México desde el espacio: cómo obtener, procesar y calibrar una serie de tiempo de imágenes satelitales de iluminación nocturna para estimar el crecimiento económico de las entidades federativas[1]

Carlos Petricioli[2]

Resumen

En este trabajo se muestra cómo utilizar la serie de tiempo de imágenes satelitales de iluminación nocturna de la tierra de la NOAA. Se realiza la metodología necesaria para hacer comparable intertemporalemente e intersatelitalmente (para algunos años se cuenta con información de dos satélites) para finalmente poder utilizar los datos de intensidad de iluminación visible desde el espacio, medida en Número Digital con valores de 0 a 63, para poder estimar el crecimiento del Producto Interno Bruto de las Entidades Federativas de México. Estas imágenes son promedios anuales de iluminación global nocturna que se conforman de pixeles que representan 30 arcosegundos o un kilómetro cuadrado cada uno, la serie completa alcanza alrededor de 290 millones de pixeles sólo para México. Es por esto que se pretende explicar cómo trabajar con estos datos desde la descarga. Para empezar, se presenta una literatura extensa de cómo utilizar series de tiempo espaciales y literatura sobre el uso que se la ha dado a ésta serie de iluminación espacial. Además, se generó una serie de tiempo de iluminación nocturna calibrada para poder hacer un mejor uso de ésta. Adicionalmente, se estimó una serie de tiempo de índices de Gini espaciales de desigualdad de iluminación a nivel Estatal con el cual se puede medir la desigualdad en todos los Municipios dentro de los Estados, una métrica totalmente nueva en las cuentas nacionales. Finalmente, se extiende su aplicación presentando el marco teórico necesario para hacer un modelo del crecimiento del PIB Estatal durante los años de ésta serie y utilizando la iluminación nocturna como variable *proxy*. En el modelado se obtuvieron resultados muy deseables gracias a la calibración. Se encontró que la iluminación nocturna representa una buena *proxy* para corregir el crecimiento económico que reporta el INEGI correspondiente a las Entidades. En algunos casos, se encontró que el INEGI puede llegar a sobre o sub estimar el crecimiento de largo plazo correspondiente a 10 años en las Entidades hasta en un 15 %. Esta tesis plantea una opción viable para corregir este sesgo.

Tópicos: Análisis geoespacial, DMSP-OLS, Night-lights, GIS, Intercalibración satelital, Estimación PIB subnacional, NLDI, Gini, Proxy, Crecimiento Económico, Minería de Datos, R, Rgeos.

[1] Última actualización: 11 de junio de 2015.

[2] Agradezco los comentarios de Adolfo de Unánue quien asesoró y dio la idea de este proyecto.

carpetri@gmail.com mexico-nocturno.carlospetricioli.com carlospetricioli.com

Capítulo 1.

Introducción

Lo primero que cualquiera se debe de preguntar al leer este trabajo es ¿por qué usar imágenes satelitales de iluminación nocturna? Responder es sencillo si se analiza detalladamente. Para empezar, los datos de iluminación son una medida objetiva y altamente sistemática en comparación con la mayoría de los indicadores comunes como el producto interno bruto (PIB), los censos de población y vivienda, estimaciones e indicadores de pobreza e indicadores de marginación y bienestar social. Además, muchas de las decisiones públicas y privadas se toman día con día con base en aquellas variables que no tienen la característica de ser tan objetivas y sistemáticas como lo son las luces y que, de hecho, tienen como base estimaciones y aproximaciones que contienen errores de medición y supuestos teóricos que las hacen menos confiables. Estos indicadores suelen utilizarse para comparaciones que, en su mayoría, no deberían de hacerse por la amplitud de fuentes de las que provienen; se generan en distintos países, organizaciones, periodos y con distintas metodologías y calidades. Inevitablemente, estos problemas generan sesgos que comúnmente se subestiman. Por el contrario, la serie de datos de iluminación nocturna satelital consiste de datos que se toman de un mismo satélite, haciéndolas comparables en todo el mundo y a una muy alta resolución (ver capítulo 3 para más detalle sobre esto). Es importante señalar que dado que la tierra es una esfera y los datos se proyectan en un plano (imagen), hay algo de distorsión a medida que los pixeles, que corresponden a un lugar específico de la tierra, se alejan del Ecuador; sin embargo, esto no es un problema porque ésta distorsión es perfectamente cuantificable. Por el contrario, el hecho de que sean datos *espaciales* representa

una ventaja ante indicadores, como los que se mencionaron anteriormente, que no están directamente relacionados a coordenadas terrestres bien definidas como es en el caso de las luces satelitales. En resumen y para responder la pregunta, se pretende utilizar los datos de iluminación nocturna satelital porque representan una medida relativamente más precisa y objetiva de medir en el mundo debido a que la resolución permite hacer comparaciones tan finas como a nivel *manzana urbana*[1] a un mismo tiempo y con una relación espacial. Es por lo anterior y porque la iluminación, sin considerar algunas pequeñas excepciones como incendios, reflejos lunares y relámpagos, se genera únicamente por el hombre por lo que se plantea utilizar la información de iluminación tomada del espacio como una media de estimación.

Por otra parte, medir la actividad económica en México es una tarea importante que puede servir, entre otras cosas, como una medida confiable de bienestar en el país. Como dicen Henderson, et. al. en su artículo [24],

> "El producto interno bruto (PIB) es la variable más importante en el análisis del crecimiento económico. Los problemas conceptuales para definir el PIB, más aún usarlo como medida de bienestar, son temas de los cursos de introducción a economía. Igual de relevante, sin embargo, es el problema de que el PIB por sí sólo suele estar mal medido, especialmente en países en desarrollo. En cuanto a los países desarrollados, en muchos países una mucho menor fracción de la actividad económica se lleva a cabo en el sector formal, el grado de integración económica y regularización de precios entre regiones es menor y, más significativo aún, la infraestructura estadística del gobierno es más débil. Estos factores generan que calcular el PIB nominal (valor agregado total, en precios domésticos) sea difícil." [2]

[1] "Una manzana urbana es un espacio geográfico que está constituido por un grupo de viviendas, edificios, predios, lotes o terrenos de uso habitacional, comercial, industrial, de servicios, entre otros." Ver INEGI en [28].

[2] En inglés: "Gross Domestic Product (GDP) is the most important variable in analyses of economic growth. The conceptual problems in defining GDP, let alone using it as a measure of welfare, are the stuff of introductory economics courses. Just as serious, however, is the problem that GDP itself is often badly measured, especially in developing countries. Relative to developed countries, in many developing countries a much smaller fraction of economic activity is conducted within the formal sector, the degree of economic integration and price equalization across regions is lower, and, most significantly, the government statistical infrastructure is weaker. These factors make the calculation of nominal GDP (total value added, in domestic prices) difficult."

Con base en la metodología que realizó Henderson, esta tesis propone un marco estadístico para poder dar una solución a cómo medir el crecimiento del PIB real utilizando la información de iluminación generada desde del espacio. Esto, además, porque un problema adicional que surge al momento de medir el crecimiento económico, es que el gobierno no genera datos a nivel sub-nacional de la misma manera en la que genera datos nacionales y sin embargo, una buena parte del desarrollo y crecimiento económico se da a este nivel. De hecho, en México se cuenta con información poblacional a niveles muy desagregados hasta nivel manzana urbana y sin embargo no se cuenta con una medida del PIB a estos niveles. Por esta razón y dado que la iluminación nocturna representa una fuente de información muy rica y confiable de regiones que miden a partir de un kilómetro cuadrado a la altura del Ecuador, se plantea generar una medida del PIB a nivel Estatal por medio de estimaciones de la actividad económica medida con los datos de iluminación nocturna. Esto, además, sin considerar que la información de los satélites se tiene con una frecuencia mucho mayor que la mayoría de las cuentas nacionales en México.

Para contrastar un poco, en la literatura económica es común utilizar variables *proxy* para evitar el problema de medición en las cuentas nacionales importantes y poder tener mediciones consistentes para cuando no hay datos disponibles. Por ejemplo, dicen Henderson et. al. en [24] que la Reserva Federal basaba su índice de producción en una encuesta que medía la electricidad que se le administraba a consumidores industriales. De manera similar, los historiadores económicos han utilizado estas herramientas en tiempos en los que no se contaba con cuentas nacionales, como Good [20], que en 1994 estimó el producto en 22 regiones del imperio de Haubsburgo para el periodo de 1870 a 1910 utilizando *proxys* como el número de cartas enviadas per cápita.

El objetivo más acotado de este trabajo pretende utilizar datos de iluminación nocturna tomadas desde el espacio exterior por medio de satélites como variable *proxy* para poder estimar el crecimiento económico de México y así determinar qué tan buena herramienta es la iluminación nocturna como variable *proxy* para estimar el crecimiento económico del país. Como objetivos adicionales, pretende explicar a detalle y servir de herramienta libre sobre cómo obtener los datos espaciales de iluminación y explicar cómo utilizar las herramientas necesarias para hacer el análisis estadístico y geoespacial correspondiente para explotarlos.

Este artículo es muy distinto de lo que se puede encontrar en la literatura principalmente por dos razones. La primera, es de los pocos que utilizan los datos de iluminación

nocturna a un nivel sub nacional; y la segunda, este es el único que se ha hecho específicamente para México que utilice la serie completa a nivel Estatal de lo que se ha publicado. Adicionalmente, dentro de la literatura revisada, este trabajo es el primero en el que se utiliza toda la serie completa de imágenes satelitales de iluminación, desde 1992 hasta 2013 (ver más detalles en la sección 2.3 del capítulo 2).

Una vez que se explicó la motivación para utilizar las luces, en el capítulo 2 se presenta un pequeño contexto teórico de qué se entiende como Estadística Espacial, qué tipo de datos se utilizan en esta rama de la estadística y cuáles son las técnicas de modelado más comunes y las complicaciones más recurrentes que suele haber en el uso de datos espaciales. Más adelante, en el mismo capítulo, se explica qué los datos espaciales dependen de los Sistemas de Información Geográfica para su correcta interpretación y uso.

Adicionalmente en el capítulo 2 se presenta una revisión extensa de la literatura relacionada al uso específico de la serie de iluminación nocturna satelital visible del espacio que se utiliza en este trabajo. Se puede ver que se ha utilizado desde hace más tiempo de lo que se esperaba, en una gran variedad de aplicaciones y que, sin embargo, no se ha explotado del todo su potencial. Este trabajo representa un primer intento para lograrlo.

Para terminar el capítulo 2 se presentan unas definiciones del Producto Interno Bruto y el Producto Interno Bruto Estatal para contextualizar con un poco más de profundidad el alcance que puede tener este trabajo.

Más adelante, en el capítulo 3 se presentan los detalles completos de la serie de iluminación nocturna satelital. Se define qué se entiende por los datos, cómo se obtienen y por qué es que se eligió utilizar esta serie en específico. Además, se presenta una solución a un problema importante que presenta esta serie. Como se detallará, los datos de la serie de iluminación nocturna no son directamente comparables entre distintos años ni entre distintos satélites. Es por esto que se plantea una metodología completa para poder resolver este tema.

Por otra parte y como complemento al trabajo realizado en este proyecto, en el capítulo 3, se explica qué es un índice de Gini de desigualdad y para estimar una serie de desigualdad de iluminación nocturna satelital para todos los años de la serie de iluminación correspondiente a cada una de las Entidades Federativas en México. Esta medida puede ser muy útil debido a que no se cuenta con muchos índices de este estilo desagregados a este nivel por lo que se considera un valor agregado importante.

En el capítulo 4 se presenta, por una parte, el comportamiento de la serie de iluminación

nocturna satelital en un comparativo internacional para México contra los países miembros de la Organización para la Cooperación y Desarrollo Económico (OCDE). Por otro, el comportamiento de la serie dentro de los Estados de México y a través de los años. En este capítulo se incluye también el comparativo de los resultados obtenidos para la serie de desigualdad espacial de Gini de iluminación nocturna satelital visible desde el espacio.

Todo lo anterior para poder contextualizar el modelo teórico que se presenta en el capítulo 5 en donde se explica cómo se puede utilizar la serie de iluminación nocturna satelital como variable *proxy* para poder estimar el crecimiento económico en las Entidades Federativas de México. Se presentan los resultados del modelado empírico, contrastando con algunas variables importantes como el porcentaje de viviendas que cuentan con electricidad y como el índice de desigualdad de Gini que se estimó en este trabajo. Con lo anterior, se comparan los resultados obtenidos con los datos de las cuentas nacionales en México.

Finalmente en el capítulo 6 se plantean las principales conclusiones, algunas consideraciones adicionales y se plantean algunas alternativas para continuar con este estudio.

Capítulo 2.
Marco teórico y literatura

2.1. Estadística espacial

Para poder utilizar la base de datos de la NOAA, es fundamental entender algunos de los métodos que existen para analizar datos espaciales cuantitativos. Un conjunto de datos se entiende como espacial si cada elemento único tiene relacionado una referencia geográfica que puede ser en coordenadas de longitud y latitud o en distintas proyecciones del globo (para más detalle ver Cliff [8]). Es decir, los datos espaciales se distinguen por observaciones obtenidas en ubicaciones espaciales $s_1, s_2, \ldots, s_n$ donde s_i pueden ser coordenadas en el plano $\mathbb{R}^2$ o $\mathbb{R}^3$ (longitud, latitud y relieve).

Para esto, se denomina análisis espacial en la literatura de los Sistemas de Información Geográfica (GIS por sus siglas en inglés) y de Ciencia de Información Geográfica (GISc por sus siglas en inglés) al conjunto de técnicas que usan datos espaciales como base del estudio[1]. En el análisis espacial se pueden realizar distintos tipos de modelado como modelado cartográfico, cuando los datos espaciales se representan en un plano y se generan resultados que se visualizan en otro plano y como el análisis estadístico espacial, que se sostiene en técnicas estadísticas en general aplicadas a datos espaciales. Los modelos espaciales pretenden simular la correlación entre observaciones en diferentes posiciones en el espacio.

En particular, la estadística espacial es un subconjunto de la estadística que se espe-

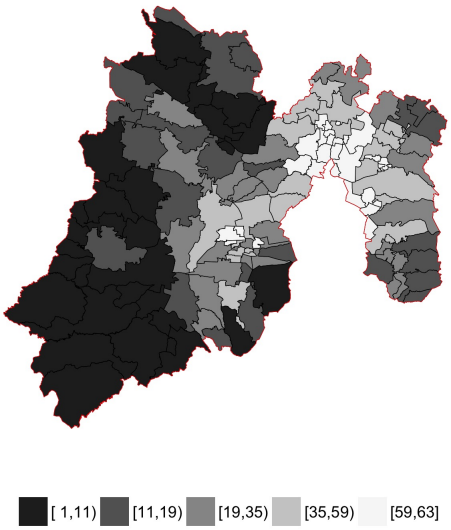
[1]Se puede ver más detalles sobre GIS en el libro de Longley, et. al. [34]

cializa en el análisis e interpretación de datos georeferenciados. Algunas de las preguntas típicas en la esta estadística espacial incluyen: cómo varia una variable en el espacio, qué controla ésta variación espacial, en dónde se puede encontrar muestras para describir la varianza espacial, cuántas observaciones se necesitan para representar el comportamiento espacial de cierta variable e incluso predecir cuál es valor de una variable en una nueva ubicación (Goovaerts en [21]). La estadística espacial tiene aplicaciones en áreas como climatología, ecología, salud y bienes raíces, dado que en estas áreas es común contar con información espacial que se quiere modelar.

Finalmente, para fines de este trabajo, se utilizarán tres tipos de enfoques de análisis espacial. El primero, el análisis de datos espaciales de variación discreta que se caracteriza por datos de la variable de interés sobre una región espacial que en este caso, serán las variables del INEGI correspondientes a las Entidades Federativas de México o los datos de iluminación relacionados a cierta área. Es decir, para cada M región de interés se tiene cada punto $s = (long, lat)$ en M en el plano. Se puede definir como un proceso $Y(s)$ que es un vector aleatorio en una ubicación $s \in \mathbb{R}^r$ donde s se refiere a una región M sobre el dominio $D \subset \mathbb{R}^r$ (usualmente $r \in 2, 3$) que se encuentra particionado en un número finito de unidades de área con fronteras bien definidas. Por ejemplo, la imagen 2.1 a continuación muestra datos espaciales de variación discreta, cada polígono se refiere a un municipio en el Estado de México y el color corresponde al grupo de iluminación nocturna al que pertenece cada uno. Se puede notar que los municipios más iluminados (colores más claros representan mayor iluminación) son los que están más pegados al Distrito Federal[2].

[2]En este trabajo se utilizaron como referencia los municipios en México vigentes en el Marco Geoestadístico 2010 versión 5.0, en donde no se encuentran incorporados cinco municipios debido a que fueron creados posterior a junio de 2010, que es la fecha de referencia de la versión 5.0 del Marco Geoestadístico. "El municipio de Bacalar del estado de Quintana Roo creado por el decreto 421, publicado el 17 de febrero de 2011 en el Periódico Oficial de la entidad. Cuatro municipios del estado de Chiapas creados por el Decreto 008, publicado el miércoles 23 de noviembre de 2011 en el Periódico Oficial No. 337 de la entidad, por el que se establece la Tercer Reforma a la Constitución Política del Estado de Chiapas y se crean los municipios de Mezcalapa, El Parral, Emiliano Zapata y Belisario Domínguez." Para más detalle consultar el Marco Geoestadístico del INEGI en [28].

Imagen 2.1: Quintiles de iluminación en el Estado de México, 2010



Fuente: Elaboración propia con datos de la NOAA [1] y polígonos de municipios del INEGI [28]

El segundo enfoque, el análisis de datos de variación continua, comprende procesos estocásticos cuyos valores pueden ser conocidos en todos los puntos del área de estudio. Es decir, en vez que cada observación se refiera a un área delimitada, como en el caso anterior, aquí se cuenta con datos para cada par de coordenadas. El tercer enfoque, análisis de procesos espaciales de puntos, se debe pensar como un conjunto de datos en una retícula, donde se acota el área M como un conjunto espacial donde se presentan observaciones. Los datos en la retícula se relacionan a su respectiva longitud y latitud que generan un conjunto de puntos distribuidos en un área fija cuya ubicación fue generada por un mecanismo estocástico.

En la estadística espacial, el proceso de modelado típicamente sigue la siguiente estructura. Primero, se diseña el método de muestreo y de procesamiento de datos. Para fines de este proyecto, es el trabajo de calibrar los satélites para que se capturen datos sólo en las tardes (diario de 20:00 a 22:00 aproximadamente) y filtrando las observaciones libres de nubes por medio de los sensores DMSP-OLS que hace la NOAA (ver capítulo 3 para más detalles). Segundo, el proceso de recolección de datos, análisis y pre-procesamiento. La NOAA limpia las observaciones de forma que no haya observaciones ruidosas, por ejemplo

por iluminación que proviene de relámpagos o incendios. Tercero, análisis de los datos y estimación del modelo. Cuarto, implementar el modelo y evaluar su desempeño. Finalmente, el quinto, producir y distribuir los resultados. Los pasos tercero a quinto son el trabajo de esta tesis.

Algunas técnicas de modelado espacial incluyen (ver Hengl [25]):

- Regresión con coordenadas.

  En este modelado se asume que la variable a modelar es función de las coordenadas para ciertas coordenadas dadas. En la literatura se denomina tendencia de superficies e interpolación en movimiento de superficies, dependiendo si la función se estima para todos los puntos de la muestra o para una vecindad local respectivamente. La regresión se puede escribir como

  $$Z(s) = f(x, y) + \epsilon, \tag{2.1}$$

  y las predicciones se pueden hacer con

  $$\hat{z}(s_0) = \sum_{r,s \in n} a_{r,s} x^r y^s = a^T s_0, \tag{2.2}$$

  en donde $r + s < p$ es el número de transformaciones de las coordenadas y p es la dimensión de la superficie.

  Los coeficientes del modelo se pueden estimar por medio de la solución de mínimos cuadrados, si se supone un error cuadrático.

- Splines

  Como los splines tradicionales, se busca ajustar un polinomio por partes al conjunto de datos. Existen varios tipos de splines que consideran diferentes técnicas de suavizamiento. Los más comunes son splines regularizados con tensión y suavizamiento. En este caso, los coeficientes se ajustan con la ecuación

  $$\hat{z}(s_0) = a_1 + \sum_{i=1}^{n} w_i R(v_i), \tag{2.3}$$

  en donde a_1 es una constante y $R(v_i)$ es una función de radio que se determina con

las ecuaciones

$$R(v_i) = -(E_1(v_i) + \log(v_i) + C_E), \tag{2.4}$$

$$v_i = \left[\psi \frac{h_0}{2}\right]^2, \tag{2.5}$$

en donde $E_1(v_i) = \int_{-\infty}^{v_i} e^t/t \, \mathrm{d}t$ es la exponencial integral, $C_E = 0.577215$ es la *constante de Euler*, ψ es un parámetro de tensión y h_0 es la distancia entre la estimación y el punto a interpolar. Los coeficientes a_i y w_i se obtienen con la solución al sistema dado por las ecuaciones (2.6) y (2.7)

$$\sum_{i=1}^{n} w_i = 0, \tag{2.6}$$

$$a_1 + \sum_{i=1}^{n} w_i \left[R(v_i) + \delta_{i,j} \frac{\bar{\omega}_0}{\bar{\omega}_i}\right] = z(s_i) \quad \forall j = 1, \ldots, n; \tag{2.7}$$

en donde $\frac{\bar{\omega}_0}{\bar{\omega}_i} > 0$ son factores de peso para el suavizamiento en cada punto s_i. El parámetro de tensión controla la distancia sobre la que los puntos dados influyen la superficie resultante, mientras que el parámetro de suavizamiento controla las desviaciones verticales de la superficie a los puntos dados. Estos splines regularizados con tensión y suavizamiento pueden ser equivalentes al método de *kriging* (ver más adelante), en donde las coordenadas se utilizan para explicar la parte determinística de la varianza. Pueden producir resultados muy similares.

- *Kriging* En la estadística espacial se suele utilizar coloquialmente como *interpolación espacial*. La versión estándar se denomina *kriging ordinario*. Las predicciones se basan en el siguiente modelo:

$$Z(s) = \mu + \epsilon'(s) \tag{2.8}$$

donde μ es una función constante estacionaria (media global) y $\epsilon'(s)$ es la parte estocástica de la variación. Las predicciones se hacen como en la ecuación (2.9)

$$\hat{z}_*(s_0) = \sum_{i=0}^{n} w_i(s_0) z_i(s_i) = \lambda_0^T z, \tag{2.9}$$

en donde λ_0 es el vector de pesos de *kriging* (w_i), z es un vector de tamaño n de observaciones espaciales. De alguna forma se está ponderando cada región. Intuitivamente, se estiman los pesos de forma que reflejen la estructura de autocorrelación espacial. Para analizar los datos se utiliza la semivarianza, que se refiere a variaciones entre valores vecinos con la siguiente ecuación:

$$\lambda(h) = \frac{1}{2}\mathbb{E}\left[(z(s_i) - z(s_i + h))^2\right], \tag{2.10}$$

en donde $z(s_i)$ es el valor de la variable objetivo a un punto s_i y $z(s_i + h)$ es el valor en la vecindad de distancia $s_i + h$. Suponemos que hay n observaciones por lo que se tienen $n(n-1)/2$ pares para calcular la semivarianza. Se puede graficar las semivarianzas contra sus respectivas distancias, esto produce un variograma. Sin embargo, estas suelen ser complicadas de interpretar por lo que, comúnmente se promedian. Lo común es que las semivarianzas sean pequeñas a distancias cortas y que se estabilizan a distancias mayores.

