom, stávajú sa v našom riešení dátum a čas záznamov irelevantné až na jedinú podmienku, že nemôžeme trénovať predikčný model na záznamoch z toho istého dňa, pre ktorý predikujeme hodnotu vyprodukovanej energie, pretože tieto záznamy by ani v skutočných podmienkach neboli prístupné. V trénovacej množine sa ale môžu vyskytnúť historické záznamy z dní, ktorých dátum je neskorší než dátum záznamu, pre ktorý predikujeme. Uvedomujeme si, že tieto záznamy by v skutočných podmienkach tak isto neboli prístupné, no rozhodli sme sa ich používať z dôvodu nedostatku historických záznamov. Máme historické záznamy pre tri fotovoltaické elektrárne a po odstránení chybných záznamov máme dostupných 336 dní pre prvú, 334 pre druhú a 449 pre tretiu. Keby sme dodržiavali časovú postupnosť dátumov, a netrénovali predikčný model na záznamoch z dní neskorších než deň, pre ktorý predikujeme, nemali by sme dostatok dát pre výber dostatočne podobnej trénovacej množiny aj pre meranie štatistiky presnosti predikcie s dostatočnou výpovednou hodnotou. Keďže ale predikcia produkcie FVE v našom riešení nezávisí od časovej postupnosti záznamov, môžeme s historickými záznamami pracovať len ako so záznamami opisujúcimi meteorologické podmienky, ktoré mohli nastať v ktorýkoľvek deň alebo rok aj v minulosti a kompenzujeme tým nedostatok historických záznamov. Berieme na vedomie, že toto riešenie nedostatku dát nie je teoreticky úplne správne, pretože v skutočnosti neexistujú záznamy „z budúcnosti“, ale v praxi majú prevádzkovatelia fotovoltaických elektrární k dispozícií viac historických záznamov než 334 alebo 449 dní, ak samozrejme nie je samotná FVE funkčná menej ako 334, respektíve 449 dní. Na naše výsledky sa môžeme pozerať ako na presnosť predikcie FVE po tom, čo má FVE k dispozícií aspoň 334, respektíve 449 dní historických záznamov.

### Výber záznamov podľa podobnosti.

Ako sme už viac krát napísali, do trénovacej množiny vyberáme najpodobnejšie záznamy so záznamom, pre ktorý chceme predikovať produkciu FVE. Vytvorili sme jednoduchý spôsob ohodnotenia podobnosti záznamov. Pre každý záznam v množine záznamov, z ktorých vyberáme záznamy do trénovacej množiny, vypočítame jeho hodnotu podobnosti, respektíve rozdielnosti, pretože číselná hodnota je väčšia je väčšia pre viac rozdielne záznamy a najpodobnejšie záznamy majú hodnotu blízku nule. Podľa tejto hodnoty podobnosti všetky ohodnotené záznamy usporiadame vzostupne a vyberieme prvých *n* záznamov do trénovacej množiny.

Naše prvé experimenty s výberom najpodobnejších dát hodnotili záznamy podľa podobnosti v hodnote GHI, ktorá je najpodstatnejšou pri produkcii FVE. Náš výpočet hodnoty podobnosti vysvetlíme v nasledujúcich odsekoch.

Máme množinu potenciálnych záznamov () o veľkosti na výber záznamov do trénovacej množiny. Máme záznam , pre ktorý predikujeme produkciu FVE za meteorologických podmienok, ktoré opisuje. Máme škálu GHI (), ktorá je absolútnou hodnotou rozdielu maximálnej a minimálnej hodnoty GHI v množine :

Hodnotu rozdielnosti každého záznamu z množiny , kde ohodnotíme nasledovne. Vypočítame absolútnu hodnotu rozdielu v hodnote GHI medzi a .Túto hodnotu normalizujeme na škálu GHI ():

