

# Alternativas en la estimación del número básico de reproducción de una enfermedad.

Autores:

Gustavo Nicolás Páez<sup>1\*</sup>, Juan Felipe Cerón, Santiago Cortés<sup>2</sup>, Adolfo J. Quiroz<sup>3</sup>, José Fernando Zea, Camila Franco, Érica Cruz, Gina Vargas, Carlos Castañeda<sup>4\*</sup>.

## Abstract

Se consideran cuatro métodos para la estimación de la tasa de contagio de virus, que ofrecen la ventaja de proporcionar intervalos de confianza para la tasa y para el número de nuevas infecciones ocurridas. Uno de estos métodos, propuesto en Thompson et al. (2013), es un método de estimación bayesiana, bien establecido en la práctica epidemiológica. De los modelos nuevos propuestos aquí, es un modelo de espacio de estados, mientras que los otros dos modelos propuestos se basan en la simulación, partiendo de una modelación condicional Poisson del proceso infeccioso. Se discuten ventajas y dificultades de los métodos considerados, tanto desde el punto de vista teórico como desde el punto de vista del ajuste de parámetros en la práctica.

## Introducción

Durante la pandemia del COVID-19, la tasa efectiva de reproducción “R” se ha convertido en una de las principales herramientas de política pública para el comprender la transmisión de la enfermedad y, posteriormente, planificar. La importancia de esta métrica radica en su capacidad de estimar, durante el periodo de infección, cuantas personas susceptibles pueden ser contagiadas por una persona enferma. Debido a este potencial, varios gobiernos (Zhao et al. (2020), Dickens et al. (2020), “The R number” (2020)), incluido el de Colombia, usan este valor para modular las restricciones de movilidad de las personas con miras a evitar incrementos en las tasas de contagio que saturen la capacidad hospitalaria.

Pese a su popularidad, no existe un criterio unificado para el cálculo de esta medida y los modelos actuales son sensibles tanto a los parámetros biológicos de la enfermedad como al contexto social donde se desarrollan (Adam 2020). Por tal motivo, este estudio tiene como objetivo el desarrollo de cuatro modelos estadísticos para el cálculo de la tasa efectiva de reproducción que capturan la complejidad biológica y social del comportamiento de la infección.

De esta forma, se parte del modelo de Cori et al. (2013) que al ser extendido por Thompson et al. (2013) e implementado en múltiples paquetes estadísticos como Python o R, ha gozado de una amplia popularidad en los tiempos de pandemia. Una vez se han analizado los supuestos, ventajas y limitaciones de este modelo, este estudio propone tres modelos alternativos que, desde la teoría,

---

<sup>1</sup> Myanmar Development Institute, Myanmar. Información de contacto: [gn.paez145@uniandes.edu.co](mailto:gn.paez145@uniandes.edu.co)

<sup>2</sup> Factored.ai, Colombia

<sup>3</sup> Universidad de los Andes, Colombia

<sup>4</sup> Instituto Nacional de Salud, Colombia. Información de contacto de Carlos Castañeda: [ccastanedao@ins.gov.co](mailto:ccastanedao@ins.gov.co)

corrigen las principales falencias del modelo previamente mencionado. Adicionalmente, el estudio presenta ocho escenarios de pandemias simuladas donde los distintos modelos buscan estimar los parámetros teóricos. Este ejercicio permite realizar una comparación empírica de las ventajas y desventajas de la aplicación de cada modelo. Finalmente, una vez se ha concluido la evaluación tanto teórica como empírica de estos modelos, este estudio presenta los resultados de los cálculos de la tasa de reproducción para los primeros cinco meses de la pandemia en Colombia.

## Modelos de estimación de las tasas de reproducción

Esta sección tiene como objetivo describir desde la teoría y la calibración, cuatro modelos alternativos para la estimación de la tasa de reproducción. Para tal fin, la sección se divide en tres partes. La primera parte se enfoca en conceptualizar la tasa de reproducción y los alcances de esta definición. La segunda parte plantea dos de los modelos, los cuales parten del concepto del “tiempo-desde-infección” y mediante técnicas de series de tiempo estiman las tasas de reproducción. Finalmente, la tercera parte plantea dos modelos basados en procesos de simulación estocástica a partir de los cuales las dinámicas de contagio en la enfermedad son descritas y donde las tasas de reproducción se derivan vía la calibración de estos modelos para ajustarse a los casos observados.

### Tasas de reproducción

Uno de los elementos más importantes para comprender la evolución de una epidemia es la tasa de reproducción. Este indicador busca medir cuantos individuos susceptibles pueden ser infectados por una persona enferma durante el tiempo en que ésta se encuentra en un estado infeccioso. Ahora bien, este valor se puede dividir en tres componentes (van den Driessche, 2017):

1. Duración del periodo de infección.
2. Probabilidad que la infección pase de una persona infecciosa a una persona susceptible.
3. Número promedio de veces, por unidad de tiempo, que un individuo infeccioso se encuentra a un individuo susceptible durante el día.

Esta descomposición permite evidenciar que este índice es influenciado por factores biológicos de la enfermedad (que determinan por cuanto tiempo la persona es infecciosa y parte de la facilidad en que la enfermedad pasa entre personas) y factores sociales (que determinan el tipo y frecuencia de las interacciones entre personas). De ahí se derivan dos elementos relevantes para tener en cuenta para la modelación de la tasa. Por un lado, esta tasa varía según el contexto y la idiosincrasia de la población de estudio. Debido a esto, extrapolar valores a zonas geográficas distintas puede ser inapropiado. Por otro lado, la variación también es temporal y es afectada por elementos tales como políticas de cuarentena, distanciamiento social y cambios en la población susceptible.

Autores como Fraser (2007) y Cori et al. (2013) proponen diferenciar dos tipos de tasas de reproducción. La primera tasa, identificada como la tasa de reproducción efectiva,  $\mathcal{R}(t)$ , está asociada a el número medio de personas que alguien que empieza su infección en el día  $t$ , infectará si los tres factores previamente mencionados se mantienen constantes durante el periodo de infección. Por otro lado, está la tasa de reproducción de caso,  $\mathcal{R}_C(t)$ , la cual busca medir el número de personas que fueron infectados por alguien que terminó su periodo de infección en el día  $t$ . De esta forma, el primer indicador tiene un énfasis prospectivo y de característica contrafactual, a partir del cual se plantean escenarios tales como: *ceteris paribus* la tasa de reproducción es  $\mathcal{R}(t)$ , y por ende, si no se ejercen cambios en el contexto social, el número de infectados que se estiman será  $X$ . Por otro lado, el  $\mathcal{R}_C(t)$  presenta un enfoque retrospectivo y de monitoreo a partir del cual se puede visualizar, de forma directa, la influencia que han tenido cambios sociales (usualmente impulsados por políticas públicas) en el desarrollo de la enfermedad. Finalmente, ambos indicadores

permiten entender si la epidemia se está expandiendo (cuando estos indicadores son mayores a 1) o se está reduciendo (cuando estos indicadores son menores a 1). Hechas estas aclaraciones, las siguientes secciones presentan los modelos mediante los cuales estos indicadores se van a medir.

Es preciso aclarar que, para los cuatro modelos que se presentan a continuación, la definición matemática asociada a un símbolo dado puede presentar variaciones. Sin embargo, esto se hace para resaltar aquellas variables que definen mecánicas de transmisión análogas entre modelos.

## Modelos de tiempo desde infección

En esta sección se presentan dos modelos basados en los cálculos originales de Kermack-McKendrick (1927). Esta estrategia de modelación se elige dado que sus elementos son intuitivos (Fraser, 2007) y han sido sistemáticamente usados para la evaluación de las epidemias más recientes (Thompson et al., 2019). Por tal motivo, el principal modelo de referencia de esta sección es el planteado por Cori et al. (2013) y extendido por Thompson et al. (2013) que sintetiza el estado actual de esta serie de modelos. Dada su forma de estimación, este modelo se va a denominar modelo Bayesiano. Una vez sus elementos principales han sido descritos, se pasa a discutir las limitaciones del modelo y proponer un modelo alternativo, denominado modelo Estado Espacio, y discutir sus ventajas y limitaciones.

### Modelo Bayesiano

Sea  $I_N(t)$  el número de nuevos individuos infecciosos en el periodo ( $t$ ) siendo un valor estocástico que depende del número de personas infecciosas en periodos anteriores, ajustado por el potencial de estos individuos para infectar a alguien. De esta forma, se define la ecuación de renovación como<sup>5</sup>:

$$\mathbb{E}[I_N(t)] = \sum_{\tau=1}^{\infty} \beta(t, \tau) I_N(t - \tau) \quad (1)$$

La ecuación (1) debe entenderse como la esperanza condicional de  $I_N(t)$  dadas las cantidades de nuevos infecciosos en períodos anteriores. A manera de ejemplo, de esta ecuación surge la noción que, por una persona que se infectó hace 10 periodos ( $I_N(t - 10)$ ), se espera que  $\beta(t, 10)$  personas sean infecciosas el periodo  $t$ . Ahora bien, se considera que el factor de ajuste  $\beta(t, \tau)$ , se puede descomponer, de forma multiplicativa, en un factor que depende del tiempo,  $\phi(t)$ , y el cual está influenciado por las características sociales asociadas a la tasa de reproducción, y un factor dependiente del rezago,  $\omega(\tau)$ , asociado netamente al comportamiento biológico de la enfermedad. De esta forma, se tiene que:

$$\beta(t, \tau) = \phi(t)\omega(\tau) \quad (2)$$

A partir de la lógica sugerida por Wallinga y Teunis (2004), este modelo interpreta  $\omega(\tau)$  como el intervalo serial de la enfermedad. Es decir,  $\omega$  está asociado a una distribución de probabilidad donde  $\omega(\tau)$  es la probabilidad de que transcurran  $\tau$  unidades de tiempo desde que una persona se vuelve infecciosa hasta que una segunda persona, infectada por la primera, se vuelve infecciosa. De esta interpretación Wallinga y Teunis (2004) derivan que el número de personas que pueden ser infectadas por un individuo es:

---

<sup>5</sup> A menos que se aclare lo contrario, los valores esperados presentados en este modelo son condicionados a la historia. Sin embargo, esto no se hace explícito en la notación para evitar que esta se sobrecargue de caracteres.

$$\mathcal{R}(t) = \sum_{\tau=1}^{\infty} \beta(t, \tau) = \phi(t) \quad (3)$$

Al combinar estas ecuaciones, se deriva que

$$\mathcal{R}(t) = \frac{\mathbb{E}[I_N(t)]}{\sum_{\tau=1}^{\infty} \omega(\tau) I_N(t - \tau)} \quad (4)$$

y

$$\mathbb{E}[I_N(t)] = \mathcal{R}(t) \sum_{\tau=1}^{\infty} \omega(\tau) I_N(t - \tau) \quad (5)$$

Finalmente, y siguiendo la lógica de Fraser (2007),

$$\mathcal{R}_c(t) = \sum_{\tau=1}^{\infty} \omega(\tau) \mathcal{R}(t - \tau) \quad (6)$$

Ahora bien, asumiendo conocimiento previo del intervalo serial de la enfermedad, el problema de calibración se reduce a la estimación de  $\mathbb{E}[I_N(t)]$ .

