

# Seminarska naloga Gorenc in Primožič

Primož Gorenc in Marija Primožič

9/1/2021

## Contents

1. Predstavitev raziskovalnega problema/domneve . . . . .	1
2. Opis pridobitve podatkov in opis podatkov . . . . .	2
3. Opis statistične metodologije . . . . .	3
4. Predstavitev rezultatov . . . . .	7
5. Vprašanja ob članku in razumevanje metodologije . . . . .	11
6. Neznana statistična metodologija . . . . .	11
7. Možnosti za dodatno/drugačno analizo podatkov . . . . .	11
Viri in literatura . . . . .	12

## 1. Predstavitev raziskovalnega problema/domneve

Za najino seminarsko nalogo sva predelala članek Contribution of EMI and GPR proximal sensing data in soil water content assessment by using linear mixed effects models and geostatistical approaches, ki je bil napisan s strani Emanuele Bacca, Daniela De Benedetto in Anna Maria Stellaci. Članek se ukvarja z merjenjem razlik v vlažnosti talne povrhnjice oz po angleško topsoil water content spatial variation (. Ta tema postaja vedno bolj pomembna v sodobnem svetu, kjer raste pomen takoimenovanega natančnega kmetijstva oziroma precision farming, kjer z obdelavo podatkov iz meritev želijo maksimirati učinkovitost v kmetijstvu. Prav tako raste tudi pomen vode s vedno večjim pomanjkanjem le te ob vedno bolj pogostih sušah. Merjenje vlage v zemlji je zato ključne pomena pri optimizaciji namakanja v modernem kmetijstvu.

Tradicionalno se je to vzorčilo z direktnimi vzorci zemlje in laboratorijskimi analizami le teh. A ta vzorčenja so počasna, časovno intenzivna in tudi uničujoča do prsti. Poleg tega zahtevajo veliko količino laboratorijskega dela. Obstajajo pa tudi nove metode napovedovanja vlažnosti prsti s pomočjo novih metod vzorčenja. V članku se posvetijo dveh novim metodam elektromagnetni indukciji in zemeljski radar. Ti dve metodi nista invazivni in dajeta podatke o vlažnosti zemlje takoj z relativno nizkimi stroški.

Elektromagnetna indukcija in ostale naprave, ki merijo prevodnost zemlje so lahko uporabne za mapiranje prevodnosti tal. Ker je eden izmed glavnih prevodnikov v tleh voda je zato uporabna tudi za merjenje količine te, na podlagi meritev o prevodnosti tal. Ravno zaradi te možnosti so začeli raziskovalci uporabljati podatke iz elektromagnetne indukcije za merjenje vlažnosti prsti. (Truman, Manning, Sauer, & Schneider, 2016) Zemeljski radar je je nedestruktivna geofizična tehnika, z visoko učinkovitostjo merjenja vlažnosti prsti. Deluje s visokofrekvenčnim pulzom, katerega del se odbije nazaj do radarja. Hitrost potovanja pulza je odvisna od električne prevodnosti zemlje, ta pa pogjena z vlažnostjo, zato lahko tudi s tem radarjem merimo vlažnost prsti. (Xinbo, et al., 2019)

Raziskovalna problema s katerimi se soočajo raziskovalci v članku sta dva. Prvi je primerjava prej omenjenih metod merjenja vlažnosti z elektromagnetno indukcijo in z zemeljskim radarjem. Želijo preveriti če sta metodi ustrezni, koliko natančni dejansko sta, ter katera je boljša metoda. Primerjali bodo kako natančni so rezultati če izmerijo vlažnost samo z električno indukcijo, če izmerijo samo z zemeljskim radarjem in če uporabijo kombinacijo podatkov iz obeh. Drugi raziskovalni problem pa je katera izmed linearnih metod je najboljše za izdelavo za napovedovanje in mapiranje vlažnosti prsti. Tu bodo primerjali 3 različne metode in sicer navadni

kriging, linearni mešani model in kriging z zunanjim vplivom. Preverili bodo natančnost napovedovanja z vsemi temi metodami in jih med seboj primerjali po natančnosti napovedi.

## 2. Opis pridobitve podatkov in opis podatkov

Vsi podatki so pridobljeni iz poskusa. Ta se je odvijal na ravni podlagi dolžine 40 metrov in širine 20 metrov v eksperimentalni kmetiji CREA (Council for Agricultural Research and Economics) blizu italjanskega mesta Bari. Poskus je bil namenoma narejen na sorazmerno majhni površini, da se minimizirajo ostali faktorji razlik v prsti. Gre večinoma za glineno prst, ki lahko slabi signal, zato so pred pričetkom eksperimenta izmerili še upornost zemlje in slabljenje radarske energije. 116 različnih vzorcev je bilo pobranih z razmikom 4 metre in na teh vzorcih je bilo z gravimetrično metodo izračunana količina vode. Število vzorcev je bilo namenoma zelo veliko, zato da bi bili zagotovljeni optimalni pogoji za izbrane metode. Izbrani površini so takoj po vzorčenju izmerili vlažnost prsti še z elektromagnetno indukcijo in dvema različnima zemeljskima radarjema (eden s frekvenco 250 MHz in drugi z frekvencama 600 in 1600 MHz). S temi so šli čez celotno površino z razmikom enega metra med posameznimi meritvami. Na spodnji sliki je narisana shema za lažjo predstavo vzorčenja. Z rdečo piko so označene lokacije, kjer je bil pobran vzorec za gravimetrično metodo po zelenih črtah pa so se gibali z obema senzorjema (elektromagnetno indukcijo in zemeljskim radarjem).

