

A multiple criteria approach for constructing a pandemic impact assessment composite indicator: The case of COVID-19 in Portugal

José Rui FIGUEIRA^{a,*}, Henrique M. OLIVEIRA^b, Ana Paula SERRO^c, Rogério COLAÇO^d,
Filipe FROES^e, Carlos ROBALO CORDEIRO^e, António DINIZ^e, Miguel GUIMARÃES^e

^a*CEGIST, Instituto Superior Técnico, Universidade de Lisboa, Portugal*

^b*CAMGSD, Instituto Superior Técnico, Universidade de Lisboa, Portugal*

^c*CQE, Instituto Superior Técnico, Universidade de Lisboa, Portugal*

^d*IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Portugal*

^e*GCOM, Ordem dos Médicos, Portugal*

Abstract

The COVID-19 pandemic has caused impressive damages and disruptions in social, economic, and health systems (among others), and posed unprecedented challenges to public health and policy/decision-makers concerning the design and implementation of measures to mitigate its strong negative impacts. The Portuguese health authorities are currently using some decision analysis-like techniques to assess the impact of this pandemic and implementing measures for each county, region, or the whole country. Such decision tools led to some criticism and many stakeholders asked for novel approaches, in particular those having in consideration dynamical changes in the pandemic behavior arising, e.g., from new virus variants or vaccines. A multidisciplinary team formed by researchers of the COVID-19 Committee of Instituto Superior Técnico at Universidade de Lisboa (CCIST analysts team) and medical doctors from the Crisis Office of the Portuguese Medical Association (GCOM experts team) gathered efforts and worked together in order to propose a new tool to help politicians and decision-makers in the combat of the pandemic. This paper presents the main steps and elements, which led to the construction of a pandemic impact assessment composite indicator, applied to the particular case of COVID-19 in Portugal. A multiple criteria approach based on an additive multi-attribute value theory (MAVT) aggregation model was used to construct the pandemic assessment composite indicator (PACI). The parameters of the additive model were built through a sociotechnical co-constructive interactive process between CCIST and GCOM team members. The deck of cards method was the technical tool adopted to help in building the value functions and the assessment of the criteria weights. The final tool was presented in a press conference and had a strong impact on Portuguese media, as well as on the main health decision-making stakeholders in the country.

Keywords: Multiple criteria analysis, Composite indicator, Multi-attribute value theory (MAVT), Robustness and validation analyses, Deck of cards method, Multidisciplinary team work.

*Corresponding author at: CEGIST, Instituto Superior Técnico, Av. Rovisco Pais, 1, 1049-001 Lisboa, Portugal.
Email address: figueira@tecnico.ulisboa.pt (José Rui FIGUEIRA)

Contents

1	Introduction	3
2	Concepts, definitions, and notation	6
2.1	Basic data	6
2.2	MAVT additive model	6
2.3	Chromatic ordinal classification model	7
3	Modeling aspects	8
3.1	Criteria model	8
3.2	Parameters of the aggregation model and the chromatic system	10
3.2.1	Value functions (interval scales)	10
3.2.2	Weighting coefficients (ratio scales)	16
3.3	Illustrative example	18
3.4	Chromatic classification system (ordinal scales)	19
3.5	Graphical model for visualization and communication	20
4	Results, robustness analyses, and recommendations	21
4.1	Implementation issues and verification tests	21
4.2	Results	21
4.2.1	Pandemic evolution	22
4.2.2	Cumulative contribution of each criterion	23
4.2.3	The impact of vaccination	24
4.3	Robustness analyses	25
4.3.1	Simulation on the weights	25
4.3.2	Exact robustness analysis	25
4.4	Validation, adjustments, and comments	27
4.5	Measures and recommendations	28
5	Conclusions	30
	Acknowledgements	31
	References	32
	Appendix	35

1. Introduction

A pandemic causes impressive damages and disruptions in social, economic, and health systems (among others) and has strong implications in the life of populations worldwide, leading not only to very serious physical and mental health problems, but also to poverty and hunger. The COVID-19 is the most recent pandemic, which is posing unprecedented challenges to public health and policy/decision-makers, namely in design and implementation of measures to mitigate its negative impacts.

In Portugal a team from the COVID-19 Crisis Office of Portuguese Medical Association (GCOM experts team) and a team from the COVID-19 Committee of Instituto Superior Técnico (CCIST analysts team), did several efforts to help in fighting against COVID-19 and mitigate its negative impact on the life of people. After a first step in which they acted separately, then joined efforts and created synergies to better achieve these goals.

A tool called “Risk Matrix” (RM), later changed to accommodate some *ad hoc* rules, has been used for the Portuguese health authorities to help in the decision process. Based on this tool, a risk state (with a colour associated) is assigned to each county. The RM has been subject to some criticism, mainly because it is incomplete and cannot give an adequate idea about the evolution of the pandemic in the country. The term “Risk Matrix” has a meaning different from the well-known decision aiding tool with the same name in the field of Decision Analysis. In addition, the term “matrix” has nothing to do with the mathematical object with the same name. Here, the RM is a two dimensional referential accompanied with a visual chromatic system (from light green to dark red), where the abscissa axis represents the raw values of the transmission rate ($R(t)$) and the ordinates axis the value of average incidence of new positive cases over the last seven days. In addition, two cut-off lines (one horizontal and one vertical) are used as criteria for separating the referential in four regions: being the southwest region the lowest risk region impact; the northeast region, the highest impact risk region; and, the other two regions (northwest and southeast), the regions with intermediate risk impact. The different territorial units of the country, namely the counties and regions, were colored with this system and some measures have been assigned to each colour.

The RM main drawbacks can be succinctly presented as follows:

1. Despite the usefulness and advantages of the visual chromatic system for communication purposes, it suffers from a major pitfall, which renders very difficult to see the evolution of the pandemic over the time-line (plotting each daily situation in the referential and linking all the successive points by a line leads to a very confusing evolution curve).
2. Despite the importance of the dimensions used in the referential (incidence and transmission), they are only a part of the problem. Both are related with the activity of the pandemic, but more dimensions should be considered, namely some related with the severity of the pandemic;
3. The impact of the risk seems not being appropriately modeled since moving from an $R(t) = 0.1$ to an $R(t) = 0.2$ as the same risk impact as moving from an $R(t) = 0.9$ to an $R(t) = 1.0$, which does not represent what the population, in general, feels about the impact of the transmission (the same reasoning can be applied to the incidence dimension).
4. There is no differentiation of the contributions of incidence and transmission for the overall risk impact, they count equally (this can be acceptable, but it is not always the case).

An improved RM containing more cut-off lines than the one used by the Portuguese health authorities, and that allows making a finer analysis of the situation, was previously proposed by the CCIST team, but it still continues to suffer from the same drawbacks as the original RM. In parallel the GCOM team also proposed an improved RM, different from the one used by CCIST team. Indeed, to overcome some of the drawbacks of the original RM, the GCOM team proposed to use the two-dimensional referential system, but considering several indicators in the abscissa axis and also several indicators in the ordinate axis. This was a first important step towards an MCDA-based indicator. The two sets of dimensions (called ahead pillars) of this new RM are the activity and the severity of the pandemic. Unfortunately, the way the activity and the severity indicators were considered was questionable, and despite this RM had led to a more complete and finer analysis of the problem it still represents the drawbacks (1), (3), and (4). This GCOM proposal was rendered public in the first week of June 2021. In the beginning of July 2021, the two teams (CCIST and GCOM), started joining efforts in order to propose the composite indicator presented in this paper. This new proposal had a strong impact in Portugal, especially in the *media* and health policy/decision-makers actors. In the end of July 2021, the RM of the Portuguese health authorities was changed to include some *ad hoc rules*, based on the severity aspects of the pandemic, as the one proposed in our pandemic assessment composite indicator (PACI).

What was missing in the proposed RM approaches? A more adequate system to characterize the state of pandemic impact and a system for recommending the most suitable measures to mitigate the pandemic impact. Therefore, the *main decision problem* we were faced was to build a state indicator of the pandemic impact for a given territorial unit (country, region, county, etc), with the purpose of assigning mitigating measures/recommendations to each state (going from the less to the most restrictive ones). As *fundamental observation*, it is important to follow the recent evolution of the pandemic impact for a better planning when a given territory unit moves from a given state to another one. In addition, it is extremely important to assess, in some way, the impact of the Portuguese *vaccination plan*. Each territorial unit is assessed on the daily basis, regarding a set of criteria (also called, in our case, indicators or dimensions) grouped in two main points of view: the pillar of activity and the pillar of the severity of the pandemic. In its very nature, this problem statement can be viewed as belonging to the field of multiple criteria decision aiding/analysis (MCDA). For more details, the reader can consult the books by Belton and Stewart (2002) and Roy (1996). The problem is known in the literature as an ordinal classification (or sorting) MCDA problem (Doumpos and Zopounidis, 2002; Zopounidis and Doumpos, 2002). There are several ways of building a state composite (aggregation) indicator (see El Gibari et al., 2019, for a recent survey in this topic). The main MCDA approaches for designing composite indicators are the following:

1. *Scoring based approaches*, as for example, multi-attribute utility/value theory (MAUT/MAVT) aggregation models (e.g., Dyer, 2016; Keeney and Raiffa, 1993; Tsoukiàs and Figueira, 2006), analytical hierarchy process (e.g., Saaty, 2016), fuzzy sets techniques (e.g., Dubois and Perny, 2016), and fuzzy measure based aggregation functions (Grabisch and Labreuche, 2016).
2. *Outranking approaches*, as for example, ELECTRE methods (Figueira et al., 2016), PROMETHEE methods, (Brans and De Smet, 2016), and other outranking techniques (Martel and Matarazzo, 2016).

3. *Rule based systems*, as for example, decision rule methods (Greco et al., 2016), and verbal decision analysis (Moshkovich et al., 2016).

According to the definition of our decision problem, outranking approaches and rule based systems are powerful MCDA techniques for ordinal classification and could be adequate tools to build a state composite indicator (ordinal scale). However, the need of analysing an evolution of the pandemic (see the fundamental observation stated before) requires the construction of a more richer scale (of a cardinal nature). It is true that outranking approaches and rule based systems can be adapted to produce such a cardinal scale (see, for example, Figueira et al. 2021), but this is a complex process, rather difficult to explain to the main actors and the public in general. Consequently, the most adequate approach for dealing with our problem is a scoring-based approach. Since the model must be enough simple for interaction and communication purposes with the experts and the public in general, without losing its adequacy to the reality, we focused our study only on MAVT methods, discarding thus other more complex scoring-based methods. After a more deeper analysis we finally decided to keep the simple additive MAVT approach. It was adequate to model our problem and render the communication easy. At this point, another question raised: in which way should we built the additive model? There are two ways of answering such a question:

1. Through a constructive learning approach (machine learning like approaches), as for example, UTA type methods (Siskos et al., 2016), or an adaptation of more sophisticated techniques as the GRIP method (Figueira et al., 2009) with representative functions.
2. Through a co-constructive sociotechnical interactive process between analysts and policy/decision-makers or experts using, for example, the classical MAVT method (Keeney and Raiffa, 1993), the MACBETH method (Bana e Costa et al., 2016), or the deck of cards method (Corrente et al., 2021).

In every co-constructive sociotechnical process the analyst must be familiar with the technicalities of the method. In addition, the policy/decision-makers or experts must understand the basic questions for assessing their judgments. The deck of cards method by Corrente et al. (2021) revealed to be an adequate tool. Its adequacy comes from some important aspects: time limitation to produce a meaningful indicator, easy to be understood by the experts, easy to communicate with the public in general, and easy to reproduce the calculations for a reader with an elementary background on mathematics.

In this study we explored the potential of MAVT theory through the deck of cards method (Corrente et al., 2021) in an innovative application: the construction of a cardinal impact assessment composite indicator of the pandemic, used both to observe the evolution of the pandemic and to form a state ordinal indicator with measures/recommendations associated with each state, with the purpose of its application to the case of COVID-19 in Portugal.

The paper is organized as follows. Section 2 introduces the basic mathematical background needed along the paper. Section 3 presents the main models used to perform this study (criteria model, aggregation model, and graphical model). Section 4 displays the results, derives, and discusses robust conclusions, and it also provides some recommendations. Finally, Section 5, presents the main conclusions and points out some avenues for future research.

2. Concepts, definitions, and notation

This section introduces the main concepts, definitions, and notation used along the paper. It comprises the criteria model basic data, the MAVT additive model, and the chromatic classification system.

2.1. Basic data

The basic data can be introduced as follows. Let $T = \{t_1, \dots, t_i, \dots, t_m\}$ denote a set of actions or *time periods* (in general, days) used for observing the pandemic state in a given territory unit (country, region, district, etc), and $G = \{g_1, \dots, g_j, \dots, g_n\}$ denote the set of relevant *criteria* (our problem dimensions or indicators) identified with the experts for assessing the actions or time periods. The *performance* $g_j(t_i) = x_{jt_i} \in E_j$ represents the impact level of activity or severity over the action or time period $t_i \in T$, according to criterion g_j , being E_j the (continuous or discrete) scale of this criterion, for $j = 1, \dots, n$. We will assume, without any loss of generality, that, for each criterion, the higher the performance level, the higher the impact on the pandemic. The set of criteria has been built according to some desirable properties (see Keeney, 1992).

