

Spatial modeling of confirmed cases of COVID-19 in Chile

Nicolás Hamann, BSc¹, and Carola Blazquez, PhD¹

¹Universidad Andres Bello, Viña del Mar, Chile, n.hamannoyarce@uandresbello.edu, cblazquez@unab.cl

Abstract- This study utilizes global and local spatial modeling techniques to comprehend the factors that may contribute to the expansion of the SARS-CoV-2 virus at the commune level in Chile. The results show that the global regression model (Ordinary Least Squares) presents a poor fit with an adjusted R^2 of 0.33. However, the local modeling regressions (Multiscale Geographically Weighted Regression, MGWR) obtain a better fit in the explanation of the factors that affect the incidence rate of the disease COVID-19 with a R^2 of 0.90 and presents a AICc lower than the models obtained by the Geographically Weighted Regression. Therefore, MGWR significantly improves the performance in comparison with the common global regression model. The mobility indices, meteorological factors such as wind speed and precipitation, and the environmental pollution variables proved to be related with the incidence rate of COVID-19. The use of local regression models yields relevant information to health professionals and legislators to help effectively control this virus and other similar viruses with similar spread characteristics in Chile.

Keywords- Pandemic, spatial regression models, COVID-19, Chile

Modelamiento espacial de casos confirmados por COVID-19 en Chile

Nicolás Hamann, BSc¹, and Carola Blazquez, PhD¹

¹Universidad Andres Bello, Viña del Mar, Chile, n.hamannoyarce@uandresbello.edu, cblazquez@unab.cl

Resumen— Este estudio utiliza técnicas de modelamiento espacial a nivel local y global para entender los factores que podrían contribuir a la expansión del virus SARS-CoV-2 a nivel comunal en Chile. Los resultados muestran que el modelo de regresión global (Mínimos cuadrados ordinarios) presenta un ajuste pobre con un R^2 ajustado = 0,33. Sin embargo, los modelos de regresión local (regresión ponderada geográficamente) obtienen el mejor ajuste en la explicación de los factores que afectan la tasa de incidencia de la enfermedad COVID-19 con un $R^2 = 0,90$ y además presenta un AICc más bajo que los modelos obtenidos por regresión espacialmente ponderada. Por lo tanto, regresión ponderada geográficamente multiescala mejora bastante el rendimiento en comparación de un modelo de regresión global común. Los índices de movilidad, factores meteorológicos tales como la velocidad del viento y precipitación, y las variables de contaminación ambiental probaron estar relacionadas con la tasa de incidencia COVID-19. El uso de los modelos de regresión locales puede entregar información relevante a profesionales de la salud y legisladores para ayudar a controlar este virus de manera más efectiva y otras enfermedades con similares características de contagio en Chile.

Palabras Claves— pandemia, modelos de regresión espacial, COVID-19, Chile.

I. INTRODUCCIÓN

COVID-19 es una enfermedad respiratoria altamente contagiosa causada por el síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2) [1]. Este virus no es el primer coronavirus que la humanidad ha tenido que enfrentar. Existen más de siete coronavirus humanos y la diferencia es que este virus es mucho más infeccioso. A pesar de tener una mortalidad menor y tener síntomas menos graves, su forma de contagio y la capacidad de contagiar a otros humanos, incluso sin presentar síntomas, transformó este virus en una pandemia mundial, causando estragos en la salud y economía global [2].

El primer brote del virus SARS-CoV-2 se registró el 17 de noviembre de 2019 en el centro de Wuhan, China. La Organización Mundial de la Salud (OMS) declaró una pandemia por este virus sólo tres meses después del primer caso registrado. Este virus se extendió por el mundo llegando a Chile el 3 de marzo de 2020 con el primer caso confirmado tras testear positivo con la enfermedad COVID-19. En Chile, hasta el 29 de abril de 2021 habían ocurrido 1.418.062 casos de COVID-19 (1.198.245 con confirmación de laboratorio y 219.817 probables, sin confirmación de laboratorio) con una tasa de 7.287,7 por 100.000 habitantes. Según casos confirmados por laboratorio, las mayores tasas de incidencia acumulada por 100.000 habitantes se concentraron en la Región de Magallanes

(15.670,4), Región de Tarapacá (10.722,5) y Región de Los Ríos (9.187,5) [3].

