
Modelización matemática de la
propagación de enfermedades humanas.
Aplicación al caso de la reciente epidemia
de Ébola.
Estimación de parámetros.

Irene Serrano García

*Tutores: Benjamin Ivorra
Ángel Manuel Ramos*



6 de septiembre de 2015

Máster en Tratamiento Estadístico Computacional de la Información
Curso 2014-2015

Agradecimientos

A las primeras personas que quiero agradecer son a mis tutores Benjamin Ivorra y Ángel M. Ramos, pues sin su ayuda y conocimientos no hubiese sido posible realizar este trabajo. Gracias por todo el tiempo que han dedicado a que lograra realizar el trabajo.

Quisiera agradecer también a los profesores que me han impartido clase. Por todo lo que me han enseñado y me han transmitido estos años. Gracias a su ayuda y consejos he podido desarrollar mis conocimientos y ponerlos en práctica.

Gracias a mis compañeros de clase, que me han ayudado a alcanzar los objetivos del curso. Sobre todo a las personas con las que he compartido trabajos prácticos, su ayuda e interés han sido cruciales en el curso.

No puedo olvidar a mi familia, a mis padres y a mi hermano que siempre me han animado. Agradecer todo lo que hacen por mí, todo el apoyo que siempre me han dado en las decisiones que he tomado. Sin su ayuda no hubiese logrado todo lo que he conseguido, no hubiese soportado los momentos duros de la carrera, ni los que han venido detrás. Gracias por todo.

Y por último agradecer a Alfonso su cariño y paciencia. Darle las gracias por hacerme la vida más fácil, por escucharme y entenderme, por animarme y ver el lado bueno de las cosas y por hacer que sonría cada día.

Resumen

La enfermedad del virus del Ébola es una enfermedad humana letal que en 2014 requiso atención especial por el brote que surgió en algunos de países africanos y la posibilidad que se propagase a otros continentes. El modelo matemático $Be - CoDiS$ simula la propagación de la enfermedad entre los países; por tanto, sirve para estudiar la evolución de la enfermedad considerando el flujo migratorio entre los países y los efectos de las medidas de control. En este trabajo se estudia y se analiza el modelo $Be - CoDiS$. Primero, se describe la formulación matemática de cada componente del modelo. Segundo, se intentan reconstruir los datos de las personas infectadas por la enfermedad, y las personas fallecidas en Guinea, Liberia y Sierra leona, para tener la información completa. Posteriormente se pasa a analizar dos parámetros del modelo; uno la tasa de contagio efectiva de personas infectados, y otro, el parámetro que simula las medidas de control. Se busca el mejor método de estimación de éstos dos para que cuando se simule, se consiga reflejar con mayor precisión la realidad. Finalmente se estudia el error que comete el modelo. Se trata de minimizarlo ajustando ciertos parámetros, pero que los resultados no se alejen de la realidad.

Abstract

Ebola virus disease is a lethal human disease that required a particular attention in 2014 due to outbreak in some African countries and possible spread to other continents. The mathematical model $Be - CoDiS$ simulates the disease spread between countries. It used to study the evolution of human disease and it considers the migratory flux between countries and control measure effects. In this work we study and analyze the model $Be - CoDiS$. First, we describe the mathematical formulation of each component of the model. Second, we try to complete Ebola data about confirmed cases of Ebola virus disease (EVD) and death due to EVD. Next, we analyze two parameters of the model; one of these is the disease fatality effective contact rate of a infectious person, and the other is the parameter which simulates the efficiency of the control measures. We look for the best estimation method which provides results similar to reality. Finally, we study the model error. We try to minimize it and analyze the results returned by the model.

Índice general

1. Introducción y objetivos	6
2. Introducción histórica del Ébola	8
3. Función biológica del Ébola	11
4. Descripción del modelo epidemiológico	14
4.1. Propagación dentro de los países	15
4.2. Propagación entre países	17
4.3. Salidas que proporciona el modelo	19
4.4. Parámetros estimados del modelo	20
5. Reconstrucción de los datos	23
6. Estimación de los parámetros $\beta_I(i)$ y κ_i	32
7. Minimización del error	38
7.1. Algoritmo genético	40
7.1.1. Operadores básicos	41
7.1.2. Teorema fundamental de los algoritmos genéticos	42
8. Conclusiones	50

1 Introducción y objetivos

El Ébola es una enfermedad letal humana que en 2014 requirió una atención especial por parte de las autoridades internacionales de la salud debido al brote surgido en algunos países africanos. Esta epidemia fue centro de atención mundial por su posible propagación a otros continentes, como ocurrió al aparecer personas infectadas en Estados Unidos y España. La epidemia continuó durante el año 2015 y actualmente todavía se considera activa.

Los modelos y las simulaciones son unas importantes herramientas de decisión que pueden ser usados para controlar y erradicar las enfermedades humanas o animales. Así, se adaptan a las características de la enfermedad para simular situaciones reales. Los modelos matemáticos son de gran ayuda en la predicción de la posible evolución del virus del Ébola para dar algunas recomendaciones sobre factores en los que enfocar los esfuerzos para erradicar dicha enfermedad.

En este contexto, fue propuesto el modelo temporal y espacial *Be-CoDiS* (Between-Countries Disease Spread) por Benjamin Ivorra, Diène Ngom y Ángel M. Ramos para estudiar la evolución de esta enfermedad entre países. En su artículo [4], los autores presenta la primera versión del modelo epidemiológico que servirá para la simulación de la propagación de la enfermedad en un área. En este trabajo se utilizó la implementación del software en MATLAB que se puede descargar de forma gratuita en <http://www.mat.ucm.es/momat/software.htm>

Las principales ideas que introduce el modelo son las siguientes:

- (a) el estudio de la propagación en el país y entre países,
- (b) el uso de bases de datos reales,
- (c) la estimación de coeficientes dinámicos en el tiempo, acordes con los indicadores de los países (por ejemplo, situación económica, condiciones climáticas, etc)

La principal característica del modelo es que al mismo tiempo considera los siguientes efectos: flujo migratorio entre países, efectos de las medidas de control y estimaciones de coeficientes dinámicos en el tiempo para cada país.

Este trabajo de fin de máster trata de hacer estudios sobre el modelo *Be-CoDiS* para ver si cambiando algunos métodos de estimación de parámetros, o cambiando el valor de algunos parámetros, se consiguen resultados distintos.

La Organización Mundial de la Salud (OMS) reporta frecuentemente el número de casos detectados y personas fallecidas por la enfermedad del Ébola en todos los países afectados. Principalmente reporta Guinea, Liberia y Sierra Leona [6,7]. Estos datos nos interesan para poder comparar los resultados que aporta el modelo con la realidad.

El primero de nuestros objetivos sería completar los datos reportados puesto que no se tiene la información a nivel diario. Así, en la sección 5 se estudiarán y compararán varios métodos de imputación.

El segundo objetivo de este trabajo es estudiar la estimación de ciertos parámetros del modelo con distintos métodos. La tasa de contagio de la enfermedad de una persona infecciosa y el parámetro que simula la eficiencia de las medidas de control en el modelo *Be – CoDiS*, formulado en el artículo, se estiman mediante un algoritmo de regresión no lineal. Lo que se va a hacer en la sección 6 es estimarlos mediante otros dos métodos distintos: interpolación lineal y optimización no lineal; y posteriormente, comparar los resultados de los tres métodos.

Por último en la sección 7, se realizará el tercer objetivo que es estudiar el error que comete el modelo, ver técnicas de minimización de éste y comprobar la similitud de los datos proporcionados con la realidad.

2 Introducción histórica del Ébola

El virus del Ébola se detectó por primera vez en 1976; hubo dos brotes independientes; uno, en Zaire (actual República Democrática del Congo) y otro, en Sudán. La aldea de Zaire en la que se produjo el brote está situada cerca del río Ébola, de ahí el nombre del virus.

El primero episodio tuvo lugar en Zaire y afectó a 318 personas, con una letalidad del 90% de los enfermos, mientras que el segundo se presentó en la zona sur de Sudán, abarcando las áreas de Nzara, Maridi y Lirangu, con un total de 250 casos y una letalidad promedio de apenas 60%, aunque en la región más afectada llegó al 80%.

La mayor extensión del virus se produjo de forma intrahospitalaria, por contacto persona a persona y por la reutilización de agujas contaminadas, provocando pánico entre el personal.

El 5 de noviembre de 1976, uno de los científicos del Microbiological Research Establishment de Porton (Inglaterra), al estar trabajando en un homogenizado del hígado extraído a una cobaya inoculada con Ébola, se pinchó accidentalmente un pulgar a través de su guante de alta seguridad. El investigador quedó en observación durante cinco días y al sexto enfermó.

En el año 1979 se produce un nuevo brote en Sudán con 34 casos y 22 fallecidos. Hacia la década de los 90 se presentan casos en Filipinas, Virginia y Texas, así mismo, durante los años 1994 al 2000, Gabón registró el mayor número de casos, con más de 350 personas infectadas y alrededor de 280 fallecidos. En el año 2007 Uganda registra un nuevo brote de la enfermedad por el virus del Ébola con 149 infectados y 37 muertos. Este mismo país, en el año 2012, declara una nueva epidemia con 24 enfermos, de los que fallecieron 17, según datos de la OMS.

La epidemia actual comenzó en Guinea durante diciembre de 2013. La OMS fue notificada oficialmente de la rápida evolución del brote el 23 de marzo de 2014 cuando ya había también casos en varios distritos de Liberia. Durante mayo, el foco de la epidemia en Guinea se expandió a distritos vecinos de Sierra Leona. A partir de julio, hubo un gran aumento en el número de casos no solamente en los focos epidémicos de los tres países sino también en otros sitios distantes del epicentro. El 8 de agosto, la OMS declaró a la epidemia como una *emergencia de salud pública de preocupación internacional*.

En marzo de 2014 se registró el inicio del último brote, que se inició en Guinea y se extendió a Sierra Leona, Liberia y Nigeria. En la actualidad, la Organización Mundial de la Salud reporta 28.055 casos (nuevos, confirmados, probables y posibles) y 11.288 fallecidos (a 28 de Agosto de 2015), de acuerdo a los datos comunicados por los Ministerios de Salud de los países afectados.

Debido a una respuesta exitosa para controlar la propagación de la enfermedad, Senegal y Nigeria fueron declarados libres de Ébola el 17 de octubre y 19 de octubre de 2014 respectivamente. Esto se logra al pasar 42 días desde que el último paciente da negativo en las pruebas de laboratorio. De los países con transmisión localizada, Mali y los Estados Unidos de América continúan vigilando posibles contactos. En España, los 83 contactos de una enfermera infectada en Madrid completaron el período de seguimiento de 21 días.

La enfermedad por virus Ébola (EVE) puede llegar a tener una tasa de letalidad del 90%. A pesar de que la actual epidemia de EVE en África Occidental no tiene precedentes en escala, el curso clínico de la infección y la transmisión del virus son similares a los de los brotes anteriores. El período de incubación, duración de la enfermedad y la tasa de letalidad están todos dentro de los rangos reportados para las epidemias anteriores. La combinación de los signos y síntomas registrados entre el inicio de los síntomas y la presentación clínica es también similar a la de otros informes. La epidemia actual es excepcionalmente grande, probablemente no por las características biológicas del virus, sino más bien debido a los atributos de las poblaciones afectadas y debido a que los esfuerzos de control han sido insuficientes para detener la propagación de la infección.

Ciertas características de las poblaciones afectadas pueden haber dado lugar a la rápida difusión geográfica de la infección. Las poblaciones de Guinea, Liberia y Sierra Leona están altamente interconectadas, con mucho tráfico transfronterizo en el epicentro y conexiones relativamente sencillas por carretera entre las ciudades y las aldeas rurales, y entre las capitales nacionales densamente pobladas. La población entremezclada ha facilitado la propagación de la infección, pero, a pesar de esto, una gran epidemia podría haber sido evitable. El determinante crítico del tamaño de la epidemia pudo ser, en este caso, la lentitud en la aplicación de medidas de control rigurosas.

En lo relativo a su transmisión, el virus del Ébola se introduce en la población humana por contacto con órganos, sangre, secreciones u otros líquidos corporales de animales infectados. Posteriormente, el virus se propaga en la comunidad mediante la transmisión de persona a persona, por contacto directo de personas infectadas, o por contacto indirecto con materiales contaminados. Los pacientes son contagiosos mientras el virus esté presente en la sangre y las secreciones.

La enfermedad suele caracterizarse por la aparición súbita de fiebre, debilidad intensa y dolores musculares, de cabeza y de garganta, lo cual va seguido de vómitos, diarrea, erupciones cutáneas, disfunción renal y hepática y, en algunos casos, hemorragias internas y externas. Los resultados de laboratorio muestran disminución del número de leucocitos y plaquetas, así como elevación de las enzimas hepáticas. El período de incubación (intervalo desde la infección hasta la aparición de los síntomas) oscila entre 2 y 21 días.

En la actualidad no hay vacuna contra el Ébola. Los casos graves requieren cuidados intensivos. Los enfermos suelen estar deshidratados y necesitar rehidratación por vía intravenosa u oral con soluciones que contengan electrolitos. Tampoco hay ningún tratamiento específico, aunque se están evaluando nuevos tratamientos farmacológicos.

3 Función biológica del Ébola

Unas herramientas de decisión útiles en el control y en la erradicación de enfermedades son la modelización y la simulación. Cada enfermedad tiene sus propias características y es necesario adaptar los modelos de simulación a éstas para poder considerar una situación real.

Una función de la epidemiología es describir cómo la enfermedad se propaga, identificar las causas o factores de riesgo de la enfermedad para descubrir el motivo por el que alguien se contagia. También construye y prueba teorías y así evalúa programas de control y prevención.

Algunos de los propósitos de los modelos epidemiológicos son los siguientes:

- El proceso de formulación del modelo aclara hipótesis, variables y parámetros.
- Los modelos permiten la exploración de los efectos de las distintas hipótesis.
- Los modelos son herramientas experimentales para probar teorías y conjeturas.
- Los modelos pueden ser usados para estimar parámetros clave en el ajuste de los datos.
- Los modelos pueden ser usados para comparar enfermedades de diferentes tipos o de diferentes temporadas o sobre diferentes poblaciones.
- Los modelos pueden ser usados para evaluar teóricamente, comparar y optimizar los programas de control.
- Los modelos pueden ser útiles para evaluar la sensibilidad de los resultados y poder así ajustar los valores de los parámetros.
- Los modelos pueden sugerir datos críticos que necesitan ser recogidos.
- Los modelos pueden contribuir al diseño y análisis de estudios epidemiológicos.
- Los modelos pueden ser usados para identificar tendencias hacer predicciones generales, o estimar la incertidumbre de estas predicciones.

Aunque los modelos epidemiológicos son muy beneficiosos, a pesar de todo también tienen limitaciones. Algunas de estas son las siguientes:

- Un modelo epidemiológico es una simplificación extrema de la realidad, no es la realidad.
- Los modelos determinísticos son modelos matemáticos donde las mismas entradas producirán invariablemente las mismas salidas, no contemplándose la

existencia del azar ni el principio de incertidumbre. Están estrechamente relacionados con la creación de entornos simulados para el estudio de situaciones hipotéticas. Así, estos modelos no reflejan la incertidumbre de la propagación de la enfermedad.

