

**Георги Симеонов Симеонов**

**МОДЕЛИРАНЕ И АНАЛИЗ НА ГОЛЕМИ ДАННИ  
ЗА ЕПИДЕМИЯТА ОТ COVID-19**

**АВТОРЕФЕРАТ**

на дисертация за присъждане  
на образователна и научна степен "доктор"

Област на висше образование:

4. Природни науки, Математика и Информатика

Професионално направление:

4.6. Информатика и компютърни науки

Докторска програма: Информатика

Научен ръководител: проф. дмн Петър Станчев

Научен консултант: проф. дмн Огнян Кунчев

София, 2021 г.

Дисертационният труд съдържа 125 страници, от които 110 страници основен текст, 4 основни глави и библиография към всяка отделна глава общо 85 заглавия.

Дисертационният труд е обсъден и насочен за защита на заседание на разширено звено към секция "Софтуерни технологии и информационни системи" на ИМИ – БАН, назначено със заповед номер ....., проведено на .....

Защитата на дисертационния труд ще се състои на ..... от ..... часа в аудитория ..... на ИМИ-БАН на открито заседание на научно жури в състав:

1. ....
2. ....
3. ....
4. ....
5. ....

Материалите по защитата са на разположение на интересуващите се в библиотеката на ИМИ-БАН.

Автор:  
Георги Симеонов Симеонов

Заглавие:  
"Моделиране и анализ на Големи Данни за епидемията от COVID-19"

## **Увод и обща характеристика на дисертационния труд**

В исторически план математическото моделиране в епидемиологията се появява в трудовете на Бернули около 1766 г. [5], но доскоро то не беше общоприето като математически инструмент, който да се използва и прилага широко в политиките от областта на общественото здраве, а вместо това се разглеждаше като много тясна специализирана изследователска област достъпна само за приложни математици и теоретици биолози.

Нещата се промениха с появата на пандемията на HIV, когато математически модели бяха използвани за първи път за прогнозиране на бъдещото разпространение на епидемията, и при анализа на влиянието на промяната в поведението на населението върху честотата на случаите на заразени с HIV.

Истинският пробив в математическото моделиране като приложим инструмент в сферата на общественото здраве дойде обаче, поради опасенията, че вирусът на едра шарка може да бъде използван при умишлено заразяване, водещо до опустошителни огнища сред популации на съвременното общество, които са само частично имунизирани.

Настоящата Дисертация представлява принос към нововъзникващата област на Цифровата и Изчислителната Епидемиология. Последната може да бъде определена като моделиране на "Големи данни" на заразни болести, по-специално на COVID-19, и може да бъде определено като създаване и изучаване на нови математически модели за анализ на епидемични огнища, включително техните компютърни имплементации и симулации в рамките на концепцията за Големи данни.

### **Актуалност и значение на научни проблеми, разгледани в дисертацията**

Инфекциозните заболявания предизвикват милиони смъртни случаи всяка година и са основната причина за човешката смъртност. Добре познати са смъртоносните болести като пневмония, туберкулоза, диарийни заболявания, малария, морбили, ХИВ / СПИН, а наскоро и COVID-19.

Необходимостта от централизиран контрол на епидемичните огнища може да се счита за основен двигател за развитието на съвременната епидемиология, по-специално за развитието на новата област наречена **Цифрова и Изчислителна Епидемиология** (Digital and Computational Epidemiology).

През последното десетилетие благодарение на положените усилия в световен мащаб се ускорява развитието на глобална мрежа за наблюдение, предоставяща епидемиологични Големи данни, за борба с пандемията на възникващи и повторно възникващи инфекциозни заболявания. Тази тенденция се разви много силно напоследък поради пандемията Covid-19.

Учени от различни области, вариращи от медицина, вирусология, имунология, генетика, молекулярна биология до компютърни науки, статистиката и приложната математика се обединиха за бърза оценка на потенциално спешни ситуации.

За решаване на тези задачи, математическото моделиране играе централна роля в усилията, насочени към прогнозиране, оценка и контрол на потенциалните огнища. За да се разбере по-добре и да се моделира разпространението на заразните болести, трябва да се отчита влиянието на множество променливи, вариращи от нивото на микростоприемник, до взаимодействията между патоген и стоприемник. Също така, преобладаващите екологични, социални, икономически и демографски фактори по целия свят трябва да бъдат анализирани и задълбочено проучени.

Поради тази причина е важно да сме наясно с основните подходи, които се използват за наблюдение и моделиране на динамиката на инфекциозните болести, и с основните концепции, които определят тяхното прилагане и практика в областта на наличните Големи данни. Вижте [33], [34], [35].

Големият въпрос, определящ актуалността на темата е: Как може да се разработи политика за обществено здравеопазване срещу заплахи с патогени, които *не циркулират в момента*? Няма начин за провеждане на епидемиологични разследвания, а пък единствените налични данни в случая на едра шарка са отпреди ерата на ликвидирането (на шарката). Следователно, за да създадем политика, трябва да се комбинират знания от *исторически данни за едра шарка с данни за днешното общество*, и евентуалните намеси трябва да бъдат тествани въз основа на тази налична информация. Математическо моделиране предоставя гъвкав инструмент за решаване на проблема и беше използван за анализ на възможни стратегии за ваксинация и други интервенции. Именно математическото моделиране определя днешните парадигми на епидемиологията на инфекциозните болести. Ако използваме съвременни термини, за анализ на разпространението на едра шарка и планиране на политики за нейното ограничаване, епидемиолозите и математическите модели са решили епидемиологичния проблем, ка-

то са приложили подход който принадлежи към парадигмата за Големи данни и Машинното обучение - масово използване на исторически данни от различни държави, генериране на прогнози (или, прогнозни сценарии) от различни алтернативни модели и от "автоматизирано обучение" от историческите данни. С оглед на горните наблюдения, основните цели на настоящата работа е създаване на нови подходящи математически модели и тяхната имплементация, за прилагането им към Големи данни в съвременната Епидемиология, която може да се разглежда и като наука, основана на Големи Данни във всички аспекти на това понятие.<sup>1</sup> Подробно обсъждане на ролята на големите данни при решаването на проблеми на епидемиологията свързани с готовността на съвременните общества да контролират разпространението на пандемии е публикувано още през 2015 г. можете да прочетете тук [32]. На Фигура 1 се визуализира мястото на Големите данни в епидемиологията: разнообразие (variety), обем (volume), скорост (velocity) и достоверност (veracity).

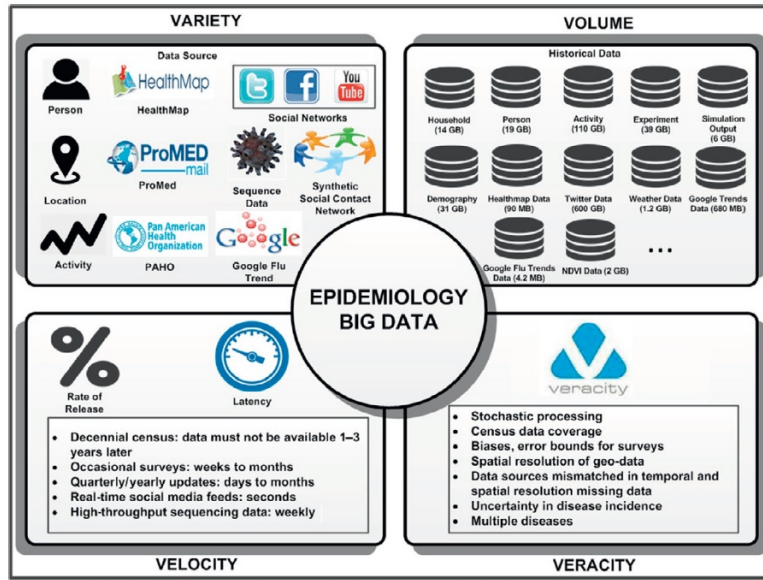
## Състояние на изследвания проблем

Една от основните цели на настоящата дисертация е създаването на модели за анализ на разпространението на болестта Covid-19, както и да представи тяхната софтуерна разработка, като веб-базиран онлайн инструмент за анализ в реално време с интерактивни визуализации. За да дадем представа за състоянието на изследвания проблем, ние предоставяме списък с най-популярните инструменти за анализ в световен мащаб на разпространението на COVID-19, предоставящи веб-базиран интерфейс на онлайн инструменти за анализи:

1. The tool "Projections" by the IHME Institute is available at the link <https://covid19.healthdata.org/bulgaria?view=total-deaths&tab=trend>
2. Tool Delphi from the Massachusetts Institute of Technology, <https://www.covidanalytics.io/home>
3. Tool at Imperial College London, <https://www.covidsim.org>
4. The Los Alamos National Laboratory (LANL), <https://covid-19.bsv.gateway.org/>

---

<sup>1</sup>За последните концепции на Отворени Данни и Големи данни, към които се отнасяме [30] и [31]



Фигура 1: Image courtesy: S.M. Shamimul Hasan, Virginia Tech.

- 
5. The SI-KJalpha model from the University of Southern California, <https://github.com/scc-usc/ReCOVER-COVID-19>

## Цели и задачи на настоящата работа

Основната цел на дисертацията е разработка на нови математически модели и представяне на техните софтуерни реализации под формата на уеб базирани инструменти. Задачите, които са решени за постигане на горната цел са:

1. Изграждане на модел за краткосрочна прогноза на епидемиологичните криви, модел TVBG-SEIR, базиран на сплайни с четири възли, представляващи коефициентите на "трансмисия (предаване)" и "премахване" на модела SEIR. Анализира се краткосрочната глобална еволюция на контролираните епидемии чрез въвеждане на мерки за ограничаване/отваряне от здравните власти. Включването на различни сценарии за прогнозиране с ограничаване (lockdown) варира във времето и позволява да се анализира не само първичната епидемична вълна, но и също възникващата вторична вълна и следващи вълни. Вижте Глава 2.
2. Разработка на уеб инструмент за изграждане на сценарии за COVID-19 (накратко SBT-COVID-19), базиран на софтуерните платформи Jupyter и Vokeh. Използва се софтуерната архитектура на Jupyter Notebooks с Vokeh. Основен принцип на работа е два програмни процеса, единият базиран на клиентска инфраструктура и другият на ядрото (Kernel). Инструментът може да се използва като софтуер за подпомагане на взимането на решения (Decision Support system - DSS) от (здравните) власти за изследване на различни сценарии и политики, като се контролира/променя мащабът на мерките за ограничения/задържане (домашна и социална изолация/карантина, ограничаване на пътувания и други), и се оценява тяхната ефективност. Инструментът SBT-COVID-19 позволява да се прецени колко дълго трябва да се поддържат мерките за ограничаване. Вижте глава 2 за описание на функционалностите на инструмента и глава 4 с описание на софтуера.
3. Изграждане на модел за дългосрочна прогноза, базиран на ATVBG-SEIR модела относно оценката на продължителността на епидемията от COVID-19 в една държава, отчитаща различни сценарии. Включихме ваксини в модела, които се извършват съгласно план

за ваксинация, предоставен на месечна база. Също така, ние моделирахме сезонния ефект. Алгоритъмът взема предвид основното ограничение на здравната система, което е броят на леглата за интензивно лечение (ICU), предназначени за пациенти с COVID-19. Вижте глава 3.

4. Разработване на уеб базиран инструмент за сценарии за мерки на ограничаване/отваряне, базиран на софтуерните платформи Jupyter и Vokeh; инструментът е достъпен онлайн на <http://atvbg-seir.eu>. Той използва алгоритъма, имплементиращ методологията на модела ATVBG-SEIR. Резултатите демонстрират ефикасността на инструмента, като го прилагат към данни от COVID-19 от Австрия, България, Германия, Италия, Великобритания и САЩ. Вижте глава 3 за описанието на функционалностите на инструмента и глава 4 за описание на софтуерното решение.



## Апробация на резултатите

### Списък на публикациите по дисертацията

1. Kounchev O., Simeonov G., Kuncheva Zh., 2021. *Estimation of the Duration of Covid-19 Epidemic in a Single Country, with or without Vaccinations. The Case of Bulgaria and Germany*, Comptes rendus de l'Academie bulgare des Sciences, Vol. 74, No. 5, pp. 677-686, DOI: 10.7546 / CRABS.2021.05.05
2. Kounchev O., Simeonov G., Kuncheva Zh., 2021. *Scenarios for the spread of COVID-19 analyzed by the TVBG-SEIR spline model*, Biomath 10 (2021), 2103087, <http://dx.doi.org/10.11145/j.biomath.2021.03.087>
3. Kounchev O., Simeonov G., Kuncheva Zh., *How Long 'Lockdowns' Are Needed to End the COVID-19 Epidemic in a Single Country, with or without Vaccinations*, (January 14, 2021). Available at SSRN: <https://ssrn.com/abstract=3766521> or <http://dx.doi.org/10.2139/ssrn.3766521>
4. Kounchev O., Simeonov G., Kuncheva Zh., 2020, *The TVBG-SEIR spline model for analysis of COVID-19 spread, and a Tool for prediction scenarios*, arXiv:2004.11338, <https://arxiv.org/pdf/2004.11338>
5. Stanchev, P., Ancheva, H., Pavlov, R., Simeonov, G., *The eleventh national information day: Open science, Open Data, Open Access, Bulgarian Open Science Cloud*, DiPP2020, 2020-September, pp. 275–281

### Доклади

1. Georgi Simeonov, *ATVBG-SEIR Scenarios Tool for Estimation of Covid-19 duration with and without vaccinations*, (2021), Пролетна научна сесия на Софийски университет "Св. Кл. Охридски"(ФМИ) – 27.03.2021 г., Секция: Covid-19 математически модели и прогнози, <https://www.fmi.uni-sofia.bg/bg/proletna-nauchna-sesiy-a-na-fmi-2021>
2. O. Kounchev and G. Simeonov, *How Long 'Lockdowns' Are Needed to End the COVID-19 Epidemic in a Single Country, with or without Vaccinations*, (2021), Интердисциплинарен семинар по биоматематика и научни изчисления – 18.02.2021 г., <https://math.bas.bg/?p=10124>, <https://researchseminars.org/seminar/BMNI>

3. Георги Сименов, *NI4OS-Еуроге – национални инициативи за отворена наука в Европа към Европейски облак за отворена наука (EOSC)*, Годишна отчетна сесия за 2020 г., секция СофТИС, ИМИ-БАН.

### Цитирания

1. Kyurkchiev, Nikolay, et al. "A new modifications of the SIR/SEIR models with" intervention polynomial factor". Methodological aspects."International Journal of Differential Equations and Applications 20.1 (2021).
2. Kyurkchiev, N., Kyurkchiev, V., Iliev, A., & Rahnev, A. ANOTHER LOOK AT SIR/SIRD/SEIR/GSEIR MODELS: NEW TRENDS, METHODOLOGICAL ASPECTS.
3. Kyurkchiev, Nikolay, Anton Iliev, and Asen Rahnev. A Look at the New Logistic Models with"Polynomial Variable Transfer". Alemania: LAP LAMBERT Academic Publishing, 2020. [Book]
4. Margenov, Svetozar, et al. "Mathematical and computer modeling of COVID-19 transmission dynamics in Bulgaria by time-depended inverse SEIR model."AIP Conference Proceedings. Vol. 2333. No. 1. AIP Publishing LLC, 2021.
5. Giffin, Andrew. "Methods for Causal Inference on Spatial Data with Environmental and Public Health Applications."(2020).
6. Georgiev, Slavi G., and Lubin G. Vulkov. "Coefficient Identification for SEIR Model and Economic Forecasting in the Propagation of COVID–19."BGSIAM'20 (2020): 14.
7. Ivanova, Miroslava, and Lilko Dospatliev. "DATA ANALYTICS AND SIR MODELING OF COVID-19 IN BULGARIA."International Journal of Applied Mathematics 33.6 (2020): 1099.
8. Loli Piccolomini, Elena, and Fabiana Zama. "Monitoring Italian COVID-19 spread by a forced SEIRD model."PloS one 15.8 (2020): e0237417.
9. Thakur, Lakshman, and Mikhail Bragin. "Data Interpolation by Optimal Splines with Free Knots Using Linear Programming."(2021).

