

Zastosowanie epidemiologicznego modelu SIR do pandemii w Polsce

Autor:

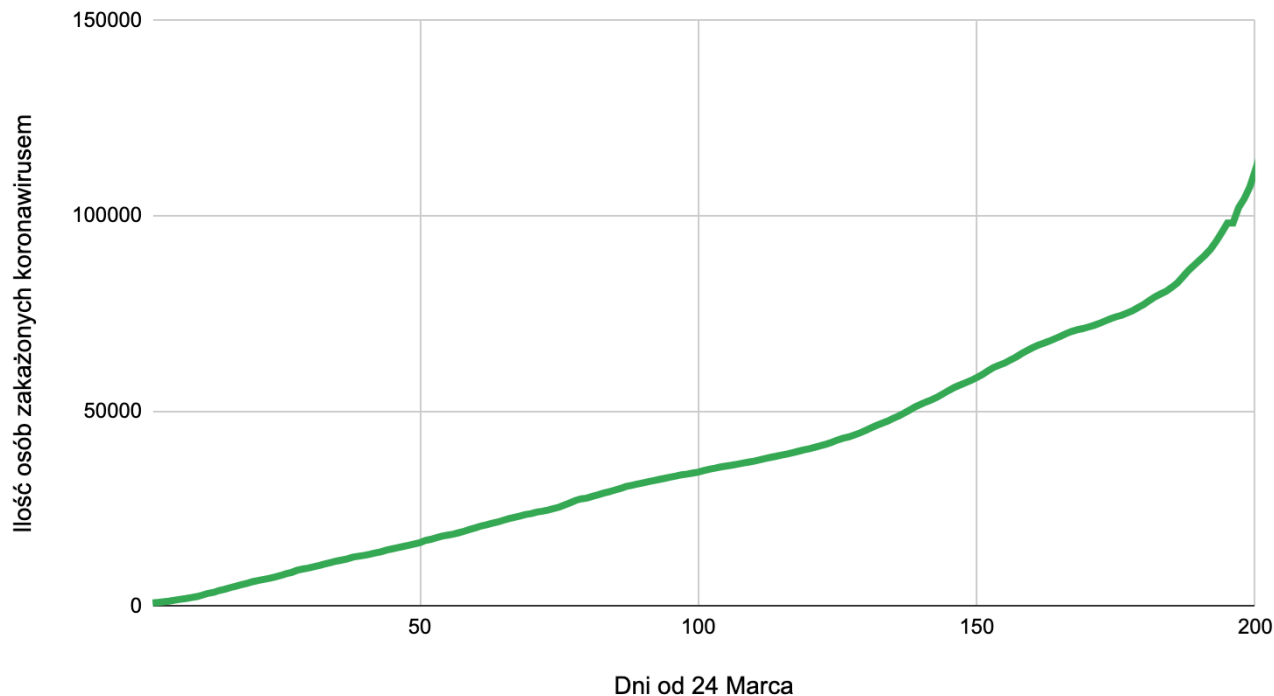
Justyna Trzeba

Klasa III liceum

IV Liceum Ogólnokształcące im. Emilii Szanieckiej w Łodzi

Obserwacja zjawiska: Przyrost osób zainfekowanych COVID-em w Polsce

Zmiany w liczba odnotowanych przypadków zakażenia koronawirusem w Polsce od 24 Marca przez następne 200 dni



Problem badawczy:

1. W jakim stopniu epidemiologiczny model SIR poprawnie przewiduje przebieg pandemii koronawirusa w Polsce?
2. Jakimi metodami można ulepszyć ten model tak aby zwiększyć jego efektywność?

Hipoteza:

SIR model jest najprostszym modelem epidemiologicznym , dlatego jego przewidywania są niepoprawne.

Podstawowe informacje o Modelu SIR

Zmiennymi zależnymi w modelu SIR są trzy różne grupy , na które została podzielona populacja Polski :

1. $S(t)$:(ang. susceptible) Ilość osób podatnych na zakażenie wirusem.
2. $I(t)$: (ang. infected) Ilość osób zainfekowanych, którzy chorują i roznoszą infekcję
3. $R(t)$: (ang. removed) Ilość osób ozdrowiałych. Do tej grupy zaliczamy osoby, które wyzdrowiały i nabyły odporność po przejściu infekcji, osoby zaszczepione, jak również osoby zmarłe.

Zmienną niezależną jest czas oznaczony jako t , mierzony w dniach.