- Un modelo es estocástico cuando al menos una variable del mismo es tomada como un dato al azar y las relaciones entre variables se toman por medio de funciones probabilísticas. Por tanto, estos modelos incorporan oportunidad, pero normalmente son costosos computacionalmente y más difíciles de analizar que los correspondientes modelos determinísticos.

Los tres estados epidemiológicos de los modelos clásicos son: susceptible, infectado y recuperado. Los movimientos de un estado a otro están determinados por ecuaciones diferenciales.

Pero en el caso del Ébola, si una persona está infectada, la persona pasa sucesivamente por los siguientes estados:

- $E \rightarrow$ Infectado: la persona está infectada pero no puede infectar a los demás y no tiene síntomas clínicos visibles. La duración media en este estado es de 11,4 días.
- $I \rightarrow$ Infeccioso: la persona puede infectar a otras y empieza a desarrollar los síntomas clínicos. La duración media es de 5 días y se denomina periodo infeccioso.
- $H \rightarrow$ Hospitalizado: la persona es hospitalizada y puede infectar a otras personas. La duración media en este estado es de 4,5 días. La probabilidad de infectar de los enfermos hospitalizados es 25 veces menor que la probabilidad de infectar de las personas en estado infeccioso. Después de aproximadamente 4,2 días, entre el 25% y el 72,8% de los hospitalizados mueren debido al EVE y pasan al estado D . Pero si la persona hospitalizada sobrevive una media de 5 días, la persona sobrevive al EVE y pasa al estado R .
- $D \rightarrow$ Muerto: la persona que no ha sobrevivido al EVD . Los cadáveres de las personas infectadas pueden seguir siendo contagiosos hasta que son quemados.
- $B \rightarrow$ Incinerado: la persona ha muerto a causa del EVD , su cadáver es incinerado y ya no se considera infeccioso.

- $R \rightarrow$ Recuperado: la persona ha sobrevivido al EVD y ya no es infecciosa. Estas personas desarrollan una inmunidad natural al EVE .

Una vez que la persona se ha infectado de EVE es hospitalizada, y las medidas de control que se toman para controlar la propagación es la siguiente:

- Aislación: el infectado es aislado de otras personas, solo los profesionales sanitarios pueden tener contacto con el enfermo.
- Cuarentena: los movimientos de personas en el área de origen se restringen, esto es para evitar que la enfermedad se propague.
- Seguimiento: se intenta identificar contactos potencialmente infecciosos los cuales pueden ser la causa de la propagación del EVD a otra persona.
- Mejorar las condiciones sanitarias: los funerales de los cadáveres infectados son controlados por el personal sanitario para reducir el contacto entre los cadáveres y las personas susceptibles.

4 Descripción del modelo epidemiológico

El modelo *Be – CoDiS* es usado para evaluar la propagación de la enfermedad dentro de un mismo país, y entre varios países durante un periodo de tiempo.

Las principales notaciones usadas para describir *Be – CoDiS* son las siguientes:

Notación	Descripción (unidad de medida)
T_{max}	Número máximo de días de simulación (día)
Δt	Medida del paso de discretización del tiempo(día)
$\beta_I(i)$	Tasa de contacto de la enfermedad de una persona en el estado <i>I</i> en el país <i>i</i> (día ⁻¹)
$\beta_H(i)$	Tasa de contacto de la enfermedad de una persona en el estado <i>H</i> en el país <i>i</i> (día ⁻¹)
$\beta_D(i)$	Tasa de contacto de la enfermedad de una persona en el estado <i>D</i> en el país <i>i</i> (día ⁻¹)
γ_E	Tasa de transición de una persona en estado <i>E</i> (día ⁻¹)
γ_I	Tasa de transición de una persona en estado <i>I</i> (día ⁻¹)
γ_{HR}	Tasa de transición de una persona en estado <i>H</i> a estado <i>R</i> (día ⁻¹)
γ_{HD}	Tasa de transición de una persona en estado <i>H</i> a estado <i>D</i> (día ⁻¹)
C_0	Número de días que una persona permanece hospitalizado (día)
$\mu_m(i)$	Tasa de mortalidad natural del país <i>i</i> (día ⁻¹)
$\mu_n(i)$	Tasa de natalidad natural del país <i>i</i> (día ⁻¹)
$\omega(i)$	Porcentaje de letalidad de la enfermedad en el país <i>i</i> (%)
κ_i	Eficacia de las medidas de control en el país <i>i</i> (día ⁻¹)
$\lambda(i)$	Primer día que se aplican las medidas de control en el país <i>i</i> (día)
$m_I(i, t) / m_H(i, t)$	Eficacia de las medidas de control(%) en el país <i>i</i> al tiempo <i>t</i> aplicadas a las personas en estado <i>I</i> o <i>H</i> , respectivamente
$m_{tr}(i, j, t)$	Eficacia de las medidas de control(%) aplicadas a las personas que migran del país <i>i</i> al <i>j</i> en el tiempo <i>t</i>
$\tau(i, j)$	Tasa de migración diaria desde el país <i>i</i> al país <i>j</i> (día ⁻¹)
N_{CO}	Número de países de la región estudiada
$NP(i, t)$	Número de personas en el país <i>i</i> en el tiempo <i>t</i>
$S(i, t) / E(i, t) / I(i, t) / H(i, t) / R(i, t) / D(i, t) / B(i, t)$	Número de personas en el estado <i>S, E, I, H, R, D, B</i> en el país <i>i</i> en el tiempo <i>t</i>

Al inicio de la simulación, $t = 0$, en los países libres de la enfermedad solo viven personas en estado susceptible. La simulación se detiene cuando $t = T_{max}$, siendo

T_{max} un parámetro especificado por el usuario. Las medidas de control se tienen en cuenta en el modelo, pudiéndose activar o desactivar para cuantificar la efectividad de éstas.

4.1. Propagación dentro de los países

La propagación dentro de un determinado país i es modelada usando un modelo determinístico. Se asume que en cada instante de tiempo, la población de un país está homogéneamente distribuida, así la distribución espacial de la enfermedad dentro de un país puede ser omitida. También se asume que los nacimientos son personas en estado susceptible.

Bajo estas suposiciones y sin considerar la interacción entre países, el modelo para cada país i es el siguiente:

$$\begin{aligned}
\frac{dS(i,t)}{dt} &= \frac{S(i,t)(m_I(i,t)\beta_I(i)I(i,t) + m_H(i,t)\beta_h(i)H(i,t))}{NP(i,t)} \\
&\quad - \frac{S(i,t)(m_D(i,t)\beta_D(i)D(i,t))}{NP(i,t)} - \mu_m(i)S(i,t) \\
&\quad + \mu_m(i)(S(i,t) + E(i,t) + I(i,t) + H(i,t) + R(i,t)) \\
\frac{dE(i,t)}{dt} &= \frac{S(i,t)(m_I(i,t)\beta_I(i)I(i,t) + m_H(i,t)\beta_H(i)H(i,t))}{NP(i,t)} \\
&\quad + \frac{S(i,t)(m_D(i,t)\beta_D(i)D(i,t))}{NP(i,t)} \\
&\quad - \mu_m(i)E(i,t) - \gamma_{E\chi_{E_{fit}}}(E(i,t)) \\
\frac{dI(i,t)}{dt} &= \gamma_{E\chi_{E_{fit}}} - (\mu_m(i) + \gamma_I)I(i,t) \\
\frac{dH(i,t)}{dt} &= \gamma_I I(i,t) - (\mu_m(i) + (1 - \omega(i))\gamma_{HR} + \omega(i)\gamma_{HD})H(i,t) \\
\frac{dR(i,t)}{dt} &= (1 - \omega(i))\gamma_{HR}H(i,t) - \mu_m(i)R(i,t) \\
\frac{dD(i,t)}{dt} &= \omega(i)\gamma_{HD}H(i,t) - \gamma_D D(i,t) \\
\frac{dB(i,t)}{dt} &= \gamma_D D(i,t)
\end{aligned}$$

donde :

- $i \in \{1, \dots, N_{CO}\}$,
- $N_{CO} \in \mathbb{N}$ es el número de países,
- $NP(i, t) = S(i, t) + E(i, t) + I(i, t) + H(i, t) + R(i, t) + D(i, t) + B(i, t)$ es el número de personas vivas y también muertas o incineradas a causa del Ébola en el país i en el tiempo t ,
- $\mu_n(i) \in [0, 1]$ es la tasa de natalidad (día⁻¹) en el país i : número de nacimientos por día y per cápita,
- $\mu_m(i) \in [0, 1]$ es la tasa de mortalidad (día⁻¹) en el país i : número de muertes por día y per cápita (o, equivalentemente, el inverso de la esperanza de vida media (día) de una persona),
- $E_{fit} \in [0, +\infty)$ es un parámetro de tolerancia definido por

$$\chi_{E_{fit}}(x) = \begin{cases} x & \text{si } x \geq E_{fit}, \\ 2x - E_{fit} & \text{si } \frac{E_{fit}}{2} \leq x \leq E_{fit}, \\ 0 & \text{en caso contrario.} \end{cases}$$

. Se trata de un filtro usado para evitar propagaciones artificiales de la epidemia debido a valores de x despreciables,

- $\omega(i) \in \{0, 1\}$ es el porcentaje de letalidad de la enfermedad en el país i : porcentaje de personas que no sobreviven a la enfermedad,
- $\beta_I(i) \in \mathbb{R}^+$ es la tasa de contacto efectivo (día⁻¹) de una persona en estado I en el país i : la media de contactos efectivos de una persona en estado I por día antes de aplicar las medidas de control,
- $\beta_H(i) \in \mathbb{R}^+$ es la tasa de contacto efectivo (día⁻¹) de una persona en estado H en el país i ,
- $\beta_D(i) \in \mathbb{R}^+$ es la tasa de contacto efectivo (día⁻¹) de una persona en estado D en el país i ,

- $\gamma_E, \gamma_I, \gamma_{HR}, \gamma_{HD}, \gamma_D \in \mathbb{R}^+$ denotan la tasa de transición (día⁻¹) del estado E, I, H, H o D al estado I, H, R, D o B , respectivamente: el número de personas por día y per cápita que pasa de un estado al otro,
- $m_I(i, t), m_H(i, t), m_D(i, t) \in \{0, 1\}$ (%) son funciones que representan la efectividad de las medidas de control aplicadas a las personas no hospitalizadas, las hospitalizadas y los cadáveres infectados respectivamente, en el país i en el tiempo t para erradicar los brotes.

Focalizando en las medidas de control, las cuales consisten en aislar a las personas y las áreas de riesgo y mejorar las condiciones sanitarias de los funerales, se multiplican las tasas de contacto ($\beta_I(i), \beta_H(i)$ y $\beta_D(i)$) por una función decreciente que simula la reducción del número de contactos efectivos como mejora por las medidas de control.

Así, se consideran las funciones:

$$m_I(i, t) = m_H(i, t) = m_D(i, t) = \exp(-\kappa_i \max(t - \lambda(i), 0)),$$

donde $\kappa_i \in [0, +\infty)$ (día⁻¹) simula la eficacia de las medidas de control (valores mayores implican valores menores de las tasas de contacto) y $\lambda(i) \in \mathbb{R} \cup \{+\infty\}$ (día) denota el primer día que se aplican esas medidas de control.

Este sistema de ecuaciones diferenciales se completa con los datos iniciales $S(i, 0), E(i, 0), I(i, 0), H(i, 0), R(i, 0), D(i, 0)$ y $B(i, 0) \in \mathbb{R} \quad \forall i = 1, \dots, N_{CO}$.

4.2. Propagación entre países

La propagación de la enfermedad entre países es modelada usando un modelo determinístico espacial de tipo “Individual-Based”. Los países son clasificados en uno de los siguientes estados:

- $F \rightarrow$ Libre de enfermedad
- $O \rightarrow$ Con brotes de la epidemia

Se asume que en cada tiempo t un país i está en estado O si $I(i, t) + H(i, t) \geq 1$ (hay al menos una persona infectada en el país).

La única manera de introducir la enfermedad del país i en estado O al país j es que las personas viajen de i a j . Por tanto, se considera la matriz $(\sigma(i, j))_{i, j=1}^{N_{CO}}$ donde $\sigma(i, j) \in [0, 1]$ es el ratio de transferencia de personas desde el país i al j , el cual se expresa en % de población de i por unidad de tiempo. Debido a las medidas de control en ambos países, se asume que los ratios pueden variar con el tiempo. Se asume también, que las únicas personas que pueden viajar y traspasar la enfermedad

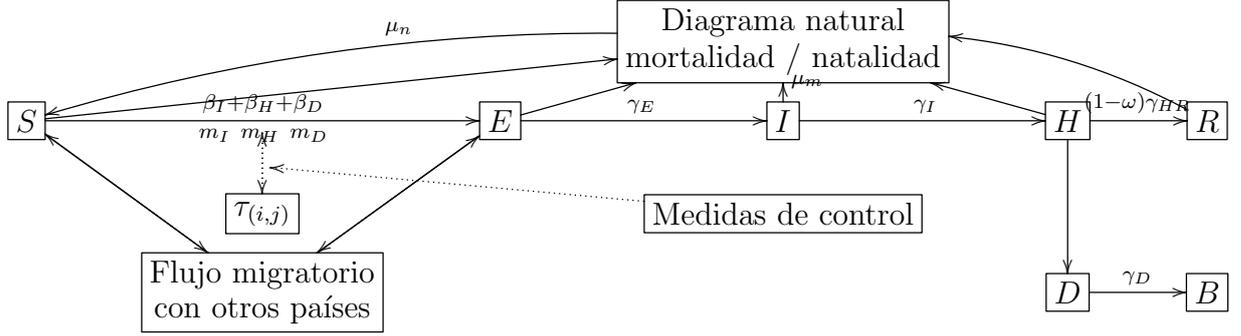
son los estados S y E .

Así, se considera el modelo siguiente:

$$\begin{aligned}
\frac{dS(i,t)}{dt} &= -\frac{S(i,t)(m_I(i,t)\beta_I(i)I(i,t) + m_H(i,t)\beta_H(i))H(i,t)}{NP(i,t)} \\
&\quad -\frac{S(i,t)(m_D(i,t)\beta_D(i)D(i,t))}{NP(i,t)} - \mu_m(i)S(i,t) \\
&\quad + \mu_n(i)(S(i,t) + E(i,t) + I(i,t) + H(i,t) + R(i,t)) \\
&\quad + \sum_{i \neq j} m_{tr}(j,i,t)\sigma(j,i)S(j,t) - \sum_{i \neq j} m_{tr}(i,j,t)\sigma(i,j)S(i,t) \\
\frac{dE(i,t)}{dt} &= \frac{S(i,t)(m_I(i,t)\beta_I(i)I(i,t) + m_H(i,t)\beta_H(i)H(i,t))}{NP(i,t)} \\
&\quad + \frac{S(i,t)(m_D(i,t)\beta_D(i)D(i,t))}{NP(i,t)} - \mu_m(i)E(i,t) - \gamma_E \chi_{\epsilon_{fit}}(E(i,t)) \\
&\quad + \sum_{i \neq j} m_{tr}(j,i,t)\sigma(j,i)\chi_{\epsilon_{fit}}(E(j,t)) \\
&\quad - \sum_{i \neq j} m_{tr}(i,j,t)\sigma(i,j)\chi_{\epsilon_{fit}}(E(i,t)) \\
\frac{dI(i,t)}{dt} &= \gamma_E \chi_{E_{fit}}(E(i,t)) - (\mu_m(i) + \gamma_I)I(i,t) \\
\frac{dH(i,t)}{dt} &= \gamma_I I(i,t) - (\mu_m(i,t) + (1 - \omega(i))\gamma_{HR} + \omega(i)\gamma_{HD})H(i,t) \\
\frac{dR(i,t)}{dt} &= (1 - \omega(i))\gamma_{HR}H(i,t) - \mu_m(i)R(i,t) \\
\frac{dD(i,t)}{dt} &= \omega(i)\gamma_{HD}H(i,t) - \gamma_D D(i,t) \\
\frac{dB(i,t)}{dt} &= \gamma_D D(i,t)
\end{aligned}$$

El sistema de ecuaciones diferenciales es completado con los datos iniciales $S(i,0)$, $E(i,0)$, $I(i,0)$, $H(i,0)$, $R(i,0)$, $D(i,0)$ y $B(i,0) \in [0, +\infty) \quad \forall i = 1, \dots, N_{CO}$.

