Entendamos el COVID-19 en México

Octavio Miramontes
Instituto de Física
Universidad Nacional Autónoma de México

16 de junio de 2020

Resumen

En el presente trabajo se acompañan los datos diarios de la epidemia COVID-19 en México. Para ello se discute uno de los modelo epidemiológico más básicos y se muestra como puede ser resuelto para analizar las cifras oficiales de los casos positivos identificados en la República Mexicana. Esto permite explorar la dinámica de las predicciones de su evolución, en un corto plazo. Adicionalmente, se explican los conceptos epidemiológicos más básicos y se hace énfasis en mostrar las herramientas computacionales que se requieren para ello. Debido a su simplicidad, los resultados de esta investigación no deben ser tomados como base para recomendaciones de salud pública sino más bien como un ejercicio para fines educativos y de divulgación.

Índice

1. Introducción **2**
1.1. Un poco de historia . . . 2
1.2. La epidemiología matemática . . . 2
1.3. El modelo SIR y la pandemia de 1918 . . . 5
1.4. Covid-19 en China . . . 7
1.5. Covid-19 en Europa . . . 8

2. México, COVID-19 y los modelos tipo SIR **9**
2.1. ¿Predecir una epidemia con datos parciales? . . . 10
2.2. Una crítica a la capacidad predictiva de los modelos tipo SIR . . . 14
2.3. Evolución de casos positivos detectados en México . . . 15
2.4. Pruebas masivas en México . . . 18
2.5. Aplanar la curva, movilidad y contacto social . . . 19
2.6. ¿Se ha aplanado la cuerva en México? . . . 19

3. Modelos epidemiológicos modernos **22**

4. Conclusiones **25**

5. Anexo A - Programación de los modelos SIR y SEIR **27**

6. Anexo B - Ajuste de funciones a los datos numéricos del modelo SIR y SEIR **29**

7. Anexo C - Extracción de datos de una gráfica **34**

1. Introducción

Actualmente existe una propagación a nivel global (pandemia) del coronavirus SARS-CoV-2 causante de la enfermedad COVID-19. Enfermedad contagiosa nueva en humanos para la cual no tenemos respuesta inmune ni vacuna. Se cree que su origen son animales silvestres (posiblemente murciélagos) de la localidad de Wuhan en China, donde aparentemente se originó a finales del año 2019. Sin embargo no se ha identificado al individuo cero (aquel primer infectado) por lo cual su origen exacto, en el tiempo y en el espacio, permanece incierto [1]. La enfermedad fue transportada rápidamente a escala global por individuos inicialmente sin síntomas viajando desde China a Europa y Estados Unidos principalmente. En México se detectó por primera vez en febrero en un individuo que estuvo de viaje en Italia [2].

COVID-19

Hasta el 23 de mayo del 2020, el número de personas infectadas (positivas) con el SARS-CoV-2 asciende a 5,482 millones en todo el mundo y se han registrado 346,072 fallecimientos [3]. Los síntomas principales de la enfermedad son la tos seca, la fiebre, estornudos, dolores de cabeza, dolor de garganta y malestar físico principalmente. Se transmite principalmente en pequeñas partículas de fluidos contaminados, mediante el estornudo o el contacto de la piel en las manos y su posterior contaminación en la cara, ojos y nariz. Sus síntomas se manifiestan aproximadamente a la semana de la infección (puede variar entre dos días a dos semanas). Dado que hasta la fecha no existe una vacuna ni medicamentos comprobados para la mitigación de la enfermedad, la únicas medidas eficaces son el distanciamiento físico y la higiene profunda.

1.1. Un poco de historia

La historia de las enfermedades contagiosas en humanos es amplia. Existen casos dramáticos de pandemias en zonas geográficos amplias y que han afectado a millones de personas, como la viruela (causada por dos variedades de virales). Se tiene evidencia de epidemias de viruela en casos tan antiguos como en las momias de Egipto o en la Europa del siglo XVIII, donde alrededor de 400,000 murieron al año debido a esta enfermedad.

Una de las pandemias más devastadoras fue la llamada "muerte negra" o "la plaga" que mató alrededor de entre 75 y 200 millones de personas en el siglo XIV, en diversas regiones de Europa, África y Asia. Esta enfermedad aparentemente se debió a la bacteria *Yersinia pestis*.

Otro caso es el de la mal llamada "gripe española" o Pandemia de gripe de 1918, que infectó alrededor de 500 millones de personas de las cuales mató alrededor de entre 17 y 50 millones en varias partes del mundo (ver Figura 1). Más recientemente tenemos la pandemia del 2019 conocida como pandemia de la influenza H1N1. Pandemias actuales y vigentes son la COVID-19 y la del SIDA.

1.2. La epidemiología matemática

Uno de los primeros intentos conocidos para explicar con un modelo matemático la dinámica de una epidemia se debe al físico, matemático y médico Daniel Bernoulli (1700-1782), quien en 1766 publicó un artículo en el que propuso estrategias óptimas para la vacunación contra la viruela, basadas en un modelo matemático. Sin embargo, un modelo más robusto y que se emplea aún hoy para explicar los aspectos básicos de la dinámica de una epidemia se debe, entre otros, a William O. Kermack (1898-1970) y a Anderson G. McKendrick (1876-1943); quienes desarrollaron el modelo

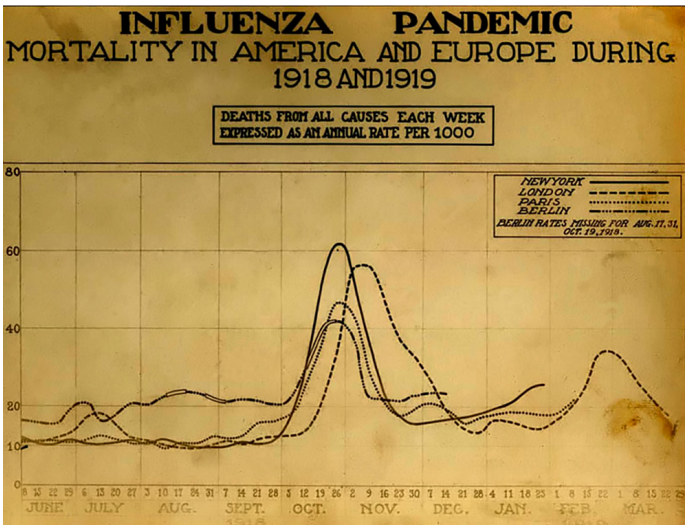

Figura 1: Gráfica de la mortalidad de la Pandemia de gripe de 1918. Nótese el pico de defunciones entre Octubre y Noviembre en cuatro ciudades. Cortesía del Museo Nacional de Salud y Medicina, EEUU.

que lleva sus nombres: modelo Kermack-McKendrick [4] (véase también [5]). Este modelo también se conoce como modelo SIR debido a que divide a los individuos en tres categorías temporales: los susceptibles $S(t)$, los infectados $I(t)$ y los recuperados $R(t)$ y estas cantidades se relacionan en un conjunto de tres ecuaciones diferenciales acopladas (ver ecuaciones 1-3):

$$dS/dt = -kSI \tag{1}$$

$$dI/dt = kSI - lI \tag{2}$$

$$dR/dt = lI \tag{3}$$

donde k y l son cantidades mayores a cero y representan la tasa media de recuperación (k) y la tasa media de infección (l). Estas dos cantidades se relacionan junto con N (la población total) para definir una cantidad clave:

$$R_0 = N(\frac{l}{k}) \tag{4}$$

R_0 es conocida como la tasa reproductiva de la enfermedad. Si R_0 >1 significa que la enfermedad se propaga exponencialmente, si R_0 <1 significa que la infección se extinguirá en un tiempo corto.

El sistema de ecuaciones Kermack-McKendrick no admite soluciones analíticas; entonces se resuelven numéricamente (ver Figura 2 y el Anexo A, donde también se describe el modelo SEIR). Para nuestros propósitos, la gráfica de los individuos infectados, digamos la curva en la gráfica derecha de la figura 2, puede ajustarse a funciones "pico" bien conocidas (ver anexo B). Ejercicio que nos da varias funciones para el ajuste, ordenadas de acuerdo al coeficiente de regresión r^2 y que se muestran en la tabla 1.

El ajuste de curvas es verdaderamente bueno. Los coeficientes de regresión r^2 son muy cercanos a 1.0 y en los dos primeros casos son casi idénticos. Lo anterior se atestigua en la gráfica de la curva $I(t)$ del modelo SIR en la Figura 3 donde se han dibujado las tres curvas superpuestas y su límite de confianza al 95 %. Son indistinguibles entre sí.

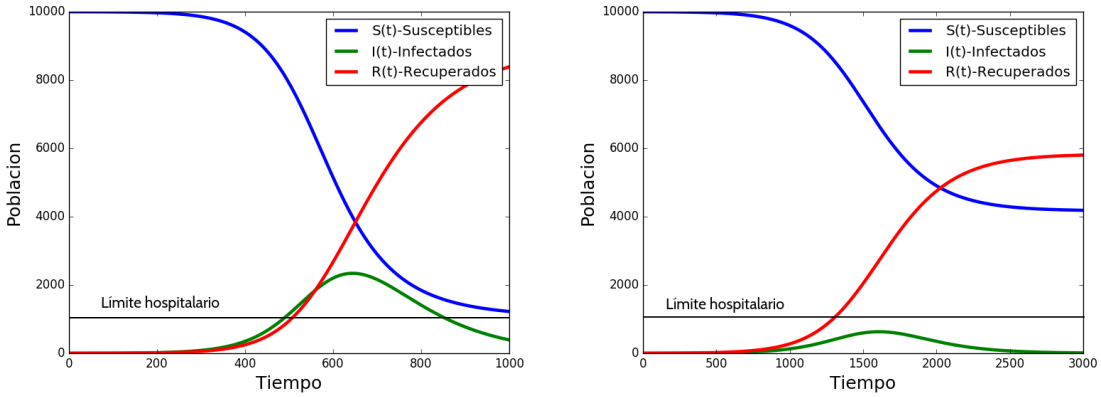

Figura 2: Dos gráficas que muestran la evolución temporal de una infección en una hipotética ciudad de 10,000 habitantes, según el modelo SIR. La infección inicia con un individuo enfermo que propaga por contagio la enfermedad entre 9,999 individuos susceptibles. La población de individuos infectados tiene una curva (en verde) caracterizado por un máximo, que es grande o menor dependiendo de la agresividad del patógeno. A la izquierda R_0 = 2.5 y a la derecha R_0 = 1.5. Se ha dibujado una línea horizontal hipotéticamente a los 1000 que indica la capacidad hospitalaria en esa ciudad. En un caso se verá rebasada; en eso consiste la tan hablada "aplanar la curva", ya que la propagación de la infección puede disminuirse mediante intervención.

Función	Nombre	r^2
ADS	Doble sigmoide asimétrica	0.9999956766
LgstcPowPk	Logísitica con potencia	0.9999657106
AsymLgstc	Logísitica asimétrica	0.9989199161

Tabla 1: Tres funciones ajustadas a la curva $I(t)$ (en verde) en la gráfica derecha de la Figura 2 de un modelo SIR. Véase el anexo B para más detalles

Una manera alternativa de representar los datos de una epidemia, es mediante la curva de datos acumulados. Es útil porque muestra una asíntota en el número total de infectados. La forma de la curva es de tipo "transición", es decir tipo "S". En la figura 4 se exhibe el ajuste a dos funciones conocidas en dinámica de poblaciones. La curva [Sigmoide] y la curva [Gompertz] (ver el Anexo B para los detalles). Cómo puede apreciarse, ambas funciones ajustan bien los datos del número acumulado de infectados $I(t)$ del modelo SIR.

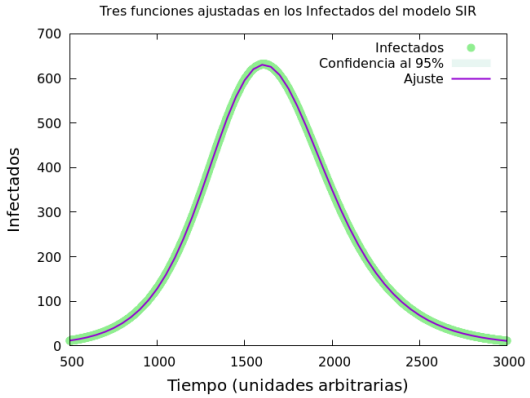

Figura 3: Gráfica de los individuos infectados $I(t)$ en el ejemplo del modelo SIR mostrado en la Figura 2 con R_0=1.5 y que aparecen como puntos color verde. Sobre ellos se han graficado las tres funciones ajustadas de la Tabla 1 y que aparecen como la línea continua magenta. El ajuste es tan bueno que los límites de confianza al 95 % quedan ocultos por la línea del ajuste.

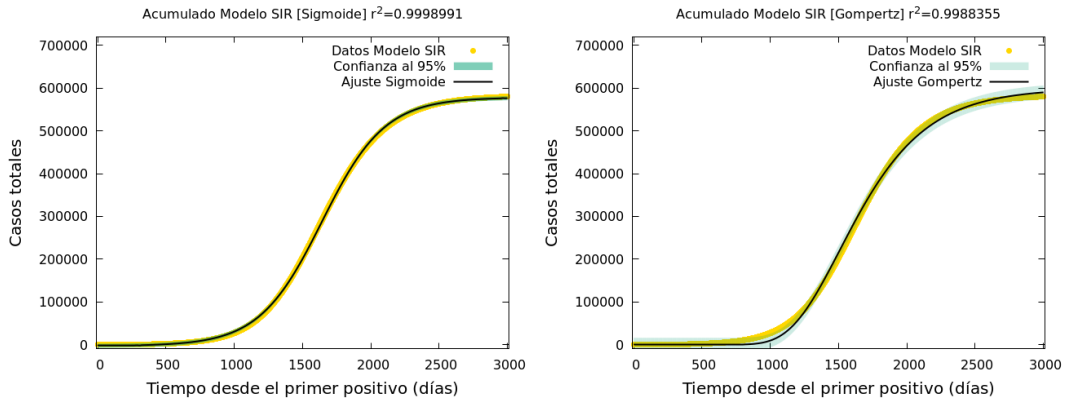

Figura 4: Dos funciones con coeficientes de regresión r^2 alto (ver el título de las gráficas) ajustados a los datos acumulados del número de infectados de una simulación del modelo SIR. Nótese la característica forma en "s" de la curva.

