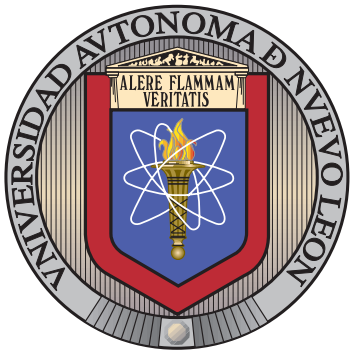
Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Posgrado en Ingeniería de Sistemas

Doctorado



Portafolio de Evidencias

de

José Alberto Benavides Vázquez

1373079

para el curso de Metodología de la investigación

con la profesora Elisa Schaeffer.

Semestre Enero 2022 - Julio 2022.

https://github.com/jbenavidesv87/metodologia

Name of the University
Name of the Faculty
Name of the Programme

RESEARCH PROJECT PROPOSAL

Name of Candidate: Applicant Full Name

Date of Admission to this programme: February 3, 2022

Full time mode

Abstract

Brief summary of the proposal and its sections.

1. Project Title

Communicate the key problem of the research in a brief sentence.

2. Introduction

The proposal length could vary, but four to seven pages are recommended in order to convince that the problem presented is relevant and could be studied and solved in the duration of the program and with the resources available.
The introduction consists of one or two paragraphs about the research problem, the related topics, the importance of the problem being studied or solved, the proposed methods to approach it, the relevance of the outcomes and their differences from the ones already existing.

3. Hypothesis

Write the hypothesis that will be tested in one sentence. This should guide the project objectives and research questions.

4. Project objectives and/or Research questions

- The two or three most important goals.
- Can be pointed in a list or a concise paragraph.

5. Background and related literature

Incorporate the background supporting the methodology and the recent studies for the research problem focusing in the main objectives or research questions.

6. Methodology

Explain in order the procedures involved in the research process, summarize the theoretical bases, explain the advantages of choosing this methods over other approaches.

7. Significance

While the Introduction mention the relevance of the research, here it need to be highlighted and demonstrated. Focus specially in the contributions that the results could bring to the state of the art and/or the solutions that can provide to the problems involved.

8. External Collaboration

If any, include authors, published papers, works in research programs or conferences where the research have been involved.

9. Project schedule

Add a schedule that includes the steps of the process, from investigation, possible publications and participation in conferences, to final products like thesis, libraries... A Gantt diagram is recommended.

References

Monash University. Writing a research proposal. https://www.monash.edu/rlo/graduate-research-writing/write-the-thesis/writing-a-research-proposal. Accessed: 2021-08-10.

quxiaofeng. A proposal form with tick, block square, paragraphs, and underlines. https://es.overleaf.com/latex/examples/a-proposal-form-with-tick-block-square-paragraphs-and-underlines/dmwhcxzxyxtp. Accessed: 2021-08-10.

University of Birmingham. How to write a research proposal. https://www.birmingham.ac.uk/schools/law/courses/research/research-proposal.aspx. Accessed: 2021-08-10.

University of Southern California. Writing a research proposal. https://libguides.usc.edu/writingguide/researchproposal. Accessed: 2021-08-10.

University of Westminster. How to write your research proposal. https://www.westminster.ac.uk/study/postgraduate/research-degrees/entry-requirements/how-to-write-your-research-proposal. Accessed: 2021-08-10.

Yale College. How to write a proposal. https://science.yalecollege.yale.edu/fellowships/how-write-proposal. Accessed: 2022-01-20.

York St. John University. Writing your research proposal. https://www.yorksj.ac.uk/study/research/apply/examples-of-research-proposals/. Accessed: 2021-08-10.

Name of the University
Name of the Faculty
Name of the Programme

RESEARCH PROJECT PROPOSAL

Name of Candidate: Applicant Full Name

Date of Admission to this programme: February 10, 2022

Full time mode

Abstract

Brief summary of the proposal and its sections.

1. Project Title

Communicate the key problem of the research in a brief sentence.

2. Introduction

The proposal length could vary, but four to seven pages are recommended in order to convince that the problem presented is relevant and could be studied and solved in the duration of the program and with the resources available.
The introduction consists of one or two paragraphs about the research problem, the related topics, the importance of the problem being studied or solved, the proposed methods to approach it, the relevance of the outcomes and their differences from the ones already existing.

3. Hypothesis

Write the hypothesis that will be tested in one sentence. This should guide the project objectives and research questions.

4. Project objectives and/or Research questions

- The two or three most important goals.
- Can be pointed in a list or a concise paragraph.

5. Background and related literature

Incorporate the background supporting the methodology and the recent studies for the research problem focusing in the main objectives or research questions.

6. Methodology

Explain in order the procedures involved in the research process, summarize the theoretical bases, explain the advantages of choosing this methods over other approaches. Reference known methods and specify the variations. It is possible to use subsections for each method used. In this section should not include figures or tables. If data or libraries are used, it is recommended that they are public in order to ease the reproductibility. If they are under license, it should be noted.

7. Significance

While the Introduction mention the relevance of the research, here it need to be highlighted and demonstrated. Focus specially in the contributions that the results could bring to the state of the art and/or the solutions that can provide to the problems involved.

8. External Collaboration

If any, include authors, published papers, works in research programs or conferences where the research have been involved.

9. Project schedule

Add a schedule that includes the steps of the process, from investigation, possible publications and participation in conferences, to final products like thesis, libraries... A Gantt diagram is recommended.

Name of the University
Name of the Faculty
Name of the Programme

RESEARCH PROJECT PROPOSAL

Name of Candidate: Applicant Full Name

Date of Admission to this programme: February 17, 2022

Full time mode

Abstract

Brief summary of the proposal and its sections.

1. Project Title

Communicate the key problem of the research in a brief sentence.

2. Introduction

The proposal length could vary, but four to seven pages are recommended in order to convince that the problem presented is relevant and could be studied and solved in the duration of the program and with the resources available.
The introduction consists of one or two paragraphs about the research problem, the related topics, the importance of the problem being studied or solved, the proposed methods to approach it, the relevance of the outcomes and their differences from the ones already existing.

3. Hypothesis

Write the hypothesis that will be tested in one sentence. This should guide the project objectives and research questions.

4. Project objectives and/or Research questions

- The two or three most important goals.
- Can be pointed in a list or a concise paragraph.

5. Background and related literature

Incorporate the background supporting the methodology and the recent studies for the research problem focusing in the main objectives or research questions.

6. Methodology

The study will focus in the **Monterrey Metropolitan Area (MMA) between 2018 and 2019**. The air quality samples in the are taken each hour from 13 sensors over the MMA that measure concentrations of **CO**, **NO**, **NO**$_2$, **O**$_3$, **SO**$_2$, **PM**$_{10}$, **PM**$_{2.5}$, and **atmospheric pressure**, **rainfall**, **relative humidity**, **solar radiation**, **temperature**, **wind velocity and direction**. This data is provided by the *Sistema Integral de Monitoreo Ambiental de Nuevo León* (SIMA) [10]. The

Mexican diseases data was obtained from the Mexican *Department of Health* [9] and contains information from all states and municipalities in Mexico such as **date of admission**, **egress date**, **age at the admission**, **gender**, **weight**, **height**, **ICD code upon arrived**, **ICD code upon diagnosed**, **reason of egress**.
The air quality data needs to be interpolated because it contains imputed records. Different temporal interpolation techniques are used and compared [5]. Also, an spatio-temporal interpolation is performed to obtain the missing data values [7]. Both data sets are processed and converted to georeferenced time series [11] that are stationary [6] in order to establish their relationship by cross correlation [4], multiple regression analysis [2], vector autorregresive approaches, causality models [8] and geographic interactions [3]. The results are ranked by metrics like R^2, the Akaike (AIC), and the Bayesian (BIC) information criteria [1]. Finally, it will be produced a web application that allow general and specialized population to interact with the data and obtain forecasts, interactions and visualizations of the models described.

7. Significance

While the Introduction mention the relevance of the research, here it need to be highlighted and demonstrated. Focus specially in the contributions that the results could bring to the state of the art and/or the solutions that can provide to the problems involved.

8. External Collaboration

If any, include authors, published papers, works in research programs or conferences where the research have been involved.

9. Project schedule

Add a schedule that includes the steps of the process, from investigation, possible publications and participation in conferences, to final products like thesis, libraries…A Gantt diagram is recommended.

References

[1] Jim Albert. Bayesian computation with R. Springer-Verlag GmbH, July 2007. ISBN 9780387713854. URL https://www.ebook.de/de/product/24984339/jim_albert_bayesian_computation_with_r.html.

[2] Peter Brockwell. Introduction to time series and forecasting. Springer, New York, 2002. ISBN 9780387216577.

[3] Alexis Comber and Wen Zeng. Spatial interpolation using areal features: A review of methods and opportunities using new forms of data with coded illustrations. Geography Compass, 13 (10), August 2019. doi: 10.1111/gec3.12465.

