

COVID-19 Lietuvoje: matematiniai ir statistiniai modeliai¹

Remigijus Leipus, Alfredas Račkauskas, Olga Štikonienė

Vilniaus Universitetas, MIF

LMD 2020

¹LMT remiamas projektas Nr. S-COV-20-4

Turinys

- ➊ *Globali profilių analizė (Tadas Danielius, Jovita Gudan, A.R.)*
- ➋ *Bangų palyginimas (Tadas Danielius)*
- ➌ *Trumpalaikės prognozės (Jurgita Markevičiūtė, Povilas Treigys)*
- ➍ *Klasterinė analizė (Jolita Bernatavičienė, Viktor Medvedev, Julius Venskus, Povilas Treigys)*
- ➎ *Testų analizė (Rūta Levulienė)*

Globali profilių analizė [Tadas Danielius, Jovita Gudan, A.R.]

- Supratimas kaip plėtojosi COVID-19 pandemija pasaulyje gali padėti valdant lokalius protrūkius ir tobulinant sveikatos politiką.

Globali profilių analizė [Tadas Danielius, Jovita Gudan, A.R.]

- ▶ Supratimas kaip plėtojosi COVID-19 pandemija pasaulyje gali padėti valdant lokalius protrūkius ir tobulinant sveikatos politiką.
- ▶ Nagrinėti pandemijos parametrai:
 - ▶ pasveikusių žmonių skaičius $R(t)$;
 - ▶ izoliuotų žmonių skaičius $Q(t)$;
 - ▶ mirusių žmonių skaičius $D(t)$.
- ▶ Visi šie parametrai yra atsitiktinės funkcijos. Todėl jų profilių analizei natūralu pritaikyti funkcinės duomenų analizės metodus.

Išvestiniai parametrai:

- ▶ mirtingumas $\kappa(t)$:

$$D(t) = \int_{T_0}^t \kappa(s)Q(s)ds + \varepsilon(t), \quad t \in [T_0, T_1], \quad (1)$$

- ▶ išgydymo greitis $\lambda(t)$:

$$R(t) = \int_{T_0}^t \lambda(s)Q(s)ds + \eta(t), \quad t \in [T_0, T_1] \quad (2)$$

čia $(\varepsilon(t))$ ir $(\eta(t))$ yra atsitiktinės funkcijos, T_0 žymi pandemijos pradžią, T_1 stebėjimų pabaigą.

Domina klausimai:

- ▶ Koks yra tipiškas Europos šalių izoliacijos (mirtingumo, pagijimo ir kt.) proceso Q modelis? Baltijos regiono?

Domina klausimai:

- ▶ Koks yra tipiškas Europos šalių izoliacijos (mirtingumo, pagijimo ir kt.) proceso Q modelis? Baltijos regiono?
- ▶ Kokios šalys yra labiausiai panašios su Lietuva pagal mirtingumą/išgijimą?

Domina klausimai:

- ▶ Koks yra tipiškas Europos šalių izoliacijos (mirtingumo, pagijimo ir kt.) proceso Q modelis? Baltijos regiono?
- ▶ Kokios šalys yra labiausiai panašios su Lietuva pagal mirtingumą/išgijimą?
- ▶ Kuriuo pandemijos laikotarpiu mirtingumas buvo labiausiai kintamas Europoje?

Domina klausimai:

- ▶ Koks yra tipiškas Europos šalių izoliacijos (mirtingumo, pagijimo ir kt.) proceso Q modelis? Baltijos regiono?
- ▶ Kokios šalys yra labiausiai panašios su Lietuva pagal mirtingumą/išgijimą?
- ▶ Kuriuo pandemijos laikotarpiu mirtingumas buvo labiausiai kintamas Europoje?
- ▶ Kuriuo pandemijos laikotarpiu mirtingumas buvo mažiausiai kintamas Europoje?

Domina klausimai:

- ▶ Koks yra tipiškas Europos šalių izoliacijos (mirtingumo, pagijimo ir kt.) proceso Q modelis? Baltijos regiono?
- ▶ Kokios šalys yra labiausiai panašios su Lietuva pagal mirtingumą/išgijimą?
- ▶ Kuriuo pandemijos laikotarpiu mirtingumas buvo labiausiai kintamas Europoje?
- ▶ Kuriuo pandemijos laikotarpiu mirtingumas buvo mažiausiai kintamas Europoje?
- ▶ Ar yra šalių kurios būtų ekstreminės mirtingumo ir išgijimo atžvilgiu?

Domina klausimai:

- ▶ Koks yra tipiškas Europos šalių izoliacijos (mirtingumo, pagijimo ir kt.) proceso Q modelis? Baltijos regiono?
- ▶ Kokios šalys yra labiausiai panašios su Lietuva pagal mirtingumą/išgijimą?
- ▶ Kuriuo pandemijos laikotarpiu mirtingumas buvo labiausiai kintamas Europoje?
- ▶ Kuriuo pandemijos laikotarpiu mirtingumas buvo mažiausiai kintamas Europoje?
- ▶ Ar yra šalių kurios būtų ekstreminės mirtingumo ir išgijimo atžvilgiu?
- ▶ Kaip ilgai tęsiasi ekstreminiai periodai?

Domina klausimai:

- ▶ Koks yra tipiškas Europos šalių izoliacijos (mirtingumo, pagijimo ir kt.) proceso Q modelis? Baltijos regiono?
- ▶ Kokios šalys yra labiausiai panašios su Lietuva pagal mirtingumą/išgijimą?
- ▶ Kuriuo pandemijos laikotarpiu mirtingumas buvo labiausiai kintamas Europoje?
- ▶ Kuriuo pandemijos laikotarpiu mirtingumas buvo mažiausiai kintamas Europoje?
- ▶ Ar yra šalių kurios būtų ekstreminės mirtingumo ir išgijimo atžvilgiu?
- ▶ Kaip ilgai tęsiasi ekstreminiai periodai?
- ▶ Kaip ir ar įvairių parametrų tendencijos skiriasi tarp Rytų ir Vakarų Europos; Lietuvos ir kitų šalių?