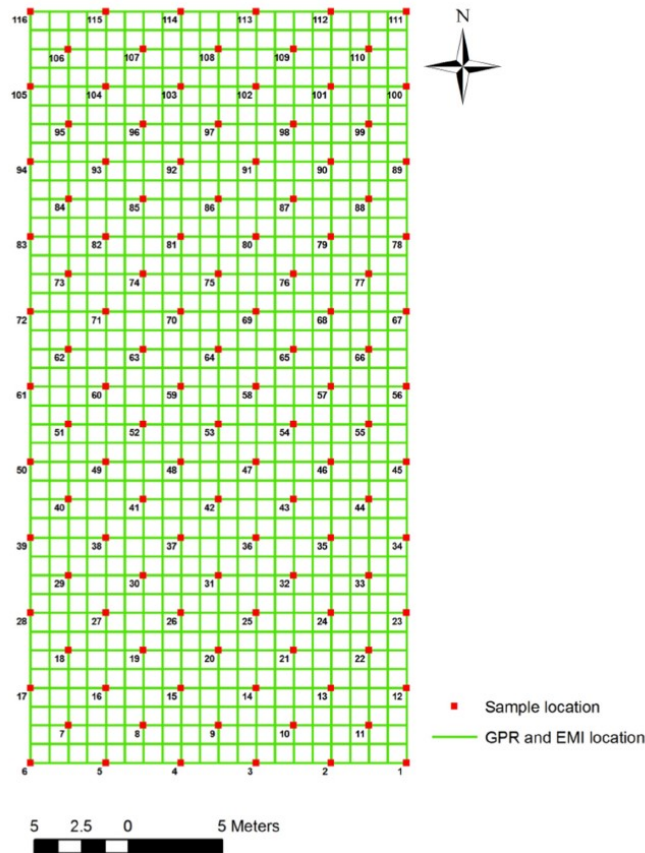


Figure 1: Shema vzorčenja

Podatki pridobljeni iz vseh teh meritev so o vlažnosti prsti. Definicija za vlažnost prsti

$$\theta = M_{water}/M_{soil} * 100$$

, kjer je izračunana vlažnost prsti,  $M_{water}$  pa masa vode v zemlji in  $M_{soil}$  masa prsti in vse skupaj pomnoženo s 100, da dobimo odstotke. (World Meteorological Organization, 2016)

Zato imajo podatki enoto  $100\text{g}^{-1}$  v dokumentu. Podatki o vlažnosti so bili zbrani z vzorci in meritvami z elektromagnetno indukcijo ter dvema različnima zemeljskima radarjema. Te izmerjene podatke o vlažnosti prsti bomo nato primerjali s pravimi in opazovali razlike.

### 3. Opis statistične metodologije

Preden so podatkom prirejali linearne modele ali pa geostatistične modele, so najprej seveda napravili deskriptivno statistično analizo, s katero so identificirali posamezne osamelce in si pregledali glavne značilnosti o porazdelitvi podatkov.

Kot že omenjeno so za samo raziskavo uporabili EMI in GPR, katerih pridobljene informacije so želeli razumeti tako ločeno, kot z obema skupaj. Podatkom o SWC in faktorjem, ki so jih pridobili z PCA, so vgradili različne modele:

- a) model, ki vključuje samo faktorje iz EMI podatkov
- b) model, ki vključuje samo faktorje iz GPR podatkov
- c) model vključuje faktorje tako iz GPR kot iz EMI podatkov.

Ničelno hipotezo so preverili u Saphiro - Wilk in Anderson-Darling statistiko. Samo razmerje med odzivno spremenljivko in faktorji, ki so jih pridobili s pomočjo PCA, pa so raziskali s Pearsonovim korelacijskim koeficientom.

#### 3.1 Uporaba modelov

Vsak model je bil analiziran z uporabo LMM(linearni mešani model) in z geostatističnimi tehnikami - kriging z zunanjim vplivom(KED - kriging with external drift), s katerimi so lahko primerjali učinkovitost posameznih pristopov. Prav tako pa so uporabili tudi navaden kriging, s katerim so ocenili SWC.

Z geostatistično analizo so želeli oceniti prostorsko heterogenost značilnosti prsti, tako, da so ustvarili mape iz zbranih podatkov. SWC(količina vode v zemlji) pa so ocenili z navadnim krigingom(OK) na različnih lokacijah - 2559 točk. Navadni kriging je ena najbolj osnovnih oblik kriginga, pri kateri je neznana vrednost( $z(x_0)$ )(od dane realizacije  $Z(x_0)$ ), ocenjena iz znanih vrednosti iz točk oz teh 2559 lokacij. torej  $z(x_i)$ ,  $i = 1, 2, \dots, N$ . Cenilko za OK lahko zapišemo kot:

$$z_{OK}(x_0) = \sum_{i=1}^N \lambda_i Z(x_i)$$

kjer  $\lambda_i$  so uteži povezane s temi N točkami oz. lokacijami. Te uteži so določene tako, da je ocena nepristranska, vrednosti zvezne in da je napaka ocene minimizirana:  $\min E[(Z^*(x_0) - Z(x_0))^2]$  pod pogojem za nepristranost ocene:

$$\sum_{i=1}^N \lambda_i = 1$$

To omogoči da je kriging natančen interpolator, saj so ocenjene vrednosti enake opazovanim vrednostim, ko se kriging lokacija sovpada z lokacijo vzorca. Pri navadnem metodi navadnega kriginga se predpostavi, da je konstantno povprečje vrednosti neznano(velja, razen, če ni znanstvenega razloga za zavrnitev te predpostavke).

#### 3.2 Natančnejše o krigingu

Sama metoda je dobila ime po južnoafriškem rudarskem inženirju Danileu Gerhardusu Kriegu, ki je razvil empirično metodo za določanje porazdelitve vsebnosti rude v kamnini na osnovi količine rudo v odvzetih vzorcih. Statistično teorijo prostorske interpolacije pa je leta 1963 zapisal francoski matematik Georges Matheron. Od takrat naprej se je metoda kriging uporabljala v vseh vedah, ki se ukvarjajo s prostorom. V sloveniji pa so jo

uporabljali v geologiji, meteorologiji in pedologiji. Delo Krigea pa je postavilo temelje v področju geostatistike. Kriging je ena najzahtevnejših metod, hkrati pa tudi najbolj uporabljana interpolacijska metoda, vendar za pravilno uporabo le - te potrebujemo veliko predznanja.

