

# Modelo de estimación del impacto del INEGI en la investigación y divulgación científica

## Model for Estimating INEGI's Impact on Scientific Research and Dissemination

José Román Herrera Morales,\* Luis Francisco Barbosa Santillán,\*\* Jorge Rafael Gutiérrez Pulido,\* María Andrade Aréchiga\* y Sara Sandoval Carrillo\*

En el ámbito de la investigación, el Instituto Nacional de Estadística y Geografía (INEGI) es reconocido como un proveedor confiable de información; sin embargo, no cuenta con un perfil de autor en sistemas especializados de información científica como *SCOPUS*, *ORCID*, *Google Académico* y otros. Esto significa que no es posible medir su impacto de forma directa haciendo uso de tales sistemas. Este trabajo propone una metodología para medirlo en el contexto científico, que considera no solo indicadores bibliométricos tradicionales sino, también, los altmétricos. Las tareas que se llevaron a cabo para tal fin son: identificación de fuentes de información científica en la web, recuperación de esta de las publicaciones y de perfiles de autores en redes sociales, así como la estandarización de los datos recuperados. Posteriormente, se procedió a la identificación de aquellos descriptores útiles para generar una métrica del impacto en la producción científica del INEGI; se analizó la contribución de cada uno de ellos aplicando técnicas de aprendizaje automático, como la regresión gradual con adición de

In the scientific context, INEGI is highly recognized as a reliable information provider. Despite this, it has not an author profile in specialized scientific information systems such as *SCOPUS*, *ORCID*, and *Google Scholar* among others. This means that its impact cannot be measured directly from them. This study proposes a methodology to measure such an impact, considering both traditional bibliometric and altmetric indicators. The tasks that were carried out for the creation of a reference database of publications that cite INEGI (BDREF) are: identification of sources of scientific information on the web, retrieval of this information from publications and authors' profiles in social networks, as well as the standardization of the data retrieved. Afterwards, we proceeded to identify those descriptors useful for generating an Impact Index to the scientific production of the INEGI (IIPCI). The significance of each of these descriptors was analyzed, applying machine learning techniques such as gradual regression with the addition of significant variables, for the construction of an opti-

\* Universidad de Colima, rherrera@uocol.mx, jrpg@uocol.mx, mandrad@uocol.mx y sary@uocol.mx, respectivamente.

\*\* Universidad Tecnológica de Puebla, luis.barbosa@utpuebla.edu.mx

variables significativas, para la construcción de un modelo de predicción optimizado; se validó la métrica empleando un subconjunto de datos representativos y se obtuvieron resultados que pueden ser equiparables; se aplicó dicha métrica al periodo 1984-2018, observándose notables incrementos del impacto científico del Instituto en cada uno de los últimos tres lapsos sexenales (20, 32 y 40 %, respectivamente). Asimismo, se tuvo la mayor tasa de crecimiento por año (26 %) del 2011 al 2012 y un máximo histórico de 6.9 unidades en el 2016.

**Palabras clave:** medición de impacto; indicadores bibliométricos; índices alométricos; minería de datos.

Recibido: 12 de febrero de 2021.

Aceptado: 6 de mayo de 2021.

mized prediction model. The IIPCI was validated using a subset of representative data and comparable results were obtained. Applying the IIPCI in the period from 1984 to 2018, notable increases were observed in each of the last three six-year periods (20%, 32% and 40% respectively). Likewise, there was the highest growth rate per year (26%) in the period 2011-2012 and a historical maximum of 6.9 units in 2016.

**Key words:** impact measurement; bibliometric index; altmetrics; data mining.



Human hands using laptop for working his business having Data Analysis Diagram/ Forrest9 / iStock

## Introducción

Para medir cualitativamente el grado de impacto que una institución, artículo, patente o libro ejerce en los trabajos que se producen dentro de la comunidad científica, se han propuesto diversas métricas; una de las más usadas son los indicadores bibliométricos. Dentro de estos destacan el número de citas, el factor de impacto (FI) y el Índice H (IH). El primero es el más utilizado en bibliometría de impacto debido a que cuantifica de forma directa la cantidad de veces que una investigación es usada para construir otras; además, es la base para obtener el FI y el IH (Velasco, 2012).

Esta propuesta que presentamos se enmarca en el proyecto 290379 aprobado en la convocatoria S0025-2016-02 del Fondo Sectorial Consejo Nacional de Ciencia y Tecnología (CONACYT)-Instituto Nacional de Estadística y Geografía (INEGI) donde, a petición del Laboratorio de Microdatos del INEGI, se busca conocer el impacto de la información que en este se genera, los servicios especializados que ofrece y responder preguntas como: ¿quiénes están utilizando fuentes de datos del Instituto para fundamentar sus investigaciones?, ¿qué tipos de publicaciones son las que más están citando al INEGI?, ¿de qué países e instituciones son los autores de estas publicaciones?, ¿cómo se ha comportado esta labor de apoyo a la investigación a lo largo del tiempo?, ¿qué tantas publicaciones científicas utilizan la información de Instituto para fundamentar su trabajo?, ¿cómo cuantificar el impacto científico que ha tenido el INEGI a través del tiempo con los datos que procesa y ofrece a la población? Las respuestas a estas interrogantes fueron plasmadas en el informe técnico de la primer etapa del proyecto (Herrera-Morales, 2018), fase que se enfocó en la identificación de referencias de uso a la información que el INEGI genera para, enseguida, rastrear menciones o citas que permitan medir su impacto no solo desde la perspectiva tradicional de la cantidad de citas que sus documentos reciben, sino empleando otros indicadores que dejan, también, conocer su impacto en otros contextos, como las redes sociales y la web. En las siguientes secciones se mencionan tanto los indicadores bi-

bliométricos tradicionales como los alométricos que fueron considerados como componentes de nuestra propuesta.

## Principales métricas de impacto científico de publicaciones y revistas

El FI es una de las más importantes que se emplean en el contexto científico para expresar la calidad de trabajos de investigación; es una razón que se obtiene con la cantidad de citas recibidas por los artículos publicados en una revista, divididos entre el número de los publicados en un determinado periodo (Garfield, 1955), el cual es, típicamente, dos años. Sin embargo, esta métrica no expresa de forma directa la calidad de un artículo de investigación sino, más bien, el de la revista donde se publicó; así, una de mayor calidad es aquella que cuenta con un FI más elevado debido a que los estudios publicados en ella fueron más leídos y citados. Desafortunada o afortunadamente para los autores, al ser aplicado este indicador sobre un grupo de trabajos, da lugar a que algunos, sin ningún mérito en cuanto a citas, se cuelguen de la fama de otros más citados, dado que se trata de un promedio, en realidad (Seglen, 1997). Por ello, es importante enfatizar que este indicador hace referencia al impacto de la revista y no al de las publicaciones que cada autor tiene.

Hirsch (2005) estableció el IH, que evalúa la producción científica de un investigador, combinando su número de publicaciones y la cantidad de citas que ha recibido. Se puede usar, también, para dar a conocer la relevancia de una institución; asimismo, detecta los autores más destacados en un área de conocimiento. No obstante, resulta inadecuado comparar entre los de diferentes áreas, pues penaliza de manera injusta a los que publican poco, pero con una gran cantidad de citas recibidas en sus trabajos. Por estas deficiencias y la queja de la comunidad científica al ser evaluado su rendimiento y el de sus instituciones con estos indicadores, se desarrolló una serie de recomendaciones sobre la adecuada evaluación de la investigación (DORA, 2018) que establecen que el FI no debe usarse

como la única manera para evaluar personal, sino que también deberían utilizarse otras métricas basadas en el contenido científico que consideren diferentes aspectos (su visibilidad y la cantidad de citas recibidas, entre otros), como los indicadores altmétricos.

En fecha reciente se han publicado trabajos donde se intenta predecir el impacto que tendrá una publicación científica empleando sofisticados algoritmos que imitan el pensamiento humano, por ejemplo, las redes neuronales profundas (Abrishami, 2019) o considerando la cantidad de citas tempranas y el FI dentro de un modelo de regresión lineal (Abramo, 2019). Sin embargo, en estas investigaciones, la métrica importante sigue siendo la cantidad de citas recibidas por las publicaciones.

### **Indicadores altmétricos para medición de impacto científico**

Como consecuencia del abuso de las bibliométricas para etiquetar qué institución o investigador merece ser premiado, se ha generado un estancamiento en el desarrollo científico debido a que se persigue a toda costa el publicar en revistas con alto FI abandonando algunas de las demás alternativas en la producción científica, como la participación en congresos o el desarrollo de proyectos complejos y, por ende, tardados, generando monocultivos científicos que conducen al empobrecimiento de la ciencia y la uniformización del pensamiento (Delgado-López-Cózar y Martín-Martín, 2019). Debido a estas problemáticas, a partir del 2010 surgieron nuevas propuestas sobre indicadores de impacto asociados a la web, denominadas *altmetrics* o métricas alternativas, y se presentan como un complemento a la evaluación de la actividad científica (Bornmann, 2014). La idea que subyace en ellas es que las menciones en blogs, noticieros, *facebook*, número de lecturas, tuits y descargas hechas en la web se consideren, también, como indicadores del impacto de las publicaciones científicas en contextos de mayor acceso y cobertura de la información para la sociedad en general (Torres-Salinas, 2013).