Este modelo completo, es al que se le llama *Be - CoDiS* y su diagrama de flujo es el siguiente:



4.3. Salidas que proporciona el modelo

Los siguientes outputs que proporciona el modelo, sirven para analizar los resultados de las simulaciones. Así, si se considera el intervalo $[0, T]$, para cada país i se tienen los siguientes valores:

- $cumul_{casos}(i, t)$: número de casos acumulados a causa del EVD en el país i el día t , el cual se calcula como:

$$cumul_{casos}(i, t) = cumul_{casos}(i, 0) + \int_0^t \gamma_I \cdot I(i, t) dt$$

- $cumul_{muertes}(i, t)$: número acumulado de muertes debidas al EVD , en el país i el día t , calculado como

$$cumul_{muertes}(i, t) = cumul_{muertes}(i, 0) + \omega(i) + \int_0^t \gamma_{HD} \cdot H(i, t) dt$$

- $RS(i)$: riesgo inicial del país i de propagar EVD a otros países, dado por:

$$RS(i) = \sum_{i \neq j} \tau(i, j) m_{tr}(i, j, 0) E(i, 0)$$

RS (personas · día⁻¹) es la cantidad diaria de personas infectadas que abandonan el país j en el tiempo $t = 0$

- $TRS(i)$: riesgo total del país i de propagar EVD a otros países, considerando el intervalos $[0, T]$, dado por:

$$TRS(i) = \sum_{t=0}^T \sum_{i \neq j} \tau(i, j) m_{tr}(i, j, t) E(i, t)$$

TRS (personas) es el número de personas infectadas enviadas a otros países durante el intervalo de tiempo considerado

- $RI(i)$: riesgo inicial de introducción del *EVD* en el país i desde otros países, dado por:

$$RI(i) = \sum_{j \neq i} \tau(j, i) m_{tr}(j, i, 0) E(i, 0)$$

RI (personas · día⁻¹) es la cantidad diaria de personas infectadas que entran el país i en el tiempo $t = 0$

- $TRI(i)$: riesgo total de introducción del *EVD* en el país i desde otros países, considerando el intervalo de tiempo $[0, T]$, dado por:

$$TRI(i) = \sum_{t=0}^T \sum_{i \neq j} \tau(j, i) m_{tr}(j, i, t) E(j, t)$$

TRI (personas) es el número de personas infectadas que recibe de otros países durante el intervalo de tiempo considerado

- $MNH(i)$: el número máximo de personas hospitalizadas al mismo tiempo en el país i durante el intervalo de tiempo $[0, T]$, calculado:

$$MNH(i) = \max_{t=0 \dots T} \{H(i, t) + R(i, t) - R(i, t - C_0)\}$$

- $Emf(i)$: el flujo migratorio total (día⁻¹). Es el porcentaje de población que abandona el país i cada día, se calcula como:

$$Emf(i) = \sum_{i \neq j} \tau(i, j)$$

4.4. Parámetros estimados del modelo

Estos parámetros se estiman gracias al conjunto de datos disponibles sobre el brote de Ébola de 2014-2015, usando las características epidemiológicas del virus y los indicadores de los países.

Se necesitan los indicadores de cada país:

- $TMP_i(t) \in \mathbb{R}$: temperatura media en el día t (°C).
- $DEN_i \in \mathbb{R}^+$: densidad de población (personas/km²).

- $NP(i, 0) \in \mathbb{N}$: número total de personas vivas, muertas o incineradas a causa del *EVD*.
- $GNI_i \in \mathbb{R}^+$: Producto Interior Bruto anual per cápita ($\text{US\$} \cdot \text{persona}^{-1} \cdot \text{año}^{-1}$). Es un indicador económico del país.
- $SAN_i \in \mathbb{R}^+$: media del gasto sanitario anual per cápita ($\text{US\$} \cdot \text{persona}^{-1} \cdot \text{año}^{-1}$). Es un indicador económico del sistema sanitario de un país, dado por la cantidad de dinero invertido por instituciones públicas y privadas en el sistema sanitario del país.
- $MLE_i \in \mathbb{R}^+$: esperanza de vida media (día). Así, $\mu_m(i) = \frac{1}{MLE_i}$
- $\mu_n(i) \in [0, 1]$: tasa de natalidad del país i . Calculado de manera análoga al anterior parámetro.

Estos datos son obtenidos de la web del Banco Mundial de Datos del año 2013: <http://data.worldbank.org>

También se tienen las estimaciones del tiempo medio que una persona está en cada estado:

- $dur_E \rightarrow$ tiempo medio (en días) que una persona está en el estado E
- $dur_I \rightarrow$ tiempo medio (en días) que una persona está en I
- $dur_{HR} \rightarrow$ tiempo medio (en días) que una persona del estado R ha pasado en el estado H
- $dur_{HD} \rightarrow$ tiempo medio (en días) que una persona del estado D ha pasado en el estado H
- $dur_D \rightarrow$ tiempo medio (en días) que una persona está en el estado D
- $varduri \rightarrow$ reducción máxima (en días) de una persona en el estado I debido a las medidas de control
- $conva \rightarrow$ duración (en días) del periodo de convalecencia

Los inversos de estos parámetros son las tasas de transición:

$$\gamma_E = \frac{1}{11,4}(\text{dia}^{-1}) \quad \gamma_I = \frac{1}{5}(\text{dia}^{-1}) \quad \gamma_{HR} = \frac{1}{5}(\text{dia}^{-1})$$

$$\gamma_{HD} = \frac{1}{4,2}(\text{dia}^{-1}) \quad \gamma_D = \frac{1}{2}(\text{dia}^{-1}).$$

Para estimar la tasa de flujo migratorio humano $\tau(i, j)$, se utilizó la información de la base de datos disponible en <http://www.global-migration.info>. Los datos

originales $\hat{\tau}(i, j)$ representan el número de personas que se desplazan del país i al j desde 2005 hasta 2010. Se calcula entonces

$$\tau(i, j) = \frac{\hat{\tau}(i, j)}{5 \cdot 365 \cdot NP(i, 0)},$$

que representa el porcentaje de personas del país i que se desplazan al país j diario.

Las tasas de contacto $\beta_H(i)$ y $\beta_D(i)$ van a depender del valor de $\beta_I(i)$. Este último, se va a estimar de varias formas y se hará un estudio del resultado obtenido, todo ello se puede ver en la sección 5.

Así, con el valor de $\beta_I(i)$ estimado, se tiene que:

- $\beta_H(i) = \frac{\beta_I(i)}{26}$ (día⁻¹), la probabilidad de ser infectado por contacto con personas en estado H es 26 veces menor que la probabilidad de ser infectados por personas en estado I .
- $\beta_D(i) = \beta_I(i)$, la probabilidad de ser infectado por contacto con personas en estado D es la misma que la probabilidad de ser infectado por personas en estado I .

El porcentaje de letalidad del *EVD* oscila entre [25, 72,8] % dependiendo de la calidad de los servicios sanitarios. Además, la tasa de letalidad máxima ha ido decreciendo debido a las medidas de control, por esto, se modela el efecto multiplicando por la tasa de letalidad máxima $m_I(i, t)$. Así, el porcentaje de letalidad se estima como:

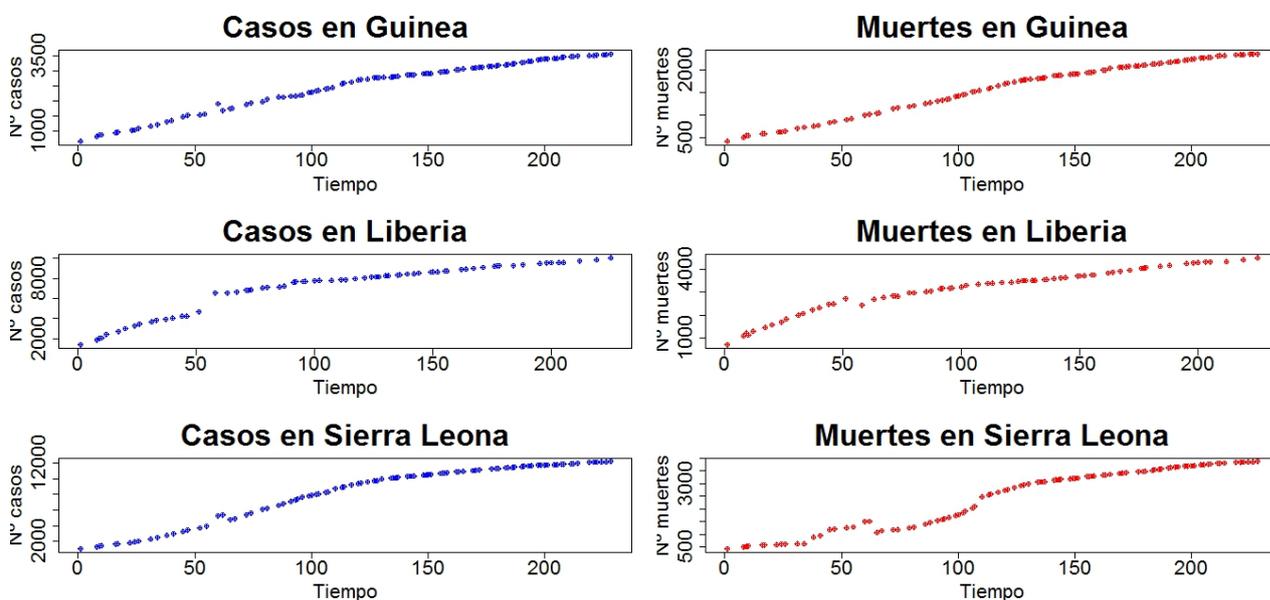
$$\omega(i) = 0,25 \frac{SAN_i}{\max_i SAN_i} + (0,728 \cdot m_I(i, t)) \cdot \left(1 - \frac{SAN_i}{\max_i SAN_i}\right)$$

5 Reconstrucción de los datos

La Organización Mundial de la Salud reporta frecuentemente el número acumulado de casos confirmados del virus del Ébola y el número acumulado de muertes que se han dado por la enfermedad de los países Guinea, Liberia y Sierra Leona.

Si se necesitan los datos con frecuencia diaria, se necesitan los datos intermedios que no están reportados. Para imputar estos valores ausentes de las series temporales se van a utilizar varios métodos y comparar los resultados.

La series reportadas tienen las siguientes gráficas:



Mediante las gráficas se puede observar que hay datos que deben ser erróneos ya que tratándose de datos acumulados, la gráfica tiene que ser creciente en todo momento y en cuatro de las seis series este hecho se viola.

El primero de los métodos que se utilizan es la regresión lineal. Así, se considera un modelo de regresión lineal simple por el cual, se quiere modelizar la relación entre los casos de personas infectadas y la variable temporal, y también modelizar la relación existente entre las muertes debidas a la EVE y la variable temporal.

El modelo de la regresión lineal simple, en el que una variable dependiente y , se expresa como función de una variable independiente x , es

$$y = \beta_0 + \beta_1 x + \epsilon$$

donde β_0 y β_1 son los coeficientes de regresión y ϵ es la componente de error aleatorio. El parámetro β_0 es la ordenada en el origen y β_1 es la pendiente de la recta de regresión y representa el cambio en la media de la distribución de y producido por un cambio unitario en x .

Los residuos representan la diferencia entre el valor observado y_i y el correspondiente valor ajustado por la regresión, \hat{y}_i . Matemáticamente el i -ésimo residuo es

$$e_i = y_i - \hat{y}_i = y_i - \beta_0 - \beta_1 x_i. \quad i = 1, \dots, n$$

siendo n el tamaño de la muestra.

Un modelo de regresión adecuado tiene que verificar las condiciones de Gauss-Markov para evitar situaciones anómalas. Sea el modelo de regresión $y = \beta_0 + \beta_1 x + \epsilon$ las condiciones son:

1^o $E[\epsilon_i] = 0 \quad \forall i = 1, \dots, n$. Así se asegura que $E[y_i] = \beta_0 + \beta_1 x_i$.

2^o $Var(\epsilon) = \sigma^2$ constante. Es la conocida como propiedad de homocedasticidad. Se trata de que la varianza de las y_i sea aproximadamente igual, dentro del rango de variación de las x_i . Es decir, que no se presenten situaciones en las que la varianza crece (decrece) con las observaciones (heterocedasticidad).

3^o $E[\epsilon_i \epsilon_j] = 0 \quad \forall i \neq j$. Los errores son incorrelados.

Para comparar los modelos se va a utilizar dos valores: el error cuadrático medio (MSE) y el coeficiente de determinación (R^2).

El MSE representa la diferencia entre el estimador y lo que se estima. Es un estimador insesgado de la varianza (σ^2) de los residuos, que es calculado por:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2} = MSE$$

El coeficiente de determinación, R^2 , se puede interpretar como la proporción de variabilidad explicada por el regresor x . Si R^2 se aproxima a 1 significa que la mayor parte de la variabilidad de y es explicada por el modelo de regresión. Viene calculado por:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

siendo \bar{y} la media muestral.

