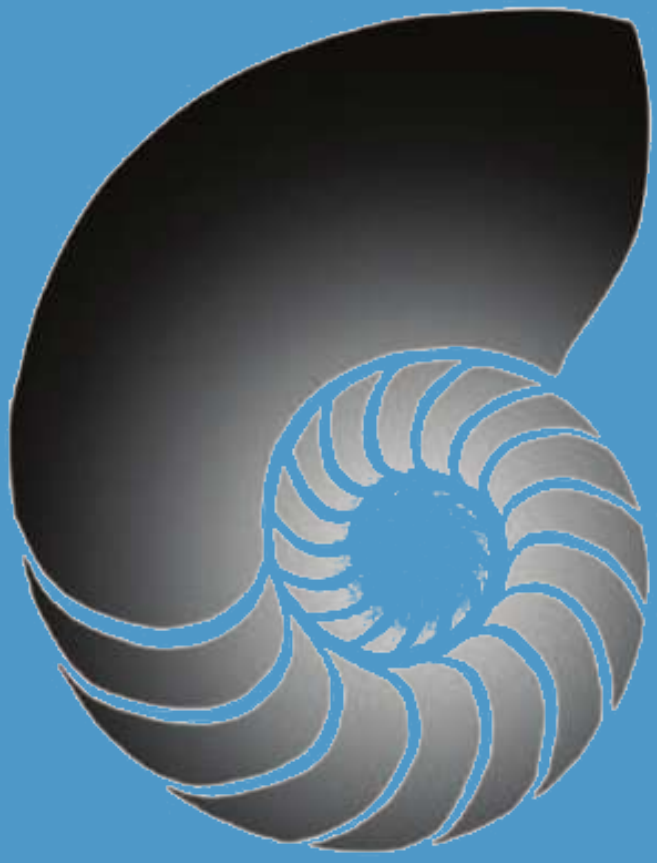
AÑO VI - NÚMERO 6
DICIEMBRE 2020

SMCCA



BOLETÍN

SOCIEDAD MEXICANA DE COMPUTACIÓN
CIENTÍFICA Y SUS APLICACIONES

BOLETÍN

SOCIEDAD MEXICANA DE COMPUTACIÓN CIENTÍFICA
Y SUS APLICACIONES

AÑO VI - NÚMERO 6

DICIEMBRE 2020

El Boletín Sociedad Mexicana de Computación Científica y sus Aplicaciones A. C. (SMCCA), Año VI, No. 6, Diciembre 2020, es una publicación anual editada por la Sociedad Mexicana de Computación Científica y sus Aplicaciones A. C., calle Luis Horacio Salinas, 545, Col. Valle de Morelos, Saltillo, Coahuila, C.P. 25013, Tel. (844) 410 1242, www.smcca.org.mx.

Editor responsable: Gerardo Tinoco Guerrero. Reserva de Derechos al Uso Exclusivo No. 04 - 2017 - 103114330600 - 203, ISSN: 2594-0457, ambos otorgados por el Instituto Nacional de Derechos de Autor. Responsable de la última actualización de este Número, Gerardo Tinoco Guerrero, Avenida Francisco J. Múgica S/N, Ciudad Universitaria, Edificio B, Morelia, Michoacán, C.P. 58030, fecha de la última modificación, 21 de diciembre de 2020.

Suscripciones al Boletín vía correo electrónico: smcca@smcca.org.mx

CONTENIDO

Artículos

Información

Carta de bienvenida

La Sociedad Mexicana de Computación Científica y sus Aplicaciones, A.C. (SMCCA) y el Comité Editorial, les dan una cordial bienvenida a la sexta edición del Boletín electrónico anual de la SMCCA, el cual tiene como objetivo mantenerlos informados de las actividades realizadas por la SMCCA y sus asociados. En el Boletín se publican noticias, eventos, artículos de divulgación y de investigación de alto nivel en el área de Cómputo Científico y sus Aplicaciones, así como resúmenes de las mejores tesis de Licenciatura en Matemáticas Aplicadas.

En esta sexta edición del boletín se presenta una breve reseña de las actividades de Epidemias y Matemáticas, llevada a cabo entre abril y julio del presente año en las plataformas Bluejeans y Zoom y transmitido en vivo por Facebook Live: https://www.facebook.com/enoan. Además, se presentan cuatro artículos de investigación por solicitud y convocatorias de futuros eventos a realizarse.

La SMCCA agradecerá que ante el interés que surja en los lectores en los temas que se presenten en nuestra publicación, éstos se conviertan en usuarios asiduos, así como en miembros activos de nuestra Sociedad. La información del registro de membresías a la SMCCA la pueden consultar en el Módulo de Registro de nuestra página http://www.smcca.org.mx.

Justino Alavez Ramírez
Presidente
Sociedad Mexicana de Computación Científica y sus Aplicaciones, A.C.

La ENOAN en tiempos del COVID-19

Humberto Madrid de la Vega
Guilmer F. González Flores
Jesús López Estrada

El COVID-19 ha trastornado todo, en especial nuestro ambiente social con severas implicaciones en los sectores de la salud, la educación y la economía. El 11 de marzo de 2020, la Organización Mundial de la Salud (OMS) declaró el brote del Coronavirus COVID-19 como una pandemia mundial. El brote evolucionó rápidamente, por lo que los gobiernos del mundo y en nuestro caso el de México, a través del Consejo de Salubridad General del Gobierno Federal, declararon una emergencia sanitaria, así como las medidas a seguir por todas las instituciones del país, para detener la propagación del COVID-19. Ante esta situación, los órganos directivos de la SMCCA, acordaron suspender la realización de la XXIX Escuela Nacional de Optimización y Análisis Numérico (ENOAN) durante el 2020 y posponer su realización para el 2021. Ante la llegada del virus, en los medios de comunicación se comenzó a hablar de modelos matemáticos y sus diversas predicciones. También algunos estudiantes seguidores de la ENOAN comenzaron a hacer preguntas sobre ellos a varios integrantes de la SMCCA. Por lo que iniciamos la aventura de organizar de manera emergente una serie de conferencias y cursos denominada como "**Epidemias y Matemáticas**". Decimos que fue aventura porque, en ese momento, nuestra experiencia en el manejo de tecnologías de comunicación era escasa y al inicio solamente teníamos aseguradas unas cuantas actividades. Para realizar las actividades, aprovechando que pertenecemos a la SMCCA, se usó la cuenta de Facebook de ENOAN, https://www.facebook.com/enoan, y se solicitó el apoyo del Departamento de Matemáticas de la Facultad de Ciencias de la UNAM, que permitió utilizar las cuentas institucionales de las plataformas *Bluejeans* y *Zoom* para tal fin. Las actividades también se transmitieron simultáneamente por *Facebook Live* y publicadas en *YouTube* en el canal *Epidemias y matemáticas*.

La finalidad de las actividades fueron de divulgación, pensando principalmente en estudiantes de licenciatura y posgrado de carreras de Ciencias e Ingenierías, aunque algunas de estas actividades tuvieron un público más amplio, tal fue el caso de la conferencia del Dr. Antonio Lazcano Araujo, distinguido biólogo de renombre internacional. El objetivo no era simplemente la modelación de la epidemia sino también la de proporcionar información sobre el virus, para lo cual se invitó a personas expertas en la evolución de los virus y de la virología. En relación a la modelación, se impartieron cursos y se presentaron conferencias sobre modelos de epidemias en general y sobre el COVID-19 y algunas conferencias sobre otras temáticas relacionadas con la pandemia. Para realizar estas actividades recurrimos a especialistas de varias instituciones de educación superior nacionales y extranjeras. Estos especialistas participan activamente en proyectos relacionados con esta pandemia.

Actividades

14 de Abril. Las actividades iniciaron con la mesa redonda titulada ***Epidemias y Matemáticas***, con la participación de la Dra. María de Lourdes Esteva Peralta y del Dr. Jesús López Estrada, ambos del Departamento de Matemáticas de la Facultad de Ciencias de la UNAM, así como la participación del Dr. Jorge X. Velasco Hernández del Instituto de Matemáticas de la UNAM, Unidad Juriquilla.

16, 21 y 26 de abril. Curso ***Introducción a la epidemiología matemática***, impartido por la Dra. María de Lourdes Esteva Peralta, Departamento de Matemáticas, Facultad de Ciencias, UNAM.

Le siguieron las siguientes conferencias:

30 de abril. ***Origen y evolución de los coronavirus***, por el Dr. Antonio Eusebio Lazcano Araujo Reyes, Miembro del El Colegio Nacional y de la Facultad de Ciencias, UNAM.

8 de mayo. ***Modelos de enfermedades infectocontagiosas con estructura de edades***, por el Dr. Cruz Vargas De-León, Maestría en Matemáticas Aplicadas de la Universidad Autónoma de Guerrero.

14 de mayo. ***Virología de los coronavirus: en busca de una vacuna para el COVID-19***, por el Dr. José Ángel Regla Nava, Research Fellow. La Jolla Institute for Immunology, San Diego (USA).

28 de mayo. ***Modelos matemáticos para la reapertura de la economía nacional durante la pandemia***, por el Dr. Ricardo Lino Mansilla Corona, Programa de Investigación Ciencia y Tecnología, Centro de Investigación Interdisciplinarias en Ciencias y Humanidades, UNAM.

3 de junio. ***Entendamos el COVID-19 en México***, por el Dr. Octavio Reymundo Miramontes Vidal, Departamento de Sistemas Complejos, Instituto de Física, UNAM.

11 de junio. ***Sistemas de auxilio al diagnóstico médico de COVID-19***, por el Dr. Boris Escalante Ramírez, Centro Virtual de Computación, UNAM, M.I. José Carlos Moreno Tagle, Facultad de Ingeniería, UNAM, y Fís. Bio. Steve Avendaño García, Facultad de Ciencias, UNAM.

16 de junio. ***Modelos matemáticos para COVID-19***, por el Dr. Roberto Alonso Sáenz Casas, Facultad de Ciencias, Universidad de Colima.

18 de junio. ***COVID-19: Modelos matemáticos, brotes locales y dispersión geográfica***, por el Dr. Gustavo Cruz Pacheco, Departamento de Matemáticas y Mecánica, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM.

23 de junio. ***Patofisiología y epidemiología del COVID-19 y algunos detalles frecuentemente ignorados***, por el Dr. Marco Arieli Herrera Valdez, Laboratorio de Fisiología de Sistemas, Departamento de Matemáticas, Facultad de Ciencias, UNAM.

Cerramos el ciclo de actividades con una terna de cursos, o suite, denominado **Epidemias, modelaje y aspectos numéricos**. La idea de esta última actividad fue presentar modelos para el estudio del COVID-19 en México, requiriéndose para ello, familiaridad con algunos modelos básicos de epidemiología matemática y solución numérica de ecuaciones diferenciales ordinarias.

25 y 30 de junio y 2 de julio. ***Fundamentos de la modelación matemática: un acercamiento a la epidemiología matemática***, por el Dr. Faustino Sánchez Garduño, Departamento de Matemáticas, Facultad de Ciencias, UNAM.

9, 14 y 16 de julio. ***Solución numérica de ecuaciones diferenciales ordinarias***, por la M. en C. María Lourdes Velasco Arregui, Departamento de Matemáticas, Facultad de Ciencias, UNAM.

23, 28 y 30 de julio. ***COVID-19 y la estimación numérica de parámetros en EDOs***, por el Dr. Jesús López Estrada, Departamento de Matemáticas, Facultad de Ciencias, UNAM.

En resumen, se llevó a cabo una mesa redonda, se impartieron 4 cursos y 9 conferencias. Fueron cuatro meses y medio de mucha actividad pero con la satisfacción de llevar información sobre esta pandemia a una buena cantidad de estudiantes, profesores, investigadores y a profesionistas de varias disciplinas, tanto en el ámbito nacional como en otros países.

Agradecemos a los investigadores que accedieron a presentar sus trabajos, también a aquellos que colaboraron proporcionando contactos y a los que difundieron la información sobre las actividades. Especialmente reconocemos la colaboración de la Dra. María de Lourdes Esteva quien fue pieza fundamental para dar los primeros pasos en esta aventura y nos proporcionó contactos muy importantes.

Cobertura de las actividades

La visibilidad de la ENOAN mediante esta actividad ha tenido un incremento considerable; de abril hasta julio, el público cautivo de la página https://www.facebook.com/enoan se incrementó 850 por ciento, al día de hoy, contamos con 4281 seguidores. La página es una referencia importante en países como Perú, Colombia, Ecuador, Bolivia, Estados Unidos, Chile, Guatemala, Honduras, Argentina y Portugal, por mencionar algunos. El 72% del público se encuentra entre 18 y 34 años y siendo el grupo de las mujeres el 49.5% de la audiencia total contra 50.5% de hombres. Se estima que el alcance es de 280.5 mil, con un promedio de 10.5 mil respuestas a cada publicación. Las transmisiones de las conferencias y cursos, así como documentos de apoyo, permanecen en la página. Al día de hoy, una conferencia cuenta con mas de 37 mil visualizaciones, sin contar ésta, el promedio es de 2.8 mil visualizaciones entre los 21 eventos restantes. Adicionalmente, se tiene una página en YouTube: Epidemias y matemáticas, en la cual se han subido los videos grabados de los eventos, esa página cuenta actualmente con 313 suscriptores.

Los videos pueden consultarse en la página de la ENOAN:
https://www.facebook.com/enoan
y en el canal de YouTube:
https://www.youtube.com/channel/UCyDnOkYMND4UM9lj_juodrA

Consideraciones finales

La experiencia obtenida con esta actividad, abre la posibilidad de desarrollar eventos en línea tales como webinars, talleres, coloquios, cursos, mesas redondas y la ENOAN 2021 misma. Recordemos que uno de los objetivos fundamentales de esta sociedad es promover y organizar toda clase de actividades académicas orientadas a la promoción y difusión de las matemáticas aplicadas y el cómputo científico en el contexto nacional.

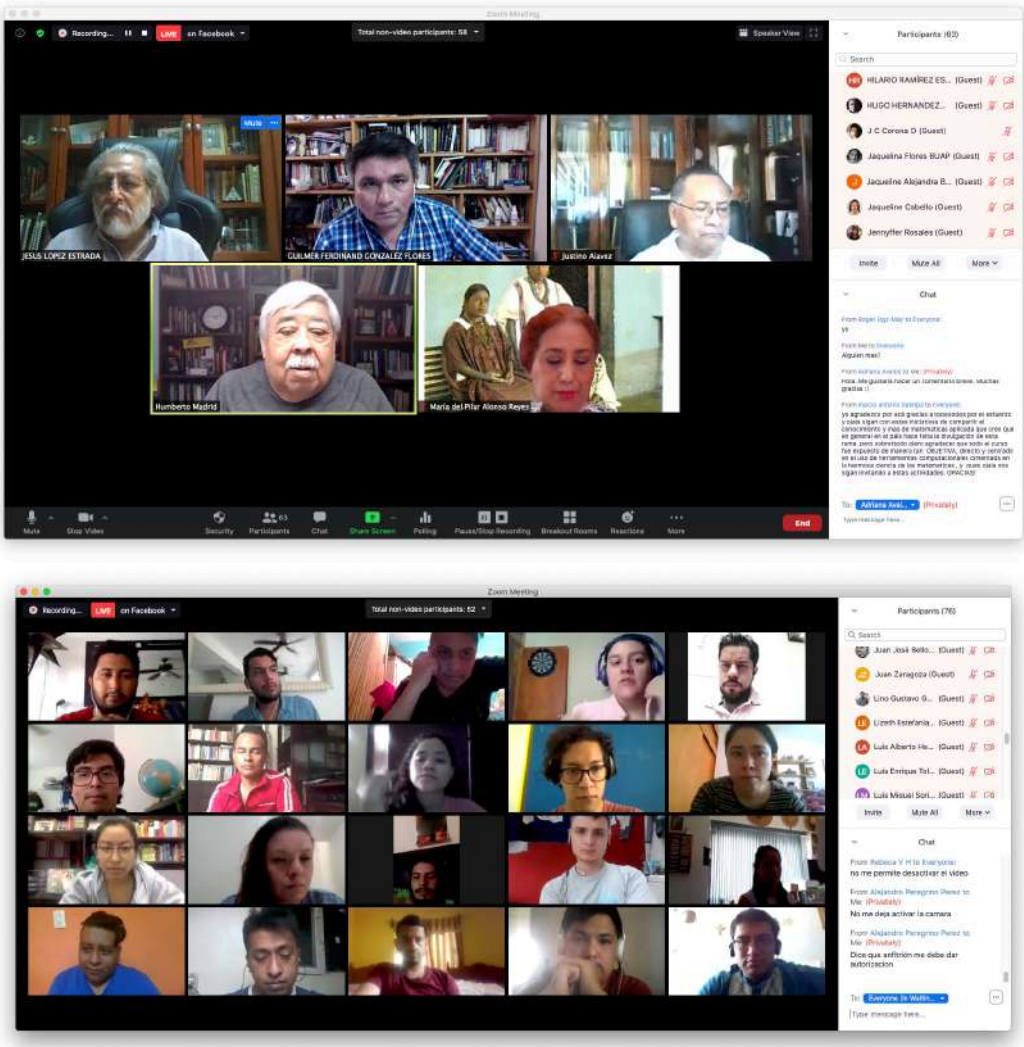


Figura 1: Algunas capturas de pantalla del ciclo de actividades.

Formulación variacional de un modelo continuo en EDP's de la angiogénesis inducida por tumor

Ana Kristhel Esteban López[1], Justino Alavez-Ramírez[1,*]
[1]División Académica de Ciencias Básicas, UJAT
*Autor por correspondencia: justino.alavez@ujat.mx

Resumen

En este trabajo retomamos el modelo matemático de Anderson y Chaplain (1998) formado por tres ecuaciones diferenciales, que describe la respuesta migratoria inicial de las células endoteliales al factor angiogénico tumoral (TAF) y a la fibronectina. Aplicamos el método variacional a los problemas de contorno mixto homogéneo y no homogéneo, que surgen del modelo antes citado, para demostrar la existencia, unicidad y dependencia continua respecto de los datos de la solución débil de dichos problemas.

Palabras clave: Células endoteliales, TAF, fibronectina, problema mixto homogéneo, formulación variacional.

1 Introducción

El cáncer no es una enfermedad nueva. Papiros egipcios que datan de aproximadamente el año 1600 a. C. ya la describían [6]. Sin embargo, fue hasta la invención del microcospio en el siglo XIX que comenzó el estudio patológico moderno de cáncer [1]. El cáncer se ha vuelto un problema global de salud pública. Según estimaciones de la Organización Mundial de la Salud (OMS) en 2015, el cáncer es la primera o la segunda causa de muerte antes de los 70 años en 91 de 172 países, y ocupa el tercer o cuarto lugar en 22 países adicionales [3]. Algunos factores de riesgo de cáncer se pueden vincular estrechamente con la herencia, los productos químicos, las radiaciones ionizantes, las infecciones o virus y los traumas. Los investigadores estudian cómo estos diferentes factores pueden interactuar de una manera multifactorial y secuencial para producir tumores malignos [1]. Los tumores pueden ser benignos o malignos. La células de los tumores malignos presentan dos características que la distinguen de las normales: se reproducen de manera descontrolada, y son capaces de invadir y colonizar tejidos y órganos distantes, en lugares donde normalmente no pueden crecer [7]. La combinación desafortunada de estas características es la que hace tan peligrosa y mortal a la mayoría de las formas del cáncer. Afortunadamente, existen muchos modelos matemáticos que permiten describir, bajo ciertas condiciones, la evolución de las células cancerígenas y el efecto que sobre ellas produce una terapia elegida con la intención de eliminar o, al menos, contener el crecimiento de un tumor [18].

La hipótesis de que el crecimiento tumoral depende de la angiogénesis fue introducida por primera vez en 1971 [4]. El término angiogénesis significa literalmente formación de nuevos vasos sanguíneos a partir de una vasculatura existente. En donde las células endoteliales (CEs) migran y proliferan, organizándose hasta formar estructuras tubulares que eventualmente se unirán, para finalmente madurar en vasos sanguíneos estables [10]. Anderson y Chaplain (1998) [2], presentaron un modelo matemático continuo que describe la formación de la red de brotes capilares en respuesta a estímulos químicos (factores angiogénicos tumorales, TAF) suministrados por un tumor sólido. El modelo también tiene en cuenta las interacciones esenciales entre las células endoteliales y la matriz extracelular mediante la inclusión de la macromolécula de la matriz fibronectina.

Todos los tejidos y órganos contienen una mezcla de células y componentes no celulares, que forman redes bien organizadas llamadas matrices extracelulares (MEC). Las MEC proporcionan no sólo estructuras físicas en las que se incrustan las células, sino que también regulan muchos procesos celulares, como el crecimiento, la migración, la diferenciación, la supervivencia, la homeostasis y la morfogénesis [17]. La MEC está conformada por una gran variedad de moléculas, las cuales interaccionan entre sí, generando una estructura tridimensional a la

cual las células se adhieren ya sea por receptores específicos o ligandos. Las macromoléculas que constituyen la MEC incluyen, entre otras, a la familia de las colágenas, que son las responsables de la resistencia mecánica de los tejidos conjuntivos, la elastina que le confiere cualidades de flexibilidad y elasticidad, proteínas de adhesión como fibronectinas y lamininas y los proteoglicanos que son esenciales para la adhesividad [7]. Las fibronectinas constituyen una familia de glucoproteínas multifuncionales, que se pueden encontrar tanto en forma insoluble, formando parte de la MEC, como en forma soluble circulando en el plasma [7]. Las células pueden ensamblar fibronectina soluble derivado de fibronectina en plasma en fibras. Alternativamente, las células pueden producir su propia fibronectina que es secretada y formada en fibras [17]. La fibronectina desempeña un papel fisiológico importante no sólo en la formación de la matriz extracelular, sino además en la adhesión y migración celular, en la coagulación de la sangre y en la cicatrización de las heridas, entre otras [9].

En cuanto al contenido de este artículo, en la sección 2, retomamos el modelo continuo dado por Anderson y Chaplain (1998) [2], el cual consiste en un sistema de ecuaciones diferenciales que describen la respuesta migratoria inicial de las células endoteliales al TAF y la fibronectina. En la sección 3, realizamos la discretización en el tiempo mediante el método de diferencias finitas, que nos permite transformar el problema continuo formado por tres ecuaciones diferenciales parciales en un problema discreto en el que en cada tiempo se debe resolver una sola ecuación diferencial parcial de segundo orden de tipo elíptico, a la cual le aplicamos el método variacional al problema de contorno mixto homogéneo, que surge del modelo, para demostrar la existencia, unicidad y dependencia continua respecto de los datos iniciales de su solución débil.