A pesar de lo complejo que resulta cuantificar este tipo de métricas alternativas, hoy en día, por fortuna, existen productos de empresas especializadas que ofrecen estos servicios. Entre los principales destacan: *Mendeley* de Elsevier, *Lagotto* de PLOS, *OurResearch* (antes *ImpactStory*), *PlumX* de SCOPUS, así como las plataformas *Crossref* y *Altmetric* (Ortega, 2020). La mayoría de estos proveedores emplean el *Document Object Identifier (DOI)* para registrar las interacciones de cada publicación en los diversos portales de internet desde donde se accede a estos recursos.

El uso de métricas alternativas (indicadores altmétricos) ha cobrado relevancia con el paso de los años; por ejemplo, Eysenbach (2011) demostró que, a partir de la divulgación de una publicación mediante tuits, se puede incrementar su número de citas. También, se ha evaluado la correlación que tienen los indicadores altmétricos con los bibliométricos, obteniéndose resultados variados, dependiendo del área de estudio. Nocera, Boyd, Boudreau, Hakim y Rais-Bahrami (2019) encontraron una muy baja correlación y Mullins *et al.* (2020), una alta. Sin embargo, todos los autores de estos trabajos coinciden en una idea en común: que las métricas alternativas captan el impacto sobre una dimensión diferente y deben ser usadas como un complemento de medición a los indicadores bibliométricos tradicionales (Bardus, 2020). En esta propuesta son consideradas como un elemento que permite medir el impacto con respecto a la contribución que realiza cada autor y publicación en los entornos de las redes sociales y en la divulgación científica en la web.

### **Proveedores de información científica**

Se tiene una amplia oferta de fuentes y servicios especializados de esta: *ScienceDirect* de SCOPUS y *Mendeley* de Elsevier, *Web of Science* de Clarivate Analytics (antes Thompson-Reuters), *Plos-One* de PLOS, *Dimensions*, *PubMed* de la Biblioteca Nacional de Medicina de Estados Unidos de América, *EBSCO Host*; para el habla hispana, están *REDALYC* y *Scielo* y, en particular para México, el Repositorio

Nacional del CONACYT. Además, están los servicios especializados de indexado y búsqueda, como *Google Académico*, *Microsoft Research Academic*, entre otros, que hacen uso de diversas herramientas y algoritmos especializados de recuperación de información científica y académica. Todos los anteriores fueron evaluados como posibles fuentes de datos y, con los seleccionados, se emplearon diversas técnicas para la recuperación de la información relevante de las publicaciones de interés para este proyecto.

## Propuesta de medición de impacto para el INEGI

El Instituto es un organismo autónomo del Estado mexicano, cuyo objetivo es la generación y difusión de datos estadísticos y geográficos del país en aspectos como el territorio, el medio ambiente, la población, la economía y el gobierno. En el ámbito de la investigación, es reconocido como un proveedor confiable de información oficial y tiene presencia a través de las publicaciones de investigadores tanto nacionales como internacionales en las que hacen referencia al Instituto. Sin embargo, se identificó que no la tiene como autor en sistemas especializados de información científica; es decir, no cuenta con un perfil en *SCOPUS*, *ORCID*, *Google Académico* y otros, lo cual se traduce en que no se puede tener una determinación de cuáles son las publicaciones que ha generado y, por lo tanto, no se puede hacer una identificación, trazabilidad y medición de su impacto *cientiométrico*.

La idea de esta propuesta considera, primero, identificar qué publicaciones utilizan al INEGI como referencia en sus investigaciones y, enseguida, recuperar e integrar diversos indicadores (tanto bibliométricos tradicionales como alométricos) que permitan medir el impacto de una forma integral, donde se consideran varios aspectos, no solo las citas. Esta información fue recopilada de varios sistemas especializados y, con ella, se formó la Base de Datos de Referencias Bibliográficas (BDREF). Nuestra metodología permite medir el impacto de las publicaciones del INEGI y hace uso de la BDREF

para crear un modelo de estimación de su impacto con respecto a su uso en los ámbitos académico y científico.

En la siguiente sección se proporciona una descripción amplia de la metodología seguida en la construcción de indicadores del impacto relacionados con la producción científica. Asimismo, se presentan los resultados y las discusiones relacionadas con ellos. Finalmente, en la última sección se muestran las conclusiones de este trabajo.

## Metodología

Esta consiste en dos etapas:

- Generación de la BDREF.
- Índice del Impacto en la Producción Científica del INEGI (IIPCI):
  - Identificación de predictores candidatos.
  - Construcción del IIPCI mediante la selección de predictores significativos.

A continuación, se detalla el proceso que se llevó a cabo en cada una de ellas.

### Generación de la BDREF

Para su formación, fue necesaria, en primera instancia, la identificación de fuentes de información especializadas, como las mencionadas arriba en la subsección *Proveedores de información científica*. Enseguida, en una selección de estas fuentes de datos, se llevó a cabo la extracción de las etiquetas documentales o atributos descriptivos de las publicaciones, mismas que se clasificaron en dos grupos: las *citantes* (de terceros que utilizan al INEGI como referente) y las *citadas* (documentos generados por el Instituto que fueron usados como referencia) (Herrera-Morales, 2019). También, se realizó la integración y homogenización de los datos para asegurar su calidad, corrigiéndose varias inconsistencias, como: registros incompletos, duplicados, parecidos con ambigüedad de datos, con caracteres especiales y codificaciones de idiomas diferen-

tes. Además, la BDREF incluye los indicadores altmétricos recuperados de cada publicación *citante* y la información sobre los perfiles de los autores recuperados de redes sociales, que servirán como un insumo de datos para la construcción del IIPCI.

## Índice del Impacto en la Producción Científica del INEGI

En esta subsección presentamos cada uno de los pasos que se siguieron para construirlo:

- Identificación de 12 predictores candidatos de la información en la BDREF.
- Obtención y normalización de la tabla de datos de cada predictor por año (periodo 1984-2018).
- Generación del Modelo Base de Impacto (MBI) con los 12 predictores, tomando una suma ponderada de estos.
- Optimización del Modelo mediante la técnica de regresión gradual con adición de variables.
- Reconstrucción del MBI.
- Comparación gráfica de los comportamientos temporales del modelo optimizado contra el MBI.

### Identificación de predictores candidatos

De la BDREF se eligieron 12 indicadores (predictores) para estimar el impacto del Instituto: publicaciones INEGI *citadas*, publicaciones *citantes* al INEGI, citas de las publicaciones *citantes*, penalización por citas imprecisas (ver cuadro A1), búsquedas del término INEGI en *Google*, Índice Altmétrico de las Publicaciones *Citantes* (IAPC), publicaciones *citantes* con DOI, publicaciones *citantes* de acceso abierto, Índice Ponderado de Autores (IPA, *ResearchGate Score* e IH), Índice Ponderado de Redes de Colaboración de Autores de Publicaciones *Citantes* (IPRCAPC), proporción de publicaciones *citantes* entre publicaciones INEGI por año, proporción por año entre el número de citas de las publicaciones *citantes* y la cantidad de publicaciones *citantes*.

Cabe mencionar que los primeros predictores son indicadores directos que fueron recopilados

y asociados a sus respectivas publicaciones, mientras que otros representan variables compuestas que llevaron un procesamiento adicional para tener un significado; este es el caso del IAPC, que contabiliza y pondera cada uno de los indicadores altmétricos recuperados para cada publicación *citante* que fue identificada (Martínez-Barajas *et al.*, 2021) (ver cuadro A2). Situación similar es la del IPRCAPC, donde se hace un análisis de los coautores de cada publicación *citante* y se les asigna un peso de acuerdo con la cantidad de autores y sus respectivas nacionalidades (Mosqueda-González *et al.*, 2021) (ver cuadro A3). Para una mayor descripción de ellos, ver el *Anexo A*.

Para obtener los valores observados por año de cada uno de los 12 predictores, se utilizaron expresiones SQL para extraer la información de la BDREF y se *normalizaron* los resultados para que todos los valores quedaran en un rango entre 0 y 1.

### Generación del MBI con los 12 predictores

Se ha utilizado la información de los predictores mencionados considerando los valores normalizados para tener un MBI como el que se expresa en la fórmula 1:

$$MBI_i = X_1 + X_2 + X_3 - X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10} + X_{11} + X_{12} \quad (1)$$

donde:

$i$  = impacto obtenido en cada uno de los años de interés (1984-2018).

En nuestro caso, hemos considerado la posible contribución de 12 predictores para construir el índice por lo que, a continuación, se hace un análisis de la relevancia de cada uno de estos.

