

UNIVERZITA PALACKÉHO V OLMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

DIPLOMOVÁ PRÁCA

2018

Bc. Jana BÍROVÁ

Prírodovedecká fakulta Univerzity Palackého v Olomouci
Katedra matematickej analýzy a aplikácií matematiky

Modelovanie šírenia epidémii

Bc. Jana Bírová

Diplomová práca



Katedra matematickej analýzy a aplikácií matematiky

Vedúci diplomovej práce: **RNDr. Tomáš Fürst, PhD.**

Vypracovala: **Bc. Jana Bírová**

Študijný program: N1103 Aplikovaná matematika

Študijný odbor: Aplikácie matematiky v ekonómii

Forma štúdia: prezenčná

Rok odovzdania: 2018

Prehlásenie

Prehlasujem, že som svoju diplomovú prácu spracovala samostatne pod vedením pána RNDr. Tomáša Fürsta, PhD. a všetky použité zdroje som uviedla v zozname literatúry.

V Olomouci dňa 20.04.2018

.....

Pod'akovanie

Týmto by som sa chcela poďakovať vedúcemu mojej diplomovej práce RNDr. Tomášovi Fürstovi, PhD. za vedenie pri písaní mojej diplomovej práce, za pomoc, podporu, ochotu, nesmiernu trpezlivosť, motiváciu a za cenné rady pri riešení problémov ktoré sa pri písaní práce vyskytli.

BIBLIOGRAFICKÁ IDENTIFIKÁCIA

Autor: Bc. Jana Bírová

Názov práce: Modelovanie šírenia epidémii

Typ práce: Diplomová práca

Pracovisko: Katedra matematickej analýzy a aplikácii matematiky

Vedúci práce: RNDr. Tomáš Füst, PhD.

Rok obhajoby práce: 2018

Abstrakt: Diplomová práca sa zaoberá epidemiologickým modelom SIR, jeho numerickým riešením v programe MatLab a nafiťovaním modelu na reálne dáta chrípky v sezóne rokov 2016-2017 v Žilinskom samosprávnom kraji. Prvá zväčša časť obsahuje teoretické informácie o modeli SIR a o riešení modelu v programe MatLab. V druhej časti sú získané informácie aplikované na reálne dáta chrípky.

Kľúčové slová: epidemiologický model, model SIR, chrípka, fitovanie dát, MatLab

Počet strán: 35

Počet príloh: 0

Jazyk: slovenský

BIBLIOGRAPHICAL IDENTIFICATION

Author: Bc. Jana Bírová

Title: Epidemic modeling

Type of thesis: Master's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: RNDr. Tomáš Fürst, PhD.

The year of presentation: 2018

Abstract: Master's thesis deals with epidemic model SIR, his numerical solution in MatLab and fitting model of real data of influenza in the season 2016-2017 in Žilina self-governing region. The first part contains theoretical information about SIR model and model solution in MatLab. In the second part obtained information are applied to real data of influenza.

Key words: epidemic model, model SIR, influenza, data fitting, Mat-Lab

Number of pages: 35

Number of appendices: 0

Language: slovak

Obsah

1	Úvod	10
2	Epidemiologické modely	11
2.1	Epidemiológia	11
2.2	Základný koncept	12
2.3	Model SIR	14
2.3.2	Riešenie v programe MatLab	20
3	Reálne dáta	23
3.1	Chrípka	23
3.2	Fitovanie reálnych dát na model SIR	23
3.3	Výhody a nevýhody modelu SIR	32
4	Záver	34

Zoznam obrázkov

1	Celkový počet prípadov eboly v rokoch 2014 - 2016	12
2	Schéma modelu SIR	15
3	Fázový diagram pre $s_0 > \frac{\beta}{\alpha}$	19
4	Fázový diagram pre $s_0 < \frac{\beta}{\alpha}$	20
5	Vývoj ochorenia	21
6	Vykreslenie dát	24
7	Simulované a syntetické dáta	29
8	Simulované a reálne dáta	31
9	Infikovaní jedinci	32

Zoznam tabuliek

1	Reálne dáta	25
2	Tabuľka	27

1 Úvod

Od nepamäti sa ľudia snažili pochopiť prečo a ako sa šíria rôzne ochorenia. Už grécky lekár Hippokratés, ktorý je niekedy považovaný za zakladateľa epidemiológie, skúmal vzťahy medzi výskytom ochorenia a vplyvom prostredia. Na vznik ochorení však bolo zo začiatku pozerané ako na trest za prílišné bohatstvo. Tento názor sa objavil v dielach filozofov Platóna a Rousseaua.

V priebehu storočí sa štúdium šírenia ochorení stalo neoddeliteľnou súčasťou lekárskeho odvetvia, z ktorého v roku 1854 vznikol vedný odbor epidemiológia. Od začiatku 20. storočia sa v epidemiológii začali uplatňovať matematické postupy, presadzované najmä Ronaldom Rossom, Andersonom Grayom McKendrickom a ďalšími. V roku 1927 publikovali A. G. McKendrick a W. O. Kermack teóriu, ktorá je považovaná za všeobecnú teóriu prenosu infekčných chorôb.

K pochopeniu a simulácii rôznych ochorení existuje množstvo druhov epidemiologických modelov. Pri písaní mojej práce som si za cieľ dala porozumenie jednému zo základných epidemiologických modelov, modelu SIR, implementovanie numerického riešenia v programe MatLab a jeho použitie na nasimulovanie ochorenia chrípky. V prvej časti práce sa budem zaoberať teoretickou stránkou epidemiologického modelu SIR a jeho riešením v programe MatLab. V druhej časti sa budem zaoberať vývojom ochorenia chrípky v Žilinskom samosprávnom kraji v sezóne rokov 2016-2017 a pokúsim sa získané dáta implementovať na model SIR. Dúfam, že pri písaní mojej práce budem schopná nadobudnuté poznatky využiť k splneniu môjho cieľa.

2 Epidemiologické modely

2.1 Epidemiológia

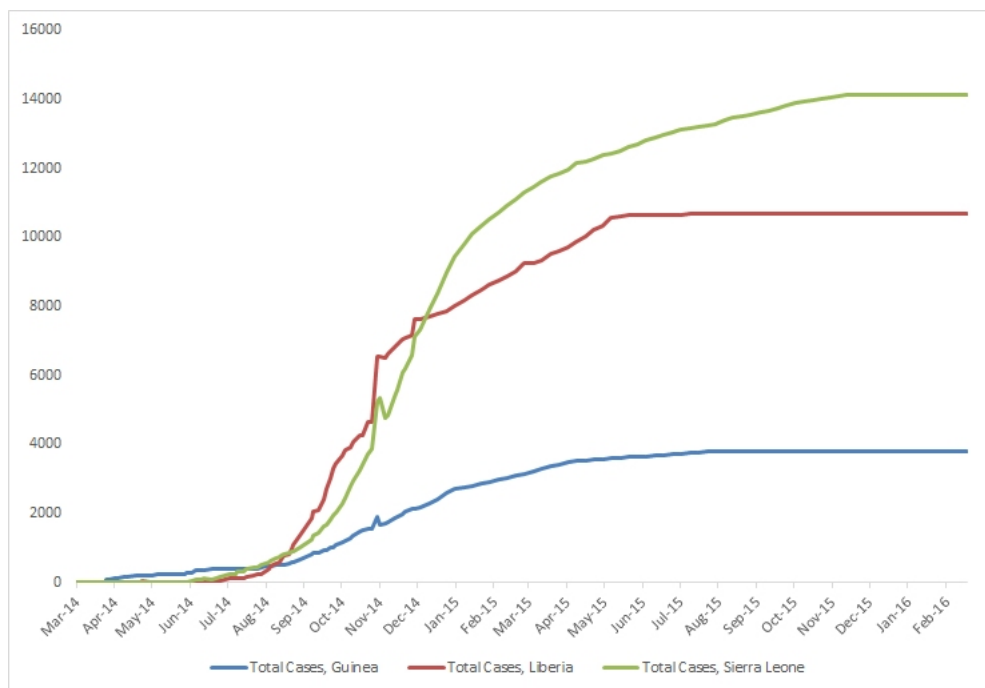
Epidemiológia je medicínsky vedný odbor, ktorý sa zaoberá teoretickými a praktickými otázkami ochrany obyvateľstva, predovšetkým pred prenosnými ochoreniami. Cieľom epidemiológie je zdravie ochraňovať, podporovať a obnovovať. Základom je epidemiologická metóda práce pozostávajúca zo zberu údajov, deskripcie, analýzy, interpretácie výsledkov, navrhovania opatrení a vyhodnocovania efektívnosti opatrení. Spolu s hygienou tvorí súbor základných disciplín preventívnej medicíny a v spolupráci s príslušnými klinickými odbormi sa podieľajú na tvorbe štúdií problémov výskytu a prevencie prenosných i neprenosných chorôb.

