

# O modelovaní šírenia CoVid-19 (II)

*“Essentially, all models are wrong, but some are useful.”*  
G. Box, Empirical Model-Building and Response Surfaces,  
Wiley, 1987, page 424.

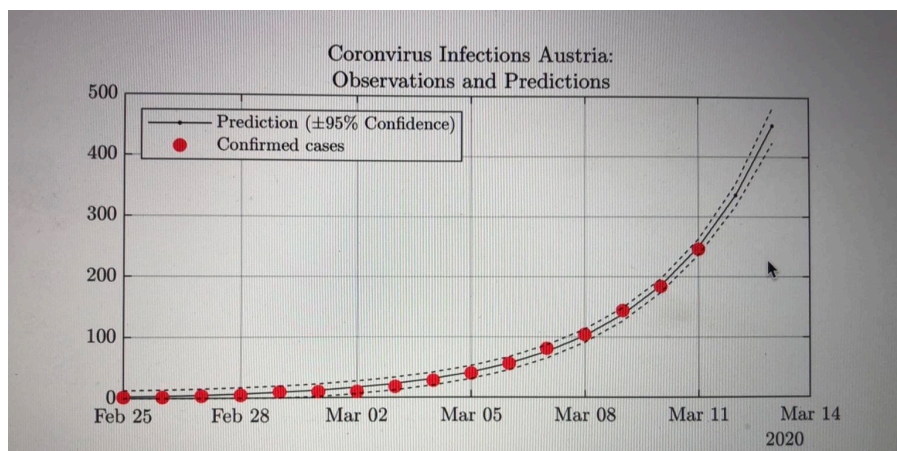
Od prvého pohľadu na problém modelovania šírenia CoVid-19 (uverejneného na tomto webe dňa 30.3.2020, <https://www.learned.sk/o-modelovani-sirenia-covid-19/>) uplynul už viac ako mesiac, preto je možné pozrieť sa na to ešte raz.

V zásade matematické modely by mali primerane dobre odrážať tú meniacu sa realitu, ktorú už poznáme, a poskytovať odhad do budúcnosti. Odhad do budúcnosti je potrebný vo všeobecnosti na rozhodovanie na základe modelu; často takéto rozhodovanie nadobúda podobu automatizovaného riadenia.

Čím lepší je súlad matematického modelu s už dostupnými dátami, tým vyššia je šanca, že tento model môže byť použitý na predpoveď budúceho vývoja procesu alebo systému aspoň v krátkodobom (a ideálne – aj v dlhodobom) časovom horizonte. Číselné výstupy z matematického modelu sa ďalej porovnávajú s pribúdajúcimi dátami, model sa upresňuje a umožňuje lepšie rozhodovanie na základe upresneného modelu. Inými slovami, na to, aby matematický model bol použiteľný, musí sa neustále priebežne adaptovať.

## Príklad rozhodovania na základe modelu

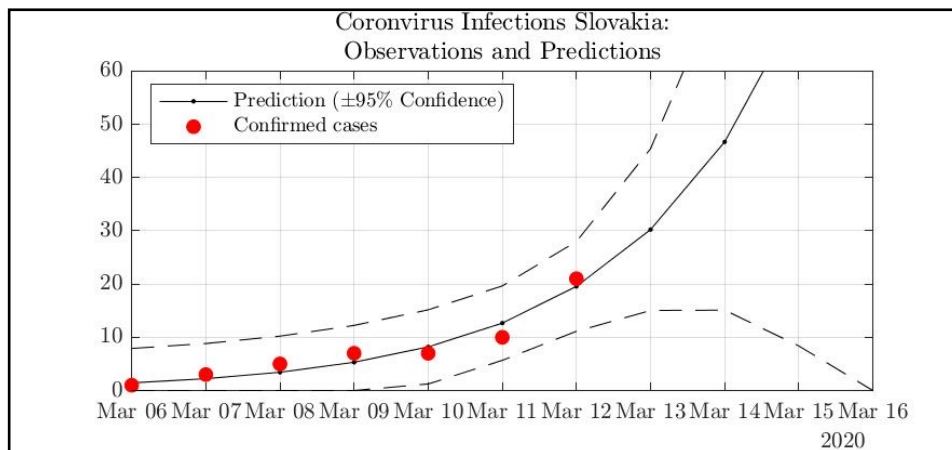
Jeden z našich študentov študuje od začiatku februára v rámci programu ERASMUS na našej sesterskej univerzite v Rakúsku. Dňa 12. marca mi napísal, že sa vracia domov kvôli vývoju situácie, lebo jeho tamojší profesor (ktorý je špičkový odborník na automatizáciu, riadenie procesov a spracovanie signálov) poradil, že bude lepšie, ak sa vráti na Slovensko: *“Zostavil model, ktorý mu už vyše týždňa vychádza. A prognóza je veľmi zlá. Preto mi povedal, že bude lepšie, ak pôjdem.”* Na moju poznámku, že ide o správny vedecký prístup, študent odpovedal: *“Áno, veľmi sa mi jeho prístup páčil. Samozrejme to zahrnul aj do prednášky – Measurement Data Processing”*, a pripojil nasledujúce foto:



