



Estimación de la resistencia a la compresión de hormigones a los 28 días utilizando Modelos de Mezclas Gaussianas

Estimation of the compressive strength of concrete at 28 days using Gaussian Mixture Models

Fecha de entrega: 4 de octubre 2024

Fecha de aceptación: 6 de enero 2025

Rodrigo Guerra

Universidad de La Serena, Departamento de Ingeniería en Obras Civiles, La Serena, Chile, rguerra@userena.cl

El objetivo del presente trabajo es estimar la resistencia a la compresión de hormigones a los 28 días utilizando Modelos de Mezclas Gaussianas. Para desarrollar el trabajo se ha utilizado una base de datos de 82 dosificaciones de distintos hormigones y sus curvas madurez-resistencia. Esta base de datos fue poblada con información obtenida desde la literatura y completada con dosificaciones realizadas en laboratorio, las cuales fueron monitoreadas durante 28 días. Con esto, se conformó una base de datos con variables como la razón agua-cemento (A/C), el contenido de cemento y la relación temperatura-resistencia en el tiempo. Se busca estimar el valor de la resistencia a los 28 días de hormigones utilizados en Ingeniería y Construcción, en un rango de resistencias a la compresión entre 15 y 45 MPa. Con ello, se pretende representar numéricamente el comportamiento de la madurez de los hormigones en el tiempo versus el incremento de resistencia a la compresión. El modelo de Regresión de Mezclas Gaussianas GMR permitió estimar la resistencia a la compresión con un error menor a 13.54%.

Palabras clave: madurez del hormigón, resistencia a la compresión, curvas madurez-resistencia, modelos de mezclas Gaussianas

The aim of this work is to estimate the compressive strength of concrete at 28 days using Gaussian Mixture Models. To carry out the work, a database of 82 dosages of different concretes and their maturity-resistance curves have been used. This database was populated with information obtained from the literature and completed with dosages carried out in the laboratory, which were monitored for 28 days. With this, a database was created with variables such as the water-cement ratio (W/C), the cement content and the temperature-resistance relationship over time. The aim is to estimate the 28-day resistance value of concrete used in Engineering and Construction, in a range of compressive strengths between 15 and 45 MPa. This is intended to numerically represent the behaviour of concrete maturity over time versus the increase in compressive strength. The Gaussian mixture regression GMR model allowed the estimation of compressive strength with an error of less than 13.54%.

Keywords: concrete maturity, compressive strength, maturity-resistance curves, Gaussian mixture models

Introducción

El hormigón es uno de los materiales más utilizados en Ingeniería y Construcción y está compuesto principalmente por cemento, agua y áridos (Zhang *et al.*, 2021; Setién *et al.*, 2003). Ocasionalmente, se adicionan a la mezcla otros materiales o aditivos para potenciar alguna de las propiedades del hormigón, con el fin de mejorar el desempeño durante su vida útil (Abellán-García y Pineda-Varón, 2022; Ngandu, 2021). Estos materiales pueden generar cambios en la constitución interna del hormigón

y, por ello, pueden influir en las propiedades tanto físicas como mecánicas (Park *et al.*, 2005). La resistencia a la compresión es una de las propiedades mecánicas más importantes, ya que con ella se asegura durabilidad, impermeabilidad y resistencia a las cargas (Abellán-García, 2021). El hormigón se instala en obra en estado fresco, lo que genera un proceso de fraguado donde el cemento, al hidratarse con el agua, genera un fenómeno exotérmico, lo que aumenta su temperatura por sobre la temperatura del ambiente (Al-Hasani *et al.*, 2023; Croppi y Lazzaroni,



2014). Con ello, comienza un proceso de endurecimiento que evoluciona generando un incremento en su resistencia día a día. Esta resistencia se evalúa mediante ensayos destructivos para estimar la carga de rotura de las probetas cilíndricas de diámetro 15 cm y altura 30 cm. Existen métodos no destructivos para estimar la resistencia de hormigones, pero tienen una importante dispersión de resultados con respecto a los valores reales de resistencia de los hormigones (Buys, 2019). Dentro de los métodos no destructivos destaca el ensayo esclerométrico o Martillo Schmidt, entre otros (NCh3565, 2018; Rudelli y Santilli, 2017).

Otro método de estimación de la resistencia de hormigones es el método de la Madurez. Los estudios de madurez en hormigones comenzaron a fines de la década de los '40 e inicio de los años '50 y estas investigaciones demostraron que existe una relación entre la temperatura de exposición de los hormigones y el tiempo, generando un incremento de la resistencia a la compresión (Nandhini y Karthikeyan, 2021). Se intensificó su investigación en la década de los '70 debido a derrumbes en EEUU de obras de hormigones a edades tempranas, que produjeron pérdidas de vidas humanas. El accidente de Fairfax County (1973) y el de Willow Island (1978) aceleraron las investigaciones y las causas se centraron en los hormigones sometidos a cargas en edades tempranas como 1 a 4 días, en zonas donde la temperatura exterior no superaba los 10°C. Esto dejó como conclusiones que los hormigones en esas condiciones no tenían la resistencia adecuada para absorber las cargas a las cuales fueron sometidos (Miller *et al.*, 2022). Lo anterior, motivó el nacimiento de la primera norma de madurez en el mundo, la que fue desarrollada en EEUU en el año 1987 y corresponde a la norma ASTM C1074 (1987). En Chile, la norma NCh170 (2016) introduce el concepto de madurez en el estudio de hormigones y el año 2018 se aprueba la norma NCh3565 (2018) que define la metodología para determinar la madurez en hormigones.

En Ingeniería y Construcción se ha buscado predecir la resistencia del hormigón a los 28 días con información en edades tempranas y, por ello, muchas de las investigaciones han abordado la determinación de la resistencia a los 28 días basada en la resistencia a los 7 días (Abdelaty, 2014; Rudelli y Santilli, 2017). Adicionalmente, otros estudios han aplicado regresiones y redes neuronales artificiales

para predecir las resistencias del hormigón basado en la razón agua-cemento A/C, el contenido de cemento, las fracciones de áridos, el tipo de cemento, la temperatura de fraguado, entre otras variables (Abellán-García, 2021).