Para tal fin, la estimación Bayesiana asume que el número de casos infecciosos nuevos en el periodo  $t + 1$  sigue una distribución de Poisson de media indicada por la ecuación (5). Adicional a esto, el modelo asume que, condicional a estar en el periodo  $t$ , y por una ventana de tiempo  $w$ , la tasa de reproducción no va a cambiar. Bajo estas premisas, se tiene que:

$$\sum_{t'=t-w}^t I_N(t') \sim \text{Poisson} \left( \mathcal{R}(t) \sum_{t'=t-w}^t \sum_{\tau=1}^{\infty} \omega(\tau) I_N(t' - \tau) \right) \quad (7)$$

Ahora bien, asumiendo una distribución *a priori* de  $\mathcal{R} \sim \text{Gamma}(\gamma_1, \gamma_2)$  donde  $\gamma_1$  es el parámetro de forma y  $\gamma_2$  es el parámetro de razón, se obtiene un modelo *gamma-Poisson* donde

$$\mathcal{R}(t) \sim \text{Gamma} \left( \gamma_1 + \sum_{t'=t-w}^t I_N(t'), \gamma_2 + \sum_{t'=t-w}^t \sum_{\tau=1}^{\infty} \omega(\tau) I_N(t' - \tau) \right) \quad (8)$$

Según Fraser (2007), las principales ventajas de este modelo son la facilidad de sus cálculos especialmente de los intervalos seriales. Sin embargo, el cálculo de estos genera problemas conceptuales para la estimación del  $\mathcal{R}(t)$ . Tal como lo plantea el autor, el cálculo del intervalo serial se ve afectado por las normas sociales. Debido a esto, asumir intervalos seriales fijos en escenarios donde las interacciones entre personas han sido afectadas por políticas públicas puede afectar la capacidad del  $\mathcal{R}(t)$  de capturar estos cambios. Adicional al problema conceptual, la necesidad de determinar una ventana de estimación  $w$  genera una suavización forzada en la cual diferentes periodos se promedian. Por esta razón,  $\mathcal{R}(t)$  no captura exclusivamente la capacidad de infección de un día determinado, sino que incluye, por construcción, una mezcla de estos valores en los periodos anexos.

### Modelo de Estado-espacio

Buscando solucionar las dos limitaciones planteadas por el modelo anterior, el modelo de Estado-espacio redefine las ecuaciones (1)-(6) para generar un cálculo en el cual  $\omega(\tau)$  capture únicamente los componentes biológicos de la enfermedad, y luego pasa de una estimación estadística Bayesiana

a una basada en series de tiempo conocida como modelo estado-espacio para evitar los problemas generados por la ventana de estimación.

Sea  $f_{Inc}$  una función de probabilidad, donde  $f_{Inc}(t)$  es la probabilidad de que un individuo se tome  $t$  días desde que fue expuesto a la enfermedad hasta el inicio de una etapa infecciosa. De la misma manera, sea  $F_{Inf}$  una función de probabilidad acumulada, donde  $F_{Inf}(t)$  es la probabilidad que una persona sea infecciosa por, a lo más,  $t$  periodos. A partir de estas definiciones, sea

$$\omega(t) = \sum_{\tau=1}^t f_{Inc}(\tau) (1 - F_{Inf}(t - \tau)) \quad (9)$$

En (9) se está redefiniendo  $\omega(t)$  siendo la probabilidad de que una persona que fue infectada hace  $t$  periodos, continúe siendo infecciosa el día de hoy. Mas aún,  $\Omega = \sum_{\tau=1}^{\infty} \omega(\tau)$  es el tiempo esperado en que una persona dura siendo infecciosa<sup>6</sup>.

Hechos estos ajustes, se tiene que, la ecuación 1, se modifica para considerar los tiempos de incubación y se define como

$$\mathbb{E} \left[ \sum_{\tau=1}^{\infty} f_{Inc}(\tau) I_N(t + \tau) \right] = \sum_{\tau=1}^{\infty} \beta(t, \tau) I_N(t - \tau) \quad (1')$$

en donde hay una relación directa entre el potencial número de infectados y el potencial de infección de las personas infectadas en el pasado. La parte izquierda de la ecuación usa la lógica de Wallinga y Teunis (2004) para estimar a partir de las personas que empiezan a ser infecciosas en el periodo  $I_N(t + \tau)$ , cuantas tuvieron un periodo de incubación  $\tau$ ; i.e. cuantas personas adquirieron la enfermedad el día  $t$ . Por otro lado, la parte derecha de la ecuación muestra cuantas personas fueron infectadas el día  $t$  por alguien que empezó a ser infeccioso el día  $t - \tau$ .

Al combinar estas ecuaciones, se deriva que

$$\mathcal{R}(t) = \frac{\mathbb{E}[\sum_{\tau=1}^{\infty} f_{Inc}(\tau) I_N(t + \tau)]}{\sum_{\tau=1}^{\infty} \omega(\tau) I_N(t - \tau)} \Omega \quad (4')$$

Nótese que este resultado presenta características análogas al modelo Bayesiano. En este caso, y acorde al concepto,  $\mathcal{R}(t)$  está asociado a la tasa de contagio en un día dado,  $\frac{\mathbb{E}[\sum_{\tau=1}^{\infty} f_{Inc}(\tau) I(t+\tau)]}{\sum_{\tau=1}^{\infty} \omega(\tau) I(t-\tau)}$  multiplicada por el número de días que la persona puede infectar, en promedio,  $\Omega$ .

$$\mathbb{E} \left[ \sum_{\tau=1}^{\infty} f_{Inc}(\tau) I_N(t + \tau) \right] = \frac{\mathcal{R}(t)}{\Omega} \sum_{\tau=1}^{\infty} \omega(\tau) I_N(t - \tau) \quad (5')$$

Ahora bien, para calibrar este modelo, se asume, al igual que en el caso anterior, que el proceso de contagio es Poisson de media dada por la ecuación (5).

$$\sum_{\tau=1}^{\infty} f_{Inc}(\tau) I_N(t + \tau) \sim \text{Poisson} \left( \frac{\mathcal{R}(t)}{\Omega} \sum_{\tau=1}^{\infty} \omega(\tau) I_N(t - \tau) \right) \quad (7')$$

Sin embargo, en vez de calibrar este modelo en un marco Bayesiano, se considera el siguiente modelo de estado espacio no Gaussiano (donde la función de enlace es, naturalmente Poisson). Para el caso del modelo Poisson, la función de enlace (link function) permite modelar la variable dependiente (número de nuevas personas infectadas) como la realización de una distribución de Poisson cuya media está en función de las variables de interés, en este caso, las nuevas personas infectadas de los periodos pasados tal que:

<sup>6</sup> Esta afirmación es una conclusión directa de la convolución sobre la función de supervivencia.

$$\mathbb{E} \left[ \sum_{\tau=1}^{\infty} \beta_{Inc}(\tau) I_N(t + \tau) \right] = e^{Z_t \alpha_t}$$

donde

$$Z_t = \left[ \ln \left( \sum_{\tau=1}^{\infty} \omega(\tau) I_N(t - \tau) \right), 1, 0 \right]$$

$$\alpha_t = \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \end{bmatrix}$$

$$T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R = \begin{bmatrix} 0 & 0 \\ \sigma_{\alpha_2} & 0 \\ 0 & \sigma_{\alpha_3} \end{bmatrix}$$

con condiciones iniciales<sup>7</sup>:

$$\alpha_1 = \begin{bmatrix} 1 \\ \mu_0 \\ \mu_0 \end{bmatrix}, P_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 10^9 & 0 \\ 0 & 0 & 10^9 \end{bmatrix}$$

De esta forma, el modelo se puede resumir en:

Ecuación de observación:

$$\mathbb{E} \left[ \sum_{\tau=1}^{\infty} \beta_{Inc}(\tau) I_N(t + \tau) \right] = e^{\alpha_{2,t}} \left( \sum_{\tau=1}^{\infty} \omega(\tau) I_N(t - \tau) \right)$$

Ecuación de transición:

$$\alpha_{t+1} = c_t + T_t \alpha_t + R_t \eta_t$$

Donde  $c_t$  es un vector conocido y a menudo omitido.  $T_t$  y  $R_t$  son sistemas de matrices descritas anteriormente que dependen de parámetros desconocidos. Igualmente  $\eta_t \sim N(0, I_k)$  y  $\alpha_1 \sim N(\alpha_1, P_1)$ .

De manera explícita,

$$\alpha_{2,t+1} = 1$$

$$\alpha_{2,t+1} = \alpha_{2,t} + \alpha_{3,t+1} + \eta_{1,t}$$

$$\alpha_{3,t+1} = \alpha_{3,t} + \eta_{2,t}$$

que describe un modelo con pendiente estocástica (Harvey, 1990) (con término estocástico  $\eta_{i,t}$ ) y para el cual los parámetros no especificados:  $\sigma_{\alpha_2}$  y  $\sigma_{\alpha_3}$  se calculan vía máxima verosimilitud.

