Predicting the course of Covid-19

Ana Cascon and William F. Shadwick

Omega Analysis and IMPA

25 January 2022

In the paper linked to the announcement for this talk, we presented what should be seen as a method for navigating waves of Covid-19.

Our focus is on making accurate (errors less than 10%) short to medium term (one week to several weeks) forecasts of Covid-19 hospitalisations.

The reason is simple:

This is the fundamental requirement for dealing with an epidemic.

Once we have accurate forecasts of demand, reliable contingency planning for healthcare is possible.

Here is an indication of how little we know about the transmission of the most familiar illness caused by a corona virus—the common cold.

An outbreak of common colds at an Antarctic base after seventeen weeks of complete isolation

BY T. R. ALLEN

Medical Officer, British Antarctic Survey

AND A. F. BRADBURNE, E. J. STOTT, C. S. GOODWIN
AND D. A. J. TYRRELL

Clinical Research Centre, Harrow, England

SUMMARY

Six of 12 men wintering at an isolated Antarctic base sequentially developed symptoms and signs of a common cold after 17 weeks of complete isolation. Examination of specimens taken from the men in relation to the outbreak has not revealed a causative agent.

The approach we have taken does not require models of infections, or of transmission, or cases or the proportion of cases that will require hospitalisation.

Modelling those things correctly would, of course, give accurate predictions of hospitalisations (and a great deal more besides).

But that is *very difficult*, as many of you know.

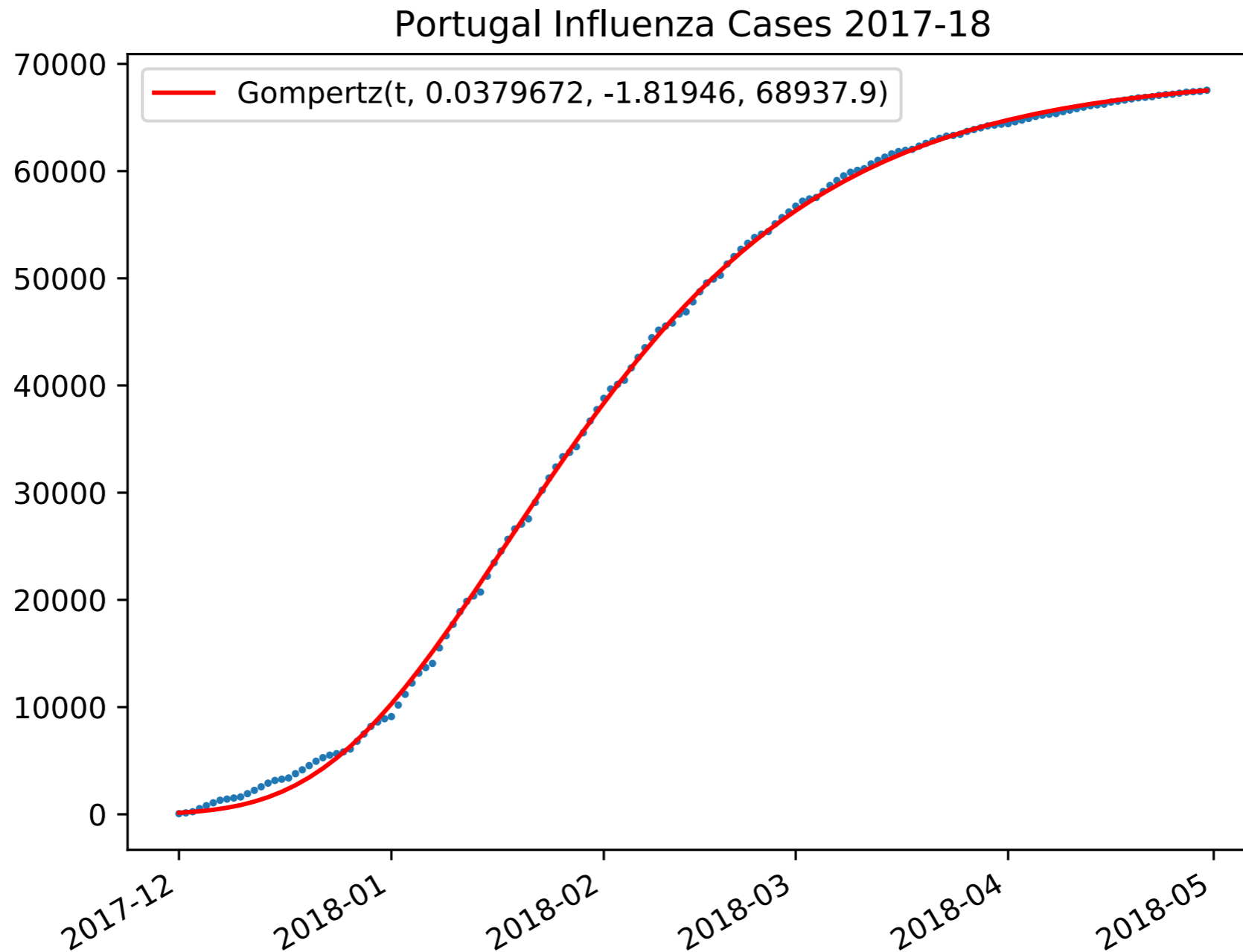Our much simpler goal can be met by a much easier approach.

It is based simply on *counting* cumulative hospital admissions, and finding a simple curve that approximates those running totals very well.

Here are 3 examples. The data are the blue dots and the curves are in red.
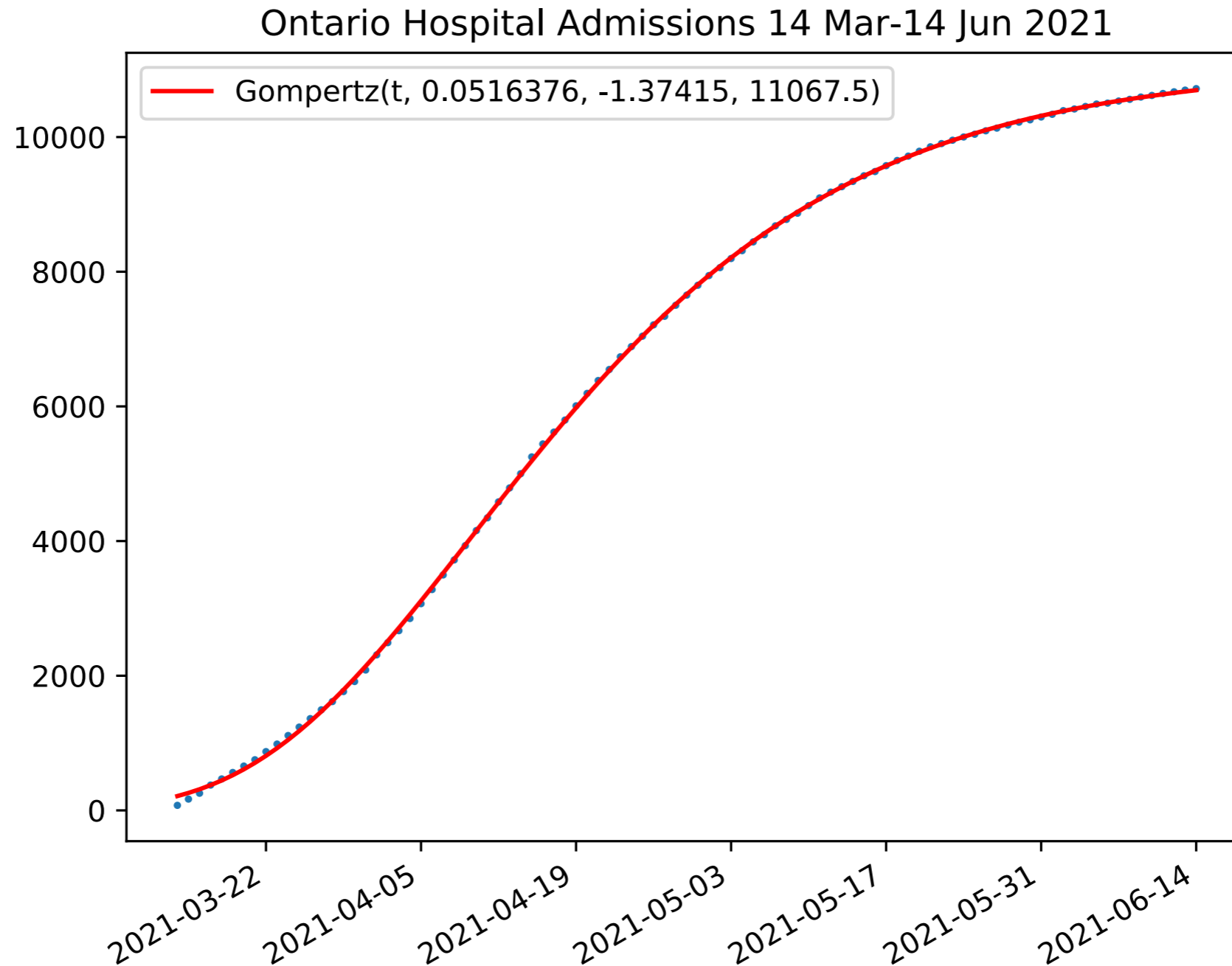
Cases in the 1853 Cholera epidemic in Aalborg, Denmark.