### Selección de los indicadores más significativos

Cuando se tienen múltiples predictores en un modelo se debe analizar que la información que aportan no esté cubierta por otro predictor, por lo tanto, es necesario optimizar este modelo. Existen diversos métodos cuantitativos para realizar este

tipo de análisis, como el de componentes principales (PCA) y selección de variables, entre otros, que se emplean cuando tenemos muchas variables explicativas y queremos saber cuáles son las que aportan algo nuevo. Para ello, primero se hace un análisis de dependencia de variables predictoras mediante el cálculo de la matriz de correlación de Pearson. Así, las  $m$  variables que presenten entre ellas valores altos de dependencia lineal se separarán de forma momentánea del modelo. Después, evaluamos la posible reincorporación de cada una de ellas, siguiendo la técnica de regresión gradual con adición de variables descrita en Guyon & Elisseeff (2003), la cual considera, en un inicio, todos los predictores disponibles en el MBI y, con un procedimiento iterativo de regresión, se procede a seleccionar el conjunto óptimo de estos, junto con cada uno de sus pesos asociados. A continuación, se describe de forma detallada la metodología.

En el primer paso, se dividen los 12 predictores (conjunto A) en dos subconjuntos, uno con los no correlacionados (conjunto B) y otro con el complemento de este, es decir, los correlacionados (conjunto C).

Después, formamos un subconjunto de entrenamiento  $E(i)$  mediante la unión del conjunto B y el primer elemento del C. Entrenamos con el algoritmo gradiente descendente (Sebastian Ruder, 2017) el subconjunto en turno  $E(i)$  para obtener un modelo de predicción candidato MRli. En el siguiente paso, calculamos y almacenamos el estadístico *T-student* asociado al predictor tomado del conjunto C, siempre y cuando su valor- $p$  sea menor a 0.05, es decir, que sea significativo para el modelo. Repetimos todo este procedimiento, pero ahora con el siguiente predictor del conjunto C y así, sucesivamente, hasta terminar con todos los elementos de este conjunto; es decir, cuando el número de iteración  $i$  sea igual al de elementos  $n(C)$ . Después, movemos del conjunto C al B el predictor  $m(i)$  que presentó el mayor valor absoluto del estadístico *T*. Calculamos el error estándar del modelo entrenado con el conjunto actualizado B. Se verifica si se cumple la condición de paro, la cual establece que el error estándar sea menor o igual al umbral de 0.1; si no sucede así, se vuelve a iterar.

Al terminar todo este proceso —es decir, cuando se cumpla la condición de paro— obtenemos un modelo optimizado y reducido con  $n(B)$  predictores, junto con sus pesos óptimos asociados (ver diagrama), para la predicción del impacto científico del INEGI, el cual presentamos en la sección de resultados.

#### Reconstrucción del MBI

El algoritmo de gradiente descendente tiene doble importancia en el método de medición del impacto, puesto que se utiliza para realizar el entrenamiento y, además, cuando ya se tiene el conjunto de predictores (conjunto D) optimizado, se usa también para la reconstrucción del MBI. El peso asignado en el algoritmo a cada variable explicativa se establece aleatoriamente en primera instancia. Este, al buscar ir reduciendo el error, va ajustando por sí mismo los pesos  $\theta$  en cada repetición. El ajuste se da de manera gradual según la tasa de aprendizaje indicada; además, es guiado por la derivada o gradiente del error, de ahí su nombre. Por último, después de haberse repetido un número determinado de veces, alcanza el mínimo global de la función de costo, también conocida como función de error. Cuando el algoritmo converge al mínimo valor de error, los pesos óptimos del modelo de regresión son alcanzados. A continuación, se escribe matemáticamente la técnica utilizada. Sea:

$$J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2} m \sum_{i=1}^m h_{\theta}(x_i - y_i)^2 \quad (2)$$

donde:

$$h_{\theta}(x) = \theta_0 X_0 + \theta_1 X_1 + \dots + \theta_n X_n \quad (3)$$

$\theta_j$  = pesos de regresión.

$J$  = función de error.

$\alpha$  = tasa de aprendizaje, con valor inicial de 0.03.

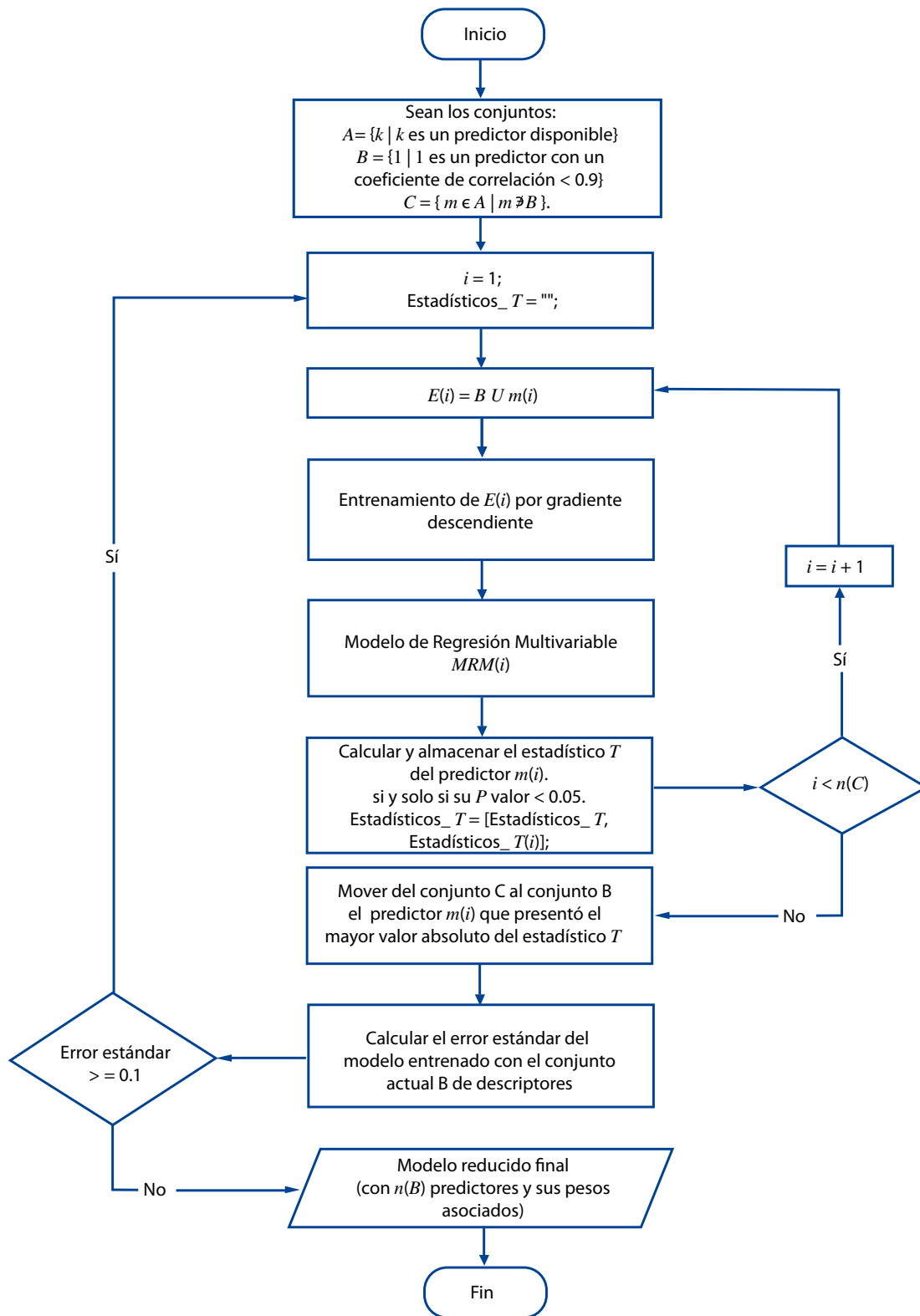
$h$  = función de regresión.

$$\text{Hacer } \{\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_j)\} \quad (4)$$

simultáneamente actualizar  $\theta_j$  repetir hasta que converja  $\theta_j$ .

Entonces, considerando los pesos óptimos generados por este algoritmo junto con los predic-

## Metodología empleada para obtener el modelo optimizado para cuantificar el impacto científico del INEGI





tores o variables asociadas, se obtiene el Modelo de Regresión Multivariable (MRM), referido en la metodología propuesta. Al final, con el modelo optimizado se podrán hacer comparaciones contra el MBI y evaluar si corresponde a una representación apropiada del original.

## Resultados

En esta sección se describen los principales hallazgos obtenidos de las dos etapas propuestas en la metodología, la obtención de la base de datos de referencias y la identificación de predictores útiles para generar los modelos IIPCI.

### Generación de la base de datos de referencias

Esta contiene la integración de las referencias de las publicaciones *citantes* al INEGI en el periodo 1984-2018 que fueron identificadas y recolectadas de las fuentes de datos seleccionadas *SCOPUS*, *REDALYC*, *Mendeley*, *Google Académico*, *Core API* y *Dimensions* (cf. arriba subsección *Generación de la BDREF de Metodología*). Para llevar a cabo su proceso de extracción, se desarrollaron herramientas de *software* que emplearon técnicas de obtención de datos para la recuperación de los atributos de las publicaciones, en particular *web scraping* o el empleo de sus respectivas API (Castrejón-Mejía *et al.*, 2020, Alvarado-Villa *et al.*, 2019, Madrigal-Martínez *et al.*, 2018).