Se utilizan los estimadores por mínimos cuadrados y así se obtienen los siguientes seis modelos de regresión, y se predicen los valores de los días en los que no se tiene dato. Así, las gráficas de la serie completa resultarían:

Casos en Guinea

Fórmula	$y = 509,93 + 8,86x$
R^2	0.9712
MSE	19079

Muertes en Guinea

Fórmula	$y = 932,85 + 12,68x$
R^2	0.9815
MSE	5930

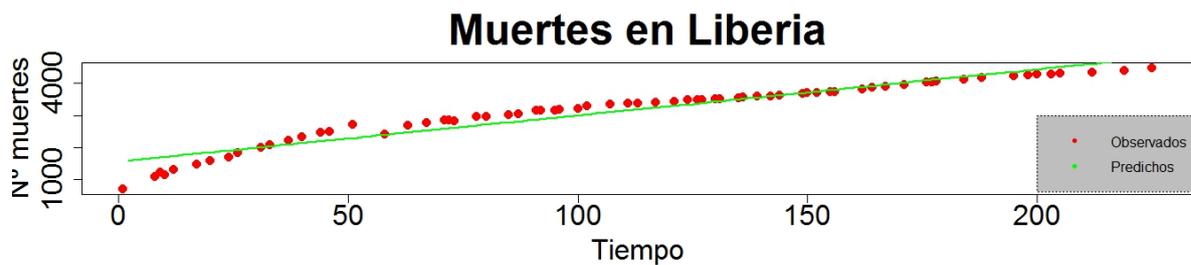
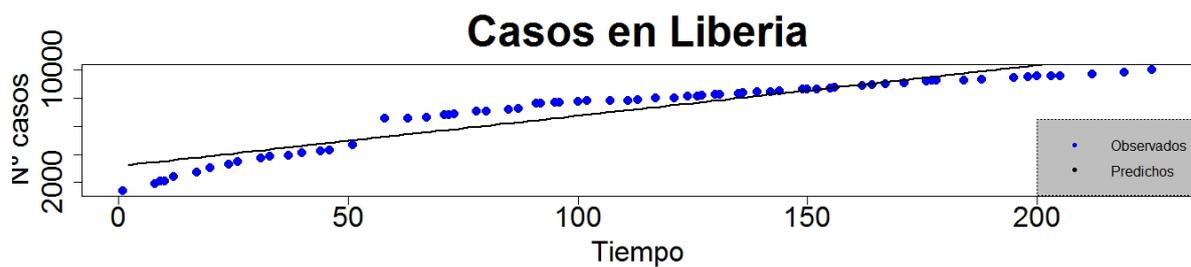


Casos en Liberia

Fórmula	$y = 3163,98 + 35,93x$
R^2	0.8786
MSE	697070

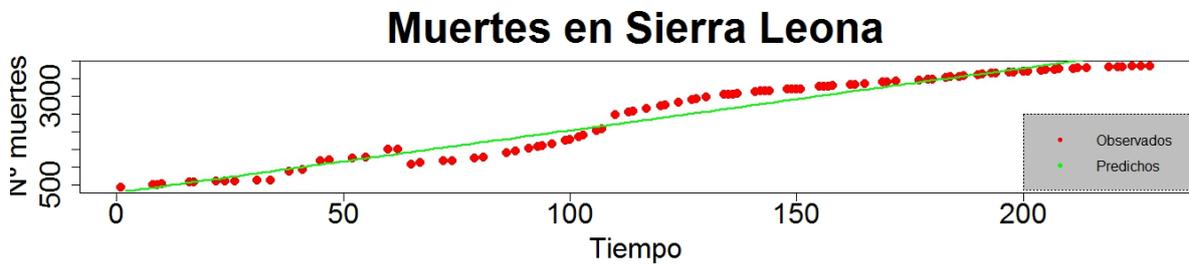
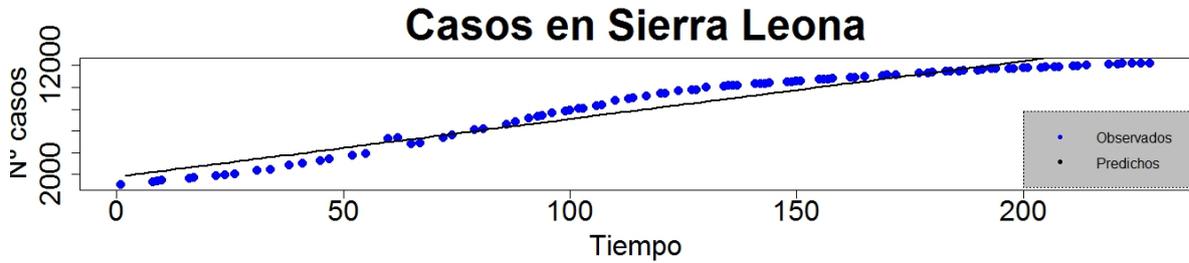
Muertes en Liberia

Fórmula	$y = 1561,16 + 14,39x$
R^2	0.9303
MSE	60708



Casos en Sierra Leona	
Fórmula	$y = 1767,55 + 53,25x$
R^2	0.9368
MSE	780492

Muertes en Sierra Leona	
Fórmula	$y = 1280,47 + 17,63x$
R^2	0.9501
MSE	66612



Estos seis modelos tienen un R^2 alto. Aun así, se observa en las gráficas completas que los datos originales tienen cierta forma curvada y no lineal.

Por esta última razón, para conseguir cierta oscilación a la recta, se considera el segundo método para imputar los valores ausentes. Usando la idea de la forma de una serie de Fourier

$$\frac{a_0}{2} \sum_n \left[a_n \cos \frac{2n\pi}{T} t + b_n \sin \frac{2n\pi}{T} t \right]$$

con $a_0 = 1$ y $n = 1, \dots, 7$, se utiliza un modelo de regresión lineal múltiple, donde se quiere estudiar la relación entre la variable original de los casos frente al tiempo y una serie de desfases, y del mismo modo la variable original de las muertes:

$$y = \beta_0 + \beta_1 t + \beta_2 \frac{1}{2} \cos \frac{2\pi}{T} t + \beta_3 \frac{1}{2} \sin \frac{2\pi}{T} t + \beta_4 \cos \frac{2\pi}{T} t + \beta_5 \sin \frac{2\pi}{T} t +$$

$$+ \beta_6 \frac{3}{2} \cos \frac{2\pi}{T} t + \beta_7 \frac{3}{2} \sin \frac{2\pi}{T} t + \beta_8 2 \cos \frac{2\pi}{T} t + \beta_9 2 \sin \frac{2\pi}{T} t +$$

$$+ \beta_{10} \frac{5}{2} \cos \frac{2\pi}{T} t + \beta_{11} \frac{5}{2} \sin \frac{2\pi}{T} t + \beta_{12} 3 \cos \frac{2\pi}{T} t + \beta_{13} 3 \sin \frac{2\pi}{T} t + \beta_{14} \frac{7}{2} \cos \frac{2\pi}{T} t + \beta_{15} \frac{7}{2} \sin \frac{2\pi}{T} t$$

Al igual que antes, valoramos los modelos de regresión resultantes a través del error cuadrático medio (MSE) y el coeficiente de determinación ajustado R_{adj}^2 . Como R^2 aumenta cuando se incluye un nuevo regresor en el modelo, sin importar la cuantía de la aportación del regresor, se usa R_{adj}^2 que solo aumenta si esta inclusión reduce la media de cuadrados de los errores. Se define como:

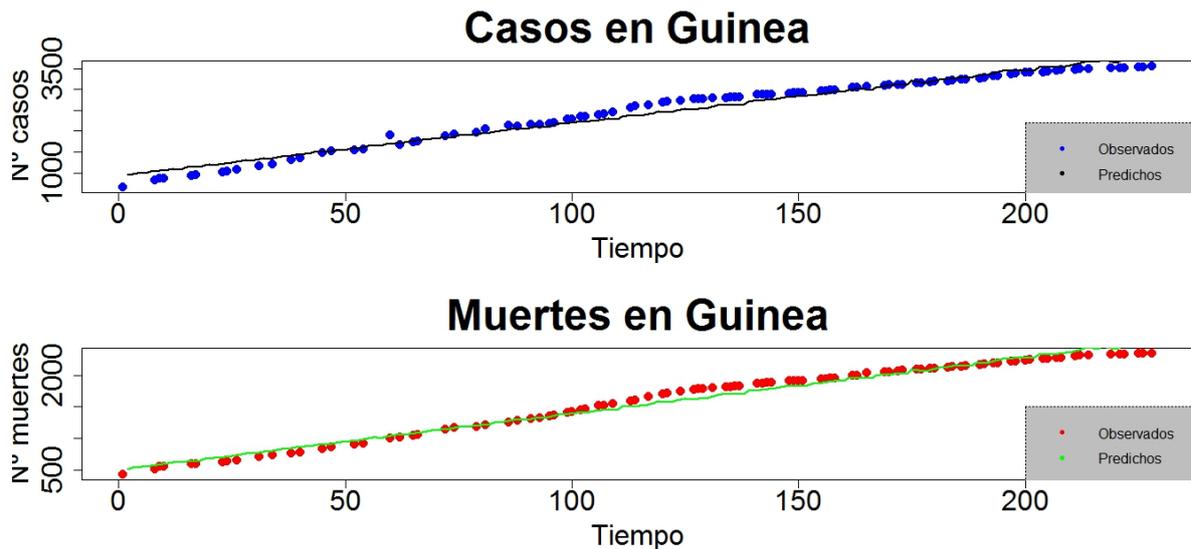
$$R_{adj}^2 = 1 - \frac{MSE}{\sum_{i=1}^n (y_i - \bar{y})^2 / (n - 1)},$$

siendo \bar{y} la media muestral.

Se consideran varios valores de T , y en el caso de Guinea, los resultados proporcionados por $T = 2$ tienen un menor MSE y un mayor R_{adj}^2 :

Casos en Guinea	
R_{adj}^2	0.9699
MSE	140.5

Muertes en Guinea	
R_{adj}^2	0.9807
MSE	78.14



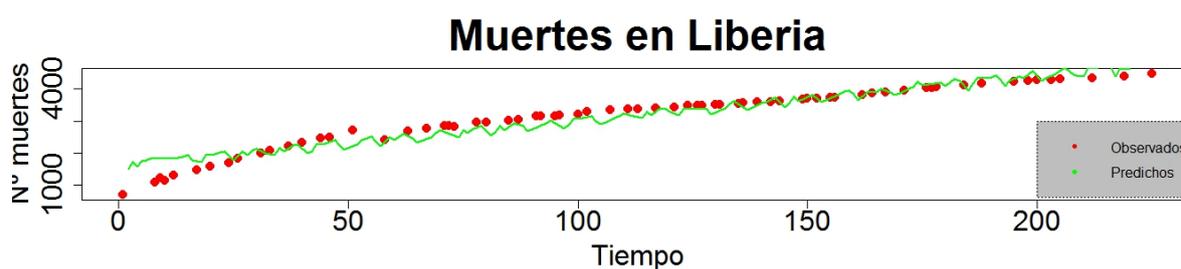
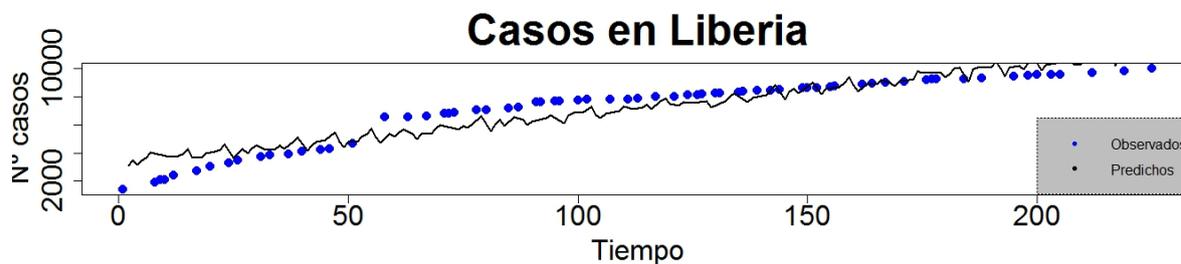
En el caso de Liberia, lo mejor resulta con $T = 4$:

Casos en Liberia

R_{adj}^2	0.8778
MSE	831.5

Muertes en Liberia

R_{adj}^2	0.9336
MSE	238.6



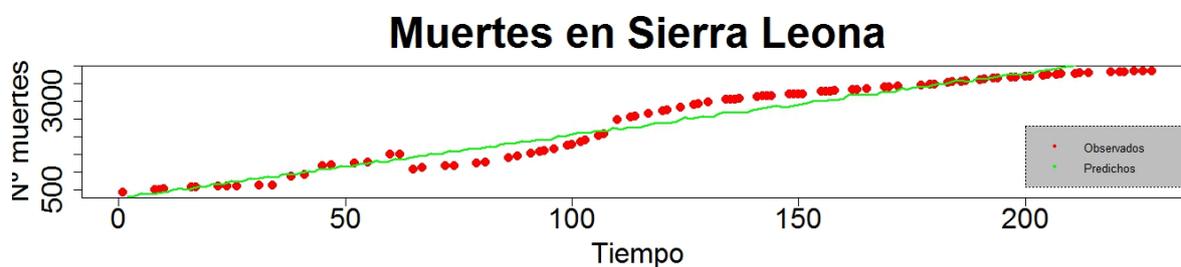
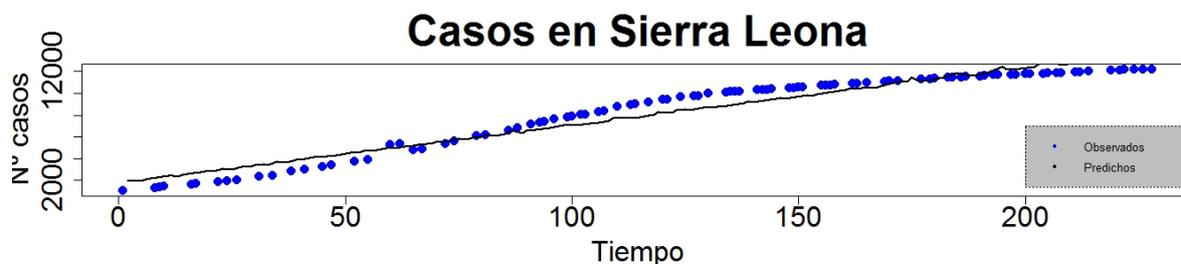
Y por último en el caso de Sierra Leona los mejores resultados los proporciona $T = 2$:

Casos en Liberia

R_{adj}^2	0.9329
MSE	905.5

Muertes en Liberia

R_{adj}^2	0.947
MSE	264.7



Se puede ver que estos modelos se ajustan un poco mejor a la forma de los datos originales. Como el objetivo es la imputación de los valores ausentes, estos modelos proporcionarían serían más apropiados para completar las series.

El tercero de los métodos utilizados consiste en considerar modelos de series temporales. Un proceso estocástico es un conjunto de variables aleatorias $\{z_t, t \in C\}$. Entonces,

una serie temporal es la realización de un proceso estocástico z_t en el que C está ordenado por puntos equidistantes en el tiempo.

Un proceso estacionario o serie estacionaria, en sentido débil, es aquella que verifica:

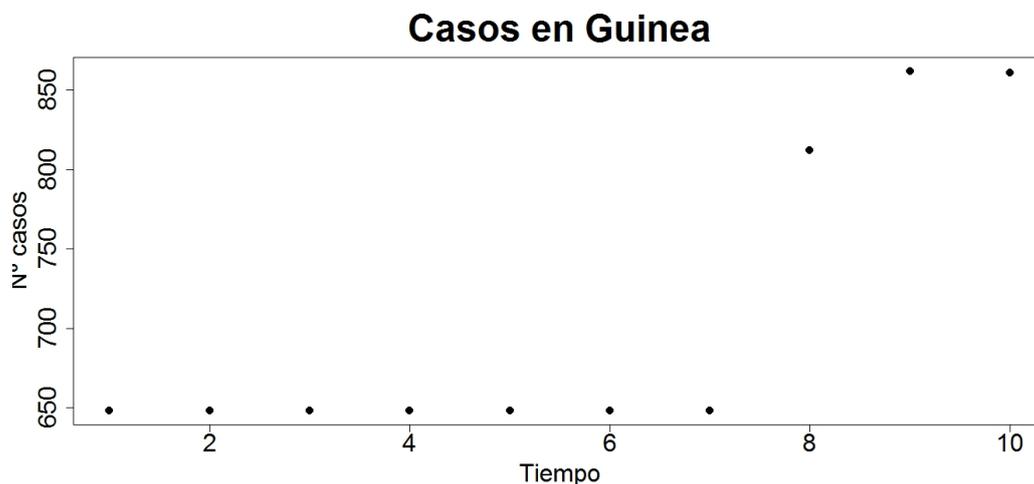
1. $E(z_t) = \mu < +\infty \forall t$
2. $Var(z_t) = \sigma^2 < +\infty \forall t$
3. $Cov(z_t, z_{t+k}) = \gamma_k \forall t \forall k$

Un proceso de ruido blanco, es un proceso estacionario que verifica las siguientes condiciones para todo t :

1. $E(z_t) = 0$
2. $Var(z_t) = \sigma^2$
3. $Cov(z_t, z_{t+k}) = 0 \forall k \neq 0$

Lo primero es considerar la primera parte de la serie temporal que no tiene valores ausentes. Por ejemplo, en el caso de la serie de los casos en Guinea, se considera el primer dato, ya que a partir de éste existen seis ausentes en la serie. Como solo tenemos un dato de partida, estos seis valores ausentes toman el valor del primer dato.