Работата по дисертацията е частично финансирана от проектите КП-06-Н32-8 и КП-06Н42-2 с ФНИ, а също така и от проект с European Open

---

Science Cloud (EOSC Secretariat).

Разработени са алгоритми и модели с програми на MATLAB както и онлайн инструменти за краткосрочни и дългосрочни прогнози съответно TVBG-SEIR и ATVVG-SEIR (Python, Jupyter).

### **Съдържание на дисертационния труд**

Дисертацията се състои от четири глави и е в обем от 125 страници. Първата глава е въведение в тематиката на математическите модели с отделения за епидемиологията. Втората глава описва краткосрочен модел. Третата глава е посветена на дългосрочен модел с ваксиниране и сезонен ефект. Четвъртата глава описва софтуерните платформи. Работата съдържа 49 фигури.

### **Благодарности**

Дисертантът изказва своите искрени благодарности на научния си консултант проф. дмн Огнян Кунчев за многобройните дискусии по различни теми от дисертацията и получените безценни съвети и неоценима подкрепа и окуражаване по време на написването на дисертационния труд. Дисертантът дължи благодарности на научния си ръководител проф. Петър Станчев и колегите от секция СофТИС. Настоящата дисертация получи финансова подкрепа от КР-06-N32-8 и КР-06N42-2 на Българският фонд „Научни изследвания“ (ФНИ) и подкрепата с грант от European Open Science Cloud (EOSC) в рамките на секретариата на EOSC.

# Глава 1

## Въведение

Инфекциозните болести причиняват повече от 13 милиона смъртни случая годишно и са основна причина за човешката смъртност. Добре познатите смъртоносни заболявания са пневмония, туберкулоза, диарийните заболявания, малария, морбили, ХИВ/СПИН и др., а наскоро и COVID-19.

Епидемиологията е изследване на разпространението и определящите фактори на свързани със здравето състояния, или събития в определени популации, а също така и приложението от това проучване за превенция и контрол на здравословни проблеми, [1].

### 1.1 Приложения на Големи данни в епидемиологията

За разлика от медицинските науки, основната област на приложение на Епидемиологията е самото население като цяло. Работата на епидемиолозите в общественото здраве включва създаване на проучвания, оценка и тълкуване на данните за общественото здраве и поддържането на системи за събиране на *Големи данни*.

За да дадем представа колко големи са големите данни в обхвата на проблемите, с които се занимаваме, споменаваме следните факти за форматите и обема на дневните данни за Covid-19:

Данните, предоставени от повечето източници и Световната здравна организация (СЗО), преобладават **Временни серии** във формат CSV

Metrics	Source	Updated	Countries
Vaccinations	Official data collated by the Our World in Data team	Daily	217
Tests & positivity	Official data collated by the Our World in Data team	Weekly	136
Hospital & ICU	Official data collated by the Our World in Data team	Weekly	35
Confirmed cases	JHU CSSE COVID-19 Data	Daily	194
Confirmed deaths	JHU CSSE COVID-19 Data	Daily	194
Reproduction rate	Arroyo-Marioli F, Bullano F, Kucinskas S, Rondón-Moreno C	Daily	185
Policy responses	Oxford COVID-19 Government Response Tracker	Daily	186
Other variables of interest	International organizations (UN, World Bank, OECD, IHME...)	Fixed	240

Фигура 1.1: Table of Our World in Data

и са в необработен формат. Те обаче обикновено се трансформират в някаква реляционна база данни, но това обикновено се случва след известно време, а не веднага. Както и при други източници на данни основни използвани формати са CSV, JSON, като има варианти за записи по редове или колони на дневните данни.

Популярният Институт за здравни показатели и оценка (IHME), който поддържа популярния инструмент за прогнози, има подробно описание на техните набори от данни, съдържащи данни за всички страни по света: Изтеглянето на прогнози за COVID-19 е достъпно на връзката <http://www.healthdata.org/covid/data-downloads>.

Декомпресираният дневен набор от данни е около 600 MB, което означава, че размерът на тримесечен период (обикновено се използва за моделите TVBG-SEIR и ATVBG-SEIR) е около 54 GB данни. **Примерни характеристики на данни за COVID-19 от проекта "Our World in Data"**: Наборите от данни, предоставени от проекта, са достъпни в основните формати: CSV, XLSX, JSON. Пълният набор от данни за COVID-19 е съвкупност от данните за COVID-19 за отделните държави. Те го актуализират ежедневно през целия период на пандемията COVID-19. Той включва набор от данни със следните характеристики, предоставени на Фигура 1.1. Пълното описание на набор от данни „Our World in Data“ за COVID-19 е достъпно на <https://github.com/owid/covid-19-data/tree/master/public/data>

Много от промените, които виждаме около нас са до голяма степен антропогенни и се случват в мащаб, по-широк и по-бърз от всякога в човешката история. Освен това регулярно се появяват **нови патогени**, което повишава значението и необходимостта за обществата да **разбират** и бъдат готови да решават **систематично** предизвикателствата на **нововъзникващи заразни болести** на различни нива. По-специално е необходимо да се разработват подходящи **Математически модели** за разпространението на новите инфекциозни заболявания, които са в състояние да се справят с именно Големите данни произтичащи от потенциалния пандемичен характер на разпространението на тези болести.

**Математическите модели** са ключови за разбирането на разпространението на епидемиите. В настоящата работа разглеждаме новоразработени математически модели за изучаване на сложната динамика на епидемии в широкомащабни популации, базирани на класическите SIR/SEIR модели в епидемиологията.

Използването на тези модели за създаване на прогнози за сценарии (наричани понякога в наши дни *projections*) за разпространението и развитието на епидемията и на здравните политики, води до проблеми, които са типични за приложенията на Големите данни. По този начин нашите методи са приложими за Големи данни в епидемиологията.

Важността на концепцията за Големи данни в епидемиологията може да се докаже чрез примери от историята. Типичен пример е обяснението на феномена на епидемиите на простудните заболявания в Индия достигащи връх извън традиционните сезони (несъответстващи на сезона). В частност така е и със случая на COVID-19, който показва, че за да даде разумно обяснение, епидемиологията трябва да бъде наука основана на Големи данни. Наистина, за да се даде научно аргументирано обяснение за хода на епидемията с COVID-19 в Индия, трябва да се извърши екстраполиране от данните за грипния сезон през предходни години, тъй като пиковият сезон на COVID-19 е през лятото (!), а не през зимата. По подобен начин, поради липсата на данни, е подходящо да се моделира "ниският (low) сезон" на COVID-19 в България въз основа на данни за грипния "нисък сезон" в други държави.

## 1.2 Математически модел с отделения на SIR / SEIR в епидемиологията

Основният фокус на настоящата работа са моделите, които принадлежат на семейството на така наречените (Compartmental) модели с "отделения". Преди да предоставим съвременното състояние на тези модели, ще представим кратко изложение на концепцията за модели с отделения, както и основни понятия и обозначения, които ще бъдат използвани по-късно.

За по-подробно кратко въведение в моделите с отделения, се позоваваме на монографията на Кийлинг-Рохани, [4] и на [5].

Класическият SEIR модел разделя обществото на индивидуумите разделено на четири отделения:

1. отделение  $C_S$  : с размер  $Sr(t)$ , съдържа "уязвими / податливи" на зараза хора в момента  $t$ .
2. отделение  $C_E$  : с размер  $Er(t)$ , което е брой "експонирани / изложени" хора в момент от време  $t$  - това са хората, които са влезли в контакт с разпространители на вируси и са "вирусоносители но не са "вирусо-разпространители"; вирусът е в скрита форма и обикновено те не показват симптоми на болест.
3. отделение  $C_I$  : с размер  $Ir(t)$ , което е брой инфекциозни случаи в момент от време  $t$  - това са хората, които са "разпространители на вируси повечето от тях проявяват някои симптоми, въпреки че може да не се проявяват и никакви симптоми (безсимптомни).
4. отделение  $C_R$  : с размер  $Rr(t)$  – числото включва броя на оздравели, поставени под карантина, с имунитет и починали лица, които ние наричаме "**отстранени**". Обикновено те идват от отделението  $C_I$  след което те или оздравяват или са починали, като и в двата случая не могат повече да бъдат разпространители на вируси. Официално тези данни се предоставят като кумулативни (събирателни) данни.

Най-опростените модели с отделения са моделите базирани на системата SIR (с липсващо  $C_E$  отделение) както и системата SEIR, които се основават на някои опростени предположения:

1. Всички в популацията с размер  $N$  са податливи на зараза.

2. Предполагаме, че популацията с размер  $N$  е перфектно и равномерно смесена.
3. Предполага се, че всеки, който се зарази, има същия капацитет на предаване на заразата, който е обобщен от коефициента на предаване  $\beta$  на моделите.
4. Коефициентът  $\beta$  който ще бъде разгледан по-долу, включва също интензивността на контактите в популацията, която също се приема, че е еднаква за всички в популацията.

Моделите от типа на SIR / SEIR позволяват просто прилагане на здравни политики и мерки, например, когато големи части от населението започват да работят от вкъщи (дистанционно). За разлика от SIR / SEIR моделите, стохастичните модели и моделите базирани на агенти изискват да се включи огромно количество информация, проследяване на всеки индивид поотделно, като опитът показва, че в много случаи крайният ефект от анализа ще бъде същият.

**Remark 1.** Поради липсата на достатъчно данни за епидемията от COVID-19, по-специално за моделиране на така наречените сезонни ефекти е много важно това, че грипните епидемии могат да се използват по аналогия за моделиране на COVID-19 епидемията, като се има предвид, че те са респираторни заболявания със сходни начини на предаване. Данните, директно сравняващи двете заболявания, обаче са оскъдни, [2].

### 1.2.1 Допълнителни предположения и характеристики на модели с отделения, в частност SIR / SEIR моделите

1. Нека да отбележим, че в полза на избора ни на SEIR модела говори фактът, че екипът на Imperial College, който има много дълъг опит в приложение на математически модели в епидемиологията, премина от модел базиран на агенти към модела SEIR през март 2020, поради своята простота и фактът, че и двата типа модели "дават като цяло подобни числа [3].
2. Важна характеристика на алгоритмите, която разглеждаме, е параметър, който е процентът на тежките случаи (хоспитализирани), които се нуждаят от Отделения за интензивно лечение (ICU).



3. Друга важна характеристика са неотчетените случаи. В Германия тези случаи се наричат *Тъмни числа*. Примерът с Китай показва, че в много представителна извадка от 9500 лица, процентът на тези, които са били болни от COVID-19 и не са показали симптоми са 80%. Следователно, груба оценка оказва, че трябва да умножим броя на официално заразените данни с коефициента 5. От друга страна, за намиране на Тъмните числа, знаменитият Институтът "Робърт Кох" в Берлин използва коефициент между 4 и 6 cf.

### 1.3 Налични официални данни за целите на моделирането

На практика нямаме "данни за реалността"  $Sr(t)$ ,  $Er(t)$ ,  $Ir(t)$ ,  $Rr(t)$ , които съответстват на описаните отдели на SEIR модела в раздел 1.2.

1. Официалните (наблюдавани) данни  $Idata(t)$ , които се представят *ежедневно* за "нови заразени случаи" с COVID-19, обикновено са за хора със сериозни симптоми. Обозначаваме с  $TotalInf(t)$  общия брой отчетени инфекциозни случаи, което е кумулативна сума от горните дневни данни  $Idata(t)$ .
2. Официално обявените данни за оздравели,  $Rdata(t)$ , съдържат кумулативния брой на случаите на оздравяване.
3. Данни за общия брой смъртни случаи поради Covid-19 - точното диагностициране на тези случаи, дали наистина като причина са починали от инфекцията с Covid-19, не е добре и строго дефинирано. Ние обозначаваме тези данни с  $Ddata(t)$ , което представлява кумулативният брой смъртни случаи до датата  $t$ .
4. Въпреки че не от самото начало на епидемията, властите предоставят в много държави данните за случаите в отделенията за интензивно лечение. Броят на тежките случаи, настанени в отделенията за интензивно лечение (ICU) изглежда най-лесно да се категоризира от всички изброени данни. Ние обозначаваме в деня  $t$  тези данни с  $ICU(t)$ .

### 1.3.1 Съвременни модели с "проекции"и аналитични инструменти

Ние предлагаме списък с някои от най-популярните инструменти в целия свят за анализ на разпространението на COVID-19, чрез уеб-базирани (безплатни) онлайн Аналитични инструменти.

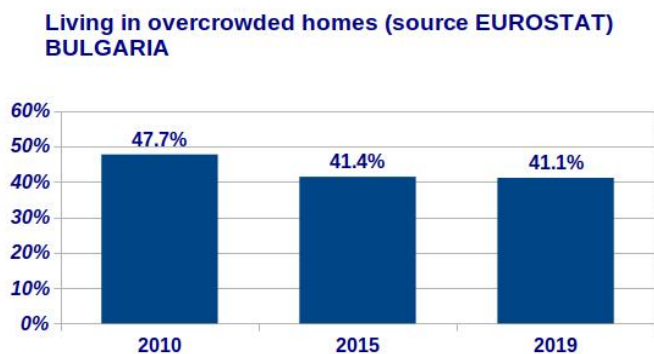
1. Инструментът "*Projection*" от Институт за здравни показатели и оценка (IHME) към университета във Вашингтон в Сиатъл <https://covid19.healthdata.org/bulgaria?view=total-deathstab=trend>. Вграден в уебсайта на Worldometers.info
2. Инструмент "*Delphi*" от Масачузетския технологичен институт, <https://www.covidanalytics.io/home>
3. Инструмент от Имперския колеж в Лондон, <https://www.covidsim.org>
4. Национална лаборатория в Лос Аламос (LANL), <https://covid-19.bsvgateway.org/>
5. моделът SI-KJalpha от Университета на Южна Калифорния, <https://github.com/scc-usc/ReCOVER-COVID-19>

## 1.4 Мултидисциплинарност на епидемиологичните модели за COVID-19

Мултидисциплинарността е много важна за успеха за справяне в днешно време с предизвикателствата в съвременната епидемиология. Ако цитираме на високо уважавания източник, Nature, [6], "Теоретиците и експериментаторите трябва да обединят сили".

1. Преди всичко **демографските** фактори са играли и играят важна роля при решаването на реалните проблеми на епидемиологията – по-специално за разбирането на детайлите за разпространението на COVID-19 в България. Трябва да се вземе в предвид, че в България 41 % от населението живее **три поколения** под един покрив заедно. Подобна е ситуацията и в други Средиземноморски страни, като Гърция, Италия, Испания. За настоящото състо-

яние на условията на живот в България и други страни, вижте Фигура 1.2<sup>1</sup>



Фигура 1.2: Процент население в България, живеещо в пренаселени домове.

2. В случая на пандемията на COVID-19 това означава, че училищните ваканции играят важна роля за регулирането на разпространението на епидемията. Картината е напълно различна в страните, където не повече от две поколения живеят заедно – Германия, Великобритания, Франция, Швейцария, скандинавските страни и др. В тези страни затварянето на училищата не играе толкова голяма роля за ограничаване на разпространението на COVID-19. Следователно, ефективността на Мерките за ограничаване силно зависят от така наречения "индекс на комуникативност" в рамките на нацията.
3. Условията **климат/време** са важни за така наречения "сезонен ефект като типичен пример е Индия.

<sup>1</sup>Фигурата е на разположение в публикацията на националния телевизионен канал БТВ в следната връзка <https://btvnovinite.bg/bulgaria/pod-edin-pokriv-pokoleni-ja-balgari-zhivejat-zaedno.html>.

## **1.5 Сезонен ефект на разпространението на Covid-19 - голямото предизвикателство**

Едно от основните предизвикателства за настоящото изследване беше създаването на прогнозни модели за сезонния ефект при разпространението на COVID-19 през лятото на 2021 и 2022. Основният проблем с моделиране на сезонния ефект в България е липсата на "чисти" данни за лятото на 2020 - огромен приток на български гастарбайтери напълно измени/развали кривата на инфекциозните случаи с COVID-19.