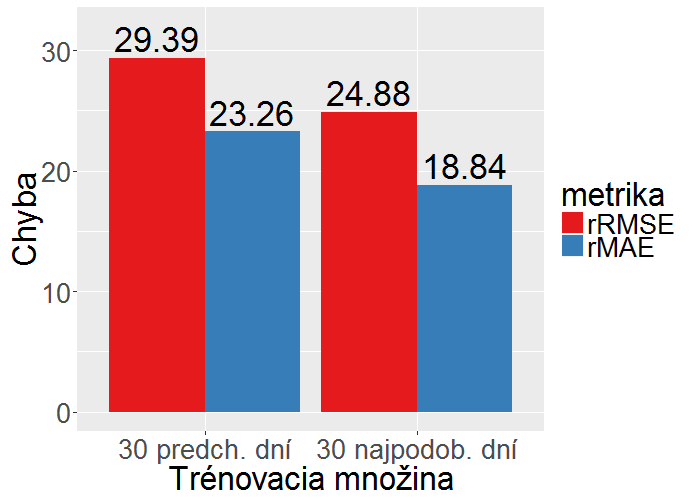
Teraz vieme usporiadať záznamy v množine podľa podobnosti so záznamom . Hodnotiť podobnosť záznamov iba podľa hodnoty GHI je nepostačujúce. Ak chceme hodnotiť podobnosť záznamov podľa viacerých hodnôt a chceme aby každá hodnota mala na podobnosť rôzny vplyv, tak ako aj produkcia FVE závisí rozdielne od týchto hodnôt, musíme vyjadriť tento vplyv a zakomponovať ho do výpočtov. Pre tento účel sme vytvorili *faktory podobnosti*. Faktor podobnosti GHI vyjadruje silu vplyvu podobnosti záznamov v hodnote GHI na celkovú podobnosť záznamov.

Pre ďalší experiment sme sa rozhodli hodnotiť podobnosť záznamov podľa GHI, teploty vzduchu a rýchlosti vetra s faktormi podobnosti 90 pre GHI, 10 pre teplotu vzduchu a 1 pre rýchlosť vetra. Tieto hodnoty sme zobrali z kapitoly 1.2 Úvod do predikcie produkcie FVE, kde píšeme že presnosť predikcie produkcie FVE závisí od presnosti výpočtu GHI na úrovni 90%, od teploty vzduchu na úrovni 10% a od rýchlosti vetra na úrovni 1%. Pri výpočte hodnoty rozdielnosti jednotlivé rozdielnosti po hodnotách GHI, teploty vzduchu () a rýchlosti vetra () vynásobíme ich faktorom podobnosti a tieto hodnoty sčítame:

Rovnako vieme ohodnotiť podobnosť záznamov aj podľa všetkých prediktorov.

Musíme ešte zmieniť, že hodnota faktorov podobnosti nemá výpovednú hodnotu sama o sebe, ale dôležitý je pomer faktorov podobnosti.

Už výberom najpodobnejších dát do trénovacej množiny podľa vyššie uvedeného príkladu s GHI, teplotou vzduchu a rýchlosťou vetra sa nám podarilo dosiahnuť nižšiu chybu predikcie oproti klasickému spôsoby výberu niekoľkých predchádzajúcich záznamov. Pri tomto experimente sme predikovali produkciu FVE za celý deň (hodinové záznamy boli sčítané). Pri klasickom spôsobe sme do trénovacej množiny vybrali 30 predchádzajúcich dní a pri výbere trénovacej množiny podľa podobnosti sme vybrali 30 najpodobnejších dní. Výsledky sú vyobrazené na nasledujúcom grafe (viď Obrázok 5), kde môžeme porovnať chybu predikcie pomocou metrík *rRMSE* a *rMAE*, ktoré sme vysvetlili v kapitole 3 Metriky presnosti predikcie*.*



