

# Análisis de datos para estimar la producción del cultivo de maíz en grano por temporal en México

Yuridia Cazares Carrillo  
Facultad de Ciencias  
Universidad Autónoma de Baja California (UABC)  
Ensenada, Baja California  
Email: yuridia.cazares@cua.uam.mx

## Resumen—

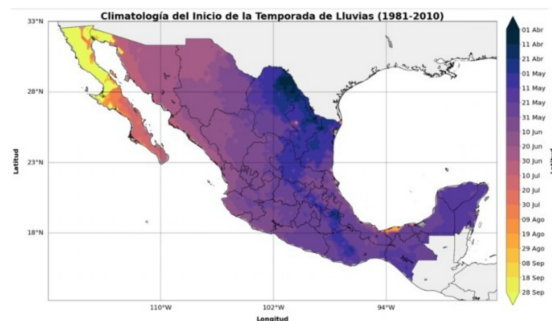
### I. INTRODUCCIÓN

En las últimas décadas, la actividad humana como la quema de combustibles fósiles, deforestación, incendios forestales y muchas otras actividades ha desencadenado un aumento significativo en las temperaturas promedio en todo el planeta. De acuerdo con la investigación realizada en 2020 por un grupo de más de 80 científicos en el informe “**Cambio climático y medioambiental en la cuenca mediterránea**”, la temperaturas podría experimentar un incremento de 2.2 grados frente a los niveles preindustriales en 2040 y los 3.8 en 2100 [1].

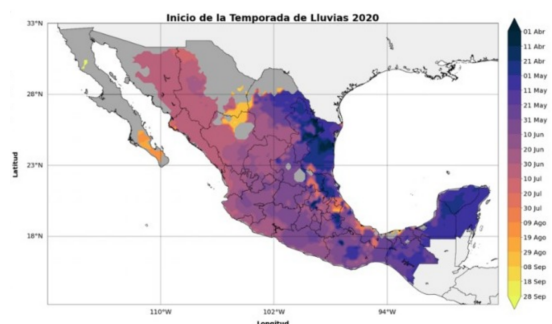
En México, el informe oficial del Reporte Anual del Clima 2020 de la Comisión Nacional del Agua (CONAGUA) reveló que la temperatura promedio está aumentando más rápido que el resto del mundo a raíz del cambio climático, con 1.4 °C por encima de la media global, esto ha ocasionado un retraso en el comienzo de la temporada de lluvias en todo el país en los últimos años, como se demuestra en los resultados de 2020 presentados en la Figura 1 [2].

Este estudio se propone analizar y examinar datos que serán extraídos del anuario estadístico de la producción agrícola y la Comisión Nacional del Agua (CONAGUA) relativos a los últimos 30 años en México. Se utilizarán técnicas de modelado estadístico como “análisis de regresión” para comprender y predecir la relación entre una variable dependiente e independiente que tiene como objetivo predecir patrones, tendencias, frecuencia y eventos futuros. Esto permitirá un análisis más detallado y preciso de cómo las variaciones en la cantidad de lluvia han impactado en el cultivo de maíz que se desarrolla bajo la modalidad de agricultura de temporal. Esta modalidad, también conocida como agricultura de secano, se basa exclusivamente en las precipitaciones estacionales y los ciclos climáticos naturales para proveer la humedad necesaria al cultivo.

Los resultados esperados nos proporcionarán información crucial para evaluar si el retraso en el comienzo de la temporada de lluvias ha tenido un impacto en los procesos de siembra y cosecha del maíz. Este análisis nos proporcionará información valiosa sobre la dirección que están tomando estas variables y nos permitirá contrarrestar estos cambios implementando algún tipo de tecnología agrícola en el futuro como



(a) Inicio de la temporada de lluvias en los años (1981-2010) a nivel nacional.



(b) Inicio de la temporada de lluvias en el año 2020 a nivel nacional.