2.1.1. Regresión kriging espacio-temporal

Los modelos de *kriging* tradicionales en $\mathbb{R}^2$ se pueden extender a una dimensión más para incluir el tiempo. A esto se le denomina geoestadística espacio-temporal (ver detalles en Kyriakidis and Journel, 1999 [31] y Gneiting et. al. [19]) En este caso, el modelado está dado por la siguiente ecuación:

$$Z(s,t) = m(s,t) + \epsilon'(s,t) + \epsilon'', \tag{2.11}$$

en donde $\epsilon'(s,t)$ es el residual espacio-temporal, típicamente *ruido blanco* con $s \in \mathbb{R}^2$ y $t \in \mathbb{R}$. En este caso, los variogramas deben ser en $\mathbb{R}^3$.

Generalmente se supone que la covarianza entre $Z(s_1,t_1)$ y $Z(s_2,t_2)$ depende de las coordenadas espacio-temporales (s_1,t_1) y (s_2,t_2) y que no hay más estructura de correlación. Sin embargo, en la práctica, la estimación y el modelado requieren de supuestos de simplificación como estacionariedad, separabilidad y simetría[3].

Definición 2.1.1. El proceso Z se dice que tiene *covarianza separable* si existen funciones

[3]En los datos espaciales de temperatura, corrientes de aire y mareas, este supuesto no se suele cumplir.

$\mathbf{COV}_E$ y $\mathbf{COV}_T$ que dependen puramente del espacio y tiempo respectivamente tal que

$$\mathbf{COV}(Z(s_1,t_1),Z(s_2,t_2)) = \mathbf{COV}_E(s_1,s_2)\mathbf{COV}_T(t_1,t_2), \tag{2.12}$$

para todo $(s_1,t_1),(s_2,t_2) \in \mathbb{R}^2 \times \mathbb{R}$.

Definición 2.1.2. El proceso Z se dice que tiene *covarianza simétrica* si es tal que

$$\mathbf{COV}(Z(s_1,t_1),Z(s_2,t_2)) = \mathbf{COV}(Z(s_2,t_2),Z(s_1,t_1)), \tag{2.13}$$

para todo $(s_1,t_1),(s_2,t_2) \in \mathbb{R}^2 \times \mathbb{R}$.

Definición 2.1.3. El proceso Z se dice que tiene *covarianza estacionaria espacial* si es tal que $\mathbf{COV}(Z(s_1,t_1),Z(s_2,t_2))$ depende del espacio s_1 y s_2 sólo por medio del vector de separación $s_1 - s_2$.

Similarmente, el proceso Z se dice que tiene *covarianza estacionaria temporal* si es tal que $\mathbf{COV}(Z(s_1,t_1),Z(s_2,t_2))$ depende del tiempo t_1 y t_2 sólo por medio del vector de rezago $t_1 - t_2$.

Finalmente, si el proceso espacio-temporal Z tiene *covarianza estacionaria espacial* y *covarianza estacionaria temporal* se dice que el proceso tiene *covarianza estacionaria*. Es decir, si existe una función C definida en $\mathbb{R}^2 \times \mathbb{R}$ tal que

$$\mathbf{COV}(Z(s_1,t_1),Z(s_2,t_2)) = C(s_1 - s_2, t_1 - t2) \tag{2.14}$$

para todo $(s_1,t_1),(s_2,t_2) \in \mathbb{R}^2 \times \mathbb{R}$. Se denomina a C como la *función de covarianza espacio-temporal* del proceso y se restringe a que $C(\cdot,0)$ y $C(0,\cdot)$ son funciones que dependen puramente del espacio y tiempo, respectivamente.

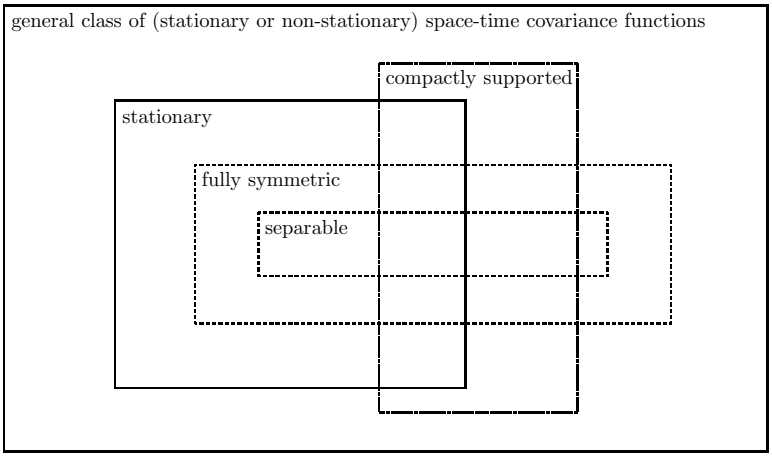
Más aún, si un proceso espacio temporal Z es estacionario con función de covarianza $C \in \mathbb{R}^2 \times \mathbb{R}$, se puede escribir la matriz como

$$\left(C(s_i - s_j, t_i - t_j)\right)_{i,j=1,\ldots,n}. \tag{2.15}$$

La imagen 2.2 a continuación resume las posibles combinaciones de supuestos que puede presentar un proceso espacio-temporal. La clase más grande es la estacionaria o no estacionaria. Es importante notar que una covarianza separable puede ser estacionaria o no. Sin

embargo, una separable es siempre simétrica y no viceversa. Esto tiene implicaciones en el modelado porque para rechazar separabilidad es suficiente con rechazar simetría.

Imagen 2.2: Diagrama de las relaciones de covarianza: separable, simétrica, estacionaria de un proceso espacio-temporal



Fuente: Diagrama tomado de Gneiting et. al. en [19].
Nota: En ocasiones se pide que la covarianza del proceso Z tenga soporte compacto de forma de que $\mathbf{COV}(Z(s_1,t_1),Z(s_2,t_2))=0$ cuando $||s_1-s_2||$ y/o $|t_1-t_2|$ sea suficientemente grande.

2.1.2. Autocorrelación espacial y temporal

Dado lo anterior, en el proceso de modelado de la estadística espacio-temporal, así como en la estadística en general, es necesario considerar que los datos pueden presentar autocorrelación temporal y espacial. La autocorrelación espacial se puede expresar informalmente como la primera ley de la geografía de Tobler [47]

> "Todo esta relacionado con todo, pero las cosas cercanas están más relacionadas que las lejanas."

En el análisis estadístico la no independencia de las observaciones puede ser un detalle importante a considerar debido a que hay algunos resultados en la teoría estadística que dependen del supuesto de independencia. En específico, en estadística espacial, las observaciones en diferentes puntos espaciales pueden no ser independientes. Por lo tanto, como

explican Cliff y Ord en su libro [7], hay que tener en cuenta que los datos espaciales pueden tener autocorrelación.

En el caso más simple, si suponemos que un proceso $Z(s,t)$ tiene covarianza estacionaria temporal, se entiende por autocorrelación espacial cuando diferentes observaciones correspondientes a posiciones relativamente cercanas pueden ser relativamente más similares que en lugares lejanos. En específico, si se tiene autocorrelación espacial positiva, los valores similares están cercanos entre sí o aglomerados; por otro lado, la negativa representa que datos parecidos están dispersos en el espacio y; finalmente, autocorrelación cercana a cero representa que los datos se distribuyen aleatoriamente en el plano.

Para saber si datos espaciales presentan este fenómeno, se pueden realizar pruebas estadísticas que evalúan si un dato es independiente de valores de la misma variable en lugares que se puedan entender como cercanos. En particular, se debe calcular la similitud entre valores en las regiones s_i y s_j y ponderar considerando qué tan cercanos son. Por lo tanto, si se tienen medidas altas en los resultados de estas pruebas, significa que los datos presentan un nivel alto de similitud principalmente en las variables a las que se les asignó mayor peso y; contrariamente, niveles bajos representan baja similitud en las mismas. Para saber si datos espaciales presentan estos fenómenos, se pueden realizar pruebas estadísticas que evalúan si un dato es independiente de valores de la misma variable en lugares que se puedan entender como cercanos. En particular, se debe calcular la similitud entre valores en las regiones s_i y s_j y ponderar considerando qué tan cercanos son. Por lo tanto, si se tienen medidas altas en los resultados de estas pruebas, significa que los datos presentan un nivel alto de similitud principalmente en las variables a las que se les asignó mayor peso y; contrariamente, niveles bajos representan baja similitud en las mismas.

Similarmente, en el caso de que suponemos que un proceso $Z(s,t)$ tiene covarianza estacionaria espacial, se entiende por autocorrelación temporal cuando diferentes observaciones correspondientes a tiempos distintos están relacionados, en este caso la cercanía no es tan relevante. En específico, si se tiene autocorrelación temporal, positiva o negativa, se dice que el proceso tiene una tendencia, positiva o negativa respectivamente. Una técnica común es tomar diferencias, de distintos órdenes, del proceso para trabajar con series estacionales. Similarmente, para saber si un proceso presenta autocorrelación se pueden hacer pruebas estadísticas, aunque, un método gráfico común consiste en analizar las funciones de autocorrelación y de autocorrelación parcial (para más detalle ver, Guerrero en [22]).

Finalmente, en el caso de que el proceso $Z(s,t)$ no tenga covarianza estacionaria espacial ni temporal, puede darse el caso de autocorrelaciones espacio-temporales. Es decir, que valores que tienen posiciones relativamente cercanas pueden ser relativamente más similares que en lugares lejanos con ciertos rezagos temporales.

2.2. Análisis de GIS

Existen distintos programas para hacer análisis geoespacial en el área de GIS. En este trabajo se utilizó principalmente el software Quantum-GIS (QGIS)[4]. Para poder realizar el procesamiento de las imágenes georeferenciadas se tomó como guía el artículo de Matt Lowe [35] en donde explica cómo realizar un preprocesamiento adecuado a imágenes satelitales en el software ArcGis[5] y se extrapoló para poder replicarlo en QGIS.

En GIS hay dos tipos de formatos típicos para realizar el análisis espacial, *Raster* y *Shape*. En este caso, las imágenes de la tierra son de tipo *Raster*. Un formato *Raster* es, simplemente, una matriz en donde cada entrada representa un pixel en la imagen y se encuentra ligado a un par de coordenadas geográficas. Para las imágenes, cada pixel del *Raster* representa un ND (Número Digital). Los archivos de tipo *Shape* son un conjunto de coordenadas que en conjunto generan polígonos en donde cada polígono puede estar relacionado a una base de datos relacional o a algún conjuro de tablas. Se puede ver un ejemplo de un archivo shape en el mapa del Estado de México en la imagen 2.1. Cada polígono es un conjunto de coordenadas que delimitan cada municipio en el Estado. El conjunto de polígonos hace al archivo *shape*.

Los datos espaciales varían mucho en cuanto a la manera en la que guardan y determinan las posiciones geográficas. Esto aplica tanto para datos espaciales obtenidos por medio de un sistema de posicionamiento global (GPS, por sus siglas en inglés) tanto como para datos espaciales de otras fuentes como las imágenes satelitales. Es por esto que se requiere tener conocimiento sobre los Sistemas de Referencias de Coordenadas (CRS, por sus siglas en inglés) para poder establecer las unidades de medición de las posiciones de las coordenadas y así poder definir distancias y describir la topología de las posiciones relativas entre observaciones. Por ejemplo, se puede hablar de CRS en grados asociados a una circunferencia o elipse, como es en el caso de la tierra, en donde se define un meridiano primal que

[4]No requiere el pago de una licencia.
[5]Requiere el pago de una licencia.

representa el origen en longitud y un *datum*, que es un CRS con respecto a un punto de origen en tres dimensiones, incluida la altura. Por otro lado, un CRS proyectado se refiere a un modelo geométrico específico representado en un plano. La mayoría de los países, por diversas razones, tienen más de un CRS por lo que es importante saber con cuál se está trabajando para poder hacer los datos comparables porque en general los distintos CRS consideran distintos modelos de elipses y meridianos de origen (ver Vivand et. al. en [4]).

Para fines de este trabajo, los datos que se estudian son observaciones en la Tierra (esfera, $\mathbb{R}^3$) que se trabajan en un plano ($\mathbb{R}^2$), por lo que es importante saber con qué CRS se está trabajado y con qué proyección para que distintos datos, como de mapas y las imágenes, sean compatibles. Sólo así los mapas, datos espaciales, pueden ser parte del análisis exploratorio y servir para ajustar un modelo espacial.

Además, todas las proyecciones inducen alguna distorsión y las proyecciones de los mapas serán diferentes por lo tanto hay que tener cuidado al momento de hacer comparaciones en los mapas porque no necesariamente se encuentran en la misma proyección. Por ejemplo, dos proyecciones comunes son:

1. Proyección de Mercator: La proyección de Mercator es un tipo de proyección cartográfica cilíndrica, que pretende representar la superficie esférica terrestre sobre una superficie cilíndrica, tangente al ecuador, que al desplegarse genera un mapa terrestre plano. Los problemas típicos son que por ejemplo Groenlandia aparece aproximadamente del tamaño de África, cuando en realidad el área de África es aproximadamente 14 veces la de Groenlandia o Alaska aparece similar en tamaño a Brasil, cuando el área de Brasil es casi 5 veces la de Alaska.

2. Proyección de Albers: La proyección de Albers, a diferencia de la anterior, preserva áreas. Es una proyección que pretende representar la superficie esférica sobre una cónica tomando dos paralelos estándar de referencia. Esta proyección es más común para los Británicos.

En la imagen 2.3 a continuación se pueden visualizar estos dos métodos para hacer proyecciones.

Imagen 2.3: Mapa de Estados Unidos con proyecciones de Mercantor y Albers

Fuente: Elaboración propia con datos del paquete `maps` [2].

Para este trabajo la proyección que se utilizó fue la de Mercantor[6] porque es la que utiliza la NOAA en sus imágenes satelitales. De hecho, se tuvieron que re-proyectar los datos de polígonos del INEGI de los Estados y Municipios para hacerlos compatibles (ver detalles del código en el apéndice B sección B.3).

2.3. Revisión de literatura

Para dar una idea general de las aplicaciones que puede tener el uso de las imágenes de iluminación nocturna espacial, esta sección resume una parte de la literatura relacionada. Las aplicaciones varían bastante e incluyen estimaciones de factores económicos, consumo de energía eléctrica, sociales y demográficas hasta índices de temas sobre medio ambiente y geografía.

Los datos de la NOAA de iluminación nocturna se han utilizado desde hace aproximadamente dos décadas. Los economistas, los más frecuentes en la literatura de iluminación nocturna espacial, intentan estimar el ingreso total o el PIB en su mayoría a nivel país y con el objetivo de analizar países con mala contabilidad nacional. Por ejemplo, Elvige et. al. en [15] en 1997 estudiaron la correlación que hay entre la iluminación nocturna satelital con actividades económicas. En específico, estimaron el área iluminada por luces antropogénicas visibles cercanas infrarrojas; es decir, luces, para 21 países usando los datos satelitales de los sensores de la DMSP-OLS. Encontraron que el área iluminada tiene un alta correlación al PIB y al consumo eléctrico aunque encontraron bastantes datos atípicos

[6] $+proj = longlat + ellps = WGS84 + no_def s + towgs84 = 0, 0, 0$

en la relación entre el área iluminada y la distribución de la población.

En otro de los primeros artículos, en el 2000, Doll, et. al. [11] usaron los datos satelitales para generar, por primera vez, un mapa del PIB con paridad de poder de compra (PPP, por sus sus siglas en inglés) a una resolución de un grado en coordenadas geográficas. Ellos usaron satélites de 1994 y 1995 se concentraron en el área iluminada de las ciudades y la combinaron con datos estadísticos para investigar el potencial de las luces y así estimar parámetros scocioeconómicos globales y para, finalmente, modelar la población urbana. También concluyeron que el área iluminada en un país está coorelacionada con el PIB y con las emisiones de dióxido de carbono (CO_2).

Un poco más adelante en el 2005, Ebener et. al. [13] realizaron un primer análisis para proponer un modelo global para predecir el ingreso per cápita nacional por medio del uso de las imágenes satelitales. Además, fueron pioneros en utilizar los datos para bajar a un nivel sub nacional para todos los países y en extrapolar los resultados a países que no cuentan con contabilidad nacional. Concluyeron que las imágenes son prometedoras en la extrapolación y para predecir el ingreso per cápita a nivel subnacional.

Además, Doll et. al. [12], también en el 2005, generaron mapas de Estados Unidos y Europa a una resolución de 5 kilómetros utilizando los datos satelitales de iluminación nocturna de 1996 y datos sub nacionales de productividad económica. Enfatizaron la ventaja que tiene presentar los datos en forma de mapas ante la forma tradicional. Encontraron correlaciones significativas con el PIB a nivel sub nacional ($R^2 = 0.979$), en Europa usaron niveles de Nomenclatura de Unidad Territorial Estadística (NUTS, por sus siglas en inglés) uno a tres, y fueron los primeros en analizar esta relación por medio de los mapas. Enfatizaron que los datos de iluminación nocturna espacial aplican más para países desarrollados, en los cuales la actividad industrial representa el 90 % de la actividad económica; mientras que en los países sub desarrollados, las mayoría de las luces provienen de los pueblos y ciudades y no de los campos de cultivo, que es de donde proviene la mayor actividad económica en estos.

Un poco más adelante Sutton et. al. [45]utilizaron los datos del año 2000 para Estados Unidos, India, China y Turquía. Estimaron los datos del PIB con los cambios en la densidad poblacional y la iluminación por medio de una regresión loglineal y tuvieron resultados razonables para predecir el PIB con la población aunque no tan satisfactorios con la suma de las luces.

Para básicamente los mismos países y aunque utilizando años distintos, se publicaron

artículos similares como los de Ghosh et.al. [17] [18] en el 2009 y 2010 que juntaron un conjunto de variables económicas y los datos de iluminación del 2006 para generar un mapa desagregado en el que estimaron la actividad económica, incluyendo la informal, en el que incluyeron a nivel sub nacional a China, India, México, Estados Unidos y a otros países a nivel nacional. Otros más como los de Zaho et.al [51] para estimar la actividad económica primaria en China. Lo mismo de Bhandari [3] en India. Todos los anteriores utilizaron sólo un año de iluminación.

No fue hasta en el 2011, que los primeros autores comenzaron a utilizar más años de la serie de los satélites. Por ejemplo Chen & Nordhaus, 2011 en [6], que utilizan la serie hasta 2008 para muchos países en categorías, según la calidad de sus estadísticas nacionales, y Kulkarni, et. al. [30], utilizando sólo la serie del 2001 al 2007, que concluyen que se puede utilizar la serie como proxy de la actividad económica para China aunque sostienen que tienen muy poco valor agregado en los países en los que se cuenta con una buena contabilidad nacional.

Fue hasta el 2012 que Henderson, et. al. en [24] y [23] que se intentó modelar el crecimiento del PIB real utilizando la serie desde 1994 hasta el 2008. Utilizan una combinación de luces por área y cuentas nacionales y subnacionales. Al igual que los anteriores, concluyen que estas estimaciones tienen muy poco valor agregado para países con buena calidad de datos sobre ingreso nacional y para los demás concluyen que sirve igual de bien que las malas estadísticas nacionales, que ya es bastante.

Los datos de iluminación nocturna satelital versión 4 de los sensores DMSP-OLS, que son los que se utilizan en este trabajo, ver más detalle en el capítulo 3, fueron desarrollados por la NOAA recientemente. En específico, se han utilizado en alrededor de 33 estudios según Huang, et. al. [27].

El reto más importante en el uso de esta serie proviene de que los datos se obtuvieron de seis distintos satélites (F10, F12, F14, F15, F16 y F18) con sensores OLS con distinta calibración. El primer artículo que trata de resolver este problema es Elvidge, et. al. [16] y otros autores como Chen, et. al. [6], Liu, et. al. [33] y Wu, et. al. [50], más adelante, son los únicos en la literatura que atacan este problema. Sin embargo, la ausencia de una calibración temporal hace muy complejo su uso así que se le dio mucho énfasis en este trabajo. Ver detalles más adelante en el capítulo 3 sección 3.3.

En resumen, en esta sección se mostró que la serie de imágenes satelitales puede tener aplicaciones en distintos temas, en particular en el contexto económico. Además, se puede

notar que, prácticamente, no se ha explotado la serie considerando todos los años (1992 a 2013). En su mayoría, los artículos suelen hacer análisis con respecto a un sólo año. Por otro lado, en los pocos artículos que utilizan más de un año de la serie de imágenes de iluminación nocturna satelital, se habla muy poco de cómo hacer comparable la serie intertemporal e intersatelitalmente y, de hecho, los pocos que hablan sobre métodos de intercalibración hacen, sin excepción, referencia y son una réplica o pequeña modificación del primer artículo de Elvidge que propone un modelo para resolver este problema.

2.4. Producto Interno Bruto

En esta sección se pretende explicar qué se entiende por el Producto Interno Bruto a nivel Estatal y cómo es que se mide. Esto porque, el Producto Interno Bruto por Entidad Federativa (PIBE) permite conocer la contribución Estatal al producto nacional y permite la evaluación de la dinámica y comparación que presentan.

El Producto Interno Bruto representa el valor de todos los bienes y servicios finales que se producen en una economía. Se puede estimar agregando todos los salarios, intereses, ganancias, las rentas, gastos, consumo, inversión, compras del gobierno y las exportaciones netas, que se componen de exportaciones menos importaciones. Ambas estimaciones deberían ser equivalentes porque el gasto de unas personas en la economía representa el gasto de otros, de forma que la suma de los ingresos debe ser igual a la suma de los gastos (ver definición del Banco Mundial en [36]). En específico, Bruto se refiere a que el valor del capital que se pierde en el proceso de producción no se deduce del valor total.

Por lo tanto hay tres maneras de medir el PIB, desde el lado del ingreso, del gasto o midiendo el valor agregado. Es decir, para medir el PIB desde el gasto, se suma el valor del consumo más la inversión más el gasto del gobierno más las exportaciones netas. Desde el ingreso, simplemente, se suma el valor de los factores de producción, trabajo y capital. Finalmente para medir el PIB desde el valor agregado se debe estimar la producción bruta y el consumo intermedio, se restan para obtener el valor agregado, este procedimiento es equivalente a medir el consumo intermedio total.