Dado que COVID-19 es una enfermedad infecciosa grave y potencialmente letal es fundamental implementar la vigilancia espacial que puede ayudar a priorizar zonas que necesiten intervenciones específicas, pruebas rápidas y asignación de recursos. En el presente estudio, se determinó la asociación entre la tasa de incidencia de COVID-19 en Chile a nivel comunal y diferentes factores tales como la movilidad de la población, los datos meteorológicos y de calidad del aire, y las condiciones socioeconómicas. Para ello, se utilizaron modelos de regresiones espaciales a nivel global (Mínimos cuadrados ordinarios, OLS por sus siglas en inglés) y local (Regresión ponderada geográficamente, GWR por sus siglas en inglés, y regresión ponderada geográficamente multiescala, MGWR por sus siglas en inglés) para así identificar aquellas variables que estén significativamente relacionadas con la tasa de incidencia de COVID-19. De esta manera, se puede comprender el comportamiento de este virus, anticipar los brotes mundiales y a su vez mejorar las políticas de salud pública en Chile. Los resultados de este estudio pueden ser un aporte relevante a profesionales de la salud pública y legisladores en la toma de decisiones preventivas para controlar esta enfermedad viral.

II. REVISIÓN BIBLIOGRÁFICA

Diversos estudios han utilizado técnicas de regresión espacial para entender la relación que existe entre los casos positivos y muertes por COVID-19 y diferentes factores desde una perspectiva espacial [4-16]. La Tabla I muestra una lista de los estudios más relevantes que fueron realizados en diversos países sobre el contagio de COVID-19. Cabe destacar que no se tiene conocimiento de un estudio similar que se haya efectuado en Chile.

III. DATOS

A. Datos sobre la tasa de incidencia de COVID-19

Para obtener datos sobre COVID-19 a nivel de comuna en Chile, se utilizaron los datos oficiales del Ministerio de Ciencias, Tecnología, Conocimiento e Innovación reportados en su repositorio web oficial (<https://github.com/MinCiencia/Datos-COVID19>). El presente estudio se limita al periodo que corresponde a la primera ola de contagios COVID-19 en Chile desde el 30 de marzo del 2020 al 14 de septiembre del 2020, como se observa en la Fig. 1. Esta figura muestra la tasa de incidencia COVID-19 que es usada en epidemiología para reflejar la expansión de la enfermedad en

una población y en un tiempo determinado. Esta tasa se calcula con (1).

$$y_i = \frac{(\sum_{j=1}^n x_{ij})}{P_i} \cdot 100,000 \quad (1)$$

Donde:

y_i : Tasa de incidencia en comuna i .

x_{ij} : Casos confirmados COVID-19 en comuna i en el día j .

P_i : Población en comuna i .

n : número de días en el periodo estudiado

TABLA I

ESTUDIOS SOBRE COVID-19 USANDO MODELOS DE REGRESIÓN ESPACIAL

Autores y año	País	Breve Descripción
Dutta et al., 2021 [4]	India	Usaron modelamiento espacial para identificar los factores que inciden en la tasa de incidencia de COVID-19.
Harris, 2020 [5]	Londres	Exploraron las variaciones espaciales en los coeficientes de regresión de las muertes por COVID-19 entre diferentes vecindarios de Londres.
Karaye & Horney, 2020 [6]	EEUU	Usaron modelos de regresión locales y globales para examinar la relación entre el número de casos de COVID-19 y vulnerabilidad social.
Li et al., 2021 [7]	China	Mediante un modelo mixto de GWR, se obtuvo la asociación de los atributos del entorno construido y COVID-19 usando efectos globales y locales.
Liu et al., 2020 [8]	China	Utilizando GWR se identifican múltiples factores asociados al contagio de COVID-19 como la temperatura.
Maiti et al., 2021 [9]	EEUU	Estudiaron los efectos espaciotemporales de diversos factores en la incidencia de COVID-19 a nivel local y global.
Midday & Roy, 2021 [10]	India	Estudiaron la relación entre la mortalidad por COVID-19 y diferentes factores usando métodos de regresión local y global (GWR y OLS).
Mollalo et al., 2020 [11]	EEUU	Investigaron la dependencia espacial con modelos de error y retardo espacial y examina no estacionariedad espacial con GWR y MGWR.
Nor et al., 2020 [12]	Malaysia	Demonstraron la existencia de un vínculo entre contaminantes ambientales como el MP2,5 y el contagio del virus COVID-19.
Sannigrahi et al., 2020 [13]	Europa	Usando modelos de regresión espacial, se investigó la asociación entre las variables socio-demográficas y los casos y muertes por COVID-19.
Urban et al. 2021 [14]	Brasil	Demonstraron que el modelo GWR es el que mejor explica la distribución espacial de COVID-19 en la ciudad de Sao Paulo.
Wang et al., 2020 [15]	India	GWR y MGWR se implementaron encontrando una posible relación entre la densidad de población, urbanización y facilidades bancarias y la tasa de incidencia COVID-19.
Xie and Zhu, 2020 [16]	China	Se investigó sobre la posible relación entre la tasa de incidencia COVID-19

