



**HAL**  
open science

# Inferência dos Estágios Iniciais da COVID-19 em Portugal

Fernando C Batista, Americo Cunha Jr

► **To cite this version:**

Fernando C Batista, Americo Cunha Jr. Inferência dos Estágios Iniciais da COVID-19 em Portugal. XXIII Congresso Brasileiro de Automática (CBA 2020), Nov 2020, Congresso Virtual, Brazil. hal-02975810v2

**HAL Id: hal-02975810**

**<https://hal.science/hal-02975810v2>**

Submitted on 4 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

# Inferência dos Estágios Iniciais da COVID-19 em Portugal

Fernando C. Batista\*, Americo Cunha Jr.\*\*

\*ESTG, CDRSP, Instituto Politécnico de Leiria, P-2411-901, Leiria, Portugal  
(e-mail: fernando.batista@ipleiria.pt).

\*\*Universidade do Estado do Rio de Janeiro, Rio de Janeiro-RJ, 20550-900 Brasil  
(e-mail: americo.cunha@uerj.br)

---

**Abstract:** Currently, the new SARS-CoV-2 coronavirus is a threat to the world population health. The apparent rapid spread of the virus has created a pandemic, and countries around the world are struggling with an increase in infected cases. In this sense, this article is part of that effort to try to better understand the initial dynamics of COVID-19 in Portugal, seeking estimate a probable date for the epidemic outbreak beginning. Using the information provided by the health authorities and using the classic logistic model, a model for the initial stages of the disease in the country is presented with the notion of its respective uncertainty. This model allows to infer an interval for the occurrence of the first case of COVID-19 in Portugal that is prior to the official notification.

**Resumo:** Atualmente, o novo coronavírus SARS-CoV-2 é uma ameaça à saúde da população mundial. A aparente rápida disseminação do vírus criou uma pandemia, e países em todo o mundo estão a lutar com um aumento de casos infetados. Neste sentido, o presente artigo insere-se nesse esforço de tentar compreender melhor a dinâmica inicial da COVID-19 em Portugal, em particular, com vistas para estimar uma provável data de início do surto epidêmico. Utilizando a informação veiculada pelas autoridades de saúde e usando o modelo logístico clássico, é apresentada uma modelagem dos estágios iniciais da doença no país, com a noção da sua respetiva incerteza. Esse modelo permite inferir um intervalo para a ocorrência do primeiro caso de COVID-19 em Portugal que é anterior à notificação oficial.

**Keywords:** COVID-19 in Portugal, SARS-CoV-2, Mathematical epidemiology, Epidemic model, Logistic regression

**Palavras-chaves:** COVID-19 em Portugal, SARS-CoV-2, Epidemiologia matemática, Modelo epidêmico, Regressão logística

---

## 1. INTRODUÇÃO

Em finais de 2019 foram registados os primeiros casos positivos de COVID-19 em Hubei (China) e de forma surpreendente foram registados os primeiros casos positivos na Europa a 25 janeiro de 2020 e a 15 de fevereiro já havia registos em todos os continentes (<https://ourworldindata.org>). A 11 de março, a doença foi caracterizada pela Organização Mundial de Saúde (OMS) como uma pandemia. Neste dia o novo coronavírus já tinha provocado mais de 120 mil infetados em 114 países (WHO, 2020 e <https://ourworldindata.org>).

Os primeiros surtos na Europa foram em Itália na região da Lombardia onde o primeiro caso oficial detetado foi no dia 20 de fevereiro. Um estudo feito nesta região italiana a 20 de março apontava para que o COVID-19 tivesse chegado muito antes de 20 de fevereiro, porque encontraram um caso com início de sintomas a 1 de janeiro de 2020 (Cereda et al. 2020). Em França, um estudo retrospectivo dos testes de PCR realizados em 58 doentes internados com pneumonia, entre 2 de dezembro de 2019 e 16 de janeiro de 2020, foi detetado o vírus SARS-CoV-2 a um doente de 42 anos que foi internado

de emergência a 27 de dezembro de 2019 e já tinha sintomas há 4 dias (Deslandes et. al., 2020).

O primeiro caso positivo oficial em Portugal foi registado no dia 2 de março na região do Porto (<https://www.dssg.pt/>), nesta data já existiam vários casos positivos nos países vizinhos. Dia 11 março de 2020 foi diagnosticado COVID-19 a dois idosos no hospital de Santa Maria em Lisboa, um dos idosos já estava internado há uma semana no hospital (Diário de Notícias, 2020). O que demonstra, pelo tempo de incubação e o aparecimento de sintomas que já haveria infeção na população junto deste idoso pelo menos desde meados de fevereiro de 2020.

Por outro lado, a Direção Geral de Saúde (DGS) informou dia 18 de abril que, “entre 21 de fevereiro e 16 de março cada pessoa infetada contagiava, em média, 2,08 pessoas”, ou seja, já admitia que havia infeções a 21 de fevereiro (DGS, 2020a).

O Delatorre et. al, fizeram uma estimativa do início da transmissão em vários países, mas não incluíram Portugal. Para a Espanha, que faz fronteira com Portugal, estimaram o dia 28 de janeiro para o início da infeção, (Delatorre E, Mir D, Gräf T, Bello G. 2020)

Numa fase inicial da doença, em março, houve vários autores que utilizaram modelos SIR para preverem a evolução da infeção em Portugal (Pais RJ and Taveira N., 2020, Teles, P. 2020). O nível de incerteza era grande e os cenários apresentados foram muito diferentes da realidade. Nenhum, da literatura consultada, entrou em conta com a possibilidade de a infeção já existir antes do primeiro caso oficial notificado.

Existe, portanto, um conjunto de evidências de que o vírus causador da COVID-19 já estava na Europa muito antes dos primeiros casos positivos oficiais. A infeção provocada pelo vírus já estaria provavelmente a provocar casos clínicos e talvez mortes, mas por alguma razão isso não provocou nenhum alarme clínico.