Ešte študent dodal, že na základe tohto modelu už od stredy 11. marca tento profesor odmietol vyučovať kontaktné a prešiel na online výučbu, *“pretože nikto si nemôže byť istý, či už náhodou nie je nositeľom”*. Tu treba poznamenať, že Rakúsko začalo zavádzať

obmedzenia až od pondelka 16. marca. Takže kolega v záujme zachovania zdravia študentov a ďalších osôb urobil správne rozhodnutie o prechode na online výučbu bez čakania na rozhodnutie úradov.

Rakúsky kolega poskytol študentom tzv. livescript (program) v MATLABe, z ktorého bolo jasné, že použil exponenciálnu funkciu ako model ( $Ae^{Bt}$ ) a regresiu ako metódu na určovanie parametrov modelu. Na moju poznámku, že sa to môže hodiť iba pre krátky začiatkový úsek, študent potvrdil, že aj rakúsky kolega na to upozorňoval, a dodal, že v prípade Slovenska mu predpovede “sedeli” až do počtu denných prípadov 61, čo ilustroval grafom s údajmi z 11. marca:



Tento príbeh ukazuje správne modelovanie procesu šírenia infekcie v začiatkovom štádiu, keď sa infekcia šíri exponenciálne, a správne rozhodovanie na základe modelu, keď rakúsky kolega zrušil kontaktnú výučbu a prešiel na online vyučovanie.

## Výstupy programu profesora Milana Batistu

Ako už bolo poznamenané pred mesiacom, model typu SIR bol aktuálne implementovaný v prostredí MATLAB profesorom mechaniky Milanom Batistom z Univerzity v Ljubljane a je uverejnený na MATLAB Central File Exchange (<https://www.mathworks.com/matlabcentral/fileexchange/74658>) pod názvom “fitVirusCOVID19”. Tento model je aktualizovaný a vylepšovaný prakticky denne. Pre tých, ktorí nepoužívajú MATLAB alebo nemajú na to čas, profesor Batista (zrejme so svojimi spolupracovníkmi) vytvoril aj webovú stránku, na ktorej sú s malým oneskorením (jeden-dva dni) dostupné aktuálne výstupy pre viaceré výbrané krajiny vrátane Slovenska aj jeho susedov: <https://www.fpp.uni-lj.si/en/research/research-laboratories-and-the-programme-team/research-programme-team/>.

Skratka “SIR” je odvodená od začiatkových písmen slov “susceptible, infected, recovered” (nakaziteľní, infikovaní, uzdravení). Sú známe viaceré modifikácie a rozšírenia, ktoré sú označované skratkami SEIR, SEIRQ, SEIRQD, atď., kde pribúdajú E (exposed – vystavení možnosti nákazy), Q (quarantined – umiestnení do karantény), D (delay – časový posun, inkubačná doba), a ďalšie. Tu však vzniká problém so vstupnými dátami. Jediný reálny použiteľný údaj, ktorý je momentálne dostupný pre prakticky všetky krajiny, je celkový počet infikovaných (total cases) v jednotlivých krajinách ku konkrétnemu dátumu (<https://ourworldindata.org/coronavirus-source-data>). Údaje pre E, Q, D, atď., momentálne nie sú.

Profesor Batista vychádzal z modelu typu SIR a zredukoval ho na jednu diferenciálnu rovnicu, v ktorej figuruje práve celkový počet infikovaných  $C(t) = I(t) + R(t)$ . Táto rovnica má parametre, ktoré sa určujú optimalizačnou metódou tak, aby výsledné riešenie čo najlepšie vystihovalo dostupné dáta. Potom priebeh tohto riešenia slúži ako aktuálna predikcia budúceho vývoja. S pribúdajúcimi dátami sa parametre modelu pre každú krajinu priebežne upresňujú – a preto sa upresňuje aj predikcia budúceho vývoja. Daný model umožňuje odhadnúť celkový počet infikovaných a celkovú dobu trvania šírenia infekcie, a práve tieto ukazovatele sú kľúčové.