1.3. El modelo SIR y la pandemia de 1918

¿Y qué tal se comporta el modelo SIR para datos reales? Curiosamente existe un registro gráfico del número de fallecidos en la Pandemia de 1918 (ver Figura 1) en cuatro ciudades: New York, Londres, París y Berlín. Es razonable suponer que el número de fallecidos es proporcional al número de individuos infectados por ello es factible realizar un ajuste de funciones a, digamos, al caso de la ciudad de New York. Una vez extraídos los datos de la gráfica (véase el Anexo C), los resultados del ajuste nos dan la tabla 2 de funciones con el coeficiente r^2 más alto, de entre 24 mejores funciones ajustadas (ver Anexo B, para más detalles). Se concluye de inmediato que la curva de fallecidos en la ciudad de NY siguen razonablemente bien la forma funcional de los individuos enfermos en el modelo SIR.

Función	r^2
AsymLgstc	0.9990799570
LgstcPowPk	0.9990437065

Tabla 2: Dos funciones ajustadas a la curva $I(t)$ para la ciudad de New York, en la gráfica de la Figura 1 (curva verde), con el mayor r^2 y que coinciden con dos de las funciones con el mejor ajuste para los datos de la curva de infectados en el modelo SIR. Nótese que los valores del coeficiente de regresión son tan parecidos que resultan casi indistinguibles entre sí. Véase el anexo B para más detalles.

Podemos a continuación graficar los ajustes de estas dos funciones sobre impuestas a los datos extraídos, esto se aprecia en la figura 5.

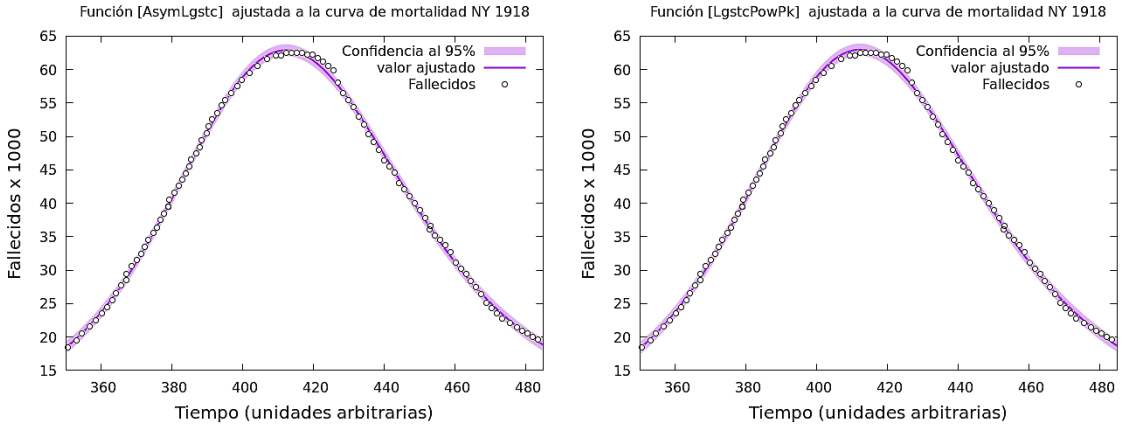

Figura 5: Dos funciones con el r^2 más alto ajustados a los datos extraídos de defunciones para la ciudad de NY durante la pandemia de gripa de 1918 (bolas blancas). En ambos casos se aprecia un ajuste muy bueno; lo que nos permite concluir razonablemente que el modelo SIR, pese a su simplicidad, pudo haber predicho estos niveles de mortalidad.

1.4. Covid-19 en China

En estos momentos existe un desacuerdo con respecto a como se originó la epidemia COVID-19 en China, concretamente en la región de Wuhan. Algunas fuentes reportan casos desde Noviembre de 2019; pero la inexistencia de pruebas especificas para un virus nuevo limitaron, en su momento, la posibilidad de un diagnostico preciso del paciente cero y con ello poder determinar su ruta exacta. China, por razones socioculturales, ha sido fuente –aunque no es el único país– del tránsito de coronavirus específicos causantes del Síndrome Respiratorio Agudo Grave (SARS) hacia los humanos. Este circuito, posible causante de enfermedades graves, fue denunciado y alertado por los expertos desde 2007 y sin embargo pasó desapercibido para las autoridades políticas, no sólo de China sino del mundo entero [6][1]. La epidemia en China proporciona los datos más completos disponible a la fecha. Eso debido al número tan alto de pacientes (ya superado al momento por los Estados Unidos y otros países), por su duración y porque ha superado el pico de mayor contagio y se encuentra en plena desaceleración, de tal manera que tenemos una evolución casi completa.

Con los datos disponibles para China [3] podemos verificar que, al igual que los datos de 1918 de New York, la curva de infectados por COVID-19 en China se ajusta muy bien a las funciones "pico" que son las mismas que se ajustan bien al modelo SIR (ver Figura 2). Esta experiencia la reportamos en la figura 6.

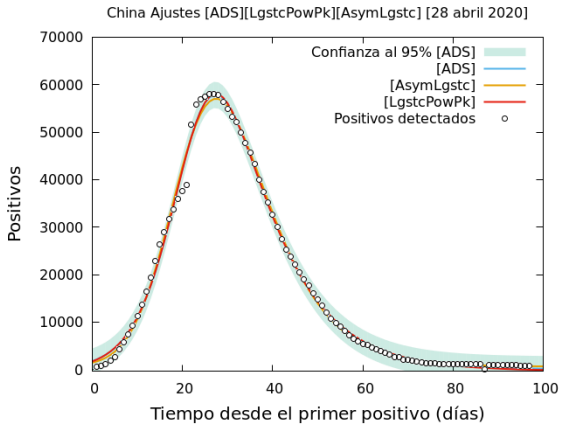

Figura 6: Las tres funciones SIR se ajustan perfectamente bien a los datos disponibles para el número de infectados diarios en China, hasta el 28 de abril del 2020 (ver Tabla 1). Esto nos da certeza que el modelo básico Kermack-McKendrick es un buen modelo para explicar esta dinámica. Véase el Anexo B para más detalles.

Una manera alternativa de confirmar que la dinámica de la epidemia en China sigue patrones matemáticos claros, es el análisis de los casos acumulados. Aquí, en lugar de las funciones anteriores, ajustamos funciones de "transición" como se aprecia en la Figura 7, donde vemos un ajuste casi perfecto.

[1] "The presence of a large reservoir of SARS-CoV-like viruses in horseshoe bats, together with the culture of eating exotic mammals in southern China, is a time bomb. The possibility of the reemergence of SARS and other novel viruses from animals or laboratories and therefore the need for preparedness should not be ignored". Ver Cheng *et al.* 2007

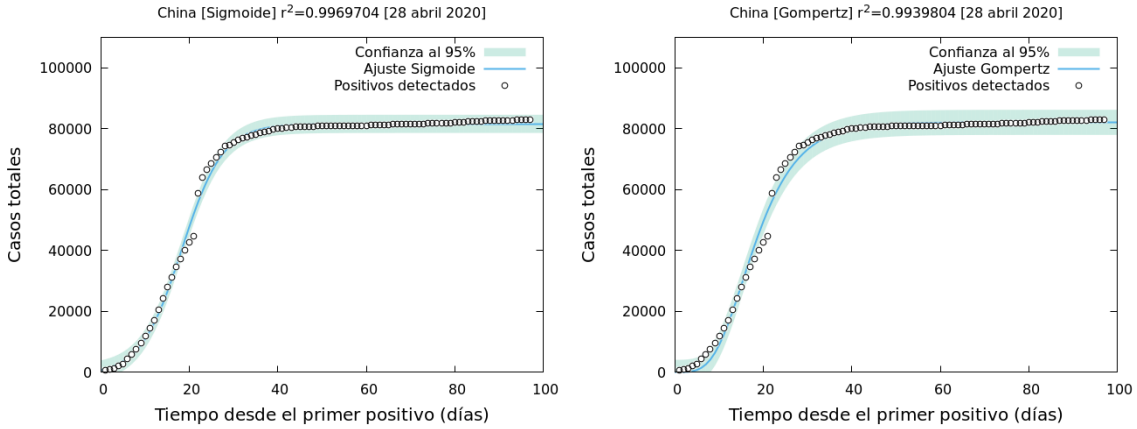

Figura 7: Dos funciones con el r^2 más alto ajustados a los datos acumulados de positivos en China.

1.5. Covid-19 en Europa

Veamos el caso de dos naciones europeas que han, aparentemente, pasado la curva máxima de positivos detectados, lo que llamaremos una epidemia "madura". Tal es el caso de España y Alemania. entre otros.

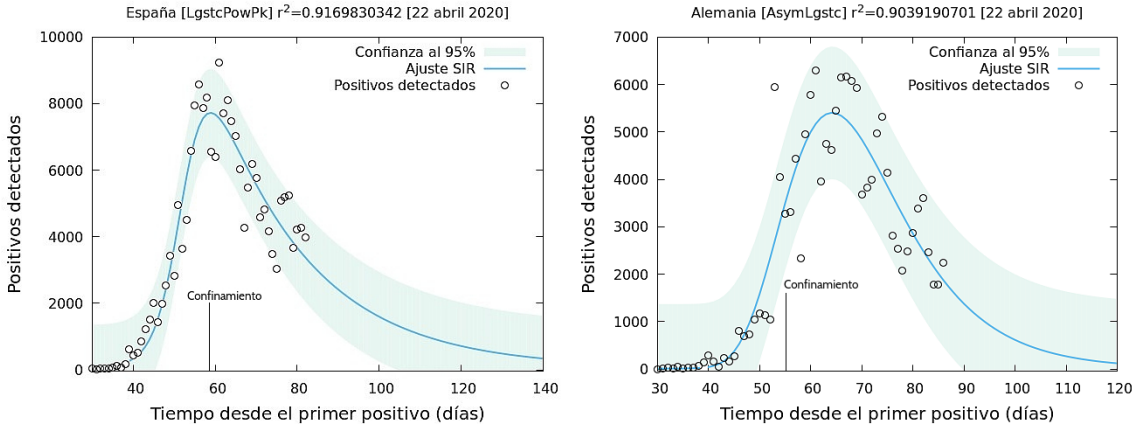

Figura 8: Funciones que asemejan al modelo SIR ajustadas a los casos de España y Alemania. En ambos casos el coeficiente de regresión más alto es bueno. Nótese que los datos reportados en ambas series de tiempo tienen una variación muy grande entre sí, lo cual no observamos en México, al menos no a la fecha. Esa dispersión tan grande puede ser característica de una Fase 3, donde hay contagios descontrolados.

España reportó su primer caso en La Gomera, Islas Canarias el 31 de enero del 2020. Posteriormente el 24 de febrero se reportaron múltiples infecciones en Tenerife. Para el 13 de marzo se reportaron casos en todas las provincias españolas lo que precipitó la declaración de emergencia el 14 de marzo. El 29 de marzo se declaró el confinamiento para todos los trabajadores en labores no esenciales. Por su parte, en Alemania, se reportó el primer caso positivo cerca de Múnich el 27 de

enero. El primer deceso ocurrió el 9 de marzo y el confinamiento general se declaró el 22 de marzo. En ambos casos, el confinamiento parece haber llegado tarde, pues el número de casos de positivos detectados era ya muy grande. En todo caso, el pico y punto de inflexión de la curva ocurrió unos días después del confinamiento, entre cinco y 10 días respectivamente ¿Habrá el confinamiento obligatorio jalado la curva hacia abajo? [7]

2. México, COVID-19 y los modelos tipo SIR

El siguiente análisis es meramente un ejercicio matemático sin validez para sugerir, criticar o apoyar políticas públicas

En México se cuenta afortunadamente con la experiencia de vigilancia epidemiológica a raíz de la pandemia de H1N1 de 2009 y ello ha permitido que la Secretaría de Salud haya podido instrumentar medidas preventivas y de preparación desde enero del 2020. Además de ello, las autoridades federales se asesoran de grupos de matemáticos expertos en modelos epidemiológicos que provienen de instituciones de prestigio en el campo como la UNAM y el CIMAT, entre otras [8]. Adicionalmente existe una estrategia sin precedente, a nivel mundial, de comunicación directa diaria de parte de las autoridades federales, para comunicar la evolución diaria de la enfermedad en México.

Un problema crucial tanto para las autoridades, académicos y población general, es saber con cierto grado de certeza al menos dos datos importantes de la epidemia en México. Uno es el día de mayor contagio, es decir, el pico de la curva y por tanto el número máximo de enfermos ese día (no debería rebasar el número de camas de hospital disponibles) y por último, la fecha de desaparición de la enfermedad ¿Podemos predecir alguna de estas fechas con suficiente antelación y exactitud? Aparentemente no y eso es una mala noticia. Se trata de un problema similar al de la predicción del clima a largo plazo. Se puede predecir, es un hecho; pero no necesariamente bien.

En estos días hemos visto el surgimiento de decenas de académicos mexicanos en la prensa clamando tener "los verdaderos datos" y por tanto haciendo críticas algunas veladas y otras francamente en confrontación hacia las autoridades de salud y las cifras que se nos proporcionan ¿Tendrán razón? Muchos de estos nuevos "expertos" han tomado el modelo SIR o derivados y han escogido hacer públicos sus ejercicios de predicción, no en los foros académicos sino en la prensa sensacionalista; alimentando con ello el sentimiento de zozobra de por sí ya experimentado por una población confinada que se siente vulnerable y confusa. Pero las autoridades de salud en México también hacen ejercicios de predicción, basados en modelos matemáticos semejantes [9]. Lo hacen porque definitivamente es la única y mejor manera de enfrentar la pandemia, es decir haciendo uso de la ciencia para conducir las políticas públicas porque esa es su responsabilidad; pero sus predicciones, lamentablemente, no están exentas de las incertidumbres que aquí exploraremos.