Onde está então, a verdadeira data de início da epidemia de COVID-19 em Portugal? Procurando uma possível resposta, o presente trabalho utiliza uma abordagem estatística paramétrica, baseada num regressor logístico, que ajusta os dados oficiais de registo de casos positivos. Três indicadores são utilizados para determinar uma curva epidémica que bem ajusta os casos de COVID-19 registrados em Portugal, sem, contudo, levar em consideração atrasos ou subnotificações. São eles o rácio entre testes positivos e testes negativos, a estabilização dos sintomas dos casos positivos registrados e o pico de início de sintomas. Esse regressor permite fazer uma projeção para trás no tempo, inferindo um intervalo para a provável data de início da epidemia portuguesa.

## 2. DADOS

Todos os dados utilizados foram extraídos dos boletins da DGS que foram agregados no serviço web *GitHub* (<https://github.com/dssg-pt/covid19pt-data>) pela organização não governamental, *Data Science for Social Good Portugal* (<https://www.dssg.pt/>). Foram usados os dados dos casos confirmados do dia 2 de março de 2020 a 6 de outubro de 2020.

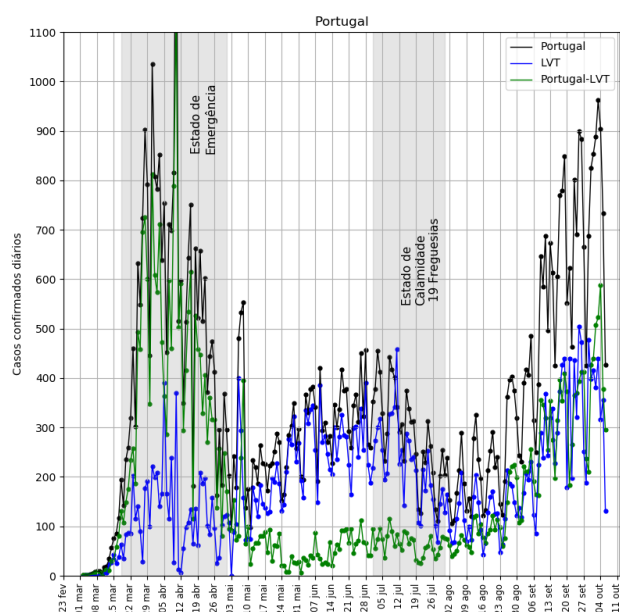


Fig. 1 - Incidência da COVID-19 em Portugal

No dia 18 de março de 2020 foi decretado o estado de emergência em Portugal, através do Decreto do Presidente da República n.º 14-A/2020, de 18 de março (Presidência da República, 2020a).

No dia 2 de abril, o Presidente da República renovou a declaração do estado de emergência até 17 de abril através do Decreto do Presidente da República n.º 17-A/2020 (Presidência da República, 2020b).

No dia 16 de abril, o Presidente da República renovou a declaração do estado de emergência até 2 de maio através do Decreto do Presidente da República n.º 20-A/2020, (Presidência da República, 2020c).

A DGS disponibilizou diariamente os dados sobre a COVID-19 em cada uma das 7 regiões do país, Norte, Centro, Lisboa e Vale do Tejo (LVT), Alentejo, Algarve, Madeira e Açores. A região Norte e a região LVT representam cada uma cerca de 35% da população total portuguesa. Estas duas regiões, pelo facto de terem uma maior densidade populacional, podem justificar a maior incidência da doença que se verificou até ao presente momento.

A COVID-19 manifestou-se de maneira diferente nestas regiões. Num primeiro momento houve uma maior incidência em todas as regiões, exceto na LVT e a partir do início de maio foi a região LVT que teve uma maior incidência. A razão disto ter acontecido pode ser o facto de a região norte ser mais industrializada e poder ter tido um contacto inicial com regiões do norte de Itália onde existiram os primeiros registos de COVID-19.

No dia 26 de junho, a Resolução do Conselho de Ministros n.º 51-A/2020 declara, na sequência da situação epidemiológica até 14 de julho, a situação de calamidade para 19 freguesias da zona norte da Área Metropolitana de Lisboa (Presidência do Conselho de Ministros, 2020).

É visível na Fig. 1 que ao longo do mês de março temos um aumento exponencial do número de casos positivos registrados. O país esteve em confinamento no Estado de Emergência de 18 de março a 2 de maio. E as 19 freguesias já referidas estiveram em Situação de Calamidade.

## 3. INDICADORES

O aumento do número de casos só estabilizou no fim do mês de março e houve a respetiva descida ao longo do mês de abril. Naturalmente esta descida já fez parte do lado descendente da curva epidémica que bem ajusta a evolução natural da epidemia. Pretendemos saber em que dia foi encontrada essa curva. Esse dia foi estimado com os três indicadores seguintes.

### 3.1 Rácio entre testes positivos e testes negativos

A mudança de tendência do rácio entre testes positivos e testes negativos é reveladora de que encontramos os pontos que fazem parte da curva epidémica procurada. Ou seja, antes desta mudança não estávamos a testar verdadeiramente uma amostra significativa da população que desenvolveu a doença. Após

esse momento verificou-se um aumento considerável de testes negativos face ao número de testes positivos e o rácio cai.

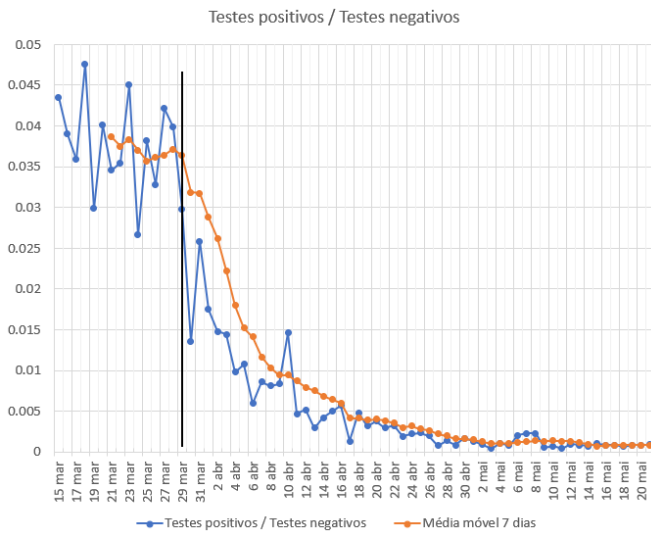


Fig. 2 - Rácio entre testes positivos e testes negativos

A DGS disponibilizou o número de testes negativos apenas a partir do dia 15 de março. Mesmo só com estes dados é possível ver na Fig. 2 uma mudança brusca do rácio no dia 29/30 de março.

