

Retrospective and prospective data-oriented evaluation of the influence of governmental measures on the spread of COVID-19

BASISGEGEVENS (voorpagina)

NAAM VAN DE HOOFDAANVRAGER:

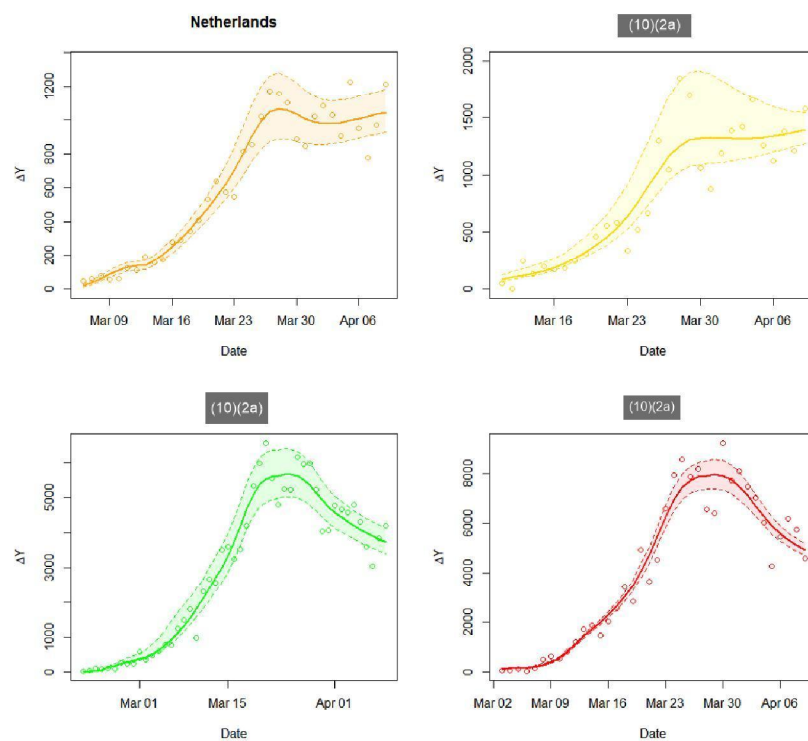
(10)(2e)

ORGANISATIE:

Eindhoven University of Technology

PROJECTTITEL:

Retrospective and prospective data-oriented evaluation of the influence of governmental measures on the spread of COVID-19



DATATEWARD:

Ik betrek een datateward bij mijn project:

Naam: (10)(2e)

Instituut: Eindhoven University of Technology

E-mail: (10)(2e)@tue.nl

Was aanwezig bij de webinar: Ja Nee

Front page

1. PROBLEEMSTELLING EN DOELSTELLING(EN):

1.1. Background: Infectious disease surveillance methods [1] and epidemic disease models [2] have been described in literature. They can detect, monitor and describe the spread of a virus among a population. Surveillance methods can be broadly categorized into two classes: test-based and model-based methods, which have been employed by various European institutes for infectious disease control, including the RIVM (The Netherlands). Epidemic disease models describe how inhabitants transit between stages, starting from being susceptible to an infection of the virus to being exposed or infectious, and then transition to different outcomes (e.g. hospitalized, death, or recovered). Both approaches describe (change in) virus spread at an aggregated or population level and they are essential to our understanding of public health and they can be used to quantify the influence of governmental interventions on COVID-19 [3].

1.2. Limitations: The outbreak of COVID-19 has revealed some critical limitations. Firstly, the daily data on confirmed number of infections, hospitalizations, and deaths are subjected to selection bias, since individuals are being collected by purposive sampling. Subpopulations at risk (e.g., health care personnel) are being oversampled, contacts of known cases are being traced and tested, and governments constantly change their testing policies during the pandemic. A second reason is the lack-of-knowledge on virus characteristics. Incubation periods, patient recovery times, immunity after infection are being determined with little precision, which affects the calculation of monitoring statistics (like) and *complicates the choice of most appropriate analysis techniques*. Thirdly, surveillance methods are not fully equipped to address the current unknown dynamics of the COVID-19. Test-based algorithms often presume a static disease process and ignore spatial and endemic heterogeneities across and within countries. Model-based algorithms have the potential of addressing such limitations, but so far they considered generic statistical models (e.g., generalized linear mixed models), instead of more specific epidemic disease models or growth curve models. Furthermore, model-based detection algorithms, are sensitive to the unknown assumptions on virus characteristics. Finally, monitoring statistics like *that are useful for controlling the spread of the virus, but may be less suitable for monitoring changes when measures are being relaxed*. Thus there is a need to address the limitations in the current state of the pandemic to be able to evaluate the consequences of eliminating governmental measures.