[4] Timothy R. Derrick and Joshua M. Thomas. Innovative analyses of human movement, chapter Time Series Analysis: The Cross-Correlation Function, pages 189–205. Kinesiology Publications, 2004.

[5] Milton Friedman. The interpolation of time series by related series. Journal of the American Statistical Association, 57(300):729–757, December 1962. doi: 10.1080/01621459.1962.10500812.

[6] Rob Hyndman. Forecasting: principles and practice. OTexts, Melbourne, 2018. ISBN 9780987507112.

[7] Lixin Li and Peter Revesz. A comparison of spatio-temporal interpolation methods. In Geographic Information Science, pages 145–160. Springer Berlin Heidelberg, 2002. doi: 10.1007/3-540-45799-2_11.

[8] F. Popescu and I. Guyon. Causality in time series. Microtome PublishingBrookline, 2013.

[9] Secretaría de Salud. Egresos Hospitalarios. Datos abiertos, August 2020. URL http://www.dgis.salud.gob.mx/contenidos/basesdedatos/da_egresoshosp_gobmx.html.

[10] Sistema Integral de Monitoreo Ambiental (SIMA), Nuevo León. aire.nl.gob.mx | Home. http://aire.nl.gob.mx/, 2020. [Accesed 14/may/2020].

[11] William Wei. Time series analysis univariate and multivariate methods. Pearson, Boston, 2019. ISBN 9780134995366.

Tarea 4

Investiga las guías sobre referencias de por lo menos tres revistas indizadas de tu área igual como los líneamientos bibliográficos referentes a las tesis de su nivel de por lo menos tres universidades extranjeras; incluye citas claras a tus fuentes. Acorde a tus hallazgos, analiza qué tan bien la bibliografía citada en tu borrador actual de la tesis cumple con lo solicitado.

Nature

- Página principal
- Guía de formato
- Referencias numeradas por orden de aparición
- Alrededor de 50 referencias sin contar las relacionadas con Metodología o información complementaria
- Sólo referenciar artículos publicados, aceptados o cargados en servidores para *preprints*
- Agregar todos los autores en las referencias, salvo cuando hay más de cinco. En ese caso usar *et al.*
- Títulos
    - Los títulos de las citas son obligatorios
    - Primera letra en mayúsculas
    - Títulos de libros y revistas en cursiva
    - Números de volúmenes en negritas
    - Citar libros requiere ciudad y editorial
- Conjutos de datos requieren un DOI, autores, título, repositorio
- Artículos sólo disponibles en línea requieren autores, título, revista, año de publicación en paréntesis, y URL o DOI
- Páginas web requieren autor, título, URL, año de publicación en paréntesis

International Journal of Forecasting

- Página principal
- Guía de estilo
- Usar generadores de bibliografía (como Mendeley)
- Citas en texto con Autor (fecha, p. 123)
- Bibliografía citadas alfabéticamente
- Las referencias en el texto y bibliografía deben coincidir
- Incluir fuentes relevantes
- Páginas web al menos URL y fecha de último acceso. DOI, autores, fecha de publicación también pueden ser incluidos
- Conjuntos de datos necesitan autor, título, repositorio, año, identificador global persistente, versión

IEEE Journal of Biomedical and Health Informatics

- Página principal
- Estructura del artículo
- Las referencias dan sustento de trabajo científico previo
- Sirven como reconocimiento al trabajo ajeno
- Agregar referencias de más va en contra de la ética de publicación

International Journal of Epidemiology

- Página principal
- Información para autores
- Las referencias deben seguir el formato de IJE
- Sólo usar notas al pie para tablas y figuras, no para referencias
- Enumeradas entre corchetes según aparecen en el texto
- Bibliografía al final
- Revistas: Autores por apellido y nombre, título en cursivas, nombre de la revista abreviada, año, volúmen, primera y última páginas
- Libros: Lugar de publicación, editorial, año
- Artículos no publicados o manuscritos no deben aparecer en la bibliografía, pero sí entre paréntesis en el texto
- Páginas web: Autores, título del sitio, título de la página, título del artículo, año de publicación, URL, fecha de última revisión
- *Preprints* sí se citan y referencian
- Fuentes de datos: Autores, título, repositorio, identificador, DOI

Methodology and Computing in Applied Probability

- Página principal
- Guía de envíos
- Plantilla de Latex
- Se pueden usar notas al pie para las referencias, que también deben incluirse en la bibliografía
- En texto, van entre paréntesis con el nombre del autor y año
- Bibliografía en orden alfabético
- Proveer DOI siempre que esté disponible

Journal of Statistical Software

- Página principal
- Guía de estilo
- Usar Latex y Bibtex

Monash University

- Guía de estilo
- Seguir lineamientos de IEEE

Yale

- Guía de estilo
- Citar todas las referencias, incluyendo datos no publicados de tutores

University of Westminster

- Guía de estilo
- Citar todas las referencias utilizadas
- Citar todas las fuentes utilizadas

Universidad Autonoma de Nuevo Leon
Facultad de Ingenieria Mecanica y Electrica
Posgrado de Ingenieria de Sistemas

RESEARCH PROJECT PROPOSAL

Name of Candidate: Jose Alberto Benavides Vazquez

Date of Admission to this programme: March 10, 2022

Full time mode

Abstract

Brief summary of the proposal and its sections.

1. Project Title

Impact of the air pollutants in the population health in the Monterrey Metropolitan Area.

2. Introduction

This study focuses on the impact that the air quality have in the population welfare primarily in the Monterrey Metropolitan Area between 2018 and 2019 obtained from the Comprehensive Environmental Monitoring System (abbreviated SIMA in Spanish from Sistema Integral de Monitoreo Ambiental) [32]. Also, it have been incorporated a database containing the hospital discharge records for Mexico over the same span of time from the open repository of the Department of Health of Mexico [12].

As the atmosphere is result of the combination of 78% of N_2, 21% of O_2 and 1% of combined helium, water vapor, oxides, methane, noble gases, sulfides, ozone, ammonia, formaldehyde, and hydroxyl radical, can be defined that the air is *polluted* when these concentrations are altered by human activities resulting in harmful environments to biological ecosystems [13].

The human activities that cause pollution have been increasing over the years. Air quality became an urban problem since 1930 when a fog in Belgium was associated with the death of 63 people and then in London during 1952 [10] when another fog caused $4,000$ deaths [28] being pneumonia the principal illness diagnosed in both cases. Its negative impact on health and general welfare concerned authorities that design regulations in order to minimize those undesirable effects. In 1981 it became an international problem when the CO concentrations were seen from space for the first time [1]. Such images show that the pollution was not only generated from expected sources, but from other unforeseen activities, and that the pollution generated in one place affected distant populations over time.

Air pollution can be caused by events that impact the atmosphere, like forest fires, increased volcanic activity, animal and vegetable decomposition, climate change [21], etc. The main difference from natural and human pollutants is that the natural ones tend to occur far from large human populations and are less often than human sources of pollution, which are generally located in or near large human populations, and their emissions are more frequent. The main human sources of air pollution are associated with transportation, fuel combustion, and industrial processes [8].

The International Classification of Diseases (ICD)[1] is a code associated to diseases by their trends and statistical values, established by the [35]. It consists, generally, in a letter and three

[1]The name of the diseases corresponding to the ICD code is obtained from https://icd.who.int/browse10/2019/en

numbers between 0 and 9. For example, the code for pneumonia due to staphylococcus is J152. This system comprehends from the values A00.0 to Z99.9, leaving the U letter for yet unknown sources. This classification was proposed by Farr at the end of the XIX century [34], and its purpose is to understand the causes of morbidity and mortality in order to improve the quality of life of the human population [34].

3. Hypothesis

The contamination levels reported can explain the statistically significant variations and trends of georeferenced data from health reports. Also, it is possible to automatize these relationships in an interactive and free web service.

4. Project objectives

- Stablish a clear and scientifically sustained relationship between air pollution and the human diseases that them can cause.
- Study the impact of the air pollution in the Metropolitan Monterrey Area during 2017 and 2019 in the health population.
- Choose or propose a methodology to calculate the causation between air pollution and associated human diseases.
- Identify the areas where they are more air pollution and stablish a relationship between the people that live nearby those areas.
- Propose a rank to determine which air pollutants have the most direct effect in the human health.
- Propose pollution emission limits, scientifically justified, that can be used by organizations interested in this matter.
- Generate interactive and free tools that let people and specialist know the levels, relationships, risks, etc. related to the air pollutants studied.

5. Research questions

- Is it possible to stablish statistically significant relationships between air pollution and human diseases?
- What are the best methods to detect the causation between air pollution and human diseases?
- How to eliminate the climate and seasonal components from the relationship between air pollution and human diseases?
- Does the environmental and geographical factors can be used to predict the amount of air pollution in an area and time?
- Which air pollutants cause more human diseases?
- It is possible to stablish short (days), medium (months) and large (years) interactions between air pollution and human diseases?

