



LATVIJAS UNIVERSITĀTE
**ĢEOGRĀFIJAS UN
ZEMES ZINĀTŅU
FAKULTĀTE**

"Laiktelpiskā pazemes ūdeņu sausuma prognozēšana
ar jauktiem modeļiem daudzslāņu sedimentācijas baseinā
klimata pārmaiņu ietekmē" (LZP-2019/1-0165)
"Dabas resursu ilgtspējīga izmantošana klimata pārmaiņu kontekstā"



FLPP
FUNDAMENTĀLIE UN
LIETIŠĀJIE PĒTĪJUMU
PROJEKTI



Latvijas Zinātnes padome

Pazemes ūdens līmeņu laikrindu aizpildīšana iekļaujot iztrūkstošo vērtību raksturu

Jānis Bikše¹ | janis.bikse@lu.lv

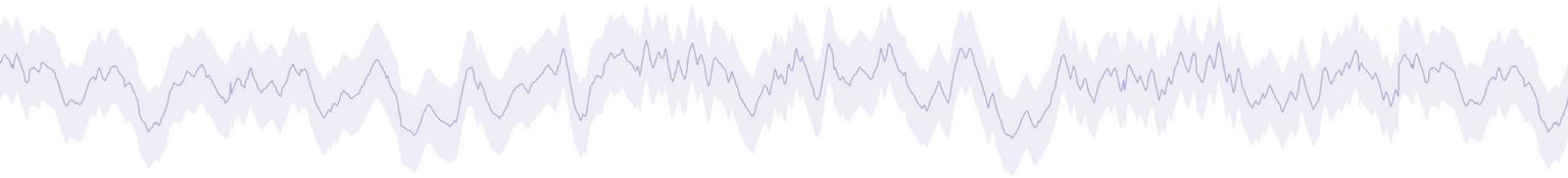
Inga Retiķe¹, Ezra Haaf², Andis Kalvāns¹

¹Faculty of Geography and Earth Sciences, University of Latvia, Jelagavas 1, Riga, Latvia

²Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden

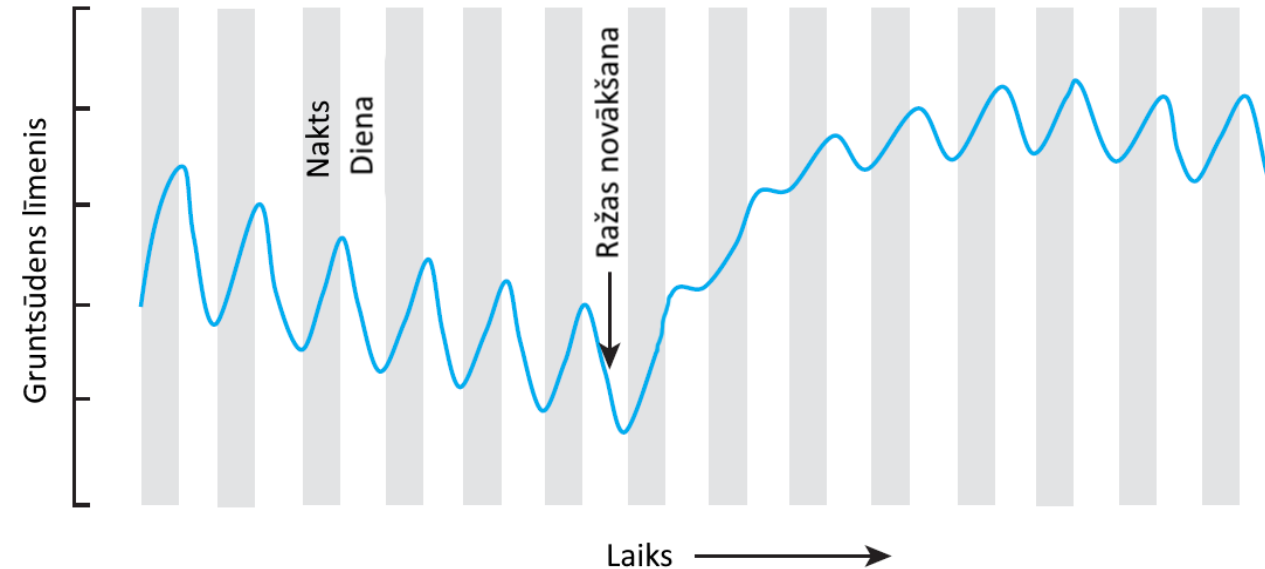
E-mail: janis.bikse@lu.lv

81. Latvijas Universitātes staptautiskā
zinātniskā konference 2023
Virtual Models for Sustainable and Safe
Living Environment



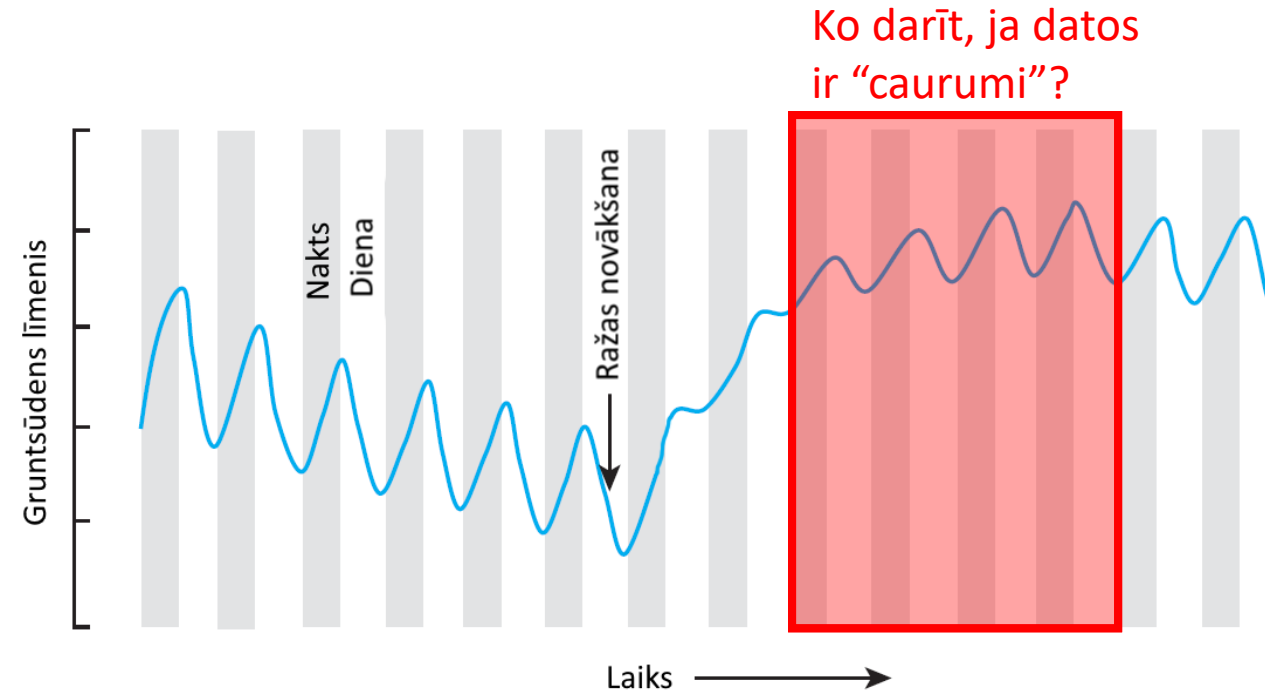
Kāpēc nepieciešami nepārtraukti novērojumi?

