Orthogonal Functions for Evaluating Social Distancing Impact on CoVID-19 Spread

Genghmun Eng
PhD Physics 1978, University of Illinois at Urbana-Champaign
geng001@socal.rr.com

June 20, 2020

Abstract

Early CoVID-19 growth often obeys: $N\{\widehat{t}\} \approx N_I \exp[+K_o\,\widehat{t}\,]$, with $K_o = [(\ln 2)/(t_{dbl})]$, where t_{dbl} is the pandemic *doubling time*, prior to society-wide *Social Distancing*. Previously, we modeled *Social Distancing* with t_{dbl} as a linear function of time, where $N[t] \approx \mathbf{1}\ \exp[+K_A\,t\,/\,(1{+}\gamma_o t)]$ is used here. Additional parameters besides $\{K_o,\,\gamma_o\}$ are needed to better model different $\rho[t] = dN[t]/dt$ shapes. Thus, a new *Orthogonal Function Model [OFM]* is developed here using these orthogonal function series:

$$N(Z) = \sum_{m=0}^{m=M_F} g_m\,L_m(Z)\,\exp[-Z]\,,$$

$$R(Z) = \sum_{m=0}^{m=M_F} c_m\,L_m(Z)\,\exp[-Z]\,,$$

where $N(Z)$ and $Z[t]$ form an implicit $N[t] \equiv N(Z[t])$ function, giving:

$$G_o \equiv [K_A/\gamma_o]\,,\ Z[t] \equiv +[G_o\,/\,(1+\gamma_o\,t)]\,,$$
$$\rho[t] \equiv [\gamma_o\,/\,G_o]\ Z^2\,R(Z)\,,$$

with $L_m(Z)$ being the *Laguerre Polynomials*. At large M_F values, nearly *arbitrary* functions for $N[t]$ and $\rho[t] = dN[t]/dt$ can be accommodated. How to determine $\{K_A,\,\gamma_o\}$ and the $\{g_m;\,m = (0, +M_F)\}$ constants from any given $N(Z)$ dataset is derived, with $\rho[t]$ set by:

$$c_{M_F-k} = \sum_{m=0}^{m=k} g_m\,.$$

The *bing.com* USA CoVID-19 data was analyzed using $M_F = (0, 1, 2)$ in the *OFM*. All results agreed to within about 10 percent, showing model robustness. Averaging over all these predictions gives the following overall estimates for the number of USA CoVID-19 cases at the pandemic end:

$< N_{\max} > = 5,009,677\ \pm\ 269,450$ (data to 5/3/20), and

$< N_{\max} > = 4,422,803\ \pm\ 162,580$ (data to 6/7/20),

which compares the pre- and post-early May *bing.com* revisions. The CoVID-19 pandemic in Italy was examined next. The $M_F = 2$ limit was inadequate to model the Italy $\rho[t]$ pandemic tail. Thus, regions with a quick CoVID-19 pandemic shutoff may have additional *Social Distancing* factors operating, beyond what can be easily modeled by just progressively lengthening pandemic *doubling times* (with *13 Figures*).

1 Introduction

The early stages of the CoVID-19 coronavirus pandemic around the world showed a nearly exponential rise in the number of infections with time. If a significant fraction of the population gets infected (*"saturated pandemic"*), exponential growth no longer applies. However, *Social Distancing* can also mitigate exponential growth, enabling pandemic shutoff with only a small fraction of the population being infected (*"dilute pandemic"*). Let $N\{\widehat{t}\}$ be the total number of CoVID-19 cases in any given region, with $\rho\{\widehat{t}\}$ being the predicted number of daily new CoVID-19 cases, so that:

$$N\{\widehat{t}\} = \int_{t'=0}^{t'=\widehat{t}} \rho\{t'\}\, dt' \,. \qquad [1.1a]$$

$$\rho\{\widehat{t}\} = dN\{\widehat{t}\} \,/\, d\widehat{t}\,, \qquad [1.1b]$$

On 3/25/2020, the *Institute of Health Metrics and Evaluation, University of Washington* (IHME) released their initial model for CoVID-19 spread[1] where:

> "*The cumulative death rate for each location is assumed to follow a parametrized Gaussian error function.*"

Since the IHME $\rho\{\widehat{t}\}$ used Gaussians, their projections assumed that the rise to the pandemic peak and its subsequent fall would be symmetric. Their implicit assumption was that the amount of *Social Distancing* was exactly what was needed to make their model predictions true. Given a sharp $\rho\{\widehat{t}\}$ rise, our concern was that the IHME model did not allow $\rho\{\widehat{t}\}$ to decrease gradually.

As a result, we developed an alternative CoVID-19 spread model, which assumed[2] *Social Distancing* gradually lengthens the CoVID-19 *doubling time*. The initial exponential growth factor $K_o = [(\ln 2)/t_{dbl}]$ was used as a starting point, where t_{dbl} was the initial *doubling time*. A new *Social Mitigation Parameter [SMP]* α_S was introduced to account for society-wide *Social Distancing* measures. A linear function was used for *doubling time* lengthening as a simple extension beyond a constant K_o, giving:

$$N\{\widehat{t}\} \approx N_I \, \exp[+K_o\, \widehat{t}\,/\,(1+\alpha_S\, \widehat{t}\,)\,]\,, \qquad [1.2a]$$

$$N\{\widehat{t} \to \infty\} \approx N_I \, \exp[+K_o\,/\,\alpha_S\,] = N^o_{\max}\,, \qquad [1.2b]$$

as an *Initial Model*[2] for the number of CoVID-19 cases, where $\widehat{t} = 0$ was the start of society-wide *Social Distancing*. Both $N_I = N\{\widehat{t} = 0\}$ and $N_F = N\{\widehat{t} = t^{data}_{end}\}$, as the most recently available data, were treated as fixed points. A minimum *root-mean-square* (*rms*) error datafit, using a logarithmic Y-axis, sets the $\{K_o, \alpha_S\}$ values. The resulting $N\{\widehat{t} \to \infty\}$ of Eq. [1.2b] is the predicted final number of CoVID-19 cases at the pandemic end.

On 4/29/2020, we sent our preprint[2] to the IHME, the Los Angeles Department of Public Health (LADPH), and to Profs. Goldenfeld and Maslov at UOI (University of Illinois at Urbana-Champaign), who were preparing a 2-day nationwide CoVID-19 remote-learning seminar for 5/6/2020 and 5/8/2020.

Also, on 4/29/2020, the IHME electronically published their 12^{th} CoVID-19 update, using their 3/25/2020 model. A graphic display of their most recent $\rho\{\widehat{t}\}$ projections showed a symmetric rise and fall. This graph was widely publicized by Dr. Alan Boyle, who was following the IHME work, summarizing it

for general audiences[3−5]. Since our Eqs. [1.2a]-[1.2b] model gave substantially different $\rho\{\widehat{t}\}$ predictions than IHME, we added a note to that effect in our pre-print, submitting the final pre-print to MedRxiv on 5/4/2020, where it was accepted and published on-line on 5/8/2020.

Concurrently, on 5/4/2020, IHME published their 13^{th} CoVID-19 update[6], where everything changed. Dr. Alan Boyle[5] summarized those changes with a note that: "*[IHME] researchers acknowledged that their previous modeling wasn't sophisticated enough*". Both IHME graphical predictions for 4/29/2020 and 5/4/2020 are shown in **Figure 1**, to highlight this change.

On 5/6/2020 and 5/8/2020, Profs. Goldenfeld and Maslov presented their UOI team's supercomputer-based $\rho\{\widehat{t}\}$ CoVID-19 projections, which also were very asymmetric. Although mathematical details for the UOI and new IHME projections are not known, virtually all $\rho\{\widehat{t}\}$ CoVID-19 projections are now asymmetric, as the developing CoVID-19 data also appears to be.

Since our *Initial Model* had only two data fitting parameters $\{K_o, \alpha_S\}$, we became concerned that those two parameters might not be sufficient to adequately describe all the different $\rho\{\widehat{t}\}$ shapes observed. To correct this potential defect, a new *Orthogonal Function Model [OFM]* is developed here to allow more accurate descriptions for a variety of $\rho\{\widehat{t}\}$ shapes, using additional mathematical techniques derived herein. This *OFM* extends Eqs. [1.2a]-[1.2b], and provides additional fitting parameters to improve $\rho\{\widehat{t}\}$ projections.

2 *Orthogonal Function Model [OFM]* Elements

The following items and methods were developed as part of this *OFM* to improve CoVID-19 projections for a variety of $\rho\{\widehat{t}\}$ data shapes.

First, the $\widehat{t} = 0$ point in Eq. [1.2a] was time-shifted so that $N[t = 0] \equiv \mathbf{1}$. This $t = 0$ point now provides an estimate for the CoVID-19 pandemic starting point, replacing the above $N\{\widehat{t}\}$ with this time-shifted version:

$$N[t] \approx \mathbf{1} \exp[+K_A\, t\,/\,(1+\gamma_o\, t)] = \exp[+G_o]\, \exp[-Z]\,, \qquad [2.1a]$$

$$G_o \equiv [K_A\,/\,\gamma_o]\,, \qquad [2.1b]$$

$$Z[t] \equiv +[G_o\,/\,(1+\gamma_o\, t)]\,, \qquad [2.1c]$$

$$N[t \to \infty,\ Z \to 0] \approx \mathbf{1}\, \exp[+K_A\,/\,\gamma_o]\ = \exp[+G_o] = N^o_{\max}\,, \qquad [2.1d]$$

$$\rho[t] = dN[t]\,/\,dt\,, \qquad [2.1e]$$

$$N[t] = \int_{t'=0}^{t'=t} \rho[t']\, dt'\,, \qquad [2.1f]$$

which enables Eq. [2.1a] to become a 1-term approximation for a larger function series. Actual data provides the $\{N_I,\, N_F\}$ values. However, these $\{t = t_I,\ N[t_I] \equiv N_I\}$ and $\{t = t_F,\ N[t_F] \equiv N_F\}$ limits are now used to set $\{K_A,\, G_o,\, \gamma_o\} > 0$, so that the $N^o_{\max}$ of Eq. [1.2b] and Eq. [2.1d] match exactly.

Second, when $Z \to 0$ in Eq. [2.1c] then $t \to \infty$; while $Z \to +\infty$ corresponds to $t \to (-1/\gamma_o)+\varepsilon$, where ε is arbitrarily small and positive. Since $N[t = 0] = \mathbf{1}$, the $t < 0$ domain has $N[t] < 1$, while setting a particular time as the $N[t] = 0$ point. Since the $\mathbf{1} > N[t] > 0$ regime has no impact on this overall analysis, virtually any decreasing function tail for the $Z \to +\infty$ limit should be allowed.

Third, instead of generalizing Eq. [2.1a] using time, it is easier to use functions of Z, where Z is given by Eq. [2.1c]. It results in these $N(Z)$ and $R(Z)$ substitutes for $N[t]$ and $\rho[t]$:

$$N(Z) \equiv \int_{Z'=Z}^{Z'=+\infty} R(Z')\, d\, Z'. \qquad [2.2]$$

Given explicit functions of Z, both $N(Z)$ and $R(Z)$ in Eq. [2.2] go from large$-Z$ to smaller$-Z$ values at longer times, eventually approaching the $Z = 0$ point. Together, $N(Z)$ and $Z[t]$ create an implicit $N(Z[t]) \equiv N[t]$ function, and $R(Z)$ and $Z[t]$ create another implicit $R(Z[t]) \equiv R[t]$ function. A standard change of variables converts them back into being explicit functions of time:

$$N[t] \equiv \int_{\hat{z}=Z[t],\ t'=t}^{\hat{z}=+\infty,\ t'=(-1/\gamma_o)} R(Z[t'])\, \tfrac{dZ}{dt'}\, d\, t' = \qquad [2.3]$$

$$\int_{t'=t}^{t'=(-1/\gamma_o)} R(Z[t'])\, [\tfrac{-G_o\, \gamma_o}{(1+\gamma_o t')^2}]\, d\, t' = (\tfrac{\gamma_o}{G_o}) \int_{t'=(-1/\gamma_o)}^{t'=t} (Z[t'])^2\, R(Z[t'])\, dt'\,.$$

It gives these equivalences between and Eq. [2.2] and Eq. [2.1f]:

$$N[t] \equiv \int_{t'=(-1/\gamma_o)}^{t'=t} \rho[t']\, dt'\,, \qquad [2.4a]$$

$$\rho[t] \equiv (\gamma_o\, /\, G_o\,)\, Z^2\, R(Z)\,, \qquad [2.4b]$$

$$Z[t] \equiv +[G_o\, /\, (1 + \gamma_o\, t)]\,. \qquad [2.4c]$$

Fourth, to allow additional data fitting parameters, the *OFM* replaces the 1-term approximation of Eq. [2.1a] with these orthogonal function series:

$$N(Z) = \sum_{m=0}^{m=M_F} g_m\, L_m(Z)\, \exp[-Z]\,, \qquad [2.5a]$$

$$R(Z) = \sum_{m=0}^{m=M_F} c_m\, L_m(Z)\, \exp[-Z]\,. \qquad [2.5b]$$

These series have $\exp[-Z]$ as their weighting function, while keeping the Eqs. [2.1b]-[2.1c] definitions for $\{Z, G_o, \gamma_o\}$. The $\{g_m;\, m = (0, +M_F)\}$ and $\{c_m;\, m = (0, +M_F)\}$ coefficients are constants that can be derived from each dataset. For a wide range of $N(Z)$ and $R(Z)$ functions, larger M_F values and more $\{L_m(Z);\, m = (0, +M_F)\}$ terms give progressively better matches to practically any *arbitrary function*. This feature is what enables improved datafits over a variety of measured $N[t]$ and $\rho[t]$ curves.

Fifth, the *OFM* uses the $\{N_I[t_I],\, N_F[t_F]\}$ data end-points to set $\{G_o,\, \gamma_o\}$ in Eq. [2.4c], and define Z, allowing the *OFM* to provide best fits over the whole data range of Z or t, while these end points are fixed in the *Initial Model*.

The difference between: (a) using the whole data range for fitting, versus (b) using the data end points for fitting, is most evident when comparing Eq. [2.1a] to Eq. [2.5a]. In Eq. [2.1a], $N[t] = G_o\, \exp[-Z]$ where G_o has a pre-set value, whereas in Eq. [2.5a], $N(Z) = g_0\, \exp[-Z]$ for $M_F = 0$, the g_0 parameter is determined by fitting over the whole data range.

Sixth, both Z and t essentially span from $\{0, +\infty\}$. Using $\exp[-Z]$ as a weighting function over that domain makes the choice of $L_m(Z)$ in Eq. [2.5a]-[2.5b] unique. They are the *Laguerre Polynomials*, with the first few being:

$$L_{-1}(Z) \equiv 0\,, \tag{2.6a}$$
$$L_0(Z) \equiv 1 = L_m(Z=0)\,, \tag{2.6b}$$
$$L_1(Z) \equiv (1-Z)\,, \tag{2.6c}$$
$$L_2(Z) \equiv (1-2\,Z+\tfrac{1}{2}Z^2)\,, \tag{2.6d}$$
$$L_3(Z) \equiv (1-3\,Z+\tfrac{3}{2}Z^2-\tfrac{1}{6}Z^3)\,, \tag{2.6e}$$
$$L_4(Z) \equiv (1-4\,Z+3\,Z^2-\tfrac{2}{3}Z^3+\tfrac{1}{24}Z^4)\,. \tag{2.6f}$$

Some important properties of the *Laguerre Polynomials* are:

$$\int_{Z=0}^{Z=+\infty} L_m(Z)\,L_n(Z)\,\exp(-Z)\,dZ = \mathbf{1}\,\delta_{m,n}\,, \tag{2.7a}$$
$$\delta_{m,n} = \begin{pmatrix} \mathbf{1} & for\ m=n \\ \mathbf{0} & otherwise \end{pmatrix}, \tag{2.7b}$$
$$\int_{Z'=Z}^{Z'=+\infty} L_m(Z')\,\exp(-Z')\,dZ' = [L_m(Z)-L_{m-1}(Z)]\,\exp(-Z)\,, \tag{2.7c}$$
$$L_m(Z)\,\exp(-Z) = \tfrac{1}{m!}\tfrac{d^m}{dZ^m}[Z^m\,e^{-Z}] = e^{-Z}\sum_{k=0}^{k=m}(-1)^k\tfrac{m!}{k!\,(m-k)!}\left[\tfrac{Z^k}{k!}\right], \tag{2.7d}$$
$$L_{m\geq 2}(Z) = [2-\tfrac{(Z+1)}{m}]\,L_{m-1}(Z) - [1-\tfrac{1}{m}]\,L_{m-2}(Z)\,, \tag{2.7e}$$

where Eq. [2.7a] defines an *orthogonal function set*. Given $\mathbf{n}$ is an integer in Eq. [2.7d], $\mathbf{n}$-*factorial* ($\mathbf{n}!$) is defined as the product:

$$\mathbf{n}! \equiv (\mathbf{n})(\mathbf{n}-1)(\mathbf{n}-2)(\mathbf{n}-3)...(3)(2)(1)\,, \tag{2.8a}$$
$$\mathbf{1}! \equiv \mathbf{0}! \equiv 1\,, \tag{2.8b}$$

along with *factorials* involving negative integers not being allowed.