Tento postup by sa dal obrazne porovnať s používaním povedzme Google Maps na navigáciu pri ceste autom alebo aj pešo. Predpokladaný čas príchodu do cieľovej destinácie, ktorý sa zobrazuje na displeji, je kontinuálne aktualizovaný na základe údajov o pohybe subjektu.

Uplynul mesiac; pozrime sa teraz, ako program profesora Batistu funguje. Nižšie na jednotlivých stranách sú uvedené výstupy programu Milana Batistu podľa údajov dostupných v dňoch 10., 17., 26. a 30. apríla. Na uvádzaných grafoch sú farbami vyznačené typické časti priebehu procesu: úvodný exponenciálny rast je na bielom pozadí, prakticky lineárna fáza je na pozadí ružovej farby, fáza spomaľovania je na žltom pozadí, fáza ukončenia procesu je na pozadí zelenej farby. Farebné označenia a spôsob určovania týchto fáz zaviedol profesor Batista pre informatívnejšiu vizualizáciu výsledkov.

Začneme našimi susedmi.

Vidieť, že v prípade Rakúska prakticky od začiatku model dáva predikciu celkového počtu infikovaných približne 15 tisíc osôb, a to ku koncu apríla alebo začiatku mája. Doba trvania procesu je potom odhadovaná na približne dva mesiace.

Podobne to vyzerá aj v prípade Českej republiky. Celkový počet infikovaných je odhadovaný na približne 8 tisíc osôb pri celkovom trvaní procesu približne dva mesiace, s ukončením niekedy koncom mája.

V prípade Maďarska celkový odhadovaný počet infikovaných vychádzal zo začiatku na približne 1,5 tisíc osôb, ale s pribúdajúcimi dátami sa začal približovať k počtu okolo 4 tisíc, s odhadovaným ukončením procesu niekedy ku koncu júna.

V prípade Poľska celkový odhadovaný počet infikovaných vychádzal na začiatku približne na 10 tisíc, ale s pribúdajúcimi dátami sa začal približovať k počtu okolo 15 tisíc, s odhadovaným ukončením procesu niekedy ku koncu júna.

Slovensko sa zrejme vďaka rýchlym opatreniam vyhlo úvodnej exponenciálnej fáze infekcie, preto prvé výstupy (10. a 17. apríla) pri malých počtoch infikovaných osôb evidentne nezodpovedali predpokladom SIR modelu. Ale ďalej už aj Slovensko "naskočilo" na tento model (výstupy z 26. a 30. apríla), podľa ktorého by sa dal očakávať celkový počet infikovaných osôb niekde okolo 2–3 tisíc, s odhadovaným ukončením procesu ku koncu júna.

Z iných krajín uvidíme, ako v predošlom príspevku, Taliansko, Španielsko, USA a UK. Vývoj odhadovaných počtov infikovaných a odhadovaného trvania procesu je zrejmy zo zodpovedajúcich grafov, ktoré sú pripojené nižšie.

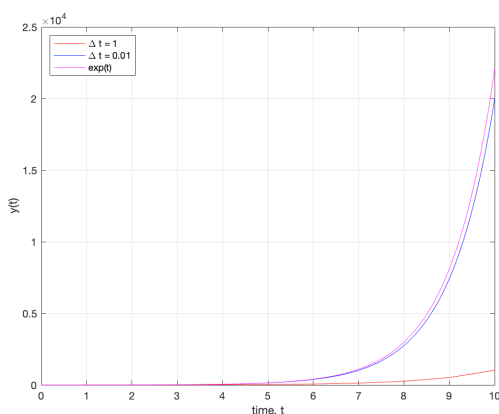
## Bližší pohľad na model IZP

Aj keď IZP vo svojich kódoch a prezentáciách používa skratku SIR, nie je to celkom tak.

Model SIR má podobu sústavy diferenciálnych rovníc, navyše sústavy nelineárnej. Tento model je odvodený za takých predpokladov, ktoré znamenajú, že sa nezaobráme individuálnymi vírusmi ani individuálnymi objektmi vírusovej infekcie, ale pozeráme sa na nich ako na spojité veličiny; čas je tiež spojitá veličina.