Danish Cholera Cases

Gompertz(t, 0.145297, -2.10774, 760.516)
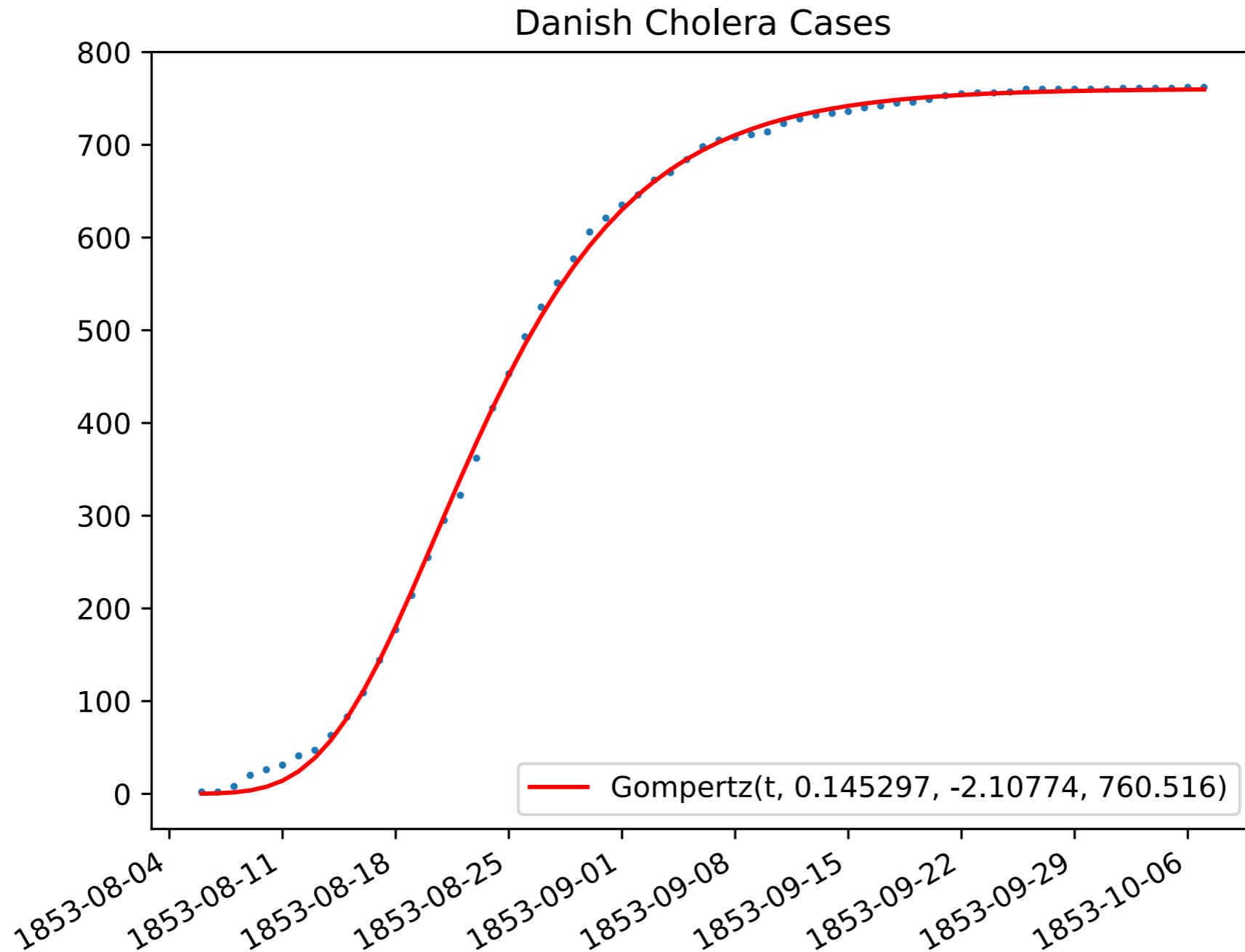
Influenza Cases recorded in the Portuguese National Health Service 2017-2018.

Portugal Influenza Cases 2017-18

Gompertz(t, 0.0379672, -1.81946, 68937.9)

Covid-19 Hospitalisations in Ontario, 14 March-14 June 2021

Ontario Hospital Admissions 14 Mar-14 Jun 2021

Gompertz(t, 0.0516376, -1.37415, 11067.5)

In each of these examples the red curve is the best fit to the cumulative data by a *Gompertz Function:*

$$X(t) = Ne^{-e^{-(at+b)}}$$

Time *t* is measured in days, starting at *t=0*. N is the (unknown) final number of events (deaths, cases, hospital admissions), a controls the rate at which the events grow and b determines the fraction of events at time *t=0*.

The three parameters can be estimated directly from event data, by non-linear regression (which simply means finding the values of the parameters that minimise the average squared error between the points on the curve and the data).

As the epidemic unfolds the parameters become progressively closer to their final values.

Decades before Newton showed *why* it was so, Galileo showed that projectiles followed parabolic trajectories, to a very good approximation.

This is an example of what is now called a *phenomenological model* — one that provides a mathematical formulation that can be useful for prediction, even though it doesn't explain the mechanism being modelled.

The Gompertz Function is a simple phenomenological model for epidemics. It doesn't explain why the growth follows these curves, but the observation that they do allows us to make forecasts.

We find the parameters that make the best Gompertz Function fit today and we use that function's values at future times as our forecast.

The accuracy of those forecasts is a matter for empirical investigation.
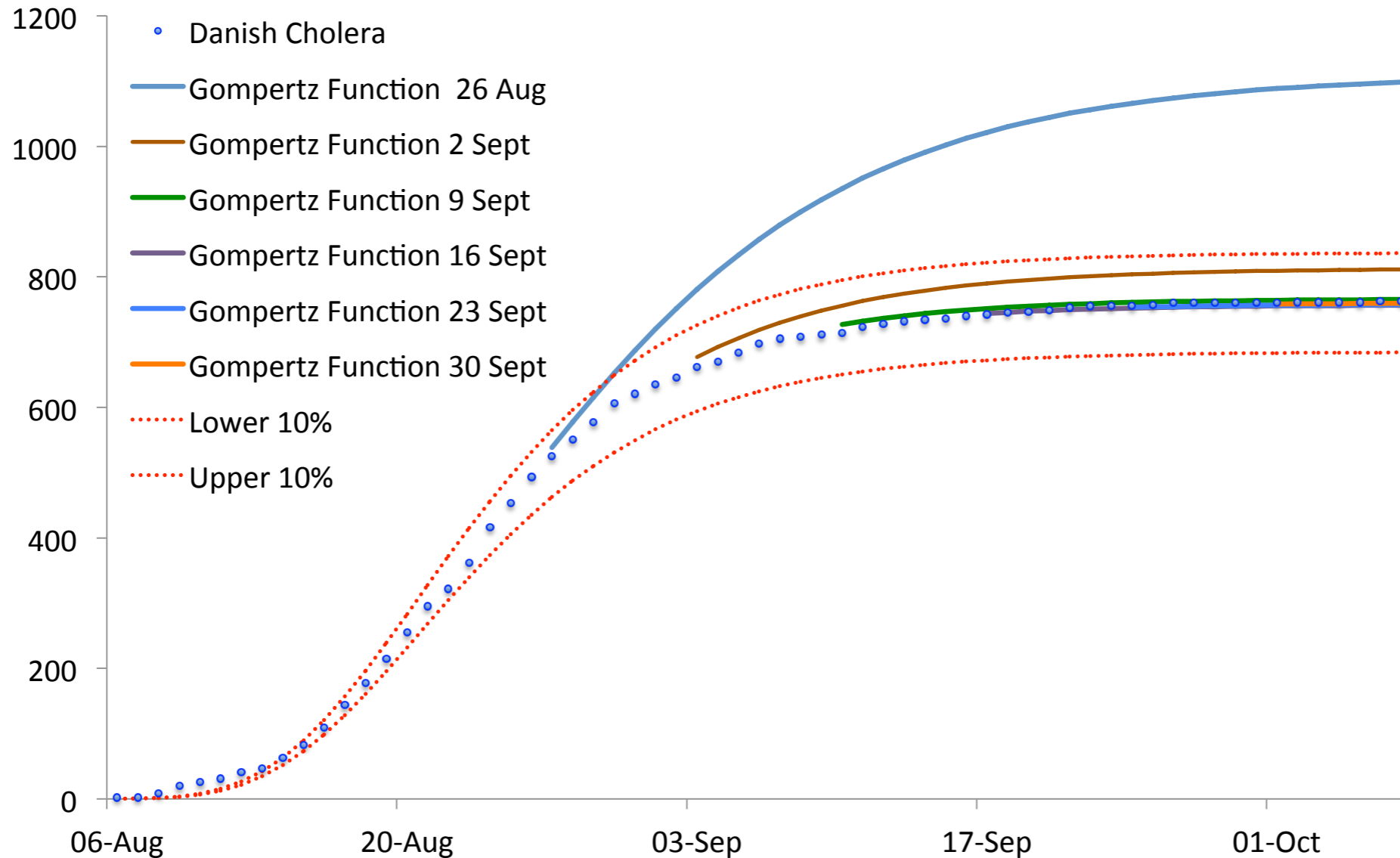
An error band of +/-10% shows how well successive weekly Gompertz Function fits predicted the future evolution of Danish Cholera cases.

