

# Estimación de un Modelo SIR Espacial de la Epidemia de COVID-19 y su Uso en Proyecciones de Corto Plazo en Perú

Kristian López Vargas<sup>\*</sup>

University of California, Santa Cruz

Gonzalo Panizo<sup>\*\*</sup>

IMCA - Universidad de Ingeniería, Perú

Diego Tocre<sup>\*\*\*</sup>

Universidad del Pacífico, Perú

Marco Gutierrez<sup>\*\*\*\*</sup>

Universidad del Pacífico, Perú

1 de julio de 2020

## Resumen

Un error común de los ejercicios de modelación de la pandemia en el Perú viene de la idea de que cada distrito vive su epidemia por separado. Contrario a dicha visión, modelar la conexión del avance de la epidemia de manera espacial puede ayudar a tener una proyección de corto plazo - por ejemplo para identificar las próximas provincias o distritos con brotes. En esta nota presentamos un modelo SIR determinístico para proyectar la expansión espacial de la epidemia en el territorio a nivel de distritos. Para eso, hemos aumentado un modelo SIR estándar para incluir la transmisión entre distritos a lo largo del territorio nacional asumiendo que la probabilidad de encuentro de una persona del distrito  $A$  con alguien del distrito  $B$  es proporcional a la conectividad entre esos distritos y al tamaño de la población del distrito  $B$ . Además la densidad poblacional y el cambio de comportamiento en el tiempo son tomados en cuenta. Un supuesto central del modelo es que una persona de un distrito se puede encontrar con otra de otro distrito, pero la probabilidad de dicho evento depende de la conectividad entre ambos distritos. Luego de estimar el modelo, detallamos la capacidad del modelo de predecir brotes distritales que con modelos de epidemias separadas no se puede.

---

<sup>\*</sup>Email: kristian@ucsc.edu

<sup>\*\*</sup>Email: gonzalo.panizo@gmail.com

<sup>\*\*\*</sup>Email: diegotocre@gmail.com

<sup>\*\*\*\*</sup>Email: a20141676@pucp.pe

# 1. Modelo

Al igual que en el modelo SIR estándar (Brauer and Castillo-Chávez, 2001; Kermack and McKendrick, 1927, 1991a,b,c; Ross, 1911; Diekmann et al., 1995; Macdonald et al., 1957), la población de un distrito,  $N_i$ , se divide en tres categorías: Susceptibles ( $S_i$ ), Infectados ( $I_i$ ) y Recuperados ( $R_i$ ).

$$N_i = S_i(t) + I_i(t) + R_i(t) \quad (1)$$

Lo nuevo en este modelo es que incorpora la interacción espacial de los infectados en el territorio. Mientras que en el modelo SIR estándar, un individuo tiene una probabilidad de encontrarse con un infectado simplemente de  $I/N$ , en este modelo el individuo representativo del distrito A tiene una probabilidad de encontrarse con otro del mismo distrito, pero también una probabilidad de encontrarse con personas de otro distrito B. Esa probabilidad está directamente relacionada a la conectividad con el distrito B, y también al tamaño del distrito B. Así, además de su propia situación inicial ( $S_{A,t_0}, I_{A,t_0}, R_{A,t_0}$ ), el crecimiento de la infección en el distrito A depende de: (1) qué tan conectado está a otros distritos, (2) qué tan grandes son los distritos con los que A está conectado y (3) la situación de infección de esos otros distritos con los que está conectado. Esto implica por ejemplo que los vecinos de distritos más grandes y con más infecciones deben tener una tasa de infección mayor.

En términos del modelo, esto le da al sistema una dinámica donde la ecuación de nuevos infectados viene dada por:

$$\beta \sum_j \left( w_{ij} \frac{I_j}{N_j} \right) S_i \quad (2)$$

donde  $\beta$  es la tasa de infección estándar de un SIR,  $w_{ij}$  es una función de probabilidad de encuentro de una persona de  $i$  con otra del distrito  $j$  asociada a la conectividad del distrito  $i$  con el distrito  $j$ .

Hay muchas formas de definir conectividad. La medida directa implicaría saber en un día cual es el desplazamiento entre distritos en el país. Una alternativa más sencilla es utilizar el hecho que la mayoría de interacciones suceden entre distritos más cercanos. Concretamente, la función  $w(\cdot)$  debe ser decreciente con la distancia entre los distritos  $i$  y  $j$ . Es decir  $w_{ij} = w(d_{ij})$ . La opción más sencilla es utilizar una función de *exponential*

*weighting* de la siguiente manera:

$$w(d_{ij}) = \frac{\exp(-\pi d_{ij})}{\sum_h \exp(-\pi d_{ih})} \quad (3)$$

donde  $\pi$  es la tasa de decaimiento de influencia en función a la distancia. Además, para tomar en cuenta que las ciudades más grandes tienen influencia a distancias más grandes, asumimos que la probabilidad de encuentro es proporcional al tamaño del distrito en habitantes. Por ejemplo, Lima tiene 10 veces más influencia que Arequipa, por ejemplo, sobre un distrito que les sea equidistante. Así obtenemos esta función de probabilidad.

$$w(d_{ij}) = \frac{N_j \exp(-\pi d_{ij})}{\sum_h N_h \exp(-\pi d_{ih})} \quad (4)$$

donde  $d_{ij}$  es la distancia entre dos distritos. Con estas modificaciones el sistema de ecuaciones diferenciales del SIR resultante se puede formular en tiempo discreto lo que resulta en el siguiente sistema de ecuaciones en diferencia:

$$\Delta S_{i,t+1} = -\beta \left[ \sum_j \left( w_{ij} \frac{I_{j,t}}{N_j} \right) \right] S_{i,t} \quad (5)$$

$$\Delta I_{i,t+1} = \beta \left[ \sum_j \left( w_{ij} \frac{I_{j,t}}{N_j} \right) \right] S_{i,t} - \gamma I_{i,t} \quad (6)$$

$$\Delta R_{i,t+1} = \gamma I_{i,t} \quad (7)$$

Podemos volver aún más realista el modelo si asumimos que el diferencial de nuevos infectados también depende de la densidad poblacional del distrito. En este sentido, se esperaría que en distritos con más pobladores por kilómetro cuadrado el contagio ocurra a un ritmo más acelerado, pues el contacto con vecinos y por tanto la probabilidad de transmitir el virus es mayor. Además, también podríamos especificar que la probabilidad de contagio dependa del tiempo de manera lineal y cuadrática. Esto incorporaría parte de los cambios en el comportamiento de las personas y de las normas promulgadas por el gobierno con respecto a la cuarentena. Una forma de incluir ambos componentes puede hacerse fácilmente modificando las ecuaciones 5 y 6 de la siguiente manera:

$$\Delta S_{i,t+1} = -(\beta_0 + \beta_1 t + \beta_2 t^2) \cdot \log(\text{densidad}_i + 1)^\delta \left[ \sum_j \left( w_{ij} \frac{I_{j,t}}{N_j} \right) \right] S_{i,t} \quad (8)$$

$$\Delta I_{i,t+1} = (\beta_0 + \beta_1 t + \beta_2 t^2) \cdot \log(\text{densidad}_i + 1)^\delta \left[ \sum_j \left( w_{ij} \frac{I_{j,t}}{N_j} \right) \right] S_{i,t} - \gamma I_{i,t}. \quad (9)$$

Este sistema de proyección se puede implementar de manera recursiva definiendo como unidad de tiempo el día.