**Wszystkie funkcje dzieli przez N, aby uprościć obliczenia.
Otrzymane ułamki wskazują na część całej populacji jaką stanowi dana grupa.**

$$\rightarrow s(t) = \frac{S(t)}{N}$$

$$\rightarrow i(t) = \frac{I(t)}{N}$$

$$\rightarrow r(t) = \frac{R(t)}{N}$$

Model SIR przyjmuje założenia, które są w większości nierealistyczne:

1. Liczebność populacji jest stała: **$S(t) + I(t) + R(t) = N = \text{const}$**
2. Populacja jest wymieszana. Oznacza to, że każda osoba, należąca do jednej z trzech grup na jakie została podzielona populacja, ma jednakowe prawdopodobieństwo spotkania osoby z innej grupy.
3. Populacja jest odizolowana, co znaczy, że: migracja do/z kraju czy przyrost naturalny jest zaniedbany.
4. Zakłada się także, że osoba, która wyzdrowiała, czyli ta która przeszła z grupy $I(t)$ do $R(t)$, ma trwałą odporność przeciwko kolejnej infekcji.

W modelu SIR rozważa się również dwa współczynniki, które są stałe przez cały badany okres

1. Współczynnik zaraźliwości, czyli tempa zarażania. Kontroluje on liczbę rzeczywistych zakażeń wynikających z każdego spotkania osób zainfekowanych i zdrowych.

$$\beta = \frac{\text{Liczba osób spotykanych przez jedną osobę w jeden dzień} \times \text{Prawdopodobieństwo zarażenia}}{\text{Cała populacja Polski}}$$

2. Współczynnik zgonów/zdrowienia, gdzie T to liczba dni, podczas których osoba zainfekowana może potencjalnie zarażać inne osoby.

$$\gamma = \frac{1}{T}$$

Schemat rozwoju infekcji według modelu SIR wraz z omówionymi współczynnikami :



Równania różniczkowe pierwszego rzędu opisujące dynamikę zmian liczby osób zdrowych, zainfekowanych i wyzdrowiałych/zmarłych wraz z wyjaśnieniami

$$\frac{dS}{dt} = -S(t) \times I(t) \times \beta \quad (1)$$

$$\frac{dI}{dt} = \beta \times I(t) \times S(t) - \gamma \times I(t) \quad (2)$$

$$\frac{dR}{dt} = \gamma \times I(t) \quad (3)$$

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \quad (4)$$

Równanie 4 opisuje założenie, że liczebność populacji (N) jest stała i nie zmienia się w czasie.

Równanie 1 opisuje zależność między tempem zmian w grupie S a bieżącą sytuacją w populacji. Ujemny iloczyn oznacza, że liczba osób zdrowych zmniejsza się na rzecz przyrostu osób zakażonych.

Równania 1 i 2 przedstawiają relacje między liczbą osób zarażonych wirusem a ilością osób podatnych na infekcję. Każdy spadek w liczbie osób zdrowych odpowiada wzrostowi w liczbie osób zakażonych.

Równanie 3 przedstawia przyrost osób wyzdrowiałych bądź zmarłych, czyli tych którzy opuścili grupę I , współczynnik zgonów/zdrowienia jest przez cały czas dodatni, dlatego wielkość tej grupy rośnie.

Metoda Eulera (Metoda ta pozwala na znalezienie przybliżonego rozwiązania)

Metoda numeryczna rozwiązywania równań różniczkowych pierwszego rzędu postaci:

$$\frac{dy}{dx} = f(x, y), \quad y(x_0) = y_0$$

Założmy, że znamy wartości x_n i y_n . Aby wyznaczyć współrzędne kolejnego punktu korzystamy z poniższych formuł rekurencyjnych, które pomagają uzyskać przybliżone oszacowanie nowej wartości x_{n+1} i y_{n+1} dla dostatecznie małego kroku h :

$$x_{n+1} = x_n + h, \quad h = \Delta x$$

$$y_{n+1} = y_n + h \times f(x_n, y_n), \quad h = \Delta x$$

Zastosowanie metody Eulera do rozwiązania równań modelu SIR:

$$\frac{dS}{dt} = -\beta SI \Rightarrow S_{t+1} = S_t - \beta \times S_t \times I_t \times \Delta t$$

$$\frac{dI}{dt} = \beta SI - \gamma I \Rightarrow I_{t+1} = I_t + (\beta \times S_t \times I_t - \gamma \times I_t) \times \Delta t$$

$$\frac{dR}{dt} = \gamma I \Rightarrow R_{t+1} = R_t + (\gamma \times I_t) \times \Delta t$$