Domina klausimai:

- ▶ Koks yra tipiškas Europos šalių izoliacijos (mirtingumo, pagijimo ir kt.) proceso Q modelis? Baltijos regiono?
- ▶ Kokios šalys yra labiausiai panašios su Lietuva pagal mirtingumą/išgijimą?
- ▶ Kuriuo pandemijos laikotarpiu mirtingumas buvo labiausiai kintamas Europoje?
- ▶ Kuriuo pandemijos laikotarpiu mirtingumas buvo mažiausiai kintamas Europoje?
- ▶ Ar yra šalių kurios būtų ekstreminės mirtingumo ir išgijimo atžvilgiu?
- ▶ Kaip ilgai tęsiasi ekstreminiai periodai?
- ▶ Kaip ir ar įvairių parametrų tendencijos skiriasi tarp Rytų ir Vakarų Europos; Lietuvos ir kitų šalių?
- ▶ Kiti klausimai

Duomenų parengimas ir žvalgomoji analizė:

$$Q := (Q(t), t \in R)$$

- ▶ $Q(t)$ - izoliuotų žmonių skaičius laiko momentu t .

Duomenų parengimas ir žvalgomoji analizė:

$$Q := (Q(t), t \in R)$$

- ▶ $Q(t)$ - izoliuotų žmonių skaičius laiko momentu t .
- ▶ Funkcija Q yra atsitiktinė:
 - (a) $Q(t) = 0$, kai $t < T_0 = 0$;

Duomenų parengimas ir žvalgomoji analizė:

$$Q := (Q(t), t \in R)$$

- ▶ $Q(t)$ - izoliuotų žmonių skaičius laiko momentu t .
- ▶ Funkcija Q yra atsitiktinė:
 - (a) $Q(t) = 0$, kai $t < T_0 = 0$;
 - (b) $Q(t)$ didėjanti intervale $[T_0, T_1]$.

Duomenų parengimas ir žvalgomoji analizė:

$$Q := (Q(t), t \in R)$$

- ▶ $Q(t)$ - izoliuotų žmonių skaičius laiko momentu t .
- ▶ Funkcija Q yra atsitiktinė:
 - (a) $Q(t) = 0$, kai $t < T_0 = 0$;
 - (b) $Q(t)$ didėjanti intervale $[T_0, T_1]$.
- ▶ Laiko momentai T_0 ir T_1 bendru atveju yra nežinomi.

Duomenų parengimas ir žvalgomoji analizė:

$$Q := (Q(t), t \in R)$$

- ▶ $Q(t)$ - izoliuotų žmonių skaičius laiko momentu t .
- ▶ Funkcija Q yra atsitiktinė:
 - (a) $Q(t) = 0$, kai $t < T_0 = 0$;
 - (b) $Q(t)$ didėjanti intervale $[T_0, T_1]$.
- ▶ Laiko momentai T_0 ir T_1 bendru atveju yra nežinomi.
- ▶ Funkcija Q kiekvienai šaliai vertinama naudojant Ramsay metodą ir funkcinę erdvę

$$L_2(R, \phi) = \left\{ f : R \rightarrow R : \int_{-\infty}^{\infty} f^2(t) \phi(t) dt < \infty \right\},$$

$$\phi(t) = \exp\{-t^2/2\}, t \in R,$$

Duomenų parengimas ir žvalgomoji analizė:

$$Q := (Q(t), t \in R)$$

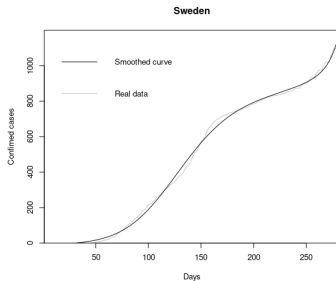
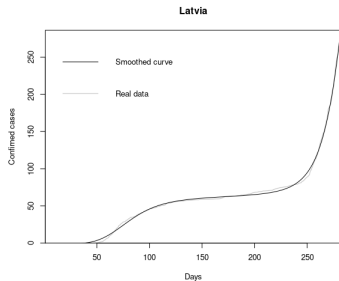
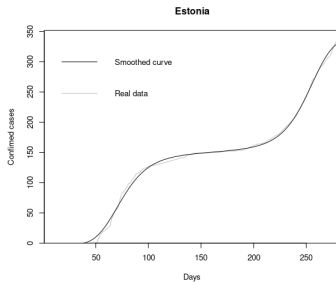
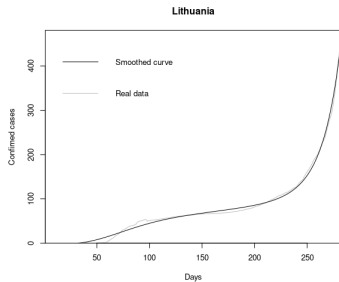
- ▶ $Q(t)$ - izoliuotų žmonių skaičius laiko momentu t .
- ▶ Funkcija Q yra atsitiktinė:
 - (a) $Q(t) = 0$, kai $t < T_0 = 0$;
 - (b) $Q(t)$ didėjanti intervale $[T_0, T_1]$.
- ▶ Laiko momentai T_0 ir T_1 bendru atveju yra nežinomi.
- ▶ Funkcija Q kiekvienai šaliai vertinama naudojant Ramsay metodą ir funkcinę erdvę

$$L_2(R, \phi) = \left\{ f : R \rightarrow R : \int_{-\infty}^{\infty} f^2(t) \phi(t) dt < \infty \right\},$$