Numerické riešenie diferenciálnych rovníc sa realizuje, zjednodušene povedané, s veľmi malým krokom diskretizácie uvažovaného časového intervalu. Ak potrebujeme hodnoty neznámej funkcie v bodoch  $t=1$ ,  $t=2$ ,  $t=3$  a  $t=4$  z intervalu  $[0, 4]$ , ktoré predstavujú prvé štyri dni, musíme použiť dostatočne malý krok diskretizácie časovej premennej (napríklad,  $\Delta t = 0.001$ ) a z vypočítaných hodnôt, ktorých bude 4000, vybrať iba štyri. Tak je to implementované v programe profesora Batistu. Práve toto mu umožňuje aproximovať dátové body spojitými riešeniami diferenciálnych rovníc a touto cestou určovať potrebné dôležité parametre procesu.

Pri pohľade do zdrojového kódu IZP však vidieť, že výpočty sú tam realizované s časovým krokom rovným jednému dňu ( $\Delta t = 1$ ), čo hneď vedie k chybným výsledkom. Toto vie každý študent matematiky, fyziky alebo technických vied po absolvovaní základného kurzu numerických metód. Napríklad, riešením rovnice  $y'(t) = y(t)$ , ktoré vyhovuje začiatočnej podmienke  $y(0) = 1$ , je exponenciálna funkcia:  $y(t) = e^t$ . Toto exaktné analytické riešenie je na nasledujúcom obrázku zobrazené fialovou čiarou. Ak použijeme najjednoduchšiu numerickú aproximáciu prvej derivácie, dostávame Eulerovu metódu pre riešenie tejto diferenciálnej rovnice s časovým krokom  $\Delta t = h$ :  $y(t_{k+1}) = y(t_k) + h y(t_k)$ . Pri dostatočne malom časovom kroku, napríklad pri  $h = 0.01$ , približné numerické riešenie (modrá čiara) je dosť blízko k exaktnému. Ale pri kroku  $h = 1$  dostávame výsledok, ktorý je zobrazený červenou čiarou a ktorý sa rozhodne nedá považovať ani za akokoľvek približné riešenie úlohy.



Tento najjednoduchší príklad ukazuje, nakoľko nekorektné čísla je možné získať nesprávnou voľbou dĺžky časového kroku pri riešení dokonca najjednoduchších lineárnych diferenciálnych rovníc. V prípade nelineárnych diferenciálnych rovníc problémy sú ešte väčšie; k tomu treba pridať aj vplyv chýb pri približnom odhadovaní parametrov modelu.

V dokumente IZP “Dopady obmedzenia mobility na šírenie vírusu Covid-19” sú však uvedené “iteračné kroky simulácie”, za ktoré IZP ďakuje “doc. Mgr. Richardovi Kollárovi, PhD., Mgr. Kataríne Boďovej, PhD. z FMFI UK v Bratislave a Dr. Zuzane Chladnej z FMPH UK v Bratislave” (citát je doslovný; IZP našiel na UK novú fakultu?). V týchto

krokoch sa používajú výrazy “počet náchylných na chorobu v čase  $t+1$ ”, “počet infikovaných v čase  $t+1$ ” a vzorec pre “počet vyliečených v čase  $t+1$ ”, ktorý by mal byť aproximáciou tretej rovnice klasického SIR modelu, iba potvrdzuje, že časový krok v modeli IZP je rovný jednému dňu. Pokusy zohľadniť iné faktory (mobilita obyvateľov, čas interakcie jedincov...) možno odvedli pozornosť od tohto principiálneho problému, ale ho nekompensovali ani nevyriešili.

Treba dodať, že jedno z vystúpení (<https://video.sme.sk/c/22375901/rozhovory-zkh-sef-analytikov-smatana-a-koronavirus-video.html>, konkrétne úsek od 5m45s po 6m05s) riaditeľa IZP Martina Smatanu, ktorý neskôr podľa správy TASR požiadal o skončenie pracovného pomeru dohodou k 15. júnu, ukázalo, že zrejme ani nevie, ako vyzerá priebeh exponenciálnej funkcie, či už  $y(t) = e^t$  alebo  $y(t) = e^{-t}$ .