En un reciente artículo, muy revelador por cierto, el grupo de investigación de Susanna Manrubia [10] ha desarrollado un modelo tipo SIR con una categoría adicional $C(t)$, los confinados. Este modelo, al cual han podido hacer una aproximación analítica de la curva de positivos, lo han empleado para analizar los datos de la pandemia en España y han concluido que las predicciones del máximo de la curva da resultados disparatados y contradictorios y esto es intrínseco a los modelos tipo SIR. Como si se tratara de una burla, los modelos SIR son extremadamente buenos para describir los datos de una pandemia. Lo hemos visto con el ejercicio hecho para la pandemia de 1918 en NY. Pero ajustar el modelo a una pandemia ya superada es muy diferente a aplicarlos a un pandemia en progreso y con ello intentar predecir. A igual que los fenómenos y modelos caóticos cuya evolución en el tiempo muestra divergencias que crecen de manera exponencial, los modelos SIR, en su evolución e intento de predicción, acumulan incertidumbres que hacen que los límites

de confianza de la predicciones divergen de manera no-lineal (algunas veces de manera exponencial). Lamentablemente este hecho no lo han tomado en cuenta los ejércitos de matemáticos, físicos, biólogos e informáticos de nuevo arribo. A continuación lo exploraremos con un ejemplo.

2.1. ¿Predecir una epidemia con datos parciales?

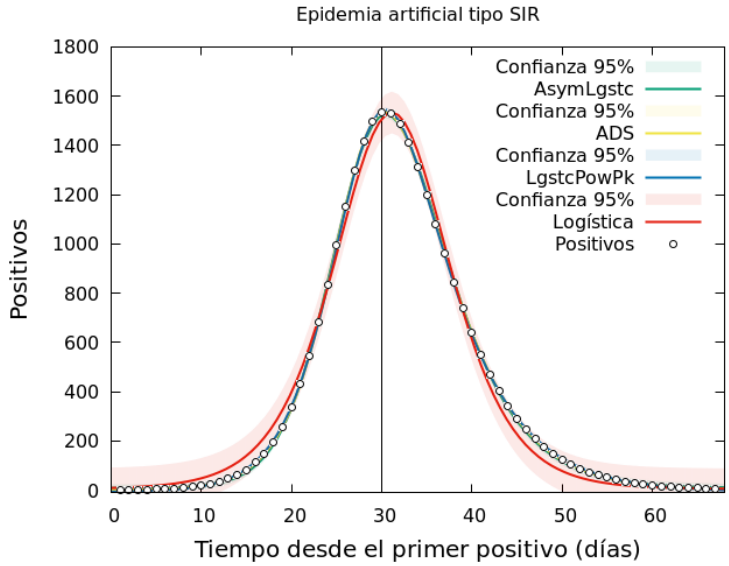

Figura 9: Cuatro funciones que se asemejan al modelo SIR de las ecuaciones (1-3) son ajustadas a una epidemia artificial. Las tres funciones asimétricas que hemos empleado anteriormente se ajustan a los datos de manera impecable. De hecho lo hacen tan bien que los límites de confianza tiene el mismo ancho que la curva de la función y por ello no son visibles; pero el caso de la función logística simétrica (en rojo) es peculiar pues no se ajusta del todo bien. Es la única cuyos límites de confianza son anchos sobre todo en la base de la campana y en el pico, en pocas palabras no explica muy bien los datos; pero su virtud se verá más adelante.

Supongamos aquí una epidemia que sigue una dinámica tipo SIR (ver las ecuaciones 1-3). Para ello hemos generado datos artificiales discretos tomados de la gráfica que se muestra en la Figura 3 (R_0=1.5) y la mostramos a continuación en la Figura 9. Nuestra epidemia tiene una duración de 67 días con un pico máximo a los 30 días donde, en ese día, se registran 1532 casos positivos. Supongamos que hemos ido acompañado el registro de positivos diarios desde el inicio de la epidemia y entonces nos preguntamos si con nuestros registros a los diez días (un tercio de los días antes del pico) podemos hacer una predicción usando las funciones descritas en la Tabla 1. Nótese que hemos agregado otra función al ajuste dado que esta describe los datos de manera razonable (r^2=0.9933007) aunque no de la mejor manera en comparación a las otras tres; se trata de una función logística simétrica (ver Anexo B, para más detalles). Lo hemos hecho porque es empleada con frecuencia y en particular ha sido empleada por un sitio de predicciones que se ha vuelto muy popular [11] y que nos brinda predicciones del pico para México.

Con este ejercicio, nos preguntamos si con datos a los 12,14,16,18,20,22,24,26 y 28 días podemos predecir con exactitud, o razonablemente bien, el pico que se avecina en el día 30 de nuestra epidemia artificial. En este escenario, no existe ninguna medida de mitigación del tipo confinamiento o

distanciamiento social, de tal manera que la tasa de infección del patógeno se mantiene inalterada.

Analizamos las gráficas que se presentan en la Figura 10. Hemos considerado los datos hasta el día 18, como si se tratara de una epidemia en evolución y no sabemos cuando ni de que magnitud sera el pico de positivos detectados. El día 18 representa el 60 % de avance en la evolución hasta el pico máximo. Podemos apreciar en un primer vistazo cuan diferentes son las predicciones que nos brindan las cuatro funciones ajustadas.

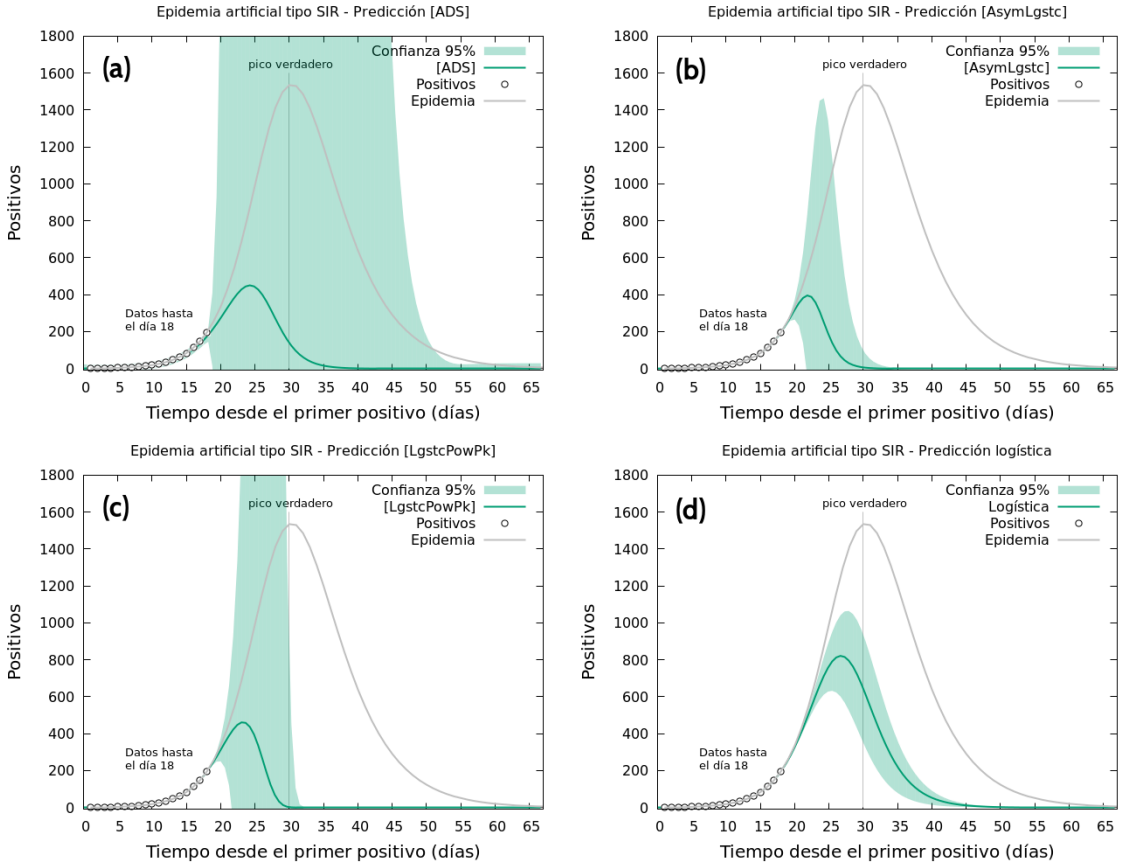

Figura 10: Cuatro funciones que se ajustan al modelo SIR son usadas para explorar la posible evolución futura de la epidemia artificial en un intento de predecir el pico máximo que ocurrirá el día 30 La línea en gris es la curva del epidemia verdadera y la línea vertical muestra el día del pico máximo. Se consideran datos "observados" hasta el día 18.

En la Figura10(a) observamos la predicción hecha con la función [ADS](Doble sigmoide asimétrica). Como ya lo mencionamos esta función tiene el coeficiente de regresión más alto para explicar la totalidad de los datos; pero usándola para predecir nos muestra un desempeño mediocre, por decir lo mejor. Simplemente los límites de confianza para la predicción de el día siguiente se han disparado a niveles absurdos. Los límites de confianza cubren prácticamente toda la epidemia. Incluso el pico máximo predicho es menor en magnitud y sucedería el día 24. Lo mismo podemos decir de los casos en (b)[Asymlgstc] y (c)[LgstcPowPk]: las magnitudes de los picos son muy menores, los límites de confianza se disparan un día o dos hacia el futuro de manera drástica y no

lineal y los picos ocurren los días 22 y 23. Por último, en (d)[Logística simétrica] podemos observar un mejor desempeño en la predicción. El pico ocurre el día 27 con mayor magnitud que en el resto, aunque menor al valor real. Pero lo importante es la evolución de los límites de confianza. En este caso su dispersión es menor y permiten predecir los valores correctos más allá del día 18 por lo menos hasta el día 24. Lo anterior lo podemos visualizar mejor en la Figura 11 donde se explora el crecimiento no lineal de los límites de confianza de las predicciones en el tiempo y la convergencia de las predicciones del pico máximo como función del tamaño del intervalo utilizado para la predicción.

Destaquemos los siguientes puntos relevantes. En la Figura 11(a) se grafica en ejes semi logarítmicos. Es decir una escala no lineal para destacar que los crecimientos tipo línea recta, en ellos, son en realidad crecimientos exponenciales o cercanos a ellos. Es decir, los límites de confianza de la predicción crecen de esta manera a medida que avanzamos en el tiempo. Concluimos que la función [ADS], la de mejor coeficiente de regresión para la totalidad de datos tiene una capacidad para estimar muy baja. Pero la función Logística (en verde) es prometedora; de hecho acompaña muy bien los datos verdaderos en la predicción a cuatro días (hasta el día 22, con datos al día 18). Por otra parte, en la Figura 11(b) hemos representado la capacidad para predecir de las cuatro funciones ajustadas en intervalos desde el día 12 hasta el día 28. Las cuatro funciones se acercan al valor verdadero (pico al día 30) desde abajo, es decir subestimando el valor del día. Pero de las cuatro funciones, la Logística (en verde) es la que mejor desempeño muestra, pues se acerca al valor verdadero mucho antes que las demás.

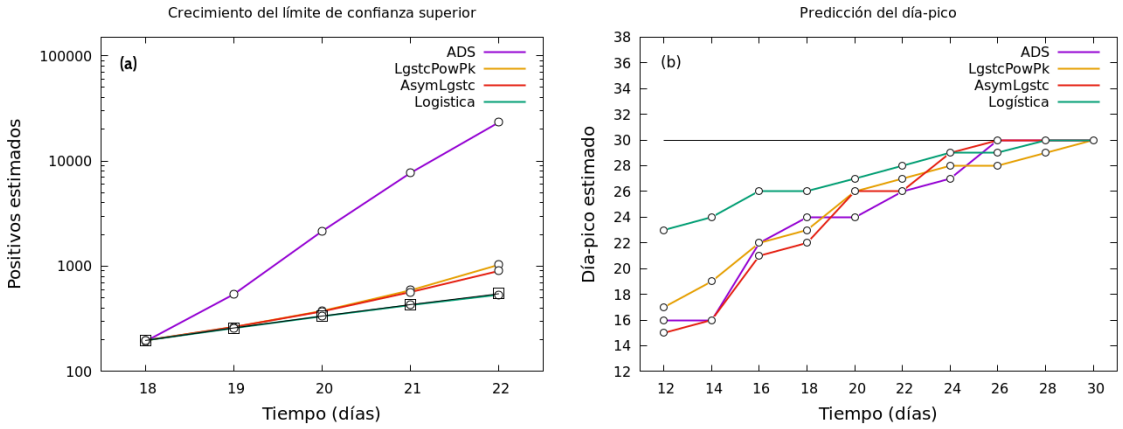

Figura 11: Cuatro funciones que se ajustan al modelo SIR son usadas para explorar aspectos predecibles de la epidemia artificial. A la izquierda (a) vemos la evolución de los límites de confianza de la predicción más allá del día 18 y hasta el día 22. Los datos disponibles son hasta el día 18, Nótese que la gráfica tiene ejes semi-log. Lo cual indica la naturaleza no lineal de este crecimiento asociado a la incertidumbre en el tiempo. Los cuadrados son los datos verdaderos de la epidemia artificial. En (b), se grafican las cuatro funciones ajustadas y la evolución de su poder predictivo a lo largo del tiempo. La línea horizontal es el valor verdadero del día en que ocurre el pico máximo.