## Глава 2

# TVBG-SEIR – краткосрочен SEIR модел, базиран на сплайни и инструмент за сценарии за прогнозиране на разпространението на COVID-19

В настоящата глава разработваме нов модел TVBG-SEIR, базиран на сплайни, за анализ на коронавирусната инфекция (COVID-19), насочен към анализ на краткосрочното глобално развитие на епидемиите, контролирани от въвеждането на мерки за Локдаун / Разхлабване от властите. Включването на различни сценарии за предсказване, които се основават на мерки с Локдаун / Разхлабване, вариращи във времето, позволява да се анализира не само първичната епидемична вълна, но и възникващата вторична вълна, а също и всякакви следващи вълни. Нека отбележим, че това, което наричаме "сценарии за прогнозиране" в наши дни много често е наричано "проекции".

Моделът е реализиран чрез уеб-базиран *Инструмент за изграждане*

на сценарии за COVID-19 (наричан за краткост SBT-COVID-19), който може да се използва като софтуер за подпомагане вземането на решения за (здравни) политики (мерки) за проучване различни сценарии, чрез контрол / промяна на мащаба на мерките за ограничаване (домашна и социална изолация / карантина, ограничения за пътуване и други) и за оценка на тяхната ефективност. По-специално инструментът SBT-COVID-19 позволява да се оцени колко дълго трябва да се поддържат мерките за Локдаун.

## 2.1 Въведение

### 2.1.1 Контекст

Настоящият модел е построен с основна цел да осигури краткосрочно планиране на мерките за ограничаване. По-специално, от съществено значение е в краткосрочен план да се оцени как скъпите, ресурсоемки мерки, прилагани от властите, като домашна и социална изолация / карантина, ограничения за пътуване и т.н., могат да допринесат за предотвратяването и контрола на инфекцията с COVID -19 и колко дълго трябва да се поддържат.

Класическите модели SIR / SEIR обаче са изследвани предимно в случая, когато основните параметри като *скорост на предаване*  $\beta$  (отразява разпространението на вируса от заразени индивиди) и *скорост на отстраняване (премахване)*  $\gamma$  (отразяващи хоспитализацията / изолационните мерки) - остават постоянни през целия период на епидемично изследване.

### 2.1.2 Обобщение на целите и методите

Важно е да се разширят класическите SIR / SEIR модели, като се създадат нови модели за динамиката на скоростта на предаване  $\beta(t)$  (понякога наричани **Бета**) и скорост на отстраняване  $\gamma(t)$  (понякога наричано **Гама**). Основната цел на настоящото изследване е да се въведе нов базиран на сплайн модел SEIR с променливи във времето  $\beta(t)$ ,  $\gamma(t)$  параметри или за краткост наричан **TVBG-SEIR модел**, който се използва за оценка на практическите последици от обществените здравни интервенции и мерки за ограничаване. Създадохме **Инструмент за генериране на сценарии за COVID-19 (SBT-COVID-19 Tool)**, базиран на

TVBG-SEIR модела, който може да се използва като инструмент за подпомагане при вземането на решения от здравните власти за определяне на мерки и политики и създаване на *прогнозни сценарии* или ("проекции"). Може да се използва за оценка на въздействието на предишни интервенции в областта на общественото здраве и за планиране на количествено и качествено въвеждане на бъдещи мерки за ограничаване за постигане на необходимите цели.

За формулирането на нашия модел използваме формулировка (анзац) с детерминистични сплайни: скоростите на предаване  $\beta(t)$  и скоростите на премахване  $\gamma(t)$  са моделирани от сплайни с два възела – *Node1*, *Node2* (едни и същи възли за  $\beta(t)$  и за  $\gamma(t)$ ) – в рамките на интервала от време – от *StartDate* до *TodayDate*. Този клас сплайни позволява правилно да се моделира динамиката, поради въвеждането на мерки за ограничаване от властите в **две стъпки**. Целта на напасването на модела TVBG-SEIR към данните, е да се идентифицират възлите на сплайните и трите стойности на  $\beta(t)$  и  $\gamma(t)$  на интервалите [*StartDate*, *Node1*], [*Node1*, *Node2*] и [*Node2*, *TodayDate*]. Предполага се, че  $\beta(t)$  и  $\gamma(t)$  са **константа** във времеви интервал [*Startdate*, *Node1*], и  $\beta(t)$  е монотонно **намаляваща**, докато  $\gamma(t)$  е монотонно **нарастваща** функция.

Изборът на само два възела на сплайните за  $\beta(t)$  и  $\gamma(t)$  изглежда подходящ за модели на исторически данни (до *Today*), но не и за много дълги периоди от време. Тези модели се използват като основа за създаване на "сценарии за прогнозиране" или ("проекции") започвайки от *Today* с прогнозна перспектива, хоризонт от около два месеца (**2m**) време. Тези модели се актуализират всеки ден (с пристигането на официалните ежедневни данни) и сценариите ("проекциите") се обновяват съответно. В един по-математически език "предсказващ сценарий" (наречан много често днес "**проекция**"<sup>1</sup>) означава избор на контролни параметри  $\beta(t)$  и  $\gamma(t)$  под формата на сплайни, определени след *Today*, които определят SEIR модела, който ще бъде детайлно дефиниран по-долу. Човек може избере тези сценарии по безкрайно много начини.

Уеб базираният инструмент SBT-COVID-19 е проектиран за визуализация на резултатите от напасването (fitting) на модел към данните за ежедневно заразените случаи и за създаване на прогнозни сценарии

<sup>1</sup>Проекцията е процес на придвижване напред във времето чрез представянето на бъдещи събития или посредством оценки, базирани на определени предположения или минали тенденции, вижте <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Projection>,

Терминът "проекция" очевидно е синоним на "сценарий за прогнозиране" и ние ще използваме и двете, особено ако е необходимо сравнение със скорошната литература.

("проекции") за ежедневно заразените случаи през хоризонта от следващите два месеца, чрез контролиране на бъдещите стойности на коефициентите  $\beta(t)$  и  $\gamma(t)$ . Този Инструмент е подробно описан във връзките:

Version 3 (with Matlab generated Figures) [http://213.191.194.141:8888/notebooks/TVBG-SEIR-Spline-model\\_v3.ipynb](http://213.191.194.141:8888/notebooks/TVBG-SEIR-Spline-model_v3.ipynb)

Version 4 (with Python generated Figures) [http://213.191.194.141:8889/notebooks/TVBG-SEIR-Spline-model\\_v4.ipynb](http://213.191.194.141:8889/notebooks/TVBG-SEIR-Spline-model_v4.ipynb)

## 2.2 Класически детерминистичен SEIR модел: понятия и означения

### 2.2.1 Определяне на класическия непрекъснат SEIR модел

Основната цел на разработването на детерминистичния модел с отделения SEIR е да се осигурят някои проследими приближения  $S(t), E(t), I(t), R(t)$  към времеви ред на "данните от реалността"  $Sr(t), Er(t), Ir(t), Rr(t)$  обяснени в раздел 1.2. Най-широко използван е моделът, базиран на система от обикновени диференциални уравнения с променливи (криви):  $S(t), E(t), I(t), R(t)$  които се дефинират, както следва:

$$S'(t) = -\frac{\beta(t) S(t) I(t)}{N} \quad (2.1)$$

$$E'(t) = \frac{\beta(t) S(t) I(t)}{N} - \sigma E(t) \quad (2.2)$$

$$I'(t) = \sigma E(t) - \gamma(t) I(t) \quad (2.3)$$

$$R'(t) = \gamma(t) I(t) \quad (2.4)$$

1. Тук  $\beta(t)I(t)/N$  изразява скоростта, при която новите индивиди (като дял от общия брой популация) са заразени от вече заразните индивиди  $I(t)$ , (вж. Кийлинг-Рохани, (2008), [8] стр.18). Тук и по-нататък  $\beta(t)$  се нарича скорост на предаване на инфекцията, която наричаме по-нататък просто *Бета*.
2. Както вече казахме, коефициентът  $\gamma(t)$  е скоростта на премахване; тя се определя от реципрочното на инфекциозния период, след който или човекът е *оздравял* (и вече не е заразен), или е *починал* (няма възможност да бъде повече заразен). Тук и по-нататък  $\gamma(t)$



се нарича премахване (отстраняване) или ще го наричаме просто *Гама*.

3. Коефициентът  $\sigma$  е латентната скорост или скоростта на "проявя на симптоми" (където  $\frac{1}{\sigma}$  е средната стойност на **инкубационен период**). В настоящият модел използваме постоянната стойност

$$\sigma = \frac{1}{5.9}$$

което представлява разумно приближение, както показват последните изследвания, [13]. (Нека да отбележим, че предишната стойност беше  $\frac{1}{5.9}$ , вижте [7].)

4. Кривата  $S(t)$  съответства на времевите серии от реални данни  $Sr(t)$ . Количеството  $\sigma E(t)$  е равно на ежедневните нови инфекциозни случаи  $Idata(t)$ . Кривата  $I(t)$  на SEIR модела обаче е равна на така наречените *активни случаи*, които се дефинират от уравнението:

$$AC(t) = Total\_Infected(t) - Rr(t)$$

Тук  $Total\_Infected(t)$  е **кумулятивната** сума от  $Idata(t)$  до датата  $t$ .

Обичайните приложения на SEIR модела са с постоянни  $\beta(t)$  и  $\gamma(t)$ . приемаме, първоначалните стойности да бъдат следните:  $S(0)$ ,  $E(0)$ ,  $I(0)$ , и  $R(0)$  са дадени и системата е решена за времето  $t \geq 0$ , където  $t$  е цяло число. Очевидно имаме уравнението:

$$S'(t) + E'(t) + I'(t) + R'(t) = 0 \quad \text{for } t \geq 0,$$

което предполага, че сумата

$$S(t) + E(t) + I(t) + R(t)$$

е константа за всеки  $t \geq 0$ . Предполага се, че има следното уравнение:

$$\begin{aligned} N &= S(0) + E(0) + I(0) + R(0) \\ &= S(t) + E(t) + I(t) + R(t) \end{aligned} \quad (2.5)$$

където  $N$  е общото население в страната  $XX$ .

Нека да напомним, че едно от най-важните свойства на SEIR модела е, че той осигурява директен начин за изразяване на основното репродуктивно число:

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \quad (2.6)$$

### 2.3 Дискретизиране на SEIR модела

На практика се използва дискретизация на непрекъснатия SEIR модел. В текущата работа използваме следната дискретизация на SEIR модела, който е получен от метода на Ойлер за апроксимация на решение с начална стойност (2.1)-(2.4):

$$S_{n+1} = S_n - \frac{\beta_n S_n I_n}{N} \quad (2.7)$$

$$E_{n+1} = E_n + \frac{\beta_n S_n I_n}{N} - \sigma E_n \quad (2.8)$$

$$I_{n+1} = I_n + \sigma E_n - \gamma_n I_n \quad (2.9)$$

$$R_{n+1} = R_n + \gamma_n I_n \quad (2.10)$$

Тук  $S_n$ ,  $E_n$ ,  $I_n$  и  $R_n$  са съответно стойностите на  $S(t)$ ,  $E(t)$ ,  $I(t)$  и  $R(t)$  в деня  $t = n$ , и началните стойности за деня  $n = 0$  са  $S_0$ ,  $E_0$ ,  $I_0$  и  $R_0$ . Горната система се решава итеративно за всички цели числа  $n \geq 0$ . Предполагаме, че размерът  $N$  на популацията остава непроменен (следователно не се вземат предвид раждаемост и смъртност). Както в непрекъснатия случай, се приема, че общата сума удовлетворява условието:

$$N = S_n + E_n + I_n + R_n \quad (2.11)$$

което прави едно от уравненията в (2.7) - (2.10) излишно.

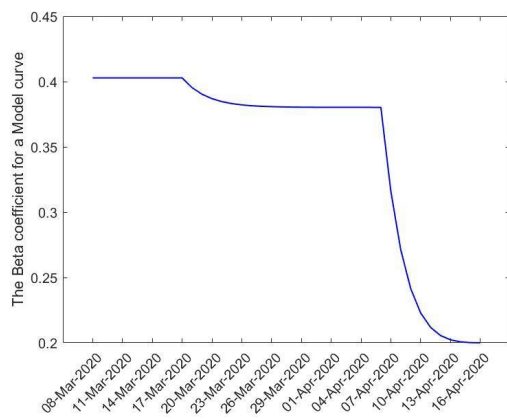
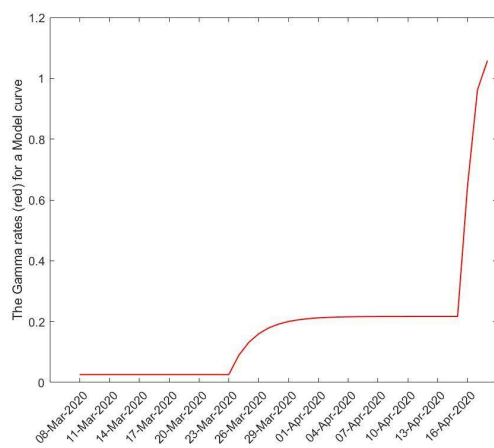
В деня  $n$  стойността  $R_n$  съответства на сумата на кумулативно възстановените плюс данните за смъртни случаи, т.е. на  $Rdata(n) + Deaths(n)$ . Обявените наблюдавани ежедневни данни за новозаразените  $Idata(n)$  съответстват на  $\sigma E_n$ , което е ясно от уравнението (2.9). По-долу използваме това съответствие, за да дефинираме квадратичната функция  $F(\Theta)$ , за напасване на нашите модели с най-малки квадрати.

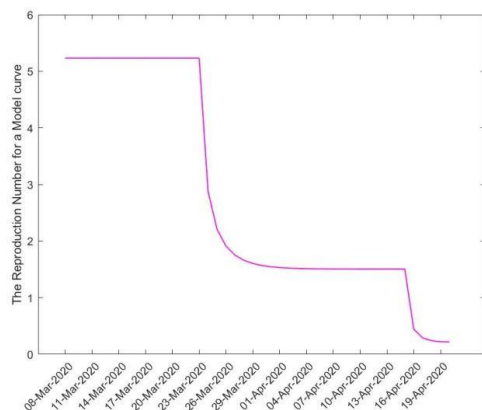
Трябва да направим важна забележка относно обозначенията, използвани в настоящия текст. Това беше ясно обяснено в [21], че в литературата авторите използват много често подвеждащи означения чрез смесване на четири различни постановки. По-специално, подвеждащо е използването на кривата  $I(t)$  на модела SEIR, която съответства на активните случаи в емпиричните данни.