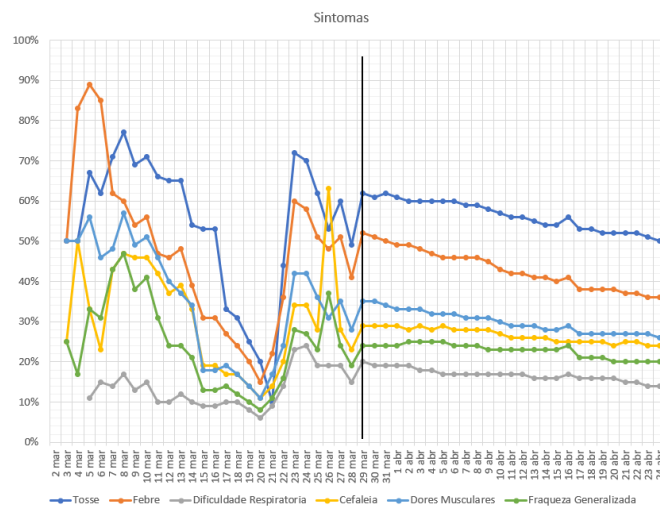


Fig. 3 Percentagem de sintomas dos casos positivos

### 3.2 Estabilização dos sintomas

No boletim diário da DGS é reportado diariamente informação percentual relativa a sintomas da grande maioria (+70%) dos casos confirmados. Os sintomas reportados são; febre, tosse, dificuldade respiratória, cefaleia, dores musculares e fraqueza generalizada (Fig. 3). Foi verificado que até dia 29 de março houve uma grande perturbação da percentagem de casos infetados diários que reportam cada um dos sintomas. A partir desta data os valores estabilizam e a evolução percentual de cada sintoma vai caindo ao longo do tempo de forma mais ou menos linear. Podemos estar perante alguma coincidência, mas esta estabilização parece ser indicadora de que, a partir do

dia 29 de março, já estamos a testar uma amostra significativa da população que está a desenvolver a doença e diariamente todos os casos tem percentualmente os mesmos sintomas que nos dias anteriores, com uma queda suave dia após dia.

Esta observação é mais um indício de que a partir do dia 29 de março os registos dos pontos positivos fazem parte de uma curva que é representativa dos dados de interesse.

### 3.3 Pico de início de sintomas

A DGS informou dia 18 de abril (DGS, 2020a) que o “Pico da incidência terá sido entre o dia 23 e 25 de março”. Este pico corresponde ao pico dos “casos observados com data de início de sintomas” que está presente no gráfico do relatório, Fig. 4 (DGS, 2020b). Esta curva tem uma antecipação dos casos positivos notificados de cerca de 5 dias, o que vem coincidir com o dia 29 de março, Fig. 4.

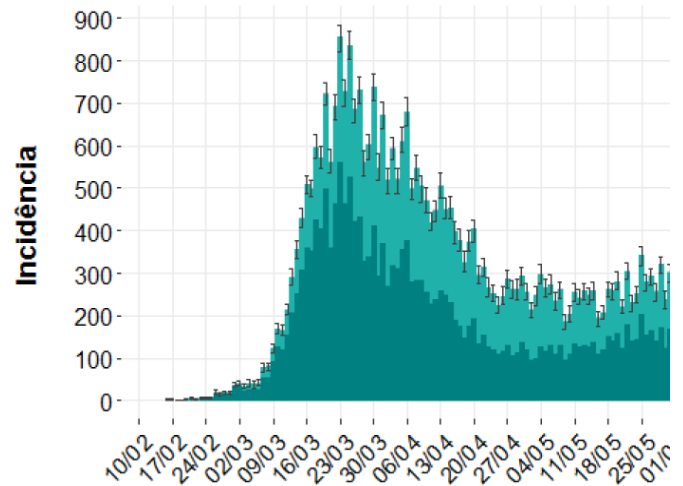


Fig. 4 Curva epidémica dos casos de infeção por SARS-CoV-2 corrigida para o atraso da notificação, em Portugal. (Verde escuro - casos observados com data de início de sintomas; verde claro - casos observados com data de início de sintomas imputada; cinzento - estimativa dos casos ocorridos, mas ainda não reportados). Autores: (DGS, 2020b)

## 4. MÉTODO

### 4.1 Média móvel

Os dados brutos da Fig. 1 têm presente a flutuação diária inerente ao processo de testagem e registo que existe ao longo do tempo. Para suavizar estas oscilações é frequente usar médias moveis de 3, 5 ou de 7 dias. No presente caso há um ciclo de aquisição de dados (testagem) com uma frequência semanal, neste sentido foi usado uma média móvel de 7 dias para suavizar o número de casos confirmados, ver Fig. 5.

### 4.2 Curva logística

Toda a problemática da análise de dados da área da epidemiologia carece de uma utilização de ferramentas de análise estatística avançada (Wasserman, L., 2004). Existem

modelos bastante complexos para abordar todo o problema (Cotta et al (2020), Tsallis C and Timakli U (2020), Kuhl (2020)), neste trabalho será utilizado um mais simples.

O modelo da curva de crescimento acumulado de Richards (Richards, 1959), a chamada curva logística generalizada, é um modelo de curva para estudos populacionais. Essa curva é utilizada em epidemiologia para previsão em tempo real de surtos de doenças (Hsieh YH. et al., 2004, Hsieh YH 2010).