Ahora bien, para adaptar esta técnica al modelo epidemiológico, se tiene que,

$$\mathcal{R}(t) = e^{\alpha_{2,t} \Omega} \quad (10)$$

Y  $\mathcal{R}_c(t)$  se calcula a través de la ecuación (6) ajustada por  $\Omega$ :

<sup>7</sup> Nótese que la varianza del prior de  $\alpha_{1,1}$  es 0 porque este parámetro no se desea que cambie en la regresión. De hecho, se desea que sea 1 desde el inicio y se mantenga. La varianza grande de los otros valores se da para resaltar que no se asume información específica sobre el prior, dándole oportunidad al modelo de ajustarse rápidamente a los datos. Finalmente,  $\mu_0$  describe  $\mathcal{R}_0$  de la literatura epidemiológica, pero se le da la libertad de variar para ajustarse al contexto específico del área de análisis.

$$\mathcal{R}_c(t) = \sum_{\tau=1}^{\infty} \omega(\tau) \frac{\mathcal{R}(t - \tau)}{\Omega} \quad (6')$$

Como se puede observar de la deducción expuesta, este procedimiento de estimación logra los mismos resultados que el modelo Bayesiano, evitando los dos problemas teóricos mencionados en la sección anterior. Ahora bien, esta propuesta trae consigo un nuevo reto, ya que se cambia la necesidad de conocer el intervalo serial por el requisito de conocer o estimar la distribución de los tiempos de incubación y de infección. Por tal motivo, si bien desde la teoría este modelo presenta mejoras con respecto al modelo anterior, en la práctica, su elección está condicionada a información que se tenga sobre estas características de la enfermedad.

## Modelos de simulación de eventos

Los dos modelos planteados en la sección anterior coinciden en calibrar un proceso de Poisson a partir del cual se modela el contagio. Sin embargo, en vez de profundizar en la dinámica del proceso, estos modelos reducen el proceso a una ecuación de enlace entre la secuencia histórica de las observaciones. Por tal motivo, en la presente sección se cambia el enfoque y se presenta un análisis centrado en la modelación del proceso. En este punto, el reto pasa de ser estadístico a probabilístico, y debido al componente de simulación entran a ser relevantes consideraciones como la capacidad computacional que demanda la estimación de parámetros y los retos asociados con la calibración de éstos. El primer modelo planteado en esta sección presenta una dinámica general, cuyo ajuste puede poner fuertes demandas sobre la capacidad computacional. Por lo que el segundo modelo limita al primero en sus primitivas desarrollando soluciones analíticas que reducen los tiempos de calibración. De esta manera, el primer modelo se denomina Poisson General, mientras que el segundo se denomina Poisson Exponencial dadas las características de las restricciones adicionales.

### Modelo Poisson General

Buscando una notación consistente con la sección anterior, se definen las siguientes entidades y eventos:

$I(t) :=$  el número de personas que están en una etapa infecciosa en el periodo  $t$ .

$I_N(t) :=$  el número de personas que están en una etapa de infección en el periodo  $t$  pero que no estaban en una etapa de infección en el periodo  $t - 1$  (i.e. nuevos casos infecciosos).

$E(t) :=$  el número de personas que están en una etapa incubación en el periodo  $t$ .

$E_N(t) :=$  el número de personas que están en una etapa de incubación en el periodo  $t$  pero que no estaban en una etapa de incubación en el periodo  $t - 1$  (i.e. nuevos casos expuestos).

Se asume que el tiempo de incubación distribuye  $\mathcal{F}_{inc}$ , y el tiempo de infección distribuye  $\mathcal{F}_{inf}$ , donde estas distribuciones son independientes del tiempo y de los individuos.

$\beta(t) :=$  el numero esperado de personas que un individuo infeccioso contagia en un día  $t^8$ .

---

<sup>8</sup> Se elige  $\beta(t)$  pues es similar en concepto al parámetro homónimo definido en la sección anterior. Sin embargo, en este caso, no se pone el rezago ya que su valor está asociado al potencial de un individuo en el periodo  $t$  y no al potencial de infección de individuos en el pasado que todavía pueden ser infecciosos en el periodo  $t$ .

### Descripción del proceso de contagio desde el punto de vista del evento

Inicie el proceso con  $E_N(0) > 0$  individuos expuestos;  $\forall \tau \geq 1, I(\tau) = I_N(\tau) = E(\tau) = 0$ ; y sea  $t = 0$ .

Paso 1:

$\forall i \in \{1, \dots, E_N(t)\}$  genere dos realizaciones  $X_{inc}(i), X_{inf}(i)$  de las distribuciones  $\mathcal{F}_{inc}$  y  $\mathcal{F}_{inf}$  respectivamente, y actualice los valores de:

- $E(h) \rightarrow E(h) + 1$  para todo  $h \in \{t + 1, \dots, t + X_{inc}(i)\}$
- $I_N(t + X_{inc}(i) + 1) \rightarrow I_N(t + X_{inc}(i) + 1) + 1$
- $I(h) \rightarrow I(h) + 1$  para todo  $h \in \{t + X_{inc}(i) + 1, \dots, t + X_{inc}(i) + X_{inf}(i)\}$

Paso 2:

Actualice  $E_N(t + 1)$  como la realización de una variable de Poisson con media  $\beta(t)I(t)$ .

Paso 3:

Actualice  $t \rightarrow t + 1$  e itere desde el paso 1.

Finalmente, para hacer los cálculos de las tasas análogas a la sección anterior, se define  $\Omega$  como el valor esperado de  $\mathcal{F}_{inf}$ . De esta forma,

$$\mathcal{R}(t) = \beta(t)\Omega \quad (4'')$$

$$\mathcal{R}_c(t) = \sum_{\tau=1}^{\infty} \omega(\tau)\mathcal{R}(t - \tau) \quad (6'')$$

### Calibración del proceso:

Dadas las distribuciones  $\mathcal{F}_{inc}$  y  $\mathcal{F}_{inf}$  y el número de casos expuestos iniciales  $E_N(0)$ , surge el reto de calibrar a través de los datos, los valores asociados a  $B = \{\beta(1), \beta(2), \dots, \beta(T)\}$ . Esta calibración se hace en dos partes. La primera parte desarrolla el criterio de ajuste y la segunda corrige por problemas asociados al sobreajuste de la serie.

Para todo  $t \in \{1, \dots, T\}$ , sea  $\widehat{I}_N(t)$  el número observado de personas que empiezan a ser infecciosas en el periodo  $t$ . Por otro lado, sea  $f(x: B)$  la fracción de simulaciones del proceso Poisson donde  $I_N(t) = x$  dados los parámetros  $B$ . De esta forma la log-verosimilitud<sup>9</sup> asociada a la realización de la serie de infección es

$$ll(B) = \sum_{t=1}^T \ln(f(\widehat{I}_N(t): B)) \quad (11)$$

En principio, sería posible estimar directamente el estimador de máxima verosimilitud a partir de la ecuación (11). Sin embargo, dada la estructura de los datos, este método tiende a sobreajustar el modelo y por ende cualquier error en la observación es asumido como un valor del modelo. Por tal motivo es preciso ajustar la verosimilitud por un factor de regularización de los  $\beta(t)$  que evite el sobreajuste. Para este objetivo, se propone la siguiente metodología. Defínase la función de costo:

$$C(\lambda; B) = -ll(B) + \frac{\lambda}{T-2} \sum_{t=3}^T \left( \frac{\beta(t-1)}{\beta(t)} - \frac{\beta(t-2)}{\beta(t-1)} \right)^2 \quad (12)$$

De esta forma, a mayor valor de  $\lambda$ , la función genera un mayor costo por cambios de orden no lineal y que puedan ser debidos a sobreajustes. Por ende, condicional al  $\lambda$ , se tiene que:

$$B_{opt}(\lambda) = \underset{B}{argMin}(C(\lambda; B)) \quad (13)$$

<sup>9</sup> La función de log-verosimilitud presentada es una aproximación dada a partir del número de simulaciones que se hagan del proceso. No obstante, dada la ley de los grandes números, es conocido que esta estimación converge a la verosimilitud del proceso.



Finalmente, si bien el factor de regularización  $\lambda$  es un meta parámetro, es posible sugerir un criterio de elección de este. Nótese que, bajo el supuesto que el proceso generador de datos es  $B_{opt}(\lambda)$ , entonces

$$2 \left( ll \left( B_{opt}(0) \right) - ll \left( B_{opt}(\lambda) \right) \right) \sim \chi_T^2 \quad (14)$$

La ecuación (14) proviene de tratar, para algún  $\lambda$  candidato,  $B_{opt}(\lambda)$ ,  $t \leq T$ , como el vector de  $\beta(t)$  verdadero y observar que, cuando se calcula la verosimilitud para  $B_{opt}(0)$  se está maximizando libremente en un universo de  $T$  parámetros y usar el teorema de distribución de cociente de verosimilitudes. Para introducir suavizamiento, se usa un criterio de selección asociado a una probabilidad alta  $p$ , por ejemplo, 95%, y se elige  $\lambda_{opt}$  entre los que cumplan  $2 \left( ll \left( B_{opt}(\lambda) \right) - ll \left( B_{opt}(\lambda_{opt}) \right) \right) \leq \chi_{T,0.95}^2$  (mientras más cercano esté el lado izquierdo de la desigualdad al cuantil 95% de una  $\chi_T^2$ , mayor es el suavizamiento introducido). Conceptualmente, este criterio permite identificar el factor de regularización que da cierto peso a la suavización (reduciendo el sobreajuste), sin permitir un distanciamiento estadístico significativo entre la serie observada (que se obtiene en el modelo completamente sobre ajustado), con la serie estimada.