2.2. MAVT additive model

The proposed model is a conjoint analysis model (see, for example, Bouyssou and Pirlot 2016), more specifically an additive MAVT model. The origins of this type of models dates back to 1969, with the seminal work by H. Raiffa, only published in 2016, in Tsoukiàs and Figueira (2006), with several comments from prominent researchers in the area. For more details about the additive model see Keeney and Raiffa (1993).

Let \succsim denote a comprehensive binary relation, over the actions in T , whose meaning is “impacts at least as much as”,. Thus, an action t' is considered to impact at least as much as an action t'' , denoted $t' \succsim t''$, if and only if, the overall value of t' , $v(t')$ is greater than or equal to the overall value of t'' , $v(t'')$, i.e., $v(t') \geq v(t'')$, where the overall value of each action is additively computed as follows:

$$v(t) = \sum_{j=1}^n w_j v_j(x_{jt}), \text{ for all } t \in T \quad (1)$$

in which w_j is the weight of criterion j , for $j = 1, \dots, n$, (assuming that $\sum_{j=1}^n w_j = 1$), and $v_j(x_{jt})$ is the value of the performance x_{jt} on criterion g_j , for all for $j = 1, \dots, n$.

The asymmetric part of the relation, $t' \succ t''$, means that t' is considered to impact strictly more than t'' , while the symmetric of the relation, $t' \sim t''$, means that t' is considered to impact equally as t'' . The three relations \succsim , \succ , and \sim are transitive.

The construction of the value function, $v_j(x_{jt})$, for criterion g_j and each action or time period $t \in T$, is done in such a way that its value increases with an increasing of the performances level of criterion j , $j = 1, \dots, n$ (this function is a non-decreasing monotonic function). Let t' and t'' denote two actions. The following conditions must be fulfilled:

1. The strict inequality $v_j(x_{jt'}) > v_j(x_{jt''})$ holds, if and only if, the impact of performance $x_{jt'}$ is considered strictly higher than the impact of performance $x_{jt''}$, on criterion g_j (it means that, t' impacts strictly more than t''), for $j = 1, \dots, n$.

2. The equality $v_j(x_{jt'}) = v_j(x_{jt''})$ holds, if and only if, the performance $x_{jt'}$ impacts the same as the performance $x_{jt''}$, on criterion g_j , (it means that t' impacts equally as t''), for $j = 1, \dots, n$.

In addition, the value functions are also used for modeling the impact of the performance differences. The higher the performance difference, the higher the strength of the value function impact. Let t' , t'' , t''' , and t'''' denote four actions. The following conditions must be fulfilled:

1. The strict inequality $v_j(x_{jt'}) - v_j(x_{jt''}) > v_j(x_{jt''''}) - v_j(x_{jt''''})$ holds, if and only if, the strength of the impact of $x_{jt'}$ over $x_{jt''}$ is strictly higher than the strength of impact of $x_{jt''''}$ over $x_{jt''''}$, on criterion g_j , for $j = 1, \dots, n$.
2. The equality $v_j(x_{jt'}) - v_j(x_{jt''}) = v_j(x_{jt''''}) - v_j(x_{jt''''})$ holds, if and only if, the strength of impact of $x_{jt'}$ over $x_{jt''}$ is the same to the strength of impact of $x_{jt''''}$ over $x_{jt''''}$, on criterion g_j , for $j = 1, \dots, n$.

In the construction of the value functions and the criteria weights we assume that the axioms of transitivity and independence hold (see Keeney and Raiffa 1993).

2.3. Chromatic ordinal classification model

The chromatic ordinal classification model is an ordinal scale with categories and colours associated with them. Let $C = \{C_1, \dots, C_r, \dots, C_s\}$ denote a set of totally ordered (and pre-defined) categories, from the best C_1 (the lowest pandemic state impact), to the worst C_s (the highest pandemic state impact): $C_1 \succ \dots \succ C_r \succ \dots \succ C_s$, where \succ means “impacts strictly more than”. The categories are used to define a set of states, as follows:

- C_1 (green): Baseline state.
- C_2 (light green): Residual state.
- C_3 (yellow): Alarm state.
- C_4 (orange): Alert state.
- C_5 (red): Critical state.
- C_6 (dark red): Break state.

There are four fundamental states, from C_2 to C_5 , with a particular set of measures/recommendations associated with them, as it can be seen in subsection 4.5. Let us point out that the colours assigned to each state will change smoothly when reaching the boundaries of the neighbouring states and that they go quickly when moving from a state to the next in the upper part of the scale, as for example from C_4 to C_5 , than when moving from a state to the next in the lower part of the scale, as for example from C_2 to C_3 . It also goes quickly from top down, i.e., in a descending way. This can be done through the way the value functions are modelled and/or the choice of the values for setting the cut-off lines with the possible definition of thresholds (see subsection 3.4, for the justification).

3. Modeling aspects

This section provides the details of the three fundamental models used in our study: the criteria model, the aggregation model, and the graphical visualization and communication model. The classification chromatic system and an illustrative example are also presented in this section.

3.1. Criteria model

A set of criteria built by the experts as the most relevant ones taking into account the two main points of view (called pillars) were used to characterize the pandemic. This set of criteria was built to fulfill several desirable properties as stated in Keeney (1992): be essential, controllable, complete, measurable, operational, decomposable, non-redundant, concise, and understandable, and were grouped as follows:

- A. Pillar I (ACT). *Activity*. This pillar was built to capture the main aspects of the COVID-19 registered or observed activity, i.e., the survival and development of the virus and its capability to still be active and cause infection on people in a given territorial unit. The following two COVID-19 activity criteria were considered to render operational this pillar.

1. Criterion g_1 - *Incidence* (INCID). The *incidence* (see Martcheva 2015) is the number of new COVID-19 positive cases presented daily, $N(\cdot)$, in the Health Official Reports. In most countries the exact daily values vary periodically along each week. In particular, in Portugal the discrete Fourier Transform exhibits a strong peak at day 7. Thus, to regularize the time series of the incidence, we consider the seven day moving average and use this variable in our computations:

$$g_1(t) = x_{1t} = \frac{\sum_{u=t-6}^t N(u)}{7} \quad (2)$$

We could use directly the raw data, but that choice would introduce artificial weekly fluctuation due to weak report at weekends.

2. Criterion g_2 - *Transmission* (TRANS). The transmission is modeled here as the rate of change of the active cases computed from the raw data of the daily incidence $N(\cdot)$ (with no moving averages). With the goal of regularizing this criterion and smoothen the weekly fluctuations, we perform the geometric average of the last seven days. Our variable is defined by using the expression below:

$$g_2(t) = x_{2t} = \left(\prod_{u=t-6}^t \frac{\sum_{v=u-6}^u N(v)}{u-1} \right)^{\frac{1}{7}} \quad (3)$$

With this formula, we have the advantage of a quicker response to the changes of incidence relative to $R(t)$ transmission rate, the usual reproduction number of an epidemic with time. Moreover, our model has the same meaning as the $R(t)$, for $t = 1$ (see Koch 2020).

B. Pillar II (SEV) - *Severity*. This pillar was built to capture on how serious are the effects of COVID-19 on Portuguese people, in particular in the health system. The following three COVID-19 severity criteria were considered to render operational this pillar.

3. Criterion g_3 - *Lethality* (LETHA). The lethality is modeled here by taking into account the ratio of deaths at a given time period u over the number of new cases fourteen days prior. Then, by considering the accumulate number of cases $N(\cdot)$ and number of deaths, $O(\cdot)$, it can be calculated using the formula:

$$\ell(t) = 100 \times \frac{O(t) - O(t-1)}{N(t-14) - N(t-15)},$$

We make the hypothesis that the average time of death after the communication of the case is fourteen days. With the goal of regularizing this variable and smooth the fluctuations we perform a moving average of the last fourteen days of this variable. The lethality formula used is given as follows:

$$g_4(t) = x_{3t} = \frac{1}{14} \sum_{u=t-13}^t \ell(u). \quad (4)$$

Another formula could be defined for modeling the lethality, but this one has been considered the most adequate in our case.

4. Criterion g_4 - *General nursery beds* (TOTAL). The total number of beds occupied by COVID-19 patients without counting the intensive care beds, $H(\cdot)$, is a raw data. The formula is a direct one:

$$g_4(t) = x_{4t} = H(t) \quad (5)$$

5. Criterion g_5 - *Intensive care beds* (INTEN). As in the previous criterion, the number of intensive care beds occupied by COVID-19 patients, $U(\cdot)$, is a raw data. The formula is also a direct one:

$$g_5(t) = x_{5t} = U(t) \quad (6)$$

All the raw data $N(\cdot)$, $O(\cdot)$, $H(\cdot)$, and $U(\cdot)$, are available at the Direccção-Geral da Saúde (DGS) web site (www.dgs.pt).

Remark 1. (*Fragility Point 1*) *Imperfect knowledge regarding the set of criteria (see Roy et al. 2014). This imperfect knowledge is mainly due to the imprecision of the tools and the procedures used to determine the raw data needed for the computation of the performance levels of the three criteria (namely $N(\cdot)$, since $O(\cdot)$, $H(\cdot)$, and $U(\cdot)$ are not suffering for a significant imprecision) and also due to the arbitrariness of the formulas chosen for such three criteria ($g_1(\cdot)$, $g_2(\cdot)$, and $g_3(\cdot)$), because other models could have been selected and well justified. Whenever, a fragility point (weakness or vulnerability) is identified, some robustness analyses are needed to guarantee the validity of the model and a confidence in the results. These robustness analyses will be presented in Section 4.*

3.2. Parameters of the aggregation model and the chromatic system

This section presents the technical aspects related with the construction of the parameters of the additive aggregation model (i.e, the value functions and the weights) as well as the chromatic classification system.

The construction of the value functions (interval scales) and the weights of criteria (ratio scales) were performed through a simplified version of the Pairwise Comparison Deck of Cards Method (here called PaCo-DCM), proposed by Corrente et al. (2021). The use of the Deck of Cards Method (DCM) in MCDA dates back to the eighties, to the procedure proposed by Simos (1989), and later revised by Figueira and Roy (2002), and used for determining the weights of criteria in outranking methods. However, Figueira and Roy (2002) mention the possibility of using the method and software for building interval and ratio scales in general. For another extension DCM and a review of applications, see Siskos and Tsotsolas (2015). Another attempt to build interval scales has been proposed by Pictet and Bollinger (2008), while Bottero et al. (2018) improved the DCM method to build more general interval scales (based on the definition of, at least, two reference levels with a precise meaning for policy/decision-maker, users, or experts), and ratio scales for determining the capacities of Choquet integral aggregation method. Finally, Dinis et al. (2021) make use of a tradeoff procedure for determining the weights of criteria for the additive MAVT model. The method used for computing the weights of criteria in this paper is very similar to the latter.

3.2.1. Value functions (interval scales)

The construction of the value functions through PaCo-DCM requires the use of pairwise comparison tables. This idea was introduced in MCDA by Saaty (2016) and later adapted and improved to accommodate qualitative judgments by Bana e Costa et al. (2016).

In what follows, we will show, step-by-step, the details of the application of PaCo-DCM. In sociotechnical processes is always important to provide some elements about the key aspects of the context that render possible and facilitate the evolution of such processes. Thus, before making such a presentation let us precise that we had a very limited time period for producing a first version of our tool (only three days) and for presenting a first prototype with meaningful results (ten days more). This has been possible since we benefit from the help of a member (a mathematician) in the CCIST team who has also a strong expertise in COVID-19 and knows very well the members of the experts from the GCOM team. In addition, the CCIST member is also very good at programming with Wolfgang Mathematica, which has been crucial for having the results of our tests and graphical tools almost instantaneously; a fundamental aspect for the interaction with the members of the GCOM team. Finally, let us point out that the members of GCOM team are medical doctors and they are familiar with the fundamentals of mathematics.

We will present the main interaction steps between the CCIST team and the GCOM team for constructing together the value function of the first criterion as a sociotechnical process, which took into account the expert's judgments and the technicalities of the PaCo-DCM tool. We will also present the details of all the computations. The reader can easily follow the way of building an interval scale with a simple, but adequate version of PaCo-DCM.

1. *The basics of PaCo-DCM for gather/assessing the experts judgments.* The method was introduced to the experts in a very simple way, by explaining them the meaning of the deck of cards used to assess their judgments through a small example. The experts were provided with two sets of cards:

- (a) A very small set of labeled cards with objects they know very well (for example, lemon, apple, and mango) and that can easily be preferentially ordered (first mango, then apples, then lemon), from the best to the worst (for the sake of simplicity assume they are totally ordered, i.e., there are no ties). All the experts agreed on the same ranking.
 - (b) A large enough set of blank cards. These blank cards are used to model the intensity or strength of preference between pairs of objects.
 - (c) Assume we have three objects o_i , o_k , and o_j . If the experts feel the strength of preferences difference between o_i and o_k is stronger than the strength of preferences difference between o_k and o_j , they put more blank cards in between o_i and o_k than in between o_k and o_j . The experts can put as many cards as they want in between two objects and they do not need to count them, just feel in their hands (in fact, in our case we used wooden balls instead of blank cards; this does not invalidate the application of the method and it is more suitable to get the judgments from the experts, because the wooden balls are easy to handle and have a better visualization effect. Let us point out that, the experts can always revise their judgments about the strength of preference and change the amount of cards in between two objects.
 - (d) We then explained the experts that:
 - No blank card in between two objects does not mean that the two objects have the same value, but that the difference is minimal (minimal here means equivalent to the value of the unit, a concept the experts would feel better later on).
 - One blank card, means that the difference of preferences is twice the unit.
 - Two blank cards, means that the difference of preferences is three times the unit.
 - And, so on.
 - (e) Finally, we explained the experts that in our case, we will model the strength or intensity of the pandemic impact instead of preferences, but the concept of preferences was extremely useful to render the experts familiar with the concept of strength of impact and with the method.
2. *(At least) two well-defined reference levels.* PaCo-DCM needs the definition of two reference levels for the construction of an interval scale. These two levels must have a precise meaning for the experts. One level is, in general, located in the lower part of the scale and the other in the upper part of the scale. This is similar to the method proposed in Bana e Costa et al. (2016) where a “neutral” and a “good” reference levels are needed to build a scale, with the assignment of the values 0 and 100, respectively. In PaCo-DCM, the values of the reference levels do not need to be set at 0 and 100. Any two values can be used in PaCo-DCM for building the interval scale. In the application of the model to the pandemic situation, the two reference levels built from the interaction with the experts were the following:
- *Baseline level:* Incidence value equal to 0. This means that no new cases have been registered over the last seven days. It does not mean the absence of the pandemic, but the fact that no new cases have been observed. The value of the baseline level was set at $v_1(0) = 0$, which is an arbitrary origin in the interval scale for the 0 preference level on the first criterion.