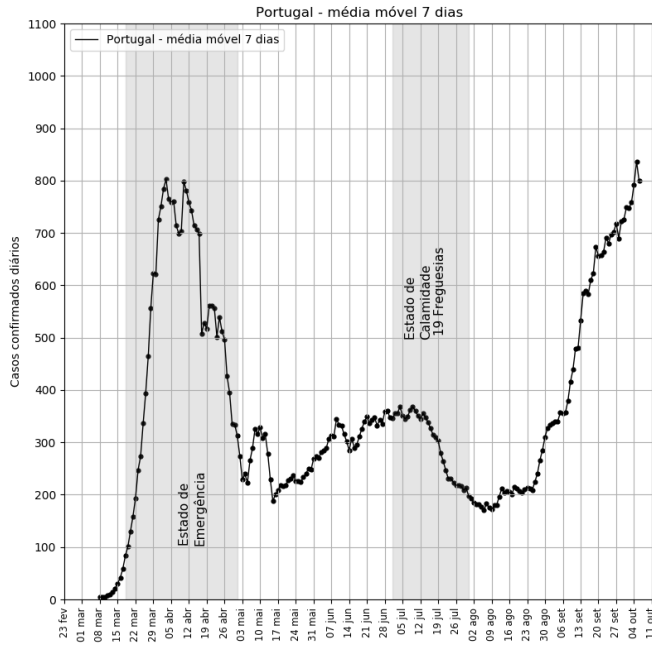


Fig. 5 - Incidência da COVID-19 em Portugal - média móvel de 7 dias

Uma das variantes da curva de crescimento acumulado de Richards encontrada na literatura é a curva logística clássica:

$$C(t) = \frac{K}{1+e^{-r(t-t_m)}}, \quad (1)$$

onde  $r, t_m$  e  $K$  são reais positivos, sendo  $r$  é a taxa de crescimento da infecção,  $t_m$  a posição no eixo horizontal do tempo do ponto de inflexão da curva e  $K$  representa o número final de infetados. Há autores que utilizam outras variantes mais complexas das curvas de Richards para modelar a propagação desta doença (Wu, K, 2020, Postnikov EB., 2020).

A curva de incidência ou a curva epidêmica, é obtida pela derivada da curva de crescimento acumulado:

$$I(t) = \frac{dC(t)}{dt} = rC(t) \left(1 - \frac{C(t)}{K}\right), \quad (2)$$

que, após uma simplificação, se escreve como:

$$I(t) = \frac{rKe^{-r(t-t_m)}}{(1+e^{-r(t-t_m)})^2}. \quad (3)$$

Será usada a equação (3) para fazer a regressão dos dados e assim estimar a curva epidêmica portuguesa. Desta forma o parâmetro,  $t_m$  representa na equação (3) a posição no eixo horizontal do tempo do ponto máximo da curva epidêmica.

### 4.3 Curva tri-logística

Ao longo do tempo da infecção de uma população, o vírus, pela dinâmica de movimentação dos infetados, dos recuperados e dos suscetíveis, vai infetar partes da população em tempos diferentes. Isto vai provocar ao longo do tempo alguma ondulação na curva epidêmica. É possível que o confinamento e o desconfinamento de partes da população num determinado momento, ajudem a explicar estas várias ondas.

Neste sentido, para modelar o mais correto possível a curva epidêmica é necessário tomar em conta esse efeito ondulatorio e assim “aplicar” ao longo do tempo várias curvas logísticas. Ou seja, a curva de crescimento acumulado total será descrita por uma soma de curvas logísticas.

No presente caso, em que os dados vão do dia 2 de março até ao dia 6 de outubro, vamos utilizar a soma de três curvas logísticas, ou seja, temos uma curva tri-logística (PS Meyer, 1994 tem um exemplo de uma curva bi-logística):

$$C(t) = C_1(t) + C_2(t) + C_3(t),$$

$$C(t) = \frac{K_1}{1+e^{-r_1(t-t_{m1})}} + \frac{K_2}{1+e^{-r_2(t-t_{m2})}} + \frac{K_3}{1+e^{-r_3(t-t_{m3})}}. \quad (4)$$

Logo a curva epidêmica é dada por

$$I(t) = \frac{r_1K_1e^{-r_1(t-t_{m1})}}{(1+e^{-r_1(t-t_{m1})})^2} + \frac{r_2K_2e^{-r_2(t-t_{m2})}}{(1+e^{-r_2(t-t_{m2})})^2} + \frac{r_3K_3e^{-r_3(t-t_{m3})}}{(1+e^{-r_3(t-t_{m3})})^2}. \quad (5)$$

### 4.4 Métricas da qualidade do ajuste

Coefficiente de determinação ( $R^2$ ) descreve a fração da variância total nos dados observados que pode ser explicada pelo modelo. O valor varia entre 0 e 1, e quanto maior for o valor, melhor será a concordância entre o modelo e os dados (Wasserman, L., 2004):

$$R^2 = \frac{\sum_{i=1}^N (I(t_i) - \bar{y})^2 - \sum_{i=1}^N (I(t_i) - y_i)^2}{\sum_{i=1}^N (I(t_i) - \bar{y})^2}. \quad (6)$$

Erro de raiz quadrática média (RMSE) representa o quadrado da medida do desvio médio entre os dados observados e os calculados pelo modelo. Pelo facto de as diferenças serem ao quadrado dá maior peso aos desvios maiores:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (I(t_i) - y_i)^2}. \quad (7)$$

Erro absoluto médio (MAE) pelo facto de não elevar ao quadrado os desvios dá o mesmo peso a todos os desvios, representará assim sempre um valor menor ou igual que o RMSE:

$$MAE = \frac{1}{N} \sum_{i=1}^N |I(t_i) - y_i|. \quad (8)$$

Nessas métricas, temos que:

- $y_i$  são os casos positivos registados no dia  $t_i$ ,
- $\bar{y}$  é o valor médio de todos os casos positivos diários,
- $I(t_i)$  é o valor da incidência calculada pelo modelo no dia  $t_i$



#### 4.5 Envelope de predição linear

Foi construído um envelope (banda) de predição para poder estimar o início da incidência (Wasserman, L., 2004):

$$I(t_i) \pm t_{N-2}(\gamma/2) \cdot S \cdot \sqrt{1 + \frac{1}{N} + \frac{(t_i - \bar{t})^2}{\sum_{j=1}^N (t_j - \bar{t})^2}}, \quad (9)$$

em que  $S$  é a raiz do quadrado médio residual, dada por:

$$S = \sqrt{\frac{\sum_{i=1}^N (I(t_i) - y_i)^2}{N-2}}, \quad (10)$$

e o  $t_{N-2}(\gamma/2)$  é o valor da distribuição para  $(1 - \gamma) \cdot 100\%$  de confiança. Como no presente caso não conhecemos a distribuição, foi assumido uma distribuição normal com uma fiabilidade de 95%.