Gompertz Function Forecasts of Ontario Covid-19 Hospitalisations
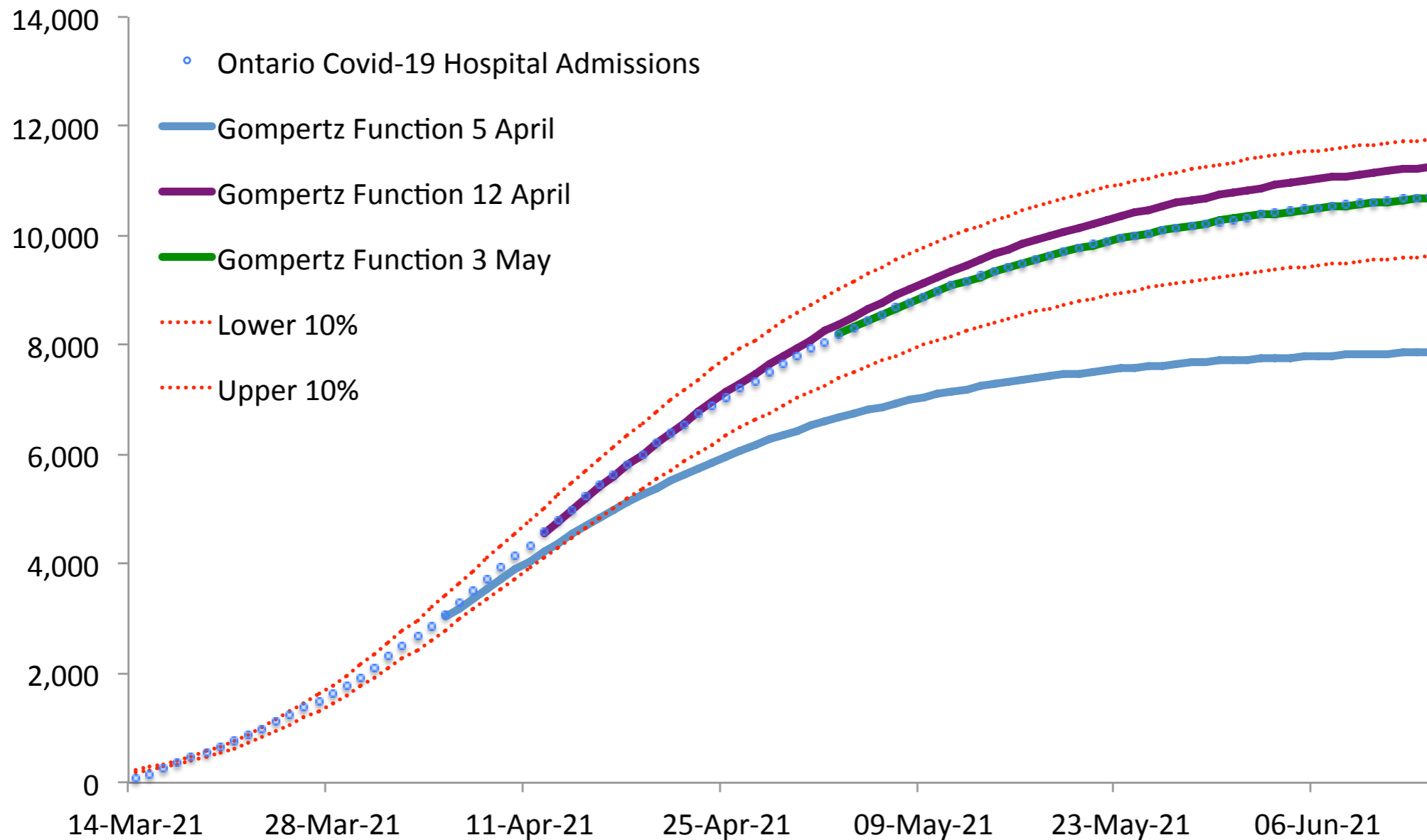
An error band of +/-10% shows how well successive weekly Gompertz Function fits predicted the future events.

Errors in Ontario Covid-19 Hospitalisation Forecasts
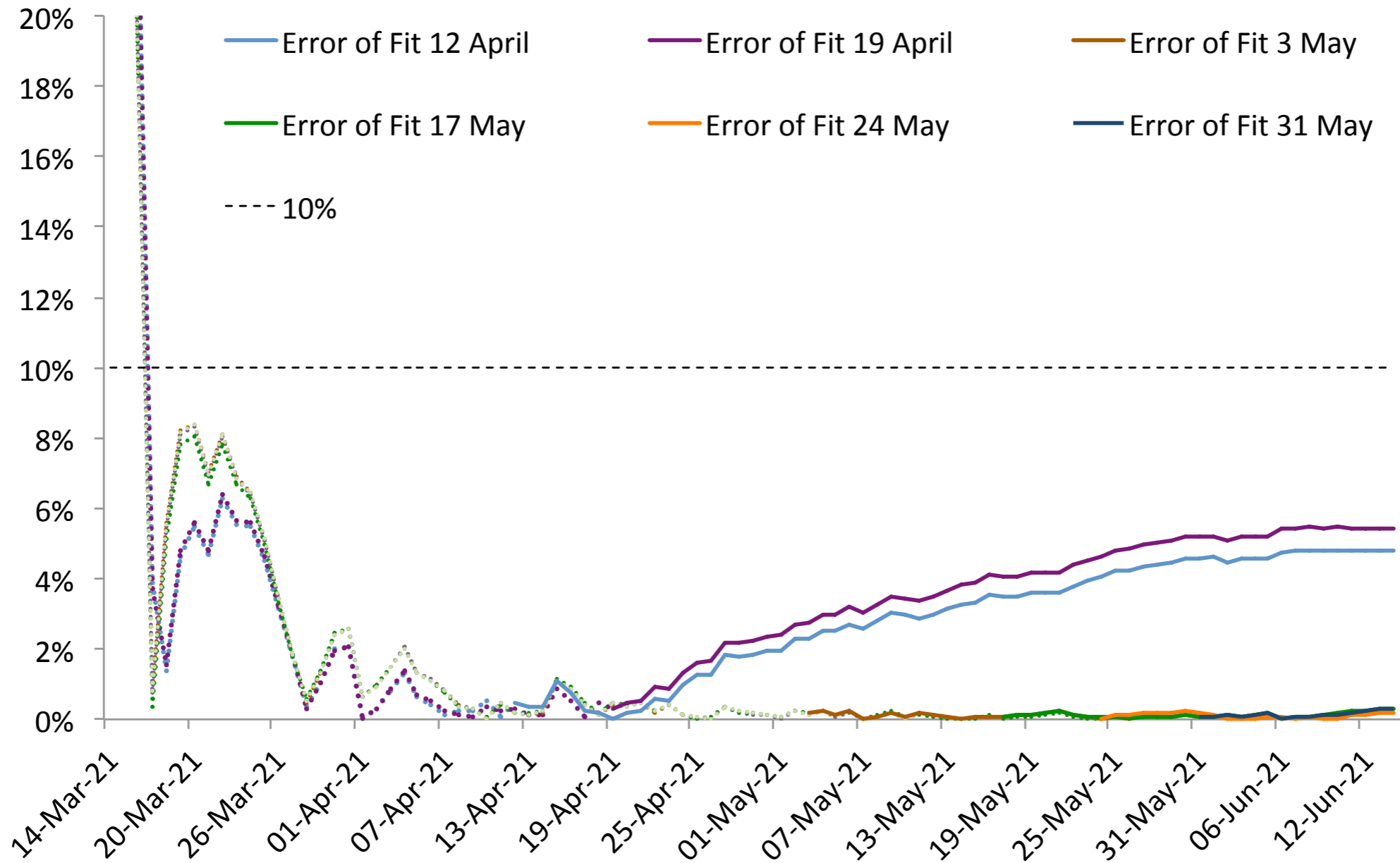
Absolute value of the difference between the observed data and the prediction as a percentage of the prediction (solid lines). The dotted lines show the fit errors in the sample data.

Galileo's model also allowed him to make inferences about the motion of a projectile. (Which you may remember from a high school physics class.)

The good fit between the cumulative event data for Covid-19 and the Gompertz function allows us to do the same thing.

The fact that they follow a Gompertz Function as the epidemic progresses tells us that:

The Gompertz Function dynamics are *features* of the time evolution of the epidemic*.*

$$X(t) = N e^{-e^{-(at+b)}}$$

N —which is the final total events—governs the scale of the epidemic but its dynamics are driven by the Gumbel distribution.
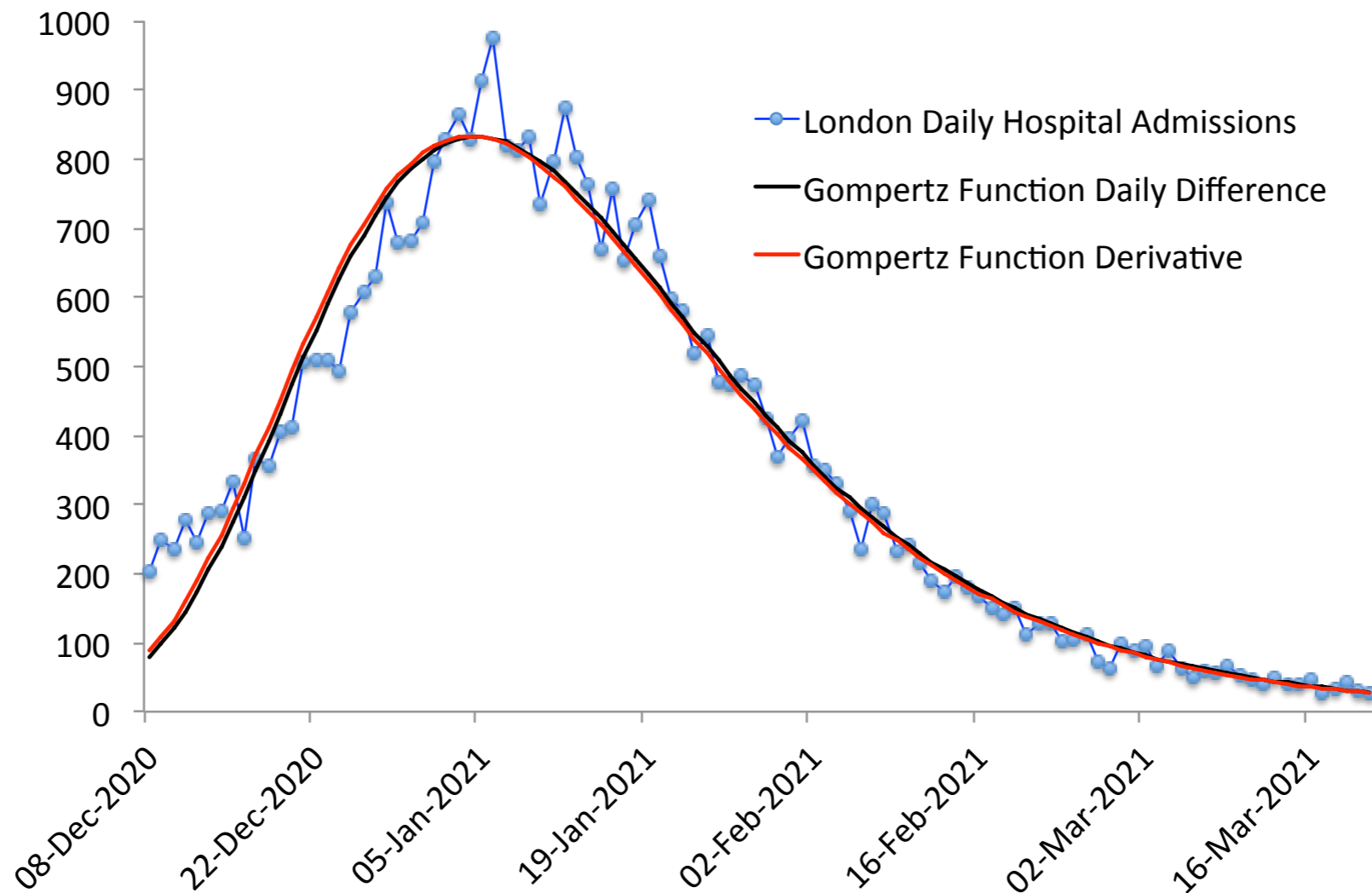
$$G(t, a, b) = e^{-e^{-(at+b)}}$$