6. Background and related literature

The first correspondence to discuss relationship between air quality and human health was registered during December 1930 in Belgium, when a three-day fog s was declared to be the cause of death of 63 people — during the fog, disease and death were observed to increase, whereas after its dissipation, the figures normalized again [10]. Similarly, in December 1952, a four-day fog in London was associated with four thousand deaths [28]. The majority of the fog-related deaths were related to respiratory diseases.

In 1959 Lawther [24] used a method to quantify the severity of respiratory illness based on their condition (negative numbers indicated they were recovered and positive numbers that they were getting worse). Then, he plotted a time series superposed with concentrations of two pollutants — SO_2 and smoke — and two weather variables — temperature and humidity; he discovered that the pollutants bore a similarity to the severity whereas the climate ones did not.

During the 1970s, short-term exposure to different pollutants was studied [7, 15, 16, 27] with experiments that implied some ethical and legal considerations, focusing on physiological analysis of the subjects. A compilation was published in by Ferris [9] who documented that the pollutants, in regulated concentrations, had a negative impact on health, primarily on respiratory diseases.

A set of methodologies are used to measure the relationships that are the focus of this study. They are **multiple regression analysis** [17, 18, 20, 25] with variations in the distributions such as logistic regressions [22], **multivariate analysis** [14, 23], **auto regressive models** [19], **causality models** [30] and **case-crossover approximations** [4, 29].

7. Methodology

The study will focus in the **Monterrey Metropolitan Area (MMA)** between 2018 and 2019. The air quality samples in the are taken each hour from 13 sensors over the MMA that measure concentrations of **CO**, **NO**, **NO**$_2$, **O**$_3$, **SO**$_2$, **PM**$_{10}$, **PM**$_{2.5}$, and **atmospheric pressure**, **rainfall**, **relative humidity**, **solar radiation**, **temperature**, **wind velocity and direction**. This data is provided by the *Sistema Integral de Monitoreo Ambiental de Nuevo León* (SIMA) [32]. The Mexican diseases data was obtained from the Mexican *Department of Health* [31] and contains information from all states and municipalities in Mexico such as **date of admission**, **egress date**, **age at the admission**, **gender**, **weight**, **height**, **ICD code upon arrived**, **ICD code upon diagnosis**, and **reason of egress**.

The air quality data needs to be interpolated because it contains imputed records. Different temporal interpolation techniques are used and compared [11]. Also, an spatio-temporal interpolation is performed to obtain the missing data values [26]. Both data sets are processed and converted to georeferenced time series [33] that are stationary [20] in order to establish their relationship by cross correlation [6], multiple regression analysis [3], vector autorregresive approaches, causality models [30] and geographic interactions [5]. The results are ranked by metrics like R^2, the Akaike (AIC), and the Bayesian (BIC) information criteria [2]. Finally, it will be produced a web application that allow general and specialized population to interact with the data and obtain forecasts, interactions and visualizations of the models described.

8. Expected results

9. Significance

While the Introduction mention the relevance of the research, here it need to be highlighted and demonstrated. Focus specially in the contributions that the results could bring to the state of the art and/or the solutions that can provide to the problems involved.

10. External Collaboration

If any, include authors, published papers, works in research programs or conferences where the research have been involved.

11. Project schedule

Add a schedule that includes the steps of the process, from investigation, possible publications and participation in conferences, to final products like thesis, libraries…A Gantt diagram is recommended.

References

[1] Hajime Akimoto. Global air quality and pollution. Science, 302(5651), 2003. doi: 10.1126/science.1092666.

[2] Jim Albert. Bayesian computation with R. Springer-Verlag GmbH, July 2007. doi: 10.1007/978-0-387-92298-0. URL https://www.ebook.de/de/product/24984339/jim_albert_bayesian_computation_with_r.html.

[3] Peter Brockwell. Introduction to time series and forecasting. Springer, New York, 2002. doi: 10.1007/978-3-319-29854-2.

[4] Eduardo Carracedo-Martínez, Aurelio Tobías, Marc Saez, Margarita Taracido, and Adolfo Figueiras. Fundamentos y aplicaciones del diseño de casos cruzados. Gaceta Sanitaria, 23: 161–165, April 2009. ISSN 0213-9111. URL http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0213-91112009000200017&nrm=iso.

[5] Alexis Comber and Wen Zeng. Spatial interpolation using areal features: A review of methods and opportunities using new forms of data with coded illustrations. Geography Compass, 13 (10), August 2019. doi: 10.1111/gec3.12465.

[6] Timothy R. Derrick and Joshua M. Thomas. Innovative analyses of human movement, chapter Time Series Analysis: The Cross-Correlation Function, pages 189–205. Kinesiology Publications, 2004.

[7] Richard Ehrlich, J. C. Findlay, J. D. Fenters, and D. E. Gardner. Health effects of short-term inhalation of nitrogen dioxide and ozone mixtures. Environmental Research, 14(2):223–231, October 1977. doi: 10.1016/0013-9351(77)90034-2.

[8] Jes Fenger. Urban air quality. Atmospheric Environment, 33(29), 1999. doi: 10.1016/s1352-2310(99)00290-3.

[9] Benjamin G. Ferris. Health effects of exposure to low levels of regulated air pollutants. Journal of the Air Pollution Control Association, 28(5):482–497, May 1978. doi: 10.1080/00022470.1978.10470621.

[10] J. Firket. Fog along the Meuse Valley. Transactions of the Faraday Society, 32:1192, 1936. doi: 10.1039/tf9363201192.

[11] Milton Friedman. The interpolation of time series by related series. Journal of the American Statistical Association, 57(300):729–757, December 1962. doi: 10.1080/01621459.1962.10500812.

[12] Gobierno de México, Secretaría de Salud. Dirección general de información en salud. https://www.gob.mx/salud/acciones-y-programas/direccion-general-de-informacion-en-salud-dgis, July 2020.

[13] Thad Godish, Wayne T. Davis, and Joshua S. Fu. Air quality. Florida, The address, 5 edition, 2015. doi: 10.1201/9781003032595.

[14] Hans W. Gottinger. Air pollution health effects in the Munich metropolitan area: Preliminary results based on a statistical model. Environment International, 9(3):207–220, January 1983. doi: 10.1016/0160-4120(83)90038-7.

[15] Jack D. Hackney, William S. Linn, Ramon D. Buckley, E. Eugene Pedersen, Sarunas K. Karuza, David C. Law, and D. Armin Fischer. Experimental studies on human health effects of air pollutants. Archives of Environmental Health: An International Journal, 30(8):373–378, August 1975. doi: 10.1080/00039896.1975.10666728.

[16] Jack D. Hackney, Frederick C. Thiede, William S. Linn, E. Eugene Pedersen, Charles E. Spier, David C. Law, and D. Armin Fischer. Experimental studies on human health effects of air pollutants. IV. Short-term physiological and dlinical effects of nitrogen dioxide exposure. Archives of Environmental Health: An International Journal, 33(4):176–181, July 1978. doi: 10.1080/00039896.1978.10667330.

[17] Victor Hasselblad. Comparison of methods for the analysis of panel studies. Technical report, Environmental Protection Agency, June 1978. https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=2000ZPAO.PDF.

[18] A. Hatzakis, K. Katsouyanni, A. Kalandidi, N. Day, and D. Trichopoulos. Short-term effects of air pollution on mortality in Athens. International Journal of Epidemiology, 15(1):73–81, March 1986. doi: 10.1093/ije/15.1.73.

[19] Steven C. Hillmer and William W. S. Wei. Time series analysis: Univariate and multivariate methods., volume 86. JSTOR, March 1991. doi: 10.2307/2289741.

[20] Rob Hyndman. Forecasting: principles and practice. OTexts, Melbourne, 2018. ISBN 9780987507112.

[21] Daniel Jacob and Darrel Winner. Effect of climate change on air quality. Atmospheric Environment, 43(1), 2009. doi: 10.1016/j.atmosenv.2008.09.051.

[22] Edward L. Korn and Alice S. Whittemore. Methods for analyzing panel studies of acute health effects of air pollution. Biometrics, 35(4):795, December 1979. doi: 10.2307/2530111.

[23] Lester B. Lave and Eugene P. Seskin. An analysis of the association between U.S. mortality and air pollution. Journal of the American Statistical Association, 68(342):284–290, June 1973. doi: 10.1080/01621459.1973.10482421.

[24] P. J. Lawther. Chronic bronchitis and air pollution. Journal (Royal Society of Health), 79(1):4–9, January 1959. doi: 10.1177/146642405907900102.

[25] Michael D. Lebowitz. A comparative analysis of the stimulus-response relationship between mortality and air pollution-weather. Environmental Research, 6(1):106–118, March 1973. doi: 10.1016/0013-9351(73)90023-6.