Lo siguiente que se realiza es unir a la serie de siete datos iguales, los tres siguientes valores reales y estudiar las series temporales de diez observaciones.



Como esta serie tiene tendencia creciente, no parece tener media constante, es decir, no parece ser estacionaria en media. Se comprueba analíticamente mediante el test de Dickey-Fuller. En líneas generales y a efectos prácticos, este test considera como hipótesis nula que la serie es no estacionaria y es necesario diferenciar. Diferenciar la serie una vez es considerar una nueva serie temporal

$$x_t = z_t - z_{t-1},$$

que consiste en cada momento tomar la diferencia entre la serie original y la serie en el instante anterior.

El resultado, en el caso de la serie de casos en Guinea, es que a un nivel de significación del 95% se acepta la hipótesis nula y por tanto se necesita diferenciar la serie una vez.

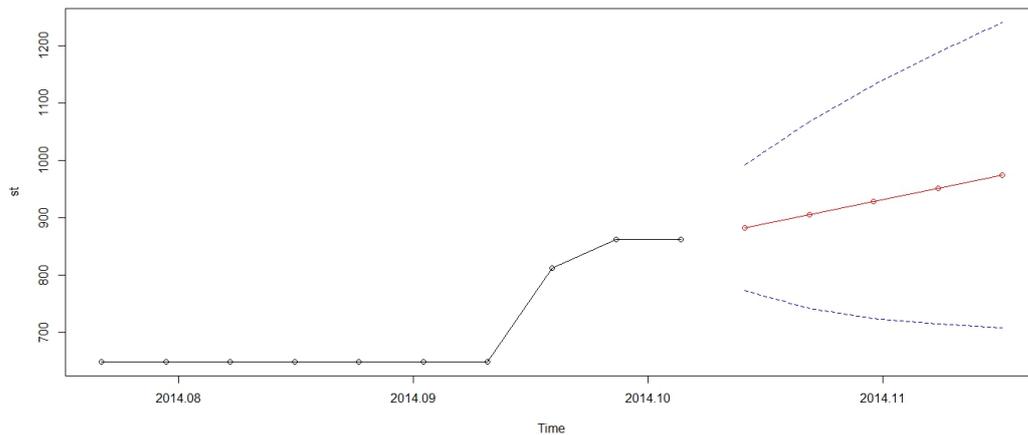
Una vez la serie es estacionaria, se estudia el correlograma de la serie. Por éste, se intuye que la serie sigue un modelo autorregresivo de orden 1, $AR(1)$. Es decir, la serie se puede expresar como

$$x_t = c + \phi x_{t-1} + a_t,$$

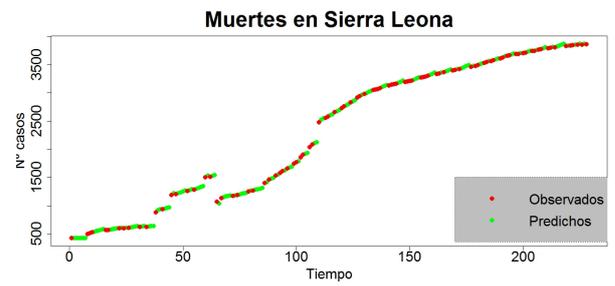
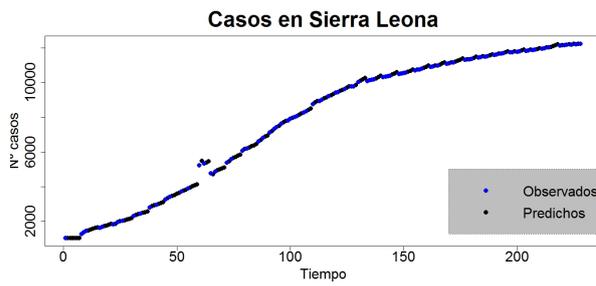
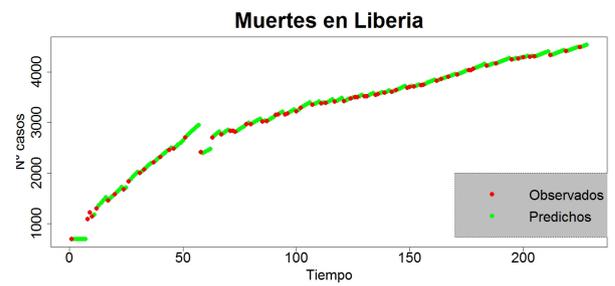
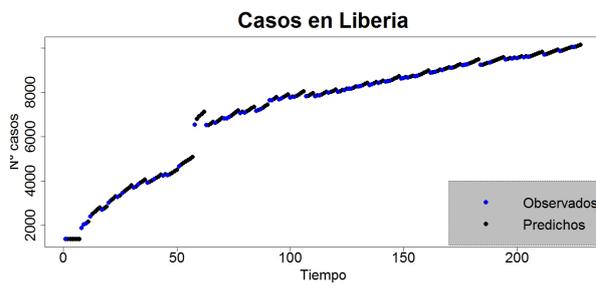
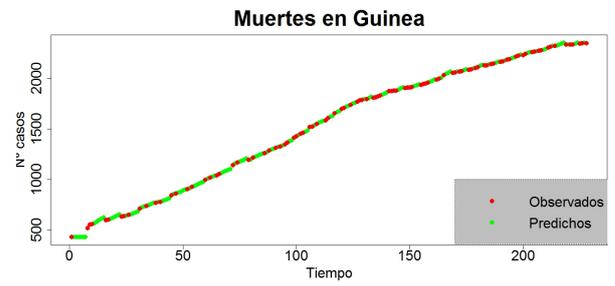
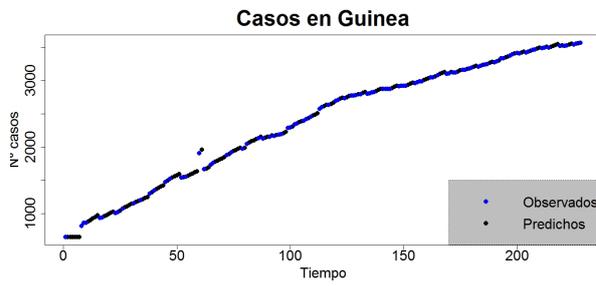
con a_t un proceso de ruido blanco. Lo estudiamos y comprobamos que tanto la constante c como el coeficiente del ϕ son significativos. Comprobando que los residuos del modelo son Ruido Blanco, se puede obtener un buen ajuste.

Así, como los siguientes cinco valores son ausentes, se predicen gracias al modelo ajustado y el resultado en la serie original sería el siguiente:

Casos en Guinea



Siguiendo la misma metodología, el siguiente paso es añadir a esta serie los siguientes valores reales, estudiar qué modelo seguirían y predecir los consecutivos valores perdidos. Y así, hasta conseguir tener todos los datos completos. Se realiza el mismo análisis para la serie de las muertes en Guinea, las dos series de Liberia y las dos series de Sierra Leona. De esta manera los gráficos de las series temporales resultantes son los siguientes:



6 Estimación de los parámetros

$\beta_I(i)$ y κ_i

El parámetro $\beta_I(i) \in \mathbb{R}^+$ es la tasa contacto efectivo de la enfermedad de una persona en estado I en el país i , es decir, el número medio de contactos efectivos de una persona en el estado I por día antes de aplicar las medidas de control.

El parámetro $\kappa_i \in [0, +\infty)$ simula la eficiencia de las medidas de control.

La variable de estimación para el ratio de transmisión $\beta_I(i)$ que se usa es:

$$\frac{GNI_i}{DEN_i},$$

donde $GNI_i \in \mathbb{R}^+$ es el Producto Nacional Bruto del país i y $DEN_i \in \mathbb{R}^+$ es su densidad de población. Se tienen los valores reales para los países Guinea, Liberia y Sierra Leona:

$$\begin{aligned} \left(\frac{GNI_{Guinea}}{DEN_{Guinea}}, \beta_I(Guinea) \right) &= (5,4944, 0,1944), \\ \left(\frac{GNI_{Liberia}}{DEN_{Liberia}}, \beta_I(Liberia) \right) &= (3,0236, 0,2645), \\ \left(\frac{GNI_{SierraLeona}}{DEN_{SierraLeona}}, \beta_I(SierraLeona) \right) &= (4,2422, 0,2605). \end{aligned}$$

En el caso de κ_i , la variable usada para estimar este parámetro de control es

$$\frac{SAN_i}{DEN_i},$$

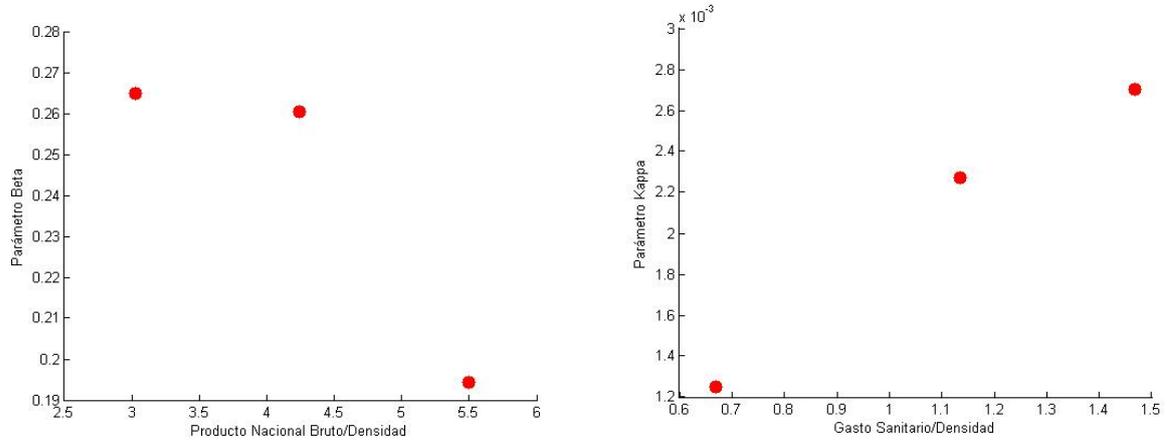
donde SAN_i indica el gasto medio anual en saludo per cápita. Los valores reales para los anteriores países son:

$$\begin{aligned} \left(\frac{SAN_{Guinea}}{DEN_{Guinea}}, \kappa_{Guinea} \right) &= (0,6697, 0,00125), \\ \left(\frac{SAN_{Liberia}}{DEN_{Liberia}}, \kappa_{Liberia} \right) &= (1,4686, 0,0027), \\ \left(\frac{SAN_{Liberia}}{DEN_{Liberia}}, \kappa_{Liberia} \right) &= (1,1344, 0,00227). \end{aligned}$$

Los gráficos de estos tres puntos para cada parámetro son los siguientes:

Parámetro $\beta_I(i)$

Parámetro κ_i



En el caso del parámetro $\beta_I(i)$, se sabe que la forma de las gráficas tiene que ser en los valores menores de $\frac{GNI_i}{DEN_i}$ estable en el valor máximo de $\beta_I(i)$, ya que los países más pobres tienen más alto riesgo de contagio, hasta que llega cierto momento que empieza a decrecer hasta el valor mínimo de $\beta_I(i)$, donde se mantiene estable, ya que llega un momento que aunque el país sea más rico, el riesgo de contagio es igual de bajo.

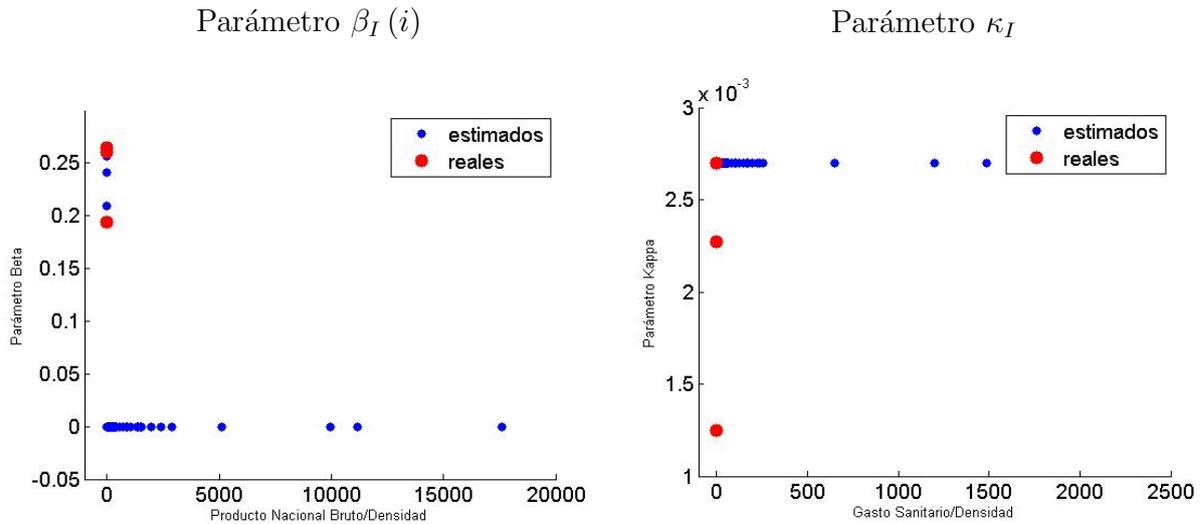
Para el caso del parámetro κ_I , la forma sería al contrario, a valores más bajos de $\frac{SAN_i}{DEN_i}$, es estable en el menor valor de la efectividad las medidas de control, llega un momento que va incrementando la efectividad a medida que aumenta el gasto sanitario por densidad de población, hasta que llega un punto de estabilización.

Los datos que se utilizarán en todas las simulaciones son los reportados por la Organización Mundial de la Salud. Como ya se ha comentado estos datos públicos están disponibles en <http://apps.who.int/ebola/en/current-situation/ebola-situation-report> en [6] y [7].

El primero de los métodos que se considera es la **interpolación lineal**. La interpolación consiste en hallar un dato dentro de un intervalo en el que conocemos los valores en los extremos. En la interpolación lineal se construyen segmentos de rectas entre los datos para estimar los valores desconocidos.