V geostatistiki izvajajo veliko meritev, dobljene vrednosti pa obravnavajo kot eno izmed možnih realizacij slučajnega procesa v prostoru. Opazovanja izvedejo na več lokacijah (rečemo na  $n$  lokacijah). Ta opazovanja predstavljajo delno realizacijo slučajnega procesa. Dobimo torej vrednosti na posameznih točkah nekega procesa, želimo pa tudi čimboljše oceniti vrednosti na vseh ostalih točkah in tako dobiti mapo nekega opazovanega prostora. Natančneje, na osnovi tega procesa naredijo statistični model za prostorsko interpolacijo vrednosti slučajne spremenljivke v poljubni točki prostora. Interpolirane vrednosti za vse točke, na katerih nismo izvedli meritev izračunamo kot linearno kombinacijo vrednosti meritve v izbrani okolici točke in jih imenujemo interpolirane vrednosti. Izračunati moramo le koeficiente linearne kombinacije, ki jih dobimo tako, da minimiziramo pričakovano vrednost kvadrata razlik med interpoliranimi in izmerjenimi vrednostmi, tako kot je zapisana formula zgoraj. Točno ta postopek imenujemo kriging.

Na spodnji sliki je prikazana ploskev, ki je pridobljena s to metodo.

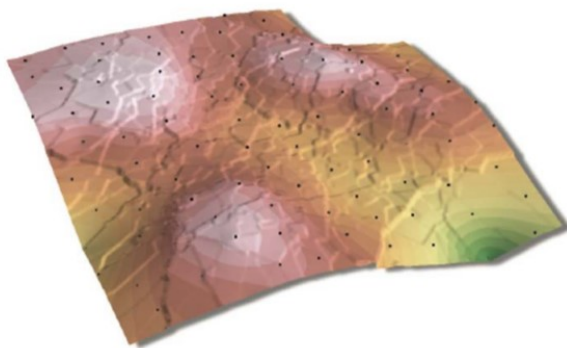


Figure 2: Ploskev, pridobljena z metodo kriging

Pri samem izračunu torej potrebujemo prostorsko spremenljivko  $Z(x)$ , z  $Z(x_i)$  pa označujemo opazovanja na različnih točkah oz. lokacijah. Če slučajna spremenljivka v prostoru ustreza določenim pogojem, potem lahko slučajno spremenljivko  $Z(x)$  zapišemo kot vsoto dveh delov:

- $Z(x) = \mu(x) + \delta(x)$ , kjer:
  - $Z(x)$  - prostorska spremenljivka
  - $\mu(x)$  - povprečje slučajne funkcije  $Z(x)$  v točki  $x$
  - $\delta(x)$  - slučajni del modela splošnega kriginga.

$\mu(x)$  predstavlja povprečje slučajne funkcije  $Z(x)$  v točki  $x$  in je ponavadi to funkcija lokacije,  $\delta(x)$  pa je kot rečeno slučajni del modela in ga želimo v postopku prostorske interpolacije s pomočjo variograma čim natančneje ovrednotiti. Torej variogram določimo za vrednost  $\delta(x)$ .

To je osnova za samo metodo kriging, za katero pa obstaja več vrst, ki se ločijo glede na to, kako definiramo povprečje  $\mu(x)$ .

Poznamo navadni kriging (OK), ki je bil uporabljen tudi pri ocenjevanju vsebnosti vode v prsti, splošni kriging, preprosti kriging in pa kriging z zunanjim vplivom (KED - kriging with external drift), ki so ga ravno tako uporabili in se mu v geostatistiki reče tudi kokriging.

Kriging metode torej določijo interpolirano vrednost kot utežno linearno premikanje povprečja vrednosti bližnjih točk. Pri določitvi uteži se mora upoštevati prostorsko odvisnost med podatki, ki se jo predstavi s semivariogramom. Izračun uteži sloni na pogoju o neprestranski oceni interpolirane vrednosti in minimalni ocenjeni variance na novi točki.

Kot zanimivost pa obstaja tudi okvir, kateri prikazuje potek prostorske napovedi za spremenljivke, ki se jih opazuje pri prsti. Temelji na kriginu.

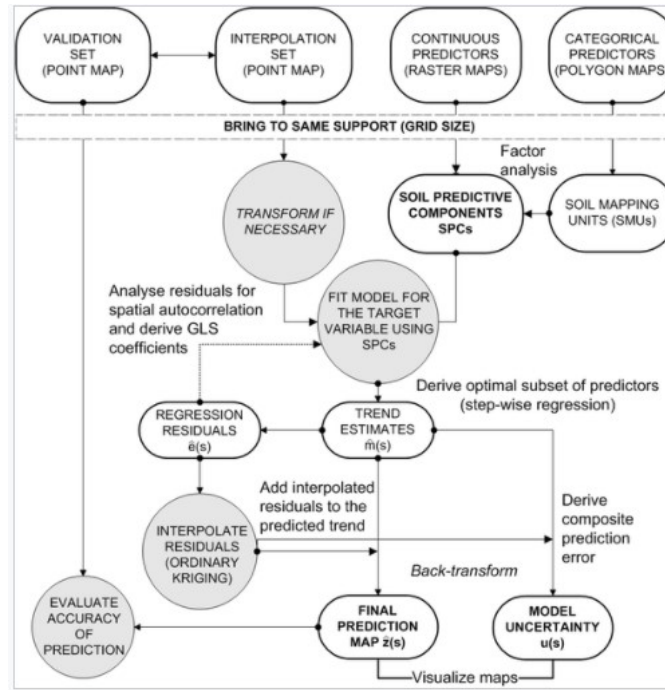


Figure 3: Postopek kriginja pri napovedovanju prostorskih spremenljivk

### 3.3 Linearni mešani model - LMM

LMM je posplošitev navadnega regresijskega pristopa, ki poleg običajnih fiksnih učinkov vključuje tudi slučajne učinke mu zato rečemo mešani linearni model. LMM zapišemo kot:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\alpha + \epsilon$$

, kjer je  $\mathbf{Y}$  proučevani slučajni vektor,  $\mathbf{X}$  matrika znanih konstant,  $\beta$  vektor neznanih regresijskih koeficientov, ki predstavlja fiksne učinke,  $\mathbf{Z}$  znana matrika,  $\alpha$  vektor slučajnih učinkov in  $\epsilon$  vektor slučajnih odstopanj.