- *Critical level:* Incidence value equal to 1125. The value of the critical level was firstly set at 1100, but after the discussion of next step, we made a slight adjustment to 1125 and decided to set $v_1(1125) = 100$, which represents the highest value before entering in a critical state. Due to the number of public health medical doctors and the capacity to trace contacts at risk (tracking), after 900 new cases there is a saturation of resources, and the experts considered that 1125 is an adequate number to model the critical level.
3. *Setting the number of value function breakpoints.* In this step we defined along with the experts the most adequate way of discretizing, by levels, the performance levels of the incidence, taking into account the initial two reference levels built in the previous step, 0 and 1100. A first discussion led us to consider only values in between 0 and 2000, more than this value would lead to a break state, even 2000 seemed a very large value. Thus, we thought to discretize the range $[0, 2000]$ into six breakpoints points 0, 450, 900, 1350, 1800, and 2000, but an amplitude of 450 in between two consecutive levels was considered quite large. Finally, we decided to discretize the range in ten points, with an amplitude of 225 between two consecutive points. The following values were finally considered, with an adjustment in the last one to be consistent with the amplitude of 225, and by considering the critical level at 1125 instead of 1100:

$\{0\} \{225\} \{450\} \{675\} \{900\} \{1125\} \{1350\} \{1575\} \{1800\} \{2025\}$.

4. *Inserting blank cards.* In this step the experts were invited to insert blank cards in between consecutive levels; it corresponds to fill the diagonal of Table 1. This process has been performed for the initial number of breakpoints (ten) and led to several adjustments resulting from the sociotechnical co-constructive interaction process between experts (GCOM team) and the analysts (CCIST team). Every change was accompanied with a figure (see, for example, Figure 1), which was an important visualization tool for helping the experts. Some consistency tests were also performed as the ones described in the next step. After building all the value functions (Appendix) and test them with past observations of the pandemic the two last levels below were discarded since a break level was set before reaching level 1575.

$\{0\}[0]\{225\}[2]\{450\}[4]\{675\}[6]\{900\}[8]\{1125\}[10]\{1350\}[13]\{1575\}[11]\{1800\}[8]\{2025\}$

However, we can observe that the number of blank cards is increasing till level 1575, then it decreases. It means the shape of the value function will move from a convex to a concave shape (it would be similar to a continuous sigmoid function). From a certain point, more new cases have almost the same impact on the pandemic as less new cases. We are referring to a part of the function where the situation would be out of control.

5. *Testing a more sophisticated version of the method.* In the PaCo-DCM method, more cells of the table can be filled from the judgments provided by the experts. The time limitation and the good understanding of the method by themselves reduced the number of interactions for checking consistency judgments. Table 1 contains all the possible comparisons with the eight levels kept (here we are not considering the two last ones). An example of a test for assessing the impact differences between the two non-consecutive levels $\{900\}$ and $\{225\}$ was performed as follows. We put the set of six blank cards, in between $\{900\}$ and $\{675\}$, and

the set of seven cards, in between {675} and {225} (the experts do not really need to know how many cards are in between these levels, but these two numbers came from the previous interaction and were provided by the experts). Then, we put a set of sixteen cards in between {900} and {225} and we asked the experts to compare the three sets of cards, asking if they felt comfortable with the third set of sixteen cards. If not, we started by removing the blank cards, one by one. We removed blank cards, till having a set of thirteen blank cards. Then, we explained the experts that thirteen cards is slightly inconsistent and showed them why. We finally, asked if they felt comfortable with a set of fourteen blank cards (six, in between {675} and {900}, plus seven, in between {225} and {675}, plus one), and they agreed. The other cells of the table can be filled by *transitivity*, i.e., by following the consistency condition presented in Corrente et al. (2021). The impact difference between two non-consecutive is determined as follows:

$$e_{ij} = e_{ik} + e_{kj} + 1 \quad \text{for all } i, k, j = 1, \dots, t \quad \text{and} \quad i < k < j \quad (7)$$

We can see that $e_{\{900\},\{225\}} = e_{\{900\},\{675\}} + e_{\{675\},\{225\}} + 1 = 6 + 7 + 1 = 14$.

	{0}	{225}	{450}	{675}	{900}	{1125}	{1350}	{1575}
{0}		0	3	8	15	24	35	49
{225}			2	7	14	23	34	48
{450}				4	11	20	31	45
{675}					6	15	26	40
{900}						8	19	33
{1125}							10	24
{1350}								13
{1350}								

Table 1: Pairwise comparison table for criterion g_1 (incidence)

6. *Computations.* The computation of the values of the breakpoints was done as follows:

- The values of the reference levels: $v_1(0) = 0$ and $v_1(1125) = 100$.
- The number of units in between them, $h = (0+1)+(2+1)+(4+1)+(6+1)+(8+1) = 25$. Remember that 0 cards does not mean the same value, but that the difference is equal to the unit. Thus, we need to add one more to all the number of blank cards in between two levels.
- The value of the unit, $\alpha = (v_1(1125) - v_1(0))/h = (100 - 0)/25 = 4$. The value of the unit is equal to four points, and now the experts could understand better the concepts of unit and value of the unit.
- The values of the breakpoints are now easy to determine: $v_1(225) = 0 + 4 \times 1 = 4$, $v_1(450) = 0 + 4 \times 4 = 16$, and so on, for the remaining: $v_1(675) = 36$, $v_1(900) = 64$, $v_1(1125) = 100$, $v_1(1350) = 140$, $v_1(1575) = 200$.

7. *The shape of the value function.* After assessing the values of the breakpoints, we can then draw a piecewise linear function as in Figure 1. Any value within each linear piece can then be obtained by linear interpolation.

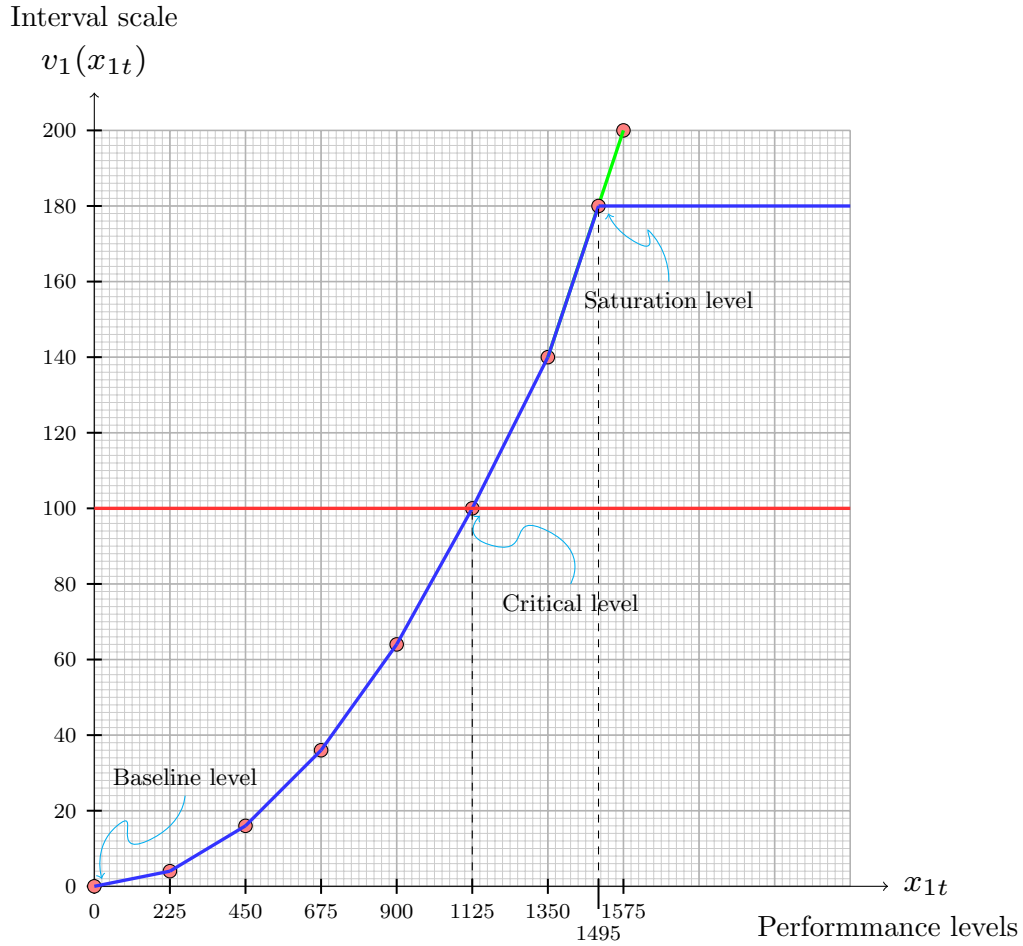


Figure 1: Shape of the value function for incidence

A marginal change in the lower part of the scale has less impact of the pandemic than the same marginal change in the upper part of the scale, as depicted in Figure 1. Before reaching the saturation level, this is a convex value function, and we can observe the marginal increasing in the value of the impact. Moving from 0 to 225 implies an increase of the indicator value, from 0 to 4, while moving from 1125 to 1350 implies an increase of the impact value from 100 to 140. The same number of additional units 225 in the upper part of the scale produces a much higher impact (40) than in the lowest part of the scale (only 4). It increases more near to the critical level, than near to the baseline level. This was a strong requirement established by the experts.

8. *The output of the model and possible approximations.* One of the outputs of the model is a piecewise linear function, whose mathematical expression can be stated as follows:

$$v_1(x_{1t}) = \begin{cases} 4x_{1t}/225 & \text{if } x_{1t} \in [0, 225[\\ 4x_{1t}/75 - 8 & \text{if } x_{1t} \in [225, 450[\\ 4x_{1t}/45 - 24 & \text{if } x_{1t} \in [450, 675[\\ 28x_{1t}/225 - 48 & \text{if } x_{1t} \in [675, 900[\\ 4x_{1t}/25 - 80 & \text{if } x_{1t} \in [900, 1125[\\ 44x_{1t}/225 - 120 & \text{if } x_{1t} \in [1125, 1350[\\ 52x_{1t}/225 - 168 & \text{if } x_{1t} \in [1350, 1495[\\ 180 & \text{if } x_{1t} \in [1495, +\infty[\end{cases} \quad (8)$$

This particular function could be approximated by a quadratic function without losing much information, but such an approximation needs to be validated by the experts:

$$\tilde{v}_1(x_{1t}) = \begin{cases} 100(x/1125)^2 & \text{if } x_{1t} \in [0, 1125] \\ 180 & \text{if } x_{1t} \in [1125, +\infty[. \end{cases}$$

The mean of the Euclidean distance between the functions \tilde{g}_1 and g_1 is given by the following expression:

$$\frac{\sqrt{\int_0^{+\infty} (g_1(\xi) - \tilde{g}_1(\xi))^2 d\xi}}{\sqrt{\int_0^{+\infty} (g_1(\xi))^2 d\xi}} = 6.40389 \times 10^{-4},$$

which is almost negligible and shows that the approximation does not lead to the loss of much information.

9. *Missing, imprecise, and inconsistent judgments.* The method described in Corrente et al. (2021) also allows to deal with missing, imprecise, and inconsistent judgments. The inconsistency analysis is performed by using linear programming, similarly to what is done in other MCDA tools, as for example, in Mousseau et al. (2003). The background of the members of the two teams and the way the sociotechnical interaction was conducted largely facilitated the process of gathering information, not requiring to use more sophisticated functionalities of the PaCo-DCM tool.
10. *The break level.* After running the model for the whole set of days of the pandemic, the experts realized they could set a maximum of 180 points for this function since all the situations beyond such a point would be equally bad and out of control. The function was thus truncated at the level 1495, which is the first level with a value of 180. After this performance level the situation collapses and all the performance levels are felt as serious as the break level.

The piecewise functions for the other four criteria, as well as the number of blank cards in between consecutive levels, are provided in the Appendix.