## Kritika modelu IZP

Výsledky predkladané Inštitútom pre zdravotnú politiku (IZP) pochopiteľne púťali a púťajú pozornosť verejnosti a nevyhli sa kritike. Anonymný autor alebo skupina autorov z portálu DSL.SK v príspevku “Štátna predpoveď šírenia koronavírusu bola veľmi amatérska, s úplne zlým výsledkom” poukázala na viaceré chyby v programovom kóde, ktorý IZP uverejnil na GitHub. Ale, ako napísali,

*“Ešte horšie ako tieto chyby vyznieva dôvod, ktorý evidentne prispel k tomuto stavu. Štát totiž v skutočnosti nevytvoril vlastný kód ale ako kľúčový kód pre samotnú simuláciu použil kód z [blogu](#) zrejme arménskeho PhD študenta Gevorga Yeghikyana.*

*<...>*

*Gevorga Yeghikyana sme kontaktovali s otázkami, prvou bolo poukázanie na použitie parametra znižovania mobility aj pri prenose infekcie medzi ľuďmi v jednej lokalite. To evidentne nebolo jeho zámerom a nakoniec podľa toho vytvoril aj rovnice. To samozrejme potvrdil, keď v reakcii na túto otázku uviedol «I have put the wrong code in the blog but never got the time to change it.»”*

IZP tiež vo svojej odpovedi, ktorá je uverejnená na webe DennikN a na webstránke IZP, reagoval aj na viaceré kritické komentáre predložené Filipom Vítekom na jeho webe.

Ani autor/autori z DSL.SK, ani F. Vítek sa nezaoberali vzťahom medzi skutočným klasickým SIR modelom a prístupom IZP.

## Záver

V súčasnosti to vyzerá tak, že použitie modelu SIR v implementácii profesora Milana Batistu dáva realistickejšie predikcie celkového počtu infikovaných a celkovej doby trvania procesu v prípade veľkého počtu krajín. Toto by sa možno dalo vysvetliť tým, že v časovom rade hodnôt kumulatívneho počtu infikovaných (na makroúrovni) sa nepriamo odrážajú všetky tie faktory, ktoré sa iní, vrátane IZP, pokúšajú zohľadňovať na mikroúrovni.

Pripravil: prof. RNDr. Igor Podlubný, DrSc.  
4.5.2020

## Odkazy:

1. Smith, D., and Moore, L.: "The SIR Model for Spread of Disease". *Journal of Online Mathematics and its Applications*, December 2001, <https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-introduction> (Retrieved May 3, 2020).
2. Hethcote, H.: The Mathematics of Infectious Diseases. *SIAM Review*, vol. 42, no. 4, pp. 599–653, [doi:10.1137/S0036144500371907](https://doi.org/10.1137/S0036144500371907).
3. Batista, M.: fitVirusCOVID19: Estimation of coronavirus COVID-19 epidemic evaluation by the SIR model. *MATLAB Central File Exchange*, <https://www.mathworks.com/matlabcentral/fileexchange/74658-fitviruscovid19> (Retrieved May 3, 2020).
4. Forecasting COVID-19 pandemic. <https://www.fpp.uni-lj.si/en/research/research-laboratories-and-the-programme-team/research-programme-team/>
5. Coronavirus Source Data, <https://ourworldindata.org/coronavirus-source-data> (Accessed May 3, 2020).
6. DSL.SK: Štátna predpoveď šírenia koronavírusu bola veľmi amatérska, s úplne zlým výsledkom, 4.4.2020, <http://www.dsl.sk/article.php?article=23646> (Retrieved May 3, 2020).
7. Vítek F.: Ako urobiť IZP model užitočným? (aneb ako si vôbec dovoľujem kritizovať IZP). 6.4.2020, <https://mocnedata.sk/vyhrady-k-izp/> (Retrieved May 3, 2020).
8. Zdrojový kód publikovaný IZP: [https://github.com/institute-of-health-policies-sk/SIRmodel\\_COVID-19](https://github.com/institute-of-health-policies-sk/SIRmodel_COVID-19) (Retrieved May 3, 2020).
9. Yeghikyan, G.: Modelling the coronavirus epidemic in a city with Python. <https://towardsdatascience.com/modelling-the-coronavirus-epidemic-spreading-in-a-city-with-python-babd14d82fa2> (Retrieved May 3, 2020).
10. IZP: Pri modelovaní scenárov sa radíme s odborníkmi, každá konštruktívna kritika je zapracovaná. 6.4.2020. <https://dennikn.sk/blog/1841096/pri-modelovani-predikcie-sa-radime-s-odbornikmi-kazda-konstruktivna-kritika-je-zpracovana/> (Retrieved May 3, 2020).
11. IZP: Dopady obmedzenia mobility na šírenie vírusu Covid-19. Aktualizácia č.2.2 (29.3.2020), [https://izp.sk/wp-content/uploads/2020/04/CRN\\_IZP\\_29-3-2020\\_final4.pdf](https://izp.sk/wp-content/uploads/2020/04/CRN_IZP_29-3-2020_final4.pdf) (Retrieved May 3, 2020).