Użycie danych epidemiologicznych z Polski

24 marca zanotowano 901 osób zakażonych koronawirusem, 10 osób wyzdrowiały i 1 śmierć z powodu zakażenia. Dlatego:

$$N = 38268000 \quad S_0 = N - (I + R) = 38268000 - (901 + 11) = 38267088$$

$$I_0 = 901$$

S to podatni na zakażenia, I zainfekowani, a R to zmarli bądź wyzdrowiali.

$$R_0 = 11$$

Według informacji udzielonej przez WHO, okres dni podczas których osoby mogą potencjalnie zarazić inne osoby to 14 dni, dlatego współczynnik zdrowienia/śmierci wynosi:

$$\gamma = \frac{1}{T} = \frac{1}{14} \approx 0.0714$$

Według przewidywań publikacji (Orzechowska & Bednarek, 2020) współczynnik zaraźliwości wynosi:

$$\beta = 0.3$$

Po podzieleniu przez N (ludność Polski):

$$s_0 = 0.999$$

$$i_0 = 2.35 \times 10^{-5}$$

$$r_0 = 2.87 \times 10^{-7}$$

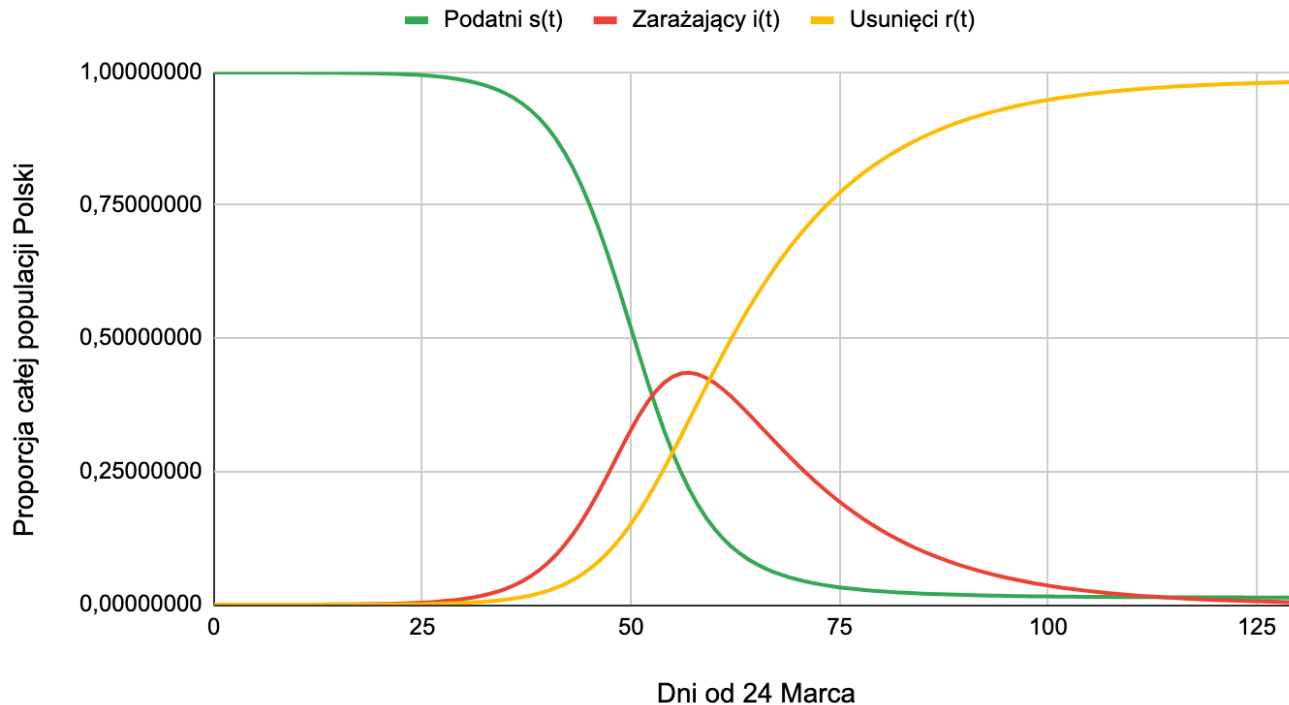
Algorytm obliczeń wprowadzony do google sheets, aby utworzyć wykres modelu SIR:

Dzień	Zdrowi (S)	Zainfekowani (I)
t	s_t	i_t
0	$s_0 = 0.999$	$i_0 = 2.35 \times 10^{-5}$
1	$s_1 = s_0 - (0.3 \times s_0 i_0) \times 1$	$i_1 = i_0 + (0.3 \times s_0 i_0 - 0.0714 \times i_0) \times 1$
2	$s_2 = s_1 - (0.3 \times s_1 i_1) \times 1$	$i_2 = i_1 + (0.3 \times s_1 i_1 - 0.0714 \times i_1) \times 1$
...
n	$s_n = s_{n-1} - (0.4 \times s_{n-1} i_{n-1}) \times 1$	$i_n = i_{n-1} + (0.4 \times s_{n-1} i_{n-1} - 0.0714 \times i_{n-1}) \times 1$