The Gompertz Function increases rapidly but its acceleration peaks at time

$$T_1(a, b) = \frac{log(\frac{2}{3+\sqrt{5}}) - b}{a}$$

At this time only about 7% of the final events have occurred.

$$X(T_1) = e^{-(\frac{3+\sqrt{5}}{2})} N \approx 0.0729 N$$

The velocity is very asymmetric, with a rapid rise and much slower decline.

London Daily Hospital Admissions

Gompertz Function Daily Difference

Gompertz Function Derivative

It reaches its maximum at

$$T_2(a, b) = \frac{-b}{a}$$

At peak velocity, approximately 37% of N total events have occurred.

$$X(T_2) = \frac{N}{e} \approx 0.368N$$

This has immediate consequences for the Herd Immunity Threshold.

The Gompertz Function's first derivative is a good approximation to daily increments.

$$\dot{X}(t+1) \approx X(t+1) - X(t)$$

Once we reach peak velocity the derivative starts decreasing—so the number of new events, which is just the daily increment, is smaller than it was the day before.

This means that the reproduction number must be less than 1.

We have illustrated the Gompertz Function fits to hospitalisations but they are equally good for Cases and Deaths.

If cases are a good proxy for infections, then once the velocity peaks,the reproduction number R_t must be less than 1.

In any epidemic that follows a Gompertz Function, the Herd Immunity Threshold is reached with about 37% of the susceptible population infected.

This is what we see in the Danish Cholera example. It continues to follow the Gompertz Function as cases (and R_t) decline rapidly to zero.

But that's not what we saw in the initial Covid-19 outbreak.

All of the outbreaks passed the peak velocity, and then…

We saw a transition from Gompertz Function growth to *linear growth* in cumulative events. Here's the transition for hospitalisations in London.
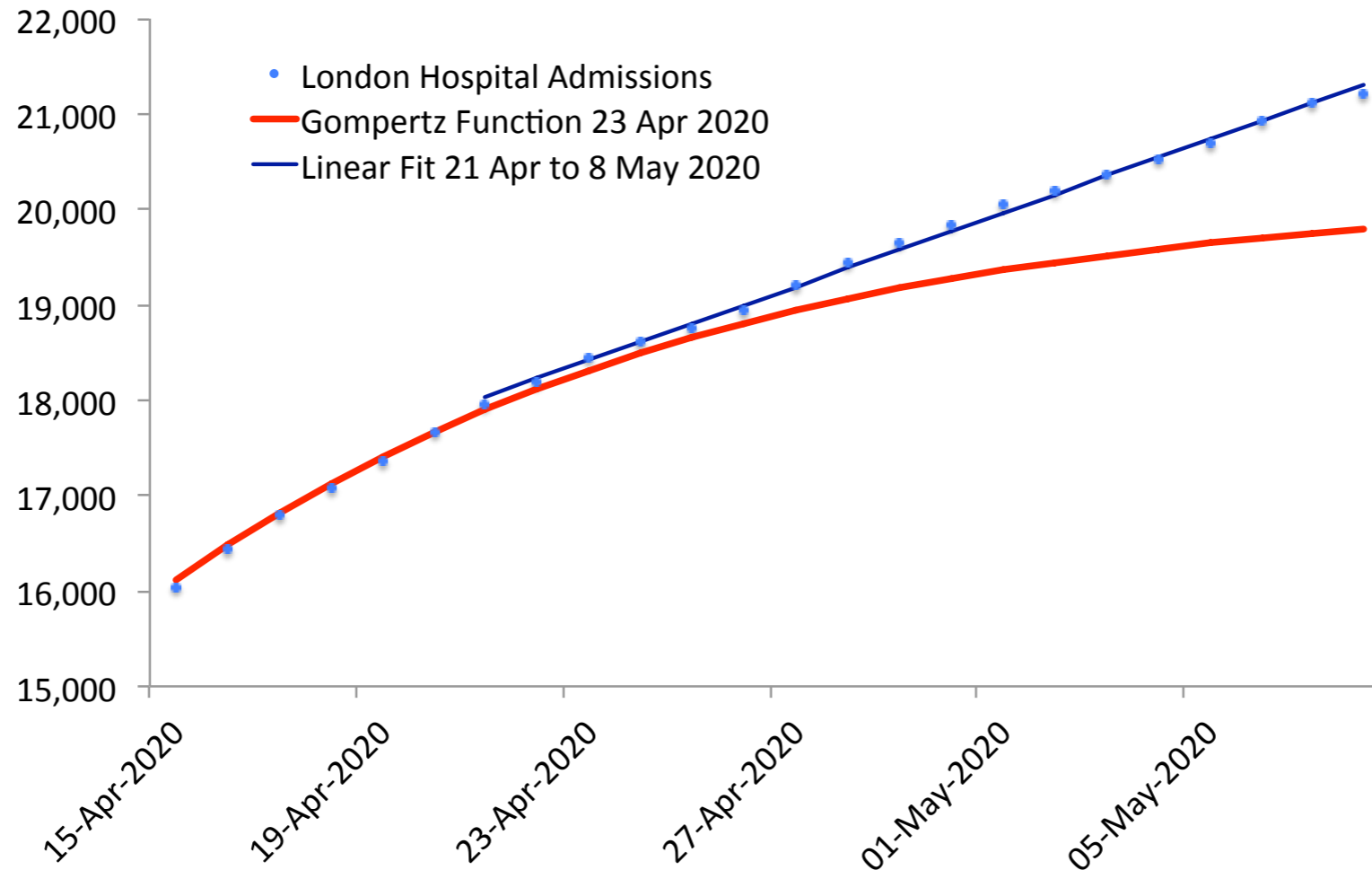
The linear approximation is *extremely* good with r^2 >0.99 and very small errors between the data and the linear regression line.

When the cumulative is growing (approximately) linearly, we must have the same number of new events added in each time period.

If cases are a good proxy for infections: this means that $R_t = 1$. This is the condition for endemicity.

We saw the transition to linearity repeated again and again in Covid-19 data.

Then we discovered that there is an example of a yearly cycle of transitions between Gompertz Function growth and linear growth:
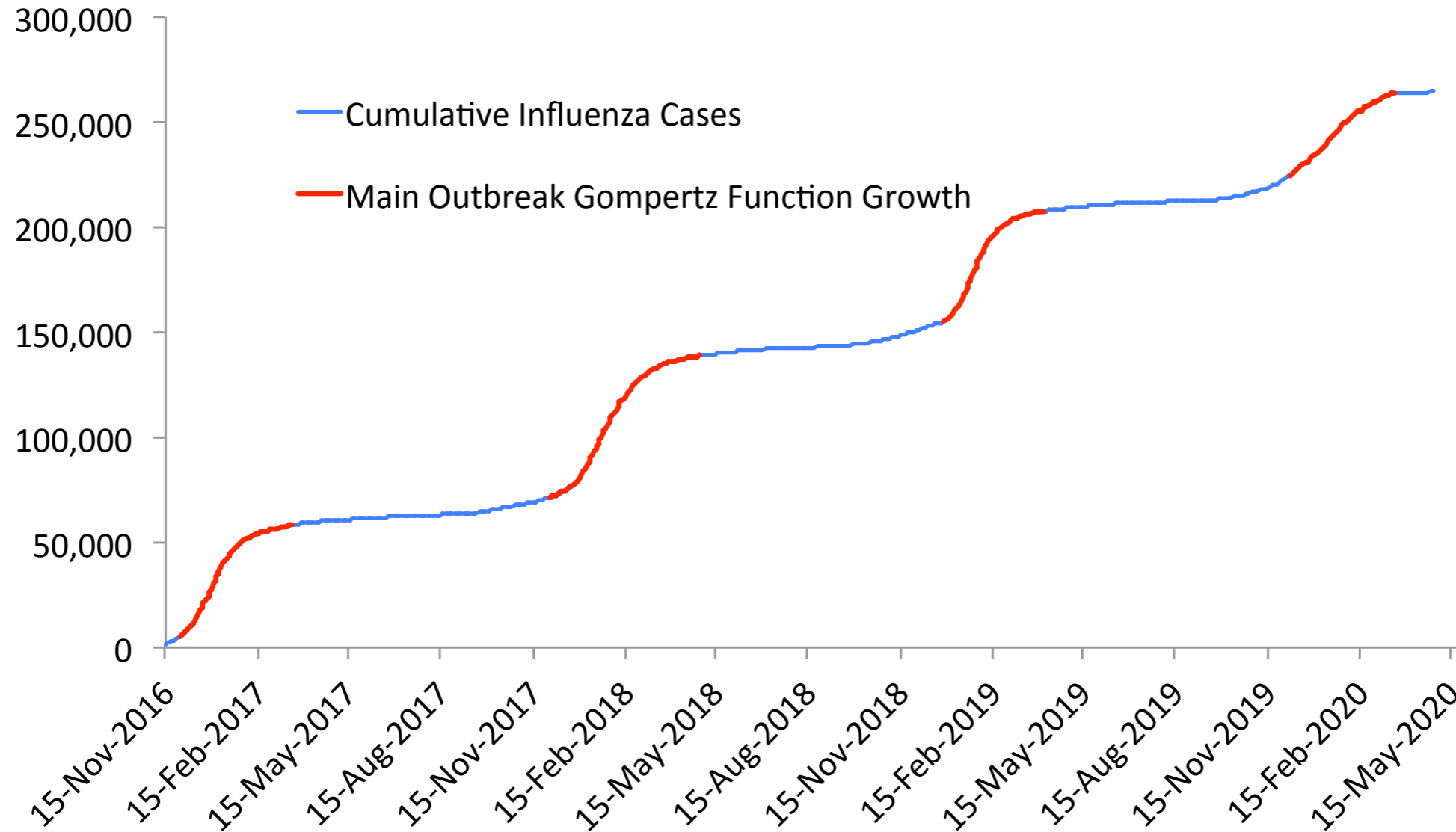
Endemic Influenza.