En México el INEGI estima el PIB con el método de valor agregado por medio del Sistema de Cuentas Nacionales (SCN) que establece la Organización de las Naciones Unidas (ONU). Para hacerlo, se clasifican las actividades de la producción en la economía como primarias, secundarias y terciarias, que incluyen primarias: actividades de agricultura,

ganadería, pesca, aprovechamiento forestal, pesca y caza; secundarias: actividades manufactureras, de construcción, minería, electricidad, agua y suministro de gas; y las terciarias: servicios, comercios y la actividad del gobierno.

Históricamente, los cálculos desarrollados en México sobre el PIBE, se pueden ubicar en cuatro periodos: el primero abarca desde la década de los ochenta hasta la mitad de los años noventa. En ese tiempo se desarrollaron cálculos del PIBE, en valores corrientes, a precios productor y conforme a las recomendaciones metodológicas del SCN de 1968. A partir del año 1998, se inició el segundo periodo siguiendo los criterios metodológicos del SCN, 1993. En esta etapa el desglose de resultados alcanzados se divulgó a dieciocho sectores de actividad económica: nueve grandes divisiones y nueve divisiones industriales, publicándose una serie para los años 1993-2006. En el año 2009, se presentó la tercera con periodicidad anual donde se actualizó el año base del SCN de México de 1993 al año 2003. Finalmente, en el 2014, el método de cálculo del PIBE incorpora los conceptos y metodologías de contabilidad nacional sugeridas por el Manual del SCN 2008.

Para su cálculo, se ordena la información censal por entidad federativa, se excluyen los sectores de actividad económica cuyas unidades de observación corresponden a empresas (construcción, transportes, servicios financieros y de seguros, etcétera) y se calcula, por actividad económica, los componentes de las cuentas de producción, es decir el valor bruto de producción, consumo intermedio y valor agregado bruto. En caso de haber diferencias entre la sumatoria de los valores estatales y los resultados nacionales, estas se alinean con los valores nacionales (ver INEGI en [29] para más detalles).

Capítulo 3.
Datos de iluminación nocturna espacial

La National Oceanic and Atmospheric Administration (NOAA) [37] ha recolectado imágenes satelitales del mundo desde hace varios años. Dentro de la información que tiene se incluyen datos diarios de iluminación nocturna y, de hecho, ha procesado ésta información diaria de iluminación en imágenes que resumen la iluminación de cada año en una sola imagen referenciada por medio de coordenadas (longitud y latitud). Aunque existe más de una base de datos de imágenes de iluminación nocturna y georeferenciada, en este trabajo se utiliza la base de imágenes que genera la Defense Meteorological Satellite Program Operational Linescan System (DMPS-OLS)[1] [1] porque es la serie más larga disponible, contiene información desde 1992 hasta 2013. Es importante notar que la intensidad de las luces nocturnas reflejan el uso de luces exteriores e interiores y que, en general, el consumo de prácticamente todos los bienes y servicios en la tarde requiere de luces, es por esto que se piensa que a medida que el ingreso per cápita sube, también lo hace el uso de luces por persona.

La DMPS-OLS, *nighttime lights time series* versión 4 de la serie de luces nocturnas[2] se divide en dos bases distintas.

1. Promedios de luz visible, luces estables y coberturas sin nubes, 1992 a 2013. Esta base contiene 35 archivos de datos, cada uno de alrededor de 300 Megabytes (MB) comprimidos y con 3 Gigabytes (GB) descomprimidos. Cada paquete incluye el pro-

[1] Para más detalle ver: http://ngdc.noaa.gov/eog/sensors/ols.html
[2] http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

medio visible en bruto, el promedio limpiado por cobertura de nubes, los datos que se obtuvieron sin nubes.

2. Promedio de luz x Pct. Esta base contiene 35 archivos de datos, cada uno de alrededor de 500-800 MB comprimidos y aproximadamente 3 GB descomprimidos. Este producto se deriva del numero digital promedio de la banda visible Número Digital (ND) de las detecciones de luces sin nubes multiplicado por la frecuencia de la detección de luz. La inclusión del porcentaje de frecuencias de detección normaliza los ND's por variaciones en la persistencia de luz (i.e. el valor de una luz que es detectada solamente la mitad del tiempo se ve reducida por un 50 %). Esta serie contiene las detecciones por fuegos y ruido de fondo.

Los sensores de los satélites se llaman Operational Linescan System (OLS). Consiste de dos sensores de banda ancha, uno para las longitudes de onda visibles y cercanas infrarrojas (400 a 1100 nm)[3] y otro para la longitud de onda térmica infrarroja (10.5 - 12.6 μm)[4]. El OLS es un escáner radio métrico con un campo de visión de alrededor de 3 mil kilómetros y captura imágenes a una resolución de 0.56 kilómetros que se suavizan más adelante en pixeles de 5x5 para reducir el uso de memoria[5].

3.1. Selección de la base de iluminación

Como enfatiza Pestaloozi en su artículo [40],

> "Crear imágenes satelitales globales de buena calidad desde datos sub-orbitales de satélites es un compromiso de una dificultad monumental y requiere un gran esfuerzo de procesamiento."[6]

De hecho, sólo se puede utilizar una parte de los datos que capturan los sensores OLS. Al final, cada imagen corresponde a un área del mundo entre 180 y -180 grados de longitud y de -65 a 75 grados de latitud. Es decir, como detallan Henderson, et. al. en [24], algunas partes cercanas a los polos se eliminan. Sin embargo, ellos estiman que sólo alrededor de 10

[3] Se refiere a nano-metros. 1 nm = 10^{-9}m

[4] Se refiere a micro metros. 1μm=10^{-6} m

[5] Ver a Doll en [10] para más detalle.

[6] En inglés: "Creating good-quality global images from the sub-orbital data of the satellites is an undertaking of monumental difficulty and requires a huge processing effort."

mil personas (0.0002 %) de la población mundial vive ahí por lo que no es tan significativo para el análisis.

Una vez que se descargan y descomprimen estos archivos se tienen dos tipos de archivos más, un tif y un tfw. El archivo tif es una extensión para imágenes pero en este caso la imagen tiene asociado a cada pixel un par de coordenadas (longitud y latitud), la información de las coordenadas, el tamaño de cada pixel y la información de rotación se encuentra en el archivo tfw.

Específicamente, esta base contiene información de tres tipos de iluminación. Cada conjunto tiene la estructura: $F_{sat,anio}$ donde *sat* se refiere al tipo de satélite que se utilizó para la medición y *anio* al correspondiente año. Los tres tipos de imágenes son:

- $F_{sat,anio}$ *Cloud free*: Contiene imágenes tomadas libres de nubes. Tiene cobertura del número de observaciones que están en cada cuadrícula de 30 arco segundos[7]. Esta imagen se puede utilizar para identificar zonas en donde hay pocas observaciones donde la calidad se reduce. Para algunos años hay lugares con cero observaciones.

- $F_{sat,anio}$ *Average vis.*: Las imágenes contienen el promedio de la banda de número digital sin filtro alguno. El rango de valores está entre 0 y 63. Los lugares que no tienen observaciones libres de nubes se representan con el valor de 255.

- $F_{sat,anio}$ *Stable lights*: Las imágenes estables están *pre-procesadas* y contienen las luces de las ciudades, pueblos y otros lugares con iluminación persistente, incluyendo quemas de gas. Se filtran acontecimientos efímeros como incendios. Además se identifica el *ruido* y se remplaza por valor de 0. El número digital contiene valores entre 0 y 63. Los lugares que no tienen observaciones libres de nubes se representan con el valor de 255.

Para tener una idea de cómo se comparan los datos de iluminación en las distintas imágenes que tiene la NOAA, Pestaloozi en su artículo [40], realizó histogramas de las distintas versiones para el año 2006. En su artículo, se observa que la distribución de las distintas series se comporta muy similarmente. Es por esto que, además de que es la más popular en la literatura y debido a que ya se encuentran pre-procesados de la forma más adecuada con respecto a los objetivos del estudio, las imágenes que se utilizaron para este

[7] alrededor de 860 metros cuadrados.

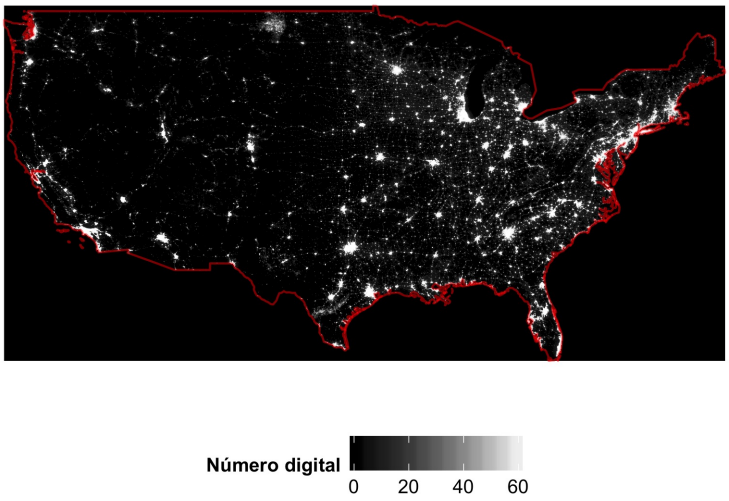
trabajo son las de iluminación nocturna estable. Se puede consultar en el Apéndice B el código para descargar y descomprimir los datos de los satélites. A partir de ahora, se referirá a los datos de la serie de iluminación nocturna espacial estable versión 4 publicada por la NOAA en [1] como los datos de iluminación nocturna espacial, ver tabla 3.1 más adelante para la lista de las imágenes que se incluyen en la serie.

3.2. Filtro de datos en las imágenes satelitales

Las imágenes de iluminación nocturna son *rasters* de alta resolución en donde cada pixel a la altura del Ecuador representa alrededor de un kilómetro cuadrado. Sin embargo, trabajar con imágenes de hasta 3 GB no es una tarea trivial y, dado que en este trabajo se trabaja con toda la serie de imágenes, a diferencia de la mayoría de los casos en la literatura (ver más detalles en la sección 2.3 del capítulo 2), trabajar directamente sobre las imágenes no fue una opción viable. Se requiere recortar las imágenes de los satélites para que se puedan trabajar en una computadora tradicional en términos de espacio de almacenamiento y de uso de memoria. Es por esto que, para fines de este trabajo, se programó un algoritmo que recorta las imágenes para extraer los datos que se buscó analizar. La idea principal del algoritmo es muy sencilla y consiste en seguir los siguientes pasos:

1. *Recortar las imágenes.* Se tomó cada imagen de la serie completa de satélites (F101992 a F182013, ver tabla 3.1 para la serie completa más adelante) en formato Raster. Para hacer esto se utilizó un polígono, shapefile, para delimitar la zona de interés. En este caso, fueron varios polígonos, se utilizó uno de cada país miembro de la OCDE para hacer un comparativo internacional, un polígono de México para extraer los datos del país y, finalmente, uno de cada Estado de México. El resultado de recortar las imágenes satelitales es un buen primer paso; sin embargo no es del todo satisfactorio. En la imagen 3.1 a continuación se muestra el resultado de recortar la imagen del satélite F18 del año 2013. En ésta, se puede notar que el recorte sólo tomó los límites de las coordenadas de Estados Unidos, formando un rectángulo, y no recortó los datos que se encuentran en la parte continental. Esto representa un problema porque es demasiada información con iluminación, $ND = 0$.

Imagen 3.1: Recorte de la imagen de iluminación nocturna de USA, F182013

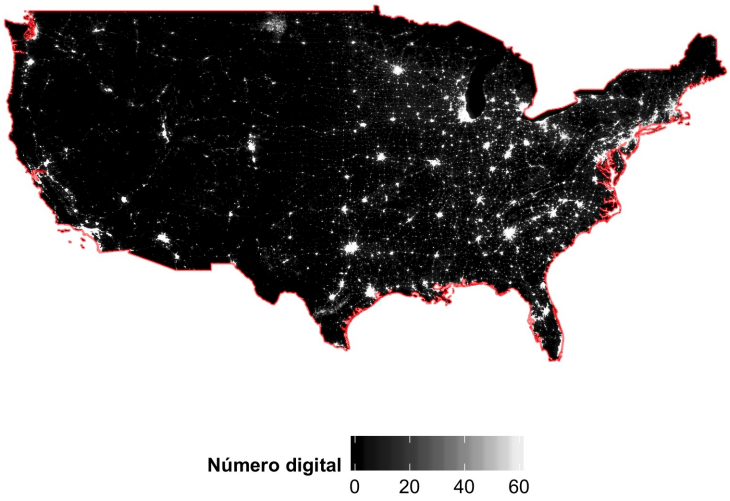


Fuente: Elaboración propia con datos de la NOAA [1] y polígonos del World Borders Dataset [44].

2. Filtrar los datos. Este procedimiento fue un poco más complejo dado que las imágenes ya recortadas contienen, en el caso de México para una sola imagen, alrededor de 8.3 millones de observaciones únicas[8]; es decir, para toda la serie de México se tienen alrededor de 290.5 millones de NDs únicos, cada uno con su respectivo par de coordenadas, satélite, año y ND correspondientes. El objetivo de este paso, fue descartar toda la información que no se encuentra dentro de cada polígono de interés, para no sesgar el análisis. Se tomó la imagen ya recortada como en la imagen anterior, se pasaron todos los datos a una tabla y se hizo una intersección utilizando las coordenadas relativas a cada pixel, para obtener los datos que se encuentran exclusivamente dentro del polígono. El resultado de este procedimiento se puede apreciar en la imagen 3.2 a continuación.

[8]En el caso de Canadá, que fue el país más grande que se analiza en este trabajo, son 52.7 millones de observaciones únicas.

Imagen 3.2: Recorte filtrado de la imagen de iluminación nocturna de USA, F182013



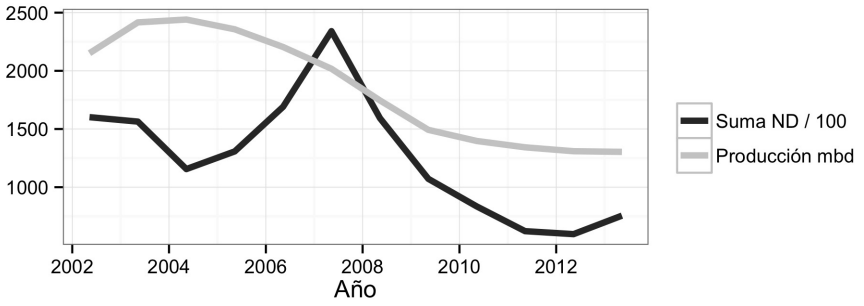
Fuente: Elaboración propia con datos de la NOAA [1] y polígonos del World Borders Dataset [44].

En el paso dos, es importante tener en consideración que hay países en los que se puede observar datos de iluminación nocturna en el mar y que posiblemente estén relacionadas con actividades pesqueras y que se podría estar perdiendo información valiosa en términos de actividad económica. Sin embargo, se decidió omitir esta consideración.

En la literatura de iluminación nocturna espacial, se suele filtrar los datos de las imágenes satelitales para remover las observaciones correspondientes al fenómeno de iluminación por quema de gases. Sin embargo, para este trabajo se tiene la hipótesis de que las quemas de gas pueden estar relacionadas con la extracción petrolera de México que puede tener un impacto considerable en la actividad económica del país dado que es históricamente un extractor petrolero importante internacionalmente. De hecho, Elvidge, et. al. en [16] en un comparativo internacional, posiciona a México en el lugar doce en términos de detección de quema de gas y en primero que más ha incrementado la quema a nivel internacional en los años 2007 a 2008. Aunque no se ha incrementado la producción petrolera de la misma manera que la quema de gases según los autores, se piensa que esto podría explicarse por el declive de algunos yacimientos petrolíferos importantes, como se puede ver en la imagen 3.3 a continuación en donde se muestra la caída en la producción del yacimiento de Canta-

rell, contrastado con la suma de la iluminación nocturna satelital correspondiente al área identificada por la NOAA como zona de quema de gas en México.

Imagen 3.3: Declive de la producción petrolera de Cantarell



Fuente: Elaboración propia con datos de la NOAA [1] y datos de PEMEX [38].

De hecho, la quema de gases identificados por la NOAA corresponde justo a la zona de extracción petrolera de Cantarell y que, como se encuentra en el mar principalmente, el algoritmo para identificar los pixeles que se vio en la sección 3.2 no los identificó con ningún Municipio porque se encuentran en el mar y los excluyó. Es por esto que, a diferencia la literatura, para este trabajo se decidió agregar los datos de la iluminación nocturna satelital correspondiente a la quema de gases que identificó la NOAA. Para ser consistente con los demás datos, se decidió agregar estos a un Municipio ficticio que se denominó *Cantarell* y que corresponde al Estado de Campeche por que es un Estado atípico en términos de PIB Estatal per cápita debido a que una gran parte de los ingresos petroleros provienen del yacimiento petrolero de Cantarell.

3.3. Intercalibración

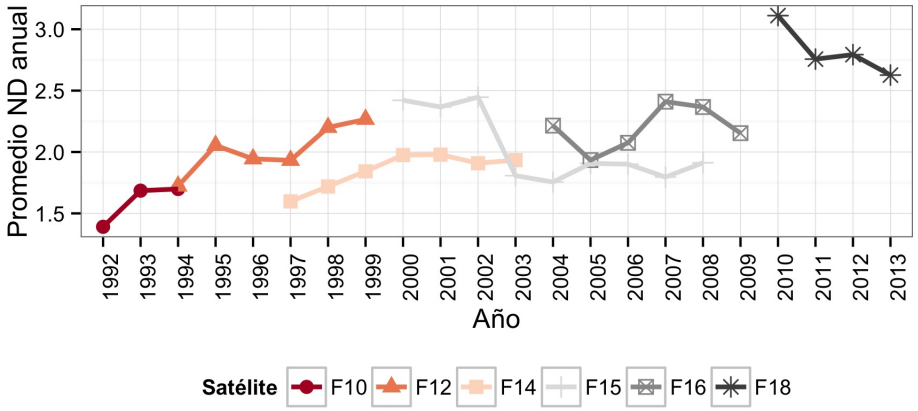
Una vez que se recortan las imágenes, como se puede notar de la tabla 3.1 a continuación, hay años para los que se cuenta con información de dos satélites y por lo tanto en este trabajo se le dio mucho interés en realizar calibración para poder hacer comparables los datos entre los distintos años para un mismo satélite y los datos para un mismo año en distintos satélites.

Tabla 3.1: Satétiles por año

Año	Satélites	Año	Satélites	Año	Satélites	Año	Satélites
1992	F10	1998	F12 F14	2004	F15 F16	2010	F18
1993	F10	1999	F12 F14	2005	F15 F16	2011	F18
1994	F10 F12	2000	F14 F15	2006	F15 F16	2012	F18
1995	F12	2001	F14 F15	2007	F15 F16	2013	F18
1996	F12	2002	F14 F15	2008	F15 F16		
1997	F12 F14	2003	F14 F15	2009	F16		

En la imagen 3.4 a continuación se puede apreciar que no es tan similar la información de distintos satélites para un mismo año y que, incluso, para un mismo satélite hay una varianza significativa, en especial en el satélite F15 se puede apreciar una caída sospechosa en el 2002. Esto porque cada sensor en los satélites tiene diferentes configuraciones y, aun en el mismo satélite, los sensores se deterioran con el tiempo así que hay que ser precavidos al hacer estas comparaciones.

Imagen 3.4: Serie de iluminación nocturna en México por satélite



Fuente: Elaboración propia con datos de la NOAA [1]

Dado que los sensores de los satélites OLS no tienen una calibración que haga comparable los distintos satélites para mismos años, se deben de calibrar con algún método

empírico. Como se mencionó en la revisión de literatura (ver capítulo 2 sección 2.3), los únicos autores que hicieron algo para tratar este problema son Liu, et. al. [33], Li, et. al. [32], Wu, et. al. [50], Elvidge, et. al. [16] y Zhao et. al. [51]. En todos se sigue un método de calibración de región invariante elaborado por primera vez por Elvidge, et. al. en [16]. Desafortunadamente, en Liu, et. al. [33], se propone un método de calibración que aplicó sólo para China hasta los satélites del 2008; en Li, et. al. [32], también para China, tomaron sólo los datos para los años en los que se cuenta dos satélites e hicieron una regresión con la cual descartó los *outliers* como cambios sospechosos en los datos y se quedaron con los demás datos para el ajuste. Wu, et. al. en [50] también utilizaron el método de la región invariante, al igual que Elvidge, et. al. [16]. Ambos presentaron los resultados explícitos para poder replicar la calibración la realizaron para todo el mundo y utilizaron toda la serie de satélites versión 4. Es por esto que se decidió replicar el método de región invariante aplicado para México, tomando una región nacional para el ajuste y así poder compararlo con la calibración que propusieron estos autores. Un artículo adicional, el Henderson, et. al. [24] quita el problema de la calibración estimando con efectos fijos de cada satélite en cada año. En particular todos los anteriores se basaron y citaron al artículo de Elvidge, et. al. [16].

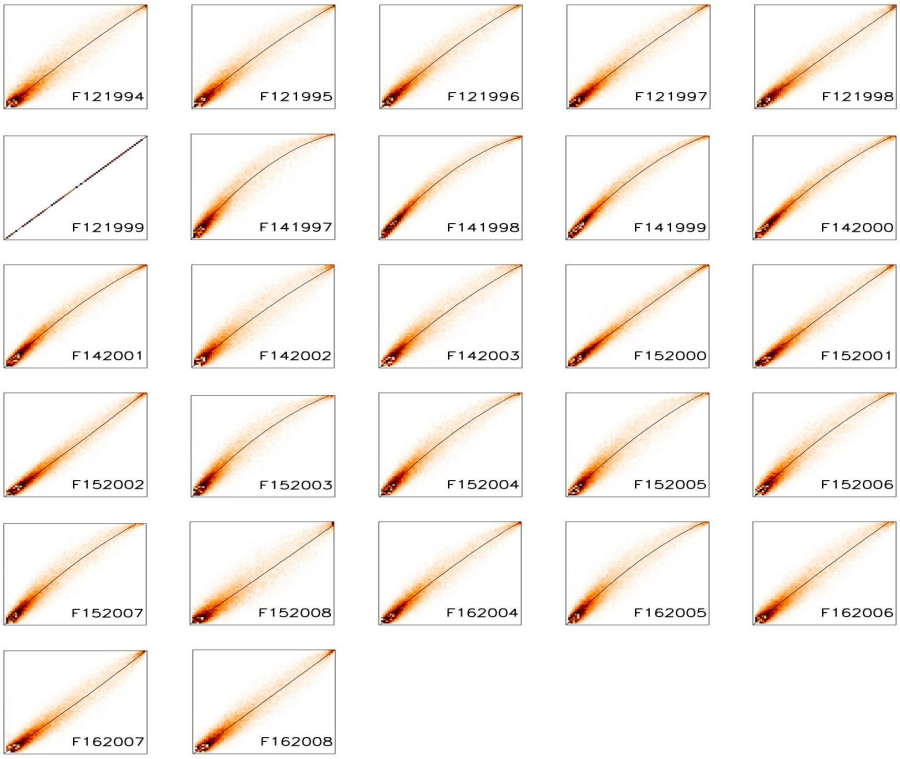
3.3.1. Intercalibración de Elvidge

Más específicamente, Elvidge, et. al. en [16] proponen hacer una intercalibración satelital por medio de un proceso empírico en el que, con base en la idea de que la iluminación en ciertas áreas densamente pobladas, como ciudades y pueblos, cambia muy poco en el tiempo, normalizan los datos de iluminación tomando un satélite de referencia y ajustando un polinomio de segundo orden a los datos de los demás satélites contra los datos del satélite tomado como referencia con el objetivo de que todos estén en el mismo rango que el satélite que se tomó como referencia.