## 2. Estrategia de calibración del modelo

Con los objetivos de (i) conocer los valores que toman los parámetros del modelo SIR que utilizamos (y así tener una mejor idea de la comportamiento de la expansión del virus) y (ii) predecir las cifras de infectados para los siguientes periodos, hemos realizado una regresión no lineal de mínimos cuadrados sobre el número de infectados a nivel distrital en cada periodo de tiempo y basada en el modelo SIR detallado anteriormente:

$$I_{i,t} = (1 - \gamma) I_{i,t-1} + (\beta_0 + \beta_1 t + \beta_2 t^2) \left[ \sum_j \left( \frac{N_j \exp(-\pi d_{ij})}{\sum_h N_h \exp(-\pi d_{ih})} \frac{I_{j,t-1}}{N_j} \right) \right] S_{i,t-1} + \varepsilon_{i,t}, \quad (10)$$

donde  $\varepsilon$  es un término estocástico que se distribuye  $N(0, \sigma^2)$ . Los parámetros a estimar o imputar son  $\gamma$ ,  $\beta$  y  $\pi$ . La estimación se realizó en Stata a través del comando nl.

## 3. Datos

Se utilizaron los datos a nivel individual de casos positivos de COVID-19 en todo el Perú, desde el 5 de marzo hasta el 17 de junio del año 2020. Estos datos han sido colapsados a nivel distrital. Una limitación importante de los datos es que no se reportan los casos recuperados, sino que solo se cuentan con los casos que alguna vez dieron positivos, pero no se tiene un seguimiento en la base de datos si es que se recuperaron o no en determinada fecha. Esto supone un problema para la aplicación del modelo SIR que proponemos, ya que los recuperados suponen una parte importante del análisis. Con lo que sí contamos es con datos de recuperados a nivel nacional recopilados por la Universidad Johns Hopkins. Por ende, para solucionar este problema, calculamos tasas de recuperación diaria dentro del periodo de análisis y la descontamos homogéneamente a los distritos de estudio. De este modo, calculamos 85,001 casos activos el 17 de junio.

Para construir la matriz de conectividad entre distritos, lo ideal sería contar con distancias por vía terrestre y en kilómetros cuadrados o tiempos de viaje en auto de un distrito a otro. Esto puede extraerse a partir de los APIs de Google Maps y de Waze. Al momento de empezar a trabajar en este proyecto, no pudimos recolectar las distancias

entre distritos, pero sí entre provincias, a través del API de Google Maps. Usando estas distancias entre provincias y calculando distancias euclidianas entre distritos (haciendo uso de los centroides de cada distrito), pudimos construir aproximaciones a distancias entre distritos por vía terrestre y en kilómetros cuadrados. Para un mayor detalle sobre el cálculo de estas distancias y el uso del API de Google Maps, véase el Anexo [A](#).

## 4. Desempeño para proyectar

Hemos realizado la estimación de: (1) el modelo estándar, (2) del modelo que incorpora densidades poblacionales por distrito y (3) el que incorpora densidades poblacionales más tendencias temporales cuadráticas en la probabilidad de contagio. Con la finalidad de evaluar el desempeño de estos modelos para proyectar a futuro, es decir el poder predictivo del modelo fuera de la muestra, utilizamos datos hasta el 30 de Abril, que es una fecha ubicada aproximadamente a la mitad del periodo entre el inicio de la pandemia y la fecha hasta la cual teníamos datos de infectados al iniciar este proyecto. Además, del 30 de Abril a la fecha en la cual se hicieron los datos individuales de infectados públicos (23 de mayo) hay un poco más de 3 semanas de diferencia, por lo que haciendo este análisis en esta fecha nos puede dar una idea de cuánto podría haberse refinado las políticas focalizadas y la discusión académica si los datos abiertos se hubieran hecho públicos semanas antes.

Para todas las estimaciones, no estimamos el valor de  $\gamma$ , debido a que hemos hecho una suposición sobre la distribución observada de recuperados a través de las localidades, sino que le imputamos un valor de 4%, que es muy similar al que calculamos analizando los datos agregados a nivel nacional recopilados por la Universidad Johns Hopkins. Los estimados para los tres modelos se muestran en el Cuadro [1](#).

La ventaja de usar este modelo espacial es poder predecir el nivel de infectados no solo a nivel nacional, sino también a nivel distrital o provincial. De esta manera, pueden refinarse las políticas focalizadas que se lleven a cabo, priorizando a los distritos, provincias o regiones con mayor riesgo de acelerar el número de infectados activos. Así pues, presentamos en las figuras [1](#) y [2](#) las predicciones que el modelo con densidades poblacionales y tendencias temporales hace hasta 30 días después del 30 de Abril para dos grupos de provincias: (i) las 18 con mayor número de infectados al 17 de Junio y (ii) las 18 con menos de cinco infectados al 30 de Abril y mayor crecimiento de infectados 15 días después, respectivamente.

Como se observa, la predicción del número de infectados activos para las provincias con mayor número de infectados es muy cercano a lo observado, por lo que las