- Indeksi (sausuma, dinamikas u.c.), kam ir prasības pret novērojumu datu kvalitāti (nav iztrūkstošas vērtības)
- Ilgtermiņa rindu “saglābšana”, tendenču analīze
- Datu salīdzināšana starp stacijām
- Izmantošana noteiktos modeļos



Kāpēc nepieciešami nepārtraukti novērojumi?

- Indeksi (sausuma, dinamikas u.c.), kam ir prasības pret novērojumu datu kvalitāti (nav iztrūkstošas vērtības)
- Ilgtermiņa rindu “saglābšana”
- Datu salīdzināšana starp stacijām
- Izmantošana noteiktos modeļos



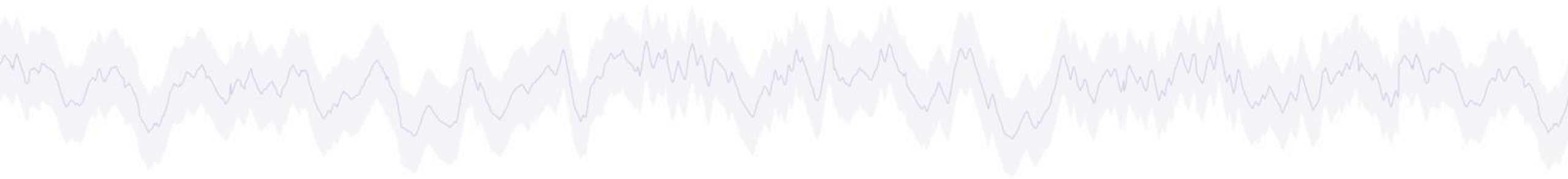
Izplatīta pieeja – ja datos ir “caurumi”, tad neizmantojam

Īss kopsavilkums

Liels mērķis – analizēt pazemes ūdens (PŪ) līmeņu dinamiku, tās saistību ar citiem faktoriem (ģeoloģija, sateces baseins, klimats)

BET – ir nepieciešami kvalitatīvi, regulāri dati pētījuma veikšanai

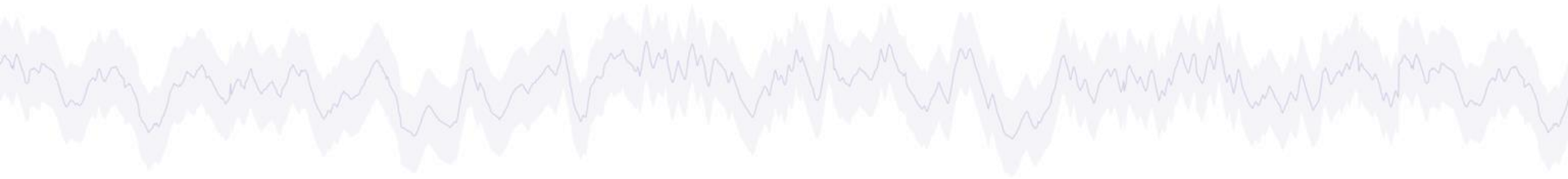
PŪ datos ir daudz iztrūkstošu vērtību un ir metodes, kā tās aizvietot/aizpildīt, BET – kā novērtēt aizpildīšanas kvalitāti?



Īss kopsavilkums

Šajā prezentācijā demonstrēts kā:

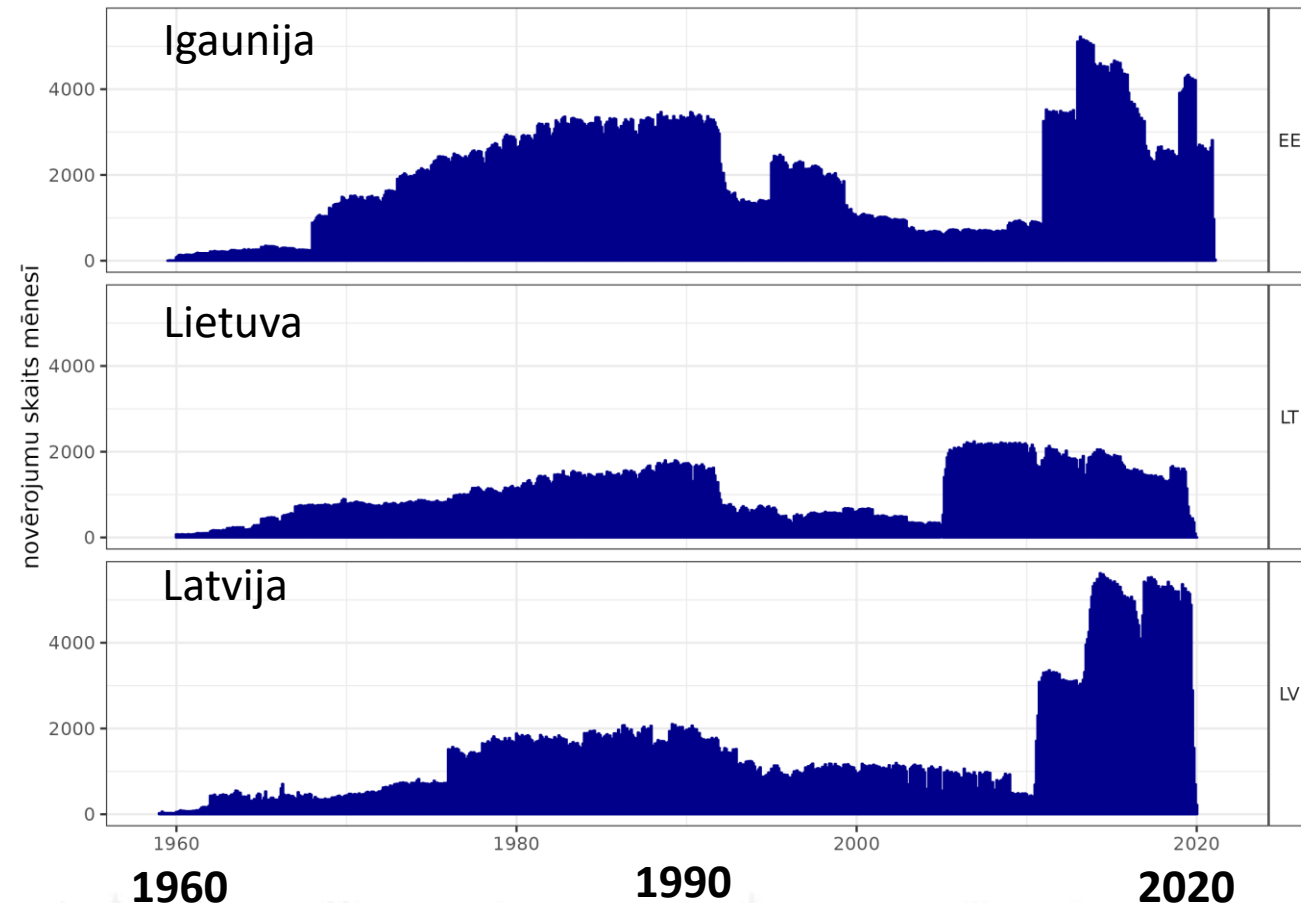
- 1) aizpildīt iztrūkstošās vērtības PŪ līmeņu novērojumos
- 2) novērtēt PŪ līmeņu aizpildīšanu veikspēju, ņemot vērā iztrūkstošo vērtību raksturu