La BDREF (basada en el modelo de datos relacional) tiene una estructura *ad hoc* diseñada para almacenar la información descriptiva general sobre cada uno de estos documentos *citantes* y *citados*, entre los que destacan las etiquetas del título y la

fecha de la publicación, el idioma, los autores, la revista y la editorial, así como algunos atributos de identificación estandarizados, ya sea para la publicación, la revista o los autores (como el DOI, ISSN, ISBN y *ORCID*, entre otros).

Entre los principales hallazgos que podemos mencionar sobre la BDREF se encuentran la identificación de 44 235 registros de publicaciones, de las cuales 28 984 son *citantes* (ver cuadro 1) y 15 251, *citadas*.

Considerando un análisis sobre el tipo de documento del conjunto de publicaciones *citantes*, se encontró que 82 % corresponde a artículos publicados en *journals*; el resto se dividió entre capítulos de libro (6.5 %), libros (2.4 %), tesis en diferentes grados (0.3 % de doctorado, 1.0 % de maestría y 1.7 % de pregrado) y otros formatos (cerca de 4 %).

Respecto al idioma predominante encontrado en la escritura de las publicaciones *citantes*, encontramos que el español tiene 53.1 % de los registros, seguido del inglés con 44.1 %; el resto lo cubren otros idiomas, como francés, portugués y alemán, con una participación menor a 1 % cada uno. Si se verifica que las publicaciones *citantes* cuentan con un identificador universal (DOI), resulta que 51 % sí lo tiene; el restante no fue identificado.

En cuanto a las publicaciones *citadas* (INEGI), se cuantificó un subconjunto de estas, considerando aquellas en las que se identificó con precisión su fecha de publicación, resultando 13 595, de las cuales el total cuenta con un *Universal Product Code (UPC)* asignado por el mismo Instituto. Acerca del DOI, ningún registro de publicaciones del INEGI lo tiene. Sin embargo, de este subconjunto de registros, se pudo precisar que 45.73 % está identificado con un código ISBN, mientras que 11.64 % cuenta con ISSN.

Cuadro 1

### Publicaciones *citantes* recuperadas en cada fuente de datos seleccionada

Fuente de datos	CORE	Dimensions	Google Académico	Mendeley	REDALYC	SCOPUS
Cantidad	610	12 800	3 647	470	9 721	1 736

Referente a los perfiles académicos de los autores, se logró la identificación en el sistema SCOPUS de 15 054 de los 49 325 registrados en la BDREF, empleando para ello el SCOPUS-ID, que fue un atributo identificador que nos permitió hacer el *match* entre los investigadores que teníamos registrados en la base de datos. De los identificados se extrajeron los datos de su institución de adscripción e información relevante de su perfil académico. Después, teniendo certeza en los datos del nombre del autor y su adscripción, se recuperaron 4 764 perfiles utilizando el portal *ResearchGate*, que incluyen información complementaria de ellos, junto con el Índice de Impacto *RG Score*, que asigna la misma plataforma. Adicionalmente, se identificaron 1 184 perfiles en el portal de *Publons* y 223 en *Twitter*.

## Modelado del Índice de Impacto INEGI

En esta subsección se evalúa la pertinencia de cada uno de los 12 predictores como posibles miembros del modelo de predicción optimizado (IIPCI), basado en algoritmos de aprendizaje automático y la técnica de regresión gradual con adición de variables. Se consideran como insumos de entrada los valores obtenidos para cada uno de los predictores extraídos de la BDREF, mismos que se muestran en el Anexo B.

### Análisis de los predictores

De acuerdo con el proceso propuesto arriba en la subsección *Reconstrucción del MBI*, se obtienen los predictores que están correlacionados entre los 12 disponibles mediante la matriz de correlación de Pearson. Además, se validan visualmente con la gráfica 1a aquellos que presentaron correlación mayor a 0.9 y se aprecia cómo siguen una tendencia similar los siguientes predictores: IAPC (PubCitantes\_Altmetric), publicaciones *citantes* con DOI (PubCitantes\_DOI), publicaciones *citantes* de acceso abierto (PubCitantes\_OpenAccess), IPA (AutoresCitantes\_Score) y el IPRCAPC (AutoresCitantes\_Colaboraciones).

En contraste, los que no presentaron una marcada correlación fueron: publicaciones INEGI *cita-*

*das* (INEGI\_Citados), publicaciones *citantes* al INEGI (Pub\_Citantes), citas de las publicaciones *citantes* (Citas\_Pub\_citantes), penalización por citas imprecisas (Cita\_imprecisa) y búsquedas del término INEGI en *Google* (Tendencias\_Google). Su comportamiento a través del tiempo se muestra en la gráfica 1b, donde se observa un incremento abrupto para el predictor Tendencias\_Google en el 2004 debido a que fue cuando *Google* inició el registro de tendencias de búsqueda.

Asimismo, en la gráfica 1c se aprecia el comportamiento de los predictores compuestos, que son la proporción de publicaciones *citantes* entre publicaciones INEGI por año (INEGI\_Citados) y la proporción por año entre el número de citas de las publicaciones *citantes* entre la cantidad de publicaciones *citantes* (Pub\_Citantes); se percibe que no existe mayor correlación entre ellos.

### Modelo matemático optimizado (IIPCI)

Teniendo el conjunto de 12 predictores candidatos (A) clasificados en dos conjuntos, los altamente correlacionados (C) y los no correlacionados (B), se aplicaron los pasos descritos en la metodología propuesta (ver el diagrama y subsección *Reconstrucción del MBI de Metodología*) para llegar a un modelo con ocho predictores significativos. El modelo matemático optimizado obtenido para IIPCI se muestra en la fórmula 5:

$$IIPCI_i = X_1 + 2.9X_2 + 1.4X_3 - 0.9X_4 + 0.9X_5 + 2.9X_8 + 0.7X_{11} + X_{12} \quad (5)$$

donde:

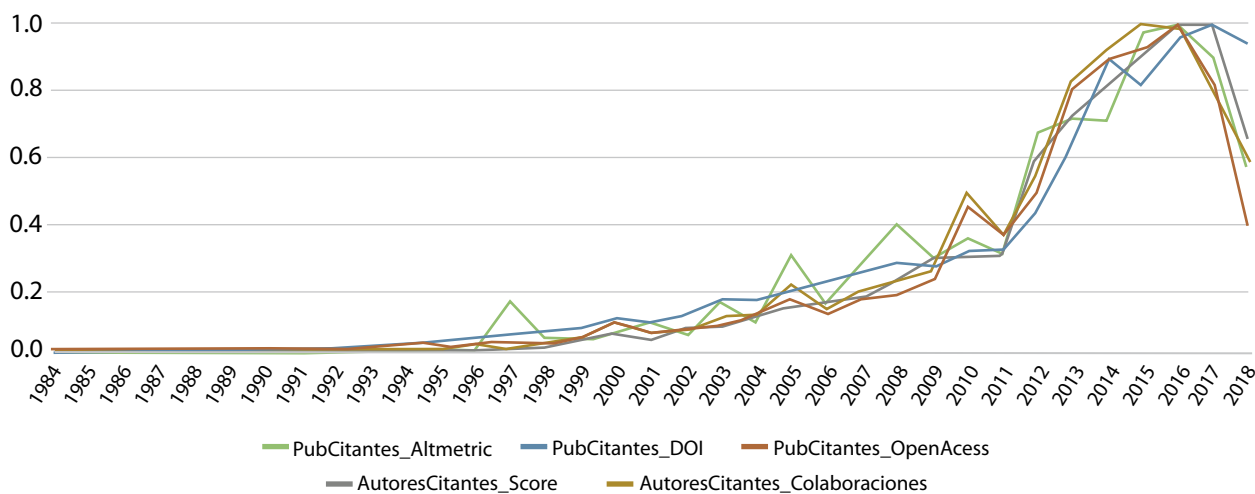
$i$  = año de muestreo en el periodo 1984-2018.  
 IIPCI = Índice de Impacto en la Producción Científica del INEGI.

Nótese cómo el cuarto término del modelo se resta de los demás, lo cual se debe a que representa los errores cometidos por el que hace las citas.