## 2.4 Дефиниране на TVBG-SEIR модела

### 2.4.1 Техническо описание на TVBG-SEIR модела

1. Ние обозначаваме *StartDate* с  $T_1$ ; това съответства на датата, когато бъдат обявени първите случаи на COVID-19, в крайна сметка можем да изберем  $T_1$  да бъде дата, когато започва по-стръмният растеж на епидемията. Ние обозначаваме с  $T_4$  датата *EndDate* (обикновено се избира Today).
2. Избираме два вътрешни възела в интервала  $[T_1, T_4]$  за интерполационни сплайни, моделиращи коефициентите  $\beta(t), \gamma(t)$ :  $Node1 = T_2$  and  $Node2 = T_3$ . Това съответства на две стъпки от въвеждането на наложени ограничителни мерки от властите на страната *XX*. Обикновено датата  $T_2$  може да е Датата на първите ограничителни мерки или дата, близка до нея, и  $T_3$  може да е Дата на вторите ограничителни мерки или дата, близка до нея.
3. Моделът трябва да отразява естественото очакване, че след като има наложени официални ограничения, то те ще доведат до съществена промяна в скоростта на предаване и отстраняване, макар и не веднага. Предполагаме, че  $\beta(t)$  е монотонно намаляваща с времето, което съответства на естественото очакване, че колкото по-рестриктивни са мерките толкова по-малка е скоростта на предаване. Съответно скоростта  $\gamma(t)$  се приема, че е монотонно нарастваща, отговарящо на очакванията, че колкото по-силни са мерките, толкова по-голяма е скоростта на отстраняване.
4. Предполагаме, че  $\beta(t)$  и  $\gamma(t)$  са постоянни между началната дата  $T_1$  и първия възел  $T_2$ , т.е.  $\beta(T_1) = \beta(T_2)$  и  $\gamma(T_1) = \gamma(T_2)$ . Това съответства на спокойния живот на обществото (без мерки за ограничаване), когато  $\beta(t)$  и  $\gamma(t)$  са почти константни.
5. За да бъдем по-точни, сплайните, които използваме, не са обичайните полиномиални, а така наречаните експоненциални сплайни зависещи от параметър в експонентата, която прави бърз спад до следващата целева стойност на  $\beta(t)$  съответно това прави бързо увеличаване до целева стойност на  $\gamma(t)$ . На фигурите 2.1, 2.2 предоставяме примери за динамика на  $\beta(t)$  и  $\gamma(t)$ :
6. **Базово репродуктивно число**  $\mathcal{R}_0$  е ключова епидемиологична стойност за всички модели на епидемии, вижте [8], [9], [10], [11].

Фигура 2.1: моделна крива на  $\beta(t)$ .Фигура 2.2: моделна крива на  $\gamma(t)$ .



Фигура 2.3: Пример за променящо се във времето базово репродукционно число получено по формулата:  $\mathcal{R}_0 = \frac{\beta(t)}{\gamma(t)}$ .

Следвайки [11] (формула (2.4)), за случаите на SEIR модели с постоянни скорости  $\beta(t)$  and  $\gamma(t)$ , базовото репродуктивно число  $\mathcal{R}_0$  се определя от формулата:

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

където сме приели, че естествените нива на раждаемост и смъртност са малки и също равни. Поради горната забележка, можем да екстраполираме горната формула за всички интересни моменти от време, като положим:

$$\mathcal{R}_0 = \frac{\beta(t)}{\gamma(t)}$$

На фигура 2.3 представяме базовото репродуктивно число  $\mathcal{R}_0$ , получено по последната формула за някои специфични модели TVBG-SEIR.

7. Стойностите на  $\beta(t)$  и  $\gamma(t)$ , се определят като интерполационни сплайни на подинтервалите, определени от началната дата  $T_1$ , датите на възлите  $T_2$ ,  $T_3$ , и крайната дата  $T_4$ . Цялата конфигурация

се определя от общо осем параметъра, които са събрани в  $\Theta$ , зададени като:

$$\Theta = \{T_2, T_3, \beta(T_2), \beta(T_3), \beta(T_4), \gamma(T_2), \gamma(T_3), \gamma(T_4)\}$$

8. Данните, които използваме за напасването на дискретния модел TVBG-SEIR, са официалните данни за ежедневно новозаразени случаи  $Idata(t)$  (или техните кумулативни вектори  $cum(Idata)(t)$  и кумулативните данни за оздравели и фатални случаи.
9. Накрая, напасваме модела към данните, като оптимизираме позициите на двата възела  $T_2, T_3$ , и нивата на  $\beta(t)$  и  $\gamma(t)$ , т.е. чрез прилагане на минимизиране на най-малките квадрати. А именно, минимизираме следния квадратичен функционал  $F(\Theta)$  като променяме набора от параметри на  $\Theta$ :

$$F(\Theta) = \sum_{j=1}^n (cum(Idata)(t_j) - \sigma \cdot cum(E)(t_j))^2 + \\ + (Rdata(t_j) + Deaths(t) - R(t_j))^2$$

Тук означаваме с  $cum(E)(t)$  кумулативния вектор на решението  $E_n$  на дискретната SEIR система до датата  $t$ .

10. Минимизирането на  $F(\Theta)$  се извършва чрез пробно вземане на два възела  $T_2 < T_3$  на сплайни в интервал  $[T_1, T_4]$ ; стойности на интерполация за сплайните,

$$\beta(T_2), \beta(T_3), \beta(T_4), \gamma(T_2), \gamma(T_3), \gamma(T_4)$$

също се променят.

11. Както казахме, резултатът от минимизирането на квадратичния функционал  $F(\Theta)$  е набор от параметри  $\Theta$  и за намиране на минимума, може да има множество решения с близки стойности на функционала  $F(\Theta)$ . Когато намерим модел, базиран на параметрите  $\Theta$  преминаваме към създаване на сценарии за прогнозиране. Избираме някаква дата  $T_6$  която наричаме **Horizon**, нека да бъде най-много 2 (два) месеца от  $Today = T_4$ , и определяме  $T_6 = T_4 + 2 \text{ months}$

По-нататък следващ сценарий се определя, като се избере допълнителен възел  $T_5$ , който е *Трета ограничителна мярка дата* и параметрите

$$\{\beta(T_5), \beta(T_6), \gamma(T_5), \gamma(T_6)\}$$

Слагаме

$$\beta(T_5) = \text{Coef1} * \beta(T_4)$$

$$\beta(T_6) = \text{Coef11} * \beta(T_5)$$

$$\gamma(T_5) = \text{Coef2} * \gamma(T_4)$$

$$\gamma(T_6) = \text{Coef22} * \gamma(T_5)$$

Коефициентите Coef1, Coef11, Coef2, Coef22 се използват допълнително, за да контролират и представят създаването на нашия сценарий в инструмента SBT-COVID-19.

## 2.5 Анализ на разпространението на COVID-19 в България през октомври-декември 2020 г.

В настоящия раздел предлагаме приложение към българските данни за COVID-19 и сценарии за прогнозиране ("проекции"), генерирани от инструмента SBT-COVID-19. Ще използваме "сценарий за прогнозиране" вместо терминология "проекция".

В инструмента SBT-COVID-19 можете да намерите онлайн резултатите за анализ на данните за COVID-19 и за генериране на сценарии за България, за периода март-август 2020 г. (вижте [12]). Там показахме възможността за следваща вълна от инфекция.

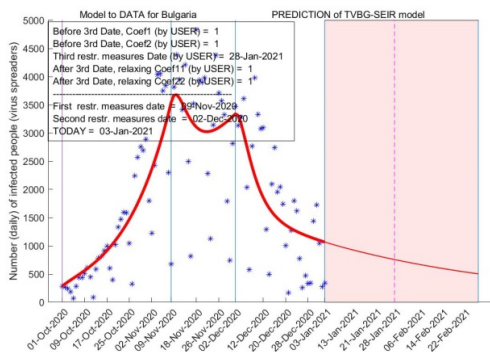
Тук ние демонстрираме как се *генерират сценарии за прогнозиране* ("проекции") въз основа на българските данни за периода 1 октомври 2020 г. - 3 януари 2021 г. В резултатите, предоставени по-долу, е ясно видимо, че моделът правилно отразява вълната на епидемията през октомври-ноември и нейния спад в края на декември поради ограничителни мерки на затваряне (локдаун), наложени на 25, ноември 2020. Също така наемква за появата на следваща вълна за *определени сценарии*, които съответстват на специални избори за сплайни за  $\beta(t)$ ,  $\gamma(t)$  след дата *Today*, която моделира отпускане на мерките за ограничаване.

**Definition 2.** *Под сценарий разбираме избор на коефициентите: Coef1, Coef2, които определят дали ограничителните мерки се отпускат (т.е. да изберем 0.2, 0.4, 0.6, 0.8), запазване без промяна на мерките (= 1.0) или затягане на мерки (т.е. избираме за коефициенти 1.2, 1.4, 1.6, 1.8),*

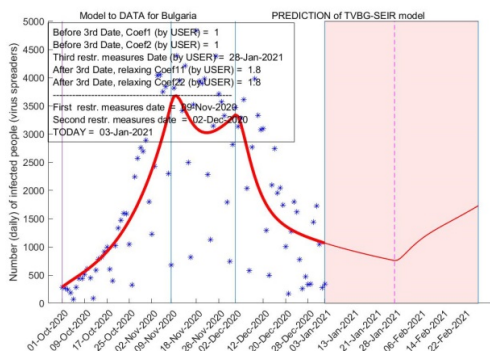
които определят параметрите  $\beta(t), \gamma(t)$  на епидемията след Today дата, както и на коефициентите Coef11, Coef22, които указват отпускане (ако са избрани 1.2, 1.6, 1.8) от двата вида мерки след датата на HORIZON (за които имаме три възможни избора, а именно, 5, 15, 25 дни от дата Today).

1. За България разгледаваме данните от *StartDate* която е  $T_1 = 1-Oct-2020$ , до крайна дата *Today*, равна на  $T_4 = 3-Jan-2021$ . Трета дата на ограниченията =  $T_5 = 28-Jan-2021$ , и за *Horizon* датата е  $T_6 = 22-Feb-2021$ .
2. Както обяснихме в раздел 2.4 минимизирането на функционала  $F(\Theta)$  се състои в разглеждане на множество двойки възли  $T_2, T_3$  (около 150 за тримесечен период) за сплайните  $\beta(t), \gamma(t)$ . Избираме двойката  $T_2, T_3$  и съответните параметри  $\Theta$  (които дефинират **Model1**) за които минимума на  $F(\Theta)$  на  $F(\Theta)$  се постига. Съществуват обаче и други вектори на параметрите  $\Theta$ , за които функционалът  $F(\Theta)$  достига стойности много близки до оптималната стойност  $F(\Theta)$ . Ние ги обозначаваме с  $\underline{\Theta} = \underline{\Theta}^1, \underline{\Theta}^2, \underline{\Theta}^3$ , и т.н. Тези вектори дефинират параметри  $\beta^{(j)}, \gamma^{(j)}$ , или еквивалентно, модели, които обозначаваме с *Model1, Model2, Model3*, и т.н.
3. За оптимистичен модел, *Model1*, намираме  $T_2 = 9-Nov-2020$ ,  $T_3 = 2-Dec-2020$ , с  $F(\underline{\Theta}) = 30.96$ , отгук и съотношението  $\max(F(\Theta)) / F(\underline{\Theta})$  е около 9. Фигура 2.4 показва най-простия сценарий за прогнозиране започвайки от *Today* =  $T_4 = 3-Jan-2021$ . В Легендата на Фигурата, Coef1 = 1 and Coef2 = 1 означава, че няма да има промяна на мерките от властите, стартиращи от *Today* и завършващи на *Трета дата на ограничения* =  $T_5 = 28-Jan-2021$ . По-нататък, Coef11 = 1 и Coef22 = 1 означава, че няма да последва облекчаване на мерките започвайки от *28-Jan-2020*.
4. Втората вълна е неизбежна, ако са по-облекчени мерките наложени от здравните власти: а именно, отпускане на двете мерки, т.е. Coef11 = Coef22 = 1.8 след *28-Jan-2021* ще генерира силна следваща вълна на инфекции, както се вижда от фигура 2.5
5. Подобни са заключенията с *Model3* за България (с  $Fval = 31.60$ , като се появява и следваща вълна.
6. За *Model2* (с  $Fval = 31.78$ ) имаме **най-оптимистичният** сценарий, тъй като можем частично отпускане и на двете мерки след





Фигура 2.4: Model1: Най-простият сценарий за прогнозиране, започващ от  $Today = T4 = 3\text{-Jan-}2021$ .



Фигура 2.5: Втората вълна се появява след отпускане както на Бета мерките, така и на Гама мерките, чрез коефициентите  $Coef11 = Coef22 = 1.8$  след 28-Jan-2021.

28-Jan-2021 (т.е.  $\text{Coef11} = \text{Coef22} = 1.4$ ), и няма да се появи втора вълна.

Както споменахме по-горе, можем да използваме примерно десет модела  $Model2, Model3, \dots, Model11$  и да генерираме техните криви  $I(t)$  за получаване на оценка на доверителни интервали за всеки момент от време  $t$ . Въпреки това, горното представяне използващо оптимистични и песимистични сценарии е най-просто и ясно.

## 2.6 Приложение към италиански и немски данни

Ние представяме подобни резултати, получени от нашия *инструмента SBT-COVID-19* за италианските данни. Резултатите за Италия, разглеждани до  $Today = 1-Jan-2021$  са подобни на българските. За  $Model0$  имаме  $T_2 = 6-Nov-2020$ ,  $T_3 = 20-Nov-2020$ , и

$$F_{val} = F(\Theta) = 153.09.$$

Представяме също така и анализ на германските данни посредством сценарии, генерирани от *инструмента SBT-COVID-19*.

Общото наблюдение е, че за разлика от данните за България и Италия, данните за Германия през периода март-май, 2020, са показали много силна тенденция към експлодиране в следваща вълна.

## Глава 3

# ATVBG-SEIR: SEIR модели за дългосрочни прогнози базирани на сплайни

Основната цел на настоящата глава е да представи ATVBG-SEIR модела за генериране на дългосрочни сценарии за прогнозиране ("проекции"), които ще ни помогнат да отговорим на въпроса, колко дълго са необходими локдауни за прекратяване на COVID-19 епидемия в една отделна държава, със или без ваксинации. Нашата методология се основава на конкретен SEIR модел, наречен ATVBG-SEIR модел, който е обяснен по-долу. Ние също така включваме ваксинации в модела, които се извършват според ваксинационен план, предварително зададен на месечна база. Разработен е допълнително модел модифициращ основния ATVBG-SEIR, който отразява и така наречения сезонен ефект (летния сезон). Наличен е и уеб базиран инструмент за сценарии за локдауни (достъпен онлайн на адрес <http://atvbg-seir.eu>), въз основа на алгоритъм, прилагащ методологията.

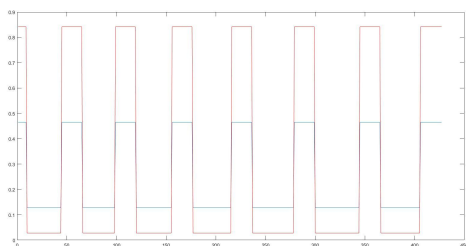
### 3.1 Моделът ATVBG-SEIR

За да направим адекватен SEIR модел и подходящ алгоритъм, ние вземаме предвид основните характеристики на епидемията COVID-19:

1. Нашият подход за моделиране на динамиката на епидемията е да използваме един *силно опростен* SEIR модел с променливи параметри, който ние наричаме *Alternating Time Varying Beta-Gamma-SEIR* (съкратено като **ATVBG-SEIR**) (който силно напомня за управленията от типа bang-bang в теорията на Оптималното управление). Моделът има само два редуващи се режима: **Стриктен режим** (*имитиращ строги мерки за ограничаване*) и **Релаксиран режим** (*имитиращ отслабени мерки за ограничаване*) с различни стойности за  $\beta, \gamma$ , които са редуващи се константи, както е обяснено по-долу.
2. Основното ограничение е горната граница на **леглата за интензивно отделение (ICUs)**, запазени за пациенти с COVID-19, които в случая на България не надвишават 1100, докато в Германия те са най-малко 8000 (с общ лимит от 40000 ICU), 2,764 в Италия, 26,900 в САЩ, и др.<sup>1</sup>
3. За простота избираме *ATVBG-SEIR модел*, който да използва фиксирани периоди от **около 21 дни** (или повече) за *Стриктни режими*, докато периодите за *Релаксирани режими* се определят итеративно от нашия алгоритъм при достигане на горната граница на ICU леглата в интензивните отделения.
4. Най-трудоемката и не очевидна задача при конструирането на ATVBG-SEIR модела е да се идентифицират строгият режим и релаксиращият режим.
5. Избрахме няколко **стриктни** режима за България. Един от тях има базово репродукционно число  $\mathcal{R}_0 = 0,9$  и се получава от данните за периода *01.4.2020 – 15.05.2020*.
6. Също така разгледахме и **стриктни** режими с репродукционни числа, които са много по-малки от единица (т.е. имаме затихване на епидемията):  $\mathcal{R}_0 \approx 0,8, 0,5, 0,2$

ПРИМЕР: По-долу представяме пример за *ATVBG-SEIR* модел с параметри за периода от *12-Dec-2020* до *11-Feb-2022*. Показана е графиките

<sup>1</sup>Не изглежда, че има много точни данни за горната граница на леглата в интензивните отделения, запазени за случаите на Covid-19, тъй като различните източници показват различен брой.