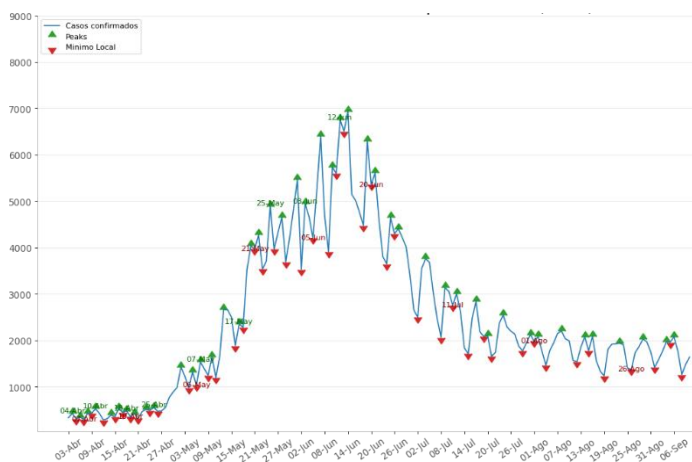


Fig. 1 Número de casos confirmados en Chile durante el periodo estudiado. Fuente: Elaboración propia.

B. Datos Socioeconómicos

El Índice de Desarrollo Comunal (IDC) es un índice que mide distintos aspectos de la realidad a nivel comunal en Chile. Este índice es el producto de un estudio realizado en el 2020 por el Instituto Chileno de Estudios Municipales y el Instituto de Estudios del Hábitat de la Universidad Autónoma de Chile [17]. Para crear este índice, se utilizaron tres dimensiones: Salud y bienestar social, economía y recursos y educación. IDC entrega valores cercanos a uno cuando la comuna posee un mayor desempeño en las tres dimensiones y cercanos a cero cuando tiene un peor desempeño. De esta manera, se puede realizar una comparación de la realidad nacional a nivel comunal.

C. Datos sobre Contaminación Ambiental

Los datos sobre contaminación ambiental fueron proporcionados por el Registro de Emisiones y Transferencias de Contaminantes (RETC) del Ministerio del Medio Ambiente de su página web (<https://retc.mma.gob.cl/>). RETC reporta las emisiones al aire asociadas a la combustión de leña residencial, incendios forestales, incendios urbanos y quemas agrícolas como también las emisiones de transporte en ruta para todas las comunas a lo largo de Chile. Entre los tipos de contaminantes reportados se incluye el material particulado 2.5 (MP2,5), óxidos de nitrógeno (NOx) y dióxido de azufre (SO2).

D. Datos Meteorológicos

Los datos meteorológicos fueron recopilados de las bases de datos compiladas por el Centro de Ciencia del Clima y Resiliencia (CR)2. Estos datos corresponden a series de tiempo espacialmente distribuidas y registradas a través de instrumentos in situ o remotos que pueden ser descargados de su página web (<https://www.cr2.cl/acerca-del-cr2/>).

E. Datos de Movilidad

Los datos de movilidad fueron descargados del repositorio COVID-19 del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. Estos datos consideran el movimiento de los teléfonos móviles conectados a la red de Telefónica en el territorio nacional, de manera agrupada y

anónima. Se utilizaron las siguientes tres series de tiempo en este estudio:

1) *Interno (IM interno)*: Esta serie de tiempo representa la evolución del índice de movilidad interno en la comuna y es una medida de los viajes que ocurren al interior de dicha unidad administrativa.

2) *Externo (IM externo)*: Esta serie de tiempo representa la evolución del índice de movilidad externo a la comuna y es una medida de los viajes que tienen origen al interior de la comuna y destino al exterior de la comuna, y además de los viajes que tienen origen al exterior de la comuna y destino al interior de la comuna.

3) *Suma (IM suma)*: Esta serie de tiempo representa la evolución del índice de movilidad que es la suma de los dos índices anteriores ($IM = IM\ interno + IM\ externo$).

La Tabla II muestra estadísticas básicas descriptivas de las variables utilizadas en este estudio, la cual está compuesta de 32 variables con datos para las 346 comunas de Chile. Todas las variables fueron recopiladas en una estructura de datos del módulo Pandas de Python llamada dataframe.

IV. METODOLOGÍA

A. Identificación de variables estadísticamente relevantes

Se utilizaron dos criterios principales para la elección de variables significativas. Primero, se consideraron las variables que en la literatura se ha comprobado su relación con la tasa de incidencia COVID-19. Segundo, se seleccionaron las variables esenciales utilizando del método de reducción de dimensionalidad (LASSO) que elimina las variables estadísticamente insignificantes o redundantes minimizando la suma cuadrada residual (valor actual - valor predicho, también como norma L1) [18]. LASSO descarta las variables redundantes que tienen coeficientes iguales a cero y selecciona aquellas que poseen coeficientes superiores a cero.