Seventh, when data are used to determine the $\{g_m;\ m=(0,M_F)\}$ constants for the Eq. [2.5a] $N(Z)$ analytic approximation, an equivalently precise $R(Z)$ is set by Eq. [2.2] and Eq. [2.5b], with its $\{c_m;\ m=(0,M_F)\}$ constants being:

$$c_{M_F-k} = \sum_{m=0}^{m=k} g_m\,. \tag{2.9}$$

This simple form of Eq. [2.9] arises from the fact that $L_m(Z=0)=1$. Also, Eq. [2.5a] and Eq. [2.9] combine to give:

$$N(Z=0) \equiv N(0) \equiv N[t\to\infty] = c_0 \equiv \sum_{m=0}^{m=M_F} g_m\,, \tag{2.10}$$

as a new predicted total number of CoVID-19 cases at pandemic end.

Eighth, the $\{g_m;\ m=(0,M_F)\}$ constants can be arranged in a $\overrightarrow{g}-$vector form, with comparable constants for $R(Z)$ from Eq. [2.2] arranged in a $\overrightarrow{C}-$vector form. It allows Eq. [2.9] to be written as:

$$\overrightarrow{C} = \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \overrightarrow{g} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix}. \tag{2.11}$$

Once the $\{g_m;\ m=(0,M_F)\}$ constants are found, the c_0 value in Eq. [2.11] becomes the $\{M_F+1\}$-term replacement value for the predicted total number of CoVID-19 cases at the pandemic end, which refines the initial $N^o_{\max}$ value of Eq. [1.2b] or Eq. [2.1d]. How to determine $\{K_A,\,G_o,\,\gamma_o\}$ and the $\{g_m;\ m=(0,M_F)\}$ constants in Eq. [2.5a] from a given set of data, is derived next.

3 Finding $\{K_A, \gamma_o\}$ for $Z[t]$ from Data

For a given dataset, the *OFM* begins with using Eq. [1.2a] to set $\{K_o, \alpha_S\}$, as in our *Initial Model*. Society-wide *Social Distancing* is assumed to occur at or before the time t_I, where N_I cases are already observed. Since the most recently available data at t_F has N_F cases, Eq. [2.1a] becomes:

$$N[t_I] \approx \mathbf{1} \exp[+K_A\, t_I\,/\,(1+\gamma_o\, t_I)] = N_I\,, \qquad [3.1a]$$

$$N[t_F] \approx \mathbf{1} \exp[+K_A\, t_F\,/\,(1+\gamma_o\, t_F)] = N_F\,, \qquad [3.1b]$$

$$N[t \to \infty] \approx \mathbf{1} \exp[+K_A\,/\,\gamma_o] = N^o_{\max}\,, \qquad [3.1c]$$

which using the new $t = 0$ point for the *OFM*. Evaluating $N[t < t_I]$ for $t < t_I$ estimates what the pandemic prior history might have been, had society-wide *Social Distancing* already been in place. Evaluating $N[t > t_F]$ for $t > t_F$ estimates how the pandemic evolves assuming these *Social Distancing* measures remain in place. The prior Eq. [1.2a] gave:

$$N[t_F] = N_I\, \exp\{+K_o\,(t_F - t_I)/\,[1+\alpha_S\,(t_F - t_I)]\}\,, \qquad [3.2]$$

with the $t_F \to 0$ limit of Eq. [3.2] giving:

$$N_I\, \exp[-K_o\, t_I\,/\,(1-\alpha_S\, t_I)] \equiv \mathbf{1}\,, \qquad [3.3a]$$

$$t_I\, = \ln(N_I)\,/[K_o + \alpha_S\, \ln(N_I)]\,. \qquad [3.3b]$$

Here, Eq. [3.3b] sets the precise t_I time shift needed to convert from Eq. [1.2a] to Eq. [2.1a], which is easier to generalize. In addition, the $t = 0$ point of Eq. [2.1a] gives $N[t \to 0] = \mathbf{1}$ as an estimate for the pandemic starting point.

Since t_I and $(t_F - t_I)$ sets t_F, the Eqs. [3.1a]-[3.1b] fully determine $\{K_A,\, \gamma_o\}$, without needing any recalculations on the original dataset. Taking various ratios of Eq. [3.1b] to Eq. [3.1a] gives:

$$\ln(N_F)\,/\,\ln(N_I) = \tfrac{t_F}{t_I}\,\tfrac{(1+\gamma_o\, t_I)}{(1+\gamma_o\, t_F)}\,, \qquad [3.4a]$$

$$\ln[N_F\,/\,N_I] = K_A\,[\tfrac{t_F}{(1+\gamma_o\, t_F)} - \tfrac{t_I}{(1+\gamma_o\, t_I)}]\,, \qquad [3.4b]$$

as separable equations to first find γ_o, then K_A, with these results:

$$\gamma_o = \{\,[\ln(N_I)\,/\,t_I] - [\ln(N_F)\,/\,t_F]\,\}\,/\,[\,\ln(N_F) - \ln(N_I)\,]\,, \qquad [3.5a]$$

$$K_A = [\,(1\,/\,t_I) - (1\,/\,t_F)\,]\,/\,\{\,[1\,/\,\ln(N_I)] - [1\,/\,\ln(N_F)]\,\}\,, \qquad [3.5b]$$

which sets the $Z[t]$ function in Eq. [2.1c] or Eq. [2.4c].

4 Determining the g_m Constants from Data

When data for $N_{data}(Z)$ are given over the whole $Z = \{0^+, \infty^-\}$ range, the g_n constants for Eq. [2.5a] are exactly determined via:

$$N_{data}(Z) = \sum_{m=0}^{m=M_F} g_m\, L_m(Z)\, \exp[-Z]\,. \qquad [4.1a]$$

$$\int_{Z=0}^{Z=+\infty} L_n(Z)\, N_{data}(Z)\, dZ = \qquad [4.1b]$$

$$\sum_{m=0}^{m=M_F} g_m \int_{Z=0}^{Z=+\infty} L_n(Z)\, L_m(Z)\, \exp[-Z]\, dZ \equiv g_n,$$

where the *Laguerre Polynomial* orthogonality condition of Eq. [2.7a] forces the Eq. [4.1b] sum to reduce to one term.

When the $N_{data}(Z)$ only spans a finite range of: $t_I < t < t_F$ and $Z_{\min} < Z < Z_{\max}$, an extrapolation of $N_{data}(Z)$ for $(Z < Z_{\min})$ and $(Z > Z_{\max})$ is needed. One method could set $N_{data}(Z < Z_{\min}) \equiv 0$ and $N_{data}(Z > Z_{\max}) \equiv 0$, which results in these Eqs. [4.1a]-[4.1b] cognates:

$$N_{data}(Z) = \sum_{m=0}^{m=M_F} \widehat{g}_m \, L_m(Z) \, \exp[-Z] \, , \tag{4.2a}$$

$$\int_{Z=Z_{\min}}^{Z=Z_{\max}} L_n(Z) \, N_{data}(Z) \, dZ = \tag{4.2b}$$

$$\sum_{m=0}^{m=M_F} \widehat{g}_m \int_{Z=0}^{Z=+\infty} L_n(Z) \, L_m(Z) \, \exp[-Z] \, dZ \equiv \widehat{g}_n \, .$$

Its advantages are: (a) for $m \neq n$, every $\widehat{g}_m$ and $\widehat{g}_n$ are independent, as in orthogonal functions; and (b) these $\widehat{g}_m$ values provide new estimates for the $N_{data}(Z < Z_{\min})$ and $N_{data}(Z > Z_{\max})$ regimes. But since $N_{data}(Z < Z_{\min})$ and $N_{data}(Z > Z_{\max})$ were originally assumed to vanish, this method is inconsistent. Alternatively, adding reasonable "*tails*" to the data could extend the original $N_{data}(Z)$ domain, but those functions are not always known.

The third path, used here, takes the Eq. [4.1a] "*final answer*" as a *self-consistent* extrapolation for $(Z < Z_{\min})$ and $(Z > Z_{\max})$, while retaining the $N_{data}(Z)$ values for the $(Z_{\max} \geq Z \geq Z_{\min})$ regime. It replaces Eq. [4.1b] with:

$$g_n \equiv \sum_{m=0}^{m=M_F} g_m \int_{Z=0}^{Z=+\infty} L_n(Z) \, L_m(Z) \, \exp[-Z] \, dZ = \tag{4.3a}$$

$$\int_{Z=Z_{\max}}^{Z=+\infty} L_n(Z) N(Z) dZ + \int_{Z=Z_{\min}}^{Z=Z_{\max}} L_n(Z) N_{data}(Z) dZ + \int_{Z=0}^{Z=Z_{\min}} L_n(Z) N(Z) dZ \, ,$$

$$N(Z) = \sum_{m=0}^{m=M_F} g_m \, L_m(Z) \, \exp[-Z] \, . \tag{4.3b}$$

The $\{g_m; \; m = (0, M_F)\}$ now appears on both sides of each Eq. [4.3a] g_n-equation, which is handled as follows. Defining:

$$Q_n \equiv \int_{Z=Z_{\min}}^{Z=Z_{\max}} L_n(Z) \, N_{data}(Z) \, dZ \, , \tag{4.4a}$$

$$K_{m,n} \equiv \int_{Z=Z_{\min}}^{Z=Z_{\max}} L_m(Z) \, L_n(Z) \, \exp(-Z) \, dZ = K_{n,m} \, , \tag{4.4b}$$

Eqs. [4.3a]-[4.3b] can be re-written as a 3×3 matrix $\underline{\mathbf{M}}_3$, which relates a data-driven $\overrightarrow{Q}_3$-vector to a resultant $\overrightarrow{g}_3$-vector:

$$\overrightarrow{Q}_3 \; = \; \underline{\mathbf{M}}_3 \; \overrightarrow{g}_3 \, , \tag{4.5a}$$

$$\begin{pmatrix} Q_o \\ Q_1 \\ Q_2 \end{pmatrix} = \begin{pmatrix} K_{0,0} & K_{0,1} & K_{0,2} \\ K_{1,0} & K_{1,1} & K_{1,2} \\ K_{2,0} & K_{2,2} & K_{2,2} \end{pmatrix} \begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix} . \tag{4.5b}$$

$$(\underline{\mathbf{M}}_3)^{-\mathbf{1}} \; \overrightarrow{Q}_3 \; \equiv \overrightarrow{g}_3 \, , \tag{4.5c}$$

where $(\underline{\mathbf{M}}_3)^{-\mathbf{1}}$ is the matrix inverse of $\underline{\mathbf{M}}_3$. When $\{Z_{\min}, \, Z_{\max}\} \rightarrow \{0, +\infty\}$, this $\underline{\mathbf{M}}_3$ becomes the Identity Matrix. The following $k_{m,n}(Z)$ integrals set $K_{m,n}$:

$$k_{m,n}(Z) = \int_{Z'=Z}^{Z'=+\infty} L_m(Z') \, L_n(Z') \, \exp(-Z') \, dZ' = k_{n,m}(Z) \, , \tag{4.6a}$$

$$K_{m,n} \equiv k_{m,n}(Z_{\min}) - k_{m,n}(Z_{\max}) = K_{n,m} \, . \tag{4.6b}$$

The $k_{m,n}(Z)$ integrals can be determined using Eq. [2.7c], which gives:

$$k_{0,0}(Z) = 1\exp(-Z)\,, \tag{4.7a}$$
$$k_{1,1}(Z) = \{1\ + Z^2\}\exp(-Z)\,, \tag{4.7b}$$
$$k_{2,2}(Z) = \{1 + 2\,Z^2 - Z^3 + \tfrac{1}{4}Z^4\}\exp(-Z)\,, \tag{4.7c}$$
$$k_{0,1}(Z) = (-Z)\ \exp(-Z)\,, \tag{4.7d}$$
$$k_{0,2}(Z) = (-Z)\,\{1 - \tfrac{1}{2}Z\}\exp(-Z)\,, \tag{4.7e}$$
$$k_{1,2}(Z) = (-Z)\,\{1 - Z + \tfrac{1}{2}Z^2\}\exp(-Z)\,. \tag{4.7f}$$

To extract $\{g_0,\, g_1,\, g_2\}$, the 3×3 symmetric $\underline{\mathbf{M}}_3$ matrix needs inversion:

$$\underline{\mathbf{M}} = \begin{pmatrix} a & d & f \\ d & b & e \\ f & e & c \end{pmatrix}\,, \tag{4.8a}$$
$$\det[\underline{\mathbf{M}}] \equiv [abc - ae^2 - bf^2 - cd^2 + 2\,def]\,, \tag{4.8b}$$
$$\det[\underline{\mathbf{M}}]\ (\underline{\mathbf{M}})^{-1} \equiv \begin{pmatrix} [bc - e^2] & -(cd - ef) & -(bf - de) \\ -(cd - ef) & [ac - f^2] & -(ae - df) \\ -(bf - de) & -(ae - df) & [ab - d^2] \end{pmatrix}\,, \tag{4.8c}$$

which determines $\{g_0,\, g_1,\, g_2\}$ from the $\{Q_0,\, Q_1,\, Q_2\}$ data. A best fit $N(Z)$ for $Z = \{0^+, \infty^-\}$ results, along with an equivalent fit for $R(Z)$ using Eq. [2.9].

Instead of having to find the best $\{g_0,\, g_1,\, g_2\}$ triplet, one could find the best $\{g_0',\, g_1'\}$ by just using using $\{Q_0,\, Q_1\}$ and an $\underline{\mathbf{M}}_2$ sub-matrix; or one could find the best $\{g_0^+\}$ by itself by just using $\{Q_0\}$ and an $\underline{\mathbf{M}}_1$ sub-matrix:

$$\overrightarrow{Q}_2 \ = \ \underline{\mathbf{M}}_2\ \overrightarrow{g}_2\,, \tag{4.9a}$$
$$\begin{pmatrix} Q_o \\ Q_1 \end{pmatrix} = \begin{pmatrix} K_{0,0} & K_{0,1} \\ K_{1,0} & K_{1,1} \end{pmatrix} \begin{pmatrix} g_0' \\ g_1' \end{pmatrix}\,, \tag{4.9b}$$
$$\begin{pmatrix} g_0' \\ g_1' \end{pmatrix} = \begin{pmatrix} K_{0,0} & K_{0,1} \\ K_{1,0} & K_{1,1} \end{pmatrix}^{-1} \begin{pmatrix} Q_o \\ Q_1 \end{pmatrix} = \frac{1}{[K_{0,0}\,K_{1,1} - K_{0,1}\,K_{1,0}]} \begin{pmatrix} K_{1,1} & K_{1,0} \\ K_{0,1} & K_{0,0} \end{pmatrix} \begin{pmatrix} Q_o \\ Q_1 \end{pmatrix}, \tag{4.9c}$$
$$\overrightarrow{Q}_1 \ = \ \underline{\mathbf{M}}_1\ \overrightarrow{g}_1\,, \tag{4.9d}$$
$$(\ Q_o\ ) = (\ K_{0,0}\ )\ (\ g_o^+\ )\,, \tag{4.9e}$$
$$(\ g_o^+\ ) = (\ K_{0,0}\ )^{-1}\ (\ Q_o\ )\,. \tag{4.9f}$$

When the $N_{data}(Z)$ is comprised of $j = \{1, 2, ...J\}$ discrete values between $\{Z_{\min},\, Z_{\max}\}$, with each Z_j having an $N_{data}^{(j)}(Z_j)$ value, the Eq. [4.4a] integral needs to be replaced by a sum. Let:

$$Z_j \equiv +[G_o\,/\,(1 + \gamma_o\,t_j)]\,, \tag{4.10}$$

with $Z_0 = Z_1$ and $Z_{J+1} = Z_J$, the Q_n replacement for Eq. [4.4a] is:

$$Q_n \equiv \sum_{j=1}^{j=J} L_n(Z_j)\ N_{data}^{(j)}(Z_j)\ \Delta_j\,. \tag{4.11a}$$
$$\Delta_j \equiv \tfrac{1}{2}|Z_{j+1} - Z_{j-1}|\,, \tag{4.11b}$$

with the $N[t]$ and $\rho[t]$ being set by Eq. [2.3] and Eqs. [2.4a]-[2.4c].

Finally, the Eqs. [4.7a]-[4.7f] $k_{m,n}(Z)$ integrals are easy to compute for $0 \le m \le 2$ and $0 \le n \le 2$. But the general case is not well-known or tabulated in many *Tables of Integrals*. The key is how to express a product of two *Laguerre Polynomials* efficiently as a sum over a larger set of single *Laguerre Polynomials*, so as to convert the Eq. [4.6a] integrals into the Eq. [2.7c] form.