Фигура 3.1: Пример за параметри  $\beta(t)$  (синьо) and  $\gamma(t)$  (червено) за ATVBG-SEIR модела, за периода от 12-Дец-2020 до 11-Feb-2022.

на параметрите:  $\beta(t)$  (синьо) и  $\gamma(t)$  (червено). За всеки ден двойката числа  $(\beta(t), \gamma(t))$  алтернативно взема следните двойки стойности

$$StrictRegime = (0.4649, 0.8415)$$

$$RelaxRegime = (0.1276, 0.0279)$$

т.е.  $\beta(t)$  взема последователно стойностите 0.4649 и 0.1276, а  $\gamma(t)$  взема последователно стойностите 0.8415 и 0.0279. В този пример периодите на *Strict режим* са **21 дни**, както се вижда на Фигура 3.1

## 3.2 Активни случаи в непрекъснатия SEIR модел

Основната идея при разработването на SEIR модела е да се осигурят някои **проследими** приближения  $S(t)$ ,  $E(t)$ ,  $I(t)$ ,  $R(t)$  към реалните данни. В раздел 2.2 чрез системата уравнения (2.1)-(2.4) въведохме класическия диференциален SEIR модел, следвайки обичайните референции [14], [18], [20], с постоянни  $\beta(t)$  и  $\gamma(t)$ . Приемаме, че началните стойности  $S(0)$ ,  $E(0)$ ,  $I(0)$ , and  $R(0)$  са дадени и системата е решена за моменти от време  $t \geq 0$ . Както вече обсъдихме, вижте (2.5), приема се, че е в сила следното уравнение:

$$N = S(t) + E(t) + I(t) + R(t)$$

където  $N$  е общото население в страната  $XX$ . Важно за нас е емпиричното наблюдение, че критичните ICU случаи (в интензивно отделение)

са стабилен процент от активните случаи. Следователно трябва да намерим израз за Активните случаи в модела SEIR. В наличните емпирични данни Активните случаи в деня  $t$  са дадени от формула:

$$\begin{aligned} AC(t) = \text{Active Cases}(t) &= \text{Total number of infected}(t) \\ &- \text{Total number of Recovered}(t) \\ &- \text{Total number of Fatalities}(t) \end{aligned} \quad (3.1)$$

Получаваме следния резултат, като използваме теорията на ОДУ, [22],[23]:

**Theorem 3.** Да приемем, че променливите параметри  $\beta(t), \gamma(t)$  от модела SEIR имат краен брой точки на прекъсване, които обозначаваме  $0 < t_1 < t_2 < \dots < t_p < T$ . След това, за всеки момент от времето  $t \neq t_j$  за  $j = 1, 2, \dots, p$  и  $0 \leq t \leq T$  определя:

$$AC(t) = I(t) + C \quad (3.2)$$

където  $C$  е константа.

### 3.3 Дискретизиране на SEIR модела и активни случаи

На практика се използва дискретизация на непрекъснатия SEIR модел. Разгледахме дискретизацията на SEIR модела (2.7) - (2.10) в раздел 2.3. Той е много интуитивен и фактически произлиза от метода на Ойлер за апроксимация на решение с начална стойност (2.1) - (2.4) в раздел 2.2 (вж. също [20]),

$$\begin{aligned} S_{n+1} &= S_n - \frac{\beta_n S_n I_n}{N} \\ E_{n+1} &= E_n + \frac{\beta_n S_n I_n}{N} - \sigma E_n \\ I_{n+1} &= I_n + \sigma E_n - \gamma_n I_n \\ R_{n+1} &= R_n + \gamma_n I_n \end{aligned}$$

Приемаме, както е в (2.11):

$$N = S_n + E_n + I_n + R_n$$

Сега представяме следната дефиниция на така наречените **Модифицирани активни случаи**:

**Definition 4.** Определяме **Модифицирани активни случаи за SEIR модела**: За  $N = 1$  полагаме:

$$MAC_1 = TotalInf(1) - R_2 \quad (3.3)$$

и за всяко цяло число  $P \geq 2$  задаваме:

$$MAC_P = TotalInf(1) + \sigma \sum_{n=2}^P E_n - R_{P+1} \quad (3.4)$$

В сила е следният важен резултат, а именно:

**Theorem 5.** **Модифицираните активни случаи** удовлетворяват равенството

$$MAC_P - MAC_1 = I_{P+1} - I_2 \quad (3.5)$$

или еквивалентно,

$$MAC_P = I_{P+1} + TotalInf(1) - R_2 - I_2.$$

Важно е да отбележим следното съответствие между официални данни и дискретния SEIR модел:

1. Означението на кривата  $I(t)$  в SEIR модела донякъде подвежда. Това означение обикновено се смесва с ежедневно новоинфектираните случаи. Всъщност, истината е, че в деня  $t$  ежедневните новоинфектирани случаи  $New\_Infected$  са равни на

$$NewInfected_t = \sigma E_t$$

където  $E_t$  е променливата на дискретния SEIR модел.

2. Видяхме по-горе, че в *непрекъснатия* SEIR модел кривата  $I(t)$  съвпада с активните случаи в деня  $t$ , с точност до константа, да припомним теорема 5. От друга страна, в *дискретния* SEIR модел дискретната крива  $I_t$  съвпада до константа с модифицираните активни случаи  $MAC_{t-1}$ .

## 3.4 Моделиране на ваксинациите

Възможно е да се въведат без големи усилия ваксинациите в нашите непрекъснати и дискретни SEIR модели.

Има множество статии, посветени на моделирането на ваксинацията чрез SIR и SEIR модели, както и за оптимален контрол на процеса

на ваксинация. Тук трябва да споменем следните източници: Biswas et al. (2014), [15], Neilan и Lenhart (2010), [24], Brauer and Castillo-Chavez (2001) [17], Brauer et al. (2008) [16].

Ще обозначим с  $u(t)$  процента на възприемчивите лица  $S(t)$ , които се ваксинират за единица време. Друга формулировка е да се използва директно числото  $V(t)$  на  $V(t) = u(t)S(t)$ . По-специално, ако ваксинираме 30000 хора на месец, тогава ние ще трябва да извадим  $30000/30 = 1000$  души в деня  $t$  от числото на възприемчивите индивиди  $S(t)$ . Ще пропуснем формулировката на непрекъснатия SEIR модел с ваксинации и тук представяме само неговия дискретен аналог:

$$S_{n+1} = S_n - \beta_n S_n I_n / N - V_n \quad (3.6)$$

$$E_{n+1} = E_n + \beta_n S_n I_n / N - \sigma E_n \quad (3.7)$$

$$I_{n+1} = I_n + \sigma E_n - \gamma_n I_n \quad (3.8)$$

$$R_{n+1} = R_n + \gamma_n I_n + V_n \quad (3.9)$$

### 3.5 Определяне на броя на ICU и Алгоритъм към методологията

По-долу обсъждаме най-важните елементи от нашия алгоритъм:

1. Емпиричните данни до момента показват, че броят на ICU, заети от критични случаи на COVID-19 са много силно свързани с така наречения **Активни случаи**.
2. Според **емпиричните** наблюдения броят на пациентите при COVID-19, настанени в ICU леглата за интензивни грижи варира, например за България става въпрос за 0,55%, а за Германия около 1,35% от активните случаи. При прилагането на нашия алгоритъм ние спазихме тези проценти. Нашата основна цел е да отчетем броя на ICU случаите на COVID-19 в рамките на **ATVBG-SEIR модела**. Съответно, полагаме

$$ICU(t) = 0.55\% * AC(t)$$

в ATVBG-SEIR модела за България. Тук процентът 0,55% е бил получен като средна стойност за достатъчно дълъг период от време.

3. Друг важен момент в подкрепа на адекватността на резултатите е, че по мнението на много медицински експерти, заразените случаи



### 3.6. ОСНОВНИ РЕЗУЛТАТИ, ПОЛУЧЕНИ С ИНСТРУМЕНТА ЗА СЦЕНАРИИ ЗА ЛОКДАУНИ БЕЗ ОТЧИТАНЕ НА СЕЗОНЕН ЕФЕКТ 39

са поне 3 – 6 пъти повече от официално регистрираните. В нашия модел ние вземаме това в предвид.

4. Според експертите епидемията приключва, ако 70% от населението придобие имунитет чрез ваксинация или чрез преболедуване. В нашия алгоритъм сме задали това условие като **правило за спиране** на алгоритъма, което при изпълнение автоматично определя продължителността на итерациите съгласно ATVBG-SEIR модела.

## 3.6 Основни резултати, получени с инструмента за сценарии за локдауни без отчитане на сезонен ефект

### 3.6.1 Резултати за България:

Следвайки *оптимистичния сценарий* за България (дефиниран в раздел 3.7 по-долу), а именно

$$RelaxRegime = (\beta = 0.0436, \gamma = 0.0215), \quad \mathcal{R}_0 = 2.0,$$

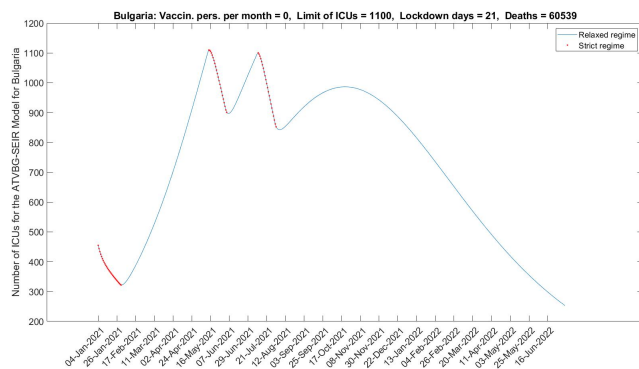
$$StrictRegime = (\beta = 0.0199, \gamma = 0.0288), \quad \mathcal{R}_0 = 0.6915$$

без ваксинации епидемията завършва на *06-Jul-2022*, с два основни локдауна през пролетта/лято, докато 70% от населението получи имунитет, и прогнозният общ брой смъртни случаи е 60539, вижте фигура 3.2.

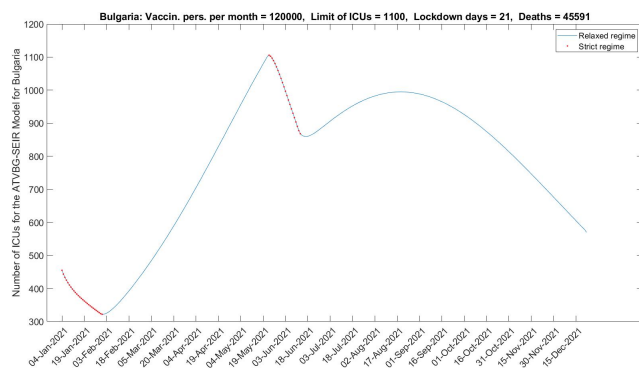
За същия *оптимистичен сценарий*, както се вижда от Фигура 3.3, с план за ваксинация 120000 лица (240000 ваксинации) на месец, епидемията завършва на *26-Dec-2021*, с един голям локдаун, докато 70% от населението получи имунитет, и приблизително общ брой смъртни случаи 45591: От друга страна, ако изберем *песимистичния сценарий* (в раздел 3.7 по-долу), а именно,

$$RelaxRegime = (\beta = 0.0902, \gamma = 0.0159), \quad \mathcal{R}_0 = 5.7$$

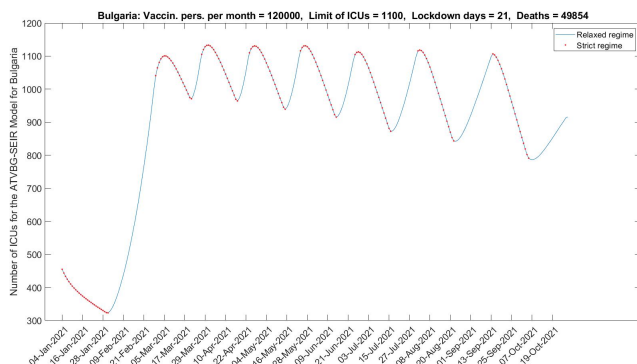
$$StrictRegime = (\beta = 0.0218, \gamma = 0.0270), \quad \mathcal{R}_0 = 0.8056$$



Фигура 3.2: Оптимистичен сценарий за България - кривата на ICU; без ваксинации.



Фигура 3.3: Оптимистичен сценарий за България - кривата на ICU; с ваксинации.



Фигура 3.4: Песимистичен сценарий за България: кривата на ICU; с ваксинации

### 3.6.2 Оценка на броя на смъртните случаи за оптимистични/песимистични сценарии

За да се получи груба оценка на броя на смъртните случаи, т.е. крива  $Deaths(t)$  за ATVBG-SEIR модела, припомняме, че той не участва директно в модела, но е включен в премахнатите случаи, в кривата  $R(t)$ . Може да избегнем разглеждането на по-сложен модел от типа SEIRD (където променливата  $D$  съответства на кривата  $Deaths(t)$ ) чрез въвеждане на емпирично съотношение. Тествахме различни подходи и установихме, че коефициентът на смъртност зададен от съотношението основано на емпиричните данни:

$$\mathcal{D}_0 \approx \frac{Deaths(t)}{TotalInf(t)} \quad (3.10)$$

е доста стабилен. Основният статистически анализ показва че във всички случаи на страните, които разгледахме, горепосоченият коефициент на смъртност има най-малко стандартно отклонение в сравнение с други съотношения, като например

$$Deaths(t) / CumulativeICU(t) \quad \text{и} \quad Deaths(t) / Recovered(t).$$

За нас е важно, че за дати  $t$  след *Today* ще можем да използваме ATVBG-SEIR модела, а именно, ще изчислим  $TotalInf(t)$  чрез използване на кривата  $E(t)$  на дискретния модел ATVBG-SEIR, тъй като  $TotalInf(t)$

съвпада със сума от  $\sigma E(t)$ . Следователно можем да направим прогнозен сценарий на кривата  $Deaths(t)$ ; получаваме оценка на кривата на смъртните случаи в бъдеще чрез формулата:

$$Deaths(t) = \mathcal{D}_0 * TotalInf(t)$$

като използваме това съотношение практически само на последната дата  $t$  от прогнозираната епидемия.

Стойността на  $\mathcal{D}_0$  зависи от конкретната държава, като ние взимаме предвид и коефициента на ескалация, показващ реалния брой на заразни. Следователно получаваме, че  $\mathcal{D}_0$  за България има стойност средно около 0,012 до януари 2021 г. За същия период, за Германия, стойността му е 0,011; за Австрия - 0,008; за САЩ - 0,012, за Италия - 0,033, за Великобритания - 0,032.

Трябва да кажем, че това, което получаваме, е оценка на кривата  $Deaths(t)$ , която не отчита, евентуална адекватна ваксинация, където на първа линия са тези, които са най-уязвими към COVID-19 (с тежки съпътстващи заболявания). Несъмнено, при отчитане на този фактор смъртните случаи определено ще бъдат по-малко.

### 3.7 Напасване (fitting) на класическия модел SEIR към българските данни

Необходимо е сериозно изследване, за да се намерят подходящи строг режим и релаксиран режим, които ще представляват частите от сплайните с променливи параметри  $\beta(t)$ ,  $\gamma(t)$  на ATVBG-SEIR модела.

Опитахме се да намерим параметрите  $\beta$ ,  $\gamma$  на класическия SEIR модел по време на Строги мерки през месеците април-август 2020. Обаче самият факт, че полученият основен коефициент на репродукция  $\mathcal{R}_0$  винаги е по-голям от единица, показва, че нещо не е наред. Единственото логично обяснение е, че през този период имаше много голям входящ поток на работна емиграция и студенти (предимно български граждани) от чужбина.