[26] Lixin Li and Peter Revesz. A comparison of spatio-temporal interpolation methods. In Geographic Information Science, pages 145–160. Springer Berlin Heidelberg, 2002. doi: 10.1007/3-540-45799-2_11.

[27] William S. Linn, Ramon D. Buckley, Charles E. Spier, Raymond L. Blessey, Michael P. Jones, D. Armin Fischer, and Jack D. Hackney. Health effects of ozone exposure in asthmatics. American Review Of Respiratory Disease, 117(5):835–843, May 1978. doi: 10.1164/arrd.1978.117.5.835.

[28] W. P. D. Logan. Mortality in the London fog incident, 1952. Lancet (London, England), 1:336–8, February 1953. doi: 10.1016/s0140-6736(53)91012-5.

[29] M. Maclure. The case-crossover design: A method for studying transient effects on the risk of acute events. American journal of epidemiology, 133:144–53, January 1991. doi: 10.1093/oxfordjournals.aje.a115853.

[30] F. Popescu and I. Guyon. Causality in time series. Microtome PublishingBrookline, 2013.

[31] Secretaría de Salud. Egresos Hospitalarios. Datos abiertos, August 2020. URL http://www.dgis.salud.gob.mx/contenidos/basesdedatos/da_egresoshosp_gobmx.html.

[32] Sistema Integral de Monitoreo Ambiental (SIMA), Nuevo León. aire.nl.gob.mx | Home. http://aire.nl.gob.mx/, 2020. [Accesed 14/may/2020].

[33] William Wei. Time series analysis univariate and multivariate methods. Pearson, Boston, 2019. doi: 10.2307/2289741.

[34] World Health Organization. International statistical classification of diseases and related health problems. World Health Organization, Geneva, 2011. ISBN 9789241548342.

[35] World Health Organization. International classification of diseases. https://www.who.int/classifications/icd/en/, 2020.

Tarea 6

Investiga las indicaciones sobre el reportaje de la solución propuesta y resultados experimentales de por lo menos tres revistas indizadas de tu área y aquellas sobre tesis de su nivel de estudios de por lo menos tres universidades extranjeras; incluye citas claras a tus fuentes. Acorde a tus hallazgos, prepara un "checklist" sobre aspectos de la solución propuesta y experimentos por incluir en el anteproyecto para que cumpla con lo solicitado.

Nature

- Página principal
- Guía de formato
- Acceso público a los datos, materiales, código y protocolos de investigación para replicar resultados

International Journal of Forecasting

- Página principal
- Guía de estilo
- Los resultados deben ser suscintos.

IEEE Journal of Biomedical and Health Informatics

- Página principal
- Estructura del artículo
- Los resultados deben corresponder a los realmente obtenidos en la investigación
- Se aconseja el uso de figuras para mostrar tendencias
- Se recomiendan las tablas para mostrar valores exactos

International Journal of Epidemiology

- Página principal
- Información para autores
- Las referencias deben seguir el formato de IJE
- Los resultados deben evidenciar la significancia del trabajo
- Hacer disponibles los datos para generar resultados
- Reportar tiempos de respuesta (para epidemias)
- Reportar intervalos de confianza (si los hay)
- Especificar si los resultados están relacionados con la hipótesis, se incluyeron posteriormente o tienen otra derivación

Methodology and Computing in Applied Probability

- Página principal
- Guía de envíos
- Plantilla de Latex
- Evitar enviar resultados modificados malintencionadamente que puedan afectar a la confianza en la revista

- Presentar clara y honestamente los resultados
- Prohibido falsear, fabricar o manipular los datos para maquillar resultados
- Obtención y preprocesamiento de datos según lineamientos del área de estudio
- Ser capaz de verificar validez de los resultados, principalmente proporcionando orígenes de datos

Journal of Statistical Software

- Página principal
- Guía de estilo
- Escalar imágenes muy grandes mediante Latex
- Utilizar imágenes en PDF a menos que sean demasiado grandes, en ese caso preferir PNG de resolución que permita su clara visualización

Monash University

- Guía de estilo
- Mostrar resultados de todos los experimentos y las pruebas estadísticas que los validan
- Inlcuir estadística descriptiva de los datos
- Apoyarse en tablas e imágenes
- Discutir tablas e imágenes en texto

Yale

- Guía de estilo
- Mostrar resultados relacionados a los objetivos especificados

Checklist

- [ ] Resultados breves, claros, honestos, correspondientes a la metodología
- [ ] Publicar resultados relacionados con la hipótesis y objetivos específicos
- [ ] Especificar si los resultados son derivados del estudio
- [ ] Compartir datos, código y la metodología del procesamiento sin falsear información
- [ ] Reportar estadística descriptiva básica y tendencias
- [ ] Especificar intervalos de confianza y pruebas estadísticas
- [ ] Apoyarse con el uso de imágenes y tablas cuando sea necesario (favorecer Latex nativos)
- [ ] Explicar y referenciar tablas e imágenes agregadas

Universidad Autonoma de Nuevo Leon
Facultad de Ingenieria Mecanica y Electrica
Posgrado de Ingenieria de Sistemas

RESEARCH PROJECT PROPOSAL

Name of Candidate: Jose Alberto Benavides Vazquez

Date of Admission to this programme: March 31, 2022

Full time mode

1. Project Title

Impact of the air pollutants in the population health in the Monterrey Metropolitan Area.

2. Introduction

This study focuses on the impact that the air quality have in the population welfare primarily in the Monterrey Metropolitan Area between 2018 and 2019 obtained from the Comprehensive Environmental Monitoring System (abbreviated SIMA in Spanish from Sistema Integral de Monitoreo Ambiental) [32]. Also, it have been incorporated a database containing the hospital discharge records for Mexico over the same span of time from the open repository of the Department of Health of Mexico [12].

As the atmosphere is result of the combination of 78% of N_2, 21% of O_2 and 1% of combined helium, water vapor, oxides, methane, noble gases, sulfides, ozone, ammonia, formaldehyde, and hydroxyl radical, can be defined that the air is *polluted* when these concentrations are altered by human activities resulting in harmful environments to biological ecosystems [13].

The human activities that cause pollution have been increasing over the years. Air quality became an urban problem since 1930 when a fog in Belgium was associated with the death of 63 people and then in London during 1952 [10] when another fog caused $4,000$ deaths [28] being pneumonia the principal illness diagnosed in both cases. Its negative impact on health and general welfare concerned authorities that design regulations in order to minimize those undesirable effects. In 1981 it became an international problem when the CO concentrations were seen from space for the first time [1]. Such images show that the pollution was not only generated from expected sources, but from other unforeseen activities, and that the pollution generated in one place affected distant populations over time.

Air pollution can be caused by events that impact the atmosphere, like forest fires, increased volcanic activity, animal and vegetable decomposition, climate change [21], etc. The main difference from natural and human pollutants is that the natural ones tend to occur far from large human populations and are less often than human sources of pollution, which are generally located in or near large human populations, and their emissions are more frequent. The main human sources of air pollution are associated with transportation, fuel combustion, and industrial processes [8].

The International Classification of Diseases (ICD)[1] is a code associated to diseases by their trends and statistical values, established by the [35]. It consists, generally, in a letter and three numbers between 0 and 9. For example, the code for `pneumonia due to staphylococcus` is `J152`. This system comprehends from the values `A00.0` to `Z99.9`, leaving the `U` letter for yet unknown sources. This classification was proposed by Farr at the end of the XIX century [34],

[1] The name of the diseases corresponding to the ICD code is obtained from https://icd.who.int/browse10/2019/en

and its purpose is to understand the causes of morbidity and mortality in order to improve the quality of life of the human population [34].

3. Hypothesis

The contamination levels reported can explain the statistically significant variations and trends of georeferenced data from health reports. Also, it is possible to automatize these relationships in an interactive and free web service.

4. Project objectives

- Stablish a clear and scientifically sustained relationship between air pollution and the human diseases that them can cause.
- Study the impact of the air pollution in the Metropolitan Monterrey Area during 2017 and 2019 in the health population.
- Choose or propose a methodology to calculate the causation between air pollution and associated human diseases.
- Identify the areas where they are more air pollution and stablish a relationship between the people that live nearby those areas.
- Propose a rank to determine which air pollutants have the most direct effect in the human health.
- Propose pollution emission limits, scientifically justified, that can be used by organizations interested in this matter.
- Generate interactive and free tools that let people and specialist know the levels, relationships, risks, etc. related to the air pollutants studied.

5. Research questions

- Is it possible to stablish statistically significant relationships between air pollution and human diseases?
- What are the best methods to detect the causation between air pollution and human diseases?
- How to eliminate the climate and seasonal components from the relationship between air pollution and human diseases?
- Does the environmental and geographical factors can be used to predict the amount of air pollution in an area and time?
- Which air pollutants cause more human diseases?
- It is possible to stablish short (days), medium (months) and large (years) interactions between air pollution and human diseases?