Por último exploraremos la capacidad de predicción cuando exploramos los datos acumulados de positivos. Esto lo representamos en la Figura 12 donde dos funciones con coeficientes de regresión altos se han ajustado. La curva de datos acumulados tiene la típica forma "S" y aunque no brinda información precisa sobre donde esta el pico, si brinda información sobre fechas importantes

como el fin de la epidemia y el número total de positivos al final de esta. Algo también importante a destacar es que los datos presentados de manera acumulada son menos "ruidosos" que la serie de los casos diarios. Las curvas ajustadas son la Sigmoide y la Gompertz (Ver Anexo B para más detalles). Ahora analicemos un escenario de posible predicción del número total de infectados haciendo uso únicamente de intervalos de datos incompletos, desde los 12 días hasta los 66 días de duración de la epidemia artificial. Este ejercicio se muestra en la Figura 13.

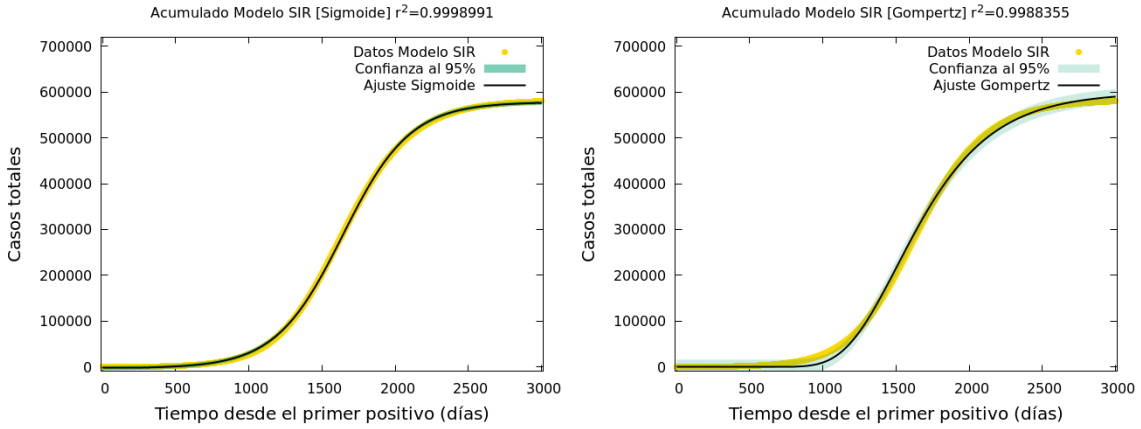

Figura 12: Dos funciones con coeficiente de regresión alto, ajustadas a los datos acumulados de positivos en una epidemia artificial tipo SIR.

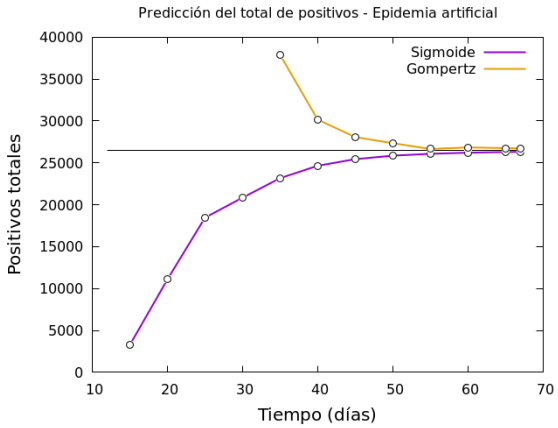

Figura 13: Dos funciones con coeficiente de regresión alto, ajustadas a los datos acumulados de positivos en una epidemia artificial tipo SIR para varios intervalos de tiempo con la intención de predecir el número total de positivos. La línea horizontal es el total de positivos al día 67 y son 26,533 individuos contabilizados. En el caso de la función Gompertz, los valores predichos se disparan de manera no lineal cuando el intervalo de datos considerados no es grande. Por ello su utilidad es limitada.

Un detalle resalta a los ojos de inmediato: en esta última gráfica una función se acerca por arriba y otra por debajo de manera no lineal. Y esto nos da la oportunidad de establecer un intervalo

de predicción temporal para los números de la epidemia artificial usando simultáneamente ambas funciones. Un hecho a destacar además es que la función Gompertz se dispara a valores absurdamente altos cuando se tienen pocos datos para la predicción. Como no se sabe de antemano, en una epidemia real, en que parte de la curva estamos, la función Gompertz tiene aparentemente un valor predictivo pobre. Otros trabajos reportados han hecho uso de los datos acumulados y funciones tipo "S" ajustadas a ellos para analizar su capacidad predictiva [12].

2.2. Una crítica a la capacidad predictiva de los modelos tipo SIR

Ya hemos visto en lo descrito anteriormente que los ajustes de funciones a las curvas de los modelos SIR tiene un valor predictivo cuestionable debido a que las incertidumbres de la estimación crecen de manera no lineal [10]. Pero los modelos tipo SIR o SEIR parametrizados sufren de los mismo ¿que camino nos queda, es mejor un método a otro? En primer lugar, funciones ajustadas requieren de parámetros que no tienen un significado biológico evidente. Son como valores de una caja negra. Se ajustan muy bien a los datos completos de una epidemia; pero son pobres en su capacidad de predicción. Pero la otra alternativa no es del todo mejor. Los errores e incertidumbres en la estimación de sus parámetros se traducirá en una capacidad pobre para estimar una epidemia pues se propagarán de manera no lineal como ya se mencionó, aunque son muy buenos para ajustar los datos de una epidemia ya madurada y casi a punto de terminar [13, 14]. Veamos por ejemplo el modelo tipo SIR desarrollado por el Gobierno de la Ciudad de México [9]. Este tiene nueve ecuaciones diferenciales acopladas y 12 parámetros a estimar. Se han hecho intentos por estudiar como los errores se propagan y esa práctica, ya común, es muy necesaria [15]. Pero son 12 parámetros con sus incertidumbres y aunque el modelo es el mejor camino, no está exento de problemas.

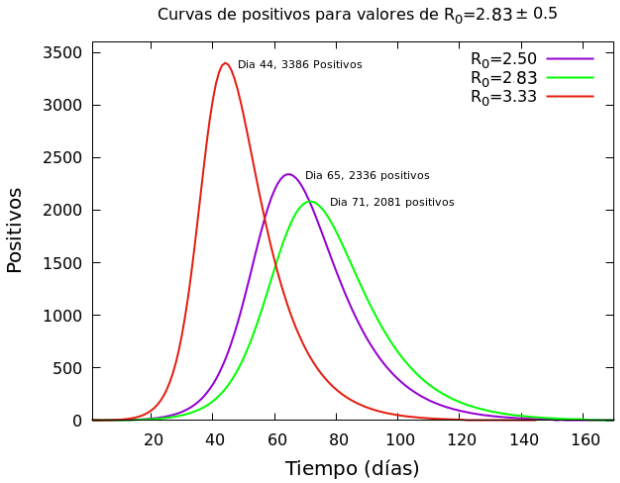

Figura 14: Curva de infectados (positivos) para el modelo SIR simple, para un valor de R_0 y su incertidumbre de ± 0.5. N=10,000.

Continuando con las incertidumbres de los modelos tipo SIR, la determinación del parámetro básico de reproducción de la enfermedad (R_0) es extremadamente importante; pero no es sencillo ni exacto. Varios estudios concluyeron que este valor, en promedio, era de 3.28; pero la incertidumbre es muy alta, inicialmente se había estimado el valor entre 1.5 y 6.68 [16]. En el caso de la Ciudad de México, los científicos encargados del modelo han estimado que el valor es de 2.83 [9] usando datos

proporcionados por el INDRE, antes del 22 de marzo. Las medidas de intervención como distancia social alteran el valor de R_0, una complicación adicional. Como ejemplo de la relación no lineal de los efectos de los errores en la determinación del valor de R_0 veamos la Figura 14 donde hemos graficado el valor 2.83±0.5, es decir entre [2.33 y 3.33], veamos como esta variación afecta el pico de infectados en el modelo SIR simple.

La incertidumbre en la determinación del parámetro R_0 se propaga de manera no lineal y no es proporcional al tamaño del error. La epidemia hipotética tiene un R_0 de 2.5. Con este valor, el pico de máximo contagio ocurre el día 65 con 2336 infectados. En el límite inferior de la incertidumbre, con R_0=2.33, el máximo contagio ocurre el día 71, una semana después y hay 2081 contagiados. Para el límite superior, cuando R_0=3.33, el máximo contagio es el día 44, 21 días antes y con 3386 infectados, que son más de mil casos adicionales. Es común entonces que a estos casos extremos se les llame "escenario optimista" y "peor caso", respectivamente. A pesar de lo dicho anteriormente intentaremos, como ejercicio educativo, hacer estimaciones en el caso del COVID-19 en México, haciendo uso del ajuste de funciones con datos parciales (también conocido como data-driven).

2.3. Evolución de casos positivos detectados en México

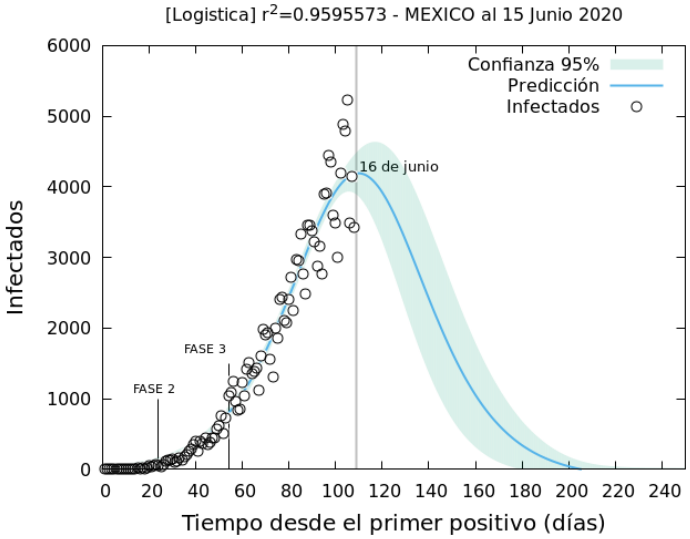

Figura 15: Función logística ajustada a la curva de evolución del número diario de positivos detectados en México. La línea vertical en gris marca la fecha predicha del pico: 16 de Junio, con datos reales hasta el 15 de junio. Los límites de confianza al 95 % marcan las fechas del pico en el intervalo [11 de junio - 21 de junio]

Con los reportes técnicos diarios proporcionados por la Secretaría de Salud de México podemos repetir el ejercicio que hicimos anteriormente para la ciudad de New York durante la pandemia de 1918. Este ejercicio lo presentamos en una versión anterior de este documento [17] y donde ajustamos cuatro funciones pico y discutimos sus bondades para el ajuste. A continuación nos limitamos a la función [Logística] con datos hasta el 15 de junio. La función logística parece tener mejor capacidad de predecir pues sus límites de confianza son pequeños comparados con las otras funciones exploradas. Anteriormente se había predicho el pico alrededor del día 27 de junio. En

esta actualización, el pico se ha corrido hacia el 16 de junio, siempre y cuando la población en general respete la petición para permanecer en casa, como lo veremos más adelante.

Dado que al 15 de junio es el día 109 desde el primer caso reportado en México, se ha acumulado suficiente tiempo de predicciones como para probar la robustez de nuestro método. Hemos visto que el pico de una epidemia artificial puede predecirse correctamente con cierta anticipación –aunque no mucha– y que la forma que adopta el patrón de predicciones (ver Figura 11b) es inicialmente casi linear y luego se acuesta en la cercanía del valor del pico. Este patrón es el de una función de transición y un buen ajuste lo da la función sigmoide. Podemos entonces analizar si el patrón de fechas de pico predichas se ajusta también a un patrón sigmoidal y lo vemos en la figura 16. Donde el valor asintótico es el 16 de junio (intervalo de confianza entre el 12 y 18 de junio). Recordemos que esto es válido únicamente si se mantiene el confinamiento. Sin embargo los eventos recientes en Guadalajara no son esperanzadores.

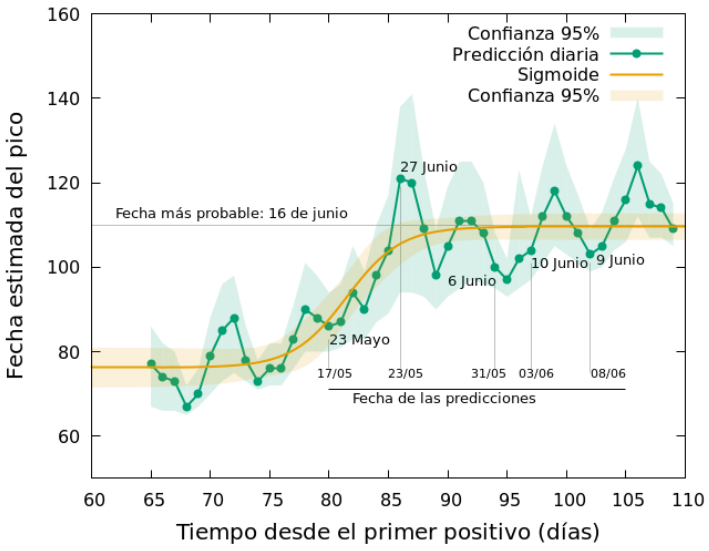

Figura 16: Función sigmoide ajustada a las fluctuaciones en la fecha del pico predicho. Este método sugiere como fecha más probable para el pico, el día 16 de junio, que coincide con la predicción que hemos hecho anteriormente.

Para no quedarnos con la curiosidad, repitamos aquí la predicción hecha por la Secretaría de Salud de México del día pico al 17 de abril, usando las funciones ajustadas. Esto se aprecia en la Figura 17 donde hemos graficado las cuatro funciones sobrepuestas a los datos conocidos. Tenemos resultados dispares.

Función	Día pico estimado
ADS	-
LgstcPowPk	7 mayo
AsymLgstc	23 mayo
Logística	28 abril

Tabla 3: Cuatro funciones ajustadas y los días picos predichos.

