

---

# ESTIMACIÓN DE LOS CASOS DE COVID-19 Y BROTOS EN TIEMPO REAL A TRAVÉS DE LLAMADAS TELEFÓNICAS

---

EZEQUIEL ÁLVAREZ \*  
DANIELA OBANDO \*\*  
SEBASTIÁN CRESPO \*\*  
ENIO GARCÍA \*\*  
NICOLÁS KREPLAK \*\*  
FRANCO MÁRSICO \*\*

## Resumen

Uno de los principales problemas, para controlar la propagación de la epidemia de COVID-19, es el retraso en la confirmación de los casos. Tener información sobre la evolución de la epidemia o los brotes, es crucial en la toma de decisiones para las políticas de Salud Pública. El presente trabajo propone un algoritmo que estima el número de casos de COVID-19, en tiempo real, utilizando los datos de las llamadas telefónicas a una línea de COVID-19. Modelando las llamadas como fondo (proporcional a la población) más señal (proporcional a los infectados), se ajustó el indicador en la Provincia de Buenos Aires (Argentina) con un coeficiente de determinación  $R^2 > 0,85$ . El resultado permitió estimar el número de casos dado el número de llamadas de un distrito específico, días antes que los resultados del laboratorio estuviesen disponibles. El algoritmo fue validado con datos reales y permitió hacer seguimiento temprano. Además, se realizó un seguimiento detallado de las incertidumbres en las estimaciones, ya que la alarma utiliza la importancia de los observables como indicador principal para detectar una anomalía. Se presenta en detalle el caso de Villa Azul, Quilmes, donde resultó crucial para controlar un brote a tiempo. Las herramientas fueron diseñadas urgentemente con los datos disponibles, muchas de ellas, están actualmente en desarrollo y, por lo tanto, tienen ciertas limitaciones con posibilidad de mejoras.

**Palabras Clave:** Vigilancia epidemiológica, Vigilancia molecular, Sistemas de alarma temprana, COVID-19.

---

\*International Center for Advanced Studies (ICAS), ICIFI-CONICET y ECyT-UNSAM, Campus Miguelete, Buenos Aires, Argentina.

\*\*Ministerio de Salud de la Provincia de Buenos Aires, La Plata, Buenos Aires, Argentina.

---

## 1. Introducción

---

La epidemia de COVID-19 está causando daños globales en prácticamente todos los aspectos de la sociedad mundial, desde principios de 2020. Aunque se está haciendo un gran esfuerzo en muchos campos de la ciencia para mitigar sus efectos, la enfermedad se está propagando continuamente y, en muchas regiones, una segunda ola está causando grandes preocupaciones. Las dificultades para controlar la epidemia se deben, en parte, a la combinación crucial de ser altamente contagiosa [7], tener un largo período de incubación [15], durante el cual, es posible el contagio unos días antes de la aparición de los síntomas [10], tener casos leves o asintomáticos [7] y, también, porque el diagnóstico puede tardar días desde que se contacta con el sistema sanitario. En particular, esto último, hace que los brotes se propaguen y la epidemia evolucione mientras se procesan los resultados de laboratorio. Este efecto es más importante en los países de ingresos bajos y medianos debido a problemas operativos y logísticos, generalmente causados por las desigualdades tecnológicas y económicas [1, 21].

En las catástrofes masivas los Sistemas de Alerta Temprana (SATs) juegan un papel clave para la mitigación de daños [9, 14, 20], disminuyendo los tiempos de respuesta por parte de los equipos de emergencia y mejorando su efectividad. La principal estrategia de los SATs en la vigilancia de enfermedades infecciosas es la incorporación de la información producida lo más cercanamente posible al momento de la infección [8, 13]. En este caso, la aparición de los síntomas y su detección por parte de la percepción de salud individual y comunitaria, es la primera señal detectable de un brote. Los sistemas de alerta temprana basados en la vigilancia sindrómica se han aplicado en la vigilancia epidemiológica para la identificación y confirmación temprana de los brotes [11, 12, 16, 18, 19]. Una de las principales características del sistema de alerta temprana es la utilización de la información sanitaria proporcionada por la población para activar las alarmas locales. Hoy en día, con el amplio uso de aplicaciones de telefonía celular y líneas telefónicas específicas del sistema de salud, constantemente se generan importantes bases de datos con información sobre la vigilancia sindrómica [5]. La geolocalización juega un papel principal en la definición espacial y temporal de los brotes detectados por los sistemas de alerta temprana [4].

En la Provincia de Buenos Aires, Argentina, la línea telefónica 148 de COVID-19 es uno de los primeros contactos entre una persona que se cree

infectada y el Sistema de Salud. El equipo capacitado de Atención Médica recibe y responde a las preguntas de las personas generando, simultáneamente, una base de datos de vigilancia sindrómica. Si la persona tiene síntomas que podrían indicar una infección por COVID-19, se le indica que siga el protocolo correspondiente. Es importante señalar que, esa base de datos de síndromes, se utilizó como insumo para la estimación de casos y la detección de brotes en la provincia de Buenos Aires.

En este trabajo, se presenta un método para mitigar los efectos de la epidemia mediante la estimación del número de casos de COVID-19 sin tener que esperar las confirmaciones de laboratorio. Esto proporciona al sistema de salud una herramienta para reaccionar con anticipación y evaluar las políticas de salud pública actuales o futuras. En la Sección 2 se describen los datos de la línea COVID y se presentan los detalles del modelo matemático para estimar el número de casos utilizando los datos de las llamadas telefónicas. En la Sección 3.1 se muestra cómo funciona el modelo en la Provincia de Buenos Aires y cómo puede ser usado para rastrear la epidemia en-vivo. En la Sección 3.2 se presenta la Alarma de brote temprano y sus detalles a través del caso de Villa Azul, Quilmes. Las limitaciones y mejoras actuales del modelo se discuten en la Sección 4 y, finalmente, se presentan las conclusiones en la Sección 5.