1.3. Solution: A common scientific approach is to consider an existing infectious disease method and start improving or enhancing this method to accommodate all limitations. This leads to more complex epidemic disease methods, which are most likely more sensitive to (virus characteristic) assumptions. Although this is a solid scientific approach, we rather propose a data-oriented approach, where we use multiple data analyses methods simultaneously to capture and understand the information present in the collected data and to signal and explore possible spatial-temporal changes in the spread of the virus. We will implement these methods on the daily numbers of confirmed infections, hospitalizations, and deaths from different countries, areas and sub-populations, and complement it with other (publically) available data sources (e.g. telephony data). Our data-oriented approach may be viewed as a dashboard with a multitude of different signals that will tell us when, where, and what changes are happening in the spread of the virus (in line with goals of the Dutch government [4]) and how it relates to governmental measures both retrospectively and prospectively.

1.4. Goals: We will develop, validate, and apply statistical quantities using various infectious disease models that are sensitive to signalling changes in temporal-spatial dynamics of virus spread, with the goal to quickly identify the results of (implementing and) eliminating governmental measures in (sub)populations both locally and nationally. Our quantities will extend and complement current quantities that are in use at RIVM, to help RIVM to quickly track changes and maintain control of the spread of the virus.

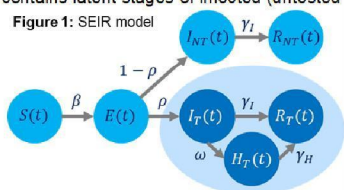
2. PLAN VAN AANPAK:

Our main goal is to develop and implement a dashboard with dynamical indicators of change (DIOCs) that can quickly detect spatial-temporal changes in the virus spread. A DIOC is a statistical quantity, like , *that is calculated from* infectious data. It is dynamic since it is daily updated and it is possibly altered by new information becoming available during the pandemic. We will develop and validate DIOCs that focus on the growth rate and effective contact rate, because governments can only implement physical constraints.

Our approach is composed of two parts. The first part will study the behavior of various novel DIOCs, which are derived from epidemic disease models and sigmoidal growth curve analysis, using several simulation scenarios. Setting criteria for our DIOCs, to signal changes, will be based on ideas from the field of disease detection and surveillance [1]. The second part will study the behavior of all DIOCs together, to understand their coherence, to sharpen criteria, and to build the dashboard with software package [R], suitable for RIVM. Both parts will use DIOCs retrospectively and prospectively to detect changes in the virus spread related to (implementation and elimination of) governmental measures. Our approaches will validate DIOCs in practice and take into account changes in test policies, to be able to isolate changes due to the virus transmission.

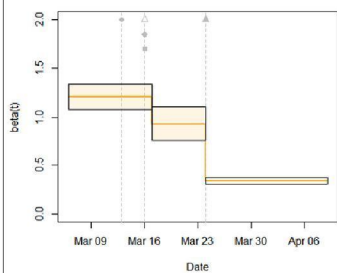
PART I.1. DIOCs for epidemic disease modeling: Recently, we have developed a data-oriented

approach, based on the SEIR (Susceptible-Exposed-Infectious-Removed) model (Figure 1) to create a non-increasing daily contact-rate profile (Figure 2). Our SEIR model contains latent stages of infected (untested) and removed (untested) individuals at day . Susceptible () and exposed () individuals at day aren't observed either. Exposed individuals are in a transition phase, where they have been in contact with the virus but they are not yet contagious. Incubation time distributions in SEIR models are typically assumed exponential, but literature on COVID-19 suggests a Weibull distribution [5], which we have implemented. Individuals are contagious in the infected stage, but we assumed that individuals are not contagious in hospitalized () and removed stages. We implemented an exponential recovery distribution.



Available information for estimation of our SEIR model comes from observed stages (,) on infected and hospitalized individuals, that were tested. Since we deal with a non-standard SEIR model, we developed an iterative Poisson regression analysis applied to the newly confirmed infected individuals per day. We estimated a piece-wise constant daily effective-contact rate , which would drop in value when observed data suggests this. Figure 2 shows the estimated daily contact rates (colored horizontal lines) for the Netherlands with 95% confidence bands (horizontal black lines). Vertical gray lines indicate the moments of governmental interventions (triangle = lockdown; square = closing schools; circle = banning events; diamond = closing restaurants). The moments of interventions nicely overlap with the change points in daily contact rates, indicating that our approach is capable of detecting changes. The reproduction number is less suitable here, since it is a more complicated function of the transition rates , , and 's.