### Modelo Poisson Exponencial

Si bien el proceso de Poisson General permite la inclusión de múltiples funciones de distribución, el proceso de simulación puede requerir una gran capacidad computacional y depende de grandes muestras para que las aproximaciones hechas en el proceso de calibración tengan valor estadístico. Debido a estas limitaciones, en la presente sección se particulariza el modelo anterior para poder generar soluciones analíticas que preserven la mayoría del potencial del modelo de simulación reduciendo la demanda computacional. Para tal fin, se hacen las siguientes modificaciones.

$\tau_1 :=$  Sea  $e^{-\tau_1}$  la probabilidad que una persona que esté incubando el virus en el periodo  $t$  pase a ser infecciosa en el periodo  $t + 1$ .

$\tau_2 :=$  Sea  $1 - e^{-\tau_2}$  la probabilidad que una persona que esté infecciosa en el periodo  $t$  deje de serlo en el periodo  $t + 1$ .

Conceptualmente, estas dos funciones limitan  $\mathcal{F}_{inc}$  y  $\mathcal{F}_{inf}$  a dos procesos con propiedad de no memoria. A partir de este cambio, las siguientes derivaciones surgen de forma directa:

$$\mathbb{E}[I(t)] = e^{-\tau_2} \mathbb{E}[I(t-1)] + (1 - e^{-\tau_2}) \mathbb{E}[E(t-1)] \quad (15)$$

$$\mathbb{E}[E(t)] = \beta(t-1) \mathbb{E}[I(t-1)] + e^{-\tau_1} \mathbb{E}[E(t-1)] \quad (16)$$

$$\mathbb{E}[I_N(t)] = (1 - e^{-\tau_1}) \mathbb{E}[E(t-1)] \quad (17)$$

$$\mathbb{E}[E_N(t)] = \beta(t-1) \mathbb{E}[I(t-1)] \quad (18)$$

De esta forma, una vez se definen las condiciones iniciales, las ecuaciones 15-18 describen la dinámica del sistema y permiten calcular rápidamente el estado de este en cualquier periodo.

Mas aún, dado que el proceso es aditivo y sin memoria,  $I_N(t+1) \sim Poisson(\mathbb{E}[I(t)])$ . Finalmente, dado que este proceso es una instancia específica del Poisson General, las ecuaciones (4'') y (6'') siguen describiendo la generación de las tasas de reproducción.

### Calibración del proceso

Dado los resultados analíticos de la sección anterior se puede calcular la log-verosimilitud de la serie así:

$$ll(B) = \sum_{t=1}^T \left( \mathbb{E}[I(t)] \ln(\widehat{I_N(t)}) - \mathbb{E}[I(t)] - \ln(\widehat{I_N(t)!}) \right) \quad (11')$$

De este punto en adelante, las ecuaciones (12)-(14) se mantienen como en el modelo Poisson General.

### Generación de intervalos de confianza

A diferencia de los modelos de tiempo-desde-infección, los modelos de simulación no traen una fórmula incluida para la generación de intervalos de confianza sobre los parámetros estimados. Por esta razón, se ha implementado un procedimiento del cálculo de estos intervalos basado en la técnica de remuestreo (Bootstrap). Para ese estudio, el remuestreo es una adaptación de la técnica planteada por Davison y Hinckley (1999).

Como antes, para  $t = 1, \dots, T$ ,  $\widehat{I}_N(t)$  representa el número de individuos infecciosos observados a tiempo  $t$  e  $I_N(t)$  es el número correspondiente obtenido al minimizar el costo dado en la ecuación (12). Para cada  $t$  en el rango dado, definimos el residuo

$$e(t) = I_N(t) - \widehat{I}_N(t) \quad (19)$$

Para un entero  $v > 0$ , y  $v \leq t \leq T - v$ , se define el promedio local (móvil)

$$\bar{I}_v(t) = \frac{1}{2v+1} \sum_{\tau=t-v}^{t+v} \widehat{I}_N(\tau) \quad (20)$$

y el estimador local de varianza

$$s^2(t) = \frac{1}{2v} \sum_{\tau=t-v}^{t+v} (\widehat{I}_N(\tau) - \bar{I}_v(t))^2 \quad (21)$$

Para todas las simulaciones presentadas en este estudio,  $v$  se fija a 2. Sin embargo, distintos ejercicios hechos por el equipo de trabajo mostraron que la estimación es robusta a cambios en este parámetro.

Para los valores extremos de  $t$ , las ecuaciones (20) y (21) se adaptan omitiendo los términos necesarios. Como la serie de  $\widehat{I}_N(t)$  es heterocedástica, los residuos se estandarizan haciendo

$$r(t) = e(t)/s(t) \quad (22)$$

El remuestreo procede de la siguiente manera. El entero  $b_T$  (típicamente  $T = 1000$ ) denota el número de realizaciones del remuestreo. Para cada  $b$  en  $1, \dots, b_T$  se muestrea  $r_b^*(1), \dots, r_b^*(T)$  al azar, con reemplazo, de  $\{r(1), \dots, r(T)\}$  y así se determinan las observaciones de la  $b$ -ésima muestra bootstrap como

$$\widehat{I}_b^*(t) = I_N(t) + r_b^*(t)s(t) \quad (23)$$

A partir de los  $\widehat{I}_b^*(t)$  se ajusta nuevamente el modelo, usando otra vez (11) y (12), obteniendo, de esa manera, para cada  $t$ , un estimador  $I_b^*(t)$ . Para cada  $t$ , la colección de estimadores  $I_1^*(t), \dots, I_{b_T}^*(t)$  permite extraer cuantiles que definen el intervalo de confianza para  $I_N(t)$ .

Finalmente, para cada una de las muestras de Bootstrap se procede a hacer la estimación de los parámetros del modelo de contagio. Nótese que cada etapa de la estimación de los datos del proceso es continua y, por ende, al aplicar el teorema de la envolvente, las propiedades de los diferentes  $\beta(t)$  generados por el remuestreo son apropiadas para definir los intervalos de confianza a partir de las mismas.

## Análisis de simulación

Esta sección compara los modelos descritos a partir de la calidad de su ajuste sobre ocho escenarios teóricos asumiendo formas funcionales del  $\mathcal{R}(t)$  que han sido simulados siguiendo la lógica del proceso de infección. Las simulaciones provienen de un modelo discreto en el cual los individuos

pasan de una etapa de incubación a una de infección en el siguiente periodo con probabilidad  $F_{inc}$ , y de una infecciosa a una no infecciosa con probabilidad  $F_{inf}$ . Además, siguiendo el planteamiento del Modelo Poisson General,

$$E_N(t + 1) \sim \text{Poisson}(\beta(t)I(t)).$$

De este modo, se simulan series de incidencia  $I_N$  dada una serie de transmisibilidad  $\beta(t)$  o equivalentemente, de una serie de números de reproducción  $\mathcal{R}(t)$ .

La selección de hiperparámetros de suavización (ventanas temporales para el modelo bayesiano, y  $\lambda$  para el Poisson General) fue realizada a partir de inspección visual. Al final de esta sección se presentan las observaciones pertinentes a este proceso.

Adicional a lo anterior, los parámetros usados para estas simulaciones fueron:

1. Ventana del modelo Bayesiano: 7
2. Número máximo de rezagos: 7 en las ecuaciones (7) y (7')
3. Para el proceso de simulación,  $F_{inc}$  y  $F_{inf}$  se estiman como distribuciones geométricas con parámetros 2 y 7 respectivamente.
  - a. Dado el truncamiento en el número de rezagos para los modelos de series de tiempo, solo los primeros siete valores de las distribuciones son tenidos en cuenta y sus probabilidades son estandarizadas tal que las probabilidades sumen 1 y las medias se mantengan.
  - b. El intervalo serial se calcula de las distribuciones  $F_{inc}$  y  $F_{inf}$  originales y luego de esto se hace el truncamiento y la respectiva estandarización.

Finalmente, el modelo Poisson general no hace parte de las simulaciones dado que la capacidad computacional requerida hace que, para una serie suficientemente larga, la calibración pueda tardar varias horas o incluso días, haciéndolo virtualmente inútil para decisiones de política pública. Por tal motivo, los modelos representados son Bayesiano, Estado Espacio y Poisson exponencial .

### Escenarios de simulación

Con el fin de profundizar en las características de cada modelo, se han planteado una serie de simulaciones que buscan modelar diferentes dinámicas de transmisión. En la práctica, la verdadera serie  $\mathcal{R}(t)$  corresponderá a una combinación de algunas (y más) de las dinámicas básicas. Sin embargo, estudiar estos casos por separado permite visualizar las ventajas y limitaciones de cada modelo. Los escenarios considerados son:

1.  $\mathcal{R}(t)$  constante.
2.  $\mathcal{R}(t)$  linealmente creciente.
3.  $\mathcal{R}(t)$  linealmente decreciente.
4.  $\mathcal{R}(t)$  escalonada creciente.
5.  $\mathcal{R}(t)$  escalonada decreciente. Esta simulación se asemeja a los resultados obtenidos por varios modelos ajustados a la epidemia de COVID-19 en Colombia hasta el mes de abril.
6.  $\mathcal{R}(t)$  constante a excepción de dos grandes picos de infección.
7.  $\mathcal{R}(t)$  se desvanece abruptamente.

8. La serie observada de incidencias presenta valores bajos con respecto a su valor esperado.

A continuación, se presentan las series estimadas en cada uno de los casos. También la serie de nuevos infectados correspondiente a cada simulación, y su valor esperado dados los parámetros de la simulación (gráficas del lado izquierdo).