Existen estudios que abordan la estimación de la resistencia del hormigón mediante la madurez, pero pocos trabajos han abordado la implementación de Modelos de Mezclas Gaussianas (GMM por su sigla en inglés “Gaussian mixture models”) en dicha predicción. En la literatura se encuentran estudios asociados al empleo de modelos de mezclas Gaussianas en el área de hormigones como es el caso de Zhang *et al.* (2021) quienes investigaron sobre la construcción de relaciones difusas entre la resistencia a la compresión y las características de la imagen microestructural 3D para la hidratación del cemento, utilizando redes neuronales de función de base radial polinomial basadas en modelos de mezclas Gaussianas. En la misma línea, Chen *et al.* (2023) utilizaron modelos de mezclas Gaussianas para analizar emisiones acústicas para la identificación del tipo de grieta en columnas de hormigón corroídas bajo carga excéntrica. De igual forma, el trabajo de Farhidzadeh *et al.* (2013) para la identificación de daños y clasificación del modo de fisura en estructuras de hormigón armado, y lo desarrollado por Mayakuntla *et al.* (2023) respecto a la clasificación de la gravedad de la corrosión en estructuras de hormigón basada en modelos de mezclas Gaussianas mediante imágenes ultrasónicas, muestran el interés en la aplicación de estas metodologías en el área de la Ingeniería y Construcción.

Materiales y métodos

Modelos de mezclas Gaussianas

Machine Learning es una clase de inteligencia artificial (IA) que se refiere a un conjunto de metodologías que son capaces de detectar automáticamente patrones en los datos experimentales, sin considerar una ecuación determinada, y que luego pueden usarse para desarrollar modelos de pronóstico y así respaldar la toma de decisiones en condiciones inciertas (Sun *et al.*, 2021; Thai, 2022; Mohtasham Moein *et al.*, 2023). Es decir, nos permite hacer predicciones óptimas dada toda la información disponible, aunque esa información pueda ser incompleta o ambigua (Bishop, 2006). Lo más importante es que proporciona a los sistemas informáticos la capacidad de aprender y mejorar



por sí mismos, adaptándose a sus entornos, en lugar de ser programados explícitamente (Thai, 2022; Mohtasham Moein *et al.*, 2023).

Hay tres tipos principales de aprendizaje: supervisado (*supervised*), no supervisado (*unsupervised*) y por refuerzo (*reinforcement*). El aprendizaje descriptivo o no supervisado se asocia con problemas mucho menos definidos, donde el objetivo es descubrir relaciones subyacentes en los datos. Es decir, en el aprendizaje no supervisado, hay un conjunto de ejemplos de aprendizaje en los que sólo se conoce la cantidad de entradas y no hay información disponible sobre la salida correcta (Mohtasham Moein *et al.*, 2023). Entre los algoritmos más utilizados en el aprendizaje no supervisado se encuentran Hidden Markov models, k-means, hierarchical clustering y Gaussian mixture models (Luk *et al.*, 2010; Farhidzadeh *et al.*, 2013; Mayakuntla *et al.*, 2023).

En los últimos años la tendencia es utilizar métodos de Machine Learning para predecir las propiedades mecánicas de materiales de construcción, y en especial del hormigón (Mohtasham Moein *et al.*, 2023). Machine Learning se utiliza en ingeniería estructural en diversos campos como la evaluación del desempeño sísmico, modelado de resistencia a la tracción y resistencia a la compresión, identificación de sistemas estructurales y control de vibraciones, por nombrar algunos (Mohtasham Moein *et al.*, 2023).

La distribución de una variable normal está completamente determinada por dos parámetros, su media (μ) y su desviación estándar (σ). Dicha distribución decrece de forma simétrica a ambos lados en función del valor de la desviación estándar. Para una distribución Gaussiana de una variable x , la función de densidad de probabilidad $F(x)$, para $-\infty \leq x \leq \infty$, es la mostrada en la fórmula (1) (Vidya Sagar, 2018)

$$F(x) = G(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

Si bien la distribución Gaussiana tiene algunas propiedades analíticas relevantes, adolece de limitaciones importantes cuando se trata de modelar algunos conjuntos de datos reales (Bishop, 2006). En la Figura 1 se puede ver que el conjunto de datos reales (línea negra) forma tres grupos dominantes y que una distribución Gaussiana simple no

puede capturar esta estructura, mientras que una superposición lineal de varias Gaussianas (líneas roja, azul y verde) proporciona una mejor caracterización del conjunto de datos reales.

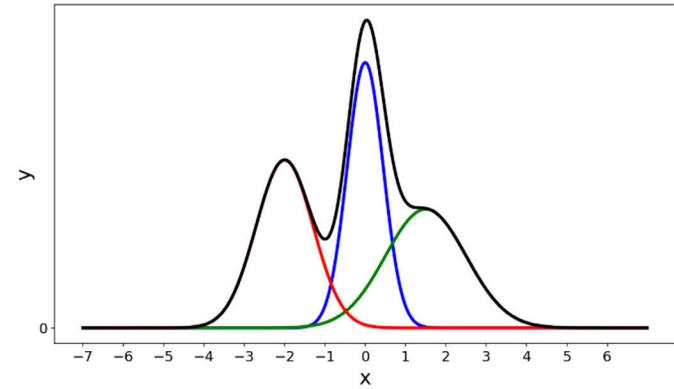


Figura 1: Densidad de probabilidad normal. Aproximación de datos reales (línea negra) con distribución Gaussiana multivariante.

Utilizando un número suficiente de gaussianas y ajustando sus medias y covarianzas, así como los coeficientes en la combinación lineal, casi cualquier densidad continua puede aproximarse con una precisión arbitraria. Por lo tanto, se considera una superposición de K densidades Gaussianas de la forma:

$$p(x) = \sum_{k=1}^K \pi_k N(x|\mu_k, \Sigma_k) \quad (2)$$

Lo anterior, se llama mezcla de Gaussianas $N(x|\mu_k, \Sigma_k)$ y cada distribución Gaussiana se denomina componente de la mezcla y tiene su propia media μ_k y covarianza Σ_k (Bishop, 2006). Los parámetros π_k en la fórmula (2) se denominan coeficientes de mezcla. Si se integra ambos lados de (2) con respecto a x , y se observa que tanto $p(x)$ como las componentes Gaussianas individuales están normalizadas, se obtiene:

$$\sum_{k=1}^K \pi_k = 1 \quad (3)$$

Esto implica que $\pi_k \geq 0$ para todo k , es decir, que $0 \leq \pi_k \leq 1$. En muchas aplicaciones, sus parámetros están determinados por estimadores de máxima verosimilitud, normalmente utilizando el algoritmo EM, por sus siglas en inglés “expectation-maximization” (Bishop, 2006).