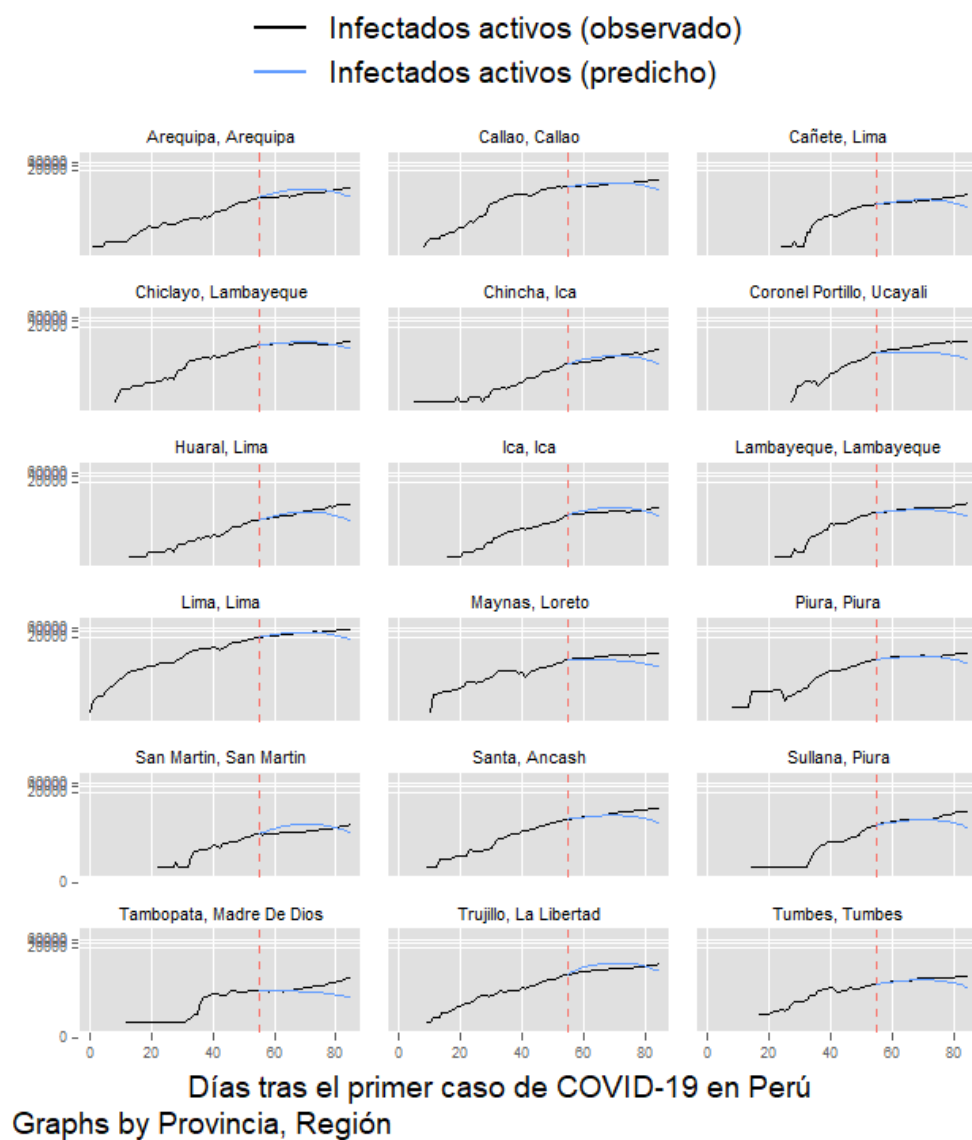


Figura 1: Predicción del número de infectados activos en las 18 provincias con mayor número de infectados al 17 de junio (escala logarítmica)

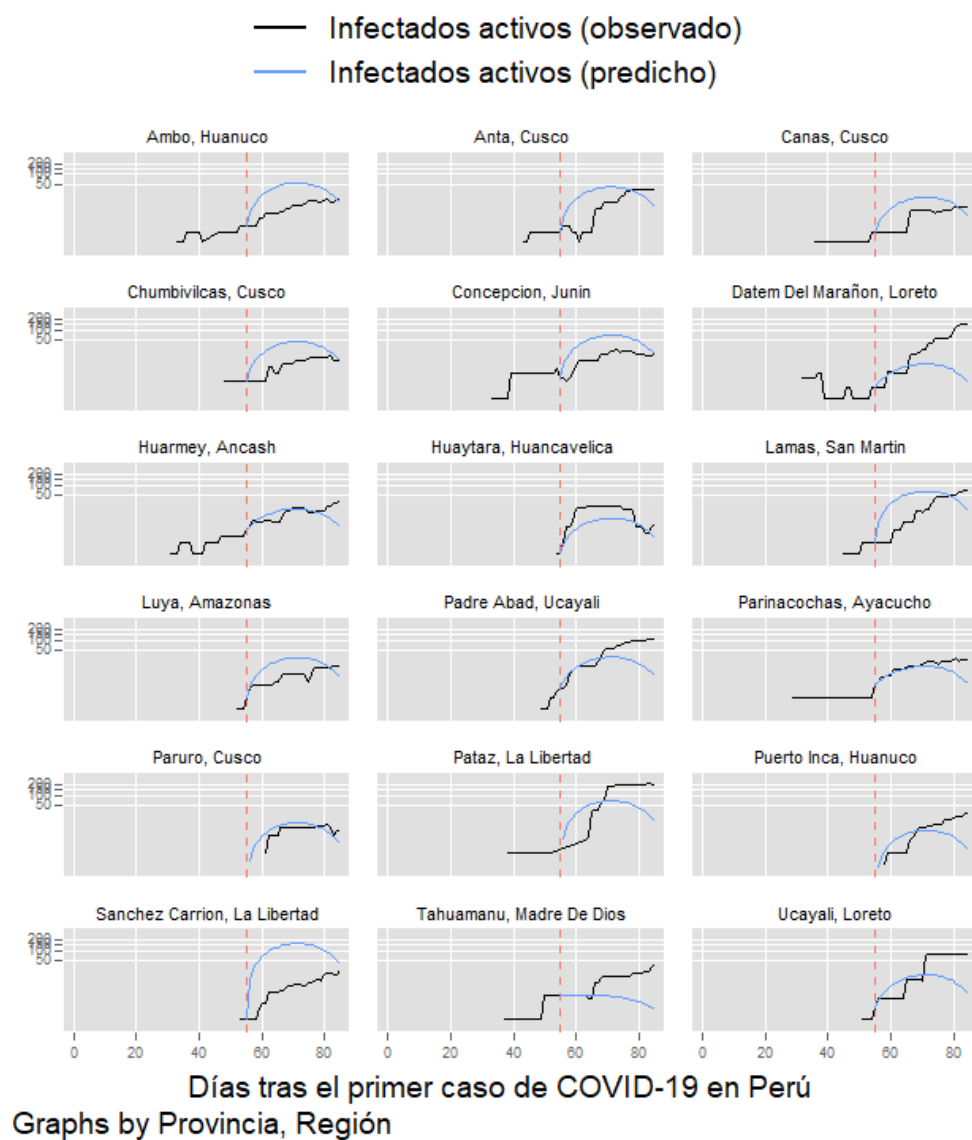


Figura 2: Predicción del número de infectados activos en las 18 provincias con pocos infectados y mayor crecimiento de la infección (escala logarítmica)

Cuadro 1: Estimados del modelo SIR espacial hasta el 30 de abril

Parámetro	Estimado	Error estándar	IC Inferior	IC Superior
<i>A. Estándar (<math>R^2=0.9837</math>)</i>				
$\pi$	0.0007	0.0000	0.0006	0.0007
$\beta$	$3.91 \cdot 10^{-16}$	$4.36 \cdot 10^{-18}$	$3.82 \cdot 10^{-16}$	$3.99 \cdot 10^{-16}$
<i>B. Incorporando densidades poblacionales (<math>R^2=0.9839</math>)</i>				
$\pi$	0.0003	0.0000	0.0002	0.0003
$\beta$	$2.57 \cdot 10^{-17}$	$2.59 \cdot 10^{-18}$	$2.06 \cdot 10^{-17}$	$3.08 \cdot 10^{-17}$
$\delta$	1.1161	0.0418	1.0342	1.1979
<i>C. Densidades poblacionales y tendencias temporales (<math>R^2=0.9839</math>)</i>				
$\pi$	0.0003	0.0000	0.0002	0.0003
$\beta_0$	$-1.56 \cdot 10^{-17}$	$5.42 \cdot 10^{-18}$	$-2.62 \cdot 10^{-17}$	$-4.97 \cdot 10^{-18}$
$\beta_1$	$1.93 \cdot 10^{-18}$	$2.97 \cdot 10^{-19}$	$1.35 \cdot 10^{-18}$	$2.51 \cdot 10^{-18}$
$\beta_2$	$-2.29 \cdot 10^{-20}$	$3.33 \cdot 10^{-21}$	$-2.94 \cdot 10^{-20}$	$-1.64 \cdot 10^{-20}$
$\delta$	1.1618	0.0421	1.0792	1.2444

predicciones podrían haber sido buenas guías de acción hasta 30 días después de haberse estimado el modelo. En el caso de las provincias con pocos infectados y mayor crecimiento infeccioso, la predicción del modelo no es tan precisa como para el grupo anterior de provincias, pero igual tampoco se aleja mucho de lo observado y podría funcionar también como una guía de acción. No sorprende que para estas provincias con menos infectados la predicción no sea tan precisa, puesto que contribuyen con menos variabilidad a la estimación que las provincias anteriores.

Asimismo, para tener una idea de la predicción del modelo para la totalidad de provincias, en la figura 3 se muestra una comparación entre lo observado y predicho a través de dos mapas nacionales coloreados a nivel provincial según deciles del número de infectados activos. El mapa con las cifras predichas también parte del modelo con densidades poblacionales y tendencias temporales.

Las figuras análogas a las mostradas en esta sección sobre la base de los otros dos modelos pueden encontrarse en el Anexo REF.