1. През периода 6 април 2020 г. – 15 май 2020 г. имяхме много строги мерки в България. Калибрирахме SEIR модела и получихме следните параметри:

$$\beta = 0.0517, \gamma = 0.0156$$

които дават стойност на базисното репродуктивното число

$$\mathcal{R}_0 = \frac{\beta}{\gamma} = 3.3118$$

2. През периода *15 май 2020 г. – 31 юли 2020 г.* имахме някои разхлабващи мерки, като отворени ресторанти и други, но НЯМА училища, а точно тогава ние имаме и летен сезонен ефект, получихме следните параметри за напаснатия (fitted) SEIR модел:

$$\beta = 0.0434, \gamma = 0.0254$$

което дава базисно репродуктивно число

$$\mathcal{R}_0 = \frac{\beta}{\gamma} = 1.7092 > 1$$

което е по-голямо от едно. Отново обяснението е, че много сезонни работници и български студенти дошли от чужбина са пренесли заразата.

3. Също така за периода *15 май 2020 г. – 31 август 2020 г.* ние получаваме:

$$\beta = 0.0495, \gamma = 0.0327$$

което определя

$$\mathcal{R}_0 = \frac{\beta}{\gamma} = 1.5142 > 1.$$

4. За период *1 June, 2020-15 Sept., 2020*, имаме:

$$\beta = 0.0685, \gamma = 0.0469$$

Следователно получаваме,

$$\mathcal{R}_0 = \frac{\beta}{\gamma} = 1.4601 > 1$$

5. Въпреки това, само близо до септември получихме базисно репродуктивно число  $\mathcal{R}_0 < 1$ , което може да се обясни с много силен сезонен ефект. За периода *1 август, 2020-20 септември, 2020*, получихме параметрите:

$$\beta = 0.0252, \gamma = 0.0363, \quad \mathcal{R}_0 = 0.6958 < 1$$

6. От друга страна, ние избрахме като **стриктен** режим този който получихме през зимата, за *1 декември, 2020-4 януари, 2020*, а именно

$$\beta = 0.0208, \gamma = 0.0291, \quad \mathcal{R}_0 = 0.7134 < 1$$

7. Избрахме за **релаксиран** режим този в периода *15 Септември, 2020 - 15 октомври, 2020*, а именно

$$\beta = 0.0436, \gamma = 0.0215, \quad \mathcal{R}_0 = 2.0$$

Въз основа на горното дефинирахме един възможен **релаксиран** режим определен като:

$$RelaxRegime = (\beta = 0.0902, \gamma = 0.0159)$$

За **Строг режим** проверихме различни периоди. За период на официално въведени строги мерки (няма училища, няма ресторанти, и др.), *1 декември - 29 декември, 2020*, калибрахме SEIR модела и получихме следните параметрите:

$$(\beta = 0.0199, \gamma = 0.0288)$$

с базисно репродуктивно число

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \approx 0.691.$$

Следователно, полагаме

$$StrictRegime = (\beta = 0.0199, \gamma = 0.0288)$$

В нашето изследване за намиране на подходящи параметри за строги и релаксирани режими, ние проверихме различни под-интервали от периода *1 април - 15 септември, 2020*.

За периода с частични мерки (няма училища, но отворени ресторанти и т.н.), *1 август - 15 септември, 2020*, получихме

$$(\beta = 0.0312, \gamma = 0.0367), \quad \mathcal{R}_0 = 0.8524$$

Освен това, за периода с облекчени мерки (отворени училища, отворени ресторанти, и др.), *1 октомври-30 октомври, 2020*, получихме следните параметри на калибрираният класически модел SEIR

$$(\beta = 0.0871, \gamma = 0.0162), \quad \mathcal{R}_0 = 5.4$$

и за периода със същите облекчени мерки, *15 септември-15 октомври, 2020*, получихме

$$(\beta = 0.0436, \gamma = 0.0215), \quad \mathcal{R}_0 = 2.0$$

В резултат на горното може да се дефинират няколко модела, които могат да бъдат разглеждани в диапазона от най-оптимистични до най-песимистични:

За **България**:

Сценарият **оптимистичен** е със следните параметри:

$$RelaxRegime = (\beta = 0.0436, \gamma = 0.0215), \quad \mathcal{R}_0 = 2.0$$

$$StrictRegime = (\beta = 0.0199, \gamma = 0.0288), \quad \mathcal{R}_0 = 0.6915$$

**Умереният** сценарий:

$$RelaxRegime = (\beta = 0.0599, \gamma = 0.0141), \quad \mathcal{R}_0 = 4.239$$

$$StrictRegime = (\beta = 0.0218, \gamma = 0.0270), \quad \mathcal{R}_0 = 0.8056$$

**Песимистичен** сценарий:

$$RelaxRegime = (\beta = 0.0902, \gamma = 0.0159), \quad \mathcal{R}_0 = 5.7$$

$$StrictRegime = (\beta = 0.0218, \gamma = 0.0270), \quad \mathcal{R}_0 = 0.8056$$

Ние внедрихме методологията на модела ATVBG-SEIR в Lockdown Инструмент, при който крайният потребител може да играе с няколко параметъра, което е най-добрият начин да получи разбиране за силните страни на модела, за да прави прогнози с него, наличен на <http://atvbg-seir.eu>.

В горния модел има два основни фактора, които могат значително да намалят продължителността на епидемията и броя на смъртните случаи: 1. Сезонният ефект, който моделираме в раздела по-долу. 2. Приоритетната имунизацията

### 3.8 Моделиране на сезонния ефект за разпространение на COVID-19

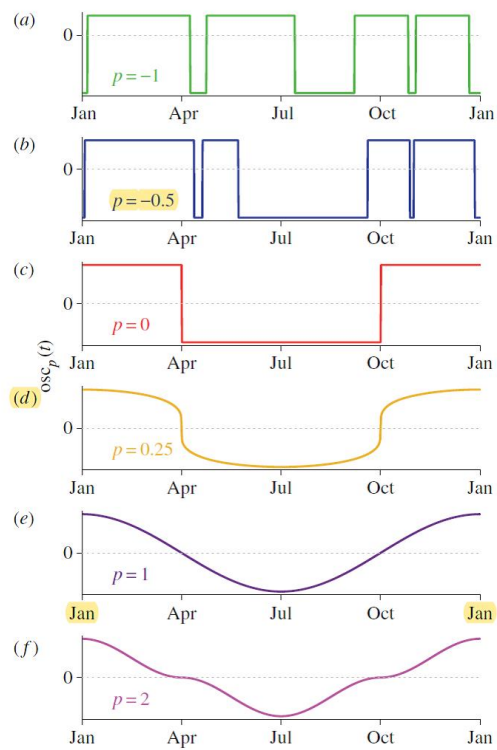
Поради липсата на достатъчно информация в момента, най-добрият подход към моделирането на сезонното поведение на разпространението на Covid-19 изглежда е в пряка аналогия със сезонното поведение на стандартния грип, вижте обзорната статия [27].

Голям брой научни трудове са изследвали богатите нелинейни ефекти, причинени от периодично вариращи проценти на контакти  $\beta(t)$  в *епидемични модели*; съществуват няколко отлични обзора, вижте напр. [25], [26].

В публикацията [28], следващият модел е използван:

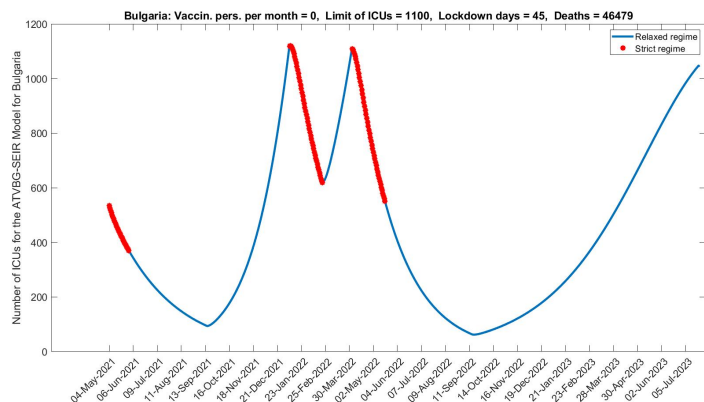
$$\beta(t) = \beta_0 (1 + \beta_1)^{\phi(t)}.$$

Други използват различни прости модели, като в [29], които предоставят визуализирането на Фигура 3.5:



Фигура 3.5: Различни сезонни модели.





Фигура 3.6: Инструмент за сценарий за локдауни за България, от 4 май 2021 г.

1. В случай, че не ни е необходима точна оценка, а само груба, можем да използваме просто изместване на кривите през летния сезон с около 4,5 – 5 месеци. Съответно, грубата оценка на броя на инфекциозните случаи и броя на смъртни случаи също не е трудно да се изчисли.
2. Въпреки това, друг по-сложен, но все още лесен за изпълнение подход за отчитане на сезонността дава разумно приближаване до реалната картина, и разширява модела ATVBG-SEIR. А именно, предполагаме, че през "ниския" сезон 15-May-2021 - 15-Яер-2021, "летният ефект" е **много подобен** на ефекта, причинен от мерките за стриктно ограничаване, т.е. използваме параметри  $\beta(t)$  и  $\gamma(t)$  в ATVBG-SEIR модела, които са същите като в *StrictRegime* = ( $\beta = 0.0199$ ,  $\gamma = 0.0288$ ), предоставени по-горе в раздела 3.7.
3. Чрез прилагане на инструмента за сценарий за локдауни за България, започвайки на 4 май, 2021 г., получаваме крива на ICU на фигура 3.6.
4. Виждаме, че се появява "период на забавяне" от около 4 – 5 месеци за лятото на 2021 и 2022 г.

## Глава 4

# Софтуерни платформи, използвани за уеб-базираните инструменти

В настоящата глава предоставяме повече подробности за прилагането на уеб-базираните инструменти за Covid-19 и използваните при разработката софтуерни платформи и програмни среди.

### Кратък преглед на проекта Jupyter

Jupyter е проект на общност за разработване на софтуер с отворен код и услуги за интерактивни изчисления. Проектът Jupyter стартира като отделен проект от IPython разработен от Фернандо Перес през 2011 г. като подобрен интерпретатор на Python. Интерактивен Python, а именно IPython първоначално е разработен за езика Python като командна обвивка за интерактивни изчисления. Системата Jupyter поддържа над 100 езика за програмиране (наречени „ядра“ (kernels) в екосистемата Jupyter) включително Python, Java, R, Julia, Matlab, Octave, Scheme, Scala и много други. Jupyter изпълнява ядрото на IPython, като могат да бъдат инсталирани допълнителни ядра. Софтуерните приложения по

---

проекта Jupyter са предназначени да поддържат интерактивни програми за данни и научни изчисления. Списък с основните компоненти, които се съдържат в проекта Jupyter са:

- Jupyter Notebook - Уеб базиран интерфейс към различни програмни среди на Python, Julia, R и много други.
- JupyterLab - Модерен уеб базиран интегриран интерфейс и IDE среда за различните notebooks (т.е. уеб документи с изпълним код), редактори, конзоли и др.
- Jupyter Client - Това е услуга, която съдържа референтната реализация на протокола Jupyter. Също така е клиентска библиотека за стартиране, управление и комуникация с ядра (kernels) на Jupyter.
- Jupyter kernels - Това са среди за изпълнение за всеки език за програмиране на Jupyter.
- IPykernel - Софтуерен пакет, който предоставя ядрото на IPython на Jupyter.
- QtConsole - Qt базиран терминал за ядра на Jupyter, подобен на IPython.
- nbviewer - HTML viewer за Jupyter notebooks.
- nbconvert - Конвертира файловете на бележника на Jupyter в други формати.

За повече подробности в онлайн книгата "Jupyter Notebook ecosystem вижте <https://jupyter4edu.github.io/jupyter-edu-book/jupyter.html>.

От друга страна, IPython предлага повече функции в сравнение със стандартния Python:

- Действа като основно ядро за Jupyter notebook и други инструменти от потребителски интерфейс (front-end) на проекта Jupyter.
- Притежава способност за самоанализ на обекта. Интроспекцията е способността да се проверяват свойствата на обект по време на изпълнение.
- Маркиране на синтаксис.

- Съхранява историята на взаимодействията.
- Завършване на ключови думи, променливи и имена на функции.
- Magic командна система, полезна за контролиране на средата на Python и изпълнение на задачи на ОС.
- Възможност за вграждане в други програми на Python.
- Осигурява достъп до инструмента за отстраняване на грешки в Python debugger.

## Кратък преглед на Jupyter Notebook

Jupyter Notebook е разработен като концепция за изчислим документ, вижте <https://ipython-books.github.io/chapter-3-mastering-the-jupyter-notebook/>. Тези „Notebook“ или „Notebook документи“ се използват и като уеб приложения, могат да създават и споделят изчислими документи, които съдържат изпълним програмен код на живо, визуализации и интерактивни уеб елементи. Jupyter Notebook е отличен инструмент предоставящ възможностите на уеб средата за извършване на анализ на научни данни в реално време.

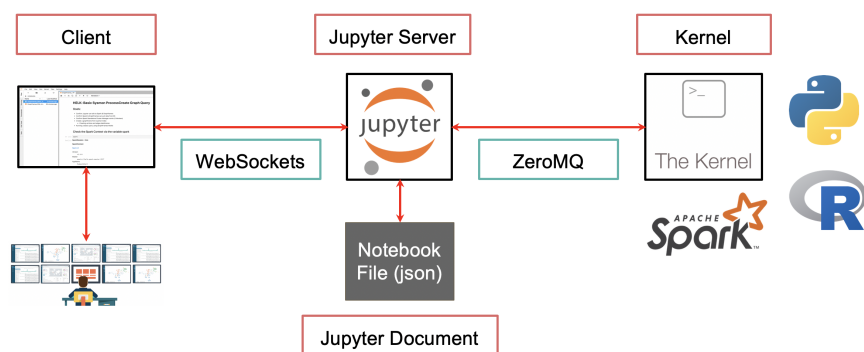
Екосистемата на Notebook: Jupyter Notebook са представени като документи на JavaScript Object Notation (JSON). JSON е езиково независим текстов файлов формат за представяне на структурирани документи. Каквито са Notebook (документите), които могат да бъдат обработвани от всеки език за програмиране и те могат да бъдат преобразувани в други формати като Markdown, HTML, LaTeX/PDF и други.

Около Notebook (документа) има цяла екосистема от различни инструменти. Notebook се използват за създаване на слайдове, учебни материали, публикации в блогове, научни статии и дори книги.

## Софтуерна архитектура на Jupyter Notebook

Jupyter реализира двупроцесов модел с ядро и клиент. Клиентът е интерфейсът, предлагащ на потребителя възможността да изпраща код до ядрото. Ядрото изпълнява кода и връща резултата на клиента за показване. В терминологията Read-Evaluate-Print Loop (REPL) ядрото прилага Evaluate, докато клиентът прилага Read и Print на процеса.

Клиентът може да бъде Qt widget, ако изпълняваме конзолата Qt, или браузър, ако стартираме Jupyter Notebook. В Jupyter Notebook ядрото получава цели клетки наведнъж, така че няма представа за самият



Фигура 4.1: Jupyter Software Architecture

Notebook. Съществува силно отделяне между линейния документ, съдържащ Notebook, и основното ядро (kernel).

Всички комуникационни процедури между различните процеси се изпълняват върху протокола за съобщения ZeroMQ (или ZMQ) (<http://zeromq.org>).

Notebook комуникира с основното ядро, използвайки WebSocket, базиран на TCP протокол, внедрен в съвременните уеб браузъри.

Свързване на множество клиенти към едно ядро: Във Notebook, въвеждането на `connect-info` в клетка дава информацията, от която се нуждаем, за да свържем нов клиент (например конзола Qt) към основното ядро.