Como se quiere que los parámetros sigan lo descrito anteriormente, a partir de cierto valor de $\frac{GNI_i}{DEN_i}$, para el parámetro $\beta_I(i)$ toma el valor más bajo y se mantiene constante, y para cierto valor de $\frac{SAN_i}{DEN_i}$, κ_I se estabiliza en su valor más alto.

El resultado gráfico de este ajuste es el siguiente:



Los puntos denominados como “reales” son los correspondientes a Guinea, Liberia y Sierra Leona. Con esta estimación de los parámetros para los demás países, comprobamos los resultados que aporta el modelo *Be – CoDiS*:

Outputs

Model alone running time: 9.3851 (s) | Number of iterations/Maximum: 866/4158

Final Date: 19-Apr-2016
 Cumulative number of reported cases: 28524
 Cumulative number of deaths: 11804
 Maximum number of hospitalized persons at the same time: 2234
 Mean model R0 value [Min,Max]: 2.0868 [1.6074,2.3994]
 Number of infected countries: 4
 Characteristics of infected countries:

	Code	DateInf	R0	FCases	FDeaths	ICases	IDeaths	MNH	Dur	TRS
Guinea	60	'06-Dec-2013'	1.8	3919	2411	3	1	180	703	1.1013
Liberia	92	'04-Jun-2014'	2.3	11963	5328	0	0	981	581	0.37089
Nigeria	115	'24-May-2015'	2.4	18	6	0	0	2	175	0.0004381
Sierra Leone	143	'27-Jul-2014'	2.4	12624	4059	0	0	1046	618	0.16227

Be-CODIS total running time: 11 (s)

Aparte de los tres países infectados que hemos tenido en cuenta todo el trabajo, aparece como infectado Gambia con un solo caso.

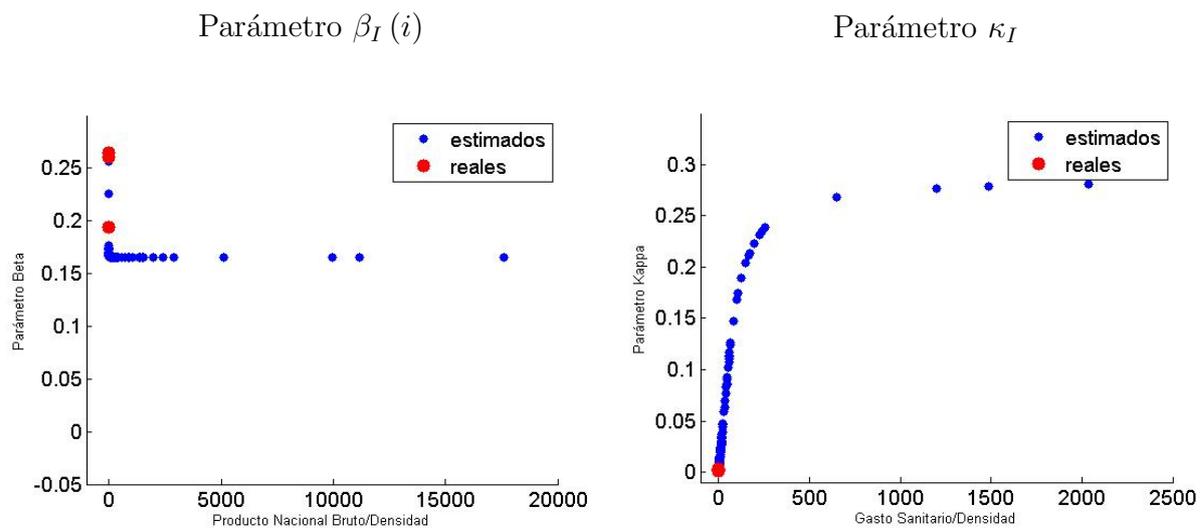
El segundo método que se tiene en cuenta es una **regresión no lineal**. Considerando la forma que debería tener la gráfica de la dependencia entre $\frac{GNI_i}{DEN_i}$ y $\beta_I(i)$ y la forma de la dependencia entre $\frac{SAN_i}{DEN_i}$ y κ_I , se considera un modelo de la siguiente forma

$$y = \alpha_1 \arctan \left(\alpha_2 \cdot \frac{1}{x} + \alpha_3 \right) + \alpha_4,$$

donde x es la variable independiente o regresora, y es la variable dependiente o variable respuesta y α_j $j \in \{1, 2, 3, 4\}$ son los coeficientes de la regresión no lineal.

Para encontrar estos coeficientes de regresión, se usa la función de MATLAB *nlinfit*, programada para realizar ajustes no lineales. Esta función devuelve la estimación de los coeficientes de la regresión no lineal de la variable respuesta y en relación con el regresor x usando el modelo especificado. Los coeficientes son estimados usando una estimación iterativa por mínimos cuadrados, a partir de unos coeficientes iniciales especificados.

El resultado gráfico obtenido mediante esta estimación es el siguiente:



La forma de estas gráficas es más parecida a lo razonado anteriormente que usando la interpolación lineal.

Los resultados que nos ofrece el modelo epidemiológico $Be - CoDiS$ con este ajuste es el siguiente:

Outputs

Model alone running time: 6.5554 (s) | Number of iterations/Maximum: 866/4158

Final Date: 19-Apr-2016
Cumulative number of reported cases: 28607
Cumulative number of deaths: 11851
Maximum number of hospitalized persons at the same time: 2240
Mean model R0 value [Min,Max]: 2.0888 [1.6731,2.3998]
Number of infected countries: 8
Characteristics of infected countries:

	Code	DateInf	R0	FCases	FDeaths	ICases	IDeaths	MNH	Dur	TRS
United Kingdom	57	'06-Nov-2015'	1.5	1	0	0	0	1	0	0.00015337
Guinea	60	'06-Dec-2013'	1.8	3919	2411	3	1	180	703	1.1009
Gambia	61	'12-Jan-2015'	2.4	77	46	0	0	6	291	0.0093254
Liberia	92	'04-Jun-2014'	2.3	11963	5328	0	0	981	581	0.37071
Nigeria	115	'19-May-2015'	2.5	21	7	0	0	2	189	0.00046547
Senegal	141	'11-Feb-2016'	1.8	1	0	0	0	1	0	0.00016529
Sierra Leone	143	'27-Jul-2014'	2.4	12624	4059	0	0	1046	618	0.16222
United States of America	168	'03-Oct-2015'	1.5	1	0	0	0	1	0	8.6548e-05

Be-CODIS total running time: 8 (s)

En este caso, son ocho los países que llegan a ser infectados. A parte de los analizados Guinea, Liberia y Sierra Leona, se infectan: Reino Unido con un solo caso, Gambia con 77 casos, Nigeria con 21 casos, Senegal con un solo caso y Estados Unidos de América con un solo caso.

Estos resultados son más parecidos a la realidad de la epidemia de Ébola de 2014-2015 que se originó en diciembre de 2013 en Guinea, y se extendió posteriormente a Liberia, Sierra Leona, Nigeria, Senegal, Estados Unidos, España, Mali y Reino Unido. Aún así, aparece Gambia que no ha llegado a infectarse en la realidad y no aparece España, que ha llegado a tener un infectado.

El tercer método considerado para la estimación, es la **minimización del error cuadrático medio**, método de optimización con función objetivo no lineal. Considerando que las gráficas deberían seguir la fórmula de la regresión no lineal, se busca minimizar, a través de la función objetivo la diferencia entre los valores reales y los estimados a través del modelo.

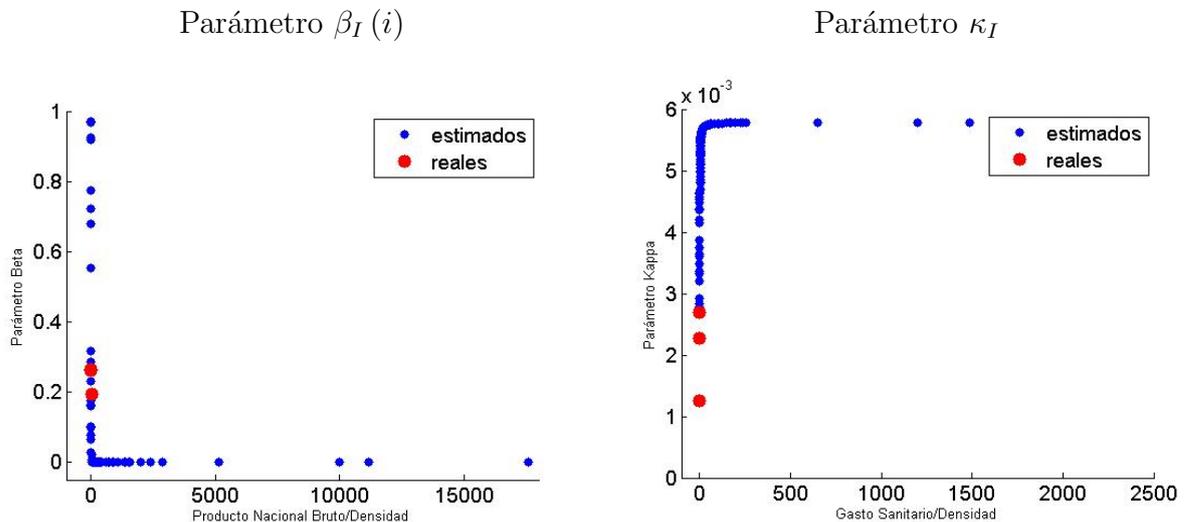
$$\text{mín} \sum_{j=1}^3 \left(y_j - \left(\alpha_1 \arctan \left(\alpha_2 \cdot \frac{1}{x_j} + \alpha_3 \right) + \alpha_4 \right) \right)^2$$

donde j recorre los tres puntos de la muestra. Como única restricción se pide que la diferencia en uno de los tres puntos sea nula.

Para resolver este problema, se usa el paquete *optimtool* de optimización de MATLAB. La función *fmincon* intenta encontrar el mínimo de un problema de programación no lineal y, eligiendo un algoritmo de punto interior con valores iniciales de los coeficientes

$$(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0,1, 0,1, 0,1, 0,1)$$

se obtiene las siguientes gráficas de dependencia:



De esta forma, el resultado que se obtiene a través del modelo *Be – CoDiS* es:

```

Outputs
*****

Model alone running time: 9.4785 (s) | Number of iterations/Maximum: 866/4158

Final Date: 19-Apr-2016
Cumulative number of reported cases: 28507
Cumulative number of deaths: 11798
Maximum number of hospitalized persons at the same time: 2233
Mean model R0 value [Min,Max]: 2.0867 [1.6127,2.3994]
Number of infected countries: 4
Characteristics of infected countries:

```

	Code	DateInf	R0	FCases	FDeaths	ICases	IDeaths	MNH	Dur	TRS
Guinea	60	'06-Dec-2013'	1.8	3919	2411	3	1	180	703	1.1013
Gambia	61	'16-Apr-2016'	1.7	1	0	0	0	1	0	0.00018126
Liberia	92	'04-Jun-2014'	2.3	11963	5328	0	0	981	581	0.37095
Sierra Leone	143	'27-Jul-2014'	2.4	12624	4059	0	0	1046	618	0.16228

```

Be-CODIS total running time: 10 (s)

```

En este caso el número de países infectados es cuatro, los tres principales Guinea, Liberia y Sierra Leona, y Gambia con un solo caso. Este resultado es similar al que proporciona el ajuste a través de la interpolación lineal.

Se compara los tres métodos a través de la salida del modelo:

	Interpolación lineal	Regresión no lineal	Minimización
Fecha final	19/04/2016	19/04/2016	19/04/2016
Casos acumulados	28524	28607	28507
Muertes acumuladas	11804	11851	11798
Máximo hospitalizados	2234	2240	2233
Media <i>RO</i>	2,0868	2,0888	2,0867
Valor <i>RO</i> [máx, mín]	[1,6074, 2,3994]	[1,6731, 2,3998]	[1,6127, 2,3994]
Nº países infectados	4	8	4

El número de casos acumulados y muertes acumuladas difieren un poco en los tres métodos, esto es debido a los distintos países que se obtienen en cada método. Con el método *nlinfit* se obtienen más países, países los cuales en la realidad han llegado a ser infectados.

En los tres países principales del brote (Guinea, Liberia y Sierra Leona), los datos de los casos acumulados y las muertes acumuladas que se obtienen por el modelo a partir de los tres métodos son iguales. Debido a esto, se considera que el mejor método de estimación de ambos parámetros es la regresión no lineal, ya que se parece más a la realidad.

7 Minimización del error

Se considera la parte del modelo relativa a la propagación de la enfermedad entre países. En esta parte, existen ciertos parámetros relativos al movimiento de personas entre países.

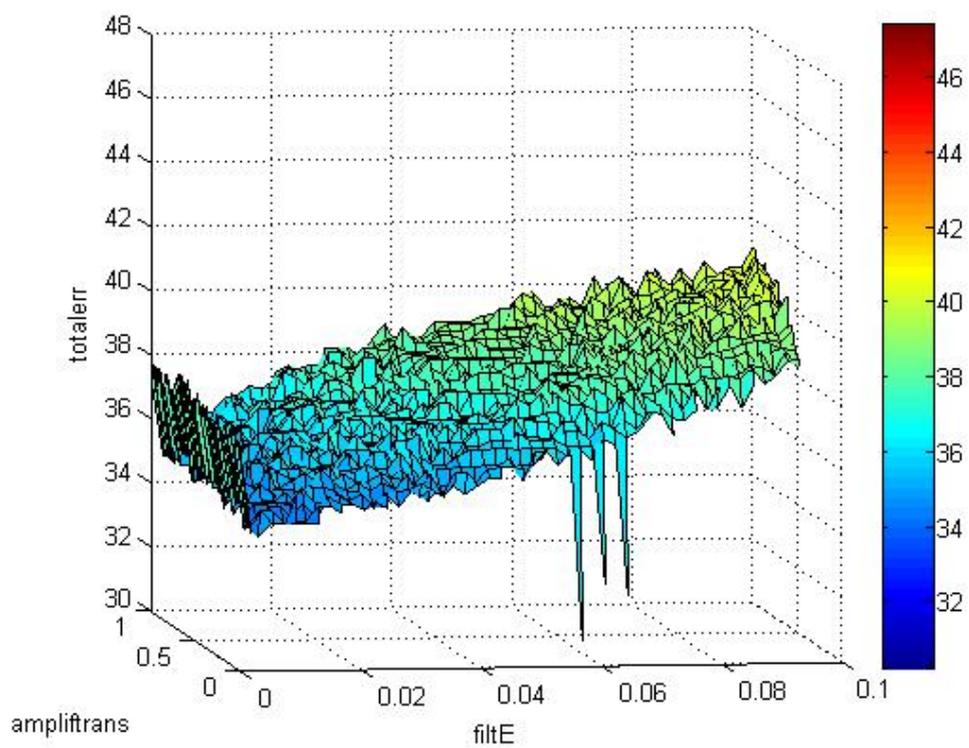
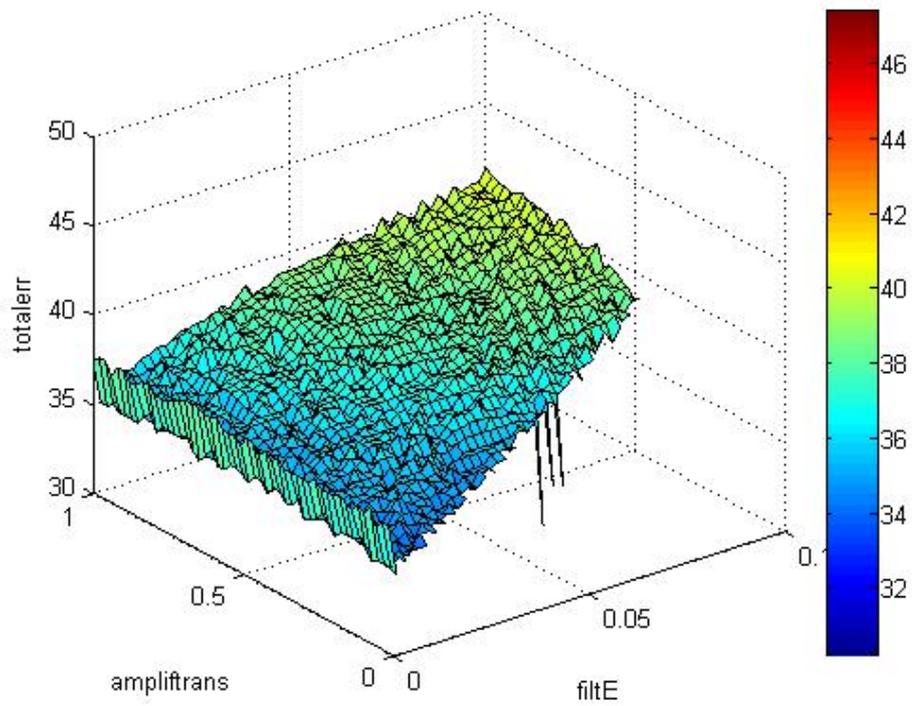
- $filtE$, parámetro de tolerancia, filtro numérico usado para evitar el proceso de infección en países con número bajo de individuos pertenecientes al estado E
- $filtT$, parámetro de tolerancia, filtro numérico para evitar el proceso de circulación entre los países con menor individuos pertenecientes al estado E
- $ampliftrans$, parámetro que multiplica el porcentaje de población migratoria por día de un país a otro. Se usa para incrementar o disminuir el flujo de movimiento.