6. Background and related literature

The first correspondence to discuss relationship between air quality and human health was registered during December 1930 in Belgium, when a three-day fog s was declared to be the cause of death of 63 people — during the fog, disease and death were observed to increase, whereas after its dissipation, the figures normalized again [10]. Similarly, in December 1952, a four-day fog in London was associated with four thousand deaths [28]. The majority of the fog-related deaths were related to respiratory diseases.

In 1959 Lawther [24] used a method to quantify the severity of respiratory illness based on their condition (negative numbers indicated they were recovered and positive numbers that they were getting worse). Then, he plotted a time series superposed with concentrations of two pollutants — SO_2 and smoke — and two weather variables — temperature and humidity; he discovered that the pollutants bore a similarity to the severity whereas the climate ones did not.

During the 1970s, short-term exposure to different pollutants was studied [7, 15, 16, 27] with experiments that implied some ethical and legal considerations, focusing on physiological analysis of the subjects. A compilation was published in by Ferris [9] who documented that the

pollutants, in regulated concentrations, had a negative impact on health, primarily on respiratory diseases.

A set of methodologies are used to measure the relationships that are the focus of this study. They are **multiple regression analysis** [17, 18, 20, 25] with variations in the distributions such as logistic regressions [22], **multivariate analysis** [14, 23], **auto regressive models** [19], **causality models** [30] and **case-crossover approximations** [4, 29].

7. Methodology

The study will focus in the **Monterrey Metropolitan Area (MMA)** between 2018 and 2019. The air quality samples in the are taken each hour from 13 sensors over the MMA that measure concentrations of **CO**, **NO**, **NO**$_2$, **O**$_3$, **SO**$_2$, **PM**$_{10}$, **PM**$_{2.5}$, and **atmospheric pressure**, **rainfall**, **relative humidity**, **solar radiation**, **temperature**, **wind velocity and direction**. This data is provided by the *Sistema Integral de Monitoreo Ambiental de Nuevo León* (SIMA) [32]. The Mexican diseases data was obtained from the Mexican *Department of Health* [31] and contains information from all states and municipalities in Mexico such as **date of admission**, **egress date**, **age at the admission**, **gender**, **weight**, **height**, **ICD code upon arrived**, **ICD code upon diagnosis**, and **reason of egress**.

The air quality data needs to be interpolated because it contains imputed records. Different temporal interpolation techniques are used and compared [11]. Also, an spatio-temporal interpolation is performed to obtain the missing data values [26]. Both data sets are processed and converted to georeferenced time series [33] that are stationary [20] in order to establish their relationship by cross correlation [6], multiple regression analysis [3], vector autorregresive approaches, causality models [30] and geographic interactions [5]. The results are ranked by metrics like R^2, the Akaike (AIC), and the Bayesian (BIC) information criteria [2]. Finally, it will be produced a web application that allow general and specialized population to interact with the data and obtain forecasts, interactions and visualizations of the models described.

8. Expected results

Univariate and multivariate analysis are performed on the time series from both the data sources of air quality and human diseases variables. Descriptive statitstical results are expected to inform levels of pollutants and its levels according to local legislations (shown in Figure 1) or timespans where more cases are reported (found in Figure 2). Causal analysis are based in causal diagrams (like the one on Figure 3). The causal tests have confidence intervals where $\alpha = 0.05$. For example, Figure 4 shows the statisticaly significant results of a Granger test on x causing y variables in green. When there is a causal relationship between variables, matemathical models are able to find the relationships between the variables. Here, a vector autorregresive model is used to calculate the interactions between variables and its temporal lags. This results in equations where a variable can be explained from the most significant interactions and its coeficients. As an example, the `J` diseases depends on the interactions

$$Y_J = -1.96a_{\text{J:1}} - 3.81a_{\text{J:6}} - 0.89a_{\text{PM10:14}} - 6.71a_{\text{J:7}} - 1.31a_{\text{PM10:6}} \dots$$

where a are the values of the variables and lags written as subindex in the form `variable:lag`.

9. Significance

The relationship of air pollutants and human diseases have been widely studied. From those studies it is evident that the forecasting that can be produced it is relevant to the scientific community and other decision making agents. Nevertheless, the causation interactions of those factors are primarily studied in focus groups experiments where the ethical concerns limits the exploration. Mathematical causality models can bring a better understanding of the network of interactions between pollutants, diseases and confounding factors like weather conditions such

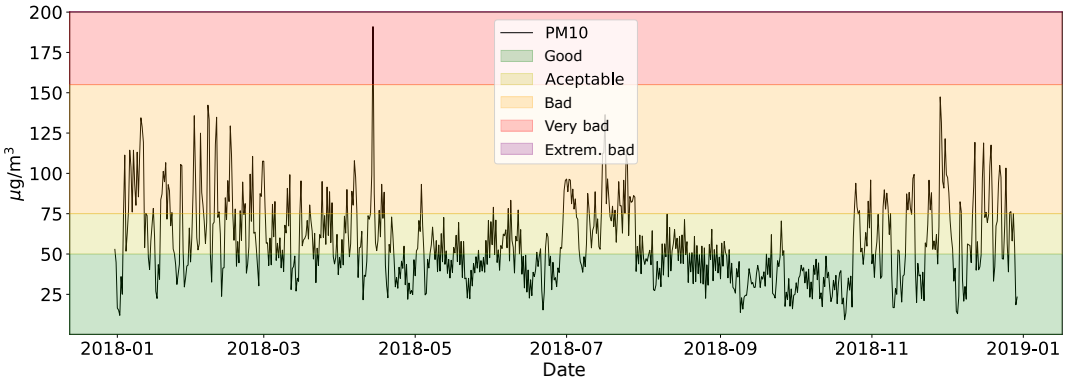


Figure 1: PM_{10} 12-hour mean concentration for the MMA in 2018 displayed over the categorical ranges.

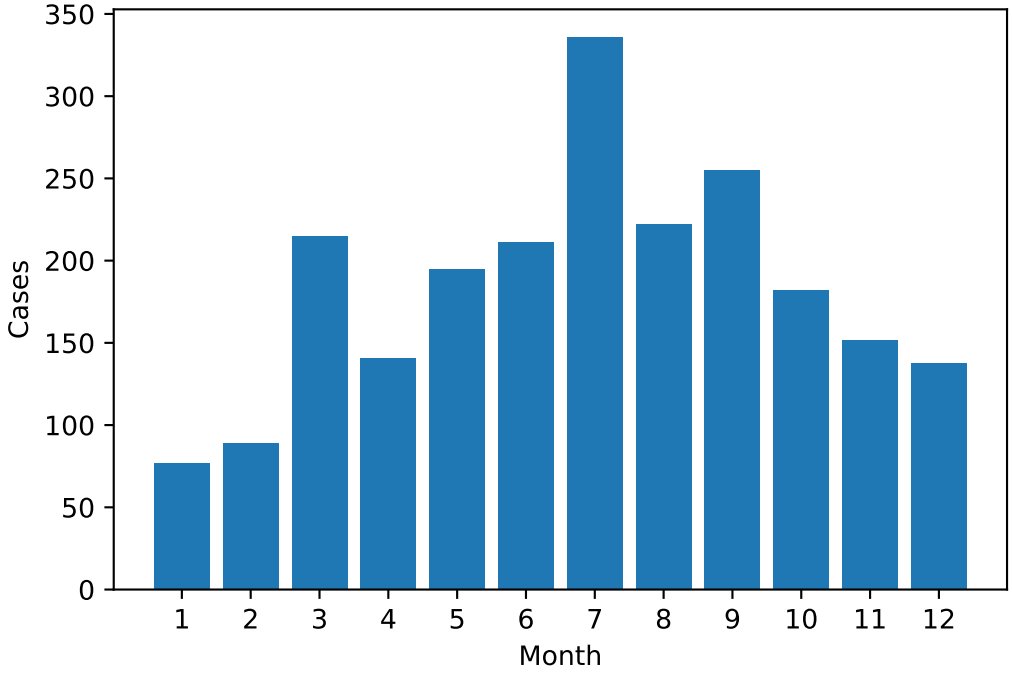


Figure 2: Bar diagram of the preprocessed records of patients from the MMA in 2017.

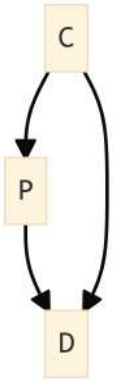


Figure 3: Example of a causal diagram where pollutant P cause disease D, and both are caused by a confounder C.