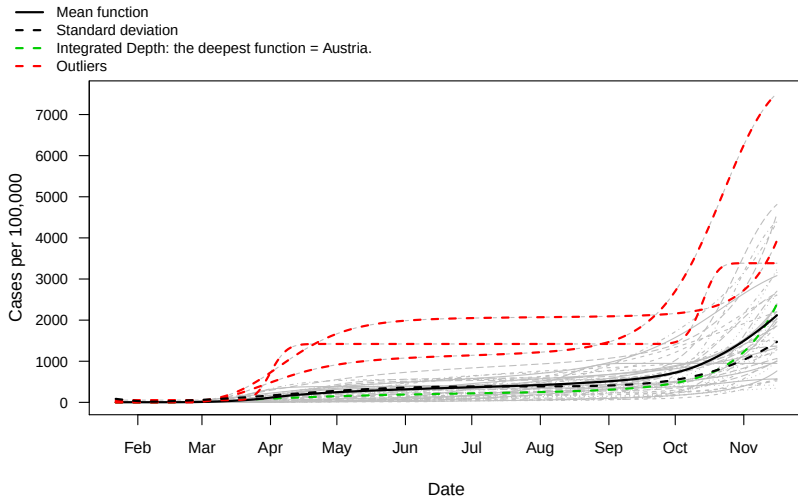
$\phi(t) = \exp\{-t^2/2\}$, $t \in R$, su skaliarine daugyba

$$\langle f, g \rangle := \int_{-\infty}^{\infty} f(t)g(t)\phi(t)dt.$$

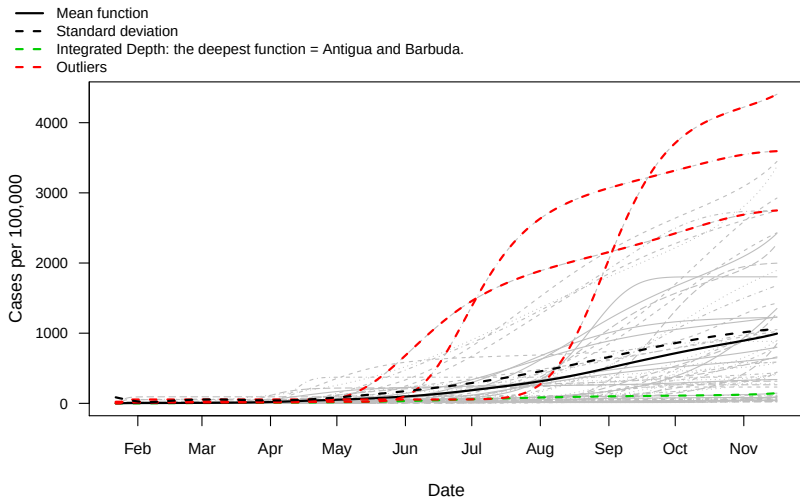
Pavyzdys



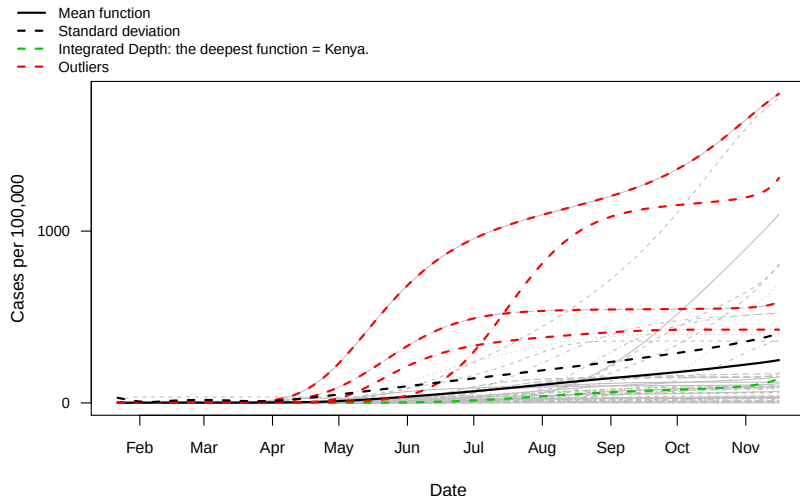
Pavyzdys: Europa



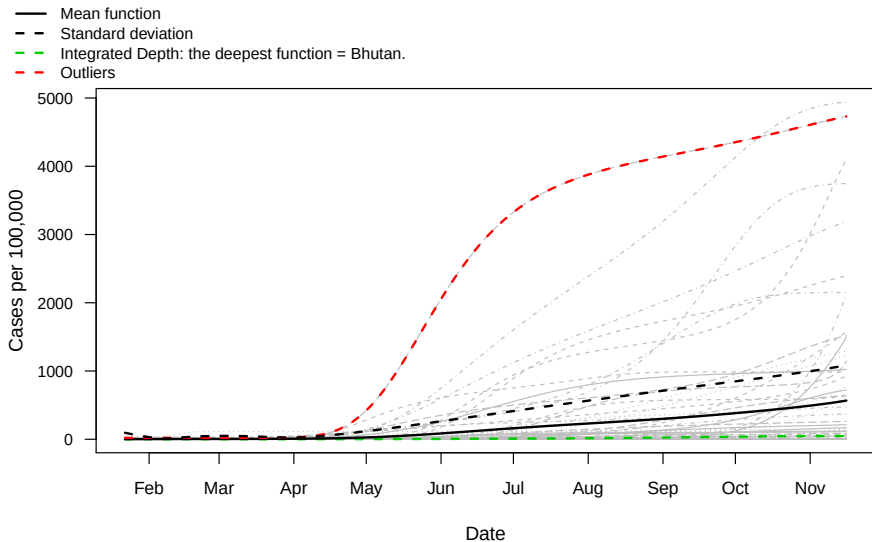
Išskirtys: Holy See, San Marino, Andora



Išskirtys: Chile, France French Guiana, Netherlands Aruba



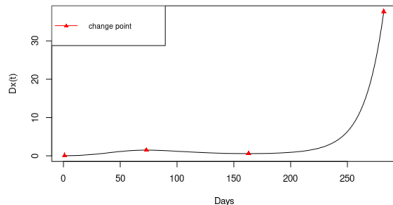
Išskirtys: France Mayotte, Djibouti, South Africa, Sao Tome and Principe