2 Modelo matemático

Partimos del modelo matemático de Anderson y Chaplain (1998) [2]. Este modelo describe cómo las células endoteliales que emergen de un vaso padre, responden y migran a través de la motilidad aleatoria, de la quimiotaxis a través de gradientes del factor angiogénico tumoral (TAF) liberado por el tumor, y de la haptotaxis a través de gradientes de fibronectina en la matriz extracelular. En este modelo, la densidad de las células endoteliales (en o cerca de una punta de brote capilar) por unidad de área se denota por n, que está influenciada por la motilidad aleatoria, la quimiotaxis y la haptotaxis. La concentración de TAF y la concentración de fibronectina están representadas por c y f, respectivamente. La quimiotaxis está en respuesta a los gradientes de TAF y la haptotaxis está en respuesta a los gradientes de fibronectina. El sistema de ecuaciones diferenciales parciales en el modelo está dado por

$$\begin{aligned} \frac{\partial n}{\partial t} &= D_n \nabla^2 n - \nabla \cdot \left(\frac{\chi_0 k_1}{k_1 + c} n \nabla c \right) - \nabla \cdot (\rho_0 n \nabla f) \\ \frac{\partial f}{\partial t} &= wn - \mu n f \\ \frac{\partial c}{\partial t} &= -\lambda n c \end{aligned} \tag{1}$$

donde, como se dijo antes, n es la densidad de las células endoteliales, c es la concentración de TAF y f es la concentración de fibronetina. D_n es el coeficiente de motilidad aleatoria de la célula, χ_0 es el coeficiente quimiotáctico, ρ_0 es el coeficiente haptotáctico, k_1, λ, w y μ son constantes positivas.

Considerando una geometría bidimensional en la que las ecuaciones del modelo están definidas en el dominio espacial cuadrado de lado L (un cuadrado de tejido corneal), $[0, L] \times [0, L]$, se reescala la distancia del vaso parental al tumor con $\tilde{x} = x/L$ y $\tilde{y} = y/L$, con lo cual el vaso parental permanece en $x = 0$ y el tumor en $x = 1$. Adimensionalizando el tiempo con $\tau = L^2/D_c$, donde D_c es el coeficiente de difusión del TAF; la densidad inicial de células endoteliales con n^0 y las concentraciones iniciales del TAF y fibronectina con c^0 y f^0, respectivamente. Haciendo el cambio de variables:

$$\tilde{c} = \frac{c}{c^0}, \quad \tilde{f} = \frac{f}{f^0}, \quad \tilde{n} = \frac{n}{n^0}, \quad \tilde{t} = \frac{t}{\tau}, \tag{2}$$

se obtiene el siguiente sistema adimensionalizado:

$$\begin{aligned} \frac{\partial \tilde{n}}{\partial \tilde{t}} &= D \nabla^2 \tilde{n} - \nabla \cdot \left(\frac{\chi}{1 + \alpha \tilde{c}} \tilde{n} \nabla \tilde{c} \right) - \nabla \cdot \left(\rho \tilde{n} \nabla \tilde{f} \right) \\ \frac{\partial \tilde{f}}{\partial \tilde{t}} &= \beta \tilde{n} - \gamma \tilde{n} \tilde{f} \\ \frac{\partial \tilde{c}}{\partial \tilde{t}} &= -\eta \tilde{n} \tilde{c}, \end{aligned} \tag{3}$$

donde $D = D_n/D_c$, $\chi = \chi_0 c^0/D_c$, $\alpha = c^0/k_1$, $\rho = \rho_0 f^0/D_c$, $\beta = \omega L^2 n^0/f^0 D_c$, $\gamma = \mu L^2 n^0/D_c$ y $\eta = \lambda L^2 n^0/D_c$. El modelo adimensionalizado (3) con condiciones iniciales, parámetros y condiciones de contorno apropiadas, es el que usan Anderson y Chaplain [2] para simular la migración (evolución) de las células endoteliales hacia la fuente de la señal tumoral. Es muy importante notar que solamente la primera ecuación del sistema (3), es la que contiene derivadas parciales espaciales, mientras que la segunda y tercera ecuación solo contienen la primera derivada en el tiempo. Así que al discretizar estas dos últimas ecuaciones con respecto al tiempo t, darán lugar a ecuaciones algebraicas. Por otro lado, el dominio espacial del modelo adimensionalizado (3) es el conjunto $\Omega = \{(x,y) \in \mathbb{R}^2 : 0 < x < 1,\ 0 < y < 1\}$ y $\Gamma = \partial\Omega = \cup_{m=1}^{4} \Gamma_m$, donde $\Gamma_1 = \{(x,y) \in \Gamma : 0 \leq x < 1,\ y = 0\}$, $\Gamma_2 = \{(x,y) \in \Gamma : x = 1,\ 0 \leq y < 1\}$, $\Gamma_3 = \{(x,y) \in \Gamma : 0 < x \leq 1,\ y = 1\}$ y $\Gamma_4 = \{(x,y) \in \Gamma : x = 0,\ 0 < y \leq 1\}$.

3 Discretización en el tiempo y formulación variacional del modelo adimensionalizado

Discretizamos las derivadas parciales que aparecen en el lado izquierdo del sistema (3), usando un esquema progresivo para aproximar las derivadas parciales con respecto al tiempo t. Para simplificar notaciones, escribiremos n, f y c en lugar de $\tilde{n}$, $\tilde{f}$ y $\tilde{c}$, respectivamente, teniendo siempre en cuenta las relaciones dadas en (2). Así tenemos por ejemplo, $\frac{n(x,y,t+\Delta t)-n(x,y,t)}{\Delta t} \cong \frac{\partial n(x,y,t)}{\partial t}$, $\Delta t \neq 0$. Consideramos un intervalo de tiempo $[0,T]$, N entero positivo, $\Delta t = T/N$ y puntos $t_k = k\Delta t$, $k = 0, 1, \ldots, N$, [5, 13]; y denotemos por $n^k = n(x,y,t_k)$, la primera ecuación del sistema (3) se discretiza en el tiempo como

$$n^k - \Delta t D \nabla^2 n^k = n^{k-1} - \Delta t \nabla \cdot \left(\frac{\chi}{1 + \alpha c^{k-1}} n^{k-1} \nabla c^{k-1} \right) - \Delta t \nabla \cdot \left(\rho n^{k-1} \nabla f^{k-1} \right) \tag{4}$$

para $k = 1, 2, \ldots, N$ y n^0 dado. Similarmente, para la segunda y tercera ecuación de (3), de donde resulta el esquema discreto:

$$\begin{aligned} -\Delta t D \nabla^2 n^k + n^k &= F \\ f^k &= \Delta t \beta n^{k-1} + \left(1 - \Delta t \gamma n^{k-1}\right) f^{k-1} \\ c^k &= \left(1 - \Delta t \eta n^{k-1}\right) c^{k-1}, \end{aligned}$$

o bien si utilizamos un esquema de discretizacion en el tiempo hacia adelante, resulta el esquema:

$$\begin{aligned} \Delta t D \nabla^2 n^k + n^k &= F \\ f^k &= \frac{f^{k-1} + \Delta t \beta n^k}{1 + \Delta t \gamma n^k} \\ c^k &= \frac{c^{k-1}}{1 + \Delta t \eta n^k}, \end{aligned}$$

para $k = 1, 2, \ldots, N$, y n^0, f^0, c^0 dados, donde $F = n^{k-1} - \Delta t \nabla \cdot \left(\frac{\chi}{1+\alpha c^{k-1}} n^{k-1} \nabla c^{k-1} \right) - \Delta t \nabla \cdot \left(\rho n^{k-1} \nabla f^{k-1} \right)$.

Condiciones iniciales. El primer evento de angiogénesis inducida por tumor es la secreción de TAF por las células tumorales, que se difunde en la matriz extracelular y se establece un gradiente de concentración entre el tumor y el vaso progenitor. Anderson y Chaplain (1998) [2] consideran un campo de concentración inicial de TAF de la forma $c(x,y,0) = \exp(-(1-x)^2/\epsilon_1)$, $(x,y) \in \Omega$, $\epsilon_1 = 0.45$. Una vez que el TAF ha alcanzado el vaso sanguíneo principal, las células endoteliales dentro del vaso se forman en grupos que finalmente se convierten en brotes. Los autores antes citados suponen que inicialmente se forman tres grupos a lo largo del eje y $(0 \leq y \leq 1)$ en $x \approx 0$, con el tumor localizado en $(1, 1/2)$ y el vaso progenitor de las células endoteliales también en el eje y $(0 \leq y \leq 1,\ x = 0)$. El grupo inicial de células endoteliales se genera con $n(x,y,0) = \exp(-x^2/\epsilon_3)\,\mathrm{sen}^2(7\pi y)$ si $y \in [0.15, 0.28] \cup [0.43, 0.58] \cup [0.72, 0.85]$ y $n(x,y,0) = 0$ en otro caso; $\epsilon_3 = 0.001$. Una vez que las células endoteliales han sido activadas por el TAF, degradan la lámina basal del vaso principal. Este daño inicial resulta en una mayor capacidad de permeabilidad del vaso que permite que la fibronectina plasmática de la sangre se escape del vaso parental y se difunda en el tejido corneal. Posteriormente, esta fibronectina plasmática se une a la matriz extracelular del tejido corneal, creando una alta concentración inicial de fibronectina en y alrededor del buque matriz. Los autores antes citados toman el perfil de concentración inicial de fibronectina como $f(x,y,0) = k\exp(-x^2/\epsilon_2)$, $(x,y) \in \Omega$, $k = 0.75$ y $\epsilon_2 = 0.45$.

Condiciones de contorno. Veamos ahora las condiciones de contorno que se deben satisfacer en la frontera Γ de Ω, para que se pueda plantear el problema de existencia, unicidad y dependencia continua de los datos del problema de la solución del sistema (3). Anderson y Chaplain (1998) [2] proponen una condición de no flujo en la frontera Γ de la forma $\zeta \cdot (-D_n \nabla n + n\left[\chi(c)\nabla c + \rho_0 \nabla f\right]) = 0$, donde ζ es el vector normal unitario exterior a la frontera Γ. Orme y Chaplain (1997) [11], consideran condiciones de frontera mixta. Por ejemplo, una vez que el proceso de angiogénesis ha comenzado, es decir, la membrana basal del vaso parental se rompe para formar brotes, y como se concentra la atención exclusivamente en las células endoteliales que están cerca del brote, ya que éstas son las que emigran hacia el tumor, entonces se puede suponer que la densidad inicial de las células endoteliales es $n(0, y, t) = 1$ en Γ_4 y condiciones de flujo cero en las demás fronteras, es decir, $\frac{\partial n}{\partial x}(1, y, t) = 0$ en Γ_2, $\frac{\partial n}{\partial y}(x, 0, t) = 0$ en Γ_1 y $\frac{\partial n}{\partial y}(x, 1, t) = 0$ en Γ_3, ya que como lo indican Paweletz y Knierim (1989) [12], las células endoteliales no cruzan la membrana basal en ninguna otra parte. Por otro lado, ya que el TAF está siendo secretado por el tumor, su concentración permanece constante en ese punto y decae en cero en el vaso parental, por lo tanto, se consideran las condiciones de frontera: $c(1, y, t) = 1$ en Γ_2, $c(0, y, t) = 0$ en Γ_4, $\frac{\partial c}{\partial y}(x, 0, t) = 0$ en Γ_1 y $\frac{\partial c}{\partial y}(x, 1, t) = 0$ en Γ_3. Análogamente, para la concentración de fibronectina, ya que las células endoteliales liberan fibronectina, por lo que las condiciones de frontera para este caso serán: $f(0, y, t) = 1$ en Γ_4, $\frac{\partial f}{\partial x}(1, y, t) = 0$ en Γ_2, $\frac{\partial f}{\partial y}(x, 0, t) = 0$ en Γ_1 y $\frac{\partial f}{\partial y}(x, 1, t) = 0$ en Γ_3.

Si hacemos $r = \Delta t D > 0$ y $u = n^k$, la ecuación (4) toma la forma $-r\nabla^2 u + u = F$, que es una ecuación diferencial parcial de segundo orden de tipo elíptico.

3.1. Formulación variacional del problema mixto

Como observamos antes, Orme y Chaplain (1997) [11] sugieren estudiar un problema mixto. Para ello, dividimos la frontera Γ de Ω en dos partes disjuntas, es decir, $\Gamma = \Gamma_D \cup \Gamma_N$ con $\Gamma_D \cap \Gamma_N = \emptyset$. El problema mixto lo formulamos como sigue:

Problema mixto no homogéneo. Sean $F : \overline{\Omega} \to \mathbb{R}$, $g : \overline{\Gamma}_D \to \mathbb{R}$ y $h : \overline{\Gamma}_N \to \mathbb{R}$ continuas y $r > 0$, determinar $u \in C^2(\Omega)$ y continua en $\overline{\Omega}$ tal que

$$\begin{array}{rcccl} -r\nabla^2 u + u & = & F & \text{en} & \Omega \\ u & = & g & \text{sobre} & \Gamma_D \\ \dfrac{\partial u}{\partial \mathbf{n}} & = & h & \text{sobre} & \Gamma_N, \end{array} \tag{5}$$

donde $\mathbf{n}$ es el vector normal exterior unitario sobre la frontera Γ_N.

Una forma de estudiar el problema mixto no homogéneo (5) es transformándolo en un problema equivalente que sea de tipo Dirichlet homogéneo sobre Γ_D, mediante un cambio de variable apropiado [8, 13]. Para ello, notemos que si $u \in H^1(\Omega)$, el teorema de la traza A.1 garantiza que existe un mapeo $\gamma_0 : H^1(\Omega) \to L^2(\Omega)$ tal que $\gamma_0(u) = u|_{\Gamma_D}$. Más explícitamente, el rango de γ_0 es el espacio $H^{1/2}(\Gamma_D)$. Así, podemos suponer que $g \in H^{1/2}(\Gamma_D)$. Así que de nuevo por el teorema de la traza A.1, existe $\widehat{g} \in H^1(\Omega)$ tal que $\widehat{g}|_{\Gamma_D} = \gamma_0(\widehat{g}) = g$. Proponemos el cambio de variable

$$w = u - \widehat{g}. \tag{6}$$

Transformemos ahora la primera y tercera ecuación de (5) en términos de w:

$$F = -r\nabla^2 u + u = -r\nabla^2 w - r\nabla^2 \widehat{g} + w + \widehat{g},$$

de donde

$$-r\nabla^2 w + w = \widehat{F}, \quad \widehat{F} = F + r\nabla^2 \widehat{g} - \widehat{g}. \tag{7}$$

Y

$$h = \frac{\partial u}{\partial \mathbf{n}} = (\nabla u) \cdot \mathbf{n} = (\nabla w) \cdot \mathbf{n} + (\nabla \widehat{g}) \cdot \mathbf{n} = \frac{\partial w}{\partial \mathbf{n}} + \frac{\partial \widehat{g}}{\partial \mathbf{n}},$$

de donde

$$\frac{\partial w}{\partial \mathbf{n}} = \widehat{h}, \quad \widehat{h} = h - \frac{\partial \widehat{g}}{\partial \mathbf{n}}. \tag{8}$$

Así, bajo el cambio de variable (6) y teniendo en cuenta (7) y (8), el problema mixto no homogéneo (5) es equivalente al problema:

Problema mixto homogéneo. Sean $\widehat{F} : \overline{\Omega} \to \mathbb{R}$ y $\widehat{h} : \overline{\Gamma}_N \to \mathbb{R}$ continuas y $r > 0$, determinar $w \in C^2(\Omega)$ y continua en $\overline{\Omega}$ tal que

$$\begin{array}{rcl} -r\nabla^2 w + w &=& \widehat{F} \quad \text{en} \quad \Omega \\ w &=& 0 \quad \text{sobre} \quad \Gamma_D \\ \dfrac{\partial w}{\partial \mathbf{n}} &=& \widehat{h} \quad \text{sobre} \quad \Gamma_N. \end{array} \tag{9}$$

Para realizar la formulación variacional del problema mixto homogéneo (9), las condiciones de frontera sugieren de manera natural buscar soluciones débiles en el espacio de Sobolev $V = H^1_{0,\Gamma_D}(\Omega) = \{v \in H^1(\Omega) : v = 0 \text{ sobre } \Gamma_D\}$ Puesto que se satisface la desigualdad de Poincaré (teorema A.3) en V, V es un espacio de Hilbert con la norma

$$\|v\|_V = \left(\int_\Omega |\nabla v|^2\right)^{1/2} = \left(\sum_{i=1}^{2} \int_\Omega \left(\frac{\partial v}{\partial x_i}\right)^2\right)^{1/2}, \quad v \in V.$$

Multiplicando ambos lados de la primera ecuación de (9) por una función de prueba $v \in V$ e integrando sobre Ω tenemos:

$$-r \int_\Omega \left(\nabla^2 w\right) v + \int_\Omega wv = \int_\Omega \widehat{F} v. \tag{10}$$

Puesto que

$$\left(\nabla^2 w\right) v = v\frac{\partial^2 w}{\partial x^2} + v\frac{\partial^2 w}{\partial y^2}. \tag{11}$$

Entonces, bajo el supuesto que Ω satisface las hipótesis del teorema de Green A.2, se sigue que

$$\begin{array}{rcl} \displaystyle\int_\Omega v\frac{\partial^2 w}{\partial x^2} &=& \displaystyle\int_\Gamma v|_\Gamma \left.\frac{\partial w}{\partial x}\right|_\Gamma n_1 - \int_\Omega \frac{\partial v}{\partial x}\frac{\partial w}{\partial x} \\ &=& \displaystyle\int_{\Gamma_D} v|_{\Gamma_D} \left.\frac{\partial w}{\partial x}\right|_{\Gamma_D} n_1 + \int_{\Gamma_N} v|_{\Gamma_N} \left.\frac{\partial w}{\partial x}\right|_{\Gamma_N} n_1 - \int_\Omega \frac{\partial v}{\partial x}\frac{\partial w}{\partial x}, \end{array}$$

de donde,

$$\int_\Omega v\frac{\partial^2 w}{\partial x^2} = \int_{\Gamma_N} v\frac{\partial w}{\partial (n_1, 0)} - \int_\Omega \frac{\partial v}{\partial x}\frac{\partial w}{\partial x}. \tag{12}$$

Similarmente,

$$\int_\Omega v\frac{\partial^2 w}{\partial y^2} = \int_{\Gamma_N} v\frac{\partial w}{\partial (0, n_2)} - \int_\Omega \frac{\partial v}{\partial y}\frac{\partial w}{\partial y}. \tag{13}$$

Tomando $\mathbf{n} = (n_1, n_2)$ que es el vector normal exterior unitario a la frontera Γ_N e integrando (11) sobre Ω, y teniendo en cuenta (12) y (13), obtenemos

$$\begin{array}{rcl} \displaystyle\int_\Omega \left(\nabla^2 w\right) v &=& \displaystyle\int_{\Gamma_N} v\left[\frac{\partial w}{\partial (n_1, 0)} + \frac{\partial w}{\partial (0, n_2)}\right] - \int_\Omega \left[\frac{\partial v}{\partial x}\frac{\partial w}{\partial x} + \frac{\partial v}{\partial y}\frac{\partial w}{\partial y}\right] \\ &=& \displaystyle\int_{\Gamma_N} v\frac{\partial w}{\partial \mathbf{n}} - \int_\Omega \nabla w \cdot \nabla v, \end{array}$$

de donde

$$\int_\Omega \left(\nabla^2 w\right) v = \int_{\Gamma_N} v\widehat{h} - \int_\Omega \nabla w \cdot \nabla v. \tag{14}$$

Sustituyendo (14) en (10) y agrupando términos, resulta

$$r \int_\Omega \nabla w \cdot \nabla v + \int_\Omega wv = r \int_{\Gamma_N} \widehat{h} v + \int_\Omega \widehat{F} v.$$

Por lo tanto, la formulación variacional o débil del problema mixto homogéneo (9) se formula como sigue:

Problema variacional mixto homogéneo. Sean $\widehat{F} \in L^2(\Omega)$, $\widehat{h} \in L^2(\Gamma_N)$ y $r > 0$, determinar $w \in V$ tal que

$$r \int_\Omega \nabla w \cdot \nabla v + \int_\Omega wv = r \int_{\Gamma_N} \widehat{h} v + \int_\Omega \widehat{F} v, \quad \forall\, v \in V. \tag{15}$$

La existencia, unicidad y dependencia continua de la solución con respecto a los datos y parámetros del problema variacional mixto homogéneo (15), lo damos en el siguiente teorema vía el teorema de Lax-Milgram B.1.

Teorema 3.1 (Existencia y unicidad)**.** *Sea Ω un conjunto abierto y acotado de clase C^2 en $\mathbb{R}^2$ con frontera $\Gamma = \Gamma_D \cup \Gamma_N$, $\Gamma_D \cap \Gamma_N = \emptyset$. Si $\widehat{F} \in L^2(\Omega)$, $\widehat{h} \in L^2(\Gamma_N)$ y $r > 0$, entonces el problema variacional mixto homogéneo* (15) *tiene una única solución $w \in V$. Además, existe una constante $C > 0$ tal que*

$$\|w\|_V \leq \frac{C}{r}\left(\|\widehat{F}\|_{L^2(\Omega)} + \|\widehat{h}\|_{L^2(\Omega_N)}\right). \tag{16}$$

Más aún, la funcional

$$J(v) = \frac{r}{2}\int_\Omega |\nabla v|^2 + \frac{1}{2}\int_\Omega v^2 - r\int_{\Gamma_N} \widehat{h}v - \int_\Omega \widehat{F}v, \quad \forall\, v \in V \tag{17}$$

alcanza su mínimo en w.