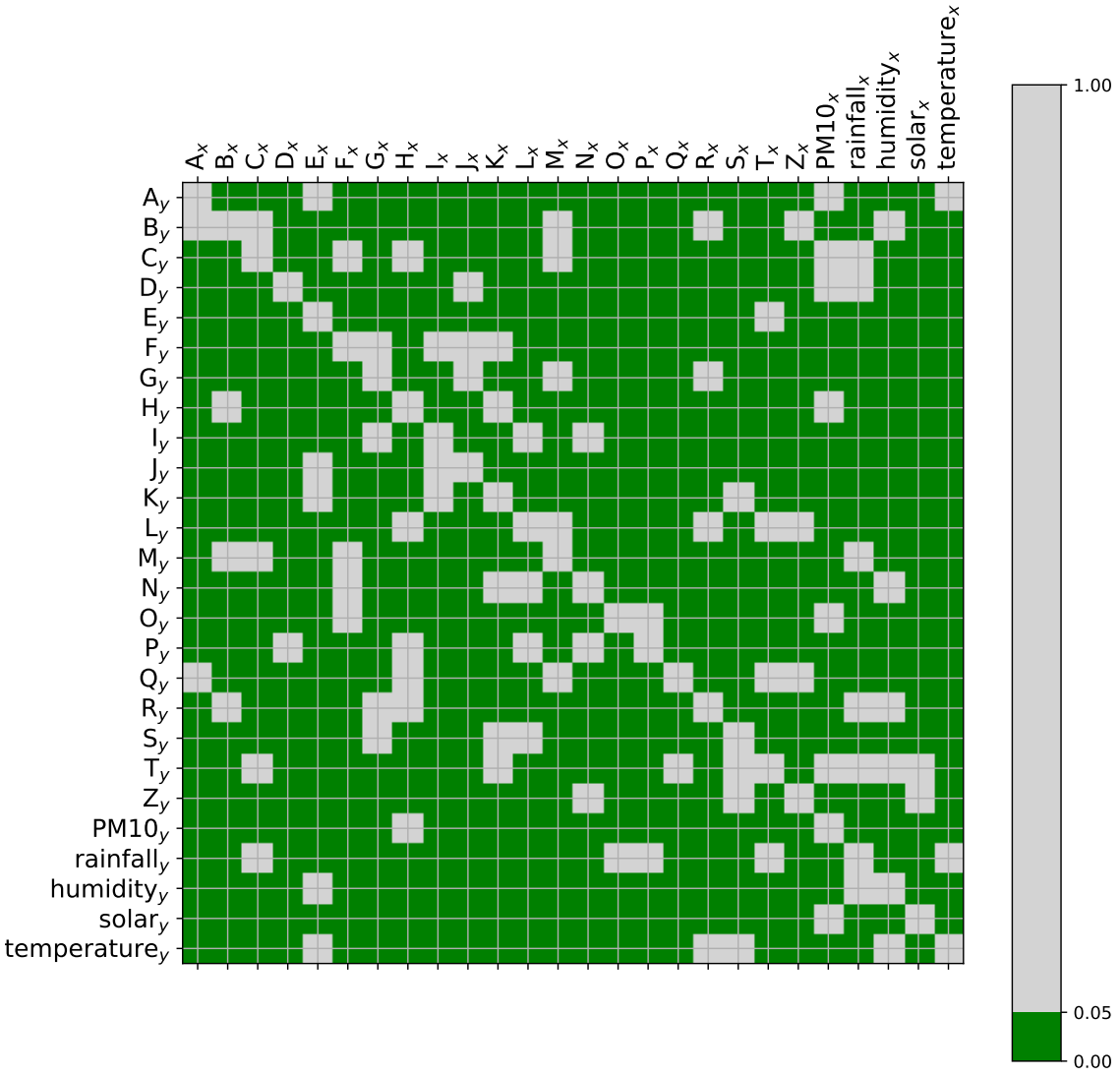


Figure 4: Granger causality test for x (horizontal) variables causing y variables (vertical).

as temperature, atmospheric pressure and so on, supporting the possibility of simulations that extends the ethical concerns limits discused.

10. Project schedule

The Gantt diagram of this project can be accessed in shorturl.at/kvHU0.

References

[1] Hajime Akimoto. Global air quality and pollution. Science, 302(5651), 2003. doi: 10.1126/science.1092666.

[2] Jim Albert. Bayesian computation with R. Springer-Verlag GmbH, July 2007. doi: 10.1007/978-0-387-92298-0. URL https://www.ebook.de/de/product/24984339/jim_albert_bayesian_computation_with_r.html.

[3] Peter Brockwell. Introduction to time series and forecasting. Springer, New York, 2002. doi: 10.1007/978-3-319-29854-2.

[4] Eduardo Carracedo-Martínez, Aurelio Tobías, Marc Saez, Margarita Taracido, and Adolfo Figueiras. Fundamentos y aplicaciones del diseño de casos cruzados. Gaceta Sanitaria, 23: 161–165, April 2009. ISSN 0213-9111. URL http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0213-91112009000200017&nrm=iso.

[5] Alexis Comber and Wen Zeng. Spatial interpolation using areal features: A review of methods and opportunities using new forms of data with coded illustrations. Geography Compass, 13 (10), August 2019. doi: 10.1111/gec3.12465.

[6] Timothy R. Derrick and Joshua M. Thomas. Innovative analyses of human movement, chapter Time Series Analysis: The Cross-Correlation Function, pages 189–205. Kinesiology Publications, 2004.

[7] Richard Ehrlich, J. C. Findlay, J. D. Fenters, and D. E. Gardner. Health effects of short-term inhalation of nitrogen dioxide and ozone mixtures. Environmental Research, 14(2):223–231, October 1977. doi: 10.1016/0013-9351(77)90034-2.

[8] Jes Fenger. Urban air quality. Atmospheric Environment, 33(29), 1999. doi: 10.1016/s1352-2310(99)00290-3.

[9] Benjamin G. Ferris. Health effects of exposure to low levels of regulated air pollutants. Journal of the Air Pollution Control Association, 28(5):482–497, May 1978. doi: 10.1080/00022470.1978.10470621.

[10] J. Firket. Fog along the Meuse Valley. Transactions of the Faraday Society, 32:1192, 1936. doi: 10.1039/tf9363201192.

[11] Milton Friedman. The interpolation of time series by related series. Journal of the American Statistical Association, 57(300):729–757, December 1962. doi: 10.1080/01621459.1962.10500812.

[12] Gobierno de México, Secretaría de Salud. Dirección general de información en salud. https://www.gob.mx/salud/acciones-y-programas/direccion-general-de-informacion-en-salud-dgis, July 2020.

[13] Thad Godish, Wayne T. Davis, and Joshua S. Fu. Air quality. Florida, The address, 5 edition, 2015. doi: 10.1201/9781003032595.

[14] Hans W. Gottinger. Air pollution health effects in the Munich metropolitan area: Preliminary results based on a statistical model. Environment International, 9(3):207–220, January 1983. doi: 10.1016/0160-4120(83)90038-7.

[15] Jack D. Hackney, William S. Linn, Ramon D. Buckley, E. Eugene Pedersen, Sarunas K. Karuza, David C. Law, and D. Armin Fischer. Experimental studies on human health effects of air pollutants. Archives of Environmental Health: An International Journal, 30(8):373–378, August 1975. doi: 10.1080/00039896.1975.10666728.

[16] Jack D. Hackney, Frederick C. Thiede, William S. Linn, E. Eugene Pedersen, Charles E. Spier, David C. Law, and D. Armin Fischer. Experimental studies on human health effects of air pollutants. IV. Short-term physiological and dlinical effects of nitrogen dioxide exposure. Archives of Environmental Health: An International Journal, 33(4):176–181, July 1978. doi: 10.1080/00039896.1978.10667330.

[17] Victor Hasselblad. Comparison of methods for the analysis of panel studies. Technical report, Environmental Protection Agency, June 1978. https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=2000ZPAO.PDF.

[18] A. Hatzakis, K. Katsouyanni, A. Kalandidi, N. Day, and D. Trichopoulos. Short-term effects of air pollution on mortality in Athens. International Journal of Epidemiology, 15(1):73–81, March 1986. doi: 10.1093/ije/15.1.73.

[19] Steven C. Hillmer and William W. S. Wei. Time series analysis: Univariate and multivariate methods., volume 86. JSTOR, March 1991. doi: 10.2307/2289741.

[20] Rob Hyndman. Forecasting: principles and practice. OTexts, Melbourne, 2018. ISBN 9780987507112.

[21] Daniel Jacob and Darrel Winner. Effect of climate change on air quality. Atmospheric Environment, 43(1), 2009. doi: 10.1016/j.atmosenv.2008.09.051.

[22] Edward L. Korn and Alice S. Whittemore. Methods for analyzing panel studies of acute health effects of air pollution. Biometrics, 35(4):795, December 1979. doi: 10.2307/2530111.

[23] Lester B. Lave and Eugene P. Seskin. An analysis of the association between U.S. mortality and air pollution. Journal of the American Statistical Association, 68(342):284–290, June 1973. doi: 10.1080/01621459.1973.10482421.

[24] P. J. Lawther. Chronic bronchitis and air pollution. Journal (Royal Society of Health), 79(1):4–9, January 1959. doi: 10.1177/146642405907900102.

