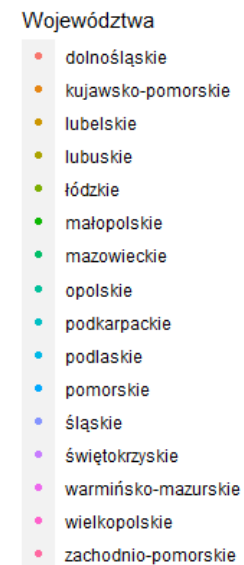
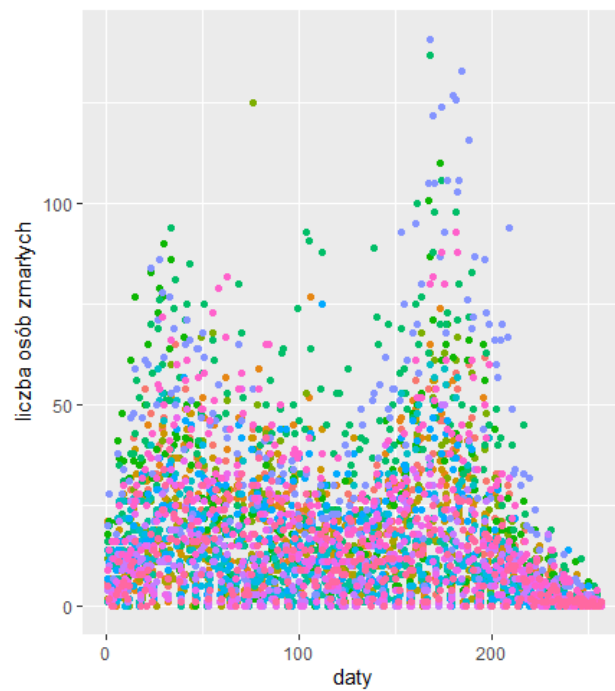
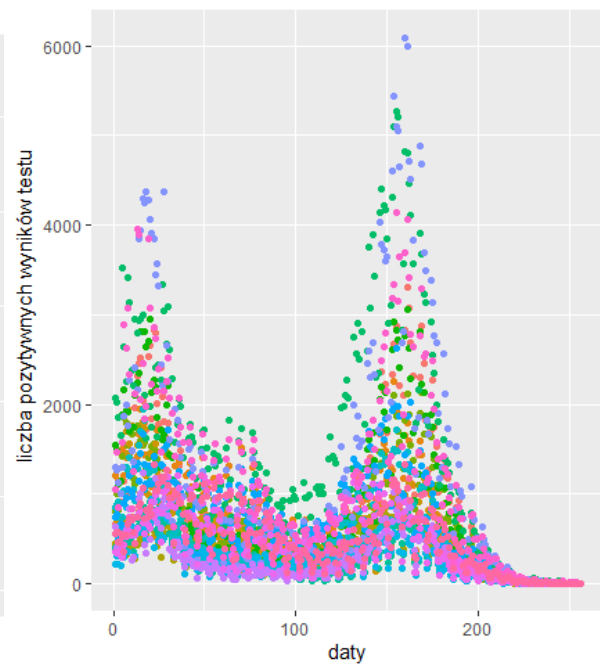
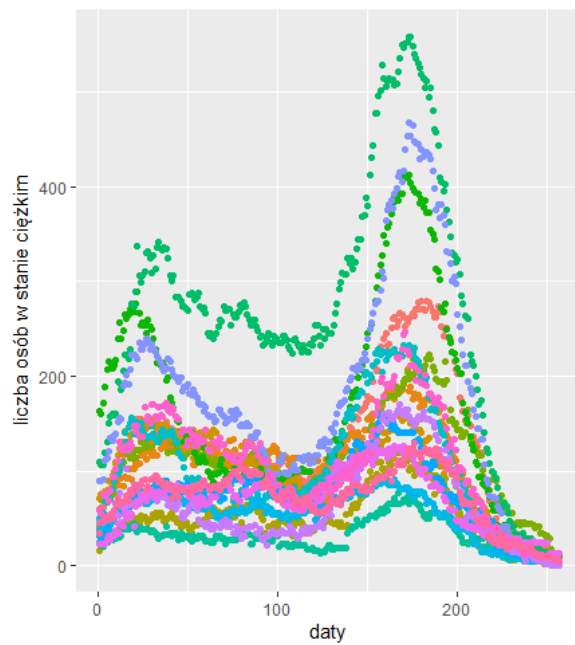
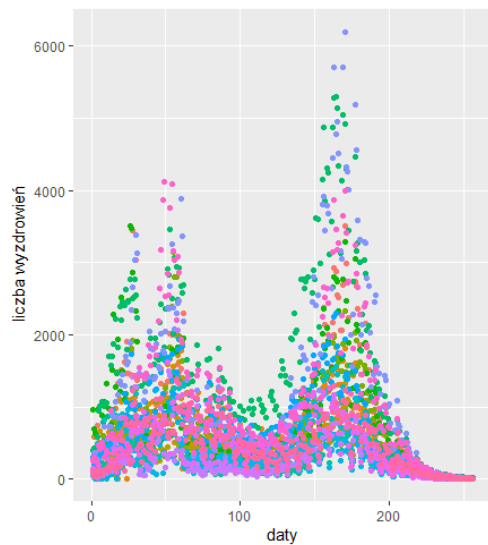
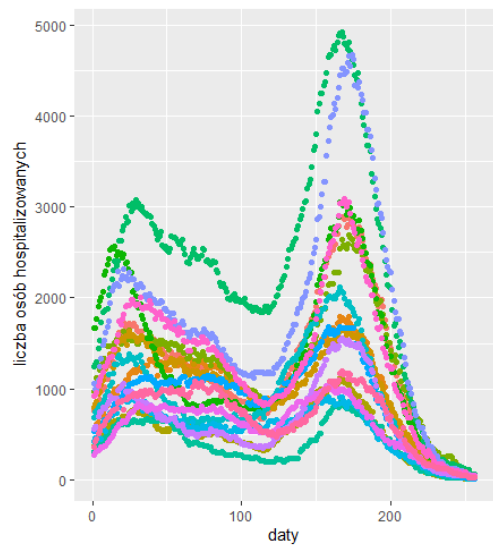
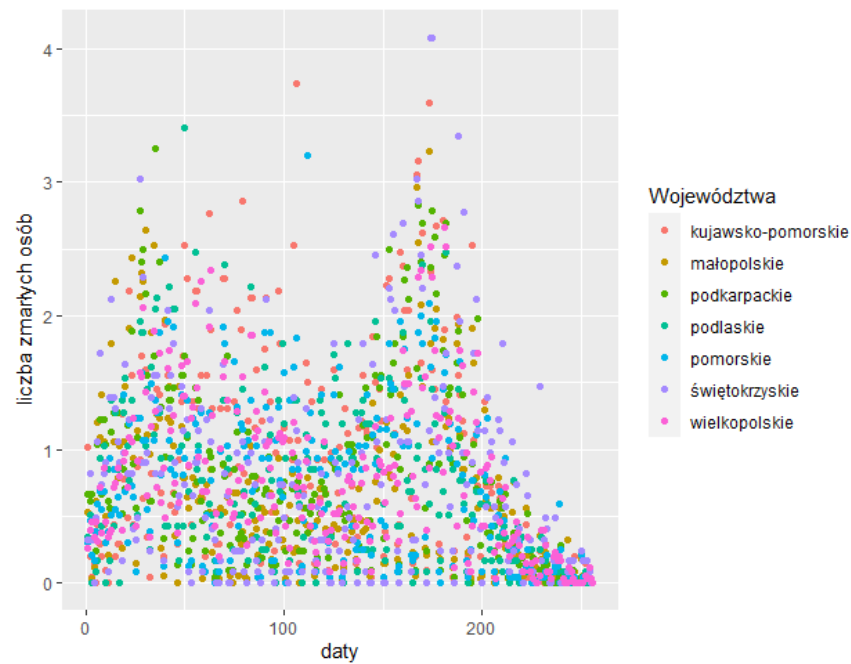
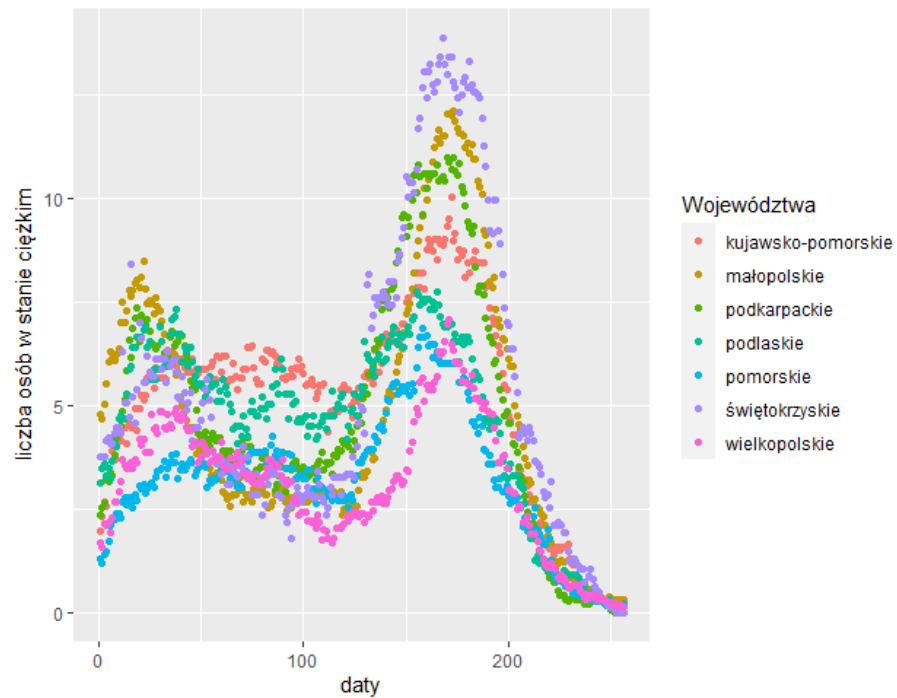
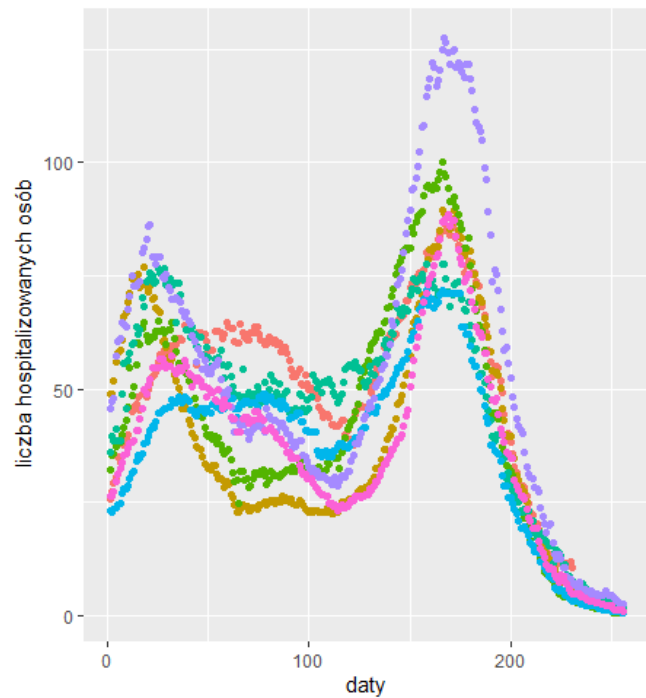