- Veljajo še naslednje lastnosti:
- $E[\alpha] = 0$
- $var(\alpha) = \sigma_s^2 \mathbf{H}(\phi)$
- $E[\epsilon] = 0$
- $\alpha$  in  $\epsilon$  sta normalno porazdeljena, ne nujno enaka razsežna vektorja,  $\alpha \sim \text{Norm}(0, \sigma^2 I_n)$  in  $\epsilon \sim \text{Norm}(0, \tau^2 I_m)$
- $\alpha$  in  $\epsilon$  sta neodvisna slučajna vektorja.

Vemo, da je razlika z linearnim regresijskim modelom v členu  $\mathbf{Z}\alpha$ , ki lahko zavzame veliko različnih oblik in s tem dobimo veliko različnih modelov (ANOVA, longitudinalni, marginalni, ...)

Standardna metoda za ocenjevanje parametrov v LMM je metoda največjega verjetja. Ta metoda pri ocenjevanju variančnih komponent upošteva fiksne učinke modela kot znane vrednosti in ne upošteva izgube prostostnih stopinj, do katerih pride zaradi ocenjevanja fiksnih učinkov modela. Posledično lahko vrne pristranske cenilke. To sta dokazala Neyman in Scott, ki sta pokazala da v primeru naraščanja št. parametrov

z večanjem velikosti vzorca, cenilka ni nujno dosledna. Zaradi zgoraj omenjenih težav, se je razvila restringirana metoda največjega verjetja - RMNV, ki so jo tudi tukaj uporabili.

Zakaj torej RMNV? Ker omogoča pridobivanje cenilk zgolj za parametre, ki nas zanimajo.

V tej študiji so ocenili in primerjali različne LMM, ki so jih ocenili s funkcijo lme v knjižnici v R-ju.

### 3.4 Kriging z zunanjim vplivom( KED - Kriging with external drift)

Je posebna vrsta kriginga z nestacionarnimi pogoji, ki dovoljuje uporabo sekundarnih informacij za upoštevanje prostorskih spremenljivk. Te sekundarne spremenljivke so izbrane zaradi njihove pomembne korelacije s spremenljivkami, ki nas še posebej zanimajo. Tukaj je KED uporabljena kot notranja slučajna funkcija reda k: IRF-k(intrinsic random functions of order k).

Pri tej vrsti kriginga imamo zopet prostorsko spremenljivko  $Z$  in njeno vrednost v točki oz. na lokaciji  $x$ , ki jo izračunamo kot:

$$Z(x) = m(x) + Y(X)$$

, pri čemer je  $m(x)$  deterministična funkcija za zunanji vpliv in  $Y(x)$  je prostorska slučajna spremenljivka s povprečjem 0. Prva predpostavka je da je  $m(x) = \sum_{l=0}^k a_l f^l(x_\alpha)$ , kjer so  $a_l$  neznani koeficienti,  $f^l(x)$  pa so poznane funkcije koordinat( $x = (x,y)$ ). Običajno so  $f^l(x) = x^l$  monomi koordinat. Tudi  $m(x)$  lahko razbijemo na dve vsoti, ki se imenujeta notranji in zunanji vpliv.

V primeru z IRF - k, so podatki transformirajo s pomočjo linearnih kombinacij

$$Z(\lambda) = \sum_{\alpha=1}^N \lambda_\alpha Z(\mathbf{x}_\alpha)$$

, z naslednjo omejitvijo:

$$\sum_{\alpha=1}^N \lambda_\alpha f^l(x_\alpha)$$

, za  $l = 1, \dots, k$ , z  $x_0$  kot neznan ocenjevalno točko. Definirana  $Z(\lambda, x) = \sum \lambda_\alpha Z(x_\alpha + x)$ , če je  $Z(\lambda, x)$  stacionarna slučajna funkcija glede na  $x$ , potem je  $Z(x)$  notranja slučajna funkcija reda k. Korelacijska struktura, povezana z naključnim delom je izražena s posplošeno kovariančno funkcijo z razdaljo  $h$  med dvema opazovanjema  $x_\alpha$  in  $x_\beta$ ,  $K(x_\alpha - x_\beta)$ , označena kot  $K(\mathbf{h})$ . Druga predpostavka je, da je korelacijska struktura  $K(\mathbf{h})$  lahko izražena kot linearna kombinacija danega nabora znanih osnovnih struktur z nekimi pogoji za koeficiente. Posplošena kovarianca je polinomski model posplošene kovariance, s pogoji za koeficient b:

$$K(|\mathbf{h}|) = C_0 \delta(|\mathbf{h}|) - b_0 |\mathbf{h}| + b_1 |\mathbf{h}|^2 \log |\mathbf{h}| + b_2 |\mathbf{h}|^3$$

kjer  $\delta(|\mathbf{h}|) = 0$  za  $|\mathbf{h}| > 0$ , drugače pa  $\delta(|\mathbf{h}|) = 1$ . Koeficienti  $c_0, b_0, b_1$  in  $c_2$ , v dvorazsežnem prostoru morejo zadoščati naslednjim pogojem:

$$c_0 \geq 0, b_0 \geq 0, b_1 \geq 0, c_2 \geq \frac{3}{2} \sqrt{b_0 b_1}.$$

V notranji slučajni funkciji kriging je strukturna analiza narejena v dveh delih: najprej se določi red - k predmika in nato se oceni posplošeno kovarianco  $K(\mathbf{h})$  tako, da se priredi parametrični model. Za določitev stopinj odnašanja in najboljšo kombinacijo za pomožne spremenljivke, so po naraščajoči velikosti razvrščene napake najmanjših kvadratov za vsako točko posebej in za različne možnosti; najmanjši rang, povprečen za različne ciljne točke, se sovpada z optimalnim modelom za odnašanje. Uporabljeno je bilo tudi prečno preverjanje(ang. cross validation), da se je izbrala najbolj optimalna posplošena kovarianca. Tisti model, ki je imal standardizirano napako najbližje 1, je bil izbran za najbolj primerne.