The 'Influenza Observatory'

The Portuguese National Health Service reports daily influenza case data starting in November 2016.

This gave us the opportunity to observe the annual Influenza cycle over several years in detail.
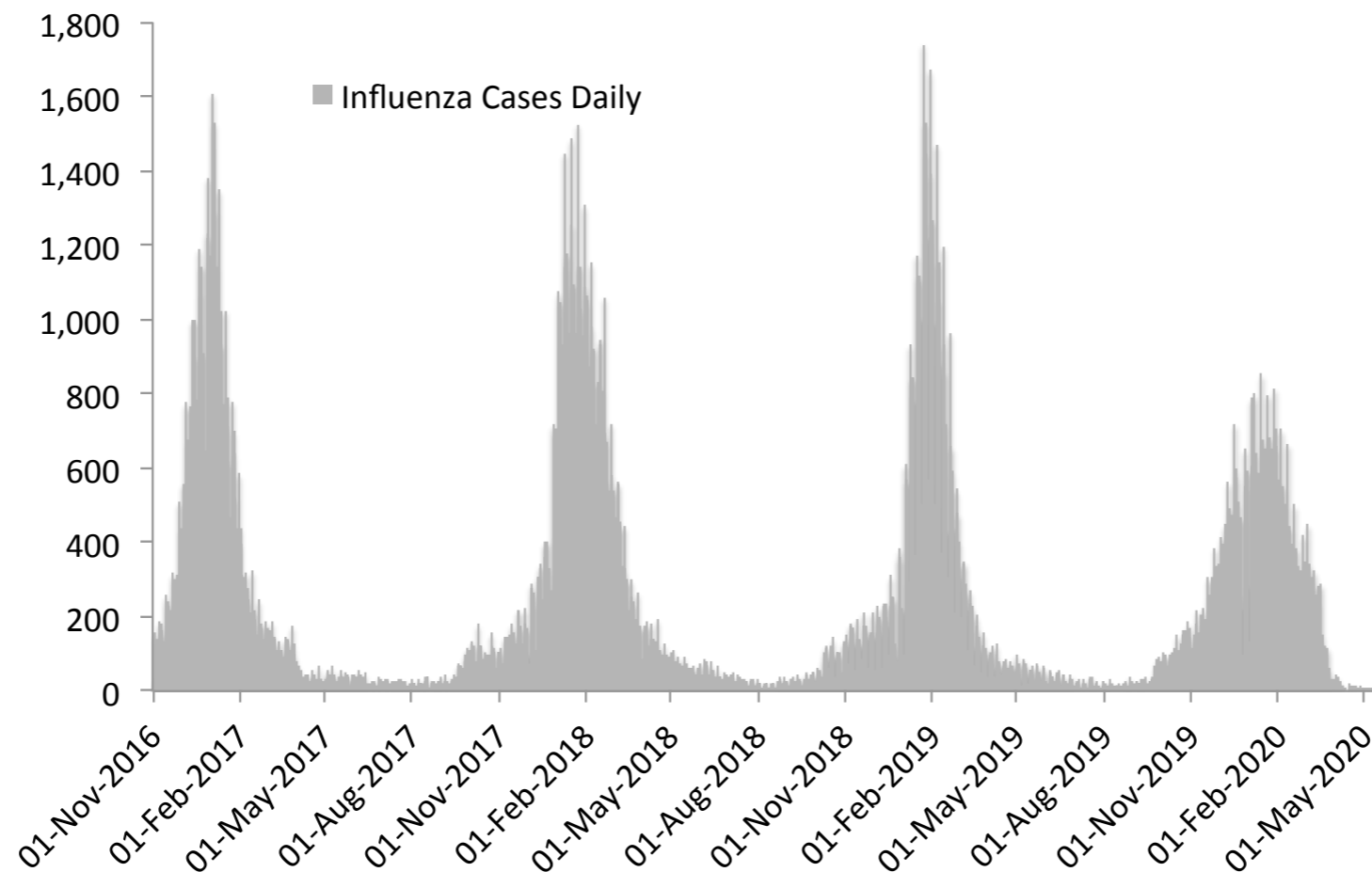
Influenza cases in Portugal Nov 2016 to May 2020.

The annual peak outbreaks all follow Gompertz Function growth.

They happen every year in the influenza cycle for the North Temperate Zone described by Edgar Hope-Simpson.

Each is preceded and followed by linear growth.

There's also some 'fine structure' which again follows Hope-Simpson
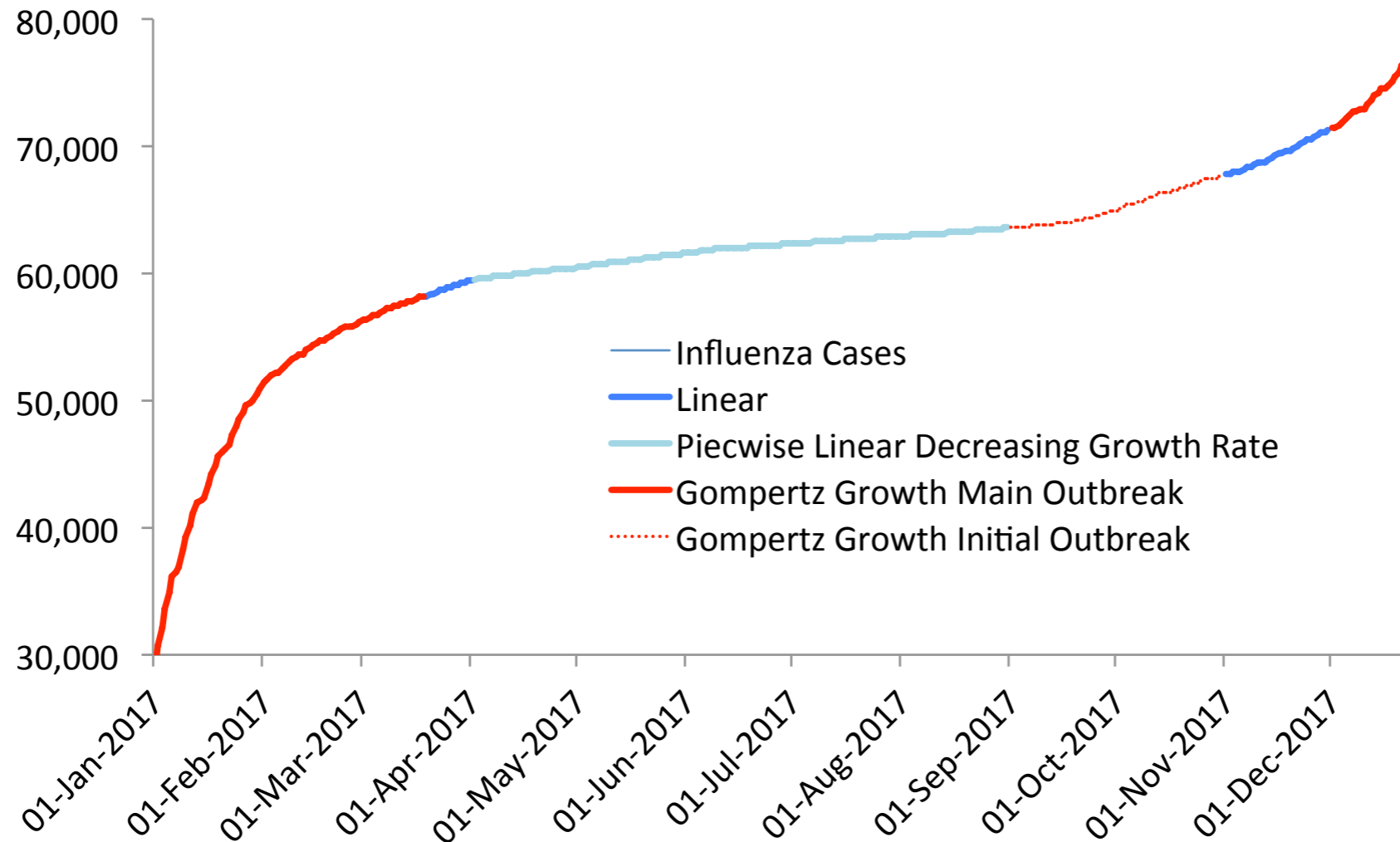
The November entry 'line' has a higher slope than the April exit 'line'. Following the summer off-season where the growth rate decreases, there's a period of Gompertz Function growth in early autumn that accounts for this.