Predykcja liczby osób  
hospitalizowanych oraz osób w stanie  
ciężkim dla drugiej i trzeciej fali  
COVID-19

Patrycja Hęćka

# 1. Analizowane dane.





## 2. Rozszerzanie bazy.

Rozważmy wektory  $x_i(t) : i = 1, \dots, n; t \in T$  oraz załóżmy, że obserwacje  $y_{ik}$  są dostępne dla każdego węzła  $t_{i1}, t_{i2}, \dots, t_{im_i} \in T$ . Wówczas każdą obserwację  $y_{ik}$  możemy zapisać następująco:

$$y_{ik} = x_i(t_{ik}) + \epsilon_{ik}, i = 1, \dots, n; k = 1, \dots, m_i, \quad (1)$$

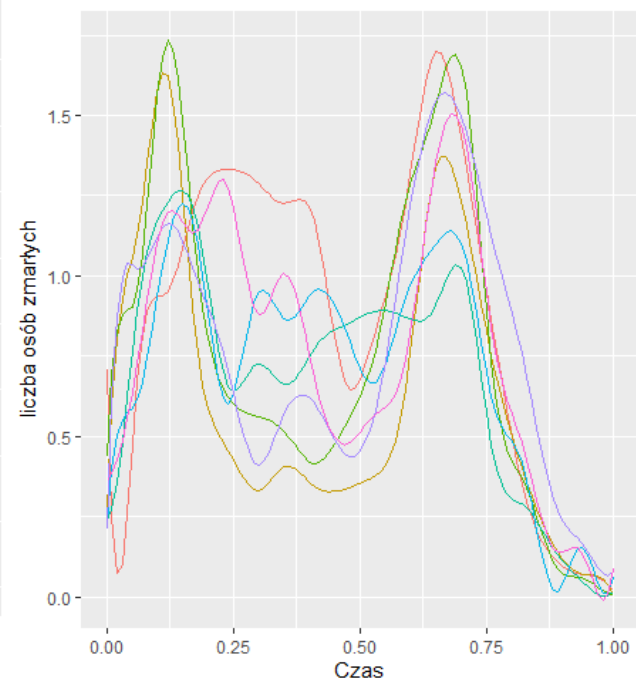
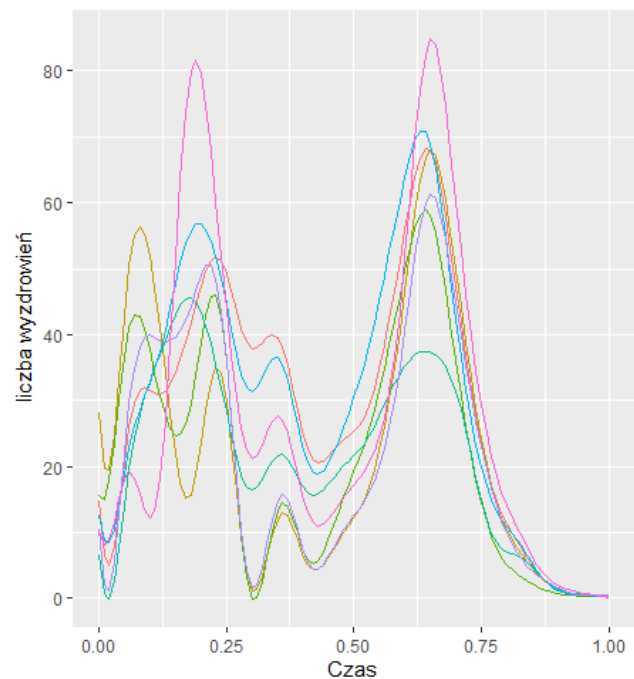
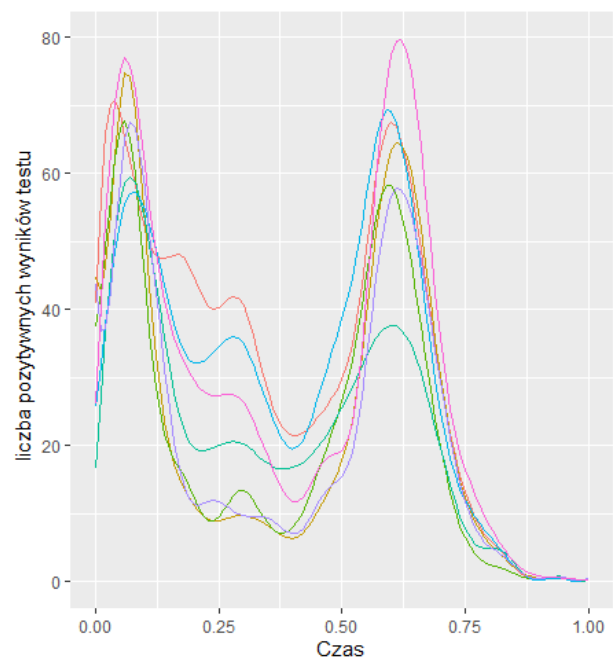
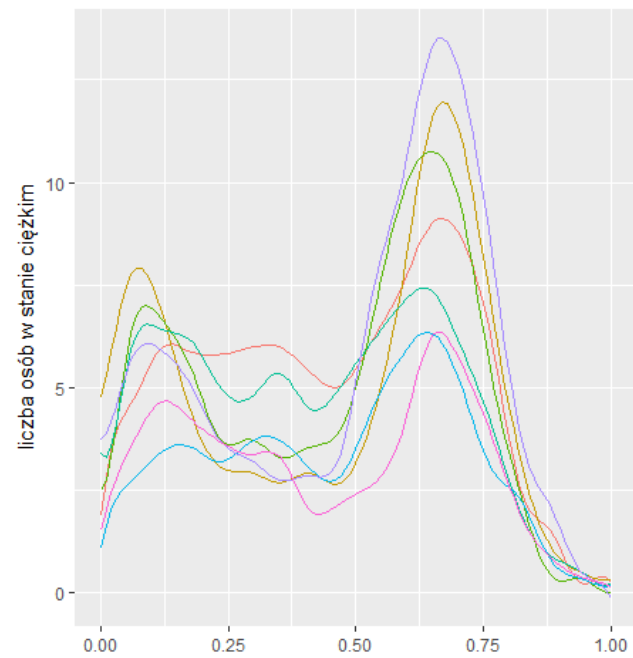
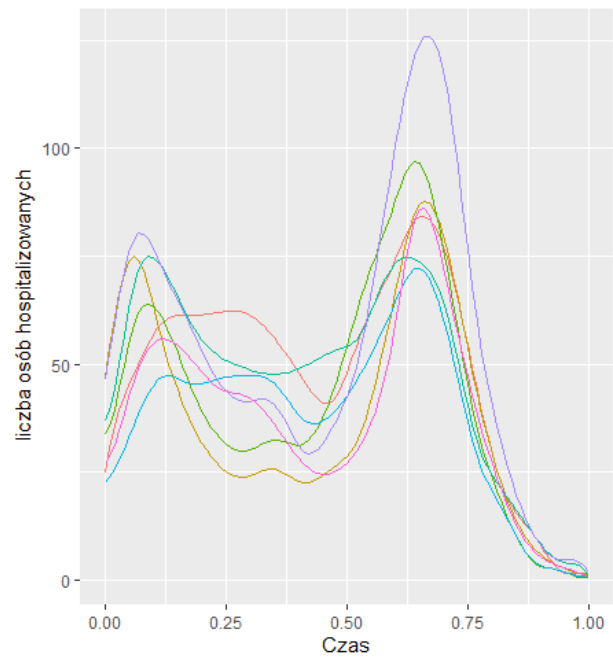
Założmy, że próbkowe krzywe należą do skończenie-wymiarowej przestrzeni generowanej przez zestaw funkcji bazowych  $\{\phi_1(t), \dots, \phi_p(t)\}$ . Wtedy możemy przedstawić każdą krzywą  $x_i$  poprzez liniową rozbudowę postaci:

$$x_i(t) = \sum_{j=1}^p \alpha_{ij} \phi_j(t), i = 1, \dots, n. \quad (2)$$

$$\hat{\alpha}_i = (\Phi_i' \Phi_i)^{-1} \Phi_i' y_i.$$

$$\Phi_i = (\phi_j(t_{ik}))_{m_i \times p}$$

$$\hat{y}_i = \Phi_i \hat{\alpha}_i = \Phi_i (\Phi_i' \Phi_i)^{-1} \Phi_i' y_i.$$