Remark 2. (*Fragility Point 2*) *Subjectivity in building the value functions. There is some obvious subjectivity in the construction of the value function since the experts are not precise instruments as high tech thermometers. In addition, there is not a true value function for modeling the incidence; this function is a construct, which can be more or less adequate to the situation. This is another kind of fragility point of our model, which justifies the use of robustness analyses as we will present in Section 4.*

3.2.2. Weighting coefficients (ratio scales)

The assignment of a value for each criterion weight was also performed through PaCo-DCM, but the interaction protocol with the experts and the nature of the judgments was presented to them in different way. The weights are interpreted here as scaling factors or substitution rates. The dialogue with the experts was conducted as follows:

1. *Constructing dummy projects.* A set of five dummy projects, one *per* criterion, representing the swings between the baseline level and the critical level were built as follows (see also Dinis et al. 2021).
 - $p_1 = (1125, 0, 0, 0, 0) \equiv (100, 0, 0, 0, 0)$. This project represents the impact on the pandemic of the swing (regarding the first criterion) from the baseline level to the critical level, keeping the remaining criteria at their baseline levels.
 - $p_2 = (0, 1, 0, 0, 0) \equiv (0, 100, 0, 0, 0)$. The meaning of this project is similar to the one provided for the first project. A transmission rate equal to one is considered adequate by the experts to represent the critical level.
 - $p_3 = (0, 0, 3.6, 0, 0) \equiv (0, 0, 100, 0, 0)$. The meaning of this project is similar to the one provided for the first project. For defining this project, the experts considered that the half of the maximum value of $g_3(t)$ along the pandemic in Portugal corresponds to 100 points. Thus, $\max\{g_3(t)\} = 7.19148$. Consequently, 100 points correspond to the value of lethality of $3.59574 \approx 3.6$.
 - $p_4 = (0, 0, 0, 2500, 0) \equiv (0, 0, 0, 100, 0)$. The meaning of this project is similar to the one provided for the first project. The 2500 represent 15% of the total number of beds, and its an adequate number for defining the critical level.
 - $p_5 = (0, 0, 0, 0, 200) \equiv (0, 0, 0, 0, 200)$. The meaning of this project is similar to the one provided for the first project. The 200 beds represent 80% of the difference between the current number of beds and the number of beds existing before the pandemic, and this number was defined by the experts as adequate to represent the critical level.

The concept of swings is in line with the swing weighting technique by von Winterfeldt and Edwards (1986) and the use of two reference levels with the concepts of “neutral” and “good” by Bana e Costa et al. (2016).

2. *Ranking the dummy projects with possible ties.* The experts received five cards, one with each one of the previous projects and the analyst’ team asked them to provide a ranking of these five cards, with possible ties, according to the impact that the swings have in the pandemic. The project(s) leading to the highest impact was (were) placed in the first position, the

one(s) with the second more impact on the second, and so on. The following ranking has been proposed by the experts.

$$\{p_1\} \{p_3, p_4, p_5\} \{p_2\}$$

The analysts explained to the experts that the project in the first position will get the highest weight, the ones in the second position the second highest weight, and the project in the last position the lowest weight.

3. *Inserting blank cards.* The experts were invited to insert blank cards in between consecutive positions to differentiate the role each weight (swing) will have in the impact of pandemic, after explaining them the meaning of swings and substitution rates. The following set of blank cards was provided by the experts.

$$\{p_1\} [2] \{p_3, p_4, p_5\} [3] \{p_2\}$$

Of course, as in the case of the value functions a more sophisticated PaCo-DCM procedure could be used for such a purpose, but the experts felt comfortable with the information they provided.

4. *Assessing the value of the substitution rates.* This was the most difficult question for the experts. We need to establish a relation between the weight of the criterion in the first position of the ranking (incidence) and the weight of the criterion in the last position of the ranking (transmission). In PaCo-DCM this is called the z -ratio, used to build a ratio scale. After a long discussion and several attempts, the experts provided the following relation between the two weights: $z = \hat{w}_1/\hat{w}_2 = 2$. We are using \hat{w}_j , for the non-normalized weight of criterion g_j , for $j = 1, \dots, 5$.
5. *Calculations.* The computations are similar to the ones performed for the value functions:
 - The values of the non-normalized weights of the projects in the first and last positions of the ranking, i.e., $\hat{w}_1 = 2$ and $\hat{w}_5 = 1$.
 - The number of units in between them, i.e., $h = (2 + 1) + (3 + 1) = 7$.
 - The value of the unit, $\alpha = (\hat{w}_1 - \hat{w}_5)/h = (2 - 1)/7 = 0.14286$.
 - The non-normalized weights: $\hat{w}_2 = 1$, $\hat{w}_3 = \hat{w}_4 = \hat{w}_5 = 1.42858$, and $\hat{w}_1 = 2$.
 - The normalized weights: $w_2 = 1/7.28574 = 0.13725$, $w_3 = w_4 = w_5 = 1.42858/7.28574 = 0.19608$, and $w_1 = 2/7.28574 = 0.27451$.
6. *Final adjustments.* After adjusting the model results to the real pandemic data and some discussions with the experts, the following weights have been proposed to be used in the model: $w_2 = 0.141$, $w_3 = w_4 = w_5 = 0.193$, and $w_1 = 0.280$.

Remark 3. (*Fragility Point 3*) *Subjectivity in building the weights of criteria. The justification is in line with the one provided in Remark 2, which also requires the use of robustness analyses (see Section 4).*

3.3. Illustrative example

This is an illustrative example with five actions, i.e., five different time moments of the pandemic. Moment $t = 0$ is four days before of the day of the press conference with the *media* (July 14, 2021) at the Portuguese Medical Association (in Lisbon). The other moments, t , were set with respect to the number of the days before $t = 0$ and corresponds to the first lock down in Portugal (March 20, 2020), one of the lowest activity and severity periods (July 31, 2020), Christmas (December 24, 2020), and some days after the second lock down (January 24, 2021). Table 2 presents the activity as severity performance levels for the five considered criteria, according to the two pillars.

t	Date	Pilar I (ACT)		Pilar II (SEV)		
		x_1 (INCID)	x_2 (TRANS)	x_3 (LETHA)	x_4 (TOTAL)	x_5 (INTEN)
-474	2020-03-20	194	1.301	4.160	128	41
-343	2020-07-31	197	0.978	1.140	340	41
-197	2020-12-24	3574	0.987	2.180	2348	505
-166	2021-01-24	12341	1.039	3.460	5375	742
0	2021-07-10	3658	1.042	0.382	488	144

Table 2: Performances levels for five moments of the pandemic

From the data in Table 2 and by applying the piecewise linear value functions previously constructed, we obtained the results shown in Table 3. The last column of this table provides the overall value of each moment of the pandemic, after applying the formula of Model (1).

t	Date	Pilar I (ACT)		Pilar II (SEV)			Value
		$v_1(x_{1t})$ $w_1 = 0.280$	$v_2(x_{2t})$ $w_2 = 0.141$	$v_3(x_{3t})$ $w_3 = 0.193$	$v_4(x_{4t})$ $w_4 = 0.193$	$v_5(x_{5t})$ $w_5 = 0.193$	
-471	2020-03-20	3.441	180.00	115.571	1.0240	4.300	49.6800
-342	2020-07-31	3.503	60.900	31.6990	2.7200	4.300	17.0400
-197	2020-12-24	180.0	76.702	60.5640	89.056	180.0	124.832
-166	2021-01-24	180.0	180.00	96.1120	180.00	180.0	163.810
0	2021-07-10	180.0	180.00	10.6240	3.9040	52.80	88.7700

Table 3: Value functions scores for the five moments of the pandemic

We can observe that our pandemic indicator, PACI, reached its highest value in January 2021 and the lowest in July 2020. The four first moments of this example are displayed in Figure 6b (Appendix) and used to test the experts and some anonymous people about the validity of the indicator.

3.4. Chromatic classification system (ordinal scales)

The chromatic classification system is a tool that makes use of colours for better visualizing the ordinal scale built with the experts. The colours selected for our model were inspired by the ones used in the RM since the Portuguese people were already familiar with them.

Five fundamental states were defined with the experts: residual, alert, alarm, critical, and break. In addition, two more states were considered at the extremes, a baseline (the very lowest one) and a saturation or break state (the highest). All these states are zones defined in between two consecutive levels or cut-off lines:

- *Baseline level* (cut-off line value = 0). The five performance baseline levels are presented in the following list: [0, 0, 0, 0, 0]. Each baseline level has a precise meaning for the experts. In this case, the first two mean there are no pandemic activity recorded over the last seven days, which does not mean, of course, the pandemic is extinct, but simply that we did not registered activity over the last seven days. The other three values, mean there are no deaths over the last seven days and there are no hospitalized COVID-19 patients.
- *Residual level* (cut-off line value = 10). The five performance residual levels are presented in the following list: [338, 0.93, 0.36, 750, 60]. The experts were confronted with this list, the values of each level on each value function, and its adequacy to represent an overall value of 10. These elements were validated by the experts.
- *Alert level* (cut-off line value = 40). The five performance alert levels are presented in the following list: [707, 0.963, 1.43, 1571, 126]. The discussion with the experts was performed as in the previous case.
- *Alarm level* (cut-off line value = 80). The five performance alarm levels are presented in the following list: [1000, 0.989, 2.89, 2222, 178]. The discussion with the experts was performed as in the residual and alert cases.
- *Critical level* (cut-off line value = 100). The five performance critical levels are presented in the following list: [1125, 1, 3.6, 2500, 2727, 200]. As the baseline levels, these levels are reference levels for the experts and have a particular meaning.
- *Break level* (cut-off line value = 120). The five performance break levels are presented in the following list: [1227, 1.009, 4.31, 218]. The discussion with the experts was performed as in the residual, alert, and alarm cases.
- *Saturation level* (cut-off line value = 180). The five performance saturation levels are presented in the following list: [1506, 1.034, 6.47, 3346, 268]. In this case, all the experts agreed that any performance higher than the ones presented in the list will be considered as serious as the ones in the list. This corresponds to what we considered a saturation level.

The five fundamental states can be represented as in the Figure 2. The transition between colours or states is not necessarily done abruptly.

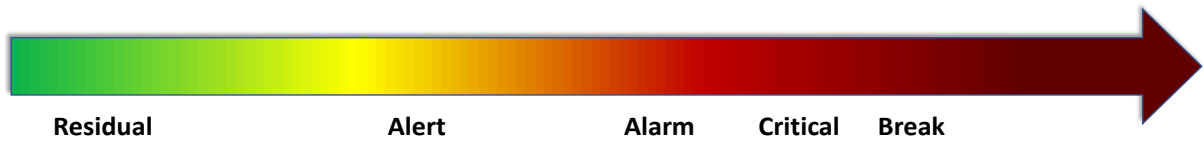


Figure 2: Chromatic classification system

A smooth transition can be considered since the policy/decision-makers cannot necessarily make the decisions automatically, after moving to a different state. It is important to see the evolution of the pandemic in the next few days, after definitely moving from the current to a new state, and implement the measures/recommendations of this new state. Indeed, a lower and an upper threshold for each cut-off line could be considered instead for render operational a smooth transition.

As it can also be seen in the figure, and given the way the value functions were built, whenever we move to the next state, there is less room to make the decisions, i.e., it moves quickly, for example, from the alarm state to the critical state, than from the residual state to the alert state. This feature was a strong requirement by the experts. The main states can be briefly defined as follows:

- *Residual*: Absent or minimal pandemic activity without any impact on health structures (i.e., at the normal operating level) and without compromising the system tolerance.
- *Alert*: Mild pandemic activity, still without impact on the normal activity of health structures, but reaching the usual flexibility, adaptability and safety tolerance threshold (e.g., increase in the emergency room visits and/or in the occupancy rate of hospital admissions).
- *Alarm*: Moderate pandemic activity, already impacting the normal activity of health structures, with reallocation of technical and human resources and commitment to other health needs, reaching the functional reserve threshold.
- *Critical*: Strong pandemic activity, having already exceeded the system’s reserve threshold, conditioning effort and disruption in the activity of health structures allocated almost exclusively to the pandemic.
- *Break*: Very strong pandemic activity and imminent collapse of health structures.

Remark 4. (*Fragility Point 4*) *Cut-off lines subjectivity.* This is in line with the previous two fragility points. The definition of the cut-off lines is subjective, since they result from a co-constructive interactive process with the experts. However, the fact of defining thresholds for modeling a smooth transition between successive states can mitigate the subjectivity behind the definition of these cut-off lines.

3.5. Graphical model for visualization and communication

One of the main features of our model is the visualization functionalities to render easy the communication with the public in general. Apart from other minor graphical functionalities, four types of graphical tools were developed:

1. A graphic which displays the evolution of the indicator behaviour with state colours and cut-off lines to separate each state (see Figure 3).
2. An animation graphical tool with the cumulative contribution of each criterion to the pandemic (see Figure 4).
3. A graphical representation of the (positive) impact of vaccination plan to mitigate the progress of the pandemic in the country (see Figure 5).
4. A state chromatic line, as in Figure 2, with the corresponding measures by type, as presented in Section 4.5.

More details about these graphical tools will be provided in the next section.

4. Results, robustness analyses, and recommendations

This section is devoted to the implementation issues and verifications tests, results presentation, their validation, robustness analyses, final comments and definition of the measures and recommendations of the chromatic model for helping policy/decision-makers and the public in general.