Epidemiológovia vo svojej práci používajú širokú škálu štúdií - od pozorovaní až po riadené experimenty. Ich činnosť je charakterizovaná ako deskriptívna, analytická a experimentálna. Epidemiologické štúdie sa zameriavajú, pokiaľ je to možné, na objektívne posudzovanie dopadu vonkajších vplyvov, biologických činiteľov, stresu alebo chemikálií na chorobnosť a úmrtnosť. Rozpoznávanie príčinných vzťahov medzi týmito vplyvmi a ich dôsledkami je dôležitou náplňou epidemiologickej praxe. Osobitné postavenie v epidemiológii má štúdium príčin vzniku ochorení. Epidemiológia sa zaoberá odvodzovaním hypotéz, ktoré vysvetľujú znaky rozloženia ochorení v zmysle špecifických charakteristík alebo skúseností. Tieto overuje na základe účelovo navrhnutých štúdií a zozbieraných údajov. V určitom zmysle takto napomáha klasifikácii chorých osôb [5], [7].

Epidémia je rýchle šírenie infekčných ochorení u veľkého počtu ľudí v danej populácii v krátkom časovom období. Epidémia môže byť obmedzená na jedno miesto, ak sa však rozšíri do iných krajín alebo kontinentov a postihuje značný počet ľudí, môže sa to označiť za pandémiu. Epidémie niektorých chorôb, ako je napríklad chrípka, sú definované tak, že dosahujú určité definované zvýšenie výskytu nad základnú mieru [9], [10].

Ako príklad vypuknutia epidémie si môžeme uviesť prepuknutie ochorenia eboly v roku 2014 v západnej Afrike, ktorá trvala až do marca 2016. Počas celého tohto obdobia sa ochorením nakazilo 28 646 ľudí a vyžiadalo si 11 323 životov [12].

Obr. 1: Celkový počet prípadov eboly v rokoch 2014 - 2016



Príkladom, kedy k prepuknutiu epidémie nedošlo je chrípková sezóna v rokoch 2007-2008 v Slovenskej republike, kedy aktivitu chrípky môžeme charakterizovať ako nízku. V celej krajine nebola zaznamenaná typická zimná epidémia, ochorenia sa vyskytovali len sporadicky. Celkovo sa ochorením chrípky v tejto sezóne nakazilo 238 764 obyvateľov Slovenskej republiky [17].

2.2 Základný koncept

Táto sekcia vychádza z [2], [6]. Najpoužívanejším modelom v problematike epidémii je Kermackov-McKendrikov model [11]. Napriek tomu, že niektoré skutočnosti zjednodušuje, je jeho analýza užitočná, pretože pomáha predvídať dynamiku šírenia infekčných chorôb. Medzi faktory, ktoré neuvažuje, patria niektoré elementárne skutočnosti spojené s existenciou a funkciou akejkoľvek spoločnosti. Jedná sa najmä o pôrodnosť a úmrtnosť zdravých jedincov, prípadne o dĺžku choroby, či je vôbec porovnateľná s dobou jednotlivých životných etáp človeka. Ďalej nezohľadňuje vekovú štruktúru v sledovanej populácii, inkubačnú dobu ochorenia a ďalšie kritéria.

Epidemiologické modely sú navrhnuté ako sústava diferenciálnych rovníc, kde sledovanú populáciu považujeme za dostatočne veľkú na to, aby sme s počtom ľudí v sledovanej populácii mohli pracovať ako so spojitou veličinou. Ich podstatou je rozdelenie celkovej populácie do disjunktných skupín, kde jedinec patrí do danej skupiny na základe toho, v akom zdravotnom stave sa vzhľadom na šíriacu sa infekčnú chorobu nachádza. Prechod medzi jednotlivými skupinami je podmienený tzv. *koeficientmi prechodu*, ktoré vplývajú na priebeh choroby.

Veľkosť populácie považujeme za nemennú. Takéto zjednodušenie je prípustné, ak zmena veľkosti populácie je bezvýznamná vzhľadom k rýchlosti priebehu choroby alebo ak zavedieme predpoklad, že počet narodených a zomretých je v rovnováhe.

Teraz si tento model zostavíme. Vychádzame z dôležitého predpokladu, že v uzavretej populácii o N jedincoch sa v časovom okamžiku t nachádzajú nasledujúce skupiny:

$S(t)$ - počet zdravých, chorobou ohrozených jedincov,
 $I(t)$ - počet infikovaných, respektíve osôb, ktoré aktívne chorobu prenášajú
 $R(t)$ - počet rezistentných osôb, ktorí už ochorenie nemajú (vyliečení, zomretí).

Ďalej vychádzame z týchto predpokladov:

1. Choroba sa šíri kontaktom medzi infikovanými, zdravými a ohrozenými jedincami.
2. Choroba nemá latentné obdobie, čo znamená, že sa choroba začne vyvíjať bezprostredne po nakazení infikovaným.
3. Populácia je homogénna. Všetci ohrození jedinci sú teda rovnako ohrození a všetci infikovaní sú rovnako infekční. Pravdepodobnosť stretnutia akýchkoľvek dvoch jedincov je v populácii rovnaká.
4. Populácia je autonómna, má teda konštantnú veľkosť. Neberieme do úvahy ani narodenie nových jedincov, ani migráciu. Všetci zomretí sú zahrnutí do skupiny osôb $R(t)$, ktorí už chorobu absolvovali.

Posledný predpoklad môžeme napísať v tvare:

$$S(t) + I(t) + R(t) = N = konst. \quad (1)$$

Vyššie uvedené predpoklady sú síce istým spôsobom obmedzujúce, ale v mnohom sa príliš nelíšia od skutočnosti. Príkladom je podmienka konštantnej veľkosti populácie - pri priebehu mnoha ochorení je reálna, pretože veľké množstvo chorôb má natoľko rýchly priebeh, že prírastok aj úbytok jedincov v

populácii za daný časový úsek je úplne zanedbateľný.

Neuvažovanie latentného obdobia umožňuje počítať priebeh bez meškania, čo je značným uľahčením pri výpočtoch. Zjednodušením je aj predpoklad, že stretnutie jedincov navzájom je u všetkých rovnako pravdepodobné.

2.3 Model SIR

Model založený na existencii troch vyššie uvedených kategórií osôb v populácii nazývame modelom SIR. Matematická konštrukcia vychádza z nasledujúcich predpokladov:

- Prírastok infikovaných jedincov je úmerný počtu ohrozených a infikovaných jedincov, tj. $\propto \alpha S(t)I(t)$, kde α je priemerný počet jedincov, ktorých infikuje jeden infikovaný jedinec za časovú jednotku. Ohrozených osôb rovnakou rýchlosťou ubúda.
- Rýchlosť, s akou ubúda infikovaných jedincov, či už vyliečením alebo úmrtím, je úmerná počtu infikovaných osôb, t.j. $\propto \beta I(t)$, teda β je miera zotavenia.
- Populácia je natoľko veľká, že vyvolané zmeny môžeme považovať za spojité.

