

Comportamiento epidemiológico COVID-19 usando el método de Descomposición en Modos Dinámicos para la ciudad de Santiago de Cali.

Autor: Fabian David Navarro Cano.
fabilandav06@javerianacali.edu.co

Director: Abel Alvarez Bustos.
abel.alvarez@javerianacali.edu.co

Pontificia Universidad Javeriana Cali Colombia
Matemáticas Aplicadas
18 de Enero del 2022

Resumen

El SARS-CoV-2 es un virus que produce una enfermedad respiratoria conocida como coronavirus del 2019 (COVID-19). Este pertenece a la familia de los coronavirus, siendo un tipo de virus que infecta a seres humanos y algunos animales alrededor del mundo. De hecho, la infección por SARS-CoV-2 fue identificada por primera vez en personas expuestas a un mercado de mariscos en el sector de Wuhan, China, sobre el mes de noviembre del 2019, [9]. Desde entonces, ha sido una de las problemáticas de salud pública más grandes de la historia de la humanidad en los últimos siglos, debido a que este virus tiene una rápida tasa de transmisión y contagio. Es así como, la Organización Mundial de la Salud (OMS) emitió un estado de pandemia en cuestión de pocos meses, ocasionando grandes pérdidas económicas a nivel mundial. Dada la dimensión de esta problemática, se ha despertado el interés por parte de diferentes equipos científicos con el objetivo de identificar el comportamiento de la propagación, a fin de hallar una forma pertinente para su tratamiento.

En virtud de esta situación, en este proyecto de grado se presentará un método emergente impulsado por datos, denominado Descomposición en Modos Dinámicos (DMD), el cual permitirá conocer la dinámica epidemiológica de manera espacio-temporal sobre un periodo corto de tiempo [11, 4], utilizando los reportes generados dentro del área metropolitana de Santiago de Cali, Colombia, ver [1]. En este sentido, se emplearán las herramientas computacionales MATLAB y R para ejecutar las diferentes simulaciones que serán de utilidad para comprender el comportamiento de este virus; como de igual manera, se analizarán algunos conceptos complementarios y auxiliares que serán pertinentes para tener un adecuado manejo del método DMD, como son la Descomposición en Valores Singulares (SVD), el Análisis de Componentes Principales (PCA), el Análisis de Componentes Independientes (ICA) y finalmente el Análisis de Koopman.

Palabras Claves: *Descomposición en Modos Dinámicos, Coronavirus, DMD, COVID-19, SARS-CoV-2.*

1. Introducción

Los avances del ser humano se encuentran en interacción con el medio que los rodea, lo cual lleva a concebir diferentes estrategias para avanzar en ámbitos personales, sociales y económicos. Sin embargo, las acciones realizadas para alcanzar estos avances, de alguna manera acarrearán diversos problemas sociales y medioambientales, [7, 8]. De hecho, actualmente, una de las problemáticas en el entorno del ser humano ha ocasionado un virus que produce una grave enfermedad respiratoria a nivel mundial, llevan-

do a cada país a una crisis de salud y escasez en ámbitos socioeconómicos de grandes proporciones.

Esta afectación mundial está dada por el virus SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) siendo este, un tipo de virus que produce una enfermedad respiratoria denominada coronavirus del 2019 o COVID-19; esta enfermedad zoonótica (enfermedad que se transmite entre animales y seres humanos) fue identificada por primera vez en algunas personas en la ciudad de Wuhan, China, dentro de un

mercado de productos húmedos. Tal problema de salud tuvo su primer registro el día 17 de noviembre del 2019, lo que derivó en la actualidad (marzo del 2021) en un registro con casos confirmados de más de 125 millones de personas en todos los continentes. La gravedad de esta enfermedad fue tal que conllevó a la Organización Mundial de la Salud (OMS), a declarar el inicio de un estado de pandemia el día 11 de marzo del 2020, cuando presentó una tasa de mortalidad del 3% al 7% aproximadamente [14, 6, 2].