### 3. Multiple Function-on-Function Linear Model

$$y_i(t) = \alpha(t) + \sum_{j=1}^J \int_T x_{ij}(s) \beta_j(s, t) ds + \epsilon_i(t), i = 1, \dots, n,$$

$$x_{ij}(t) = \bar{x}_j(t) + \sum_{l=1}^{n-1} \xi_{il}^{x_j} f_l^{x_j}(t),$$

$$y_i(t) = \bar{y}(t) + \sum_{l=1}^{n-1} \xi_{il}^y f_l^y(t),$$

$$\xi_{ik}^y = \sum_{j=1}^J \sum_{l=1}^{n-1} b_{kl}^{x_j} \xi_{il}^{x_j} + \epsilon_{ik}, i = 1, \dots, n; k = 1, \dots, n - 1.$$

$$\hat{y}_i(s) = \bar{y}(s) + \sum_{k=1}^K \hat{\xi}_{ik}^y f_k^y(s) = \bar{y}(s) + \sum_{k=1}^K \left( \sum_{j=1}^J \sum_{l \in L_{kj}} \hat{b}_{kl}^{x_j} \hat{\xi}_{il}^{x_j} \right) f_k^y(s)$$



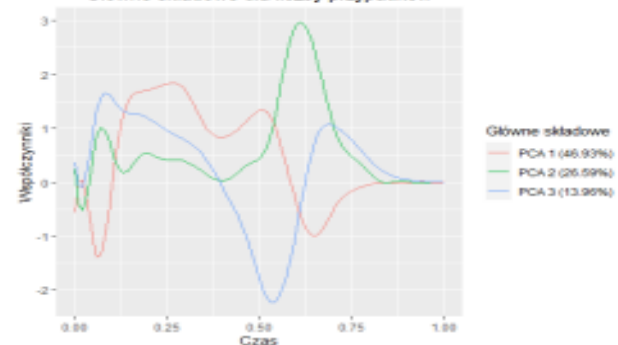
## 4. Analiza głównych składowych funkcjonalnych

Pierwsze główne składowe wyjaśniają kolejno 46.93%, 42.27%, 38.78%, 41.17%, 44.30% zmienności predyktorów X1, X2, X3, Y1, Y2 z próby treningowej.

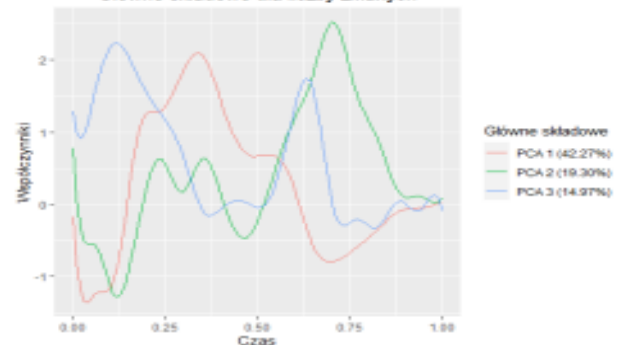
Drugie główne składowe wyjaśniają kolejno: 26.59%, 19.30%, 26.79%, 29.41%, 31.70% wariancji.

Trzecie główne składowe objaśniają znacznie mniej, odpowiednio: 13.96%, 14.97%, 18.91%, 15.45%, 12.71%.

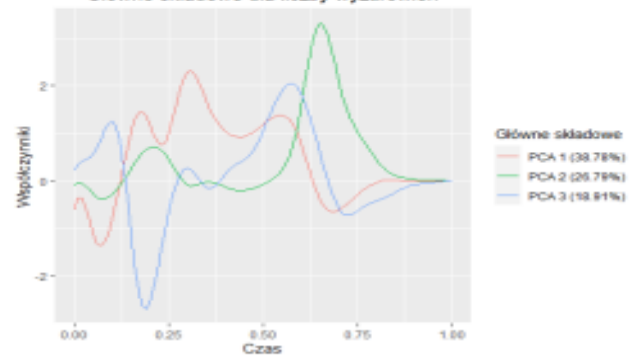
Główne składowe dla liczby przypadków



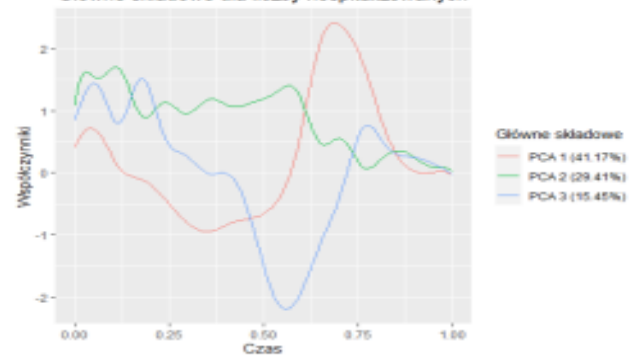
Główne składowe dla liczby zmarłych



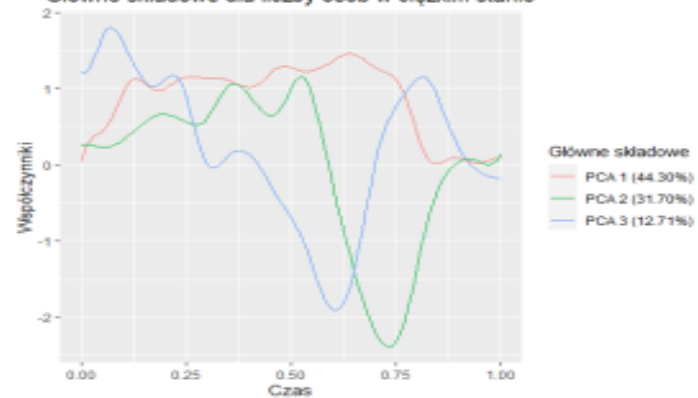
Główne składowe dla liczby wyzdrowień

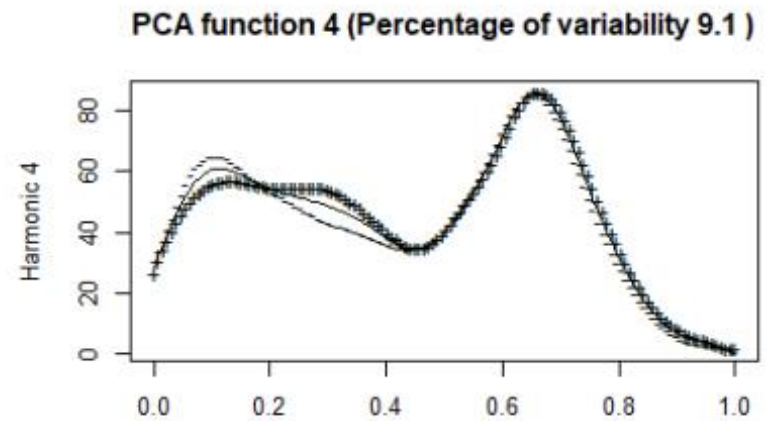
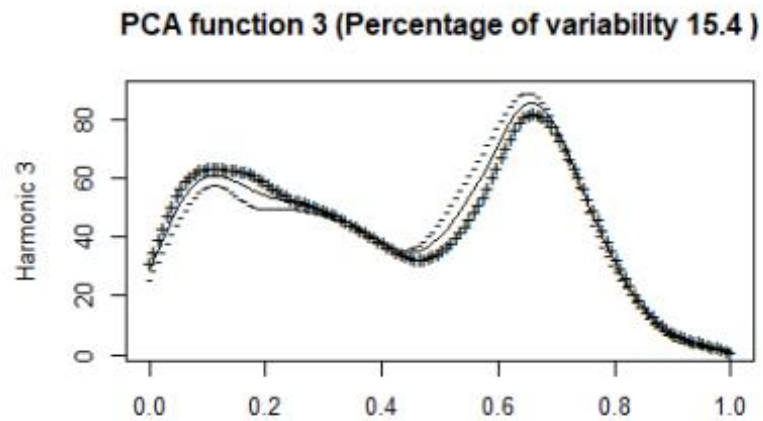
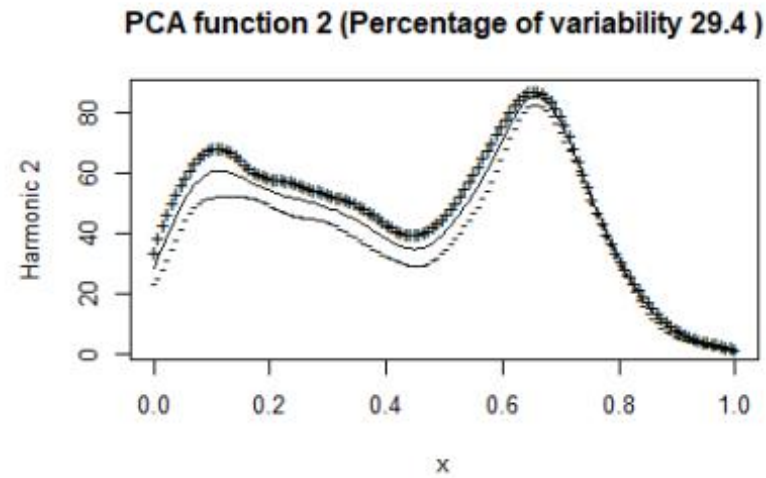
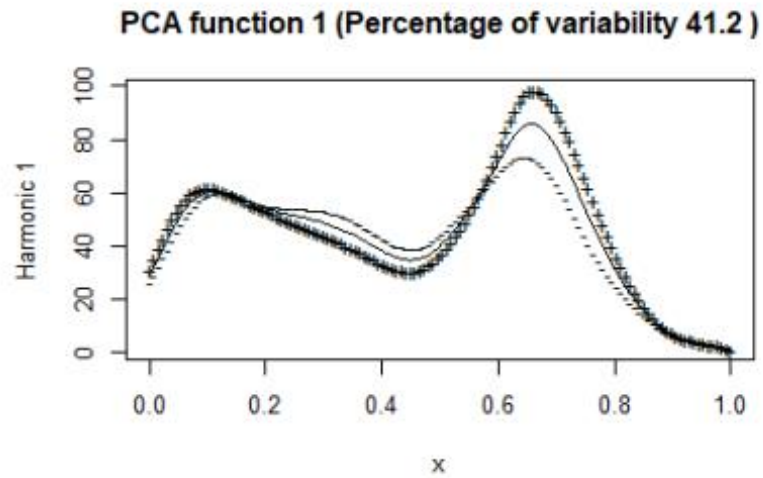