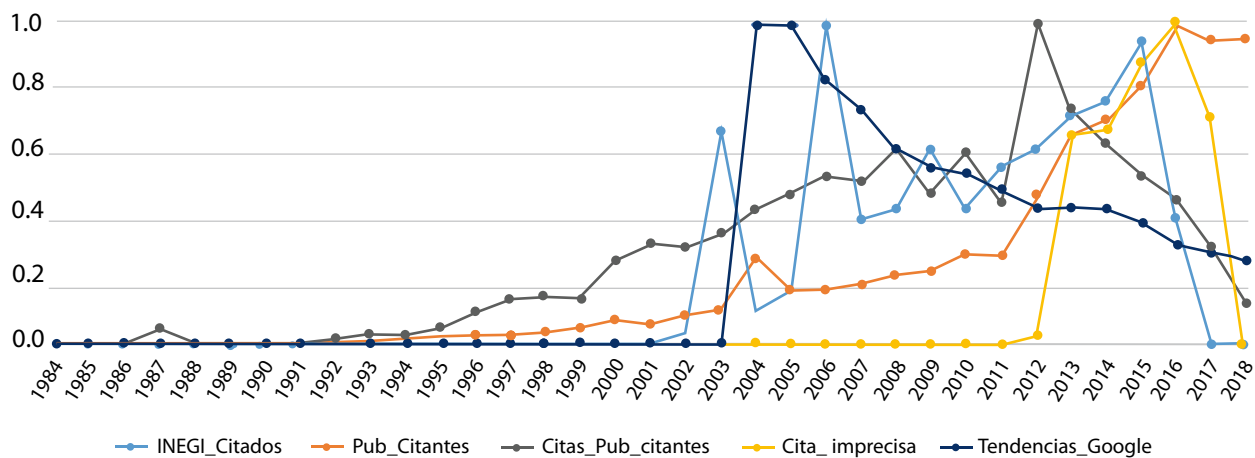
Reportamos, también, los resultados estadísticos del análisis de regresión obtenidos de cada

### Tendencia de los valores normalizados de los predictores

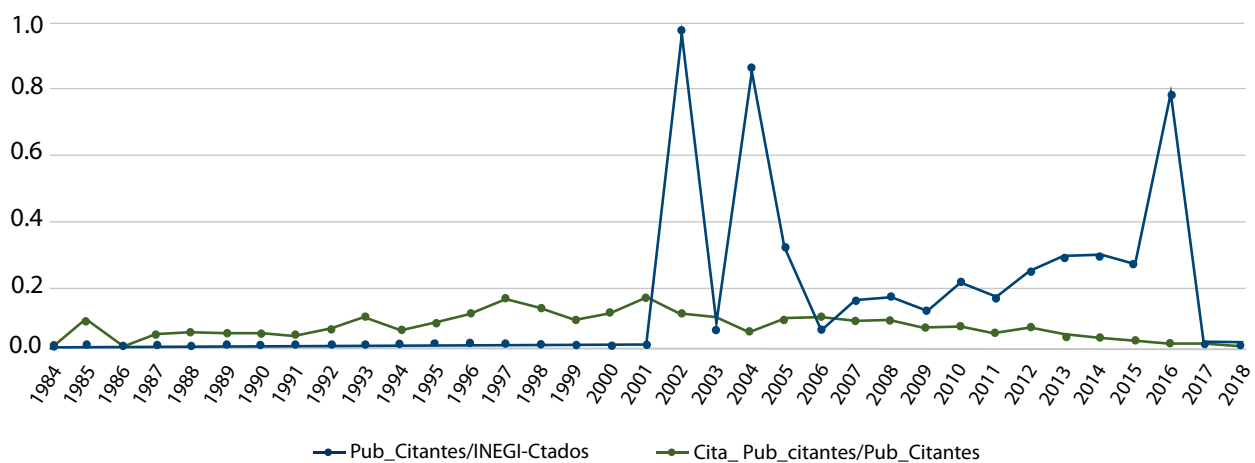
#### a) Comportamiento temporal de los cinco predictores correlacionados



#### b) Comportamiento temporal de predictores no correlacionados



#### c) Comportamiento temporal de los predictores compuestos



uno de los ocho predictores miembros del modelo (ver cuadro 2). En todos los casos, el valor- $p$  ( $p$ -value) asociado fue menor a 0.05, lo cual garantiza una confianza mínima de 95 % e indica que el predictor es significativo.

En el cuadro 2 se aprecia que no aparecen los valores para  $X_6$ ,  $X_7$ ,  $X_9$ ,  $X_{10}$  debido a que no cumplie-

ron con el criterio de regresión aplicado. También, se reporta gráficamente la evolución temporal que presenta año con año el índice IIPCI a partir de 1984 (ver gráfica 2).

Como se puede observar en la gráfica 2, se obtienen comportamientos equivalentes tanto con el MBI usando los 12 predictores como con el modelo

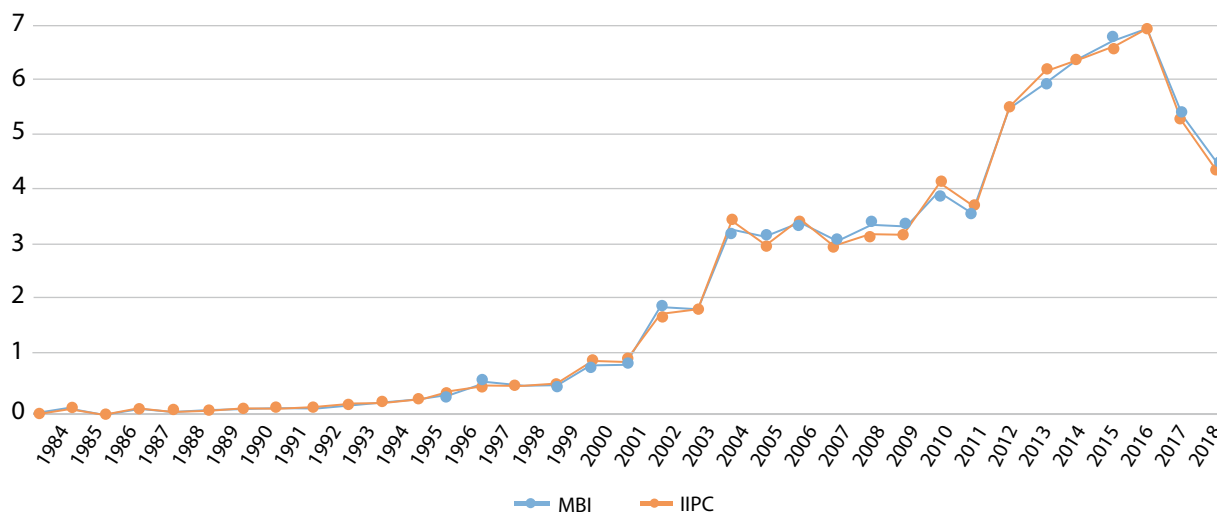
Cuadro 2

**Resultados estadísticos del análisis de regresión para los predictores significativos del modelo IIPCI**

Parámetro	Coefficiente estimado	Error estándar	T-student	p-value
Intercepción	-0.031	0.029	-1.056	0.30011034
INEGI_Citados ( $X_1$ )	0.998	0.105	9.541	2.6871E-10
Pub_Citantes ( $X_2$ )	2.952	0.168	17.522	1.2628E-16
Citas_Pub_citantes ( $X_3$ )	1.459	0.188	7.778	1.7959E-08
Cita_imprecisa ( $X_4$ )	-0.947	0.225	-4.203	0.00024315
Tendencias_Google ( $X_5$ )	0.916	0.086	10.633	2.4406E-11
PubCitantes_OpenAccess ( $X_8$ )	2.965	0.340	8.717	1.8205E-09
Pub_Citantes/INEGI_Citados ( $X_{11}$ )	0.781	0.091	8.605	2.3782E-09
Citas_Pub_Citantes/Pub_Citantes ( $X_{12}$ )	1.032	0.115	8.953	1.0429E-09

Gráfica 2

**Comparación del IIPCI considerando el MBI (con 12 predictores) vs. IIPCI optimizado (con ocho predictores)**



IIPC que solo emplea ocho de estos; prácticamente, las dos curvas siguen el mismo patrón de conducta, por lo tanto, resulta válido emplear el modelo optimizado IIPC para representar a través del tiempo el comportamiento del impacto científico del INEGI. Asimismo, debemos mencionar que no estamos comparando contra otros modelos de reducción de dimensionalidad —como PCA—, sino contra una optimización del mismo modelo base MBI.

Otra forma de apreciar el impacto del INEGI a través del tiempo es midiendo la zona bajo la curva del IIPCI, calculando el porcentaje del área que ocupa en cada uno de los sexenios recientes con respecto a la total. En la gráfica 3 se observa un evidente incremento de este por cada sexenio que transcurre; en particular, en el último (2012-2018) es donde se ha tenido el más alto con un poco más de 40 % del total estimado. Además, se tuvo la mayor pendiente positiva o tasa de crecimiento por año que corresponde a 26 % en el

periodo 2011-2012 y un pico máximo histórico de 6.9 sobre 7.0 unidades en el 2016.