## 5. Mejoras Recomendadas

Las mejoras potenciales de este modelo vienen principalmente de obtener mejores datos insumo. En primer lugar, se debe contar con una base de infectados a nivel de distrito que tenga información de casos recuperados.

Por otro lado, la conectividad se puede aproximar o modelar de varias maneras. Una



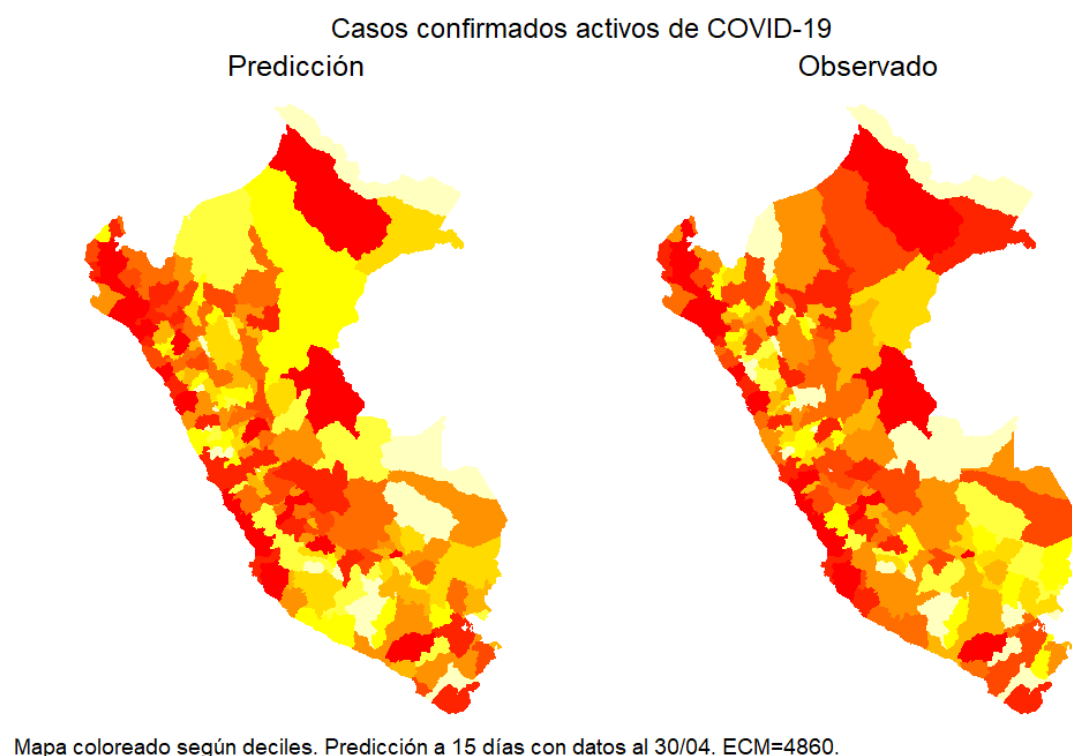


Figura 3: Comparación entre la predicción del modelo y lo observado a 15 días de la estimación con datos al 30 de Abril

de ellas es utilizar data celular agregada de transiciones de personas entre distritos. Actualmente contamos con datos parciales para construir dicha medida.

Asimismo, se puede utilizar las distancias y tiempos de viaje en automóvil de un distrito a otro. Actualmente hemos calculado estos tiempos de viaje entre provincias haciendo uso del API de Google Maps. No obstante, es posible calcular directamente la distancia entre distritos haciendo uso de este mismo API, pero por razones de tiempo y presupuesto no lo hemos podido hacer a la fecha de esta publicación.

## Referencias

- F. Brauer and C. Castillo-Chávez. *Mathematical models in population biology and epidemiology*, volume 40 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 2001.
- O. Diekmann, J. Heesterbeek, and J. Metz. The legacy of kermack and mckendrick. In D. Mollison, editor, *Epidemic models: Their structure and relation to data*, pages 95–115. Cambridge University Press, Cambridge, 1995.
- W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.
- W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics—i. *Bulletin of Mathematical Biology*, 53(1):33–55, Mar 1991a. ISSN 1522-9602. doi: 10.1007/BF02464423. URL <https://doi.org/10.1007/BF02464423>.
- W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics—ii. the problem of endemicity. *Bulletin of Mathematical Biology*, 53(1):57–87, Mar 1991b. ISSN 1522-9602. doi: 10.1007/BF02464424. URL <https://doi.org/10.1007/BF02464424>.
- W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics—iii. further studies of the problem of endemicity. *Bulletin of Mathematical Biology*, 53(1):89–118, Mar 1991c. ISSN 1522-9602. doi: 10.1007/BF02464425. URL <https://doi.org/10.1007/BF02464425>.
- G. Macdonald et al. The epidemiology and control of malaria. *The Epidemiology and Control of Malaria.*, 1957.
- R. Ross. *The prevention of malaria (with Addendum)*. John Murray, London, 1911.

## A. Cálculo de las distancias entre distritos

### A.1. Sobre el uso del API de Google Maps y las distancias entre provincias

Para la construcción, se utilizó el API “Distance Matrix” de Google Maps. Esta nos permite calcular la distancia en kilómetros entre una localidad y otra dado un medio de transporte. En este caso, el medio fue un automóvil. El punto de partida y llegada no es necesariamente el centroide de la provincia, pero suele coincidir con ellos o estar muy cercano a ellos. Google Maps busca los puntos con los que se puede calcular una distancia de viaje. Estos pueden ser la entrada de la provincia o de alguna localidad cercana que tenga una carretera por la que algún automóvil pueda recorrer.

Usando una lista de provincias provista por la página web del INEI<sup>1</sup>, se creó un script en Python que calculó la distancia en automóvil desde una provincia hasta las demás. Esto se hizo para todas las provincias de la lista. Cabe resaltar que se limpió esa lista antes de su uso, eliminando las observaciones repetidas y las tildes para evitar que el API encontrara un error al momento de encontrar alguna provincia. Asimismo, la matriz final expresó las provincias en términos de su UBIGEO en lugar de su usar sus propios nombres.

Google Maps no siempre puede calcular la ruta de una provincia a otra, pues eso está sujeto al medio de transporte que se haya escogido. De hecho, sobre las 196 provincias de estudio, 13 no tenían datos de distancias en kilómetros debido a que no se reportaba en Google Maps. Para esto, extrapolamos estas distancias utilizando los datos de los distritos más cercanos que sí contaban con esta información.

Usar esta API implica un costo de US\$5 por cada 1000 cálculos y se cuenta con un crédito gratuito mensual de US\$200, así como uno anual de US\$300, por lo que es limitado el uso que se le puede dar. Sin embargo, este es uno de los más eficientes para esta tarea.