Demostración. Con el propósito de aplicar el teorema de Lax-Milgram B.1, definimos: $B(w,v) = r\int_\Omega \nabla w \cdot \nabla v + \int_\Omega wv$, $(w,v) \in V\times V$ y $\langle \mathcal{F}, v\rangle = r\int_{\Gamma_N} \widehat{h}v + \int_\Omega \widehat{F}v$, $v \in V$. El objetivo es determinar $w \in V$ tal que $B(w,v) = \langle \mathcal{F}, v\rangle$, para todo $v \in V$. En efecto. a) Es claro que $\mathcal{F}$ es una funcional lineal en V.
b) Veamos que $\mathcal{F}$ es continua en V. Aplicando primero la desigualdad de Schwarz para integrales (véase página 63 de [15]), luego el teorema de la traza A.1 y finalmente la desigualdad de Poincaré A.3, obtenemos:

$$\begin{aligned} |\langle \mathcal{F}, v\rangle| &\leq r\left|\int_{\Gamma_N} \widehat{h}v\right| + \left|\int_\Omega \widehat{F}v\right| \\ &\leq r\left(\int_{\Gamma_N} \widehat{h}^2\right)^{1/2}\left(\int_{\Gamma_N} v^2\right)^{1/2} + \left(\int_\Omega \widehat{F}^2\right)^{1/2}\left(\int_\Omega v^2\right)^{1/2} \\ &= \|\widehat{h}\|_{L^2(\Gamma_N)}\|v\|_{L^2(\Gamma_N)} + \|\widehat{F}\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)}, \end{aligned}$$

de donde,

$$|\langle \mathcal{F}, v\rangle| \leq C_T\|\widehat{h}\|_{L^2(\Gamma_N)}\|v\|_{H^1(\Omega)} + C_P\|\widehat{F}\|_{L^2(\Omega)}\|v\|_*. \tag{18}$$

La desigualdad de Poincaré A.3, garantiza que la norma $\|v\|_* = \sum_{i=1}^2 \|\frac{\partial v}{\partial x_i}\|_{L^2(\Omega)}$ es equivalente a la norma $\|\cdot\|_{H^1(\Omega)}$ en V. Así, existe $\alpha > 0$ tal que $\|v\|_* \leq \alpha\|v\|_{H^1(\Omega)}$ para todo $v \in V$. Por otro lado, la misma desigualdad de Poincaré A.3, nos dice que las normas $\|\cdot\|_V$ y $\|\cdot\|_{H^1(\Omega)}$ son equivalentes, por lo que también existe $\beta > 0$ tal que $\|v\|_{H^1(\Omega)} \leq \beta\|v\|_V$ para todo $v \in V$. Concluimos de (18) que

$$|\langle \mathcal{F}, v\rangle| \leq \beta C_T\|\widehat{h}\|_{L^2(\Gamma_N)}\|v\|_V + \beta\alpha C_P\|\widehat{F}\|_{L^2(\Omega)}\|v\|_V = M\|v\|_V \quad \forall\, v \in V, \tag{19}$$

donde $M = \beta C_T\|\widehat{h}\|_{L^2(\Gamma_N)} + \beta\alpha C_P\|\widehat{F}\|_{L^2(\Omega)} > 0$, $C_T > 0$ es la constante de la traza y $C_P > 0$ es la constante de Poincaré. Por consiguiente, $\mathcal{F}$ es continua en V.
c) También es directo verificar que B es una forma bilineal y simétrica sobre $V \times V$.
d) Veamos que B es continua. Notemos primero que

$$|B(w,v)| \leq r\left|\int_\Omega \nabla w \cdot \nabla v\right| + \left|\int_\Omega wv\right|. \tag{20}$$

Trabajemos con la primera integral. Aplicamos la desigualdad de Schwarz para integrales (véase página 63 de [15]), luego la desigualdad de Schwarz para n pares de números reales o complejos (véase pág. 16 de [14]), tenemos:

$$\begin{aligned} \left|\int_\Omega \nabla w \cdot \nabla v\right| &\leq \left|\int_\Omega \frac{\partial w}{\partial x}\frac{\partial v}{\partial x}\right| + \left|\int_\Omega \frac{\partial w}{\partial y}\frac{\partial v}{\partial y}\right| \leq \int_\Omega \left|\frac{\partial w}{\partial x}\right|\left|\frac{\partial v}{\partial x}\right| + \int_\Omega \left|\frac{\partial w}{\partial y}\right|\left|\frac{\partial v}{\partial y}\right| \\ &\leq \left[\int_\Omega \left(\frac{\partial w}{\partial x}\right)^2\right]^{1/2}\left[\int_\Omega \left(\frac{\partial v}{\partial x}\right)^2\right]^{1/2} + \left[\int_\Omega \left(\frac{\partial w}{\partial y}\right)^2\right]^{1/2}\left[\int_\Omega \left(\frac{\partial v}{\partial y}\right)^2\right]^{1/2} \\ &\leq \left(\int_\Omega \left(\frac{\partial w}{\partial x}\right)^2 + \int_\Omega \left(\frac{\partial w}{\partial y}\right)^2\right)^{1/2}\left(\int_\Omega \left(\frac{\partial v}{\partial x}\right)^2 + \int_\Omega \left(\frac{\partial v}{\partial y}\right)^2\right)^{1/2}, \end{aligned}$$

de donde

$$\left|\int_\Omega \nabla w \cdot \nabla v\right| \leq \|w\|_V\|v\|_V, \quad \forall\,(w,v) \in V \times V. \tag{21}$$

Aplicando de nuevo la desigualdad de Schwarz para integrales y luego la desigualdad de Poincaré A.3, a la segunda integral de (20), resulta:

$$\begin{aligned} \left|\int_\Omega wv\right| &\leq \|w\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)} \leq C_P^2\|w\|_*\|v\|_* \\ &\leq \alpha^2 C_P^2\|w\|_{H^1(\Omega)}\|v\|_{H^1(\Omega)} \leq \beta^2\alpha^2 C_P^2\|w\|_V\|v\|_V, \end{aligned} \tag{22}$$

donde $\|\cdot\|_*$, $C_P > 0$, $\alpha > 0$ y $\beta > 0$, son como en (19). Sustituyendo (22) y (21) en (20), se sigue que $|B(w,v)| \leq (r + \beta^2\alpha^2 C_P^2)\,\|w\|_V\|v\|_V$, para todo $(w,v) \in V \times V$, lo que prueba que B es continua.
e) La forma bilineal B es coerciva ya que

$$B(v,v) = r\int_\Omega |\nabla v|^2 + \int_\Omega v^2 \geq r\int_\Omega |\nabla v|^2 = r\|v\|_V^2, \quad \forall\, v \in V.$$

Como se satisfacen todas la hipótesis del teorema de Lax-Milgram (teorema B.1), concluimos que el problema variacional mixto homogéneo (15), tiene una única solución $w \in V$.
f) Notemos que la constante de coercividad es $r > 0$, por lo que se sigue también del teorema de Lax-Milgram B.1 y (19) que

$$\begin{aligned} \|w\|_V &\leq \frac{1}{r}\sup\{|\langle \mathcal{F}, v\rangle| : v \in V,\ \|v\|_V \leq 1\} \\ &\leq \frac{1}{r}\sup\left\{\beta C_T\|\widehat{h}\|_{L^2(\Gamma_N)}\|v\|_V + \beta\alpha C_P\|\widehat{F}\|_{L^2(\Omega)}\|v\|_V : v \in V,\ \|v\|_V \leq 1\right\} \\ &\leq \frac{1}{r}\left(\beta C_T\|\widehat{h}\|_{L^2(\Gamma_N)} + \beta\alpha C_P\|\widehat{F}\|_{L^2(\Omega)}\right) \leq \frac{C}{r}\left(\|\widehat{h}\|_{L^2(\Gamma_N)} + \|\widehat{F}\|_{L^2(\Omega)}\right), \end{aligned}$$

donde $C = \beta\,\text{máx}\,\{C_T, \alpha C_P\} > 0$. Esto muestra la dependencia continua de la solución w con respecto a los datos $\widehat{F}$, $\widehat{h}$ y r del problema. También con esto se prueba la desigualdad (16).
g) Finalmente, como B es simétrica en $V \times V$, se sigue una vez más del teorema de Lax-Milgram B.1 que la funcional $J : V \to \mathbb{R}$ por (17) alcanza su mínimo en w. □

4 Conclusiones

La aportación más relevante del trabajo consistió en aplicar el método variacional para demostrar, en el teorema 3.1, la existencia, unicidad y dependencia continua de la solución débil del problema mixto homogéneo (9) con respecto a los datos y parámetros del problema. Como consecuencia, también se garantizó la existencia de la solución débil del problema no homogéneo (5).

Agradecimientos. A los árbitros por todas las sugerencias y recomendaciones para mejorar el manuscrito. A CONACYT por la beca de manutención que otorgó al primer autor durante el desarrollo del proyecto.

Apéndice

A Espacios de Sobolev y desigualdad de Poincaré

Definición A.1. *Sea $L^p(\Omega)$, $1 \leq p < \infty$, el espacio de las clases de todas las funciones $\varphi : \Omega \to \mathbb{R}$ medibles y tales que $|\varphi|^p$ es integrable sobre Ω. Si $\varphi \in L^p(\Omega)$, se define la norma de φ como $\|\varphi\|_{L^p(\Omega)} = \left(\int_\Omega |\varphi|^p d\boldsymbol{x}\right)^{1/p}$, $1 \leq p < \infty$. $L^\infty(\Omega)$ es el espacio de todas las clases de funciones $\varphi : \Omega \to \mathbb{R}$ medibles y esencialmente acotadas sobre Ω, véase [16]. Si $\varphi \in L^\infty(\Omega)$ se define la norma de φ como $\|\varphi\|_{L^\infty(\Omega)} = \sup^0\{|\varphi(\boldsymbol{x})| : \boldsymbol{x} \in \Omega\}$, donde $\sup^0$ denota el supremo esencial de φ sobre Ω.*

$L^2(\Omega)$ es un espacio de Hilbert con el producto interno $\langle\varphi,\psi\rangle_{L^2(\Omega)} = \int_\Omega \varphi\psi\, d\boldsymbol{x}$ para todo $\varphi, \psi \in L^2(\Omega)$.

Definición A.2. *Se llama espacio de Sobolev $H^1(\Omega)$ al espacio de las funciones $u \in L^2(\Omega)$ cuyas derivadas parciales (en el sentido de las distribuciones) pertenecen a $L^2(\Omega)$, esto es, $H^1(\Omega) = \left\{u \in L^2(\Omega) : \frac{\partial u}{\partial x_i} \in L^2(\Omega),\ 1 \leq i \leq n\right\}$.*

El espacio $H^1(\Omega)$ dotado del producto interno

$$\langle u, v\rangle_{H^1(\Omega)} = \langle u, v\rangle_{L^2(\Omega)} + \sum_{i=1}^{n} \left\langle \frac{\partial u}{\partial x_i}, \frac{\partial v}{\partial x_i} \right\rangle_{L^2(\Omega)} \quad \forall u, v \in H^1(\Omega)$$

el cual induce la norma

$$\|u\|_{H^1(\Omega)} = \left(\|u\|^2_{L^2(\Omega)} + \sum_{i=1}^{n} \left\| \frac{\partial u}{\partial x_i} \right\|^2_{L^2(\Omega)} \right)^{1/2} \quad \forall u \in H^1(\Omega)$$

es un espacio de Hilbert.

Si denotamos por $\mathcal{D}(\Omega)$ al conjunto de todas las funciones de prueba sobre Ω, es decir, funciones de clase $C^\infty(\Omega)$ y de soporte compacto contenido en Ω, entonces $\mathcal{D}(\Omega) \subset H^1(\Omega)$. Es especialmente útil el espacio $H^1_0(\Omega) = \overline{\mathcal{D}(\Omega)}$, es decir, la adherencia respecto de la norma de $H^1(\Omega)$, del espacio de las funciones de prueba $\mathcal{D}(\Omega)$. $H^1_0(\Omega)$ con la norma que hereda de $H^1(\Omega)$ es también un espacio de Hilbert. Dicho de un modo un tanto impreciso, el espacio $H^1_0(\Omega)$ es el formado por las funciones de $H^1(\Omega)$ que se anulan sobre la frontera de Ω. Decimos de un modo un tanto impreciso dado que la frontera de Ω tiene medida nula, y dos funciones de $L^2(\Omega)$ que son iguales salvo en un conjunto de medida cero son, como funciones de $L^2(\Omega)$, iguales. Para eliminar esta ambigüedad se introduce el concepto de traza de una función de $H^1(\Omega)$, véase por ejemplo [8]. Extendemos la definición del espacio de Sobolev para funciones en $L^p(\Omega)$, $p \geq 1$, como sigue:

Definición A.3. *Sea $m \geq 1$ entero y $1 \leq p \leq \infty$. El espacio de Sobolev $W^{m,p}(\Omega)$ se define como $W^{m,p}(\Omega) = \{u \in L^p(\Omega) : \partial^\alpha u \in L^p(\Omega) \ \forall\, |\alpha| \leq m,\ \alpha \in \mathbb{N}^n\}$ dotada de la norma*

$$\|u\|_{m,p,\Omega} = \sum_{|\alpha|\leq m} \|\partial^\alpha u\|_{L^p(\Omega)} \quad \forall\, u \in W^{m,p}(\Omega)$$

o equivalentemente, para $1 < p < \infty$, de la norma

$$\|u\|_{m,p,\Omega} = \left(\sum_{|\alpha|\leq m} \int_\Omega |\partial^\alpha u|^p \right)^{1/p} = \left(\sum_{|\alpha|\leq m} \|\partial^\alpha u\|^p_{L^p(\Omega)} \right)^{1/p} \quad \forall\, u \in W^{m,p}(\Omega).$$

Un caso de especial importancia ocurre cuando $p = 2$, donde se obtiene el espacio $H^m(\Omega) = W^{m,2}(\Omega) = \{u \in L^2(\Omega) : \partial^\alpha u \in L^2(\Omega) \ \forall\, |\alpha| \leq m,\ \alpha \in \mathbb{N}^n\}$ dotada de la norma

$$\|u\|_{m,\Omega} = \|u\|_{m,2,\Omega} = \left(\sum_{|\alpha|\leq m} \int_\Omega |\partial^\alpha u|^2 \right)^{1/2}. \tag{23}$$

El espacio $H^m(\Omega)$ posee un producto interno natural definido por

$$\langle u, v\rangle_{m,\Omega} = \sum_{|\alpha|\leq m} \int_\Omega \partial^\alpha u \partial^\alpha v, \quad \forall\, u, v \in H^m(\Omega),$$

que induce la norma (23), por lo que $H^m(\Omega)$ es un espacio de Hilbert. Como antes, introducimos un importante subespacio de $W^{m,p}(\Omega)$. Si $1 \leq p < \infty$, entonces $\mathcal{D}(\Omega)$ es denso en $L^p(\Omega)$. También, si $\varphi \in \mathcal{D}(\Omega)$, entonces $\partial^\alpha \varphi \in \mathcal{D}(\Omega)$ para todo $\alpha \in \mathbb{N}^n$, por lo que $\mathcal{D}(\Omega) \subset W^{m,p}(\Omega)$ para cualquier m y p. Si $1 \leq p < \infty$, se define el espacio $W^{m,p}_0(\Omega)$ como la adherencia de $\mathcal{D}(\Omega)$ en $W^{m,p}(\Omega)$. Así $W^{m,p}_0(\Omega)$ es un subespacio cerrado de $W^{m,p}(\Omega)$ y sus elementos pueden ser aproximados en la norma de $W^{m,p}(\Omega)$ por funciones de clase C^∞ con soporte compacto contenido en Ω. Cuando $p = 2$, los espacios $W^{m,p}_0(\Omega)$ se denotan como $H^m_0(\Omega)$. En general, $W^{m,p}_0(\Omega)$ es un subespacio estricto de $W^{m,p}(\Omega)$, salvo cuando $\Omega = \mathbb{R}^n$ (véase [8]).

Teorema A.1 (de la traza, Kesavan (1989) [8])**.** *Sea $\Omega \subset \mathbb{R}^n$ un conjunto abierto y acotado de clase C^{m+1} con frontera Γ. Entonces existe un mapeo traza $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_{m-1})$ de $H^m(\Omega)$ en $(L^2(\Omega))^m$ tal que*

a) Si $v \in C^\infty(\overline{\Omega})$, entonces $\gamma_0(v) = v|_\Gamma, \gamma_1(v) = \frac{\partial v}{\partial \mathbf{n}}\big|_\Gamma, \ldots, \gamma_{m-1}(v) = \frac{\partial^{m-1} v}{\partial \mathbf{n}^{m-1}}\Big|_\Gamma$, donde $\mathbf{n}$ es el vector normal exterior unitario sobre la frontera Γ.

b) El rango de γ es el espacio $\prod_{j=0}^{m-1} H^{m-j-1/2}(\Gamma)$.

c) El núcleo de γ es el espacio $H_0^m(\Omega)$.

Teorema A.2 (de Green, Kesavan (1989) [8]). *Sea $\Omega \subset \mathbb{R}^n$ un conjunto abierto y acotado de clase C^1 que yace sobre el mismo lado de su frontera Γ. Sean $u, v \in H^1(\Omega)$. Entonces para $1 \leq i \leq n$, $\int_\Omega u \frac{\partial v}{\partial x_i} = \int_\Gamma u|_\Gamma v|_\Gamma n_i - \int_\Omega \frac{\partial u}{\partial x_i} v$, donde $\mathbf{n}(\boldsymbol{x}) = (n_1(\boldsymbol{x}), n_2(\boldsymbol{x}), \ldots, n_n(\boldsymbol{x}))$ es el vector normal exterior unitario sobre la frontera Γ.*

Teorema A.3 (Desigualdad de Poincaré, Kesavan (1989) [8]). *Sea Ω un conjunto abierto y acotado en $\mathbb{R}^n$. Entonces existe una constante positiva $C = C(\Omega, p)$ tal que $\|u\|_{L^p(\Omega)} \leq C \sum_{i=1}^n \left\| \frac{\partial u}{\partial x_i} \right\|_{L^p(\Omega)}$, $u \in W_0^{1,p}(\Omega)$. En particular, $u \mapsto \sum_{i=1}^n \left\| \frac{\partial u}{\partial x_i} \right\|_{L^p(\Omega)}$ define una norma sobre $W_0^{1,p}(\Omega)$, la cual es equivalente a la norma $\|\cdot\|_{1,p,\Omega}$. Sobre $H_0^1(\Omega)$, la forma bilineal $\langle u, v \rangle = \int_\Omega \sum_{i=1}^n \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i}$ define un producto interno que induce una norma equivalente a la norma $\|\cdot\|_{1,\Omega}$.*

B Formulación variacional abstracta y el teorema de Lax-Milgram

Una función $f \in L^2(\Omega)$ vista como una distribución sobre $\mathcal{D}(\Omega)$, se extiende a una forma lineal y continua sobre $H_0^1(\Omega)$, por medio de la aplicación $f : H_0^1(\Omega) \to \mathbb{R}$ por $\langle f, u \rangle = \int_\Omega f(\boldsymbol{x}) u(\boldsymbol{x}) d\boldsymbol{x}$, para todo $u \in H_0^1(\Omega)$.

Definición B.1 (Problema variacional). *Sea $(H, \|\cdot\|)$ un espacio de Hilbert, $f : H \to \mathbb{R}$ una forma lineal y continua y $B : H \times H \to \mathbb{R}$ una forma bilineal. Por problema variacional entendemos el problema de determinar $u \in H$ tal que*

$$B(u, v) = \langle f, v \rangle \quad \text{para todo } v \in H. \tag{24}$$

La existencia, unicidad y dependencia continua respecto de los datos iniciales de la solución de (24), se obtiene a través del teorema de Lax-Milgram.

Definición B.2. *Sea B una forma bilineal sobre un espacio normado $(H, \|\cdot\|)$. B es continua si existe $M > 0$ tal que $|B(u, v)| \leq M\|u\|\|v\|$ para todo $u, v \in H$, y es coerciva o H-elíptica si existe $m > 0$ tal que $|B(u, u)| \geq m\|u\|^2$ para todo $u \in H$. B es simétrica si $B(u, v) = B(v, u)$ para todo $u, v \in H$.*

Teorema B.1 (de Lax-Milgram, Kesavan (1989) [8]). *Sea $(H, \|\cdot\|)$ un espacio de Hilbert, $f : H \to \mathbb{R}$ una forma lineal y continua, y $B : H \times H \to \mathbb{R}$ una forma bilineal continua y coerciva, entonces el problema variacional (24) tiene una única solución u en H. Además, $\|u\| \leq \frac{1}{m}\|f\|_*$, donde $\|f\|_* = \sup\{|\langle f, v \rangle| : v \in H \text{ y } \|v\| \leq 1\}$. Si B es también simétrica, entonces la funcional $J : H \to \mathbb{R}$ por $J(v) = \frac{1}{2}B(v, v) - \langle f, v \rangle$ para todo $v \in H$ alcanza su mínimo en u.*

Referencias

[1] Aibar, S., Celano, C., Chambi, M.C., Estrada, S., Gandur, N., Gange, P., González, C., González, O., Grance, G., Junin, M., Kohen, N., Molina, J., Núñez, M.G., Sáenz, M., Troncoso, M., Vallejos, A. *Manual de Enfermería Oncológica*. Argentina: Instituto Nacional del Cáncer, 2008.

[2] Anderson, A.R.A. and Chaplain, M.A.J. Continuous and discrete mathematical models of tumor-induced angiogenesis. *Bulletin of Mathematical Biology*, 60:857–900, 1998.

[3] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A. Global cancer statistics 2018: GLOBOCAN Estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6):394–424, 2018.

[4] Folkman, J. Tumor angiogenesis: therapeutic implications. *The New England Journal of Medicine*, 20:1182–1186, 1971.

[5] Glowinski, R. *Numerical Methods for Fluids (Part 3): Finite Element Methods for Incompressible Viscous Flow*. Elsevier Science, 2003.

[6] Graña, A. Breve evolución histórica del cáncer. *Carcinos*, 5(1):26–31, 2015.

[7] Jiménez García, L.F., Merchant Larios, H. *Biología celular y molecular*. México: Pearson Educación, 2003.

[8] Kesavan, S. *Topics in functional analysis and applications*. India: Wiley, 1989.

[9] Kornblihtt, A.R., Pesce, C.G., Alonso, C.R., Cramer, P., Srebrow, A., Werbajh, S. and Muro, A.F. The fibronectin gene as a model for splicing and transcription studies. *The FASEB Journal*, 10(2):248–257, 1996.

[10] Martínez-Ezquerro, J.D., Herrera, L.A. Angiogénesis: VEGF/VEGFRs como blancos terapéuticos en el tratamiento contra el cáncer. *Cancerología*, 1:83–96, 2006.

[11] Orme, M.E. and Chaplain, M.A.J. Two-dimesnional models of tumor angiogenesis and anti-angiogenesis strategies. *IMA Journal of Mathematics Applied in Medicine and Biology*, 14:189–205, 1997.

[12] Paweletz, N. and Knierim, M. Tumor-related angiogenesis. *Critical Reviews in Oncology Hematology*, 9(3):197–242, 1989.

[13] Reddy, J.N. *Applied functional analysis and variational methods in engineering*. USA: McGraw-Hill, 1986.

[14] Rudin, W. *Principios de Análisis Matemático*. México: McGraw-Hill, 3a. edition, 1980.

[15] Rudin, W. *Real and complex analysis*. USA: McGraw-Hill, 3rd. edition, 1987.

[16] Taylor, A.E. and Lay, D.C. *Introduction to functional analysis*. USA: John Wiley, 2nd. edition, 1980.

[17] Theocharis, A.D., Skandalis, S.S., Gialeli, Ch. and Karamanos, N.K. Extracellular matrix structure. *Advanced Drug Delivery Reviews*, pages 1–76, 2015.

[18] Zapata Peña, J., Ortiz, A.C. Uso de modelos matemáticos para la descripción del crecimiento de tumores cancerosos. *NOVA-Publicación Científica en Ciencias Biomédicas*, 8(14):140–147, 2010.

Metodología de reproducibilidad para estudios estadísticos de una población

Perla E. Castillo Flores[1*],
José Luis Fraga Almanza[1], Rina B. Ojeda Castañeda[2]
[1]Facultad de Ciencias Físico Matemáticas, UAdeC
[2]Centro de Investigación en Matemáticas Aplicadas, UAdeC
[*]Autor por correspondencia: perlacastilloflores@uadec.edu.mx

Resumen

En este artículo se presenta una metodología basada en la Ciencia de Datos que proporciona herramientas estadísticas y computacionales para llevar a cabo una estructuración óptima, limpieza y reproducibilidad de datos provenientes de encuestas u otros instrumentos de medición, con el fin de realizar análisis exploratorios confiables para extraer la información subyacente de los datos, que permita avalar la toma de decisiones y el establecimiento de políticas públicas frente a problemáticas de interés social. El ejemplo de aplicación de esta metodología corresponde al estudio estadístico realizado para identificar las 10 principales causas de defunción en la población infantil de los estados de Coahuila y Chihuahua, México.

Palabras clave: Reproducibilidad de datos; Limpieza de datos; Gestión de datos; Obesidad infantil; Mortalidad infantil.