En todas las simulaciones se ha considerado y se considera $filtE = filtT$.

En general, los valores iniciales que usa el modelo no están disponibles, al igual que la solución explícita del sistema de ecuaciones diferenciales que define cómo varían los estados a lo largo del tiempo según se propaga la enfermedad. Por tanto, para conseguir una aproximación de la solución, se usa una solución numérica adecuada. Por lo que, se comete cierto error, al que se denomina *totalerr*.

El objetivo es minimizar este error dependiente de los tres parámetros anteriormente descritos. Para eso, se estudia la forma que tiene este error, la dependencia que tiene respecto los parámetros. También, al estar este error ponderado según la cantidad de infectados de los países, se quiere ver si esta ponderación hace que este error aumente o disminuya.

Para conocer la forma de dependencia, cómo varía el valor de *totalerr* con respecto a las otras dos variables, se grafica la superficie que representa $filtE$, $ampliftrans$ y *totalerr*. Se obtiene lo siguiente:



Se observa que los valores van disminuyendo a medida que disminuye $filtE$ hasta llegar un momento que vuelven a aumentar. Pero existen tres mínimos relativos pronunciados en medio de la superficie. Por tanto, para encontrar el valor mínimo absoluto se va a utilizar un algoritmo genético.

7.1. Algoritmo genético

El algoritmo genético es una técnica de búsqueda basada en la teoría de la evolución de Darwin, en los mecanismos de la genética y de la selección natural. Es la única metaheurística que trabaja simultáneamente con conjuntos de soluciones factibles, que las considerará como individuos de una población que se cruza, se reproduce y puede incluso mutar para sobrevivir.

Un algoritmo genético consiste en una función matemática que toma como entradas a los individuos de una población y devuelve como salidas cuales de ellos deben generar descendencia para una nueva generación.

Versiones más complejas de algoritmos genéticos generan un ciclo iterativo que directamente toma a la especie (el total de los ejemplares) y crea una nueva generación que reemplaza a la antigua una cantidad de veces determinada por su propio diseño. Una de sus características principales es la de ir perfeccionando su propia heurística en el proceso de ejecución, por lo que no requiere largos períodos de entrenamiento especializado por parte del ser humano.

Se supone que los individuos (posibles soluciones del problema), pueden representarse como un conjunto de parámetros, los cuales agrupados, forman una ristra de valores (a menudo referida como cromosoma). Si bien, el alfabeto utilizado para representar los individuos no debe necesariamente estar constituido por el $\{0, 1\}$, buena parte de la teoría en la que se fundamentan los algoritmos genéticos utiliza dicho alfabeto.

En resumen, el algoritmo durante la fase reproductiva selecciona los individuos de la población para cruzarse y producir descendientes, que constituirán, una vez mutados, la siguiente generación de individuos. La selección de padres se efectúa al azar usando un procedimiento que favorezca a los individuos mejor adaptados, ya que, a cada individuo se le asigna una probabilidad de ser seleccionado que es proporcional a su función de aptitud. Según dicho esquema, los individuos bien adaptados se escogerán probablemente varias veces por generación, mientras que, los pobremente adaptados al problema, no se escogerán más que de vez en cuando.

7.1.1. Operadores básicos

Con los algoritmos genéticos se considera un conjunto de soluciones iniciales como individuos de una especie que evolucionan para conseguir adaptarse y mejorar dicha especie. La calidad de los individuos se mide por la función de aptitud. Las relaciones entre los individuos se ajustan a unos operadores que se analizan a continuación.

Operador de selección

No todos los individuos sobreviven durante un determinado tiempo, solo lo hacen los más aptos, es decir, los que tengan una mejor función de aptitud. Una posibilidad para realizar esto sería considerando que sólo sobreviven la mitad de los individuos siendo la probabilidad de supervivencia de cada individuo proporcional a su función de aptitud. La elección de los individuos se realiza por técnicas de Monte Carlo.

Operador de cruce

Los individuos seleccionados, se pueden cruzar para mejorar la especie. La forma en que se cruzan permite transmitir las cualidades a las generaciones futuras. Es lo que va a distinguir unos algoritmos de otros.

Operador de mutación

Aunque globalmente se mejore, hay determinadas características que no pueden modificarse solo con los dos operadores anteriores. Así, este operador perturba alguna de las soluciones obtenidas.

Puede ser que esta perturbación empeore la aptitud del individuo mutado, pero si este es el caso, los otros dos operadores anularán esta disminución y solo sobrevivirán aquellas mutaciones positivas., Este operador mutación se aplica muy poco, es decir, con una probabilidad muy baja.

Concepto de esquema

Considerando las distintas soluciones del problema de optimización combinatoria como cadenas de caracteres existe un paralelismo con conceptos de genética: cada caracter es un gen, la cadena de caracteres es un cromosoma. De esta forma, las modificaciones de las distintas soluciones se pueden interpretar como modificaciones genéticas producidas por la reproducción cruce y mutación de los genes de los cromosomas.

Con el fin de distinguir las características de un individuo concreto de una generación y las de toda una generación en concreto, se introduce el concepto de esquema.

Formalmente se introduce la siguiente notación:

- V alfabeto $\{0, 1\}$

- k número de caracteres de una cadena
- n tamaño de cada generación de individuos
- t índice de la generación
- $A(t)$ individuos de la generación t -ésima
- $V^+ = V \cup *$ alfabeto extendido
- $H \in V^{+k}$ esquema

Para comparar los esquemas se introducen los siguientes conceptos:

- El orden de un esquema es número de caracteres fijos de dicho esquema. Se representa por $o(H)$
- La longitud de un esquema es la distancia entre las posiciones extremas de los caracteres fijos. Se representa por $\delta(H)$

En el caso de problemas de minimización hay que modificar la función de aptitud para que sean más probables los de menor valor.

7.1.2. Teorema fundamental de los algoritmos genéticos

El teorema estudia la evolución temporal de los esquemas con el fin de analizar el comportamiento de los algoritmos genéticos. Estudia la evolución de un esquema genético H con el tiempo, con este objetivo introduce el número medio de individuos de una población $A(t)$ pertenecientes al esquema H como $m(H, t)$.

Considerando que hay n individuos en la generación t , es decir, $|A(t)| = n$, la probabilidad de que sobreviva en la siguiente generación el individuo i -ésimo con aptitud f_i es

$$p_i = \frac{f_i}{\sum_j f_j}$$

Se denomina $f_T = \sum_{i=1}^n f_i$ la suma de todas las aptitudes de generación t -ésima.

Considerando inicialmente solo el operador selección, que genera un número aleatorio para cada individuo de la población $A(t)$, se investiga cuántos de los $m(H, t)$ individuos del esquema H sobrevivirán en la siguiente generación:

$$m(H, t+1) = \sum_{i=1}^n \{\text{Prob}\{i\text{-ésimo Individuo generado} \in H\}\}$$

La probabilidad de generar un individuo del esquema H es

$$\frac{\sum_{j \in H} f_j}{f_T}$$

y esta probabilidad es independiente del índice i , por lo que la expresión anterior es

$$m(H, t + 1) = n \frac{\sum_{j \in H} f_j}{f_T}$$

Introduciendo la aptitud media de cada individuo de la población \bar{f} y la aptitud media de cada individuo del esquema H como $f(H)$:

$$\bar{f} = \frac{\sum_{i=1}^n f_i}{n} = \frac{f_T}{n}$$

$$f(H) = \frac{\sum_{f \in H} f_j}{m(H, t)}$$

se concluye

$$m(H, t + 1) = n \frac{m(H, t) f(H)}{f_T} = m(H, t) \frac{f(H)}{\bar{f}}$$

Entonces, un esquema con cuya amplitud media supere a la media de la generación incrementará su presencia en las generaciones futuras, por los que sobrevivirán los esquemas más aptos.

Con todo esto, se aplica un algoritmo genético para encontrar la combinación de *filtE* y *ampliftrns* que minimice *totalerr*. Si se considera

$$0 < \text{filtE} < 1 \quad 0,5 < \text{ampliftrns} < 100$$

la solución óptima es

$$\text{filtE} = 5,1680 \cdot 10^{-04} \quad \text{ampliftrns} = 0,5427$$

y

$$\text{totalerr} = 34,3966$$

Por si el algoritmo genético se ha encajado en algún mínimo relativo, para probar si se puede lograr alcanzar un mínimo absoluto, se utiliza la función de MATLAB *fmincon* con un algoritmo de punto interior y como punto inicial la solución obtenida del algoritmo genético. Pero se consigue el mismo resultado, esta optimización última no logra dar otra combinación de parámetros con menor valor.

La salida del modelo con esta combinación de valores es la siguiente:

```

Outputs
*****

Model alone running time: 9.1558 (s) | Number of iterations/Maximum: 913/4158

Final Date: 05-Jun-2016
Cumulative number of reported cases: 27124
Cumulative number of deaths: 11278
Maximum number of hospitalized persons at the same time: 2116
Mean model R0 value [Min,Max]: 2.083 [1.6702,2.4209]
Number of infected countries: 5
Characteristics of infected countries:

```

	Code	DateInf	R0	FCases	FDeaths	ICases	IDeaths	MNH	Dur	TRS
Guinea	60	'06-Dec-2013'	1.8	3937	2422	3	1	181	703	0.29108
Gambia	61	'19-Feb-2015'	2.5	78	47	0	0	6	291	0.0024326
Liberia	92	'15-Jul-2014'	2.3	11143	4962	0	0	913	578	0.089625
Nigeria	115	'12-Jul-2015'	2.5	21	6	0	0	2	187	0.00012789
Sierra Leone	143	'13-Sep-2014'	2.4	11945	3841	0	0	990	616	0.038982

	PEc	PEd	NEc	NEd
Spain	100	100	1	0
United Kingdom	100	100	1	0
Guinea	15	19	57	57
Gambia	93	90	49	29
Liberia	43	40	719	283
Mali	100	100	8	6
Nigeria	68	60	8	2
Senegal	100	100	1	0
Sierra Leone	36	30	305	77
United States of America	100	100	4	1

```

Model total error regarding historic data: 34.3966

```

Los países que llegan a tener alguna persona infectada son Guinea, Liberia, Sierra Leona, Gambia y Nigeria. En la realidad, Gambia no llega a infectarse, y el modelo ha dado a Gambia con 78 casos de infección y 47 muertos. Nigeria, que en el modelo sí se detecta como país infectado, 21 casos infecciosos y 6 muertes, en la actualidad ha llegado a tener 20 casos de infección y 8 muertes, bastante similar.

Como el algoritmo genético tiene una parte aleatoria, se vuelve a simular para ver si esta aleatoriedad afecta en la búsqueda del mínimo. Así, la segunda vez que se consideró con las mismas condiciones iniciales, se obtuvieron los mismos resultados.

Como al estudiar el comportamiento de *totalerr* frente a ambas variables se observó que, respecto *ampliftrans* los valores más bajos se consiguen con valores menores, se considera ahora el algoritmo genético con las siguientes condiciones:

$$0 < filtE < 1 \quad 0 < ampliftrans < 1$$

En este caso, la solución óptima que alcanza es

$$filtE = 5,9046 \cdot 10^{-4} \quad ampliftrans = 0,1654$$

y

$$totalerr = 34,0969$$

De esta manera, hemos conseguido alcanzar otro mínimo relativo, menor que el anterior encontrado. Como antes, volvemos a intentar mejorar mediante la programación

no lineal, pero no se consigue nada distinto.