Pazemes ūdens līmeņu novērojumi

Baltijas valstīs valsts monitoringa tīklos kopumā veikti 3'389'494 (derīgi) PŪ līmeņu novērojumi:

- Igaunijā 1'475'538
 - Lietuvā 776'475
 - Latvijā 1'137'481
- Tie nosedz laika posmu sākot no 1959. gada, bet novērojumu daudzums pa gadiem mainās



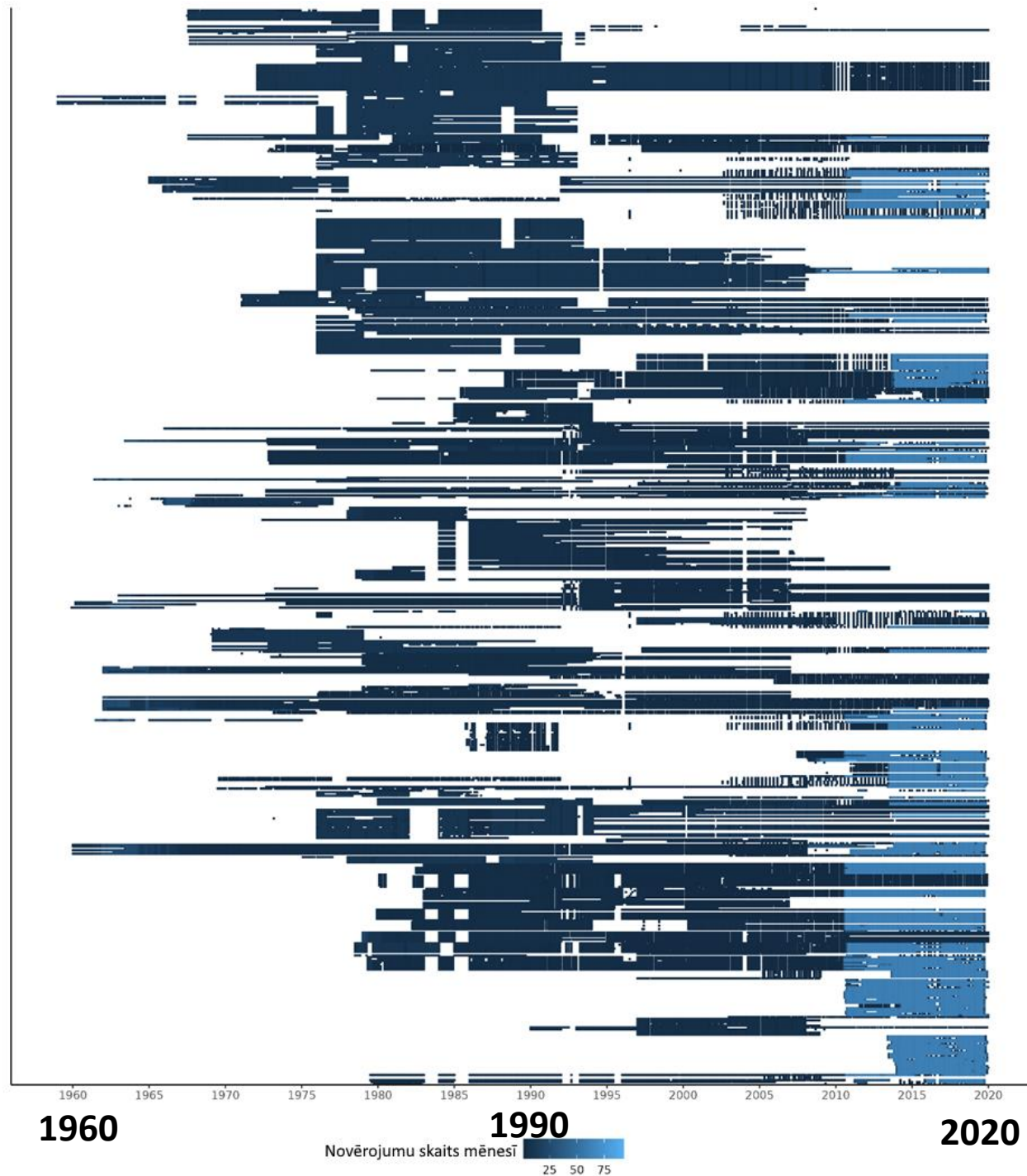
Novērojumu (ne)regularitāte

Daudzos urbemos ir novērojumi:

- pārtraukumi
- periodi ar dažādu novērojumu regularitāti
- novērojumu apsīkšana, nesasniedzot mūsdienas (dažādi temporālie pārklājumi)

Problēma: Nepieciešamība pēc regulāriem (ikdienas) datiem!

Latvijas monitoringa urbumu dati:



Novērojumu datu interaktīva labošana

Urbumu līmeņu pārbaude Latvijai

Info!

Atjaunot urbumu sarakstu

Stacija: Aistere

Urbums piils: Aistere, 332 | ns: 2503

Parametrs:
 Pazemes udens līmenis no zemes virsmas
 Pazemes udens līmenis vjl
 udens stabs virs automatiska limena deveja
 Statiskais līmenis

Veids:
 Automatiskie
 Izejas dati
 Kontroles
 Lauka
 Manuālie

grafiku parametri:
 Līnijas
 Punkti
 Reverse vertical

visu veidu grafikiem:
 free scale
 fixed scale

update 5g

Aistere, 332
ns: 2503
horizonts: D 2 br - D 3 am
dziļums [m]: 291.3
filtrs no-līdz: 241-282
m vjl.: 28.146
urb. gads: 1971

karte tabDati tabUnikalie g1 g1-plotly g2-pa-veidiem g2-plotly g3-gradienti g3-plotly g4-kontrole g5-labot g5-plotly g6-visiUrb g6-plotly

Pieeja publicēta:

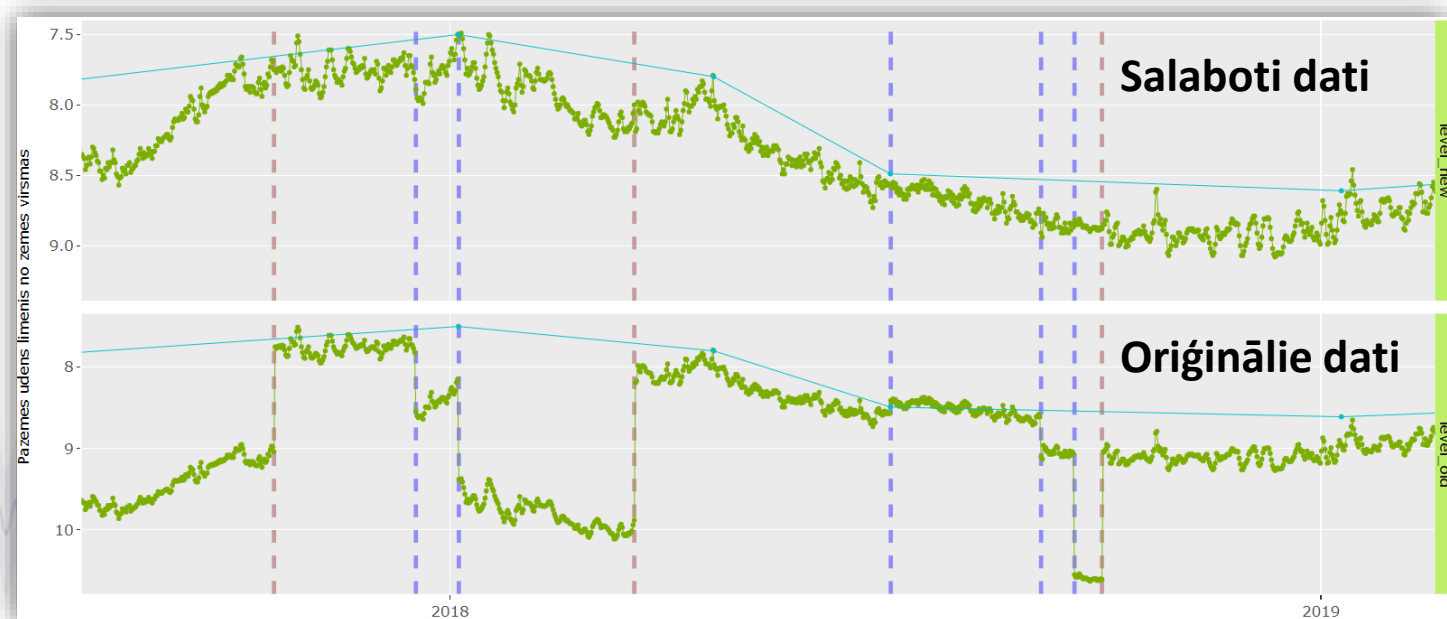


Journal of Hydrology
Volume 605, February 2022, 127294

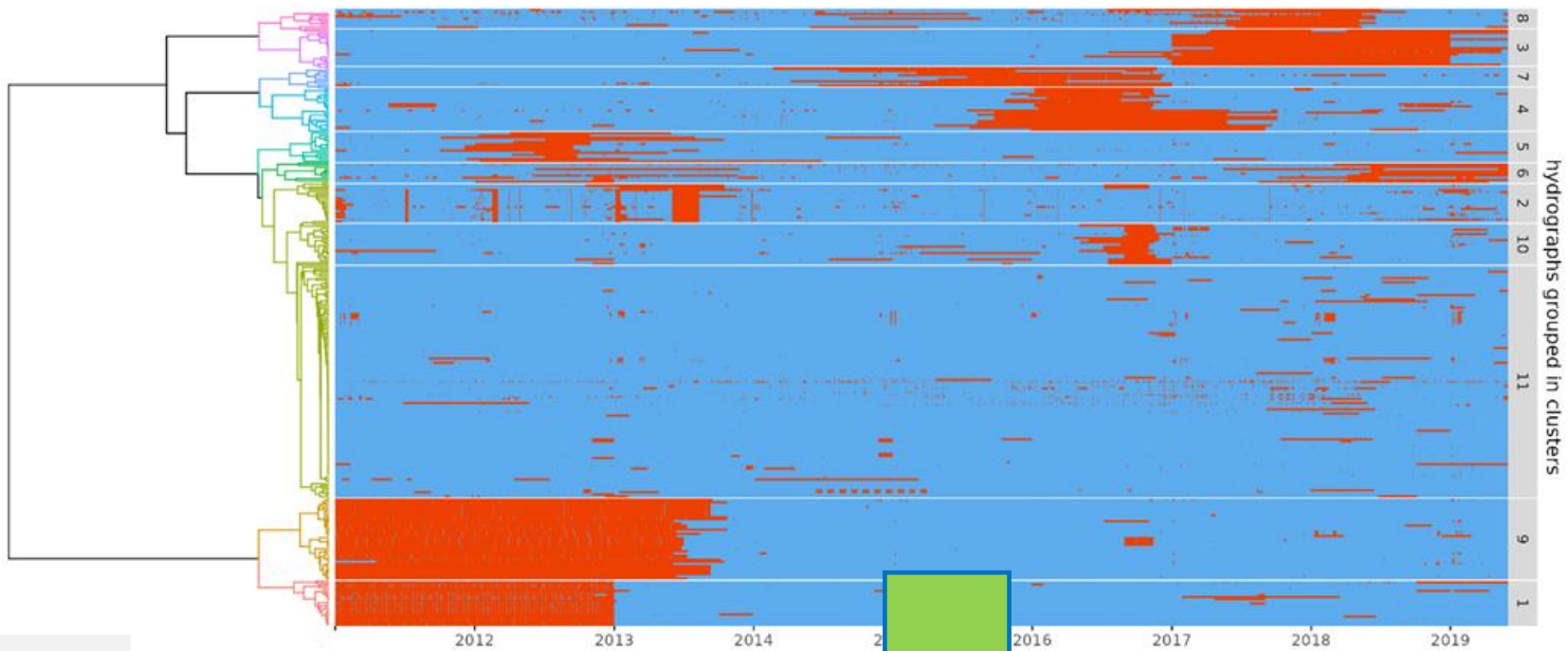


Rescue of groundwater level time series: How to visually identify and treat errors

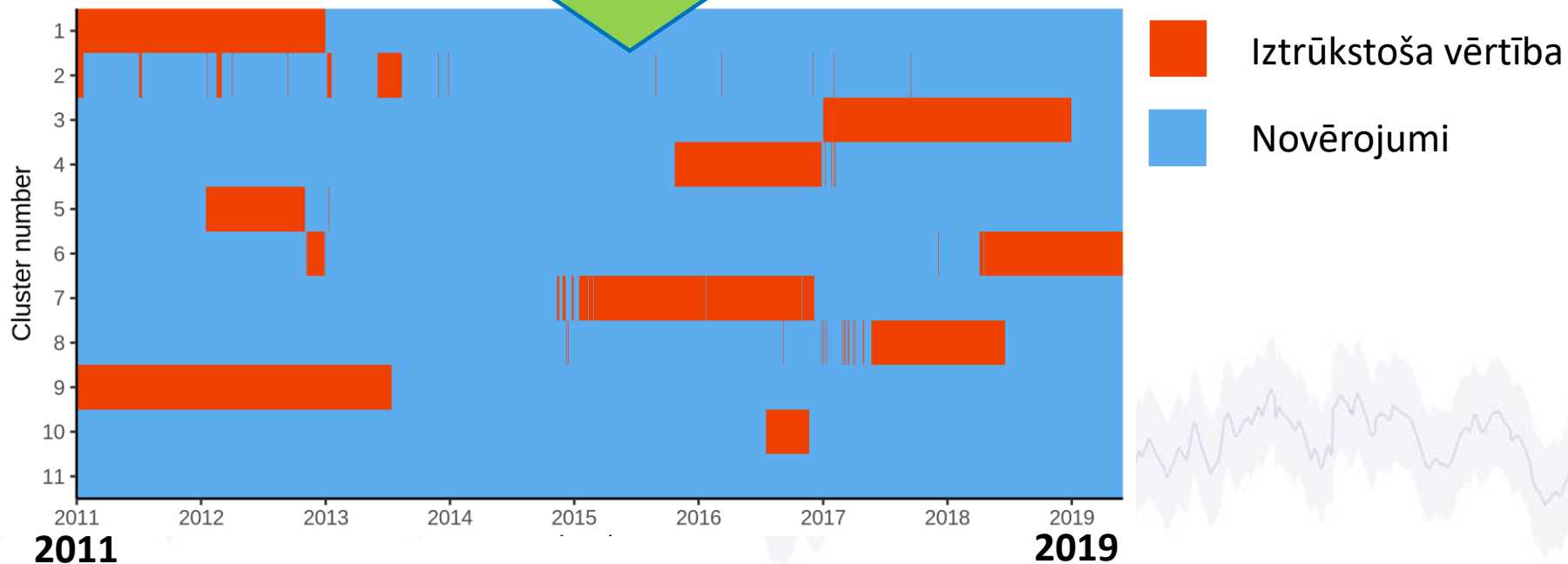
Inga Retike ^{a, *}, Jānis Bikše ^{a, *}, Andis Kalvāns ^{a, *}, Aija Dēliņa ^{a, *}, Zanīta Avotniece ^{a, *}, Willem Jan Zaadnoordijk ^{b, *}, Marta Jemeljanova ^{a, *}, Konrāds Popovs ^{a, *}, Alise Babre ^{a, *}, Artjoms Zelenkevičs ^{a, *}, Artūrs Baikovs ^{a, *}