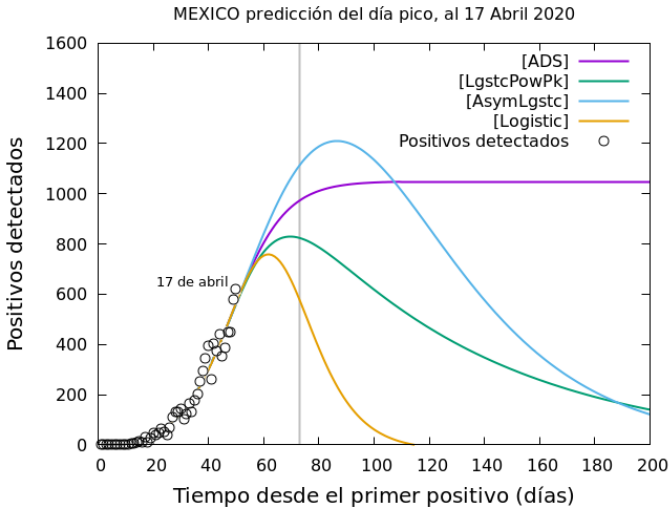

Figura 17: Cuatro funciones ajustadas a la curva de evolución del número de positivos detectados en México, con datos hasta el 17 de abril. La línea gris vertical marca el 10 de mayo.

De entrada, la función [ADS] no converge en un tiempo razonable y por ello la descartamos de inmediato. Las otras tres funciones predicen el día pico en diferentes fechas con una diferencia entre ellas de casi un mes [28 abril- 23 mayo](ver Tabla 3). El punto medio en este intervalo es justamente el 10 de mayo. Pero esta es una conclusión más bien *ad hoc*. La conclusión verdadera es que el método de "data-driven" (por ejemplo [11]) es pobre para estimar con horizontes de tiempo grandes.

Sin embargo, ilustremos otra predicción: el saldo total de positivos detectados hacia el final de la epidemia con los datos disponibles hasta el 15 de junio. En un ejercicio anterior [17], al 13 de mayo observamos que hacia el fin de la epidemia el número total de casos predichos se acercaba asintóticamente al valor de 91,000 individuos positivos, es decir alrededor de 9,100 defunciones; sin embargo, la falta a la petición de frenar la movilidad ha producido un aumento en el número de casos totales predichos. Ahora, al 15 de junio podemos predecir alrededor de 259,000 los casos totales. Nótese que hemos dejado de fuera la función Gompertz pues esta, como se ilustró en la Figura 13, no es de utilidad cuando los datos disponibles son limitados.

Recientemente (13 de mayo) el Gobierno de España ha dado a conocer que el porcentaje estimado significativo de la población, en territorio español, que resultó infectada y se mostró asintomática o con síntomas leves es del 5 % de la población total [18] (población total es 46,754,778 al mes de mayo de 2020) es decir alrededor de 2,337,739 casos. Se trata de una cifra muy baja para intentar un escenario de "inmunidad por manada"; pero la cifra en sí es reveladora pues muestra que los casos "ocultos" es un porcentaje muy bajo, al contrario de lo que los detractores señalaban sin pruebas. Al 13 de mayo, España contaba con 272,646 casos detectados, entonces el número de "ocultos" es alrededor de 10 veces esta cifra. Un estudio de la Universidad Politécnica de Cataluña había estimado este porcentaje desde el 29 de abril [19]. En el caso de México, que inició las acciones de mitigación y pudo controlar de mejor manera la expansión del contagio, la cifra de casos "ocultos" no detectados posiblemente sea similar o menor al 5 % de la población total. Las autoridades sanitarias en México han estimado (9 de abril) el número de casos "ocultos" en un factor de 10 o 12 de los casos detectados [20]. Tomando en cuenta lo anterior, hacia finales de la epidemia, el número

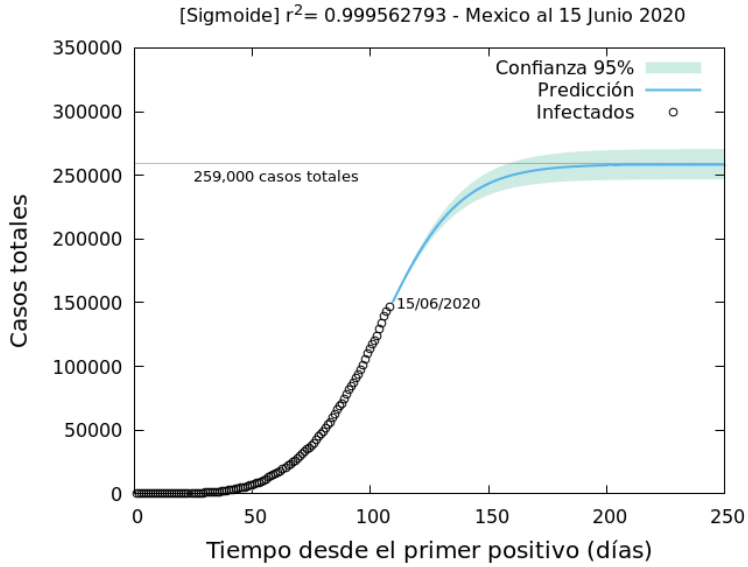

Figura 18: Predicción del número total de casos al finalizar la epidemia hacia el mes de septiembre del 2020. Para entonces se habrá alcanzado la cifra de 259,000 casos totales medidos (línea gris). Predicción con datos disponibles hasta el 15 de junio. La línea gris horizontal marca el número total predicho.

de total de infectados en México se aproximará, posiblemente, a los 2,590,000 casos totales entre asintomáticos y sintomáticos. En un último ejercicio, sabemos que la población de México al 17 de mayo del 2020 es 128,932,800 individuos aproximadamente, entonces 2,590,000 representa alrededor del 2.0 % de la población. Esta es una cifra muy inferior a la de España y tampoco permitiría en México implementar una estrategia de inmunidad grupal.

País o región	Positivos prom (%)	No. Pruebas	Población total
España [18]	5	60,897	46,754,778
Madrid Comunidad [18]	11.3	3,185	6,661,949
Cataluña [18]	5.9	6,318	7,619,494
Andalucia [18]	2.7	9,726	8,476,718
Castilla-La Mancha [18]	10.8	5,066	2,106,331
Francia [21]	4.4	-	67,028,048
Ile-de-France (París) [21]	9.9	-	12,174,880
Los Angeles County [22]	4.3	863	10,039,107
Santa Clara County [23]	2.8	3,330	1,927,852

Tabla 4: Resultados de pruebas de prevalencia de inmunidad en diversas regiones del mundo.

2.4. Pruebas masivas en México

Se ha hecho un eco muy grande en la prensa acerca de la necesidad de hacer pruebas masivas en México, a manera en como lo han hecho países del primer mundo. Veamos en la tabla 4 algunos

de esos ejercicios, llamados de prevalencia de anticuerpos (pruebas rápidas) y veamos si la estimación de un factor de entre 10 y 12 de asintomáticos para México por cada contagiado detectado, anunciada por el Subsecretario Hugo López-Gatell, es razonable. En primer lugar notemos que en la tabla aparecen regiones del mundo que fueron severamente afectadas por el virus, como es el caso de España, Francia y EEUU. Tengamos en cuenta que el número de asintomáticos es proporcional a lo severo de la propagación del patógeno. Dado que México ha resultado, a la fecha 15 de junio, con una epidemia mucho más favorable que estas regiones, es de esperarse que el número de asintomáticos en México sea menor que la de las regiones severamente afectadas. El promedio de los porcentajes mostrados en la tabla 4 es de apenas 4.1 % (no se contabiliza las regiones internas de Francia y España para no repetir). Otro estudio concluye que el promedio de varias regiones en Europa y EEUU es de 11 % (incluyendo el caso de la región más afectada del mundo: New York) [24]. Concluimos con facilidad que el número de asintomáticos (los tan llamados casos "ocultos") es de entre 4.1 y 11 %, en casos de propagación severos. No existe ningún argumento científico que nos lleve a pensar que el número de asintomáticos en México es un factor de 30, como se ha manejado por parte de la prensa sensacionalista. Para que eso ocurriera, la severidad de la epidemia en México debería ser aún en proporciones mayores a lo que se vivió en New York, con un dramático colapso del sistema de salud y eso claramente no es el caso. Independientemente de esta opinión anterior, es muy importante que en México se concluya el estudio de prevalencia de anticuerpos y se de a conocer el resultado de esas pruebas, aún cuando las cifras estimadas por la Secretaría de Salud de México parezcan razonables [20].

2.5. Aplanar la curva, movilidad y contacto social

En la actualidad se han instrumentado varias medidas de intervención para mitigar la epidemia en México, lo que se ha dado por llamar "aplanar la curva" [4].

- Escuelas y universidades cerradas.
- Labores no esenciales suspendidas.
- Práctica de "Susana distancia".
- Sugerencia de uso de cubreboca e higiene profunda.
- Confinamiento voluntario en casa.

De todas las medidas, quizás la más efectiva es la del confinamiento voluntario; sin embargo, los datos de movilidad que proporcionan las empresas que registran esos datos [25], indican que la movilidad en México (entre el 16 de abril y el 03 de junio) no se ha detenido lo suficiente, de hecho se ha incrementado en comparación con otros países y esto es preocupante porque es la mejor manera de disminuir la tasa de contagio y evitar que la epidemia tenga una duración de hasta 6 meses con un incremento en la mortalidad (ver Tabla 5).

2.6. ¿Se ha aplanado la cuerva en México?

La respuesta es un rotundo sí. En México las acciones del gobierno federal y estatales han dado frutos sumamente positivos ¿Cómo sabemos que se ha "aplanado la curva"? Lo sabemos mediante dos métodos. Uno cualitativo y otro cuantitativo. Veamos. En la Figura 19 hacemos una comparación de los casos positivos diarios reportados por España, Estados Unidos y México. Escogemos estos tres países porque tienen una relación cercana con México y porque representan tres casos: (a) EEUU donde se tomaron medidas mínimas para mitigar la epidemia y actualmente encabeza

País	Recreación (%)	Alimentos y Farmacias (%)	Esparcimiento (%)
	16Ab / 9Ma / 16Ma / 3Ju	16Ab / 9Ma /16Ma /3Ju	16Ab / 9Ma / 16Ma / 3Ju
Colombia	-84 / -71 / -68 /-57	-65 / -46 / -43 /-29	-77 / -61 / -59 / -47
Argentina	-84 / -75 / -72 / -67	-34 / -20 / -19 / -18	-80 / -91 / -89 / -86
Chile	-75 / -58 / -68 / -63	-51 / -31 / -45 / -37	-74 / -63 / -69 / -63
Mexico	-66 / -56 / -55 / -52	-26 / -17 / -16 / -14	-54 / -51 / -49 / -45
Brasil	-59 / -47 / -55 / -47	-5 / +10 / -8 / +3	-59 / -52 / -56 / -43
Descripción	Restaurantes, bares, cafés	Mercados, mercerías, farmacias	Parques, playas, jardines

Tabla 5: Datos de disminución de la movilidad en ciertas categorías de lugares, según lo proporciona Google Inc., recopilando datos de los GPS de los teléfonos celulares Android.

el número de contagios y muertes en el mundo, (b) España donde se tomaron medidas de mitigación dramáticas (confinamiento obligatorio y vigilado por la policía); pero muy tardías y que en su momento estuvo entre los países europeos con el mayor número de casos y (c) México donde se establecieron estrategias de mitigación tempranamente y se preparó la infraestructura sanitaria con anticipación. La comparación es clara, son tres casos donde hubo mitigación mínima, mitigación tardía y mitigación anticipada. Las pendientes de crecimiento de las curvas nos revelan que México (c) aplanó su curva de crecimiento de la epidemia de manera exitosa. Datos hasta el 25 de mayo (ver la Figura 2).

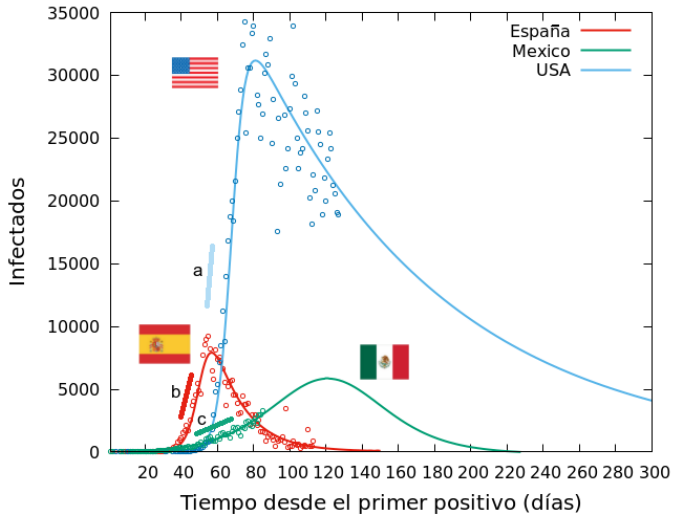

Figura 19: Tres pendiente para el crecimiento de una curva de infecciones. México (c) tiene, comparativamente a EEUU (a) y España (b), la menor de las pendientes y eso es resultado de un programa exitoso de mitigación. En México sí se ha logrado un aplanado de la curva. Datos hasta el 25 de mayo. Las líneas continuas son las curvas estimadas.

Como evidencia cuantitativa tenemos la medida histórica del parámetro de crecimiento R_0 de la epidemia. Como ya lo mencionamos, es un parámetro sumamente difícil de estimar; pero es de gran importancia (aquí usamos el método propuesto en [26]). Las medidas de confinamiento y distancia social están destinadas a reducir los contagios y con ello a reducir el crecimiento de R_0

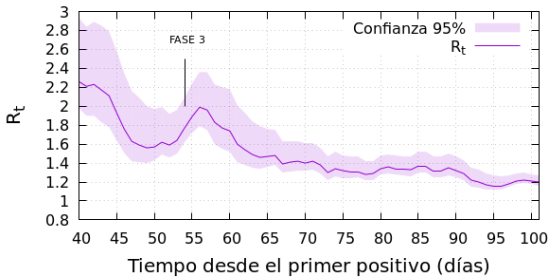

Figura 20: Evolución temporal del parámetro R_t. Se ha mantenido una tendencia consistente a la baja rápidamente y despues de la declaración de la Fase 3, una bajada lenta hacia el valor R_t=1. A partir del 25 de mayo, aparentemente, el valor es muy cercano a uno, lo cual significa que la epidemia estaría entrando en una meseta de desaceleración, es decir la cercanía del pico de máximos contagios (pero es necesario ser cauto y esperar más días con más datos para confirmar este valor de R_t<1). La evolución decreciente del valor de R_t significa un aplanado exitoso de la curva.