### A.2. Aproximación de las distancias entre distritos

Para convertir las distancias entre provincias a distancias entre distritos aplicando la siguiente aproximación. Sea  $I$  la capital de la provincia a la que pertenece el distrito  $i$  y  $J$  la capital de la provincia a la que pertenece el distrito  $j$ , entonces la distancia del

---

<sup>1</sup>Esta puede descargarse [aquí](#)

distrito  $i$  al distrito  $j$  en kilómetros sobre vía terrestre se calcula de la siguiente manera:

$$d_{ij} = D_{iI} * 111 + d_{IJ} + D_{Jj} * 111, \quad (11)$$

donde  $d$  denota distancia en kilómetros por vía terrestre y  $D$  hace referencia a la distancia euclidiana entre dos distritos. Esta formula supone que una correcta aproximación a la distancia entre dos distritos es la suma de la distancia euclidiana entre el distrito de origen y la capital de su provincia (convertida a kilómetros), la distancia en kilómetros entre la capital de la provincia del distrito de origen y la análoga para el distrito de destino (obtenidas del API) y la la distancia euclidiana entre el distrito de destino y la capital de su provincia (convertida a kilómetros). Para esto, calculamos las distancias euclidianas entre distritos empleando las coordenadas de los centroides de cada distrito. Además, las convertimos de euclidianas a en kilómetros multiplicándolas por 111 que, según una muestra de distancias en kilómetros entre distritos a las que tenemos acceso, es el promedio de los ratios  $d/D$ .

## B. Figuras para los otros modelos

### B.1. Según el modelo estándar

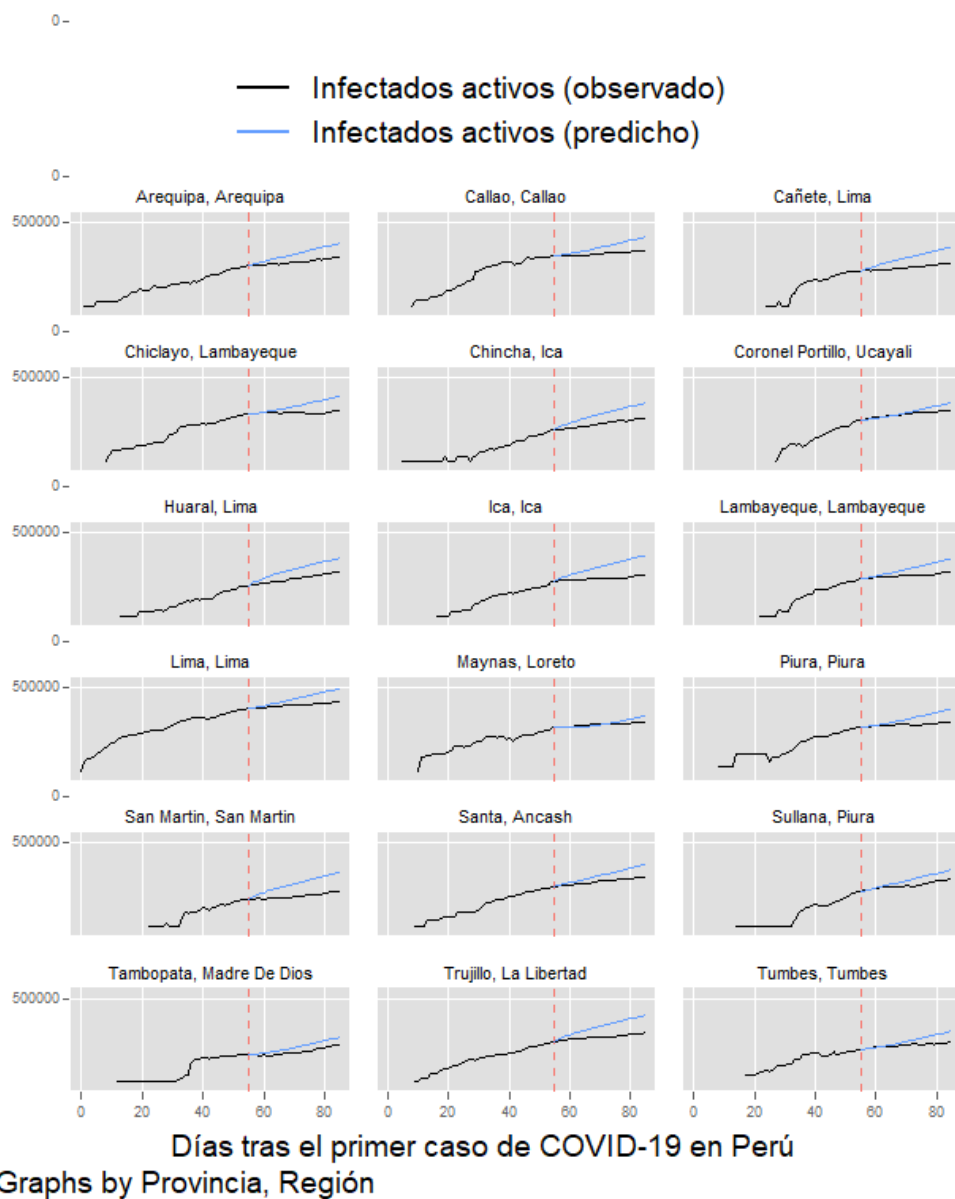


Figura 4: Predicción del número de infectados activos en las 18 provincias con mayor número de infectados al 17 de junio (escala logarítmica)

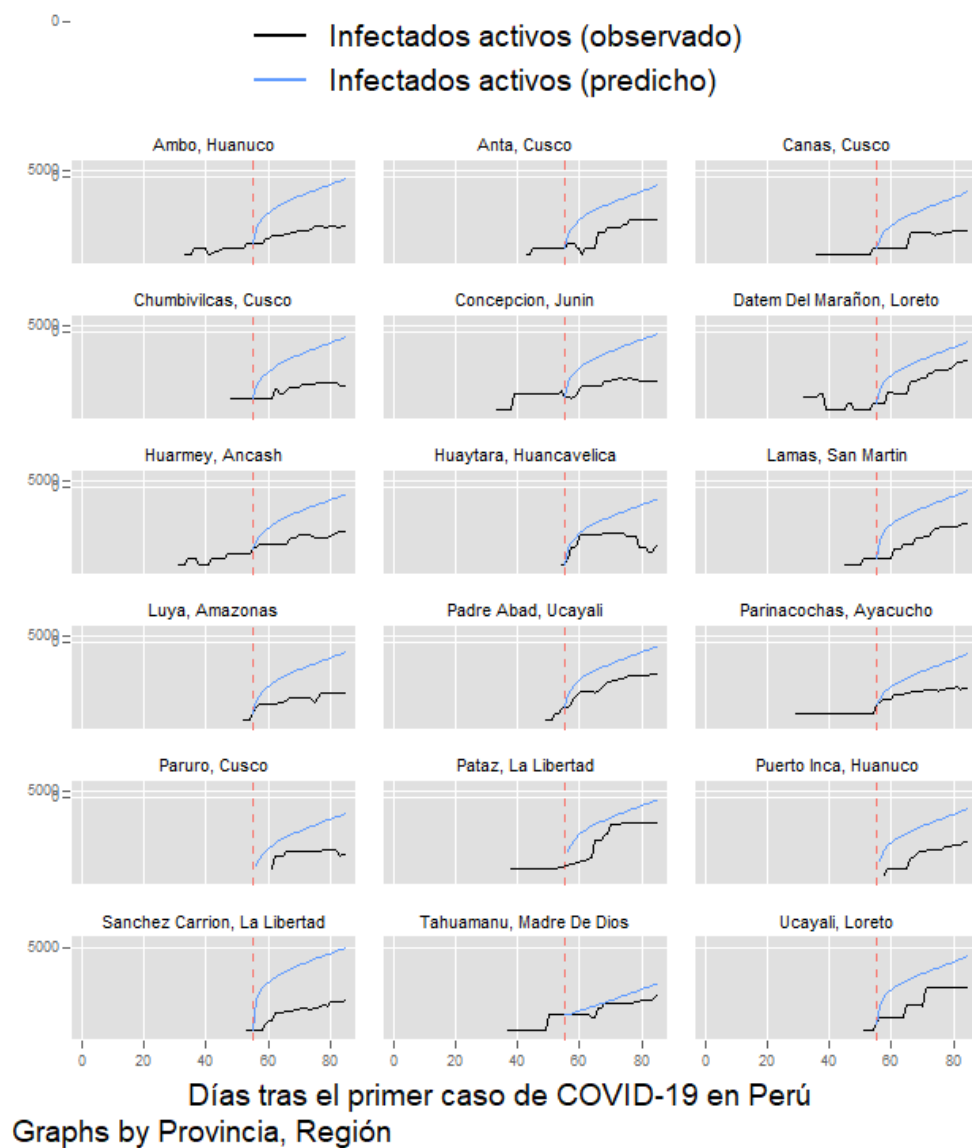


Figura 5: Predicción del número de infectados activos en las 18 provincias con pocos infectados y mayor crecimiento de la infección (escala logarítmica)