Figura 1. Climatología del comienzo de la temporada de lluvias en México. Los colores más cálidos (rosa-amarillo) representa el inicio de temporada de lluvia en los meses (julio a septiembre) mientras que las áreas con colores más fríos (violeta-azul) representa el inicio de temporada de lluvia en los meses (abril a junio).

invernaderos, malla sombra, macro túnel u otra tecnología de riego que complemente la modalidad “por temporal”. Asimismo, diseñar posibles estrategias de adaptación y mitigación que podrían ayudar a reducir la vulnerabilidad del cultivo frente a los desafíos climáticos, dado que México es un país con una gran dependencia del maíz, siendo éste un alimento básico en la dieta mexicana.

### II. CARACTERIZACIÓN DE LOS DATOS

En este estudio, se utilizaron datos recopilados del anuario estadístico de la producción agrícola y los resúmenes men-

suales de temperaturas y lluvias de la comisión nacional del agua (CONAGUA) durante un periodo de 38 años (1985-2022) para analizar la producción de maíz en grano sembrada por temporal, usando la modalidad de cielo abierto.

Los detalles de las variables utilizadas se presentan en la Tabla I y II.

Las variable 'Entidad' de tipo discreta representa la entidad federativa. La variable 'Sembrada' de tipo continuo representa la extensión de terreno en hectáreas en la que se cultivó maíz en grano, mientras que 'Cosechada' refleja la cantidad cosechada exitosamente de maíz. Por su parte, 'Siniestra' representa las pérdidas en relación al total sembrado.

Tabla I  
MUESTRA DE PRODUCCIÓN AGRÍCOLA DEL AÑO 1985

Entidad	Superficie (ha)		
	Sembrada	Cosechada	Siniestra
Jalisco	726,554	708,307	18,247
Chiapas	672,749	652,489	20,26
Estado de México	601,099	601,099	0,0
Puebla	573,648	545,251	28,397
Veracruz	503,165	477,152	26,013
Guerrero	478,637	468,074	10,563
Zacatecas	455,939	385,718	70,221
Oaxaca	437,465	398,624	38,841
Michoacán	429,889	408,849	21,04
Guanajuato	410,924	322,099	88,825
...	...	...	...

La variable 'Precipitación' representa el total de lluvia en el área de la entidad con su respectiva media.

Tabla II  
MUESTRA DE PRECIPITACIÓN DEL AÑO 1985

Entidad	(mm)	
	Precipitación	Media
Jalisco	769	64,0833
Chiapas	1849,9	154,1583
Estado de México	777,2	64,7666
Puebla	1080,9	90,0750
Veracruz	1286,9	107,2416
Guerrero	1046,5	87,2083
Zacatecas	417,5	34,7916
Oaxaca	1386,5	115,5416
Michoacán	765,3	63,775
Guanajuato	533,9	44,4916
...	...	...

### III. TRABAJOS RELACIONADOS

En un estudio realizado por Shakoar [6] se aplicó un enfoque ricardiano para probar las relaciones entre los ingresos agrícolas netos (NFR) y el clima en la región árida el método es un enfoque transversal que estudia la producción agrícola. Explica cómo los cambios en las variables climáticas afectan los ingresos agrícolas netos. Cabe mencionar que solo se utilizaron datos climáticos anuales y no datos mensuales o estacionales.

El artículo *Impactos potenciales del cambio climático en la agricultura: un caso de estudio de la producción de café en*

*Veracruz, México* [3], adopta un modelo de regresión múltiple que integra los determinantes climáticos y de la producción de café en Veracruz. Su objetivo fue establecer una función de coproducción que diera la información sobre cómo esta actividad responde a cambios en las variables económicas y climáticas.

Los trabajos relacionados antes revisados se enfocan en dar resultados utilizando una variable económica, el primer estudio trabaja un enfoque ricardiano la cual examina cómo el clima en diferentes lugares afecta la renta neta o el valor de las tierras agrícolas, la ventaja más importante del modelo ricardiano reside en su capacidad para tener en cuenta adaptaciones privadas. En el modelo ricardiano, los agricultores maximizan sus ganancias bajo el cambio climático, cambiando la combinación de cultivos, las fechas de siembra y cosecha y siguiendo una serie de prácticas agronómicas. Sin embargo, en este estudio solo se tomarán en cuenta variables climáticas y de producción del cultivo de maíz, el segundo estudio adopta un modelo de regresión múltiple, tomando variables económicas, climáticas y de producción. En cambio, los siguientes muestran un enfoque de predicción utilizando algoritmos de aprendizaje autónomo y técnicas de regresión.