Figure 2: Daily growth-rate profile of the Netherlands



We will use our daily effective-contact rate estimate as a DIOC to detect changes in virus spread by relaxing the monotonicity restriction on the profile, in particular looking for increases that may warn us about losing control over the pandemic when measures are being relaxed or when we see that behavior of individuals change (in combination with telephony data). When more and more information becomes available (incubation times, recovery times, etc.) we will incorporate this information in the SEIR model and obtain a more accurate DIOCs. However, the daily effective-contact

rate has not been studied properly yet to be used as DIOC. We need to study in more detail influences of our assumptions (on distributions and contagious stages), sensitivity in detecting changes, and potential delays in detecting changes. Furthermore, we need to study how our approach works in smaller areas (provinces) and how we can include subpopulation characteristics in the daily contact rate (e.g. age-group).

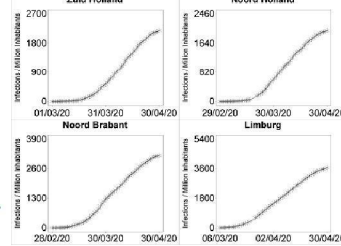
PART I.2. DIOCs for sigmoidal growth curves: Population dynamics and biological growth have been studied with many different sigmoidal curves [6], with Verhulst's logistic growth curve being the earliest and easiest contribution. Some sigmoidal curves connect directly to epidemic disease models (e.g., Verhulst's model describes a Susceptible-Infected model), but in general they have a more limited scope. They describe change over time for a particular outcome (e.g., infections, hospitalizations, or deaths), , without making any further assumptions on transition between stages. A large family of sigmoidal growth curves is described by its growth rate (i.e. the derivative of growth at day) having the following form:

$$(1) \quad \beta(t) = \frac{r}{1 + e^{-r(t-t_0)}}$$

with an unknown parameter for the contact rate, , unknown growth parameters, and the unknown maximum outcome. Verhulst's model is a special case with . Estimation of the full system in (1) is difficult when and should be both estimated (numerical issues), thus we restrict to models with . Table 1 contains parameter estimates of model (1) on the number of infections for several countries (; data April 30, 2020), showing model (1) can be fitted to data of COVID-19 in different countries. Large differences in expected maxima and growth rates are visible across countries, showing sensitivity in detecting changes in spread of the virus. Figure 3 shows an excellent fit of model (1) to four Dutch provinces.

Table 1	r	$\log_2(\beta)$	M
(10)(2a)	0.736 [0.726; 0.747]	0.479 [0.384; 0.574]	55306 [54889; 55724]
	0.715 [0.705; 0.725]	0.546 [0.459; 0.634]	79491 [77758; 81224]
	0.579 [0.550; 0.608]	1.017 [0.805; 1.229]	12045 [11580; 12510]
	0.782 [0.776; 0.788]	0.349 [0.289; 0.409]	135858 [135574; 136142]
	0.720 [0.715; 0.724]	1.101 [1.052; 1.150]	170303 [170020; 170586]
	0.676 [0.668; 0.685]	1.038 [0.953; 1.123]	107896 [107283; 108509]
	0.651 [0.647; 0.656]	1.632 [1.582; 1.682]	228775 [228093; 229456]
	0.715 [0.702; 0.727]	0.586 [0.475; 0.696]	44519 [44154; 44884]
	0.609 [0.597; 0.622]	1.363 [1.259; 1.467]	10794 [10783; 10804]
	0.684 [0.664; 0.704]	0.286 [0.128; 0.444]	41860 [38636; 45084]
	0.664 [0.662; 0.666]	2.133 [2.107; 2.158]	1491805 [1486473; 1497138]
	0.746 [0.741; 0.752]	0.635 [0.576; 0.693]	225039 [223178; 226900]

The goal is to formulate DIOCs for model (1), to determine spatial-temporal growth changes in the number of infections, hospitalizations, and deaths. This research is in line with test-based algorithms in disease surveillance, but DIOCs for model (1) are novel. We have been using on COVID-19's number of infections to determine the **turning point** of the virus spread (which we reported to be March 31 for the Netherlands), with *the estimated maximum at day*. It is an important statistic at the latest phase of the virus spread, when it gets close to one. If it starts deviating from one again, we have a signal that the virus spread is increasing again. We will study its behavior under system (1) and under epidemic disease models to know how to use it for monitoring spatial-temporal changes. We will also use the time-varying growth rate estimate *at day* to detect changes in growth, since it may abruptly change in value in case of sudden changes in virus spread (as we have seen with Verhulst model). We will conduct simulation studies to quantify criteria to flag out-of-control situations and to signal violations in the expected growth patterns.