This problem was originally solved by G. N. Watson[7] in 1938, and simplified by J. Gillis and G. Weiss[8] in 1960. It is a sum of terms, where each coefficient contains four different *factorials* involving integers. Their key result is:

$$L_r(Z)\, L_s(Z) = \sum_{t=|r-s|}^{t=(r+s)} C_{rst}\; L_t(Z)\,, \qquad [4.12a]$$

$$C_{rst} = \int_{X=0}^{X=+\infty} L_r(X)\, L_s(X)\, L_t(X)\, \exp(-X)\; dX\,, \qquad [4.12b]$$

$$C_{rst} \equiv \frac{(-1)^p}{2^p} \sum_{n=0}^{n=(r+s)} (2^{2n}) \frac{(r+s-n)!}{(r-n)!\,(s-n)!\,(2n-p)!\,(p-n)!}, \qquad [4.12c]$$

$$p \equiv (r+s-t), \qquad [4.12d]$$

where ALL terms in the sum for $n = \{0, (r+s)\}$ also have an implicit requirement that none of the integer arguments for any of the *factorials* can be negative. Thus, all terms with negative arguments for the *factorial* must be omitted. Nowadays, this calculation can be done on a computer, but it would have been difficult in 1960, and nearly impossible in 1938.

5 USA: *Orthogonal Function Model* Results

This USA analysis only uses data after mid-March 2020, when several State Governors instituted mandatory *Mitigation Measures*. The widely available *bing.com* CoVID-19 data[9] for the USA had these limits:

$$N_I[t_I \leftrightarrow 3/21/2020]_{day\#1} = \{25,722\}\,, \qquad [5.1a]$$

$$N_F[t_F \leftrightarrow 5/03/2020]_{day\#44} = \{1,183,653\}\,, \qquad [5.1b]$$

with $(t_F - t_I) = 43$ *days*. Our *Initial Model* of Eq. [1.2a] sets these parameter values for the USA :

$$K_o = \{0.3248758\,/\,day\}\,, \qquad [5.2a]$$

$$\alpha_S = \{0.06159\,/\,day\}\,, \qquad [5.2b]$$

$$N[t \to \infty] \approx N_I\, \exp[+K_o\,/\,\alpha_S] = N^o_{\max} \equiv \{5,024,900\}\,. \qquad [5.2c]$$

Using Eq. [3.3b] for t_I and t_F sets:

$$t_I = \ln(N_I)\,/\,[K_o + \alpha_S\, \ln(N_I)] = \{10.685885\, days\}\,, \qquad [5.3b]$$

$$t_F = t_I + \{43\ days\} = \{53.685885\, days\}\,, \qquad [5.3c]$$

for use in the *OFM*. **Figures 2-3** show how this *Initial Model*, by itself, compares to the USA CoVID-19 data. **Figure 2** uses a logarithmic Y-axis for the predicted total number of CoVID-19 cases, and **Figure 3** shows the daily new CoVID-19 case predictions on a linear Y-axis plot.

The daily new case data exhibits large day-to-day variations, likely due to reporting delays, among other factors. This *Initial Model* for the USA has a predicted maximum of ˜31,760 new cases per day at Day 37.686 on 4/17/2020, along with ˜6,757 new cases per day still occurring at Day 200 on 9/26/2020.

The time axis in **Figure 2** is different than in our previous paper[2], due to the time shift of Eq. [2.1a], where the new $t = 0$ point estimates the CoVID-19 pandemic starting point being on 3/10/2020. Even if *Social Distancing* had been in effect at the start of the pandemic, **Figure 2** shows that the $N_I[t_I] = \{25,722\}$ level still could have been reached in $10 - 11$ days.

Figures 3 compares the measured data for the total number of CoVID-19 cases after *Social Distancing* started, to the early-time portion of this *Initial Model*. That comparison shows that the early-time data starts off a little below the curve; the later-time data rises a bit above the curve; and the final-time data again matches the curve, since it is a fixed point for this analysis.

These predictions assume: (I) The present *Mitigation Measures* are continued; (II) No "*second wave*" of infection or re-infection occurs; and (III) No further *Mitigation Measures* are taken to reduce the number of CoVID-19 cases.

These *Initial Model* results are first refined by applying the Eq. [2.1a] time shift, with Eqs. [3.5a]-[3.5b] setting these $\{\gamma_o,\, K_A,\, G_0\}$ values:

$$\gamma_o = \{0.1801634\ /day\}\,, \tag{5.4a}$$
$$K_A = \{2.779906\ /day\}\,, \tag{5.4b}$$
$$G_o \equiv [K_A\,/\,\gamma_o] = \{15.4299153\}\,, \tag{5.4c}$$

where Eq. [3.1c] also gives:

$$\underset{t\to\infty}{Lim}\{\mathbf{1}\exp[\tfrac{+K_A\,t}{(1+\gamma_o\,t)}]\} = \mathbf{1}\exp[+\tfrac{K_A}{\gamma_o}] = N^o_{\rm max} \approx \{5,024,900\}\,, \tag{5.5}$$

which matches Eq. [5.2c], as it should. Then:

$$Z[t] = +[G_o\,/\,(1+\gamma_o\,t)] = [15.4299153\,/\,(1+0.1801634\ t)\,, \tag{5.6}$$

defines Z for the *OFM*, where:

$$Z_{\rm min}[t_F \approx 53.686] = +[G_o\,/\,(1+\gamma_o\,t_F)] = 1.4458002586\ , \tag{5.7a}$$
$$Z_{\rm max}[t_I \approx 10.686] = +[G_o\,/\,(1+\gamma_o\,t_I)] = 5.2748142044\ . \tag{5.7b}$$

The resultant Eq. [4.5b] $\underline{\mathbf{M}}_3$ matrix of $K_{m,n}$ entries is:

$$\underline{\mathbf{M}}_3 = \begin{pmatrix} K_{0,0} & K_{0,1} & K_{0,2} \\ K_{1,0} & K_{1,1} & K_{1,2} \\ K_{2,0} & K_{2,2} & K_{2,2} \end{pmatrix} = \begin{pmatrix} +0.23044 & -0.31357 & -0.13858 \\ -0.31357 & +0.58041 & +0.05609 \\ -0.13858 & +0.05609 & +0.23635 \end{pmatrix}. \tag{5.8}$$

It has a rather small $\det[\underline{\mathbf{M}}_3] = 0.001375$ value, with an inverse of:

$$(\underline{\mathbf{M}}_3)^{-\mathbf{1}} \equiv \begin{pmatrix} 97.478 & 48.246 & 45.707 \\ 48.246 & 25.643 & 22.204 \\ 45.707 & 22.204 & 25.762 \end{pmatrix}\ . \tag{5.9}$$

A convolution of $L_m(Z)$ functions with the measured $\overrightarrow{Q}_3$ dataset vector of Eqs. [4.11a]-[4.11b], along with the above $(\underline{\mathbf{M}}_3)^{-\mathbf{1}}$, gives this final $\overrightarrow{g}-$vector:

$$(\underline{\mathbf{M}}_3)^{-\mathbf{1}}\ \overrightarrow{Q}_3 \equiv (\underline{\mathbf{M}}_3)^{-\mathbf{1}} \begin{pmatrix} Q_o \\ Q_1 \\ Q_2 \end{pmatrix} = (\underline{\mathbf{M}}_3)^{-\mathbf{1}} \begin{pmatrix} +1,169,103 \\ -1,576,445 \\ -722,430 \end{pmatrix} \equiv$$
$$\overrightarrow{g}_3 = \begin{pmatrix} g_o \\ g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} +4,884,354 \\ -60,065 \\ -178,415 \end{pmatrix}\,, \tag{5.10}$$

determining the constants needed for $N(Z)$ in Eq. [2.5a]. The coefficients for $R(Z)$, which sets the predicted number of daily new CoVID-19 cases, are:

$$\overrightarrow{C}_3 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}\ \overrightarrow{g}_3 = \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} +4,645,874 \\ -238,480 \\ -178,415 \end{pmatrix}, \tag{5.11}$$

determining the constants needed for $R(Z)$ in Eq. [2.5b]. Using these $\{g_0,\, g_1,\, g_2\}$ values along with Eq. [2.11] gives:

$$N(0) \equiv N[t\to\infty] = c_0 \equiv \{4,645,874\}\,, \tag{5.12}$$

as a new predicted total number of CoVID-19 cases at the pandemic end for the *OFM*, which is a ~7.54% or 379,026 reduction in the number of cases, compared to the *Initial Model* $N^o_{\max}$ value of Eq. [5.5].

Using Eq. [2.4c] for $Z[t]$, and substituting the Eq. [5.12] $\overrightarrow{C}_3$ values into Eq. [2.5b] gives $R(Z)$. The $\rho[t]$ in Eq. [2.4a] is derived from $R(Z)$ using Eq. [2.4b], with the resulting *OFM* $\rho[t]$ plotted in **Figure 4**, using a linear Y-axis, along with the $t > t_I$ raw data for the daily new CoVID-19 cases.

Raw data for $t < t_I$ was not included in these analyses, because they cover the exponential rise period, prior to *Social Distancing*. Those data are not applicable to estimating *Social Distancing* effects.

However, the **Figure 4** *OFM* provides an extrapolation for those $t < t_I$ times, which shows what an exponential rise plus lengthening *doubling times* would have looked like, if both had been operating continuously from the CoVID-19 pandemic start. The companion $N[t]$ analytic result, plotted using a logarithmic Y-axis, along with the $t > t_I$ raw data for the total number of CoVID-19 cases, is show in **Figure 5**.

Comparing the size and timing of the $\rho[t]$ pandemic peak, and its Day 200 value, between the *Initial Model* (**Figs. 2-3**) and *OFM* (**Figs. 4-5**), gives:

$$\begin{pmatrix} \text{Parameter} & \textit{Initial Model} & \textit{Orthog. Func. Model} \\ N[t \to \infty] & 5,024,900 & 4,645,874 \\ \max\{\rho[t_p]\} & 31,760 \ / \ day & 32,069 \ / \ day \\ [t_p] \ \textit{Date} & 4/17/2020 & 4/15/2020 \\ \overset{200}{\underset{Day}{}}\rho[t] & 6,757 & 5,962 \ / \ day \end{pmatrix} . \qquad [5.13]$$

This table shows that the *OFM* predicts fewer total cases ($N^o_{\max}$ *vs* c_o) and fewer daily new CoVID-19 cases at Day 200, as well as giving an earlier and higher pandemic peak prediction.

While the above analysis used $M_F = 2$ with Eq. [5.10] $\overrightarrow{g}_3$ setting the best $\{g_0, g_1, g_2\}$ values, the *OFM* also provides estimates for the simpler $M_F = \{0, 1\}$ cases, as outlined by Eqs. [4.9a]-[4.9f]. For $M_F = 1$, the best two $\{g'_0, g'_1\}$ values are gotten by only using $\{Q_0, Q_1\}$ and an $\underline{\mathbf{M}}_2$ sub-matrix of $\underline{\mathbf{M}}_3$. For $M_F = 0$, the best $\{g^+_0\}$ by itself is derived by using $\{Q_0\}$ and the $\underline{\mathbf{M}}_1$ sub-matrix. These alternative estimates give:

$$\overrightarrow{g}_2 = \begin{pmatrix} g'_0 \\ g'_1 \end{pmatrix} = \begin{pmatrix} +0.23044 & -0.31357 \\ -0.31357 & +0.58041 \end{pmatrix}^{-1} \begin{pmatrix} +1,169,103 \\ -1,576,445 \end{pmatrix} = \begin{pmatrix} 5,200,870 \\ 93,713 \end{pmatrix}, \qquad [5.14a]$$

$$\overrightarrow{g}_1 = \begin{pmatrix} g^+_0 \end{pmatrix} = \begin{pmatrix} +0.23044 \end{pmatrix}^{-1} \begin{pmatrix} +1169103 \end{pmatrix} = \begin{pmatrix} 5,073,351 \end{pmatrix}. \qquad [5.14b]$$

These additional calculations give the following progression of estimates for $N[t \to \infty]$, which is the final number of CoVID-19 cases at the pandemic end:

$$\begin{pmatrix} N^o_{\max} \\ g^+_o \\ c'_o = g'_0 + g'_1 \\ c_o = g_0 + g_1 + g_2 \end{pmatrix} = \begin{pmatrix} 5,024,900 \\ 5,073,351 \\ 5,294,583 \\ 4,645,874 \end{pmatrix}, \qquad [5.15]$$

These Eq. [5.15] results show that the $N[t \to \infty]$ projected final number of CoVID-19 cases remains fairly stable, even as the number of data fitting

parameters is increased from 0 to 3. The average and 1σ standard deviation among these $N[t \rightarrow \infty]$ projections is:

$$< N_{\max} >= 5,009,677 \ \pm \ 269,450\,, \qquad [5.16]$$

where 1σ is ˜5.4% of the overall average.

Comparing the results among **Figs. 2-5** highlights several items:

(a) All $\rho[t]$ functions have a sharp rise, and a much slower decreasing tail.

(b) The overall fit-to-data, as given in **Fig. 3** and **Fig. 5**, shows that the extra parameters in the *OFM* can fit the $\rho[t]$ shape better.

(c) The *OFM* helps to estimate the uncertainty in the *Initial Model*, which Eq. [5.16] showed was ˜5.4%.

(d) These results, taken together, exhibit only a relatively small change in the $N[t \rightarrow \infty]$ limits. Thus, the *Initial Model* function captures much of the progression to pandemic shutoff.

The $\rho[t]$ tail may still differ from these predictions, due to factors such as:

(i) The CoVID-19 dynamics may change in the long-term low $\rho[t]$ regime;

(ii) A "second wave" or multiple waves of $\rho[t]$ rise and fall may occur; both of which are beyond the scope of this CoVID-19 pandemic modeling;

(iii) Using just an exponential rise at the CoVID-19 pandemic start, plus lengthening *doubling times*, may limit how much mitigation can be easily modeled using only a few adjustable parameters.

Figure 4 provides some evidence for the above (iii) possibility. While lengthening the *doubling time* enables pandemic shutoff in the long time *dilute pandemic* limit; **Figure 4** also shows that this model tends to approach final pandemic shutoff rather slowly.

6 USA Data: The *bing.com* Change

This analysis of the *bing.com* USA data begins at mid-March 2020, when mandatory *Mitigation Measures* were instituted. However, in early-May, *bing.com* changed their entire database, revising all numerical values back to the start of their reporting history.

The revised *bing.com* USA data from mid-March through early-June is analyzed next, which had these values:

$$N'_I[t_I \leftrightarrow 3/21/2020]_{day\#1} = \{23,710\}\,, \qquad [6.1a]$$

$$N'_J[t_J \leftrightarrow 5/03/2020]_{day\#44} = \{1,177,014\}\,, \qquad [6.1b]$$

$$N'_F[t_F \leftrightarrow 6/07/2020]_{day\#79} = \{1,920,628\}\,, \qquad [6.1c]$$

covering $(t'_F - t'_I) = 78$ *days*, as compared to the original *bing.com* data, which was used in above analyses, and only spanned $(t_F - t_I) = 43$ *days*. The Eqs. [6.1a]-[6.1b] revised $\{day\,\#1,\ day\,\#44\}$ values are {˜7.82%, ˜0.56%} lower than the original Eqs. [5.1a]-[5.1b] data.

Applying the *Initial Model* to this revised dataset gives:

$$K'_o = \{0.3471686\,/\,day\}\,, \qquad [6.2a]$$

$$\alpha'_S = \{0.06618\,/\,day\}\,, \qquad [6.2b]$$

$$N[t \rightarrow \infty] \approx N'_I\ \exp[+K'_o\,/\,\alpha'_S] = N^1_{\max} \equiv \{4,499,494\}\,. \qquad [6.2c]$$

Using Eq. [3.3b] gives these t'_I and t'_F results:

$$t'_I \;= \ln(N'_I) \,/\, [K'_o + \alpha'_S \,\ln(N'_I)] = \{9.9361076\, days\}\,, \qquad [6.3a]$$
$$t'_F \;= t_I + \{78\ days\} = \{87.9361076\, days\}\,, \qquad [6.3b]$$

for use in the *OFM*. The Eq. [6.2c] calculated $N^1_{\rm max}$ value is ˜10.456% lower than the prior $N^o_{\rm max}$ of Eq. [5.5]. Since the *Initial Model* uses an *rms* best fit on logarithmic axes for $N[t]$, it emphasizes differences at low $N[t]$ values, where the revised *bing.com* data changes were larger. Thus, some of the ˜10.456% change in $N^1_{\rm max}$ may be due to the revised *bing.com* data, but the longer $(t'_F - t'_I)$ data interval also contributes to modifying the $\{K'_o,\, \alpha'_S\}$ values.

The *Initial Model* datafit for the revised USA data is shown in **Figures 6-7**, and is a better datafit than the *Initial Model* results of **Figures 2-3**. Comparing the *OFM* result of Eq. [5.12], which gave $N[t \to \infty] = \{4,645,874\}$, to the *Initial Model* result of Eq. [6.2c] shows that they differ by just ˜3.25%.