1 Introducción

En las diferentes dependencias del gobierno de México, como, por ejemplo, la Secretaría de Salud, cada día se está generando un gran número de datos relacionados con el estado de salud de una población y los servicios de salud a los que tiene acceso. Esto debido a la necesidad de contar con información que les permita realizar sus actividades, tomar decisiones y establecer políticas públicas de manera eficaz y eficiente, sin embargo, para que esos datos les resulten útiles, deben ser procesados, analizados y presentados al público interesado de la forma más clara y confiable posible.

En general los datos masivos que se generan en cualquier institución o empresa, pública o privada, pueden ser procesados, analizados y presentados utilizando distintas metodologías y herramientas dependiendo de cuál sea su finalidad o utilidad.

En la literatura clásica de la estadística, la mayor parte de la teoría que se presenta, se centra en el modelado de datos, la predicción y la inferencia estadística, asumiendo que los datos se encuentran en el estado correcto para su análisis, por lo que, al mencionar las etapas del procesamiento de los datos, éstas se reducen a la obtención y clasificación de los datos de entrada, el proceso de operaciones necesarias para convertir los datos en información significativa, la presentación y diseminación de los resultados. Sin embargo, como mencionan de Jonge y Van der Loo (2013) en [3], es muy raro que los datos sin procesar, con los que se trabaja estén en el formato correcto, no tengan errores, estén completos y tengan todas las etiquetas y códigos correctos para realizar el modelado de datos, la predicción y la inferencia estadística. Por tal motivo, los usuarios y analistas de datos deben tener cuidado, tanto con las fuentes generadoras como con la información que se encuentra en éstas, desde el momento de accesar a ellas, ya que pueden tener estructuras poco entendibles o bien contener datos basura o faltantes, que les dificultará su manejo, análisis e interpretación [11].

Para llevar a cabo una buena explotación del conocimiento subyacente de grandes datos con fines de análisis para la toma de decisiones, es necesario que los datos tengan la menor cantidad posible de errores, para reducir

dificultades al momento de procesarlos y que tengan la calidad necesaria para su interpretación con un alto grado de robustez y confianza. Para lograr estos requerimientos se necesita, que una vez que se ha accesado a ellos, aplicar una metodología estadística y computacional que permita organizarlos, almacenarlos y limpiarlos, mediante un procedimiento que sea repetible y documentable. Esta metodología deberá regir los procesos de acceso, depuración, complementación, y análisis de los datos que se encuentren en casi cualquier base de datos de forma confiable.

En este artículo se expone la propuesta de una metodología de reproducibilidad desarrollada para accesar, manejar y analizar bases de datos de salud de acceso libre, proporcionadas por la Dirección General de Información en Salud (DGIS) [4] del gobierno de México, con el fin de presentar resultados estadísticos de las principales causas de mortalidad de la población infantil del estado de Coahuila, así como del estado de Chihuahua, con fines comparativos. La metodología que se presenta se basa en el uso adecuado de tres herramientas computacionales: BASH, para crear una jerarquía de directorios adecuada, descomprimir archivos de forma automatizada y hacer una limpieza de archivos basura, MySQL como gestionador de bases de datos y el lenguaje R para el procesamiento de datos, comunicación con las bases de datos y el análisis estadístico de éstos. En general la metodología está basada en los principios de la Ciencia de Datos y se ofrece como una guía para realizar proyectos de análisis de datos en grandes cantidades, con el fin de que su desarrollo se vuelva óptimo, confiable y dinámico.

2 Descripción del problema

Un problema grave que surge cuando se llevan a cabo análisis exploratorios de datos, sean estos descriptivos, inferenciales o bien con fines de modelación del fenómeno o proceso a estudiar, es partir de datos sucios, incompletos, no robustos y poco confiables. Al no contarse con una metodología que permita establecer las reglas y mecanismos para realizar los diferentes procesos de acceso, estructuración y organización de los datos, así como la depuración de los mismos, los procesos posteriores para obtener los análisis estadísticos deseados, resultarán erróneos o poco robustos o no confiables. La materia prima de todo análisis exploratorio son los datos que provengan de bases bien estructuradas y limpias de errores para que se garantice que cualquier usuario interesado en reproducir los análisis o ampliarlos puedan hacerlo sin ningún problema.

Ante esta problemática se propone una metodología estadística y computacional que permite organizar, almacenar y limpiar los datos requeridos en un procedimiento estadístico y que además produzca su repetitividad y su documentación con el fin de ser transferida como herramienta para acceder, analizar y depurar casi cualquier base de datos de forma confiable.

3 Metodología

Dado un conjunto de datos, es muy importante prepararlo y limpiarlo antes de realizar cualquier análisis, de lo contrario se dispondrá de datos basura que podrían causar muchos problemas potenciales sobre todo cuando se trabaja con gran cantidad de datos [11]. Si bien es un camino largo y laborioso preparar las bases de datos, la aplicación de una metodología desde el inicio del proceso, proporcionará a los investigadores grandes beneficios, ya que permitirá automatizar las actividades de búsqueda y descarga de los datos, así como la identificación de los tipos de formatos utilizados dependiendo de la fuente, su limpieza y manipulación, para en las siguientes subsecciones hacer análisis o modelación y obtener conclusiones. En la figura 1, se presentan de forma gráfica las fases de la metodología propuesta, misma que se desarrolló e implantó en la plataforma de un sistema operativo UNIX y es útil para bases de datos estructuradas.

3.1. Definición de la población de estudio y búsqueda de las fuentes de información

El primer paso de la metodología consiste en definir el objetivo del estudio exploratorio y la población a ser analizada, para posteriormente buscar las fuentes de información que proporcionen los datos requeridos para lograr el objetivo de estudio en la población muestral. Se requiere hacer una búsqueda exhaustiva para encontrar las mejores y más confiables fuentes de datos para el objetivo de estudio. Este primer paso pareciera trivial, pero no lo es, es más importante conocer la confiabilidad de la fuente que el hecho de tener muchos datos.

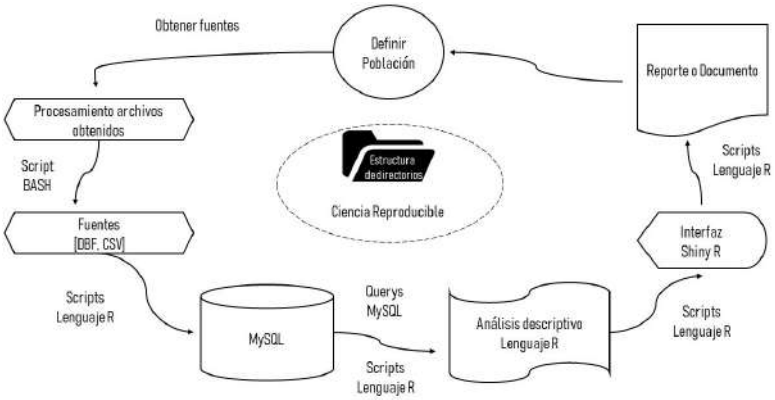


Figura 1: Metodología aplicada, la cual se utilizó para organizar, almacenar y limpiar las bases de datos. Fuente: Elaboración propia.

3.2. Procesamiento de Archivos Fuentes

En la figura 2 se muestran las cuatro tareas de los scripts en Bash [10] para llevar a cabo el procesamiento de archivos. La tarea 1 permitirá tener un espacio de trabajo para guardar contenido de forma ordenada con el fin de que el acceso a los datos, figuras y demás documentos sea más sencillo, crear la jerarquía se vuelve una tarea rápida pues se puede lograr con una sóla línea de comando. La tarea 2, con *wget* [6] permite descargar archivos de la Web con tan sólo disponer de su URL y al mismo tiempo renombrarlos y colocarlos en los directorios correspondientes. Cuando se descargan los archivos fuente generalmente se encuentra como archivos ZIP que deben ser descomprimidos, de ello se encarga la tarea 3. Finalmente, la tarea 4 es eliminar todo el contenido que no sea útil. Para cada tarea, BASH evitará hacer trabajo de más y ahorrar tiempo volviéndose una herramienta sumamente valiosa, sin embargo, debido a que es una herramienta de los sistemas UNIX, la presente metodología no puede ser implementada en sistemas operativos de Microsoft.

Figura 2: Procesamiento de archivos mediante BASH. Fuente: Elaboración propia.

3.3. Exportación a MySQL de las Bases de Datos

Las fuentes de información elegidas proporcionan archivos que deben ser procesados para importar los datos contenidos en ellos a bases de datos con un modelo relacional y que R y MySQL permiten llevar a cabo. Primero deben crearse contenedores vacíos en MySQL que luego serán llenados con los datos contenidos en los archivos obtenidos. La creación de los contenedores se realiza mediante un script en MySQL que indica la creación de cada base de datos con el comando *CREATE DATABASE* [9]. Una vez que se han creado las bases de datos vacías éstas deben poblarse, la tarea puede automatizarse escribiendo un conjunto de scripts en lenguaje R. Para el entorno de trabajo se usó el software RStudio y el paquete RMySQL [8] como interfaz entre las bases de datos y R. La finalidad es mantener las bases de datos ya capturadas, limpias y en un formato específico de forma que los usuarios puedan acceder a ellas y realizar el análisis correspondiente, con ello se logrará que las bases siempre

estén disponibles en el formato original para correcciones o nuevos análisis. Para establecer la conexión a la base de datos se usa la función *dbConnect* [2] de la librería RMySQL que permite crear funciones con los comandos e instrucciones requeridos. En los scripts de R deben definirse los paquetes a cargar, funciones y variables a ser usadas para trabajar con la información [8]. Hay diferentes funciones que permiten leer los archivos que se tienen para obtener todos sus datos dependiendo del tipo de archivo, el más común es *read_ csv()* del paquete **readr**. Una vez que se ha creado la conexión con MySQL y los archivos pueden ser leídos, se pueblan las bases de datos creando las tablas correspondientes con *dbWriteTable()* [2]. Desde R se permite crear las bases de datos y poblarlas en las ubicaciones adecuadas según la jerarquía de directorios creada previamente. Una vez que se han importado las bases de datos a MySQL es importante ordenarlas, limpiarlas y procesarlas. Primero se debe asegurar que los datos se encuentren ordenados de forma que cada columna sea una variable y cada fila una observación, de esta forma se conseguirá un formato que consiga reducir el costo de cómputo y optimizar los resultados de los algoritmos [7]. Luego, se deben detectar los datos que son basura y desecharlos. MySQL permite realizar la limpieza de datos mediante distintas funciones que logran cambiar nombres a variables, identificar duplicados, borrar datos, etc., [5].

3.4. Análisis de Datos

En esta segunda etapa ya con las bases de datos en el formato adecuado y limpias, R permitirá transformar los datos reduciendo las observaciones a aquellas que sean de interés, ya sea creando nuevas variables que sean función de las ya existentes y/o calculando estadísticos de resumen como recuentos y medias, también permite cambiar nombres a las variables, reordenar observaciones y/o generar información mediante la visualización y el modelado de los datos. Gestionar las bases de datos mediante RMySQL permite cargarlas en una variable en R y realizar las modificaciones necesarias sin afectar a la base de datos original. Para comenzar un análisis de datos se escriben scripts en R en los que se contienen los procedimientos deseados para el análisis de los datos. Los datos específicos se obtienen mediante consultas (Querys) en MySQL por lo que será necesario crear conexión entre R y MySQL en cada script. Una forma de iniciar con el análisis exploratorio de datos es mediante la visualización, R contiene múltiples paquetes y funciones encargados de esta acción dependiendo del tipo de datos que se estén trabajando, luego pueden generarse hipótesis y probarlas.

3.5. Interfaz Shiny

Luego de que se tiene el análisis de los datos, una forma de interactuar con los resultados sin modificar el código es a través de Shiny [12], paquete de R que permite crear aplicaciones web interactivas (apps). El paquete debe ser instalado en la mayoría de los casos ya que no forma parte del core básico del lenguaje R, una forma de obtenerlo es mediante la instrucción install.package esto hará que R se comunique al CRAN (The Comprehensive R Archive Network) para obtener la librería en su versión más reciente y estable.

La figura 3 muestra la estructura de archivos al trabajar con shiny.

Figura 3: Estructura de archivos shiny. Fuente: Elaboración propia.

La figura 3 es el resultado de la creación de un proyecto en shiny por medio de RStudio (*File ->NewProject ->NewDirectory ->ProyectType: Shiny Web Aplication*). Es importante mencionar la naturaleza de cada uno de los archivos mostrados en la figura 3. El archivo app.R es la unión de la aplicación web y del archivo servidor encargado de procesar todos los datos involucrados en la aplicación, los directorios restantes son opcionales pero se recomienda crear los necesarios para tener un mejor orden de clasificación de los archivos. Un dato relevante es que el nombre del directorio que contiene todos y cada uno de los archivos y sub directorios de la estructura Shiny debe ser el nombre de la aplicación.

3.6. Documentación

La parte crítica de cualquier proyecto de análisis de datos es la comunicación, pues no importa que tan buenos sean los modelos y gráficos que permitan entender los datos, a menos que se puedan comunicar los resultados de su análisis a otras personas [7]. El proyecto debe ser reproducible y para ello se deben documentar los paquetes necesarios, datos y código utilizado, se recomienda cargar los paquetes al inicio de cada script en R para que sea sencillo identificarlos.

4 Un Ejemplo de Aplicación de la Metodología

Debido a que hoy en día la obesidad infantil y sus comorbilidades han cobrado relevancia en México por su alta prevalencia [1], se buscan identificar las principales causas de defunción en los infantes y si alguna o algunas de ellas pudieran estar relacionadas con el exceso de peso al ser enfermedades crónicas asociadas. Para ello se identificaron las bases de datos de defunciones que proporciona la Dirección General de Información en Salud (DGIS) [4] y se fijó como principal objetivo identificar las 10 principales causas de defunción en niños y adolescentes (menores de 16 años) del estado de Coahuila y de Chihuahua. Para obtener los resultados del análisis se siguió la metodología descrita previamente, y en esta sección se presentan algunos de los resultados obtenidos, sin embargo, no se muestra la parte de shiny ni la del documento reproducible por motivos de espacio.

4.1. Aplicación de la Metodología

Para la búsqueda de las fuentes de información y la definición de la población, se formuló la pregunta: dado el problema de obesidad infantil y las enfermedades crónicas asociadas, ¿qué datos son útiles para identificar las enfermedades por las que mayormente mueren los niños y pudieran estar asociadas a la obesidad?, luego, se definió como población a todos los menores de 16 años del estado de Coahuila y Chihuahua, ante lo que se indentificó que la DGIS [4] cuenta con un Subsistema Epidemiológico y Estadístico de Defunciones (SEED) que integra información de mortalidad del país a través de bases de datos abiertos desde el año 1998 hasta el 2018 incluyendo catálogos y descriptores de campos de cada base de datos posterior al 2012, los cuales sirven como guías que contienen los nombres de campos, tipos de datos y rango de valores. Sólo las bases posteriores al 2012 garantizan calidad al mencionar que se encuentran en la tercera forma normal, lo cual quiere decir que no hay redundancia en los datos y que las bases tiene un buen diseño en el que todos los datos son planos (strings, números y caracteres) sin estructuras multidimensionales como un arreglo o vector por lo que pueden ser leídas por cualquier gestor de base de datos.

Mediante un script en Bash se creó jerarquía de directorios que se muestra en la figura 4.

Figura 4: Jerarquía de directorios. Fuente: Elaboración propia.

La función del directorio Bibliografía fue trivial. El directorio de Descargas permitió guardar los archivos tal y como los ofrecen las fuentes. El directorio Edición se creó para guardar el documento escrito en LaTex del proyecto y para el cual se necesitaron los subdirectorios Figuras, donde se archivaron los gráficos hechos en R y Jabref para crear una base de datos de la bibliografía utilizada. Una vez que los archivos originales fueron almacenados en el

directorio de descargas un archivo en Bash permitió extraer sólo lo necesario y almacenarlo según se trataba de una base de datos, un catálogo o un descriptor de campos en el subdirectorio de la fuente correspondiente, en este caso en SEED. El directorio de procesamiento almacenó los scripts necesarios para el procesamiento y análisis de datos en las herramientas Bash, R y MySQL.

Luego de crear la jerarquía de directorios, se realizó un script en Bash que permitió traer desde la fuente todas las bases de datos disponibles, guardarlas en el directorio Descargas y renombrarlas con nombres cortos (instrucciones utilizadas: wget, mv). Los archivos que se descargaron están en formato ZIP por lo que fue necesario escribir un script en BASH para descomprimir de forma recursiva los archivos obtenidos y posteriormente hacer una limpieza eliminando los archivos innecesarios dejando solo los .csv y .pdf que correspondían a las bases de datos, catálogos y descriptores de campos, mandándolos a los subdirectorios correspondientes.

Se crearon contenedores vaciós en MySQL para cada una de las bases de datos, con los nombres *defun98, defun99, ..., defun17, defun18*, los cuales luego fueron poblados mediante scripts en R permitiendo así tener las bases capturadas y limpias en las tablas correspondientes para que diferentes usuarios puedan disponer de ellas siempre que lo requieran a través de MySQL y R.

Con la función dbConnect de la librería RMySQL en R, se creó la conexión con MySQL para disponer de las bases de datos, la ventaja de utilizar R para llevar a cabo las consultas necesarias para el análisis es que se permite manipular a las bases de datos a conveniencia sin alterarlas directamente, por lo que no pierden su forma original con la que fueron pobladas. Una vez creada la conexión se pudieron leer los archivos requeridos a través de *read.csv()*. Una vez creada la conexión, se inició el análisis de los datos a través de Querys en SQL que permitieron obtener información y gráficos.

Se empezó a trabajar con bases de datos de diferentes años, sin embargo, se encontraron inconsistencias en los años 1998-2011. Los errorres encontrados fueron: falta de homogeneización en los nombres de las variables y datos inconsistentes, por ejemplo, se tenían registros de enfermedades específicas de los recién nacidos en edades posteriores al año de vida. En las bases de datos consecutivas este problema fue corregido. En la figura 5 se muestra el error antes mencionado y se observa como se distribuye el porcentaje de defunciones por la causa P220 (Síndrome de dificultad respiratoria del recién nacido) en intervalos de edad. En el año 2010, figura 5a, los resultados sugerían que hubo población que murió por dicha causa en edades posteriores al intervalo de 0-5 años de edad, lo cual es erróneo debido a las propias características de la causa P220, es decir, dicha causa no debería aparecer en edades posteriores al año de vida, tal y como sucede en el año 2015, figura 5b. Lo mismo ocurrió con otras causas del recién nacido, siendo esto un indicador de que los datos para años previos al 2012 no se encontraban bien estructurados. Además, partir del año 2012 fue posible analizar en qué tiempo murieron los infantes sin tener que redondear al año de vida, pues el dominio de la variable edad permitió ingresar minutos, horas, días y meses de vida para los menores de un año. Se decidió realizar un análisis descriptivo de las bases 2012 al 2017.

(a) 2010 **(b)** 2015

Figura 5: Cantidad de defunciones a causa del síndrome de dificultad respiratoria del recién nacido (P220) por intervalos de edad en el estado de Coahuila. Fuente: Elaboración propia.

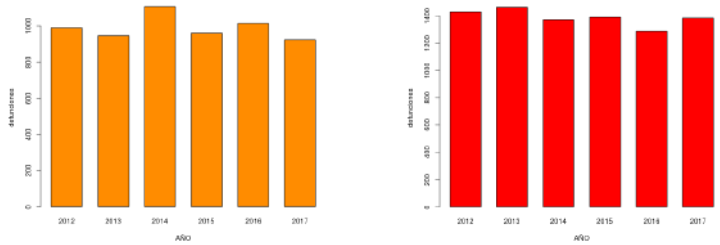
Finalmente, mediante scripts R se llevaron a cabo las consultas pertinentes a través de Querys para encontrar las 10 principales causas de defunción en los estados de Coahuila y Chihuahua en los infantes, creando gráficos y tablas de descripción de las causas, a continuación, se muestra el análisis descriptivo.

4.2. Resultados del Análisis Descriptivo

Como se indicó al final de la sección 4.1, los resultados, y la interpretación de éstos, que se presentan en esta sección están basados únicamente en el análisis de las tablas de frecuencias, graficas sectoriales e histogramas

de variables relacionadas con la mortalidad infantil y las principales causas de ésta en los estados de Coahuila y Chihuahua; no corresponden a inferencias realizadas a través de métodos formales como serían el caso de las pruebas estadísticas de comparación o de métodos de asociación de variables que permiten medir la existencia de correlación entre ellas, que sería la siguiente fase en cualquier estudio estadístico de algún fenómeno poblacional.

En los años analizados, figura 6, el número de defunciones se mantuvo casi constante para cada estado, en el caso de Coahuila, figura 6a, el promedio de defunciones fue de 989 y en Chihuahua el promedio fue de 1386, figura 6b. También se observa que, en Coahuila, en el año 2014, hubo un ligero incremento en el total de defunciones para el intervalo de edad analizado. Los datos para el último año del que se tiene registro (2017) indican que en Coahuila el número de defunciones bajó en comparación al año previo, sin embargo, en Chihuahua aumentó el número de defunciones en comparación al 2016.

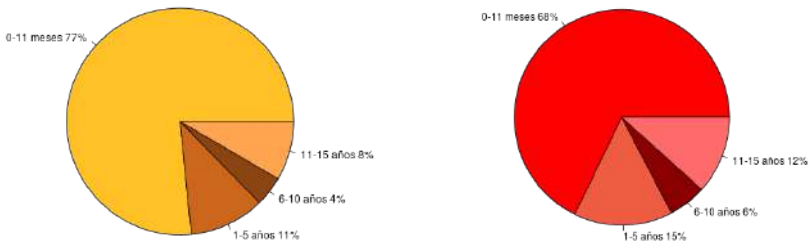


(a) Coahuila **(b)** Chihuahua

Figura 6: Total de defunciones por año. Fuente: Elaboración propia.

Analizando el número de defunciones por intervalos de edad, se notó que no había mucha variación entre sí, pues se mantenían números muy similares y la mayoría de las causas permanecían constantes en cada año en los dos estados, por ello, aquí se presentan únicamente como ejemplo los gráficos y tablas del año 2015, por ser el año intermedio.

En la figura 7 se observa el número de defunciones por intervalos de edad y en la figura 8 se observa el número de defunciones por sexo, ambos en el año 2015 en el estado de Coahuila y Chihuahua. Los intervalos de edad considerados fueron los siguientes: 0-11 meses debido a las causas propias de los recién nacidos, de 1 a 5 años, de 6 a 10 años y de 10 a 15 años, figura 7. Se observa que alrededor del 70 % de las defunciones corresponden a recién nacidos. Si un niño sobrevive a los primeros 11 meses de vida el riesgo a morir decrece notoriamente, destacando que el menor riesgo de perecer está en el rango de edad de 6 a 10 años. En la figura 8 se observa que en ambos estados existe ligeramente mayor prevalencia de defunciones en niños que en niñas.