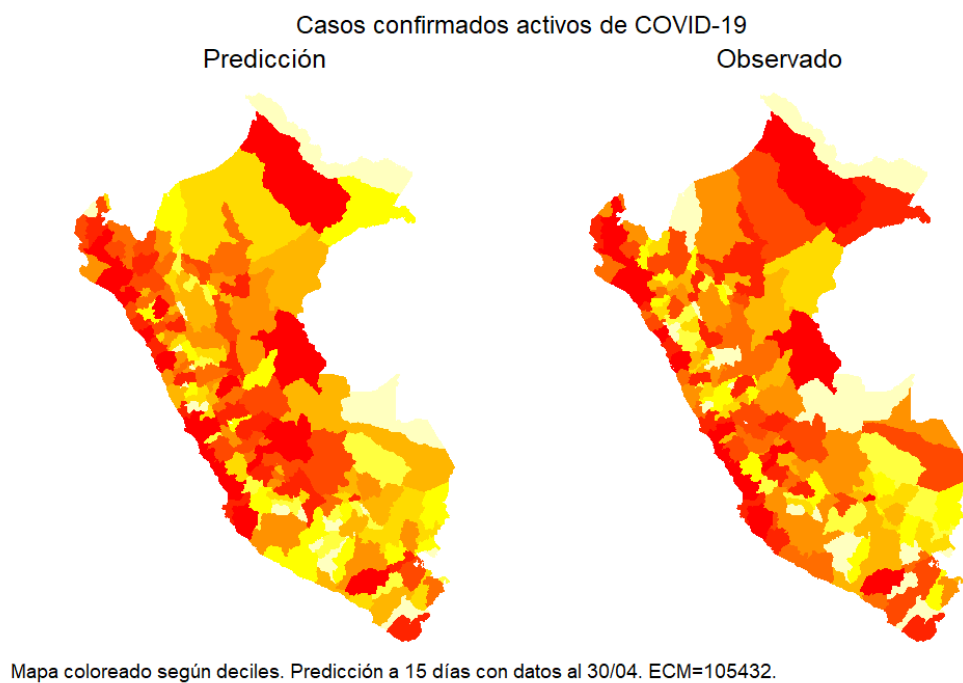


Figura 6: Comparación entre la predicción del modelo y lo observado a 15 días de la estimación con datos al 30 de Abril

## B.2. Según el modelo con densidades poblacionales y sin tendencias temporales

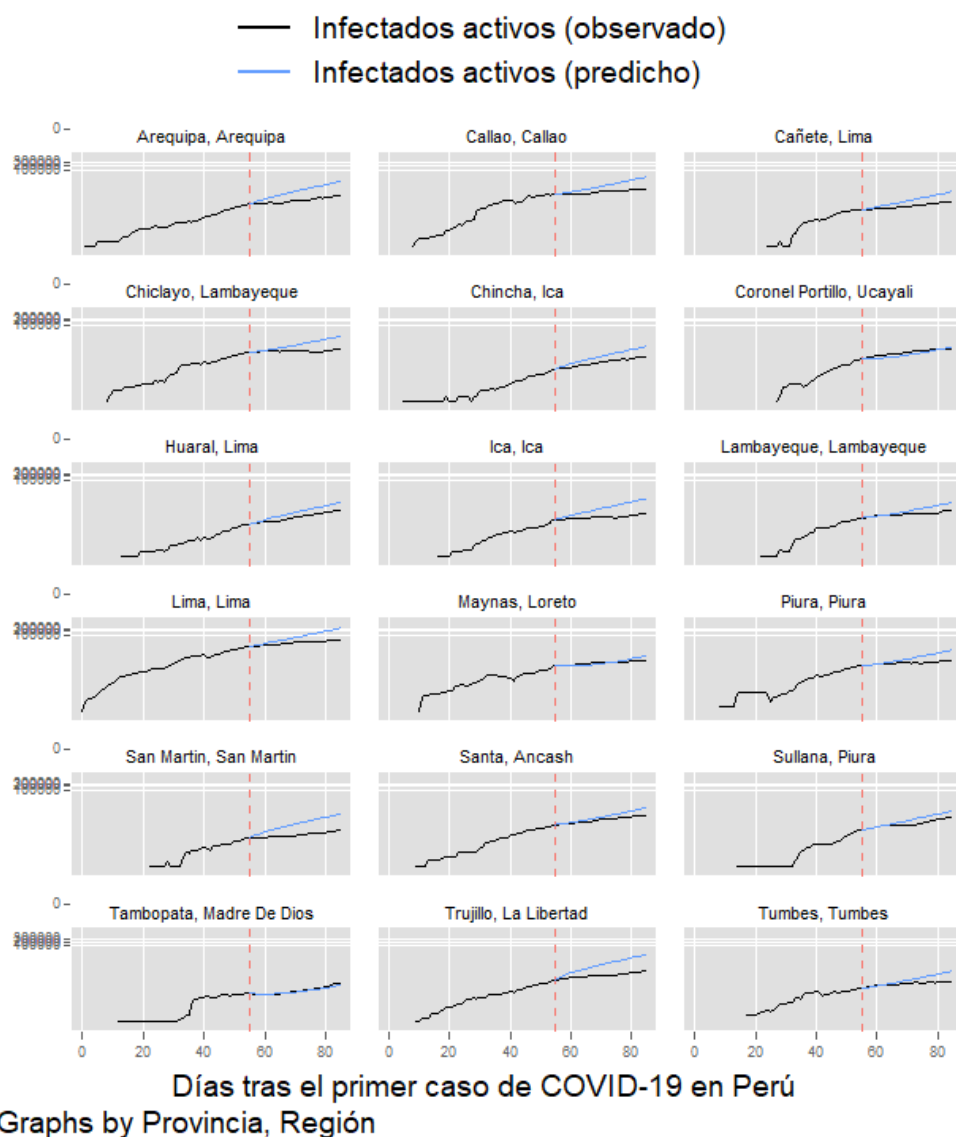


Figura 7: Predicción del número de infectados activos en las 18 provincias con mayor número de infectados al 17 de junio (escala logarítmica)