---

## 2. Métodos

---

En esta sección, se describe el modelo matemático implementado para relacionar las llamadas telefónicas a una línea COVID-19 con los casos confirmados por el laboratorio por distrito y por día. En los siguientes párrafos se esboza el funcionamiento de la línea COVID y, luego, se describen los detalles del modelo.

### 2.1. COVID - línea 148 en la provincia de Buenos Aires

La Provincia de Buenos Aires (PBA) es la provincia más poblada de Argentina, con más de 17 millones de habitantes. Alrededor de 13 millones de personas viven en el área metropolitana que rodea la Ciudad de Buenos Aires. Es importante destacar que, los 4 millones restantes, viven en una vasta área y con baja densidad de población conocida como el interior de la provincia. Esta heterogeneidad demográfica conduce a un sistema de salud hiper-centralizado. Con el fin de atender la creciente demanda de asistencia médica causada por el COVID-19, las autoridades de salud pública implementaron, en febrero de 2020, un teléfono COVID específico al que se accede marcando 148. El obje-

tivo de esta línea COVID es atender todas las inquietudes de la comunidad relacionadas con la pandemia, lo que incluye preguntas, dudas, informes de síntomas y referencia al sistema de atención sanitaria, entre otros.

La línea COVID creció en la cantidad de personal a medida que la epidemia se propagó en la PBA. La solicitud de llamadas pasó de unos pocos cientos por día en marzo hasta aproximadamente 20 mil por día a finales de agosto. Hasta finales de junio, el sistema no se saturó y se atendieron todas las llamadas que requerían asistencia. Lo cual permite considerar que, durante este periodo, un indicador procedente de esta línea COVID sería relativamente insesgado. Esto es especialmente cierto si se compara con otros indicadores como las pruebas o el procesamiento en laboratorio, que estaban cambiando su comportamiento considerablemente a medida que la epidemia se propagaba. El periodo comprendido entre el 1 de abril al 26 de junio de 202, la línea COVID mostró ser relativamente estable a los cambios importantes.

Cuando la gente llama a la línea COVID 148, entra en un menú de voz automático en el que una de las opciones corresponde a síntomas similares a los de COVID. Cuando el usuario entra en esta opción, su llamada es atendida por un operador capacitado y, aplicando un breve cuestionario sobre su experiencia, indica si la llamada no pasa el umbral para ser registrada o si corresponde a una de las dos categorías registradas: contacto cercano y caso sospechoso. Si la llamada corresponde a alguna de estas categorías, el operador registra sus datos y, en particular, el distrito desde el que llama. En la Figura 1 se muestra el flujo de trabajo de la línea COVID. En la fase inicial de la implantación del sistema, el registro no incluía información fiable sobre la dirección exacta del usuario. Este hecho crucial llevó a desarrollar el sistema, explicado en la Sección 2.2, restringiendo la información sobre el usuario sólo a su distrito. Se prevén futuras actualizaciones del sistema que proporcionen una localización más precisa de la llamada, sin embargo, el trabajo actual, se restringe al distrito del llamante y sólo una vez que su llamada fue tomada por un operador entrenado por COVID.

## 2.2. Modelo matemático para estimar los casos de llamadas telefónicas a la línea 148 COVID

En esta sección se presenta el modelo matemático para estimar los nuevos infectados usando los datos de la llamada telefónica, aplicado a la PBA. La estructura mantiene las mismas líneas que [2], pero con diferentes propósitos y diferentes filtros en el conjunto de datos.

El conjunto de datos consideró llamadas de muchos distritos y durante un rango de tiempo determinado a una línea COVID. Cada una de estas llamadas



Figura 1: Flujo de trabajo de la línea 148 de COVID.

Nota: A medida que las personas llaman a la línea COVID, según su percepción de la salud, los operadores formados en COVID-19 determinan si corresponden a casos sospechosos o de contacto cercano. En tal caso, su historial pasa al equipo de vigilancia epidemiológica y se ordena un hisopado para análisis de COVID-19. Algunos días después, el resultado del hisopado se añade al registro correspondiente. El flujo de información ordinario se esquematiza con flechas negras. El algoritmo descrito en este documento funciona con la primera parte de la información que se entrega en el flujo de trabajo a medida que los operadores determinan que el caso se corresponde con los criterios correspondientes (flecha roja).

puede ser:

*fondo*: personas con síntomas similares, pero no infectadas

*señal*: personas infectadas con COVID-19.

En este caso, el “*fondo*” se corresponde a aquellas llamadas no atribuibles a casos reales de COVID-19 y que, por lo tanto, pueden ser caracterizadas como “*ruido*”. En cambio, la “*señal*” es lo que realmente se quiere medir/detectar y se corresponde a las llamadas relacionadas con casos de COVID-19.

Bajo supuestos razonables de homogeneidad en el espacio y el tiempo, es posible modelar que las llamadas de *fondo* en cada distrito y ventana de tiempo son proporcionales a la población total del distrito y a la duración de la ventana de tiempo. Mientras que, las llamadas de *señal*, son proporcionales al número total de personas infectadas en el distrito cuyo registro se abre en la correspondiente ventana de tiempo, aunque su confirmación de laboratorio pueda estar disponible en un momento posterior. Por lo tanto, si se divide todo el conjunto de datos en trozos correspondientes en el espacio a los distritos en la PBA, y en el tiempo a las ventanas de tiempo de  $\Delta t^{(j)}$  días, que pueden ser elegidos arbitrariamente.