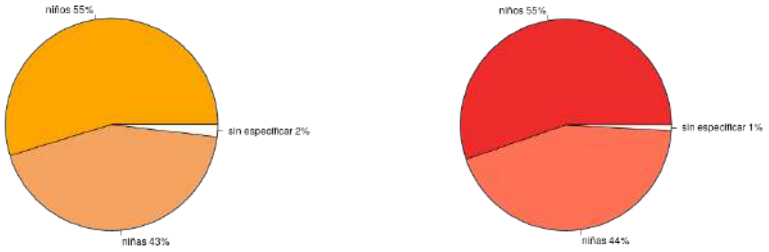


(a) Coahuila **(b)** Chihuahua

Figura 7: Total de defunciones por intervalos de edad, año 2015. Fuente: Elaboración propia.

En la tabla 1 se muestra la descripción de las 10 principales causas de defunción en menores de 16 años observadas en la figura 9.

Se ve en la figura 9 que las principales causas de defunción fueron: P220, P369, Q249, P072 y P239, de las cuáles, tres son propias de los recién nacidos. En el caso de Chihuahua se sumó la causa J189. La principal causa es P220 con frecuencias muy altas en comparación al resto de las causas. Como este primer análisis es marcado por la tendencia en enfermedades propias de los recién nacidos debido a que es la etapa en la que se registran la mayoría de las muertes, se llevó a cabo un segundo análisis de las causas de mortalidad en el resto de los intervalos de edad, esto se presenta en la tabla 2.



(a) Coahuila **(b)** Chihuahua

Figura 8: Total de defunciones por sexo, año 2015. Fuente: Elaboración propia.

	a) Coahuila		
	Causa	Cantidad	Descripción
1	P220	140	Síndrome de dificultad respiratoria del recién nacido.
2	P369	64	Sepsis bacteriana del recién nacido, no especificada.
3	Q249	52	Malformación congénita del corazón, no especificada.
4	P239	31	Neumonía congénita, organismo no especificado.
5	P072	26	Inmaturidad extrema.
6	W840	23	Obstrucción no especificada de la respiración en vivienda.
7	Q039	17	Hidrocéfalo congénito, no especificado.
8	A419	15	Sepsis, no especificada.
9	W849	14	Obstrucción no especificada de la respiración en lugar no especificado.
10	J988	13	Otros trastornos respiratorios especificados.
	b) Chihuahua		
	Causa	Cantidad	Descripción
1	P220	149	Síndrome de dificultad respiratoria del recién nacido.
2	P369	111	Sepsis bacteriana del recién nacido, no especificada.
3	Q249	59	Malformación congénita del corazón, no especificada.
4	J189	41	Neumonía, no especificada.
5	P072	39	Inmaturidad extrema.
6	X599	29	Exposición a factores no especificados que causan otras lesiones y las no especificadas.
7	W780	28	Inhalación de contenidos gástricos en vivienda.
8	P239	26	Neumonía congénita, organismo no especificado.
9	A419	21	Sepsis, no especificada.
10	W789	21	Inhalación de contenidos gástricos en lugar no especificado.

Tabla 1: 10 principales causas de defunción en infantes en el año 2015.

(a) Coahuila **(b)** Chihuahua

Figura 9: 10 principales causas de defunción en infantes en el año 2015. Fuente: Elaboración propia.

	a) 0-11 meses		
	Causa	Cantidad	Descripción
1	P220	140	Síndrome de dificultad respiratoria del recién nacido.
2	P369	64	Sepsis bacteriana del recién nacido, no especificada.
3	Q249	48	Malformación congénita del corazón, no especificada.
4	P239	31	Neumonía congénita, organismo no especificado.
5	P072	26	Inmaturidad extrema.
6	W840	19	Obstrucción no especificada de la respiración en vivienda.
7	P240	13	Aspiración neonatal de meconio.
8	Q039	13	Hidrocéfalo congénito, no especificado.
9	P77X	12	Enterocolitis necrotizante del feto y del recién nacido.
10	J988	11	Otros trastornos respiratorios especificados.
	b) 1-5 años		
	Causa	Cantidad	Descripción
1	Q039	4	Hidrocéfalo congénito, no especificado.
2	V092	3	Peatón lesionado en accidente de tránsito que involucra otros vehículos de motor, y los no especificados.
3	Q249	3	Malformación congénita del corazón, no especificada.
4	W849	3	Obstrucción no especificada de la respiración en lugar no especificado.
5	W840	3	Obstrucción no especificada de la respiración en vivienda.
6	X599	3	Exposición a factores no especificados que causan otras lesiones y las no especificadas.
7	A419	2	Sepsis, no especificada.
8	C910	2	Leucemia linfoblástica aguda [LLA].
9	J960	2	Insuficiencia respiratoria aguda.
10	C749	2	Tumor maligno de la glándula suprarrenal, parte no especificada.
	c) 6-10 años		
	Causa	Cantidad	Descripción
1	G809	4	Parálisis cerebral, sin otra especificación.
2	G039	2	Meningitis, no especificada.
3	G049	2	Encefalitis, mielitis y encefalomielitis, no especificadas.
4	V093	2	Peatón lesionado en accidente de tránsito no especificado.
5	N049	2	Síndrome nefrótico, no especificada.
6	V878	2	Persona lesionada en otros accidentes especificados de transporte de vehículo de motor sin colisión (tránsito).
7	X094	2	Exposición a humos, fuegos o llamas no especificados en calles y carreteras.
8	X950	2	Agresión con disparo de otras armas de fuego, y las no especificadas en vivienda.
9	X470	2	Envenenamiento accidental por, y exposición a otros gases y vapores en vivienda.
10	I490	1	Fibrilación y aleteo ventricular.
	d) 11-15 años		
	Causa	Cantidad	Descripción
1	G809	6	Parálisis cerebral, sin otra especificación.
2	C910	6	Leucemia linfoblástica aguda [LLA].
3	V892	6	Persona lesionada en accidente de tránsito, de vehículo de motor no especificado.
4	X700	6	Lesión autoinfligida intencionalmente por ahorcamiento, estrangulamiento o sofocación en vivienda.
5	E43X	3	Desnutrición proteicocalórica severa, no especificada.
6	C859	2	Linfoma no Hodgkin, no especificado.
7	C419	2	Tumor maligno del hueso y del cartílago articular, no especificado.
8	C402	2	Tumor maligno de los huesos largos del miembro inferior.
9	A419	2	Sepsis, no especificada.
10	G409	2	Epilepsia, tipo no especificado.

Tabla 2: 10 principales causas de defunción en el estado de Coahuila en el año 2015, por intervalos de edad. Fuente: Elaboración propia.

Las causas de defunción para recién nacidos observadas en la tabla 2 se mantuvieron casi constantes respecto a años previos. Las causas asociadas a lesiones o accidentes de distinta índole cobraron un alto número de vidas en los infantes de cualquier edad. En este año las causas G809 (Parálisis cerebral) y C910 (Leucemia Linfoblásticas) afectaron a infantes de los distintos intervalos de edad, aunque con mayor frecuencia cobraron la vida de aquellos en el intervalo de 11 a 15 años de edad donde fueron las dos principales causa de defunción, además G809 fue también la principal causa de defunción en el intervalo de 6 a 10 años de edad. Q039 (hidrocéfalo congénito) fue la principal causa de defunción en el intervalo de 1 a 5 años de edad, en el mismo intervalo aparecieron dos causas por obstrucción de la respiración W849 y W840. Se observó en el intervalo de los infantes mayores que la causa E43X (Desnutrición proteicocalórica severa, no especificada) apareció en la lista de las principales causas por primera vez fuera del intervalo de 1 a 5 años de edad.

Del análisis del resto de las bases de datos 2012-2017 se concluye que el Síndrome de dificultad respiratoria para el recién nacido (P220) fue la principal causa de defunción en los estados analizados, abarcando alrededor del 12 % del total de defunciones en el rango de 0-5 años en Coahuila y alrededor del 9 % en Chihuahua. Se detectó que las causas que cobran más vidas en recién nacidos son aquellas relacionadas con afecciones del sistema respiratorio tales como infecciones (neumonías), obstrucción respiratoria u otro tipo de trastornos repiratorios, en el caso de infantes mayores se mantienen causas relacionadas con lesiones, tumores y leucemia.

En general, el mayor número de muertes sucede dentro de los primeros 5 años de vida y disminuye para edades posteriores.

Sólo en 2016 se observó entre las causas de defunción a la diabetes mellitus, la cuál pudiera estar relacionada a la obesidad infantil, ante ello se nota que el exceso de peso no tiene afectaciones mortales en este rango de edades pues las enfermedades relacionadas pueden empezar a desarrollarse en edades tempranas pero es hasta el futuro cuando se notan las repercusiones y la mortalidad debida a estas.

5 Conclusiones

La metodología estadística y computacional desarrollada y presentada en este trabajo es de gran utilidad para todos los profesionales interesados en realizar análisis estadísticos de fenómenos demográficos, de salud, económicos y sociales, así como otra gran diversidad de procesos administrativos, a partir de bases de datos generadas con la información obtenida de encuestas u otros instrumentos de medición. Sirve como guía para llevar a cabo las requeridas etapas anteriores a la realización del análisis de exploración de los datos con el fin de garantizar la robustez y confiabilidad de éste. Las técnicas y métodos que conforman esta metodología son fundamentales para automatizar las tareas de gestión, organización, estructuración y limpieza de los datos. Además, en esta metodología se conjuntan funciones que proporciona el ambiente de software de R, para llevar a cabo análisis estadísticos descriptivos, inferenciales, de modelación, reproducibilidad y documentación [2]. Las funciones y paquetes considerados en esta metodología para la gestión y análisis de datos, son herramientas propias del sistema UNIX.

Del ejemplo de aplicación se pueden concluir dos cosas, que la estructura y depuración de una base de datos garantizan la calidad de éstos; que cada base de datos puede presentar diferentes tipos de errores y se debe estar preparado para ello, eligiendo el gestionador de bases de datos que se adecue a las características y necesidades de éstas para así seleccionar la mejor. Específicamente se deben buscar fuentes que proporcionen bases de datos bien documentadas con catálogos, descripción de su metodología, modelo y estructura que evitará invertir mucho tiempo para conocer su contenido. Las bases de datos que ofrece la DGIS tienen errores a pesar de venir de una fuente confiable. Los principales errores se encontraron en las bases previas al 2012, ya que son bases no estructuradas, y tienen inconsistencia en los nombres de las variables con respecto a las bases posteriores a esta fecha.

Referencias

[1] Ávila et al. Estado de nutrición en población escolar mexicana que cursa el nivel de primaria. *Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán*, 2016.

[2] CRAN R PROJECT. *R Database Interface: Package DBI*, Dec. 2019. Version 1.1.0. https://cran.r-project.org/web/packages/DBI/DBI.pdf.

[3] De Jonge E., Van der Loo. *An introduction to data cleaning with R.* Statistics Netherlands Grafimedia, 2013. ISSN 1572-0314.

[4] DGIS. Defunciones. datos abiertos, Feb. 2020. Visitado 19-10-2020.

[5] DuBois Paul. *MySQL.* Addison-Wesley Professional, 5th edition, 2013.

[6] Free Software Foundation, Inc. *GNU Wget 1.20 Manual*, November 2018. Visitado 19-11-2020.

[7] Grolemund G. and Wickham H. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O Reilly, 1 st edition, 2017.

[8] Jiménez Chura Adolfo C. RMySQL para el análisis de datos de postulantes e ingresantes del área biomédicas a la Universidad Nacional del Altiplano (Puno Perú). *Revista de Investigaciones Altoandinas*, 19(2):201–210, 2017.

[9] MySQL. *Section 3.3.3 Creating and Selecting a Database (MySQL 8.0 Reference Manual)*, 2020. https://dev.mysql.com/doc/refman/8.0/en/creating-database.html.

[10] Ramey C. and Fox B. *Bash Reference Manual*. Free Software Foundation, Inc., 12 May 2019.

[11] Ricardo Catherine. *Bases Datos*. Mc Graw Hill, 2009.

[12] RStudio Inc. Shiny from r studio. https://shiny.rstudio.com/tutorial/, 2020.

Clúster de PCs tipo Beowulf utilizado en un problema de segmentación de imágenes médicas

José Luis Fraga Almanza[1*], Carlos Eduardo Rodríguez García[1]
Roberto Costancio Torres Ramírez[1]
[1]Facultad de Ciencias Físico Matemáticas, UAdeC
[*]Autor por correspondencia: josefraga@uadec.edu.mx

Resumen

En este trabajo se muestran las etapas principales de la implantación de un Clúster tipo Beowulf, en la Facultad de Ciencias Físico Matemáticas de la Universidad Autónoma de Coahuila. En éste se implementó el algoritmo de agrupamiento de datos K-Means, con la finalidad de distribuir (no de paralelizar) el proceso de clasificación de píxeles en las imágenes. También se realizó un preprocesamiento para la selección de los centroides iniciales requeridos por el algoritmo para una mejor segmentación de imágenes, que como se mostrará logra mejores resultados para el análisis de imágenes digitales médicas, como la segmentación de las mamografías que buscan microcalcificaciones dentro de la mama, que las obtenidas al implantarse de forma secuencial.

Palabras clave: Clustering, K-Means, MPI4Py, Microcalcificaciones.

1 Introducción

El uso intensivo de las computadoras personales (PCs) para dar solución a problemas científicos se dio desde la década de los 70s del siglo pasado, y aun cuando las capacidades computacionales emergían de manera exponencial, la mayoría del equipo a pesar de ser útil, dadas las nuevas necesidades, se volvía rápidamente obsoleto. Esta situación originó la necesidad de diseñar y construir procesadores cada vez más rápidos, de bajo costo y redes tipo ethernet altamente eficientes, aprovechando los avances en la tecnología. Estos desarrollos favorecieron un cambio en la relación precio/prestación en favor del uso de un conjunto de PCs interconectadas a una red ethernet, en lugar de un único procesador de alta velocidad, para resolver un problema dado en común [2]. A este conjunto se le llamó clúster.

En el año 1994, en la revista National Aeronautics and Space Administration (NASA), Becker y Sterling, presentaron el desarrollo de un tipo particular de clúster llamado Beowulf, [13], utilizando una metodología de multicomputadoras para aplicaciones paralelas/distribuidas (APD). Un clúster de este tipo consiste básicamente de un servidor y uno o más clientes (o esclavos) conectados por medio de una red ethernet y sin ningún hardware específico. En sus inicios el primer sistema operativo que se utilizó en este tipo de clústeres fue GNU/Linux, esto no quiere decir que sea el único, se pueden utilizar otros sistemas operativos incluyendo los privativos (Sun Solaris, HP-UX, Microsoft Windows y IBM AIX). Las características que tienen los sistemas operativos de código abierto [12], son la clave para que sean utilizados generalmente en los clústeres tipo Beowulf.

En este trabajo se presentan las etapas y pasos que se realizaron para la instalación y configuración del clúster Beowulf utilizando el sistema operativo FreeBSD, que es un descendiente de la Berkeley Software Distribution (BSD) y derivado de UNIX, [14]. La implementación del clúster se llevó a cabo en la Facultad de Ciencias Físico Matemáticas (FCFM) de la Universidad Autónoma de Coahuila (UAdeC), con la finalidad de implementar el algoritmo K-Means para realizar el análisis de imágenes médicas, específicamente mamografías digitales que se usan para identificar microcalcificaciones en la misma.

2 Desarrollo del clúster tipo Beowulf

A continuación se presentan los pasos realizados, ver Figura 1, para la conexión y configuración de los elementos que forman el clúster tipo Beowulf de la FCFM de la UAdeC.

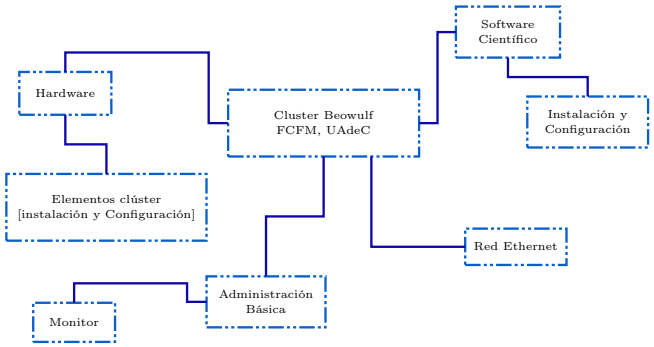


Figura 1: Pasos en el desarrollo del Clúster tipo Beowulf, FCFM, UAdeC. Fuente: Elaboración propia.

2.1. Adecuación del hardware actual

La FCFM de la UAdeC contaba con doce computadoras activas de tipo genéricas. Una de ellas se caracterizaba por tener una placa base marca Intel con un procesador intel core 2 duo, memoria RAM de 4GB y una capacidad de disco de 500GB. Estas características permitieron adecuarla para que funcionara como el nodo maestro, es decir, como la computadora donde se instalaron las librerías, el software y las configuraciones necesarias para la comunicación y réplica de flujos de datos a las once computadoras restantes que al contar con placas base Intel y tipo de procesador Xeon E5 de 16 núcleos, memoria RAM de 16GB y un disco de capacidad de 500GB, por lo que podían convertirse en los nodos esclavos que realizarán las tareas intensivas en la resolución del problema que tuvieran en común.

También se contaba con una red ethernet de alta velocidad con un switch de la marca 3COM de 24 puertos con capacidad 10/100/1000Mbs y con módulos de fibra óptica, la cual permitió la interconexión del equipo maestro y los equipos esclavos. La instalación física de los equipos se realizó a través de un rack que cuenta con ventilación independiente de la marca Sun Microsystems. En la Figura 2 se muestran los equipos de cómputo conectados y listos para instalar el software requerido. Como respaldo de energía se cuenta con dos Smart-UPS 3000XL de la marca APC, que es en donde están conectados la mayor parte de los equipos del clúster.

Figura 2: Instalación física de los elementos del clúster de la FCFM de la UAdeC. Fuente: Elaboración propia.

La configuración física implica atornillar los equipos en el rack, conectar los cables de energía y los cables de

red a las tarjetas de red de los equipos, pero también deben ser conectadas al switch principal, para formar la red ethernet que se utiliza en el clúster.

2.2. Red ethernet

Una vez que se preparó y adecuó el hardware como se indicó en la sección 2.1, se llevó a cabo la conexión.

La configuración lógica de la red ethernet del clúster de la FCFM de la UAdeC es la comunicación entre el nodo maestro y los nodos esclavos, ésta se llevó a cabo por medio de direcciones del tipo *Internet Protocol* (IP) privadas (10.10.1.0/24). La conexión física y lógica de las PCs, es la parte medular del clúster. La conexión y configuración de red que se tiene actualmente se puede observar en la Figura 3, de esta manera el clúster realiza todas las tareas que se le envíen.

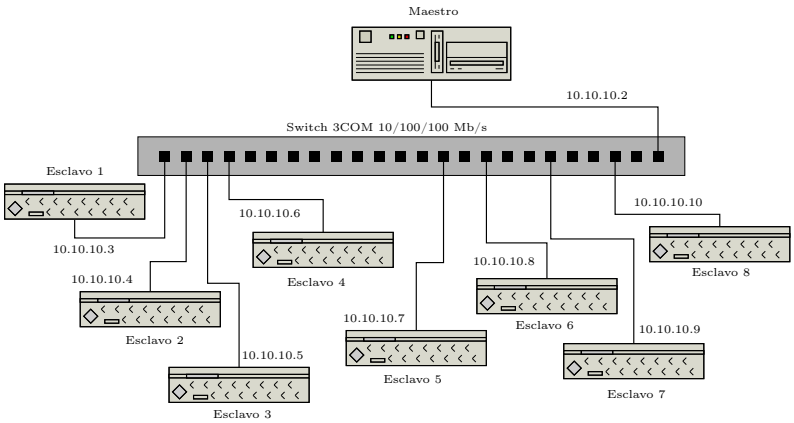


Figura 3: Red ethernet del clúster FCFM de la UAdeC. Fuente: Elaboración propia.

Los nodos del clúster deberán estar registrados en el archivo */etc/hosts* del nodo maestro con sus respectivas direcciones IP y sus nombres (host name). De la misma manera que se configuró el nodo maestro, se debe hacer una configuración equivalente en todos y cada uno de los nodos esclavos, es preferible que se tenga instalado y configurado un servicio de *DHCP* para que la asignación de direcciones IPs en el clúster sea de manera dinámica (esto es realmente útil cuando se tiene una gran cantidad de nodos en el clúster).

2.3. Instalación del software

El software utilizado en el clúster es del tipo *software libre*. La instalación se hizo mediante el manejador de paquetes que en este caso se llama *pkg*, que es una herramienta de apoyo para llevar a cabo la tarea de descargar, instalar y configurar el programa requerido. Los lenguajes de programación escenciales en el clúster son C, C++ y Fortran, que están contenidos en el paquete llamado *GNU Compiler Collection* (GCC), el cual es utilizado para llevar a cabo instalaciones de software adicional, por ejemplo, cuando sólo se cuenta con su código fuente, pero también es utilizado por *pkg* para realizar tareas de configuración e instalación de dependencias requeridas por algún otro software adicional, la instalación se realiza de la siguiente manera:

```
pkg install lang/gcc
```

A continuación se presenta una lista que es un subconjunto de los lenguajes, programas y librerías que se requieren adicionalmente para llevar a cabo la segmentación de las imágenes digitales.

1. MPI
2. Linear Algebra Package (LAPACK)
3. Scalable LAPACK (ScaLAPACK)

4. Python
5. MPI4Py

MPI es un modelo de comunicación ampliamente usado en computación paralela y distribuida, en [7], se puede encontrar una buena introducción a este modelo de comunicación. Es un estándar que define la sintaxis y la semántica de las funciones, contenidas en una biblioteca de paso de mensajes diseñada para ser usada en programas que hagan uso de multiples procesadores. Se instala con la siguiente instrucción:

```
pkg install net/mpich
```

LAPACK es una colección de subrutinas escritas en FORTRAN que sirven para resolver problemas matemáticos y que son parte del álgebra lineal numérica. Las subrutinas de LAPACK se basan en otras subrutinas más sencilla que se conocen por la sigla BLAS (Basic Linear Algebra Subprograms). La instalación de LAPACK se hace de la siguiente forma:

```
pkg install math/lapack
```

ScaLAPACK es un conjunto de rutinas LAPACK rediseñadas para la computación paralela de memoria distribuida. Está diseñado para computación heterogénea y es portátil en cualquier computadora que admita MPI, depende de PBLAS de la misma forma que LAPACK depende de BLAS. La instalación se realiza de la siguente forma:

```
pkg install math/scalapack
```

Python es un lenguaje de programación de alto nivel, orientado a objetos, con una semántica dinámica integrada, es conocido por su gran variedad de usos, por ejemplo, inteligencia artificial, big data, cómputo científico, etc. La instalación es la siguiente.

```
pkg install py37-pip
```

Adicionalmente para la implementación del clúster es necesario instalar *Python Intel*, que es una distribución relativamente nueva de *Python*, desarrollada por *Intel* para la *Deep Learning & Vision Tools*. Los detalles de esta distribución de *Python* se pueden encontrar en [8]. La instalación se lleva a cabo mediante un archivo ejecutable llamado *bash_setup_intel_python.sh*, para detalles de instalación consultar en [8].

MPI4Py es un estándar de *MPI* para *Pyhton*, permite a los programas escritos en *Python* ejecutarse en computadoras con múltiples procesadores, en [10] se explica a detalle todo sobre éste modelo de programación en Python.