### 3.5 Primerjava modelov

LMM modele so vrednotili s primerjavo o informiranosti, analizo ostankov in korelacijo med opazovano in napovedano vrednostjo vsebnosti vode v prsti z uporabo Pearsonovega korelacijskega koeficienta, kar so izračunali s pomočjo r-ja. Primerjavo KED modelov so napravili s prečnim preverjanjem. Izračunali so povprečno standardizirano napako, s katero so ocenili nepristranskost napovedi, prav tako pa je morala biti približno 0. S korenem povprečne kvadrirane standardizirane napake so ocenili natančnost povprečne kvadrirane napovedne napake, ki mora biti približno 1. Izračunali pa so še koren povprečne kvadrirane napake, s katero so preverili kako dobra je njihova napoved, pri katerem so bili boljši tisti, ki so imeli manjše vrednosti, saj to pomeni da so napovedane vrednosti bližje dejanski vrednostim. Da so izračunali stopnjo povezanosti med napovedanimi in opazovanimi vrednostmi, so izračunali Pearsonove koeficiente. Za primerjavo med različnimi pristopi k modeliranju napovedi (splošen kriging, LMM, in kriging z zunanjim vplivom), so uporabili izračune korena povprečne kvadrirane napake.

## 4. Predstavitev rezultatov

V spodnji tabeli 1 imamo predstavljene rezultate iz vzorcev, troje iz meritev z gravimetrično metodo. Povprečna vlažnost znaša  $24,89 \text{ g } 100 \text{ g}^{-1}$  in varira med 21 in  $27,09 \text{ g } 100 \text{ g}^{-1}$ .

Table 1: Podatki iz meritev

Spremenljivke	Vrednosti
število vzorcev	113.00
povprečna vlažnost	24.89
stand. odklon	1.11
mediana	24.95
minimum	21.00
maksimum	27.09

Spodaj sta naslikana še grafa porazdelitve rezultatov vlažnosti in zdi se, da je vlažnost normalno porazdeljena okoli povprečja. To je bilo preverjeno tudi z Anderson-Darling statistiko, katere p vrednost znaša 0,055 in s tem potrejuje približno normalno porazdelitev.

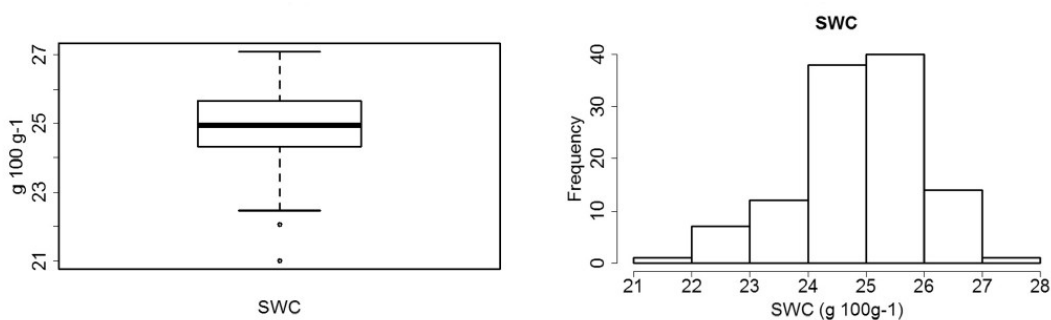


Figure 4: Grafa porazdelitve rezultatov

### 4.1 Primerjava rezultatov elektromagnetne indukcije, zemeljskega radarja in skupnega merjenja

Rezultati različnih meritev so bili primerjani z linearnimi mešanimi modeli, kjer so v raziskavi sestavili 3 modele: 1. model samo iz dejavnikov iz elektromagnetne indukcije, 2. model samo iz dejavnikov iz

zemeljskega radarja in 3. dejavniki iz obeh senzorjev. Rezultati so predstavljeni v tabeli 2. Za 1. model je znašal Akaike informacijski kriterij (merilnik za napako napovedi) 316,5, kar je bil najvišji izmed kriterijev, zato je ta model najslabše prilegal, kar vidimo tudi po najnižjem Piersonovem korelacijskem koeficientu. Pri 2. modelu, torej tem iz dejavnikov iz zemeljskega faktorja so štirje od šestih dejavnikov imeli pomemben učinek pri ocenjevanju vlažnosti prsti. Te dejavniki so bili obe dejavnika pri 250 MHz, prvi od treh dejavnikov pri 600 MHz in dejavnik pri 1600 MHz. Ta model ima nižji AIC in nekoliko višji korelacijski koeficient, zato je prileganje nekoliko boljše. Prvi 3. modeli pa imamo faktorje iz obeh senzorjev, najnižji AIC in najvišji korelacijski koeficient, zato je prileganje najboljše. Vsi trije modeli imajo aritmetično sredino 0 in standardni odklon blizu 1, prav tako njihova distribucija ni distveno drugačna od normalne. Tretji model, ki je vključeval faktorje iz obeh meritev je bil iz teh podatkov najboljši in je bil zato izbran za nadaljno analizo primerjav različnih načinov modeliranja, katere rezultati sledijo v naslednjem poglavju. Vsi trije modeli so bili vključeni tudi pri krigingu z zunanjim vplivom, kjer sta imela prva dva modela zelo podoben korelacijski koeficient, zopet pa je bil nekoliko boljši 3. model, kjer je bil koeficient korelacije največji tudi pri krigingu z zunanjim vplivom in je znašal 0,72. Tudi pri krigingu z zunanjim vplivom, je bil zato 3. model izbran za nadaljno analizo.

Table 2: Primerjava modelov iz meritev

vrsta modela	AIC	kor. koef. LMM	stand. odklon	kor. koef. KED
model iz elektromag. indukcije	316.5	0.51	0.9507	0.689
model iz zemeljskega radarja	302.1	0.657	0.8337	0.682
model iz obeh meritev	293.3	0.702	0.787	0.72

#### 4.2 Primerjava navadnega kriginga, linearne mešanega modela in kriginga z zunanjim vplivom

Kot rečeno primerjamo tudi različne linearne modele in sicer model navadnega kriginga (OK), linearne mešanega modela (LMM) in kriginga z zunanjim vplivom (KED). Z uporabo treh različnih modelov so bile dobljene 3 različne slike vlažnosti prsti. Te tri slike so prikazane skupaj na sliki Rezultati modelov OK, LMM in KED. Iz slik je razvidno, da izgleda slika z navadnim krigingom najbolj gladka in urejena, kar se zgodi zaradi karakteristik navadnega kriginga. Na vseh slikah pa vidimo, da severni oziroma še posebej severozahodni del površine je bolj vlažen kot ostali (tisti del kjer je več rdečih pik), južni del pa je nekoliko bolj suh.