Para realizar la calibración; primero, los autores decidieron tomar un año de mucha iluminación nocturna como referencia, en su caso los datos del satélite F121999. Con este satélite, se enfocaron en buscar alguna zona que tuviera un amplio rango de niveles de número digital y que tuviera mínima varianza en la serie ínter-satelital. En el siguiente paso de este proceso se encontraron que hay *clusters* (aglomeraciones) de muy alta iluminación nocturna, incluyendo los datos con $ND = 63$, y de muy baja saturación, lo cual representa

un reto al momento de calibrar porque es importante tener una muestra que tenga observaciones para todos los niveles de iluminación para poder tener un rango amplio de números digitales para comparar. Es decir, se fijaron en todos los pixeles de cada año comparados contra los del año de referencia y las visualizaron en una gráfica de dispersión con la idea de encontrar el de mínima varianza intertemporal. Se puede observar en la imagen 3.5 la dispersión para la ciudad que eligieron. En su estudio, los autores encontraron que la ciudad de Sicilia en Italia era la que presentaba mejores datos con estas características. El objetivo de la visualización fue verificar que se tuviera una dispersión razonable de números digitales por satélite, analizar el comportamiento de cada ND contra los demás satélite y poder identificar los datos atípicos (*outliers*).

Imagen 3.5: Comparación ND intersatelital de Sicilia, Elvidge.



Fuente: Imagen del artículo de Elvidge et. al. [16].

Una vez que encontraron una zona que consideraron óptima para hacer la calibración,

ajustaron la siguiente regresión con base en el comportamiento que vieron en la imagen de dispersión. Los resultados de los coeficientes se pueden consultar en la tabla 3.2 más adelante.

$$ND_{calibra} = C_0 + C_1 ND + C_2 ND^2 \tag{3.1}$$

Tabla 3.2: Coeficientes de intercalibración, Elvidge

Satélite	C_0	C_1	C_2	Satélite	C_0	C_1	C_2
F121994	0.1651	1.1244	-0.0018	F152001	-0.4365	1.0850	-0.0009
F121995	0.4103	1.2116	-0.0035	F152002	-0.2173	0.9715	0.0008
F121996	0.2228	1.2700	-0.0040	F152003	-0.0000	1.5238	-0.0079
F121997	-0.0008	1.1651	-0.0023	F152004	-0.3657	1.3772	-0.0056
F121998	0.1535	1.0451	-0.0009	F152005	-0.6201	1.3504	-0.0049
F121999	0.0000	1.0000	0.0000	F152006	-0.6005	1.3551	-0.0049
F141997	0.0291	1.6568	-0.0103	F152007	-0.1615	1.3960	-0.0054
F141998	0.1831	1.5980	-0.0096	F152008	0.5031	0.9370	0.0004
F141999	-0.1674	1.5116	-0.0078	F162004	-0.4436	1.2081	-0.0030
F142000	0.1061	1.3877	-0.0059	F162005	-0.2375	1.4249	-0.0063
F142001	-0.2595	1.3467	-0.0053	F162006	0.0287	1.1338	-0.0013
F142002	0.4486	1.1983	-0.0035	F162007	0.3210	0.9216	0.0013
F142003	-0.2768	1.2838	-0.0044	F162008	-0.1203	1.0155	-0.0001
F152000	0.1029	1.0845	-0.0010				

Fuente: Datos tomados de Elvidge et. al. [16].

Sin embargo, para fines de este trabajo, los resultados de intercalibración que proponen Elvidge et. al. tienen varios problemas. Para empezar, la calibración que ellos hicieron no incluye todos los satélites disponibles, lo cual es un problema porque los últimos datos que consideran ya son bastante obsoletos (2008). Otra observación menor pero importante, es que al incluir el coeficiente C_0 se le permite a la regresión asignar valores negativos. Los coeficientes negativos incluyen 13 satélites como se puede apreciar en la tabla 3.2 de coeficientes ajustados. Más aun, al momento de ajustar, los números digitales negativos representan un porcentaje muy alto de los datos, cercano al 75 % para los años con se cuenta con coeficientes $C_0 < 0$, se puede ver en la tabla 3.3. Esto es un problema porque

más adelante en el paso de estimación del PIB, se pueden dar casos en los que se tiene ND calibrado negativo para los valores en los que el ND vale cero y esto puede representar un problema de modelado y los autores no mencionan nada al respecto.

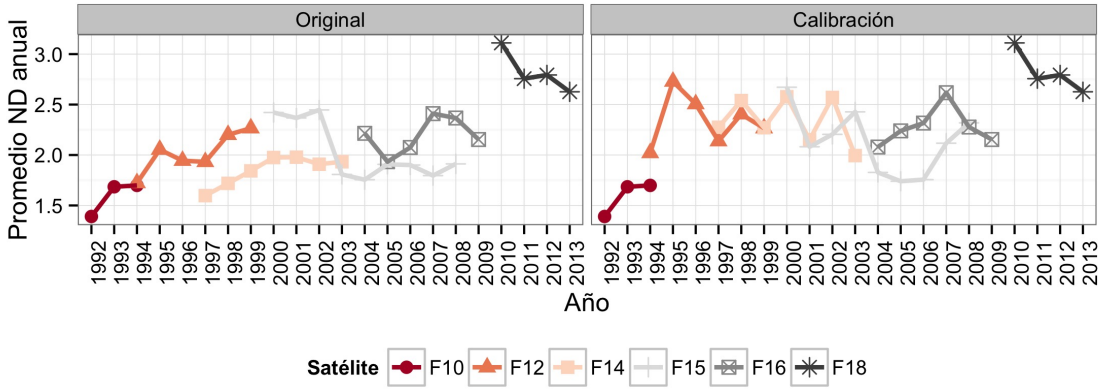
Tabla 3.3: Porcentaje de Números Digitales negativos en México, calibración Elvidge

Satélite	% $ND_{cal} < 0$	Satélite	% $ND_{cal} < 0$
F121997	81.3	F152005	73.6
F141999	77.7	F152006	73.8
F142001	77.0	F152007	75.9
F142003	77.9	F162004	74.5
F152001	76.2	F162005	75.3
F152002	75.1	F162008	74.9
F152004	75.3		

Fuente: Elaboración propia.

A pesar de las consideraciones anteriores, en este trabajo se realizó la calibración que hicieron Elvidge, et. al. en [16] para poder realmente comprobar el ajuste. El procedimiento fue simple, se tomaron los coeficientes que publicaron en su artículo y se calibraron todos los números digitales de la serie completa para México, pixel por pixel, y en los casos en los que no se tiene un coeficiente (1992, 1993, 2009 a 2013) se dejó el dato original. Los resultados de esta calibración se pueden ver en la imagen 3.6 a continuación, en donde se muestra el promedio de la serie en México para cada uno de los satélites antes de la calibración y la misma pero con los datos calibrados según el ajuste que propusieron los autores. De la imagen se puede apreciar que los coeficientes de la calibración que estimaron no parecen ajustar tan bien para México. Esto probablemente se deba a que se fijaron en un área muy específica que al parecer no es tan generalizable como ellos esperaban. Más técnicamente, esto es un caso típico de sobre ajuste en el modelado.

Imagen 3.6: Serie de iluminación nocturna en México por satélite, calibración Elvidge



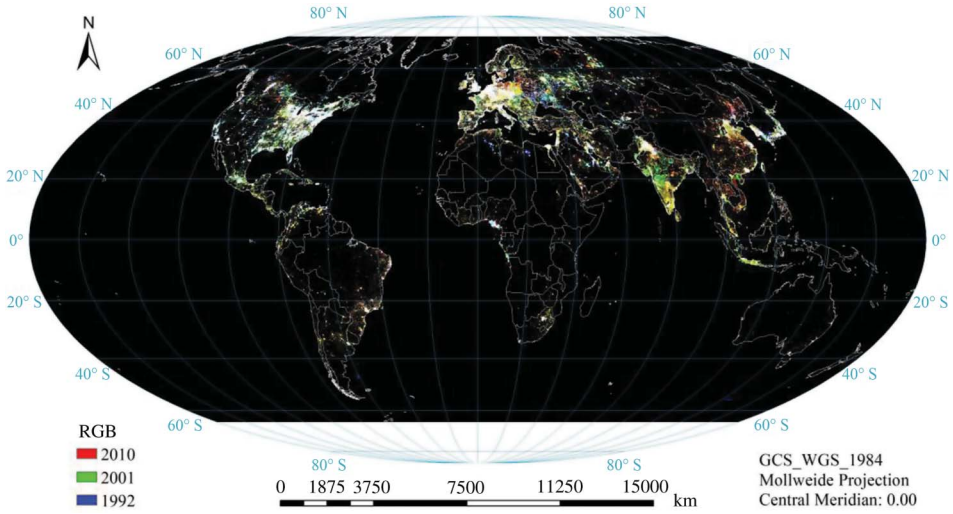
Fuente: Elaboración propia con datos de la NOAA [1] y coeficientes de Elvidge et. al. [16]

3.3.2. Intercalibración de Wu

Considerando que Chen, et. al. en [6] argumentan que los resultados del modelado con el método de Elvidge es un poco menos preciso, se optó por seguir la metodología de Wu. En el artículo de Wu, et. al. [50] hicieron una calibración similar a la que hicieron Elvidge, et. al.. En este caso, para hacer el ajuste utilizaron la serie de iluminación nocturna de la NOAA con los satélites F101992 al F182010 aunque sospechosamente dejaron fuera el satélite F152008, que de hecho es el que más varianza tiene en el rango que consideraron (ver más adelante en la imagen 3.8), sin comentar nada al respecto. Para la calibración, siguiendo la idea de Elvidge, utilizaron el método de la región invariante. Es decir, se fijaron primero en la iluminación correspondiente a tres años, 1992, 2001 y 2010, sin especificar de qué satélite se trataba en el caso del 2001, que como se ve en la tabla 3.1, la serie contiene datos de los satélites F14 y F15. Con estos tres años, se fijaron en regiones en donde el número digital no cambiara demasiado con los años pero que tuviera suficiente dispersión en términos de número digital para evitar las aglomeraciones de niveles altos y bajos que se mencionaron anteriormente. Con estos criterios y dado que se encuentran ubicados en lugares geográficos muy distintos, seleccionaron las ciudades de Mauritius, Puerto Rico, y Okinawa, Japón como las regiones invariantes. Además, para evitar el problema de la distorsión del área de las regiones que se alejan del Ecuador, decidieron proyectar la serie

de iluminación a la de Mollweide, que al igual que la de Albers, preserva área. En la imagen 3.7 se puede apreciar la proyección que ellos utilizaron y las regiones en colores más blancos representan las regiones que cambiaron menos en los años seleccionados.

Imagen 3.7: Comparación del ND anual en el mundo, Wu



Fuente: Imagen del artículo de Wu, et. al. [50].

Como referencia para la intercalibración utilizaron datos del 2006 ponderados. Es decir, utilizaron los datos de una imagen satelital que generó la NOAA en con la iluminación del 2006 en la que, los datos de iluminación anual, se ponderaron por el tiempo que se mantuvieron iluminados durante el año. Por ejemplo, si para ciertas coordenadas, la iluminación captada, sin considerar las observaciones fuera de nubes, fue la mitad del año, entonces el número digital correspondiente se ponderó con un factor de 50 %. Con estos datos y las regiones invariantes, la regresión que ellos ajustaron fue la ecuación (3.2) a continuación:

$$ND_{calibra} + 1 = C_0(ND + 1)^{C_1}, \tag{3.2}$$

o lo que es equivalente

$$\ln(ND_{calibra} + 1) = \ln(C_0) + C_1 \ln(ND + 1), \tag{3.3}$$

en donde, $ND_{calibra}$ es el valor del número digital después de la calibración, C_0 y C_1 son los coeficientes del modelo. En este caso, se le sumó uno al número digital para evitar que el logaritmo sea infinito. Además, en el proceso de modelado, ellos removieron los datos donde el $ND = 0$ que representan lugares que no se encuentran iluminados. Los resultados de los coeficientes que calcularon se pueden ver en la tabla 3.4 a continuación.

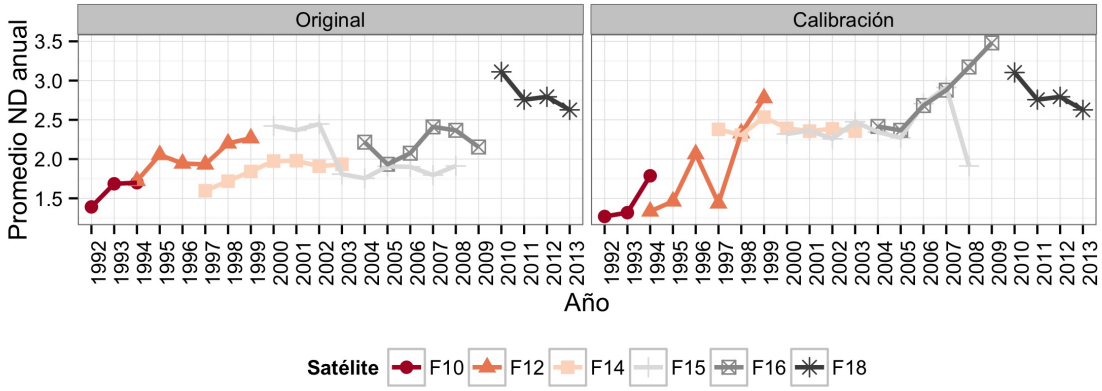
Tabla 3.4: Coeficientes de intercalibración, Wu

Satélite	C_0	C_1	Satélite	C_0	C_1	Satélite	C_0	C_1
F101992	0.8959	1.0310	F141999	1.0347	1.0904	F152006	0.9926	1.1226
F101993	0.6821	1.1181	F142000	0.9885	1.0702	F152007	1.1823	1.0850
F101994	0.9127	1.0640	F142001	0.9282	1.0928	F162004	0.7638	1.1507
F121994	0.4225	1.3025	F142002	0.9748	1.0857	F162005	0.6984	1.2292
F121995	0.3413	1.3604	F142003	0.9144	1.1062	F162006	0.9028	1.1306
F121996	0.9247	1.0576	F152000	0.8028	1.0855	F162007	0.8864	1.1112
F121997	0.3912	1.3182	F152001	0.8678	1.0646	F162008	0.9971	1.0977
F121998	0.9734	1.0312	F152002	0.7706	1.0920	F162009	1.4637	0.9858
F121999	1.2743	0.9539	F152003	0.9852	1.1141	F182010	0.8114	1.0849
F141997	1.3041	0.9986	F152004	0.8640	1.1671			
F141998	0.9824	1.1070	F152005	0.5918	1.2894			

Fuente: Datos tomados de Wu et. al. [50].

En este trabajo se realizó también la calibración que hicieron Wu, et. al. en [50] para poder comparar el comportamiento del ajuste para los datos en México. El procedimiento fue el mismo que en la sección anterior con los datos de Elvidge, et. al. En este caso, no se tiene un coeficiente para el satélite F152008, como se mencionó, ni tampoco de los últimos años así que se dejó el dato original. Los resultados de esta calibración se pueden ver en la imagen 3.8 a continuación, en donde se muestran el promedio de la serie en México para cada uno de los satélites antes de la calibración y la misma pero con los datos calibrados según el ajuste que propusieron Wu, et. al.. De la imagen se puede apreciar que los coeficientes de la calibración que estimaron parecen ajustar bastante mejor que los de Elvidge para México. Sin embargo, se puede notar la caída en la observación del 2008 que no está calibrada y que los datos del 2011 en adelante no se calibraron.

Imagen 3.8: Serie de iluminación nocturna en México por satélite, calibración Wu



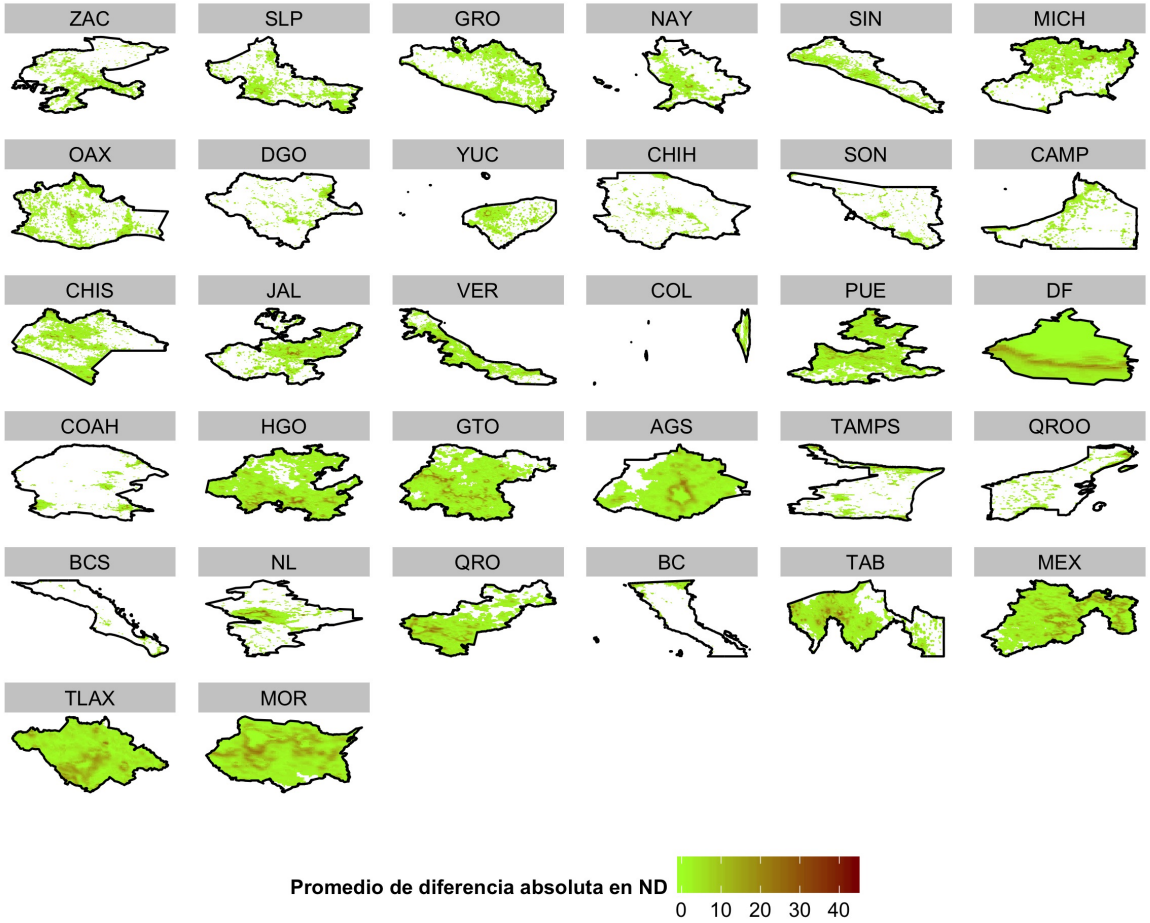
Fuente: Elaboración propia con datos de la NOAA [1] y coeficientes de Wu et. al. [50]

3.3.3. Intercalibración para México

Con base en el método de la región invariante, en esta sección se muestran los resultados de una intercalibración realizada para México con datos de México nada más. En este caso, la región invariante será un Estado del país. El satélite de referencia será el del 2010 ponderado por la NOAA, al igual que el artículo de Wu, et. al.. Para identificar la región, siguiendo la idea de comparar los tres años, se tomaron los satélites F109992, F142001 y F182010 y se calcularon diferencias en valor absoluto del 2001 con el de 1992 y del 2010 con el 2001 para cada pixel y después se promediaron estos dos y se agregó a nivel Estado. La imagen 3.9 a continuación resume este procedimiento en la que se puede observar el promedio de las dos diferencias absolutas por pixel para cada Estado. Al igual que en el artículo de Wu, et. al [50], se filtraron los datos con $ND = 0$. En la imagen se puede notar que los Estados del primer renglón como Zacatecas, San Luis Potosí, Guerrero, Nayarit, Sinaloa y Michoacán son los que presentan menor varianza comparado estos años.

Imagen 3.9: Región invariante en México

Los Estados están ordenados por renglón de menor a mayor diferencia promedio Estatal.

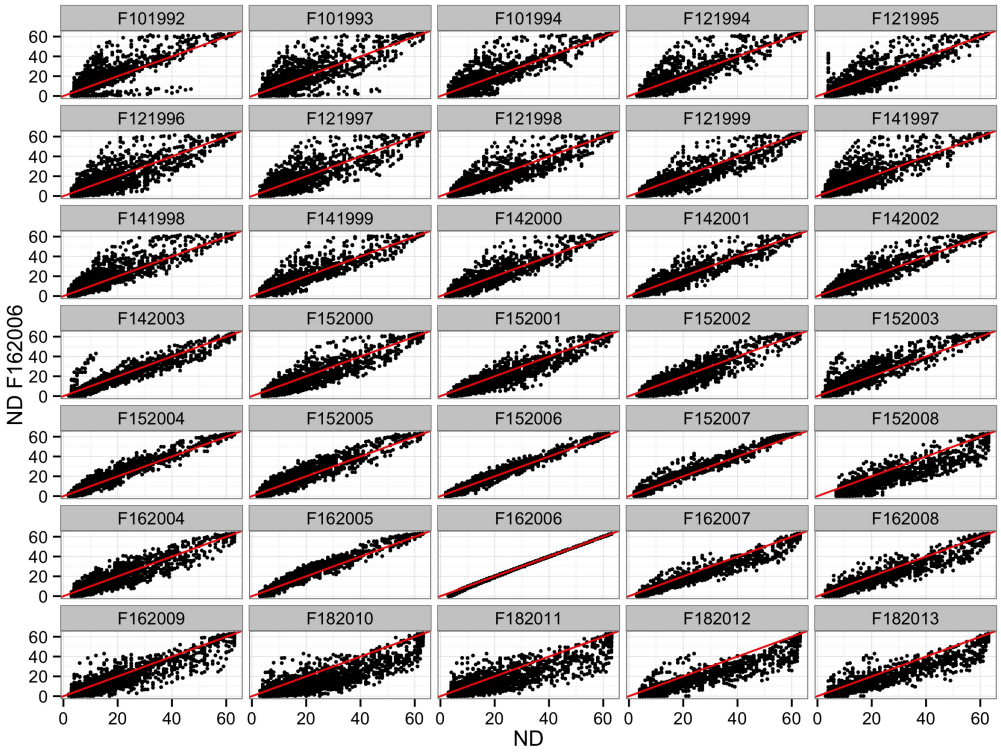


Fuente: Elaboración propia con datos de la NOAA [1] **Nota:** Las regiones en blanco corresponden a pixeles filtrados con $ND = 0$.