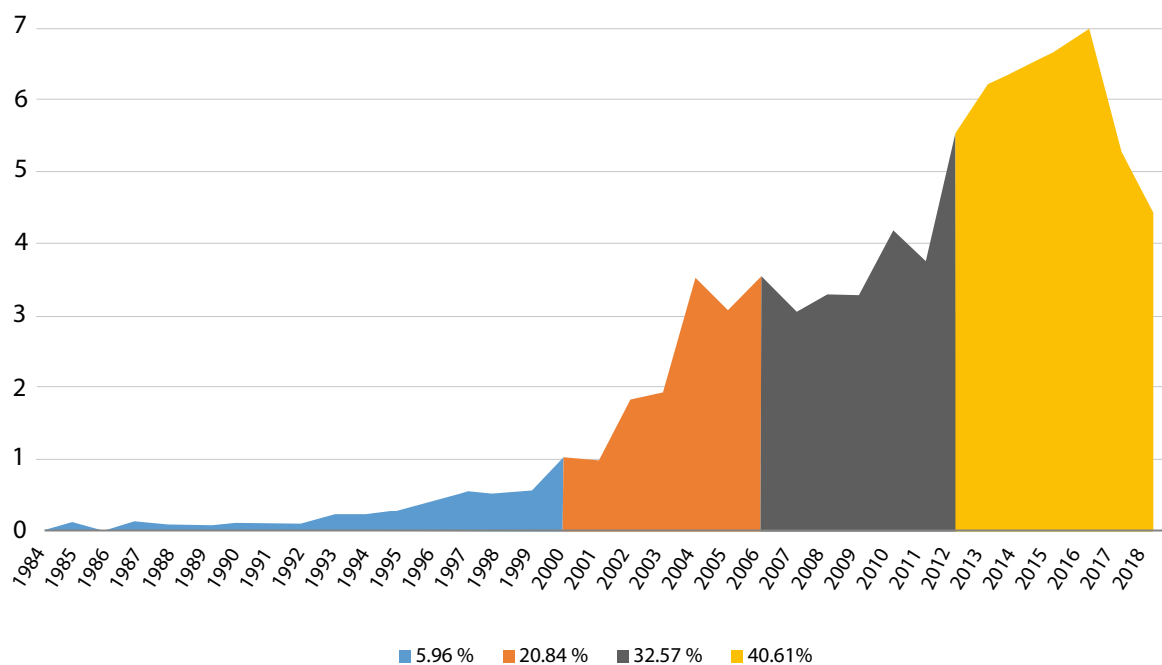
## Conclusiones

En este trabajo se ha presentado una metodología de medición del impacto científico que ha tenido el INEGI en los ámbitos académicos y de divulgación científica, que consiste en un modelo de estimación basado en un índice compuesto de diversos indicadores y métricas relacionadas con el uso de información generada por el Instituto, como citas o referencias en publicaciones científicas, al cual denominamos Índice de Impacto en la Producción Científica del INEGI y no es comparable con otro tipo de métrica individual conocida o empleado en sistemas especializados de información científica.

Para llegar al diseño de este, se realizaron varias tareas, entre las que destacan la formación

Gráfica 3

### IIPCI representado a lo largo del tiempo agrupado por los últimos periodos sexenales



de una base de datos de referencias de publicaciones *citantes* al INEGI (BDREF) mediante la identificación de repositorios digitales, minado de sus atributos referenciales y métricas asociadas a las publicaciones *citantes*, filtrado y depurado de registros duplicados e inconsistentes y la estandarización de datos.

El análisis de la información almacenada en la BDREF incluyó diversos patrones descriptivos evaluados a través del tiempo, como: cantidad de citas a publicaciones del INEGI, cantidad de publicaciones que citan al Instituto, cantidad de citas que obtuvieron las publicaciones *citantes* al INEGI, impacto ponderado de métricas alométricas asociadas a publicaciones *citantes* a este organismo, entre las que se incluyen descargas, me gusta (*likes*), lecturas del resumen, menciones en *Twitter*, etc., así como penalizaciones aplicadas por falta de referencia formal, citas incorrectas o ambiguas al INEGI en las publicaciones citantes, cantidad de publicaciones *citantes* a esta institución que son de acceso abierto, cantidad de publicaciones *citantes* al INEGI que tienen un identificador universal DOI, impacto asociado a perfiles de autores con presencia en redes sociales académicas, Índice de Tendencias de Búsqueda en *Google Trends*, impacto de redes de colaboración entre coautores de publicaciones *citantes* al Instituto. Con esta serie de predictores se generó un modelo donde se analizó su respectiva contribución mediante la técnica de aprendizaje artificial denominada regresión gradual con adición de variables significativas.

Al tratarse el Índice de Impacto en la Producción Científica del INEGI de un modelo predictivo de estimación del impacto para el Instituto que fue entrenado y optimizado con 12 predictores extraídos mediante un proceso exhaustivo de minado de datos, se prevé que en los años siguientes se pueda tener dificultades para tener la disponibilidad de todos estos indicadores que se requieren para la alimentación del modelo, por lo que nos dimos a la tarea de buscar un subconjunto de datos de fácil disponibilidad y que pudiera emular de forma apropiada la entrada de información al modelo de estimación IIPCI con lo cual obtuvimos uno

optimizado que utiliza menos predictores y que pudiera ser usado en un escenario más acotado, como el Laboratorio de Microdatos del INEGI.

Como contribución, este modelo puede ser replicable y aplicable para medir el impacto de diversas instituciones generadoras de contenido científico siguiendo la metodología descrita y requiriendo, para ello, recopilar los datos respectivos de cada uno de los indicadores considerados en el modelo. Con los resultados que se obtengan se podrían hacer comparaciones equiparables entre varios organismos.

A diferencia del modelo de regresión lineal que utilizan Abramo *et al.* (2019), el cual solo considera dos variables (número de citas tempranas y FI de la revista) para medir y predecir el impacto científico de una publicación, el nuestro resultó más completo e integral dado que emplea ocho variables para su estimación (ver *Anexo A*) considerando tanto predictores clásicos como alométricos.

Los resultados derivados de este análisis asisten al INEGI en determinar qué información de sus programas estadísticos (censos, encuestas, etc.) es más utilizada y cuál es menos empleada en el campo científico. Esto puede contribuirle en la toma de decisiones estratégicas u operativas en el país al ubicar qué censos y encuestas debe fortalecer, así como orientar su periodicidad, entre otros aspectos. Aunque nuestra propuesta está pensada para el INEGI, puede ser aplicada a otras instituciones, incluyendo autores individuales que sean generadores de contenido con una problemática similar.

Finalmente, como trabajo futuro, se pueden generar modelos predictivos del impacto por áreas del conocimiento que maneja el INEGI, como económica y sociodemográfica; gobierno; seguridad pública e impartición de justicia; geográfica; medio ambiente; así como ordenamiento territorial y urbano. Para ello, se puede aplicar un análisis semántico latente de los resúmenes de los artículos *citantes* del Instituto y, con ello, presentar resultados del impacto, detallados por área de conocimiento.

## Fuentes

- Abramo, G., C. A D'Angelo & G. Felici. "Predicting publication long-term impact through a combination of early citations and journal impact factor", en: *Journal of Informetrics*. 13(1). 2019, pp. 32-49.
- Abrihami, A. & S. Aliakbary. "Predicting citation counts based on deep neural network learning techniques", en: *Journal of Informetrics*. 13(2). 2019, pp. 485-499.
- Altmetric. *How is the Altmetric Attention Score calculated?* 2020 (DE) último acceso el 20 de junio de 2020 en <https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-score-calculated->
- Alvarado-Villa, D. A. & M. C. Santa Ana. *Acceso y recuperación de metadatos de publicaciones digitales científicas utilizando los servicios de Dimensions*. Tesis para obtener el grado de Ingeniero de *Software*. Facultad de Telemática, Universidad de Colima, 2019.
- Alvarez, J. J., F. Almenárez Mendoza & M. Labrador. "An accurate way to cross reference users across Social Networks", en: *Southeast Con 2017*. Charlotte, NC, 2017, pp. 1-6 (DE) <https://doi.org/10.1109/SECON.2017.7925366>
- Bardus, M., R. El Rassi, M. Chahrouh, E. W. Akl, A. S. Raslan, L. I. Meho y E. A. Akl. "The Use of Social Media to Increase the Impact of Health Research: Systematic Review", en: *Journal of Medical Internet Research*. 22(7). e15607, 2020.
- Bornmann, L. "Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics", en: *Journal of Informetrics*. 8(4). 2014, pp. 895-903 (DE) <https://doi.org/10.1016/j.joi.2014.09.005>
- Castrejón-Mejía, O. E. *Aplicación de minería de datos basado en servicios de información de SCOPUS*. Tesis para obtener el grado de Ingeniero de *Software*. Facultad de Telemática, Universidad de Colima, 2020.
- Delgado-López-Cózar, E. & A. Martín-Martín. "El factor de impacto de las revistas científicas sigue siendo ese número que devora la ciencia española: ¿hasta cuándo?", en: *Anuario ThinkEPI*, 13. 2019 (DE) <https://doi.org/10.3145/thinkepi.2019.e13e09>
- DORA. *Declaración de San Francisco sobre la evaluación de la investigación*. Traducción por Beatriz Parda-Peláez. 2018 (DE) <https://sfdora.org/read/es>
- Eysenbach, G. "Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact", en: *Journal of Medical Internet Research*. 13(4). 2011, pp. 123 (DE) <https://doi.org/10.2196/jmir.2012>
- Garfield, E. "Citation indexes for science; a new dimension in documentation through association of ideas", en: *Science*. 122(3159). 1955, pp. 108-111.
- Guyon, I. & A. Elisseeff. "An introduction to variable and feature selection", en: *Journal of Machine Learning Research*. 3(Mar), 2003, pp. 1157-1182.
- Herrera-Morales, J. R. *Primer informe técnico del proyecto 290379 del Fondo Sectorial CONACyT-INEGI: medición del uso de los servicios del Laboratorio de Microdatos y procesamiento remoto en investigaciones académicas y en la fundamentación de la definición, operación y evaluación de política pública*. México, 2019.
- Hirsch, J. E. "An index to quantify an individual's scientific research output", en: *Proceedings of the National Academy of Sciences*. 102(46). 2005, pp. 16569-16572 (DE) <https://doi.org/10.1073/pnas.0507655102>
- James, J. "Data Never Sleeps 7", en: *Domosphere*. 2019 (DE) último acceso el 20 de junio de 2020 en <https://www.domo.com/learn/datanever-sleeps-7>
- Madrigal-Martínez, K. F., J. R. Herrera-Morales, A. R. Gallardo, S. Sandoval-Carrillo y P. C. Santana-Mancilla. "Aplicación de minería de datos para extracción de documentos académicos del repositorio de REDALYC", en: *Coloquio de investigación multidisciplinaria*. ISSN 2007-8102, vol. 6(1). 2018, pp. 1822-1829.
- Martínez-Barajas, M. A. *Recolección e integración de indicadores alométricos asociados a publicaciones científicas que citan al INEGI*. Tesis para obtener el grado de Ingeniero de *Software*. Facultad de Telemática, Universidad de Colima, 2021.
- Mosqueda-González, B. A. *Extracción de perfiles académicos y su impacto en redes sociales*. Tesis para obtener el grado de Ingeniero de *Software*. Facultad de Telemática, Universidad de Colima, 2021.
- Mullins, Ch., C. J. Boyd & B. L. Corey. "Examining the Correlation Between Altmetric Score and Citations in the General Surgery Literature", en: *Journal of Surgical Research*. 248. 2020, pp. 159-164 (DE) <https://doi.org/10.1016/j.jss.2019.11.008>
- Nocera, A. P., C. J. Boyd, H. Boudreau, O. Hakim & S. Rais-Bahrami. "Examining the Correlation Between Altmetric Score and Citations in the Urology Literature", en: *Urology*. 134. 2019, pp. 45-50 (DE) <https://doi.org/10.1016/j.urology.2019.09.014>
- Ortega, J. L. "Altmetrics data providers: A meta-analysis review of the coverage of metrics and publication", en: *El Profesional de la Información*. 29(1). 2020 (DE) <https://doi.org/10.3145/epi.2020.ene.07>
- Ruder, S. "An overview of gradient descent optimization algorithms", en: arXiv preprint arXiv:1609.04747v2. 2017.
- Seglen, Per O. "Why the impact factor of journals should not be used for evaluating research", en: *British Medical Journal*. V. 314. 1997, pp. 498-502 (DE) <https://doi.org/10.1136/bmj.314.7079.497>
- Torres-Salinas, D., Á. Cabezas-Clavijo & E. Jiménez-Contreras. "Altmetrics: New Indicators for Scientific Communication in Web 2.0", en: *Comunicar*. 21(41). 2013, pp. 53-60 (DE) <https://doi.org/10.3916/c41-2013-05>
- Velasco, B., J. M. E. Bouza, J. M. Pinilla & J. A. San Román. "La utilización de los indicadores bibliométricos para evaluar la actividad investigadora", en: *Aula Abierta*. 40(2). 2012, pp. 75-84.