[25] Michael D. Lebowitz. A comparative analysis of the stimulus-response relationship between mortality and air pollution-weather. Environmental Research, 6(1):106–118, March 1973. doi: 10.1016/0013-9351(73)90023-6.

[26] Lixin Li and Peter Revesz. A comparison of spatio-temporal interpolation methods. In Geographic Information Science, pages 145–160. Springer Berlin Heidelberg, 2002. doi: 10.1007/3-540-45799-2_11.

[27] William S. Linn, Ramon D. Buckley, Charles E. Spier, Raymond L. Blessey, Michael P. Jones, D. Armin Fischer, and Jack D. Hackney. Health effects of ozone exposure in asthmatics. American Review Of Respiratory Disease, 117(5):835–843, May 1978. doi: 10.1164/arrd.1978.117.5.835.

[28] W. P. D. Logan. Mortality in the London fog incident, 1952. Lancet (London, England), 1:336–8, February 1953. doi: 10.1016/s0140-6736(53)91012-5.

[29] M. Maclure. The case-crossover design: A method for studying transient effects on the risk of acute events. American journal of epidemiology, 133:144–53, January 1991. doi: 10.1093/oxfordjournals.aje.a115853.

[30] F. Popescu and I. Guyon. Causality in time series. Microtome PublishingBrookline, 2013.

[31] Secretaría de Salud. Egresos Hospitalarios. Datos abiertos, August 2020. URL http://www.dgis.salud.gob.mx/contenidos/basesdedatos/da_egresoshosp_gobmx.html.

[32] Sistema Integral de Monitoreo Ambiental (SIMA), Nuevo León. aire.nl.gob.mx | Home. http://aire.nl.gob.mx/, 2020. [Accesed 14/may/2020].

[33] William Wei. Time series analysis univariate and multivariate methods. Pearson, Boston, 2019. doi: 10.2307/2289741.

[34] World Health Organization. International statistical classification of diseases and related health problems. World Health Organization, Geneva, 2011. ISBN 9789241548342.

[35] World Health Organization. International classification of diseases. https://www.who.int/classifications/icd/en/, 2020.

Tarea 8

Crea una ID ORCID y perfiles de Scopus, Google Scholar y ResearchGate; si ya los tienes, actualízalos. Incluyelos en tu CV junto con tu repositorio de GitHub y actualiza el CV también. Identifica y sigue a por lo menos cinco expertos de tu área en Google Scholar; revisa si tienen cuentas de Twitter y/o GitHub que puedas seguir. Evidencia todas estas actividades en tu entregable de la semana.

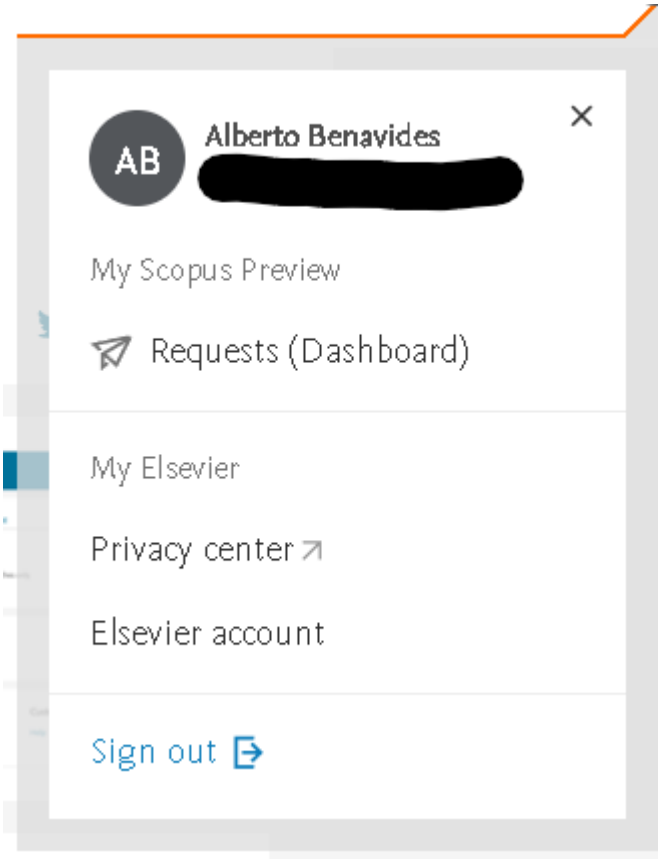
CV

Curriculum vitae

ORCID

Perfil

Scopus



ResearchGate

Perfil

Google Scholar

Perfil

Perfiles destacados

- Graciela González Farías
- Judea Pearl
- Christofer Walters
- Josh Angrist
- Rob J. Hyndman

Tarea 9

Investiga sobre repositorios de *pre-prints* de tu área, sobre la política de *pre-prints* de por lo menos tres revistas indizadas potenciales para publicar los resultados de tu proyecto. Investiga también potenciales mailing lists de tu área particular. Documenta los hallazgos en tu entregable de la semana.

Servidores para *pre-prints*

Recomiendo revisar la entrada de Wikipedia que enlista los repositorios de *pre-prints* y ordenarla por tamaño.

arXiv

- Página web
- Repositorio perteneciente a la Universidad de Cornell
- Áreas de interés: Ciencias computacionales, estadística, ciencia (ingeniería) de sistemas, economía (por lo de pronóstico y causalidad)

Preprints

- Página web
- Forma parte de MDPI, una editorial de acceso abierto multidisciplinaria
- Áreas de interés: Inteligencia artificial, matemáticas computacionales, manipulación de datos, ciencias computacionales, probabilidad y estadística

TechRxiv

- Página web
- Repositorio afiliado a IEEE
- Áreas de interés: Ciencias computacionales, redes, procesamiento computacional, geociencia, análisis de señales

Política para *pre-prints*

Nature

- Página web
- Recomienda publicar *pre-prints*
- Afiliada a Research Square
- Los autores pueden elegir la licencia de su trabajo
- Si una publicación se acepta, el autor debe actualizar el *pre-print* con la información de la publicación
- Notificar a los medios masivos de comunicación que el trabajo aún no ha sido revisado por pares

Elsevier

- Página web
- Se pueden publicar *pre-prints* en cualquier momento
- Para trabajos aceptados,
    - es recomendable asociar *pre-print* con el DOI

    - Actualizar *pre-prints* con el manuscrito aceptado

IEEE

- Página web
- Está bien publicar en TechRxiv o arXiv
- En cuanto se envía un artículo, se deben pasar los derechos de autor a IEEE

Listas de correos

- Kaggle: Ciencia de datos, bases de datos, concursos con participaciones y repositorios abiertos
- Data Machina: Aprendizaje máquina, el último artículo a la fecha trata sobre caudalidad
- KDnuggets: Todo lo que me interesa, Data X (donde X es ciencia, minería, análisis...)
- O'Reilly: Programación, datos e inteligencia artificial

Tarea 10

Prepara en LaTeX con beamer diapositivas sobre tu anteproyecto. Ensaya la presentación grabándote y toma tiempo de cuánto tardaste en presentarla. No incluyas la grabación en el entregable, pero sí un screenshot de su duración.

- Plantilla presentación
- Presentación

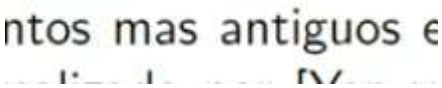
Retroalimentación a presentación

Autoevaluación: 4

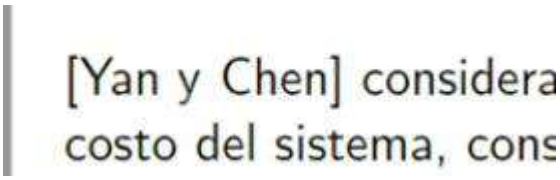
Hubo detalles cuyo nombre olvidé y sí me hubiera servido tener algún detalle mnemotécnico para recordarlo.

Eder Ismael

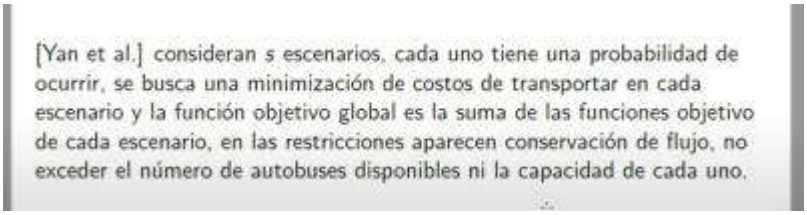
más (varias veces)



Esto debería ir sin corchetes y con número o entre paréntesis, la cita.



Me parece que esto lo puedes escribir como una fórmula. Casi todos la entenderían.



Esto mismo creo que podría aplicarse en el resto de las diapos.

Estaría padre que les vayas dando nombres de variables al as descripciones.

Quizás te puedas apoyar de imágenes con camioncitos, usuarios, puntos de partida, etc.

Incluso un grafo chiquito como ejemplo ayuda a evitar poner tanto texto.

Por lo demás, me parece que tienes una muy buena duración de presentación.