(a partir de ahora le llamaremos R_t pues varía en el tiempo). La evolución decreciente del valor de R_t es señal de un aplanado exitoso de la curva de infecciones que es el resultado, a su vez, de una baja movilidad. En ausencia de instrumentos farmacéuticos, el confinamiento sigue siendo la única solución y debe mantenerse lo máximo posible para aplanar la curva. Una consecuencia muy buena de este aplanado es que los servicios sanitarios en México nunca llegaron a un punto de saturación y la disponibilidad de camas y salas de UTI ha sido constante, muy por lo contrario con los escenarios vistos en los EEUU y España. Como ilustración de lo anterior, veamos la figura 21 donde se reporta la disponibilidad de camas en los hospitales de México al 14 de junio, es decir en la víspera del pico de la epidemia. Afortunadamente no existe saturación y ello es una muestra de un manejo adecuado de la epidemia.

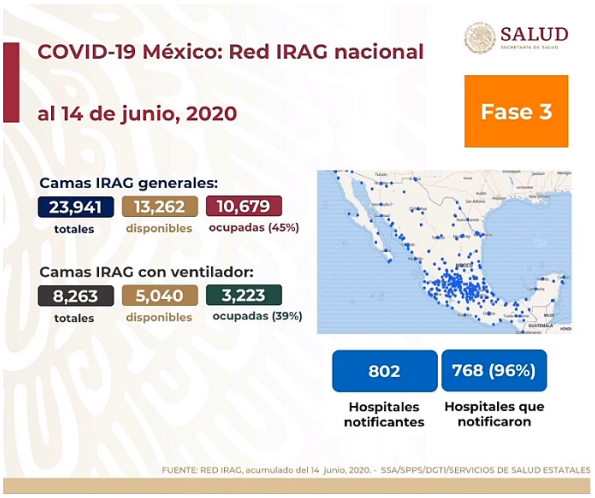

Figura 21: Camas disponibles y Unidades de Cuidado Intensivo, reportadas por la Secretaría de Salud de México, al 14 de junio.

La medición precisa de movilidad humana es un hecho posible gracias a las tecnologías de satélites y de telefonía celular que, sin embargo, hasta ahora se han incorporado de manera muy marginal a los modelos epidemiológicos. En la siguiente sección revisaremos este punto importante; pues es el futuro de los modelos epidemiológicos y el monitoreo de una epidemia.

3. Modelos epidemiológicos modernos

Los modelos epidemiológicos Kermack-McKendrick, pese a su "antigüedad" son modelos vigente y en uso común. Después de los modelos SIS y SIR, se han realizado extensiones a otras versiones más realistas que, básicamente, consisten de extender el número y tipo de compartimentos en los que la población puede clasificarse. Así, por ejemplo, el modelo SEIR agrega la categoría "Expuestos" y el modelo agrega en consecuencia una ecuación más al formalismo SIR. El modelo SEIS, elimina la categoría de recuperados debido a que podría no existir inmunidad y los individuos pueden infectarse de inmediato. El modelo SIQR agrega la categoría de individuos en cuarentena ("Quarentine") y así podemos imaginar el agregar más y más categorías, Si bien todas estas extensiones del modelo original tienen una utilidad, se reconoce que pueden ser aún mas realistas si, por ejemplo, los conjuntos de individuos no son homogéneos como sucede en la realidad y para ello es necesario, por ejemplo, introducir el espacio explicito para estudiar fenómenos como la percolación y su efecto en la propagación espacial de infecciones [27].

Los primeros modelos de infecciones con espacio utilizaron redes de autómatas con dinámica SIR [28], también se han empleado modelos con formalismos de mapeos acoplados [29, 30]. A medida que los modelos fueron incorporando teoría de redes complejas, el grado de realismo fue incrementándose [31], hasta llegar a modelos basados en agentes [31], otros incorporando movilidad explicita [32] y finalmente efecto de mundo pequeño [33]. Este último modelo lo detallaremos a continuación, a manera de ejemplo.

Un turistas alemán de visita en La Gomera, Canarias, resultó positivo el 31 de enero de 2020 a una prueba de COVID-19, iniciándose con ello el registro del virus en territorio español, como se mencionó anteriormente. Entre Europa y Wuhan en China, foco de la infección, hay una distancia de unos 10,000 kilómetros que un vuelo comercial puede recorrer en unas 24 horas. El que sea suficiente una sola persona para transportar una enfermad infecciosa a distancias continentales e iniciar una pandemia de proporciones globales, es un factor que los modelos epidemiológicos toman en cuenta; pero no siempre fue así. El primer modelo epidemiológico que incorporó este ingrediente –que en teoría de redes complejas se conoce como efecto de mundo pequeño– además de una dinámica de movilidad y contactos sociales fue desarrollado en 2002 por Miramontes y Luque [33].

Este modelo incorpora el espacio explícitamente, es decir, además de una dinámica en el tiempo, como en el modelo Kermack-McKendrick, se toma en cuenta que la epidemia se propaga en el espacio. Para ello se introduce una cuadrícula discreta bidimensional en la que se puede fijar, por ejemplo, una población; pero se puede tener una resolución aún más fina y definir individuos directamente. En un modelo más sencillo, como el que discutimos, los individuos tienen varios atributos, por ejemplo su localización en todo momento y su estado de salud que puede ser sano y susceptible (S) o infectado (I), por ello este tipo de modelos simples y compartimentalizados se llaman modelos SIS. Los individuos pueden moverse aleatoriamente en su vecindad a los cuadros de la cuadrícula que están a su alrededor y que estén vacíos (ver la Figura 22). Este tipo de movimiento se conoce como caminatas aleatorias y existe una teoría matemática muy desarrollada que explica sus propiedades, como por ejemplo, el llamado tiempo de primer encuentro, que nos da la probabilidad media de que dos individuos cualesquiera se encuentren eventualmente.

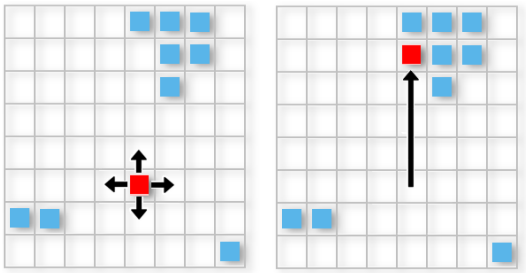

Figura 22: En la cuadrícula de la izquierda se representa el espacio bidimensional donde se mueven los individuos (agentes) y en donde se propaga la infección. Existen varios individuos sanos (en azul) y un individuo infectado (en rojo). Los individuos pueden moverse aleatoriamente a los cuadros adyacentes en cada unidad de tiempo. A la derecha, el individuo infectado ha dado un salto de larga distancia (un atajo, como si hubiera tomado un tren o un avión) y por ello será capaz de infectar rápidamente a los individuos sanos en el grupo que se localiza en la zona superior de la cuadrícula.

Los individuos susceptibles se pueden enfermar al tener contacto con un infectado. Existe un tiempo de recuperación τ después del cual los infectados se curan y pasan al estado de susceptibles.

Dado que es un modelo con espacio explicito, es posible hablar de la densidad de los contactos sociales en ese espacio; pero de manera importante, es posible tener en cuenta que cada individuos tiene una comunidad social con la que interactúan con mayor frecuencia (su familia, sus amigos) y que esta red de interacciones sociales tiene una estructura muy especial que la convierte en lo que conocemos como red compleja: algunos pocos individuos tienen una comunidad social muy grande, mientras que la mayoría tiene menos o pocos contactos. Además, un individuo puede tener lazos muy fuertes en su comunidad local; pero puede de pronto tener conexión con algún otro individuo distante en una comunicad lejana. Este fenómeno de los atajos sociales en redes se conoce, como lo mencionamos anteriormente, como efecto de mundo pequeño y puede agregarse en el modelo que describimos, como una probabilidad p de que un individuo se mueva a cualquier sitio de la red en un paso de tiempo (ver la Figura 22). El resultado principal de este modelo simple nos dice que existe una densidad critica de individuos infectados δ (percolación) por abajo de la cual una epidemia se extingue y por arriba de la cual se da persistencia. El valor de esta densidad crítica disminuye si se permite a los individuos viajar grandes distancias, bastando muy pocos viajeros infectados para hacer que la epidemia se propague a todo el espacio y persista por siempre (ver la Figura 23).

Si la densidad poblacional es baja y por debajo de a densidad crítica, aunque la probabilidad de viajes largos sea grande, el modelo predice correctamente que la epidemia se extingue por falta de contactos sociales por ello es muy importante las estrategias de distanciamiento social y cuarentena de los enfermos, es decir que la movilidad sea la menor posible (ver la Figura 24).

Un modelo pionero y sencillo como el que hemos discutido hace suposiciones básicas que son en la actualidad superadas por modelos de agentes más realistas y de gran escala. Por ejemplo, la cuadrícula espacial es una capa sobre un territorio de dimensiones muy grandes donde se representan densidades de población que vienen de datos del Censo de Población y por ello el número de individuos se asemeja a la realidad. Además, se pueden incorporar diversas hipótesis sobre el movimiento de tal manera que no sea del todo aleatorio sino dirigido sobre rutas de transporte como calles, metro, carreteras, vías de tren o rutas aéreas. Una adición más reciente, que se incorpora a los modelos epidemiológicos de agentes, es el monitoreo anónimo y agregado de datos de des-

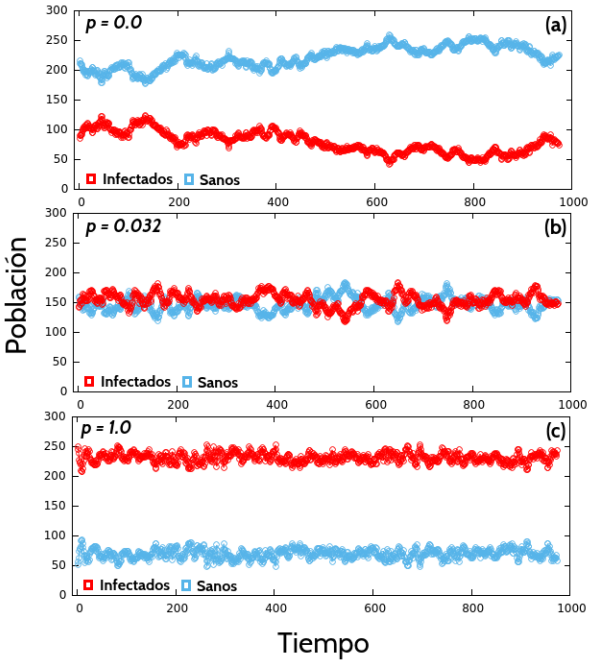

Figura 23: Dinámica de una epidemia donde la densidad de población (conectividad social) esta por arriba de la densidad crítica y por tanto la infección existe por siempre, como en el caso de las gripes comunes en el mundo real. La población se divide en susceptibles (S, en azul) e infectados (I, en rojo). (a) La dinámica de movilidad se restringe a movilidad local (p=0) y con ello la población de infectados se mantiene siempre; pero es una porción mínima de la población, los sanos son mayoría. (b) En este escenario, con una probabilidad de saltos de larga distancia tan baja como p=0.032, las poblaciones de sanos e infectados se iguala. (c) El caso extremo donde todos los individuos pueden realizar viajes de larga distancia (p=1.0) se traduce en que la población de infectados supera a la población de sanos.

plazamiento de personas que son capturados en los dispositivos móviles de millones de usuarios alrededor del mundo.

A pesar de su simplicidad, este modelo de agentes no se equivoca en su predicción fundamental: disminuir la movilidad de los individuos para disminuir la tasa de infecciones cuando no existe inmunidad previa ni vacuna.

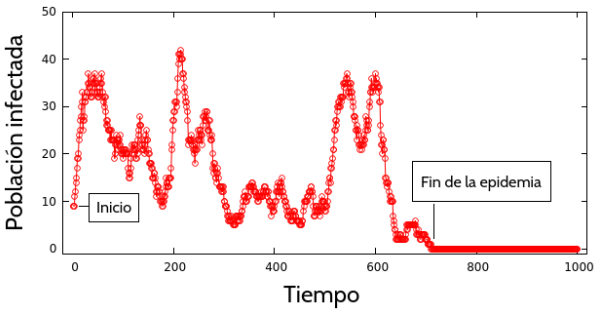

Figura 24: Un escenario de distanciamiento social donde la densidad de población en la cuadrícula del modelo esta por debajo de la densidad crítica y la probabilidad de saltos largos es grande (p=1.0). En este caso la densidad de contactos sociales es muy baja y por ello, después de un transitorio largo, caracterizado por fluctuaciones erráticas y de gran amplitud, la epidemia cede y se extingue. Nótese que estas fluctuaciones, en momentos, parecen acercar a la población hacia la extinción de a epidemia; pero luego se esta se reactiva nuevamente debido a que la distribución de los individuos en el espacio no es homogénea y existen procesos de contagio locales que pueden amplificarse repetidamente. De hecho, este fenómeno de rebrote (con un confinamiento laxo) de una epidemia puede verse claramente en la gráfica de los casos en Londres durante la pandemia de 1918.