Next, the *OFM* is applied to further refine this *Initial Model* prediction. Those results are shown in **Figure 8** and **Figure 9**, which were derived as follows. First, the Eqs. [2.1a]-[2.1d] time-shift was done:

$$\gamma'_o = \{0.183266685\ /day\}\,, \qquad [6.4a]$$
$$K'_A = \{2.96074425\ /day\}\,, \qquad [6.4b]$$
$$G'_o \equiv [K'_A \,/\, \gamma'_o] = \{15.31947555\}\,, \qquad [6.4c]$$
$$N[t] \approx \mathbf{1}\exp[+K'_A\, t \,/\, (1+\gamma'_o\, t)] = \exp[+G'_o]\,\exp[-Z] = N^1_{\rm max}\,, \qquad [6.4d]$$
$$Z[t] \equiv +[G'_o \,/\, (1+\gamma'_o\, t)] = [\,15.31947555 \,/\, (1 + 0.183266685\ t\,)\,]\,, \qquad [6.4e]$$

for this dataset. Next, using Eqs. [6.3a]-[6.3b] for $\{t'_I,\, t'_F\}$ gives:

$$Z_{\rm min}[t'_F \approx 87.936] = +[G'_o \,/\, (1+\gamma'_o\, t'_F)] = 0.851312775\ , \qquad [6.5a]$$
$$Z_{\rm max}[t'_I \approx 10.686] = +[G'_o \,/\, (1+\gamma'_o\, t'_I)] = 5.245823369\ . \qquad [6.5b]$$

The $\underline{\mathbf{M}}_3$ matrix of $K_{m,n}$ entries, as set by the $\{Z_{\rm min},\, Z_{\rm max}\}$ values, is:

$$\underline{\mathbf{M}}_3 = \begin{pmatrix} K_{0,0} & K_{0,1} & K_{0,2} \\ K_{1,0} & K_{1,1} & K_{1,2} \\ K_{2,0} & K_{2,2} & K_{2,2} \end{pmatrix} = \begin{pmatrix} +0.421584 & -0.335744 & -0.253505 \\ -0.335744 & +0.585931 & +0.077270 \\ -0.253505 & +0.077270 & +0.306047 \end{pmatrix}, \qquad [6.6]$$

and it has an inverse of:

$$(\underline{\mathbf{M}}_3)^{-\mathbf{1}} \;\equiv \begin{pmatrix} 12.309\,835 & 5.905\,615\,4 & 8.705\,429\,3 \\ 5.905\,615\,4 & 4.598\,672\,3 & 3.730\,671\,8 \\ 8.705\,429\,3 & 3.730\,671\,8 & 9.536\,424\,4 \end{pmatrix}. \qquad [6.7]$$

The $\overrightarrow{Q}_3-$vector for this dataset gives this updated $\overrightarrow{g}_3-$vector:

$$(\underline{\mathbf{M}}_3)^{-\mathbf{1}}\ \overrightarrow{Q}_3 \equiv (\underline{\mathbf{M}}_3)^{-\mathbf{1}} \begin{pmatrix} Q_o \\ Q_1 \\ Q_2 \end{pmatrix} = (\underline{\mathbf{M}}_3)^{-\mathbf{1}} \begin{pmatrix} +1,896,161 \\ -1,507,039 \\ -1,156,648 \end{pmatrix} \equiv$$
$$\overrightarrow{g}_3 = \begin{pmatrix} g_o \\ g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} +4,372,319 \\ -47,455 \\ -145,659 \end{pmatrix}. \qquad [6.8]$$

where $c_0 = (g_0 + g_1 + g_2) = (4,179,205) = N[t \to \infty]$ is the new *OFM* predicted total number of CoVID-19 cases, which is down from the *Initial Model* value of $\exp[+K'_A \,/\, \gamma'_o] = N^1_{\rm max} = (4,499,494)$ from Eq. [6.2c]. This ˜7.12% reduction is similar to the ˜7.42% change between Eq. [5.12] and Eq. [5.5]. A similar analysis for $\overrightarrow{g}_2$ and $\overrightarrow{g}_1$, using Eqs. [4.9a]-[4.9f], gives this summary:

$$\begin{pmatrix} N^1_{\max} \\ g^+_o \\ c'_o = g'_0 + g'_1 \\ c_o = g_0 + g_1 + g_2 \end{pmatrix} = \begin{pmatrix} 4,499,494 \\ 4,497,699 \\ 4,514,812 \\ 4,179,205 \end{pmatrix} . \qquad [6.9]$$

The $N[t \to \infty]$ projected final number of CoVID-19 cases in Eq. [6.9] remains fairly stable, even as the number of data fitting parameters is increased from 0 to 3. This result is similar to the Eq. [5.15] analysis of the original *bing.com* data, which spanned only $(t_F - t_I) = 43$ *days*. The average and 1σ standard deviation among these Eq. [6.9] calculations for $N[t \to \infty]$ is:

$$< N_{\max} > = 4,422,803 \ \pm \ 162,580 , \qquad [6.10]$$

where 1σ is ˜3.68% of the overall average. It is somewhat lower than the ˜5.14% value of Eq. [5.16]. Thus, having $(t'_F - t'_I) = 78$ *days* of data for analysis reduces the overall uncertainty in these projections.

Comparing Eq. [6.9] to Eq. [5.15] also shows the following trends. The 1-term calculations, using either $\{N_{\max}\}$ or just $\{g^+_o\}$ by itself, give similar results. The 2-term calculations, using $\{g'_0, g'_1\}$ gives $\lesssim 10\%$ higher results, while using $\{g_0, g_1, g_2\}$ gives $\lesssim 10\%$ lower results. This oscillation around the average value of Eq. [6.10] shows that the *Initial Model* of Eq. [2.1a] and Eq. [6.4d] capture much of how *Social Distancing* enables pandemic shutoff.

Comparing the **Fig. 6** *Initial Model* and the **Fig. 8** *OFM* for the pandemic peak size, timing, and Day 200 values gives :

$$\begin{pmatrix} \text{Parameter} & \textit{Initial Model} & \textit{Orthog. Func. Model} \\ N[t \to \infty] & 4,499,494 & 4,179,205 \\ \max\{\rho[t_p]\} & 30,727 \ / \ day & 30,909 \ / \ day \\ [t_p] \ Date & 4/15/2020 & 4/13/2020 \\ \overset{200}{Day} \rho[t] & 5,783 & 5,140 \ / \ day \end{pmatrix} . \qquad [6.11]$$

The $\rho[t]$ at 200-days nearly scales with $N[t \to \infty]$, while the *OFM* predicts a higher and earlier $\rho[t]$ pandemic peak. Comparing the revised *bing.com* 78-day dataset up through 6/7/2020 of Eq. [6.11], to the original *bing.com* 43-day dataset up through 5/3/2020 of Eq. [5.13] gives:

$$\begin{pmatrix} \textbf{6/7/2020 vs. 5/3/2020} & \textit{Initial Model} & \textit{Orthog. Func. Model} \\ N[t \to \infty] & -8.95\,\% & -9.00\,\% \\ \max\{\rho[t_p]\} & -9.67\,\% & -9.64\,\% \\ [t_p] \ Date & -2 \ days & -2 \ days \\ \overset{200}{Day} \rho[t] & -8.56\,\% & -8.62\,\% \end{pmatrix} . \qquad [6.12]$$

Both the *Initial Model* and the *OFM* found a comparable amount of change between the two datasets; likely due to the revised *bing.com* values being lower, along with the larger dataset enabling increased modeling precision.

The *Initial Model* and the *OFM* also provide self-consistent CoVID-19 predictions over the two different time periods. Each model held its predictive power to within $< 10\%$ for over a month {43 *days vs.* 78 *days*}, without needing recalculations or parameter value changes, which provides a strong data-driven validation of the potential utility of these models. When the *Initial Model* is a somewhat good fit, this *Orthogonal Function Model* provides even better fits.

7 Italy: Revised *bing.com* Data Analysis

This Italy analysis uses data beginning on Feb. 23, 2020, from the revised *bing.com* CoVID-19 database[9], which has these values:

$$N_I[t_I \leftrightarrow 2/23/2020]_{day\#1} = \{150\}\,, \tag{7.1a}$$
$$N_F[t_F \leftrightarrow 6/15/2020]_{day\#114} = \{237, 290\}\,, \tag{7.1b}$$

with $(t_F - t_I) = 113\ days$. The number of daily new CoVID-19 cases shows a sharp post-peak decrease for Italy, in contrast the the above USA data. That sharp decrease provides a near-worst case test for the *OFM*. The *Initial Model* best fit on a logarithmic Y-axis, gives these initial parameters:

$$K_o = \{0.665772\,/\,day\}\,, \tag{7.2a}$$
$$\alpha_S = \{0.08153\,/\,day\}\,. \tag{7.2b}$$

Using Eq. [3.3b] for t_I and t_F gives:

$$N[t \to \infty] \approx N_I\,\exp[+K_o\,/\,\alpha_S] = N^o_{\rm max} \equiv \{527, 875\}\,, \tag{7.3a}$$
$$t_I\ = \ln(N_I)\,/\,[K_o + \alpha_S\,\ln(N_I)] = \{4.66414\,days\}\,, \tag{7.3b}$$
$$t_F\ = t_I + \{113\ days\} = \{117.66414\,days\}\,, \tag{7.3c}$$

for use in the *OFM*. The revised *bing.com* Italy data and the *Initial Model* datafit are shown in **Figure 10** and its inset. The *Initial Model* is not a good fit due to the high curvature of the data on the logarithmic Y-axis, which is similar to our previous[2] results for Italy. The *OFM* is applied next.

Using Eqs. [3.5a]-[3.5b] sets these $\{\gamma_o,\ K_A,\ G_0\}$ values:

$$\gamma_o = \{0.1316771\ /day\}\,, \tag{7.4a}$$
$$K_A = \{1.734075\ /day\}\,, \tag{7.4b}$$
$$G_o \equiv [K_A\,/\,\gamma_o] = \{13.1691513\}\,, \tag{7.4c}$$

where Eq. [3.1c] also gives:

$$\underset{t\to\infty}{Lim}\{\mathbf{1}\exp[\tfrac{+K_A\,t}{(1+\gamma_o\,t)}]\} = \mathbf{1}\exp[+\tfrac{K_A}{\gamma_o}] = N^o_{\rm max} \equiv \{527, 875\}\,, \tag{7.5}$$

which matches Eq. [7.3a], as it should. Then:

$$Z[t] = +[G_o\,/\,(1+\gamma_o\,t)] = [13.1691513\,/\,(1+0.1316771\ t)\,, \tag{7.6}$$

defines Z for the *OFM*, where:

$$Z_{\rm min}[t_F \approx 117.664] = +[G_o\,/\,(1+\gamma_o\,t_F)] = 0.7995757039\ , \tag{7.7a}$$
$$Z_{\rm max}[t_I \approx 4.664] = +[G_o\,/\,(1+\gamma_o\,t_I)] = 8.1659787106\ . \tag{7.7b}$$

The resultant symmetric matrix $\underline{\mathbf{M}}_3$ of $K_{m,n}$ entries is:

$$\underline{\mathbf{M}}_3 = \begin{pmatrix} K_{0,0} & K_{0,1} & K_{0,2} \\ K_{1,0} & K_{1,1} & K_{1,2} \\ K_{2,0} & K_{2,2} & K_{2,2} \end{pmatrix} = \begin{pmatrix} +.4492355 & -.3571046 & -.2228851 \\ -.3571046 & +.7176744 & -.1261928 \\ -.2228851 & -.1261928 & +.6411040 \end{pmatrix} \tag{7.8}$$

It has an $(\underline{\mathbf{M}}_3)$ inverse of:

$$(\underline{\mathbf{M}}_3)^{-\mathbf{1}}\ \equiv \begin{pmatrix} 7.\,159\,042\,3 & 4.\,143\,277\,7 & 3.\,304\,449\,2 \\ 4.\,143\,277\,7 & 3.\,841\,256\,5 & 2.\,196\,544\,9 \\ 3.\,304\,449\,2 & 2.\,196\,544\,9 & 3.\,140\,988\,9 \end{pmatrix}\ . \tag{7.9}$$

A convolution of $L_m(Z)$ functions with the measured data sets $\overrightarrow{Q}_3$ using Eqs. [4.11a]-[4.11b], with $(\underline{\mathbf{M}}_3)^{-\mathbf{1}}$ of Eq. [4.5c] giving this final $\overrightarrow{g}-$vector:

$$(\underline{\mathbf{M}}_3)^{-\mathbf{1}}\ \overrightarrow{Q}_3 \equiv (\underline{\mathbf{M}}_3)^{-\mathbf{1}} \begin{pmatrix} Q_o \\ Q_1 \\ Q_2 \end{pmatrix} = (\underline{\mathbf{M}}_3)^{-\mathbf{1}} \begin{pmatrix} +298,131 \\ -206,004 \\ -190,856 \end{pmatrix}\ \equiv$$

$$\overrightarrow{g}_{\,3} = \begin{pmatrix} g_o \\ g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} 650,127 \\ 24,702 \\ -66,815 \end{pmatrix} . \qquad [7.10]$$

The coefficients for $R(Z)$, which set the predicted number of daily new CoVID-19 cases for the *OFM*, are given by:

$$\overrightarrow{C}_{\,3} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \overrightarrow{g}_{\,3} = \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} +608,013 \\ -42,113 \\ -66,815 \end{pmatrix} . \qquad [7.11]$$

Using these $\{g_0,\, g_1,\, g_2\}$ values along with Eq. [2.11] gives:

$$N(0) \equiv N[t \to \infty] = c_0 \equiv \{608,013\} , \qquad [7.12]$$

as a new predicted total number of CoVID-19 cases at the pandemic end. It is a ˜15.18% or 80,138 increase in number of cases, compared to the *Initial Model* $N^o_{\rm max}$ value of Eq. [7.3a].

A graph of $\rho[t]$ for the predicted number of daily new CoVID-19 cases is shown in **Figure 11**, using Eqs. [2.4b] and [2.5b]. For this fast pandemic shutoff case, the *OFM* improvement over the *Initial Model* is not large. When the initial $[\exp(-Z)]$ function is not a good fit, which is likely for quicker pandemic shutoffs, a lot of terms, beyond the $M_F = 2$ value used here, are needed in Eq. [2.5a] for a good fit. An alternative method for choosing the initial $[\exp(-Z)]$ function is examined next, to see if additional improvements result for that case.

8 Italy: An Alternative Starting Function

There is a wide latitude in the choice of an initial $[\exp(-Z)]$ function for the Eqs. [2.5a]-[2.5b] orthogonal function expansions. However, when the *Initial Model* is not a good fit, the common practice of minimizing *rms* error using a logarithmic Y-axis for the *Initial Model* may not be optimal, since the *Orthogonal Function Model [OFM]* creates best fits using a linear Y-axis.