B. Análisis de regresión

Como parte del análisis de regresión, primero, se realizó un análisis de exploración usando una regresión global que investiga las relaciones entre un conjunto de variables independientes y la variable dependiente minimizando la suma de los cuadrados de los residuales (OLS, por sus siglas en inglés, Ordinary Least Square). OLS se usó para tener una referencia de comparación con los otros modelos usados en este estudio, y así verificar si explican estadísticamente de mejor manera las relaciones entre la tasa de incidencia COVID-19 y sus potenciales factores asociados. OLS está definida por (2) [19, 20].

$$y_i = \beta_0 + \sum_{k=1}^n \beta_k x_{ik} + \delta_i \quad (2)$$

donde y_i representa la estimación de la tasa de incidencia COVID-19 en la comuna i , x_{ik} es la observación i de la variable k , β_k es el coeficiente de regresión global para la variable k , β_0 es el intercepto y δ_i es el error.

TABLA II
ESTADÍSTICA DESCRIPTIVA DE LAS VARIABLES

Variables	Media	Desviación Estándar	Mínimo	Máximo
COVID-19				
Tasa incidencia (1/100,000)	1671,31	1413,63	0	8362,37
Socioeconómicos				
Índice de bienestar	0,763	0,149	0,19	0,99
Índice de económico	0,123	0,096	0,02	0,78
Índice de educación	0,650	0,092	0,00	0,90
Índice de desarrollo comunal (IDC)	0,379	0,119	0,03	0,87
Ranking nacional IDC	173,89	99,88	1,0	345,00
Contaminación Ambiental				
Arsénico	3,39e-05	3,72e-05	0	2,72e-4
Carbono negro	30,98	65,96	0	670,35
Compuestos orgánicos volátiles	2973,14	8035,81	0	89868,10
Dióxido de azufre (SO ₂)	12,85	43,37	0	561,12
Dióxido de carbono (CO ₂)	36122,30	74544,00	0	684637,00
Material particulado 10 (MP10)	362,78	770,75	0	6926,50
Material particulado 2,5 (MP2,5)	340,14	720,29	0	6733,12
Material particulado	430,06	952,82	0	9531,30
Mercurio	2,14e-05	2,34e-05	0	1,71e-04
Metano (CH ₄)	293,84	680,63	0	7450,01
Monóxido de carbono (CO)	6157,46	14415,70	0	156173,00
Óxidos de nitrógeno (NO _x)	77,99	174,24	0	1843,72
Nitrógeno amoniacal (NH ₃)	27,18	60,11	0	587,12
Óxido nitroso (N ₂ O)	3,54	7,54	0	68,50
Plomo	1,06e-05	1,16e-05	0	8,48e-05
Meteorológicos				
Velocidad del viento (m/s)	2,87	1,31	0,088	8,70
Temperatura media (°C)	14,19	2,97	0	21,34
Precipitación (mm)	1,24	1,49	0	7,52
Movilidad				
IM interno	3,07	2,35	0	10,45
IM externo	4,62	4,13	0	39,86
IM Suma	7,69	4,65	0	42,84

OLS asume que las observaciones son independientes y constantes en el área de estudio (i.e., estacionariedad espacial), y además se espera que los términos relacionados con el error no son correlacionados [19]. En cambio, la regresión ponderada geográficamente (GWR, por sus siglas en inglés, Geographically Weighted Regression) permite que los parámetros varíen espacialmente, es decir, las variables son no estacionarias y modela las relaciones locales entre estos predictores y la variable dependiente o de interés. La ventaja de GWR es que calcula los coeficientes de la regresión dependiendo de su locación geográfica y entrega información local sobre la relación entre la variable de interés y algún factor en un punto geográfico específico [10]. GWR extiende la regresión global definida en (2) e introduce estimadores locales

para agregar un factor geográfico a la regresión expresado por (3) [21].

$$y_i = \xi_{i0} + \sum_{k=1}^n \xi_k(\mu_i, v_i)x_{ik} + \delta_i \quad (3)$$

donde los estimadores locales son representados por (μ_i, v_i) que es una matriz que representa la ubicación de la comuna i ; x_{ik} es la observación i de la variable k , $\xi_k(\mu_i, v_i)$ representa el coeficiente local para el predictor k en la comuna i en la ubicación (μ_i, v_i) .