Dzień	Usunięci (R)
t	r_t
0	$r_0 = 2.87 \times 10^{-7}$
1	$r_1 = r_0 + (0.0714 \times i_0) \times 1$
2	$r_2 = r_1 + (0.0714 \times i_1) \times 1$
...	...
n	$r_n = r_{n-1} + (0.0714 \times i_{n-1}) \times 1$

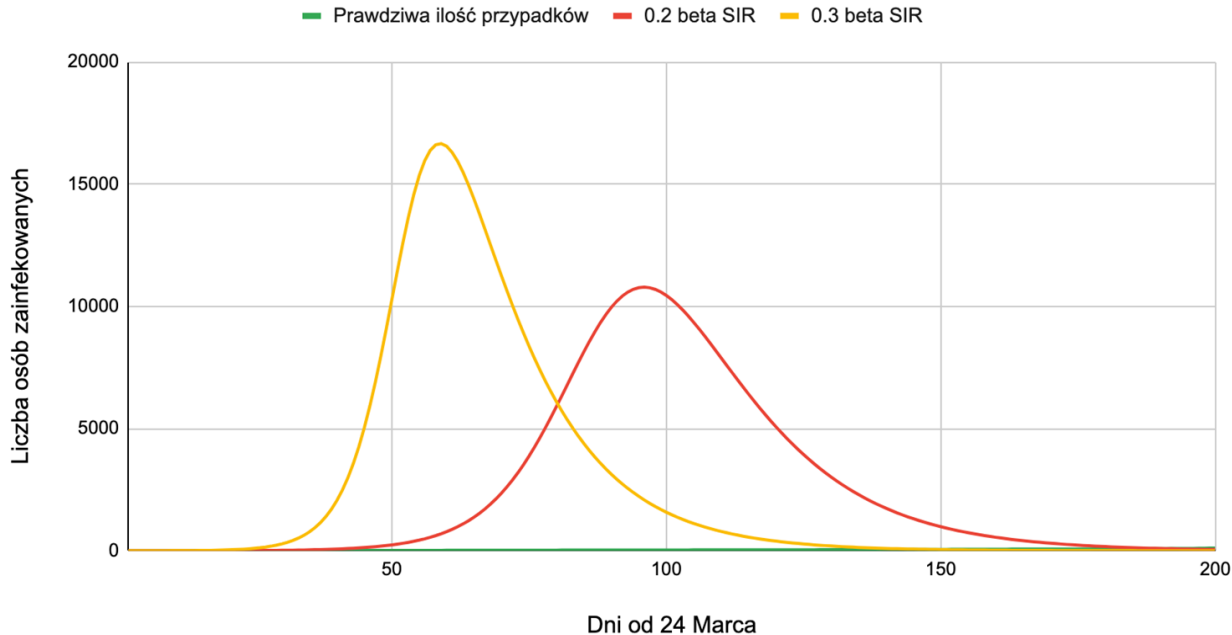
Przeprowadzenie eksperymentu: wykres modelu SIR ukazujący zmiany w czasie w poszczególnych grupach populacji Polski (osób zdrowych, zainfekowanych i wyzdrowiałych/zmarłych)

Model SIR zaaplikowany do danych z pandemii koronawirusa w Polsce



Weryfikacja hipotezy: Porównanie przewidywań modelu SIR do prawdziwych danych z Polski

Porównanie przewidywanej liczby zakażeń koronawirusem w Polsce przez model SIR do rzeczywistych danych



Zmiany w liczbie osób zarażonych koronawirusem w Polsce proponowany przez model SIR z dwoma wartościami beta nie pokrywa się z wykresem prawdziwych przypadków. To sugeruje, że przewidywania modelu SIR są nieprawidłowe.

Propozycja metody ulepszającej przewidywania modelu SIR

Proponuję posłużenie się metodą iteracji wyszukiwania dziesiętnego, w którym systematycznie zwiększam liczbę miejsc po przecinku (do 4 miejsc) w celu znalezienia wartości parametru beta, tak aby jak najdokładniej dopasowywał model SIR do danych z Polski.