Minimizing the *rms* error between the *Initial Model* and data using a linear Y-axis is done to provide an alternative $[\exp(-Z)]$ function. This alternative starting point gives these parameter values, replacing Eqs. [7.2b]-[7.2c]:

$$K^L_o = \{1.1863559\,/\,day\} , \qquad [8.1a]$$
$$\alpha^L_S = \{0.15220\,/\,day\} . \qquad [8.1b]$$

The $N(t_I) = N_I$ and $N(t_F) = N_F$ values are still needed to properly set the above $\{K^L_A,\, \gamma^L_o\}$ values for $Z[t]$. Using Eq. [3.3b] for t_I and t_F gives:

$$N[t \to \infty] \approx N_I \, \exp[+K^L_o \,/\, \alpha^L_S] = N^L_{\rm max} \equiv \{364,161\} , \qquad [8.2a]$$
$$t^L_I \;= \ln(N_I) \,/\, [K^L_o + \alpha^L_S \, \ln(N_I)] = \{2.570908\ days\} , \qquad [8.2b]$$
$$t^L_F \;= t^L_I + \{113\ days\} = \{115.570908\ days\} , \qquad [8.2c]$$

for use in the *OFM*, while still using this linear Y-axis initial fit. **Figure 12** and its inset show how this alternative *Initial Model* compares to the Italy CoVID-19 data. Using Eqs. [3.5a]-[3.5b] sets these new $\{\gamma^L_o,\, K^L_A,\, G^L_0\}$ values:

$$\gamma^L_o = \{0.2500379\ /day\} , \qquad [8.3a]$$
$$K^L_A = \{3.2018233\ /day\} , \qquad [8.3b]$$
$$G^L_o \equiv [K^L_A \,/\, \gamma^L_o] = \{12.8053524\} , \qquad [8.3c]$$

where Eq. [3.1c] also gives:

$$\underset{t\to\infty}{Lim}\{\mathbf{1}\exp[\tfrac{+K_A^L\,t}{(1+\gamma_o^L\,t)}]\} = \mathbf{1}\exp[+\tfrac{K_A^L}{\gamma_o^L}] = N_{\max}^L \equiv \{364, 161\}\,, \qquad [8.4]$$

which matches Eq. [8.2a], as it should. Then:

$$Z^L[t] = +[G_o^L\,/\,(1+\gamma_o^L\,t)] = [12.8053524\,/\,(1+0.2500379\;t\,)\,, \qquad [8.5]$$

defines Z^L for this alternative fit analysis, where:

$$Z_{\min}^L[t_F^L \approx 115.5709] = +[G_o^L\,/\,(1+\gamma_o^L\,t_F^L)] = 0.428314107\ , \qquad [8.6a]$$

$$Z_{\max}^L[t_I^L \approx 2.5709] = +[G_o^L\,/\,(1+\gamma_o^L\,t_I^L)] = 7.79471714\ . \qquad [8.6b]$$

The resultant symmetric matrix $\underline{\mathbf{M}}_3$ of $K_{m,n}$ entries is:

$$\underline{\mathbf{M}}_3 = \begin{pmatrix} K_{0,0} & K_{0,1} & K_{0,2} \\ K_{1,0} & K_{1,1} & K_{1,2} \\ K_{2,0} & K_{2,2} & K_{2,2} \end{pmatrix} = \begin{pmatrix} +.6511948 & -.2758817 & -.2286253 \\ -.2758817 & +.7457076 & -.1094322 \\ -.2286253 & -.1094322 & +.6094403 \end{pmatrix}. \qquad [8.7]$$

It has an $(\underline{\mathbf{M}}_3)$ inverse of:

$$(\underline{\mathbf{M}}_3)^{-\mathbf{1}} \equiv \begin{pmatrix} 2.341\,468\,6 & 1.022\,082\,7 & 1.061\,904\,9 \\ 1.022\,082\,7 & 1.823\,453\,8 & 0.710\,846\,59 \\ 1.061\,904\,9 & 0.710\,846\,59 & 2.166\,853\,5 \end{pmatrix}. \qquad [8.8]$$

A convolution of $L_m(Z^L)$ functions with the measured data sets $\overrightarrow{Q}_3$ using Eqs. [4.11a]-[4.11b], with $(\underline{\mathbf{M}}_3)^{-\mathbf{1}}$ of Eq. [4.5c] giving this final $\overrightarrow{g}-$vector:

$$(\underline{\mathbf{M}}_3)^{-\mathbf{1}}\ \overrightarrow{Q}_3 \equiv (\underline{\mathbf{M}}_3)^{-\mathbf{1}} \begin{pmatrix} Q_o \\ Q_1 \\ Q_2 \end{pmatrix} = (\underline{\mathbf{M}}_3)^{-\mathbf{1}} \begin{pmatrix} +188,281 \\ -10,274 \\ -61,526 \end{pmatrix} \equiv$$

$$\overrightarrow{g}_3 = \begin{pmatrix} g_o \\ g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} 365,018 \\ 129,969 \\ 59,315 \end{pmatrix}. \qquad [8.9]$$

The coefficients for $R(Z)$, which set the predicted number of daily new CoVID-19 cases for the *OFM*, are given by:

$$\overrightarrow{C}_3 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \overrightarrow{g}_3 = \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} +554,303 \\ 189,284 \\ 59,315 \end{pmatrix}. \qquad [8.10]$$

Using these $\{g_0,\, g_1,\, g_2\}$ values along with Eq. [2.11] gives:

$$N(0) \equiv N[t\to\infty] = c_0 \equiv \{554,303\}\,, \qquad [8.11]$$

as a new predicted total number of CoVID-19 cases at the pandemic end. It is a ˜5.00% or 26,428 increase in number of cases, compared to the *Initial Model* $N_{\max}^o$ value of Eq. [7.3a]. A graph of $\rho[t]$ for the predicted number of daily new CoVID-19 cases is shown in **Figure 13**, using Eqs. [2.4b] and [2.5b]. A tabulated summary for all of these Italy calculations is:

Parameter	*Initial Model*	*Orthog. Func.*	*Init.Model Re-do*	*Orthog.Func. Re-do*
$N[t\to\infty]$	527,875	608,013	364,161	554,303
$\max\{\rho[t_p]\}$	2,860 / *day*	3,886 / *day*	3,848 / *day*	4,073 / *day*
$[t_p]$ *Date*	4/2/2020	3/29/2020	3/14/2020	3/26/2020

.

The *Initial Model* shapes for $\rho[t]$ were very different, depending on whether that initial datafit was performed by minimizing *rms* error using a logarithmic Y-axis (**Figure 10**) or a linear Y-axis (**Figure 12**, *Initial Model Re-do*) as expected. However, comparing the two *OFM* (**Figure 11** vs **Figure 13**) calculations, shows that their overall $\rho[t]$ shapes are quite similar.

While the max$\{\rho[t_p]\}$ calculated pandemic peaks generally increase, they are all below the data near-peak values of ~$4,800 - 6,500$ cases/day shown in **Figs. 10-13**. Thus, for quick pandemic shutoffs, the *Initial Model* [$\exp(-Z)$] function is less important than needing more M_F terms. When the *Initial Model* is not a good fit, the *OFM* only gives limited improvements for $M_F = 2$.

9 Summary and Conclusions

The early stages of the CoVID-19 coronavirus pandemic began with a nearly exponential rise in the number of infections with time. Let $N[t]$ be the total number of CoVID-19 cases vs time. Our *Initial Model*[2] used this basic function:

$$N[t] \approx \mathbf{1} \exp[+K_A\, t\,/\,(1+\gamma_o\, t)] = \exp[+G_o]\, \exp[-Z]\,, \qquad [9.1a]$$

$$Z[t] \equiv +[G_o\,/\,(1+\gamma_o\, t)]\,,\; G_o \equiv [K_A\,/\,\gamma_o]\,, \qquad [9.1b]$$

to model *Social Distancing* effects by progressively lengthening the *doubling time* for the pandemic growth. The $\gamma_o = 0$ limit of Eq. [9.1a] corresponds to a purely exponential rise. This *Initial Model* enables calculation of a pandemic shutoff with only a small fraction of the total population becoming infected (*"dilute pandemic"*).

To allow more data fitting parameters than just $\{K_A,\, \gamma_o\}$, an *Orthogonal Function Model [OFM]* was developed, using these orthogonal function series:

$$N(Z) = \sum_{m=0}^{m=M_F} g_m\, L_m(Z)\, \exp[-Z]\,, \qquad [9.2a]$$

$$R(Z) = \sum_{m=0}^{m=M_F} c_m\, L_m(Z)\, \exp[-Z]\,, \qquad [9.2b]$$

$$c_{M_F-k} = \sum_{m=0}^{m=k} g_m\,, \qquad [9.2c]$$

where $N[t] = N(Z[t])$. The $\{g_m;\, m = (0, +M_F)\}$ are a set of constants that are determined from each dataset. Using $\exp[-Z]$ as a weighting function, with $L_m(Z)$ as an orthonormal function set on the $Z = \{0^+, \infty^-\}$ interval, the choice of $L_m(Z)$ becomes unique. They are the *Laguerre Polynomials*, with several important properties given in Eqs. [2.6a]-[2.7e].

The expected number of daily new CoVID-19 cases, $\rho[t]$, is given by:

$$N[t] \equiv \int_{t'=(-1/\gamma_o)}^{t'=t} \rho[t']\, dt'\,, \qquad [9.3a]$$

$$\rho[t] \equiv (\gamma_o\,/\,G_o)\, Z^2\, R(Z)\,. \qquad [9.3b]$$

For a wide range of $N(Z)$ data, larger M_F and more $\{L_m(Z);\, m = (0, +M_F)\}$ terms gives progressively better matches to almost any *arbitrary function*, enabling improved data fitting for a variety of $N[t]$ and $\rho[t]$ shapes.

Methods are developed here to derive $\{K_A,\, \gamma_o\}$, and determine the $\{g_m;\, m = (0, +M_F)\}$ and $\{c_m;\, m = (0, +M_F)\}$ constants from any given $N[t]$ dataset. Whereas our *Initial Model* was an $M_F = 0$ case, the $M_F = 2$ case was used here for data analysis, as an *OFM* example.

These methods were applied to the CoVID-19 pandemic data for the USA. Analysis results using the original *bing.com* up data to ~5/3/2020 are given in

Figures 2-5 and Eq. [5.13]. During early-May, *bing.com* revised their entire database, all the way back to their earliest values. This revised USA *bing.com* data, which included an extended time period into June 2020, was also analyzed, with results given in **Figures 6-9** and Eq. [6.11].

For the USA, the *Initial Model* and *OFM* results differed by only ˜10%, showing that the *Initial Model* was a somewhat good fit, while the *OFM* is a better fit. Comparing our calculations using the 43-day 5/3/2020 original *bing.com* dataset to the 78-day 6/7/2020 revised *bing.com* dataset, showed that our early-May USA projections predicted the June data to within $\lesssim 10\%$ for the same model. Thus, both models provided self-consistent CoVID-19 projections, holding their predictive power for over a month $\{43\ days\ vs.\ 78\ days\}$, without recalculations or parameter value changes.

The Italy CoVID-19 pandemic data was studied next, as a worst-case test of the *OFM*. The post-May 2020 revised *bing.com* database was used, with results presented in **Figures 10-13**. Italy had a much sharper CoVID-19 pandemic shutoff for $\rho[t]$ compared to the USA. While the *OFM* can give substantial improvements, here $M_F = 2$ does not provide enough extra parameters, to convert an *Initial Model* result that was not a good fit, into a substantially better fit. A larger M_F and additional orthogonal function terms are needed.

Even then, the long-term tail can be inaccurate, since both the *Initial Model* and the *OFM* extension have natural $\rho[t]$ asymptotic limits of $\rho[t]\tilde{}[1/t^2]$. Larger M_F values could allow multiple terms to cancel, but a polynomial-like tail of $\rho[t]\tilde{}[1/t^P]$, with $P \geq 2$, would likely remain, making it difficult to estimate the functional form of the CoVID-19 tail for quick pandemic shutoffs.

Overall, both the *Initial Model* and this *Orthogonal Function Model* show how progressively lengthening the pandemic *doubling time* enables CoVID-19 pandemic shutoff, even in the *dilute pandemic* limit. However, there may a natural limit to how fast this one mitigation factor can achieve pandemic shutoff. For cases like Italy, other *Social Distancing* factors may be operating that enable and enhance quick CoVID-19 pandemic shutoff, which are **not** effectively being modeled by just lengthening the pandemic *doubling times*.

10 List of Figures

Figure 1: Comparison of IHME CoVID-19 Projections, 29 April 2020 vs 4 May 2020. CDC CoVID-19 Website highlighted IHME Projections prior to the IHME May 2020 update.

Figure 2: *Initial Model* for USA CoVID-19 Projections using data up to 5/3/2020. Predicted Number of Daily CoVID-19 Cases has a peak of 31,760 cases/day on 4/17; with 5,024,900 cases total; and ˜6,757 new cases/day at Day 200 on 9/26/2020.

Figure 3: *Initial Model* for USA CoVID-19 Projections vs data up to 5/3/2020. Original *bing.com* data up to 5/3/2020 are shown, prior to their *new reporting method*. Data starts slightly below, then goes slightly above the *Initial Model* prediction line.

Figure 4: *Orthogonal Function Model*, USA CoVID-19 Projections, data to 5/3/2020. Predicted Number of Daily CoVID-19 Cases has a peak of 32,069 cases/day on 4/15; with 4,645,874 cases total; and ~5,962 new cases/day at Day 200 on 9/26/2020.

Figure 5: *Orthogonal Function Model*, USA CoVID-19 Projections, data to 5/3/2020. Original *bing.com* data up to 5/3/2020 are shown, prior to their *new reporting method*. *Orthogonal Function Model* matches data a bit better than the *Initial Model*.

Figure 6: *Initial Model* for USA CoVID-19 Projections vs data up to 6/7/2020. Predicted Number of Daily CoVID-19 Cases has a peak of 30,727 cases/day on 4/15; with 4,499,494 cases total; and ~5,783 new cases/day at Day 200 on 9/27/2020.

Figure 7: *Initial Model* for USA CoVID-19 Projections vs data up to 6/7/2020. Revised *bing.com* data, circa 5/3/2020, changed all values back to the pandemic start. *Initial Model* appears to be a good datafit by itself.

Figure 8: *Orthogonal Function Model*, USA CoVID-19 Projections, data to 6/7/2020. Revised *bing.com* data; daily# of CoVID-19 Cases Peak at 30,909 cases/day on 4/13/2020; with 4,179,205 cases total; and ~5,140 new cases/day at Day 200 on 9/27/2020.

Figure 9: *Orthogonal Function Model*, USA CoVID-19 Projections, data to 6/7/2020. Revised *bing.com* data, posted circa 5/3/2020, changed values back to the pandemic start. *Orthogonal Function Model* matches the data a bit better than the *Initial Model*.

Figure 10: *Initial Model* for Italy CoVID-19 Projections vs data up to 6/15/2020. *Initial Model* matches Total Number of Cases at data start and data end, but best fit using a logarithmic Y-axis does not give a good fit for Predicted Number of Daily CoVID-19 cases.

Figure 11: *Orthogonal Function Model* for Italy CoVID-19 data up to 6/15/2020. *Orthogonal Function Model* gives improved datafit, but 3-terms in orthogonal function series is insufficient to accurately predict a rapidly decreasing Number of Daily CoVID-19 cases.

Figure 12: *Initial Model* re-do, Italy CoVID-19 data to 6/15/2020. New starting point is a best fit function on a linear Y-axis, instead of having a best fit using a logarithmic Y-axis. Alternative method may allow a few-term orthogonal function series to better match the data.

Figure 13: *Orthogonal Function Model* re-do, Italy CoVID-19 data to 6/15/2020. *Orthogonal Function Model* re-do gives a slightly better small-series fit. Other *Social Distancing* impacts likely exist besides just lengthening pandemic *doubling times*.

11 References

1. https://www.MedRxiv.org/ ID=MedRxiv/2020/043752v1, 03.25.2020, "*Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months*",

IHME COVID-19 Health Service Utilization Forecasting Team.

2. https://www.MedRxiv.org/content/10.1101/2020.05.04.20091207v1, https://doi.org/10.1101/2020.05.04.20091207, "*Initial Model for the Impact of Social Distancing on CoVID-19 Spread*", Genghmun Eng

3. https:// www.geekwire.com / 2020 / univ-washington--epidemiologists-predict-80000-covid-19-deaths-u-s-july, "*Univ. of Washington researchers predict 80,000 COVID-19 deaths in U.S. by July*", Alan Boyle, *GeekWire*, March 26, 2020.

4. https://www.yahoo.com/finance/news/coronavirus-modelers-raise-projected-u-041641553.html, "*Coronavirus modelers raise projected U.S. death toll and lengthen state-by-state recovery timeline*", Alan Boyle, *GeekWire*, April 27, 2020.

5. https://finance.yahoo.com/news/ pandemic-projection-puts-u-death-220824741.html, "*New pandemic projection puts U.S. death toll at nearly 135,000, due to less social distancing*", Alan Boyle, *GeekWire*, May 4, 2020.

6. http://www.healthdata.org/covid/updates "*COVID-19: What's New for May 4, 2020: Updated IHME COVID-19 projections: Predicting the Next Phase of the Epidemic*", IHME COVID-19 Health Service Utilization Forecasting Team.

7. G. N. Watson, "*A Note of the Polynomials of Hermite and Laguerre*", Journal of the London Mathematical Society, **13**(1938), pp. 29-32.

8. J. Gillis and G. Weiss, "*Products of Laguerre Polynomials*", Math. Comput., **14**(69), Jan. 1960, pp. 60-63

9. www.bing.com/covid: '*Bing COVID-19 Tracker*', and https://www.bing.com/covid?form=CPVD07.