Regresión de mezclas Gaussianas

La Regresión de Mezclas Gaussianas (GMR por su sigla en inglés “Gaussian mixture regression”) es un enfoque de regresión que modela distribuciones de probabilidad en lugar de funciones. Por tanto, es posible modelar mapeos multimodales (Fabisch, 2021). Ghahramani y Jordan (1994) propusieron por primera vez la regresión de mezcla Gaussiana mediante el algoritmo EM. Posteriormente, Calinon *et al.* (2007) introdujeron el término regresión de mezcla Gaussiana en el contexto del aprendizaje por imitación para trayectorias de robots y le siguieron muchas publicaciones que utilizan GMR en este dominio como el trabajo de Stulp y Sigaud (2015).

Para la aplicación de GMR primero se entrena una distribución de probabilidad conjunta $p(x,y)$ de la entrada x y la salida y , mediante el algoritmo EM, y luego se calcula la distribución condicional $p(y|x)$ para hacer predicciones. Por tanto, el entrenamiento es el mismo procedimiento que en un modelo de mezcla Gaussiana estándar. Fabisch (2021) destaca que durante la fase de entrenamiento se prepara un modelo de mezcla Gaussiana a través de EM.

$$p(x, y) = \sum_{k=1}^K \pi_k N_k \left(x, y \mid \mu_{xy_k}, \Sigma_{xy_k} \right) \quad (4)$$

Donde $N_k \left(x, y \mid \mu_{xy_k}, \Sigma_{xy_k} \right)$ son distribuciones Gaussianas con media μ_{xy_k} y covarianza Σ_{xy_k} , K es el número de Gaussianas y $\pi_k \in [0,1]$ que suman uno. En la etapa de predicción, la regresión de mezcla Gaussiana se puede utilizar para predecir distribuciones de variables, calculando la distribución condicional $p(y|x)$. La distribución condicional de cada Gaussiano individual es:

$$N \left(x, y \mid \mu_{xy}, \Sigma_{xy} \right) \quad (5)$$

$$\mu_{xy} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad \Sigma_{xy} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \quad (6)$$

Esta queda definida por las fórmulas (7) y (8).

$$\mu_{y|x} = \mu_y + \Sigma_{yx} \Sigma_{xx}^{-1} (x - \mu_x) \quad (7)$$

$$\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \quad (8)$$

En un modelo de mezcla Gaussiana se calcula la distribución condicional de cada Gaussiano individual y sus anteriores según la fórmula (9).

$$\pi_{y|x_k} = \frac{N_k \left(x \mid \mu_{x_k}, \Sigma_{x_k} \right)}{\sum_{l=1}^K N_l \left(x \mid \mu_{x_l}, \Sigma_{x_l} \right)} \quad (9)$$

Para luego obtener la distribución condicional:

$$p(y|x) = \sum_{k=1}^K \pi_{y|x_k} N_k \left(y \mid \mu_{y|x_k}, \Sigma_{y|x_k} \right) \quad (10)$$

GMR puede ser usado para predecir la distribución de variables, a través del cálculo de la distribución condicional $p(y|x)$ (Fabisch, 2021). Si GMR es usado para predecir valores de y desde valores de x , se denomina regresión o *forward analysis*, y si se requiere predecir valores de x desde valores de y , se denomina *inverse analysis*, esta habilidad predictiva de GMR es muy relevante (Kaneko, 2021).

Relación madurez - resistencia en hormigones

La norma ASTM C1074 (1987) permite al usuario expresar el índice de madurez, ya sea en términos del factor temperatura-tiempo usando la ecuación de Nurse-Saul o en términos de la edad equivalente a una temperatura específica usando la ecuación de Arrhenius (Abdel-Jawad, 2006). En Chile, la norma ha definido el método de Nurse-Saul como la ecuación a aplicar en la determinación de la Madurez en hormigones (NCh3565, 2018). La fórmula (11) está asociada al modelo de madurez de Nurse-Saul.

$$M(t) = \sum_0^t (T - T_0) \cdot \Delta t \quad (11)$$

Donde $M(t)$ es la función de Madurez ($^{\circ}\text{C}\cdot\text{h}$), T es el promedio de la temperatura en cada intervalo de tiempo ($^{\circ}\text{C}$), T_0 es la temperatura Datum ($^{\circ}\text{C}$) y Δt es el intervalo de tiempo (h), que la norma define en 30 min. La temperatura Datum representa la temperatura a la cual el hormigón comienza a ganar resistencia. En Chile, el valor para la temperatura datum recomendado es 0°C para las regiones con temperaturas cálidas la mayoría del año y



de -10°C para regiones del extremo sur o zonas de bajas temperaturas (NCh3565, 2018). La integración del área bajo la curva mostrada en la Figura 2, es la representación gráfica de la madurez de Nurse-Saul.

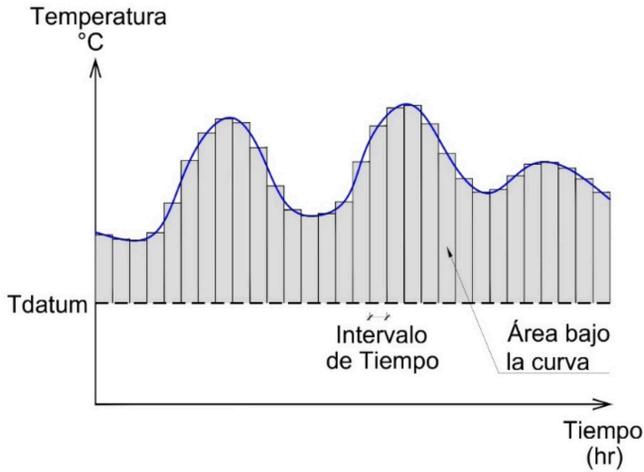


Figura 2: Representación gráfica de la madurez

El seguimiento en el tiempo de la temperatura permite la recolección de la información a través del curado de la probeta, para este estudio la medición se realizó en intervalos de 15 min durante 28 días. La recolección de datos de temperatura se realizó con un sensor DS18b20. El sensor de temperatura DS18b20 se conectó mediante una protoboard a una placa Arduino Uno para la recolección de los datos de temperatura como se muestra en las Figuras 3(a) a 3(c).