Metoda wyszukiwania dziesiętnego

Zaczynam od wartości z jednym miejscem po przecinku i zwiększam wartości o odstęp 0.1: 0.1, 0.2, 0.3, 0.3... Z wcześniejszych wykresów obserwuję, że wartości 0.2 i 0.3 generują niepoprawny model, dlatego wybieram wartość beta równą 0.1. Ze względu na duży błąd procentowy i bezwzględny przy tej wartości (ukazanej w tabeli w następnym slajdzie), wnioskuję, że współczynnik ten powinien być równy wartości pomiędzy 0 a 0.1.

Kontynuuję zwiększanie miejsc po przecinku: zwiększam wartości o przedział 0.01: 0.01, 0.02, 0.03... Błąd procentowy i bezwzględny wskazuje, że wartość beta 0.09 jest najbardziej poprawna, dlatego zachowuję wartość 0.09 i kontynuuję zwiększanie miejsc po przecinku używając przedziału 0.001: 0.091, 0.092, 0.093...

Wartość, która okazuje się najbardziej dokładna jest 0.096. Stosując iteracje powyższej procedury poszukuje wartości beta do czterech miejsc po przecinku stosując odstęp 0.0001: 0.0961, 0.0962, 0.0963...

Błąd procentowy w dwustym dniu badania okazuje się być najmniejszy dla wartości 0.0966, dlatego stwierdzam tą wartość za najbardziej dokładną do 4 miejsc po przecinku, a wykres obrazujący jest przedstawiony w kolejnych slajdach.

Zastosowanie iteracji wyszukiwania dziesiętnego aby dopasować wartość współczynnika beta

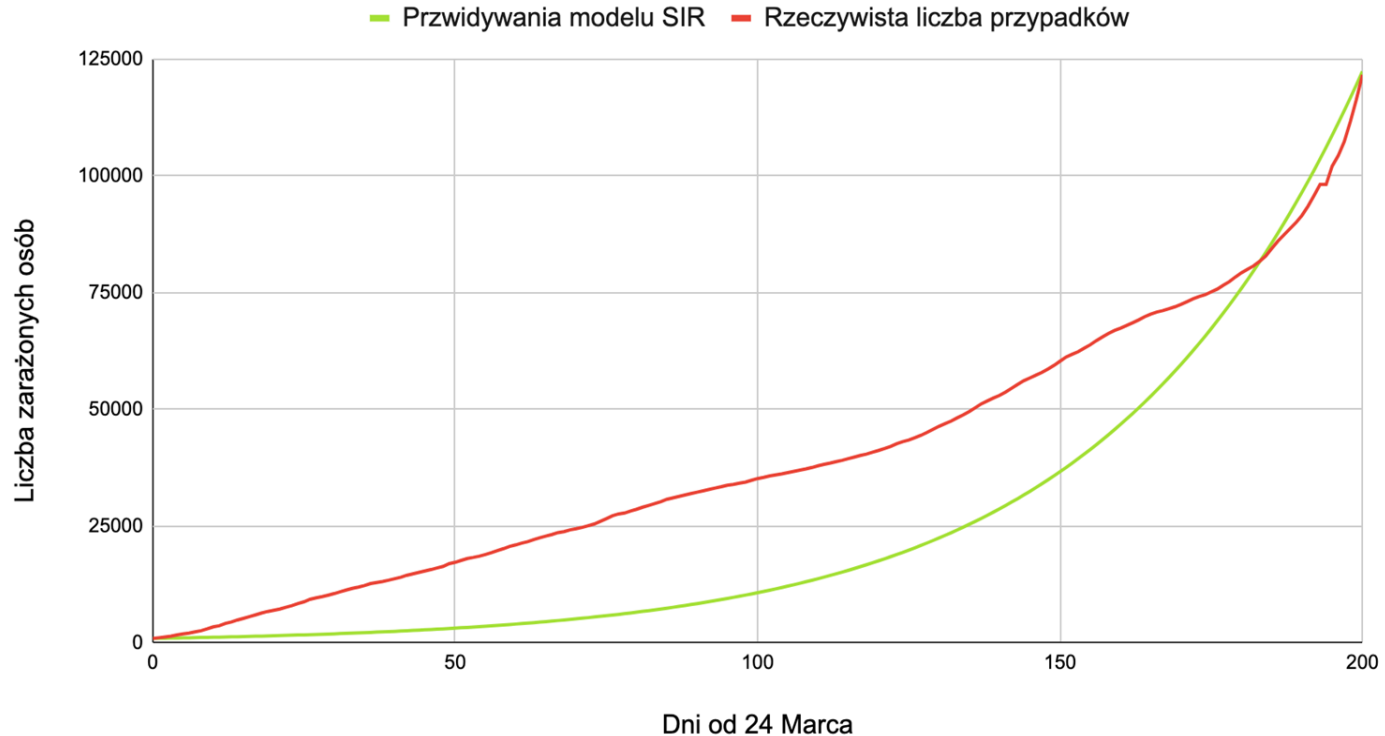
Wartość parametru β	Przewidywana wartość $i(200)$ używając tej wartości β	Rzeczywista wartość w dniu 200	Błąd procentowy ¹ (%)	Maksymalny błąd bezwzględny ² w okresie 200 dni
0.1	218 766	121638	79.9	207138
0.08	4879	121638	96.0	854333
0.09	34582	121638	71.6	787097
0.095	90372	121638	25.7	631120
0.096	109248	121638	10.2	574995
0.0965	120070	121638	1.3	542448
0.0966	122355	121638	0.6	535548
0.0967	124683	121638	2.5	528513