Ah, bueno, al final podrías cortar desde Youtube el final, para que termine donde dicen que ya terminó.

Aquí explican cómo: https://support.google.com/youtube/answer/9057455?hl=en

Luis Ángel

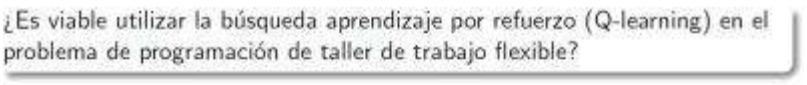
En las diapositivas que tienes sólo texto, no lo leas y, si lo vas a leer, al menos destaca las palabras más relevantes en negritas. Si logras eso, algunas de esas diapositivas podrías cambiarlas por esquemas. La diapositiva con la comparación entre las tres formas de resolver el modelo con tres triángulos y la siguiente están geniales. Luego los ejemplos me parecen abundantes. A parte de eso, creo que hubiera estado bien ver una sección de Trabajo a futuro. A mí me ha servido para constatar mi avance. Este semestre siento que hice poco, pero poder tachar algunos puntos de ahí, da tranquilidad.

Arnoldo

Como título breve, en lugar de ese acrónimo, podrías poner Q-learning Taller Trabajo Flexible.

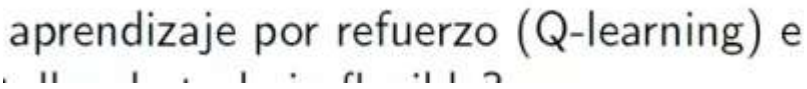
Sólo pusiste punto final al último elemento. Creo que es mejor ponerlo a todos o a ninguno.

No son necesarios, pero me gusta que uses imágenes como ésta para acompañar diapositivas tan vacías. Yo no lo hago, porque no suelo saber qué poner.

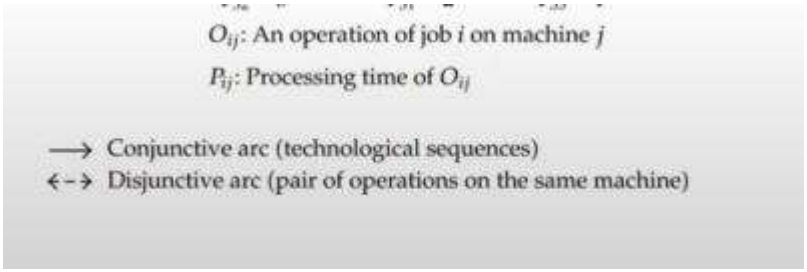


qué

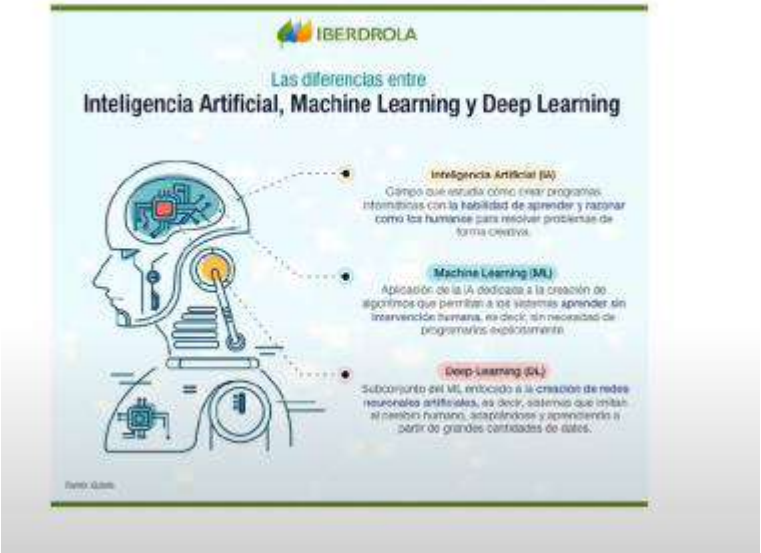
No sé si sea necesario siempre escribir "aprendizaje por refuerzo (Q-learning)", sino sólo Q-learning, quizás.



Igual que yo, andas de pocho, jaja.



Esto es una referencia de una referencia de una referencia. Quizás podrías hacer tú el diagrama.



Títulos sin punto final.

Tiempo y recursos.

Si esto está centrado, mejor ponlo alineado a la izquierda.

Huang *et al.* (2022) presentan el método clasificación de corte (Cut Ranking) para seleccionar los cortes en un problema de ramificación y corte para programación entera mixta (mixed-integer programming).

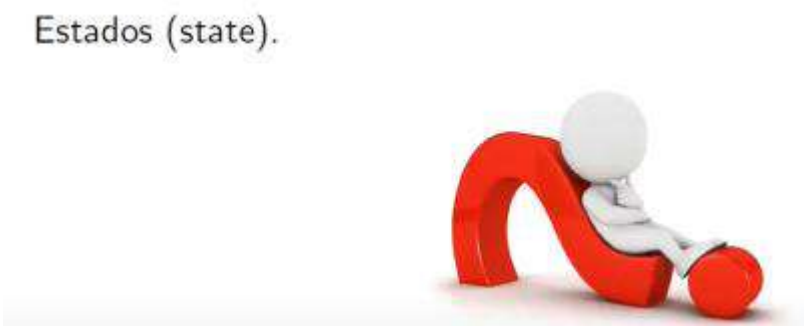
Todas las diapositivas que planteas como trabajo relacionado, podrías resumirlas en un cuadro.

Puedes hacer este diagrama con una mejor calidad.

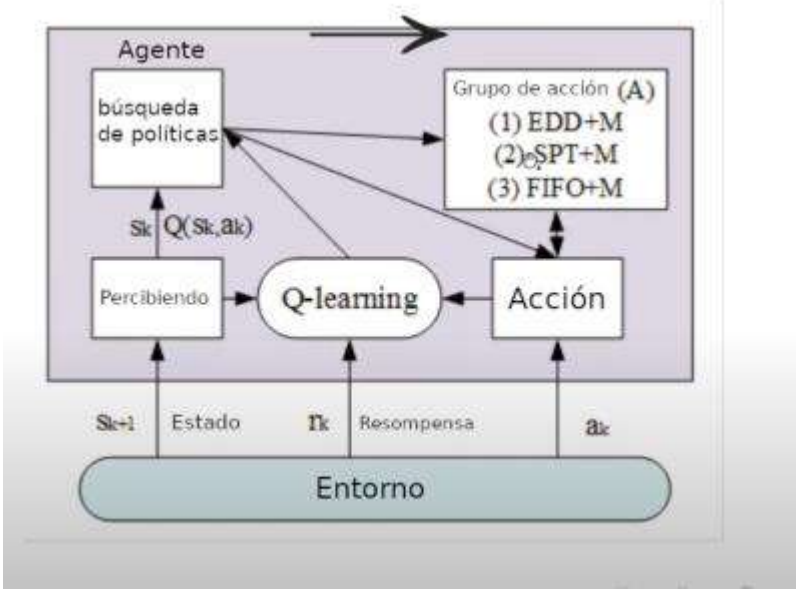
Usa un entorno de algoritmo para esto:

- Inicializar $Q(s,a)$ arbitrariamente
- Establecer el parámetro γ y α
- Repetir (para cada episodio):
  1. Inicializar s
  2. Repetir (para cada paso del episodio):
     - Elegir a desde s usando una de las políticas de Q.
     - Tomar la acción a, observar r, s'
     - $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma\, max(Q(s',a') - Q(s,a))]$
     - $s \leftarrow s'$
     - Hasta que s' sea terminal.

Ahí ya no le vi la relación a esta misma imagen de antes, jaja.

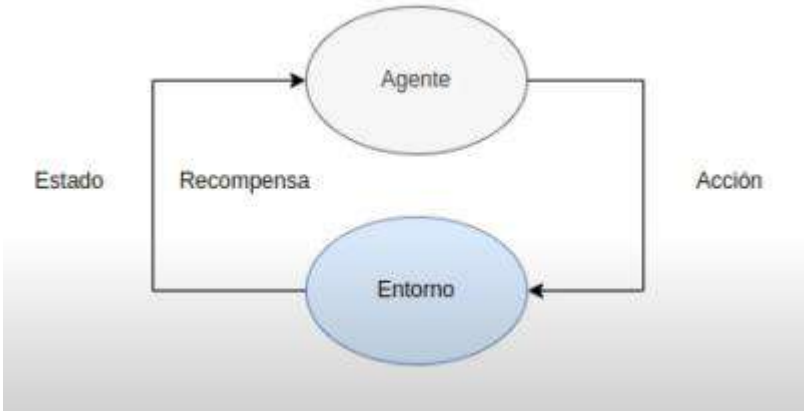


Puedes hacer este diagrama con una mejor calidad.



1. [15:45]

También éste:



Mayúscula al inicio de cada oración, puntos finales consistentes.