Pre zjednodušenie si označíme $s(t) = S(t)/N$, čo je pomer ohrozených jedincov v čase t k celkovej sledovanej populácii. Ďalej si označíme $i(t) = I(t)/N$, čo je pomer infikovaných jedincov v čase t k celkovej sledovanej populácii. A nakoniec si označíme $r(t) = R(t)/N$, čo je pomer rezistentných jedincov v čase t k celkovej sledovanej populácii.

Za týchto podmienok je matematický model definovaný troma diferenciálnymi rovnicami popisujúcimi dynamiku jednotlivých kategórií osôb.

$$s'(t) = -\tilde{\alpha}s(t)i(t) \tag{2}$$

$$i'(t) = \tilde{\alpha}s(t)i(t) - \beta i(t) \tag{3}$$

$$r'(t) = \beta i(t) \tag{4}$$

Tieto rovnice ďalej dopĺňajú počiatočné podmienky v tvare:

$$s(0) = s_0 > 0 \quad (5)$$

$$i(0) = i_0 > 0 \quad (6)$$

$$r(0) = r_0 = 0 \quad (7)$$

Tento systém je nelineárny a nemá analytické riešenie. Aj napriek tomu však môžu byť analyticky odvodené významné výsledky.

Všimnime si, že:

$$\frac{ds}{dt} + \frac{di}{dt} + \frac{dr}{dt} = 0, \quad (8)$$

z toho vyplýva, že:

$$s(t) + i(t) + r(t) = 1. \quad (9)$$

Vyššie uvedená rovnica nám naznačuje, že sa stačí zaoberať len dvomi z troch premenných rovnice a tretiu vieme dopočítať z rovnice (9).

Obr. 2: Schéma modelu SIR



Pre úplnú špecifikáciu modelu by mali byť šípky na Obr. 2 označené prechodovými rýchlosťami. Medzi s a i je prechodová rýchlosť $\tilde{\alpha}i$, kde α je miera rýchlosti prenosu. Medzi i a r je prechodová rýchlosť β .

Dynamika infekčnej triedy je závislá od nasledujúceho pomeru:

$$R_0 = \frac{\tilde{\alpha}}{\beta}, \quad \text{kde} \quad \tilde{\alpha} = \frac{\alpha}{N}, \quad (10)$$

nazývaný aj *základné reprodukčné číslo*, ktoré je interpretované nasledovne: berieme do úvahy rýchlosť ochorenia keď práve začína, keď sa vyskytuje iba niekoľko prípadov ochorenia a zvyšok sledovanej populácie je v skupine

zdravých, chorobou ohrozených jedincov - čo sa v epidémiológii nazýva *naivná populácia* - a berieme do úvahy zdravých jedincov, ktorí sa nakazia v rannom štádiu prepuknutia ochorenia. Základné reprodukčné číslo je definované ako priemerný počet ďalších jedincov, ktorí prejdú ochorením, predtým ako sa zotavia. V našom modeli vieme vypočítať R_0 priamo.

Vyjadrením i z 3. diferenciálnej rovnice modelu (rovnic (4)) a dosadením do 1. diferenciálnej rovnice modelu (rovnic (2)), separovaním premenných a integrovaním dostaneme:

$$\begin{aligned} s'(t) &= -s(t)r'(t)\frac{\tilde{\alpha}}{\beta} \\ s'(t) &= -R_0s(t)r'(t) \\ \frac{s'(t)}{s(t)} &= -R_0r'(t) \end{aligned}$$

vieme, že platí

$$\int_0^t \frac{s'(t)}{s(t)} dt = \ln|s(t)|, \quad (11)$$

potom

$$\begin{aligned} \int_0^t \frac{s'(t)}{s(t)} dt &= \int_0^t -R_0r'(t) dt \\ \frac{\ln|s(t)|}{\ln|s(0)|} &= -R_0(r(t) - r(0)) \\ \frac{s(t)}{s(0)} &= e^{-R_0(r(t)-r(0))} \\ s(t) &= s(0)e^{-R_0(r(t)-r(0))} \end{aligned} \quad (12)$$

Veta 2.3.1. *Nech $(s(t), i(t))$ je riešenie rovníc (2), (3) v $T = \{(s, i) \mid s \geq 0, i \geq 0, s + i \leq 1\}$. Ak $R_0s_0 \leq 1$, potom $i(t)$ klesá k 0 pre $t \rightarrow \infty$. Ak $R_0s_0 > 1$, potom $i(t)$ najskôr rastie na maximálnu hodnotu rovnú $1 - \frac{1}{R_0} - \frac{\log(R_0s_0)}{R_0}$ a potom klesá k 0 pre $t \rightarrow \infty$. $s(t)$ je klesajúca funkcia a limitná hodnota s_∞ je*

koreň $(0, \frac{1}{R_0})$ rovnice

$$1 - s_\infty + \frac{\log \frac{s_\infty}{s_0}}{R_0} = 0$$

[16].

Na základe vyššie uvedenej vety môžeme v prípade limity pre $t \rightarrow \infty$ označiť $i_\infty = 0$ a podľa rovnice (9) môžeme napísať

$$s_\infty + r_\infty = 1. \quad (13)$$

A teda v prípade limity $t \rightarrow +\infty$, podiel rezistentných jedincov sa riadi touto transcendentálnou rovnicou ¹

$$r_\infty = 1 - s(0)e^{-R_0(r_\infty - r(0))} \quad (14)$$

Rovnica (14) ukazuje, že na konci epidémie, pokiaľ $s(0) = 0$, nie všetci jedinci z populácie sú rezistentní, takže niektorí musia zostať v skupine ohrozených. To znamená, že koniec epidémie je spôsobený skôr poklesom počtu infikovaných jedincov než absolútnym nedostatkom ohrozených jedincov. Úloha základného reprodukčného čísla je teda veľmi dôležitá.

Hlavnou otázkou pri vypuknutí akejkoľvek epidémie je, či sa pre dané parametre modelu a počiatkové podmienky bude nákaza šíriť a následne, ako bude šírenie prebiehať. Ďalej nás zaujíma, ako vážna epidémia bude, teda aké maximálne hodnoty nadobudne stav skupiny infikovaných a ako sa bude vyvíjať stav skupiny rezistentných voči chorobe.

Po prepísaní rovnice (4) pre infikovaných dostaneme:

$$\frac{di}{dt} = (R_0 s(t) - 1)\beta i(t), \quad (15)$$

to znamená, že ak:

$$R_0 > \frac{1}{s(0)}, \quad (16)$$

potom:

¹Transcendentálna rovnica je rovnica obsahujúca transcendentálnu funkciu premennej. Takéto rovnice často nemajú uzavreté riešenie.

$$\frac{di}{dt}(0) > 0, \quad (17)$$

teda, že hneď na začiatku narastie počet infikovaných. Naopak, ak

$$R_0 < \frac{1}{s(0)}, \quad (18)$$

potom

$$\frac{di}{dt}(0) < 0, \quad (19)$$

to znamená, že nezávisle od počiatkovej veľkosti náchylnej populácie ochorenie nemôže nikdy spôsobiť epidémiu [1], [2].