#### 5. ESTIMATIVA DA CURVA EPIDÊMICA INICIAL

Utilizando os indicadores, podemos com alguma certeza admitir que todos os dados a partir do dia 29 de março fazem parte da “curva epidêmica estacionária”. Na Fig. 5 podemos ver que há uma alteração de tendência no dia 15 de maio. Observando a Fig 1 vimos que a infeção que estava centrada em todo o país, exceto na região LVT, passou a estar presente essencialmente na região LVT.

Para estimar a “curva epidêmica estacionária” inicial (Curva Epidêmica 1) foram assim utilizados apenas os dados de 29 de março a 15 de maio. Utilizando o regressor da equação (3) temos três parâmetros que devemos encontrar,  $r$ ,  $t_m$  e  $K$ . Olhando para os dados podemos reparar que todos esses pontos estão do lado descendente da curva, o que quer dizer que o pico da infeção se encontra no dia 29 de março ou antes.

Como não é possível saber, com os dados públicos, a localização do pico da “curva epidêmica estacionária”, serão admitidos três cenários que correspondem a três datas de pico:

- Cenário 1 - pico no próprio dia 29 de março;
- Cenário 2 - pico no dia 24 de março;
- Cenário 3 - pico no dia 19 março.

Fixando assim o parâmetro  $t_m$ , o regressor irá estimar a taxa de infeção  $r$  e o número final de infetados  $K$ .

#### 6. ESTIMATIVA DA CURVA EPIDÊMICA TOTAL

Para além da curva inicial de 29 de março a 15 de maio temos mais duas curvas, uma que vai até dia 4 de agosto e outra que irá apanhar os restantes dias. Para estimar a curva epidêmica destas três ondas (Curva Epidêmica 2) é usada a curva epidêmica (5) que deriva da curva tri-logística (4).

Teremos assim 8 parâmetros para determinar,  $r_1, K_1, r_2, t_{m2}, K_2, r_3, t_{m3}$  e  $K_3$ , o valor de  $t_{m1}$  terá o mesmo valor que o  $t_m$  de cada cenário da Curva Epidêmica 1.

Para facilitar o ajuste dos 8 parâmetros foram usados, como valores iniciais de cálculo para o regressor, os dados da tabela 1 e para cada uma das outras duas ondas, foram estimados os respetivos 3 parâmetros de forma individual, para ter assim 8 parâmetros próximos da solução do regressor.

### 7. RESULTADOS

#### 7.1 Curva Epidêmica 1

Podemos observar na Fig. 6 a Curva Epidêmica 1 do Cenário 1 e o respetivo envelope de predição de 95%. O regressor usou apenas os pontos do dia 29 de março até ao dia 15 de maio para calcular os três parâmetros. Os outros dois cenários podem ser vistos nas Figs 7 e 8. Como era de esperar a Curva Epidêmica 1 do Cenário 3 projeta o início da epidemia para mais tarde, Fig. 8.

Na tabela 1 estão apresentados os valores para os parâmetros dos três cenários da Curva Epidêmica 1 determinados pelo regressor (3).

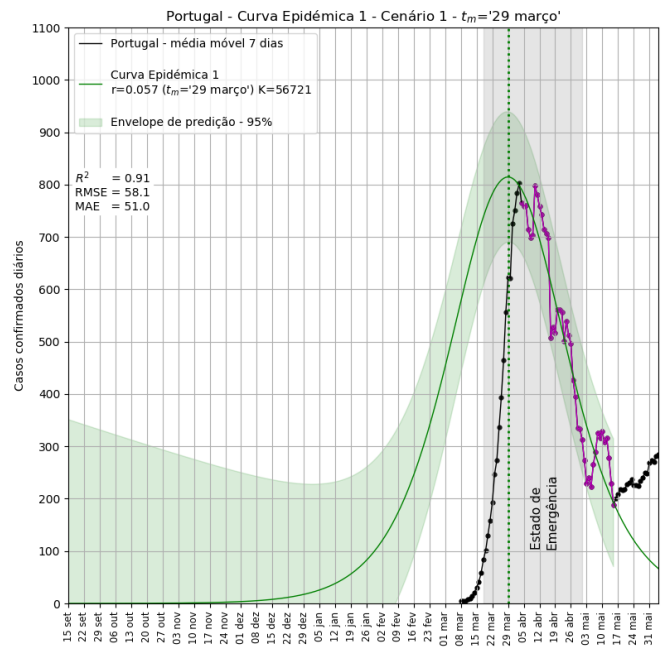


Fig. 6 Curva Epidêmica 1 - Cenário 1 - dia 29 de março

A incerteza dos parâmetros, longe dos pontos conhecidos, leva a que os limites superiores e inferiores de todos os envelopes dos três Cenários tenham uma grande amplitude.

Na Fig. 9 foi colocada a média dos três regressores de cada um dos três cenários da “Curva Epidêmica 1”. Foi também colocada a média dos três envelopes de predição. Com esse “preditor médio” temos um indicativo que o primeiro caso em Portugal pode ter começado em janeiro, antes do dia 26.

#### 7.2 Curva Epidêmica 2

O regressor (5) usou os pontos do dia 29 de março até ao dia 6 de outubro para calcular os oito parâmetros. Os parâmetros

$r_1$  e  $K_1$  da curva inicial estão na Tabela 2 e as respectivas curvas nas Figs. 10, 11 e 12.