If we zoom in, the autumn Gompertz Function growth is apparent.
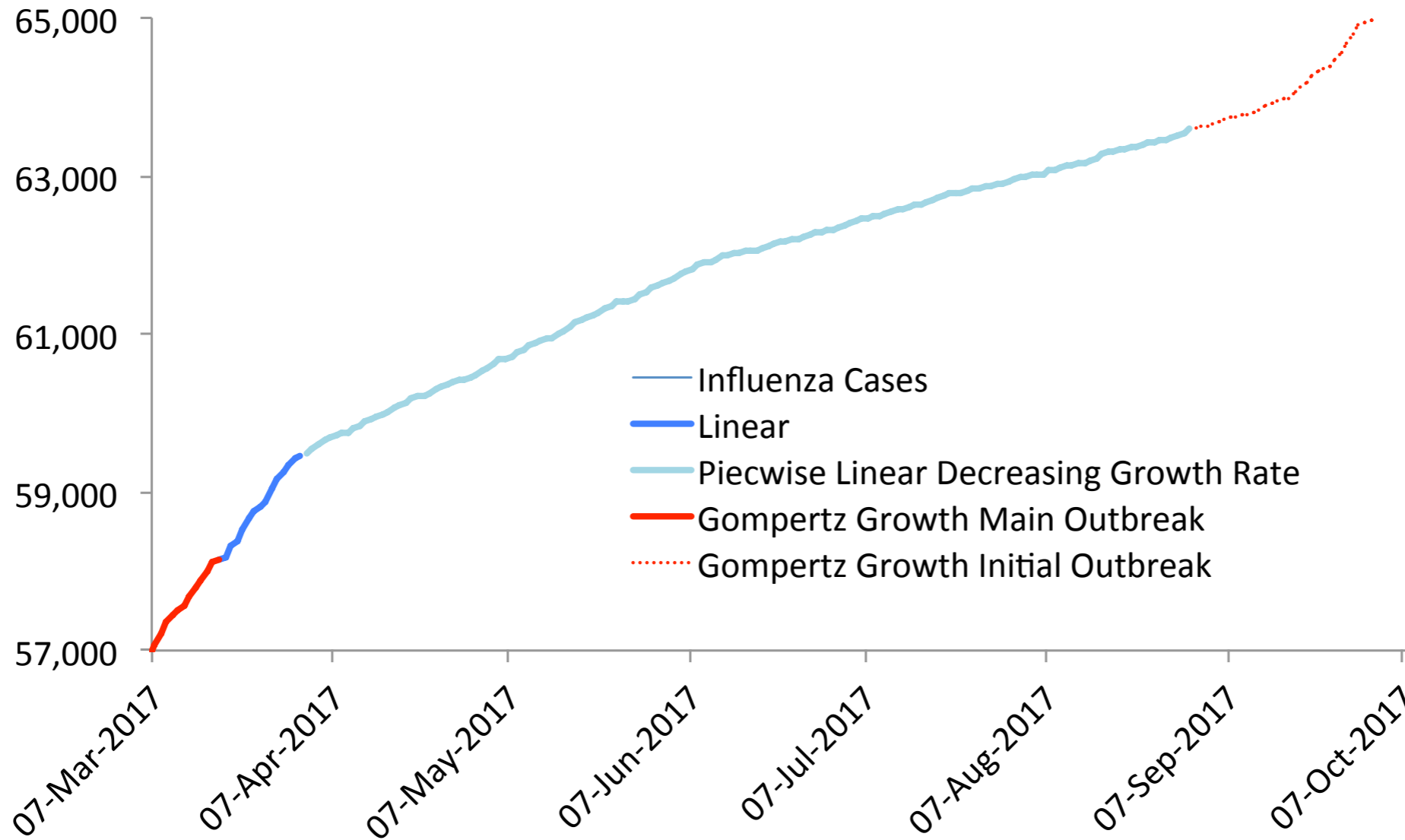
This is repeated every year in the Portuguese data set.

The annual Influenza Cycle in the Portugal

Portuguese Influenza Growth Cycle		
Year	Gompertz Function Growth	Linear Growth
2016-2017	28 Nov 16-7 Mar 17	8 Mar 17-31 Aug 17
2017-2018	1 Sep 17-31 Oct 17 1 Dec-17-30 Apr 18	1 Nov 17-30 Nov 17 1 May 18-14 Sep 18
2018-2019	15 Sep 18-14 Nov 18 28 Dec 18-21 Mar 19	15 Nov 18-27 Dec 18 22 Mar 19-31 Aug 19
2019-2020	1 Sep 19-16 Nov 19 8 Dec 19-21 Mar 20	17 Nov 18-7 Dec 19 22 Mar 20-31 Aug 20
2020-2021	1 Sep 20-31 Oct 20 1 Jan 21-21 Mar 21	1 Nov 20-31 Dec 20

We expect to see the same cycle in the Northern Temperate Zone.
We'll show that Covid-19 provides some indirect evidence for this.

The Extended Gompertz Function Model

In our paper we have illustrated this alternating cycle for Sweden, London, Isle-de-France, Ontario and Portugal in the Northern Temperate zone, for the Brazilian State of Rio de Janeiro in the Southern Tropical zone and the State of Amazonas for an equatorial example.

It occurs equally in cases, hospitalisations, ICU admissions and deaths.

Here are some samples of the alternating phases.

The Extended Gompertz Function Model

Alternating Gompertz Function and Linear Growth for Covid-19		
Area and Data type	Gompertz Growth	Linear Growth
Sweden Hosp	3 Mar 20-17 May 20 1 Oct 20-5 Nov 20 1 Dec 20-7 Feb 21 19 Mar 21-24 May 21	18 May 20-30 Sep 20 6 Nov 20-30 Nov 20 8 Feb 21-18 Mar 21
London Hosp	19 Mar 20-23 Apr 20 1 Sep 20-31 Oct 20 8 Dec 20-21 Mar 21	24 Apr 20-31 Aug 20 1 Nov 20-7 Dec 20
Isle de France Hosp	18 Mar 20-11 May 20 9 Aug 20-16 Oct 20 11 Nov 20-4 Jan 21 8 Mar 21-6 Jun 21	12 May 20-8 Aug 20 17 Oct 20-10 Nov 20 5 Jan 21-7 Mar 21
Ontario Hosp	1 Mar 20-14 Jun 20 2 Sep 20-31 Oct 20 28 Nov 20-14 Feb 21 14 Mar 21-14 Jun 21	15 Jun 20-1 Sep 20 1 Nov 20-27 Nov 20 15 Feb 21-13 Mar 21
Portugal ICU	14 Mar 20-15 May 20 8 Sep 20-11 Nov 20 8 Jan 21-31 Mar 21	16 May 20-7 Sep 20 12 Nov 20-7 Jan 21
RJ Hosp	15 Mar 20-30 Jun 20 26 Dec 20-29 Jan 21 18 Feb 21-21 Apr 21 5 May 21-14 Jun 21	1 Jul 20-25 Dec 20 30 Jan 21-17 Feb 21 22 Apr 21-4 May 21
Amazonas Deaths	1 Apr 20 -14 Jun 20 23 Dec 20-10 Mar 21	15 Jun 20-22 Dec 20

End of Table

The Extended Gompertz Function Model

In every case we saw Gompertz Function growth resume in the normal Hope-Simpson seasonal period (including the autumn one we observed in the Portuguese Influenza cycle) as the second round of Covid-19 outbreaks took place.

In addition, both Isle-de-France and Ontario had large 'out of season' outbreaks in the spring of 2021.

In England, there was a small 'out of season' outbreak from 29 May to 15 July 2021.

The Extended Gompertz Function Model

Covid-19 outbreaks follow a cycle of alternating Gompertz Function Growth and linear growth in cumulative events.

Each of the Gompertz Function growth phases is predictable as illustrated above.

Linear extrapolation gives accurate predictions in the linear phases.

This reduces the forecasting problem to:

1) Detecting the transition between phases, most importantly from linear to Gompertz Function growth.

2) Providing realistic bounds on growth in the early Gompertz Function phase or reducing the time required for the Gompertz Function fits to achieve predictive power.

Covid-19 hospitalisations in England grew linearly in November 2021. Using the data to 5 December, we detected a departure from linear growth.The data has a two day lag so we knew this on 7 December.

For the sample of daily admissions from 15 November to 5 December 2021 we used a form of Extreme Value Theory[*] analysis to calculate bounds for growth of hospital admissions in the initial phase of the Omicron outbreak.

Our bounds predicted that total admissions between 6 December and 31 December 2021 would be bounded above by 30,613 admissions. The actual total for that period was 29,288.

*Extreme Value analysis allows us to model the (unobserved) tail of the sample distribution of daily admissions. We use it to predict 3 Levels that depend only on the sample.

Level 1 is the average daily admissions conditional on exceeding the sample maximum.
Level 2 is the average daily admissions conditional on exceeding Level 1.
Level 3 is the average daily admissions conditional on exceeding Level 3.

It is an empirical observation that cumulative admissions are contained between linear growth at Level 1 admissions per day and linear growth at Level 3 admissions per day until the Gompertz Function achieves a high level of predictive power.