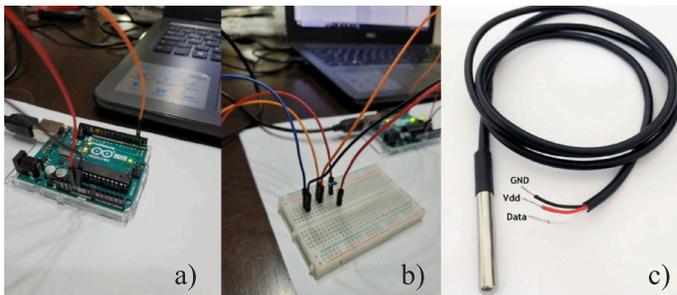


Figura 3: a) y b) Conexión y c) sensor de temperatura DS18b20

Como ejemplo, se presentan los resultados de madurez de un hormigón G-36, es decir con resistencia a los 28 días de 36 MPa, que tiene una razón agua-cemento de 0.46 y la cantidad de cemento es 395 kg/m^3 . La curva mostrada en la Figura 4 corresponde al hormigón G-36 y está basada en el factor temperatura-tiempo a 1, 3, 5, 7, 14, 21 y 28 días. Esta es la curva de calibración de la relación

madurez-resistencia y de ella se puede estimar el valor de la resistencia en el tiempo, asociándolo al valor de la madurez desarrollada por el hormigón hasta dicho instante de evaluación.

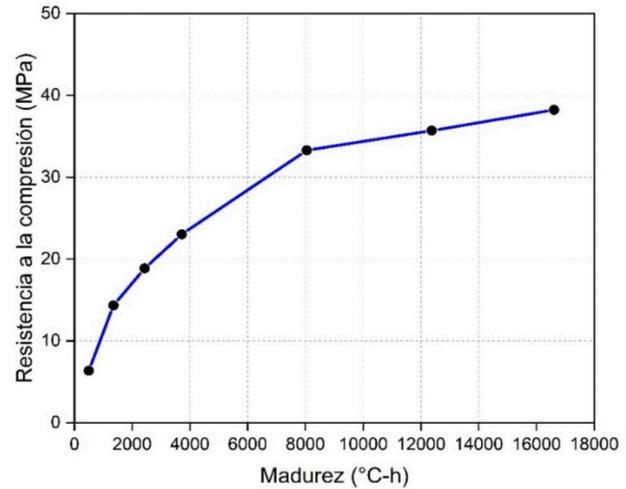


Figura 4: Curva Madurez – Resistencia hormigón de laboratorio

Para la recolección de información en este trabajo se implementó un sistema de adquisición de datos con seis sensores DS18b20 midiendo en paralelo. En la Figura 5 se muestra el sistema de recolección de datos que se debe implementar con la placa Arduino, la protoboard y los seis sensores de temperatura.

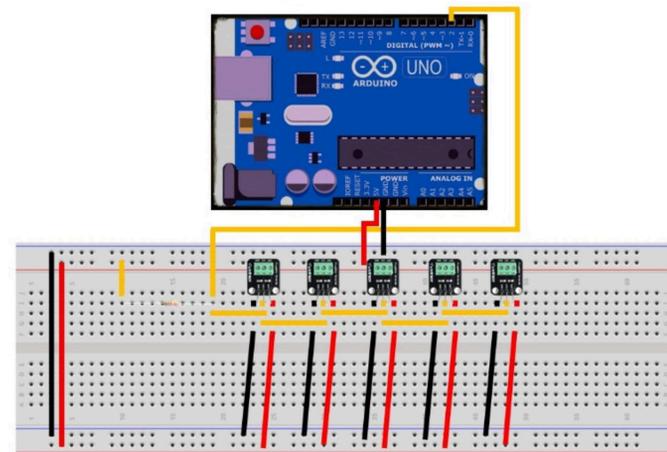


Figura 5: Esquema del circuito Arduino para la recolección de datos

Dataset de trabajo y variables de entrada

Para la generación del dataset de trabajo se ha alimentado esta base de datos desde dos fuentes. Una de ellas



corresponde a la información disponible como fuente abierta en la literatura técnica del área hormigones (41 datos) y la otra fuente de datos son los generados mediante la dosificación de mezclas de laboratorio (41 datos). Los datos recolectados han sido obtenidos de tesis doctorales y de pregrado, artículos en revistas internacionales y trabajos de laboratorios ejecutados en el Departamento de Ingeniería en Obras Civiles de la Universidad de La Serena. Para ello, se ha recopilado información asociada a la curva característica de hormigones que relaciona Madurez y Resistencia a la compresión y datos relevantes sobre sus dosificaciones como cantidad de cemento, razón agua – cemento, entre otros parámetros.

Los datos generados en laboratorio responden a dosificaciones de hormigones de prueba donde se modificaron gradualmente las cantidades de materiales constituyentes del hormigón. Esto permitió controlar las cantidades de cada componente del hormigón y sus temperaturas de curado. Sin embargo, algunas de las características de los componentes se han mantenido constantes como es el tipo de árido y el tipo de cemento. Para ello, se ha variado la cantidad de cemento de 302 a 445 kg/m³ y por consecuencia la razón agua –cemento varía de 0.40 a 0.60. En las Figuras 6(a) a 6(c) se puede ver la confección y curado de probetas cilíndricas de hormigón en laboratorio.



Figura 6: a) Moldes, b) fabricación y c) curado de probetas cilíndricas de hormigón

Estos hormigones fueron curados con muestras en piscina a temperaturas cercanas a los 23°C durante los 28 días y otro set de probetas gemelas fue curado a temperatura ambiente, lo que en Chile corresponde aproximadamente a 15°C en la ciudad de La Serena, entre los meses de mayo a agosto en que se realizó este estudio. Para cada una de las dosificaciones se tomaron probetas cilíndricas, las cuales

fueron ensayadas a compresión a los 7 días, la que será considerada como variable de entrada en el modelo, y a los 28 días, la que será considerada como la resistencia o valor target. Se elaboró una probeta adicional para insertar el sensor de temperatura y así registrar la temperatura durante los 28 días en intervalos de 15 min. Lo anterior, permitió tomar datos para conformar la curva Madurez – Resistencia.