4.1. Implementation issues and verification tests

Our application was coded in the software Wolfram Mathematica, version 12.0. All the functionalities of the Mathematica code were verified in several small examples with particular (extreme and pathological). This step includes verifications in the debugging, input of criteria performance levels parameters, calculation of the criteria performance levels, calculation of the value functions and weights parameters, calculation of the comprehensive values for each time unit, all the graphical models outputs, robustness analyses, as well as the checking whether all the logical structure of the models are correctly represented in the computer. The entire application has been designed to translate all the three models (criteria model, MAVT aggregation model, and graphical visualization and communication model) successfully in their entirety, as well as some additional functionalities for validation, simulation, and other robustness analyses purposes.

An Excel version of the model was also implemented with less functionalities. The computations of the Excel PACI model are available for public domain at a web site of Instituto Superior Técnico ([indicadorcovid19.tecnico.ulisboa .pt/](http://indicadorcovid19.tecnico.ulisboa.pt/)) and a web site of the Portuguese Medical Association (ordemosmedicos.pt/iap/). This software automatically computes the daily changes of all the five criteria performance levels and the actual transmission rate, which is computed by using the Robert Koch Institute formula (see Koch 2020).

4.2. Results

The results provide information on three main aspects: the pandemic evolution, the cumulative contribution of each criterion for such an evolution, and the impact of the vaccination plan. Let us remark that PACI can be used in a forecasting perspective if we have good prediction techniques for the raw data used in the five criteria formulas.

4.2.1. Pandemic evolution

Before running the MAVT-based model with our set of criteria, we tested it with the two criteria of the RM, by setting the baseline, critical, and break levels as we did for our PACI tool, and considering linear value functions (since moving from an $R(t) = 0.1$ to an $R(t) = 0.2$ has the same impact as moving from an $R(t) = 0.9$ to an $R(t) = 1.0$, and this is true for any location along the $R(t)$ scale; the same reasoning applies to the incidence criterion). We also considered equal weights for each criterion. For other cut-off lines we set them as in our model. The evolution line can be observed in Figure 6a (Appendix). It seems clear that this MAVT-based model does not represent adequately what Portuguese people feels about the impact of the pandemic. We can observe that the impact is always rather high, above the cut-off line of the alarm state. For the experts and the public in general the low impact of some moments of the pandemic, as for example, August 2020, cannot be clearly seen in this evolution line.

In Figure 3, we can observe the global evolution of the PACI values in Portugal, along with the cut-off lines for separating the chromatic states.

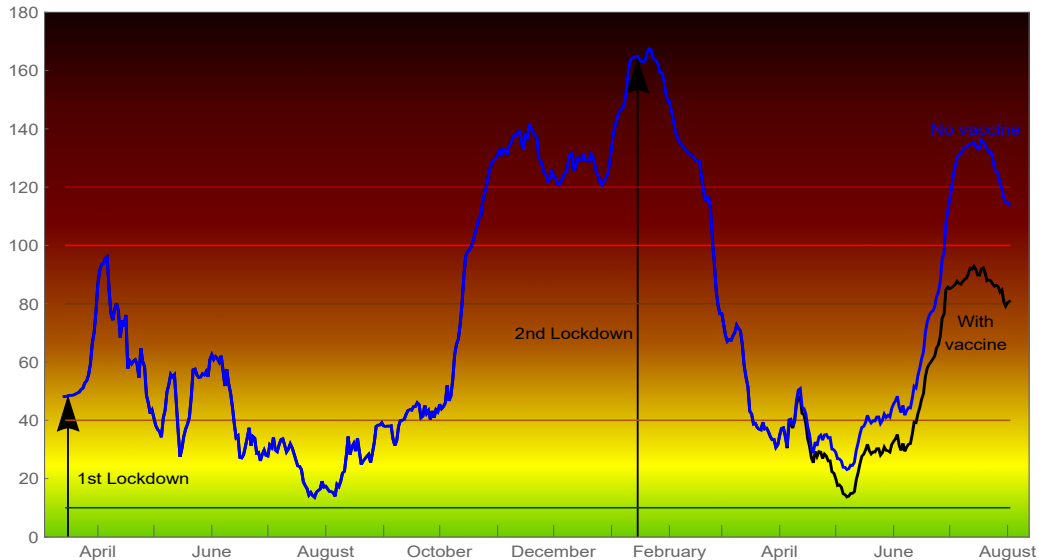


Figure 3: Evolution of the PACI till 2021-08-02

We can see that the indicator gives precisely the intuitive human perception of the impact of the pandemic in Portugal, namely the three main waves. The lockdown occurred exactly during the first wave, when the PACI value was close to 50 points. The inertia of the system brought the PACI to a maximum on April 6, 2020, with 95 points. The Summer of 2020 was relatively mild with minimal values of the PACI value near to 14, between July 26 and August 6, 2020. We reached intolerable values, between November 2020 and January 2021. In terms of pure pandemic impact, we can see that the Fall of 2020 and the Winter of 2021 correspond to the same pandemic wave. The day of maximum impact of the pandemic in Portugal, with 167.5 points, was precisely January 21, 2021, which preceded the reinforcement of the lockdown measures/recommendations on January 22, 2021. The sharp drop from January 25, 2021 was caused by the lockdown of January 15, 2021. The mentioned descent was greatly reinforced from February 6, 2021, just when

we started to feel the effects of closing the schools on the values of incidence and transmission rate. After April, 2021 due to the effect of vaccination there is a clear detachment of the curves obtained considering or not this parameter. Taking into account the effect of vaccination, it can be observed that the effect of the lockdown was maximal on May 7, 2021, with 13.4 points corresponding to the absolute minimum of the indicator since the appearance of COVID-19 in Portugal. After that date, we noticed an increase of the indicator, related to the surge of the Delta variant in the country. With a local maximum of 92.3 on July 9, 2021. In the absence of lockdown lifting, after that date, there is a slow tendency of decay of the impact, which can be related to the positive evolution of the vaccination in Portugal.

4.2.2. Cumulative contribution of each criterion

In Figure 4 we can observe the individual contributions of each criterion to the PACI value in Portugal.

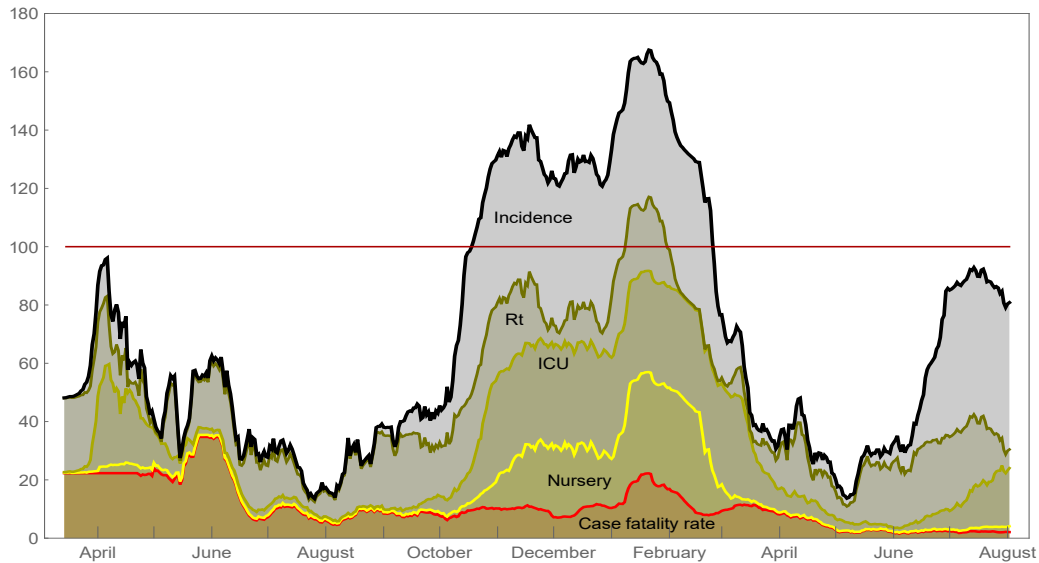


Figure 4: Cumulative contribution for the PACI till 21-08-04

It is noteworthy the long term and gradual decrease of the contribution of the lethality (case fatality rate line) to the indicator. We note that the contribution of this criterion increased dramatically during the January crisis. The contribution of the occupancy of general nursery beds (nursery line) is very significant during the second pandemic wave in Portugal, namely between October 2020 and February 2021. This partial contribution reacts strongly when the spread of the disease is out of control. It is interesting to notice that the intensive care beds (ICU line) criterion contribution to the PACI is significant at each pandemic wave. The intensive care beds occupancy grows after the increase of incidence (incidence line) with a delay of 10-12 days. Nevertheless, the relation between the two criteria (incidence and intensive care beds occupancy) is very clear and appears at every pandemic wave, including in the last wave after the beginning of June.

Contrarily to the intensive care beds criterion contribution, the growth rate contribution (R_t line) appears at an early stage of each wave and is the first alarm sign of a future increase of the

incidence, which is natural and expected. For instance, in the first wave of 2020, the PACI was mainly due to the growth rate and the case fatality rate for the first days of the introduction of infection by COVID-19 in Portugal. The same effect is clear in the second wave, in October, and in the last wave before June 2021.

Finally, the incidence contribution to the PACI was severe in the months between October 2020 and February 2021. The softening of governmental control measures/recommendations in Portugal on Christmas 2020 was done when the contribution of the incidence was high. The introduction and the effects of the Delta variant are visible in the contribution of the incidence to the PACI in June 2021. Fortunately, this increase in the incidence contribution was balanced by the drop in the case fatality rate, general nursery occupancy, and intensive care beds occupancy relative to the values before the introduction of the new variant and generalization of the vaccination in Portugal.

4.2.3. The impact of vaccination

Figure 5 shows a comparison of the values of PACI for actual data/parameters (lower curve) and an estimation of the indicator computed without the introduction of vaccination (upper curve). The upper curve was computed with the same observed incidence and growth rate of the actual PACI (lower curve), i.e., with the two criteria of Pillar I, but with no reduction of the severity of the disease (see the methodology described below), i.e., without the three criteria of Pillar II. The upper curve is, naturally, a lower bound estimate of the indicator without vaccination, since the immunization process related to vaccination reduces as well the incidence and growth rate.

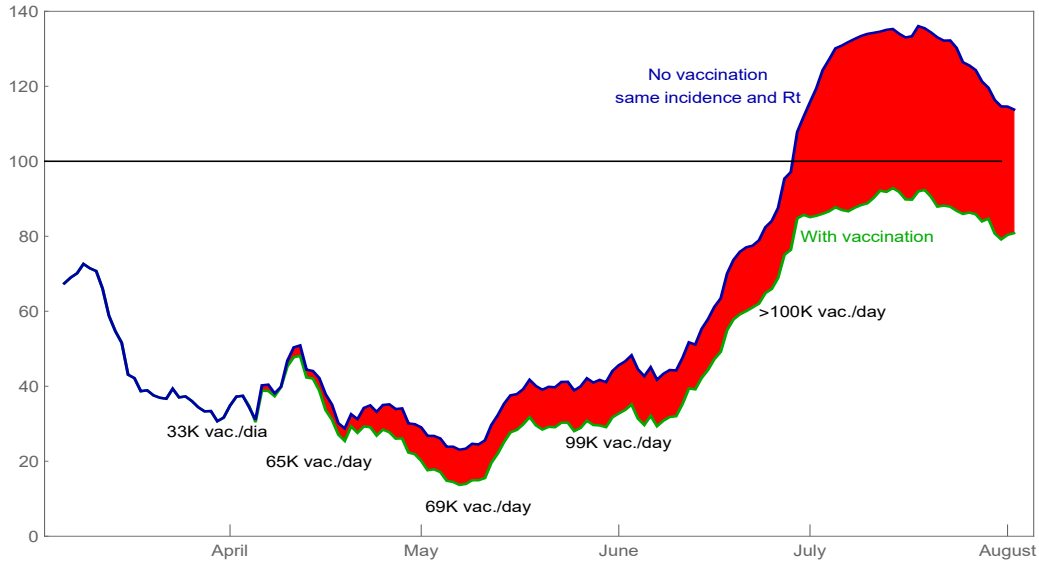


Figure 5: The vaccination impact till 21-08-04

The methodology to obtain this lower bound estimate of the indicator without vaccination, is to consider the averages of the proportion of intensive care beds and general nursery occupancies relative to the incidence, and fix that rate instead of using the actual data. Finally, for the case of fatality rate, we used the average of the first 390 days of the pandemic in Portugal. At this point, day 390, there is no difference between the values of the actual PACI indicator and the values of

the lower estimate PACI. We notice a severe increase of the difference between the curves with time. Naturally, since the vaccination also reduces the incidence and growth rate, after some time, more precisely in July 15, 2021, the indicator PACI starts to decrease.

4.3. Robustness analyses

The robustness analysis conducted in this study are in line with the definition proposed in Roy (2010), where a change in the results was observed after a (simultaneously) change of all the parameters that are affected by some fragility aspects (see Remarks 1, 2, and 3). There are two major types of robustness analyses, one based in simulation (also called pseudo-robustness analysis) and other based on an exact characterization of the effects of the changes in parameters. We detail these two types of robustness analyses in the next paragraphs.

4.3.1. Simulation on the weights

The fragility point of Remark 3, which is related with the subjectivity when assessing the weights of criteria, is one of the most critical fragility points in practice. A simulation analysis (also called pseudo-robustness) is dominated by an exact robustness analysis, as it was done in the next paragraph, but it has the advantage of being able to produce a large set of lines, whose shape gives us an idea of the evolution of our indicator (just for an illustrative purpose, see the last figure in Appendix, i.e., Figure 8, with a $\pm 20\%$ change on the weights used in the application of PACI to the Portuguese pandemic situation).