2.4. Administración básica del clúster

La tarea principal que se lleva a cabo en el uso del clúster es su administración, esto implica hacer lo siguiente.

- Crear cuentas de usuarios.
- Preparar el directorio Net File System (NFS): */export/aplicaciones*.
- Monitorear el estado del clúster con Ganglia.

Las *cuentas de usuario* se crean utilizando el comando *useradd* de UNIX, previamente el administrador deberá crear los grupos de trabajo para dar una jerarquía adecuada en el clúster. En este caso existen los grupos: *investigadores* y *alumnos*, son grupos con ciertos privilegios ante el sistema operativo para poder clasificar los procesos enviados a los nodos esclavos.

El *NFS* es un sistema que permite accesar localmente a un sistema de archivos que se encuentra en un dispositivo físico remoto, es decir, *NFS* permite tener acceso inmediato a los archivos de otra computadora como si fueran archivos de una computadora local. Este sistema utiliza los protocolos RPC (Remote Procedure Call), la preparación de su configuración en el nodo maestro que exporta el sistema de archivos permite definir quiénes pueden acceder al sistema de archivos y con qué privilegios (lectura, escritura, etc.). La configuración se realiza mediante la edición del archivo ubicado y llamado */etc/exports*. La sintaxis básica de este archivo es como sigue:

```
<sistema de archivos> <regla\_IP> (opciones). <regla\_IP> (opciones),
```

donde *sistema de archivos* es el directorio asociado al sistema de archivos que se quiere exportar, *regla_IP* es una regla que define las IPs que se pueden montar de manera remota con NFS, el sistema de archivos y *opciones* definen en NFS los privilegios que se otorgarán.

Para monitorear el estado del clúster se utiliza una interfaz web llamada *Ganglia*, es una herramienta que hace uso de un servidor web como Apache [9]. A continuación se muestra un ejemplo de Ganglia funcionando en el clúster de la FCFM de la UAdeC.

Figura 4: Ganglia, herramienta para monitoreo del clúster FCFM de la UAdeC. Fuente: Elaboración propia.

En la Figura 4 se observa en la parte superior un conjunto de pestañas que representan diferentes visualizaciones del clúster para observar el estado en que se encuentra su funcionamiento. También se muestra el nombre *apolo* que se le dió al clúster en su configuración, en la parte media se presentan algunas gráficas de diferentes elementos que conforman el clúster como lo es su memoria RAM, el CPU, el flujo de Red y otras más. En el ejemplo se muestra que en este momento se encuentran 66 procesadores activos en cuatro PCs y que un nodo no esta funcionando bien y por lo tanto Ganglia lo reporta como un nodo inactivo.

3 Problema de agrupamiento de datos

El agrupamiento de datos, es considerado uno de los problemas relevantes en el aprendizaje no supervisado [6]. Es decir, como en todos los problemas de este tipo, lo que se trata de hacer es encontrar grupos en un conjunto de datos. De manera más general, puede definirse el agrupamiento de datos como *el proceso de organización de objetos que son muy similares de alguna manera*, ver Figura 5.

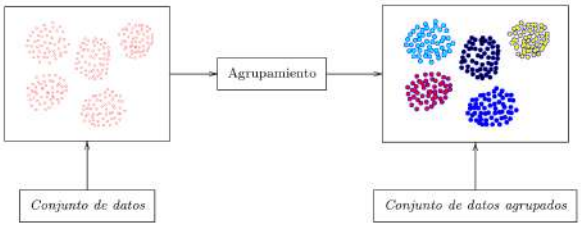


Figura 5: Agrupamiento de puntos en $\mathbb{R}^2$. Fuente: Elaboración propia.

Definición 3.1 (Problema de agrupamiento de datos)**.** *Dado un conjunto D de datos con n objetos en un espacio m-dimensional, agrupar datos es particionar los mismos en k grupos tales que los puntos dentro de un grupo son más similares entre ellos que con otros grupos, dicha similitud se mide atendiendo a alguna función distancia.*

3.1. El algoritmo K-Means

Una solución al problema de agrupamiento de datos se da al aplicar algoritmos de agrupamiento, como el K-Means. Este algoritmo tiene caractarísticas simples y una gran cantidad de variantes como se describe en [1]. A continuación se mencionan algunos aspectos del algoritmo K-Means implementado en el Clúster, para mayor detalle, consultar en [6, 3].

Algoritmo K-Means

Dado un conjunto de datos $D = \{x_1, x_2, \ldots, x_n\}$ y un k número de clústers a formar, entonces

1. Seleccionar un conjunto arbitrario o aleatorio de centroides iniciales en D para los k clústers: $c_1, c_2, \ldots, c_k$.
2. Para cada x_j de D, encontar el centroide c_j más cercano a x_i y asignamos x_j al clúster C_j

$$C_j = \min_{1 \leq j \leq k} ||x_i - c_j||.$$

3. Para cada uno de los clústers recalcular su centroide basado en los elementos que están contenidos en el clúster.
4. Repetir los pasos 2 y 3 hasta que

$$q_j = \sum_{v \in \pi_j} \|a_v - m_j\|_2^2$$

donde q_j es una partición de los k grupos.

$$\min_{\Pi} \quad Q(\Pi) = \sum_{j=1}^{k} q_j = \sum_{j=1}^{k} \sum_{v \in \pi_j} \|a_v - m_j\|_2^2.$$

El algoritmo K-Means proporciona agrupamientos buenos y malos. Enseguida se presenta un ejemplo de buen agrupamiento y otro de mal agrupamiento.

(a) Buen agrupamiento. **(b)** Mal agrupamiento.

Figura 6: Ejemplos de agrupamiento de puntos en $\mathbb{R}^2$ con K-Means. Fuente: Elaboración propia.

Se observa que en el caso del buen agrupamiento se puede trazar una recta que divida al plano en regiones, cada una de ellas conteniendo a uno de los grupos. En el caso de un mal agrupamiento es imposible trazar una recta con esas caracterísiticas. Esta es la clave para saber cuando K-Means dará un buen agrupamiento.

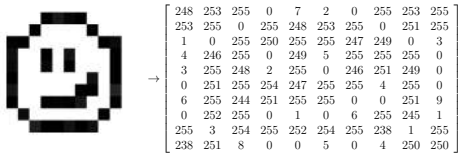
A pesar de ser utilizado ampliamente en una gama de aplicaciones, el algoritmo K-Means no está exento de limitaciones. Algunos de los inconvenientes que presenta han sido ampliamente descritos en la literatura [11]. Existen técnicas o variantes de K-Means que intentan superar estas limitaciones, por ejemplo en el trabajo realizado por Fraga, Mederos y Madrid [4, 5] se propone una variante a la hora de seleccionar los centroides iniciales, misma que se utiliza en este trabajo.

3.2. Imágenes en escala de grises

En una imagen digital a escala de grises, cada pixel puede definir solamente una tonalidad de gris (luminosidad) y el número de pixeles define la cantidad de información que contiene la imagen. Las tonalidades están comprendidas en el intervalo $[0, 255]$ donde el tono negro es el valor de 0 y el blanco es el valor de 255, ver Figura 7. Una representación matemática de una imagen en escala de grises en una computadora puede definirse asociándole a cada pixel una terna (x, y, t), donde x, y representa la posición del pixel y t su tonalidad, por lo tanto una imagen digital es un conjunto representado en tres dimensiones.

Figura 7: Escala de grises de 0 a 255. Fuente: [6].

Una imagen se representa por una matriz. En la Figura 8 se observa un ejemplo de una imagen digital en escala de grises, representada en la computadora.



$$\rightarrow \begin{bmatrix} 248 & 253 & 255 & 0 & 7 & 2 & 0 & 255 & 253 & 255 \\ 253 & 255 & 0 & 255 & 248 & 253 & 255 & 0 & 251 & 255 \\ 1 & 0 & 255 & 250 & 255 & 255 & 247 & 249 & 0 & 3 \\ 4 & 246 & 255 & 0 & 249 & 5 & 255 & 255 & 255 & 0 \\ 3 & 255 & 248 & 2 & 255 & 0 & 246 & 251 & 249 & 0 \\ 0 & 251 & 255 & 254 & 247 & 255 & 255 & 4 & 255 & 0 \\ 6 & 255 & 244 & 251 & 255 & 255 & 0 & 0 & 251 & 9 \\ 0 & 252 & 255 & 0 & 1 & 0 & 6 & 255 & 245 & 1 \\ 255 & 3 & 254 & 255 & 252 & 254 & 255 & 238 & 1 & 255 \\ 238 & 251 & 8 & 0 & 0 & 5 & 0 & 4 & 250 & 250 \end{bmatrix}$$

Figura 8: Representación de una imagen en una computadora. Fuente: Elaboración propia.

Si la imagen digital es de $m \times n$ entonces la representación de los datos para ser agrupados con K-Means es una matriz de orden $mn \times 3$ donde las primeras dos columnas son las posiciones x, y en la matriz y la tercer columna es la tonalidad. De esta manera cada renglón de la matriz de datos representa un pixel de la imagen.

3.3. Segmentación de imágenes

La segmentación de imágenes es un tipo particular de agrupamiento de datos, en este sentido se puede utilizar K-Means y llevar a cabo este tipo de agrupamiento. Cada segmento o grupo de la imagen tiene las siguientes propiedades:

1. Las componentes x y y de cada pixel están cercanas entre sí, es decir, están cercanas geométricamente.
2. Las componentes t de cada pixel tienen tonos de gris cercanos entre sí.

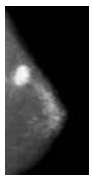


Figura 9: Imagen con resolución 1024Mp. Fuente: Proyecto en desarrollo, División de Ciencias Exactas y Naturales de la UNISON.

A continuación se presentan los resultados proporcionados por K-Means de la segmentación de la Figura 9 que tiene un tamaño de 1280×800 pixeles, por lo tanto una matriz de tamaño 1024000×3, dando distintos valores de k y con centroides iniciales arbitrarios, esto es, aplicando el algoritmo K-Means sin hacer alguna selección especial de los centroides iniciales.

Un preprocesamiento (selección de centroides iniciales de manera adecuada al problema) como el que se presenta en [5] fue utilizado posteriormente para segmentar la misma imagen, obteniendose resultados mucho más significativos que utilizando el algoritmo sin él, como se puede ver en la Figura 11.

(a) Valor de $k = 3$. **(b)** Valor de $k = 4$.

Figura 10: Segmentación de mamografía con K-Means con centroides iniciales arbitrarios. Fuente: Elaboración propia.

Las segmentaciones obtenidas en la Figura10 significan, que se buscan dentro de la mamografía, tres y cuatro tonalidades diferentes en la escala de grises. Por ejemplo en la Figura 10a solo se aprecia el contorno color blanco de la mamografía y dos tonalidades diferentes en el fondo. En la Figura 10b se aprecia que dentro de la mamografía hay cuatro tonalidades diferentes, blanco, negro y dos tonalidades de tipo diferente de gris, sin embargo, esto no proporciona información relevante, ya que se busca la existencia de microcalcificaciones dentro de la mamografía.

Por tal motivo, es necesario realizar el preprocesamiento descrito en [5], que realice una elección adecuada de los centroides iniciales, para así hacer evidente la existencia de microcalcificaciones dentro de la mamografía.

(a) Valor de $k = 3$ y preprocesamiento. **(b)** Valor de $k = 4$ y preprocesamiento.

Figura 11: Mamografía segmentada con K-Means, eligiendo centroides iniciales adecuados. Fuente: Elaboración propia.

En las Figuras 11a y 11b se observan mayores detalles dentro de la mamografía segmentada, esto significa que la segmentación ha mejorado. En ellas se puede distinguir la existencia de microcalcificaciones, por ejemplo, en la Figura 11a esas microcalcificaciones son de color blanco dentro de la mama, es decir, tonalidades que pertenecen a un mismo grupo. Aumentar el valor de k es encontrar más características (grupos) en la mamografía que se está segmentando. Esto no significa que incrementar el valor de k hará mucho mejor la segmentación, eso dependerá del tipo de problema que se esté trabajando.

Utilizando el mismo preprocesamiento pero ahora escribiendo una versión distribuida del programa con MPI4Py e implementado en el clúster de la FCFM, se obtiene la segmentación que se muestra en la Figura 12.

(a) $k = 3$ y MPI4Py. **(b)** $k = 4$ y MPI4Py.

Figura 12: Segmentación de una mamografia con K-Means, en forma distribuida. Fuente: Elaboración propia.

La forma distribuida ha mejorado significativamente la segmentación en la mamografía. En las Figuras 12a y 12b se observa que las microcalcificaciones están en color blanco y en el resto de la mamografía se aprecian diferencias de tonalidades de grises dibujando patrones distintos. Esto se debe a la elección adecuada de los centroides iniciales y por que en cada nodo del clúster se calculan centroides iniciales diferentes. El nodo maestro

obtiene un vector de centroides diferentes y siguiendo el preprocesamiento descrito en [5] se vuelve a realizar una nueva segmentación.

El programa que realiza la segmentación con K-Means en forma distribuida en el clúster se llama *segkmeans-preale.py* que utiliza las bibliotecas OpenCV y Numpy, su diagrama de flujo es:

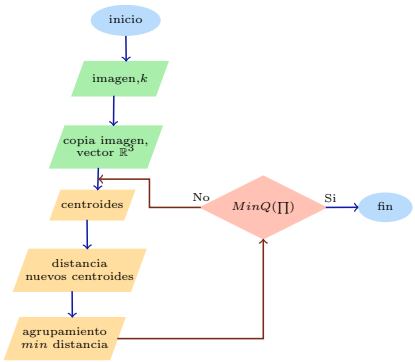


Figura 13: Diagrama de Flujo programa: segkmeanspreale.py. Fuente: Elaboración propia.

La biblioteca MPI4Py es utilizada para distribuir los procesos en el clúster con la instrucción *from mpi4py import MPI* y se comunica por medio de MPI.COMM_WORLD con todos los procesos enviados al clúster. El esquema de envió de procesos sería como se puede observar en la Figura 14.

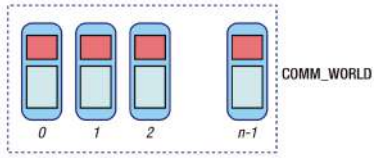


Figura 14: Concepto de comunicación de procesos con MPI. Fuente: [10].

La ejecución del programa se realiza con:

mpiexec -f archivonodos -n 4 python3 -i mamografia.jpg -k 3

Para más detalles de cómo distribuir un programa utilizando MPI4Py revisar [10], documento base para llevar acabo el experimento que se presenta en este trabajo.

4 Conclusiones

El experimento de la segmentación de la imagen de una mamografía digital con la finalidad de identificar microcalcificaciones en la mama, con resultados altamente satisfactorios permiten ver la gran utilidad y ventaja de tener un clúster como herramienta computacional de apoyo en el análisis de este tipo de imágenes médicas, ya que en un momento dado se podrían eliminar las biopsias que resultan ser procedimientos invasivos para los pacientes.

La construcción de un clúster tipo Beowulf representó ser para la FCFM de la UAdeC, una inversión de muy bajo costo en recursos financieros, pero de una muy alta utilidad para los investigadores y alumnos, quienes podrán utilizarlo para realizar su trabajo. Con él se puede hacer búsquedas de soluciones a problemas complejos, es una opción económica, sustentable y generadora de conocimiento tanto para investigadores como para los alumnos. Las ventajas que tienen este tipo de clústeres es que no requieren de hardware especial, ya que están compuestos de piezas que se pueden encontrar fácilmente en el mercado y a un bajo costo. El software libre es una gran oportunidad de aprendizaje tanto para alumnos como para investigadores de la FCFM, ya que proporciona la libertad que necesita el usuario para no depender de un presupuesto económico, ni estar sujeto a términos y condiciones de uso como lo impone el software privativo. Con este tipo de sofware se cuenta con programas, librerias y compiladores especializados en el área del cómputo científico.

Un hecho importante de los clústeres tipo Beowulf es que las soluciones obtenidas al utilizarlos, se logran a menor costo, y por tanto existe un gran impacto económico, lo que justifica seguir construyendo supercomputadoras de este tipo.

Referencias

[1] Anil K. Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8):651 – 666, 2010. Award winning papers from the 19th International Conference on Pattern Recognition (ICPR).

[2] Daillidis Christos. *Establishig Linux Clusters for high-performance computing*. PhD thesis, Naval Postgraduate School, 2004.

[3] Eldén, Lars. *Matrix methods in data mining and pattern recognition*. SIAM, 2007.

[4] Fraga Almanza, José Luis . El Algoritmo K-Means y algunas aplicaciones. Technical report, Universidad Autónoma de Coahuila, 2011.

[5] Fraga Almanza, José Luis and Mederos Madrazo, Boris de Jesús and Madrid de la Vega, Humberto . *Segmentación de imágenes en escala de grises con K-Means*, chapter Cápitulo 6, pages 47–59. Tópicos Recientes sobre Aplicaciones de las Matemáticas en México, 2015.

[6] Gan, Guojun and Ma, Chaoqun and Wu, Jianhong. *Data clustering: theory, algorithms, and applications*, volume 20. Society for Industrial and Applied Mathematics, 2007.

[7] Hidrobo, Francisco and Hoeger, Herbert. Introduccón a mpi. Centro Nacional de Cálculo Científico Universidad de los Andes, CeCalCULA Mérida Venezuela, 2005.

[8] Intel inside. intel distribution for python. https://software.intel.com/content/www/us/en/develop/tools/distribution-for-python.html, 2020. Visto 19-11-2020.

[9] Massie, Matt and Li, Bernard and Nicholes, Brad and Vuksan, Vladimir and Alexander, Robert and Buchbinder, Jeff and Costa, Frederiko and Dean, Alex and Josephsen, Dave and Phaal, Peter and Pocock, Daniel. *Monitoring with Ganglia*. O'Reilly Media, Inc., 1st edition, 2012.

[10] Pajankar, Ashwin. *Parallel Programming in Python 3*, pages 87–97. Apress, Berkeley, CA, 2017.

[11] Pérez, J and Henriques, MF and Pazos, R and Cruz, L and Reyes, G and Salinas, J and Mexicano, A. Mejora al algoritmo de agrupamiento K-means mediante un nuevo criterio de convergencia y su aplicación a bases de datos poblacionales de cáncer. *II Taller Latino Iberoamericano de Investigación de Operaciones*, 2007.

[12] Prieto, S.S. and Población, Ó.G. and TOME, A.G. *Unix y Linux. Guía práctica, 3a edición*. RA-MA S.A. Editorial y Publicaciones, 2004.

[13] Sinisterra, María Mercedes and Díaz Henao, Tania Marcela and Ruiz López, Erik Giancarlo. Clúster de balanceo de carga y alta disponibilidad para servicios web y mail. *Informador Técnico*, 76:93, dic. 2012.

[14] Wikipedia. Berkeley software distribution. https://es.wikipedia.org/wiki/Berkeley_Software_Distribution, 2020. Visto por última vez el 15 nov 2020 a las 16:43.

A numerical approximation for build the stress/strain diagram with just the maximum load obtained

J. Alberto Guzmán-Torres[1*] Francisco J. Domínguez-Mota
Gerardo Tinoco-Guerrero Elia M. Alonso-Guzmán
Marco A. Navarrete-Seras
[1]Facultad de Ingeniería Civil, UMSNH
[*]Correspondence author: jaguzman@umich.mx

Abstract

One of the fundamental features in the design of rigid pavements is the rupture modulus, which value is obtained experimentally using a bending test and indicates the maximum load that a concrete beam can support. To determine this value is expensive and impractical since the tests require rather large stress values, there is always a risk of producing brittle fractures in the probes used. For this reason, it is proposed in this paper an implementation of numerical estimation for the deformations due to the maximum load through a simple interpolation model that allows the generation of the stress/strain diagram using the only data available: the maximum load. A primary focus of analysis in this field of study is the approach in which the materials are analyzed. In this paper, the research is performed under the consideration of an isothermal process, i.e., at a constant temperature. The results show that it is possible to calculate the maximum deformation value in the beam previous to the failure. The analysis was performed in one dimension and, it is possible to generate the stress/strain diagram and distinguish the elastic and elastoplastic areas of the concrete.

Keywords: Finite Element Analysis; Concrete; Blends; Approximation

1 Introduction

Nowadays, concrete is one of the most used materials in construction. It is used for various important reasons like its ability to be moldable in almost any shape compared to other materials, nevertheless, it requires fast and reliable quality control tests [19].

That is why the laboratories of materials which are responsible for checking quality control are an important factor for this kind of materials, forming part of the important decisions as they continue with the implementation of a product, change it or modify it and for this reason, the area of material resistance is particularly important. Usually, one of the most requested destructive tests is the simple compression test, in which, to determine the $f'c$ (specified compressive stress) values is the goal. Other important tests are the Mr (rupture modulus) and the Test Method for Splitting Tensile Strength. The flexion test in concrete is applied to prismatic or rectangular specimens to obtain the break modulus values [12]. A thorough analysis of the test results is required due to the importance of these parameters in the design of rigid pavements, on which the bearing surface is made of concrete [6]. Obtaining the deformations that are generated in this kind of element is not a simple task, because it requires adequate instrumentation to measure a small displacement that occurs in a short lapse of time previous to a brittle fracture. Therefore, we need to implement an efficient method to obtain accurate information, for this reason, numerical analysis such as those given by the finite element analysis [2] can be useful to calculate the deformations generated by the applied load. The finite element analysis requires to discretize the structure making a complete analysis of the behavior of the structure [8]. The modeling is essential to study the fundamentals processes that occur in materials to predicted real performance on the structures of concrete and to use both to improve the

design [16]. In this study, the analysis was carried out under the consideration of an isothermal process, and this is because Young's modulus, the maximum load supported, and the elastic parameters involved in the bending test are notably sensitive to temperature changes. Also, the numerical methods in conjunction with lab results are fundamental to improve the analysis. This paper contains information about $f'c$ (specified compressive stress) values of a different blend of concrete elaborated under controlled conditions in the lab, The aggregates used in this study were obtained from the "joyitas" material bank located in the city of Morelia Michoacán. The cement used was a product of the Holcim trademark, the simple compressive strength of the design was 35 Mpa. For this analysis, prismatic elements were made with the elaborated blend and they were used in bending stress test to obtain the maximum load that can be held.

The study is based on the elaboration of the deformation stress/strain diagram obtaining only the maximum load in the bending test made in the laboratory. These data have become a benchmark for structures as rigid pavements. The dynamic problems of fracture can be classified into stationary and those with propagation cracks. In both cases, the tension fields and displacements around the crack are time-dependent. This value is 35 Mpa, where the design of the blend was based on the methodology American Concrete Institute. Prismatic specimens of 15 cm $\times$ 15 cm $\times$ 60 cm were made and they were demolded and cured, where the beams are doubly supported, there are studies for dynamic cases as mentioned in [13] or for beam cantilever cases [11]. The dynamic problems of fracture can be classified into stationary and those with propagation cracks. In both cases, the tension fields and displacements around the crack depending on the time. [7] and [5] studied stress fields and displacement of dynamic problems of stationary cracks in isotropic materials [1], Nishioka and Atluri [21] also, studied the tension fields and displacement of the dynamic propagation of cracks with high speed in isotropic materials. The results were obtained monitoring for different ages of tests which were 60, 90, 140 and 180 days taking the arithmetic mean as the predominant value and also the maximum load recorded by the laboratory equipment as the value for each numerical analysis [10].