JupyterHub: JupyterHub, достъпен на <https://jupyterhub.readthedocs.io/en/latest/>, е библиотека на Python, която може да се използва за обслужване на Notebook на набор от крайни потребители, например студенти от конкретен клас или членове на лаборатория в изследователска група. Той обработва удостоверяване на потребителя и други подробности на ниско ниво.

Сигурност в Notebook: Възможно е нападателят да постави злонамерен код в Jupyter Notebook. Тъй като Notebook могат да съдържат скрит JavaScript код в клетъчен изход, теоретично е възможно зловредният код да се изпълнява тайно, когато потребителят отвори Notebook. Поради тази причина Jupyter има модел на защита, при който HTML и

JavaScript кодът в Notebook може да бъде доверен или ненадежден. Изходите, генерирани от потребителя, винаги имат доверие. Въпреки това, изходите, които вече са били налице, когато потребителят за първи път е отворил съществуващ Notebook, са ненадеждни.

Моделът на защита се основава на криптографски подпис, присъстващ във всеки Notebook. Този подпис се генерира с помощта на секретен ключ, собственост на всеки потребител.

## История и развитие на системата IPython и Jupyter

IPython първоначално е разработен от Фернандо Перес през 2001 г. Текущата му версия е IPython 7.19.0, която изисква версия на Python 3.4 или по-нова. IPython 6.0 беше първата версия, поддържаща Python 3. Потребителите с Python 2.7 трябва да работят с IPython версия 2.0 до 5.7

Концепцията за изчислителни документи (Notebooks) започна през 80-те години на миналия век, когато бяха пуснати MATLAB и Mathematica. Тези графични интерфейси към интерактивната обвивка имаха функции като форматиране на текст, добавяне на графики, таблица и добавяне на математически символи. SAGE Notebook също е уеб базиран Notebook.

Създателите на IPython започнаха работа по интерфейса на бележника за черупката на IPython през 2005 г. IPython Notebook скоро добави поддръжка на други езици като R и Julia. През 2014 г. Perez стартира проекта Jupyter като отделен проект от IPython, тъй като проектът IPython ставаше голям с добавени към него инструменти като сървър за Notebook и конзола Qt.

От IPython 4.0 всички допълнителни компоненти бяха преместени в проекта Jupyter и добавиха поддръжка на други езици към IPython Notebook. IPython продължава да се фокусира върху подобряването на своята подобрена функция за интерпретатор. Той също така предоставя основно ядро на интерфейса (front-end) на Jupyter Notebook.

## Кратък преглед на Vokeh като библиотеката и като програмна среда

Vokeh библиотеката и Vokeh средата са конкурентни на Jupyter Notebook,

вижте <https://docs.bokeh.org/en/latest/docs/reference/server.html>.

Bokeh обаче е не само конкурентен, но може да се комбинира и да работи съвместно с Jupyter Notebook.

По този начин използвахме Bokeh за интерфейс като част на внедряването - това включва библиотеки с приспособления, контроли, интерактивни елементи на потребителския интерфейс:

„Bokeh.server“ е аналогът на Jupyter Notebook. Той предлага персонализирано приложение Bokeh Server TornadoCore (уеб сървър), популярно сред уеб софтуерните разработки.

Архитектурата на Bokeh е такава, че „моделни обекти“ на високо ниво (представляващи неща като графики, диапазони, оси, глифове и т.н.) се създават в Python и след това се преобразуват във формат JSON, който се консумира от клиентската библиотека, BokehJS.

Сам по себе си този гъвкав и отделен дизайн предлага предимства, например е лесно други езици (R, Scala, Lua и т.н.) да управляват точно същите графики и визуализации на Bokeh в брауъра.

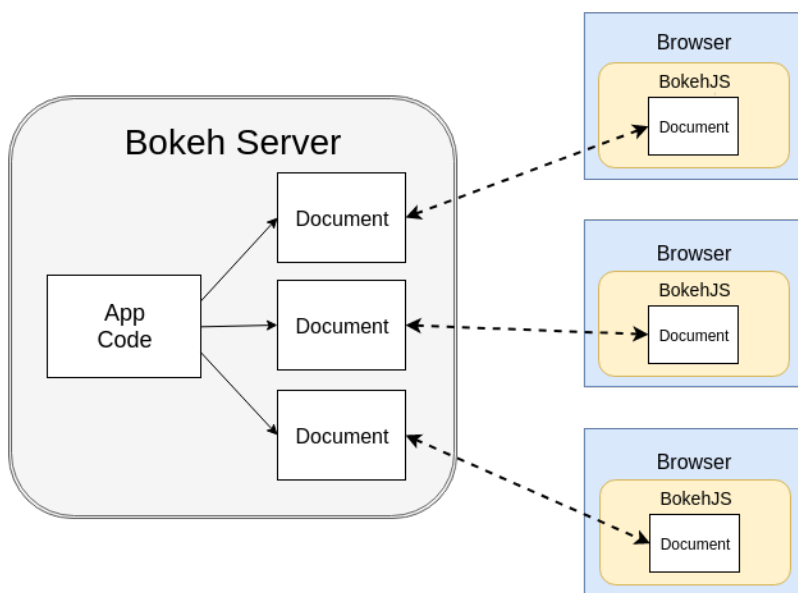
Възможността за синхронизиране между Python и брауъра е основната цел на сървъра Bokeh.

Най-гъвкавият начин за създаване на интерактивни визуализации на данни с помощта на сървъра Bokeh е създаването на приложения Bokeh и обслужването им със сървър Bokeh. В този сценарий сървър Bokeh използва кода на приложението за създаване на сесии и документи за всички клиенти (обикновено брауъри), които се свързват. На фигурата по-долу: Bokeh сървър (вляво) използва код на приложение за създаване на Bokeh документи. Всяка нова връзка от брауър (вдясно) води до това, че сървърът Bokeh създава нов документ, само за тази сесия.

Диаграмата на сървъра Bokeh е дадена на Фигура 4.2.

В нашето изпълнение създадохме приложение на Bokeh, което се изпълнява НЕ под сървър Bokeh, а в Jupyter Notebook. Това наложи създаването на интеграционен буфер и персонализирана конфигурация. Кодът на приложението се изпълнява в сървъра Bokeh всеки път, когато се прави нова връзка, за да се създаде новият документ Bokeh (известен още като Notebook), който ще бъде синхронизиран с брауъра. Кодът на приложението също така задава всички обратни извиквания, които трябва да се изпълняват винаги, когато се променят свойства като стойности на приспособленията (widgets).

Стартиране на сървър Bokeh (приложение): вижте [https://docs.bokeh.org/en/latest/docs/user\\_guide/server.html](https://docs.bokeh.org/en/latest/docs/user_guide/server.html) Основната цел на



Фигура 4.2: Bokeh server



сървърът Bokeh е да синхронизира данните между основната среда на Python и библиотеката BokehJS, работеща в брауъра.

Сървърът Bokeh улеснява създаването на интерактивни уеб приложения, които свързват събитията на интерфейса на предния край с изпълняващия код на Python.

Bokeh създава модели на високо ниво на Python, като графики, диапозони, оси и глифове, след което преобразува тези обекти в JSON, за да ги предаде на своята клиентска библиотека, BokehJS.

За повече информация относно последното, вижте описанието на BokehJS. Този гъвкав и отделен дизайн предлага някои предимства. Например, лесно е да имате други езици, като R или Scala, да управлявате графики на Bokeh и визуализации в брауъра.

Поддържането на тези модели в синхрон между средата Python и брауъра ще осигури допълнителни мощни възможности:

- отговаря на потребителски интерфейс и събития с инструменти в брауъра с изчисления или заявки, използвайки пълната възможности на Python.
- автоматично изпраща актуализации от страна на сървъра към елементите на потребителския интерфейс, като например widgets или графики в брауъра
- използва периодични, изчаквания и асинхронни обратни извиквания, за да стимулирате поточните актуализации

Това е мястото, където Bokeh сървърът влиза в приложение. Документацията за Bokeh е достъпна на връзката <https://github.com/bokeh/bokeh>.

Bokeh освен като сървър може да се използва и като интерактивна библиотека за визуализация за съвременните уеб брауъри.

Той осигурява елегантна конструкция с многофункционална красива графика, интерактивност при големи или поточни набори от данни. Bokeh може да помогне на всеки, който би искал бързо и лесно да прави интерактивни визуализации, табла за управление и приложения за данни.

С помощта на Bokeh може да създават визуализации, работещи с JavaScript без да се пише код на JavaScript.

Bokeh е достъпен и през дистрибуцията на Anaconda Python. Bokeh е спонсориран проект на NumFOCUS, вижте <https://docs.bokeh.org>

`/en/latest/index.html`.

Следните препратки са много полезни за първоначално запознаване с Jupyter и Bokeh:

1. Karlijn Willems, Jupyter Notebook Tutorial: The Definitive Guide, at <https://www.datacamp.com/community/tutorials/tutorial-jupyter-notebook>
2. Tutorialspoint - Jupyter Tutorial, at [https://www.tutorialspoint.com/jupyter/ipython\\_introduction.htm](https://www.tutorialspoint.com/jupyter/ipython_introduction.htm)

## Глава 5

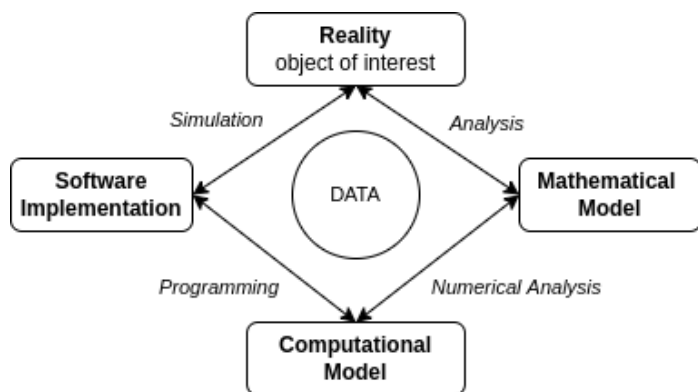
# Софтуерна реализация на инструментите

В настоящата глава се предоставя повече подробности за разработката на софтуер на създадените веб базирани инструменти и са дадени някои ключови концепции.

Дисертацията има принос към компютърната епидемиология чрез моделиране на заразни заболявания, в частност на COVID-19. Проучване на нови модели за анализиране на епидемични огнища, включително и създаване на техните софтуерни реализации като веб-базирани интерактивни инструменти за анализ, прогнози, сценарии за подпомагане на вземането на решения.

### 5.1 Преглед на някои ключови концепции

При създаването на модели на процеси, случващи се в реалния свят, моделите се основават на разбиране на процеса, посредством анализиране на наблюдаваните данни. Математическият модел е опростена математическа конструкция, свързана с част от реалността. В други случаи можем да направим разлика между моделиране на данни и математическо моделиране. Човек може да открие и съпостави интересен модел от данни, но да не може или с малка възможност да обясни процеса чрез изграждане на математически модел и това моделиране се отна-



ся до моделиране на данни. Калибриране (намиране на стойности за параметри) за моделите се изпълняват програмно симулации/модели за генериране на данни и напасване с реалните наблюдавани данни. Софтуерните модели са свързани с процеса на разработка и има много различни методологии и софтуерни подходи.

## 5.2 Софтуерен подход

Разработка с тестове или Test-Driven Development (TDD) подход е използван за създаването на инструментите. Описовните стъпки са описани:

1. тестов код с условия, ако се проваля (false). При успех, тестваният код вече покрива необходимата функционалност.
2. създаване на код, за да премине успешно теста. Всички предишни тестове трябва да са успешни. Новият код се добавя към съществуващата функционалност.
3. преработване кода, ако е необходимо, подобрения на кода, без да се променя поведението или каквато и да е функционалност на този етап.
4. Повтаряне на предходните стъпки 1-2-3

## 5.3 Софтуерно приложение

Създадените инструменти могат да се използват от потребителите като обикновени уеб приложения, освен това инструментите имат и специален режим с по-сложна структура като изчислими документи, известни още като IPython (Jupyter) Notebooks и са интерактивни среди за изпълнение, съставени от клетки или група клетки. Всяка клетка може да съдържа както данни, така и изпълним код. Моделът за проектиране на Model View Controller (MVC) (например widgets в Vokeh) широко приложимо и за почти всички приложения с потребителски интерфейс.

## 5.4 Софтуерна среда

Уеб базираните инструменти се съхраняват и изчисляват от страна на сървъра. Достъпът се осигурява през уеб браузъри. Средата позволява паралелно на множество потребители да провеждат независимо свои собствени експерименти. При необходимост могат да се активират допълнителни ограничения за сигурност за потребителски права, нива на достъп и т.н. Създадените уеб базирани инструменти могат да се използват и като софтуер като услуга в облачни изчислителни платформи предоставящи предимства като мащабируемост, достъпност и балансиране на натоварването.

Библиотеката Vokeh е избрана и използвана, тъй като нейната широка гама от widgets (контроли), инструменти за графика и UI елементи управлявани със събития, различни състояния, позволява да се работват интерактивни уеб приложения, които свързват събития от потребителски интерфейс с изпълнявания код на Python.

## 5.5 Източници на данни

Източниците на данни се предоставят като набори от данни/колекции, съхранявани в различни формати: CSV, XLSX, JSON. За някои набори от данни е предоставен API интерфейс / REST. Серийните данни са необходими за обработка в различен ред на колоните и се използват различни етикети. Има източници на данни, съхраняващи данни в редове вместо в колони. Също така се използват различни формати на дати. Всичко това изисква допълнителна обработка на данни.

## 5.6 Етапи на обработка на данни

- Извличане на данни – използване на програмен интерфейс за свързване и извличане на набори от данни за всяка отделна държава или за достатъчно големи административни области.
- Обработка на данни – трансформиране в единен сериен формат за входни данни за разработените инструменти TVBG и ATVBG.
- Проверка на данните – проверка на записите за пропуски/липсващи непоследователни данни
- Изчистване на данни – / Премахване или коригиране на липса на данни, ако е възможно се използва изглаждане и усредняване на данните.
- Генериране на данни – за да се намерят най-добрите сценарии за прогнозиране, е необходимо да се стартират модели на много итерации от порядъка на хиляди пъти, което е тежка изчислителна задача и на този етап се произведат значително количество данни.

## 5.7 Входни данни

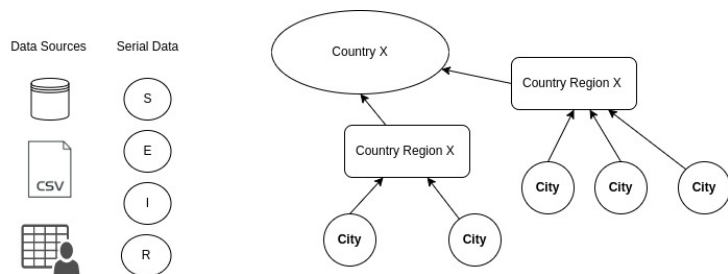
Форматът CSV е избран като основен формат на данни за създадите уеб базирани инструменти. Това е широко използван формат за серийни данни / времеви редове. При необходимост форматът може лесно да бъде променен на всеки друг подходящ формат за серийни данни.