PART II. Evaluating DIOCs jointly and implementation: DIOCs may not signal change by itself, but the combined DIOCs may show a change in virus spread. This requires knowledge on the dependency structure between DIOCs. Note that we have several different DIOCs applied to different data (infections, hospitalizations, and deaths) and to subpopulations and regions (e.g., provinces). We will use copulas to determine this dependency structure. A copula function is independent of the marginal distributions of the separate DIOCs (studied in Part I), and it can be studied after we have created criteria for single DIOCs. We will study common copula families [7] and investigate which family best describes the dependency structure. We will use Bayesian selection and goodness-of-fit methods to compare different copula families [8]. The selected copula is then used to create a multi-dimensional out-of-control region for detecting changes in growth and virus spread. The same simulation studies conducted earlier are used to evaluate performance of the full system of DIOCs. Applying it retrospectively and prospectively on COVID-19, we will demonstrate how it may perform in practice.

We have experience in creating [R] codes for third-parties. For instance, we are building a statistical web-based analysis software package for microbiological analysis in pharmaceutical industry. Our codes will be made publicly available and will be developed together with RIVM.

3. HAALBAARHEID VAN HET PROJECT:

3.1. Time schedule: We request funding for 18 months. Part I will be conducted in the first 6 months with 2 fte post-docs and 0.5 fte (0.3 UD + 0.2 HGL) supervision (to compensate for teaching obligations in Q1, 2020-2021) at TU/e to be able to use our DIOCs real-time on data of COVID-19 and help RIVM detect possible changes after the summer. Part 2 will take 1 year using 1 fte post-doc and 0.2 fte supervision (0.1 UD + 0.1 HGL) at TU/e for research, software codes for the dashboard, and retrospective validation on COVID-19 data. RIVM will contribute in-kind to this project with 0.6 fte (research, discussions, meetings, and implementation).

3.2. Motivation feasibility: Researchers of TU/e and RIVM have broad experience in process monitoring and surveillance and in biological growth modeling. Prof. Dr. Van den Heuvel has been working in the field of statistics for epidemiological research and has a large group of statisticians working on the topics proposed in this proposal. Prof. Dr. Wallinga is a renowned researcher in epidemic disease modeling and has support of a large group of researchers at RIVM. Collaboration was initiated by RIVM to join forces during the crisis.

4. RELEVANTIE VOOR DE PRAKTIJK:

Our proposal is a direct solution for the theme "Onderzoek naar de effectiviteit en impact van maatregelen/strategieën in respons op de coronacrisis" of the third focus area "Maatschappelijke dynamiek". It supports the initiative of the Dutch government to build a dashboard that can track and detect changes in the virus spread. Part I of our proposal is urgent and cannot be executed later and it is a direct extension of our recently conducted research projects on COVID-19.

5. DEELNAME VAN DE STAKEHOLDER(S) (e.g. patiënten, zorgprofessionals, etc.):

Researchers from TU/e and RIVM (in-kind contribution of 0.6 fte/year over 18 months) will jointly execute the proposal. RIVM will share their data with TU/e to enable the execution of the research.

6. LITERAATUURREFERENTIES (optioneel):

[1] Unkel *et al.*, Statistical methods for the prospective detection of infectious disease outbreaks: a review, *Journal of the Royal Statistical Society A*, 2012, 175(1):49-82 [2] Held *et al.*, Handbook of infectious disease data analysis. Boca Raton, FL: CRC Press, 2020. [3] Pan *et al.*, Association of Public Health

Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA* 2020; 323: 1915–23. [4] <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2020/05/20/kamerbrief-stand-van-zaken-covid-19/Kamerbrief+update+stand+van+zaken+Covid+19.pdf> [5] Backer *et al.*, The incubation period of 2019-nCoV infections among travellers from Wuhan, China. *Eurosurveillance* 2020; 25: 2000062. [6] Tsoularis *et al.*, Analysis of logistic growth models, *Mathematical Biosciences*, 2002, 179: 21-55. [7] Nelsen RB, *An introduction to copula's*, Springer, New York, 2006. [8] Genest *et al.*, Goodness-of-fit for copulas: A review and a power study, *Insurance Mathematics and Economics*, 2009, 44(2): 199-213.