Each result obtained comes from static linear analysis with finite elements in a flat stress condition, considering infinitesimal deformations and applying a vertical displacement on the upper half face of the beam [14]. This paper contains information about the results of rupture modulus of concrete tested at different ages and one way to generate the stress/strain diagram using numerical methods.

2 Description of the Test

American Standard Test Method C 78-00 [4] establishes the test method for determination of the concrete bending strength, using by third-point loading method. When using molded specimens, turn the test specimen on its side concerning its position as molded and center it on the support blocks. When using sawed specimens, place the specimen so that the tension face corresponds to the top or bottom of the specimen as cut from the parent material. In Figure (1) the description of the elements involved in the test is shown.

In the bending test, it can be determined important values that characterize the concrete such as the rupture modulus, which is given by the Eq. 1

$$R = \frac{P\ L}{b\ d^2}, \tag{1}$$

where, R is the rupture modulus in kpa, P is the maximum load applied in kg, L is the distance between supports in cm, b is the average width of the specimen in cm, d is the average thickness in cm.

The parameters involved in Eq. (1) can be obtained through direct methods of measurement on the specimen except for P; the value of P is the maximum load obtained by the test. The information that can be obtained from this test is often limited to the instrumentation available, and doing post-processing turns out complicated, since only the rupture modulus is measured. One of the aims of this paper is to provide a tool for post-processing using the displacement values in the test; these deformations are very difficult to calculate in practice due to its cost. The use of other instruments as micrometers is impractical because the failure occurs in a fragile way (i.e., happens very fast), and the micrometer can be damaged when the failure occurs, making the readings almost impossible to read. Another problem is the fact that the deformations are located in a minimal range ($\sim 10^{-4}m$).

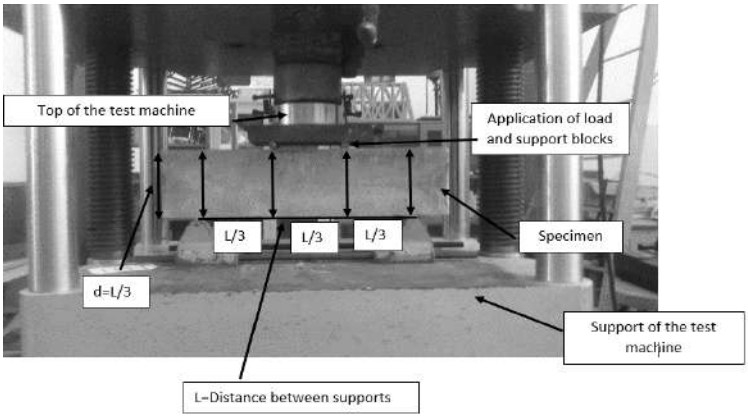


Figure 1: Flexural strength test with load in the thirds of the beam

This is the main reason for using the Finite Element analysis that can solve this problem and help to get enough information about a test that only provides scarce information and it allows to generate the stress/strain diagram to have a general overview of the behavior of the element in the the study.

2.1 Procedure Test "in situ"

This paper makes the analysis, the model, the calculation of the deformations and the calculation of the static elasticity module for a specific case, which was studied at the Civil Engineering School of the Universidad Michoacana de San Nicolás de Hidalgo located in Morelia, Michoacán, México. The known feature for this concrete is the resistance to simple design compression whose value is 39.2266 MPa. The test comprises setting loads to the element in 4 stages, the first stage comprises an applied load of 250 kg, the second of 500 kg, 750 kg and 1000 kg are the last 2 stages of load. Once the load is applied, the stabilization of the readings in the micrometers is expected to obtain the deformation data, the data are registered until the last load. For this case, just a maximum load of 1000 kg is placed.

One aim is to compare the analytical solution in tests measured in situ, and verify the approximation, to establish the consistency of the model. The model performs the calculation of the deformations and the deformation stress/strain diagram just with the maximum load got in the bending test in the laboratory, and for this purpose, the test was performed to establish if the model represents the measurements made in real experiments.

In Figure (2) the implementation for get the deformations generated is shown, which one result from the applied load.

Figure 2: Procedure and implementation of the test

Table (1) presents the loads measured in different stages.

Loading stages	load (kg)
1st load	250
2nd load	500
3rd load	750
4th load	1000

Table 1: Applied loads stages

3 Finite Element analysis

In this article, two-dimensional Euler-Bernoulli beam bilinear finite elements to produce a matching of the local and global coordinates are used. This methodology is characterized by suitable piecewise polynomials and, the problem analyzed is adequate for these conditions.

The deformation of a beam must have not continuous slope as well as continuous deflection at any two neighboring beam elements. To satisfy this continuity requirement each node has both deflection, υ and slope θ_i as nodal variables. In this case, any two neighboring beam elements have common deflection and slope at the shared nodal point.

The Euler Bernoulli beam equation is based on the assumption that the plane normal to the neutral axis before deformation remains normal to the neutral axis after deformation [18].

In this work, an efficient numerical approach to calculate the deformation generated by an applied load on a beam is proposed. The load is taken as the maximum value registered by the bending test. The process starts by proposing the number of nodes in the beam that will be discretized, the moment of inertia is calculated according to its cross-section, and the elasticity module is proposed for the initial tests. Then, the domain is discretized, and the connectivity table of the elements involved is generated, the number of elements will depend on the number of nodes previously established.

Evaluating the system $K * U = F$ and substituting the respective values the stiffness matrix for the system is obtained, as shown in the following expression:

$$K = \begin{bmatrix} 1.6554e+09 & 2.0693e+07 & -1.6554e+09 & 2.0693e+07 \\ 2.0693e+07 & 3.4488e+05 & -2.0693e+07 & 1.7244e+05 \\ -1.6554e+09 & -2.0693e+07 & 1.6554e+09 & -2.0693e+07 \\ 2.0693e+07 & 1.7244e+05 & -2.0693e+07 & 3.4488e+05 \end{bmatrix}.$$

From this information the assembly of the global stiffness matrix is carried out, which was solved considering the expression $K * U = F$ the system has the following form:

$$\begin{bmatrix} 1.6554e+09 & 2.0693e+07 & -1.6554e+09 & 2.0693e+07 \\ 2.0693e+07 & 3.4488e+05 & -2.0693e+07 & 1.7244e+05 \\ -1.6554e+09 & -2.0693e+07 & 1.6554e+09 & -2.0693e+07 \\ 2.0693e+07 & 1.7244e+05 & -2.0693e+07 & 3.4488e+05 \end{bmatrix} \begin{Bmatrix} \upsilon_1 \\ \theta_1 \\ \upsilon_2 \\ \theta_2 \\ \vdots \\ \upsilon_i \\ \theta_i \end{Bmatrix} = \begin{Bmatrix} V_1 \\ M_1 \\ V_2 \\ M_2 \\ \vdots \\ V_i \\ M_i. \end{Bmatrix}$$

Lastly, the system is solved, the deformations are obtained efficiently and, with this data, it is possible to build different diagrams.

It is very important to notice that the calculation of the maximum deformation is derived from the analysis of the previously mentioned set of equations, which shows that the inputs of the vector of terms independent of each other correspond to values of V_i and M_i, deformations and moments respectively. In this case by convention twenty-five nodes are computed, i.e., $n = arbitrary\ odd\ value$ so that $F \in \mathbb{R}^{2nx2n}$ and the entries of the vector F will be of value $2n$ i.e., has an independent vector of fifty entries, where the entry n corresponds to

the maximum deflection and the entry $n+1$ corresponds to the maximum moment, considering that the initial entry is equal to zero ($F(1,1)=0$), and thus, these increase until the highest value, from the entry $(n+1)$ the values begin to decrease with the same magnitude with which it grew until reaching zero in the last of the entries of the system, that is $F(2n,1)=0$. In this case, it is considered that the number of nodes chosen is sufficient for the flexibility of the solution since elevating the number of nodes for this case does not impact the solution and the representation, unless the objective would be to carry out a deep discretization. The information can be represented as $V_{max}=F(n,1)$ and $M_{max}=(F(n+1),1)$. The aforementioned discretization was implemented in MATLAB 2018b [20] using a personal computer (Windows 10 Pro, processor Intel (R) Core(TM) i5-3210M CPU @ 2.50GHz RAM: 8 GB). The crucial steps to solve this system are shown, and it also shows the idealization of the mesh with the finite element analysis in Figure (3) to solve the problem.

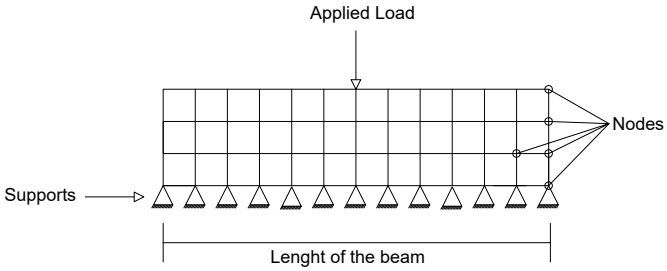


Figure 3: Idealized mesh

The methodology aforementioned is used for analyzing the behavior of a concrete element, and this kind of analysis it's important for the trend in different materials [22]. This element was tested to flexion, and the information obtained was introduced as input data for the finite element analysis.

Usually, this method requires many amounts of operations and in this work is proposed an algorithm elaborated in MATLAB, which solves this problem in an efficient way [9]. One of the principal problem with this kind of analysis is the sensibility and precision of results and this depends on the number of nodes that discretized the structure of the system.

This algorithm allows to input values that equal to features such as inertia moment, maximum load, which is obtained in the flexural strength test, length of the element, in this case, the length of the beam, the number of nodes discretizing the system and, if possible, the modulus of elasticity. The main structure of the aforementioned algorithm.

4 Numerical Tests

The blends of concrete used in the bending tests were made according to the ACI methodology [17], and the design resistance to simple compression chosen 35 Mpa. The blends are prepared during the dry season. The specimens were developed and tested under controlled conditions in the laboratory, mainly the flexure test must be carried out at uniform speed, in such a way that the stress increase of the extreme fibers does not exceed 980 Kpa/min or 10 kgf/cm^2 per minute, allowing higher speeds 50% before to the estimated load of rupture.

One must determine the average width, the thickness, and the location of the failure line, using the average of three measures: one in the middle, and two on the edges of the specimen. All of them must be precise, up to 1 mm. The resulting data as shown in table (2).

For the finite element analysis, distribution of twenty-five nodes equispaced along the beam and a moment of inertia of $60e^{-6}$ m was considered, this moment is calculated as a function of the sectional area and an elasticity

Test days	Maximum load registered(kn)	Rupture modulus(Mpa)
60	39.8640	5.42
90	41.1879	5.60
140	41.9334	5.70
180	43.1100	5.86

Table 2: Results of the flexure test in the laboratory

modulus of 35.9253 Mpa, this modulus is obtained from the resonant frequency test made in laboratory [3]. The number of nodes depends on the approximation required, in this case, twenty-five nodes are used, so the computing time is relatively short for this number, and the maximum deformation generated by the maximum load can be represented with one unique value.

For this problem, it is considered a value of time $t = 0$ for the quasi-static case. Since the crack occurs very fast in the bending test, it is to say, the time to get the crack is up to three seconds, which is quite short to measure deformations. Different analyses were performed corresponding to four test ages, the first test shows the maximum load to 60 days, the second age is for 90 days with the maximum load obtained in that test and it is the same case for the other test ages. For this reason, four diagrams are shown, one for each test age with a maximum load for each one.

It is important to mention that the results measured with the approximations obtained are data that were not generated in the test due to the restriction and difficulty of equipment implementation for taking data in this kind of test, therefore the results obtained are approximations and estimation of data that do not exist in physical tests, however, due to the conditions of this type of elements, their performance is partially known and thus the results generate a good approximation to the aforementioned performance, i.e., the approximation to an analytical solution is generated, in addition to the availability of data to be analyzed.

5 Results

For the construction of the previous graphs, the quasi-static case was considered, where even though the tests are compared at different ages, time is not considered a factor that directly influences the calculation of the deformations, this procedure is known as batch calculation. In the batch calculation, it is considered the set data that is in that stage, and it is compared with the set data that is in another stage. In this case, the test ages are considered as the different stages.

The deformations were calculated using the finite element analysis, for the construction of the previously diagrams all data obtained in the numerical solution is used, for diagrams shown in the Figures (4 - 6), it can be seen the length of the beam in the horizontal axis, which is 0.60 meters, such measure is the length of all specimens. In the vertical axis, the displacements (in meters) calculated with the FEM are shown. These Figures show the deformations that occurred by the effect of an applied load, and it is essential to say that the maximum deformation occurs with the maximum load registered in the test of bending in the laboratory. Notice that as the days go by the resistance increases, therefore, the deformations increase for each case, and we can conclude that the deformations have a direct connection with the applied load. The computed deflection and slip profiles have been compared with test results of the continuous composite beam at the specific load states, however, this is not informative to verify the computational results. [15]. At the early loading stages, both curves will increase linearly.

Figure (6) shows all deformations obtained by the numerical solution.

Based on the displacements generated through FEM, it is possible to generate the stress/strain diagram, since there is the necessary information to do so. It is important to mention that in this part the interpretation of the results obtained by FEM will be the factor to be able to generate the diagrams successfully. Once the information obtained is worked on, the stress/strain diagram is generated for each of the test ages. It is important to highlight how these diagrams were generated, the procedure is described below.

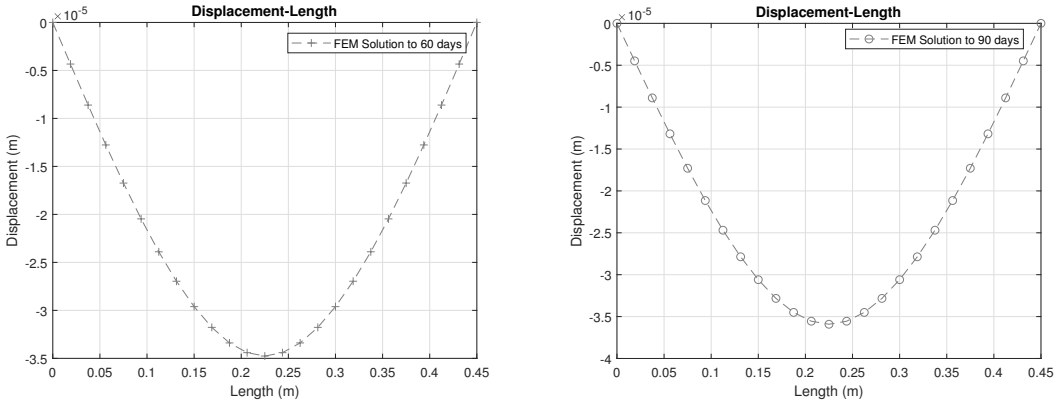


Figure 4: Numerical solution for 60 and 90 days

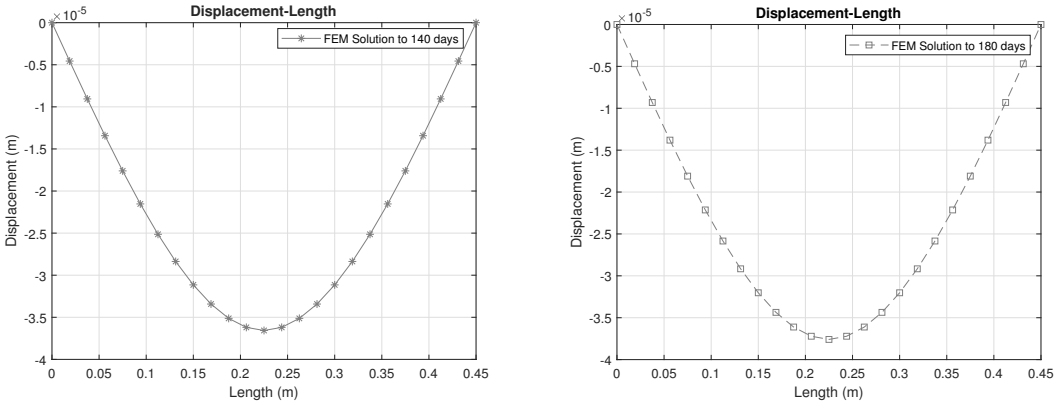


Figure 5: Plots of numerical solution for 140 and 180 days

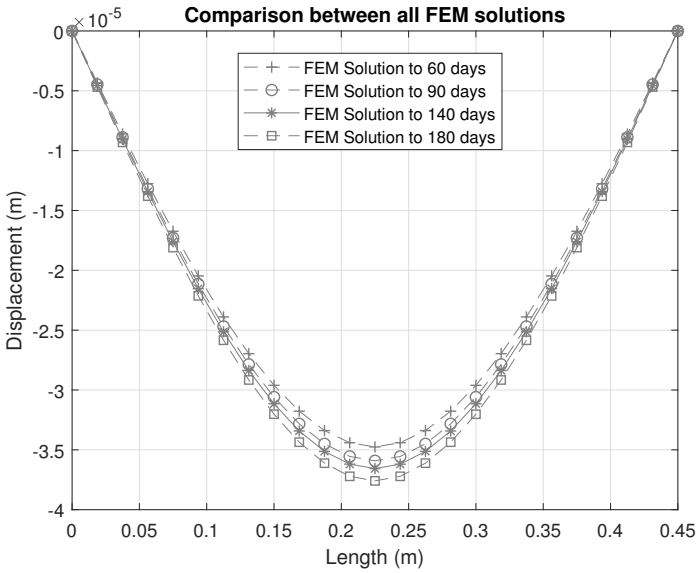


Figure 6: Plot of displacement for all test days

First, the straight line is made between the initial deformation which is a zero value, and the final deformation, this straight line is shown in the stress/strain diagram was traced joining the origin and the point corresponding to the maximum deformation, which is the cause of the maximum load applied.

For the aforementioned diagram, it is impossible to give more information as these are the only data obtained from this test, which is elaborated in the laboratory. Using the finite element analysis, it is possible to generate the second trajectory in each one of the test ages, these make an approximation to the deformation of each element, these deformations were calculated by the previous process and with this data, it's possible to generate the corresponding stress/strain diagram. Using the deformations obtained by the FEM, it is possible to construct a continuous linear function in fragments, which represents the connection of each one of the deformations calculated, since the initial deformation, which is zero until the final deformation, and in the same way, the value its equal to zero. The interesting part is that the intermediary values are the FEM data.

This data is converted in a linear function, whit known values, the initial and final deformations, therefore the constructed lines could be named such as C' function, i.e., a continuous function in pieces. One of the principal application with this data it's the analysis of the elastic zone and the plastic zone for each of the beams, this information can be identified and thus perform further analysis, that allows generating a piece of crucial information about of development and performance of this important material.

The results obtained from the analysis are shown below in Figures (7 - 8).

This paper is characterized by numerical analysis, which is based on unique data, which is the maximum load obtained from the bending test. The displacements generated by the load through the finite element analysis are approximated from it and it is possible to generate the stress/strain diagram with the analysis of equations type C1. The governing equation is the Eule-Bernoulli fourth-order partial differential equation. An efficient way to make an approximation for the determination of the displacements generated by the load is to use the formulation of Lagrange, where it is possible to use a polynomial of a lower order, for this case, a polynomial of second-order is the solution.

$$f_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}f(x_0) + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}f(x_1) + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}f(x_2)$$

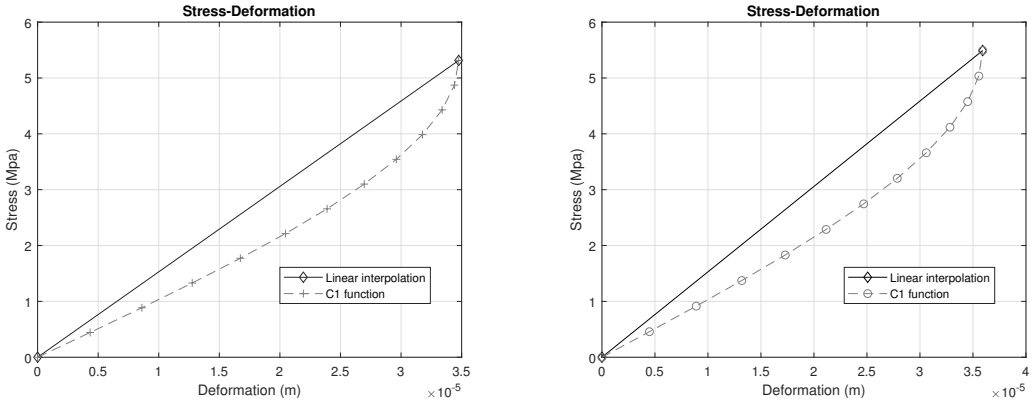


Figure 7: Solution of the linear interpolation model for 60 days and 90 days

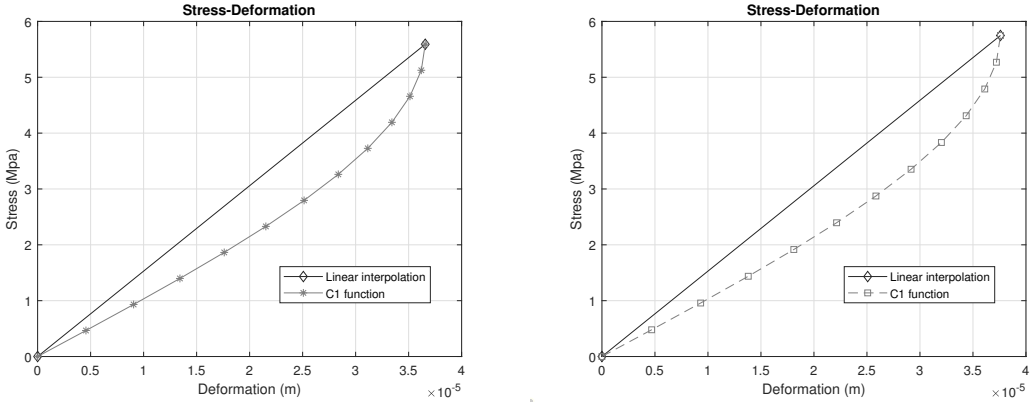


Figure 8: Solution of the linear interpolation model for 140 days and 180 days

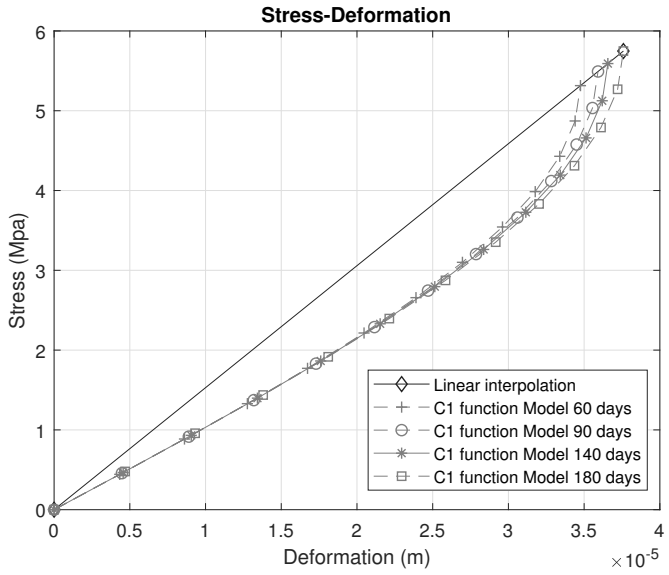


Figure 9: Comparison of the model at different test ages vs linear interpolation

Where x_0 and $f(x_0)$ are considered as the origin (0,0), the variables x_2 and $f(x_2)$ correspond to the length of the element and to the last displacement respectively, which in this case the length of the element is the length of the beam subjected to the bending test and the last displacement is equal to 0 since it is considered that the beam could reach its original form before it reaches the maximum load.