4. Conclusiones

La pandemia COVID-19 parece que ha tomado al mundo por sorpresa. A pesar de que existen formalismos matemáticos muy maduros para predecir y acompañar pandemias, el mundo asiste perplejo a la muerte de cientos de miles y a una de las mayores crisis económicas de que se tenga memoria. De manera sigilosa los dirigentes políticos, con algunas pocas excepciones, guiados por el paradigma del libre mercado, se han avocado a desmantelar durante los útimos 30 años el sistema de salud pública. Adicionalmente se ha transferido la producción y distribución de resultados científicos producidos por instituciones públicas a unas pocas empresas monopólicas que los guardan bajo "paywalls". Por si fuera poco, algunos dirigentes de naciones poderosas (notablemente, los EEUU, Gran Bretaña y Brasil; por cierto son los países que encabezan la tasa de muertes) se caracterizan por tener una actitud despectiva hacia la opinión de sus científicos y por lo tanto incorporan a regañadientes algunas pocas de las recomendaciones emitidas por expertos, en una actitud que bien podría ser llamada de genocida. Frente a esto, varias instituciones científicas y educativas del mundo entero han hecho un llamado a sus comunidades científicas locales para dedicar esfuerzos a los múltiples aspectos de la pandemia actual. El presente estudio nace a raíz de este llamado. No tanto como una opinión a nivel de experto; pero si como un esfuerzo para alentar a los estudiantes de nivel licenciatura a acercarse a la matemática y modelado básico de la propagación de enfermedades.

Como conclusiones de este proyecto, podemos mencionar que los modelos tipo SIR y derivados (SEIR, SQIR, etc.) son modelos vigentes para explicar los aspectos cualitativos de una epidemia; pero hemos visto que también pueden ser útiles para iluminar aspectos cuantitativos. Un esfuerzo mayor debe hacerse para desarrollar una nueva generación de modelos que incorporen movilidad y conducta social a nivel individual.

En México, afortunadamente existe una comunidad de expertos epidemiólogos que han diseñado una estrategia para enfrentar los estragos de la epidemia y hacer recomendaciones exitosas

de política pública. La epidemia en México se encuentra en crecimiento y cercano a la zona del pico de la curva. Este pico máximo de positivos detectados deberá ocurrir entre el 12 de junio y el 18 de junio, con fecha más probable el 16 de junio. Se espera que la epidemia tenga una duración de entre 160 y 250 días, de mantenerse el mismo ritmo de confinamiento (desafortunadamente la celebración del 10 de mayo ha causado un verdadero estrago a los esfuerzos por mitigar la propagación del virus). Nuestro ejercicio estima en alrededor de 259,000 el número total de positivos identificados y alrededor de 25,900 el número de defunciones al termino de la enfermedad, entre agosto y octubre. Conjeturamos que la población inmunizada en México hacía finales de octubre será un número reducido cercano al 2.0 % lo cual no permitirá la práctica de "inmunidad de rebaño". Este ejercicio, exclusivamente para fines docentes, considera los resultados de datos reportados por las autoridades de la Secretaría de Salud hasta el 15 de junio del 2020.

5. Anexo A - Programación de los modelos SIR y SEIR

En este trabajo hemos usado el siguiente código Python [34] para estudiar numéricamente el modelo Kermack-McKendrick (SIR). Los datos numéricos obtenidos con las variaciones de las poblaciones S,I y R se escriben en disco y se grafican los resultados:

```
# El modelo Kermack-McKendrick (SIR)
# cargar las bibliotecas numericas y de graficacion
import numpy as np
import matplotlib
import matplotlib.pyplot as plt

# definicion del metodo de Euler
def km_model(init_vals, params, t):
    S_0, I_0, R_0 = init_vals
    S, I, R = [S_0], [I_0], [R_0]
    l, k = params
    deltat = t[2] - t[1] # Pasos de tiempo constantes
    for t_ in t[:-1]:
        next_S = S[-1] - (k * S[-1] * I[-1]) * deltat
        next_I = I[-1] + (k * S[-1] * I[-1] - l * I[-1]) * deltat
        next_R = R[-1] + (l * next_I) * deltat
        S.append(next_S)
        I.append(next_I)
        R.append(next_R)
    return np.stack([S,I,R]).T

# definicion de valores iniciales y parametros
# init_vals[S(0),I(0),R(0)]
init_vals = [9999, 1, 0]
params = [1e-1, 0.00002]
t_max = 300
deltat = 0.1
t = np.linspace(0, t_max, int(t_max/deltat))
km_results = km_model(init_vals, params, t)

# Escritura de valores a disco
outF = open("datos.txt", "w")
for line in km_results:
  print >>outF, line
outF.close()
print('Los valores de las poblaciones S,I,R, se han escrito en Disco')
print('Ver el archivo datos.text')

# Grafica de resultados
plt.figure(figsize=(8,6))
plt.plot(km_results, linewidth=3.3)
plt.legend(['S(t)-Susceptibles', 'I(t)-Infectados', 'R(t)-Recuperados'])
plt.xlabel('Tiempo',fontsize=18)
plt.ylabel('Poblacion',fontsize=18)
plt.show()
```

Un modelo posterior conocido SEIR, agrega un tipo nuevo de población, los "Expuestos" y es usado ampliamente para simular epidemias en las que existe un tiempo de latencia entre la exposición al patógeno y se esta usando ampliamente para simular la COVID-19 [35–37]. A continuación

presentamos un código Python para su simulación [38]. Basta decir, para nuestro propósito, que la curva de infectados en el modelo SEIR se ajusta perfectamente a las funciones tipo campana [ADS][LgstcPowPk] y [AsymLgstc]; entonces nuestro ejercicio con el modelo SIR es pertinente.

```
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import sys

def base_seir_model(init_vals, params, t):
    S_0, E_0, I_0, R_0 = init_vals
    S, E, I, R = [S_0], [E_0], [I_0], [R_0]
    alpha, beta, gamma = params
    dt = t[1] - t[0]
    for _ in t[1:]:
        next_S = S[-1] - (beta*S[-1]*I[-1])*dt
        next_E = E[-1] + (beta*S[-1]*I[-1] - alpha*E[-1])*dt
        next_I = I[-1] + (alpha*E[-1] - gamma*I[-1])*dt
        next_R = R[-1] + (gamma*I[-1])*dt
        S.append(next_S)
        E.append(next_E)
        I.append(next_I)
        R.append(next_R)
    return np.stack([S, E, I, R]).T

# Definir parametros
N = 10000 # poblacion total
init_vals = [1 - 0.0002, 0.0001, 0.0001, 0] #poblacion normalizada

alpha = 0.2
beta = 1.75
gamma = 0.5
params = alpha, beta, gamma
t_max = 100
dt = 0.1
t = np.linspace(0, t_max, int(t_max/dt)+1)

# Simulacion
results = base_seir_model(init_vals, params, t)
print(results)
print('R_0=', beta/gamma)
outF = open("datos.txt", "w")
for line in results:
  print >>outF, line
outF.close()

# Resultados
plt.figure(figsize=(8,6))
plt.plot(N*results, linewidth=3.3)
plt.legend(['S(t)', 'E(t)', 'I(t)', 'R(t)'])
plt.xlabel('Tiempo',fontsize=18)
plt.ylabel('Poblacion',fontsize=18)
plt.show()
```

6. Anexo B - Ajuste de funciones a los datos numéricos del modelo SIR y SEIR

El ajuste de funciones a datos numéricos es un proceso siempre polémico. Dada una colección de datos, el número de funciones arbitrarias que pueden ajustarse es virtualmente ilimitado. Un criterio simple y aceptado para seleccionar funciones es calcular un Coeficiente de Regresión (Coeficiente de determinación) o simplemente r^2, el cual nos da una medida de la bondad del ajuste. Desde luego que existen métodos estadísticos más sofisticados, como el Criterio de Información de Akaike [39] ; pero nos limitaremos al uso de r^2. No discutiremos este tema aquí. Basta mencionar que el universo de funciones que pueden ser ajustadas sólo tiene sentido porque conocemos las ecuaciones diferenciales que describen el fenómeno. Es decir tenemos una hipótesis atrás. Con esto en mente, sabemos además que la forma funcional de la evolución de la población de individuos infectados $I(t)$ es de tipo "pico" es decir, contiene un máximo. Esto ya nos dice de entrada que sólo funciones "pico" deben ser empleadas para el ajuste. Finalmente, la evolución temprana de los infectados tiene un crecimiento exponencial, entonces las funciones que ajustemos deben ser semejantes a aquellas empleadas en dinámica de poblaciones y que son tipo exponenciales y logísticas y sus combinaciones. El ajuste lo hemos hecho usando un ajuste no-lineal basado en el método de Levenberg–Marquardt [40] implementado en un software comercial llamado *TableCurve 2D* [41]. Una versión gratuita limitada esta disponible.

A raíz del ajuste a los datos extraídos numéricamente del modelo, hemos concluido, como se ha dicho en la tabla 1, que existen cuatro funciones "pico" con r^2 alto. El software típicamente puede ajustar más de 3000 funciones a un conjunto de datos, de las cuales 23 son funciones predefinidas tipo "pico". Estas son:

1) **[ADS]**. La Doble Sigmoide Asimétrica es una función pico con dos términos exponenciales que modulan una subida, un pico y una bajada y tiene seis grados de libertad (a,b,c,d,e,f):

$$I(t) = a + \frac{b}{1 + exp\left[-\frac{(t-c)+(d/2)}{e}\right]}\left[1 - \frac{1}{1 + exp\left[-\frac{(t-c)-(d/2)}{f}\right]}\right] \tag{5}$$

El ejercicio reportado en la Figura 3 arroja estos valores para los parámetros del ajuste: $a = 0{,}299686373; b = 4253{,}515262; c = 869{,}5777920; d = 92{,}55029820; e = 96{,}90866393; f = 166{,}2990276$.

Un código mínimo en C, para su implementación es el siguiente:

```
#include <math.h>

/*--------------------------------------------------------------------*/
double eqn(double x)
/*--------------------------------------------------------------------*
   r2=0.9999981080307787
   a= 0.2996863727186627
   b= 4253.515262130173
   c= 869.5777920315184
   d= 92.55029819798507
   e= 96.90866393425107
   f= 166.2990276147703
```

```
 *-------------------------------------------------------------*/
{
  double y;
  y=0.2996863727186627+(4253.515262130173/(1.0+exp(-(x-
    869.5777920315184+92.55029819798507/2.0)/
    96.90866393425107)))*(1.0-1.0/(1.0+exp(-(x-
    869.5777920315184-92.55029819798507/2.0)/
    166.2990276147703)));
  return(y);
}
```

2) **[LgstcPowPk]**. Es una función con varios elementos exponenciales y potenciales multiplicativos que modulan una subida, un pico y una bajada; poseé cinco grados de libertad: (a,b,c,d,e):

$$I(t) = a + b\left[(1+\eta)^{-(e+1)/e}\right]\left[\frac{\eta(e+1)^{(e+1)/e}}{e}\right], \tag{6}$$

$$\eta = exp\left[\frac{t + dln(e) - c}{d}\right] \tag{7}$$

El ejercicio reportado en la Figura 3 arroja estos valores para los parámetros del ajuste: $a = -3{,}061320838435103; b = 1550{,}033030201599; c = 913{,}1032581514686, d = 105{,}9574679234311, e = 1{,}757043252897661$.

Un código mínimo en C, para su implementación es el siguiente:

```
#include <math.h>

/*-------------------------------------------------------------*/
double eqn(double x)
/*-------------------------------------------------------------*
   r2=0.9999043421077948
   a= -3.061320838435103
   b= 1550.033030201599
   c= 913.1032581514686
   d= 105.9574679234311
   e= 1.757043252897661
 *-------------------------------------------------------------*/
{
  double y;
  double n;
  n=exp((x+105.9574679234311*log(1.757043252897661)-
    913.1032581514686)/105.9574679234311);
  y=-3.061320838435103+1550.033030201599*pow(1.0+n,-
    (1.757043252897661+1.0)/1.757043252897661)*n*
    pow(1.757043252897661+1.0,(1.757043252897661+
    1.0)/1.757043252897661)/1.757043252897661;
  return(y);
}
```

3) **[AsymLgstc]**. Es una función con varios elementos exponenciales y potenciales multiplicativos que modulan una subida, un pico y una bajada; poseé cinco grados de libertad: (a,b,c,d,e):

$$I(t) = a + b\eta(1+\eta)^{-(e+1)}\left[e^{-e}\right]\left[(e+1)^{(e+1)}\right], \tag{8}$$

$$\eta = exp\left[-\frac{t + dln(e) - c}{d}\right] \tag{9}$$

El ejercicio reportado en la Figura 3 arroja estos valores para los parámetros del ajuste:
$a = 4{,}449365019007865; b = 1520{,}439490107339; c = 914{,}2041176950709; d = 154{,}2941627736599; e = 2{,}11213143731454$.