Por otra parte, para encontrar qué Estado tiene una dispersión de Número Digital razonable que permita hacer una calibración útil, se realizó una gráfica de dispersión como lo hicieron Elvidge et. al. en la imagen 3.5 para cada uno de los Estados seleccionados con menor diferencia temporal; es decir, Zacatecas, San Luis Potosí, Guerrero, Nayarit,

Sinaloa y Michoacán. Después de analizarlas, se eligió a Nayarit como la región invariante. Se puede ver la imagen de la dispersión de Nayarit en la imagen 3.10 a continuación. Se puede apreciar que la dispersión de ND cubre todos los niveles del más bajo al más alto y que los años que no se incluyeron en los artículos anteriores son los que presentan mayor variabilidad.

Imagen 3.10: Dispersión de ND contra ND de calibración, Estado de Nayarit



Fuente: Elaboración propia con datos de la NOAA [1]

El ajuste se realizó con la metodología de Wu, et. al. por medio de la siguiente ecuación (3.4):

$$\ln(ND_{calibra} + 1) = \alpha + \beta \ln(ND + 1), \tag{3.4}$$

en donde, $ND_{calibra}$ es el valor del número digital después de la calibración, α y β son los coeficientes del modelo. Al igual que antes, se le sumó uno al número digital para evitar que

el logaritmo sea infinito y se filtraron los valores $ND = 0$. Los resultados de los coeficientes que calcularon se pueden ver en la tabla 3.5 a continuación.

Tabla 3.5: Coeficientes de la calibración satelital intertemporal

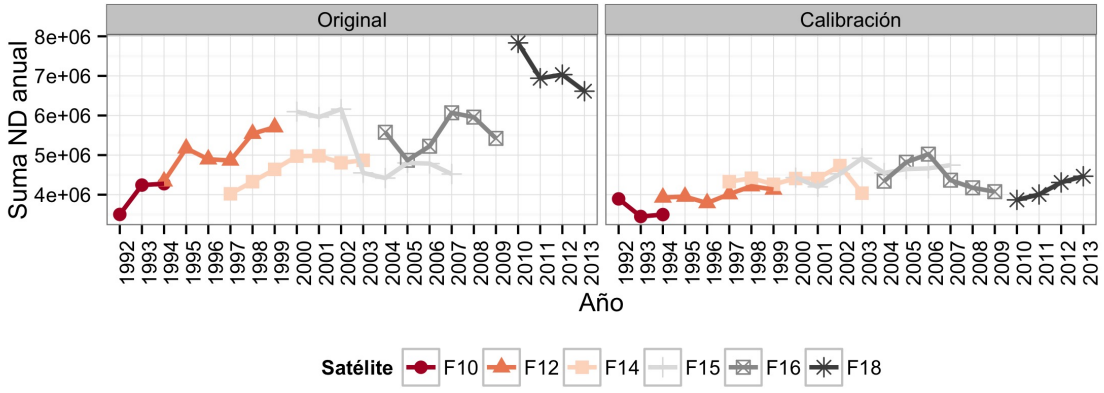
Satélite	α	β	Satélite	α	β	Satélite	α	β
F101992	0.194	1.562	F142000	0.192	1.532	F162004	0.168	1.537
F101993	0.180	1.524	F142001	0.216	1.485	F162005	0.210	1.539
F101994	0.157	1.573	F142002	0.237	1.484	F162006	0.174	1.576
F121994	0.155	1.567	F142003	0.212	1.469	F162007	0.145	1.550
F121995	0.113	1.643	F152000	0.104	1.656	F162008	0.174	1.484
F121996	0.177	1.508	F152001	0.123	1.597	F162009	0.232	1.400
F121997	0.202	1.472	F152002	0.117	1.628	F182010	0.130	1.464
F121998	0.136	1.576	F152003	0.230	1.537	F182011	0.167	1.428
F121999	0.125	1.582	F152004	0.250	1.498	F182012	0.124	1.530
F141997	0.250	1.494	F152005	0.264	1.461	F182013	0.153	1.488
F141998	0.261	1.471	F152006	0.276	1.445			
F141999	0.220	1.493	F152007	0.264	1.479			

Fuente: Elaboración propia.

Nota: Todos los coeficientes son significativos con un valor $p < 2^{-16}$. Ver detalles en el anexo D.

Finalmente, se puede apreciar de la imagen 3.11 a continuación el comportamiento de la serie ajustada con la calibración tomando a Nayarit como región invariante.

Imagen 3.11: Serie de iluminación nocturna en México por satélite, calibración propia



Fuente: Elaboración propia con datos de la NOAA [1]

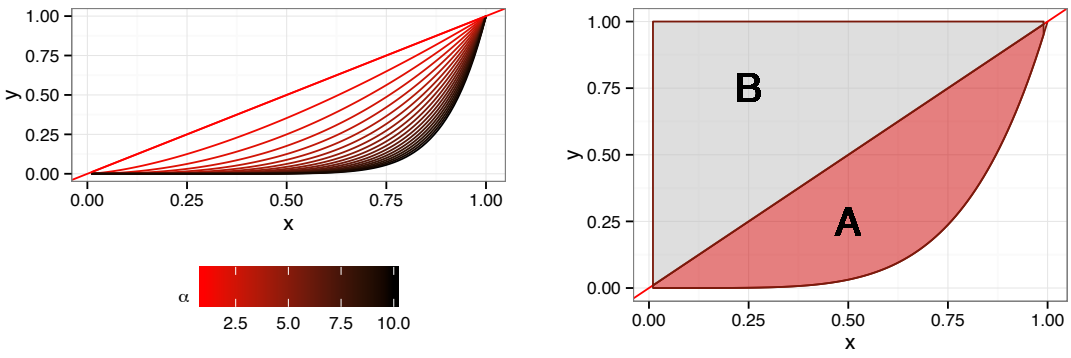
El último paso para terminar el pre-procesamiento fue un promedio simple de los distintos satélites para los distintos años.

3.4. Índice de Gini de iluminación nocturna espacial

Finalmente, para complementar la información que se puede obtener de los datos de iluminación nocturna espacial, en esta sección se desarrolla un índice de Gini espacial siguiendo la metodología descrita por Elvidge, et al. en [14]. Se desea hacer este índice porque es otro método para poder estudiar la distribución de las luces y su dispersión, o en términos económicos, desigualdad. La idea es que algunos países se encuentran mucho más centralizados que otros, o lo mismo aplica para los Estados en el caso de México, en términos de iluminación nocturna satelital mientras que países más democráticos presentan distribuciones más uniformes de iluminación en el territorio. En particular, se busca tener una medida explícita y empírica del desarrollo humano con sólo imágenes satelitales e información poblacional. Además, según Elvidge, et. al. [14], cada vez es más reconocido que la distribución del ingreso en la población se coorrelaciona con la felicidad de la sociedad. Sin embargo, hacer mediciones para este tipo de análisis suele ser complejo así que la iluminación satelital es una opción viable.

Para comparar el nivel de desigualdad se puede utilizar un índice de Gini que no es más que una medida de dispersión relativa que, comúnmente, se utiliza en economía para medir las diferencias entre la distribución del ingreso en la población. La manera más intuitiva de entender qué es un índice de Gini es por medio de su curva de Lorenz, en la que se muestra la distribución acumulada de la función que desea medir con respecto a la función acumulada de la población. En el caso original, se trató de la del ingreso y población. La interpretación es que para cada punto del eje x, en nuestro caso la población acumulada al nivel x_0, le corresponde y_0 porcentaje de la variable y acumulada, en nuestro caso, iluminación nocturna. En el caso de perfecta igualdad, la curva de Lorenz representa una línea recta que significa que para cada porcentaje de la población le corresponde exactamente el mismo porcentaje de iluminación. En el caso extremo contrario, la curva se ve como una letra L reflejada en un espejo (⅃). Significa que todo el ingreso se acumula hasta el último individuo de la población. La imagen 3.12 a continuación en la derecha representa el cálculo del índice, que se detalla más adelante en la ecuación 3.5, y que es equivalente a dividir el área de A entre el área de B.

Imagen 3.12: Índice de Gini



Fuente: Elaboración propia.

Es decir, el índice de Gini se calcula como el área que hay debajo de la línea de perfecta equidad y la curva de Lorenz dividido entre el área que hay en la curva de perfecta igualdad. Por ejemplo, si se tiene que la función de la curva de Lorenz se describe como $f(x) = x^{\alpha}$ con $\alpha \geq 1$ para toda $x \in [0, 1]$, la imagen 3.12 a la izquierda muestra cómo se ve una curva

de $f(x)$ para distintos valores de α. Más específicamente, el coeficiente de gini G se calcula como

$$G = \frac{\int_0^1 (t - t^\alpha)\, \mathrm{d}t}{\int_0^1 t\, \mathrm{d}t} = \frac{\frac{1}{2} - (\alpha + 1)^{-1}}{\frac{1}{2}}, \tag{3.5}$$

en donde x representaría la distribución acumulada de la población y $f(x)$ la distribución acumulada de la iluminación nocturna satelital. Con esto, si el coeficiente de gini es uno, entonces el denominador es igual al numerador o en palabras, hay perfecta desigualdad, la curva es una $\mathbf{J}$. Este caso se da cuando $\alpha \to \infty$. Lo contrario es cuando $\alpha \to 1$. Además al ser un índice se puede comparar entre distintos países perfectamente.

Con esto en mente, para este trabajo se calculó un índice de gini de desigualdad de iluminación nocturna satelital para el caso de México y para cada uno de los Estados de la República (ver detalles del resultado del índice en el capítulo 4) que se puede interpretar como un índice de desarrollo.

Para generar el índice de gini de iluminación nocturna satelital, se necesitó una serie de tiempo de población municipal que no está disponible en la contabilidad nacional. Es por esto que, para este trabajo, se tomó, por una parte, la población correspondiente a cada municipio en el país con datos del Censo de Población y Vivienda del INEGI [29] 2010; por otra, la serie de tiempo, desde 1992 hasta el 2013, de población correspondiente a cada Estado en México con datos del Consejo Nacional de Población (CONAPO) [9] y se tomó el año en el que coinciden los datos, 2010, para calcular el porcentaje que representa la población de cada municipio en cada Estado y con este porcentaje, se hizo el supuesto que la estructura en los municipios se mantiene relativamente constante, y se proyectó, con toda la serie completa, la población para cada municipio en cada año desde 1992 hasta 2013. La ventaja de hacer este supuesto es que la estructura de crecimiento poblacional se mantiene dado que está implícita en los datos a nivel estatal y; sin embargo, aunque no sea muy útil para analizar la movilidad social entre municipios, no se pierde toda la estructura dado que también hay movilidad entre Estados. Con esto ya fue posible hacer el cálculo de la distribución acumulada en cada Estado sumando por Municipio. En cuanto a los datos de iluminación nocturna satelital, dado que ya se indexaron todos los pixeles correspondientes a México con clave de municipio y coordenadas geográficas, fue simple calcular la distribución acumulada de la iluminación nocturna en cada región. Con esto fue suficiente para poder hacer el cálculo del índice de Gini y tener una buena medida de desigualdad de iluminación en México (ver resultados en el capítulo 4).

Capítulo 4.
Iluminación nocturna en México

En este capítulo se realiza un análisis descriptivo del comportamiento de la serie de iluminación nocturna visible del espacio para México. Para empezar, se presenta una comparación internacional con países miembros de la OCDE. Más adelante, se analiza el comportamiento de la serie en el tiempo e identifican las principales diferencias en las Entidades Federativas de México. A lo largo de este capítulo se utilizan algunas variables de estadística nacional para contrastarlas con los datos de iluminación, entre estas se incluyen el Producto Interno Bruto, la Población total y la densidad poblacional. En particular, la sección 4.1 describe la composición del Número Digital para ubicarlo en un contexto internacional, la sección 4.2 presenta la composición del ND en los distintos Estados de la República y su evolución durante las fechas que se analizan en este trabajo y se ilustra su relación con algunas variables de interés. Finalmente, en la sección 4.3 se presentan los resultados del índice de Gini de iluminación nocturna satelital para las Entidades Federativas.

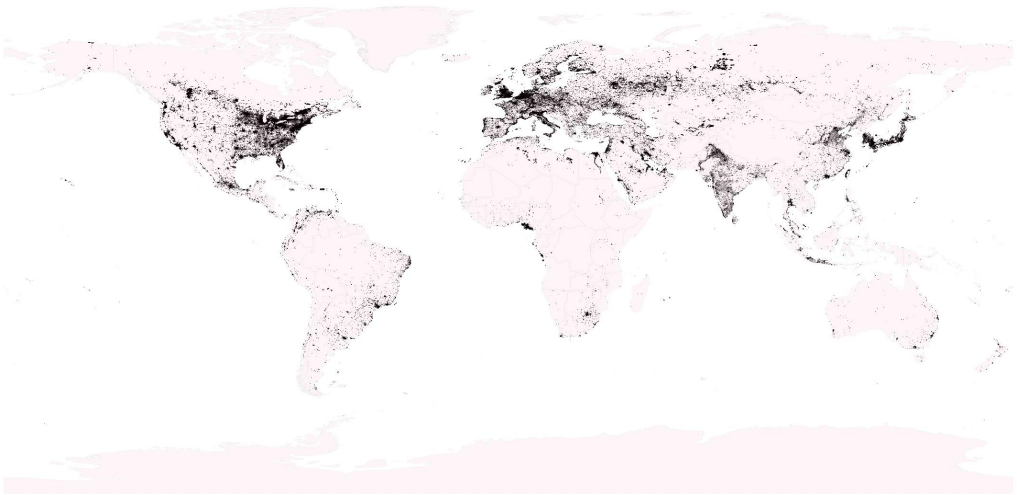
4.1. Comparativo internacional de la iluminación nocturna satelital

Para poder entender cómo se distingue México en el mundo, en esta sección se decidió comparar las composiciones del Número Digital anual a nivel nacional de algunos de los países en vías de desarrollo y desarrollados con los países de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) con el objetivo de hacer una comparación diferente

de la que se puede encontrar en la literatura (ver detalles en el capítulo 2), que compara todo tipo de países, desde subdesarrollados hasta países desarrollados. Es por esto que para hacer la comparación se consideró sólo la parte de las imágenes correspondiente a las coordenadas de México y los países de la OCDE para poder comparar el comportamiento del Número Digital a nivel nacional en cada país miembro.

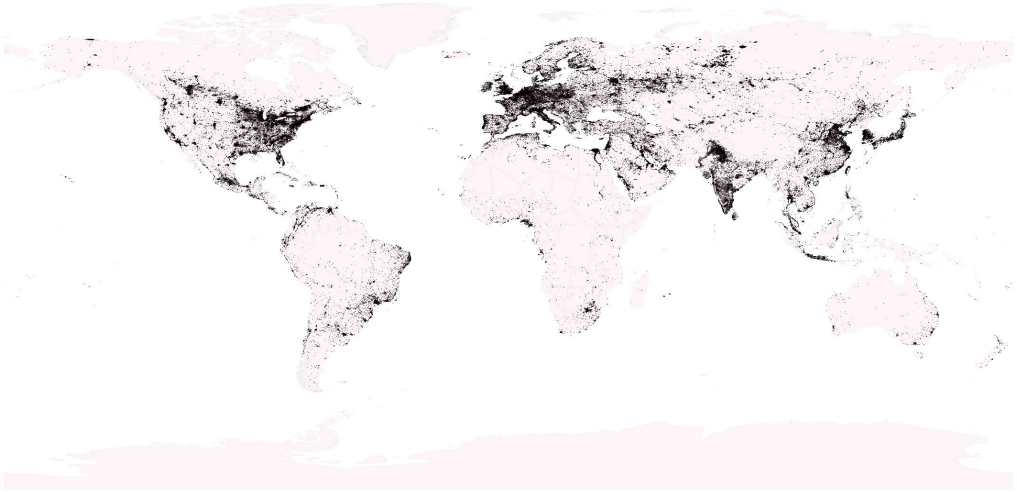
Sin embargo, antes que esto y para tener una perspectiva global del crecimiento de la iluminación nocturna satelital en el tiempo, en las imágenes del mundo 4.1 y 4.2 a continuación se puede ver el negativo de la iluminación mundial en 1992 y 2013, respectivamente, tal como se descargan junto con un mapa del mundo[1] en el fondo para facilitar la visualización. En las imágenes, los colores más obscuros corresponden a mayor iluminación nocturna. Se puede notar que ha habido un crecimiento importante y evidente sobre todo en países de Europa, Asia y América Latina.

Imagen 4.1: Negativo de la iluminación nocturna en 1992



[1]El archivo de polígonos con los países del mundo se puede ver en [44].

Imagen 4.2: Negativo de la iluminación nocturna en 2013

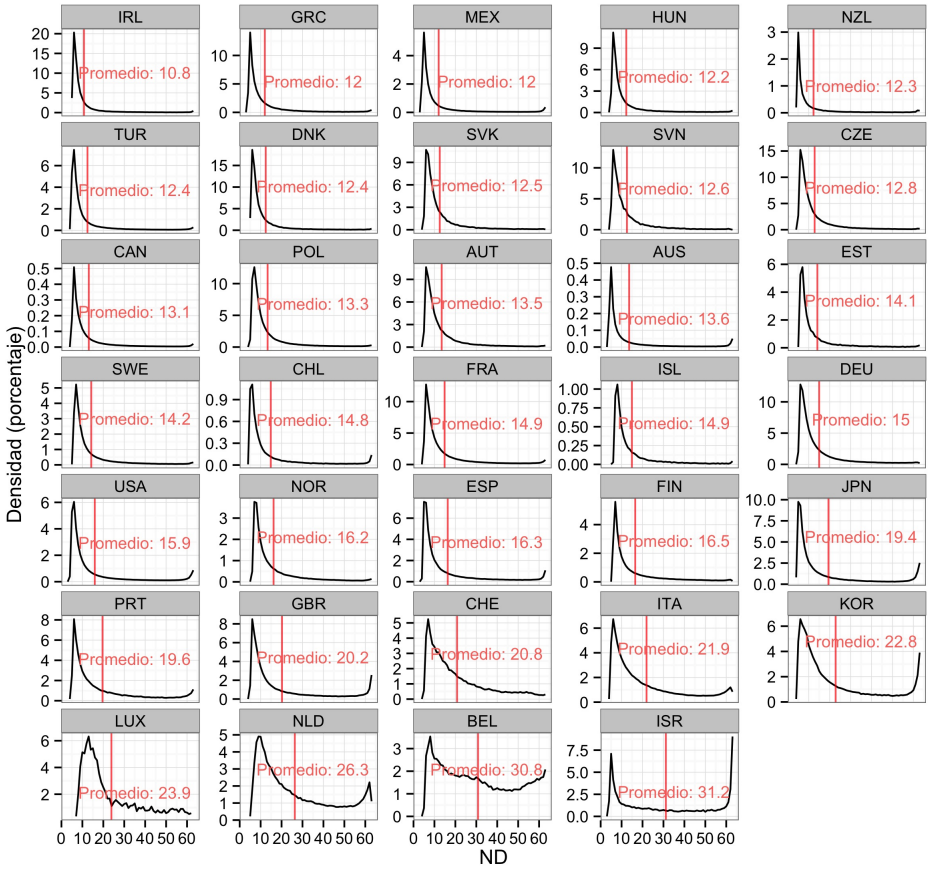


Para regresar a la comparación que se quiere presentar, se puede ver que hay diferencias significativas entre los países en desarrollo contra países desarrollados y lo mismo con los subdesarrollados, como es en el caso de los países del centro de África comparados con los de América Latina y los anteriores contra los de Europa, Estados Unidos y Canadá. Para facilitar esta comparación, se puede ver la imagen 4.3 a continuación en donde se compara la distribución de la iluminación nocturna satelital medida con el número digital para los treinta y cuatro países miembros de la OCDE. La imagen está ordenada con respecto a la media de número digital para hacer un contraste más objetivo. En particular, de la imagen, se puede apreciar que México (MEX) ocupa el penúltimo lugar, empatado con Grecia (GRC), en términos del promedio de $ND > 0$ sólo por arriba de Irlanda (IRL) con promedios de 12 (México y Grecia) y de 11 (Irlanda). Además se puede notar que la distribución en los países es radicalmente diferente. Por ejemplo México, es un país extremadamente más obscuro en promedio pero además, en la cola más alta de la distribución (en niveles de ND cercanos a 63), México presenta un alza que no se puede notar en otros países similares en área geográfica como lo son Australia (AUS) y Francia (FRA) aunque tienen un promedio más alto de $ND > 0$ con 14 y 15 respectivamente. Esto quiere decir que en México la desigualdad de la iluminación es incluso más marcada que en

los miembros de la OCDE. Para contrastar ésta diferencia se puede ver el índice de Gini espacial ($Gini_E$) de la OCDE en la tabla 4.1, en donde se observa que México tiene de los niveles más altos de desigualdad de iluminación (72 %) y se encuentra por debajo sólo de Turquía, que tiene un índice de 75 % que presenta datos muy similares de las demás variables a las de México, y Noruega (NOR) que tiene un índice de 75 % pero es el segundo país más rico en PIB per cápita.

Imagen 4.3: Distribución del ND en la OCDE, 2013

Países ordenados con el promedio de ND.



Fuente: Elaboración propia con datos de la NOAA [1].

Notas: Las distribuciones no suman cien porque no se incluyen los valores con ND= 0.
El promedio se calculó con los valores ND> 0.

Más aun, en la tabla 4.1 se puede ver la distribución del ND agrupada por niveles de intensidad de iluminación nocturna satelital en los países miembros de la OCDE junto con algunas variables relevantes para el contraste. La tabla se encuentra ordenada con respecto al PIB per cápita de cada país medido en dólares del 2013. Lo primero que salta de la tabla es que México ocupa el primer lugar en pobreza de PIB per cápita con 10 mil dólares anuales contra Luxemburgo que es el país más rico en PIB per cápita de la OCDE con 111 mil dólares y que además tiene una distribución de ND mucho más cargada a niveles altos de iluminación que los demás países.

También, se puede notar que Estados Unidos (USA), que es de los países que tienen niveles más bajos de densidad poblacional con 35 personas por kilómetro cuadrado, presenta los mayores niveles en iluminación correspondiente a los niveles más altos, cerca del 1 % de NDs se encuentran entre 62 y 63. Además Estados Unidos se posiciona sólo por debajo de países densamente poblados como los Países Bajos (NLD), Israel (ISR) y Corea (KOR) que tienen 497, 363 y 503 personas por kilómetro cuadrado respectivamente y que tienen porcentaje de NDs en los niveles más altos de iluminación nocturna (entre 62 y 63) de 1.1 %, 9 % y 3.1 % cada uno.