Teraz si ukážeme príklad na nasledujúcich obrázkoch znázorňujúcich fázové portréty v jednotlivých prípadoch. Na ich vykreslenie opäť použijeme program MatLab:

Algoritmus nám numericky vyrieši model a pre rôzne počiatkové podmienky vykreslí fázový portrét a v prípade, ak $alf < bet$ vykreslí aj "hraničnú" hodnotu.

```

alf = 1;
bet = 9;
close all

Tmax = 100;

for i0 = 0:.1:1

    [T Y] = ode45(@(t,y) SIR(t,y,alf,bet),[0 Tmax], [i0 1-i0]);
    figure(1)
    plot(Y(:,2),Y(:,1), 'b')
    hold on

end

axis([0 1 0 1])
ylabel('infected fraction')
```

```

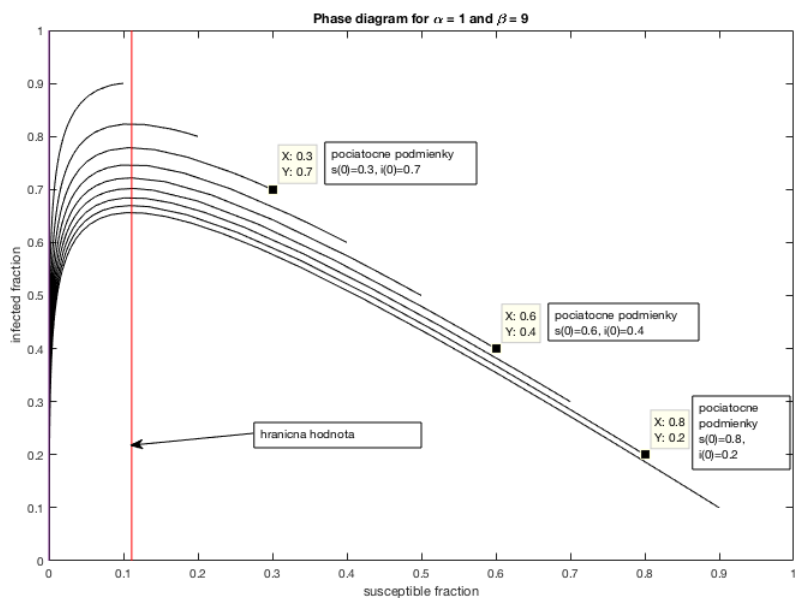
xlabel('susceptible fraction')
title(['Phase diagram for \alpha = ' num2str(alf)
      ' and \beta = ' num2str(bet)])

if alf < bet
    pom = alf/bet;
    xx = [pom; pom];
    yy = [0; 1];
    plot(xx,yy,'r')
end

```

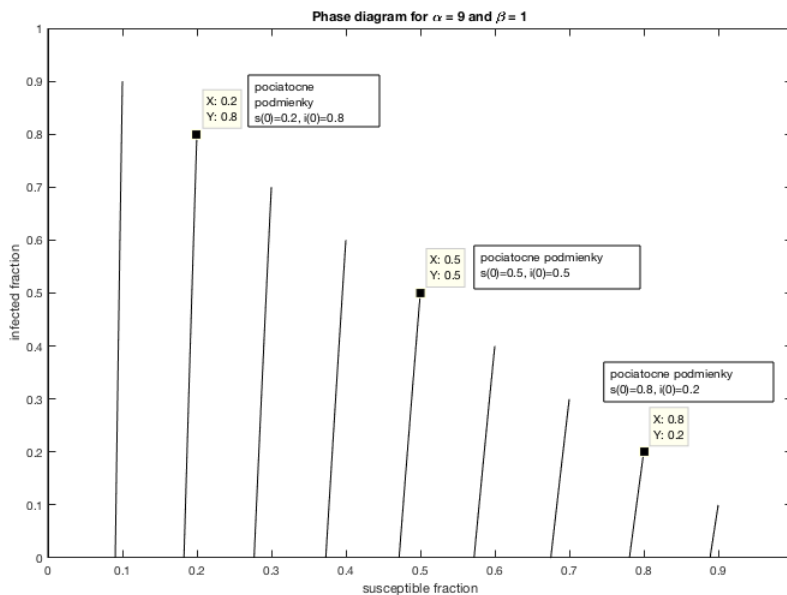
Obr. 3 nám znázorňuje fázový diagram dynamiky vývoja v čase pre rôzne počiatočné podmienky $s(0)$ a $i(0)$ v prípade, keď $s_0 > \frac{\beta}{\alpha}$. Vidíme, že najskôr počet infikovaných rastie, infekcia sa šíri, a po dosiahnutí "hraničnej" hodnoty začína klesať.

Obr. 3: Fázový diagram pre $s_0 > \frac{\beta}{\alpha}$



Naopak, na Obr. 4 vidíme, že v prípade ak $s_0 < \frac{\beta}{\alpha}$ (v tomto prípade $\alpha = 9$, $\beta = 1$), k šíreniu infekcie nedôjde, epidémia neprepukne.

Obr. 4: Fázový diagram pre $s_0 < \frac{\beta}{\alpha}$



2.3.2 Riešenie v programe MatLab

Pomocou programu MatLab si vykreslíme základné riešenie modelu SIR pri zadaných parametroch α a β a pri zadaných vstupných hodnotách.

```
function dy = F(t,y)
dy = [0; 0; 0];
alf = 0.02;
bet = 0.6;

dy(1) = - alf*y(1)*y(2);
dy(2) = alf*y(1)*y(2) - bet*y(2);
dy(3) = bet*y(2);
end
```

```

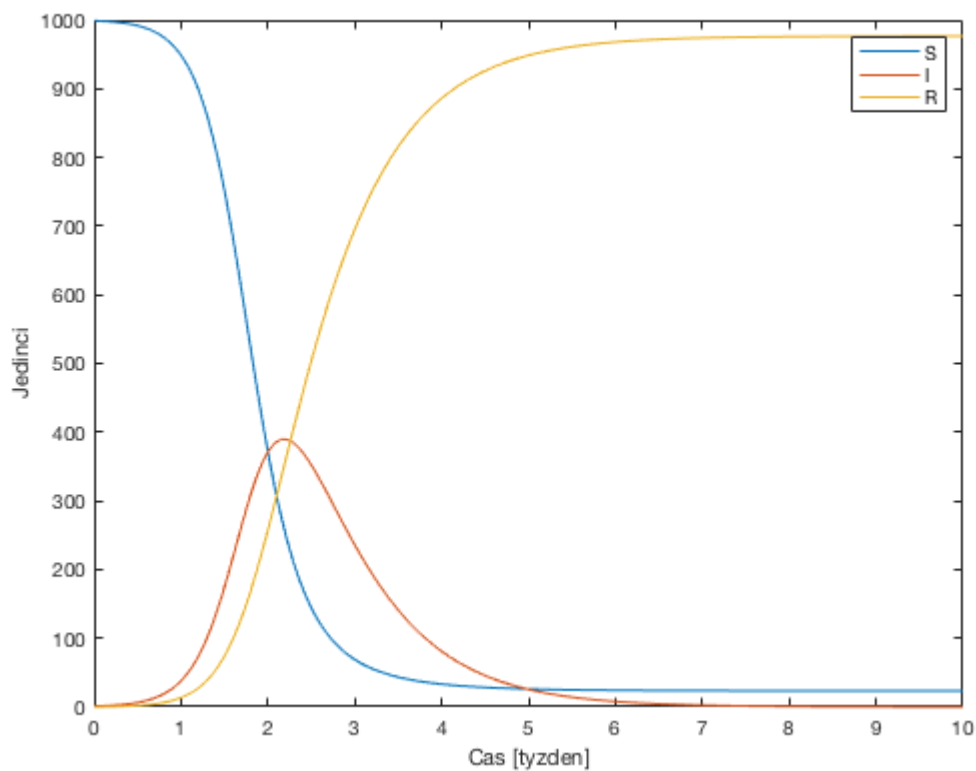
options = odeset('RelTol', 1e-4, 'NonNegative', [1 2 3]);
[t,y] = ode45('F', [0 10], [999 1 0], options);
plot(t,y);
xlabel('Cas');
ylabel('Jedinci');
legend('S', 'I', 'R');

```

Funkcia F nám reprezentuje model SIR. Model je vyriešený pomocou riešiča *ode45*, pri zadanej tolerancii relatívnej chyby a zadanej podmienke nezáporných hodnôt jednotlivých skupín jedincov, spolu so zadaným časom a počiatočnými hodnotami $S(0)$, $I(0)$ a $R(0)$ [8].