De esta manera, se definen los parámetros:  $N_p$ , es la población del distrito correspondiente;  $N_I$ , es el número de infectados confirmados en el mismo distrito cuyo registro fue abierto durante la ventana de tiempo correspondiente;  $n_c$  es el ajuste al número total de llamadas; y  $N_c$ , no presente en la ecuación, es el número total de llamadas realmente realizadas. Se debe observar, por lo tanto, que este conjunto de ecuaciones (una para cada ventana  $j$ ), puede extenderse en función de la longitud de la ventana de tiempo elegida. Una vez planteado este conjunto de ecuaciones  $j = 1, \dots, k$ , es posible ajustar los mejores valores de los coeficientes  $\theta_p, \theta_I$  que minimizan la distancia cuadrada entre  $n_c$  y  $N_c$ . Es importante considerar que sólo hay dos coeficientes ( $\theta_p, \theta_I$ ) que deben ajustarse a todas las ecuaciones diferentes para cada ventana. Se plantea, entonces, la siguiente ecuación para todos los periodos  $j$ :

$$n_c^{(j)} = \theta_p \Delta t^{(j)} N_p^{(j)} + \theta_I N_I^{(j)} \quad (1)$$

Este ajuste funciona mejor si todas las ventanas corresponden a períodos en los que los métodos de prueba no han cambiado drásticamente, como puede ser, por ejemplo, si el número de pruebas diarias se modifica considerablemente, o si se consideran nuevos síntomas como umbral para la prueba, entre otros. La razón de exigir esto, es tener un equilibrio coherente entre el número de infectados notificados y el número de llamadas en todas las ventanas. Con este

objetivo, es mejor ajustar los parámetros cada vez que se produzcan cambios importantes en los métodos de prueba y de notificación.

Una vez ajustados los parámetros  $\theta_{p,I}$  en la Ecuación 1, incluyendo su incertidumbre del ajuste, es posible estimar el número de nuevos infectados en una determinada ventana como:

$$n_I^{(j)} = \frac{1}{\theta_I} (N_c^{(j)} - \theta_p \Delta t^{(j)} N_p^{(j)}) \quad (2)$$

Obsérvese que, el lado derecho, requiere datos que se obtienen en el mismo día y, por lo tanto, se puede estimar el número de casos  $n_I$  en-vivo, sin necesidad de esperar los resultados del laboratorio. A su vez, el algoritmo permite estimar el número total de casos nuevos en cada tramo, pero no determinar cuál de las llamadas corresponde a los nuevos casos. La incertidumbre en la estimación de  $n_I$  se calcula aplicando la fórmula de propagación de errores en la Ecuación 2. Si las variables están correlacionadas, como por ejemplo  $\theta_p$  y  $\theta_I$ , se debe tener esto en cuenta, sin embargo, en el caso utilizado en este estudio, no considera esta correlación en comparación con otros términos. Para los parámetros  $\theta_{p,I}$  se utilizó la incertidumbre proveniente del ajuste, para  $N_c$ , se utilizó la incertidumbre de Poisson; y para  $N_p$  se debe decidir si se añade una incertidumbre sistemática o sólo se utiliza Poisson, como se realizó en este trabajo. Como se discute más adelante, las incertidumbres en las estimaciones juegan un papel central en el diseño de la alarma de brotes tempranos y, por lo tanto, deben ser manejadas con cuidado, especialmente las sistemáticas si están presentes.

Para aplicar este algoritmo en la PBA, se utilizó el conjunto de datos de las llamadas telefónicas a la línea 148 COVID. Se trabajó con todas las llamadas telefónicas que entraron en la línea COVID y que alcanzaron el umbral de ser un contacto cercano o un caso sospechoso. La razón de la granularidad seleccionada para las alarmas es porque el distrito desde el que llama el usuario está registrado por el operador. Aunque la dirección también está registrada, la mayoría de las veces, en la práctica poseen muchas ambigüedades, palabras mal escritas, u otros errores no intencionados dan como resultado que sólo alrededor de 50 %-70 % de los casos puede ser reconstruido correctamente. Se consideró el conjunto de datos de las llamadas entre el 1 de abril y el 26 de junio, ya que, posteriormente, el centro de llamadas se sobrecargó, produciendo sesgos intratables. A lo largo de este período, se ajustaron los datos varias veces en diferentes conjuntos de datos, obteniendo resultados bastante similares y con un coeficiente de determinación siempre satisfactorio  $R^2 > 0,85$ . En particular, a medida que aumentaban los casos, se obtuvieron estimaciones más precisas para  $\theta_I$ , como es de esperar.

Para mostrar la robustez de las hipótesis, se muestra cómo funciona este modelo con los datos del 1 de mayo al 26 de junio, divididos en dos ventanas de tiempo de igual duración. Se consideraron todos los distritos en la PBA cuyo número de llamadas en estos trozos fue superior a 100. Después de este filtrado, se contó con 43 trozos, es decir, 43 puntos de datos. Después de realizar el ajuste indicado en la Ecuación 1 se obtiene:

$$\theta_p = (5,16 \pm 1,59) \times 10^{-6} \text{ llamadas por habitante por día}$$

$$\theta_I = 0,69 \pm 0,05 \text{ llamadas por persona infectada}$$

Vale la pena notar que, los valores precisos de estos coeficientes ajustados, tienen una fuerte dependencia del proceso de filtrado de llamadas y de la arquitectura del sistema de llamadas. En particular, estos valores difieren de los presentados en [2], porque se consideró un nivel diferente de filtrado para obtener el distrito de cada usuario. El ajuste de este conjunto de datos arroja un coeficiente de determinación  $R^2 = 0,91$ , que indica la solidez de las hipótesis implicadas. La Figura 2 muestra la comparación entre los datos y el ajuste para el número de llamadas telefónicas, tal como se plantea en la Ecuación 1.