Slike primerjav so imele tudi primerjavo standardnih odklonov (v prikeru krigingov) oziroma standardne napake (v primeru linearne mešanega modela), a na te je najbolj vplival efekt meje, torej zaradi manjšega števila merjenj v robovih površine je standardna napaka tam večja, zato teh slik nisva vključila v seminarsko, saj ne prikažejo najboljše podatkov. Poleg tega te napake oziroma odkloni niso povsem primerljivi med sabo.



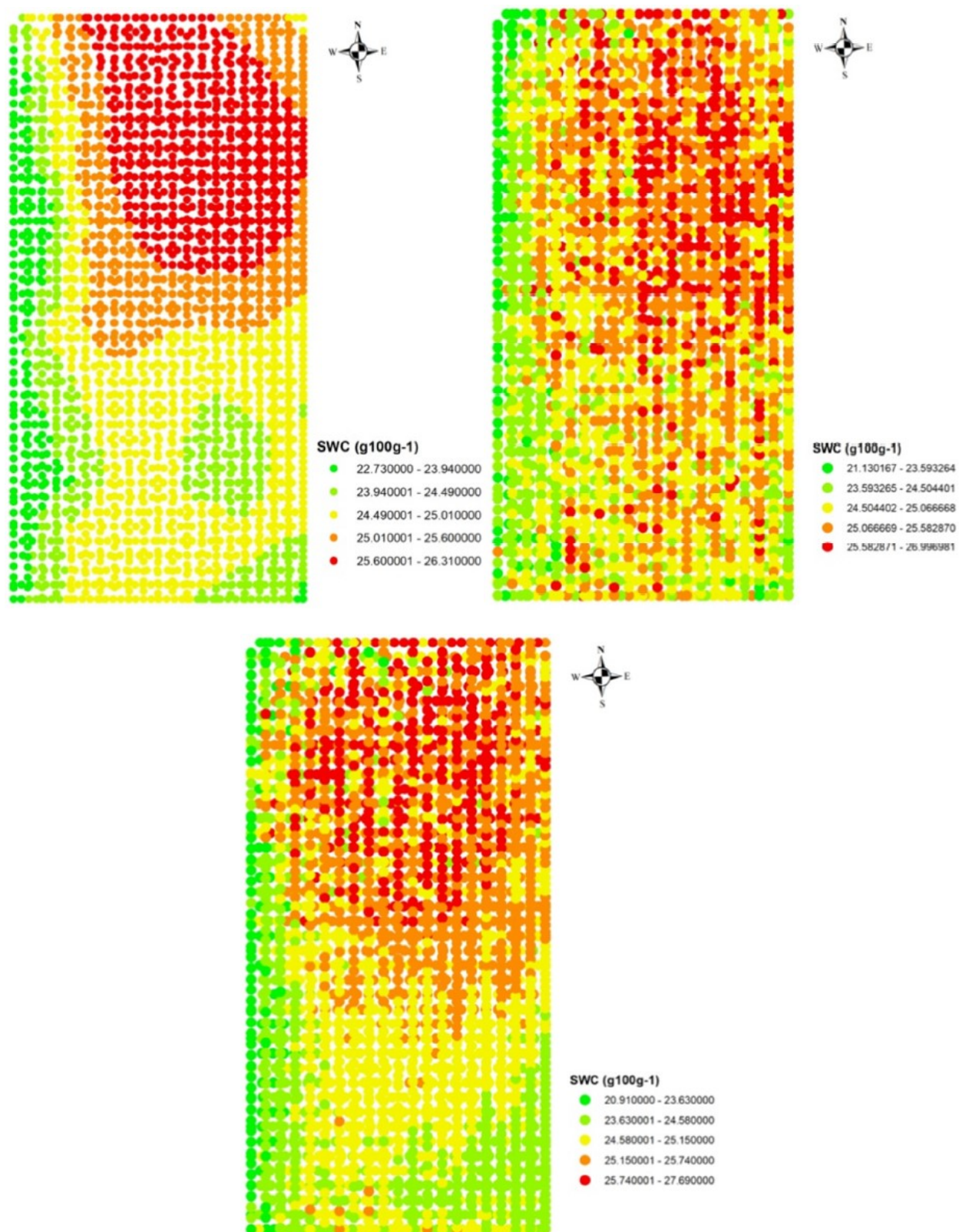


Figure 5: Rezultati meritev modelov OK, LMM in KED

### 4.3 Primerjava korelacije med modeli OK, LMM in KED

V raziskavi se je primerjala tudi korelacija med različnimi modeli. Tu je kriging z zunanjim vplivom bil najboljši model. To je bilo dokazano najprej z merjenjem korena povprečne standardne napake, ki je bil najnižji pri krigingu z zunanjim vplivom in je znašal 0,76. Pri navadnem krigingu je bila ta napaka nekoliko večja in sicer 0,83. Dodatno je bila preverjena tudi korelacija, kjer je imel KED tudi največji korelacijski koreficient, ki je znašal  $r=0.72$  in več točk je bilo v bližini 1:1 črte, kar lahko vidimo tudi na spodnji sliki, ki vse tri grafe primerja. Kriging z zunanjim vplivom je na diagramu b in je razvidno, da je korelacija nekoliko večja kot pri ostalih.