## Anexo A

### Predictores empleados en el IIPCI

$X_1$ . INEGI\_Citados. Se refiere al número de publicaciones que el mismo INEGI produce, derivado de los proyectos estadísticos que realiza periódicamente, como los censos y las encuestas.

$X_2$ . Pub\_Citantes. Representa la cantidad de publicaciones académicas por año que tomaron la información del INEGI como base para fundamentar sus trabajos.

$X_3$ . Citas\_Pub\_citantes. Se refiere al número de citas que tienen las Pub\_Citantes.

$X_4$ . Cita\_imprecisa. Son aquellas citas incluidas en Pub\_Citantes que desafortunadamente no siguen las reglas convencionales para citar como referencia a una obra del INEGI. Se identificaron variantes incorrectas en la forma de citar (ver cuadro A1).

Cuadro A1

#### Penalización por tipos de imprecisión de la cita al INEGI de acuerdo con el grado de severidad de la falta

Tipo	Peso
Cita en el resumen	1
Cita en tabla o figura	2
Cita en el texto	3
Cita ambigua	4
No hay cita	5
No se identificó o la cita es correcta	0

Así, por cada publicación *citante* que incurra en alguna de las situaciones mostradas en el cuadro A1, se disminuye el índice de impacto, penalizando según el peso correspondiente.

$X_5$ . Tendencias\_Google. Este patrón se refiere al número de veces por año que la palabra INEGI ha sido consultada dentro de Google. Con el servicio de *Google Trends*, se puede explorar la relevancia o tendencia de cualquier expresión de búsqueda,

obteniendo el número de veces que se ha empleado en un periodo y es manejado como porcentajes, considerando 100 % el día en que mayores incidencias se hayan realizado en el lapso seleccionado.

$X_6$ . PubCitantes\_Altmetric. Se refiere a un valor resultante que considera todos los indicadores altmétricos de cada una de las Pub\_Citantes, de acuerdo con los pesos expresados en el cuadro A2:

$$X_6 = \sum_{k=1}^m \sum_{j=1}^n (W_{(j)} Alt_{(j)}) \quad (1A)$$

donde:

$W_{(j)}$  = peso ponderado del altmétrico  $j$ .

$Alt_{(j)}$  = frecuencia del altmétrico  $j$ .

$m$  = número total de publicaciones *citantes* en cada año.

$k$  = índice de iteración de la publicación *citante* en la secuencia.

$n$  = número total de indicadores altmétricos de la  $k$ -ésima publicación *citante*.

$j$  = índice de iteración del indicador altmétrico en la secuencia.

Los pesos asignados a cada altmétrico se establecieron considerando las recomendaciones que da *Almetric Company* (Altmetric, 2020) y se detallan en el cuadro A2.

Cuadro A2

#### Pesos empleados para los indicadores altmétricos de las publicaciones

Tipo de indicador altmétrico	Peso
Lecturas en <i>Mendeley</i>	0.1
Lecturas del resumen	0.001
Mención en <i>Wikipedia</i>	3
Descargas	0.2
<i>Likes</i>	0.05
Menciones en blogs	5.5
Menciones en periódicos digitales	8-6
Vistas del texto completo	0.01
Tuits	1.0

Fuente: Altmetric, 2020.



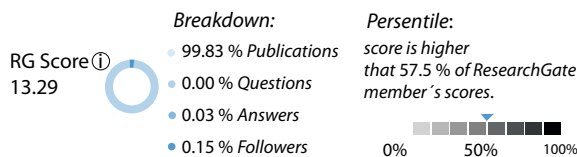
$X_7$ . PubCitantes\_DOI. Número de Pub\_Citantes que tienen asignado un identificador único global (DOI).

$X_8$ . PubCitantes\_OpenAccess. Número de Pub\_Citantes que son de acceso abierto (*Open Access*).

$X_9$ . AutoresCitantes\_Score. Este patrón es un indicador compuesto que hace referencia a la calidad del autor. Se obtiene sumando dos puntajes: el IH y el asignado por *ResearchGate* (*RG Score*) que toma en cuenta el número de publicaciones, preguntas, respuestas y número de seguidores que tiene el autor en su plataforma (ver gráfica A1).

Gráfica A1

**Presentación del puntaje establecido por *ResearchGate* para calificar la productividad de un autor**



$X_{10}$ . AutoresCitantes\_Colaboraciones. Se refiere a un patrón compuesto que sirve para representar el grado de colaboración académica entre coautores. Al considerar las redes de colaboración entre investigadores, se puede cuantificar el impacto del

INEGI, asignando diferentes pesos de acuerdo con el tipo de colaboración que tengan los autores de una publicación. Por ejemplo, si una Pub\_Citantes está hecha por un solo autor nacional, tendrá menor impacto que otra donde se tengan colaboraciones internacionales. Para obtener el grado de contribución que este patrón ejerce sobre el impacto del INEGI, se toman en cuenta los criterios para clasificación de las redes de colaboración de acuerdo con la adscripción de los coautores descritos en el cuadro A3.

Predictores compuestos. Combinar predictores para generar nuevos es válido, siempre y cuando tenga significado esta agrupación. Nos basamos en el concepto de factor de impacto, por ser una medida consolidada y ampliamente utilizada en el enfoque bibliométrico tradicional. Procedemos, entonces, a relacionar el número de citas que tuvieron las publicaciones, entre el número de ellas.

$X_{11}$ . FI\_INEGI\_Citados. Se refiere a la proporción directa de la cantidad de Pub\_Citantes entre el número de documentos INEGI\_Citados en el mismo año.

$X_{12}$ . FI\_Pub\_Citantes. Es la proporción de citas de los *citantes*, que se calcula con la cantidad de citas que obtuvieron las Citas\_Pub\_citantes entre el total de estas (Pub\_Citantes).