Es posible encontrar en la literatura distintos autores que han trabajado en la predicción de valores de la resistencia a los 28 días, mediante Machine Learning, empleando distintas variables de entrada y métodos como Redes Neuronales, entre otros (Abellán-García, 2021; Ngandu, 2021; Abellán-García y Pineda-Varón, 2022).

Se ha tenido en consideración variables inputs del modelo alimentadas por la cantidad de cemento, razón agua cemento, madurez a los 3 y 7 días, la temperatura media de curado y la resistencia a la compresión a los 7 días. La variable output o target es la resistencia a la compresión a los 28 días. Los datos se han tabulado de forma de facilitar su entendimiento y se muestran sólo algunas de sus filas en la Tabla 1.

Tabla 1: Dataset utilizado en el estudio

A/C	Cemento, kg	Madurez 3 días, °C-h	Madurez 7 días, °C-h	Resistencia 7 días, MPa	T media curado, °C	Resistencia 28 días, MPa
0.41	390	2013	4509	16.96	29.798	29.13
0.47	336	2564	6539	31.70	41.624	37.20
0.45	320	1129	3178	26.60	20.385	43.20
...
0.57	395	1137.29	3640.40	13.01	21.145	24.14
0.59	395	1035.65	2065.63	9.37	12.389	18.52
0.59	395	1230.46	3847.75	11.22	22.372	22.58

Para considerar la madurez a los 28 días se ha introducido un valor medio de la temperatura de curado (T_{meq}), que resulta al dividir la madurez a los 28 días por 672 horas totales. En la Figura 7 es posible ver la evolución de la temperatura durante el curado para un hormigón con A/C igual a 0.46, lo que permite calcular T_{meq} igual a 24.716°C.

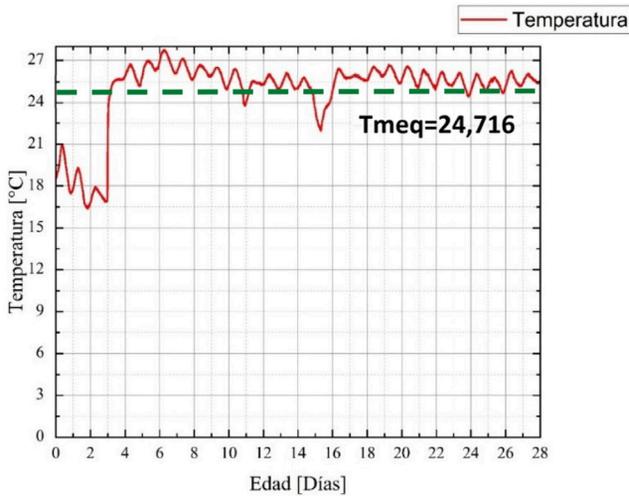


Figura 7: Curva temperatura-tiempo para una probeta de hormigón de laboratorio, A/C = 0.46

El dataset compuesto por 6 inputs y 1 output se ha dividido en un arreglo de entrenamiento y otro de testeo. Además, las distintas variables han sido estandarizadas para obtener distribuciones con una media igual a cero y una desviación estándar igual a 1.

Generación y evaluación del modelo GMR con Python

Python es un lenguaje muy utilizado en el trabajo de modelos de mezclas Gaussianas y Regresiones de Mezclas Gaussianas GMR (Fabisch, 2021; Kaneko, 2021, 2022). Los trabajos en esta línea han permitido simplificar la generación de regresiones multivariantes utilizando mezclas Gaussianas, creando librerías muy completas que permiten su utilización como complementos de las librerías Python scikit-learn.

Fabisch (2021) presenta la librería *gmr* aplicable a regresión multivariante y entrega el fundamento teórico del análisis asociado. Esta información está disponible en <https://github.com/AlexanderFabisch/gmr>. En la misma línea, Kaneko (2021) introduce la librería *dcekit* y otros materiales complementarios disponibles en <https://github.com/hkaneko1985/dcekit>.

El modelo GMR generado en este trabajo considera el código Python, librería *dcekit*, el formato de impresión de gráficas y la determinación de estadígrafos como R^2 , $RMSE$ y MAE ; se han realizado algunas modificaciones

de idioma y formato, con el fin de realizar la regresión de mezclas Gaussianas (Kaneko, 2021).

Indicadores estadísticos

La evaluación del desempeño del modelo GMR se realizó a través del estudio de tres indicadores estadísticos o estadígrafos. Para ello se utilizó el error absoluto medio (MAE), la raíz del error cuadrático medio ($RMSE$) y el coeficiente de determinación múltiple (R^2). Estos indicadores estadísticos se muestran en las fórmulas (12) a (14).

$$MAE = \frac{1}{n} \sum_{i=1}^n (|y_{est} - y_{target}|) \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (|y_{est} - y_{target}|)^2} \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{est} - y_{target})^2}{\sum_{i=1}^n y_{target}^2} \quad (14)$$

Donde n es el número de datos, y_{est} corresponde a los valores de resistencia estimados o predichos por el modelo GMR e y_{target} son los valores reales de resistencia determinados mediante ensayos de laboratorio a los 28 días.

El indicador estadístico MAE se calcula como las diferencias absolutas entre los valores target y los valores estimados por el modelo GMR. El $RMSE$ se determina como la raíz cuadrada del error residual promedio mediante la comparación de los valores estimados por el modelo con los valores target. Este error cuadrático medio se determina en las mismas unidades de la variable target, es decir, en MPa en nuestro caso. En general, un valor de MAE y $RMSE$ cercano a 0 demuestran un buen ajuste del modelo. Sin embargo, el $RMSE$ proporciona más peso a los datos de mayor magnitud (Abellán-García y Pineda-Varón, 2022). R^2 se define como la proporción de la varianza total de la variable explicada por la regresión. Las estadísticas de R^2 dependen de las relaciones lineales entre los valores estimados por el modelo y los valores target y, en ocasiones, pueden proporcionar resultados sesgados

cuando esta relación no es lineal o cuando los valores contienen muchos valores atípicos (Abellán-García, 2021). Un valor de R^2 cercano a la unidad, permite comprobar que el desempeño del modelo es adecuado. Un uso combinado de los indicadores estadísticos puede proporcionar una estimación objetiva de la capacidad de predicción del modelo (Abellán-García y Pineda-Varón, 2022; Abellán-García, 2021).