#### Escenario 1 $\mathcal{R}(t)$ constante

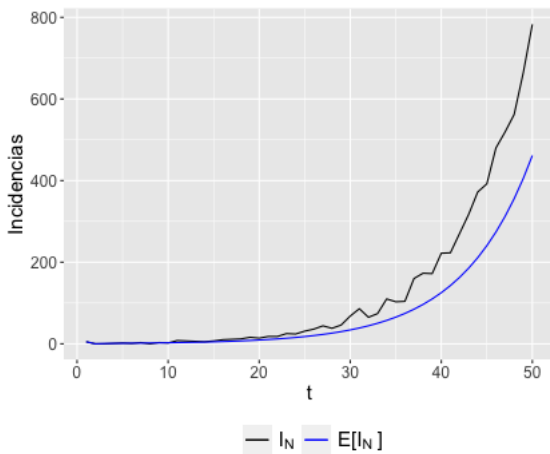


Figura 1

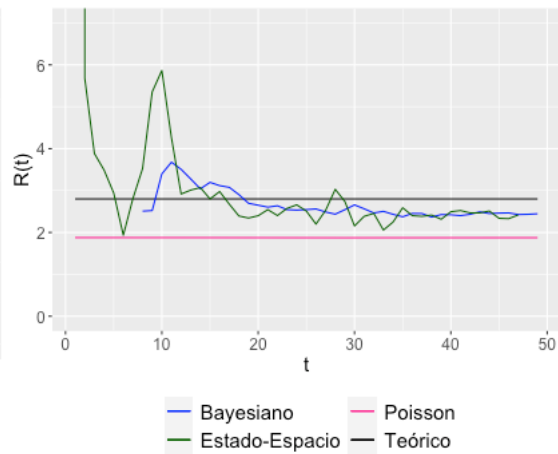


Figura 2

En este primer escenario se puede observar como el modelo Poisson exponencial es el único que desde el principio captura la tendencia de la serie, mientras que los otros dos modelos toman más tiempo en ajustarse a la serie teórica. Así mismo es importante notar como los tres modelos subestiman el valor real, siendo el modelo Poisson el que mayor subestimación presenta.

Escenario 2  $\mathcal{R}(t)$  linealmente creciente

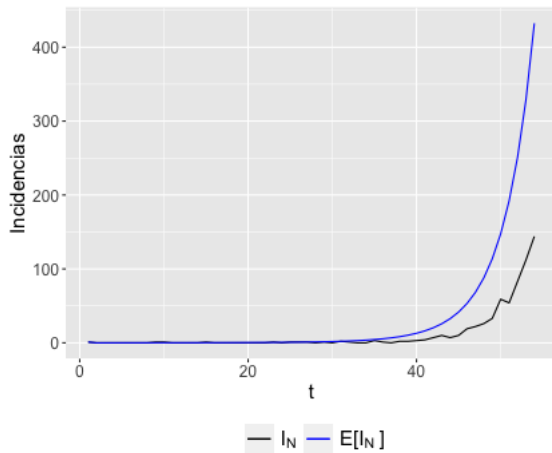


Figura 3

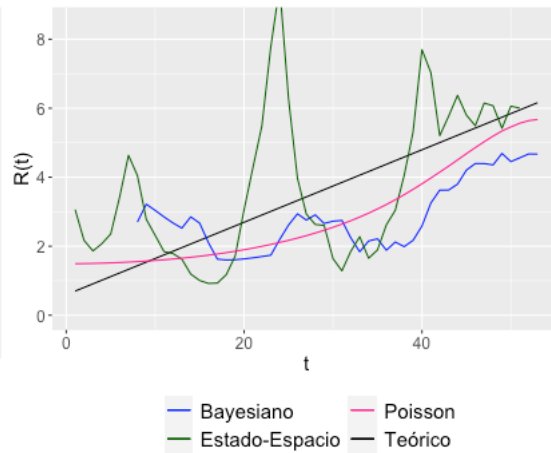


Figura 4

En el segundo escenario nuevamente se aprecia la sensibilidad del modelo de estado espacio el cual reacciona rápidamente a choques de la realización de casos. Por su parte, el modelo Bayesiano tiende a resistirse a los cambios y por eso sistemáticamente subestima el valor correcto. Nuevamente el valor más cercano es el registrado por el modelo Poisson exponencial aunque sigue subestimando en gran parte del periodo el valor teórico.

Escenario 3  $\mathcal{R}(t)$  linealmente decreciente

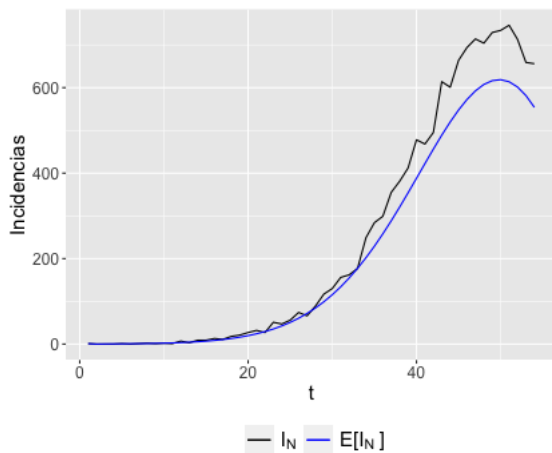


Figura 5

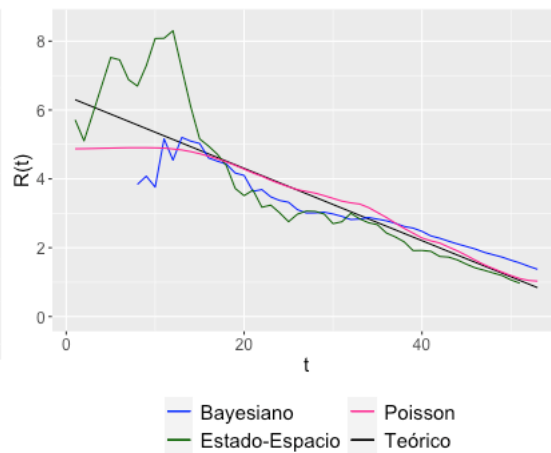


Figura 6

A diferencia de los anteriores escenarios, el escenario 3 inicia con un potencial de infección alto lo cual genera muchos casos desde el inicio y hace que los modelos se ajusten rápidamente. Dicho lo anterior, es notorio como en tramos lineales el modelo Poisson exponencial presenta mejores resultados que los anteriores. En este caso el modelo Poisson exponencial no registra subestimaciones del indicador con excepción del inicio del periodo.

Escenario 4  $\mathcal{R}(t)$  escalonada creciente

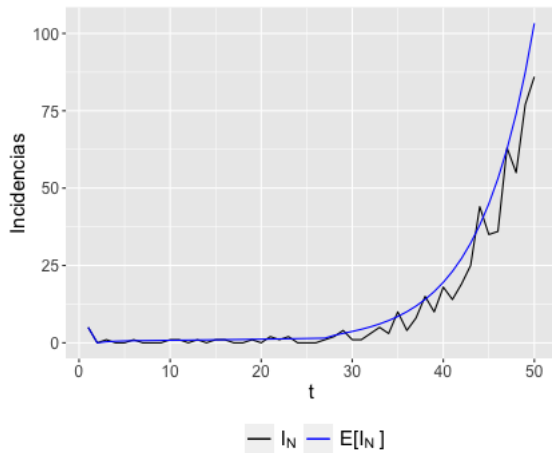


Figura 7

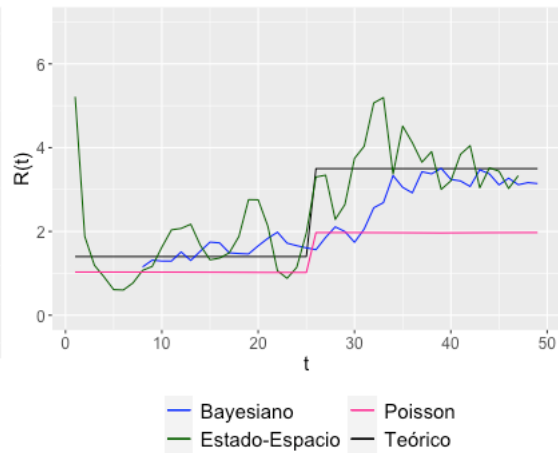


Figura 8

El escenario 4 puede entenderse como la salida de una cuarentena ya que inicia con tasas de contagio bajas y de repente los factores sociales asociados al contagio desaparecen. En este caso el modelo de estado espacio es el primero en ajustarse al cambio. El modelo de Poisson exponencial también se ajusta, pero dada la regularización que necesita, este ajuste no es suficiente. Finalmente, el modelo bayesiano se ajusta, pero toma bastante tiempo para hacerlo. Sin embargo, después del cambio estructural se observa que el Bayesiano y Estado-Espacio están más cerca de los valores teóricos que el modelo Poisson-exponencial.

Escenario 5  $\mathcal{R}(t)$  escalonada decreciente

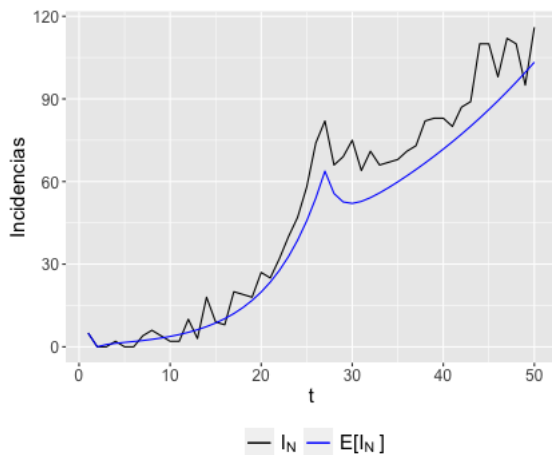


Figura 9

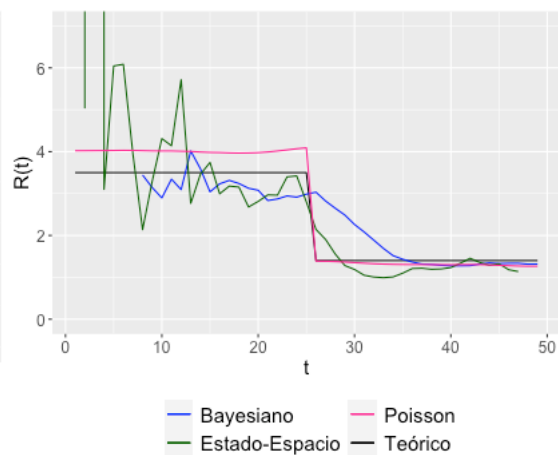


Figura 10

A la inversa del escenario anterior, para el escenario 5 se tiene un modelo Poisson-exponencial que se ajusta rápidamente, seguido por el de estado espacio y finalmente el bayesiano. Igualmente, el modelo Poisson-exponencial se ajusta casi perfectamente en términos de nivel al modelo después del cambio estructural al cabo del periodo 25. Es importante notar que en aquellos escenarios donde los  $\mathcal{R}(t)$  inician altos, el ajuste es sistemáticamente mejor, en particular en los modelos de simulación. Esto naturalmente se asocia a que, a mayor número de casos, la distribución de Poisson-exponencial se aleja de los valores 0 que pueden generar ruido en la iteración del algoritmo de simulación.