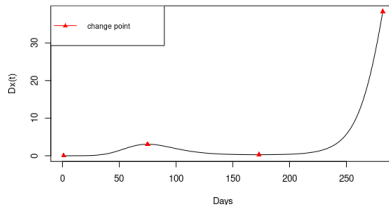
Išskirtys: Qatar

Pirmoji išvestinė: pavyzdys

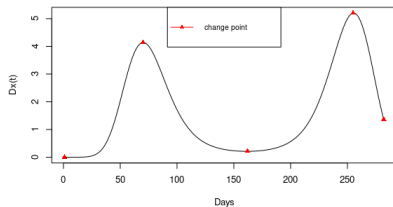
Lithuania



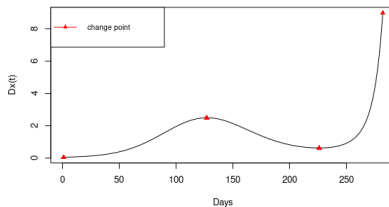
Latvia



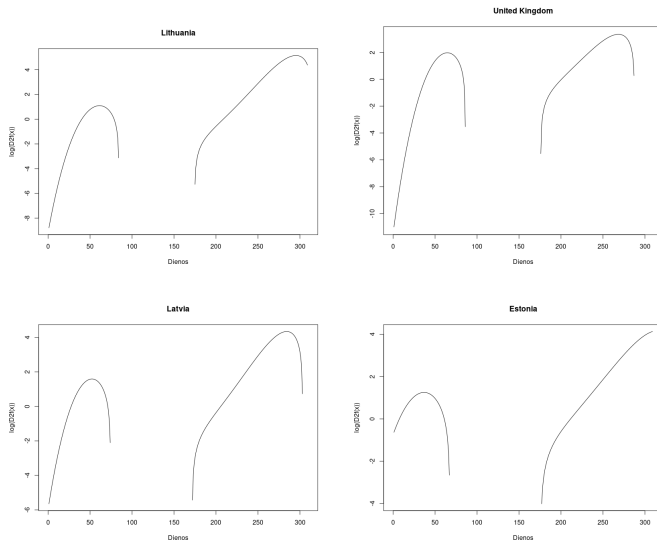
Estonia



Sweden



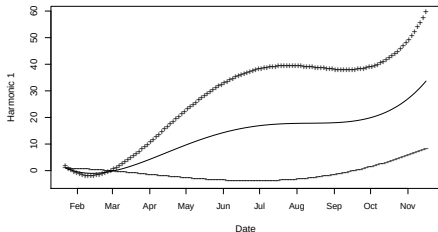
Antroji išvestinė: pavyzdys



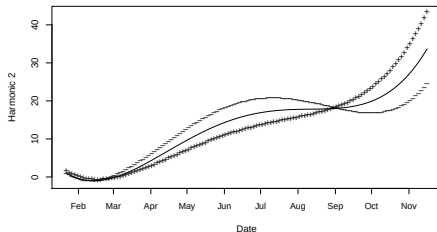
Pagrindinių komponentių analizė

- Atlikta pagrindinių komponentių analizė parodė, kad pirmos trys komponentės paaiškina 99.7% kintamumo.

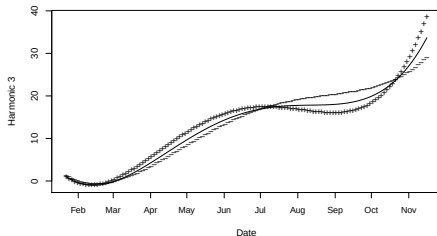
PCA function 1 (Percentage of variability 95.3)



PCA function 2 (Percentage of variability 3.9)



PCA function 3 (Percentage of variability 0.7)



- Jei $g_1(t), g_2(t), g_3(t)$ yra pirmos trys komponentės, tai $Q - EQ \sim (q_1, q_2, q_3)$:

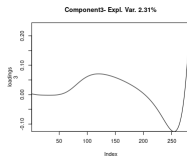
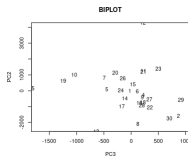
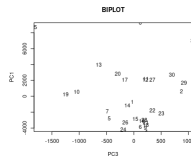
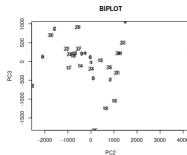
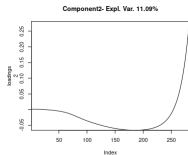
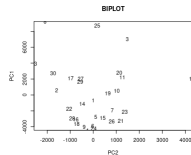
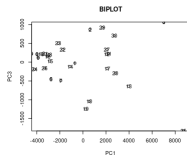
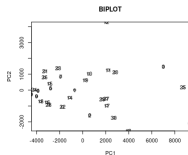
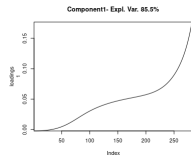
$$Q(t) \approx EQ(t) + q_1 g_1(t) + q_2 g_2(t) + q_3 g_3(t), \quad t \geq 0.$$

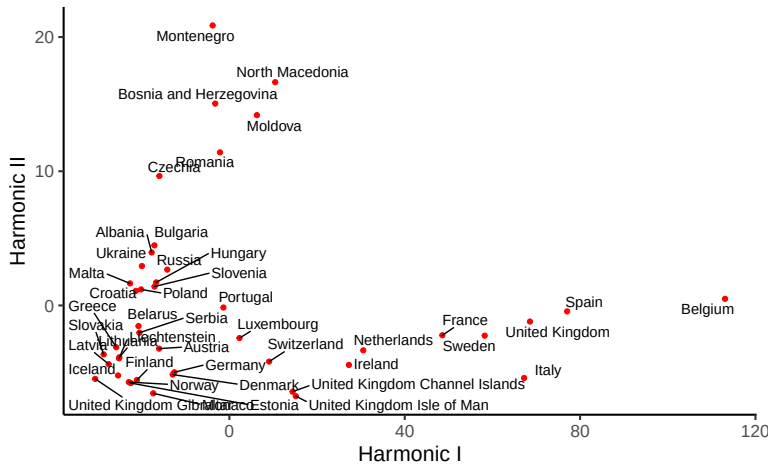
- Jei $g_1(t), g_2(t), g_3(t)$ yra pirmos trys komponentės, tai $Q - EQ \sim (q_1, q_2, q_3)$:

$$Q(t) \approx EQ(t) + q_1 g_1(t) + q_2 g_2(t) + q_3 g_3(t), \quad t \geq 0.$$

- q_i yra kreivės $Q - EQ$ projekcija į i 'tą komponentę.