En cuanto a la relación que tienen el porcentaje de Números Digitales que se ubican entre 0 y 1 con el porcentaje de población urbana, de la tabla se percibe que siguen una tendencia negativa que se comporta linealmente. Es decir, países que tienen los mayores niveles en cuanto a población urbana como Bélgica (BEL), los Países Bajos (NLD) y Alemania (DEU), con 98 %, 89 % y 75 % respectivamente, presentan los menores porcentajes de obscuridad (ND entre 0 y 1), con 1 %, 2 %, y 10 % respectivamente. Se destaca que México, que tiene el 78 % de población urbana, presenta niveles altos de obscuridad (78 %). Esto puede deberse a que es un país con una superficie continental relativamente alta en comparación con los demás miembros de la OCDE.

Además, la tabla incluye el índice de Gini espacial de iluminación nocturna satelital que calcularon Elvidge, et al. en [14] con el cual se estima la dispersión de la iluminación a lo largo de la distribución espacial de la población. Este índice no parece seguir una relación evidente con el de Gini tradicional. Se explora con mayor detalle en la sección 4.3. Finalmente, en cuanto al índice de desigualdad Gini, se puede notar que países con índices más bajos como Dinamarca (DNK), Suecia (SWE) y Finlandia (FIN) con 25 %, 26 % y 28 % respectivamente tienen una distribución de ND que está concentrada en los niveles medios de iluminación (ND entre 6 y 21) con 27 %, 12 % y 15 % respectivamente.

Tabla 4.1: Distribución de ND en la OCDE

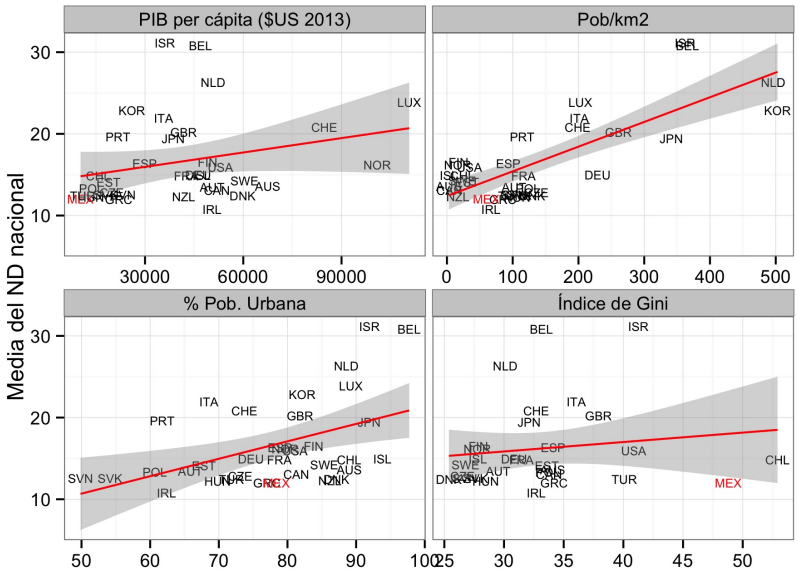
País	[0,1)	[3,5)	[6,10)	[11,21)	[21,62)	[63,64)	Prom. ND	Pob/km^2	%Pob. Urb.	Gini	PIBpc($US 2013)	Gini_E
MEX	78.03	7.76	8.22	3.15	2.61	0.36	3	60	78	49	10,307	72
TUR	65.92	5.45	18.50	5.64	4.47	0.28	4	97	72	40	10,972	75
HUN	53.01	2.76	29.82	8.83	5.65	0.27	6	107	70	28	13,485	73
POL	23.79	1.19	45.10	19.53	10.86	0.28	10	123	61	34	13,654	68
CHL	93.53	1.06	2.97	1.23	1.13	0.14	1	24	89	53	15,732	69
SVK	35.61	1.82	38.25	16.66	8.27	0.05	8	110	54	28	18,049	72
EST	69.72	0.05	18.57	6.72	5	0.18	4	31	68	34	18,877	76
CZE	11.75	2.72	51.43	22.87	11.97	0.19	11	136	73	27	19,858	68
PRT	36.05	2.66	25.24	16.40	19.43	1.12	13	114	62		21,738	64
GRC	40.89	17.46	23.22	10.91	7.60	0.41	7	85	77	34	21,966	70
SVN	26.30	3.91	41.07	19.50	9.99	0	9	102	50	27	23,295	74
KOR	4.67	5.04	28.23	26.90	32.71	3.91	22	503	82		25,977	66
ESP	50.22	7.84	20.44	10.15	10.80	1.08	8	93	79	34	29,882	71
ITA	15.16	5.06	24.69	23.40	32.29	0.84	19	201	69	36	35,686	66
ISR	26.55	8.80	12.23	11.51	32.75	9.04	23	362	92	41	36,051	71
JPN	28.42	10.56	25.91	14.06	19.27	2.53	14	341	92	32	38,634	63
GBR	33.96	1.43	28.70	16	18.15	2.56	13	261	82	38	41,781	57
NZL	91.34	3.18	3.06	1.22	1.18	0.08	1	17	86		41,824	73
FRA	23.86	3.69	39.79	18.93	13.82	0.73	11	117	79	31	42,560	63
DEU	9.66	1.99	46.71	25.11	17.41	0.21	14	229	75	31	46,251	65
BEL	0.70	0.37	14.70	22.81	61.13	2.08	31	366	98	33	46,930	59
ISL	93.91	0	3.24	1.83	1.05	0.04	1	3	94	28	47,349	76
FIN	68.52	0.01	16.08	8.05	7.67	0.07	5	18	84	28	49,151	69
IRL	28.93	3.69	50.76	10.67	5.86	0.43	8	67	62	33	50,478	62
AUT	31.21	2.10	37.65	19.31	10.29	0.21	9	102	66	30	50,511	70
NLD	1.86	0.22	19.50	31.77	47.46	1.10	26	496	89	30	50,793	70
CAN	97.74	0.18	1.30	0.47	0.32	0.02	0	4	81	34	51,964	66
USA	63.14	5.68	16.51	6.99	7.12	0.87	6	34	81	41	53,042	54
DNK	18.26	2.82	52.69	16.29	10.14	0.36	10	130	87	25	59,819	59
SWE	71.40	0.03	16.83	7.12	4.76	0.16	4	23	85	27	60,381	61
AUS	98.19	0.67	0.58	0.26	0.26	0.05	0	3	89	34	67,463	64
CHE	28.41	0.58	20.74	24.75	26.68	0.30	15	199	74	33	84,748	70
NOR	73.99	0.01	12.43	8.10	5.70	0.13	4	16	80	28	100,898	75
LUX	0.75		11.54	48.34	41.55		24	203	89		110,665	65

Fuente: Elaboración propia con datos de la NOAA [1], del Banco Mundial [36] e índice de Gini Espacial (Gini_E) de Elvidge, et. al. en [14].

Nota: No se incluye el $ND = 2$ debido a que no hay datos con este nivel. El índice de Gini_E que se calculó en este trabajo para México es de 63 %.

Por otra parte, en la imagen 4.4 más adelante se puede notar la dispersión de las variables del PIB per cápita, densidad de población, porcentaje de población urbana e índice Gini de desigualdad contra el promedio de iluminación nocturna satelital para los países miembros de la OCDE. Se destaca México en rojo para que sea fácil de identificar. De la imagen, se puede notar que hay una correlación positiva con el PIB per cápita aunque los niveles más bajos de PIB per cápita presentan más variabilidad en iluminación que los más altos. En cuanto a la densidad poblacional, se puede ver una correlación mucho más alta, los países con niveles más altos de población urbana, tienen mayor iluminación en promedio nacional. Se puede notar un comportamiento similar en cuanto a la población urbana, destacando países como Bélgica (BEL) e Israel (ISR). Finalmente, como era de esperarse, la media de iluminación nacional no parece tener correlación con el índice de desigualdad de Gini. Esto porque el índice de Gini mide la desigualdad de ingreso a través de cómo se distribuye en la población y por lo tanto es una medida de dispersión que no tiene por que estar relacionado con el promedio sino con la desviación. Para ver la relación de esta variable es más apropiado ver la tabla 4.1.

Imagen 4.4: Comparación del ND promedio en la OCDE, 2013

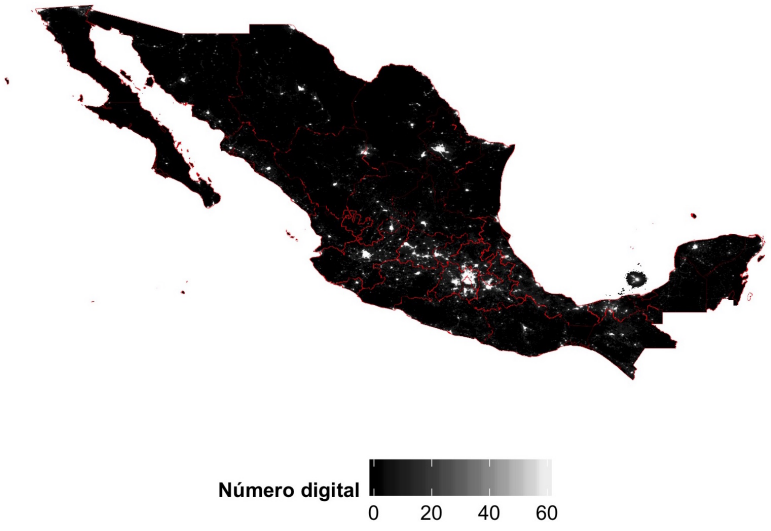


Fuente: Elaboración propia con datos de la NOAA [1] y del Banco Mundial [36].

4.2. Descripción de la iluminación nocturna en México

Una vez que se analizó México en un comparativo internacional, en esta sección se compararan las composiciones del Número Digital anual a nivel Estatal. Para empezar, se puede ver la imagen 4.5 a continuación en donde se muestran los datos de la imagen satelital de la iluminación nocturna correspondiente al 2013. Los colores más blancos corresponden a un nivel más alto de iluminación detectada por los sensores OLS de la NOAA. Se puede notar que hay una carga significativa hacia los estados del centro del país en contraste con la zona norte y las penínsulas. En particular resalta el Estado de México y el Distrito Federal como las zonas más iluminadas por un lado, y por el otro Estados muy obscuros como Guerrero y Chiapas. Además se puede notar que hay un pequeño conjunto de pixeles en el golfo de México que se encuentran aislados. Este conjunto corresponde a la zona identificada por la NOAA como de quema de gas que, como se mencionó en el capítulo 3, son los que corresponden a la zona de extracción y producción petrolera de Cantarell y que se consideran como parte de la iluminación del Estado de Campeche. En particular, en esta sección se realizaron visualizaciones como en el caso de la OCDE para facilitar las comparaciones entre los Estados.

Imagen 4.5: Iluminación nocturna satelital de México, 2013



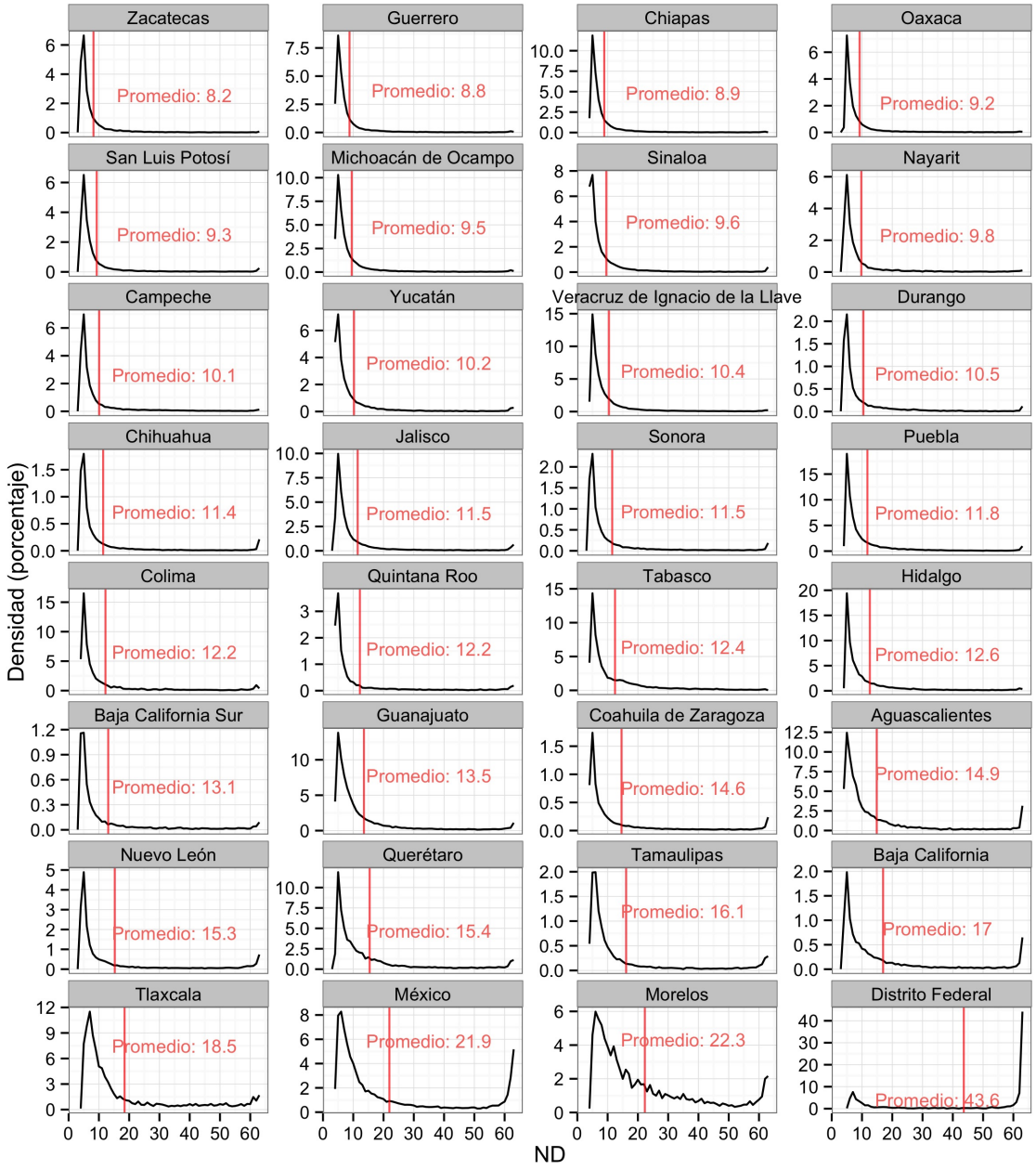
Fuente: Elaboración propia con datos de la NOAA [1].

Para hacer más explícita la comparación, se puede ver la imagen 4.6 a continuación en donde se compara la distribución de la iluminación nocturna satelital medida con el número digital para los treinta y un Estados y el Distrito Federal en el 2013. Al igual que antes, esta imagen se encuentra ordenada con respecto a la media de número digital para hacer un contraste más objetivo. Se puede apreciar que Zacatecas, Guerrero y Chiapas ocupan los últimos lugares en términos del promedio del ND sólo por debajo de Oaxaca por menos de media unidad. Por el contrario, se puede notar que los Estados de Baja California, Tlaxcala, México, Morelos y el Distrito Federal son los que tienen mayores niveles de iluminación promedio y además, tienen una densidad de Número Digital mucho más cargada a los niveles altos y con probabilidades más altas en los niveles medios.

A diferencia de la imagen de la OCDE, se puede notar que la distribución en los Estados es más similar para los lugares más bajos y similar para los lugares más altos. Por ejemplo, los estados con mayores niveles de iluminación nocturna satelital, presentan un pico en la distribución para los valores de ND más altos que los Estados con menores niveles, en promedio, no presentan de la misma manera como Guerrero o Chiapas que prácticamente tienen toda la distribución acumulada en los valores más bajos de ND. En pocas palabras, los Estados pobres en iluminación son equitativamente pobres en iluminación y los ricos son equitativamente ricos. Este fenómeno parece contradecir la idea tradicional de que México tiene niveles muy altos de desigualdad social como se vio en la tabla de la OCDE; sin embargo, esto sólo quiere decir que la desigualdad se da entre los diferentes Estados y no dentro de ellos. De hecho, se puede ver más evidentemente la desigualdad entre los Estados comparando los promedios con los de la OCDE. El país con mayor iluminación en la OCDE tiene un promedio de iluminación nocturna satelital de 31 y el más bajo 11. Mientras que en los Estados de la República, los niveles van desde 44 en el Distrito Federal hasta 8 en Zacatecas. No sólo eso, el índice de Gini tradicional es un indicador nacional así que refleja diferencias transversales a lo largo de todo el país.

Imagen 4.6: Distribución del ND en las Entidades Federativas de México, 2013

Entidades Federativas ordenadas con el promedio de ND.



Fuente: Elaboración propia con datos de la NOAA [1].

Notas: Las distribuciones no suman cien porque no se incluyen los valores con ND= 0.

El promedio se calculó con los valores ND> 0.

Una vez más, para poder hacer comparaciones explícitas, en la tabla 4.2 más adelante se puede ver la distribución del Número Digital agrupado por niveles de intensidad de iluminación nocturna. La tabla se encuentra ordenada con respecto al PIB Estatal (PIBE) per cápita medido en dólares del 2013. Para empezar, se puede notar que el Estado de Chiapas ocupa el primer lugar en pobreza de PIBE per cápita, con 4,360 dólares anuales, en contraste con Campeche que es el Estado más rico en PIBE per cápita de México, con 65,836 dólares, aunque, como se mencionó en el capítulo 3, es un Estado atípico debido que una gran parte de la extracción petrolera del país se ubica en Campeche y no es un Estado grande en población, de hecho, es el segundo menos poblado con 16 personas por kilómetro cuadrado, sólo por arriba de Baja California Sur que tiene 10. También se ve que Estados como México y Morelos, que tienen 7,205 y 8,037 dólares de PIBE per cápita respectivamente y son Estados geográficamente muy cercanos al Distrito Federal, tienen de los porcentajes más altos de iluminación nocturna entre 62 y 63 con 2.16 % y 5.18 % respectivamente. Lo anterior muestra la fuerte centralización del ingreso que se vive en el país al menos visto desde el espacio. Aguascalientes es también un Estado que resalta por tener un porcentaje relativamente alto de iluminación en la categoría más alta con 3.15 %, ocupando el tercer lugar con respecto a estos niveles de iluminación. Cabe destacar que Aguascalientes se posiciona el quinto lugar en población por kilómetro cuadrado, y el primero si se excluyen los Estados vecinos al Distrito Federal. De la tabla también resalta que Campeche no presenta un promedio tan alto en términos de iluminación satelital, incluso considerando que en el capítulo 3 se ajustaron los datos de iluminación de Campeche para incluir la iluminación correspondiente a la quema de gas marítima que se encuentra en la zona de extracción petrolera de Cantarell. También, se puede notar que el Distrito Federal es por mucho la Entidad con mayor densidad poblacional, con una población muy cercana a 6 mil habitantes por kilómetro cuadrado, es incluso mucho mayor que la de cualquier país de la OCDE, y, además de ser la segunda Entidad más rica en PIBE per cápita (23,684 dólares), tiene la mayor parte de la distribución de iluminación cargada a los niveles más altos con casi 45 % de NDs que se encuentran entre 62 y 63. Esto fortalece la hipótesis de que la iluminación está coorelacionada con el PIB per cápita.

Tabla 4.2: Distribución de ND en los Estados de la República

Estado	[0,1)	[3,5)	[6,10)	[11,21)	[21,62)	[63,64)	Prom. ND	% Pob. Urb.	Pob/km^2	PIBEpc ($US)	$Gini_E$
Chiapas	64.20	13.61	16.29	3.79	2.23	0.03	3	49	70	4,360	46
Oaxaca	80.57	7.62	8.44	2.03	1.38	0.05	2	59	42	5,044	66
Guerrero	72.64	11.16	11.96	2.62	1.66	0.06	2	58	55	5,240	49
Tlaxcala	0.95	7.85	42.10	23.29	24.86	1.70	18	82	311	5,715	38
Michoacán	63.43	13.80	15.87	4.31	2.67	0.11	3	69	76	6,510	45
Puebla	34.11	19.91	27.22	10.66	7.65	0.96	8	74	177	6,723	50
Nayarit	79.37	9.14	7.54	2.10	1.87	0.10	2	69	229	7,105	46
México	13.47	9.86	29.27	18.74	24.34	5.18	19	85	732	7,205	36
Hidalgo	31.52	19.90	26.52	12.34	9.93	0.35	9	56	135	7,336	36
Morelos	8.93	4.84	25.01	26.59	34.14	2.16	20	84	384	8,037	30
Veracruz	47.83	16.41	23.21	7.89	4.73	0.22	5	62	110	8,401	50
Zacatecas	78.53	11.50	6.83	1.90	1.24	0.07	2	61	21	8,479	49
Guanajuato	21.91	17.92	32.50	15.57	11.66	1.11	11	70	187	8,810	46
Sinaloa	69.01	14.45	10.30	3.78	2.21	0.39	3	73	51	9,025	49
Durango	92.01	3.73	2.35	1.05	0.80	0.12	1	70	14	9,047	64
Yucatán	71.09	12.33	10.18	3.78	2.47	0.28	3	86	54	9,069	57
San Luis Potosí	78.25	9.93	8.18	2.07	1.42	0.25	2	64	44	9,083	49
Chihuahua	93.22	3.27	1.96	0.73	0.65	0.21	1	86	147	9,831	65
Jalisco	62.45	13.35	14.92	4.99	3.82	0.64	4	87	99	10,426	54
Baja California	88.87	3	3.25	2.34	1.98	0.65	2	92	473	10,566	31
Colima	44.48	21.85	19.10	6.78	7.66	0.38	7	89	124	10,602	41
Tamaulipas	87.77	2.53	5.17	1.95	2.37	0.29	2	89	43	10,797	69
Aguascalientes	24.12	17.72	28.96	15.63	11.01	3.15	11	81	223	11,353	43
Quintana Roo	87.02	6.14	3.66	1.22	1.82	0.19	2	88	30	13,160	43
Baja California Sur	94.71	2.32	1.44	0.65	0.84	0.09	1	86	10	13,326	32
Sonora	90.95	4.03	2.75	1.12	1	0.19	1	87	16	13,416	72
Querétaro	37.59	13.78	22.01	14.75	11.36	1.13	10	70	166	13,515	48
Coahuila	92.83	2.55	2.23	1.04	1.17	0.24	1	91	19	14,533	53
Tabasco	37.20	18.42	21.73	13.42	9.87	0.02	8	52	94	17,089	40
Nuevo León	80.29	8.03	5.25	2.47	3.34	0.74	3	95	77	18,230	49
Distrito Federal		0.16	22.59	8.65	24.81	44.16	44	100	5,989	23,684	13
Campeche	76.10	11.30	7.59	2.48	2.52	0.13	2	75	15	65,836	36

Fuente: Elaboración propia con datos de la NOAA [1], datos del PIB Estatal del INEGI [29], datos de la superficie geográfica Google Search y datos de población de la CONAPO [9].