The Extended Gompertz Function Model in Real Time
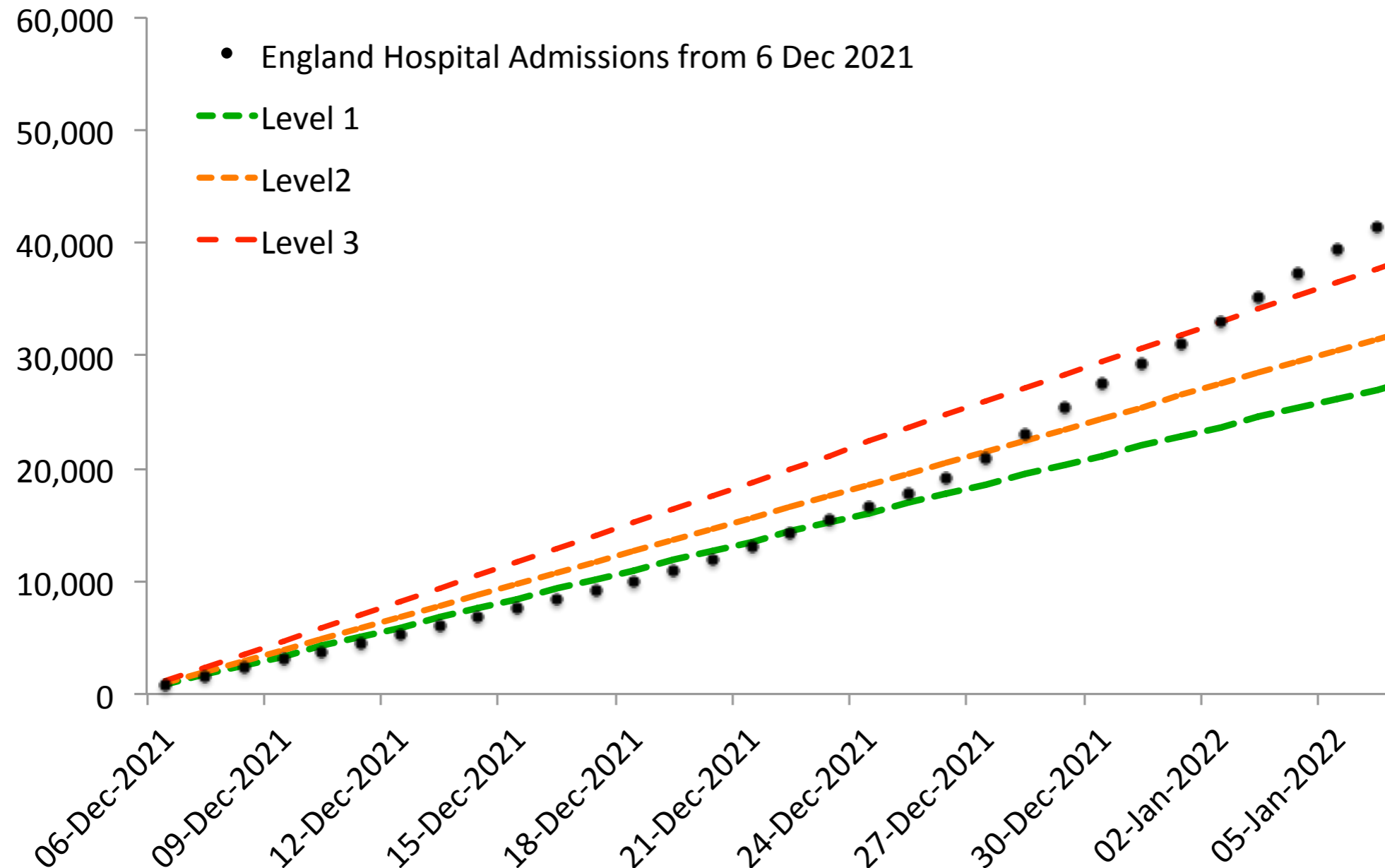
Since 29 December 2021, the prediction errors of the Gompertz Function fits to hospital admissions data have been under 10% for at least two weeks.

cascon@OmegaAnalysis.com shadwick@OmegaAnalysis.com

If the epidemic model for cases has the form

$$X(t) = NG(t)$$

where N is the unknown number of people susceptible to infection, then

$$G(t) = \frac{X}{N}$$

is the fractional number of cases. It's a non-decreasing function with values in [0,1] —therefore it's a probability distribution.

(Or maybe a parametrised family of probability distributions like the generalised Logistic distribution in the Richards Function.)

The 'Inevitability' of the Gompertz Function

Elie Cartan showed that transformation groups generate geometry and geometry generates invariants—like curvature.

The Gumbel distribution is an Extreme Value distribution—one of the exceptional distributions determined by the action of the 'location scale' group on univariate probability distributions. It's the zero curvature case.

The other Extreme Value families correspond to constant negative curvature (Weibull distributions-the short tailed case) and constant positive curvature (Fréchet distributions-the fat tailed case).

These special cases are 'attractors'. The Gumbel distribution is the attractor for 'thin-tailed' distributions, whose domains are unbounded above but which have finite moments of all orders. (e.g. Normal, Laplace, Logistic)

Our version of the EVT Theorem shows that a distribution is in the domain of attraction of an EVT attractor if and only if the limiting value of its curvature is the attractor's constant.

We showed that the value of a distribution's curvature at a given quantile is also an invariant and can be used to measure the rate of convergence of a distribution to its attractor.

The convergence for many thin tailed distributions is fast enough that after the 0.5 percentile the difference between the distribution and a Gompertz Function fit is indistinguishable at the granularity of our fits to epidemic data.

The 'Inevitability' of the Gompertz Function

The observational data tells us that the probability density function for the distribution must be asymmetric with a peak very close to the 0.368 percentile—like the Gumbel distribution.

The Extreme Value Theorem tells us that, asymptotically, every thin-tailed distribution is approximated arbitrarily well by a Gumbel distribution and for practical purposes that begins at about the 0.5 percentile.

This doesn't leave much room for a distribution to deviate from the Gumbel distribution or for the model to deviate from the Gompertz Function.

This is a method for navigating waves of Covid-19.

It makes accurate (errors less than 10%) short to medium term (one week to several weeks) forecasts of Covid-19 hospitalisations.

This is the fundamental requirement for dealing with an epidemic.

Once we have accurate forecasts of demand, reliable contingency planning for healthcare is possible.

Appendix
Ontario Covid-19 Hospitalisations

In November and December 2021, cumulative Covid-19 hospital admissions in Ontario were growing linearly. Simple linear extrapolation gave accurate predictions in this phase.
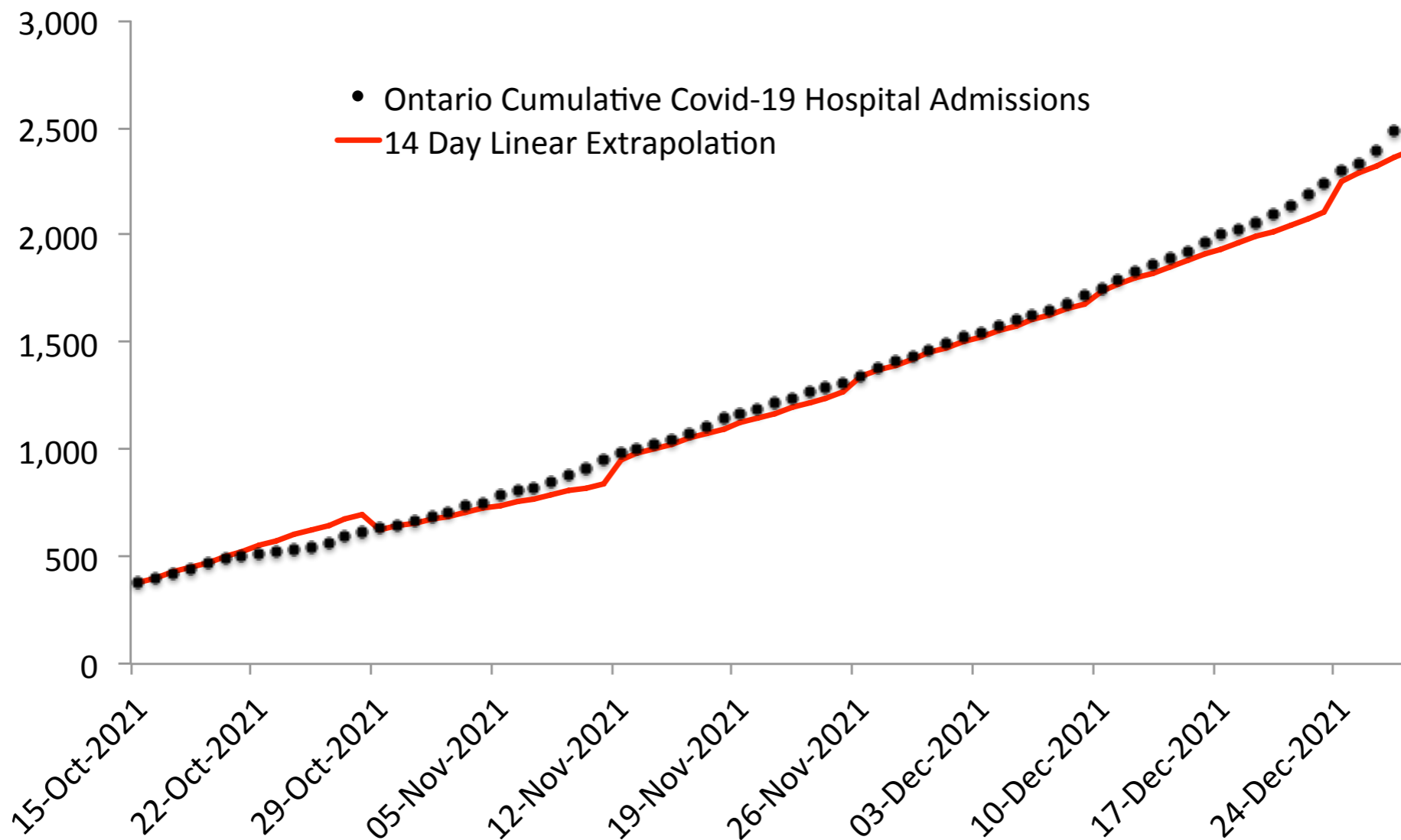
Predictions made by extrapolating the linear fit from the previous 2 weeks forward for 2 weeks had a maximum error of 13% and an average absolute error less than 4% between 15 October and 26 December 2021.