11 klāsteri ar iztrūkstošo vērtību grupām



Iztrūkstošo vērtību raksturu grupas. Izmantots kritērijs – 50% iztrūkums dienā.
Ja dienā > 50% no visām laikrindām konkrētajā klāsterī uzrāda iztrūkstošas vērtības, tad gala grupā šī diena arī kļūst par iztrūkstošo vērtību.



Iztrūkstošo vērtību aizpildīšanas novērtēšanas procedūra – divas pieejas

Typical gap patterns (TGP)

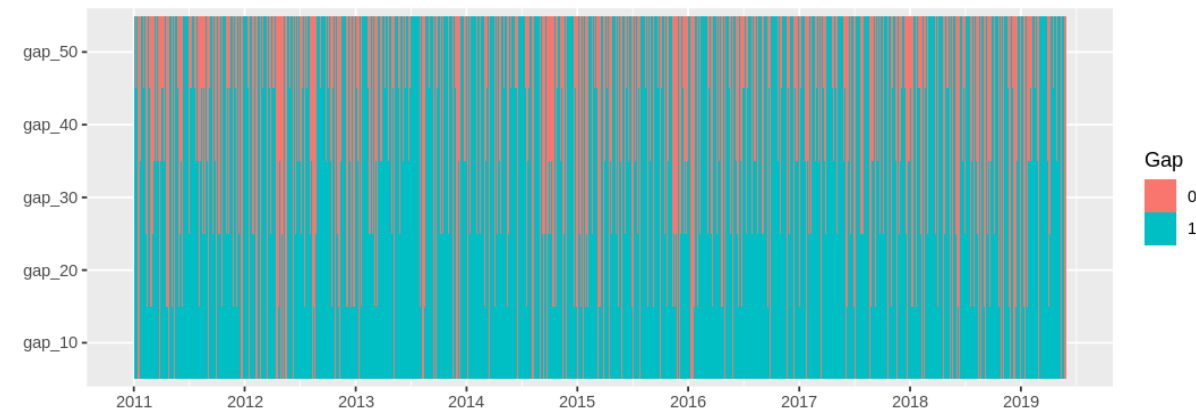
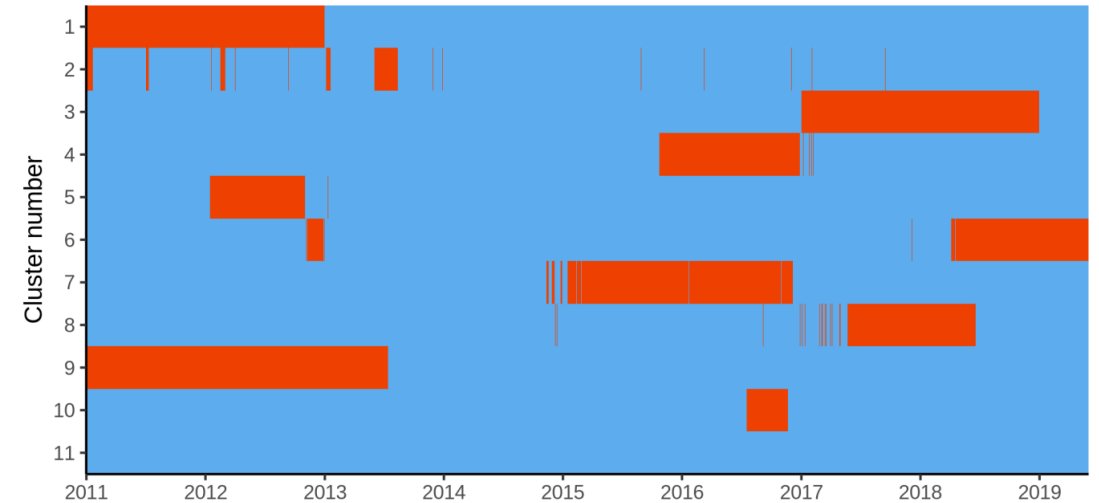
Tipiskie iztrūkstošo vērtību šabloni:

1. Katrā urbumā tiek mākslīgi “izgriezti” novērojumi balstoties uz 10 iztrūkstošo vērtību šabloniem
2. Tiek pielietots modelis iztrūkstošo datu aizpildīšanai
3. Tiek novērtēta modeļa spēja aizvietot iztrūkstošos datus, tos salīdzinot ar oriģinālajiem novērojumiem

Random gap patterns (RGP)

Pēc nejaušības principa iztrūkstošie šabloni:

Pēc nejaušības principa izgriezti “caurumi” novērojumos ar 10%, 20%, 30%, 40% un 50% iztrūkumu



Modeļi iztrūkstošo datu aizpildīšanai

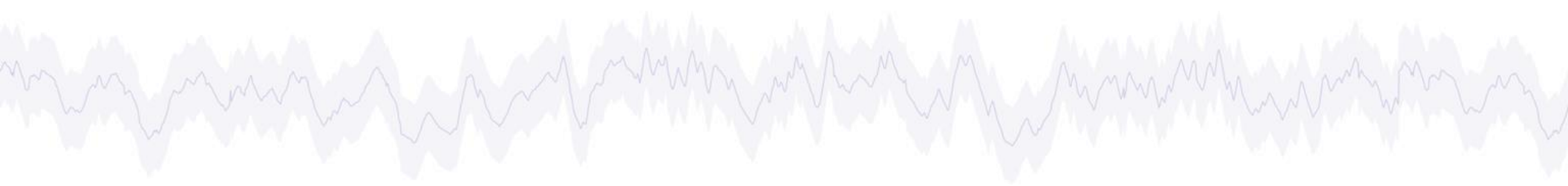
Perspektīvie modeļi

- **missForest** – balstās uz iteratīvu gadījumu mežu (random forest) algoritmu
- **imputePCA** – balstās uz iteratīvu galveno komponentu analīzes algoritmu