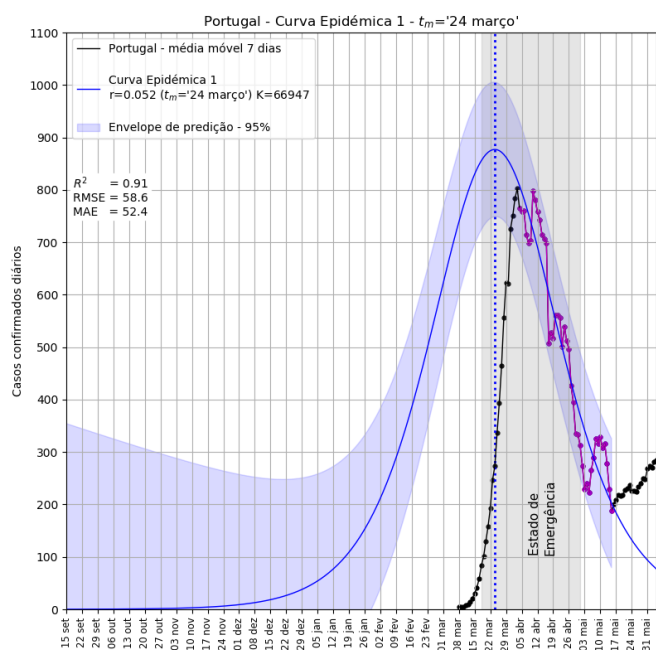


Fig. 7 Curva Epidêmica 1 - Cenário 2 - dia 24 de março

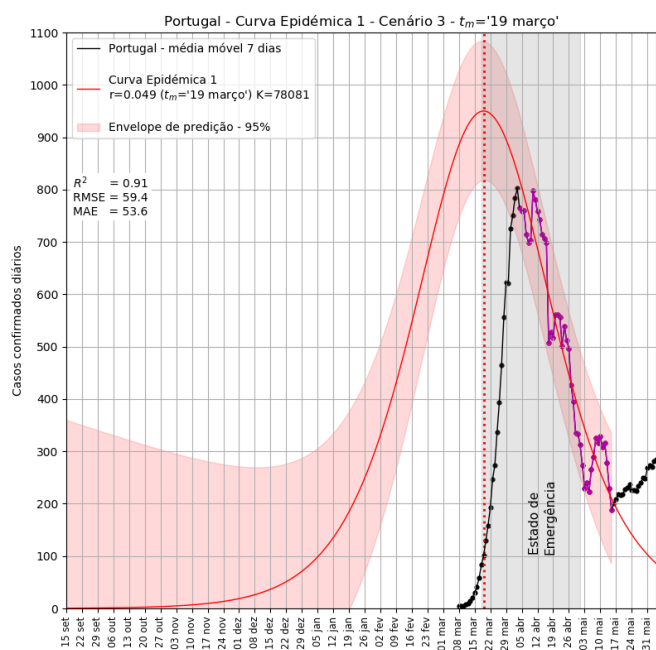


Fig. 8 Curva Epidêmica 1 - Cenário 3 - dia 19 de março

Tabela 1. Parâmetros - 3 cenários da Curva Epidêmica 1

$r$ [dia <sup>-1</sup> ]	$t_m$ [dia]	$K$ [pessoas]
0.057	(29 março)	56721
0.052	(24 março)	66947
0.049	(19 março)	78081

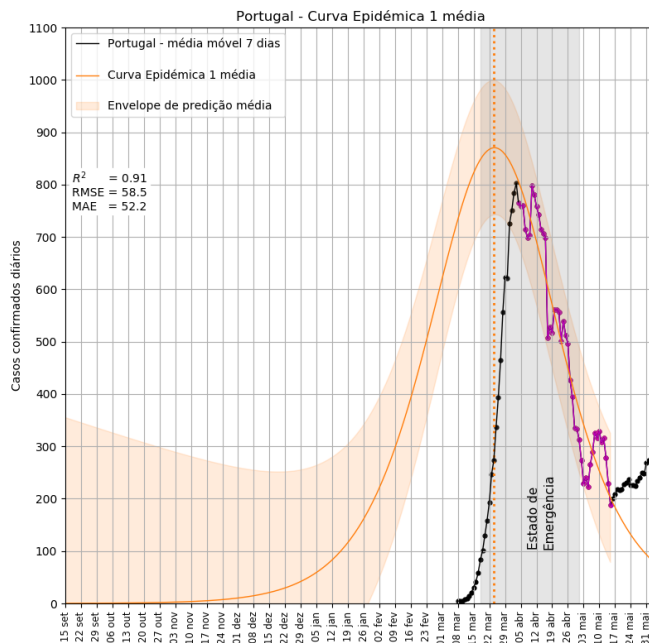


Fig. 9 Média dos três cenários da Curva Epidêmica 1

A Curva Epidêmica 2 do cenário 3, Fig.12, também projeta o início da epidemia para mais tarde do que os restantes cenários apresentados na Fig. 10 e Fig. 11.

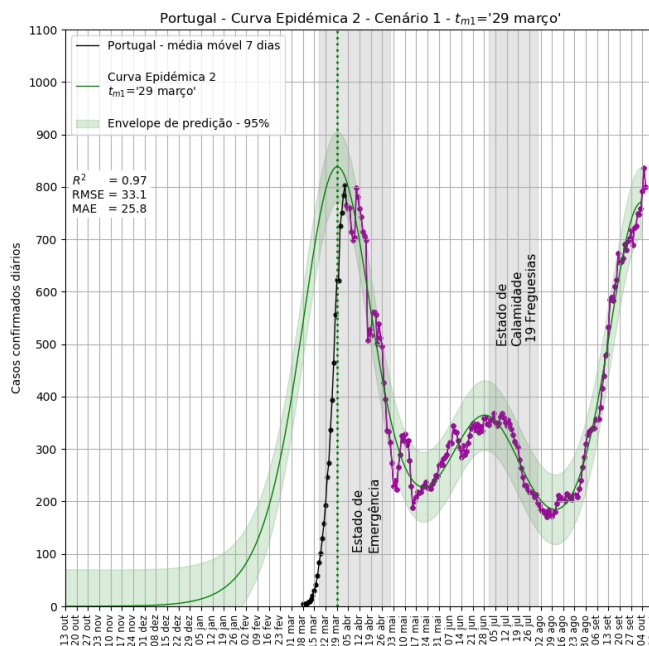


Fig. 10 Curva Epidêmica 2 - Cenário 1 - dia 29 de março

Tabela 2. Parâmetros - 3 cenários da Curva Epidêmica 2

$r_1$ [dia <sup>-1</sup> ]	$t_{m1}$ [dia]	$K_1$ [pessoas]
0,065	(29 março)	51056
0,058	(24 março)	62147
0,053	(19 março)	74119