En este sentido, al declarar el estado de pandemia a raíz del virus, se presentó un incremento exponencial de contagio en diferentes sectores del mundo que obligó a sus gobernantes a decretar aislamientos preventivos y cierres de fronteras a fin de mitigar las tasas de contagio, no obstante, realizar este tipo de estrategias, generó efectos secundarios, llevando a la población a un estado de salud grave en conjunto con una gran parálisis de la economía mundial, la cual implicaría millonarias pérdidas monetarias y una crisis socioeconómica por cuestiones de la pandemia, [9]. Dada la dimensión de este problema, se despertó el interés por parte de diferentes equipos científicos de todo el mundo, apostando por una carrera en la identificación del comportamiento de la propagación del virus con el fin de hallar una forma efectiva de tratar esta enfermedad, ver [5].

En virtud de esta situación, en este proyecto de grado se presentará un método emergente impulsado por datos, denominado Descomposición en Modos Dinámicos (DMD), el cual permitirá conocer la dinámica epidemiológica sobre un corto periodo de tiempo [12], utilizando los reportes de Covid19 que fueron generados en la zona metropolitana de Santiago de Cali, ver [1].

2. Metodología y desarrollo.

En este orden de ideas, metodológicamente este proyecto de Descomposición en Modos Dinámicos para la ciudad de Santiago de Cali esta desarrollado bajo **Tres fases** empezando por la **extracción de datos** donde se elabora una captación de la información que fue suministrada por la secretaria de salud, esta información se adquirió mediante un arreglo matricial, tal como se ve en el siguiente esquema:

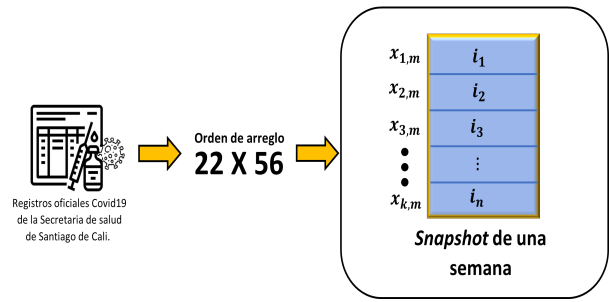


Figura 1: Arreglo matricial de los contagiados por Covid19 en sus primeras 56 semanas, donde cada semana es considerada como un *snapshot* o una pequeña representación de la enfermedad a una semana.

Ahora bien, ya con esta información comprendida se decide agrupar los *snapshots* de información, con la finalidad de obtener un esquema completo de semanas y sus comportamientos asociados tal como se ve a continuación:

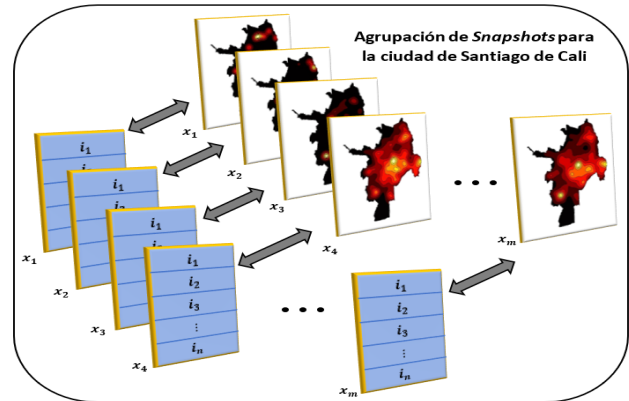


Figura 2: Conjunto total con las 56 *snapshot* que contiene el comportamiento de contagio según la semana y comuna correspondiente de Santiago de Cali.