Główne składowe dla liczby hospitalizowanych



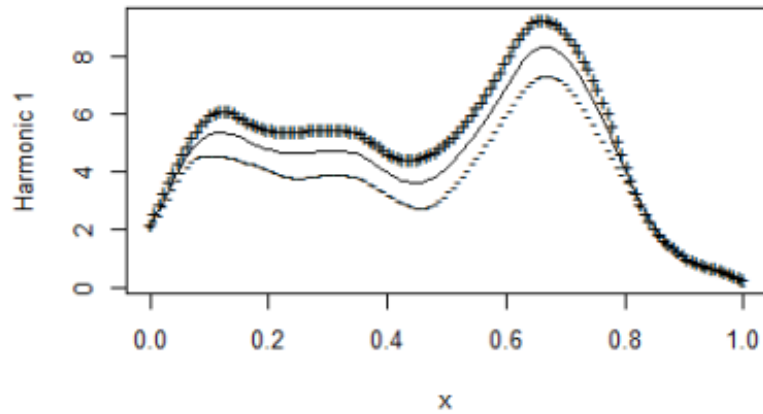
Główne składowe dla liczby osób w ciężkim stanie



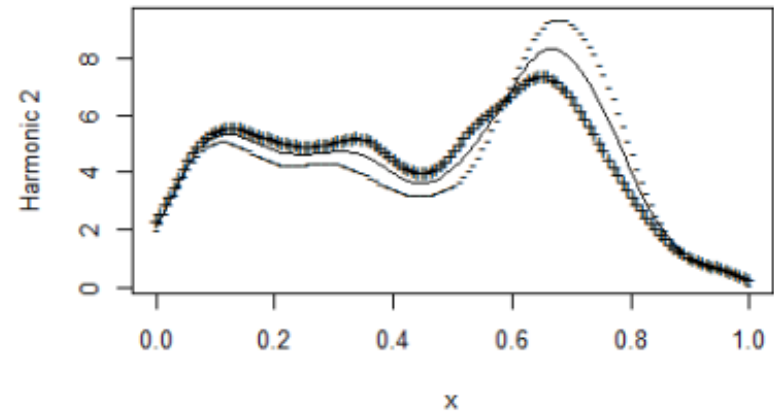


Średnia liczba osób hospitalizowanych wraz z krzywymi będącymi wynikiem dodawania i odejmowania odpowiednio przeskalowanych współczynników harmoniczných od średniej.

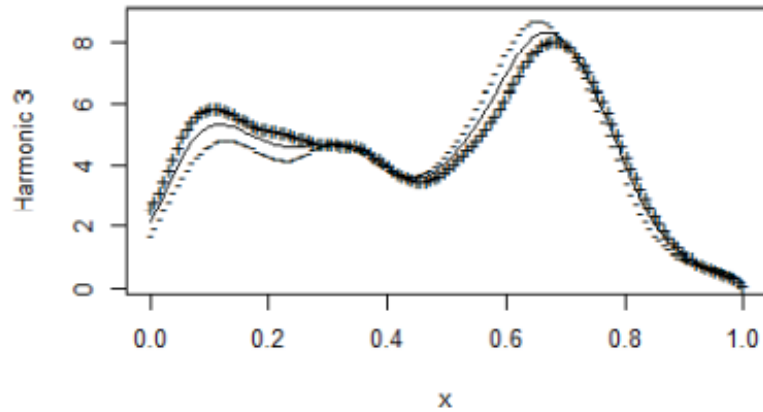
**PCA function 1 (Percentage of variability 44.3)**



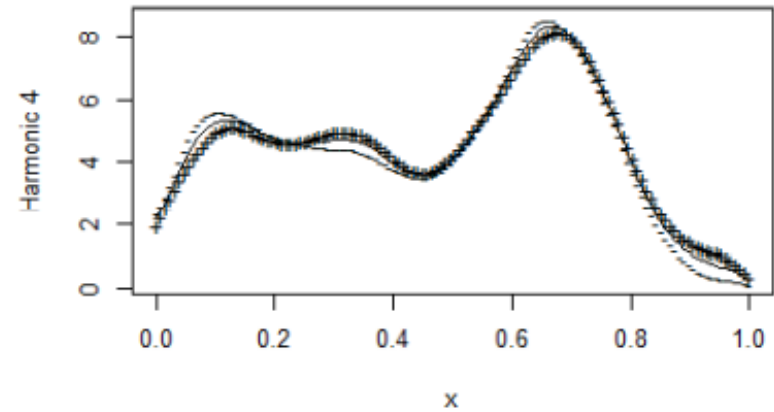
**PCA function 2 (Percentage of variability 31.7)**



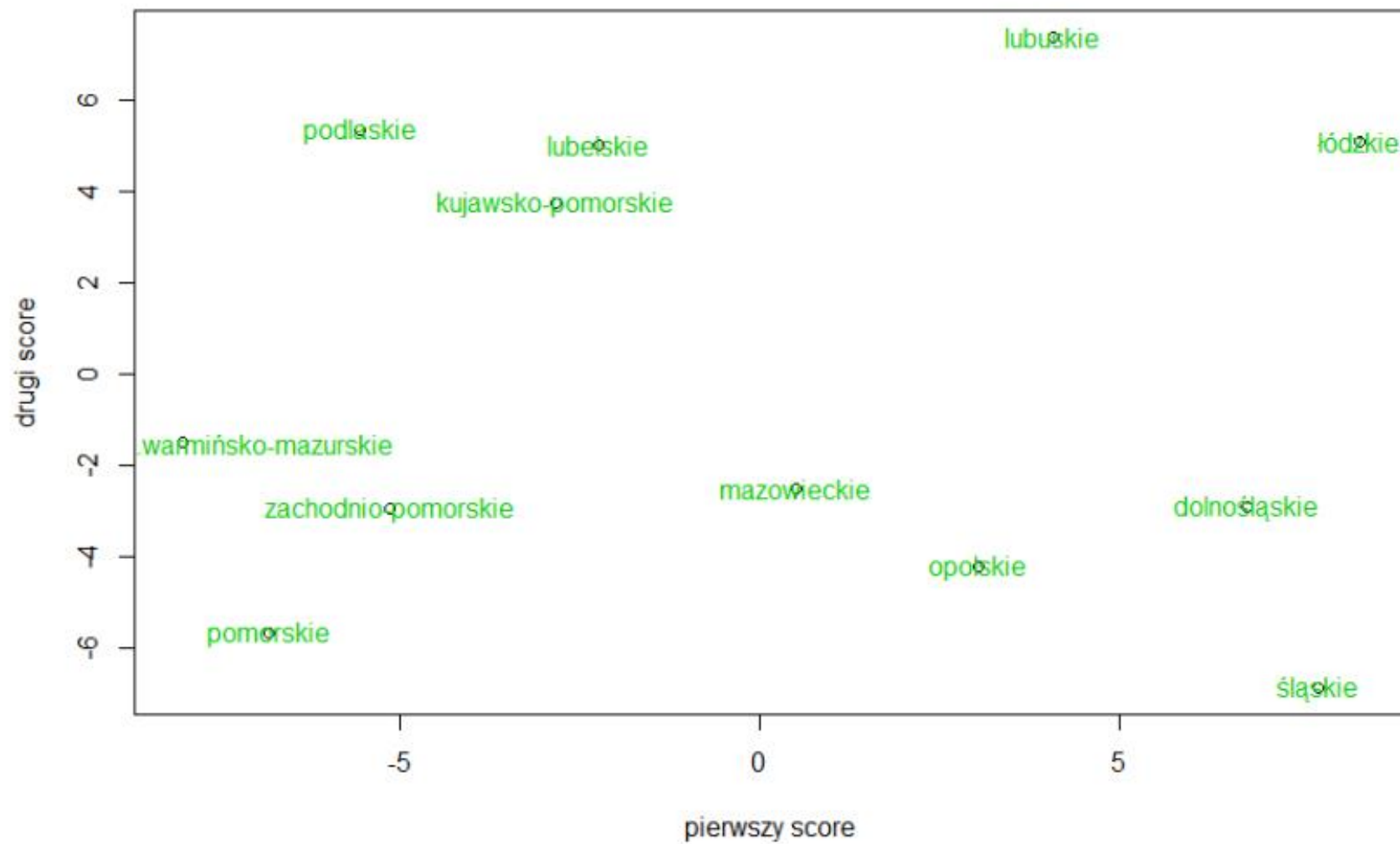
**PCA function 3 (Percentage of variability 12.7)**



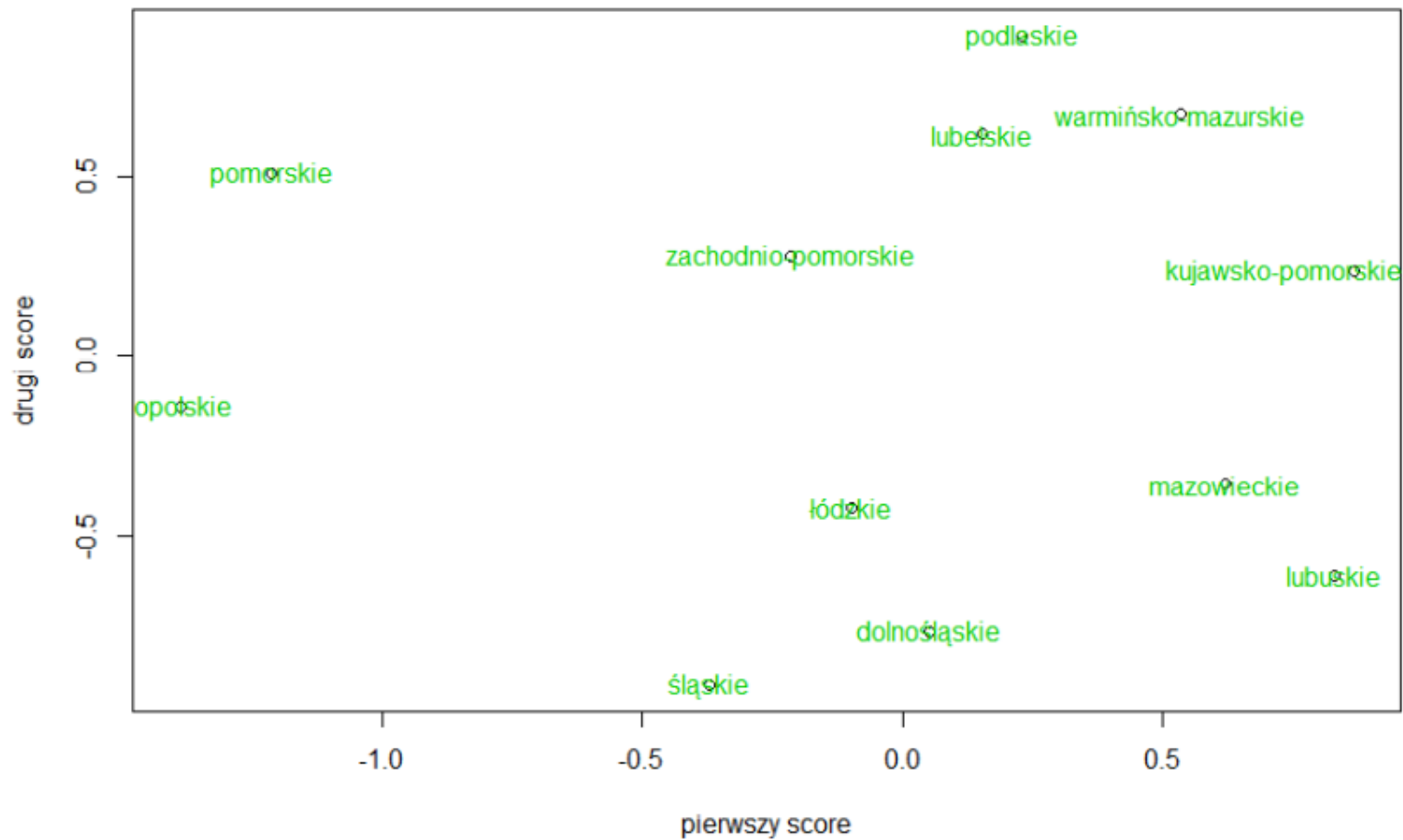
**PCA function 4 (Percentage of variability 4.9)**