Pagrindinės komponentės: Europa





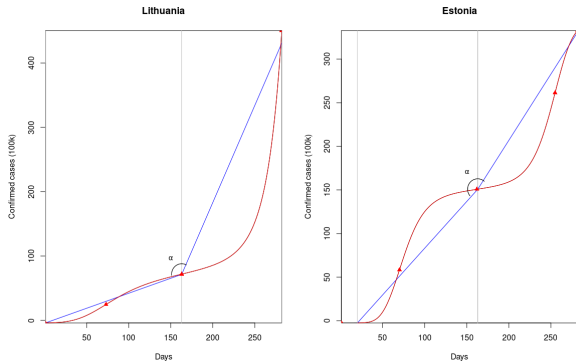
D=3

Išgijimas $\lambda(t)$

- ▶ t -fiksiuota diena;
- ▶ $R(t) = \int_{T_0}^t \lambda(s)Q(s)ds + \eta(t)$.

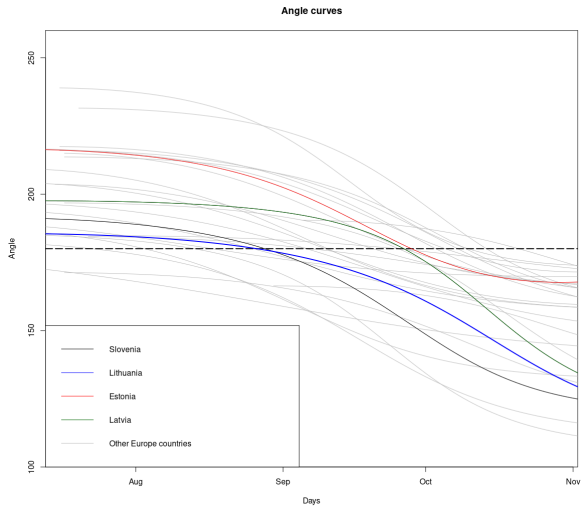
Lambda

Bangu palyginimas [Tadas Danielius]



BP

Bangu palyginimas

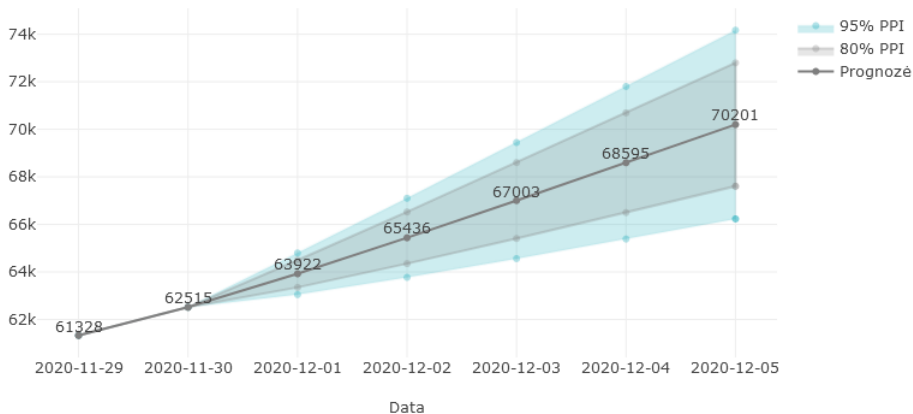


Trumpalaikės prognozės [Jurgita Markevičiūtė, Povilas Treigys]

- ▶ Nuo balandžio mėn. kasdien pradėtos teikti trumpalaikės 5 dienų COVID-19 patvirtintų atvejų prognozės. (Pristatyme naudojami duomenys iki lapkričio 30 d.)
- ▶ Prognozei naudojamas ARIMA modelis, kuris kasdien kinta.
- ▶ Prognozės skelbiamos: www.mif.vu.lt,
<https://www.vu.lt/covid-19/prognoze>,
www.covid19.projektas.vu.lt
- ▶ Vėliau pradėtos skelbti trumpalaikės prognozės
 - ▶ sergančiųjų skaičiui;
 - ▶ 14 dienų suminio skaičiaus 100 tūkst. gyventojų.
- ▶ Skaičiuojamos trumpalaikės savaitinės prognozės bendram mirčių skaičiui

Prognozų grafikai

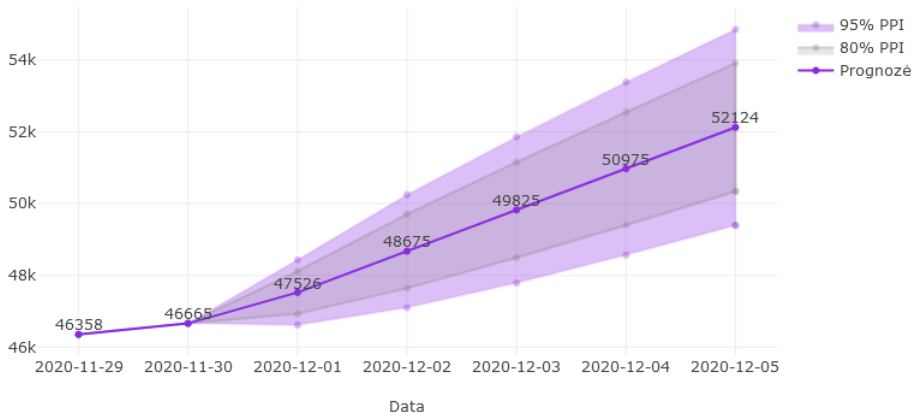
COVID-19 patvirtintų atvejų prognozė, atnaujinta 2020-12-01



PPI - prognozės pasikliautinis intervalas
(nuo mažiausios iki didžiausios prognozės reikšmės pagal 95 % arba 80 % pasitikėjimo lygį.)