GWR restringe las relaciones locales dentro de cada modelo para que varíen en la misma escala espacial, es decir, los datos son ponderados en una misma escala. En cambio, la regresión ponderada geográficamente multiescala (MGWR, por su sigla en inglés, Multiscale Geographically Weighted Regression) permite que las relaciones condicionales entre la variable dependiente y variables independientes varíen en diferentes escalas espaciales, es decir, se captura la heterogeneidad espacial dentro y entre procesos espaciales, minimiza el sobreajuste, reduce la colinealidad y el sesgo en las estimaciones de los parámetros [22, 23].

En este estudio, se compararon los rendimientos de los modelos usando las siguientes métricas: Criterio de Información de Akaike corregido (AICc), R^2 y R^2 ajustado.

V. RESULTADOS

A. Reducción de Dimensionalidad LASSO

LASSO ayuda a resolver el problema de la multicolinealidad de las variables que existe en la regresión espacialmente ponderada GWR. Como resultado de la implementación de LASSO, se seleccionaron las siguientes variables por cada tipo de datos:

- *Socioeconómicos*: Bienestar, Economía, Educación, Ranking Nacional IDC.
- *Meteorológicos*: Velocidad del Viento, Temperatura Media, Precipitación.
- *Contaminación ambiental*: MP2,5, SO2, NOx.
- *Movilidad*: IM interno y IM suma.

B. Análisis de Regresión

La Tabla III muestra los resultados de la regresión OLS, cuyo rendimiento es de $R^2 = 0,35$ y R^2 ajustado = 0,33. Esta tabla indica que las variables velocidad del viento, temperatura media, precipitaciones, índice de educación y ranking nacional son estadísticamente significativas con un $p < 0,05$.

En las Tabla IV y V, se aprecian los resultados de los modelos de regresión locales GWR y MGWR, respectivamente, para seis modelos generados con diferentes combinaciones de variables. Estas tablas presentan la media de los coeficientes de las regresiones para cada variable dependiente analizada.

TABLA III
RESULTADOS DE LA REGRESIÓN GLOBAL OLS

Variable	Estimado	Prueba t	Valor p
Vel. viento	-0,209	-3,307	0,001
Temp. media	0,259	3,995	0,000
Precip.	-0,167	-2,547	0,011
MP 2,5	0,845	1,783	0,075
SO2	0,768	1,792	0,073
NOx	-1,563	-1,941	0,052
IM Interno	-0,106	-1,610	0,107
IM Suma	-0,059	-1,045	0,296
Bienestar	-0,035	-0,409	0,683
Economía	0,057	0,719	0,472
Educación	-0,225	-3,181	0,001
IDC	-0,401	-3,368	0,001

TABLA IV
RESULTADOS DE LA REGRESIÓN LOCAL GWR

Variable	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
Vel. viento		-0,26	-0,13			-0,15
Temp. med.		-0,21	0,10			0,13
Precip.		0,21	0,69			0,86
MP 2,5			0,14		0,95	2,85
SO2						-7,43
NOx						6,06
IM Interno	-0,06		-0,06		-0,12	0,03
IM Suma	-0,03		0,00		0,02	-0,03
Bienestar	0,27			0,29		0,14
Economía	0,10			0,09		-0,13
Educación	-0,16			-0,16		-0,20
IDC	-0,03			0,027		-0,20
R^2	0,81	0,83	0,82	0,79	0,84	0,82
R^2 ajustado	0,75	0,78	0,77	0,74	0,80	0,76
AIC	577,2	513,2	534,1	578,8	489,7	568,1
AICc	632,38	553,0	570,9	619,2	534,3	629,1

TABLA V
RESULTADOS DE LA REGRESIÓN LOCAL MGWR

Variable	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
Vel. viento		-0,08	0,05			0,02
Temp. med.		-0,10	0,11			0,02
Precip.		0,063	0,62			0,34
MP 2,5			0,01		0,51	-0,07
SO2						-2,56
NOx						1,99
IM Interno	0,02		-0,08		-0,12	0,02
IM Suma	-0,05		-0,03		-0,04	-0,01
Bienestar	-0,07			-0,05		0,00
Economía	-0,04			-0,08		-0,05
Educación	-0,14			-0,11		-0,18
IDC	-0,05			-0,06		-0,06
R^2	0,88	0,82	0,84	0,87	0,84	0,90
R^2 ajustado	0,84	0,78	0,81	0,82	0,80	0,86
AIC	427,8	500,3	464,8	460,9	476,8	414,1
AICc	502,4	522,4	493,2	522,2	506,2	549,2

Al comparar los indicadores de desempeño mostrados en las Tablas III, IV y V, se puede concluir lo siguiente: primero, los modelos GWR y MGWR presentaron mejores resultados que el modelo de regresión global OLS; y segundo, los resultados del modelo 6 con la regresión local MGWR presenta indicadores R^2 y R^2 ajustado más altos y AIC y AICc más bajo

superando los resultados con la regresión local GWR. Por lo tanto, a continuación, se muestran los resultados con más detalle para dicho modelo. La Tabla VI muestra las estadísticas básicas de los coeficientes del modelo 6 usando MGWR ya que cada comuna posee su propio coeficiente a diferencia del OLS que es constante en el espacio.