Luego de esto, se tiene ahora la fase de **transformación del conjunto total** donde a partir de lo anterior se realiza una separación de la información para un óptimo funcionamiento del método DMD, ya que este modelo analiza la relación entre pares de datos (*Snapshots*) de manera espacio temporal, es decir, consiste en una medición futura \mathbf{X}_{k+1} y una medición previa \mathbf{X}_k donde $\mathbf{X} \in \mathbb{R}^n$. En la solución de estas mediciones, se considera como un mejor ajuste dado por un operador lineal matricial $\mathbf{A} \in \mathbb{R}^{n \times n}$, en relación con [3] y [11], teniendo que:

$$\mathbf{X}_{k+1} = \mathbf{A}\mathbf{X}_k, \quad (1)$$

pues bien, de acuerdo a la figura 2 y el anterior análisis se tiene el comportamiento lineal de los *snapshots* de la siguiente forma:

$$\begin{aligned} x_2 &= Ax_1 \\ x_3 &= Ax_2 \\ &\vdots \\ x_{k+1} &= Ax_k \Rightarrow \mathbf{X}_{k+1} = \mathbf{A}\mathbf{X}_k. \end{aligned}$$

De este modo, con estos estados temporales lineales anteriores, se puede construir según Proctor en [10] los siguientes arreglos matriciales para trabajar con el método DMD:

$$\mathbf{X} = \begin{bmatrix} | & | & | & \dots & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \dots & \mathbf{x}_{m-1} \\ | & | & | & \dots & | \end{bmatrix},$$

$$\mathbf{X}' = \begin{bmatrix} | & | & | & \dots & | \\ \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \dots & \mathbf{x}_m \\ | & | & | & \dots & | \end{bmatrix},$$

donde $\mathbf{X}, \mathbf{X}' \in \mathbb{R}^{n \times m-1}$, presentado de manera general que ambas matrices se pueden relacionar por el sistema dinámico de (1) utilizando como aproximación ese método DMD como se ve a continuación:

$$\mathbf{X}' \approx \mathbf{A}\mathbf{X}. \quad (2)$$

Pero ahora, ¿cómo solucionar esta aproximación lineal matricial? es por ende que se tendrá finalmente la fase de **implementación del método de Descomposición en Modos Dinámicos** donde de acuerdo al autor Kutz en [3] esta ecuación lineal matricial anterior pretende dar solución a un sistema no lineal por medio de un ajuste lineal de infinita dimensión dado por \mathbf{A} , siendo ese ajuste de la siguiente manera:

$$\mathbf{A} = \mathbf{X}'\mathbf{X}^{\dagger 1}$$

Aunque esto es en virtud, si y solo si, truncamos esta aproximación en un valor n determinado mediante el método SVD con la condición de $n \gg m$ para que el $\text{rang}(A)$ sea completo, siendo así, para ello aplicamos ese método de descomposición en \mathbf{X} para obtener el siguiente comportamiento:

$$\mathbf{X} = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^{*2}.$$

Después de haber entendido esta aproximación de A por la pseudo inversa, se procede a

ejecutar con ello para hallar los modos dinámicos del método DMD ya que estos presentan por medio de autovalores y autovectores el comportamiento de la enfermedad Covid19 de forma diagnóstica, es por ende, que a partir de lo definido se procede a ejecutar las siguientes líneas del método DMD:

$$\underbrace{\mathbf{A} = \mathbf{X}'\mathbf{X}^{\dagger}}_{\text{Aproximación}} := \mathbf{A} = \mathbf{X}' \left(\tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^* \right)^{\dagger}, \quad (3)$$

$$:= \mathbf{A} = \mathbf{X}'\tilde{\mathbf{V}}\tilde{\Sigma}^{-1}\tilde{\mathbf{U}}^*, \quad (4)$$

$$(\text{Definamos}) \tilde{\mathbf{A}} \approx \tilde{\mathbf{U}}^*\mathbf{A}\tilde{\mathbf{U}}, \quad (5)$$

$$:= \tilde{\mathbf{A}} = \tilde{\mathbf{U}}^*\mathbf{X}'\tilde{\mathbf{V}}\tilde{\Sigma}^{-1}, \quad (6)$$

$$:= \tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\Lambda, \quad (7)$$

\mathbf{W} y Λ son autovectores y autovalores respectivos de la matriz $\tilde{\mathbf{A}}$ que reconstruyen como aproximación la matriz \mathbf{A} para los modos dinámicos del método de Descomposición en Modos Dinámicos.