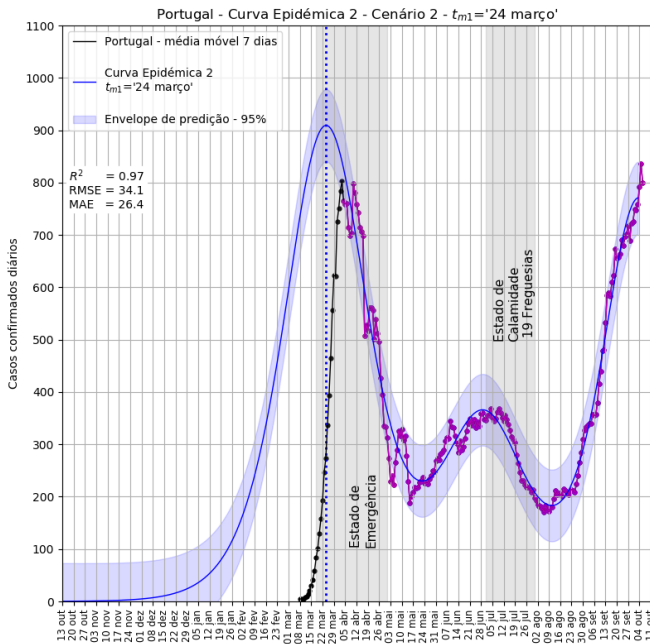


Fig. 11 Curva Epidêmica 2 - Cenário 2 - dia 24 de março

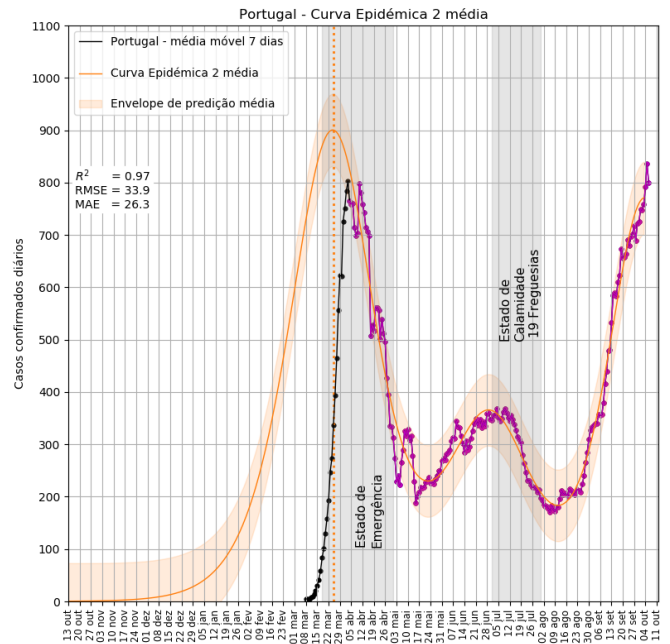


Fig. 13 Os três cenários da Curva Epidêmica 2

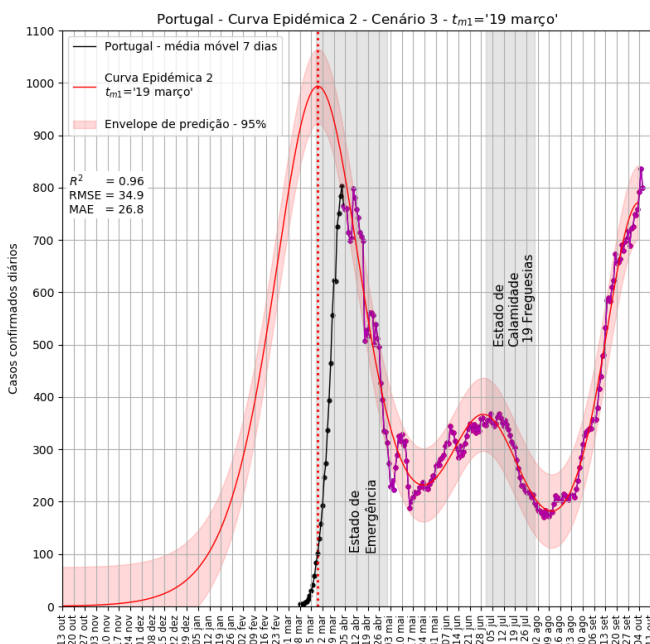


Fig. 12 Curva Epidêmica 2 - Cenário 3 - dia 19 de março

A curva gerada pelo regressor tem um coeficiente de determinação ( $R^2$ ) superior ao regressor da Curva Epidêmica 1. Podemos perceber esta diferença pelo facto de os dados terem uma menor dispersão a partir de maio.

Com esta menor dispersão o envelope criado para a Curva Epidêmica 2 é mais estreito, pois revela menor incerteza. E desta forma o limite inferior do envelope cruza o eixo do tempo numa data mais tardia, do que o correspondente limite inferior do envelope, do Cenário da Curva Epidêmica 1.

A Curva Epidêmica 1 é gerada com a informação que descreve mais correctamente o início da infeção, pois tem apenas informação pertencente à primeira onda. Apesar de esta curva ter menos pontos deve, mesmo assim, representar melhor as condições iniciais da infeção. Vamos assim olhar apenas para os envelopes da Curva Epidêmica 1, presentes nas projeções das Fig 6-9. Na tabela 3 estão apresentados os intervalos de dados pelos envelopes de cada cenário da Curva Epidêmica 1. Podemos estimar então, com alguma probabilidade, que o início da infeção se deu data entre o início de novembro de 2019 e o final de janeiro de 2020.

### 7.3 Limitações do modelo

Neste trabalho não foi tomado em conta a possível redução de contágios imposta pelo confinamento a partir do dia 18 de março. Esta medida poderá ter tido algum efeito de “achatamento” dos dados ao longo do mês de abril.

Há com certeza um atraso na notificação dos casos que não é constante. Entre a data de início de sintomas e o teste positivo, existe sempre um intervalo temporal de alguns dias que inicialmente teve uma grande variação, pela dificuldade de os serviços de saúde acompanharem a situação. Desta forma as notificações dos casos positivos, em que se baseia este trabalho, não traduzem corretamente a evolução da doença.