---

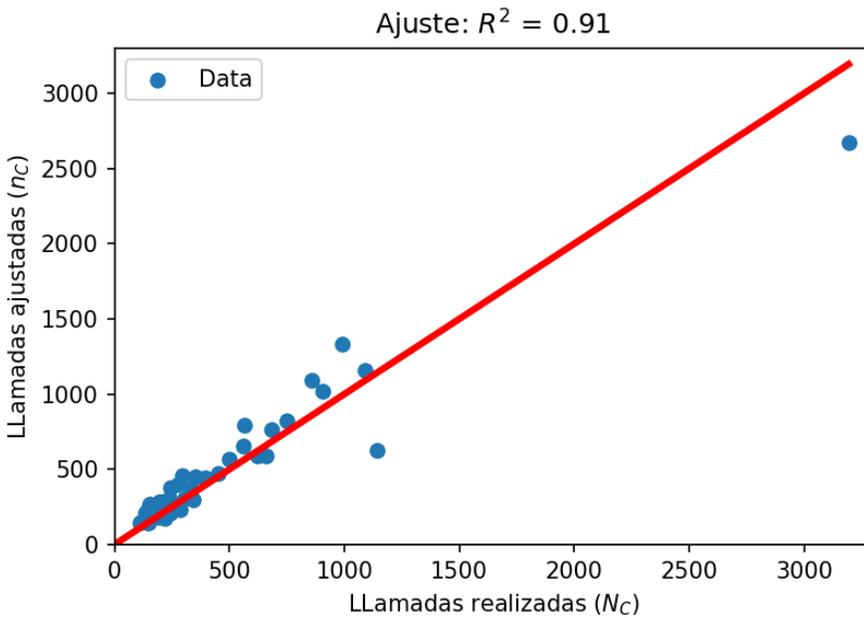
### 3. Resultados

---

#### 3.1. Seguimiento de la epidemia a través de estimaciones de modelos

El modelo matemático descrito en la Sección 2.2, proporciona un marco para estimar con muchos días de antelación el número de casos confirmados por laboratorio por día, en función de la distribución espacio-temporal de las llamadas telefónicas a la línea COVID. Se trata de un logro convincente, porque la información de la llamada telefónica está disponible en tiempo real, mientras que la confirmación de los casos por parte del laboratorio puede requerir desde unos pocos días hasta una semana desde que los pacientes informan de sus primeros síntomas. A lo largo de esta sección se muestra cómo utilizar este sistema para tener una estimación de la evolución de la epidemia en tiempo real, junto con los resultados de los casos reales en la PBA.

Como este sistema fue desarrollado y aplicado en el momento, no hubo tiempo para la validación. Sin embargo, la obtención de un  $R^2 > 0,85$  en el ajuste fue una señal satisfactoria de que el modelo funcionaba bien hasta el



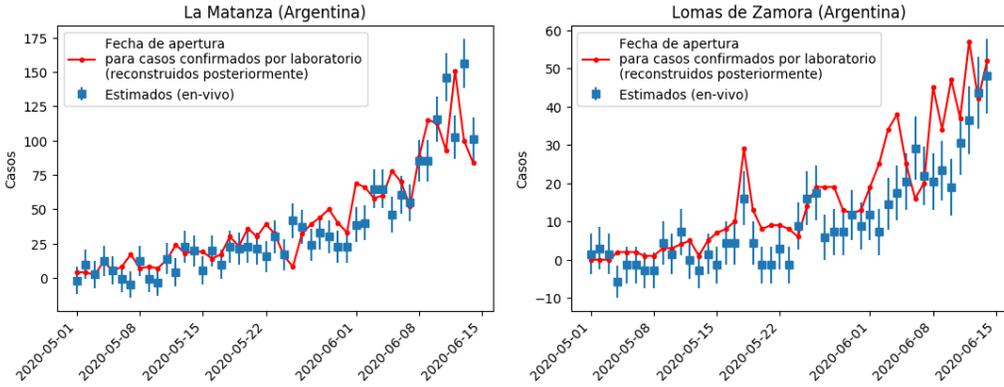


Figura 3: Comparación de datos reales versus estimación del modelo con una barra de error que representa un IC del 95 % para dos distritos ejemplo de la Provincia de Buenos Aires (La Matanza y Lomas de Zamora).

Nota: La línea roja correspondiente a los casos reales confirmados con su apertura de registro en la fecha correspondiente, se reconstruye muchos días después. En las fechas en las que la línea roja supera la estimación suele ser debido a que los operativos de DETECTAR (pruebas puerta a puerta [17]), fueron llevados a cabo. En general, el modelo arroja una muy buena estimación para monitorear la epidemia en todos los distritos afectados de la PBA.

### 3.2.1. Identificación de la formación de un brote

Con la estimación en tiempo real de los casos por día en cada distrito, es de interés desarrollar una herramienta estadística y automática que pueda disparar una alarma cuando un posible brote esté gestándose. Tener una alarma temprana sobre este tipo de características epidemiológicas, es una herramienta crucial para evitar su propagación y consecuencias drásticas.

Para detectar un posible brote hay muchos indicadores que deben ser analizados simultáneamente. Por un lado, es importante tener una estimación del número absoluto y relativo diario de casos y, por otro lado, también es importante tener una estimación de la variación diaria de estos observables. Para tener un indicador cuantitativo objetivo del potencial de un brote en una región determinada, es esencial tener una evaluación correcta de las incertidumbres en todas las estimaciones del modelo. A lo largo del sistema implementado como Alarma de brotes tempranos, se ha considerado la detección de anomalías una anomalía como un indicador central. Aquí la significancia se define como la distancia a cero del valor central del indicador, medida en unidades de su incertidumbre. O, en otras palabras,

$$\text{significancia} = \frac{\text{Valor central}}{\text{Incerteza}} \tag{3}$$

El algoritmo desarrollado, calcula diariamente la estimación del número total de casos nuevos en cada distrito en la PBA. Dado que, en la franja temporal estudiada, especialmente antes de junio, el número de casos estimados por día de muchos distritos estaba por debajo de  $\sim 5 - 10$ , consideramos incluir la estimación de los casos de los últimos dos días. Esto reduciría la relativa incertidumbre de Poisson debido a las pequeñas cifras. Se calculó que, el número de casos estimados en valor absoluto, y también en relación con 100 mil habitantes, era igualmente sensible para todos los distritos.