Obrázok 5: Porovnanie chyby predikcie poľa spôsobu výberu trénovacej množiny

## Nastavenia predikčného modelu a faktorov podobnosti

Hladanie najpresnejsich nastaveni predikcneho modelu

Hladanie najpresnejsich nastaveni – faktorov podobnosti

Tu dlhe skripty ale podarilo sa zrychli – kalitola xx

Predikcia po dnoch / hodinach

## Experimenty

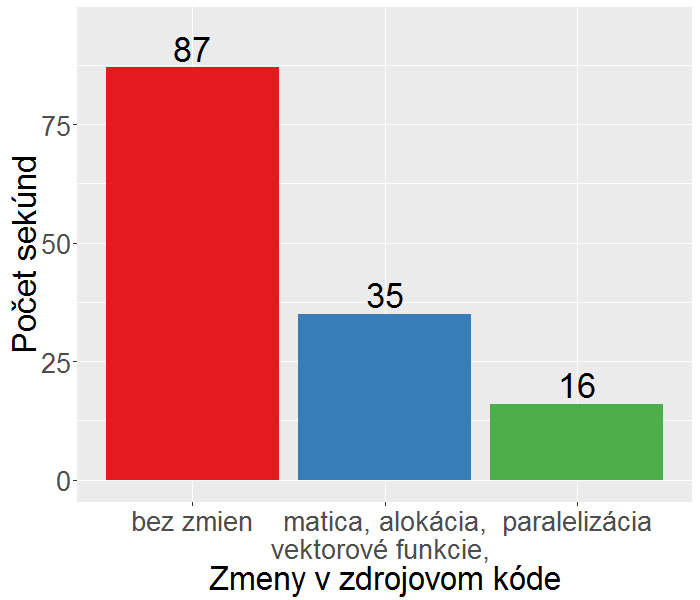
Experiment s leto/zima

Zúženie množiny dat s views

## Najpresnejšie výsledky

## Zrýchlenie predikcie v R

Keďže skripty na testovanie presnosti predikcie použitím rôznych kombinácií nastavení predikčného modelu a rôznych kombinácií faktorov podobnosti pri výbere záznamov do trénovacej množiny trvali príliš dlho, hľadali sme spôsob ako vykonávanie týchto skriptov zrýchliť. Christenson a Morris (15) merali efektívnosť zrýchlenia vykonávania jazyka R na strojoch s operačným systémom Windows. Podľa ich meraní má na 64-bitovom operačnom systéme Windows XP použitím matíc namiesto tabuliek (data.frame) dosiali v priemere zrýchlenie o hodnote 97,76%, predbežnou alokáciou pamäte po blokoch (na rozdiel od postupného alokovania pamäti) 89,73% a paralelizáciou na štyroch jadrách 49,12%.

Rozhodli sme sa preto tiež aplikovať tieto zmeny na náš zdrojový kód aj s použitím vektorových funkcií namiesto cyklov. Podarilo sa nám zrýchliť beh skriptov viac ako päť násobne. Na nasledujúcom grafe (viď Obrázok 6) je vidieť priemerný čas behov skriptov pred aplikovaním zmien zdrojového kódu, po použití matice namiesto tabuľky, predbežnej alokácií pamäte a použití vektorových funkcií namiesto cyklov, a po následnej paralelizácii. Rozdiel po aplikácií prvej sady zmien boli pochopiteľne väčšie pri dlhších skriptoch, ktoré pracovali s väčším obsahom dát. Graf znázorňuje výsledky zrýchlenia na kratšom testovacom skripte.

Obrázok 6: Porovnanie časov vykonávania skriptu.

Naše výsledky sa samozrejme líšia od výsledkov Christensona a Morrisa a sú individuálne pre každý stroj s inou špecifikáciou. My sme tieto výsledky dosiahli na operačnom systéme Windows 10 (64-bit) a procesore Intel Core i5-2450M CPU s frekvenciou 2.50 GHz s dvomi jadrami a štyrmi vláknami (ang. thread).

## Pokračovanie

Naša práca by mohla pokračovať tromi spôsobmi s cieľom zlepšiť presnosť predikcie:

1. Chyba predikcie by mohla byť nižšia použitím inej predikčnej metódy, než je náhodný les regresných stromov (napríklad support vector machine). Pre vyskúšanie iných predikčných metód a nájdenie najlepších nastavení predikčných modelov by sme ale potrebovali veľa procesorového času a predpokladáme len malý potenciál na zlepšenie presnosti výslednej predikcie oproti NLRS, pretože ako sme sa dozvedeli pri analýze metód umelej inteligencie používaných pri predikcii, NLRS dosahuje veľmi dobré výsledky v porovnaní s inými metódami ako umelá neurónová sieť a support vector machine. (2.2.4 Náhodné lesy klasifikačných alebo regresných stromov).
2. Výber najpodobnejších záznamov do trénovacej množiny zvyšuje presnosť predikcie, no nami implementovaný výber týchto záznamov je pomerne jednoduchý, kedy číselne ohodnocujeme záznamy podľa ich podobnosti, respektíve rozdielnosti so záznamom, pre ktorý predikujeme výslednú hodnotu vyprodukovanej elektrickej energie. Spôsob výberu najpodobnejších záznamov do trénovacej množiny by ale mohol byť sofistikovanejší napríklad použitím *zhľukovania (ang. clustering)* alebo iných metód používaných na analýzu dát. Lepším výberom dát do trénovacej množiny by chyba predikcie mohla klesnúť, ale predpokladáme len mierne zníženie celkovej chyby predikcie.

Že je druhou najpresnejšou som vtedy pocul, cital na nete.

1. Najväčší potenciál vidíme v aplikovaní postprocessingu. Úprava výsledkov predikcie podľa štatistiky presnosti by mohla výrazne znížiť celkovú chybu predikcie, no na implementáciu postprocessingu by sme potrebovali veľa človekohodín a samotná implementácia by bola veľmi náročná. V kombinácií s predchádzajúcim návrhom na možné pokračovanie práce by bolo možné dosiahnuť veľmi hodnotnú presnosť predikcie.