*Escenario 6  $\mathcal{R}(t)$  constante a excepción de dos picos de infección*

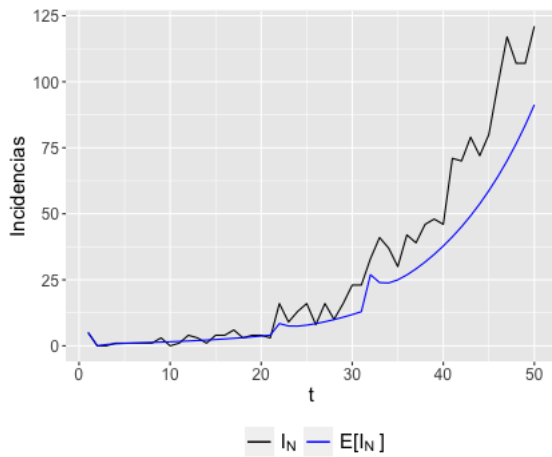


Figura 11

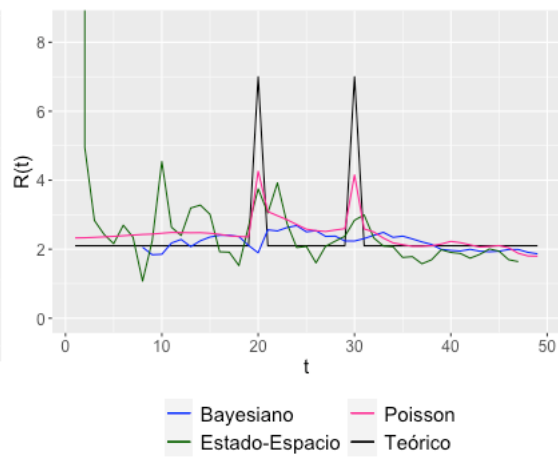


Figura 12

El escenario 6 se usa para resaltar aquellos eventos particulares donde, debido a una manifestación o un evento inesperado,  $\mathcal{R}(t)$  se incrementa inesperadamente y luego retorna al estado base. En este caso el modelo Bayesiano presenta desventaja ya que suaviza la serie demasiado y no le permite identificar dichos picos, el modelo estado-espacio presenta un valor inicial muy por encima del teórico y tampoco logra ajustarse a los picos, inclusive teniendo picos más altos en otros periodos donde no los hubo. El modelo Poisson exponencial se ajusta mejor a los niveles observados y predice mejor los picos.

*Escenario 7  $\mathcal{R}(t)$  se desvanece abruptamente*

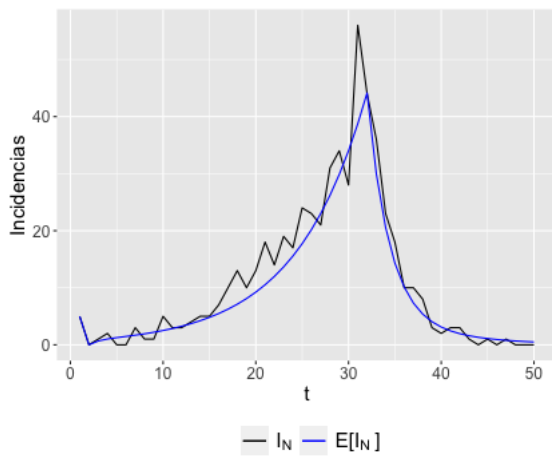


Figura 13

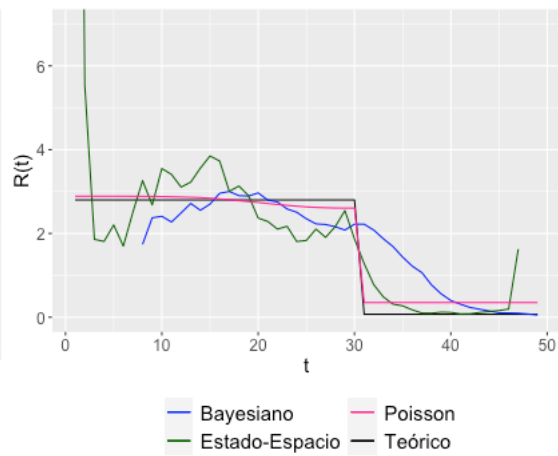


Figura 14

En el escenario 7 presenta los mismos resultados que el escenario 5.

### Escenario 8

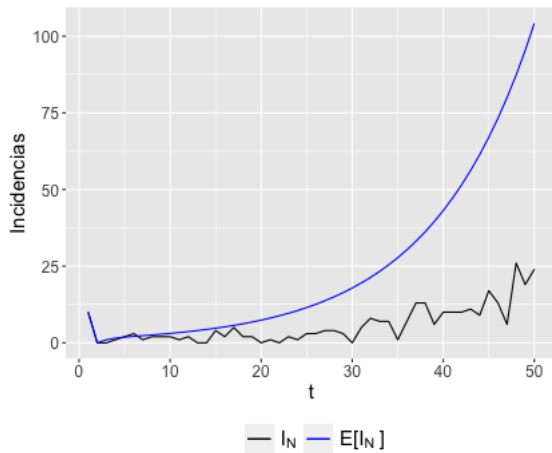


Figura 15

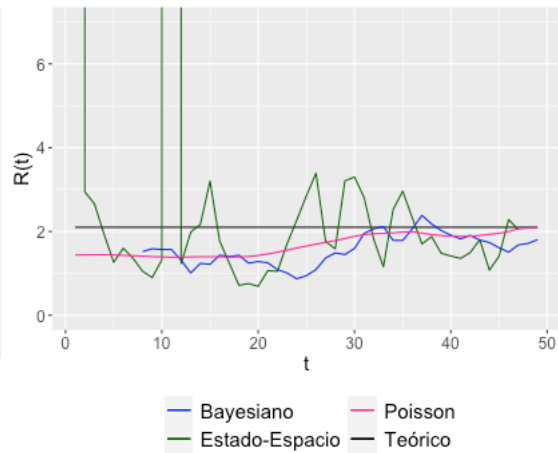


Figura 16

Finalmente, el escenario 8 muestra como ante realizaciones del proceso que estuvieron por debajo de la realización media con excepción del modelo Estado-Espacio que registra valores al inicio del periodo superiores al observado teóricamente, esto se debe a los valores pequeños en el número de casos.

### Métrica de evaluación: Error Porcentual Absoluto Medio (MAPE)

Se utilizará la métrica de El error porcentual absoluto promedio para comparar el valor teórico de  $\mathcal{R}(t)$  con su estimación según cada uno de los modelos. Esta escogencia se debe a la interpretación intuitiva de sus resultados, y a su robustez con respecto a cambios en la magnitud de  $\mathcal{R}(t)$  a lo largo del periodo de observación.

$$MAPE(\mathcal{R}, \hat{\mathcal{R}}) = \frac{1}{n} \sum_{t=1}^n \left| \frac{\mathcal{R}(t) - \hat{\mathcal{R}}(t)}{\mathcal{R}(t)} \right|.$$

La figura 17 resume la calidad del ajuste de cada modelo a cada uno de los escenarios propuestos. Se omite la representación de los casos en los cuales el MAPE del  $\mathcal{R}(t)$  estimado con respecto al teórico superaba un 200% para aumentar la resolución de los resultados más relevantes.



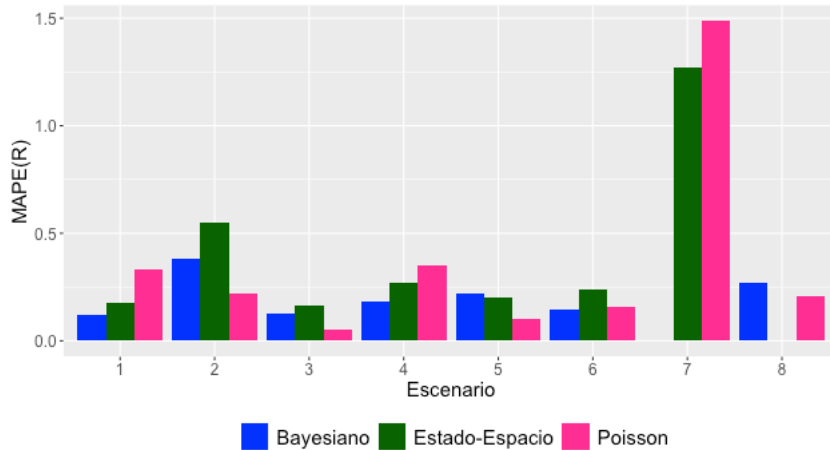


Figura 17

## Discusión

El modelo bayesiano es estable, prueba de ello es su buen ajuste en el escenario 1 donde  $\mathcal{R}(t)$  es constante. Sin embargo, esta misma cualidad hace que su desempeño en los escenarios 2 y 3 de tasa de infección creciente y decreciente linealmente no sean los más deseados. Adicionalmente, este modelo cuenta con un intervalo serial constante que requiere una ventana fija de tiempo de infección, dicha ventana es un hiperparámetro. Sin embargo, este rasgo regularizador también ocasiona un retraso proporcional en la identificación de cambios, como se observan en los escenarios escalonados (4 y 5).