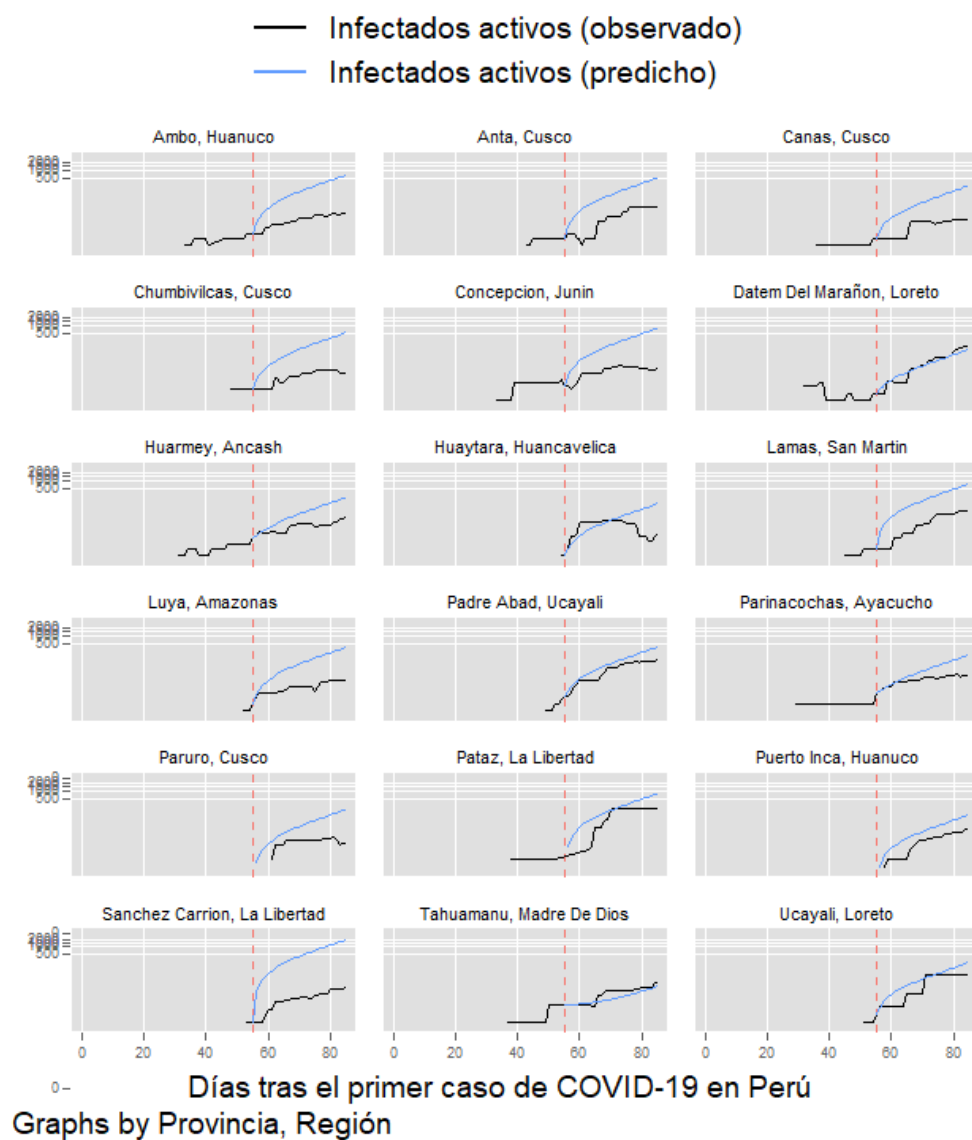


Figura 8: Predicción del número de infectados activos en las 18 provincias con pocos infectados y mayor crecimiento de la infección (escala logarítmica)

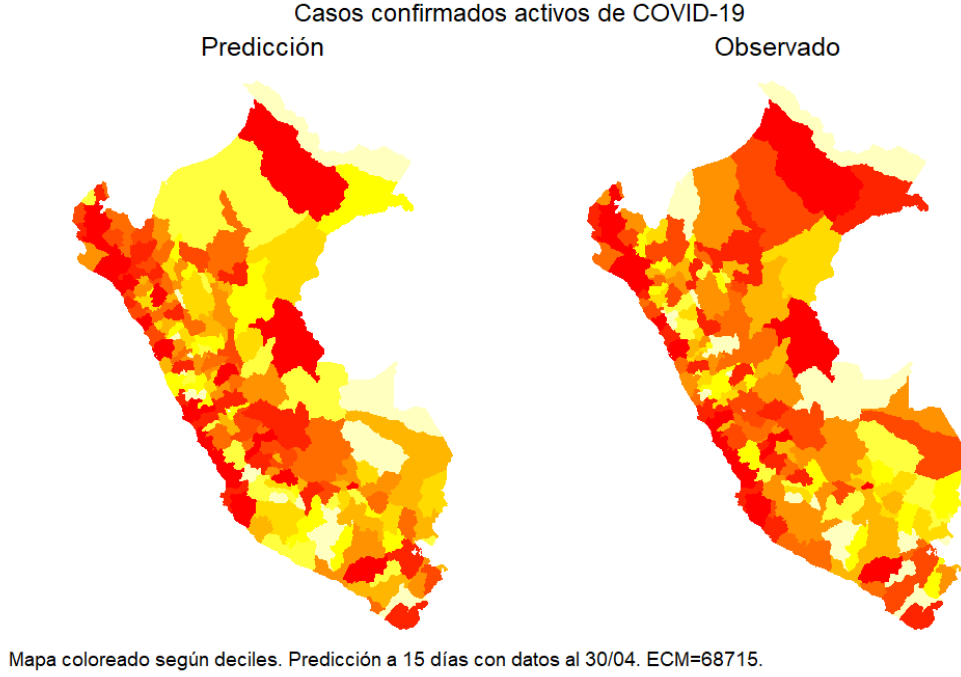


Figura 9: Comparación entre la predicción del modelo y lo observado a 15 días de la estimación con datos al 30 de Abril

### C. Comparación predictiva frente a estimar varios modelos SIR simples por cada distrito

Con la finalidad de comparar la utilidad de nuestro modelo espacial en el que el efecto contagio entre distritos es un componente importante, decidimos comparar la capacidad predictiva de estos modelos con otro escenario en el que estimemos un modelo SIR simple (sin influencia de otros distritos) para cada distrito; es decir estimamos 1874 modelos SIR que cumplen la siguiente especificación:

$$I_{i,t} = (1 - \gamma)I_{i,t-1} + \beta(I_{i,t-1}/N_i)S_{i,t-1} + \varepsilon_{i,t}, \quad (12)$$

donde, al igual que en los modelos anteriores,  $\gamma$  es la tasa de recuperación y  $\beta$  es la tasa de transmisión. Estas regresiones también fueron estimadas a partir de mínimos cuadrados no lineales.

Usando cada vector de estimados por cada distrito, hacemos la predicción hasta un periodo máximo de 30 días después del 30 de abril. Las figuras 10 y 11 muestran la predicción para ambos grupos de provincias definidos en la sección 4.

Según ambas figuras, parece claro que la predicción de las provincias con mayor número de infectados es tan buena, o incluso mejor en algunos, que el modelo espacial que proponemos en este documento. No obstante, la predicción para provincias con muy bajo número de infectados y alta tasa de crecimiento es muy deplorable. De hecho, estos modelos individuales no pueden hacer predicciones de los distritos que no han tenido variación en el número de infectados durante el periodo de estudio y esto es un problema, puesto que existen muchos distritos que hasta antes del 30 de abril no reportaban casos activos de COVID-19.<sup>2</sup> En la figura 11, una de aquellas provincias es Huaytara, a la cual imputamos una línea de predicción (azul) recta. También hicimos la comparación con varios modelos SIR individuales por distrito que incorporen además tendencias temporales lineales y cuadráticas, como las del modelo que presentamos en el documento, aunque la predicción fue incluso peor, puesto que para varias provincias el número de infectados pasaba a ser negativo (debido a la tendencia cuadrática) antes de haber pasado 15 días.