Proгнозиų grafikai

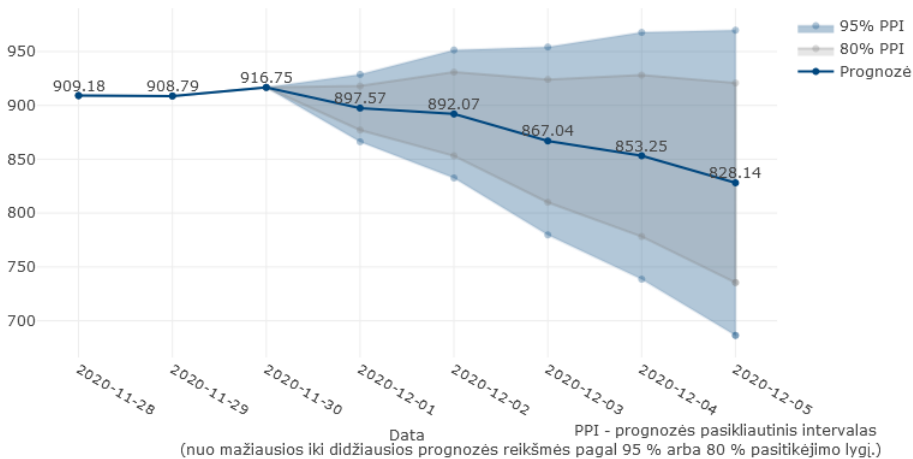
COVID-19 sergančių asmenų prognozė 2020-12-01



PPI - prognozės pasikliautinis intervalas
(nuo mažiausios iki didžiausios prognozės reikšmės pagal 95 % arba 80 % pasitikėjimo lygį.)

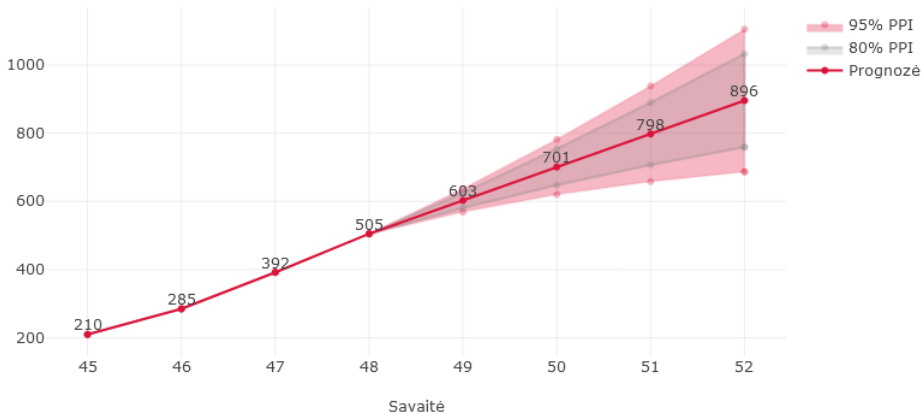
Proгнозиų grafikai

COVID-19 14 dienų suminio skaičiaus 100 tūkst. gyv. prognozė, atnaujinta 2020-12-01



Proгнозиų grafikai

Mirusiųjų dėl COVID-19 asmenų prognozė, atnaujinta 2020-12-01

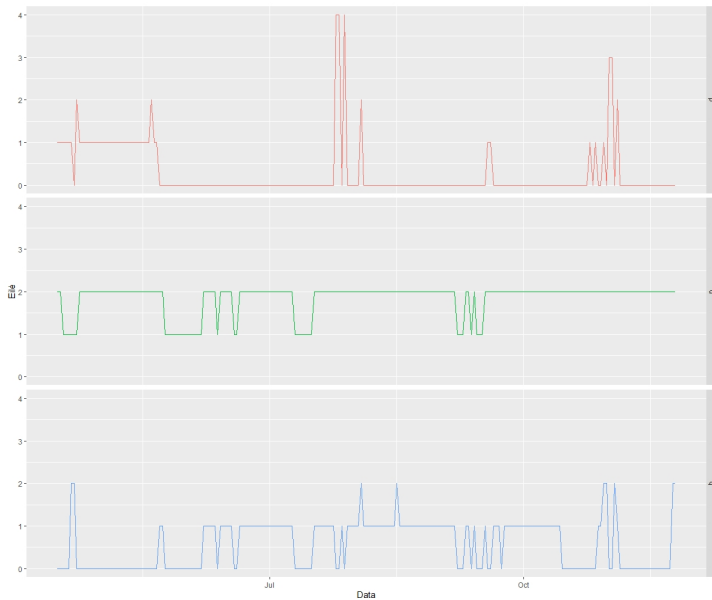


PPI - prognozės pasikliautinis intervalas
(nuo mažiausios iki didžiausios prognozės reikšmės pagal 95 % arba 80 % pasitikėjimo lygį.)

Patvirtintų atvejų prognozės analizė I

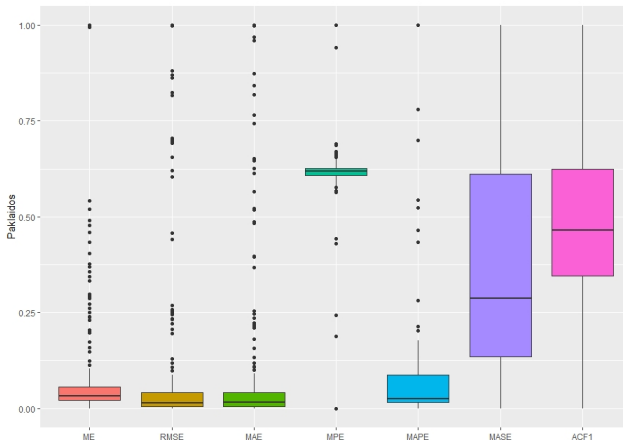
- ▶ Kasdien su atnaujintais duomenimis (papildytais ir atnaujinta istorija) sudaromas ARIMA modelis.
- ▶ Kadangi situacija kinta sparčiai, modelio sudarymui imamos 35 paskutinės dienos. (Paveiksle parodytos ARIMA modelio eilės)