• **Análisis *k-folds* cross validation**

En este trabajo existe un número reducido de observaciones (82 datos) y, por ello, el conjunto de validación o testeo terminaría siendo muy pequeño para la validación del modelo GMR. Por lo tanto, los indicadores estadísticos de validación pueden cambiar significativamente dependiendo de los datos que fueron seleccionados para la validación y los que fueron seleccionados para el entrenamiento (Abellán-García, 2021). Esto podría llevar a una evaluación poco confiable del modelo GMR. Por el contrario, se podrían elegir valores para k que generen sobreentrenamiento del modelo, logrando que los datos se adapten perfectamente en el entrenamiento y con ello el modelo estaría perdiendo su capacidad de generalización (Acuña *et al.*, 2014).

Una solución utilizada para evitar esta situación consiste en aplicar *k-folds* cross validation, cuyo esquema se muestra en la Figura 8 para un valor de $k = 5$ como ejemplo. Este procedimiento se basa en dividir los datos disponibles en k particiones, crear k modelos idénticos y entrenar a cada uno de ellos en $k - 1$ particiones mientras se evalúa con la partición restante. El indicador estadístico de validación para el modelo considerado es entonces el promedio de

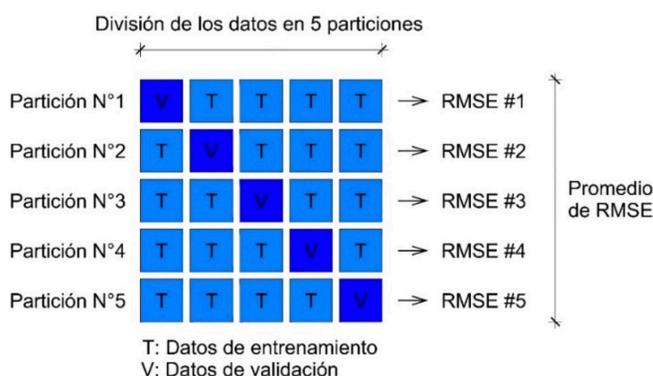


Figura 8: Esquema de aplicación validación cruzada tipo *k-fold* para $k = 5$ (Abellán-García, 2021)

los k indicadores estadísticos de validación obtenidos (Abellán-García y Pineda-Varón, 2022).

Resultados y discusión
Análisis de sensibilidad *k-folds* cross validation

Para validar los parámetros elegidos en la generación del modelo GMR óptimo, se ha realizado un análisis de sensibilidad entre el número de componentes que forman el modelo de regresión de mezclas Gaussianas y el número de *k-folds* para la validación cruzada. Para ello, se generaron distintos modelos para los valores del número de componentes GMR y *k-fold*. Este análisis de sensibilidad de los indicadores estadísticos se basó principalmente en los valores obtenidos para *RMSE* y R^2 considerando un número de componentes Gaussianas que va entre 2 y 10 y valores para *k-folds* entre 2 y 15. Los resultados obtenidos del análisis de sensibilidad para el grupo de testeo se muestran en la Figura 9.

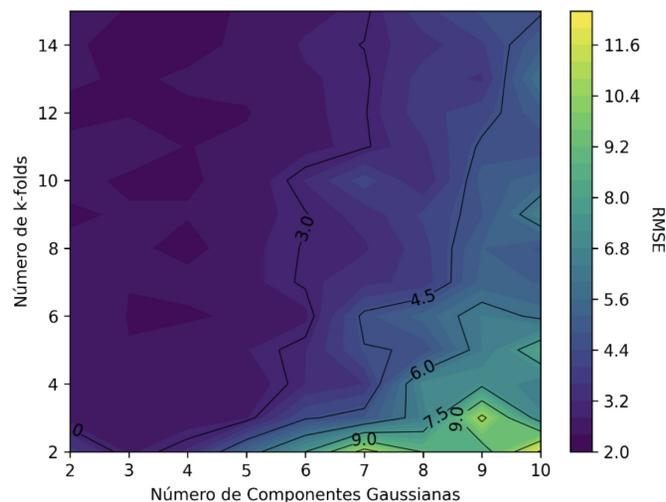


Figura 9: Análisis de sensibilidad *RMSE*

El análisis de sensibilidad nos permitió definir que el desempeño óptimo del modelo GMR se logra con 10 particiones o *k-fold* = 10, lo que representa una división de 90% de datos train y un 10% de datos test. En cuanto al número de componentes, los mejores resultados para *RMSE* y R^2 se obtienen para 3 componentes Gaussianas.

Para un número de componentes mayores a 3 se produce un aumento sostenido del valor de R^2 en los datos de entrenamiento, lo que indica un buen ajuste de los datos, pero al mismo tiempo se puede ver, en la Figura 9, un aumento del valor de *RMSE*, lo que muestra un sobreentrenamiento



u overtraining del modelo, empeorando el desempeño de los datos test.

Resultados de la aplicación del modelo GMR a la predicción de resistencias

En esta sección se muestran los resultados obtenidos con la base de datos dividida en 90% de datos train y 10% de datos test y se han considerado 3 componentes Gaussianas. Se ha agregado el coeficiente de determinación R^2 y los errores $RMSE$ y MAE obtenidos para el caso train y el caso test. En la Figura 10, se muestra el desempeño del modelo GMR de los datos train y en la Figura 11 se muestra el desempeño de los datos test. En las siguientes gráficas, las líneas de color verde delimitan un error de 10% y las líneas azules delimitan un 15% de error en las predicciones.