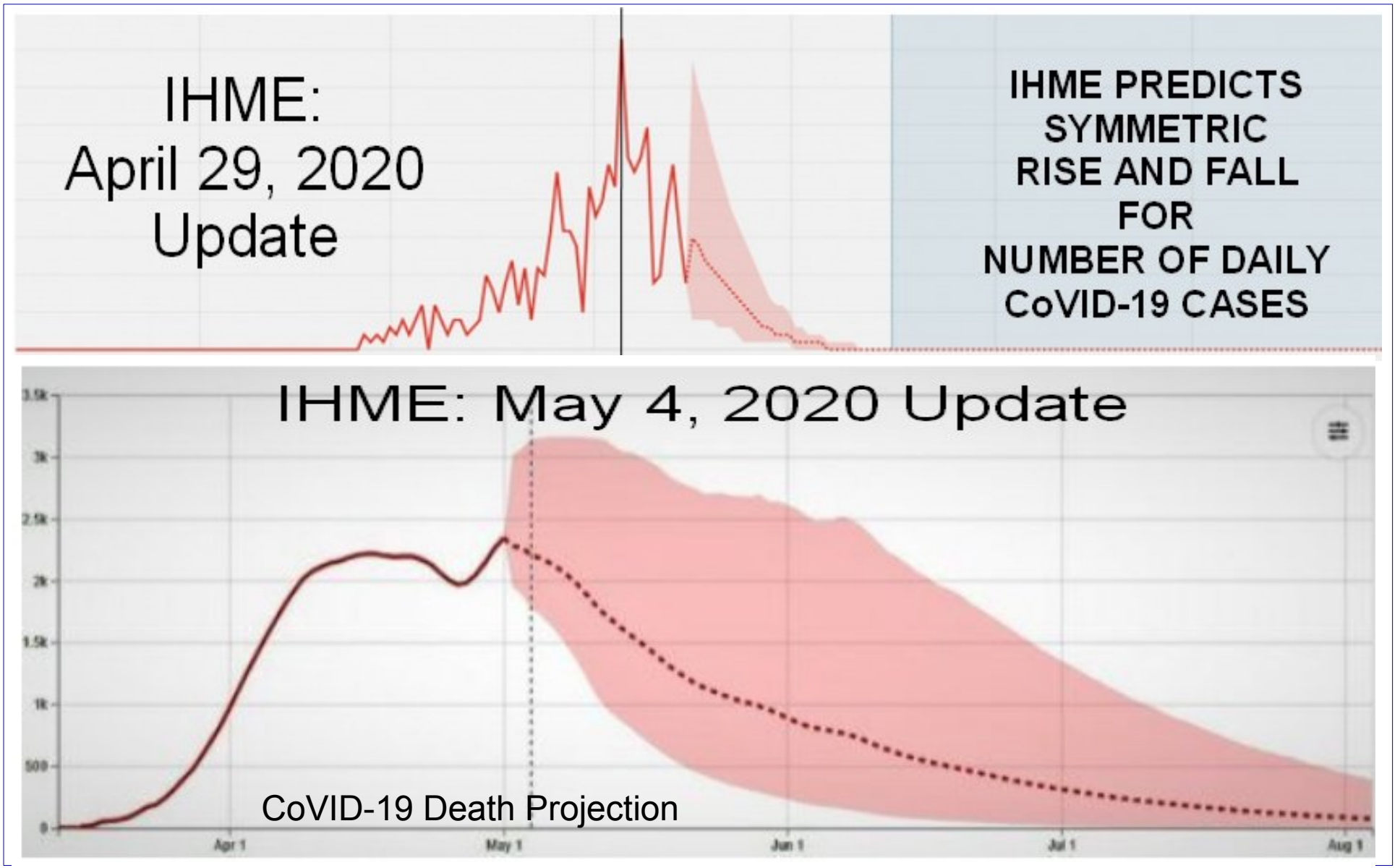

Figure 1: Comparison of IHME CoVID-19 Projections, 29 April 2020 vs 4 May 2020. CDC CoVID-19 Website highlighted IHME Projections prior to the IHME May 2020 update.

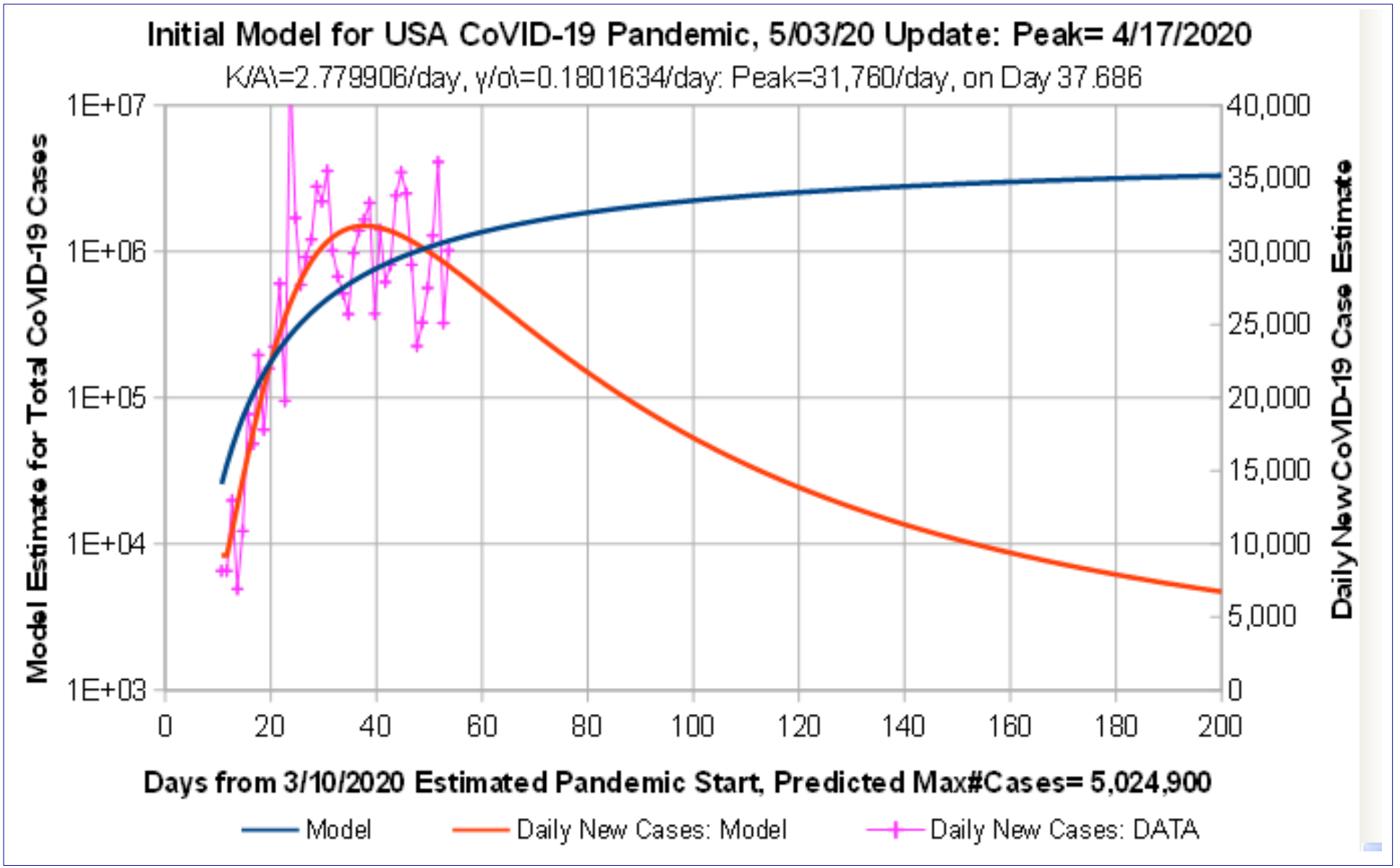

Figure 2: *Initial Model* for USA CoVID-19 Projections using data up to 5/3/2020. Predicted Number of Daily CoVID-19 Cases has a peak of 31,760 cases/day on 4/17; with 5,024,900 cases total; and ~6,757 new cases/day at Day 200 on 9/26/2020.

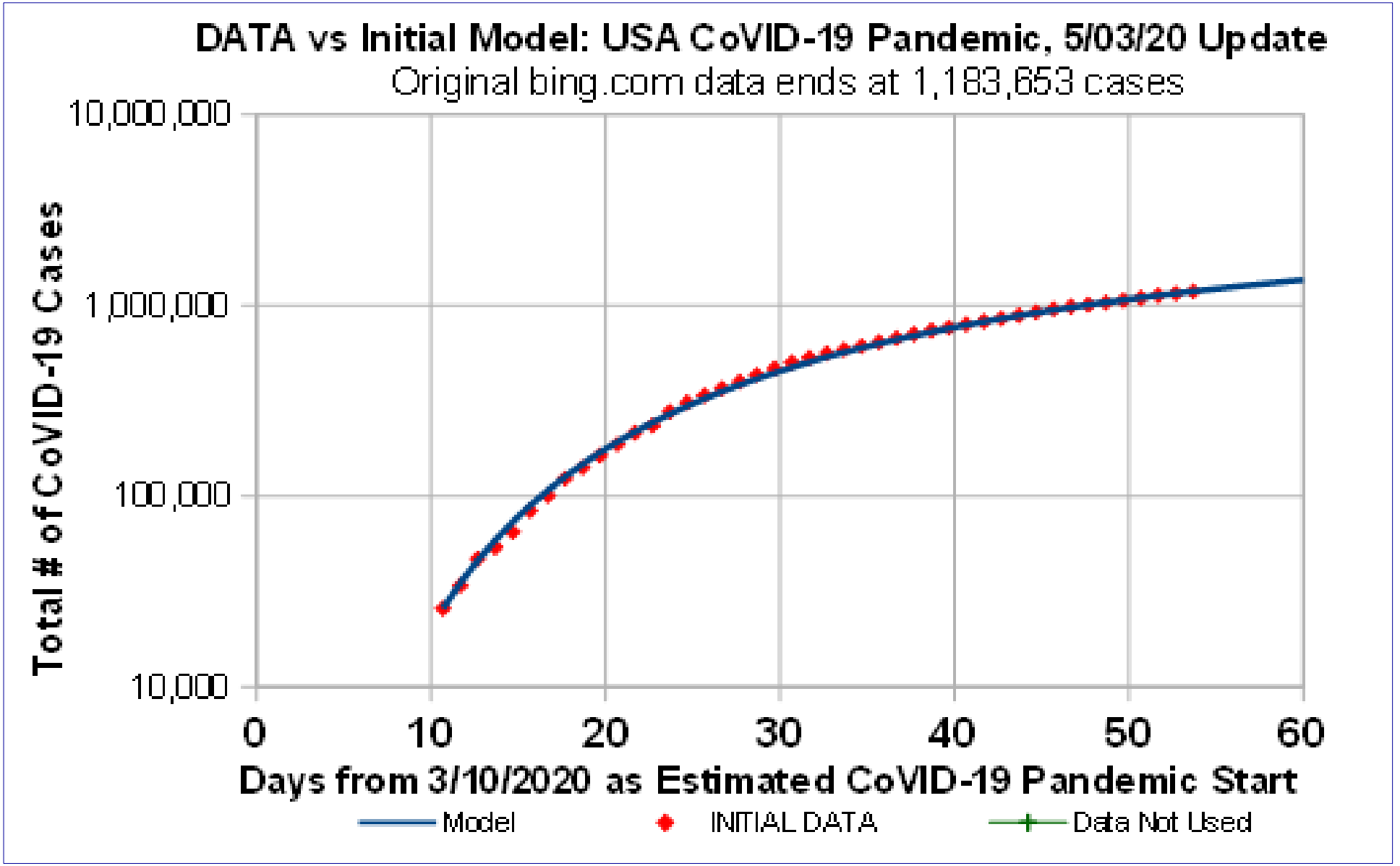

Figure 3: *Initial Model* for USA CoVID-19 Projections vs data up to 5/3/2020. Original *bing.com* data up to 5/3/2020 are shown, prior to their new reporting method. Data starts slightly below, then goes slightly above the Initial Model prediction line.

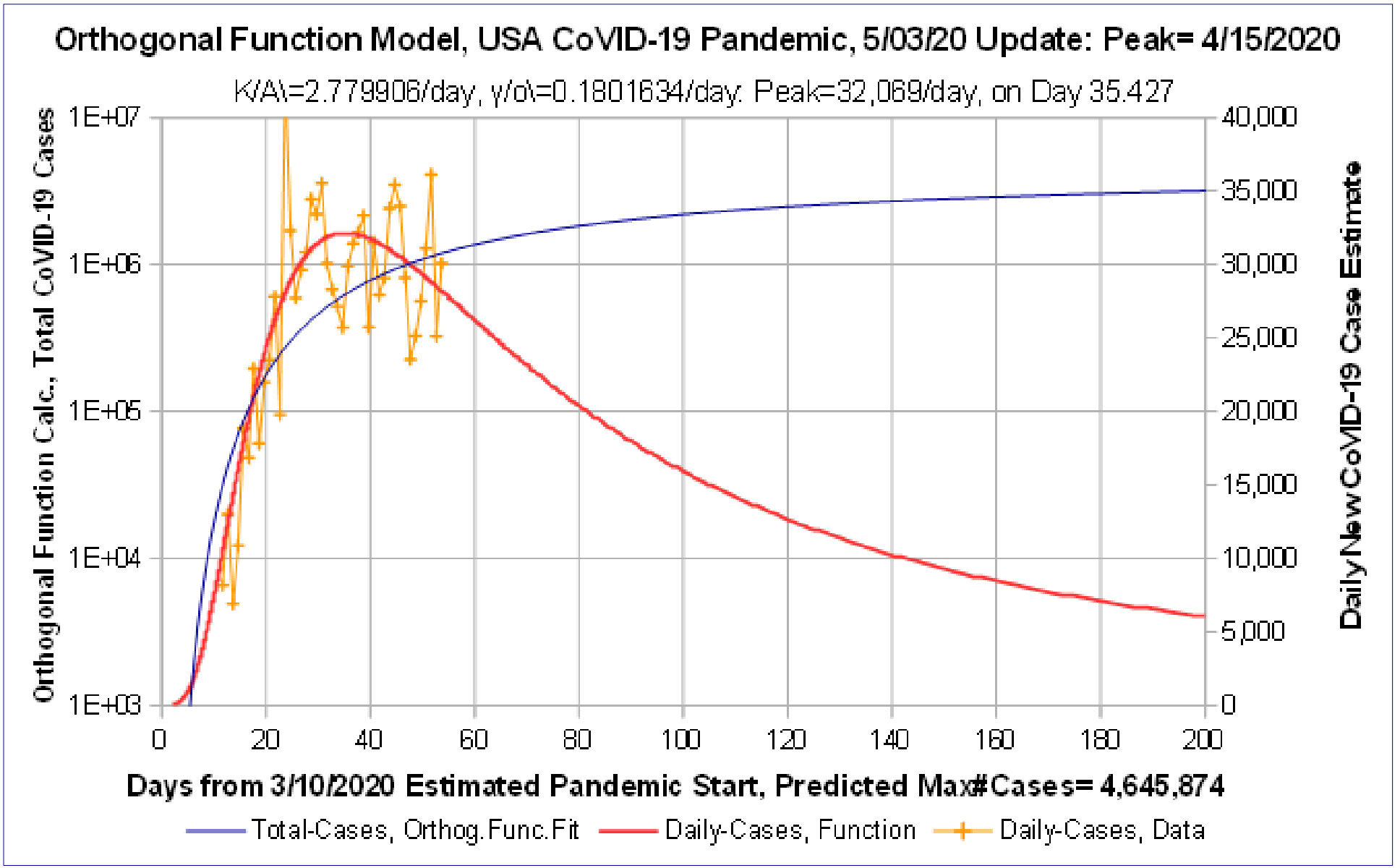

Figure 4: *Orthogonal Function Model*, USA CoVID-19 Projections, data to 5/3/2020. Predicted Number of Daily CoVID-19 Cases has a peak of 32,069 cases/day on 4/15; with 4,645,874 cases total; and ~5,962 new cases/day at Day 200 on 9/26/2020.

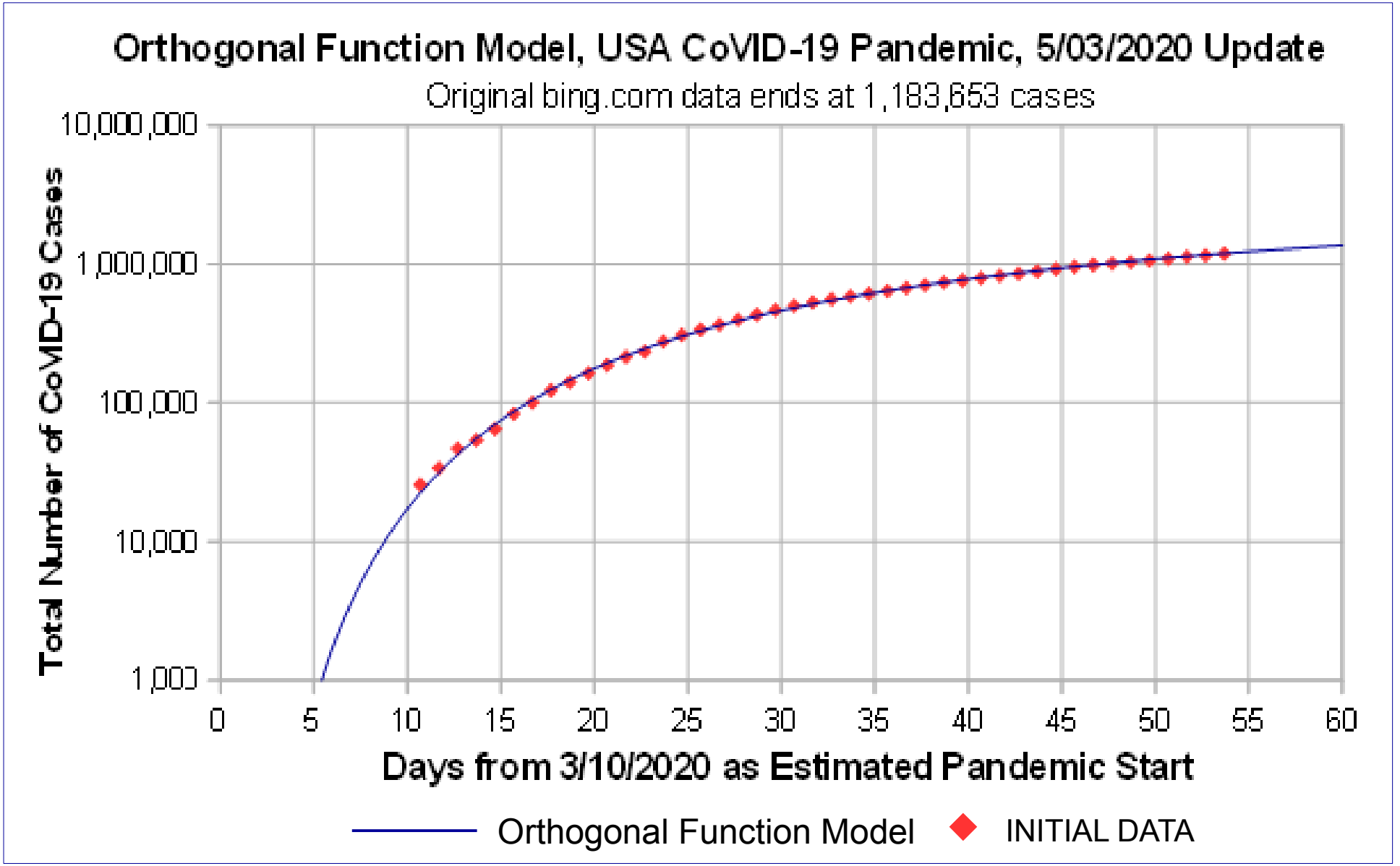

Figure 5: *Orthogonal Function Model*, USA CoVID-19 Projections, data to 5/3/2020. Original *bing.com* data up to 5/3/2020 are shown, prior to their new reporting method. *Orthogonal Function Model* matches data a bit better than the *Initial Model*.