---

<sup>2</sup>Para 1249 distritos (de 1874) no se pudo estimar la regresión debido a falta de variabilidad.

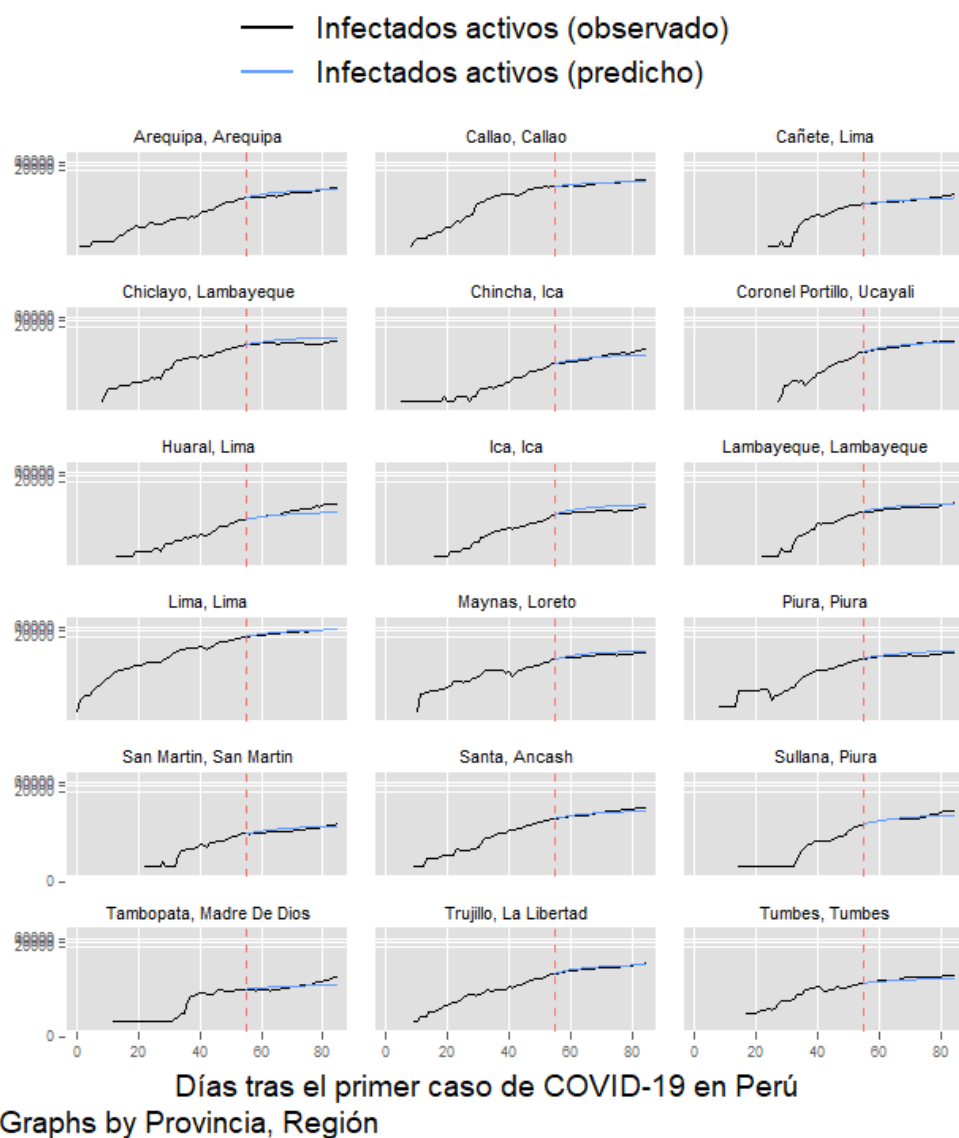
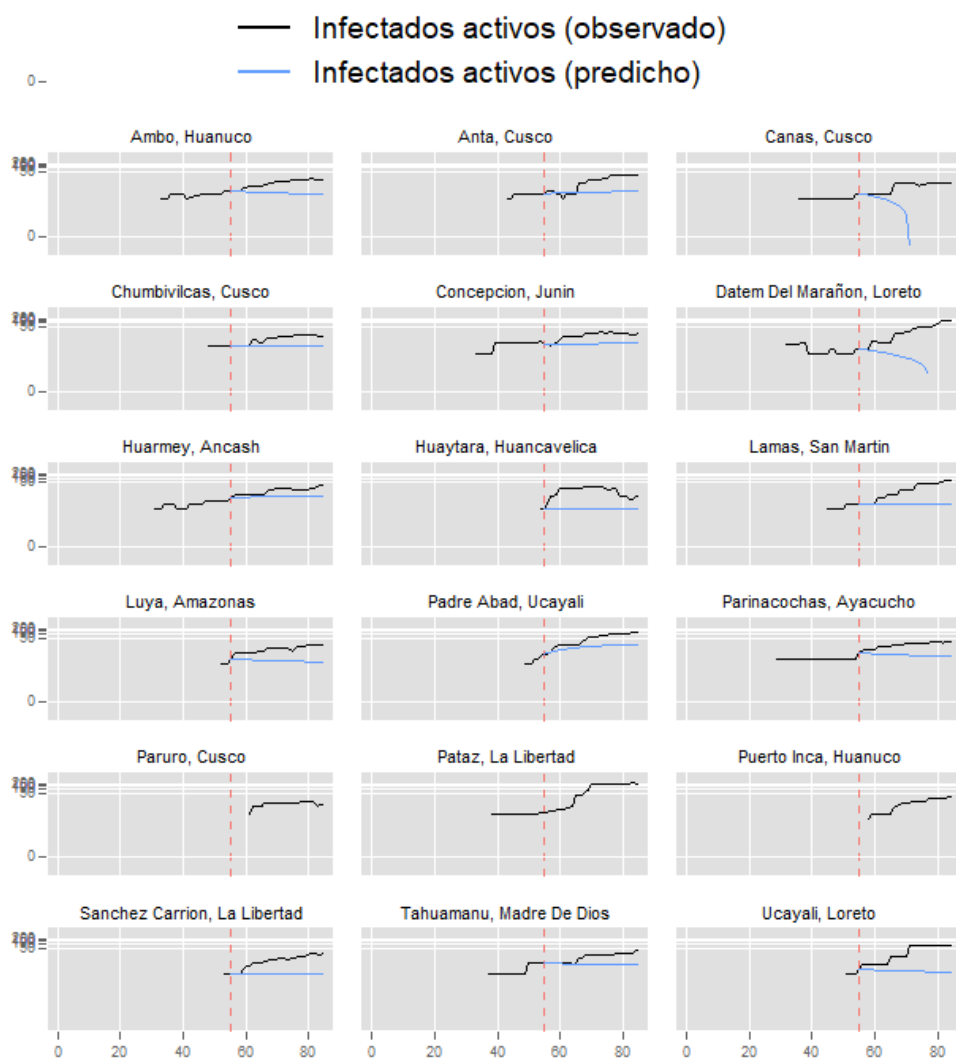


Figura 10: Predicción del número de infectados activos en las 18 provincias con mayor número de infectados al 17 de junio, usando modelos individuales por distrito (escala logarítmica)



**Días tras el primer caso de COVID-19 en Perú**  
**Graphs by Provincia, Región**

Figura 11: Predicción del número de infectados activos en las 18 provincias con pocos infectados y mayor crecimiento de la infección, usando modelos individuales por distrito (escala logarítmica)