In order to study the robustness of the results, we performed a strong change on the weights allowing a variation on the range of each one, from 0 to 1 (the sum of all weights being equal to 1). A Monte Carlo simulation allowed us to make possible this study. Figure 6c (Appendix) displays 400 lines, among the 10 000 simulations performed (the representation of more lines is time consuming and led to a software crash due to a limited memory capacity). In this figure, we can observe that the shape of all the lines is more or less the same. Indeed, this is a drastic variation of the weights, since the simulations go from an extreme unbalanced situation, where only one criterion counts for the impact on the pandemic (the one with weight equal to one), to a complete balanced situation, where all the criteria contribute equally to the impact (equal weights for all criteria). A more realistic simulation, with a perturbation of 5% below the weights values elicited with PaCo-DMC (for each day) and the same 5% above the elicited weights, produces the curves (also 400) represented in Figure 6d. It is observed that all the lines almost coincide with the shape of the curve of our PACI model. The results are quite robust to realistic variations on the weights, which shows the adequate behaviour of the PACI tool.

4.3.2. Exact robustness analysis

The exact robustness analysis was performed by taking into account all the critical fragility points of Remarks 1, 2, and 3. Thus, we made changes in the data provided by the incidence, transmission, and lethality criteria (the other two are not suffering from a strong imprecision, i.e., the number of beds in nursery and intensive care is relatively precise), in the five value functions, given the subjectivity when making their construction, and in the weights of criteria, for the same reason as the previous one.

A first strong (direct) perturbation was made on all the data/parameters related with the first three criteria and the five value functions: 10% below and above their daily performance levels/values. As for all the criteria weights, a variation of 10%, allows to build a polyhedron in a

5-dimensional space. In this polyhedron we can find, for each day, a maximum and a minimum value of the indicator by using linear programming techniques. For the maximum value, we consider a +10% change of the performance levels of the first three criteria and the values of the five value functions, and compute the maximum of the indicator formula over the polyhedron of the weights; this is done for each day (see the upper envelope curve in Figure 7a, Appendix). For computing the minimum for each day, by considering a -10% change of the performances levels of the first three criteria and values of the five value functions (see the lower envelope curve in Figure 7b, Appendix). The difference of the upper and lower curves is in average 47.12 points with a standard deviation of 11.0848 points.

Then, a more realistic robustness analyses with a $\pm 5\%$ change in the performance levels of the first three criteria and the values of the five value functions, and a similar construction of the polyhedron for the weights was performed. The results are presented in Figure 7c and Figure 7d (Appendix). The average of the difference is now 28.2366 impact points and the standard deviation is 6.9615 points.

The outline of the robustness procedure can be presented as follows:

1. Take a certain time unit, t (e.g., a given day during the pandemic).
2. Consider the first three criteria: g_1 , g_2 , and g_3 . Proceed as follows (*Fragility Point 1*):
 - (a) Consider their performance levels at time t : x_{1t} , x_{2t} , x_{3t} .
 - (b) Make a decrease of these levels. Let x_{1t}^- , x_{2t}^- , and x_{3t}^- , denote the new performance levels.
3. Consider all value functions: v_1 , v_2 , v_3 , v_4 , and v_5 . Proceed as follows (*Fragility Point 2*):
 - (a) Consider their values at time t , tanking into account the modified performance levels of the first three criteria (as we did in the previous step) and the performance levels of the last two criteria: $v_1(x_{1t}^-)$, $v_2(x_{2t}^-)$, $v_3(x_{3t}^-)$, $v_4(x_{4t})$, and $v_5(x_{5t})$.
 - (b) Make a decrease of these values. Let v_{1t}^- , v_{2t}^- , v_{3t}^- , v_{4t}^- , and v_{5t}^- , denote the new values.
 - (c) (for the sake of simplicity of notation) Consider the vector, $v^{(t)-} = (v_{1t}^-, v_{2t}^-, v_{3t}^-, v_{4t}^-, v_{5t}^-)$.
4. Consider all weights: w_1 , w_2 , w_1 , w_2 , and w_3 . Proceed as follows (*Fragility Point 3*):
 - (a) Make a decrease of the weights. Let w_1^- , w_2^- , w_3^- , w_4^- , and w_5^- , denote the lower bound values for the weights.
 - (b) Make an increase of the weights. Let w_1^+ , w_2^+ , w_3^+ , w_4^+ , and w_5^+ , denote the upper bound values for the weights.
 - (c) Remark: These changes in the weights are not indexed to the time period; they are thus valid for any t .
 - (d) Construct the polyhedron of the weights, denoted by W , as the result of the intersection on the following constraints:
 - Bounding constraints: $w_j^- \leq w_j \leq w_j^+$, for $j = 1, \dots, 5$.
 - Normalization constraint: $w_1 + w_2 + w_3 + w_4 + w_5 = 1$.
 - Consistence constraints: $0 \leq w_j \leq 1$, for $j = 1, \dots, 5$ (these constraints avoid to have negative weights of weights with values strictly greater than one).

- (e) (for the sake of simplicity of notation) Consider the vector, $w = (w_1, w_2, w_3, w_4, w_5)$. Any feasible w is an element of W , i.e., $w \in W$.

5. Solve the following linear programming problem:

$$v^-(t) = \min \left\{ w^\top v^{(t)-} : w \in W \right\},$$

where $v^-(t)$ is the lowest (which, in our case, corresponds to the best) value of the PACI model.

6. Proceed in a similar way to obtain $v^+(t)$, i.e., the worst value of the PACI model for time unit t .

4.4. Validation, adjustments, and comments

In our case, the validation consists of confront the actors, mainly the experts, but also some anonymous people, with the shape of the pandemic indicator and get some comments to validate or making some adjustments in our models. All the tests were conducted before and after performing the robustness analysis, but they have more credibility after such an analysis is performed. More precisely, we performed the following tests:

1. In a first step, we built a figure displaying different moments of the pandemic in the country, see Figure 6b (Appendix). The moments were not identified in the time line and they were not chronological ordered. We asked the experts to analyse the figure and tell us if they were able to identify these moments. We selected the following moments: the beginning of pandemic, July 2020, January 2021, and Christmas 2020. All the experts were able to easily identify all the moments Only with a little hesitation of one expert with respect to the moment related with the beginning of the pandemic was registered. The elements of the team also asked the same question to some anonymous people. We performed tests with 30 individuals. Mostly university administrative staff (10 individuals), students (10 individuals), and random people found in the streets of Lisbon (10 individuals). We gave some additional explanations about the fact that the impact was represented in a scale from 0 to 180 points as well as the minimum and the maximum of PACI values. This number of people is rather low. A more systematic and complete study would be important to get more input for validating our PACI model. Anyway, for the people that we confronted with the four pieces of our graphic of Figure 6b, only the beginning of the pandemic was a moment leading to some hesitations. The lowest and the highest impact of the pandemic were easily identified, and Christmas 2020 almost always.
2. In a second step, we have shown the experts the whole evolution curve since the beginning of the pandemic. They were able to comment and justify all the moments of the pandemic and the different critical situations, i.e., the waves that occurred during the evolution of the disease in Portugal, namely the initial growth, the Fall and Winter crisis and, finally, the surge of the Delta variant. They were also able to identify the calm situation of the Spring/Summer of 2020 and the relaxation of April 2021. Then, the same has also been done for the anonymous people. Most of the tests were positive, with 28 people being able to comment the initial stage, the Fall/Winter crisis, and the wave due to the Delta variant.

3. In a third step, we asked the experts to comment about the reasons that lead to some moments of the curve and to explain us the reasons leading to such behaviours of the PACI values. We also choose some particular moments of the curve and asked the experts to comment and justify. This is a different approach from the previous one, since the moments were not chosen by the experts, but by the analysts.
4. Finally, we asked the experts to provide some raw data characteristic of each state, run the model, and show the results. For example, the following list of performance levels [1250, 1.02, 2.8, 2235, 195] is a profile that should be considered in the critical state. After running the model, we get the value 104.2, which is within the critical state of the chromatic model. Almost all the results led to the state provided by the experts. This test lead to a very little adjustment of the weights.

4.5. Measures and recommendations

Each state of the chromatic system has some measures and/or recommendations associated with it. These measures can be of interest for the Portuguese public health authorities to protect the residents/visitors and they are not necessarily set since the beginning; they can change and be adapted to the current state of pandemic. Next, we present examples of measures/recommendations at different levels that can be associated with the various states. There are several types of measures:

1. *Legislation for organizations and people (residents and visitors)*: Legislative contextualization (laws, regulations, policies, and procedures) of public health threats.
2. *Entry and leaving the country measures*: e.g., diagnostic tests, quarantine periods, presentation of the digital certificate COVID-19.
3. *Clinical and pharmacological measures*:
 - (a) Diagnostic test for symptomatic (updated case definition) within 12 – 24h + High-risk contacts tracking with PCR (Polymerase Chain Reaction)/TRAg(Rapid Antigen) tests, within 24h.
 - (b) Vaccination.
 - (c) Genome sequencing:
 - i. ECDC (European Centre for Disease Prevention and Control) threshold (> 500 samples per week or 10% of the samples).
 - ii. Focus on infections of recovered, vaccinated, vulnerable (e.g., immunosuppressed), and severe cases (e.g., hospitalized patients).
 - iii. Realtime national recording (host, interval & vaccine type, variant).
 - (d) Tracking:
 - i. Backtracking confirmed cases.
 - ii. Not vaccinated in healthcare contexts, nursing homes or jails.
 - iii. Vulnerable even when fully vaccinated before any procedure generating aerosols, nursing homes or jails.
 - iv. Not vaccinated with higher risk of exposure in work context or schools (frequency according to level of incidence) and in outbreaks assessments.

- v. Optional (free) for CDC (Centre for Disease Prevention and Control).
 - vi. Applicable to the entire population (e.g., not vaccinated, fully vaccinated or recovered).
- (e) Digital certificate COVID-19 (PCR 72h/TRAg 48h).
 - (f) Free testing and tracking integration
4. *Internal mobility measures:*
- (a) Curfew.
 - (b) More constrained Curfew.
 - (c) Lockdown.
 - (d) Remote working and remote teaching.
 - (e) Services restrictions.
 - (f) Events canceled (depending on the type).
 - (g) Transportation restrictions.
 - (h) Local administration restrictions.
5. *Individual behavioural rules:*
- (a) Hand sanitising
 - (b) Face mask:
 - i. Infected/high-risk contacts.
 - ii. Health-care professionals/vulnerable people (e.g., immunosuppressed), optional for the remaining population.
 - iii. In closed spaces/gatherings/in open spaces when physical distancing $> 2m$ not possible.
 - iv. Always.
6. *Social behavioural rules:*
- (a) Physical safety distancing (> 2 meters).
 - (b) Indoor ventilation (it is also a social behavioural rule for groups).
7. *Other measures over non COVID-19-related activity:*
- (a) Low-medium impact.
 - (b) High impact.

Now, we can associate the measures to each state. The higher the category, the more strict shall be the measures/rules. In the alarm and alert states the policy makers should intervene quickly since the critical state can be quickly reached, given the way the value functions were constructed.

C_1 (green): *Baseline state*. Measures: 1, 3.a.b.c(all), 5.a, 5.b.i., and 6.b.

C_2 (light green): *Residual state*. Additional measures: 3.d.i.ii.iii and 5.b.ii.

C_3 (yellow): *Alarm state*. Additional measures: 3.d.iv.v, 4.d.e.f.g.h, 5.b.iii, and 6.a.

C_4 (orange): *Alert state*. Additional measures: 4.a, 5.b.iv, and 7a.

C_5 (red): *Critical state*. Additional measures: 4.b, 3.d.vi, and 7b.

C_6 (dark red): *Break state (emergency)*. Additional measures: 4c.

For implementation purposes, the measures and recommendations must be detailed. They must be adjusted during the pandemic evolution. This is not a static chromatic model with respect to the measures and recommendations associated with each state.

5. Conclusions

In this paper we proposed an innovative application of a MAVT additive model for building an assessment composite indicator (PACI) and a chromatic ordinal classification system of the utmost importance to help in the management of the pandemic COVID-19 in Portugal. This indicator was built by following a sociotechnical co-constructive interactive process between the CCIST and the GCMO teams, and to the best of our knowledge is the first MAVT model proposed to analyse the evolution and mitigate the impacts of COVID-19 in the world. It was designed with the particular purpose of answering several questions of the Portuguese people, namely: How is the pandemic evolving in the country? In which pandemic state are we currently? What measures/recommendations should we follow? What is the impact of the vaccination plan established by the government? All these questions have been fully answered, and our indicator had a strong acceptance in Portugal. It still continues to be followed and frequently mentioned in the *media*, even though the Portuguese health authorities did not adopt it officially as another indicator for effective policy/decision-making in the country. Despite the fragility points related with the data and the construction of the aggregation model itself, it comprises several advantages recognized by academics, opinion-makers, *media*, and Portuguese people in general. Besides, even the technicalities of the method, including the computations, can be reproducible for any reader with the basic knowledge on mathematics knowledge. The parameters of the aggregation model can be adjusted if justified during the pandemic evolution. This comprises the shape of the value functions, the weights, the cut-off lines, and the reference levels, in particular the critical level (if the number of beds in intensive care increases it is normal the critical level related with the fifth criterion will change accordingly). Also, the formulas of the criteria model can be adjusted or replaced by more suitable ones (this also implies changes in the aggregation model).