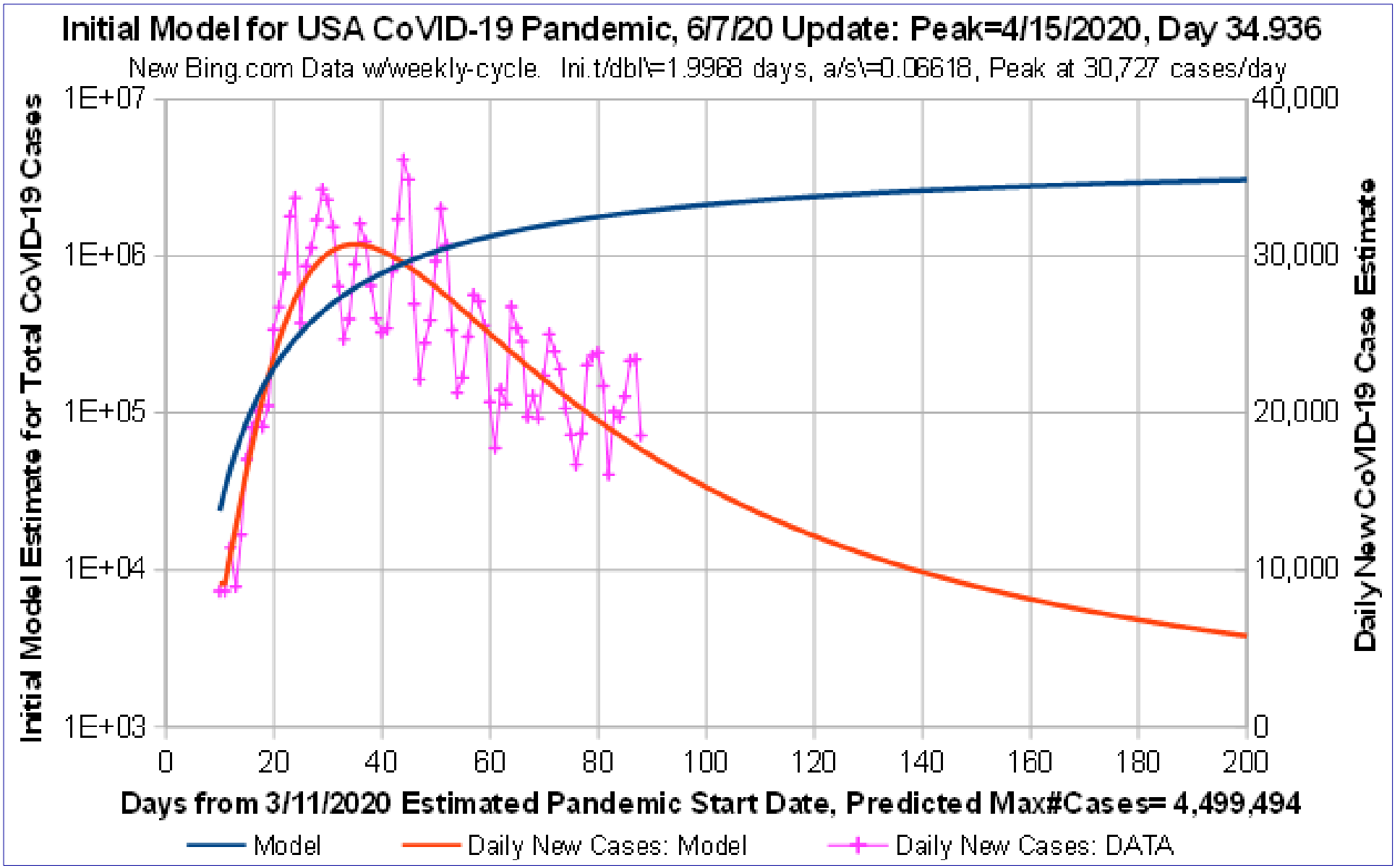

Figure 6: *Initial Model* for USA CoVID-19 Projections vs data up to 6/7/2020.
Predicted Number of Daily CoVID-19 Cases has a peak of 30,727 cases/day on 4/15; with 4,499,494 cases total; and ~5,783 new cases/day at Day 200 on 9/27/2020.

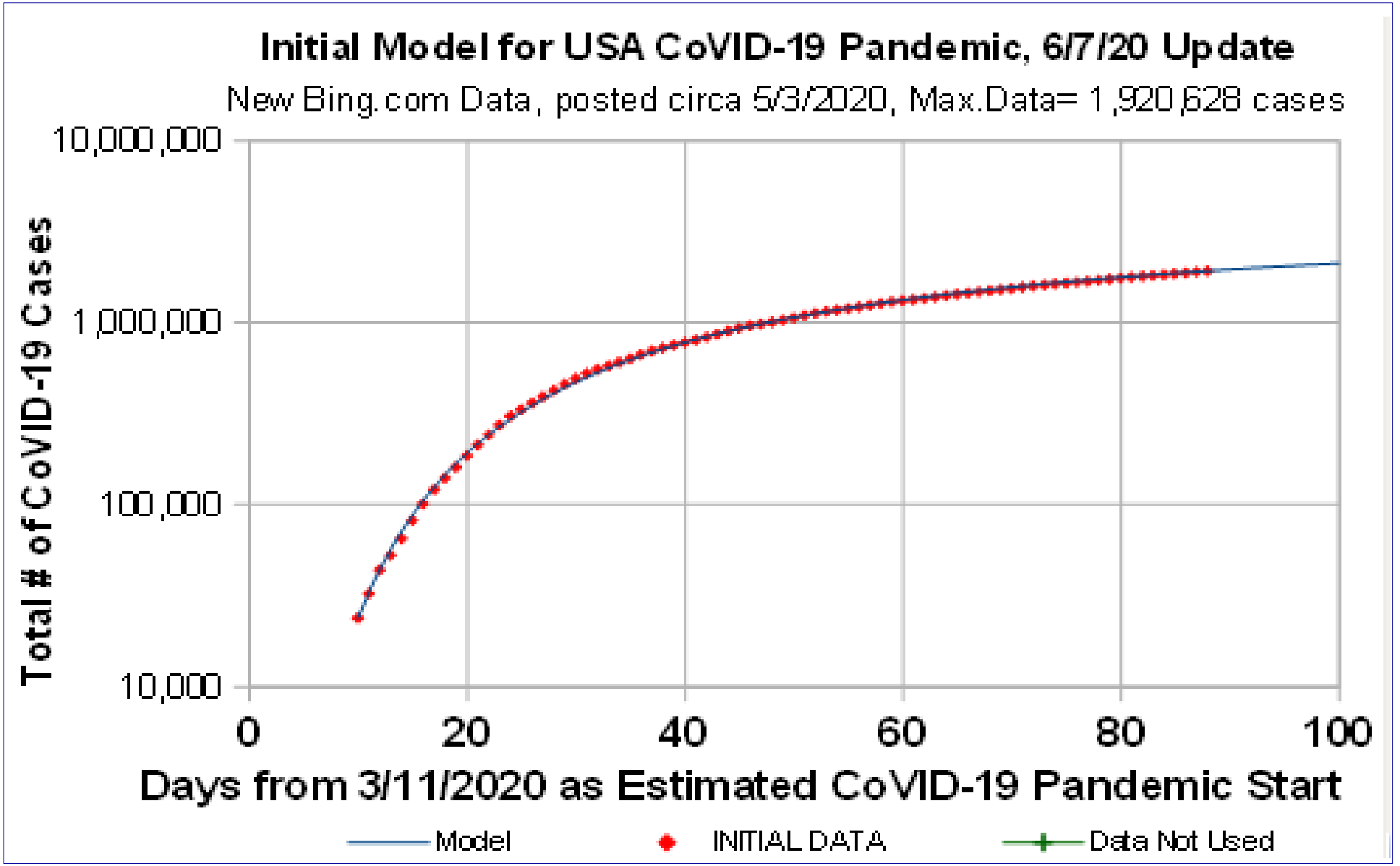

Figure 7: ***Initial Model*** **for USA CoVID-19 Projections vs data up to 6/7/2020.** Revised *bing.com* data, circa 5/3/2020, changed all values back to the pandemic start. *Initial Model* appears to be a good datafit by itself.

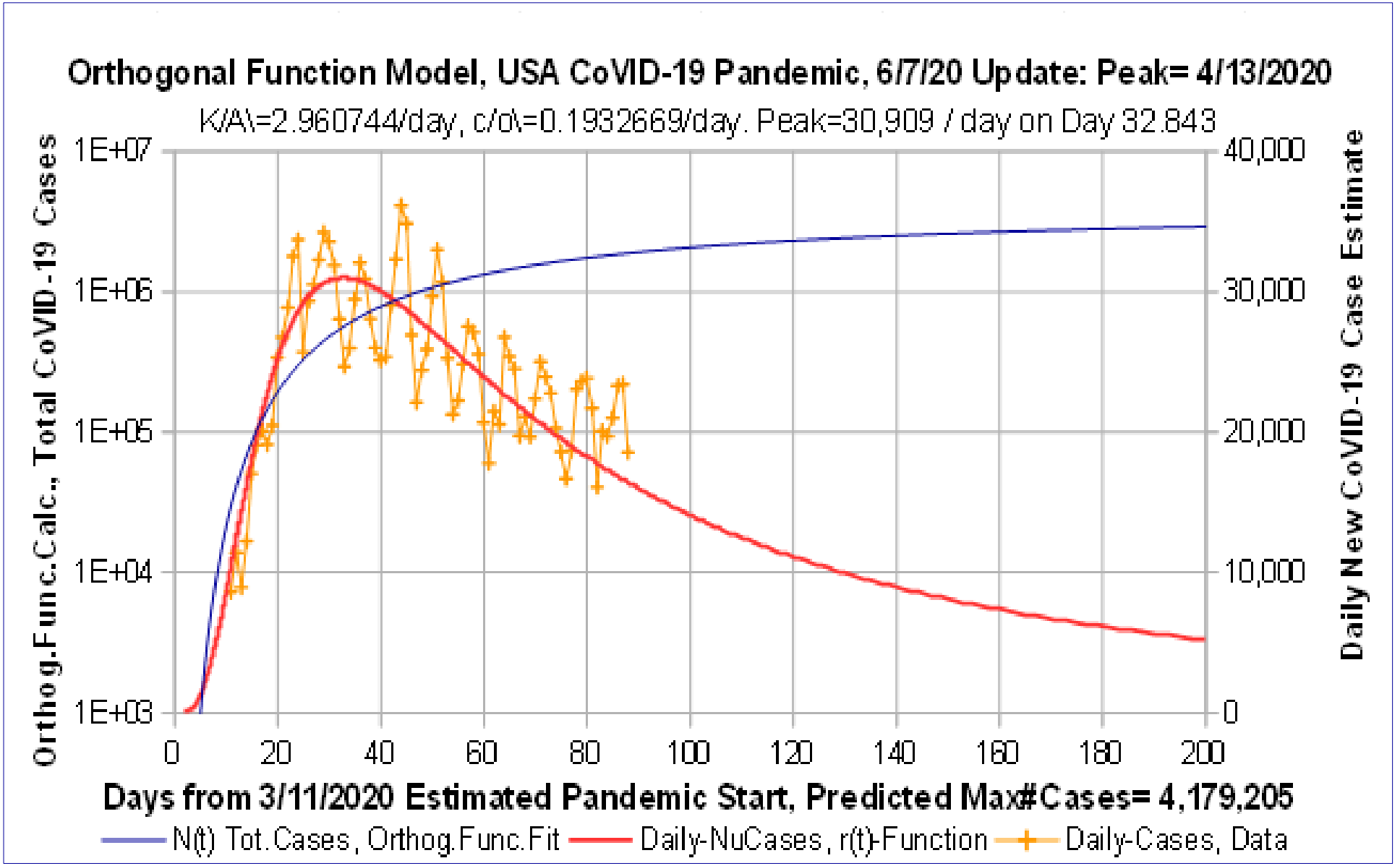

Figure 8: *Orthogonal Function Model*, USA CoVID-19 Projections, data to 6/7/2020. Revised *bing.com* data; daily# of CoVID-19 Cases Peak at 30,909 cases/day on 4/13/2020; with 4,179,205 cases total; and ~5,140 new cases/day at Day 200 on 9/27/2020.

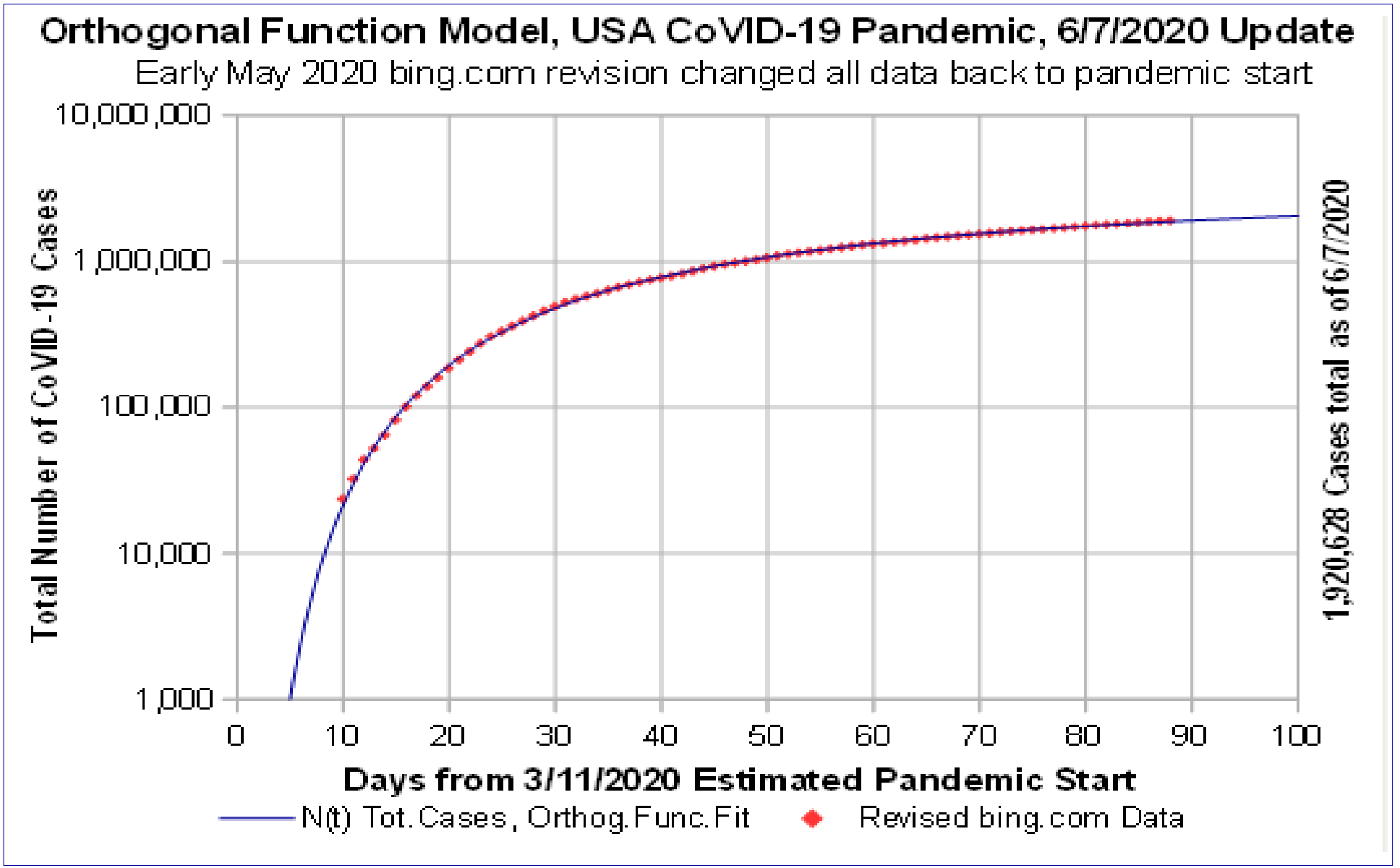

Figure 9: *Orthogonal Function Model*, USA CoVID-19 Projections, data to 6/7/2020. Revised *bing.com* data, posted circa 5/3/2020, changed values back to the pandemic start. *Orthogonal Function Model* matches the data a bit better than the *Initial Model*.