¹ Błąd procentowy = $| \text{Rzeczywista liczba przypadków} - \text{Przewidywania modelu SIR} | \div \text{Rzeczywista liczba przypadków} \times 100\%$

² Maksymalny błąd bezwzględny = $| \text{Przewidywania modelu SIR} - \text{Rzeczywista liczba przypadków} |$

Wykres ulepszzonego modelu SIR

Porównanie przewidywań modelu SIR do rzeczywistych danych z Polski



Ocena dotychczasowej metody

Zastosowana przeze mnie metoda generuje model SIR z poprawnymi przewidywaniami tylko dla ostatnich 10 dni okresu 200 dni. Podczas pozostałych dni okresu badania, liczba przewidywanych przypadków zakażeń przez model znacznie odbiega od prawdziwych danych. Dlatego, kontynuuje ulepszanie modelu.

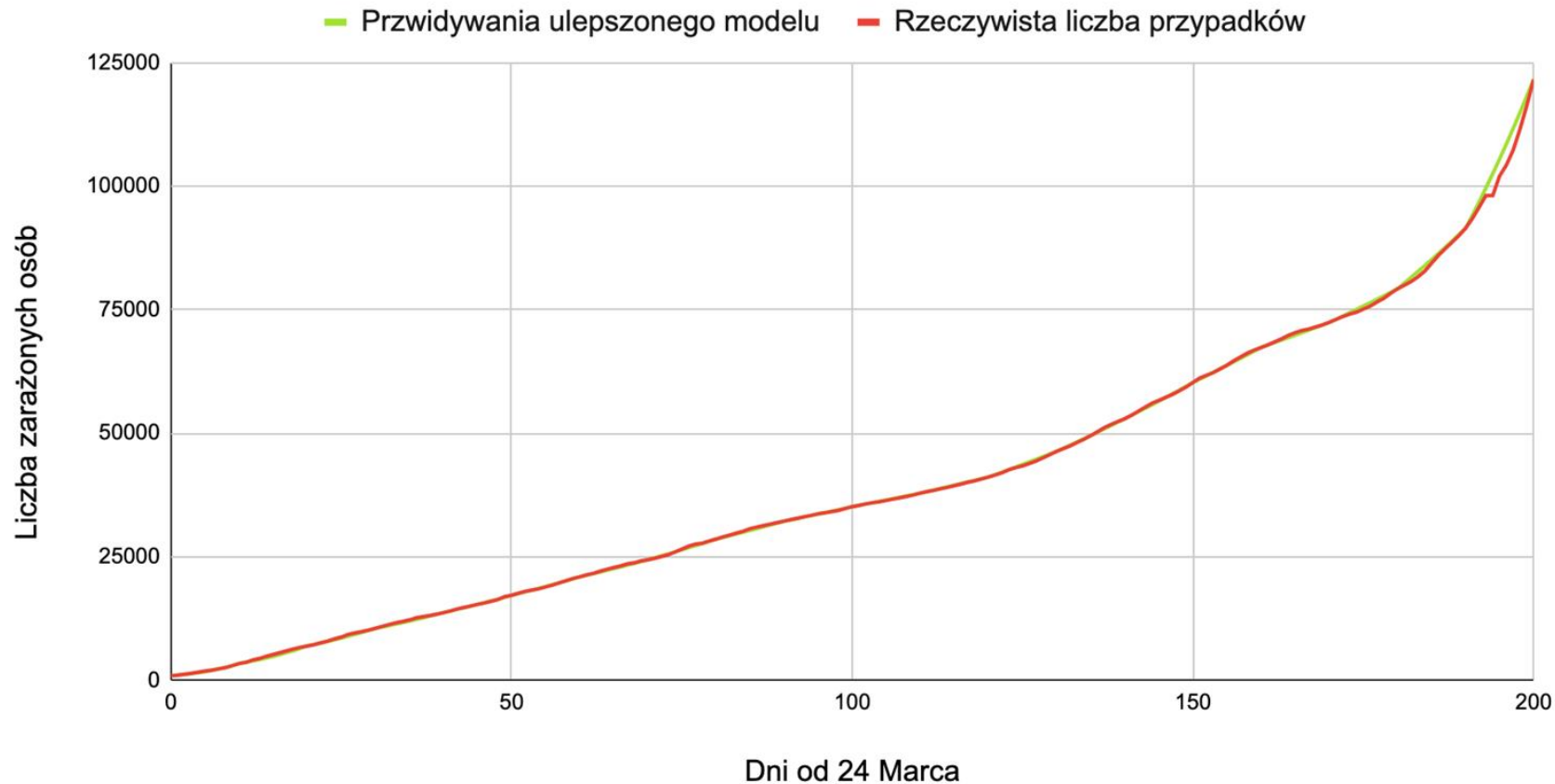
Wprowadzenie współczynnika zaraźliwości zmiennego w czasie jako metoda ulepszenia przewidywań modelu