Abi modeļi pielietoti paralelizējot darbības uz vairākiem CPU kodoliem

Vienkāršā pieeja salīdzināšanas nolūkiem:

- **Lineārā interpolācija**



Pielietotie modeļi un aprēķinu ilgums

Modelis	Modeļa parametri	Ieviesto iztrūkstošo vērtību veids	Kopējais CPU laiks [Days-H:M:S]	Vidējais CPU laiks uz vienu imputāciju [minūtes]	Izmantotais CPU kodolu skaits
missForest	maxiter=1, ntree=10	TGP	10-13:03:44	6.4	16
missForest	maxiter=1, ntree=100	TGP	71-20:12:16	43.4	16
missForest	maxiter=10, ntree=10	TGP	49-17:11:12	30.0	16
missForest	maxiter=10, ntree=100	TGP	392-11:15:12	237.1	48
missForest	maxiter=1, ntree=10	RGP	11:19:12	1.2	16
missForest	maxiter=1, ntree=100	RGP	3-17:17:20	9.8	16
missForest	maxiter=10, ntree=10	RGP	2-12:48:32	6.7	32
missForest	maxiter=10, ntree=100	RGP	22-13:41:52	59.6	32
imputePCA	ncp=3	TGP	3-20:06:00	2.3	10
imputePCA	ncp=3	RGP	0-2:20:30	0.3	5
lineārā interpolācija	-	TGP	0-0:00:22	0.0	1
lineārā interpolācija	-	RGP	0-0:00:10	0.0	1

Izmantots **HPC klāsteris (LU SMI)**

Ilgākais aprēķins uz 1 CPU prasītu ~392 dienas, bet uz HPC ar 48 CPU tas prasīja tikai ~8 dienas, jeb ~5 minūtes uz vienu aizpildīšanu

Katrā TGP = 2384 individuālas aizpildīšanas/modeļi
Katrā RGP = 545

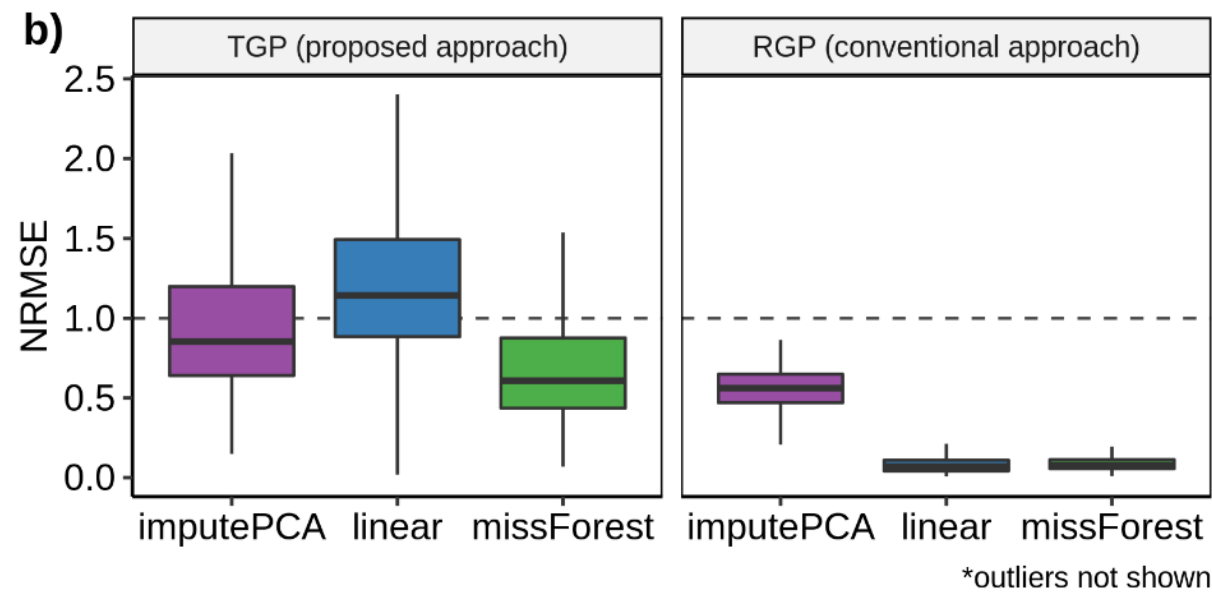
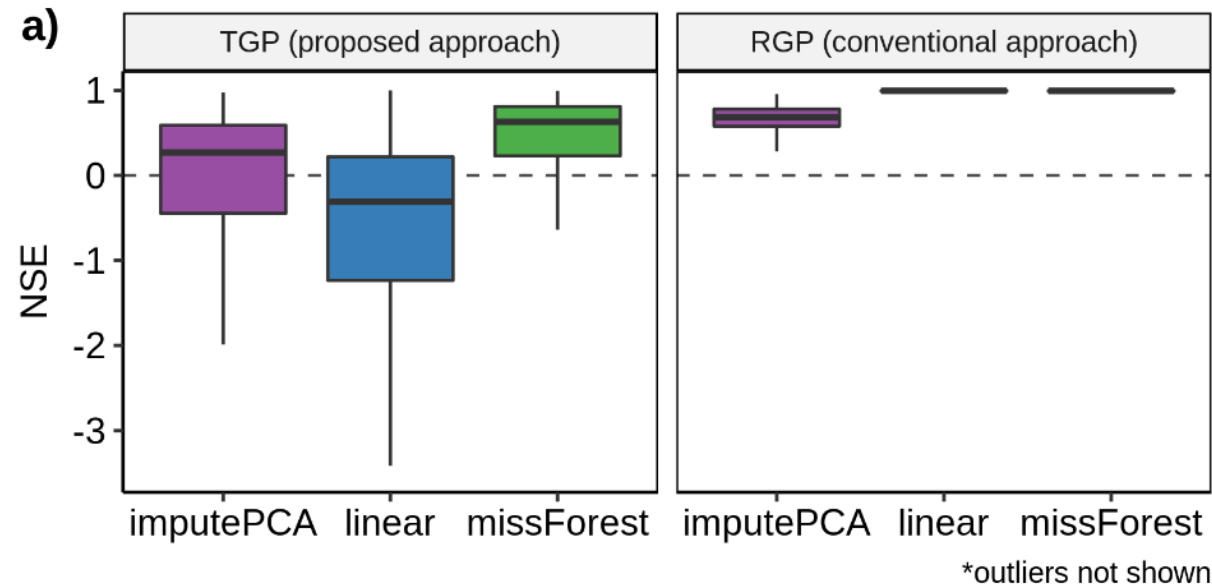
Rezultāti

Rezultātu salīdzināšanai izmantotas sekojošas metrikas:

- Normalizētā vidējā kvadrātiskā kļūda (NRMSE)
- Naša-Satklifa efektivitātes koeficients (NSE)

RGP (nejaušības princips) – fantastiski rezultāti! Bet PŪ novērojumu iztrūkumos (parasti) nav nejaušības