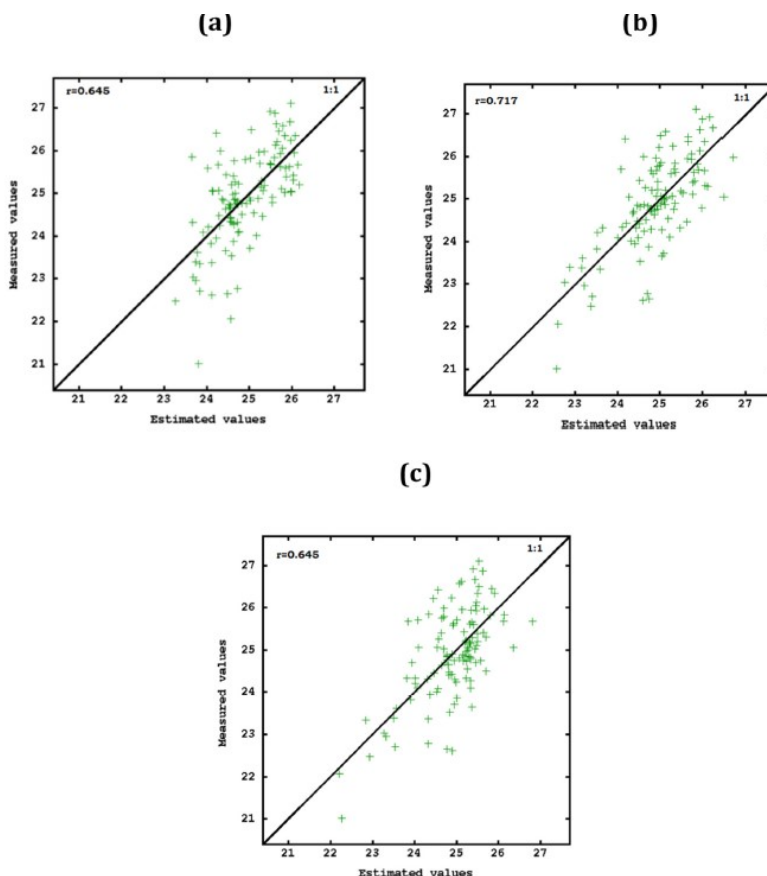


Figure 6: Primerjava razsevnih diagramov OK (a), KED (b) in LMM (c)

### 4.4 Sklepi

Iz rezultatov prejšnjih poglavij lahko sklepamo, da je najboljši model narejen iz skupnih dejavnikov iz meritev z elektromagnetno indukcijo in zemeljskega radarja skupaj. Skupni model je imel bolj natančno napoved vlažnosti prsti tako pri linearnem mešanem modelu kot pri krigingu z zunanjim dejavnikom. Sicer ta razlika ni bila velika glede na meritve le z eno napravo, a vseeno je razlika značilna.

Pri primerjavah med različnimi linearnimi modeli se je za najbolj natančnega izkazal kriging z zunanjim vplivom. Ta je še posebej boljši modeliral na podlagi podatkov iz elektromagnetne indukcije. Linearni mešani model vključuje obe kovariacije in prostorsko strukturo, zato je nepričakovano, da sta z navadnim krigingom enako učinkovita. A vseeno je bil kriging z zunanjim vplivom bolj točen kot ostala in je zato bolj primeren za modeliranje vlažnosti prsti.

## 5. Vprašanja ob članku in razumevanje metodologije

Za boljše razumevanje članka, bi lahko vključili še kakšne dele kode iz r-ja, da bi lahko pri samih statističnih testih in modeliranju lahko videli kako so dejansko izvedli določene teste. Sama razlaga modelov pa je dobro teoretično podkrepljena, vendar zelo zahtevna za branje, kar samo dokazuje ta je to strokovni članek in ni namenjen nekemu bralcu, ki ne bi imel predznanja. Uporabljenih je bilo veliko strokovnih izrazov, za katere je težko najti ustrezne slovenske strokovne izraze in jih je bilo včasih tudi potrebno zapisati po svoje. Kot najtežji del same statistične metodologije bi bilo izpostavila geostatistična pristopa, pri katerih je bilo predvsem veliko novih informacij, katerim je bilo potrebno nameniti več pozornosti in časa kot pa linearnim mešanim modelom, ki smo jih omenili že na predavanjih. Prav tako na področju geostatističnih pristopov ni bilo slovenske literature, ki bi lahko pripomogli k boljšemu razumevanju kriginga, predvsem kriginga z zunanjim vplivom. Nekoliko boljše in nazorno pa bi lahko bila tudi prikazan navzkrižni korelogram, pri katerem se da razumeti, da primerjajo dve mapi različnih modelov, pri katerih primerjajo korelacijo kot funkcijo razdalje med opazovanji in nato primerjajo mapi kako podobne so ocene za napovedane vrednosti, vendar po najinem mnenji manjka nekaj razlage, zakaj je ta postopek pomemben in kakšen je bil njegov doprinos k samim rezultatom in zaključku. Sicer pa je bila sama metodologija ustrezna, saj so uporabili kar tri modele za napovedovanje na treh različnih podatkih. S tem so dobili veliko različnih rezultatov, ki so jih lahko primerjali, ocenjevali, ter lahko opazili, kateri od modelov je najbolj primeren in daje najboljše približke(napovedane vrednosti) samim opazovanim vrednostim. Lahko bi izpostavila tudi, da so v članku predstavili veliko rezultatov in tabel iz r-ja, katere so v samem poglavju rezultatov dobro in natančno interpretirali.

## 6. Neznana statistična metodologija

V članku je bilo kar nekaj nama prej nepoznane statistične metodologije, ki sva jo podrobno opisala predvsem v poglavju 3., saj se nama zdi, da je to ključno za razumevanje kasnejših pojmov pri rezultatih. Predvsem je bilo potrebno dodatno raziskati metodo kriging s katero se do sedaj nisva spoznala, hkrati pa se nama je zdela ključna, saj je prinesla najboljše napovedi in jo avtorji posebej izpostavljajo kot najboljšo. Poelg tega sva se spopadla tudi z veliko količino tujih strokovnih izrazov, kjer ni še jasno določenih slovenskih prevodov oziroma jih je težko določiti iz angleških virov. Ena izmed metodologij, ki nama je bila tudi neznana, a je nisva prej opisala, saj ni ključnega pomena za razumevanje rezultatov pa je analiza glavnih komponent. To je bila ena izmed metod, ki so jo uporabili pri čiščenju podatkov in njihovem združevanju. Namen te metode je sintetizirati informacije iz več različnih tipov podatkov, ki so bili pridobljeni, torej iz gravimetrične metode, elektromagnetne indukcije in zemeljskega radarja. Gre za tri različne načine vzorčenja vlažnosti prsti, zato so ti podatki na začetku v različni obliki in je združevanje težavno. Analiza glavnih komponent preprečuje te težave, zmanjša tudi težave s kolinearnostjo. To metodo so uporabili na štirih različnih naborih podatkov zemeljskemu radarju za vse frekvence (250 MHz, 600 MHz, 1600 Mhz) in podatkih iz elektromagnetne indukcije. Izvedena je bila na korelacijski matriki spremenljivko, da bi dobili nekaj novih neodvisnih komponent, ki razložijo večino variacij v začetnih podatkih. Glavne komponente z večjo lastno vrednostjo od 1 so bile obrdžane po Kaiserjevem kriteriju. Nato so raziskali še informiranost posameznih spremenljivk in raziskali njihovo povezavo.