Średnia liczba osób w stanie ciężkim wraz z krzywymi będącymi wynikiem dodawania i odejmowania odpowiednio przeskalowanych współczynników harmoniczných od średniej.



Wartości pierwszego i drugiego score dla liczby osób hospitalizowanych.



Wartości pierwszego i drugiego score dla liczby osób w stanie ciężkim.

# 5. Predykcja

$$\hat{\xi}_{i1}^{y_k} = \gamma_0 + \xi_{i1}^{x_1} \gamma_1^{y_k} + \xi_{i1}^{x_2} \gamma_2^{y_k} + \xi_{i1}^{x_3} \gamma_3^{y_k} + \epsilon_i^{y_k}, \quad k = 1, 2, i = 1, \dots, 16. \quad (13)$$

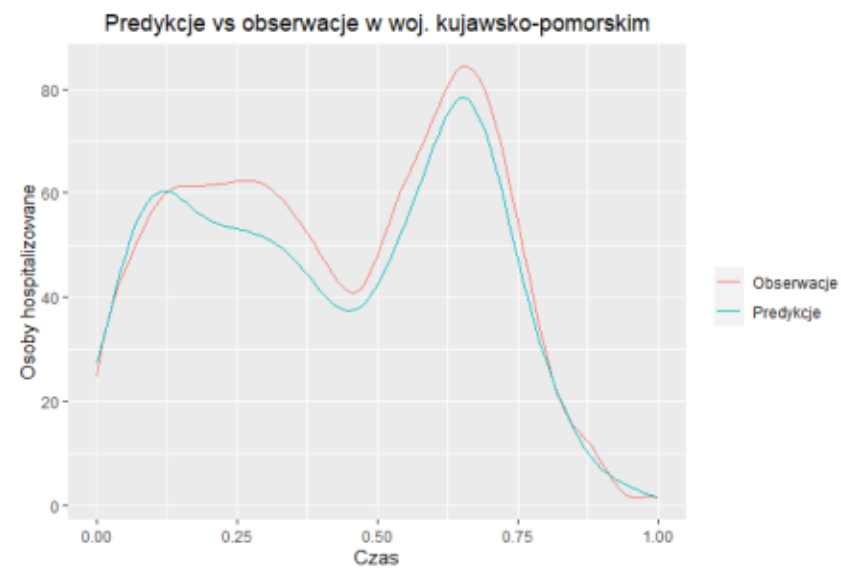
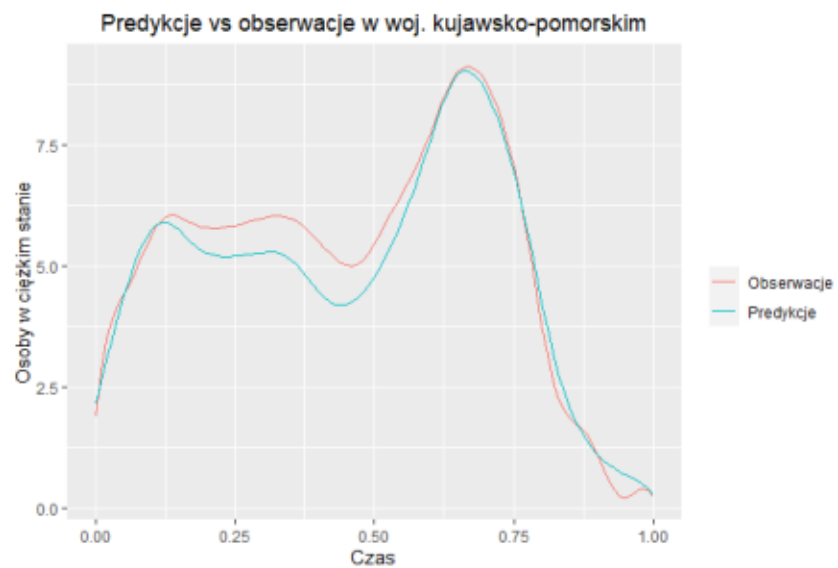
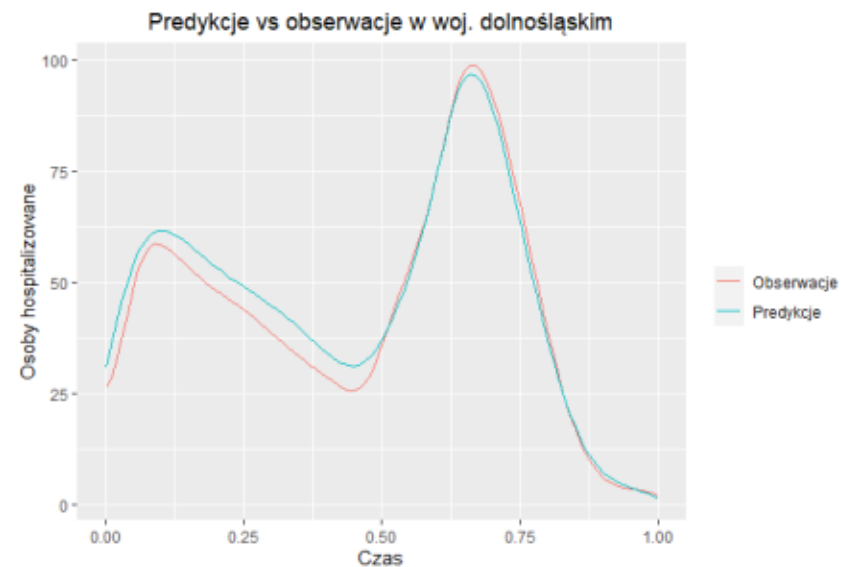
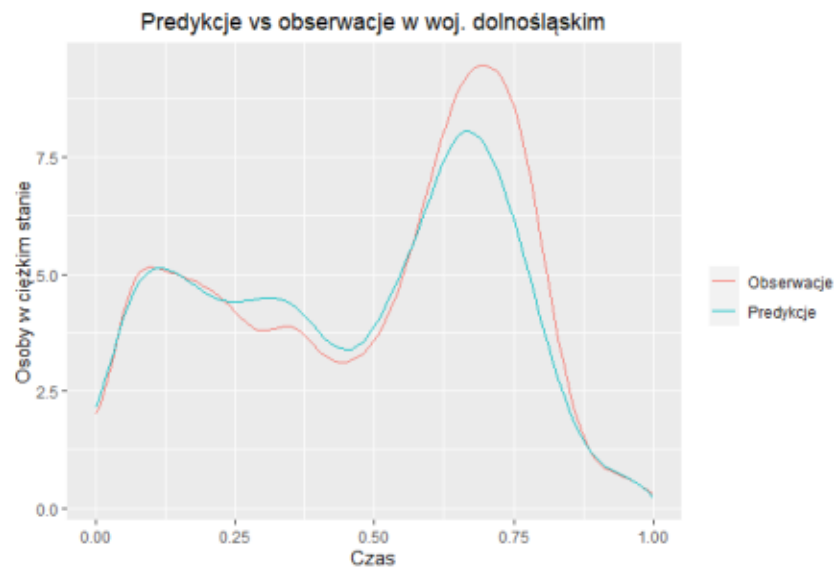
Na podstawie takich modeli dokonamy estymacji pierwszych składowych  $Y_1(t)$  i  $Y_2(t)$  z pierwszych składowych  $X_1(t)$ ,  $X_2(t)$  i  $X_3(t)$ .