Por outro lado, para sermos mais precisos, deveríamos olhar do ponto de vista biológico da infeção, ou seja, olhar para a evolução das datas de início dos sintomas que representam com maior precisão a evolução da doença. Estes dados numéricos não são públicos e também não são fáceis de obter junto da população.

Ainda que cientes das limitações apresentadas, os autores acreditam que as projeções do modelo logístico aqui



apresentado trazem informações relevantes sobre o comportamento da epidemia em Portugal, sobretudo ao apontar um possível intervalo de tempo para o início real da epidemia. Embora o intervalo estimado seja amplo e pouco informativo sobre o limite inferior, ele foi obtido com base numa inferência orientada pelos dados disponíveis, sendo portanto uma estimativa racional, que aponta o dia 27 de janeiro de 2020 como limite superior para o início da pandemia em Portugal.

**Tabela 3. Intervalo de datas estimadas para o início da infecção de cada cenário**

Cenário	Curva Epidêmica 1
1 - (29 março)	[Final de novembro, 8 fevereiro]
2 - (24 março)	[Início de novembro, 28 janeiro]
3 - (19 março)	[Meados de outubro, 19 janeiro]
Curva média	[Início de novembro, 27 janeiro]

## 6. CONCLUSÕES

O presente trabalho apresenta um modelo estatístico que ajusta bem os dados da epidemia de COVID-19 em Portugal. Este modelo não contempla efeitos de atraso e subnotificação, o que o torna uma aproximação para a curva de casos reportados simplesmente e não para a curva epidêmica verdadeira. No entanto, ele consegue reproduzir todas as fases típicas de uma epidemia, o que mostra o seu valor como uma ferramenta de análise quantitativa. Com o auxílio dos estimadores apresentados, este modelo permite estimar quantitativamente, com algum nível de incerteza, um intervalo temporal para o início real da epidemia em Portugal como sendo antes do dia 27 de janeiro de 2020, em consonância com algumas evidências encontradas na literatura.

## REFERÊNCIAS

- Cereda, D, and Other (2020). *The early phase of the COVID-19 outbreak in Lombardy, Italy*, arXiv: Populations and Evolution
- Cotta, RM, Naveira-Cotta, CP, Magal, P (2020) *Mathematical parameters of the COVID-19 epidemic in Brazil and evaluation of the impact of different public health measures*. *Biology* 9 (8), 220
- Delatorre E, Mir D, Gräf T, Bello G. (2020) *Tracking the onset date of the community spread of SARSCoV-2 in Western Countries*. Mem Inst Oswaldo Cruz.
- Deslandes A., Berti V., Tandjaoui-Lambotte Y., Alloui C., Carbonnelle E., Zahar J.R. (2020) *SARS-CoV-2 was already spreading in France in late December 2019*. Int J Antimicrob Agents.
- Diário de Notícias (2020). *Santa Maria descobre coronavírus em dois doentes já internados*. Disponível em: < <https://www.dn.pt/sociedade/santa-maria-descobre-coronavirus-em-dois-doentes-ja-internados-11914404.html>> acesso em: 24/08/2020
- DGS (2020a). *Pico da incidência terá sido entre 23 e 25 de março*. Disponível em: < <https://covid19.min-saude.pt/pico-da-incidencia-tera-sido-entre-23-e-25-de-marco/>> acesso em: 24/08/2020
- DGS (2020b). *Evolução do número de casos de COVID-19 em Portugal*. Relatório 21/08/2020. Direção Geral de Saúde de Portugal
- Kuhl, E. (2020), *Data-driven modeling of COVID-19 - Lessons learned*, Extreme Mechanics Letters, Vol 40,
- Hsieh Y.H, Lee J.Y, Chang H.L (2004). *SARS epidemiology modeling*. *Emerging infectious diseases*. 10(6):1165.
- Hsieh Y.H. (2010) *Pandemic influenza A (H1N1) during winter influenza season in the southern hemisphere*. *Influenza and Other Respiratory Viruses*. 4(4):187–197.
- Pais RJ and Taveira N. (2020) *Predicting the evolution and control of the COVID-19 pandemic in Portugal [version 1; peer review: 2 approved with reservations]*. F1000Research 2020, 9:283
- Presidência do Conselho de Ministros (2020). *Resolução do Conselho de Ministros n.º 51-A/2020*. Presidência do Conselho de Ministros. Disponível em: < <https://dre.pt/web/guest/pesquisa/-/search/136788888/details/>>
- Presidência da República (2020a). *Declara o estado de emergência*. Diário da República. Disponível em:<<https://dre.pt/pesquisa/-/search/130399862/details/>> acesso em: 24/08/2020
- Presidência da República (2020a). *Renova a declaração de estado de emergência* Diário da República. Disponível em:< <https://dre.pt/web/guest/pesquisa/-/search/131068115/details/>> acesso em: 24/08/2020
- Presidência da República (2020c). *Procede à segunda renovação da declaração de estado de emergência*. Diário da República. Disponível em: < <https://dre.pt/web/guest/home/-/dre/131908497/details/>> acesso em: 24/08/2020
- Postnikov EB. (2020) *Estimation of COVID-19 dynamics "on a back-of-envelope": Does the simplest SIR model provide quantitative parameters and predictions?* *Chaos, Solitons, and Fractals*. 135:109841.
- PS Meyer. (1994) *Bi-Logistic Growth*, *Technological Forecasting and Social Change*: 89-102
- Tsallis C and Tirnakli U (2020) *Predicting COVID-19 Peaks Around the World*. *Front. Phys*. 8:217.
- Teles, P. (2020). *Predicting the evolution of COVID-19 in Portugal using an adapted SIR model previously used in South Korea for the MERS outbreak*. arXiv.
- Richards F. A (1959). Flexible growth function for empirical use. *Journal of experimental Botany*. 10 (2):290–301.
- Wasserman, L. (2004). *All of Statistics*, Springer New York.
- WHO (2020). *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. [internet]. Disponível em: < <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>> acesso em: 24/08/2020
- Wu, K., Darcet, D., Wang, Q. et al. (2020). *Generalized logistic growth modeling of the COVID-19 outbreak: comparing the dynamics in the 29 provinces in China and in the rest of the world*. *Nonlinear Dyn*