## 7. Možnosti za dodatno/drugačno analizo podatkov

Menima, da so se avtorji precej poglobili v različne načine analize podatkov, ter jih precej podrobno testirali in predstavili, saj je vključenih veliko različnih metod predstavitve podatkov, kot so škatla z ročaji, razsevni diagram in različne slike površine in njene vodnatosti. Same analize so se lotili zelo celostno, saj so podatke zbirali na več različnih načinov, primerjali podatke iz senzorjev s tistimi iz analiz. Poleg tega so površino predhodno testirali, da bi senzorje primerno nastavili po navodilih proizvajalcev. Podatke so tudi podrobno preverili in izločili napake. Manjka tudi razlaga zakaj so izbrani točno ti trije modeli in morda ne kakšen drugi, ki bi morda podobno dobro modelirali vodnatost prsti. V primeru dostopnosti teh podatkov bi bilo zanimivo preveriti natančnost še kakšnega dodatnega modela, ki bi morda primerljivo dobro opisal vodnatost

površine. A vseeno bi bile mogoče določene nadgradnje. Variabilnost podatkov je bila v tem primeru zelo nizka, saj je imela površina zelo podobno vlažnost. Dodatna nadgradnja bi bila pri testiranju teh modelov na več različnih površinah, ali pa že na večji površini, kar bi bolj ustrezalo tudi realnim situacijam, saj so ti modeli zanimivi sploh za obsežne namakalne sisteme, kjer je seveda prst tudi precej bolj raznolika. Dodatno bi bilo smiselno tudi nadaljne testirane najbolj učinkovitega modela, torej kriginga z zunanjim vplivom, ki se je izkazal za najboljšega. Sploh natančnost tega bi bilo zanimivo preveriti na že prej omenjenih bolj raznolikih površinah. Obenem pa se poraja vprašanje zakaj ni bil linearni mešani model nič boljši kot navadni kriging, kljub večji količini podatkov, ki jih ta obdela. Avtorji sami ne dajo odgovora na to vprašanje, zato je morda smiselno to dodatno raziskati, da ni prišlo do celo kakšnih napak pri tem modelu. Ena izmed opcij za pridobivanje podatkov je tudi termalni infrardeči zemeljski radar, ki bi tudi lahko bil uporabljen za merjenje vlažnosti prsti. Tega načina članek ne omenja, ampak bi bilo morda zanimivo preveriti še dtega in natančnost podatkov s strani te metode. V raziskavi se samo zanašajo na različne mikrovalne zemeljske radarje. Poleg tega sta raziskana samo dva načina zbiranja podatkov, kjer je sicer zemeljski radar nekoliko boljši, a sta bila uporabljena dva zemeljska radarja. Sama raziskava samo preverja kateri izmed senzorjev je bolj natančen, ni pa s tem dokončnega odgovora kateri je boljši, saj ne upoštevajo kateri način je hitrejši ali pa cenejši. Vsi te faktorji so izredno pomembni pri odločitvah v praksi in ravno tu vidiva glavne možnosti za dodatno analizo podatkov. Raziskavo bi bilo potrebno razširiti na druge faktorje, saj v praksi ni nujno najpomembnejša največja natančnost, ampak tudi cena in hitrost pridobivanja podatkov. Podobno velja tudi za lienarne modele, tudi pri teh je bilo raziskana samo njihova natančnost, ne pa njihova hitrost in robustnost. Kljub temu, da je bil kriging z zunanjim vplivom najbolj natančen, se zdi da bi v praksi vseeno morda bi navaden kriging bolj uporaben, saj je vseeno enostavnejši in s tem tudi hitrejši. Ravno ti faktorji so ponavadi ključnega pomena pri odločanju o modelu v praksi.

## Viri in literatura

Barca, E., De Benedetto, D., & Stellacci, A. M. (28. 2 2019). Contribution of EMI and GPR proximal sensing data in soil water content assessment by using linear mixed effects models and geostatistical approaches. Pridobljeno iz Science Direct: <https://www.sciencedirect.com/science/article/abs/pii/S0016706118309662>

Truman, G., Manning, B., Sauer, B., & Schneider, D. (2 2016). Using Electromagnetic Induction (EMI) to Measure Soil Water. Pridobljeno iz NSW Government : [https://northwest.lis.nsw.gov.au/\\_\\_\\_data/assets/pdf\\_file/0004/590440/EM-brochure-2.pdf](https://northwest.lis.nsw.gov.au/___data/assets/pdf_file/0004/590440/EM-brochure-2.pdf)

World Meteorological Organization. (11 2016). World Meteorological Organization. Pridobljeno iz Measurement of Soil Moisture: [https://library.wmo.int/doc\\_num.php?explnum\\_id=3157#:~:text=Gravimetric%20water%20content%20is%20](https://library.wmo.int/doc_num.php?explnum_id=3157#:~:text=Gravimetric%20water%20content%20is%20)

Xinbo, L., Chen, J., Xihong, C., Qixin, L., Xin, C., & Chen, X. (2019). Measurement of soil water content using ground-penetrating radar: a review of current methods. Pridobljeno iz International Journal of Digital Earth: <https://www.tandfonline.com/action/showCitFormats?doi=10.1080%2F17538947.2017.1412520>