Teorema: Cada par (Φ, Λ) generado por el algoritmo anterior, es un **autovector** y **autovalor** de \mathbf{A} . Además, el algoritmo identifica todos los autovalores no nulos de \mathbf{A} .

Cuadro 1: Teorema del funcionamiento de autovectores y autovalores de método DMD, para ver la demostración revisar [13].

3. Algoritmo del método DMD

En resumen, los pasos y fases de la sesión anterior se condensa en el siguiente algoritmo de ejecución:

1. Organización de los datos $\{x_1, x_2, x_3, \dots, x_m\}$ en dos matrices \mathbf{X} y \mathbf{X}' .
2. Reducción por el método de SVD truncado, obteniendo $\mathbf{X} = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^*$.
3. Definición de la matriz $\tilde{\mathbf{A}} = \tilde{\mathbf{U}}^*\mathbf{X}'\tilde{\mathbf{V}}\tilde{\Sigma}^{-1}$.
4. Ejecución de la descomposición espectral de $\tilde{\mathbf{A}}$, obteniendo $\tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\Lambda$.
5. Construcción de los modos de DMD, correspondiente a cada Λ , de la forma: $\Phi = \mathbf{X}'\mathbf{V}\Sigma^{-1}\mathbf{W}$.

¹† es la Pseudoinversa de Moore-Penrose.

²* es el conjunto transpuesto.

4. Resultados

Habiendo estructurado la metodología de trabajo, se procede ahora a tener algunas estadísticas básicas para comprender así como es el funcionamiento de la información, dando como resultado lo siguiente:

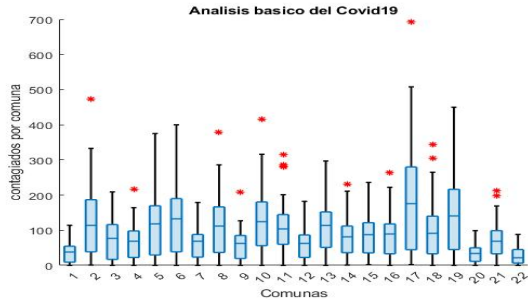


Figura 3: Diagrama de cajas que muestra una alta participación de contagio en las comunas 19, 6 y 17.

Luego, como sabemos de la anterior sesión dentro de su algoritmo de implementación, teniendo ya los *snapshots* agrupados en un conjunto total se procede de manera consiguiente a separar en sus dos matrices X y X' con el fin de darle un ingreso de información al método DMD. Esto se realiza para aplicar así la ecuación (4) como reducción y truncamiento de los datos, para así después llegar hasta los autovalores y autovectores que se aproximan en reconstrucción de la matriz A o que también es conocida como los modos dinámicos del método a partir de la ecuación (7).

Por lo tanto, los autovalores y autovectores de este resultado de ecuaciones genero los siguientes comportamientos de aporte en magnitudes gracias a su descomposición espectral realizada, viendo de la siguiente manera:

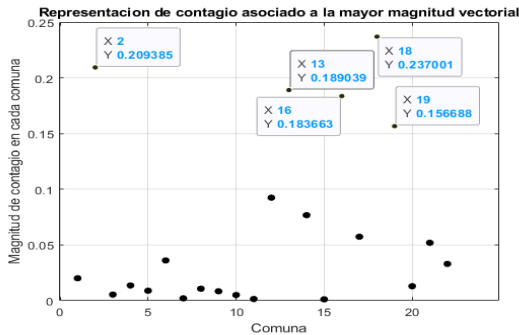


Figura 4: Valores de los mayores aportes en magnitud de contagio que generar las comunas de Cali de acuerdo al método DMD.