Rozbudowa Karhunen-Loève w terminach predyktorów głównych składowych pozwoli na predykcję  $Y_1(t)$  i  $Y_2(t)$  za pomocą następującego równania:

$$\hat{y}_{ik}(t) = \bar{y}_k(t) + \hat{\xi}_{i1}^{y_k} f_1^{y_k}(t), \quad k = 1, 2, i = 1, \dots, 16. \quad (14)$$

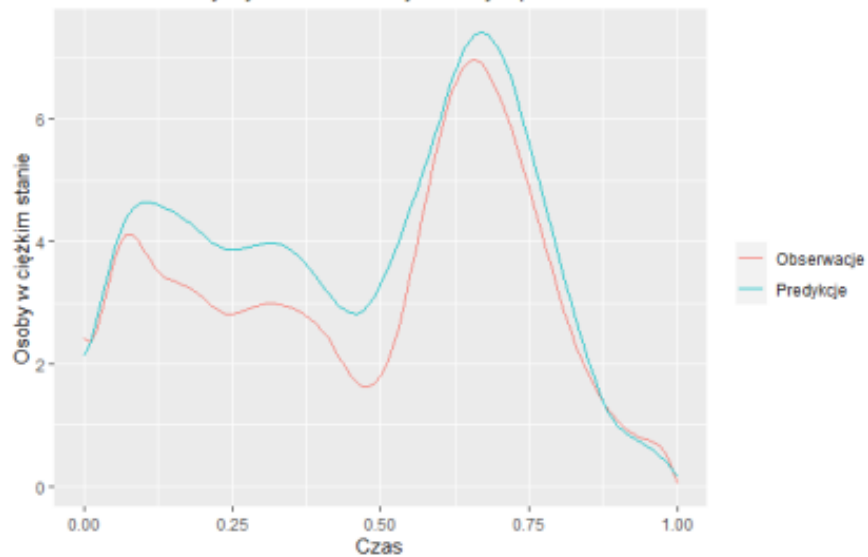
W celu przetestowania modeli na próbie treningowej posłużymy się błędem średniokwadratowym, wyrażonym równaniem:

$$MSE(y_{ik}) = \int_0^1 (y_{ik}(t) - \hat{y}_{ik}(t))^2 dt, \quad k = 1, 2, i = 1, \dots, 12. \quad (15)$$

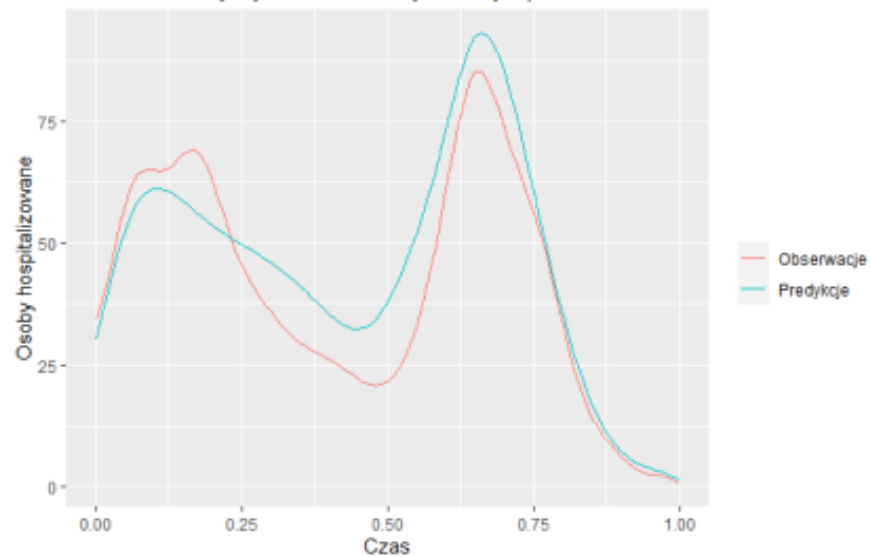




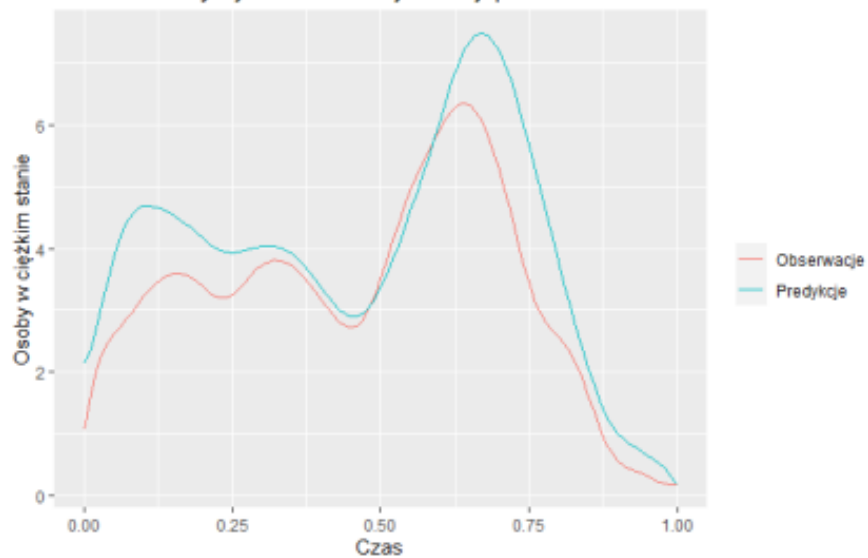
Predykcje vs obserwacje w woj. opolskim



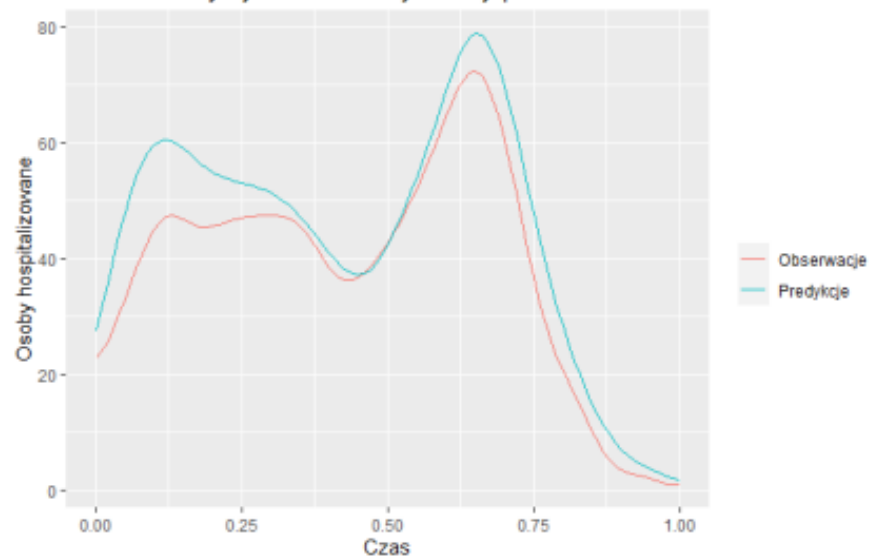
Predykcje vs obserwacje w woj. opolskim



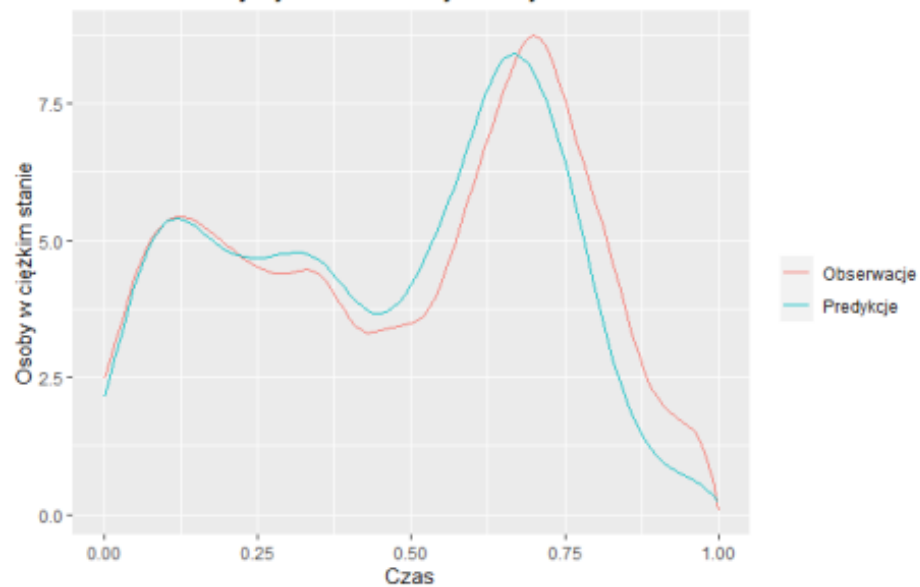
Predykcje vs obserwacje w woj. pomorskim



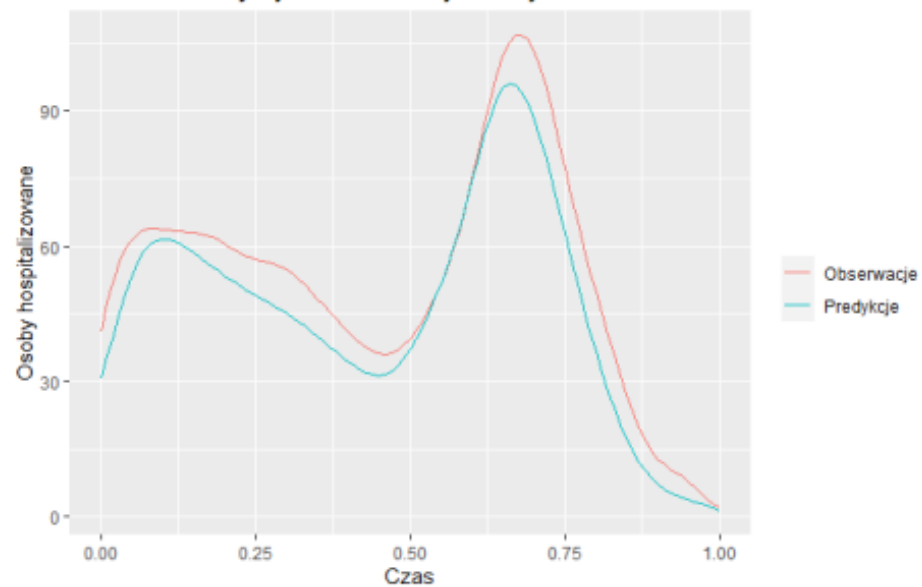
Predykcje vs obserwacje w woj. pomorskim