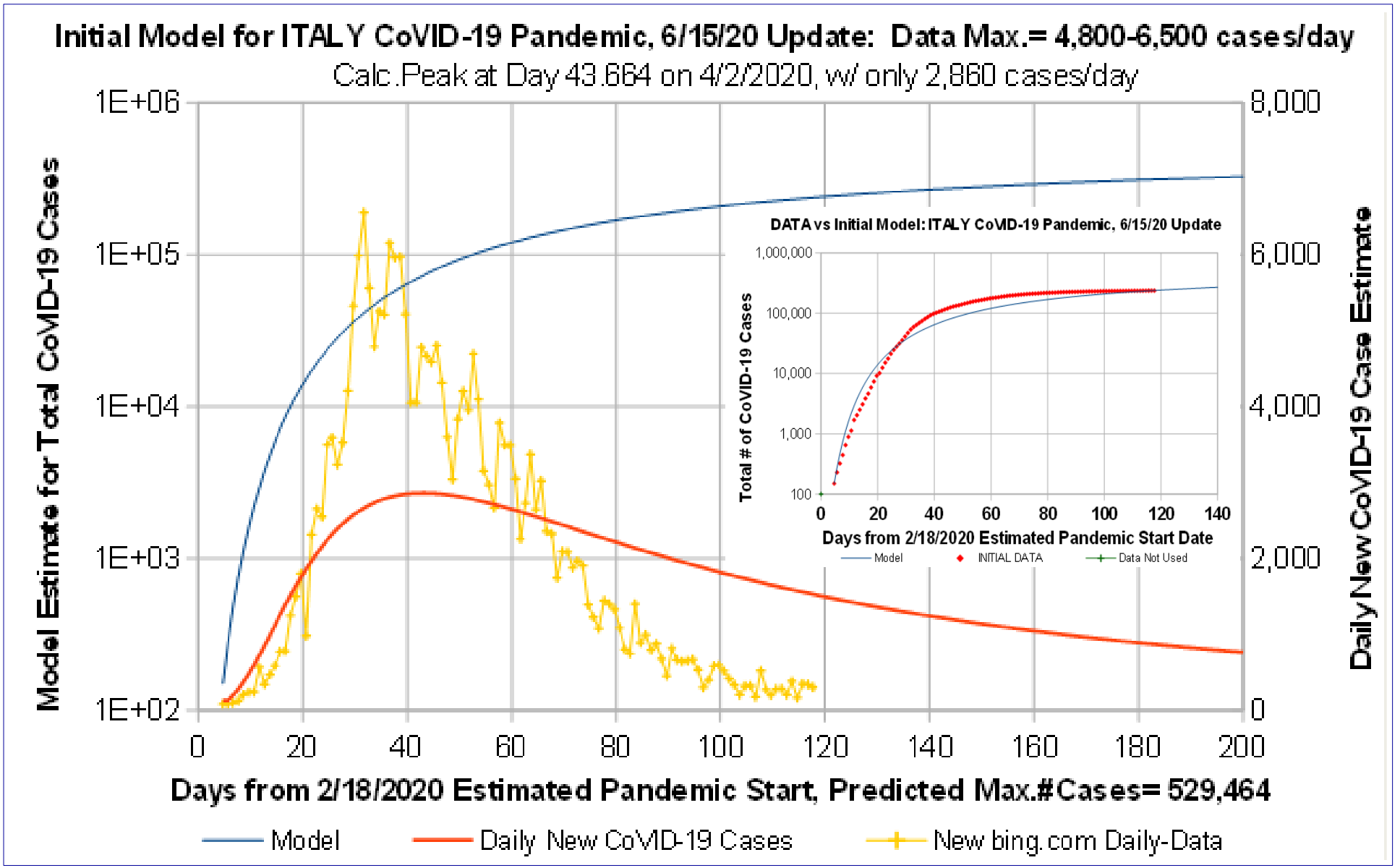

Figure 10: *Initial Model* for ITALY CoVID-19 Projections vs data up to 6/15/2020.
Initial Model matches Total Number of Cases at data start and data end, but best fit using a logarithmic Y-axis does not give a good fit for Predicted Number of Daily CoVID-19 cases.

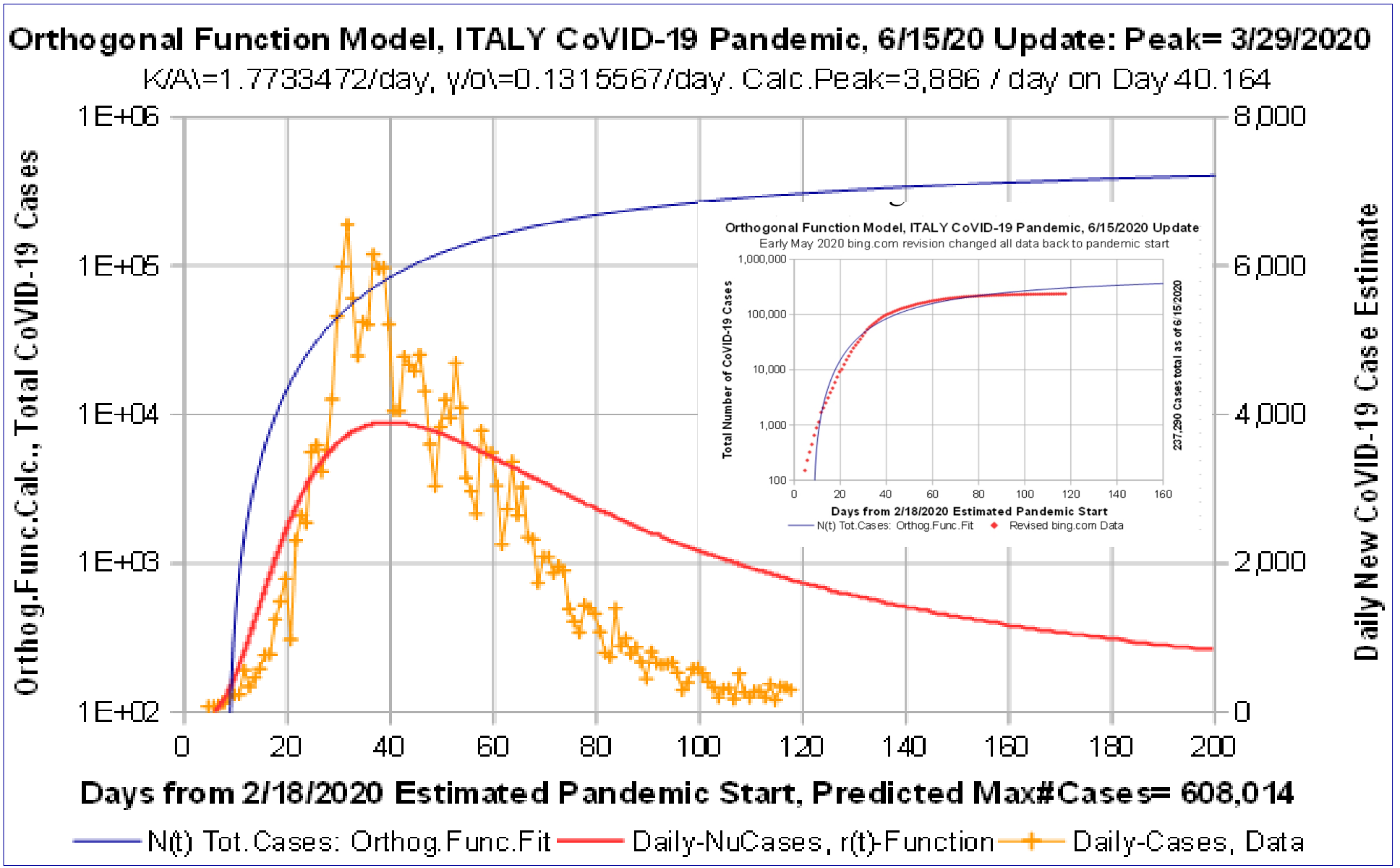

Figure 11: *Orthogonal Function Model* for ITALY CoVID-19 data up to 6/15/2020.
Orthogonal Function Model gives improved datafit, but 3-terms in orthogonal function series is insufficient to accurately predict a rapidly decreasing Number of Daily CoVID-19 cases.

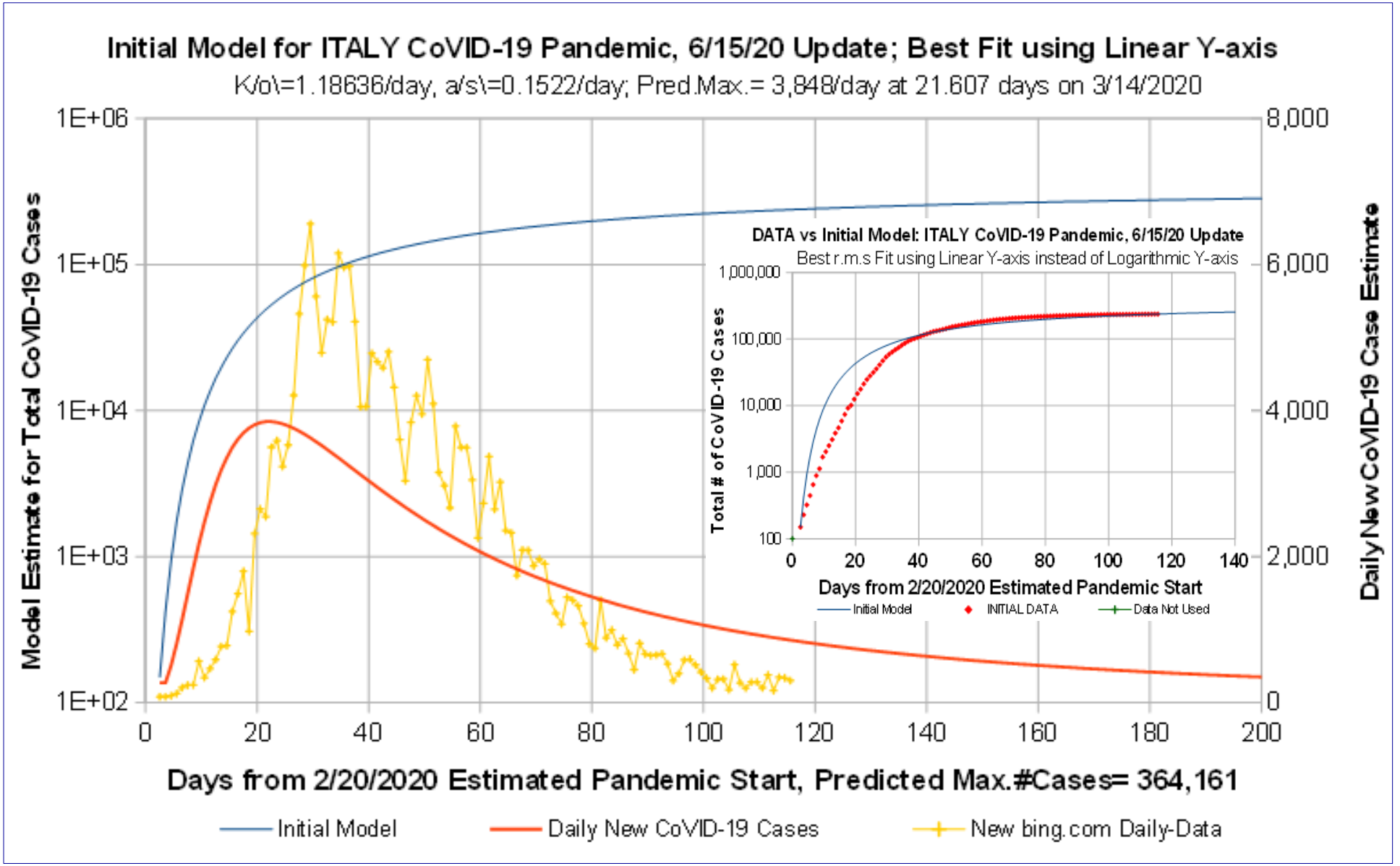

Figure 12: *Initial Model* re-do, ITALY CoVID-19 data to 6/15/2020. New starting point is a best fit function on a linear Y-axis, instead of having a best fit using a logarithmic Y-axis. Alternative method may allow a few-term orthogonal function series to better match the data.

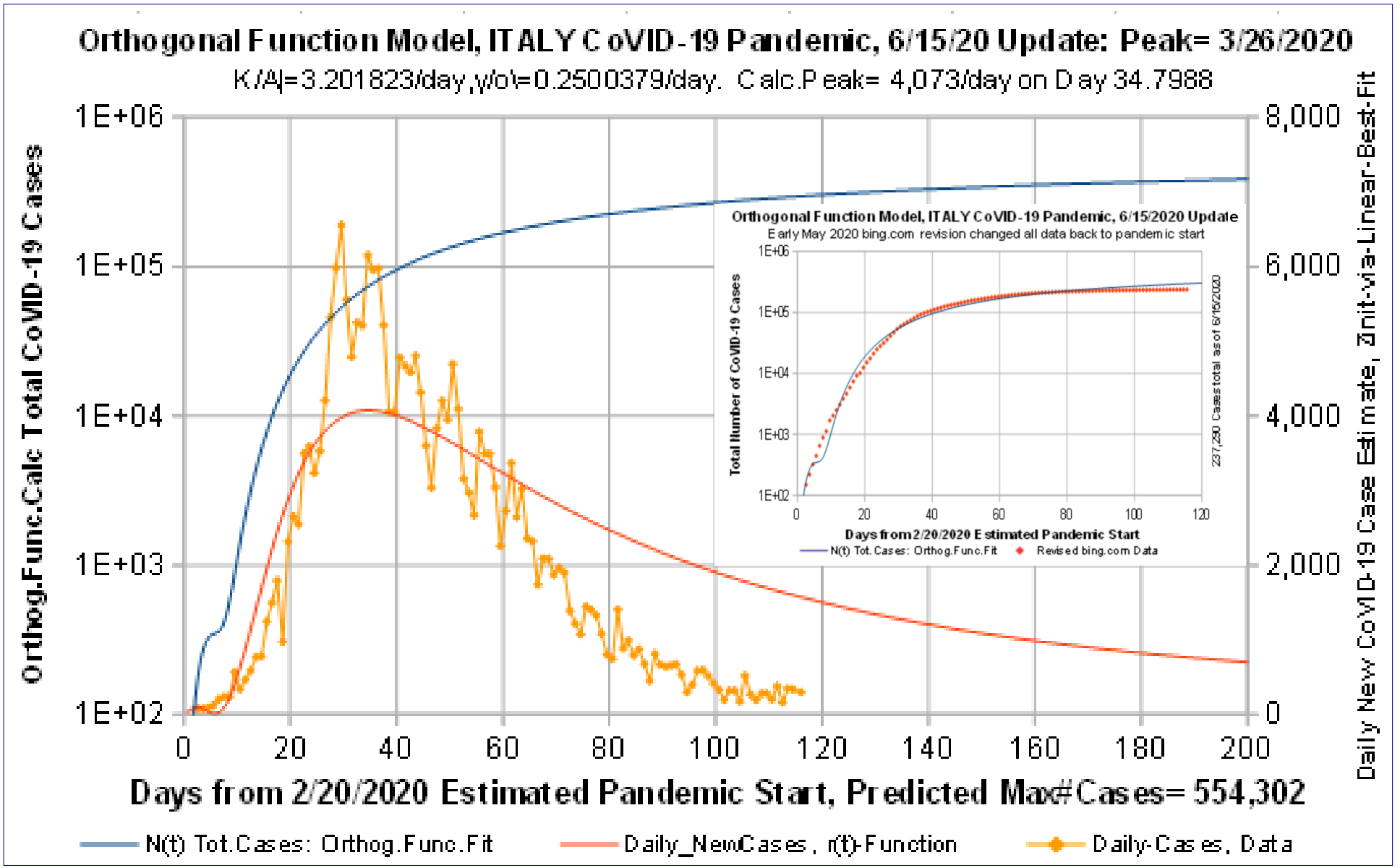

Figure 13: *Orthogonal Function Model* re-do, ITALY CoVID-19 data to 6/15/2020.
Orthogonal Function Model re-do using linear Y-axis gives a slightly better small-series fit. Other *Social Distancing* impacts likely exist besides just lengthening pandemic doubling times.