Un tercer y decisivo observable, que señala el nivel de peligro de un brote, es el aumento diario de los casos estimados. Dada la estimación diaria proporcionada por el modelo matemático, es posible reconocer una curva de rápido incremento de muchas maneras. Se ha considerado un ajuste lineal a la estimación de casos de los últimos 3 días y, la pendiente de esta recta, como una estimación del valor central del incremento diario. También se utilizó la significación como el indicador más relevante para decidir el nivel de peligro de cada distrito. En este caso, el cálculo de la barra de error en la pendiente de la línea, incluye todas las incertidumbres de la estimación de cada día incluidas en el cálculo de la incertidumbre de ajuste a través de los residuos de mínimos cuadrados. Se utilizaron 3 días para ajustar una línea porque es el tiempo mínimo necesario para ver un aumento consecutivo de dos días, mientras que todavía se está muy por delante de los resultados del laboratorio. Además, tres días es también una buena ventana de tiempo para las características específicas de COVID-19.

Esta alarma de brotes tempranos, ha proporcionado a la administración del Ministerio de Salud de la PBA herramientas muy importantes para identificar posibles brotes durante el aumento de la curva epidémica. Dado que la granularidad del algoritmo es muy pobre (distritos), el sistema debe complementarse con otros indicadores independientes, en particular los que pueden ayudar a proporcionar una localización más precisa del brote. Esto se hacía normalmente llamando manualmente a los casos registrados, y luego enviando a agentes de DETECTAR [17], para verificar si de hecho las condiciones in situ serían las previstas. La alarma de brotes tempranos ha indicado muchos brotes que han sido controlados desde mediados de abril a mediados de junio. En particular, se describirá en la Sección 3.2.2 el muy especial caso de Villa Azul, Quilmes (ha sido cubierto en los titulares de las noticias durante varias semanas [3]), y se proporcionarán los detalles de cómo la Alarma de brotes tempranos indicó el distrito de Quilmes.

### 3.2.2. Estudio de caso: Villa Azul - Quilmes

En esta sección se muestran los detalles de uno de los brotes indicados por la Alarma de brotes tempranos a mediados de mayo en el distrito de Quilmes. Este caso fue el primer gran brote en un barrio de bajos recursos en la PBA y tuvo un gran impacto en las noticias [3], no sólo por su magnitud sino también por su detección temprana, que llevó a un estricto bloqueo y aislamiento del brote para controlar su propagación a los barrios cercanos.

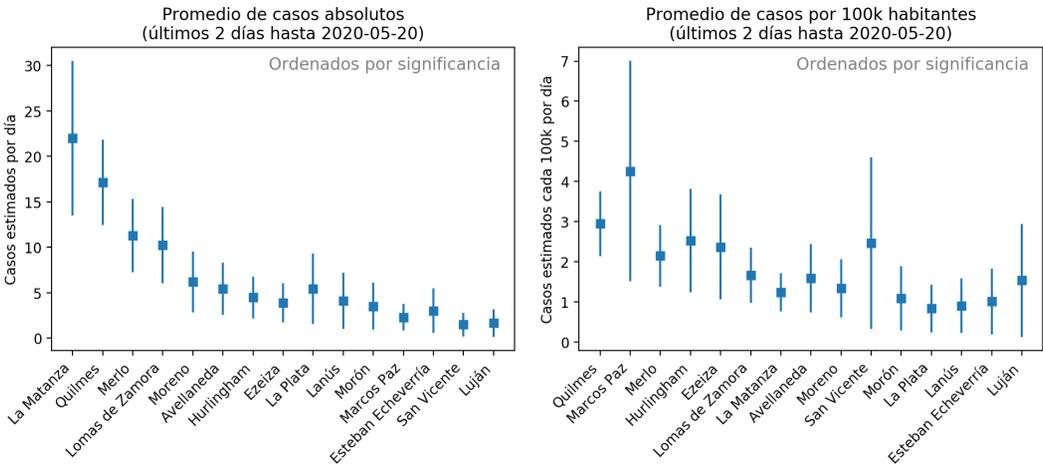


Figura 4: Estimación de casos por día usando el modelo de las llamadas telefónicas de la línea COVID de los últimos dos días.

Nota: Obsérvese que los distritos no están ordenados por su valor central, sino por su significado, que se define como la tasa entre el valor central y la incertidumbre. Es por eso que las barras de error son cruciales para proporcionar una alarma de brote temprano. Los resultados se muestran en valor absoluto (izquierda) y relativo a cada 100 mil habitantes (derecha).

Mientras que Quilmes está en la posición superior cuando se escala a relativo por cada 100 mil habitantes. El 20 de mayo la alarma indicaba un gran número de casos estimados en el distrito de Quilmes, en particular Quilmes tenía la estimación más alta en número de casos por habitantes de los dos últimos días, medida a través de la significación del indicador. En la Figura 4, se muestra el escenario del 20 de mayo en el que Quilmes es casi tan grande como La Matanza en valor absoluto, con  $\sim 1/3$  de su población. Además, el indicador del ajuste del aumento diario, también indicaba que Quilmes era el distrito de mayor importancia (véase la Figura 5). Este último indicador, sobre el ajuste de incremento diario a los últimos 3 días, puede ser visualizado en la Figura 6a, donde se trazó la estimación diaria de los últimos 7 días en

Quilmes. El ajuste se obtuvo usando los 3 últimos puntos de datos en rojo. Dado que todos estos indicadores apuntan al distrito de Quilmes, el equipo de vigilancia se encargó de localizar las llamadas telefónicas y observó un exceso procedente de Villa Azul, un barrio de bajos ingresos de Quilmes y próximo al distrito de Avellaneda.