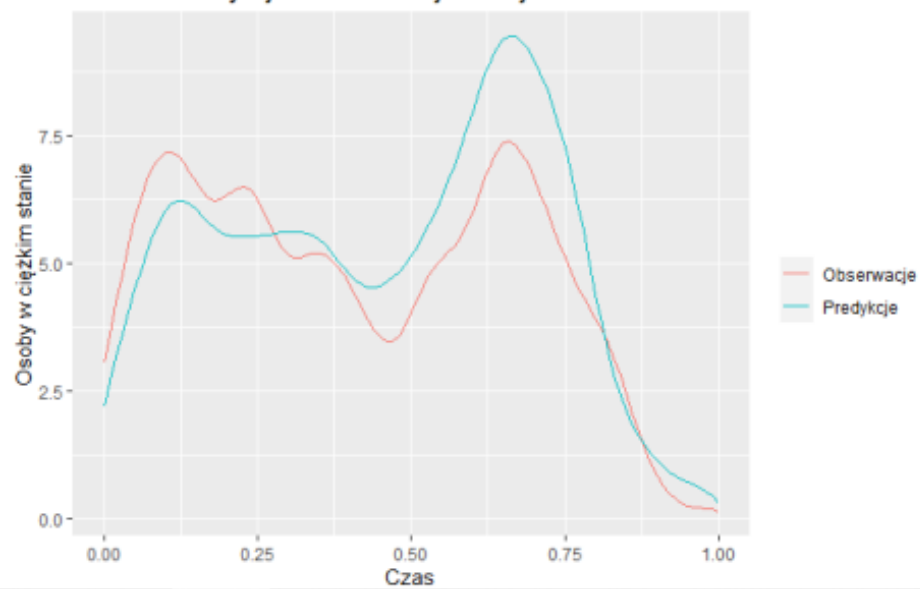
Predykcje vs obserwacje w woj. łódzkim



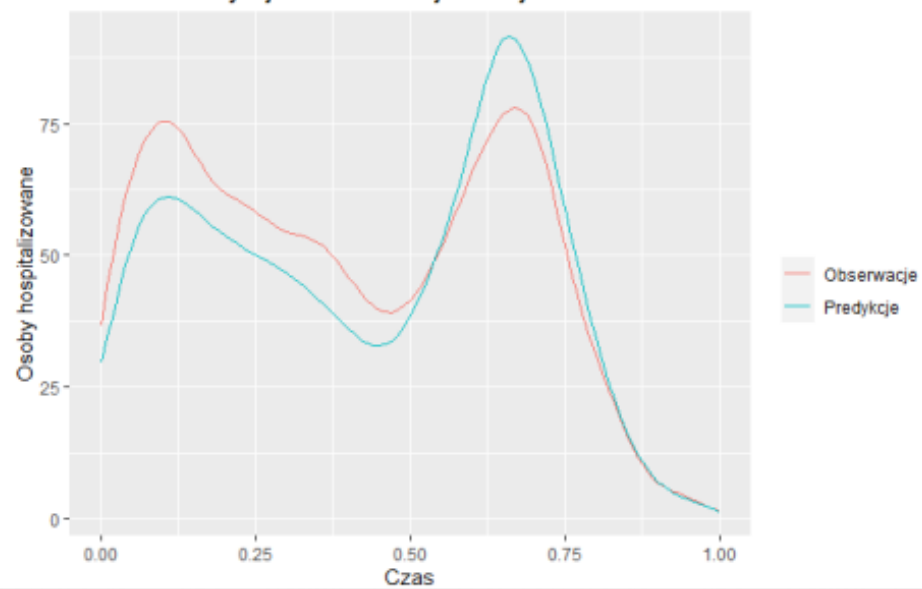
Predykcje vs obserwacje w woj. łódzkim



Predykcje vs obserwacje w woj. lubelskim

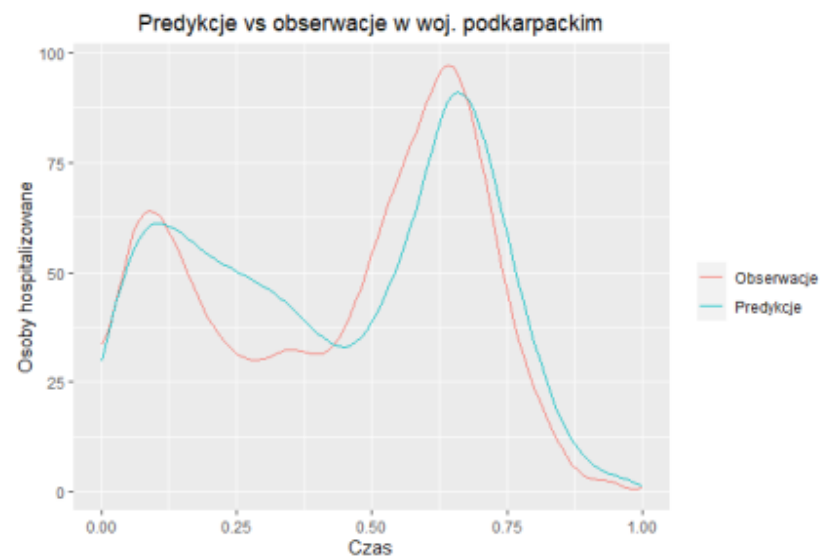
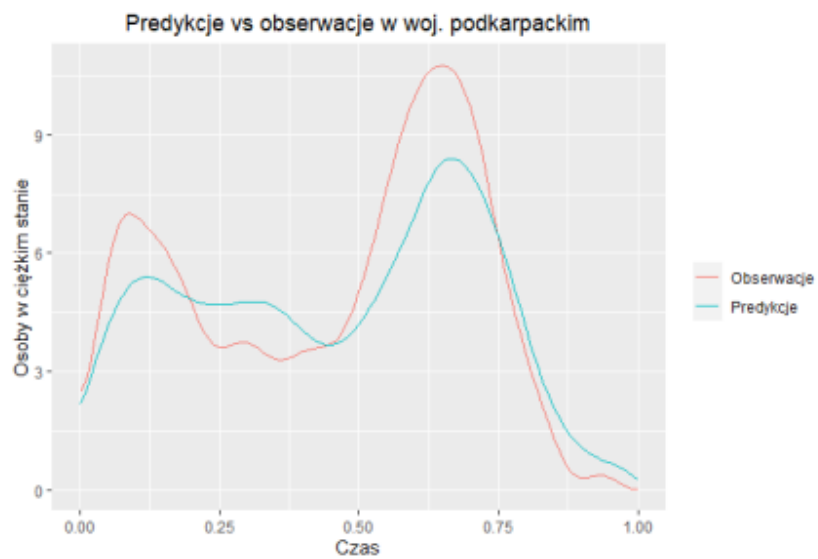
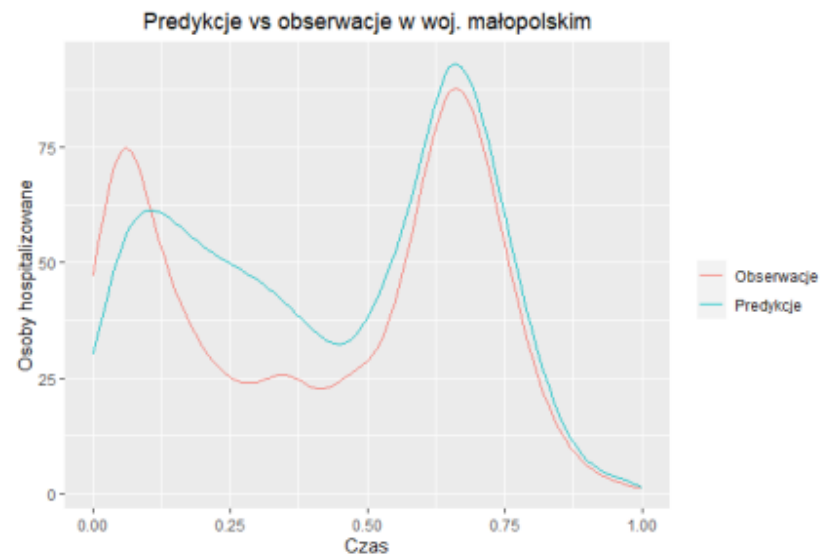
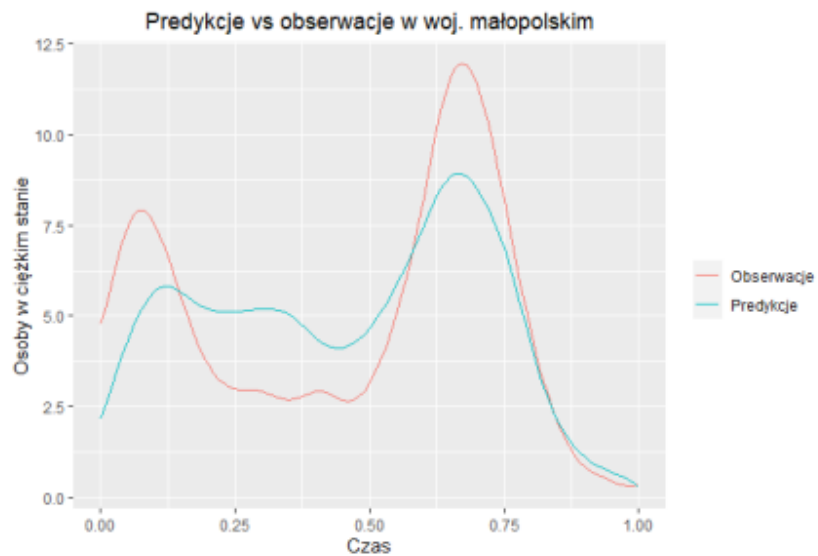