Riešič *ode45* je založený na explicitnej metóde Runge-Kutta, páre Dormand-Prince. Je to jednostupňový riešič, pri výpočte $y(t_n)$ potrebuje len riešenie v bezprostredne predchádzajúcom časovom bode $y(t_{n-1})$ [15].

Obr. 5: Vývoj ochorenia



Na Obr. 5 teda vidíme priebeh ochorenia v sledovanej populácii o veľkosti $N = 1000$, kde v čase $t = 0$ je jeden jedinec infikovaný a ostatní jedinci sú v skupine ohrozených, a pri zadaných parametroch $\tilde{\alpha} = 0.005$ [1/týždeň] a $\beta = 1.3$ [1/týždeň].

Pomocou Obr. 5 si môžeme všimnúť množstvo významných informácií. Skupina ohrozených jedincov klesá monotónne (čo je vidieť aj z rovnice (2)), ako sú ohrození jedinci infikovaní a zároveň skupina rezistentných jedincov rastie monotónne. Skupina infikovaných však najskôr rastie, ako sa jedinci nainfikujú, ako sa dostanú do skupiny rezistentných klesá a eventuálne ide k nule pre $t \rightarrow \infty$. Všimnime si, že počet ohrozených jedincov nejde úplne k nule, ale krivka pre $s(t)$ končí nad osou. Je to preto, že keď $i \rightarrow 0$ nie sú žiadny ďalší infikovaní jedinci, ktorí by nakazili zvyšných ohrozených jedincov. Rovnako skupina rezistentných jedincov nikdy nebude mať veľkosť 1 keď $t \rightarrow \infty$.

Asymptotická hodnota $r(t)$ má dôležitú praktickú interpretáciu: je to celkový zlomok jedincov, ktorí sa nakazili ochorením počas celého priebehu epidémie.

Počiatkové podmienky pre model SIR môžeme vybrať rôznymi spôsobmi, ale najbežnejší je predpoklad, že ochorenie začína iba s jedným infikovaným jedincom alebo s malým počtom jedincov c . Inými slovami, počiatkové hodnoty premenných sú $s(0) = 1 - \frac{c}{N}$, $i(0) = \frac{c}{N}$ a $r(0) = 0$.

3 Reálne dáta

3.1 Chrípka

Chrípka je infekčné ochorenia, ktoré spôsobuje vírus chrípky. Symptómy môžu byť mierne až závažné. Medzi najčastejšie symptómy patria vysoká horúčka, bolesť v krku, bolesti hlavy, kašeľ, pocit únavy a svalové bolesti. Tieto symptómy sa zvyčajne začnú dva dni po vystavení vírusu a väčšinou trvajú menej ako týždeň.

Vírus sa zvyčajne šíri vzduchom kašľom alebo kýchnutím. Môže sa tiež šíriť dotknutím sa povrchov kontaminovaných vírusom a následným dotykom úst alebo očí. Jedinec môže byť infekčný voči ostatným jedincom tak pred, ako aj v čase, keď sa prejavujú príznaky ochorenia. Odporúča sa proti chrípke očkovať. Keďže sa vírus chrípky každý rok mení, je potrebné sa očkovať každoročne.

Chrípka sa opakuje každoročne, čo má každý rok za následok 3 až 5 miliónov ochorení a približne 250 000 až 500 000 úmrtí. V 20. storočí sa vyskytli tri pandémie chrípky: španielska chrípka v roku 1918 (približne 50 miliónov úmrtí), ázijská chrípka v roku 1957 (približne dva milióny úmrtí) a Hong-Kongská chrípka v roku 1968 (približne jeden milión úmrtí)[3].

3.2 Fitovanie reálnych dát na model SIR

Pre svoju prácu som si vybrala ochorenie chrípky a to v sezóne rokov 2016 - 2017 (40. týždeň roku 2016 až 18. týždeň roku 2017) v Žilinskom samosprávnom kraji. Údaje uvedené v tabuľke 1 som získala z epidemiologického informačného systému regionálneho úradu verejného zdravotníctva Slovenskej republiky [4].

Veľkosť sledovanej populácie (počet obyvateľov Žilinského samosprávneho kraja) je $N = 690\,778$. Tretí stĺpec nám vyjadruje počet novo nakazených jedincov. Je dôležité poukázať na to, že počet novo nakazených jedincov sa nerovná počtu infikovaných.

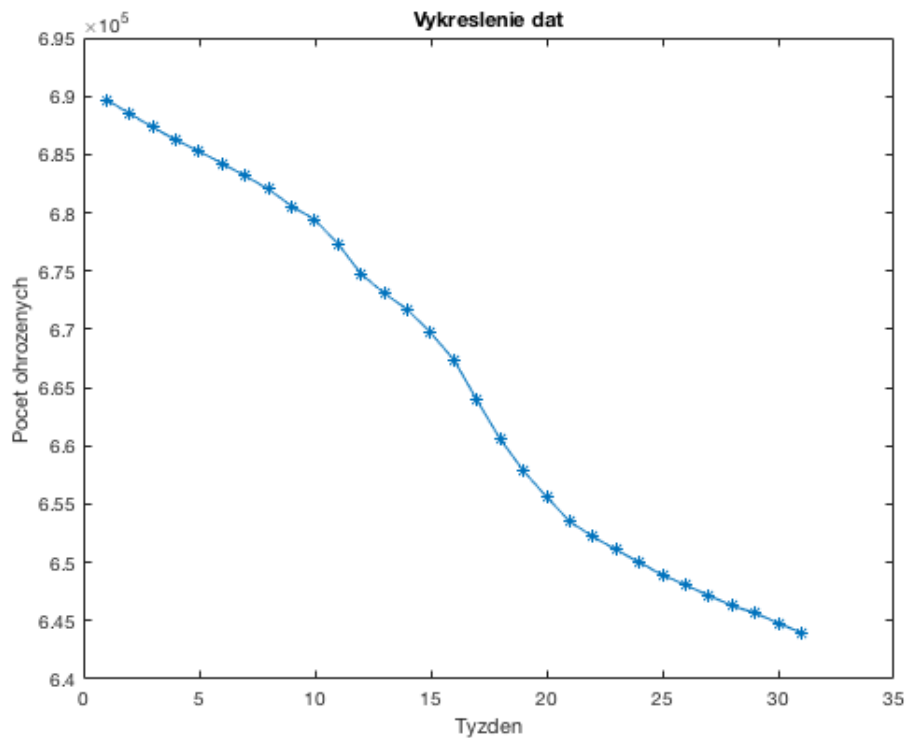
Nie je jasné, ako z počtu novo nakazených jedincov získať hodnoty $i(t)$ pretože nevieme, koľko z nakazených sa uzdraví. Vieme však, že počet ohrozených jedincov je definovaný rovnicou (2). Je teda jasné, že skupina ohrozených jedincov v čase t sa zníži o počet novo nakazených jedincov v čase t . Teda platí:

$$\begin{aligned}
S(0) &= 690778, \\
S(1) &= S(0) - 1081,76, \\
S(2) &= S(1) - 1183,99, \\
&\vdots \\
S(31) &= S(30) - 823,41.
\end{aligned}$$

Týmto spôsobom si vieme vypočítať hodnoty $S(t)$ a následne aj hodnoty $s(t)$. Tieto hodnoty sú uvedené v tabuľke 1.

Na obr. 6 sme si vykreslili vývoj počtu ohrozených jedincov v čase t . Aj v tomto prípade je vidno, že skupina ohrozených jedincov klesá monotónne (rovnica (2)).