TGP (tipiskie izstrūkstošo vērtību šabloni) – reālistiskāks novērtējums

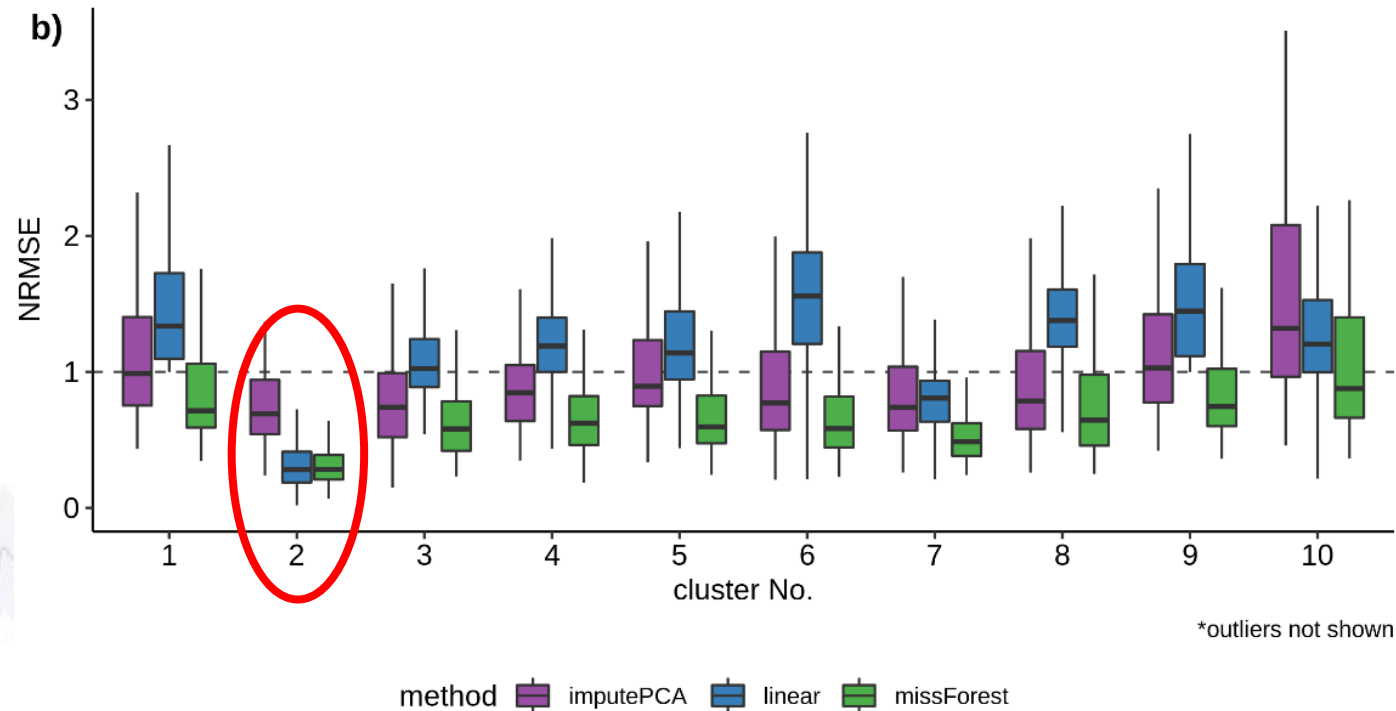
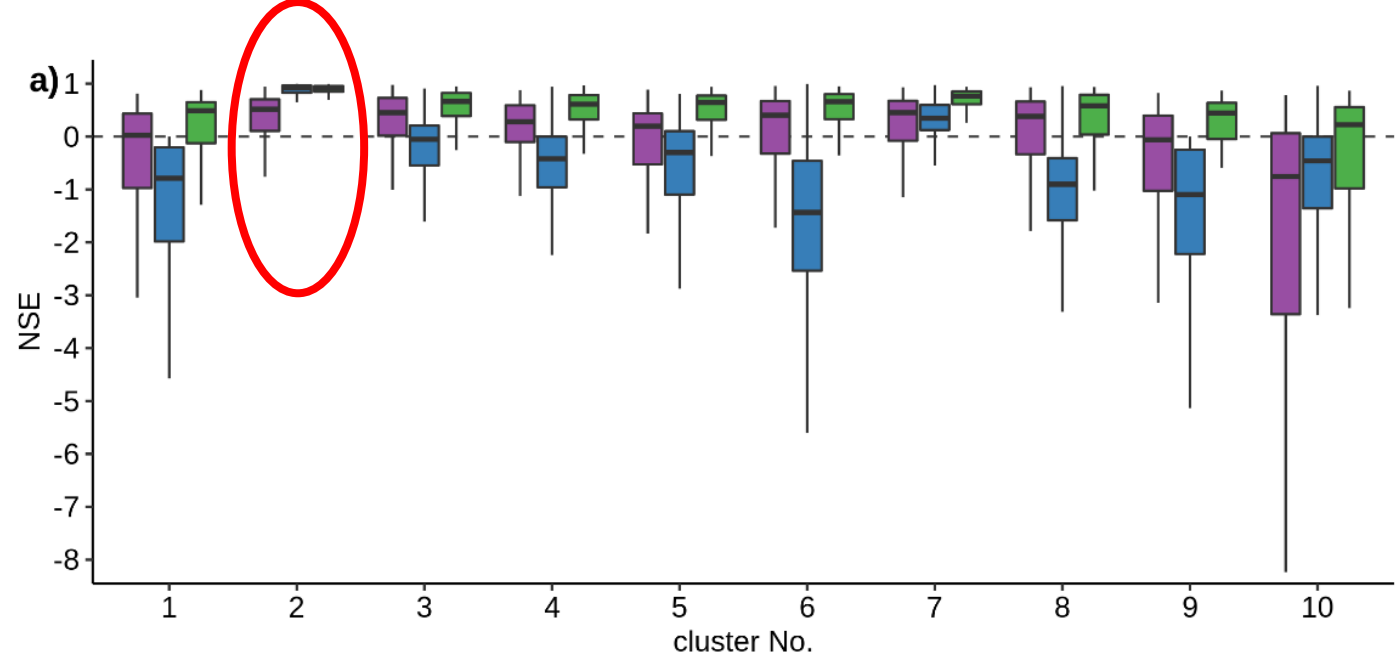
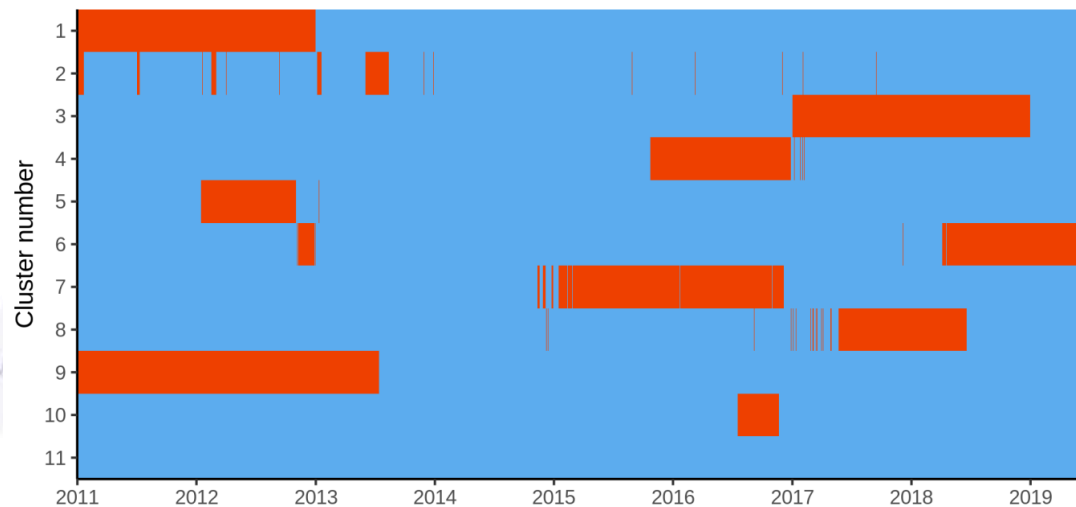