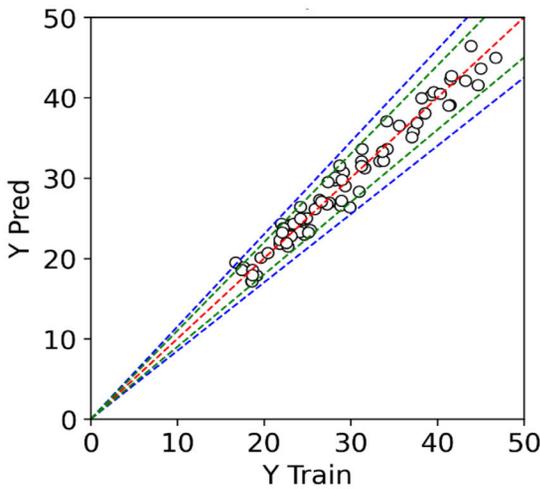


Figura 10: Correlación entre datos de entrenamiento (train) y valores estimados (pred)

El ajuste de los datos de entrenamiento reporta un valor de R^2 igual a 0.964. El valor de $RMSE$ es 1.503 MPa y el valor de MAE es 1.251. Se muestra en la Figura 11 los resultados de la aplicación del modelo entrenado previamente para los datos de testeo.

Estos datos de testeo nos permiten determinar un valor de R^2 de 0.905. El valor de $RMSE$ es 1.883 MPa y el valor de MAE es 1.571. Lo anterior, permite observar el adecuado desempeño del modelo GMR entrenado y aplicado a los datos de testeo. En la Tabla 2 se entregan los resultados obtenidos para los datos train y datos test. El modelo ha demostrado un adecuado desempeño para predecir el valor de la resistencia a la compresión a los 28 días, con un error que no supera el 13.54% y principalmente menor a 8.66%.

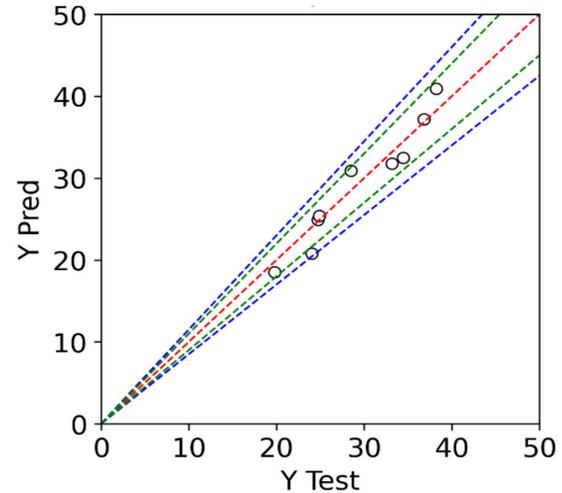


Figura 11: Correlación entre datos de testeo (test) y valores estimados (pred)

Tabla 2: Resultados indicadores estadísticos para datos train o test

Datos	R^2	$RMSE$	MAE
Train	0.964	1.503	1.251
Test	0.905	1.883	1.571

Al comparar los valores estimados mediante el modelo GMR y los valores obtenidos en laboratorio, es posible evaluar el nivel de ajuste de la propuesta y establecer la pertinencia del modelo de predicción.

Conclusiones

Este trabajo propone un modelo basado en Regresión de Mezclas Gaussianas GMR para predecir la resistencia a la compresión a los 28 días para hormigones en un rango entre 15 y 45 MPa. Para evitar el sobreentrenamiento del modelo se utilizó *k-folds cross validation* y se realizó un análisis de sensibilidad que permitió definir la mejor configuración de partición de entrenamiento con 90% de los datos y testeo con 10% de los datos. Al mismo tiempo se determinó el número de 3 componentes Gaussianas como óptimo. Los resultados combinados de los estadígrafos MAE , $RMSE$ y R^2 proporcionaron información que permitió confirmar la calidad del modelo GMR. El ajuste que se logró en los datos de entrenamiento definió un valor para R^2 de 0.964 y en el caso de los datos de testeo se obtuvo un valor de R^2 de 0.905, $RMSE$ de 1.883 MPa y MAE de 1.571. La Regresión de Mezclas Gaussianas permite hacer una



predicción adecuada de las resistencias a la compresión en hormigones con una precisión que define un error menor a 13.54%. El modelo GMR propuesto puede ser útil en la toma de decisiones para el desarrollo de dosificaciones de hormigón, incorporando condiciones como tipo de curado, temperaturas previstas en el proceso y las cantidades de los componentes del hormigón como son el agua y el cemento para una estimación preliminar.