Por otro lado, el modelo Estado-Espacio reacciona rápidamente a los cambios, sin embargo, tiene un leve rezago como se observa en los escenarios 5, 6 y 7. Lo anterior, se debe a que la variable dependiente considera tiempos de incubación a futuro. Así mismo, un aspecto transversal a las simulaciones es la alta inestabilidad que se presenta durante la fase inicial. Una solución a la inestabilidad mencionada previamente es empezar el ajuste luego de un pequeño rezago. El modelo es capaz de ajustar la tendencia de todos los escenarios propuestos, no obstante, su tendencia nunca es suave y esto se debe a la falta de un parámetro regularizador.

Finalmente, el modelo Poisson-exponencial es el que se ajusta más a los datos como lo muestra el análisis MAPE menos uno de los escenarios planteados. Sus principales ventajas incluyen:

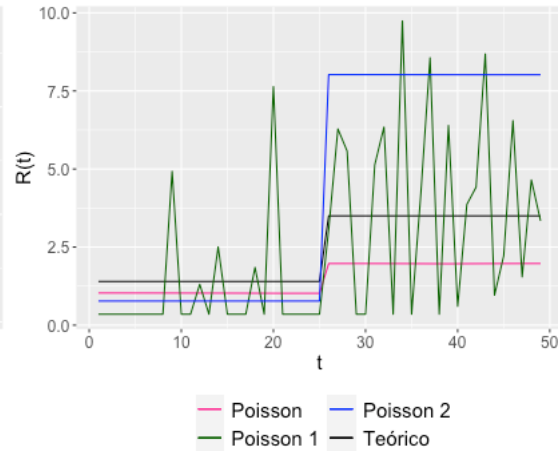
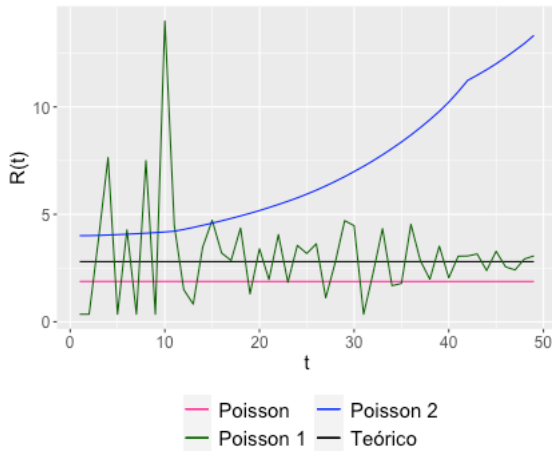
1. El ajuste visual del parámetro de regularización permite al investigador calibrar la suavidad del estimador conforme a su visión holística de la realidad de la epidemia.
2. Su capacidad de reacción a cambios súbitos se complementa con un ajuste, también preciso, de tendencias suaves (escenarios 2 y 3).
3. Aunque con cierto rezago, el modelo describe la epidemia correctamente aún en casos de bajas incidencias (escenario 8).
- 4.

Finalmente, dado que la optimización del modelo es heurística, su complejidad computacional es considerablemente mayor a la de los demás modelos expuestos. Esta se ve agravada por la necesidad de producir resultados bajo varios niveles de regularización.

## Comentario sobre el parámetro de regularización

Asumiendo que se escoge el modelo Poisson-exponencial por las ventajas señaladas anteriormente, la selección del parámetro  $\lambda$ , que controla la suavidad de la serie  $\mathcal{R}(t)$  estimada, implica un reto para el ajuste de este modelo como se señala anteriormente. Para motivar este punto, se ajustó el modelo bajo varias configuraciones de regularización a los escenarios discutidos. La siguiente figura muestra las series estimadas por el modelo con regularización ajustada visualmente y dos alternativas; la primera tiene  $\lambda = 0$  (denominada "Poisson 1"), y la segunda tiene un valor extremadamente alto (denominada "Poisson 2").

### Escenarios 1 y 4



De estas gráficas hay dos elementos claves a resaltar. Por un lado, está el hecho que algunos parámetros de calibración pueden cambiar significativamente la tendencia de la serie, como se aprecia en el escenario 1. Por otro lado, se observa como el parámetro  $\lambda$  ayuda para que la serie evite ser fuertemente afectada por el ruido de los datos, sin embargo, la elección de este puede generar problemas de sobre- o subestimación sistemática.

De esta forma se concluye que este tercer método requiere ser usado con precaución. Por una parte, si el investigador tiene un conocimiento robusto sobre las tendencias de la infección y las usa para elegir un parámetro de suavizamiento informado, el método resulta ser bastante prometedor. No obstante, si el investigador desea hacer una exploración inicial de los datos, el método puede causar errores significativos.

## Caso de estudio: Colombia

A partir de los datos de vigilancia epidemiológica publicados por el Instituto Nacional de Salud de Colombia, los tres modelos (Bayesiano, Estado Espacio, y Poisson Exponencial) se ejecutaron para entender la evolución de la enfermedad desde su inicio en Colombia hasta el 15 de Agosto del 2020. Para estos ejercicios se usaron<sup>10</sup>:

<sup>10</sup> Otras configuraciones de parámetros fueron consideradas, pero los resultados no presentaron cambios significativos.

1. Ventana del modelo Bayesiano: 5
2. Número máximo de rezagos: 7 asociados a intervalo serial, 15 asociados a infección y 5 asociados a incubación
3.  $F_{inc} \sim \text{Gamma}(3.16, 5.16)^{11}$ ,  $F_{inf} \sim \text{Weibull}(24.20, 2.98)^{12}$  y el intervalo serial se considera  $\text{Weibull}(2.23, 5.42)^{13}$
4. Para el Poisson Exponencial se usaron las medias correspondientes a distribuciones anteriores.
5. El parámetro de suavizamiento elegido sigue un criterio de ajuste  $\chi^2$  de 0.001.
6. Los intervalos de confianza de las gráficas fueron hechos a 95%.

## Resultados

### R efectivo

Periodo Marzo 1 - Abril 15

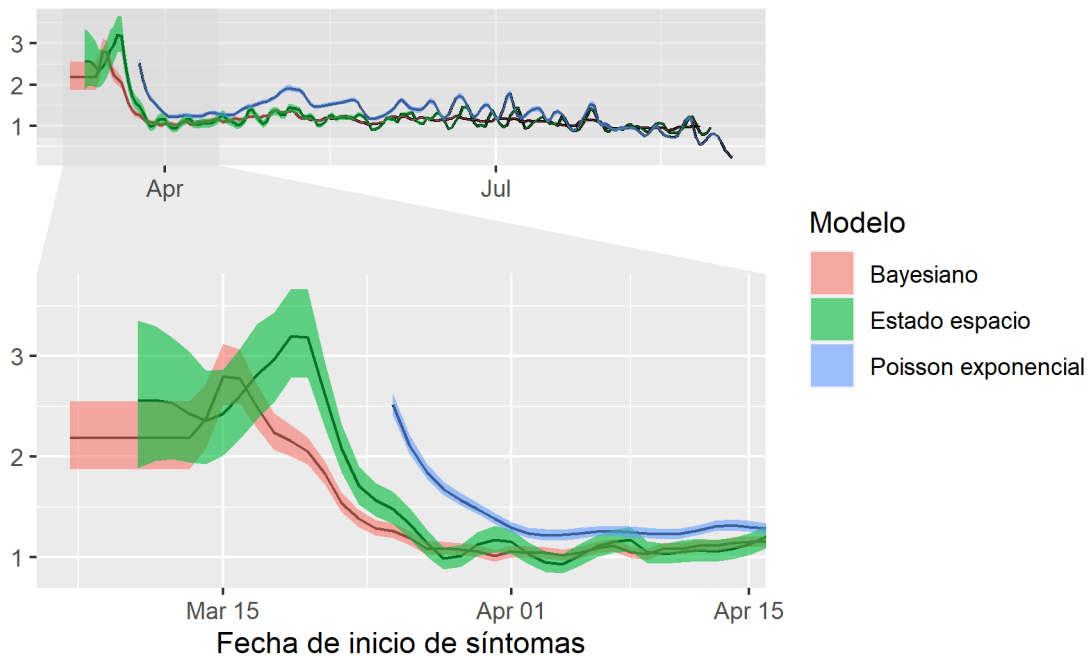


Figura 20

La figura 20 resalta la tasa efectiva a inicios de la pandemia. En este caso, los modelos consistentemente muestran una fuerte reducción del  $\mathcal{R}(t)$  en la segunda quincena de marzo. Este evento se asocia con la primera ola de cuarentenas en el país. Se aprecia como coinciden los métodos de estado espacio y bayesianos sugiriendo claramente una tasa que llegó a estar alrededor de niveles de 1. Desafortunadamente, en estos valores iniciales, el modelo Poisson Exponencial presenta mucho ruido de los datos, y por ende su interpretación no puede darse de forma directa. Aun así, el modelo permite dar una cota superior al valor real asociado al  $\mathcal{R}(t)$ .