Patvirtintų atvejų prognozės analizė II



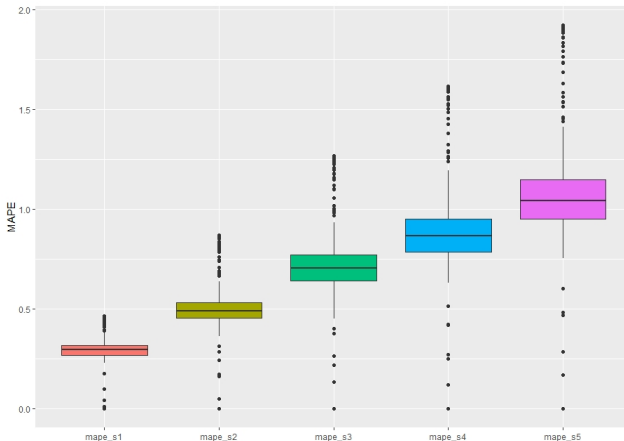
Patvirtintų atvejų prognozės analizė III

Modelių paklaidos kai kuriomis dienomis itin išaugdavo dėl susirgimų rekordų. Dinamikos palyginamumui paklaidos sunormuotos į intervalą $[0, 1]$:



Patvirtintų atvejų prognozės analizė IV

Prognozės paklaidos didėja, didėjant prognozės žingsnių skaičiui.
Paveiksle pateikta MAPE kiekvienam modeliui kiekvienam iš 5 žingsnių:



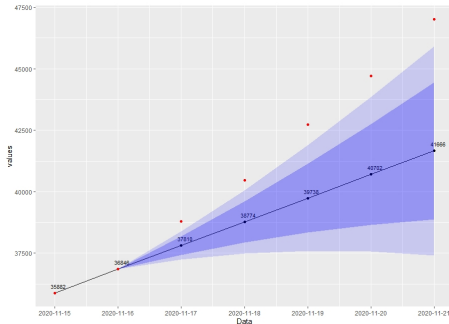
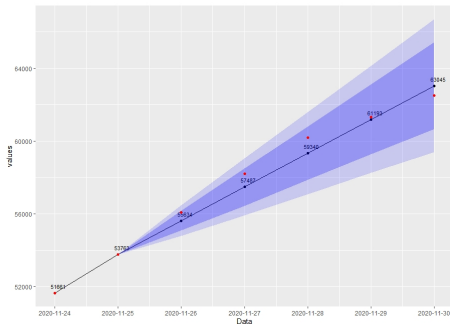
Patvirtintų atvejų prognozės analizė V

Apskaičiuota prognozės tikimybė, t.y., kiek kartų tikroji reikšmė buvo 95% bei 80% prognozės intervalo ribose kiekvienam iš žingsnių:

	80% prognozės intervalas	95% prognozės intervalas
1 žingsnio prognozė	0.71	0.83
2 žingsnių prognozė	0.66	0.82
3 žingsnių prognozė	0.64	0.82
4 žingsnių prognozė	0.64	0.79
5 žingsnių prognozė	0.63	0.79

Patvirtintų atvejų prognozės analizė VI

Juoda linija ir juodi taškai žymi prognozuojamas reikšmes, raudoni taškai žymi tikrąsias reikšmes, tamsi mėlyna juosta žymi 80% prognozės intervalą, šviesi mėlyna juosta žymi 95% prognozės intervalą.



Klasterinė analizė [Jolita Bernatavičienė, Viktor Medvedev, Julius Venskys, Povilas Treigys]

- ▶ Analizuotos 36 Europos šalys: 27 Europos Sąjungos šalys, Jungtinė karalystė, 4 EFTA šalys ir 4 šalys kandidatės į ES.

Klasterinė analizė [Jolita Bernatavičienė, Viktor Medvedev, Julius Venskys, Povilas Treigys]

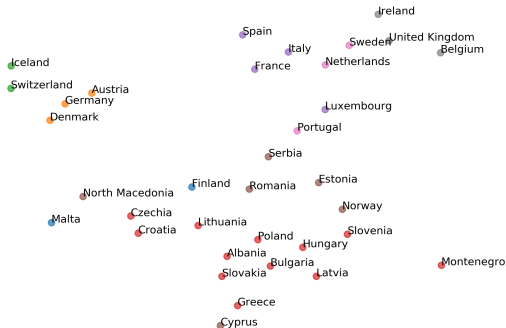
- ▶ Analizuotos 36 Europos šalys: 27 Europos Sąjungos šalys, Jungtinė karalystė, 4 EFTA šalys ir 4 šalys kandidatės į ES.
- ▶ Analizuoti įvairūs periodai. Pvz., 2020.04.01 – 2020.04.30 ir 2020.10.01 – 2020.10.25.
- ▶ Klasterizuojama pagal santykinius dydžius (100 000 gyventojų) (identifikuotus, mirties, pasveikusiųjų atvejus ir populiacijos tankį).

Klasterinė analizė [Jolita Bernatavičienė, Viktor Medvedev, Julius Venskys, Povilas Treigys]

- ▶ Analizuotos 36 Europos šalys: 27 Europos Sąjungos šalys, Jungtinė karalystė, 4 EFTA šalys ir 4 šalys kandidatės į ES.
- ▶ Analizuoti įvairūs periodai. Pvz., 2020.04.01 – 2020.04.30 ir 2020.10.01 – 2020.10.25.
- ▶ Klasterizuojama pagal santykinius dydžius (100 000 gyventojų) (identifikuotus, mirties, pasveikusiųjų atvejus ir populiacijos tankį).
- ▶ Naudoti MDS algoritmas, t-SNE algoritmas, hierarchinis klasterizavimas bei Savireguliuojantis tinklas SOM.