We detected a transition from the linear growth phase to the next Gompertz Function growth phase on 27 December 2021.

From the sample of daily admissions from 7 to 27 December 2021 we used the EVT analysis described above for England to calculate bounds for growth of hospital admissions in the initial phase of the Omicron outbreak.

*Extreme Value analysis allows us to model the (unobserved) tail of the sample distribution of daily admissions. We use it to predict 3 Levels that depend only on the sample.

Level 1 is the average daily admissions conditional on exceeding the sample maximum.
Level 2 is the average daily admissions conditional on exceeding Level 1.
Level 3 is the average daily admissions conditional on exceeding Level 3.

It is an empirical observation that cumulative admissions are contained between linear growth at Level 1 admissions per day and linear growth at Level 3 admissions per day until the Gompertz Function achieves a high level of predictive power.

Appendix
Ontario Covid-19 Hospitalisations

Cumulative hospital admissions from 28 December have remained below the upper bound. By 17 January the errors in the Gompertz Function fit predictions had already fallen below 12% and by 18 January below 6%.
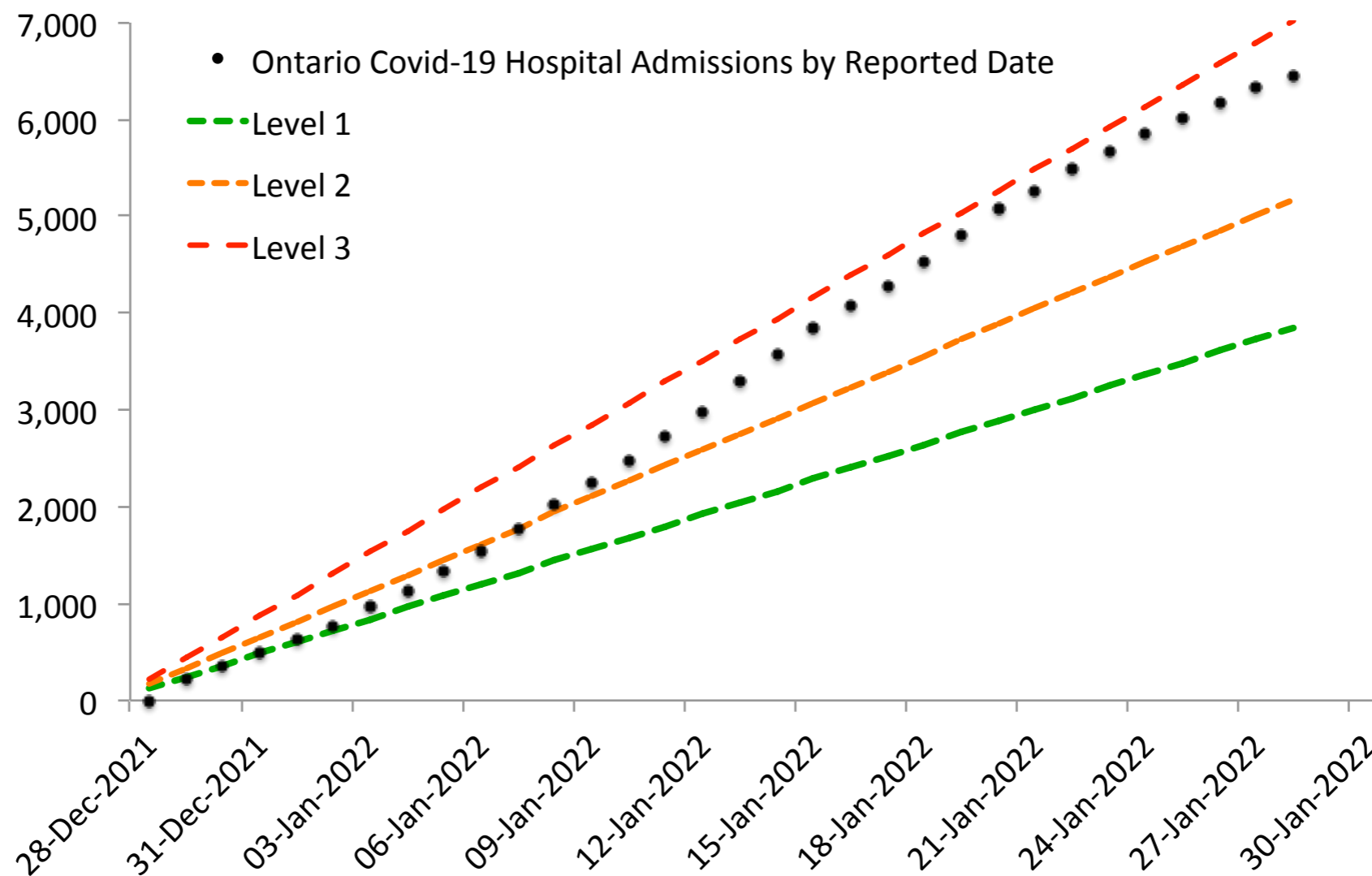
- Ontario Covid-19 Hospital Admissions by Reported Date
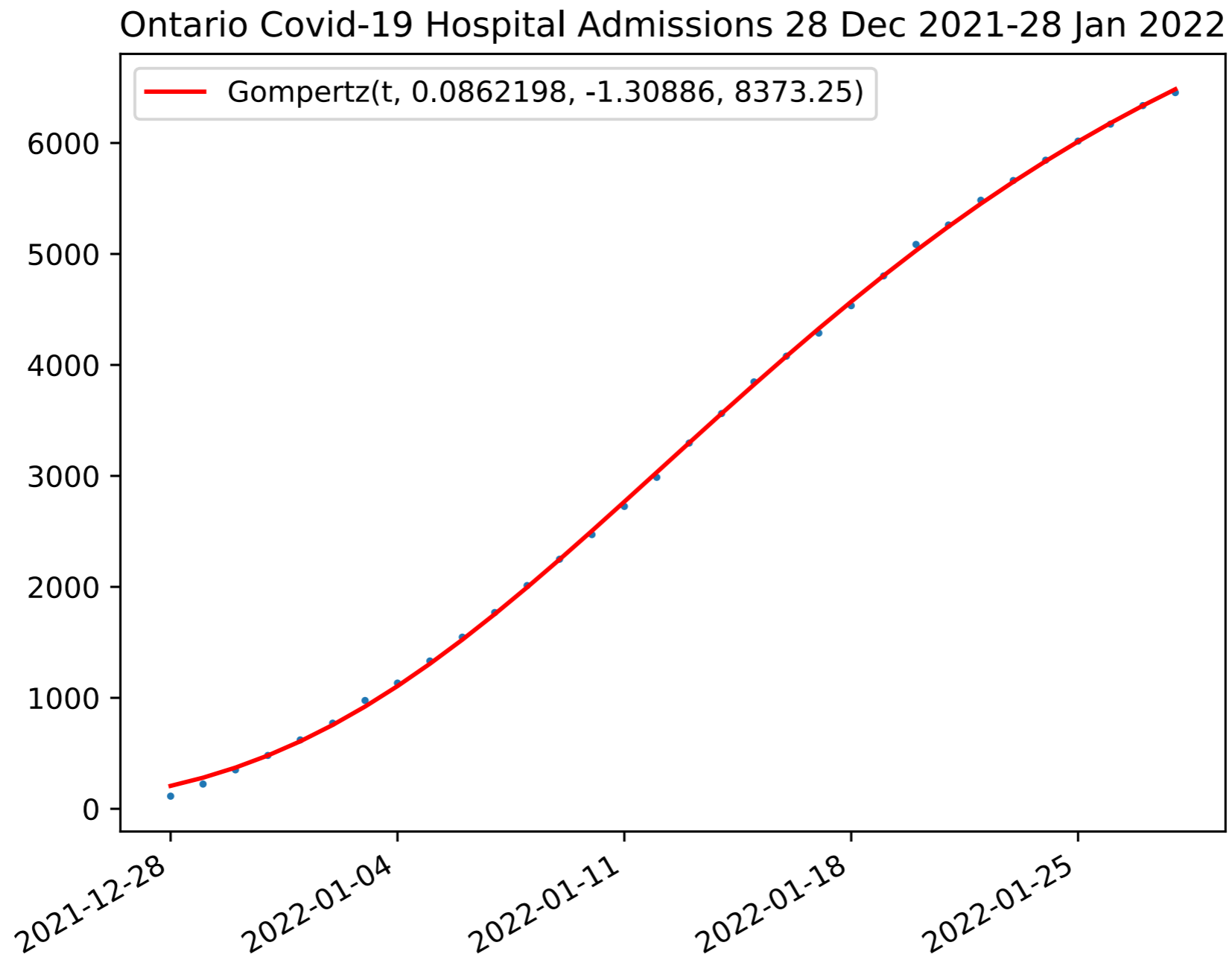- Level 1
- Level 2
- Level 3

As expected, cumulative admissions remained within the linear growth bounds until the Gompertz Function fits achieved a high level of predictive power.

Appendix
Ontario Covid-19 Hospitalisations

The model predicts that cumulative hospitalisations will reach about 8,400 by mid March unless the next linear growth phase begins before that.

Ontario Covid-19 Hospital Admissions 28 Dec 2021-28 Jan 2022

As expected, cumulative admissions remained within the linear growth bounds until the Gompertz Function fits achieved a high level of predictive power.

Acknowledgement and a note on the fits.

The Gompertz function fits reported here have been done using Python's non-linear fit routine in unoptimised code provided to us by Prof. B.A. Shadwick, University of Nebraska, Lincoln.

You can get the same results using Maple's Non-linear Fit package.

Days have been numbered from 0 for the first date in the sample. Data has not been modified, even to remove errors such as negative deaths or oscillations due to reporting gaps.

References

.

R.E. Hope-Simpson. (1981) The role of season in the epidemiology of influenza
J. Hyg., Camb. **86**, 35–47