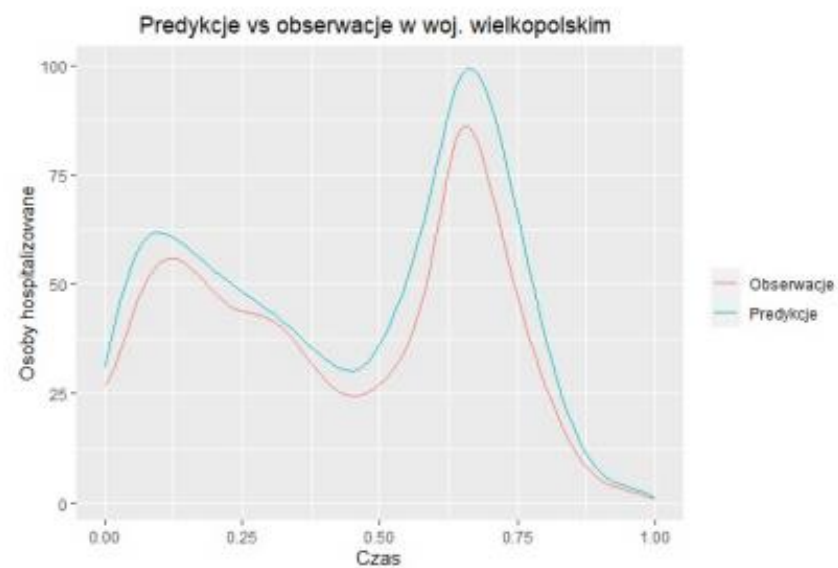
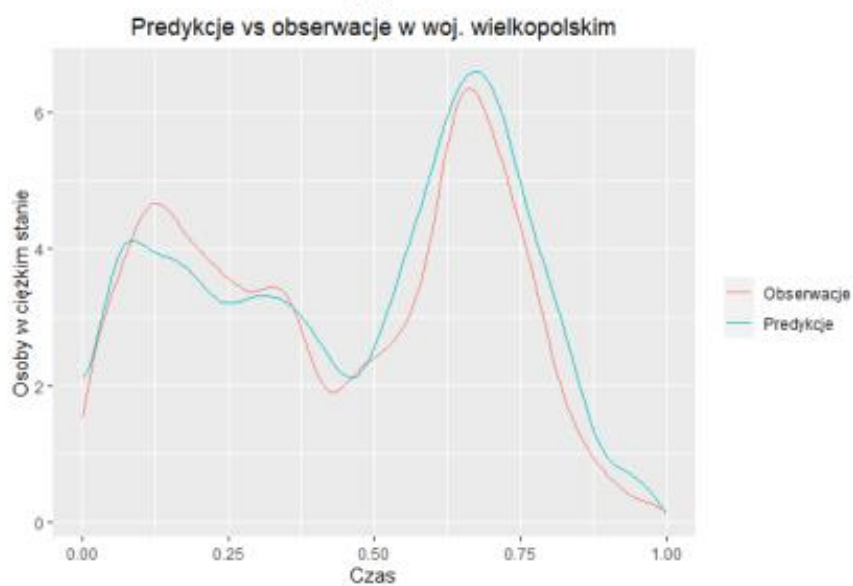
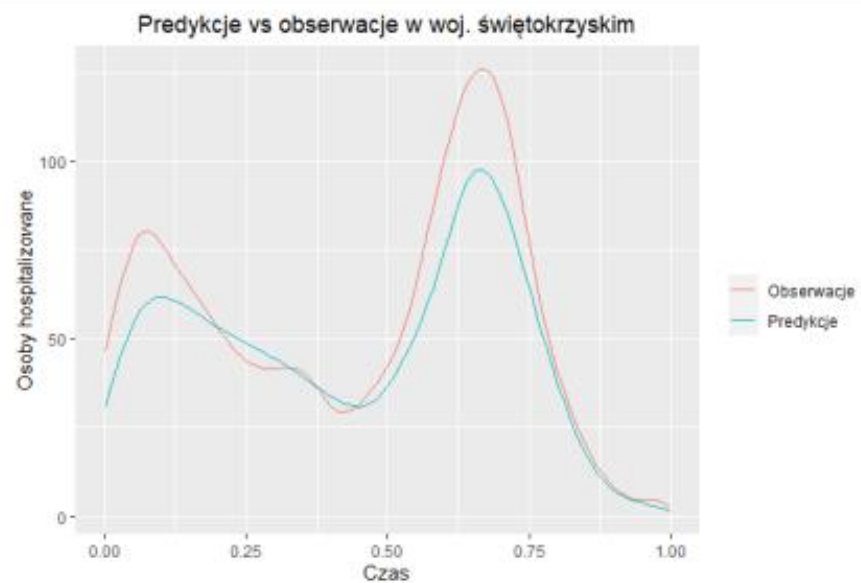
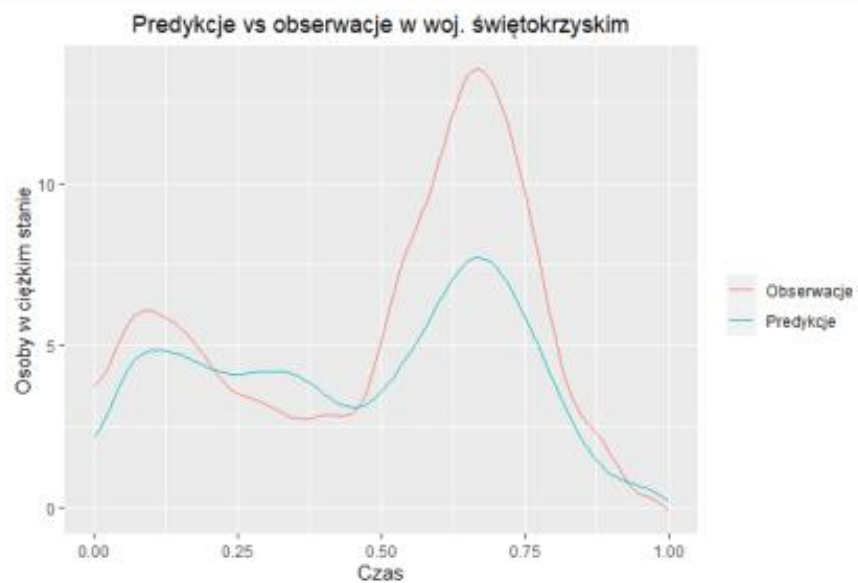
Predykcje vs obserwacje w woj. lubelskim



województwo	MSE( $y_{i1}$ )	MSE( $y_{i2}$ )
dolnośląskie	<b>4.087584</b>	0.8424024
kujawsko-pomorskie	5.801758	<b>0.5078765</b>
łódzkie	8.146502	0.7999353
lubelskie	8.434905	<b>1.1796471</b>
lubuskie	<b>10.79992</b>	0.9807714
mazowieckie	4.569325	0.6134904
opolskie	9.170555	0.7959462
podlaskie	9.612532	1.0092404
pomorskie	7.526031	0.9541513
śląskie	7.451072	1.1456427
warmińsko-mazurskie	3.906237	1.0330436
zachodnio-pomorskie	6.692964	0.9105502

Tabela 1: Wartości błędu średniokwadratowego dla  $y_{i1}$  (liczby osób hospitalizowanych) i  $y_{i2}$  (liczby osób w stanie ciężkim) dla województw z próby treningowej.





Rysunek 12: Obserwowane i przewidywane krzywe dla próby testowej.

Osoby w stanie ciężkim				
	małopolskie	podkarpackie	świętokrzyskie	wielkopolskie
czas	obs/pred	obs/pred	obs/pred	obs/pred
23.10	163/74	50/46	46/26	59/74
24.10	160/80	54/49	46/27	56/75
25.10	172/85	56/52	46/29	75/77
26.10	188/91	68/55	51/30	74/80
27.10	207/97	67/58	52/32	77/83
28.10	215/103	72/62	50/34	68/87
29.10	209/109	70/65	50/35	80/92
30.10	208/115	81/69	53/37	94/97
31.10	212/121	90/72	55/39	94/102
1.11	217/126	95/76	55/41	102/107
...	...	...	...	...
26.06	11/21	5/13	4/7	11/18
27.06	11/20	4/12	3/7	10/17
28.06	14/19	3/11	2/6	9/16
29.06	12/18	2/11	2/6	9/14
30.06	11/17	2/10	2/5	9/13
1.07	11/16	0/9	1/5	7/11
2.07	11/14	0/8	0/4	7/9
3.07	11/13	0/7	0/4	7/8
4.07	11/11	1/6	0/3	7/6
5.07	10/9	0/5	0/2	5/4

## Osoby hospitalizowane

23.10	1671/1023	681/629	559/377	924/1088
24.10	1667/1086	750/668	581/401	929/1158
25.10	1763/1151	757/707	589/426	961/1229
26.10	1907/1216	796/747	704/450	1039/1300
27.10	1972/1281	819/786	742/474	1090/1370
28.10	2003/1345	813/826	750/498	1137/1439
29.10	2072/1408	848/865	792/522	1182/1507
30.10	2250/1471	883/903	789/545	1241/1574
31.10	2289/1532	917/941	779/567	1252/1638
1.11	2366/1591	976/977	829/589	1292/1700
...	...	...	...	...
26.06	72/106	31/66	69/39	81/111
27.06	72/100	26/62	61/37	79/105
28.06	66/95	25/58	57/34	82/99
29.06	63/89	20/55	54/32	82/93
30.06	57/82	15/51	51/30	63/86

# Literatura

1. Autor zbioru danych: Michał Rogalski, COVID-19 w Polsce, Źródło: Dane zebrane na podstawie raportów podawanych przez Ministerstwo Zdrowia, danych z WSSE, PSSE, Urzędów Wojewódzkich, oraz tych uzyskanych w prośbach o dostęp do informacji publicznej. Kontakt: [contact.micalrg@gmail.com](mailto:contact.micalrg@gmail.com); [bit.ly/covid19-poland](https://bit.ly/covid19-poland); dostęp: 01.03.2022.
2. Główny Urząd Statystyczny, Powierzchnia i ludność w przekroju terytorialnym w 2021 roku, Powierzchnia, liczba ludności i gęstość zaludnienia wg stanu na 1 stycznia 2021 roku. Przekroje: województwa, powiaty, gminy, miasta. Data publikacji: 22.07.2021. Dostęp: 07.03.2022. Link: <https://stat.gov.pl/obszary-tematyczne/ludnosc/ludnosc/powierzchniai-ludnosc-w-przekroju-terytorialnym-w-2021-roku,7,18.html>.
3. Christian Acal, Manuel Escabias, Ana M. Aguilera, Mariano J. Valderrama, COVID-19 Data Imputation by Multiple Function-on-Function Principal Component Regression, *Mathematics* 2021, 9, 1237. <https://doi.org/10.3390/math9111237>, Academic Editor: JinTing Zhang, 28 May 2021.
4. Daniel Levitin, Bradley W. Vines, Regina L. Nuzzo, James O. Ramsay, Introduction to Functional Data Analysis, *Canadian Psychology*, Copyright 2007 by the Canadian Psychological Association, August 2007, Vol. 48, No. 3, 135-155, DOI: 10.1037/cp2007014.



DZIĘKUJĘ ZA UWAGĘ!