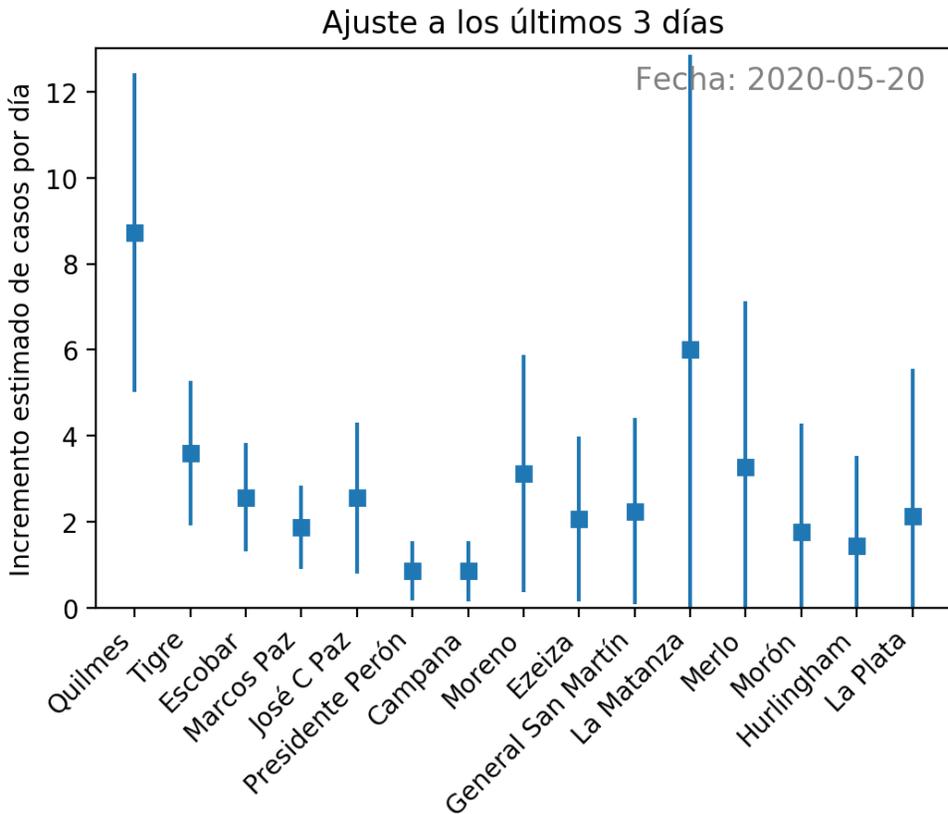


Figura 5: Pendiente de un ajuste lineal a la estimación de casos por día de los últimos tres días.

Nota: La barra de error corresponde a la inclusión de la incertidumbre en cada estimación por día y en la determinación de la pendiente en el ajuste. Esta gráfica es crucial para ordenar los distritos según la significación en esta variable. En la Figura se ve, el 20 de mayo, a Quilmes en la posición superior e indica una potencial alarma temprana de un brote, como fue consecuentemente confirmado por otros indicadores unos días después.

Estas observaciones, indicadas de antemano por la Alarma de brotes tempranos, debían ser verificadas por un indicador complementario independiente. Al día siguiente un operativo DETECTAR [17], fue enviado a Villa Azul, don-

de se verificó la situación aguda y se inició de inmediato el rastreo puerta a puerta con resultados de laboratorio urgentes. Como los primeros resultados estaban confirmando el brote en Villa Azul, la Administración de la PBA decidió un estricto encierro y aislamiento durante 14 días desde el 24 de mayo [3].

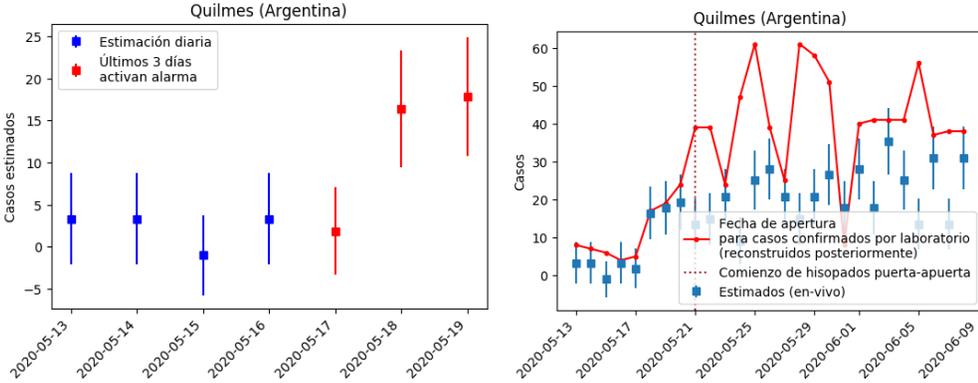


Figura 6: Izquierda: Visualización de la alarma de brote temprano el 20 de mayo para Quilmes, según lo dispuesto por el equipo de Atención de la Salud en la provincia de Buenos Aires. Derecha: Casos confirmados por el laboratorio, el día en que el Sistema de Salud aterrizó en Villa Azul para comenzar a probar puerta por puerta, y un rango más amplio de fechas para captar el panorama general del caso.

Nota: El estricto encierro en Villa Azul sin permiso de entrada ni salida duró desde el 24 de mayo hasta el 8 de junio. Como se puede ver en la trama, durante la prueba puerta por puerta, la línea sólida va por encima y no se correlaciona con los casos estimados por llamadas telefónicas, como se esperaba.