La Fig. 2 muestra los coeficientes locales de regresión para cada variable del modelo 6 usando MGWR. Mientras más rojo (azul) la asociación de la variable explicativa con la tasa de incidencia COVID-19 es mayor (menor). A partir de esta figura, se dedujo que algunos factores como el índice de movilidad interno están positivamente relacionados en el norte y sur de Chile, pues un mayor movimiento dentro de las comunas obviamente incrementa la probabilidad de contagios debido a la naturaleza de este virus. Los índices socioeconómicos (particularmente índices de economía y educación) parecieran afectar positivamente la propagación de este virus en el sur de Chile. Esto se puede explicar debido a que zonas con menor acceso a educación y menor recursos económicos no tomen las precauciones adecuadas de prevención. Factores meteorológicos como la velocidad del viento demuestran tener una fuerte relación positiva con la tasa de incidencia de COVID-19 a lo largo de Chile, por lo tanto, el viento ayuda a propagar el virus. Las emisiones de MP2,5 y SO2 parecieran tener una incidencia positiva sólo en las comunas del norte de Chile.

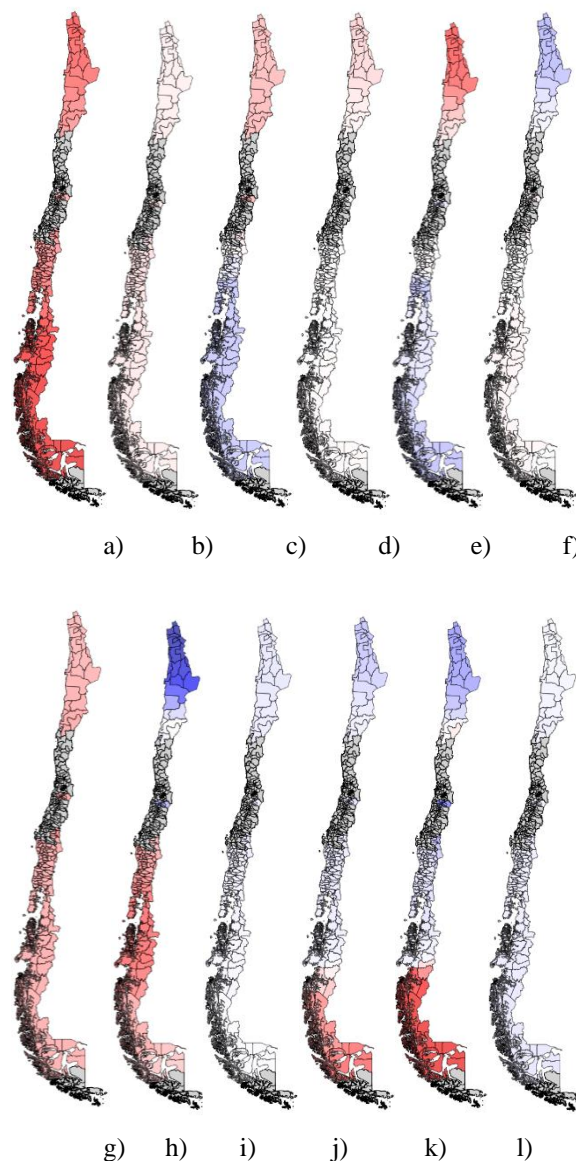


Fig. 2. Coeficientes para modelo 6 con MGWR para a) velocidad del viento, b) temperatura media, c) precipitación, d) MP2,5, e) SO2, f) NOx, g) IM interno, h) IM suma, i) bienestar, j) economía, k) educación, y l) IDC.