Rale [5] implementa un método de conjunto basándose en árboles (Gradient boosting), ya que sus observaciones tienen interacción de características no lineales, menciona que esos modelos pueden ser propensos a sobre ajustarse, pero pueden solucionarse mediante un ajuste cuidadoso de hiperparámetros.

En un estudio realizado por Jambekar [4] compara la efectividad de 3 algoritmos, regresión lineal múltiple, Random Forest Regression y Multivariate Adaptive Regression Splines (Earth), sus resultados mostraron que el rendimiento de la regresión lineal múltiple fue mejor para el conjunto de datos de maíz.

Como conclusión de los trabajos relacionados anteriores, abstraemos que es mejor utilizar un enfoque de regresión lineal múltiple, la cual se adapta con los datos a trabajar, utilizando la cosecha como la variable dependiente y objetivo principal de nuestro estudio, sin embargo, es importante considerar la dispersión de los datos a trabajar para verificar si su trayectoria es lineal.

selección de fuentes de información, integración, procesamiento y pruebas de regresión

### IV. METODOLOGÍA

**Selección de fuentes de información:** Los datos utilizados en este estudio provienen de dos fuentes: bases de datos de climatología y agricultura. Las variables clave analizadas son precipitación, la superficie sembrada, la superficie cosechada y la superficie de pérdida de la sembrada denominada siniestra en la base de datos. Las variables de agricultura se recopilaban de manera anual, dado que los datos agrícolas se registran en dos temporadas al año, mientras que los datos climatológicos son mensuales.

**Integración de datos:** Con el objetivo de homogeneizar la información, se tomó la decisión de estandarizar los valores en un formato anual. La integración de ambas bases de datos se llevó a cabo para facilitar la comprensión y uniformidad en el análisis. Se procedió a normalizar los datos utilizando la técnica de máximo-mínimo, con el propósito de optimizar y mejorar los resultados del algoritmo de regresión lineal múltiple empleado en este estudio.

**Procesamiento:** En cuanto al procesamiento de datos para el entrenamiento, se utilizaron las variables de superficie sembrada, la siniestra y la precipitación como variables independientes y como variable dependiente la superficie cosechada. Este conjunto de datos abarca un período, desde 1985 hasta el año 2020, utilizado como conjunto de entrenamiento para el algoritmo.

**Pruebas y evaluación de regresión múltiple:** Para evaluar la eficacia del modelo, se emplearon los años más recientes como conjunto de prueba (2021 y 2022). Las métricas a utilizar para evaluar la regresión es el error cuadrático medio (MSE).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Donde:

$n$  es el número total de observaciones.

$y_i$  son los valores reales u observados de la variable dependiente.

$\hat{y}_i$  son los valores predichos por el modelo.

Otra métrica estadística a utilizar para evaluar el algoritmo es el coeficiente de determinación ( $R^2$ ) la cual nos brinda la proporción de la varianza total de la variable que es explicada por el modelo de regresión, está proporciona información sobre qué tan bien se ajusta un modelo de regresión a los datos observados. El rango de la puntuación de  $R^2$  siempre está entre 0 y 1 y la puntuación más alta es un indicador de una mejor bondad de ajuste para los puntos de datos, es decir es una medida que evalúa la calidad y la precisión de un modelo en relación con los datos reales.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$y_i$ : Son los valores observados de la variable dependiente. Cada  $y_i$  representa un valor real observado en el conjunto de datos.

$\hat{y}_i$ : Son los valores predichos por el modelo. Cada  $\hat{y}_i$  representa la predicción del modelo para el valor correspondiente  $y_i$ .

$\bar{y}$ : Es la media de los valores observados. Indica el valor promedio de todos los  $y_i$  en el conjunto de datos.