La salida que proporciona el modelo es la siguiente:

```

Outputs
*****

Model alone running time: 9.4345 (s) | Number of iterations/Maximum: 984/4158

Final Date: 15-Aug-2016
Cumulative number of reported cases: 26576
Cumulative number of deaths: 11073
Maximum number of hospitalized persons at the same time: 2060
Mean model R0 value [Min,Max]: 2.0851 [1.7465,2.4762]
Number of infected countries: 5
Characteristics of infected countries:

```

	Code	DateInf	R0	FCases	FDeaths	ICases	IDeaths	MNH	Dur	TRS
Guinea	60	'06-Dec-2013'	1.7	3941	2425	3	1	181	703	0.090033
Gambia	61	'13-Jul-2015'	2.6	76	46	0	0	6	290	0.00072598
Liberia	92	'09-Sep-2014'	2.3	10897	4853	0	0	893	578	0.026951
Nigeria	115	'15-Aug-2016'	2.7	1	0	0	0	1	0	2.7473e-06
Sierra Leone	143	'24-Nov-2014'	2.4	11661	3749	0	0	966	614	0.011142

	PEc	PEd	NEc	NEd
Spain	100	100	1	0
United Kingdom	100	100	1	0
Guinea	15	19	58	57
Gambia	92	88	44	26
Liberia	42	39	628	249
Mali	100	100	8	6
Nigeria	3	0	0	0
Senegal	100	100	1	0
Sierra Leone	36	31	375	95
United States of America	100	100	4	1

```

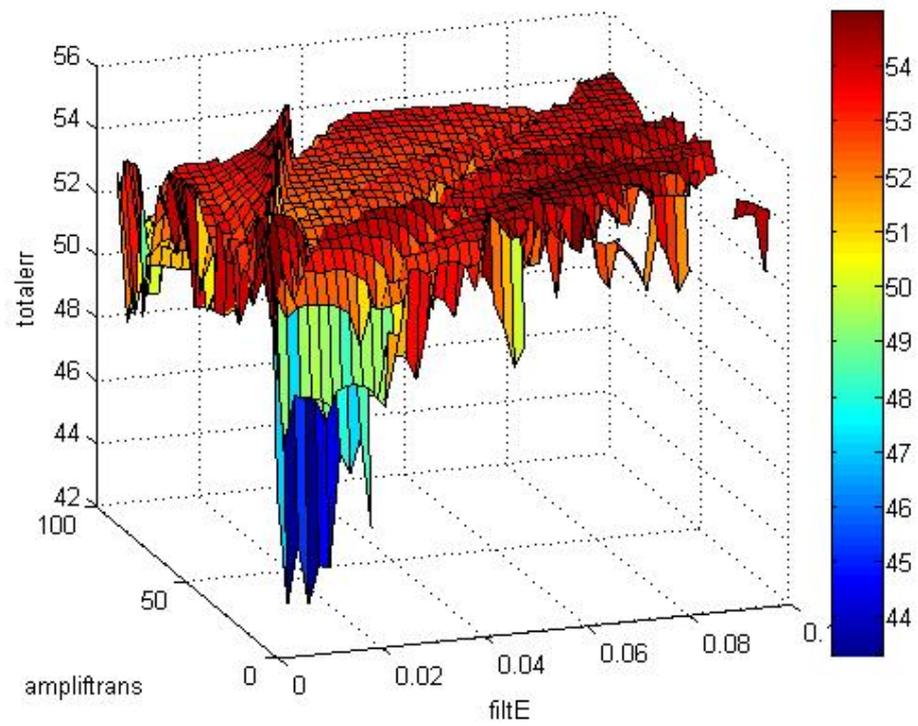
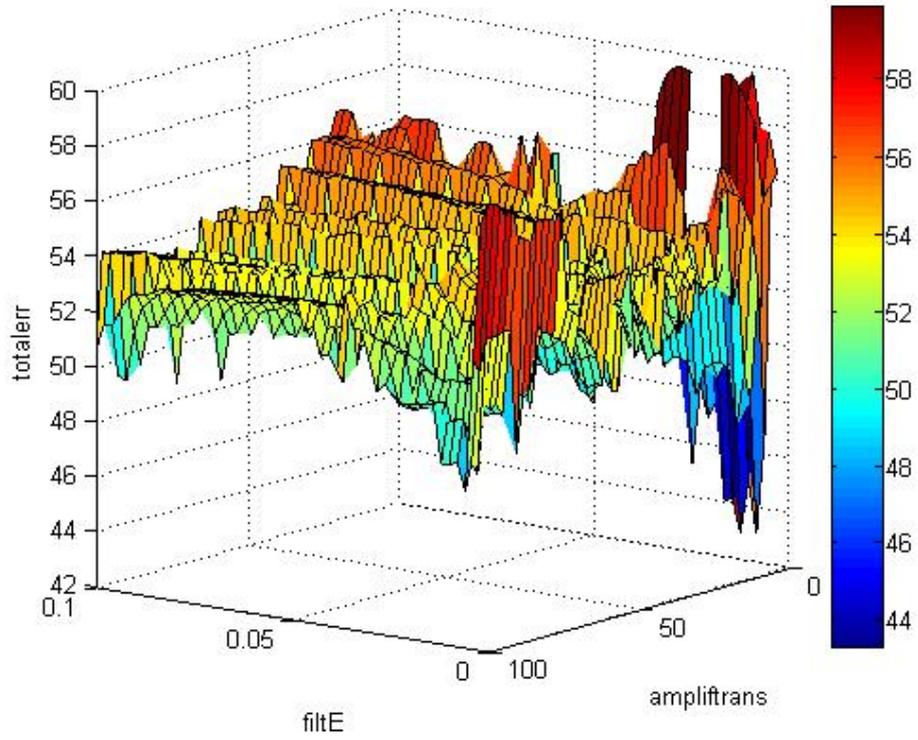
Model total error regarding historic data: 34.0969

```

Con estas condiciones, el modelo da cinco países infectados. Cuatro de ellos sí que se infectan en la realidad y además, no proporciona cinco de los infectados realmente. Por tanto, no es adecuado utilizar estos valores para hacer simulaciones y estudiar la evolución de la enfermedad.

Este error que se ha estudiado está ponderado por los países infectados. De tal manera que si un país tiene más casos de contagio, tendrá más peso en la cuantía del error que otro con menos casos infectados.

Se realiza el mismo procedimiento para ver si se obtendría un error menor sin tener en cuenta las personas infectadas en la cuantía del error. La superficie que forma graficar *filtE*, *ampliftrans* y *totalerr* sin esta ponderación es la siguiente:



Así, con el algoritmo genético en las condiciones iniciales de

$$0 < filtE < 1 \quad 0 < ampliftrans < 1$$

se llega a alcanzar un mínimo relativo cuando:

$$filtE = 6,4412 \cdot 10^{-4} \quad ampliftrans = 0,17901$$

y

$$totalerr = 68,45$$

Este error es bastante más alto que lo anterior estudiado. Aún así comprobamos, con otra simulación idéntica, si la parte aleatoria del algoritmo genético ha llegado a afectar. El resultado de la segunda simulación es el mismo, por tanto, en estas condiciones iniciales el algoritmo genético solo llega hasta este punto. Se ve la salida que aporta el modelo:

```

Outputs
*****

Model alone running time: 9.7321 (s) | Number of iterations/Maximum: 981/4158

Final Date: 12-Aug-2016
Cumulative number of reported cases: 26736
Cumulative number of deaths: 11146
Maximum number of hospitalized persons at the same time: 2074
Mean model R0 value [Min,Max]: 2.0853 [1.7473,2.4753]
Number of infected countries: 5
Characteristics of infected countries:

```

	Code	DateInf	R0	FCases	FDeaths	ICases	IDeaths	MNH	Dur	TRS
Guinea	60	'06-Dec-2013'	1.7	3941	2425	3	1	181	703	0.097448
Gambia	61	'12-Jul-2015'	2.6	77	46	0	0	6	291	0.00078897
Liberia	92	'07-Sep-2014'	2.3	11075	4932	0	0	908	578	0.029643
Nigeria	115	'11-Aug-2016'	2.7	1	0	0	0	1	0	3.0735e-06
Sierra Leone	143	'21-Nov-2014'	2.4	11642	3743	0	0	965	614	0.012075

	PEc	PEd	NEc	NEd
Spain	100	100	1	0
United Kingdom	100	100	1	0
Guinea	15	19	58	57
Gambia	92	88	44	26
Liberia	43	40	694	274
Mali	100	100	8	6
Nigeria	5	0	0	0
Senegal	100	100	1	0
Sierra Leone	36	31	379	96
United States of America	100	100	4	1

```

Model total error regarding historctic data: 68.45

```

En este caso son cuatro los países contagiados: Guinea, Liberia, Sierra Leona y Gambia. Gambia en la realidad no llega a infectarse, y el modelo no consigue detectar ningún otro país de los que en la actualidad ha tenido contagios.

Si se cambiasen las condiciones iniciales y se buscara en un rango mayor de *ampliftrans* se obtendría con

$$0 < filtE < 1 \quad 0,5 < ampliftrans < 100$$

un mínimo relativo en

$$filtE = 0,0319 \quad ampliftrans = 47,0913$$

y

$$totalerr = 49,8548$$

Error menor que en el caso anterior, pero mayor que en el caso que se ponderan los países infectados. Si se observa el gráfico de la superficie, se ve que hay valores menores que 49, pero el algoritmo genético no llega a ellos.

Así, el modelo proporciona la salida:

Model alone running time: 5.7082 (s) | Number of iterations/Maximum: 773/4158

Final Date: 17-Jan-2016
 Cumulative number of reported cases: 41639
 Cumulative number of deaths: 16745
 Maximum number of hospitalized persons at the same time: 3337
 Mean model R0 value [Min,Max]: 2.1228 [1.5174,2.4948]
 Number of infected countries: 29
 Characteristics of infected countries:

	Code	DateInf	R0	FCases	FDeaths	ICases	IDeaths	MNH	Dur	TRS
Australia	8	'15-Dec-2014'	1.2	1	0	0	0	1	0	0.0026655
Belgium	12	'11-May-2015'	1.4	1	0	0	0	1	0	0.00045771
Benin	13	'19-Aug-2015'	2	4	2	0	0	1	0	0.0012999
Canada	27	'21-Dec-2014'	1.2	1	0	0	0	1	0	0.0013199
Switzerland	28	'10-Jul-2015'	1.5	1	0	0	0	1	0	0.0019756
Germany	41	'24-Jan-2015'	1.2	1	0	0	0	1	0	0.0021502
Spain	49	'17-Nov-2014'	1.1	1	0	0	0	1	0	1.1724e-05
France	55	'22-Sep-2014'	1	1	0	0	0	1	0	0.0032225
United Kingdom	57	'19-Sep-2014'	1.1	1	0	0	0	1	0	0.005061
Ghana	59	'27-Dec-2015'	1.8	1	0	0	0	1	0	0.00036652
Guinea	60	'06-Dec-2013'	1.8	3464	2133	3	1	158	697	18.25
Gambia	61	'09-Oct-2014'	2.4	89	53	0	0	7	300	0.063919
Guinea Bissau	62	'10-Apr-2015'	1.7	2	1	0	0	1	0	0.0012023
Greenland	65	'22-Dec-2014'	1.2	1	0	0	0	1	0	0.0086819
Israel	78	'02-Oct-2015'	1.6	1	0	0	0	1	0	0.0013768
Italy	79	'05-Jan-2015'	1.2	1	0	0	0	1	0	0.00012091
Liberia	92	'10-Mar-2014'	2.3	18734	8343	0	0	1535	600	9.2165
Mexico	102	'17-Oct-2015'	1.6	1	0	0	0	1	0	8.2833e-05
Mauritania	109	'15-Feb-2015'	1.3	1	0	0	0	1	0	0.00048786
Nigeria	115	'17-Dec-2014'	2.4	26	8	0	0	3	208	0.0050676
Netherlands	117	'14-Dec-2014'	1.2	1	0	0	0	1	0	0.0043412
Norway	118	'19-Aug-2015'	1.5	1	0	0	0	1	0	0.00025775
Portugal	130	'20-Jan-2015'	1.2	1	0	0	0	1	0	0.00062867
Palestine	132	'02-Oct-2015'	1.6	1	0	0	0	1	0	0.0013768
Senegal	141	'24-Nov-2014'	1.5	3	1	0	0	1	0	0.004389
Sierra Leone	143	'15-Apr-2014'	2.4	19296	6204	0	0	1601	636	5.6179
Sweden	151	'12-May-2015'	1.4	1	0	0	0	1	0	0.00094627
United States of America	168	'04-Sep-2014'	1	1	0	0	0	1	0	0.0022012
South Africa	174	'25-May-2015'	1.4	1	0	0	0	1	0	0.00033133

Se llegarían a infectar 29 países, veinte países más que los llegados a infectar en la realidad. Esto difiere bastante de la realidad, o sea, que aunque disminuya el error

con respecto a la simulación anterior, el resultado no es nada ajustado a la realidad. Por tanto, no interesan estos valores y es preferible tomar el error ponderado, donde las salidas son más verídicas.

8 Conclusiones

En este trabajo de fin de máster, se ha presentado y estudiado el modelo epidemiológico espacio-temporal $Be - CoDiS$. La principal característica del modelo es la combinación de los efectos del flujo migratorio entre los países y las medidas de control, y el uso de coeficientes dinámicos. El modelo fue validado a corto y largo plazo en el artículo de B. Ivorra, D. Ngom y A. M. Ramos [4], y reprodujo de manera razonable la evolución de la epidemia.

La Organización Mundial de la Salud reporta frecuentemente el número de personas infectadas a causa de la enfermedad del virus del Ébola y el número de muertes en los países Guinea, Liberia y Sierra Leona. Se quiere los datos diarios de estas dos series para cada uno de los países, por tanto, en este trabajo se han desarrollado tres métodos de imputación de los valores ausentes: regresión lineal simple, regresión lineal múltiple y modelos de series temporales.

De estos tres métodos, el resultado más adecuado es considerar modelos de series temporales. Así, se ajustan los primeros datos reales a un modelo y se predicen los siguientes valores ausentes. Se repite el proceso hasta el final. Esto hace que la serie completa sea más parecida a la realidad y que sea en todo momento creciente, que así debería ser al ser datos acumulados.

Otro objetivo del trabajo, era estudiar las distintas formas de estimar dos parámetros del modelo; uno, la tasa de contagio efectivo de la enfermedad de una persona infectada y otro, el parámetro que simula la efectividad de las medidas de control. Se estudian tres métodos distintos: interpolación lineal, optimización y regresión no lineal.

Los resultados más adecuados, los más parecidos a la realidad, los proporciona el modelo $Be - CoDiS$ con estos dos parámetros estimados mediante la regresión no lineal. De esta forma, la salida del modelo da ocho países infectados (Guinea, Liberia, Sierra Leona, Nigeria, Senegal, Estados Unidos, España, Mali y Reino Unido); siete de los cuales han llegado a infectarse en la realidad. Y el número de casos de personas infectadas y de muertos debido a la enfermedad están más próximos a lo real.

El último objetivo que se tenía era estudiar el error que comete el modelo. Este error está ponderado según el número de casos infectados de cada país. Estudiándolo, sin esta ponderación el error resulta mayor. Por tanto, considerando la ponderación se logra, mediante un algoritmo genético, alcanzar un mínimo relativo. Además, el modelo solo obtiene cinco países infectados con los valores del mínimo relativo (uno de ellos en la realidad no llega a infectarse); más lejos de la realidad que con los valores que se simulaba anteriormente y se utilizaban en el artículo de B. Ivorra, D. Ngom y A. M. Ramos [4].

Bibliografía

- [1] Introducción histórica: <http://www.who.int/mediacentre/factsheets/fs103/es/>
- [2] Introducción histórica: <http://temas.sld.cu/ebola/historia/>
- [3] Introducción histórica: <http://www.quimicaviva.qb.fcen.uba.ar/v13n3/petrera.html>
- [4] B. IVORRA, D. NGOM AND A.M. RAMOS, *Be-CoDiS: A mathematical model to predict the risk of human diseases spread between countries. Validation and application to the 2014 Ebola Virus Disease epidemic. Accepted for publication in Bulletin of Mathematical Biology, ISSN: 0092-8240.*, Preprint: <http://arxiv.org/abs/1410.6153>
- [5] H. W. HETHCOTE, *The basic epidemiology models: models, expressions for R_0 , parameter estimation, and applications*
- [6] Datos reportados: <http://apps.who.int/gho/data/node.ebola-sitrep.ebola-summary?lang=en>
- [7] Datos reportados: <http://apps.who.int/ebola/en/ebola-situation-reports-archive>
- [8] Estimación de los parámetros: <http://es.mathworks.com/help/stats/nlinfit.html>
- [9] Países infectados: <http://espanol.cdc.gov/enes/vhf/ebola/outbreaks/2014-west-africa/distribution-map.html>
- [10] Países infectados: <http://www.isglobal.org/ebola>
- [11] Algoritmo genético: <http://eddyalfaro.galeon.com/geneticos.html>