Performing the above consideration x_1 is equal to the length of half of the beam since it is considered only the distance to the central axis and $f(x_1)$ is the maximum deformation, deformation obtained by the analysis of the finite element analysis. Therefore, it is possible to make an approximation of the estimation of the deformations based on a simple configuration such as Lagrange's adaptation. The result of the aforementioned approximation is shown in the Figures (10) and (11)

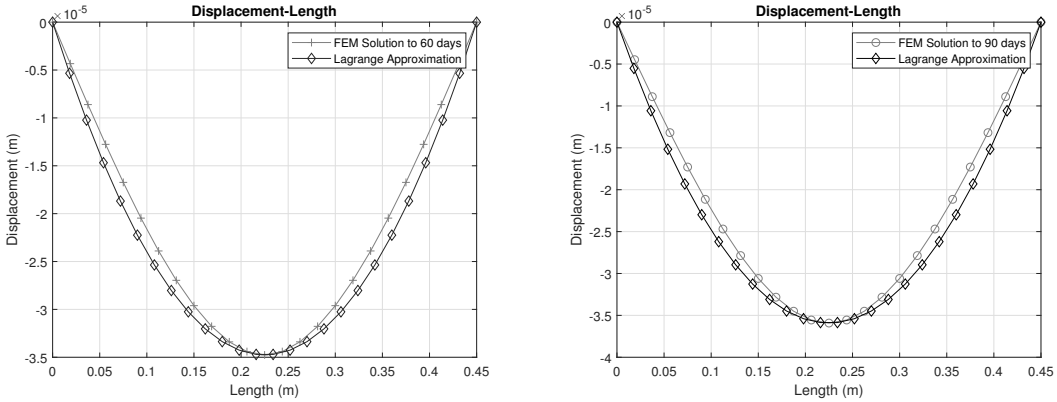


Figure 10: Numerical approximation with Lagrange for 60 and 90 days

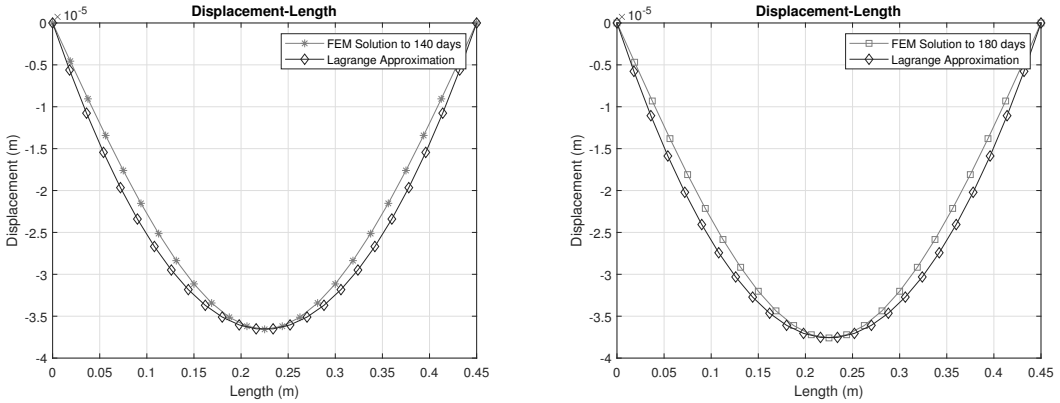


Figure 11: Numerical approximation with Lagrange for 140 and 180 days

5.1 Test in situ

The deformations obtained with the micrometers and their respective approximations are shown in Figures (12 and 13). Figures (14) and (15) show the comparative among the precision obtained by Lagrange methodology and the FEM in each applied load.

Applied load (Kg)	Deformation (mm)
250	0.006858
500	0.019050
750	0.027940
1000	0.037592

Table 3: Results of test in situ

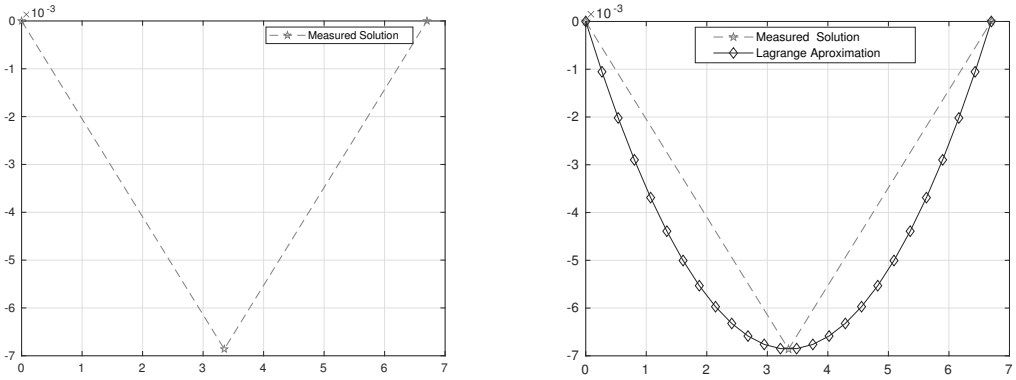


Figure 12: Measured solution and his Lagrange Approximation

6 Conclusions.

An important advantage of this procedure is an efficient way to obtain deformation stress/strain diagrams without the use of equipment such as micrometers or laser devices. Furthermore, this allows us to know each of the

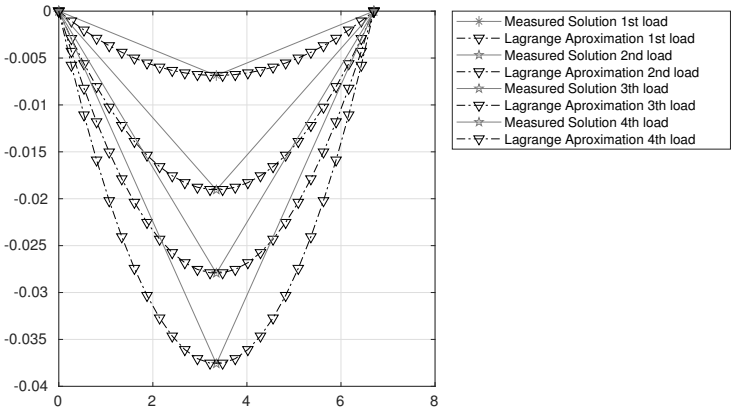


Figure 13: Measured solution and his Lagrange Approximation

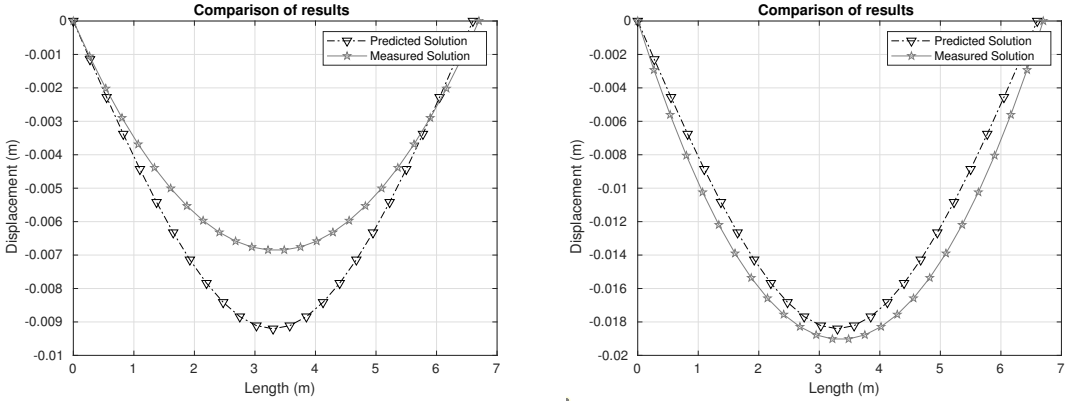


Figure 14: Behavior at 250 and 500 kg, respectively

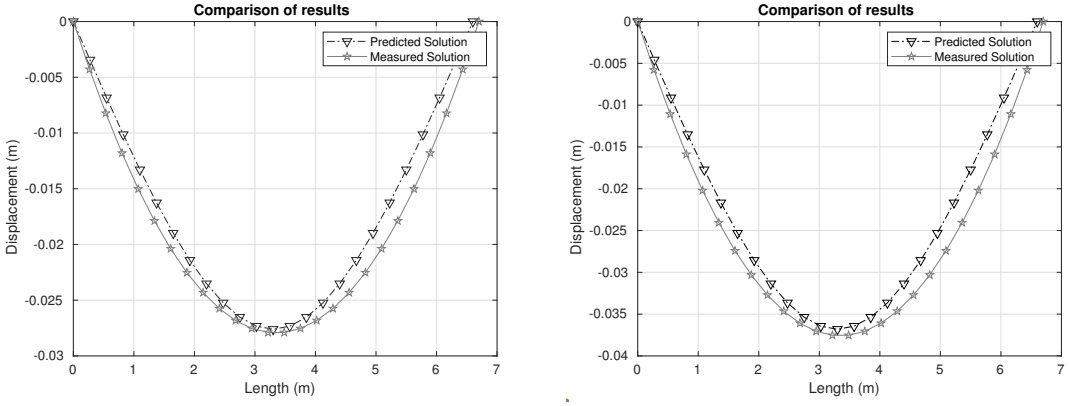


Figure 15: Behavior at 750 and 1000 kg, respectively

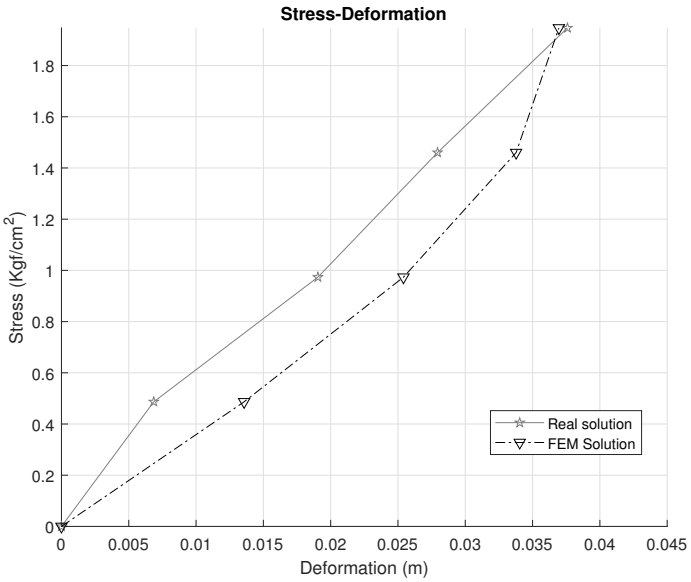


Figure 16: Stress/strain diagram for both solutions

deformations within the length of the specimen in the bending test and to consider a number n of deformations since this amount of values will be limited by the computational cost and not by physical factors. This type of data must be analyzed since deformations establish crucial parameters for slab design and rigid pavement design. Currently the analysis of this type of structure is not carried out with due detail.

The hardened concrete tests performed in the laboratory usually provide only limited specific data of the existing instrumentation, and a post-process of the information is not possible. One of the main ideas of this paper is to propose a method to compute the deformations in a concrete beam which is subject to flexure stresses using only the maximum load, by proposing a procedure to calculate the deformations generated in the beam. Thus, a stress/strain diagram can be generated to estimate the elasticity module [23].

The results obtained in this work model the behavior of an element under the application of a load. Such data can be useful for further analysis related to the performance of concrete. The methodology proposed here can be used to provide useful technical information for the construction community. Also, it is essential to mention that the results of any elastic experiment such as the bending test on concrete depend on the time of exposure to different loads because the fatigue of the object occurs

To make a stress/strain diagram using results from the laboratory, it would be necessary for a considerable number of micrometers placed along the beam to measure the deformations. The data obtained would be enough to sketch the diagram, but the performance of such measures is impractical since the instrumentation is not easy to implement. The proposal of this work is to use numerical method to solve this problem. The finite element analysis allows us to calculate the displacements generated in this test, the typology of a doubly-supported beam with a load in the center. This phenomenon is known as the problem of the Euler-Bernoulli beam equation. It is important to mention that the concrete used is a block of concrete without reinforcing steel and for this reason, the modulus of elasticity is almost zero, but using numerical analysis, we found small values for the corresponding elasticity.

Once the problem is solved with this method, the diagrams of displacements along the length of the beam can be generated. The calculation of these displacements is essential to construct the deformation stress/strain diagram since the intermediate displacements previous to the maximum displacement are used in such construction.

It is important to remark that this analysis is performed *in situ* or in the laboratory, and it provides a general overview of the performance of a certain concrete blend.

Numerical and laboratory tests allow predicting the behavior and performance of a concrete blend. It is vital to perform this analysis in an efficient way, since gathering this data may take more time and requires skill. The programming language used for this work is Matlab. One of the advantages of the proposed method is its simplicity to handle boundary conditions and it is possible to obtain extra information from the run.

The analysis with this type of methodology aims to efficiently find the numerical solution of the deformations generated by the load applied to the element. This work shows only the application of the model in a concrete mix, but it can be applied in other mixes and the numerical performance can be compared.

7 Acknowledgments.

The authors wish to thank the information provided to the materials laboratory Ing. Luis Silva Ruelas for providing the necessary equipment to obtain data. Thanks to Conacyt, CIC, and Aula-CIMNE for funding this research.

We would like to deeply thank the reviewers for their valuable comments and contributions on this work.

References

[1] G. G. Adams. Dynamic motion of two elastic half-spaces in relative sliding without slipping. *Journal of tribology*, 121(3):455–461, 1999.

[2] A. Ariaei, S. Ziaei-Rad, and M. Ghayour. Vibration analysis of beams with open and breathing cracks subjected to moving masses. *Journal of sound and vibration*, 326(3-5):709–724, 2009.

[3] C. ASTM. 125 (2000). *Standard Terminology Relating to Concrete and Concrete Aggregates, ASTM International, PA*, 2000.

[4] C. ASTM. 78-00,". *Standard Test Method for Flexural Strength of Concrete (Using Simple Beam with Third-Point Loading)," American Society for Testing and Materials*, 2000.

[5] B. Baker. Dynamic stresses created by a moving crack. *Journal of Applied Mechanics*, 29(3):449–458, 1962.

[6] F. O. Bustamante. *Estructuración de vías terrestres: vías terrestres y pavimentos*. Compañía Editorial Continental, 2002.

[7] E. Chen. Sudden appearance of a crack in a stretched finite strip. *Journal of Applied Mechanics*, 45(2):277–280, 1978.

[8] A. O. Cifuentes. Dynamic response of a beam excited by a moving mass. *Finite Elements in Analysis and Design*, 5(3):237–246, 1989.

[9] M. Cohen, A. Monteleone, and S. Potapenko. Finite element analysis of intermediate crack debonding in fibre reinforced polymer strengthened reinforced concrete beams. *Canadian Journal of Civil Engineering*, 45(10):840–851, 2018.

[10] A. Del Valle Moreno, J. Guzmán Torres, E. Alonso Guzmán, W. Martínez Molina, A. Torres Acosta, J. Terán Guillén, M. Montes Zea, A. Torres Murillo, and M. Martínez Madrid. Solicitaciones mecánicas y estáticas a concreto hidráulico simple elaborado con agregados pétreos redondeados y adicionados con fibras deshidratadas de cactus opuntia. *Publicación Técnica*, (448), 2015.

[11] S. Ghadimi and S. S. Kourehli. Multiple crack identification in euler beams using extreme learning machine. *KSCE journal of Civil Engineering*, 21(1):389–396, 2017.

[12] S. Ghadimi and S. S. Kourehli. Multi cracks detection in euler-bernoulli beam subjected to a moving mass based on acceleration responses. *Inverse Problems in Science and Engineering*, pages 1–21, 2018.

[13] H. Gökdağ. A crack identification approach for beam-like structures under moving vehicle using particle swarm optimization. *Materials Testing*, 55(2):114–120, 2013.

[14] R. Graffe and D. Linero. Simulación numérica del proceso de fractura en modo i de vigas de concreto con trayectoria de fisuración conocida mediante un modelo discreto de fisura cohesiva. *Revista ingeniería de construcción*, 25(3):399–418, 2010.

[15] P. Gu, B. Sheng, and S. Chen. Discussion of "nonlinear behaviour of steel–concrete composite bridges: finite element modelling and experimental verification". *Canadian Journal of Civil Engineering*, 39(10):1171–1172, 2012.

[16] L. Jing and J. Hudson. Numerical methods in rock mechanics. *International Journal of Rock Mechanics and Mining Sciences*, 39(4):409–427, 2002.

[17] S. H. Kosmatka, W. C. Panarese, and M. S. Bringas. *Diseño y control de mezclas de concreto*. Instituto Mexicano del Cemento y del Concreto, 1992.

[18] Y. W. Kwon and H. Bang. *The finite element method using MATLAB*. CRC press, 2018.

[19] H. Lopez-Calvo, P. Montes-Garcia, E. Alonso-Guzmán, W. Martinez-Molina, T. Bremner, and M. Thomas. Effects of corrosion inhibiting admixtures and supplementary cementitious materials combinations on the strength and certain durability properties of hpc. *Canadian Journal of Civil Engineering*, 44(11):918–926, 2017.

[20] Lowell. *MATLAB. Matlab user manual.* MA: Mathwork Inc., 2018.

[21] T. Nishioka and S. Atluri. Path-independent integrals, energy release rates, and general solutions of near-tip fields in mixed-mode dynamic fracture mechanics. *Engineering Fracture Mechanics*, 18(1):1–22, 1983.

[22] A. Nour, B. Massicotte, E. Yildiz, and V. Koval. Finite element modeling of concrete structures reinforced with internal and external fibre-reinforced polymers. *Canadian Journal of Civil Engineering*, 34(3):340–354, 2007.

[23] Y.-P. Zheng, A. Choi, H. Ling, and Y.-P. Huang. Simultaneous estimation of poisson's ratio and young's modulus using a single indentation: a finite element study. *Measurement science and technology*, 20(4):045706, 2009.

¿Quieres publicar artículos, información sobre eventos o noticias en el boletín?

La Sociedad Mexicana de Computación Científica y sus Aplicaciones A. C. (SMCCA), convoca a toda la comunidad interesada en el área de la Computación Científica y sus Aplicaciones, a presentar noticias, información sobre eventos, artículos de divulgación e investigación de alta calidad en el era, así como reportes de trabajos de tesis de nivel licenciatura y posgrado en Matemáticas Aplicadas.

Requisitos para la colaboración en el Boletín

I. Artículos de Divulgación e Investigación.
   a) Los artículos que se envíen para ser publicados deberán ser inéditos y no haber sido ni ser sometidos simultáneamente a la consideración en otras publicaciones.
   b) Los artículos a presentarse deben de ser enviados por medio de la página del Boletín: https://www.scipedia.com/sj/smcca
   c) En la página de la sociedad se puede encontrar la plantilla de LaTeX para la correcta escritura de artículos.
II. Información sobre eventos.
   a) Los eventos cuya información quiera ser publicada para promocionarlos, deberán estar relacionados con el área de las Matemáticas Aplicadas y la Computación Científica.
   b) La información debe enviarse en un archivo de imagen: PDF, JPG, PNG.
   c) La información no deberá exceder una cuartilla.
   d) Enviar la información con al menos 6 meses de anticipación a la fecha en que se llevaría a cabo.
III. Noticias.
   a) Las noticias a ser publicadas en el Boletín deben ser noticias relevantes de actividades de la SMCCA, Socios, Comunidad Científica interesada en las Matemáticas y Computación Científica.
   b) La información de las noticias debe enviarse en un archivo de imagen: PDF, JPG, PNG.
   c) La información no deberá exceder una cuartilla.

El material de colaboración, noticias e información de eventos, deberán ser dirigidos al Dr. Gerardo Tinoco Guerrero al correo electrónico de la SMCCA: smcca@smcca.org.mx

Todos los artículos son sometidos a evaluación por especialistas de instituciones nacionales e internacionales y su publicación estará sujeta a la disponibilidad de espacio en cada número. Las demás colaboraciones se someterán a corrección de estilo y su publicación estará sujeta a la disponibilidad de espacio en cada número. Sólo se aceptará el material enviado que cumpla con todos los requisitos anteriormente señalados.

El envío de cualquier colaboración al Boletín implica no solo la aceptación de lo establecido en este documento, sino también la autorización al Comité Editorial del Boletín SMCCA para incluirlo en su página electrónica, reimpresiones, colecciones y cualquier otro medio que permita lograr una mayor y mejor difusión.

Sociedad Mexicana de Computación Científica y sus Aplicaciones

Consejo directivo de la Sociedad Mexicana de Computación Científica y sus Aplicaciones 2018-2020

Presidente
Dr. Justino Alavez Ramírez.
Vicepresidente
Dra. Rina Betzabeth Ojeda Castañeda.
Secretaria de actas y acuerdos
Dra. Ma. Luisa Sandoval Solís.
Tesorero
Dr. Jorge López López.
Secretario General
Dr. Pedro Flores Pérez.
Vocal
Dr. Gerardo Tinoco Guerrero.
Vocal
Dr. Miguel Ángel Uh Zapata.

La Sociedad Mexicana de Computación Científica y sus Aplicaciones fue fundada el 16 de Mayo de 2013, para realizar actividades de investigación científica o tecnológica inscritas en el RENIECyT (Registro Nacional de Instituciones y Empresas Científicas y Tecnológicas), prestadas únicamente a los socios y asociados. Es una Asociación sin fines de lucro. Entre sus tareas fundamentales destacan: Conjuntar acciones e intereses comunes en los investigadores, profesores y estudiantes interesados en la Computación Científica y sus Aplicaciones, con el fin de fomentar la investigación de calidad, promover la actualización y el perfeccionamiento para el desarrollo científico, tecnológico y social; promover la creación, organización, acumulación y difusión de conocimientos referidos a la Computación Científica y sus Aplicaciones; promover la formación e interacción de redes y grupos de trabajo orientados hacia el desarrollo disciplinar, interdisciplinar y temático de la investigación; fomentar el desarrollo de la investigación sobre la Computación Científica y sus Aplicaciones en la República Mexicana; contribuir al mejoramiento de la enseñanza de la Computación Científica y sus Aplicaciones en la República Mexicana; promover y organizar toda clase de encuentros y eventos académicos orientados a la comunicación y discusión entre investigadores y profesores, así como también a la difusión del conocimiento hacia sectores interesados en la integración de la Computación Científica y sus Aplicaciones en los problemas de su sector.

smcca@smcca.org.mx
http://www.smcca.org.mx
https://www.scipedia.com/sj/smcca