$n$ : Es el número total de observaciones en el conjunto de datos. Indica cuantos pares de  $y_i$  y  $\hat{y}_i$  hay en el conjunto de datos.

## V. EXPERIMENTOS Y RESULTADOS

Para llevar a cabo la evaluación del modelo, se realizaron los experimentos focalizados en los estados de Jalisco, Chiapas y

Estado de México. Estos estados fueron seleccionados debido a su destacada posición como principales productores en maíz de grano en modalidad de temporal, de acuerdo a la información recopilada del anuario estadístico de la producción agrícola de México.

Durante el análisis para seleccionar los estados de prueba, se evidenció un patrón general de disminución en la producción en todos los estados; sin embargo, se identificó un comportamiento contrario en Hidalgo y Campeche, donde presentan un incremento, es por ello que se decidió incluir a Campeche para realizar pruebas en el modelo.

Cabe destacar que Chiapas y Campeche ocupan el segundo y tercer lugar de los estados donde más se presentan precipitaciones al año, Jalisco y el Estado de México se sitúan debajo de la media, según el reporte anual de la Comisión Nacional del Agua (CONAGUA).

A continuación se muestran los resultados obtenidos después de realizar el algoritmo de regresión lineal múltiple en cada estado.

Si observamos los resultados en la Tabla III, notamos que el rendimiento del algoritmo fue más exacto para el estado de Chiapas, obteniendo buenos resultados en ambas métricas de evaluación, mientras que el estado de Jalisco obtiene un buen resultado en ( $R^2$ ), pero un MSE mayor, en cambio, para el Estado de México ocurre lo inverso, un mejor resultado de MSE, pero menor bondad de ajuste en los datos ( $R^2$ ), sin embargo, para el estado de Campeche los resultados no fueron muy exitosos en ambas métricas de evaluación.

Tabla III  
COEFICIENTE DE DETERMINACIÓN ( $R^2$ ) Y ERROR CUADRÁTICO (MSE)

Estado	$R^2$	MSE
Jalisco	0.833224	0.003017
Estado de México	0.531284	0.000859
Chiapas	0.966572	0.000425
Campeche	0.785298	0.006574

En la Tabla IV se pueden observar los datos reales y las predicciones estimadas para cada estado, que concuerdan con los resultados del rendimiento que tuvieron.

Tabla IV  
PREDICCIÓN DE COSECHA

Estado	Año	Datos reales	Predicciones
Jalisco	2021	0.260915	0.236675
	2022	0.191687	0.117882
Estado de México	2021	0.471420	0.432685
	2022	0.359071	0.344261
Chiapas	2021	0.245603	0.220617
	2022	0.240914	0.225826
Campeche	2021	0.830112	0.745625
	2022	0.831288	0.753756

En la Figura (2) se muestra gráficamente los datos del estado de Jalisco, donde se observa que los datos reales tienen fluctuaciones ascendentes y descendentes, sin embargo, se observa

una tendencia decreciente constante, la predicción sigue el patrón de los datos, dando como resultado la disminución de cosecha en los datos de prueba.