Основните входни данни са ежедневно актуализирани времеви серии от:

- S – Податливи на заразяване
- E – Контактували със заразени
- I – Инфектирани, заразени
- R – Отстранени (излекувани или починали)
- ICU – брой случаи в интензивни отделения - използва се в ATVBG-SEIR

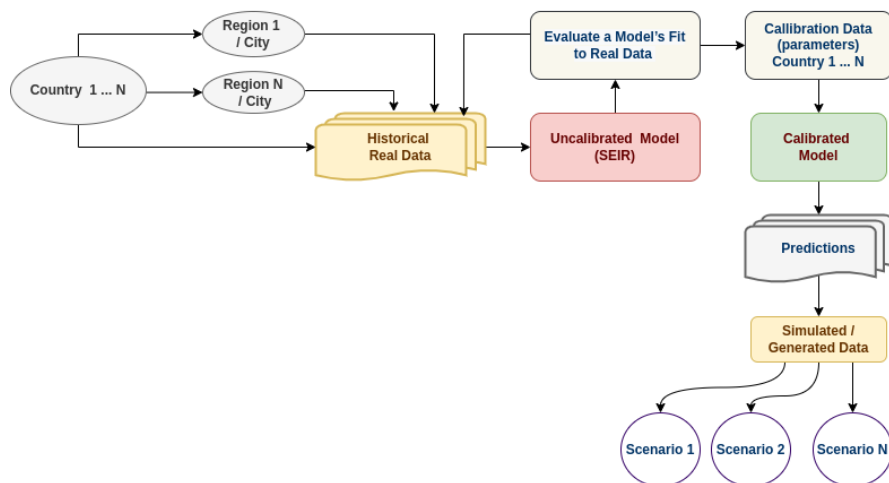
Параметрите на модела се определят от потребителя или/и чрез напасване на SEIR модела с реалните данни. За по-подробно описание вижте главата за SEIR модели. Използвани са NumPy и Pandas - специализирани библиотеки на Python за съхранение и манипулиране на серии от данни, многомерни данни. Pandas предоставя много различни IO конектори към SQL, BigQuery, CSV, HDF5, JSON и др.

Ако е необходимо, данните могат да бъдат мащабирани на различни нива. Зависи от източниците на данни и размера на регионалните единици. За страни с голяма територия и население е разумно да се направи регионално моделиране на данните като се прилагат разработените инструменти за прогнозни сценарии за отделните региони / области.



## 5.8 Основни стъпки процеси в софтуерни инструменти

Следващата фигура показва основните стъпки на изпълнение на изградените уеб инструменти. Възможно е паралелно изпълнение на отделните модели както и на ниво за всяка държава в независими изчислителни процеси.





## 5.9 TVBG-SEIR Tool уеб интерфейс

### 5.9.1 Програмни входни контролни параметри и Widgets

Country

Third restriction Date  ( The number of days between Today and Third restr. Date )

Coef1 for Transm. rate  ( Weaker Level <====> Stronger Level of Measures )

Coef2 for Removal rate  ( Weaker Level <====> Stronger Level of Measures )

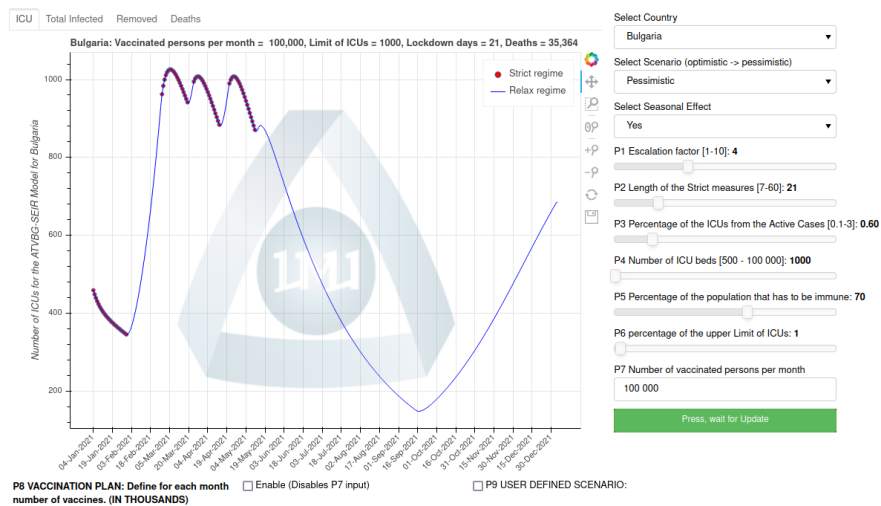
Coef11 relaxing after 3rd date  ( From Same Level ==>> To Weaker Level )

Coef22 relaxing after 3rd date  ( From Same Level ==>> To Weaker Level )

## 5.10 ATVBG-SEIR Tool веб интерфейс

### 5.10.1 Програмни входни контролни параметри и Widgets

### 5.10.2 Program input control Widgets



## Глава 6

# Приноси, заключения, бъдещи допълнителни направления

Основните приноси на настоящата работа са следните:

1. Изграждането на нови модели от тип SEIR, които са с променливи във времето параметри, базирани на сплайни. Дискретните модели са интуитивни и са лесни за разбиране за хора без опит в непрекъснати динамични системи.
2. Първият модел TVBG-SEIR се основава на сплайн модели за параметрите на SEIR (на скоростта на предаване и премахване), генерира сценарии за краткосрочно прогнозиране с двумесечен времеви хоризонт; има набор от параметри (за отпускане и затягане на мерките за ограничаване на Covid-19), които могат да бъдат променяни, за да се получат различни сценарии.
3. Вторият модел ATVBG-SEIR се различава съществено от първия, генерира сценарии за дългосрочно прогнозиране и по този начин позволява генерирането на прогнози в хоризонта за около няколко години. Той разполага с набор от параметри, които са специфични за всяка страна: ескалация, продължителност на периодите на блокиране, процент интензивни отделения, лимит на интен-

зивни отделения, два вида планове за ваксинация и няколко технически параметъра.

4. Ние разработихме два уеб базирани инструмента (софтуер като уеб услуга).
5. Първият инструмент се основава на модела TVBG-SEIR и реализира интерактивна средна стойност (визуализация на сценариите и кривите на моделите SEIR) за генериране на краткосрочни прогнози/прогнози, чрез уеб контроли за интерактивното изменение на параметрите. Той се реализира посредством Jupyter Notebook и софтуерен пакет Vokeh (Версия 2.3) за общ клиентски сървър.
6. Вторият разработен уеб-базиран инструмент използва модела ATVBG-SEIR и се занимава с генериране на дългосрочни сценарии/прогнози за прогнозиране. Тези сценарии включват персонализирани планове за ваксинация, както и подходящ модел на сезонния ефект. Има опция, позволяваща на потребителя да променя голямо разнообразие от модели (и параметри на модела) по интерактивен начин.
7. Ние сме предоставили кратко описание на софтуерното решение, базирано на Jupyter Notebook и софтуерен пакет Vokeh.
8. Ние приложихме нашите методи и модели главно към данни за България, но предоставяме и някои примери за прилагане на модели към данни за различни страни, като Германия, Австрия, Италия, Великобритания, САЩ.

**Основните изводи** на настоящата работа са:

1. Отделните модели на семейството SIR/SEIR са много гъвкав подход, който е лесен за адаптиране за анализ на данни, свързани с пандемията на Covid-19. По-специално, може да се моделира успешно въвеждането на мерки за задържане с интензивност, променяща се във времето, посредством сплайнови коефициенти в обичайната система на диференциални или диференчни уравнения.
2. Моделите тип SEIR са гъвкави и позволяват директно включване на информация за различни планове за ваксинация, поради което те предоставят адекватно моделиране на цялата картина, чрез

моделиране на комбинацията от мерки за ограничаване и ваксинации.

3. Моделът TVBG-SEIR представлява полезно средство за краткосрочно прогнозиране и може да се използва от здравни органи за краткосрочно планиране на мерки за ограничаване. Уеб базираният инструмент, изграден на основата на модела, предоставя незабавна поддръжка за планиране на политиките в динамичната пандемична ситуация.
4. Моделът ATVBG-SEIR представлява полезно средство за дългосрочно прогнозиране и може да се използва за вземане на решения в дългосрочната глобална оценка на жертвите, причинени от пандемията, както и за оценка на политиките за ваксинация.

Като **бъдещи насоки** бихме посочили изграждането на нови модели за описание на динамиката на разпространението на болестта от нови варианти на Covid-19 като Delta. Предизвикателство е също така да се моделира "смесената динамика където два варианта са доминиращи например алфа и делта варианти.

# Библиография

- [1] Last, J., 2001. *A Dictionary of Epidemiology, fourth ed.* Oxford University Press, New York.
- [2] Piroth, L., et al. (2020). *Comparison of the characteristics, morbidity, and mortality of COVID-19 and seasonal influenza: a nationwide, population-based retrospective cohort study*, The Lancet, [https://doi.org/10.1016/S2213-2600\(20\)30527-0](https://doi.org/10.1016/S2213-2600(20)30527-0)
- [3] Adam, D., *Special report: The simulations driving the world's response to COVID-19*, 03 April 2020, Nature, <https://www.nature.com/articles/d41586-020-01003-6>
- [4] Keeling, M. J. and Rohani, P., (2008), *Modelling Infectious Diseases*, Princeton University Press.
- [5] Straif-Bourgeois S., Ratard R., Kretzschmar M. *Infectious Disease Epidemiology*. Handbook of Epidemiology. 2014:2041–119. doi: 10.1007/978-0-387-09834-0\_34. PMID: PMC7176237.
- [6] Editorial, *Theorists and experimentalists must join forces*. Nat. Comput. Sci. 1, 299 (2021). <https://doi.org/10.1038/s43588-021-00082-3>
- [7] Lauer S. A., Grantz, K. H., Qifang Bi, Jones F. K., Qulu Zheng, Meredith H. R., Azman A. S., Reich N. G., Lessler J. (2020). *The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application*, Annals of Internal Medicine, Vol. 172, No. 9, May 2020, pp. 577-582, doi:10.7326/M20-0504.

- [8] Keeling, M. J. and Rohani, P., (2008). *Modelling Infectious Diseases*, Princeton University Press.
- [9] Lipsitch M., Cohen T., Cooper B., Robins J. M., Stefan Ma, Lyn James, Gowri Gopalakrishna, Suok Kai Chew, Chorh Chuan Tan, Matthew H. Samore, David Fisman, and Megan Murray, (2003), *Transmission Dynamics and Control of Severe Acute Respiratory Syndrome*, Science, 20 June 2003, Vol. 300, pp. 1966-1970, doi:10.1126/science.1086616
- [10] Wallinga, J. and Lipsitch M., (2007), *How generation intervals shape the relationship between growth rates and reproductive numbers*, Proceedings: Biological Sciences, Vol. 274, No. 1609, (Feb. 22, 2007), pp. 599-60
- [11] Heffernan J. M., Smith R. J. and Wahl L. M., (2005). *Perspectives on the basic reproductive ratio*, J. R. Soc. Interface (2005) 2, Published online (7 June 2005), pp. 281–293, doi:10.1098/rsif.2005.0042
- [12] Kounchev O., Simeonov G., Kuncheva Zh. (2021), *The TVBG-SEIR spline model for analysis of COVID-19 spread, and a Tool for prediction scenarios*, *Biomath*, Vol 10, No 1 (2021), DOI: <http://dx.doi.org/10.11145/j.biomath.2021.03.087>; previous versions at <https://arxiv.org/abs/2004.11338>
- [13] McAloon C, Collins Á, Hunt K, et al. *Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research*. *BMJ Open* 2020;10:e039652. doi:10.1136/bmjopen-2020-039652
- [14] Anderson, R. M., and May, R. M. (1991). *Infectious Diseases of Humans*. Oxford, UK: Oxford University Press.
- [15] Biswas, H. A. , L. T. Paiva and MdR de Pinho, (2014). *A SEIR model for control of infectious diseases with constraints*, *Mathematical Biosciences and Engineering*, Volume 11, Number 4, August 2014, doi: 10.3934/mbe.2014.11.761
- [16] Brauer, F., Pauline van den Driessche, Jianhong Wu (Eds.), (2008). *Mathematical Epidemiology*, Berlin: Springer, 2008, eBook ISBN 978-3-540-78911-6, <https://doi.org/10.1007/978-3-540-78911-6>

- [17] Brauer, F., Castillo-Chavez, C., and Castillo-Chavez, C. (2012). *Mathematical models in population biology and epidemiology*, (Vol. 2, p. 508), New York: Springer. eBook ISBN 978-1-4614-1686-9, <https://doi.org/10.1007/978-1-4614-1686-9>
- [18] Hethcote, H., (2000). *The Mathematics of Infectious Diseases*, SIAM Review, Dec., 2000, Vol. 42, No. 4 (Dec., 2000), pp. 599-653, <https://doi.org/10.1137/S0036144500371907>
- [19] Katok, A., and Hasselblatt, B. (1997). *Introduction to the modern theory of dynamical systems*, (No. 54). Cambridge University Press. ISBN 0-521-57557-5
- [20] Keeling M. J., and Rohani, P. (2008). *Modeling Infectious Diseases, in humans and animals*, Oxford Univ. Press, Princeton and Oxford, ISBN 978-0-691-11617-4.
- [21] Kounchev, O., Simeonov, G., Kuncheva, Zh., *The TVBG-SEIR spline model for analysis of COVID-19 spread, and a Tool for prediction scenarios*, <https://arxiv.org/abs/2004.11338>;
- [22] Kounchev, O., Simeonov, G. and Kuncheva, Zh., *How Long 'Lockdowns' Are Needed to End the COVID-19 Epidemic in a Single Country, with or without Vaccinations*, (January 14, 2021). Available at SSRN: <https://ssrn.com/abstract=3766521> or <http://dx.doi.org/10.2139/ssrn.3766521>
- [23] Kounchev, O., Simeonov, G., Kuncheva, Zh., *Estimation of the Duration of Covid-19 Epidemic in a Single Country, with or without Vaccinations. The Case of Bulgaria and Germany*, Comptes rendus de l'Academie bulgare des Sciences, Vol 74, No5, pp. 677-686.
- [24] Neilan, R. M. and Lenhart, S. (2010). *An introduction to optimal control with an application in disease modeling*, DIMACS Series in Discrete Mathematics, 75 (2010), 67-81.
- [25] Altizer, S., Dobson, A., Hosseini, P., Hudson, P., Pascual, M., Rohani, P.: Seasonality and the dynamics of infectious diseases. *Ecol. Lett.* 9, 467–484 (2006)
- [26] Grassly, N.C., Fraser, C.: Seasonal infectious disease epidemiology. *Proc. R. Soc. Lond. B Biol. Sci.* 273(1600), 2541–50 (2006)



- [27] Audi A, AlIbrahim M, Kaddoura M, Hijazi G, Yassine HM and Zaraket H (2020) Seasonality of Respiratory Viral Infections: Will COVID-19 Follow Suit? *Front. Public Health* 8:567184. doi: 10.3389/fpubh.2020.567184
- [28] David J. D. Earn, Pejman Rohani, Benjamin M. Bolker and Bryan T. Grenfell, A Simple Model for Complex Dynamical Transitions in Epidemics *Source: Science, New Series, Vol. 287, No. 5453 (Jan. 28, 2000)*, pp. 667-670
- [29] Papst Irena and Earn David J. D. 2019, Invariant predictions of epidemic patterns from radically different forms of seasonal forcing, *J. R. Soc. Interface*.162019020220190202.
- [30] Stanchev, P., Ancheva, H., Pavlov, R., Simeonov, G., The eleventh national information day: Open science, Open Data, Open Access, Bulgarian Open Science Cloud, DiPP2020, 2020, 2020-September, pp. 275–281.
- [31] Srebrov B., Kounchev O., Simeonov G., Book Chapter 19 - Big Data for the Magnetic Field Variations in Solar-Terrestrial Physics and Their Wavelet Analysis, January 2020, DOI:10.1016/B978-0-12-819154- 5.00031-X, In BOOK: Knowledge Discovery in Big Data from Astronomy and Earth Observation (pp. 347-370), ISBN:978-0-12-819154-5
- [32] Pyne, S., Vullikanti, A., and Marathe, M. (2015). *Big Data Applications in Health Sciences and Epidemiology*. *Handbook of Statistics*, 33, 171 - 202. Elsevier, <https://doi.org/10.1016/B978-0-444-63492-4.00008-3>
- [33] Liu, J. and Sh. Xia, *Computational Epidemiology: From Disease Transmission Modeling to Vaccination Decision Making*, Springer, 2020.
- [34] Kuhl, E., *Computational Epidemiology. Data-Driven Modeling of COVID-19*, Springer Verlag, 2021
- [35] Siettos, C. I., and Russo, L. (2013). Mathematical modeling of infectious disease dynamics. *Virulence*, 4(4), 295–306. <https://doi.org/10.4161/viru.24041>