NSE – jo lielāks, jo labāk (1=ideāli)

NRMSE – jo mazāks, jo labāk (0=ideāli)

Rezultāti pa klāsteriem

- Lielākajā daļā klāsteru missForest uzrāda labākos rezultātus
- Klāsterī Nr.2 lineārā interpolācija un missForest līdzvērtīgi
- Grūtības ir ar klāsteriem, kur ir gari caurumi laikrindu sākumā vai beigās (1., 6., 9.)

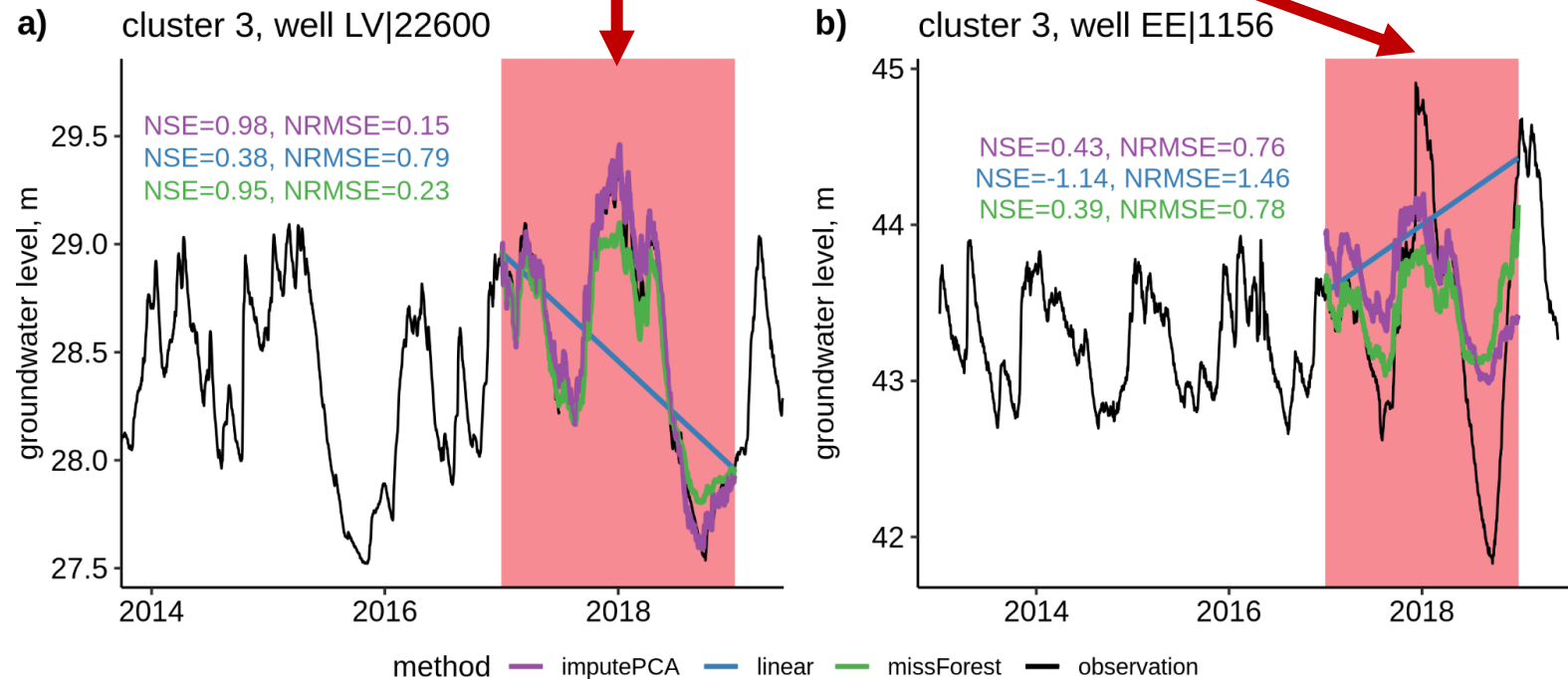


Piemērs

Izaicinājumi:

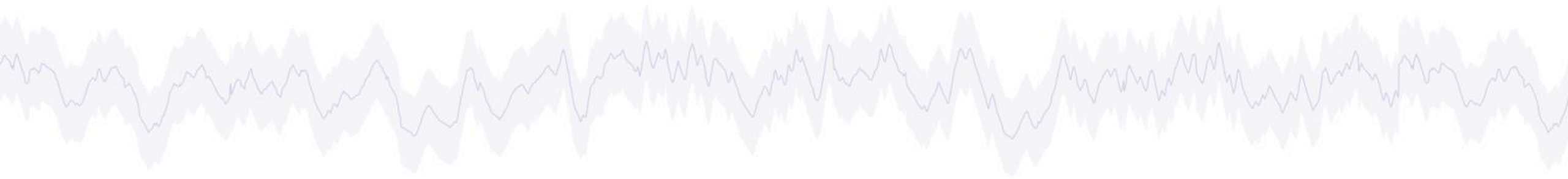
- Ekstrēmi notikumi – sausumi, piemēram, 2018.g sausums
- Antropogēnā ietekme – “neloģiskas” PŪ līmeņa izmaiņas

Izgrieztā daļa, ko aizpilda modeļi



Secinājumi

- Tipiskie iztrūkstošo vērtību šabloni sniedz reālistiskāku novērtējumu, nekā izgriezti caurumi pēc nejaušības principa
- missForest ir piemērotākais rīks no aplūkotajiem modeļiem pazemes ūdens līmeņu laika sēriju datu iztrūkumu aizpildīšanā, aizpildot pat gandrīz 2 gadus ilgus iztrūkumus diennakts datos
- Ja iztrūkstošām vērtībām ir nejaušības raksturs (klāsteris Nr.2), tad arī lineārā interpolācija sniedz tikpat labus rezultātus, kā missForest



Paldies par uzmanību!

Jānis Bikše | janis.bikse@lu.lv



LATVIJAS UNIVERSITĀTE
**ĢEOGRĀFIJAS UN
ZEMES ZINĀTŅU
FAKULTĀTE**

"Laiktelpiskā pazemes ūdeņu sausuma prognozēšana
ar jauktiem modeļiem daudzslāņu sedimentācijas baseinā
klimata pārmaiņu ietekmē" (LZP-2019/1-0165)



FLPP
FUNDAMENTĀLIE UN
LIETIŠĀJIE PĒTĪJUMU
PROJEKTI



Latvijas Zinātnes padome