Okres 200 dni od 24 marca dzielę na 20 równych przedziałów o długości 10 dni. Następnie dla każdego okresu czasu stosuję tę samą metodę systematycznego wyszukiwania dziesiętnego. Prowadzi to do otrzymania 20 różnych wartości parametru beta, które pozwalają mi na utworzenie modelu SIR z niewielkim błędem procentowym. (Świadczy to o dokładności ulepszonego modelu SIR, a zarazem o osiągnięciu większej efektywności po wprowadzonej metody wyszukiwania dziesiętnego).

Wykres ostatecznie ulepszonego modelu SIR porównany z prawdziwymi danymi ukazany jest w następnym slajdzie.

Współczynnik zaraźliwości zmienny w czasie $\pi(t)$	Przedział czasu t (Dni od 24 Marca)	Rzeczywista liczba przypadków w ostatnim dniu przedziału czasu	Przewidywanie ulepszonego modelu SIR w ostatnim dniu przedziału czasu	Błąd procentowy (%)	Maksymalny błąd bezwzględny ³ w przedziale czasu
0.2135	1-10	3383	3389	1.8×10^{-3}	116
0.1458	11-20	6934	6935	1.4×10^{-4}	367
0.1140	21-30	10511	10510	9.5×10^{-5}	387
0.0984	31-40	13693	13697	2.9×10^{-4}	320
0,0947	41-50	17204	17212	4.7×10^{-4}	118
0,0914	51-60	20931	20936	2.4×10^{-4}	116
0,0871	61-70	24395	24405	4.1×10^{-4}	262
0,0876	71-80	28577	28575	7.0×10^{-5}	330
0,0838	81-90	32227	32215	3.7×10^{-4}	358
0,0805	91-100	35146	35136	2.8×10^{-4}	140
0,0794	101-110	37891	37887	7.9×10^{-5}	106
0,0802	111-120	41162	41152	2.4×10^{-4}	110
0,0840	121-130	46346	46366	4.3×10^{-4}	324
0,0855	131-140	52961	52971	1.9×10^{-4}	266
0,0852	141-150	60281	60280	1.7×10^{-5}	302
0,0835	151-160	67372	67393	3.1×10^{-4}	327
0,0796	161-170	72453	72439	1.9×10^{-4}	507
0,0815	171-180	79240	79230	1.3×10^{-4}	718
0,0872	181-190	91514	91475	4.3×10^{-4}	1125
0,1021	191-200	121638	121653	1.2×10^{-4}	4418

Porównanie modelu SIR ze zmiennym w czasie współczynnikiem zaraźliwości (beta) do rzeczywistych danych z Polski



Wnioski i zakończenie

1. Początkowo stworzony podstawowy model SIR przewiduje rozprzestrzenianie się koronawirusa w sposób niepoprawny (co demonstruje wykres porównujący model i prawdziwe dane – slajd 14) i to potwierdza moją hipotezę.
2. Propozycja ulepszająca model, czyli zastosowanie zmiennego w czasie współczynnika zaraźliwości okazuje się efektywną metodą (biorąc pod uwagę mały błąd procentowy ulepszanego modelu). Jednakże to podejście także ma wyraźne wady:
 - 10-dniowe przedziały czasu nie wpisują się w realne zmiany obostrzeń w Polsce, lecz są one systematycznie zmieniane, aby zminimalizować błąd modelu. Dodatkowo: czym mniejsze przedziały czasu na przykład 5-dniowe okresy, a nawet mniejsze, wygenerowałyby model z mniejszymi błędami bezwzględnymi i procentowymi (jest to szczególnie zaobserwowane w ostatnich 40 dniach okresu badanego).
3. Mimo wprowadzonych ulepszeń zwiększającej efektywność modelu SIR należy pamiętać, że model ten zawiera mnóstwo założeń, które nie są realistyczne, dlatego proponowana hipoteza, mówiąca o niepoprawności przewidywań modelu SIR, nawet w świetle wprowadzonych ulepszeń, jest słuszna.