Como se evidencia, estos mayores aportes en magnitud están dados por las comunas 18, 2, 13, 16 y 19, pero que geográficamente se ven como aparece a continuación:

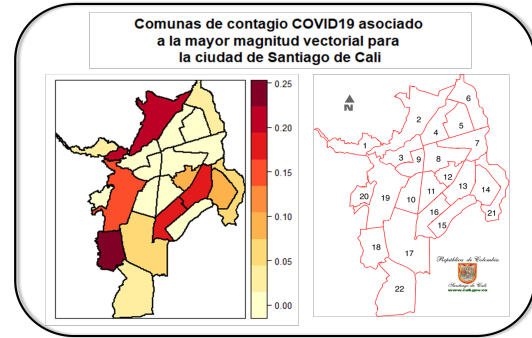


Figura 5: Mapa de los mayores aportes de magnitud de contagio que generar las comunas de Cali de acuerdo al método DMD.

Adicional ha este resultado de aporte, se indago sobre cual sería la comuna de Santiago de Cali que generará más frecuencia de aporte, es por eso que se aplico la siguiente ecuación:

$$frecuencia_j = \frac{imag(\log(\lambda_j))}{2\pi\Delta t}, \quad (8)$$

a raíz de esta ecuación, se obtuvo un esquema con las frecuencias distintivas de los autovalores de \tilde{A} , y ya con esto se selecciono la frecuencia de mayor aporte, para así obtener el panorama de frecuencia que aporta cada comuna según el siguiente esquema de valores. Coincidentalmente, estos aportes de frecuencia en comunas son iguales a los aporte de magnitud que se vieron anteriormente, para ello véase el siguiente mapa:

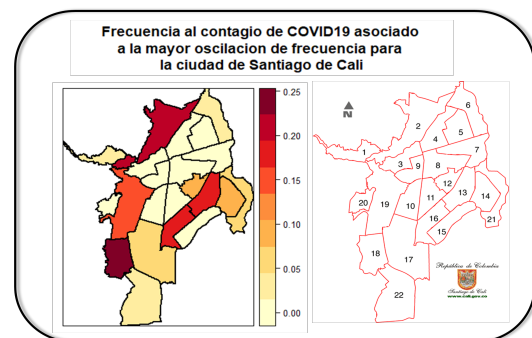


Figura 6: Mapa de los mayores aportes a frecuencia de contagio que generar las comunas de Cali de acuerdo al método DMD.

Finalmente, teniendo claro ese sector diagnóstico de la enfermedad con lo previamente ejecutado, ahora, dentro de los resultados también se aventuro a generar predicciones de contagio.

De este modo, tomando ese valor n determinado que se hablaba en la metodología donde si y solo si $n \gg m$ se tiene un sistema que $X' = f(x, \mu, t) \approx (X' \approx AX)$ permitiendo que la matriz A sea diagonalizable de la forma $A \sim PDP^{-1}$. Por tanto, tal como se ha construido, su solución predictiva sería de la siguiente manera:

$$\begin{aligned} X_2 &= AX_1 \Rightarrow X_2 = PDP^{-1}X_1 \\ X_3 &= AX_2 \Rightarrow X_3 = (PDP^{-1})(PDP^{-1})X_1 \\ &\vdots \\ X_{n+1} &= AX_n \Rightarrow X_{n+1} = PD^n \underbrace{P^{-1}X_1}_b = PD^n b. \end{aligned}$$

Por lo tanto, con la construcción de los modos dinámicos del método DMD, se tiene que $\Phi = U^*W$ permitiendo crear así una ecuación solución para reconstruir proyecciones de la manera $X_{n+1} = \Phi \Lambda^n \Phi^{-1} X_1 \Rightarrow \Phi \Lambda^n b$.

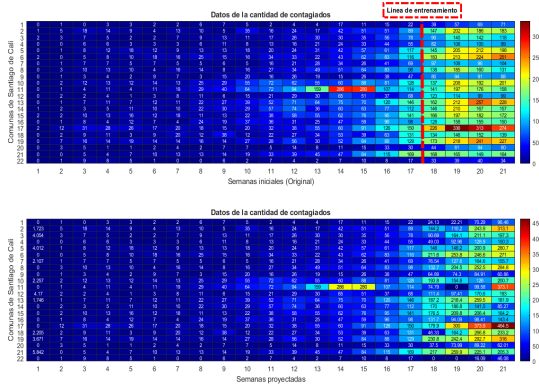


Figura 7: Mapa de calor con el número de contagiados predichos bajo el método de DMD.