Cuadro A3

Continúa

**Clasificación de redes de colaboración de acuerdo con la cantidad de coautores que son de adscripción en instituciones mexicanas o extranjeras**

Tipo	Descripción	Autores mexicanos	Autores extranjeros	Peso
RN1-Red Nacional A	Es cuando se tiene una publicación con un solo autor y es mexicano.	1	0	1
RN2-Red Nacional B	Hay dos o tres coautores, donde todos tienen adscripción en instituciones mexicanas.	{2..3}	0	2

### Clasificación de redes de colaboración de acuerdo con la cantidad de coautores que son de adscripción en instituciones mexicanas o extranjeras

Tipo	Descripción	Autores mexicanos	Autores extranjeros	Peso
RN3-Red Nacional C	Es de más de tres coautores, donde todos tienen adscripción en instituciones mexicanas.	> 3	0	3
RF1-Red Foránea A	Publicación con un solo autor de adscripción extranjera.	0	1	2
RF2-Red Foránea B	Publicación de dos o tres autores, donde todos tienen adscripción en instituciones extranjeras.	0	> 1	3
RF3-Red Foránea C	Publicación de más de tres coautores, donde todos tienen adscripción en instituciones extranjeras.	0	> 3	5
RI1-Red Internacional A	Tiene dos autores, uno nacional y otro con adscripción en el extranjero.	> 0	1	5
RI2-Red Internacional B	Se tienen tres coautores, donde al menos uno de ellos es nacional y los restantes con adscripción foránea diferente.	> 0	{1..2}	6
RI3-Red Internacional C	Se tienen cuatro autores, donde al menos uno es nacional y los otros deben ser al menos de tres diferentes adscripciones internacionales.	> 0	> 3	10

Valores de los predictores para el IIPCI obtenidos para el periodo 1984-2018g

Periodo*	Valores normalizados de los predictores INEGI_Citados ( $X_1$ ), Pub_Citantes ( $X_2$ ), Citas_Pub_citantes ( $X_3$ ), Cita_imprecisa ( $X_4$ ), Tendencias_Google ( $X_5$ ), PubCitantes_Altmetric ( $X_6$ ), PubCitantes_DOI ( $X_7$ ), PubCitantes_OpenAccess ( $X_8$ ), AutoresCitantes_Score ( $X_9$ ), AutoresCitantes_Colaboraciones ( $X_{10}$ )									
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$
2018	0.00E+00	9.55E-01	1.25E-01	4.97E-03	2.62E-01	5.64E-01	9.38E-01	3.86E-01	6.51E-01	5.83E-01
2017	0.00E+00	9.53E-01	3.06E-01	7.09E-01	2.87E-01	8.98E-01	1.00E+00	8.19E-01	1.00E+00	7.73E-01
2016	3.88E-01	1.00E+00	4.50E-01	1.00E+00	3.09E-01	1.00E+00	9.54E-01	1.00E+00	9.94E-01	9.80E-01
2015	9.54E-01	8.12E-01	5.33E-01	8.83E-01	3.81E-01	9.75E-01	8.19E-01	9.29E-01	8.96E-01	1.00E+00
2014	7.62E-01	7.06E-01	6.30E-01	6.73E-01	4.21E-01	7.09E-01	8.91E-01	8.91E-01	8.02E-01	9.27E-01
2013	7.18E-01	6.60E-01	7.40E-01	6.51E-01	4.24E-01	7.16E-01	6.31E-01	8.03E-01	7.24E-01	8.29E-01
2012	6.09E-01	4.68E-01	1.00E+00	2.48E-02	4.27E-01	6.70E-01	4.23E-01	4.93E-01	5.82E-01	5.44E-01
2011	5.53E-01	2.78E-01	4.46E-01	0.00E+00	4.84E-01	3.00E-01	3.19E-01	3.63E-01	2.95E-01	3.58E-01
2010	4.22E-01	2.83E-01	5.99E-01	0.00E+00	5.34E-01	3.50E-01	3.05E-01	4.44E-01	2.97E-01	4.86E-01
2009	6.17E-01	2.27E-01	4.70E-01	0.00E+00	5.52E-01	2.87E-01	2.68E-01	2.18E-01	2.91E-01	2.49E-01
2008	4.22E-01	2.17E-01	6.10E-01	0.00E+00	6.13E-01	3.92E-01	2.76E-01	1.76E-01	2.17E-01	2.21E-01
2007	3.88E-01	1.85E-01	5.08E-01	0.00E+00	7.35E-01	2.77E-01	2.45E-01	1.64E-01	1.73E-01	1.85E-01
2006	1.00E+00	1.66E-01	5.26E-01	0.00E+00	8.28E-01	1.51E-01	2.18E-01	1.18E-01	1.57E-01	1.35E-01
2005	1.60E-01	1.64E-01	4.68E-01	0.00E+00	9.97E-01	2.91E-01	1.88E-01	1.63E-01	1.40E-01	2.10E-01
2004	9.47E-02	2.70E-01	4.22E-01	0.00E+00	1.00E+00	9.44E-02	1.60E-01	1.10E-01	1.11E-01	1.17E-01
2003	6.65E-01	1.06E-01	3.47E-01	0.00E+00	0.00E+00	1.51E-01	1.63E-01	9.13E-02	8.34E-02	1.06E-01
2002	0.00E+00	8.58E-02	3.01E-01	0.00E+00	0.00E+00	5.53E-02	1.20E-01	7.16E-02	7.53E-02	6.60E-02
2001	0.00E+00	6.07E-02	3.17E-01	0.00E+00	0.00E+00	9.00E-02	9.72E-02	5.43E-02	3.96E-02	6.36E-02
2000	0.00E+00	7.28E-02	2.62E-01	0.00E+00	0.00E+00	5.99E-02	1.04E-01	9.30E-02	5.81E-02	9.55E-02
1999	0.00E+00	4.88E-02	1.41E-01	0.00E+00	0.00E+00	4.17E-02	7.50E-02	4.45E-02	3.45E-02	4.04E-02
1998	0.00E+00	3.50E-02	1.46E-01	0.00E+00	0.00E+00	4.67E-02	6.99E-02	2.60E-02	1.65E-02	3.25E-02
1997	0.00E+00	2.66E-02	1.40E-01	0.00E+00	0.00E+00	1.57E-01	5.91E-02	3.12E-02	1.15E-02	1.39E-02
1996	0.00E+00	2.74E-02	9.96E-02	0.00E+00	0.00E+00	1.06E-02	5.17E-02	3.12E-02	8.15E-03	2.59E-02
1995	0.00E+00	1.90E-02	4.98E-02	0.00E+00	0.00E+00	1.06E-02	4.03E-02	2.54E-02	1.85E-02	1.82E-02
1994	0.00E+00	1.46E-02	2.91E-02	0.00E+00	0.00E+00	3.38E-03	2.84E-02	2.48E-02	7.17E-03	1.23E-02

### Valores de los predictores para el IIPCI obtenidos para el periodo 1984-2018

Periodo*	Valores normalizados de los predictores INEGI_Citados ( $X_1$ ), Pub_Citantes ( $X_2$ ), Citas_Pub_citantes ( $X_3$ ), Cita_imprecisa ( $X_4$ ), Tendencias_Google ( $X_5$ ), PubCitantes_Altmetric ( $X_6$ ), PubCitantes_DOI ( $X_7$ ), PubCitantes_OpenAccess ( $X_8$ ), AutoresCitantes_Score ( $X_9$ ), AutoresCitantes_Colaboraciones ( $X_{10}$ )									
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$
1993	0.00E+00	8.24E-03	2.92E-02	0.00E+00	0.00E+00	1.01E-02	1.65E-02	1.62E-02	3.99E-03	4.97E-03
1992	0.00E+00	6.78E-03	1.47E-02	0.00E+00	0.00E+00	2.02E-03	1.53E-02	9.82E-03	6.04E-03	2.98E-03
1991	0.00E+00	5.05E-03	6.66E-03	0.00E+00	0.00E+00	2.43E-04	1.25E-02	1.79E-02	5.68E-03	8.95E-03
1990	0.00E+00	4.12E-03	7.01E-03	0.00E+00	0.00E+00	1.51E-03	9.09E-03	1.68E-02	0.00E+00	1.06E-02
1989	0.00E+00	3.32E-03	6.52E-03	0.00E+00	0.00E+00	4.52E-04	7.95E-03	5.20E-03	0.00E+00	0.00E+00
1988	0.00E+00	9.30E-04	3.01E-03	0.00E+00	0.00E+00	1.54E-04	4.55E-03	5.78E-03	2.72E-03	3.31E-04
1987	0.00E+00	3.99E-04	0.00E+00	0.00E+00	0.00E+00	2.50E-03	3.41E-03	2.89E-03	0.00E+00	1.99E-03
1986	0.00E+00	2.66E-04	0.00E+00	0.00E+00	0.00E+00	4.01E-05	1.14E-03	4.04E-03	2.40E-03	1.33E-03
1985	0.00E+00	0.00E+00	2.87E-03	0.00E+00	0.00E+00	3.76E-05	0.00E+00	3.47E-03	0.00E+00	4.31E-03
1984	0.00E+00	7.97E-04	1.40E-04	0.00E+00	0.00E+00	1.14E-05	1.14E-03	4.04E-03	0.00E+00	4.31E-03

\* No se incluyeron los años 2020 y 2019 debido a que no se cuenta con la información completa de todos los predictores, ya que el proceso de minado para integrar la BDREF culminó en mayo del 2019.