<sup>11</sup> Gao et al. (2020)

<sup>12</sup> Ling et al. (2020)

<sup>13</sup> Nishura et al. (2020)

## R efectivo

Periodo Abril 15 - Junio 15

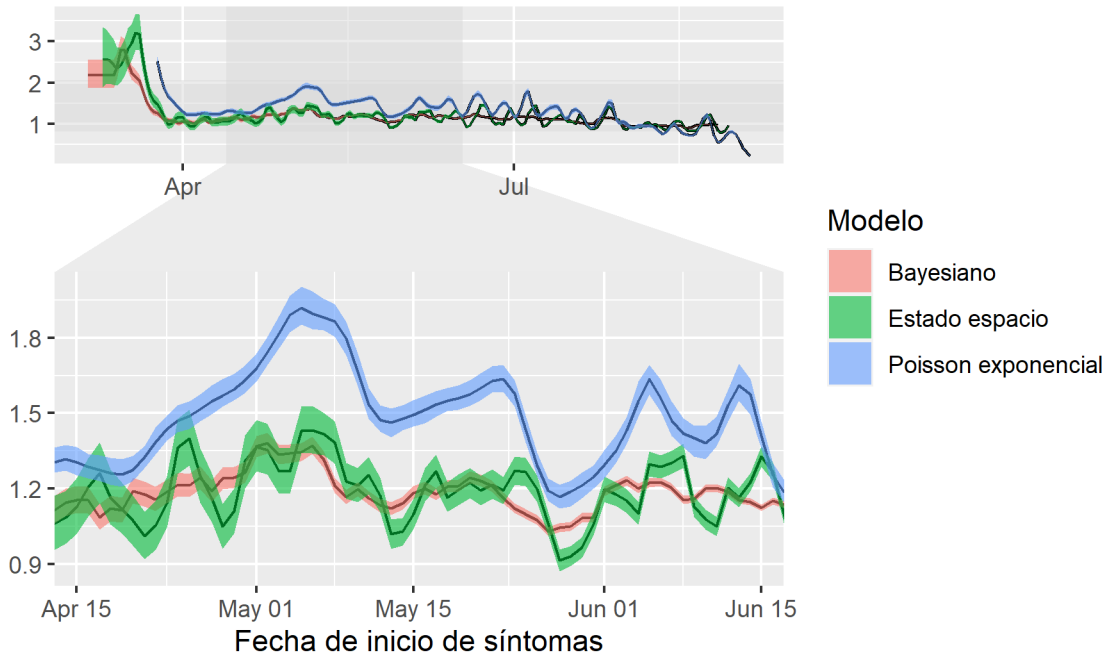


Figura 21

La figura 21 en contraste muestra una etapa de transición de cuarentenas totales a cuarentenas sectorizadas tanto espacial como ocupacionalmente y por estructura tarea de la población. En esta etapa se resalta el carácter oscilatorio de la tasa de infección, y donde se muestra que hubo un aumento significativo alrededor del primero de mayo posiblemente motivado por ser el periodo donde se hicieron varios cambios sobre las restricciones de las personas cuarentena (República, P. (2020a), República, P. (2020b), República, P. (2020c)).

## R efectivo

Periodo Junio 15 - Agosto 15

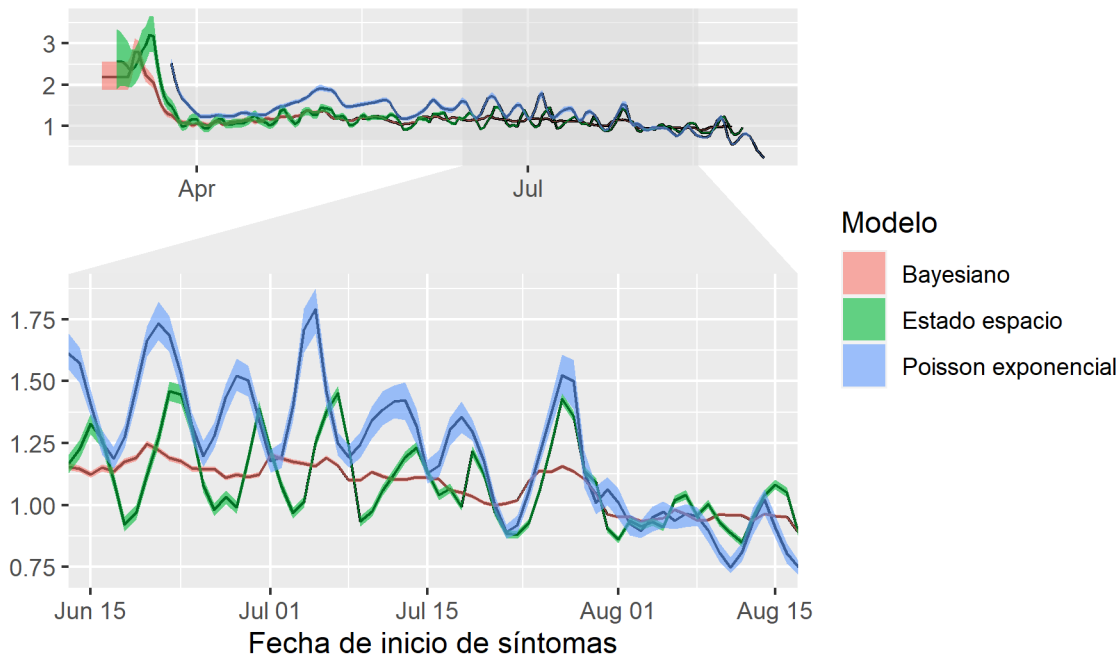


Figura 22

Finalmente, la figura 22 muestra un periodo relativamente estable, pero con unos picos específicos asociados al 19 de junio y 3 de julio (días que en el país hubo una exención del IVA a múltiples productos y por ende mucha gente salió a las tiendas a comprar). Se hace importante resaltar cómo el modelo de estado espacio es el que mejor identifica estos picos, mientras que el modelo bayesiano suaviza la curva impidiendo visualizar estas políticas. Dicho lo anterior, el modelo de estado espacio también identifica otros picos que pueden estar asociados a otros eventos puntuales, pero también pueden estar relacionados a ruido numérico. Para determinar cuál de los dos escenarios es, se requiere un estudio más detallado de los eventos políticos, económicos y sociales del país en estos días.

## Conclusiones

La tasa efectiva de reproducción es un elemento clave para el desarrollo de políticas asociadas a la prevención y mitigación de la pandemia. Sin embargo, su estimación requiere de conjuntos de supuestos los cuales pueden variar entre enfermedades y contextos sociales. Por tal motivo este estudio se ha enfocado en entender las implicaciones, alcances y limitaciones de cada uno de estos supuestos y a partir de estos generar tres nuevos modelos que mejoren la estimación de esta tasa. Tras haber realizado los ejercicios de simulación, es posible evidenciar como no existe un modelo que a-priori sea mejor que los demás y por tal motivo se recomienda el uso de varios de éstos (haciendo explícitos los supuestos que tienen), para dar a los encargados de la política pública una información completa, la cual pueda ser interpretada por el criterio de expertos para tomar las decisiones correspondientes. Finalmente, si bien los resultados presentados en el caso de estudio, interpretados a la luz de la simulación, permiten apreciar eventos asociados al análisis de la pandemia en Colombia, es clave notar que existen múltiples meta-parámetros que alimentan estos modelos, tales como la distribución del periodo de infección o de incubación. Por tal motivo, si bien

estos modelos permiten hacer mediciones que parten de una perspectiva estadística, es importante avanzar en estudios asociados a la biología del virus que ayuden a calibrar de manera precisa estos valores.

## Referencias

- Adam, D. (2020). A guide to R-the pandemic's misunderstood metric. *Nature*, 583(7816), 346-348.
- Davison, A. C. y Hinkley, D. V. (1999) *Bootstrap Methods and Their Applications*. Cambridge University Press.
- Kermack W. O. y McKendrick A. G. (1927) A Contribution to the Mathematical Theory of Epidemics *Proc Royal Soc London A* 115: 700–721.
- Wallinga, J., & Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology*, 160(6), 509-516.
- Fraser C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PloS one*, 2(8), e758. <https://doi.org/10.1371/journal.pone.0000758>
- van den Driessche P. (2017). Reproduction numbers of infectious disease models. *Infectious Disease Modelling*, 2(3), 288–303. <https://doi.org/10.1016/j.idm.2017.06.002>
- Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology*, 178(9), 1505-1512.
- Thompson, R. N., Stockwin, J. E., van Gaalen, R. D., Polonsky, J. A., Kamvar, Z. N., Demarsh, P. A., ... & Lessler, J. (2019). Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 29, 100356.
- Harvey, A (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W., ... & Wang, M. H. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International journal of infectious diseases*, 92, 214-217.
- Dickens, B. L., Koo, J. R., Lim, J. T., Park, M., Quaye, S., Sun, H., ... & Lee, V. J. (2020). Modelling lockdown and exit strategies for COVID-19 in Singapore. *The Lancet Regional Health-Western Pacific*, 100004.
- Gao, Q., Hu, Y., Dai, Z., Xiao, F., Wang, J., & Wu, J. (2020). The epidemiological characteristics of 2019 novel coronavirus diseases (COVID-19) in Jingmen, Hubei, China. *Medicine*, 99(23), e20605.
- Ling, Y., Xu, S. B., Lin, Y. X., Tian, D., Zhu, Z. Q., Dai, F. H., ... & Hu, B. J. (2020). Persistence and clearance of viral RNA in 2019 novel coronavirus disease rehabilitation patients. *Chinese medical journal*.
- Nishiura, H., Linton, N. M., & Akhmetzhanov, A. R. (2020). Serial interval of novel coronavirus (COVID-19) infections. *International journal of infectious diseases*.

- The R number and growth rate in the UK. (2020). Retrieved 18 August 2020, from <https://www.gov.uk/guidance/the-r-number-in-the-uk>
- República, P. (2020). Gobierno expide el Decreto 847 mediante el cual dicta nuevas disposiciones para el Aislamiento Preventivo Obligatorio. Retrieved 18 August 2020, from <https://id.presidencia.gov.co/Paginas/prensa/2020/Gobierno-expide-Decreto-847-mediante-el-cual-dicta-nuevas-disposiciones-para-el-Aislamiento-Preventivo-Obligatorio-200614.aspx>
- República, P. (2020). Estas son las 43 actividades exceptuadas durante el Aislamiento Preventivo Obligatorio que regirá en Colombia desde el 1° de junio, según Decreto expedido por el Gobierno Nacional. Retrieved 18 August 2020, from <https://id.presidencia.gov.co/Paginas/prensa/2020/Estas-son-43-actividades-exceptuadas-durante-Aislamiento-Preventivo-Obligatorio-que-regira-Colombia-desde-1-junio-200528.aspx>
- República, P. (2020). Abecé del Decreto 593, que amplía de 35 a 41 las actividades exceptuadas del Aislamiento Preventivo Obligatorio. Retrieved 18 August 2020, from <https://id.presidencia.gov.co/Paginas/prensa/2020/Abecedario-del-Decreto-593-que-amplia-de-35-a-41-las-actividades-exceptuadas-del-Aislamiento-Preventivo-Obligatorio-200425.aspx>