Naša práca by mohla pokračovať aj iným smerom. Nami vytvorený predikčný model vždy trénujeme na dátach z jednej elektrárne a predikcia je platná len pre túto jednu elektráreň. Do predikcie nevstupujú údaje, ktoré charakterizujú samotnú elektráreň ako napríklad inštalovaný výkon elektrárne. Predikcia je úplne nezávislá od špecifikácie elektrárne a takáto predikcia je nepoužiteľná pre iné elektrárne. Zaujal nás preto nápad, že by do predikcie vstupovali aj parametre špecifikujúce schopnosť produkcie elektrárne ako inštalovaný výkon, efektivita produkcie elektrickej energie a maximálna produkcia za hodinu alebo deň. Takto by bolo možné predikovať pre rôzne elektrárne a trénovať predikčný model aj na dátach z iných elektrární. Pre možnosť takejto predikcie by sme ale potrebovali dáta z viacerých fotovoltaických elektrární s rôznou špecifikáciou, ktoré by nám poskytli dostatočné spektrum záznamov na vytvorenie takéhoto predikčného systému.

Nevieme odhadnúť presnosť takéhoto predikčného systému. Vstupné parametre charakterizujúce FVE a možnosť trénovať predikčný model na dátach z iných elektrární by mohli mať pozitívny aj negatívny vplyv na zmenu hodnoty chyby predikcie. Za predpokladu, že by bola chyba predikcie na prijateľnej úrovni by však univerzálnosť takéhoto predikčného systému bola veľmi prínosnou. Predikovať by sme tak mohli aj pre elektrárne s malou alebo žiadnou bázou vlastných historických záznamov. Zapojením viacerých FVE do takéhoto systému by sa rozširovala báza dostupných historických záznamov, čo by prispievalo ku kvalite univerzálnosti predikcie. V prípade úspešnej implementácie takéhoto predikčného systému s malou chybou predikcie by bol globálne veľmi užitočný a mal by veľký komerčný potenciál.

# Zhodnotenie

# Technická dokumentácia

Na nasledujúcich stranách sú zdrojové kódy použité pre implementáciu predikčného modelu. Zo zdrojových kódov sú vynechané časti, ktoré priamo nesúvisia s logickým alebo programovým obsahom. Kompletné zdrojové kódy sú na priloženom CD nosiči.

## Import dát do databázy

Nasledujúce skripty v jazyku R a funkcie v PL/pgSQL slúžia na importovanie dát do databázy a je na nich viditeľná štruktúra uložených dát.



















# Zdroje

1. **Morvová, Marcela.** *Princípy metód a využitie obnoviteľných zdrojov energie.* Bratislava : Knižničné a edičné centrum FMFI UK, 2008. ISBN 978-80-89186-28-0.

2. **Letendre, Steven, Makhyoun, Miriam and Taylor, Mike.** Predicting solar power production: irradiance forecasting models, applications and future prospects. s.l. : SEPA - solar electric power association, 2014.

3. **Fakulta elektrotechniky a informatiky, Slovenská technická univerzita v Bratislave.** *Analýza možnosti predpovedania výroby elektrickej energie z fotovoltických elektrárni.* Bratislava : s.n.

4. **Diagne, Maïmouna, et al.** Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews.* 2013, November 2013, pp. 65-76.

5. **Wei, William W. S.** *Time series analysis: univariate and uultivariate methods.* s.l. : Pearson, 2006. ISBN 0-321-32216-9.

6. **Sinčák, Peter a Andrejková, Gabriela.** *Neurónové siete: inžiniersky prístup (1. diel).* Košice : ELFA-press, 1996. ISBN 80-88786-38-X.

7. **Beňušková, Ľubica.** Umelé neurónové siete. [aut. knihy] Pavol Návrat, a iní. *Umelá inteligencia.* Bratislava : Vydavateľstvo STU, 2002.

8. **Veitch, David.** *Wavelet Neural Networks and their application in the study of dynamical systems.* Department of Mathematics, University of York. 2005. Dizertácia PhD.

9. **Maroco, João, et al.** Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Research Notes.* 2011, 4:299.

10. **Thissen, U., et al.** *Using support vector machines for time series prediction.* s.l. : Chemometrics and Intelligent Laboratory Systems, 2003.

11. **Breiman, Leo.** Random Forests. *Machine Learning.* 2001, 1, pp. 5-32.

12. **Liaw, Andy and Wiener, Matthew.** Classification and Regression by randomForest. *R News.* 2002, 3, pp. 18-22.

13. **Almeida, Marcelo Pinho, Perpiñán, Oscar and Narvarte, Luis.** PV Power Forecast Using a Nonparametric PV Model. *Solar Energy.* 2015, 115, pp. 354-368.

14. **Pelland, Sophie, et al.** *Photovoltaic and Solar Forecasting: State of the Art.* s.l. : International energy agency, 2013. ISBN 978‐3‐906042‐13‐8.

15. **Christenson, Dino P. a Morris, Joshua A.** A Note on Speeding Up R for Windows. *The Political Methodologist.* 2009, Zv. 17, 1, s. 4-8.