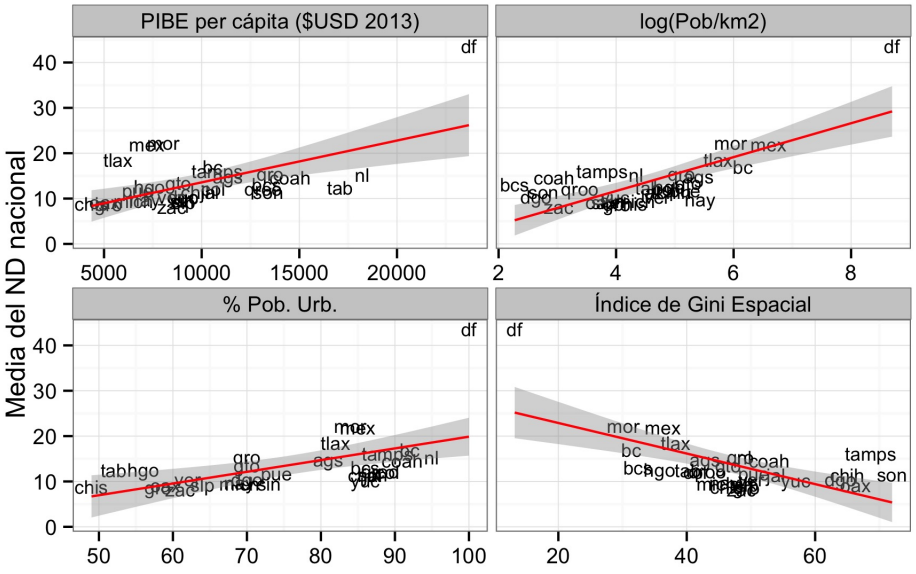
Nota: El tipo de cambio que se utilizó fue 12.25. No se incluye el $ND = 2$ debido a que no hay datos con este nivel.

Por otro lado, también de la tabla 4.2, se pueden observar los resultados del índice de Gini de iluminación nocturna para cada Entidad Federativa en el 2013. De esta columna, se comprueba que hay relativamente menos desigualdad dentro de los Estados que entre ellos como se mencionó anteriormente, prácticamente todos tienen un menor índice que el de México a nivel nacional (63 %), con la excepción de Durango (64 %), Chihuahua (65 %), Oaxaca (66 %), Tamaulipas (69 %) y Sonora(72 %), que se encuentran por arriba del promedio nacional. Además, de ésta columna, se destaca que el Distrito Federal tiene por mucho, el menor nivel de desigualdad de iluminación. De hecho, se puede notar que esta variable parece tener una correlación con el porcentaje de población urbana. Por ejemplo, Entidades como el Distrito Federal, que tiene el menor índice de desigualdad de iluminación con 13 %, Morelos (30 %), Baja California (31 %), y México (36 %) tienen los niveles más altos de población urbana, con niveles de 100 %, 84 %, 92 % y 85 % respectivamente.

Finalmente, en la imagen 4.7 más adelante se puede notar la dispersión las variables del PIBE per cápita en dólares, densidad de población, porcentaje de población urbana y el índice Gini Espacial de desigualdad en iluminación que se calculó en este trabajo, graficado contra el promedio de iluminación nocturna satelital para las Entidades Federativas. De la imagen, se puede notar que, al igual que en la OCDE, los Estados presentan una correlación positiva con el PIBE per cápita aunque los niveles más bajos de PIBE per cápita presentan más variabilidad en iluminación que los más altos con todo y que se excluye Campeche. En cuanto a la densidad poblacional, se puede ver una correlación mucho más alta, y que los Estados con niveles más altos de densidad poblacional, tienen mayor iluminación en promedio. Se puede notar un comportamiento similar en cuanto a la población urbana, destacando al Distrito Federal, el Estado de México y Morelia. Para terminar, se puede ver que la media de iluminación nacional presenta una pequeña correlación negativa con el índice de Gini de iluminación nocturna aunque posiblemente se encuentre sesgado por el Distrito Federal que tiene un promedio de Número Digital mucho más alto que los demás Estados. Sin embargo, esto no significa que haya una correlación real, de hecho, esta tendencia se explica más bien por que al momento de tomar el promedio de iluminación, la cantidad depende el área del Estado. Es decir, Estados muy grandes en área suelen tener muchos pixeles con valor $ND = 0$. Esto provoca que se tengan promedios de iluminación más pequeños en estos casos. Por otro lado, el índice de Gini de desigualdad de iluminación, al ser una medida de dispersión, depende directamente del área en cada Estado. Aquellos que tienen superficie territorial más grande suelen tener mayor superficie no iluminada y,

por tanto, índices más altos de desigualdad.

Imagen 4.7: Comparación del ND media en México, 2013



Fuente: Elaboración propia con datos de la NOAA [1] y del Banco Mundial [36].
Nota: Se excluye Campeche.

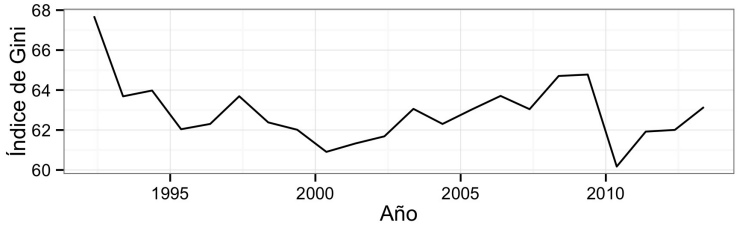
4.3. Gini de iluminación nocturna satelital, resultados

En esta sección se presentan los resultados del cálculo de la serie del índice espacial de Gini para desigualdad de iluminación nocturna satelital. Para empezar, se presentan los resultados del índice de Gini a nivel nacional para poder contrastar con los resultados obtenidos por Elvidge, et al. [14], que se vieron en la tabla 4.1 de comparativo internacional de los países miembros de la OCDE y para, finalmente, poder tener una noción de cómo se compara éste resultado con los de las Entidades Federativas de México. Para ver detalles sobre el cálculo del índice se puede ver la sección 3.4. Es importante resaltar que los datos de población que utilizaron Elvidge, et al. en [14] y los que se emplearon para el cálculo de Gini en este trabajo son distintos. Ellos utilizan una matriz de población cuyas entradas representan la población que se estima que vive en cada 10 kilómetros cuadrados

de la superficie territorial. A diferencia de ellos, para este trabajo se utilizó la población correspondiente a cada Municipio. Es decir, dado que los municipios difieren en tamaño, y en general abarcan un poco más de 10 kilómetros cuadrados de superficie, se espera que el índice sea un poco más bajo de lo que calcularon los autores debido a que se acumula más población por municipio en cada punto de las curvas de Lorenz que si se estima utilizando la matriz que ellos utilizaron. Las razones por las que se decidió utilizar estos datos en vez de la matriz de población fueron que, primero, no se pierde la estructura de relación que hay entre el área territorial y el área que se acumula de iluminación dado que en este trabajo se trabajó con los datos de iluminación nocturna por cada pixel y que, como se vio en el capítulo 3, tienen un área fija por lo el acumulado de iluminación corresponde a la misma área de superficie territorial de cada Municipio. Esto hace equivalente el cálculo del índice al que realizaron los autores sólo que a un nivel más agregado. El segundo motivo fue que, para este projecto, se pretendió utilizar datos que difieran en su fuente lo menor posible. Es decir, la prioridad fue utilizar datos del INEGI para todo lo que tienen disponible antes que cualquier otra por que, como se comentó en el capítulo 1, comparar datos que provienen de distintas fuentes puede generar sesgos en las estimaciones.

Ahora sí, en la imagen 4.8 a continuación se puede ver el resultado del cálculo del índice para México desde 1992 hasta el 2013, que son los años para los que se cuenta con datos de imágenes satelitales de iluminación nocturna. Se puede apreciar que ha habido una ligera tendencia a la baja en el índice de desigualdad que se explica debido al crecimiento de la iluminación visible desde el espacio en el mismo periodo.

Imagen 4.8: Evolución del índice de Gini de iluminación nocturna, 1992 a 2013



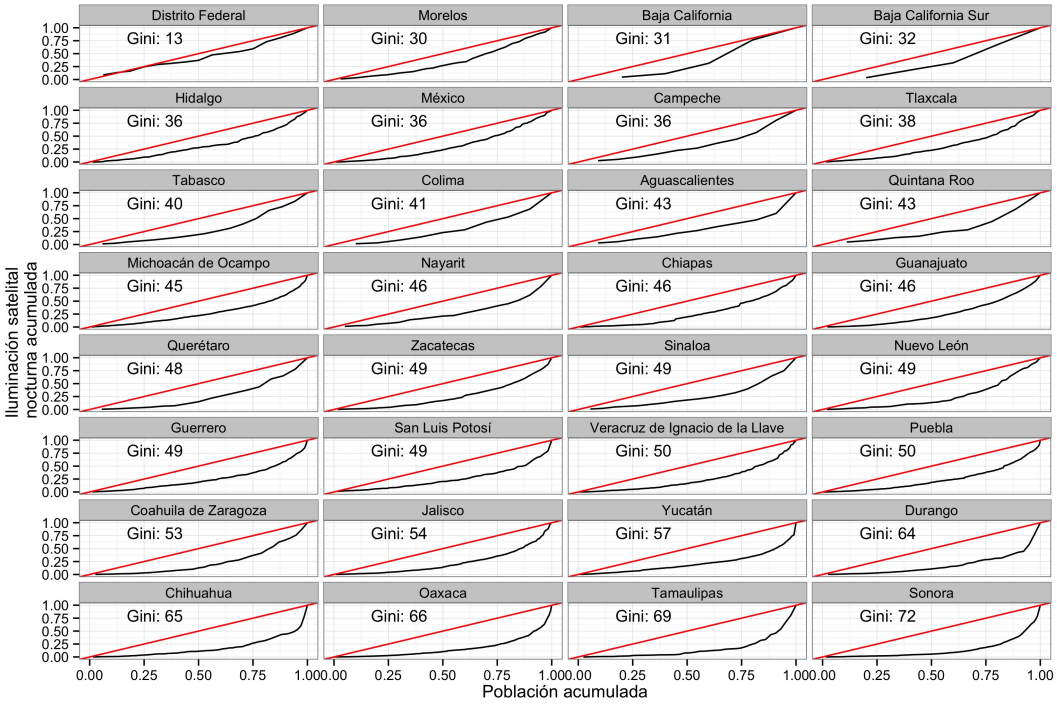
Fuente: Elaboración propia con datos de la NOAA [1] y del censo del INEGI [28].

Finalmente, para tener una comparación visual del índice de Gini de desigualdad de iluminación nocturna satelital se puede ver la imagen 4.9 a continuación en donde se

presentan las curvas de Lorenz de las Entidades Federativas en el 2013 (ver más detalle del cálculo en el capítulo 3, sección 3.4). Como se observó en la sección anterior, el Distrito Federal es la Entidad que presenta menos desigualdad de iluminación. De hecho, se puede notar que su curva de Lorenz es muy cercana a la recta de igualdad. También se puede notar, en el otro extremo, que Sonora tiene una curva de Lorenz bastante más alejada de la recta de 45 grados.

Imagen 4.9: Curvas de Lorenz en las Entidades Federativas, 2013

Orden por renglón con el índice de Gini de iluminación nocturna



Fuente: Elaboración propia con datos de la NOAA [1] y del censo del INEGI [28]

Capítulo 5.
Luces nocturnas como variable de predicción

En este capítulo se propone la metodología y el marco teórico para estimar el crecimiento económico de las Entidades Federativas de México. Se considera como base para la estimación el artículo de Henderson, et. al. [24] y se tomaron consideraciones y sugerencias que se realizan en el trabajo de Mendoza en la tesis [39] con Victor M. Guerrero como asesor. Es importante resaltar que la metodología propuesta por Henderson, et. al. en [24] la publicaron en paralelo con la de Chen, et al. en [6] incluso con notación muy similar. De hecho, los dos artículos se citan entre si y se publicaron con tan sólo unos meses de diferencia por lo que no es tan claro distinguir quién desarrolló la metodología original; sin embargo, no es relevante para los propósitos de este trabajo. En particular, en la sección 5.1 se define el contexto para poder modelar el crecimiento del PIB a nivel Estatal (se definió en la sección 2.4), por medio de la iluminación nocturna satelital. Se especifican los supuestos que requieren para modelar y se desarrolla un marco estadístico para mostrar cómo se puede estimar el crecimiento del PIBE. En la sección 5.2 se presentan los resultados empíricos y la estimación de los parámetros. Finalmente en la sección 5.3 se presentan los resultados de la estimación para el *verdadero* crecimiento del PIBE.

5.1. Iluminación nocturna para estimar actividad económica

Como se mencionó anteriormente, en esta sección se plantea el contexto económico teórico para poder utilizar la variable de iluminación nocturna satelital visible desde el espacio

como una *proxy* del crecimiento económico en las Entidades Federativas de México. Para esto, se sigue la metodología propuesta por Henderson, et. al. en [24]. Antes de empezar, para facilitar la lectura, se presenta la lista de variables que se utilizan en este contexto.

Definición 5.1.1. Lista de variables para el modelo de estimación:

X_t Iluminación nocturna satelital visible desde el espacio al tiempo t.

Y_t PIBE *verdadero* a precios del 2008 al tiempo t.

Z_t PIBE del INEGI a precios del 2008 al tiempo t.

x_t Logaritmo de la iluminación nocturna satelital visible desde el espacio al tiempo t.

y_t Logaritmo del PIBE *verdadero* a precios del 2008 al tiempo t.

z_t Logaritmo del PIBE del INEGI a precios del 2008 al tiempo t.

t Año.

j Entidad Federativa.

ϵ_x Error de medición en la estimación del PIBE como función de la iluminación nocturna satelital visible desde el espacio.

σ_x^2 Varianza del error de medición en la estimación del PIBE como función de la iluminación nocturna satelital visible desde el espacio. respecto al PIBE verdadero.

ϵ_z Error de medición en el PIBE con respecto al PIBE verdadero.

σ_z^2 Varianza del error de medición en el PIBE con respecto al PIBE verdadero.

σ_y^2 Varianza del PIBE *verdadero* a precios del 2008.

Ahora si, sea y_t el PIBE *verdadero* a precios del 2008 al tiempo t y sea z_t el PIBE al tiempo t y donde Z_t es el PIBE real medido con las cuentas nacionales, en este caso el INEGI. Sea x_t la iluminación nocturna satelital calibrada al tiempo t. Se define σ_y^2 como la varianza del ingreso real verdadero. Con esto, para cada Entidad Federativa j se asume que el modelo de la ecuación (5.1) a continuación

$$z_j = y_j + \epsilon_{z,j} \tag{5.1}$$

donde, se supondrá que $\epsilon_{z,j}$ sigue el proceso estocástico de un error de ruido blanco con varianza dada por σ_z^2. En palabras, el crecimiento real del verdadero PIBE se puede expresar como la medida del crecimiento en las cuentas nacionales del INEGI más un proceso estocástico de error con media cero.

Para la iluminación nocturna, por otro lado, se supondrá que presenta una elasticidad constante con el ingreso. Es decir, se puede expresar a la iluminación nocturna satelital como una función del ingreso que depende del parámetro β como se puede ver en la ecuación a continuación

$$X_j = Y_j^{\beta}, \tag{5.2}$$

donde β representa la elasticidad de la iluminación nocturna con respecto al ingreso. En cuanto a la estimación, se utiliza la siguiente ecuación en logaritmos

$$x_j = \beta y_j + \epsilon_{x,j}, \tag{5.3}$$

en donde se considera que el error $\epsilon_{x,j}$ también sigue una estructura de ruido blanco con una varianza dada por σ_x^2. Es decir, se piensa que error en la ecuación (5.3) como el ruido en la medición del PIBE estimado con los datos de la iluminación nocturna. Además se tiene el error de la iluminación nocturna real con respecto a lo que pueden medir los sensores OLS desde el espacio; sin embargo, esta medida de error incluye la variación entre los distintos Estados en la relación que tiene con el crecimiento del PIBE. Por tanto, es razonable pensar que el error de medición del PIBE en las cuentas nacionales no esta relacionado con el error de la iluminación observable por los satélites de la NOAA. Con este argumento, se considera que estos errores son independientes y por lo tanto se cumple que $\mathbf{COV}(\epsilon_x, \epsilon_z) = 0$.

Finalmente, dado que el objetivo es estimar una regresión en la que se modele el ingreso por medio de la iluminación nocturna satelital, se propone la siguiente ecuación (5.4)

$$z_j = \hat{\psi} x_j + e_j. \tag{5.4}$$

El parámetro $\hat{\psi}$ se puede estimar con mínimos ordinarios. Es decir, $\hat{\psi}$ está dado por

$$\hat{\psi} = \frac{\mathbf{COV}(x_j, z_j)}{\mathbf{VAR}(x_j)}, \tag{5.5}$$

en donde $\mathbf{COV}(x_j, z_j)$ y $\mathbf{VAR}(x_j)$ se pueden calcular utilizando las ecuaciones (5.14)

más adelante. Adicionalmente, se puede ver que el parámetro $\hat{\psi}$ converge en probabilidad, utilizando las ecuaciones (5.14b) y (5.14c), y que está dado por

$$\text{p lím}(\hat{\psi}) = \frac{\beta\sigma_{y_j}^2}{\beta^2\sigma_{y_j}^2 + \sigma_{x_j}^2} = \frac{1}{\beta}\left(\frac{\beta^2\sigma_{y_j}^2}{\beta^2\sigma_{y_j}^2 + \sigma_{x_j}^2}\right). \tag{5.6}$$

El parámetro $\hat{\psi}$ es una estimación del inverso de la elasticidad de la iluminación nocturna con respecto al PIB y, por construcción, es un estimado sesgado como se puede ver en la ecuación (5.6). El verdadero valor debe estar dado por β^{-1}. Sin embargo, la ecuación (5.4) define una relación para producir proxys para el crecimiento del ingreso que denotaremos como

$$\hat{z}_j = \hat{\psi}x_j.$$

Una dificultad adicional es que, aunque se pretende estimar proxys para el ingreso con base en la iluminación, el parámetro a estimar $\hat{\psi}$ es por construcción una estimación que utiliza datos del crecimiento del ingreso. Si fuera el caso en el que las cuentan nacionales tienen un error grande con respecto al ingreso verdadero, dado que suponemos que el error es independiente del de medición en las luces ($\mathbf{COV}(\epsilon_x, \epsilon_z) = 0$), el error del crecimiento del PIBE no tiene un efecto en la estimación del parámetro en la ecuación (5.4). Ahora, se puede estimar, a nivel Estatal, los valores ajustados del crecimiento real estimados por medio del parámetro la iluminación nocturna, $\hat{z}_j$. Más específicamente, $\hat{z}_j$ es un estimador del crecimiento del ingreso que se puede combinar con las cuentas nacionales, z_j, para poder generar un estimado del crecimiento del ingreso que tenga un menor error. Para hacer esto se estima el crecimiento del ingreso verdadero como una combinación lineal convexa de ambos como en la siguiente ecuación,

$$\hat{y}_j = \lambda z_j + (1-\lambda)\hat{z}_j, \tag{5.7}$$

en donde se especifican los pesos, λ y $1-\lambda$ que minimizan la varianza del error de de medición relativo a la estimación del verdadero crecimiento del ingreso. Es decir, mientras que se tenga un peso óptimo con estimación de $\hat{z}_j > 0$, los datos de iluminación nocturna mejoran la estimación del crecimiento verdadero del PIB. Dado lo anterior, la varianza del

error del crecimiento del ingreso verdadero está dado por

$$\begin{aligned}\mathbf{VAR}(\hat{y}_j - y_j) &= \mathbf{VAR}(\lambda(z_j - y_j) + (1-\lambda)(\hat{z}_j - y_j)) \\ &= \lambda^2\sigma^2_{z_j} + (1-\lambda)^2\mathbf{VAR}(\hat{z}_j - y_j). \end{aligned} \tag{5.8}$$

El término $\mathbf{VAR}(\hat{z}_j - y_j)$ se puede expresar por la ecuación a continuación

$$\begin{aligned}\mathbf{VAR}(\hat{z}_j - y_j) = \mathbf{VAR}(\hat{\psi}x_j - y_j) = \mathbf{VAR}(\hat{\psi}\beta y_j + \hat{\psi}\epsilon_{x_j} - y_j) = \\ = (\hat{\psi}\beta - 1)^2\sigma^2_{y_j} + \hat{\psi}^2\sigma^2_{x_j}. \end{aligned} \tag{5.9}$$

Como $\hat{\psi}$ converge en probabilidad al valor de la ecuación (5.6), se puede reescribir $\mathbf{VAR}(\hat{z}_j - y_j)$ como

$$\mathbf{VAR}(\hat{z}_j - y_j) = \left(\frac{\beta^2\sigma^2_{y_j}}{\beta^2\sigma^2_{y_j} + \sigma^2_{x_j}} - 1\right)^2 \sigma^2_{y_j} + \left(\frac{\beta\sigma^2_{y_j}}{\beta^2\sigma^2_{y_j} + \sigma^2_{x_j}}\right)^2 \sigma^2_{x_j} = \frac{\sigma^2_{y_j}\sigma^2_{x_j}}{\beta^2\sigma^2_{y_j} + \sigma^2_{x_j}}. \tag{5.10}$$

Por lo que, sustituyendo en la ecuación (5.8), se tiene que

$$\mathbf{VAR}(\hat{y}_j - y_j) = \lambda^2\sigma^2_{z_j} + (1-\lambda)^2\frac{\sigma^2_{y_j}\sigma^2_{x_j}}{\beta^2\sigma^2_{y_j} + \sigma^2_{x_j}}. \tag{5.11}$$

Por lo que se se puede obtener el valor óptimo de λ, minimizando la varianza $\mathbf{VAR}(\hat{y}_j - y_j)$. Es decir,

$$\min_{\lambda} \quad \lambda^2\sigma^2_{z_j} + (1-\lambda)^2\frac{\sigma^2_{y_j}\sigma^2_{x_j}}{\beta^2\sigma^2_{y_j} + \sigma^2_{x_j}}.$$

La condición de primer orden igualada a cero está dada por

$$2\lambda\sigma^2_{z_j} + 2(1-\lambda)\frac{\sigma^2_{y_j}\sigma^2_{x_j}}{\beta^2\sigma^2_{y_j} + \sigma^2_{x_j}} = 0. \tag{5.12}$$

Por lo tanto despejando se tiene que λ^* está dado por

$$\lambda^* = \frac{\sigma^2_{x_j}\sigma^2_{y_j}}{(\beta^2\sigma^2_{y_j} + \sigma^2_{x_j})\sigma^2_{z_j} + \sigma^2_{x_j}\sigma^2_{y_j}}. \tag{5.13}$$

Que es una función que depende de parámetros desconocidos, $\sigma^2_{x_j}, \sigma^2_{y_j}, \sigma^2_{z_j}$ y β, pero se

puede estimar con los valores empíricos de

$$\mathbf{VAR}(z_j) = \sigma_{y_j}^2 + \sigma_{z_j}^2 \tag{5.14a}$$

$$\mathbf{VAR}(x_j) = \beta^2\sigma_{y_j}^2 + \sigma_{x_j}^2 \tag{5.14b}$$

$$\mathbf{COV}(x_j, z_j) = \beta\sigma_{y_j}^2 \tag{5.14c}$$

$$\mathbf{COV}(y_j, x_j) = \beta\sigma_{y_j}^2. \tag{5.14d}$$

Sin embargo, como $\mathbf{COV}(x_j, z_j) = \mathbf{COV}(y_j, x_j)$ hace falta una ecuación adicional para poder estimar λ. Para esto, se supondrá que la razón de cambio de la varianza del crecimiento del PIB con respecto a la varianza total medida con la iluminación nocturna está dado por ϕ, en donde

$$\phi = \frac{\sigma_{y_j}^2}{\sigma_{y_j}^2 + \sigma_{z_j}^2}. \tag{5.15}$$

En donde, si $\phi = 1$ se tiene que $\sigma_{z_j}^2 = 0$ lo que implicaría que las cuentas del INEGI son tal cual el verdadero valor. En el otro extremo, si $\phi = 0$ implica que las cuentas nacionales no sirven de nada para medir el verdadero valor del crecimiento del ingreso. Finalmente sustituyendo esta relación, dada por ϕ, en la ecuación (5.13), se puede estimar λ^* como

$$\lambda^* = \frac{\phi\mathbf{VAR}(z_j)\mathbf{VAR}(x_j) - \mathbf{COV}^2(x_j, z_j)}{\mathbf{VAR}(z_j)\mathbf{VAR}(x_j) - \mathbf{COV}^2(x_j, z_j)} = \frac{\phi - \rho^2}{1 - \rho^2}, \tag{5.16}$$

en donde $\rho = \mathbf{CORR}(x_j, z_j)$. En este punto, Henderson, et. al. en [24] hacen una distinción porque, en su caso, suponen que los distintos países tienen diferentes calidades en cuanto a las estadísticas nacionales y por lo tanto, estiman la varianza de x para países con buenas y malas estadísticas nacionales. En este trabajo, no se hizo esta distinción dado que los datos de las Entidades Federativas corresponden todos al INEGI.