Como se puede ver, en el anterior gráfico se obtuvo una aproximación bastante buena respecto a los valores originales del conjunto de datos, por lo que el modelo predictivo se acerca lo suficiente según sus datos de entrenamiento.

5. Conclusiones

De acuerdo a los análisis en lo diagnóstico y lo predictivo, se tiene las siguientes conclusiones:

- **Diagnostico:** Dado que las comunas 18,2,13 y 16 son las de mayor aporte tanto en magnitud como frecuencia, se propone crea una política publica con la capacidad de posicionar entidades de salud eficiente con el fin de hacer control y prevención a más contagios del virus Covid19.

- **Predictivo:** El modelo evidencia una alta precisión frente a los datos originales, sin embargo, se tiene una limitante en cuantos a los datos de entrenamiento ya que son pocos dado por la condición $n \gg m$ y adicional a esto, también se considera que existe una alta susceptibilidad de la condición inicial, por lo que se sugiere indagar más en este tema.

Referencias

- [1] ARANGO, D. Evolución de los Casos de COVID-19 en Cali durante los primeros 2 meses de la Pandemia. Tech. rep., Secretaria de Salud, Santiago de Cali, Cali, Colombia, 2020.
- [2] CLERKIN, K. J., FRIED, J. A., RAIKHELKAR, J., SAYER, G., GRIFFIN, J. M., MASOUMI, A., JAIN, S. S., BURKHOFF, D., KUMARAIAH, D., RABBANI, L. R., SCHWARTZ, A., AND URIEL, N. COVID-19 and Cardiovascular Disease, 2020.
- [3] KUTZ, J. N. Data-Driven Modeling $\{\&\}$ Scientific Computation. 638.
- [4] KUTZ, J. N., BRUNTON, S. L., BRUNTON, B. W., AND PROCTOR, J. L. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.
- [5] LURIE, N., SAVILLE, M., HATCHETT, R., AND HALTON, J. Developing covid-19 vaccines at pandemic speed, 2020.
- [6] MEHTA, P., MCAULEY, D. F., BROWN, M., SANCHEZ, E., TATTERSALL, R. S., AND MANSON, J. J. COVID-19: consider cytokine storm syndromes and immunosuppression, 2020.
- [7] MILTON, K. *Ecologías: antropología, cultura y entorno*, 2007.
- [8] MURCIA, N., JAIMES, S. S., AND GÓMEZ, J. La práctica social como expresión de humanidad. *Cinta de moebio* 57, 57 (2016), 257–274.
- [9] ORGANIZATION, W. H. Coronavirus disease (COVID-19) Situation Report – 197 Data. Tech. rep., World Health Organization, 2020.
- [10] PROCTOR, J. L., BRUNTON, S. L., AND NATHAN KUTZ, J. Generalizing koopman theory to allow for inputs and control. *SIAM Journal on Applied Dynamical Systems* 17, 1 (2018), 909–930.
- [11] PROCTOR, J. L., AND ECKHOFF, P. A. Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *International Health* 11, 2 (2019), 83–92.
- [12] TISSOT, G., CORDIER, L., BENARD, N., AND NOACK, B. R. Model reduction using Dynamic Mode Decomposition, 2014.
- [13] TU, J. H., ROWLEY, C. W., LUCHTENBURG, D. M., BRUNTON, S. L., AND KUTZ, J. N. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics* 1, 2 (12 2014), 391–421.
- [14] ZU, Z. Y., JIANG, M. D., XU, P. P., CHEN, W., NI, Q. Q., LU, G. M., AND ZHANG, L. J. Coronavirus Disease 2019 (COVID-19): A Perspective from China. *Radiology* (2020), 200490.