The flexibility of the PACI mode opens several avenues for possible future research:

- This is an open model, in the sense it can easily accommodate the inclusion of more criteria and even more pillars to account for other points of view, as for example the economic impacts of the pandemic. Other aspects, as for example, the interaction effect between/among criteria are also possible, but they are more sophisticated and require adaptations of the judgments assessment techniques, as PaCo-DCM in case of some additional multiplicative terms (see Keeney and Raiffa 1993). In the case of the Choquet aggregation model, PaCo-DCM could be more or less easy to adapt (see Bottero et al. 2018), but the fact of having all the value functions between 0 and 1, would require a re-scaling.

- This model can be applied to all territorial units (country, regions, counties, sets of counties, etc.) with available data and possibly with some readjustments of the critical levels of some criteria.
- The scalability to other countries is also a possibility, but all the criteria would have to be reconsidered, as well as all the levels, in particular, the baseline and critical levels, and the cut-off lines. The comparison with other countries would be of a great importance for analysing the impact of different measures taken by different countries.
- The model can be applied to other diseases, and other problems in health, and even in different sectors where it is relevant the construction of composite indicators.
- The model has also forecasting capabilities. It only needs to have good estimates of the raw data $N(\cdot)$, $O(\cdot)$, $H(\cdot)$, and $U(\cdot)$.
- One of the most interesting avenues for future research is that of making use of constructive preference-learning techniques as some adaptations of the GRIP method (see Figueira et al. 2009) for building the composite indicator. This is a kind of machine learning based tool that infers representative value functions from examples provided by the experts. After this first study, the experts have now a much better understanding of the whole problem, and can easily provide a “good” set of examples for helping in the construction of the model parameters, value functions, weights, and even the cut-off lines.

The proposal of the PACI has been possible thanks to collaborative project between the CCIST team and the GCOM team, that still continues to carry out the research proposed in the previous listed avenues for future work. As a final remark, let us point out that this new indicator is not necessarily a competitor of the RM, both can be used at the same time to better inform the Portuguese health authorities in making their decisions.

Acknowledgements

This paper is dedicated to the memory of our friend and colleague Carlos Alves (Full Professor of Instituto Superior Técnico). The authors would like to acknowledge all the members of Crisis Office of Portuguese Medical Association (GCOM) for their valuable contributions, Professor Fernando Mira da Silva (IST) for providing the conditions to create the PACI webpage at the informatics platform of IST, Romana Borja-Santos, Ana Rodrigues, and Beatriz Santiago, for their appreciated and priceless help during all this process, and the anonymous people who helped us in validating the model. José Rui Figueira acknowledges the project PTDC/EGE-OGE/30546/2017 (hSNS: Portuguese public hospital performance assessment using a multi-criteria decision analysis framework) supported by national funds through Fundação para a Ciência e a Tecnologia (FCT), and Ana Paula Serro the project LISBOA-01-0145-FEDER-072536 (CAPTURE - Use of functionalized particles for enrichment and efficient detection of SARS-CoV-2 in clinical and environmental samples from Programa Operacional Regional de Lisboa). Henrique M. Oliveira was partially supported through FCT funding for CEMAT projects with reference UID/MAT/04459/2020.

References

- Belton, V., Stewart, T.J., 2002. *Multiple Criteria Decision Analysis: An Integrated Approach*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Bottero, M., Ferretti, V., Figueira, J.R., Greco, S., Roy, B., 2018. On the Choquet multiple criteria preference aggregation model: Theoretical and practical insights from a real-world application. *European Journal of Operational Research* 271, 120–140.
- Bouyssou, D., Pirlot, M., 2016. Conjoint measurement tools for MCDM: A brief introduction, in: Greco, S., Ehrgott, M., Figueira, J. (Eds.), *Multiple Criteria Decision Analysis*. Springer Science+Business Media New York, New York, NY, USA, pp. 97–151.
- Brans, J.-P., De Smet, Y., 2016. PROMETHEE methods, in: Greco, S., Ehrgott, M., Figueira, J. (Eds.), *Multiple Criteria Decision Analysis*. Springer Science+Business Media New York, New York, NY, USA, pp. 187–219.
- Corrente, S., Figueira, J.R., Greco, S., 2021. Pairwise comparison tables within the deck of cards method in multiple criteria decision aiding. *European Journal of Operational Research* 291, 738–756.
- Bana e Costa, C., De Corte, J.-M., Vansnick, J.-C., 2016. Mathematical foundations of MACBETH, in: Greco, S., Ehrgott, M., Figueira, J. (Eds.), *Multiple Criteria Decision Analysis*. Springer Science+Business Media New York, New York, NY, USA, pp. 421–463.
- Dinis, D., Figueira, J.R., Teixeira, A.P., 2021. A multiple criteria approach for ship risk classification: An alternative to the Paris MoU Ship Risk Profile. *CoRR abs/2107.07581*. 2107.07581.
- Doumpos, M., Zopounidis, C., 2002. *Multicriteria Decision Aid Classification Methods*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Dubois, D., Perny, P., 2016. A review of fuzzy sets in decision sciences: Achievements, limitations and perspectives, in: Greco, S., Ehrgott, M., Figueira, J. (Eds.), *Multiple Criteria Decision Analysis*. Springer Science+Business Media New York, New York, NY, USA, pp. 637–691.
- Dyer, J., 2016. Multiattribute utility theory (MAUT), in: Greco, S., Ehrgott, M., Figueira, J.R. (Eds.), *Multiple Criteria Decision Analysis*. Springer Science+Business Media New York, New York, NY, USA, pp. 285–314.
- El Gibari, S., Gómez, T., Ruiz, F., 2019. Building composite indicators using multicriteria methods: A review. *Journal of Business Economics* 89, 1–24.
- Figueira, J.R., Greco, S., Roy, B., 2021. ELECTRE-SCORE: A first outranking based method for scoring actions. *European Journal of Operational Research* DOI: 10.1016/j.ejor.2021.05.017.
- Figueira, J.R., Greco, S., Słowiński, R., 2009. Building a set of additive value functions representing a reference preorder and intensities of preference: GRIP method. *European Journal of Operational Research* 195, 460–486.

- Figueira, J.R., Mousseau, V., Roy, B., 2016. ELECTRE methods, in: Greco, S., Ehrgott, M., Figueira, J. (Eds.), *Multiple Criteria Decision Analysis*. Springer Science+Business Media New York, New York, NY, USA, pp. 155–185.
- Figueira, J.R., Roy, B., 2002. Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. *European Journal of Operational Research* 139, 317–326.
- Grabisch, M., Labreuche, Ch., 2016. Fuzzy measures and integrals in MCDA, in: Greco, S., Ehrgott, M., Figueira, J. (Eds.), *Multiple Criteria Decision Analysis*. Springer Science+Business Media New York, New York, NY, USA, pp. 553–603.
- Greco, S., Matarazzo, B., Słowiński, R., 2016. Decision rule approach, in: Greco, S., Ehrgott, M., Figueira, J. (Eds.), *Multiple Criteria Decision Analysis*. Springer Science+Business Media New York, New York, NY, USA, pp. 497–552.
- Keeney, R.L., Raiffa, H., 1993. *Decisions with Multiple Objectives*. Cambridge University Press, New York, NY, USA.
- Keeney, R.L., 1992. *Value-Focused Thinking: A Path to Creative Decisionmaking*. Harvard University Press, Cambridge, MA, USA.
- Koch, R., 2020. Erläuterung der schätzung der zeitlich variierenden reproduktionszahl r. *Memoir R/7-Tages-R* (15.5.2020).
- Martcheva, M., 2015. *An Introduction to Mathematical Epidemiology*. Springer Science+Business Media LLC New York, New York, NY, USA.
- Martel, J.-M., Matarazzo, B., 2016. Other outranking approaches, in: Greco, S., Ehrgott, M., Figueira, J.R. (Eds.), *Multiple Criteria Decision Analysis*. Springer Science+Business Media New York, New York, NY, USA, pp. 221–282.
- Moshkovich, H., Mechitov, A., Olson, D., 2016. Verbal decision analysis, in: Greco, S., Ehrgott, M., Figueira, J. (Eds.), *Multiple Criteria Decision Analysis*. Springer Science+Business Media New York, New York, NY, USA, pp. 605–636.
- Mousseau, V., Figueira, J.R., Dias, L.C., Gomes da Silva, C., J.C.N., C., 2003. Resolving inconsistencies among constraints on the parameters of an MCDA model. *European Journal of Operational Research* 147, 72–93.
- Pictet, J., Bollinger, D., 2008. Extended use of the cards procedure as a simple elicitation technique for MAVT. application to public procurement in Switzerland. *European Journal of Operational Research* 185, 1300–1307.
- Roy, B., 1996. *Multicriteria Methodology for Decision Aiding*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Roy, B., 2010. Robustness in operational research and decision aiding: A multi-faceted issue. *European Journal of Operational Research* 200, 629–638.

- Roy, B., Figueira, J.R., Almeida-Dias, J., 2014. Discriminating thresholds as a tool to cope with imperfect knowledge in multiple criteria decision aiding: Theoretical results and practical issues. *Omega - The International Journal of Management Science* 43, 9–20.
- Saaty, T.L., 2016. The Analytic Hierarchy and Analytic Network Processes for the measurement of intangible criteria and for decision-making, in: Greco, S., Ehrgott, M., Figueira, J. (Eds.), *Multiple Criteria Decision Analysis*. Springer Science+Business Media New York, New York, NY, USA, pp. 363–419.
- Simos, J., 1989. *L'Évaluation Environnementale: Un Processus Cognitif Négocié*. Ph.D. thesis. École Polytechnique Fédérale de Lausanne (EPFL), Suisse.
- Siskos, E., Tsotsolas, N., 2015. Elicitation of criteria importance weights through the Simos method: A robustness concern. *European Journal of Operational Research* 246, 543–553.
- Siskos, Y., Grigoroudis, E., Matsatsinis, N.F., 2016. UTA methods, in: Greco, S., Ehrgott, M., Figueira, J. (Eds.), *Multiple Criteria Decision Analysis*. Springer Science+Business Media New York, New York, NY, USA, pp. 315–361.
- Tsoukiàs, A., Figueira, J.R. (Eds.), 2006. RAND Memorandum – 5868 by Howard Raiffa [Special Issue], *Journal of Multi-Criteria Decision Analysis* 14 (113).
- von Winterfeldt, D., Edwards, W., 1986. *Decision Analysis and Behavioral Research*. Cambridge University Press, New York, NY, USA.
- Zopounidis, C., Doumpos, M., 2002. Multicriteria classification and sorting methods: A literature review. *European Journal of Operational Research* 138, 229–246.

Appendix

In the first part of this appendix, we provide some elements regarding the value functions for criteria, g_2 to g_4 .

1. Criterion g_2 (transmission). The performance levels, after discretizing the scale of g_2 , and the blank cards inserted in between consecutive levels, are as follows:

$$\{[0, 0.92]\}[0]\{.0.96\}[2]\{0.98\}[4]\{1.00\}[6]\{1.02\}[8]\{1.040\}$$

As in criterion g_1 , we also used some more levels to understand the evolution of the number of blank cards inserted in between consecutive levels. This was very similar as the the value function for criterion g_1 . The piecewise linear value function obtained is presented as follows:

$$v_2(x_{2t}) = \begin{cases} 4.34783x_2 & \text{if } x_{2t} \in [0.000, 0.920[\\ 600x_{2t} - 548 & \text{if } x_{2t} \in [0.920, 0.940[\\ 1000x_{2t} - 924 & \text{if } x_{2t} \in [0.940, 0.960[\\ 1400x_{2t} - 1308 & \text{if } x_{2t} \in [0.940, 0.980[\\ 1800x_{2t} - 1700 & \text{if } x_{2t} \in [0.980, 1.000[\\ 2200x_{2t} - 2100 & \text{if } x_{2t} \in [1.000, 1.020[\\ 2600x_{2t} - 2508 & \text{if } x_{2t} \in [1.020, 1.034[\\ 180 & \text{if } x_{2t} \in [1.034, +\infty[\end{cases} \quad (9)$$

As in g_1 , it also can be approximated by a quadratic function.

2. Criterion g_3 (lethality). This is a different type of value function as the ones constructed for criteria g_1 and g_2 . When discretizing the scale range of criterion g_3 and asking the experts for adding blank cards in between consecutive levels, they always considered the same number of blank cards. It means that this function is a linear function and the reason is obvious, as explained by the experts: one death is always very serious and does not depend on the place we are on the scale of this criterion, i.e., moving from one to two deaths has the same impact as moving from 49 to 50. The function can thus be presented as follows:

$$v_3(x_{3t}) = \begin{cases} \frac{200x_{3t}}{7.2} & \text{if } x_{3t} \in [0, 6.48[\\ 180 & \text{if } x_{3t} \in [6.48, +\infty[\end{cases} \quad (10)$$

The saturation level at 180 is used for making an upper level and limit the values of the indicator, and not because the number of deaths after a certain level has the same impact the number of deaths leading to the saturation/break level. The increasing of the number of deaths has always as strong impact in terms of the severity of the pandemic.