Un código mínimo en C, para su implementación es el siguiente:

```
#include <math.h>

/*--------------------------------------------------------------*/
double eqn(double x)
/*--------------------------------------------------------------*
   a= 4.449365019007865
   b= 1520.439490107339
   c= 914.2041176950709
   d= 154.2941627736599
   e= 2.11213143731454
 *--------------------------------------------------------------*/
{
  double y;
  double n;
  n=exp(-(x+2.11213143731454*log(2.11213143731454)-
    914.2041176950709)/154.2941627736599);
  y=4.449365019007865+1520.439490107339*n*pow(1.0+n,-
    2.11213143731454-1.0)*pow(2.11213143731454,-
    2.11213143731454)*pow(2.11213143731454+
    1.0,2.11213143731454+1.0);
  return(y);
}
```

4) **[Logística]**. Es una función pico con elementos exponencial; poseé cuatro grados de libertad: (a,b,c,d):

$$I(t) = a + \left[\frac{4b\eta}{(1+\eta)^2}\right], \tag{10}$$

$$\eta = exp\left[\frac{-(t-c)}{d}\right] \tag{11}$$

El ejercicio reportado en la Figura 15 arroja estos valores para los parámetros del ajuste:

$a = -26{,}88202167068127; b = 1842{,}316635171811; c = 76{,}73240968039771; d = 12{,}11319149608016$

Un código mínimo en C, para su implementación es el siguiente:

```
#include <math.h>

/*----------------------------------------------------------------*/
double eqn(double x)
/*----------------------------------------------------------------*
   a= -26.88202167068127
   b= 1842.316635171811
   c= 76.73240968039771
   d= 12.11319149608016
 *----------------------------------------------------------------*/
{
  double y;
  double n;
  n=exp(-(x-76.73240968039771)/12.11319149608016);
  y=-26.88202167068127+1842.316635171811*4.0*n/((1.0+n)*(1.0+n));
  return(y);
}
```

5) **[Gompertz]**. Es una función de transición tipo "S" con doble exponencial; poseé tres grados de libertad: (a,b,c):

$$I(t) = a\left[e^{-b(e^{(-ct)})}\right] \tag{12}$$

El ejercicio reportado en la Figura 12 arroja estos valores para los parámetros del ajuste:
$a = 26802{,}75179791457; b = 81{,}90248438633768; c = 0{,}1547713377328515$

Un código mínimo en C, para su implementación es el siguiente:

```
#include <math.h>

/*----------------------------------------------------------------*/
double eqn(double x)
/*----------------------------------------------------------------*
   a= 26802.75179791457
   b= 81.90248438633768
   c= 0.1547713377328515
 *----------------------------------------------------------------*/
{
  double y;
  Y=a*exp(-b*exp(-c*x))
  return(y);
}
```

6) **[Sigmoide]**. Es una función de transición tipo "S" con elemento exponencial; poseé cuatro grados de libertad: (a,b,c,d):

$$I(t) = a + \frac{b}{1 + e^{-(t-c)/d}} \tag{13}$$

El ejercicio reportado en la Figura 12 arroja estos valores para los parámetros del ajuste:
$a = -250{,}983164174434$; $b = 26582{,}48388386695$; $c = 31{,}06711829127314$; $d = 4{,}378613596036086$

Un código mínimo en C, para su implementación es el siguiente:

```
#include <math.h>

/*----------------------------------------------------------------*/
double eqn(double x)
/*----------------------------------------------------------------*
   a= -250.983164174434
   b= 26582.48388386695
   c= 31.06711829127314
   d= 4.378613596036086
 *----------------------------------------------------------------*/
{
  double y;
  y=-250.983164174434+26582.48388386695/(1.0+exp(-(x-
    31.06711829127314)/4.378613596036086));
  return(y);
}
```

7. Anexo C - Extracción de datos de una gráfica

Extraer puntos de una gráfica es una tarea sencilla aunque tediosa. Consideremos la siguiente figura:

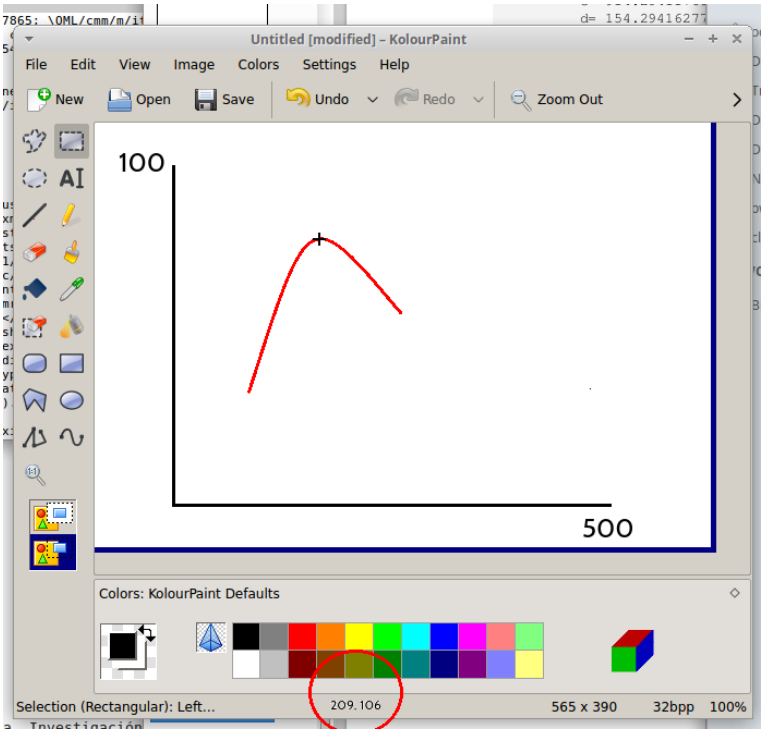

Figura 25: Ejemplo hipotético de un software que se emplea para extraer puntos de una gráfica que se encuentra en formato de imagen. En el ejemplo se ha usado KolourPaint en Linux; mismo resultado se obtiene con Paint para Windows.

En ella tenemos una gráfica hipotética en la que se muestra el curso (una cruz negra) sobre el pico de la gráfica. Abajo (círculo marcado en rojo) aparecen las coordenadas del punto (209,106) en el sistema coordenado del software (ver la Figura 25). Se anotan tantas coordenadas como se quiera. A continuación se ubican las coordenadas en el software de tres puntos que definen el sistema de coordenadas de la gráfica (0,100), (0,0) y (500,0). Haciendo una transformada lineal de intervalos, del sistema coordenado de la gráfica al sistema coordenado del software, se obtienen los valores reales de la curva. Sin embargo, como ya se mencionó, esto es sumamente tedioso y, afortunadamente existes programas gratuitos que harán esto de manera automática. Uno de ellos es DataThief III [42] que es el que hemos usado aquí (ver la Figura 26).

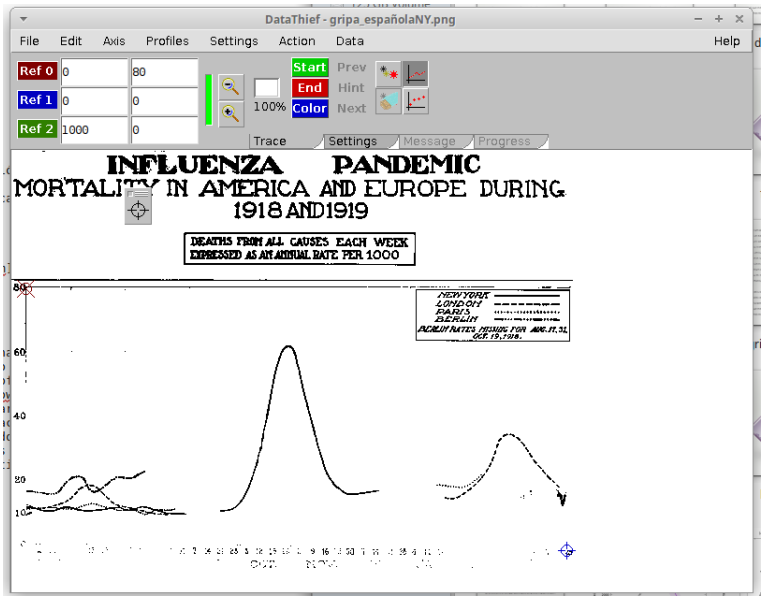

Figura 26: Ejemplo del uso de DataThieff III para obtener los puntos de la gráfica de la pandemia de 1918 en la ciudad de New York. En este caso hemos eliminado los datos de las otras curvas que se encimaban en la gráfica de NY para evitar errores. Luego de ello se obtienen automáticamente los puntos que describen la gráfica, en el sistema coordenado original.

Referencias

[1] Forster, P., Forster, L., Renfrew, C. & Forster, M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences* 202004999 (2020).

[2] Secretaría de Salud - México. Información sobre COVID-19. https://coronavirus.gob.mx/ (Accessed Abril 20, 2020).

[3] worldometer2020. COVID-19 coronavirus pandemic. https://www.worldometers.info/coronavirus/ (Accessed Abril 20, 2020).

[4] Luque, B., Ballesteros, F. & Miramontes, O. Cómo modelizar una pandemia. *Investigación y Ciencia* 48–57 (Mayo, 2020).

[5] Dietz, K. The first epidemic model: a historical note on PD En'ko. *Australian Journal of Statistics* **30**, 56–65 (1988).

[6] Cheng, V. C., Lau, S. K., Woo, P. C. & Yuen, K. Y. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clinical microbiology reviews* **20**, 660–694 (2007).

[7] Hellewell, J. *et al.* Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health* (2020).

[8] Secretaría de Salud - México. Dan a conocer integrantes del grupo científico que toma las decisiones ante Covid-19 en méxico. https://bit.ly/35uJGyJ (Accessed Mayo 04, 2020).

[9] Gobierno de la CDMX. Modelo epidemiológico Covid-19 del Gobierno de la Ciudad de México. https://modelo.covid19.cdmx.gob.mx/modelo-epidemico (Accessed Mayo 04, 2020).

[10] Castro, M., Ares, S., Cuesta, J. A. & Manrubia, S. Predictability: Can the turning point and end of an expanding epidemic be precisely forecast? https://arxiv.org/abs/2004.08842 (2020). 2004.08842.

[11] Singapore University of Technology and Design, Data-Driven Innovation Lab. Predictive monitoring of covid-19. https://ddi.sutd.edu.sg/ (Accessed Mayo 02, 2020).

[12] Ciufolini, I. & Paolozzi, A. Mathematical prediction of the time evolution of the COVID-19 pandemic in Italy by a Gauss error function and Monte Carlo simulations. *The European Physical Journal Plus* **135**, 355 (2020).

[13] Caccavo, D. Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model. *medRxiv* (2020).

[14] Fanelli, D. & Piazza, F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals* **134**, 109761 (2020).

[15] Mena, R. H. *et al.* Using the posterior predictive distribution to analyse epidemic models: COVID-19 in Mexico City. *arXiv preprint arXiv:2005.02294* (2020).

[16] Viceconte, G. & Petrosillo, N. l. *Infectious disease reports* **12** (2020).

[17] Miramontes, O. Entendamos el COVID-19 en México v1. http://scifunam.fisica.unam.mx:/mir/corona19/covid19v1.pdf (Accessed Mayo 18, 2020).

[18] MSCBS España. Los primeros resultados del ENECovid19 muestran que el 5 % de la población ha desarrollado anticuerpos frente a la enfermedad. https://www.mscbs.gob.es/gabinete/notasPrensa.do?id=4914 (Accessed mayo 10, 2020).

[19] Universidad Politécnica de Cataluña. En España se diagnostica al 10 % de los contagiados por Covid-19, según la UPC. https://rb.gy/jeprlq (Accessed mayo 13, 2020).

[20] López-Gatell, Hugo. López-Gatell: por cada caso positivo hay 10 o 12 más ocultos. https://rb.gy/fuh0bv (Accessed mayo 10, 2020).

[21] Salje, H. *et al.* Estimating the burden of SARS-CoV-2 in France. *Science* (2020).

[22] Sood, N. *et al.* Seroprevalence of SARS-CoV-2–specific antibodies among adults in Los Angeles County, California, on april 10-11, 2020. *Jama* (2020).

[23] Bendavid, E. *et al.* COVID-19 antibody seroprevalence in Santa Clara County, California. *MedRxiv* (2020).

[24] Wilson, L. Sars-cov-2, covid-19, infection fatality rate (ifr) implied by the serology, antibody, testing in new york city. *COVID-19, Infection Fatality Rate (IFR) Implied by the Serology, Antibody, Testing in New York City (May 1, 2020)* (2020).

[25] Google Inc. See how your community is moving around differently due to COVID-19. https://www.google.com/covid19/mobility/ (Accessed Abril 20, 2020).

[26] Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology* **178**, 1505–1512 (2013).

[27] Cardy, J. L. & Grassberger, P. Epidemic models and percolation. *Journal of Physics A: Mathematical and General* **18**, L267 (1985).

[28] Boccara, N. & Cheong, K. Automata network sir models for the spread of infectious diseases in populations of moving individuals. *Journal of Physics A: Mathematical and General* **25**, 2447 (1992).

[29] Fukś, H., Lawniczak, A. T. & Duchesne, R. Effects of population mixing on the spread of SIR epidemics. *The European Physical Journal B-Condensed Matter and Complex Systems* **50**, 209 (2006).

[30] Barrio, R. A., Varea, C., Govezensky, T. & José, M. V. Modeling the geographical spread of influenza A(H1N1): The case of Mexico. *Applied Mathematical Sciences* **7**, 2143–2176 (2013).

[31] Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Physical review letters* **86**, 3200 (2001).

[32] Sattenspiel, L., Dietz, K. *et al.* A structured epidemic model incorporating geographic mobility among regions. *Mathematical biosciences* **128**, 71–92 (1995).

[33] Miramontes, O. & Luque, B. Dynamical small-world behavior in an epidemical model of mobile individuals. *Physica D: Nonlinear Phenomena* **168**, 379–385 (2002).

[34] Hubbs, C. How quickly does an influenza epidemic grow? https://rb.gy/alqidc (Accessed Abril 18, 2020).

[35] World Health Organization. Bulletin of the World Health Organization. https://www.who.int/bulletin/online_first/COVID-19/en/ (Accessed Mayo 18, 2020).

[36] Goh, G. Epidemic calculator. http://gabgoh.github.io/COVID/index.html (Accessed Mayo 18, 2020).

[37] Neherlab. COVID-19 scenarios. https://covid19-scenarios.org/ (Accessed Mayo 18, 2020).

[38] DataHubbs. Social distancing to slow the coronavirus. https://www.datahubbs.com/social-distancing-to-slow-the-coronavirus/ (Accessed Mayo 18, 2020).

[39] Sakamoto, Y., Ishiguro, M. & Kitagawa, G. Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel* **81** (1986).

[40] Moré, J. J. The Levenberg-Marquardt algorithm: implementation and theory. In *Numerical analysis*, 105–116 (Springer, 1978).

[41] Systatsoftware. Tablecurve 2d. https://systatsoftware.com/products/tablecurve-2d/ (Accessed Abril 18, 2020).

[42] Tummers, B. Datathief III. 2006. https://datathief.org/ (Accessed Abril 20, 2020).