TABLA VI

ESTADÍSTICAS BÁSICAS DEL MODELO 6 USANDO MGWR

Variable	Media	Des. Std.	Mínimo	Mediana	Máximo
Vel. viento	0,05	0,02	0,04	0,04	0,11
Temp. med.	0,03	0,01	0,02	0,03	0,04
Precipitació	0,63	0,47	-0,02	0,96	1,08
MP 2,5	1,05	3,25	-1,96	0,33	14,22
SO2	1,34	6,86	-2,20	-0,26	30,15
NOx	-1,85	7,07	-31,81	-0,53	30,05
IM Interno	0,06	0,00	0,05	0,06	0,07
IM Suma	-0,04	0,13	-0,53	-0,05	0,13
Bienestar	-0,02	0,00	-0,03	-0,02	-0,00
Economía	-0,08	0,05	-0,20	-0,07	0,18
Educación	-0,18	0,14	-0,40	-0,13	0,07
IDC	-0,05	0,01	-0,08	-0,05	-0,03

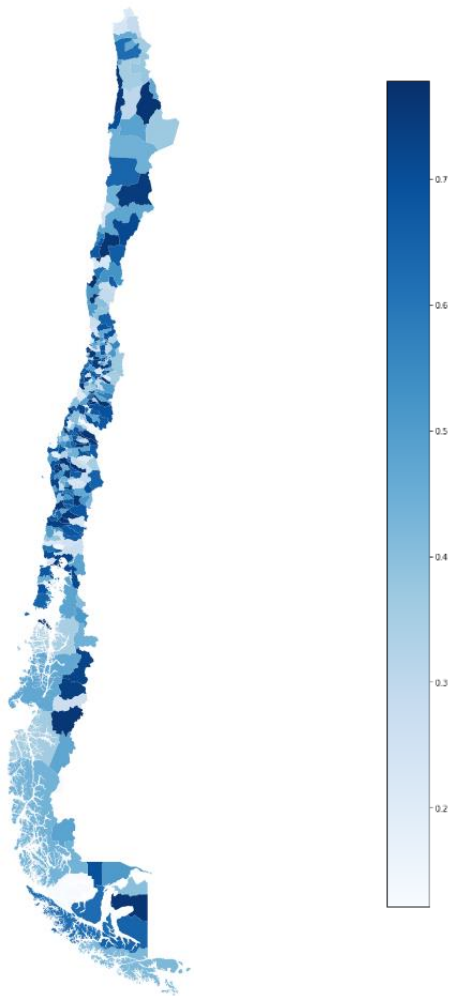


Fig. 3. Distribución del R^2 local con MGWR.

La Fig. 3 muestra la distribución espacial de los valores de R^2 local asociados a las variables independientes del modelo 6 usando MGWR. Aunque se aprecia de la figura que no existe un patrón general en la distribución a nivel local, existen diversas comunas a lo largo de Chile que tienen un valor de R^2 superior a 0,7 lo que indica un buen ajuste del modelo en esas comunas.

VI. CONCLUSIONES

En el presente estudio, se utilizaron técnicas de regresión espacial a nivel local y global para determinar el impacto de diferentes factores en esta tasa de incidencia de COVID-19 a nivel comunal en Chile. Los resultados indican que las regresiones locales GWR y MGWR presentan mejor ajuste del modelo con las variables utilizadas que la regresión espacial global OLS. Esto se debe a que los coeficientes estimados son constantes para todas comunas en Chile con OLS. En cambio, GWR y MGWR tienen la propiedad de no-estacionariedad espacial en estas estimaciones y, por ende, obtiene un coeficiente diferente para cada variable por comuna. Además,

los resultados de los modelos con MGWR presentan mejores estimaciones que con GWR ya que se incluyen diferentes escalas espaciales en su modelamiento y captura la heterogeneidad entre las variables.

Los resultados del presente estudio identifican los factores que pueden influir en la incidencia del COVID-19. Por ejemplo, la movilidad dentro de las comunas pareciera tener una fuerte correlación, lo que tiene sentido debido a la manera que este virus se transmite persona a persona. Los factores meteorológicos como la velocidad del viento y precipitaciones también parecieran estar relacionadas al incremento de esta enfermedad. Por lo tanto, se deberían tomar precauciones particularmente en la época de invierno y en las comunas con malas condiciones climáticas durante el año completo. Además, se encontró una posible relación con las emisiones de $MP_{2,5}$ y SO_2 y la incidencia de COVID-19 ya que existe la posibilidad de este virus pueda alojarse en las partículas de estas emisiones.

Como futura investigación, se utilizarán otras variables adicionales para estudiar su impacto en la tasa de incidencia de COVID-19. Además, sería interesante realizar el modelamiento espacial en unidades espaciales más pequeñas en zonas urbanas o altamente pobladas para identificar con más detalle las variables significativas.