### 3.2.3. Descripción epidemiológica y operativa de Villa Azul

Villa Azul (Quilmes) y Villa Itatí (Avellaneda) son dos barrios adyacentes de bajos recursos. El último análisis demográfico indica que, Villa Azul, tiene una población de 3.128 habitantes y Villa Itatí de 15.142. La alta densidad de edificación y vivienda, y las pequeñas calles ponen a la población en estrecho contacto. Estas características, hacen que estos barrios sean susceptibles de una rápida difusión [6]. Teniendo esto en cuenta, la detección temprana de los brotes implica un desafío principal en estos casos complejos en los que la detección y el bloqueo de la propagación deben hacerse cuando se reportan los primeros casos. En particular, la alarma temprana descrita anteriormente para el brote, ocurrió en Villa Azul permitiendo una rápida respuesta del equipo

del Sistema de Salud para mitigar y controlar su propagación a Villa Itatí.

Una vez que se produjo el estricto cierre y aislamiento, el equipo de atención social entregó agua y alimentos. No se permitió a las personas salir de la casa durante todo el aislamiento. Los equipos sanitarios de vigilancia activa comenzaron con un monitoreo de los síntomas puerta a puerta. Se analizaron los casos con manifestaciones clínicas relacionadas con COVID-19. Los casos confirmados fueron aislados dentro de sus casas en los casos en que esto era posible (si había una habitación vacía, por ejemplo) y en los casos en que no lo era, las personas fueron enviadas a un centro fuera del hospital.

---

## 4. Perspectivas y alcance

---

El desarrollo del modelo matemático para estimar el número de casos de COVID-19 se hizo con urgencia y adaptándolo a los datos disponibles. No hubo tiempo de solicitar cambios en la adquisición ni en el procesamiento de los datos. Por supuesto, el algoritmo y el sistema pueden ser mejorados en muchas direcciones.

Una de las mayores debilidades del algoritmo es la gran granularidad, que corresponde a los distritos. Las poblaciones de los distritos en el área relevante son en promedio 500 mil personas. Esta cuestión, se traduce en que la alarma de brotes tempranos deja de funcionar una vez que la densidad de casos es tal que hay más de unos pocos brotes en cada distrito. Situación que ocurrió a finales de junio en la PBA. Para una futura implementación, e está llevando a cabo una solución a este problema, obteniendo una dirección de confianza del operador entrenado por COVID que toma la llamada. Una solución más estable sería obtener esta información de la compañía telefónica, sin embargo, las regulaciones muchas veces bloquean esta posibilidad.

Por otro lado, el algoritmo tiene un beneficio muy importante que es su imparcialidad. Dado que la línea COVID funciona 24 horas los 7 días de la semana y con una metodología bastante equitativa todo el tiempo, la estimación del algoritmo no depende de la disponibilidad de pruebas o de instalaciones de prueba sobrecargadas, entre otras. Por supuesto, el sistema tiene ligeros sesgos que pueden provenir, por ejemplo, de diferentes señales de fondo debido a las diferentes características de los distritos, o al comportamiento social estacional a medida que pasan los meses. Algunos de estos sesgos pueden resolverse reajustando el modelo de vez en cuando, otros ajustando diferentes modelos en diferentes regiones. Lo importante es que el algoritmo proporciona información sobre las llamadas de *fondo* que varían en el espacio y el tiempo.

Se podrían hacer más estudios para comprender y extraer las propiedades del *fondo*, como puede ser su estacionalidad, las variaciones según las regiones, los anuncios públicos o las noticias, etc.

El punto crucial del modelo matemático, es que reconoce las anomalías debidas a comportamientos colectivos. Por lo tanto, se encuentra que el modelo matemático y los algoritmos de Alarma de brotes tempranos pueden ser útiles para muchas otras enfermedades epidemiológicas, como por ejemplo el Dengue, y otros eventos, como las catástrofes naturales, entre otros. Actualmente, se está trabajando en la mejora de este sistema en muchos aspectos, incluyendo también los algoritmos de aprendizaje automático, avances que serán publicados en un trabajo futuro.

---

## 5. Conclusiones

---

En el presente trabajo, se muestra la creación de un algoritmo de vigilancia sindrómica basado en la correlación entre las llamadas telefónicas a una línea COVID, la población de los distritos y los casos reportados. Este algoritmo funciona entendiendo que las llamadas telefónicas a una línea COVID son una parte debido a que las personas no infectadas tienen síntomas similares (*fondo*) y otra parte debido a las personas infectadas (*señal*). Observando que el fondo tiene que ser proporcional a la población del distrito, mientras que la señal es proporcional a los casos reportados, hemos ajustado nuestra suposición. El coeficiente de determinación para la Provincia de Buenos Aires (PBA) es siempre  $R^2 > 0,85$  para las diferentes muestras, lo que indica la solidez de nuestra hipótesis. Además, hemos validado nuestro modelo con datos reales.

A lo largo del manuscrito se describió el modelo, sus estimaciones y cómo se calcularon sus barras de error. También se ha mostrado cómo las estimaciones, que se obtienen en tiempo real, pueden ser utilizadas para abordar las políticas de Salud Pública sin necesidad de esperar los resultados de laboratorio, que requieren muchos más días para converger. El algoritmo funcionó en la PBA de abril a junio, ya que durante este tiempo el centro de llamadas entrenado por COVID no se sobrecargó. Por lo tanto, la estimación fue relativamente imparcial.

Se muestra cómo esta estimación puede ser usada para crear una Alarma de brotes tempranos. Además, se describe cómo la construcción de indicadores que tienen que ver con los casos diarios, y el incremento diario de casos, puede indicar brotes con antelación. La variable estadística relevante en este caso es la significación, ya que es una medida real de cuán lejos de cero están los

indicadores. Es importante destacar que este sistema puede detectar un brote y, en particular, ejemplificamos su aplicación en la detección de brotes en Villa Azul, Quilmes.