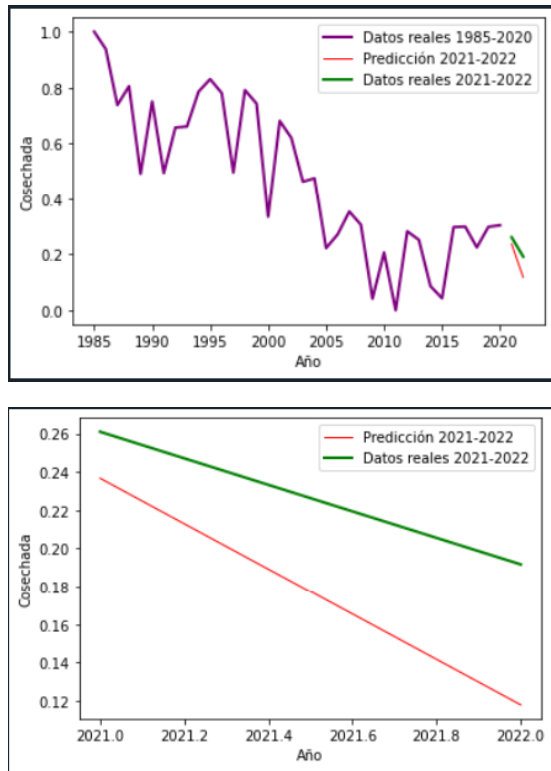


Figura 2. Regresión lineal múltiple: estado de Jalisco

En la Figura (3), se presenta la gráfica correspondiente al Estado de México. Se evidencia un patrón de variaciones ascendentes y descendentes en los datos. A pesar de ello, es destacable que durante los años 1988 y 2011 se registró una disminución de cosecha notable. Aun con estas excepciones, la tendencia general muestra una disminución constante, lo que influye en la predicción.

En la Figura (4) se visualiza la gráfica del estado de Chiapas. Se destaca un rendimiento notable en las cosechas durante el periodo de años de 1995 a 2005. Posterior a este período, se observa una estabilidad constante en su comportamiento, manteniendo una tendencia regular. Esto sugiere que la predicción de cosecha se mantiene favorable.

Para concluir, se presenta la Figura (5) que corresponde al estado de Campeche. En esta representación, se evidencia un amplio intervalo entre los datos de cada año. Aunque la tendencia general de los datos indica un aumento en las cosechas, los resultados de la predicción no se presentan muy favorables.

## VI. CONCLUSIÓN

Al concluir este estudio, se confirma que la creación de modelos de predicción basados en variables climáticas y agrícolas representa un desafío significativo. Dado que estas