Obr. 6: Vykreslenie dát



Tabuľka 1: Reálne dáta

Týžden chrípkovej sezóny	t	Počet novo nakazených	S(t)	s(t)
40	1	1081.76	689696.24	0.9984
41	2	1183.99	688512.25	0.9967
42	3	1192.97	687319.28	0.9949
43	4	1067.25	686252.03	0.9934
44	5	989.88	685262.15	0.9920
45	6	1041.69	684220.46	0.9905
46	7	1066.56	683153.9	0.9889
47	8	1141.86	682012.04	0.9873
48	9	1457.54	680554.5	0.9852
49	10	1156.32	679398.18	0.9835
50	11	2077.86	677320.32	0.9805
51	12	2597.33	674722.99	0.9767
52	13	1630.24	673092.75	0.9743
1	14	1427.15	671665.6	0.9723
2	15	1972.17	669693.43	0.9694
3	16	2311.34	667382.09	0.9661
4	17	3487.05	663895.04	0.9610
5	18	3319.88	660575.16	0.9562
6	19	2711.99	657863.17	0.9523
7	20	2247.10	655616.07	0.9490
8	21	2110.33	653505.74	0.9460
9	22	1286.92	652218.82	0.9441
10	23	1141.17	651077.65	0.9425
11	24	1058.96	650018.69	0.9409
12	25	1092.12	648926.57	0.9394
13	26	861.4	648065.17	0.9381
14	27	889.72	647175.45	0.9368
15	28	885.58	646289.87	0.9355
16	29	625.79	645664.08	0.9346
17	30	872.45	644791.63	0.9334
18	31	823.41	643968.22	0.9322

Ako sme si ukázali v časti 2.3.2, na vyriešenie modelu SIR potrebujeme vstupné hodnoty a hodnoty parametrov $\tilde{\alpha}$ a β .

Keďže vieme, že $r(1) = 0$, $N = 1$ a z tabuľky 1 poznáme $s(1)$, môžeme si pomocou rovnice (9) dopočítať aj $i(1)$. K vyriešeniu modelu však potrebujeme poznať aj hodnoty parametrov.

K nájdeniu parametrov sme zvolili postup, v ktorom si v programe MatLab najskôr zadefinujeme model SIR s globálnymi hodnotami parametrov $\tilde{\alpha}$ a β , následne pomocou funkcie *ode15s* a zadaných vstupných hodnôt vypočítame riešenie.

Ďalej použijeme funkciu nelineárnej regresie *nlinfit* na nájdenie hodnôt parametrov $\tilde{\alpha}$ a β a funkciu *ode15s* na vyriešenie modelu SIR s týmito nájdenými parametrami. Týmto postupom si nájdeme simulované dáta s rovnakými vstupnými hodnotami ako majú reálne dáta a zároveň nájdeme hľadané hodnoty parametrov $\tilde{\alpha}$ a β .

Na overenie správnosti postupu nájdenia hodnôt parametrov využijeme program MatLab. Zdefinujeme si riešenie modelu SIR pri parametroch $\tilde{\alpha} = 0.7$ a $\beta = 0.3$ a vstupných hodnotách $s(1) = 0.99844$, $i(1) = 0.00156$ a $r(1) = 0$.

```
function dy = F2(t,y)
alf = 0.7;
bet = 0.3;

dy = [0; 0; 0];

dy(1) = -alf * y(1) * y(2);
dy(2) = alf * y(1) * y(2) - bet * y(2);
dy(3) = bet * y(2);

options = odeset('RelTol', 1e-4, 'NonNegative', [1 2 3]);
[t,y] = ode15s('F2', [1:1:31], [0.99844 0.00156 0], options);
xlabel('Cas');
ylabel('Jedinci');
plot(t,y)
y
```

V tabuľke 2 sú uvedené nájdené hodnoty $s(t)$, $i(t)$ a $r(t)$.

Tabuľka 2: Tabuľka

Týžden	$s(t)$	$i(t)$	$r(t)$
1	0.9984	0.0015	0
2	0.9970	0.0023	0.0005
3	0.9950	0.0034	0.0014
4	0.9921	0.0051	0.0027
5	0.9877	0.0076	0.0046
6	0.9813	0.0112	0.0074
7	0.9719	0.0165	0.0115
8	0.9583	0.0240	0.0175
9	0.9390	0.0346	0.0262
10	0.9120	0.0491	0.0387
11	0.8756	0.0680	0.0562
12	0.8282	0.0916	0.0800
13	0.7696	0.1188	0.1115
14	0.7011	0.1473	0.1514
15	0.6264	0.1738	0.1997
16	0.5503	0.1944	0.2551
17	0.4780	0.2063	0.3155
18	0.4132	0.2087	0.3780
19	0.3576	0.2023	0.4399
20	0.3117	0.1893	0.4988
21	0.2746	0.1721	0.5531
22	0.2450	0.1529	0.6019
23	0.2217	0.1333	0.6448
24	0.2033	0.1146	0.6820
25	0.1887	0.0974	0.7138
26	0.1773	0.0820	0.74067
27	0.1682	0.0685	0.7632
28	0.1610	0.0569	0.7819
29	0.1552	0.0471	0.7975
30	0.1506	0.0388	0.8104
31	0.1469	0.0319	0.8210

Výsledkom sú syntetické dáta, pomocu ktorých overíme správnosť fungovania algoritmu a nájdených hodnôt.

```

week = xlsread('jana-birova-data-2.xlsx', 'B2:B32');
susceptible = y(:,1);
options = statset('Display', 'iter');
format long;
[values] = nlinfit(week, susceptible, 'SIRsim2', [.00000001 .00000001],
global alf bet;
alf = values(1)
bet = values(2)
[t, x] = ode15s('SIRsim', [1:1:31], [0.99844 0.00156 0]);
plot(t, x(:,1), '-bs', week, susceptible, '-rs')
xlabel('Cas'),
ylabel('Pocet ohrozenych jedincov'),
legend('Simulovane', 'Realne')
title('Simulovane a realne data')

```

Pre úplnosť je potrebné uviesť, čo znamenajú jednotlivé príkazy použité v našom algoritme.

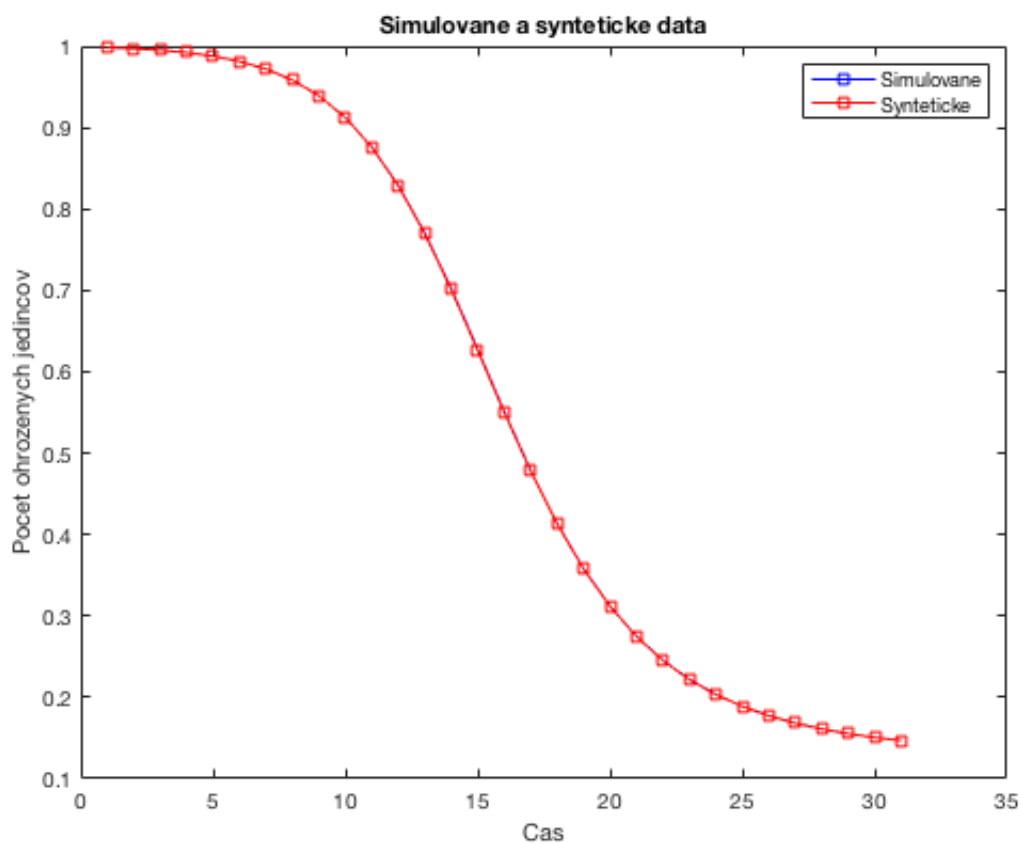
α a β nám reprezentujú parametre $\tilde{\alpha}$ a β , funkcia $F2$ nám predstavuje model SIR, teda $dy(1)$ je $s'(t)$, $dy(2)$ je $i'(t)$ a $dy(3)$ je $r'(t)$. *options* reprezentujú naše nastavenia a to *RelTol*, čo je relatívna chyba a *NonNegative*, čo je podmienka nezápornosti. Na vyriešenie použijeme riešič *ode15s* so zadanými vstupnými hodnotami, funkciou $F2$, časovým intervalom t (1:31), ktorý nám reprezentuje sezónu rokov 2016-2017 a počiatkové podmienky pre $s(1)$, $i(1)$ a $r(1)$ a s našimi nastavenými možnosťami.