3. Criterion g_4 (General nursery beds). The scale of criterion g_4 is a discrete scale, which leads also to a discrete value function. The levels selected by the experts and blank cards inserted in between consecutive levels are presented below:

$$\{0\}[0]\{500\}[2]\{1000\}[4]\{1500\}[6]\{2000\}[8]\{2500\}[10]\{3000\}[12]\{3500\}$$

The value function can be states as follows:

$$v_4(x_{4t}) = \begin{cases} 0.008x_{4t} & \text{if } x_{4t} \in \{0, 1, \dots, 498, 499\} \\ 0.024x_{4t} - 8 & \text{if } x_{4t} \in \{500, 501, \dots, 999, 999\} \\ 0.04x_{4t} - 24 & \text{if } x_{4t} \in \{1000, 1001, \dots, 1498, 1499\} \\ 0.056x_{4t} - 48 & \text{if } x_{4t} \in \{1500, 1501, \dots, 1999, 1999\} \\ 0.072x_{4t} - 80 & \text{if } x_{4t} \in \{2000, 2001, \dots, 2498, 2499\} \\ 0.088x_{4t} - 120 & \text{if } x_{4t} \in \{2500, 2501, \dots, 2998, 2999\} \\ 0.104x_{4t} - 168 & \text{if } x_{4t} \in \{3000, 3001, \dots, 3344, 3345\} \\ 180 & \text{if } x_{4t} \in \{3346, 3347, \dots\} \end{cases} \quad (11)$$

It is a discrete function, but it has the same kind of “shape” and behaviour as the functions for criteria g_1 and g_2 .

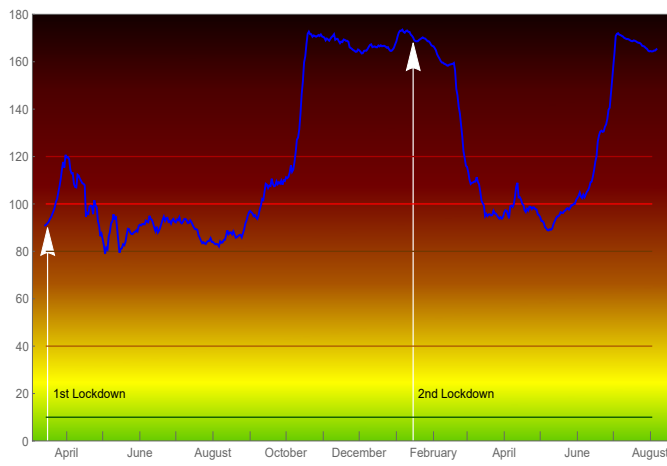
4. Criterion g_5 (Intensive care beds). As in the previous case, this scale is also a discrete one, which leads to a discrete value function too. The levels selected by the experts and the number of blank cards inserted in between consecutive levels are presented below:

$$\{0\}[0]\{40\}[2]\{80\}[4]\{120\}[6]\{160\}[8]\{200\}[10]\{240\}[12]\{280\}$$

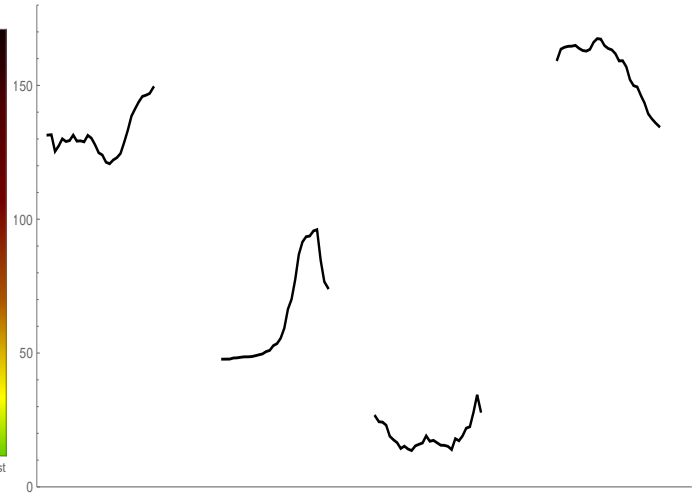
From the previous information and using PaCo-DCM we can derive the following function.

$$v_5(x_{5t}) = \begin{cases} 0.1x_{5t} & \text{if } x_{5t} \in \{0, 1, \dots, 39, 40\} \\ 0.3x_{5t} - 8 & \text{if } x_{5t} \in \{40, 41, \dots, 78, 79\} \\ 0.5x_{5t} - 24 & \text{if } x_{5t} \in \{80, 81, \dots, 118, 119\} \\ 0.7x_{5t} - 48 & \text{if } x_{5t} \in \{120, 121, \dots, 158, 159\} \\ 0.9x_{5t} - 80 & \text{if } x_{5t} \in \{160, 161, \dots, 198, 199\} \\ 1.1x_{5t} - 120 & \text{if } x_{5t} \in \{200, 201, \dots, 238, 239\} \\ 1.3x_{5t} - 168 & \text{if } x_{5t} \in \{240, 241, \dots, 266, 267\} \\ 180 & \text{if } x_{5t} \in \{268, 269, \dots\} \end{cases} \quad (12)$$

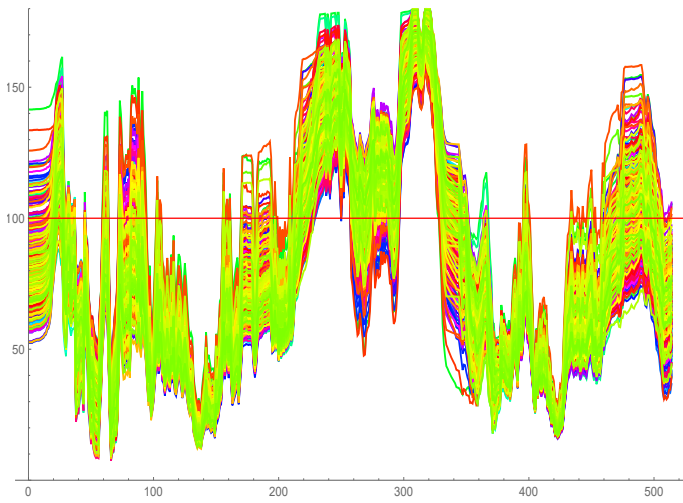
Its “shape” and behaviour are similar to the previous value function.



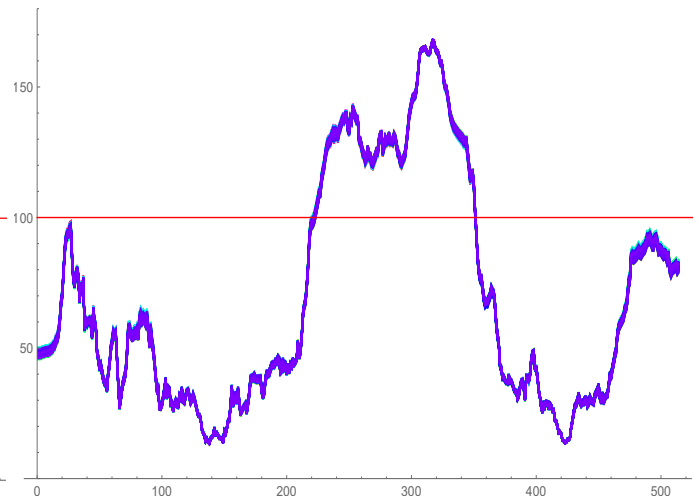
(a) RM indicator evolution (2021-08-07)



(b) Four unordered moments of the evolution line

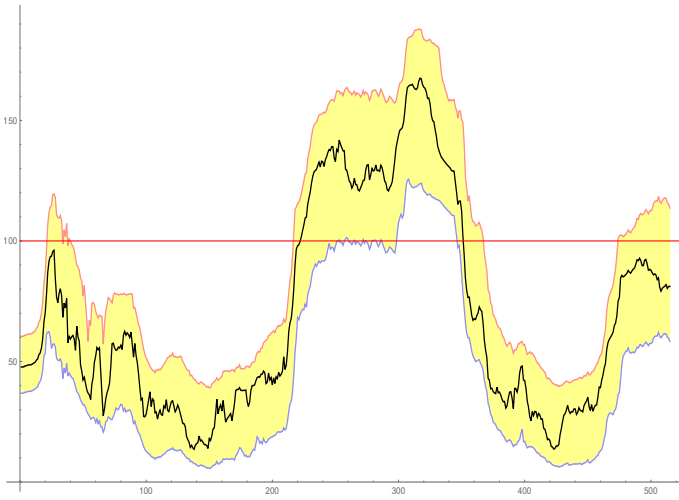


(c) Simulating a drastic change in weights (2021-08-07)

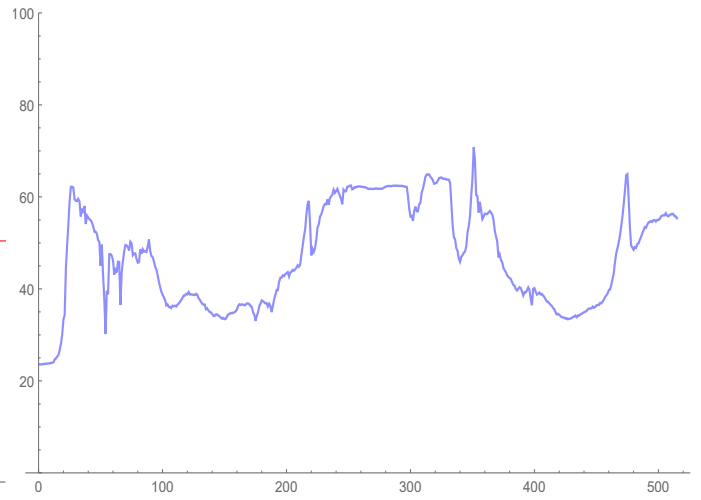


(d) Simulating a realistic change in weights (2021-08-07)

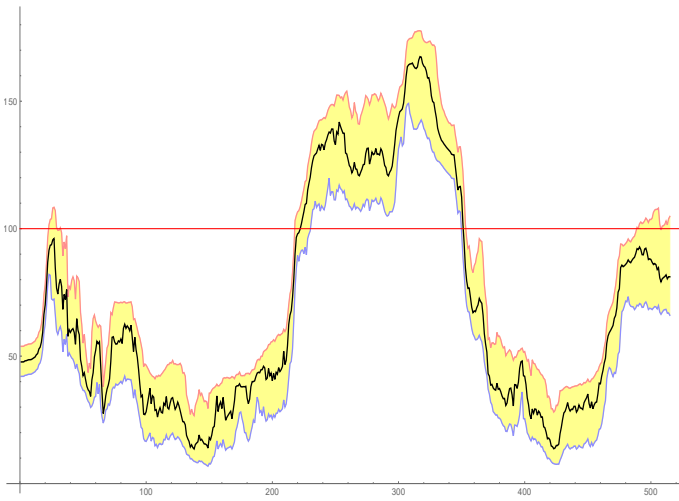
Figure 6: Validation and simulation analyses



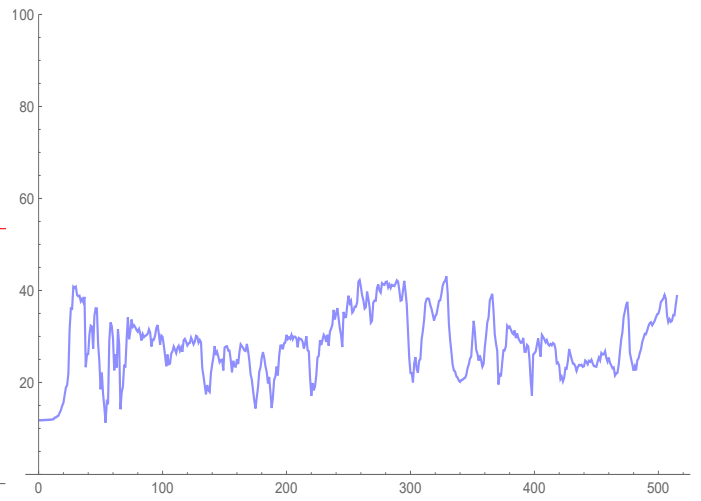
(a) Robustness of a strong variation (2021-08-07)



(b) Average of the strong variation (2021-08-07)



(c) Robustness of a realistic variation (2021-08-07)



(d) Average of the realistic variation (2021-08-07)

Figure 7: Robustness analysis

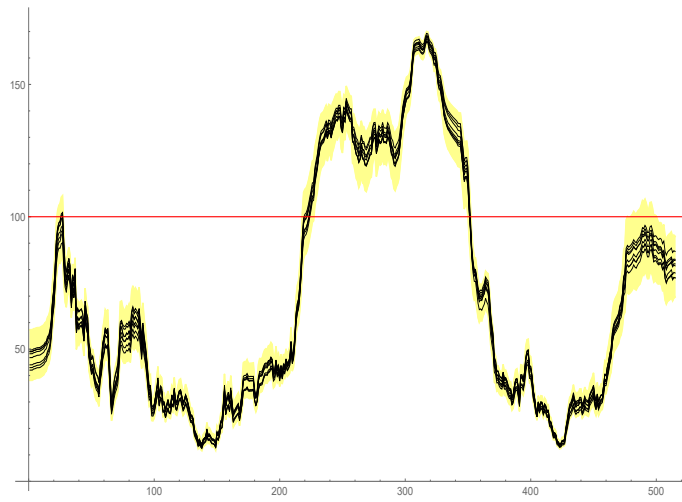


Figure 8: Simulation *versus* Robustness (2021-08-07)