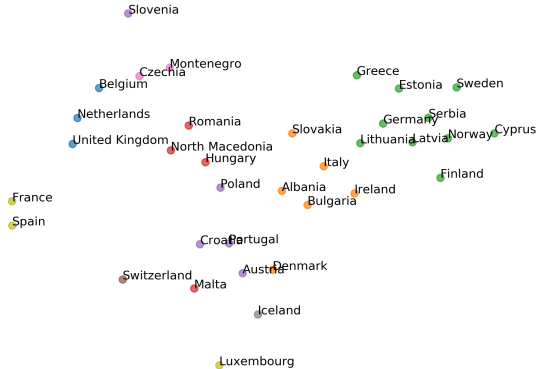
SOM + t-SNE algoritmas

T-SNE. Periodas nuo 2020-04-01 iki 2020-04-30



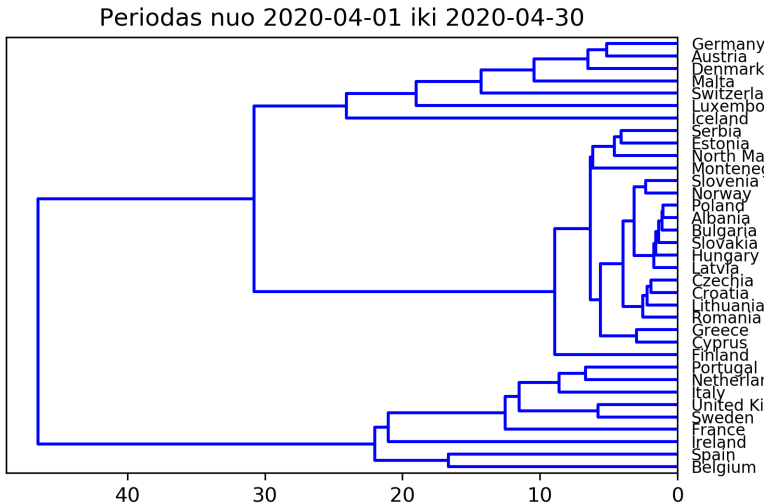
- Balandis: viename klasteryje su Lietuva 11 valstybių: Graikija, Norvegija, Slovėnija, Latvija, Vengrija, Bulgarija, Albanija, Lenkija, Kroatija ir Čekija.

T-SNE. Periodas nuo 2020-10-01 iki 2020-10-25

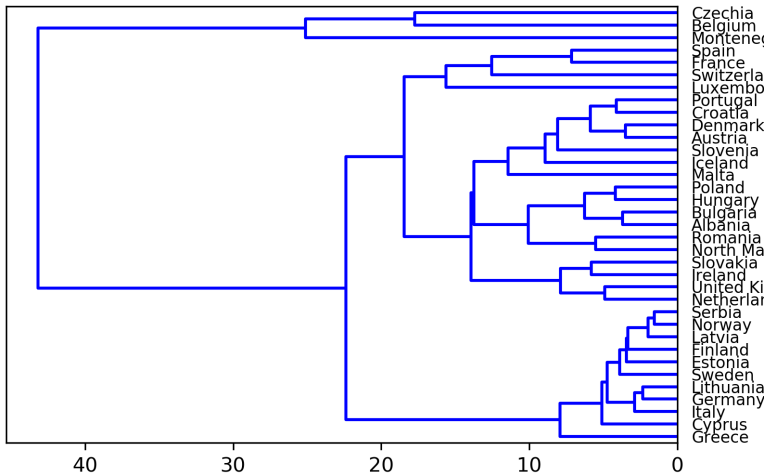


- Spalis: viename klasteryje su Lietuva yra Graikija, Latvija, Norvegija, Vokietija, Estija, Švedija, Suomija, Kipras

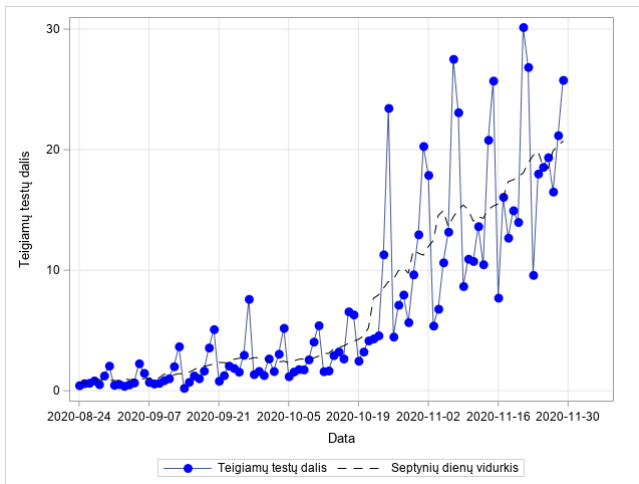
Hierarchinio klasterizavimo algoritmas



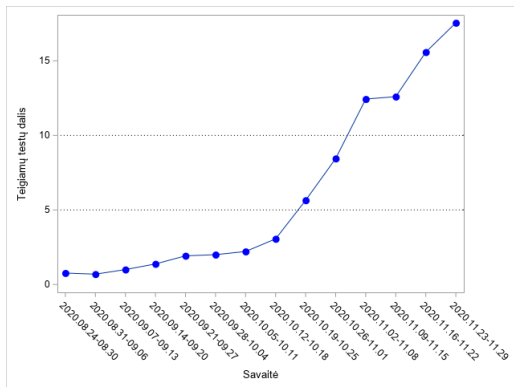
Periodas nuo 2020-10-01 iki 2020-10-25



Testų analizė [Rūta Levulienė]



Teigiamų testų dalis ir septynių dienų vidurkis.



Teigiamų testų dalis nagrinėjant suminius savaitės duomenis.

- ▶ Jeigu stebimas staigus teigiamų testų dalies padidėjimas, situacija blogėja ir reikia imtis priemonių, o jeigu skaičiai panašūs, situacija stabilizavosi.
- ▶ 2020.11.02-11.08 ir 2020.11.09-11.15 teigiamų testų dalis yra panaši, tačiau 2020.11.16-11.22 vėl stebimas staigus kilimas.