Bibliografia

Brauer F., Castillo-Chavez C. (2012) Epidemic Models. In: Mathematical Models in Population Biology and Epidemiology. Texts in Applied Mathematics, vol 40. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-1686-9_9

Callahan, J., Hoffman, K., Cox, D., O'Shea, D., Pollatsek, H., & Senechal, L. (1995). *Calculus in context: The Five College Calculus Project* (2008 ed.). Retrieved February 18, 2021, from <https://math.smith.edu/~callahan/cic/book.pdf>

CDC. (2021, February 22). Symptoms of Coronavirus. Retrieved February 28, 2021, from <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>

Froese, H. (2020, May 03). Infectious Disease Modelling: Understanding the models that are used to model Coronavirus. Retrieved March 20, 2021, from <https://towardsdatascience.com/infectious-disease-modelling-part-i-understanding-sir-28d60e29fdcf>

Gus. (2020, October 15). Statistics Poland. Retrieved March 2, 2021, from <https://stat.gov.pl/en/topics/population/population/population-size-and-structure-and-vital-statistics-in-poland-by-territorial-division-as-of-june-30-2020,3,28.html>

Hodder Education. (n.d.). Numerical methods. Retrieved March 3, 2021, from <https://www.hoddereducation.co.uk/media/Documents/Maths/OCR-Add-maths-Numerical-Methods-sample-material.pdf>

Liao, Z., Peng L., Z., Yan Z., Shengzong, L., (2020) TW-SIR: time-window based SIR for COVID-19 forecasts. *Sci Rep* 10, 22454. <https://doi.org/10.1038/s41598-020-80007-8>

Mathematics in Education and Industry (MEI). (2005, September). Methods for Advanced Mathematics (C3) Coursework. Retrieved February 20, 2021, from <http://vle.woodhouse.ac.uk/topicdocs/maths/AutographC3files/MEI Student Coursework Guide.pdf>

Mathews, D. V. (n.d.). Numerical Methods – A guide for teachers (Years 11–12). Retrieved February 25, 2021, from https://amsi.org.au/ESA_Senior_Years/PDF/PDFvcaa/numerical3j.pdf

Mathews, D. V. (n.d.). Numerical Methods – A guide for teachers (Years 11–12). Retrieved February 25, 2021, from https://amsi.org.au/ESA_Senior_Years/PDF/PDFvcaa/numerical3j.pdf

Pearson, (2019) Mathematics Analysis and Approaches for the IB diploma, 863-869

Rohani, P., Choisy, M., & Guégan, J. F. (n.d.). Mathematical Modeling of Infectious Diseases Dynamics. Retrieved March 13, 2021, from https://www.mivegec.ird.fr/images/stories/PDF_files/0507.pdf.

Sas, A. (2021, March 3). Poland: Coronavirus (COVID-19) new cases 2021. Retrieved February 31, 2021, from <https://www.statista.com/statistics/1105153/poland-coronavirus-covid-19-new-cases/>

Simon, C. M. (2020). The SIR dynamic model of infectious disease transmission and its analogy with chemical kinetics. doi: <https://doi.org/10.7717/peerj-pchem.14>

Stroyan, K. (n.d.). *Chapter 2 Using Calculus to Model Epidemics*. Retrieved February 25, 2021, from <https://homepage.divms.uiowa.edu/~stroyan/CTLC3rdEd/3rdCTLCtext/Chapters/Ch2.pdf>

Sullivan, B. (2020, April 07). SIR Model for Disease Spread - 3. Euler Integration in Spreadsheets. Retrieved January 31, 2021, from <https://www.youtube.com/watch?v=2Hsl1C1ov5c>

Trinity University (2021, January 2). Euler's Method Retrieved April 18, 2021, from <https://math.libretexts.org/@go/page/9404>

University of Colorado Boulder. (n.d.). 1.3 Prediction using SIR. 15-30. Retrieved March 17, 2021, from http://math.colorado.edu/~stade/CLS/Section_1_3.pdf

World Health Organization, (2020) WHO COVID-19 Dashboard. Geneva, Retrieved March 31, 2021, from <https://covid19.who.int/>