Las limitaciones de la Alarma de brotes temprano fueron discutidas en este documento, muchas de ellas debido a las características de los datos disponibles en el momento de su desarrollo (urgente). Se han señalado muchas maneras de mejorar su sensibilidad y precisión, en las que se está trabajando actualmente. Esta alarma también sería útil, no sólo para otras enfermedades epidemiológicas, sino también para eventos que produzcan cambios en el comportamiento colectivo, como la epidemia de Dengue, las catástrofes naturales u otras.

El algoritmo y el modelo matemático presentados han sido una de las principales herramientas en el tablero del sistema PBA Health Care durante la epidemia, y sus versiones actuales y de actualización siguen siendo utilizadas para rastrear la epidemia y detectar brotes.

**Agradecimientos:** Agradecemos el fantástico trabajo realizado por el centro de llamadas COVID 148, en particular, a R. Vaena, P. Rispoli y L. H. Molinari por útiles conversaciones. E.A. y F.M. agradecen a la Dra. I. Caridi por las discusiones. E.A. agradece a CONICET, UNSAM, CAF y Easytech-Green por el apoyo financiero y logístico durante esta investigación.

**Conflictos de interés:** Los autores declaran no tener ningún conflicto de interés.

## Referencias

- [1] O. Ahmed. Mitigating the spread of covid-19 in low-income countries. *Scientific Research and Essays*, 15(2):41–44, 2020.
- [2] E. Alvarez y F. Marsico. Covid-19 mild cases determination from correlating covid-line calls to reported cases. *arXiv preprint arXiv:2008.01176*, 2020.
- [3] Buenos Aires Times. Villa in Buenos Aires Provinc sealed off amid surge in COVID-19 cases. mayo 2020. Disponible en <https://www.batimes.com.ar/news/argentina/villa-in-buenos-aires-province-isolated-due-to-covid-19.phtml>.

- [4] D. Chen, J. Cunningham, K. Moore, y J. Tian. Spatial and temporal aberration detection methods for disease outbreaks in syndromic surveillance systems. *Annals of GIS*, 17(4):211–220, 2011.
- [5] V. Diwan, D. Agnihotri, y A. Hulth. Collecting syndromic surveillance data by mobile phone in rural india: implementation and feasibility. *Global health action*, 8(1):26608, 2015.
- [6] W. Finch y M. Hernández Finch. Poverty and covid-19: rates of incidence and deaths in the united states during the first 10 weeks of the pandemic. *Frontiers in Sociology*, 5:47, 2020.
- [7] M. Gao, L. Yang, X. Chen, Y. Deng, S. Yang, H. Xu, y X. Gao. A study on infectivity of asymptomatic sars-cov-2 carriers. *Respiratory Medicine*, 106026, 2020.
- [8] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, y L. Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.
- [9] K. Goniewicz y F. M. Burkle. Disaster early warning systems: the potential role and limitations of emerging text and data messaging mitigation capabilities. *Disaster medicine and public health preparedness*, 13(4):709–712, 2019.
- [10] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675, 2020.
- [11] K. Hope, D. Durrheim, E. d’Espaignet, y C. Dalton. Syndromic surveillance: is it a useful tool for local outbreak detection? *Journal of Epidemiology & Community Health*, página 374–374, 2006.
- [12] R. Katz, L. May, J. Baker, y E. Test. Redefining syndromic surveillance. *Journal of epidemiology and global health*, 1(1):21–31, 2011.
- [13] G. Krause, D. Altmann, D. Faensen, K. Porten, J. Benzler, T. Pfoch, y H. Claus. Survnet electronic surveillance system for infectious disease outbreaks, germany. *Emerging infectious diseases*, 13(10), 2007.
- [14] U. Kyriacos, J. Jelsma, M. James, y S. Jordan. Monitoring vital signs: development of a modified early warning scoring (mews) system for general wards in a developing country, 2014.

- [15] H. Lee, K. Kim, K. Choi, S. Hong, H. Son, y S. Ryu. Incubation period of the coronavirus disease 2019 (covid-19) in busan, south korea. *Journal of Infection and Chemotherapy*, 26(9):1011–1013, 2020.
- [16] J. Lombardo, H. Burkom, E. Elbert, S. Magruder, S. H. Lewis, W. Loschen, J. Sari, C. Sniegowski, R. Wojcik, y J. Pavlin. A systems overview of the electronic surveillance system for the early notification of community-based epidemics (ESSENCE II). *Journal of urban health*, 80(1):i32–i42, 2003.
- [17] Ministerio de Salud. Detectar - dispositivo estratégico de testeo para coronavirus en territorio argentino. *Argentina.gob.ar*, 2020. Disponible en <https://www.argentina.gob.ar/coronavirus/detectar>.
- [18] J. Pavlin. Investigation of disease outbreaks detected by “syndromic” surveillance systems. *Journal of Urban Health*, 80(1):107– 114, 2003.
- [19] M. Stoto, M. Schonlau, y L. Mariano. Syndromic surveillance: is it worth the effort? *Chance*, 17(1):19–24, 2004.
- [20] G. Texier, L. Pellegrin, C. Vignal, J.-B. Meynard, X. Deparis, y H. Chaudet. Dealing with uncertainty when using a surveillance system. *International Journal of Medical Informatics*, 104:65–73, 2017.
- [21] L. Verhagen, R. Groot, C. Lawrence, J. Taljaard, M. Cotton, y H. Rabie. Covid-19 response in low-and middle-income countries: Don’t overlook the role of mobile phone communication. *International Journal of Infectious Diseases*, 99:334–337, 2020.