Referencias

- Abdelaty, M. abd allah. (2014). Compressive strength prediction of Portland cement concrete with age using a new model. *HBRC Journal* **10**(2), 145–155. <https://doi.org/10.1016/j.hbrej.2013.09.005>
- Abdel-Jawad, Y.A. (2006). The maturity method: Modifications to improve estimation of concrete strength at later ages. *Construction and Building Materials* **20**(10), 893–900. <https://doi.org/10.1016/j.conbuildmat.2005.06.022>
- Abellán-García, J. and Pineda-Varon, F.A. (2022). Neural network-based model to predict compressive strength of concrete incorporating supplementary cementitious materials and recycled aggregates. *Revista Materia* **27**(2). <https://doi.org/10.1590/s1517-707620220002.1318>
- Abellán-García, J. (2021). Predicción basada en redes neuronales de la resistencia a compresión a los 7 días de los UHPC incorporando SCM. *Revista Materia* **26**(4). <https://doi.org/10.1590/s1517-707620210004.1380>
- Acuña, L., Torre, A.V., Moromi, I. y García, F. (2014). Uso de las redes neuronales artificiales en el modelado del ensayo de resistencia a compresión de concreto de construcción según la norma ASTM C39/C 39M. *Información Tecnológica* **25**(4), 3–12. <https://doi.org/10.4067/S0718-07642014000400002>
- Al-Hasani, L.E., Perez, G., Herndon, H.N., Park, J., Poole, J.L., Tien, I., Washburn, N.R., Cho, Y.K., Gentry, T.R. and Kurtis, K.E. (2023). Prediction of heat of hydration of cementitious systems using Gaussian process regression enables mass concrete thermal modeling. *Materials and Structures* **56**:45, <https://doi.org/10.1617/s11527-023-02134-8>
- ASTM C1074 (1987). Standard practice for estimating concrete strength by the maturity method. ASTM International, West Conshohocken, PA, USA
- Bishop, C. (2006). *Pattern recognition and Machine Learning*. Springer New York, USA. <https://doi.org/10.53759/7669/jmc202404020>
- Buys, C.E. (2019). *Investigation of a practical application of the maturity method to estimate the early-age strength of concrete*. MEng thesis, Stellenbosch University, South Africa
- Calinon, S., Guenter, F. and Billard, A. (2007). On learning, representing, and generalizing a task in a humanoid robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **37**(2), 286–298. <https://doi.org/10.1109/TSMCB.2006.886952>
- Chen, Z., Zhang, G., He, R., Tian, Z., Fu, C. and Jin, X. (2023). Acoustic emission analysis of crack type identification of corroded concrete columns under eccentric loading: A comparative analysis of RA-AF method and Gaussian mixture model. *Case Studies in Construction Materials* **18**, e02021. <https://doi.org/10.1016/j.cscm.2023.e02021>
- Croppi, J.I. y Lazzaroni, M.C. (2014). Modelo de elementos finitos para el análisis térmico de estructuras hormigonadas por etapas. *Obras y Proyectos* **16**, 51-58. <http://dx.doi.org/10.4067/S0718-28132014000200004>
- Fabisch, A. (2021). gmr: Gaussian Mixture Regression. *Journal of Open Source Software* **6**(62), 3054. <https://doi.org/10.21105/joss.03054>
- Farhidzadeh, A., Salamone, S. and Singla, P. (2013). A probabilistic approach for damage identification and crack mode classification in reinforced concrete structures. *Journal of Intelligent Material Systems and Structures* **24**(14), 1722–1735. <https://doi.org/10.1177/1045389X13484101>
- Ghahramani, Z. and Jordan, M. (1994). Supervised learning from incomplete data via an EM approach. In J. Cowan, G. Tesauro, J. Alspector (eds.), *Advances in Neural Information Processing Systems* **6**, 120–127
- Kaneko, H. (2021). Extended Gaussian mixture regression for forward and inverse analysis. *Chemometrics and Intelligent Laboratory Systems* **213**, 104325. <https://doi.org/10.1016/j.chemolab.2021.104325>
- Kaneko, H. (2022). True Gaussian mixture regression and genetic algorithm-based optimization with constraints for direct inverse analysis. *Science and Technology of Advanced Materials: Methods* **2**(1), 14–22. <https://doi.org/10.1080/27660400.2021.2024101>



- Luk, B.L., Liu, K.P., Tong, F. and Man, K.F. (2010). Impact-acoustics inspection of tile-wall bonding integrity via wavelet transform and hidden Markov models. *Journal of Sound and Vibration* **329**(10), 1954–1967. <https://doi.org/10.1016/j.jsv.2009.11.038>
- Mayakuntla, P.K., Ganguli, A. and Smyl, D. (2023). Gaussian mixture model-based classification of corrosion severity in concrete structures using ultrasonic imaging. *Journal of Nondestructive Evaluation* **42**(2), 1–20. <https://doi.org/10.1007/s10921-023-00939-9>
- Miller, D., Ho, N.M. and Talebian, N. (2022). Monitoring of in-place strength in concrete structures using maturity method – An overview. *Structures* **44**, 1081–1104. <https://doi.org/10.1016/j.istruc.2022.08.077>
- Mohtasham Moein, M., Saradar, A., Rahmati, K., Ghasemzadeh Mousavinejad, S.H., Bristow, J., Aramali, V. and Karakouzian, M. (2023). Predictive models for concrete properties using machine learning and deep learning approaches: A review. *Journal of Building Engineering* **63**, 105444. <https://doi.org/10.1016/j.jobe.2022.105444>
- Nandhini, K. and Karthikeyan, J. (2021). The early-age prediction of concrete strength using maturity models: a review. *Journal of Building Pathology and Rehabilitation* **6**:7. <https://doi.org/10.1007/s41024-020-00102-1>
- NCh3565 (2018). Hormigón - Estimación de la resistencia mecánica- Método de la madurez. Instituto Nacional de Normalización INN, Santiago, Chile
- NCh170 (2016). Hormigón - Requisitos generales. Instituto Nacional de Normalización INN, Santiago, Chile
- Ngandu, C. N. (2021). Prediction of compressive strengths for rice husks ash incorporated concrete, using neural network and reviews. *Iteckne* **18**(2), 99–107. <https://doi.org/10.15332/iteckne>
- Park, K.B., Noguchi, T. and Plawsky, J. (2005). Modeling of hydration reactions using neural networks to predict the average properties of cement paste. *Cement and Concrete Research* **35**(9), 1676–1684. <https://doi.org/10.1016/j.cemconres.2004.08.004>
- Rudeli, N. y Santilli, A. (2017). Medición de resistencia a tempranas edades del hormigón: método que mejor se ajusta para la determinación de tiempos mínimos de desencofrado de elementos verticales de hormigón. *Obras y Proyectos* **22**, 6-16. <https://dx.doi.org/10.4067/S0718-28132017000200006>
- Setién, J., Carrascal, I.A., Figueroa, J.F. y Polanco, J.A. (2003). Aplicación de una red neuronal artificial a la dosificación de hormigones premezclados. *Materiales de Construcción* **53**(270), 221–224
- Stulp, F. and Sigaud, O. (2015). Many regression algorithms, one unified model: A review. *Neural Networks* **69**, 60–79. <https://doi.org/10.1016/j.neunet.2015.05.005>
- Sun, H., Burton, H.V. and Huang, H. (2021). Machine learning applications for building structural design and performance assessment: State-of-the-art review. *Journal of Building Engineering* **33**, 101816. <https://doi.org/10.1016/j.jobe.2020.101816>
- Thai, H.T. (2022). Machine learning for structural engineering: A state-of-the-art review. *Structures* **38**, 448–491. <https://doi.org/10.1016/j.istruc.2022.02.003>
- Vidya Sagar, R. (2018). Verification of the applicability of the Gaussian mixture modelling for damage identification in reinforced concrete structures using acoustic emission testing. *Journal of Civil Structural Health Monitoring* **8**(3), 395–415. <https://doi.org/10.1007/s13349-018-0284-5>
- Zhang, L., Oh, S.K., Pedrycz, W., Yang, B. and Han, Y. (2021). Building fuzzy relationships between compressive strength and 3D microstructural image features for cement hydration using Gaussian mixture model-based polynomial radial basis function neural networks. *Applied Soft Computing* **112**, 107766. <https://doi.org/10.1016/j.asoc.2021.107766>