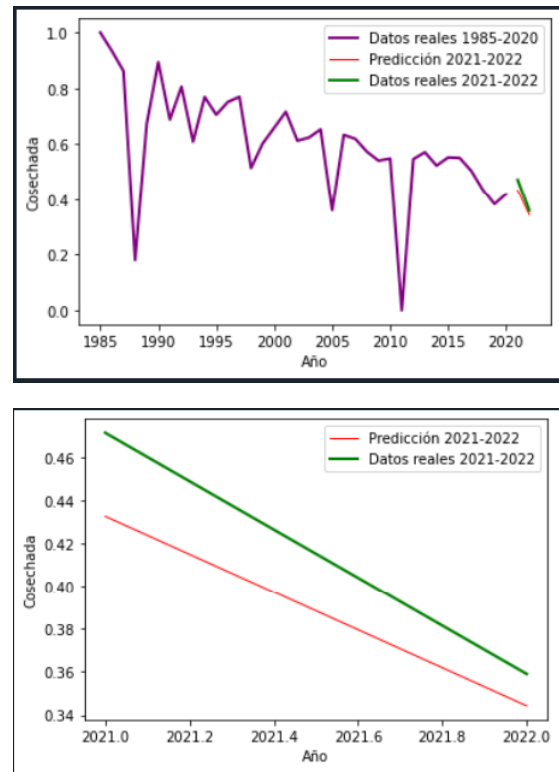


Figura 3. Regresión lineal múltiple: Estado de México

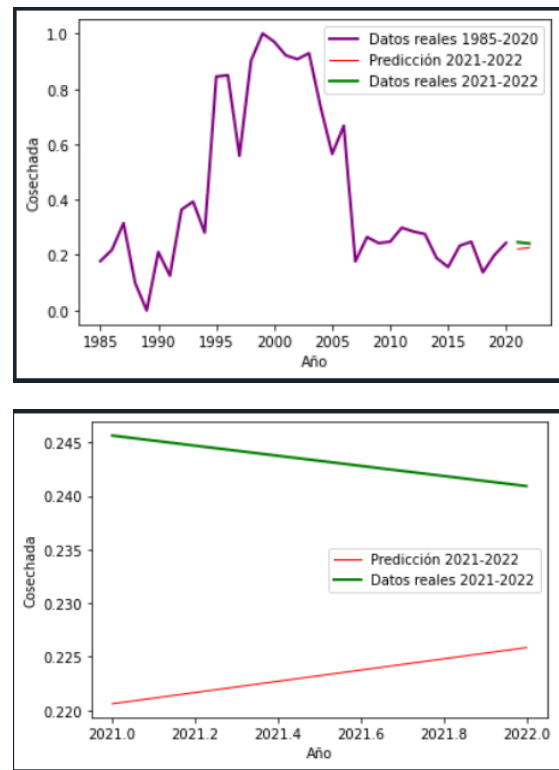


Figura 4. Regresión lineal múltiple: estado de Chiapas

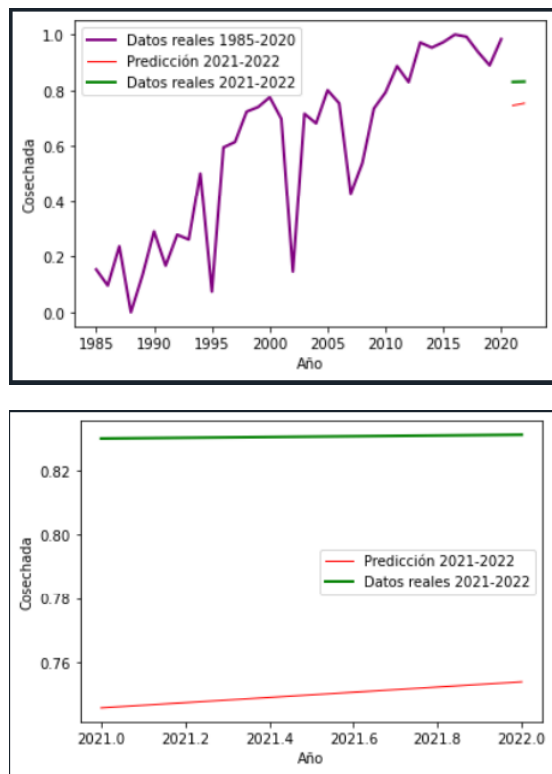


Figura 5. Regresión lineal múltiple: estado de Campeche

variables son inherentemente inestables y experimentan cambios constantes, volviendo la predicción difícil. Se sostiene que se podría potenciar la eficacia del modelo, incorporando nuevas variables independientes que contribuyan a la predicción, como la temperatura, la sequía y aspectos económicos.

Además, se puede buscar obtener datos más precisos. Dado que los datos utilizados para este estudio son anuales, se sugiere que en futuras investigaciones se considere la posibilidad de trabajar con datos mensuales. Esta estrategia permitirá tener más datos de entrenamiento, así como una menor distancia entre cada dato, lo que contribuirá a obtener una mejor eficacia del modelo.

#### REFERENCIAS

- [1] Marini K (eds.) Cramer W, J Guiot. Resumen de medecc 2020 para los responsables de la formulación de políticas. en: Cambio climático y ambiental en la cuenca mediterránea: situación actual y riesgos para el futuro. primer informe de evaluación del mediterráneo. pages 6–9, 2020.
- [2] Comisión Nacional del Agua [CONAGUA] and López M. Chable L. Loranca Y. Ledesma J. Zavala A. Martínez A Pascual, R. Reporte del Clima en México. Technical report, 2020.
- [3] C. Gay, F. Estrada, C. Conde, H. Eakin, and L. Villers. Potential impacts of climate change on agriculture: A case of study of coffee production in veracruz, mexico. *Climatic Change*, 79(3-4):259–288, December 2006.
- [4] Suvidha Jambekar, Shikha Nema, and Zia Saquib. Prediction of Crop Production in India Using Data Mining Techniques. *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 8 2018.
- [5] Neha Rale, Raxitkumar Solanki, Doina Bein, James Andro-Vasko, and Wolfgang Bein. Prediction of Crop Cultivation. *Annual Computing and Communication Workshop and Conference (CCWC)*, 1 2019.
- [6] Saboor A. Ali I. Mohsin A. Q Shakoor, U. Impacto del cambio climático en la agricultura: evidencia empírica de la región árida. 2011.