BIBLIOGRAFÍA

- [1] Organización Panamericana de la Salud, OPS. "Coronavirus." [Online] <https://www.paho.org/es/temas/coronavirus>
- [2] Mingarro, I. "Los 7 tipos de coronavirus que infectan humanos." National Geographic España. 2022. [Online] https://www.nationalgeographic.com.es/ciencia/siete-tipos-coronavirus-que-infectan-humanos_15353
- [3] Ministerio de Salud, MINSAL. "Ministerio de Salud confirma primer caso de coronavirus en Chile." 2020. [Online] <https://www.minsal.cl/salud-confirma-primero-caso-de-coronavirus-en-chile/>
- [4] I. Dutta, T. Basu, and A. Das. "Spatial analysis of COVID-19 incidence and its determinants using spatial modeling: A study on India," *Environ Chall*, vol. 4, 100096, 2021.
- [5] R. Harris, "Exploring the neighbourhood-level correlates of Covid-19 deaths in London using a difference across spatial boundaries method," *Health place*, vol. 66, 102446, 2020.
- [6] I. Karaye, and J. Horney, "The impact of social vulnerability on COVID-19 in the US: an analysis of spatially varying relationships," *Am J Prev Med*, vol. 59, no. 3, pp. 317-325, 2020.
- [7] S. Li, S. Ma, and J. Zhang, "Association of built environment attributes with the spread of COVID-19 at its initial stage in China," *Sustain Cities Soc*, vol. 67, 102752, 2021.
- [8] M. Liu, M. Liu, Z. Li, Y. Zhu, Y. Liu, X. Wang, L. Tao, and X. Guo, "The spatial clustering analysis of COVID-19 and its associated factors in mainland China at the prefecture level," *Sci Total Environ*, vol. 777, 145992, 2021.
- [9] A. Maiti, Q. Zhang, S. Sannigrahi, S. Pramanik, S. Chakraborti, A. Cerda, and F. Pilla, "Exploring spatiotemporal effects of the driving factors on COVID-19 incidences in the contiguous United States," *Sustain Cities Soc*, vol. 68, 102784, 2021.
- [10] A. Middy, and S. Roy, "Geographically varying relationships of COVID-19 mortality with different factors in India," *Sci Rep*, vol. 11, 7890, 2021.
- [11] A. Mollalo, B. Vahedi, and K. Rivera, "GIS-based spatial modeling of COVID-19 incidence rate in the continental United States," *Sci Total Environ*, 728, 138884, 2020.

- [12]N. Nor, C. Yip, N. Ibrahim, M. Jaafar, Z. Rashid, N. Mustafa, H. Hamid, K. Chandru, M. Latif, P. Saw, and C. Lin, "Particulate matter (PM2.5) as a potential SARS-CoV-2 carrier," *Sci Rep*, vol. 11, 2508, 2021.
- [13]S. Sannigrahi, F. Pilla, B. Basu, A. Basu, and A. Molter, "Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach," *Sustain Cities Soc*, vol. 62, 102418, 2020.
- [14]R. Urban, and L. Nakada, "GIS-based spatial modelling of COVID-19 death incidence in São Paulo, Brazil," *Environ Urban*, vol. 33, no. 1, pp. 229–238, 2021.
- [15]A. Wang, X. Zhong, and Y. Hurd. 2020. "Comorbidity and sociodemographic determinants in COVID-19 mortality in an US urban healthcare system. *MedRxiv*, 2020.
- [16]J. Xie, and Y. Zhu, "Association between ambient temperature and COVID-19 infection in 122 cities from China," *Sci Total Environ*, vol. 724, 138201, 2020.
- [17]J. Hernández, H. Ramírez, A. Parrao, L. Salazar, J. González, and C. Godoy, "Índice de Desarrollo Comunal. Chile 2020." [Online] https://repositorio.uaautonoma.cl/bitstream/handle/20.500.12728/6742/V11_digital_final.pdf?sequence=1
- [18]R. Tibshirani, "Regression Shrinkage and Selection via the lasso". *J Geogr Syst*, Series B, vol. 58, no.1, pp. 267–88, 1996.
- [19]L. Anselin, and D. Arribas-Bel, "Spatial fixed effects and spatial dependence in a single cross-section," *Pap Reg Sci*, vol. 92, no. 1, pp. 3-17, 2013.
- [20]M. Ward, and K. Gleditsch, *Spatial regression models*, Sage Publications, 2018.
- [21]A. Fotheringham, and T. Oshan, "Geographically weighted regression and multicollinearity: dispelling the myth," *J Geogr Syst*, vol. 18, no. 4, pp. 303–329, 2016.
- [22]A. Fotheringham, W. Yang, and W. Kang, "Multiscale Geographically Weighted Regression (MGWR)," *Ann Am Assoc Geogr*, vol. 107, no. 6, pp. 1247-1265, 2017.
- [23]T. Oshan, J. Smith, and A. Fotheringham, "Targeting the spatial context of obesity determinants via multiscale geographically weighted regression," *Int J Health Geogr*, vol. 19, no. 1, pp. 1-7, 2020.