V časti 2.3.1 sme na vyriešenie modelu SIR použili riešič *ode45*, teraz sme však zvolili riešič *ode15s* pretože sa ho odporúča použiť ak nevieme, či optimizér "nedotlačí" systém niekde, kde už je to zle podmienené. Riešič *ode15s* je založený na numerických diferenciálnych vzorcoch. Prípadne sa môžu použiť vzorce spätnej diferenciácie (BDFs, tiež známe ako Gearova metóda), ktoré sú ale zvyčajne menej účinné [13].

Ďalej sme si načítali dáta, kde *week* predstavuje sledované týždne, *susceptible* sú hodnoty $s(t)$ vypočítané v predchádzajúcom kroku. V možnostiach sme si nastavili zobrazenie iterácii. Na nájdenie hodnôt parametrov *alf* a *bet* sme použili funkciu *nlinfit* so zadanými vstupnými hodnotami, funkciou *SIRsim2*, ktorá vyrieši model SIR s globálnymi *alf* a *bet* a zvolenými počiatkovými hodnotami pre hľadané parametre. *nlinfit* je funkcia nelineárnej regresie, ktorá využíva Levenberg-Marquardt nelineárny algoritmus najmenších štvorcov [14].

Nakoniec sme opäť použili riešič *ode15s* na vyriešenie modelu s rovnakými vstupnými hodnotami s tým rozdielom, že model SIR je definovaný funkciou *SIRsim*, v ktorej máme zadané globálne *alf* a *bet*. Výsledok si vykreslíme.

Algoritmus určil hodnoty $\tilde{\alpha} = 0.699$ a $\beta = 0.3000$, na obr. 7 máme vykreslené simulované a syntetické dáta. Vidíme, že náš algoritmus funguje správne, pretože simulované a syntetické dáta sa zhodujú a rovnako sa zhodujú aj hodnoty parametrov $\tilde{\alpha}$ a β .



Obr. 7: Simulované a syntetické dáta

Keďže algoritmus funguje, tak rovnaký postup aplikujeme na naše dáta chrípky. V tomto prípade máme zadaný model SIR s globálnymi hodnotami alf a bet a jeho riešenie so zadanými vstupnými hodnotami. Ďalej si načítame zo súboru *jana – birova – data – 2.xlsx* jednotlivé sledované týždne (*week*) a vypočítané hodnoty ohrozených jedincov v týchto týždňoch (*susceptible*). Pomocou funkcie *nlinfit* nájdeme hodnoty parametrov alf a bet . So vstupnými hodnotami z našich dát a s nájdenými parametrami alf a bet vypočítame hodnoty $s(t)$, $i(t)$ a $r(t)$ a výsledok si vykreslíme.

```
function dy = F(t,y)
global alf bet;

dy = [0; 0; 0];

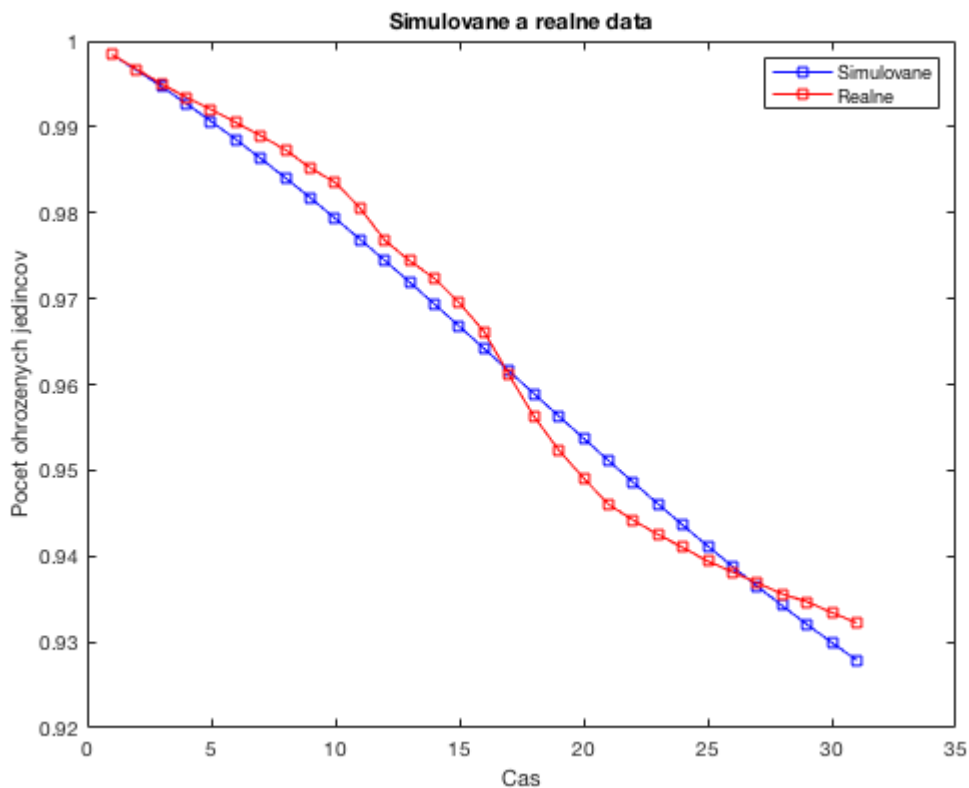
dy(1) = -alf * y(1) * y(2);
dy(2) = alf * y(1) * y(2) - bet * y(2);
dy(3) = bet * y(2);

function x = susceptible(a, y)
global alf bet;
alf = a(1);
bet = a(2);
[t, x] = ode15s('SIRsim', [1:1:31], [0.9984339976, 0.0015660024, 0]);
x = interp1 (t, x(:,1), y);

week = xlsread('jana-birova-data-2.xlsx', 'B2:B32');
susceptible = xlsread('jana-birova-data-2.xlsx', 'E2:E32');
options = statset('Display', 'iter');
format long;
[values] = nlinfit(week, susceptible, 'SIRsim2', [.00000001 .00000001],
options);
global alf bet;
alf = values(1)
bet = values(2)
[t, x] = ode15s('SIRsim', [1:1:31], [0.9984339976 0.0015660024 0]);
plot(t, x(:,1), '-bs', week, susceptible, '-rs')
xlabel('Cas'),
ylabel('Pocet ohrozenych jedincov'),
legend('Simulovane', 'Realne')
title('Simulovane a realne data')
```

Hľadané hodnoty $\tilde{\alpha}$ a β sú:

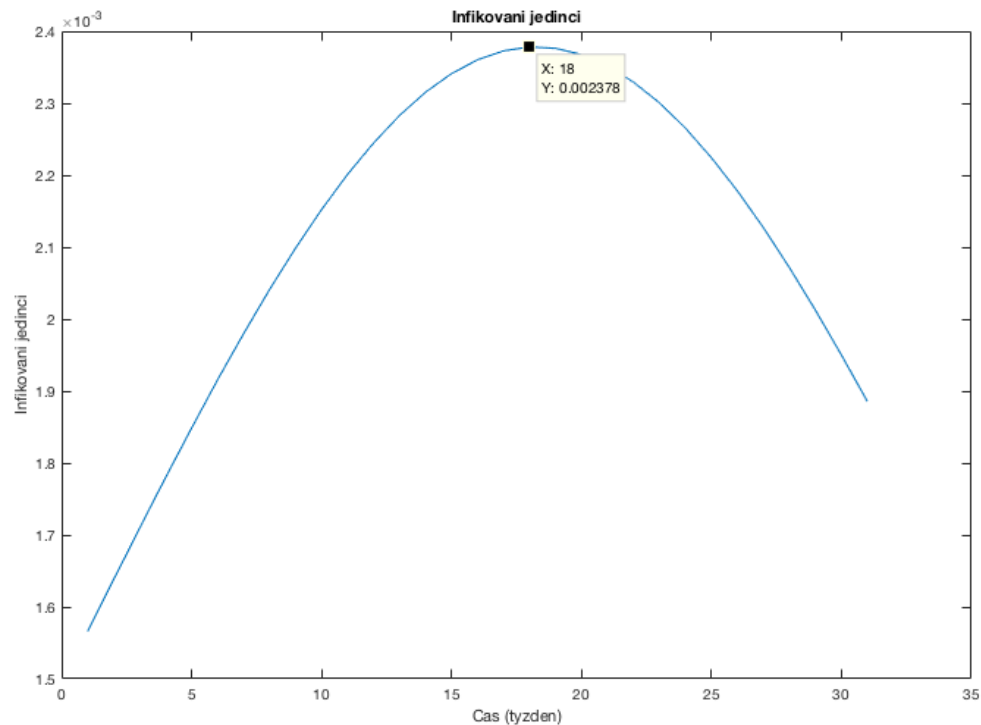
$$\tilde{\alpha} = 1.146 \text{ a } \beta = 1.098$$



Obr. 8: Simulované a reálne dáta

Na obr. 8 máme vykreslené simulované a reálne dáta, vidíme, že naše reálne dáta sú blízke simulovanému modelu.

Podľa predchádzajúceho môžeme povedať, že pomocou modelu SIR, aj keď poznáme len veľkosť sledovanej populácie, sledované časové obdobie a počet novo nakazených jedincov v tomto období, dokážeme vypočítať hodnoty parametrov $\tilde{\alpha}$ a β ako aj počet ohrozených, infikovaných a rezistentných jedincov v čase t . Rovnako vieme určiť, či sa sledované ochorenie bude alebo nebude šíriť, či prepukne v epidémiu, a teda určiť priebeh šírenia sledovaného ochorenia.



Obr. 9: Infikovaní jedinci

Na obr. 9 máme zobrazený priebeh skupiny infikovaných jedincov v sledovanom časovom období. Priebeh krivky mal typický charakter, najskôr počet infikovaných rásol, až do 18. sledovaného týždňa (5. týždeň roku 2017), kedy počet infikovaných dosiahol vrchol a následne počet infikovaných klesal.

3.3 Výhody a nevýhody modelu SIR

Ako každý model aj model SIR má svoje výhody a nevýhody. Medzi výhody môžeme zaradiť jeho jednoduchosť. Dalej to, že nezohľadňuje množstvo faktorov ako pôrodnosť a úmrtnosť v sledovanej populácii, dĺžku ochorenia, vekovú štruktúru, inkubačnú dobu ochorenia. Rovnako nezohľadňuje, či proti sledovanému ochoreniu existuje očkovanie a v prípade ak áno, aký podiel sledovanej populácie je zaočkovaný. Medzi ďalšie výhody patrí neuvažovanie latentného obdobia a teda možnosť počítať priebeh ochorenia bez meškania, čo značne zjednodušuje prácu s modelom. Jednou z najväčších výhod modelu je jednoznačne to, že aj keď nepoznáme všetky vstupné hodnoty modelu ako

parametre $\tilde{\alpha}$ a β , ako aj množstvo infikovaných jedincov, vieme nasimulovať priebeh sledovaného ochorenia.

Medzi nevýhody patrí napríklad to, že tým, že nezohľadňuje množstvo vyššie spomenutých faktorov ovplyňujúcich šírenie ochorenia nemusí simulácia zodpovedať skutočnému stavu. A ako je už z predpokladov tohto modelu zrejmé, nemôže sa použiť na všetky ochorenia.

4 Záver

Cieľom mojej diplomovej práce bolo popísať a pochopiť SIR model šírenia epidémií, implementovať numerické riešenie v prostredí MatLab a pokúsiť sa nafitovať model na reálne dáta chrípkovej epidémie v Žilinskom samosprávnom kraji v rokoch 2016-2017.

V prvej časti som sa snažila popísať SIR model šírenia chorôb a načrtnúť numerické riešenie tohto modelu v programe MatLab. V druhej časti som sa pomocou získaných poznatkov pokúsila nafitovať model SIR na reálne dáta chrípkovej epidémie v Žilinskom samosprávnom kraji.

Pre mňa osobne bola táto práca veľkým prínosom, pretože mi rozšírila znalosti a prax v práci s programom MatLab a rovnako mi rozšírila moje znalosti v oblasti epidemiologických modelov a využitia matematiky pri ich riešení, pretože počas môjho doterajšieho štúdia som nemala možnosť sa s touto časťou matematiky stretnúť.

Diplomovú prácu som sa snažila napísať vecne, ale pre čitateľa zrozumiteľne. Cieľ práce považujem za splnený. Samozrejme, v mojej diplomovej práci som sa venovala len jednému druhu epidemiologických modelov a len jednému ochoreniu, avšak táto oblasť matematiky má oveľa širšie využitie. Dúfam však, že moja diplomová práca bude aspoň malým podnetom pre vzbudenie záujmu o danú oblasť matematiky.

Literatúra

- [1] M.E.J Newman, *Networks An Introduction*, 2010, Oxford
- [2] https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology#The_SIR_model
- [3] <https://en.wikipedia.org/wiki/Influenza>
- [4] <http://www.epis.sk/AktualnyVyskyt/ChripkaAARO/sezona-2016-2017.aspx>
- [5] https://www.ruvzlv.sk/?id_menu=29866
- [6] https://en.wikipedia.org/wiki/Mathematical_modelling_of_infectious_disease
- [7] <https://en.wikipedia.org/wiki/Epidemiology>
- [8] https://www.mathworks.com/help/matlab/ordinary-differential-equations.html?s_tid=srchbrcm
- [9] *Principles of Epidemiology in Public Health Practice*, 2012, Atlanta, Georgia: Centers for Disease Control and Prevention
- [10] Green MS; Swartz T; Mayshar E; Lev B; Leventhal A; Slater PE; Shemer Js, *When is an epidemic an epidemic?*, 2002, Isr. Med. Assoc. J.
- [11] Kermack WO, McKendrick AG, *A Contribution to the Mathematical Theory of Epidemics*, 1927, Proceedings of the Royal Society A.
- [12] <https://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/cumulative-cases-graphs.html>
- [13] <https://www.mathworks.com/help/matlab/ref/ode15s.html>
- [14] https://www.mathworks.com/help/stats/nlinfit.html?s_tid=srchtitlebtdijz-2
- [15] https://www.mathworks.com/help/matlab/ref/ode45.html?s_tid=srchtitle
- [16] H. W. Hethcote, *Qualitative analyses of communicable disease models*, Math. Biosci., 28 (1976), pp. 335–356.
- [17] http://www.epis.sk/getdoc/2f10efaa-faea-4571-8f8e-8c4e397922ad/V5_SR_2008.aspx