Con lo anterior, se tiene lo necesario para poder estimar todos los parámetros del modelo. Es decir, se cuenta con un sistema de ecuaciones con el mismo número de parámetros. Para resolver el sistema, se debe comenzar suponiendo un valor para ϕ, recordando que es un índice que representa qué tan cercano es el valor de la medición del PIBE de las cuentas nacionales al verdadero valor. Una vez que se tiene el valor de ϕ se puede obtener σ_y^2 combinando la ecuación (5.14a) con la ecuación (5.15) y con el valor empírico de $\mathbf{VAR}(x_j)$. Es

decir, dado $\phi = \phi_0$, y $\mathbf{V\hat{A}R}(z_j)$, el valor empírico dado por los datos, se tiene que

$$\sigma^2_{y_j} = \mathbf{V\hat{A}R}(z_j)\phi_0, \tag{5.17a}$$

$$\sigma^2_{x_j} = \mathbf{V\hat{A}R}(z_j) - \sigma^2_{y_j}, \tag{5.17b}$$

con lo que, tomando el valor empírico de $\mathbf{C\hat{O}V}(x_j, z_j)$, se puede obtener el valor de β de la ecuación (5.14c). Es decir, β está dado por

$$\beta = \frac{\mathbf{C\hat{O}V}(x_j, z_j)}{\sigma^2_{y_j}}. \tag{5.18}$$

Finalmente, con esto es suficiente para estimar los demás parámetros como λ^*, sustituyendo en la ecuación (5.13). Por lo que ahora se procederá a la estimación empírica.

5.2. Predicción del PIBE con iluminación nocturna satelital

En esta sección se define un modelo empírico para estimar el PIB a nivel Estatal utilizando el marco teórico que proponen Henderson et al. [24] y que se definió en la sección anterior. Es importante notar que la serie de iluminación nocturna presenta cambios anuales, desviaciones de la tendencia del crecimiento balanceado y crecimiento de largo plazo como se vio en el capítulo 4.

En un contexto de variables panel para los años 2003 a 2013, se reescribe la ecuación de estimación (5.4) como

$$\tilde{e}_{j,t} = c_j + d_t + e_{j,t}, \tag{5.19}$$

para el Estado j al tiempo t. En donde, d_t controla por efectos fijos[1] en las condiciones económicas como cambios en la tecnología y costos de la energía eléctrica. El coeficiente c_j controla por efectos fijos de las diferencias culturales en el uso de la energía eléctrica a lo largo de la extensión territorial y así como la actividad económica diurna contra nocturna, diferencias entre el uso de iluminación publica contra privada y condiciones naturales para generar energía eléctrica en cada Entidad Federativa. Finalmente se puede incluir un término adicional del tiempo $\kappa_j t$, en donde el coeficiente mide las fluctuaciones en la senda de crecimiento.

[1] En el artículo de Henderson, et. al. controla con efectos de los satélites en el tiempo; sin embargo, ellos no realizaron la calibración de los satélites por lo que no tendrá la misma interpretación en este trabajo.

Desafortunadamente, la serie del PIBE del INEGI está incompleta con respecto a la serie de imágenes satelitales. La serie del PIBE cubre sólo desde el año 2003 hasta el 2013. Es por esto que se utilizaron sólo estos años en las estimaciones. Por otra parte las imágenes representan información muy valiosa en términos de datos para la estimación dado que representan un nivel muy desagregado. Para no perder este valor agregado haciendo una agregación a nivel Estatal, se ponderó la serie del PIBE a nivel Municipal utilizando el porcentaje de habitantes totales en cada municipio con respecto a los habitantes del Estado correspondiente. Con esto, se tiene datos para 11 años, y 2,456 Municipios con lo cual la serie total tiene 27,016 observaciones[2]. Esto representa un nuevo enfoque y es radicalmente distinto de lo que se ha hecho en la literatura, por ejemplo, Henderson, et. al. utilizaron una muestra grande de países y pocos años de forma que sólo trabajó con 3,015 datos.

5.2.1. Modelo base empírico

En la tabla 5.1 se puede apreciar el resumen del modelo base junto con modelos que incorporan pequeños cambios. Para empezar, en las columnas 1 a 4, se comparan los resultados de utilizar distintas versión de la serie de iluminación nocturna satelital. Es decir, en la columna 1, se puede apreciar el modelo con la serie de iluminación nocturna antes de calibrar. En la columna 2 es la misma serie ponderada por la superficie territorial de cada Municipio. En la columna 3 se utiliza la serie calibrada según la metodología del capítulo 3. En la cuarta columna se utiliza la serie de iluminación calibrada y ponderada por la superficie territorial de cada municipio. De las columnas 1 a 4, se puede apreciar que en ambos casos, con datos calibrados y no calibrados, ponderar con el área de cada Municipio no aporta a mejorar la estimación, en términos de R^2 y de error estándar. Por otra parte, de las columnas 1 y 3, que comparan los datos calibrados contra los no calibrados, se puede apreciar que la calibración mejora significativamente el modelo. El error estándar es 22 % mayor si no se utilizan los datos calibrados. Además los coeficientes de ψ y la constante son ambos significativos. Es por esto que se tomó como modelo base el modelo de la columna 3 de la tabla 5.1.

Adicionalmente, en la última columna de la tabla 5.1, se tiene el modelo base con un término adicional, la serie del índice espacial de gini de desigualdad de iluminación nocturna satelital. De la tabla 5.1 se puede notar que, como era de esperar, la R^2 se incrementa,

[2]Este dato incluye al municipio ficticio de Cantarell

lo cual fortalece al modelo porque se está capturando la variabilidad que hay entre los Estados a través del índice espacial de Gini. Sin embargo, el error estándar se mantiene prácticamente constante por lo que no se piensa que tenga valor predictivo importante.

Es importante notar, además, que los dos modelos que ponderan la iluminación satelital nocturna por medio de la superficie territorial en la tabla 5.1 tienen menos observaciones. Esto se debe a que no incluye el municipio ficticio de Cantarell que se creo para corregir por el alto PIBE que se tiene en Campeche. No se incluye debido a que no es posible imputar una superficie territorial debido a que las plataformas petroleras están en el mar principalmente y carece de sentido imputar un valor.

Tabla 5.1: Modelo base con y sin datos calibrados

	Variable dependiente:				
	log(pibe_mun)				
	(1)	(2)	(3)	(4)	(5)
Constante	1.126***	4.736***	0.581***	4.597***	
	(0.020)	(0.025)	(0.017)	(0.024)	
log(dn + 1)	0.849***				
	(0.003)				
log(dn_area + 1)		0.547***			
		(0.006)			
log(dn_cal + 1)			0.955***		0.951***
			(0.003)		(0.003)
log(dn_cal_area + 1)				0.623***	
				(0.006)	
log(gini_e)					0.184***
					(0.005)
N	27,016	26,994	27,016	26,994	27,016
R^2	0.755	0.213	0.836	0.261	0.990
Adjusted R^2	0.755	0.213	0.836	0.261	0.990
Error Std.	0.851 (df = 27014)	1.524 (df = 26992)	0.696 (df = 27014)	1.476 (df = 26992)	0.695 (df = 27014)

*p<0.1; **p<0.05; ***p<0.01

Fuente: Elaboración propia con datos de la NOAA [1], PIBE, Población Municipal, Población Estatal y superficie territorial del INEGI [29]. Tabla elaborada con el paquete `stargazer`, ver [26].

Ahora, en la tabla 5.2 más adelante se exploró la relación que tiene el PIBE con la cantidad de hogares que cuentan con electricidad en los Municipios dentro de las Entidades Federativas. Los datos de electricidad se obtuvieron del censo económico del INEGI correspondiente a los años 1995, 2000, 2005, y 2010. Se realizó una interpolación por medio de un spline natural para imputar los valores faltantes. La variable representa el porcentaje del total dentro de cada Municipio.

Tabla 5.2: Modelos de porcentaje de viviendas con electricidad

	Variable dependiente:	
	log(pibe_mun)	
	(1)	(2)
Constante	-29.170^{***}	-2.087^{***}
	(0.397)	(0.202)
log(dn_cal + 1)		0.935^{***}
		(0.003)
p_viv_elect	0.374^{***}	0.029^{***}
	(0.004)	(0.002)
N	27,016	27,016
R^2	0.232	0.837
Adjusted R^2	0.232	0.837
Error Std.	1.504 (df = 27014)	0.694 (df = 27013)

*p<0.1; **p<0.05; ***p<0.01

Nota: El % de Viviendas con electricidad se interpoló con un spline natural de los años base del censo económio del INEGI que incluye 1995, 2000, 2005 y 2010.

Fuente: Elaboración propia con datos de la NOAA [1], PIBE, Población Municipal, Población Estatal y % de Viviendas que cuentan con electricidad del INEGI [29]. Tabla elaborada con el paquete `stargazer`, ver [26].

De la tabla, 5.2 se pueden apreciar dos modelos, en la columna 1, se estimó el PIB con el porcentaje de hogares que cuentan con electricidad por Municipio y en la columna 2, se estima el modelo base incluyendo el término de iluminación nocturna satelital. Se

consideró esta variable para contrastar qué tanto poder predictivo tiene con respecto a los datos de iluminación nocturna satelital. Para empezar, comparando el modelo con sólo el porcentaje de hogares con electricidad con respecto al modelo base, se puede notar que el de electricidad en los hogares es menos apropiado debido a que presenta un peor ajuste y tiene un error estándar significativamente más alto. Por otro lado, de la columna 2, se puede notar que el porcentaje de electricidad de los hogares tiene un coeficiente muy pequeño que, aunque el error estándar del coeficiente no genera un intervalo de confianza incluya al cero, tampoco aporta al modelo. De hecho, tanto la R^2 como el error estándar se mantienen prácticamente constantes.

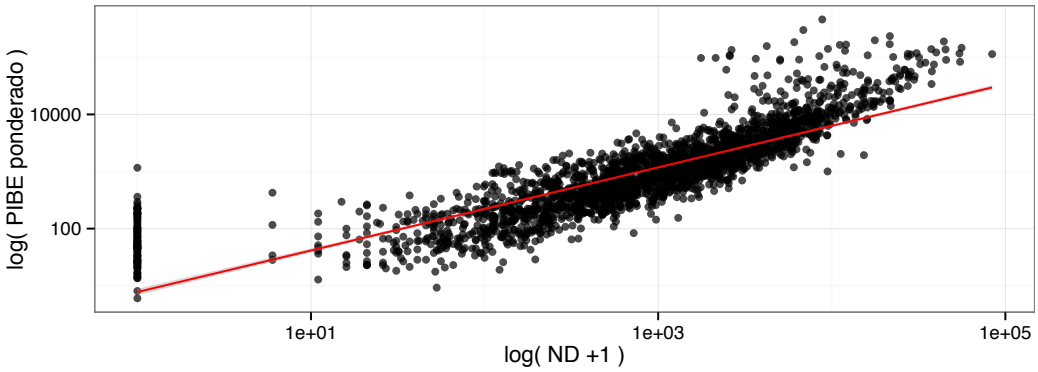
Es posible que se pueda mejorar la estimación utilizando datos de consumo de electricidad que probablemente modelen mejor el PIBE. Sin embargo, y es justo el punto de este trabajo, obtener estos datos es mucho más complejo por lo que utilizar los datos de iluminación nocturna desde el espacio es una alternativa más viable y por lo tanto una mejor opción.

De las imágenes 5.1 a 5.4 se pueden apreciar la relación que tienen el PIBE ponderado para los Municipios a precios del 2008 con la suma del Número Digital. Se puede apreciar una vez más que las imágenes 5.2 y 5.4, en donde se ponderan la iluminación con la superficie territorial, tienen una correlación mucho más débil en comparación con lo que se puede apreciar en las imágenes 5.1 y 5.3. Se puede apreciar que para el modelo base, la elasticidad que tiene la iluminación nocturna satelital con el PIBE es cercana uno, el coeficiente es $\psi = .9555$. Esta relación se aprecia muy claramente en la imagen 5.3.

Además, de la comparación entre las imágenes 5.1 y 5.3 de datos de satélites calibrados y no calibrados, se puede apreciar que el proceso de calibración está haciendo un trabajo importante. Esto es evidente de entrada porque los datos atípicos que se observan en los niveles más bajos de iluminación a la izquierda en la imagen 5.1 que corresponde a los datos de satélites originales, ya no se aprecian en la imagen 5.3.

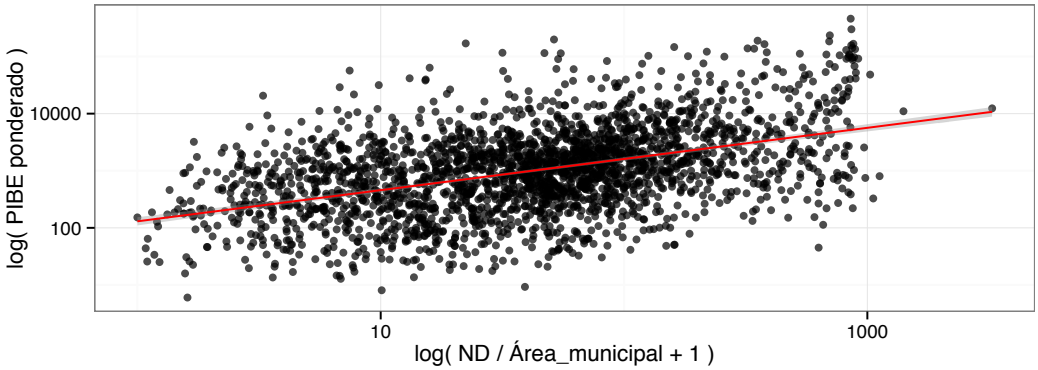
Es decir, de las imágenes 5.1 a 5.4 es evidente que la mejor alternativa para modelar el PIBE es utilizar la serie de iluminación nocturna satelital calibrada.

Imagen 5.1: Relación PIBE y ND, 2013



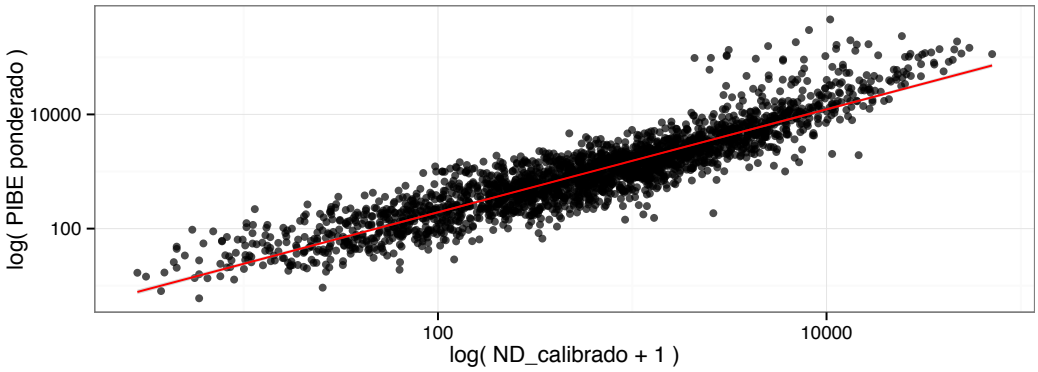
Fuente: Elaboración propia con datos de la NOAA [1], PIBE, Población Municipal y Población Estatal del INEGI [29].

Imagen 5.2: Relación PIBE y ND por superficie Municipal, 2013



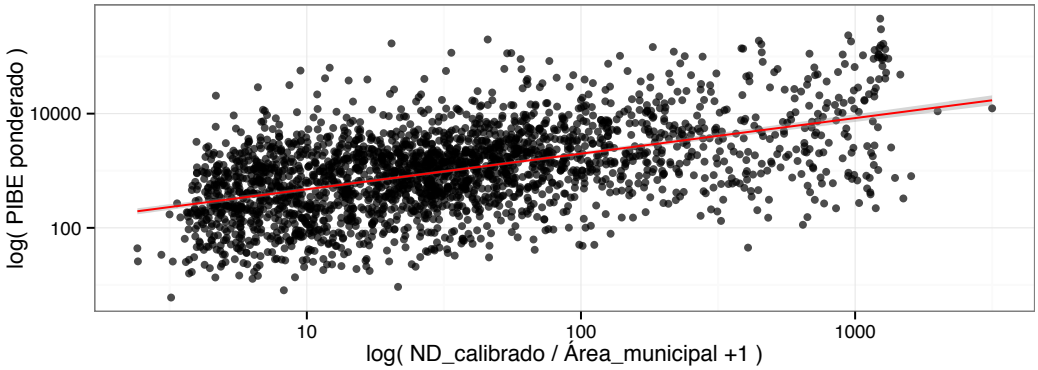
Fuente: Elaboración propia con datos de la NOAA [1], PIBE, Población Municipal, Población Estatal y superficie territorial del INEGI [29].

Imagen 5.3: Relación PIBE y ND calibrado, 2013



Fuente: Elaboración propia con datos de la NOAA [1], PIBE, Población Municipal y Población Estatal del INEGI [29].

Imagen 5.4: Relación PIBE y ND calibrado por superficie Municipal, 2013



Fuente: Elaboración propia con datos de la NOAA [1], PIBE, Población Municipal, Población Estatal y superficie territorial del INEGI [29].

Es importante notar que el modelo base que se esta considerando en este trabajo es ya, de entrada, superior al modelo base que estimaron Henderson et. al. en [24]. En su modelo se obtuvo una $R^2 = 0.769$ con todo y que se incluyeron efectos fijos de año y país comparado con la $R^2 = 0.836$ que todavía no incluye efectos fijos. Ellos no estiman el error estándar del modelo aunque estiman el de ψ, con valor de 0.03 que es muy superior

comparado contra el de 0.003 que se estimó en el modelo base. Lo anterior, aunque, los coeficientes no se comparan directamente porque ellos incorporan efectos fijos para países y años desde el modelo base. Esto habla del poder que tiene hacer la calibración además de que es valioso ponderar el PIB de forma que no se pierda información al agregar los datos a nivel país. Para seguir con la misma línea de pensamiento, se procedió a estimar el modelo incorporando efectos fijos.

5.2.2. Modelo empírico con efectos fijos

En la tabla 5.3 a continuación se muestran los resultados de incorporar efectos fijos por año, por Entidad Federativa y por ambos. En la columna 1, se puede apreciar que los coeficientes son significativos y que en general el coeficiente ψ no cambia con respecto al modelo base. Sin embargo, al final de la tabla 5.3 se puede notar que aunque hay un aumento notable en la R^2 de 0.836 a 0.990, no sucede lo mismo con el error estándar que se mantuvo relativamente constante cambia de 0.696 a 0.687. Este resultado es de esperarse dado que los efectos de cada año se mitigaron al hacer la calibración por lo que era esperado que el error no disminuyera. Sin embargo, esto es un resultado positivo debido a que fortalece la importancia de realizar la calibración.

Por otro lado, en la columna 2 de la tabla 5.3, se incorporan efectos fijos de cada Estado sin incluir los efectos anuales. En este caso, otra vez se incrementó la R^2 de 0.836 a 0.993 y, a diferencia de la columna 1, en este modelo si se redujo el error estándar de 0.696 a 0.581. Se puede apreciar que en este caso, el coeficiente estimado para $\psi = 0.907$ sigue siendo similar que en el modelo base.

Finalmente, en la columna 3 de la tabla 5.3, se incorporan efectos fijos de cada Estado y efectos fijos anuales. Se puede apreciar que para no sobreparametrizar el modelo, no se incluye un coeficiente correspondiente al año 2003. Es decir, los coeficientes de las Entidades Federativas corresponden a este año. Se puede apreciar que no hay mucha ganancia en términos de R^2 en comparación con el modelo de la columna 1, que pasó de 0.99 a 0.993, y que se redujo el error estándar todavía más que en el modelo con sólo efectos anuales, bajó de 0.687 a 0.569.

Tabla 5.3: Modelos con efectos fijos

	Variable dependiente:		
	log(pibe_mun)		
	(1)	(2)	(3)
log(dn_cal + 1)	0.956***	0.907***	0.911***
	(0.003)	(0.003)	(0.003)
anio2003	0.402***		
	(0.022)		
anio2004	0.427***		0.025
	(0.022)		(0.016)
anio2005	0.416***		0.016
	(0.022)		(0.016)
anio2006	0.489***		0.087***
	(0.022)		(0.016)
anio2007	0.566***		0.162***
	(0.022)		(0.016)
anio2008	0.628***		0.222***
	(0.021)		(0.016)
anio2009	0.592***		0.186***
	(0.021)		(0.016)
anio2010	0.597***		0.193***
	(0.022)		(0.016)
anio2011	0.659***		0.254***
	(0.021)		(0.016)
anio2012	0.732***		0.326***
	(0.021)		(0.016)
anio2013	0.755***		0.348***
	(0.021)		(0.016)
nombre_estadoAguascalientes		0.968***	0.773***
		(0.058)	(0.058)

Tabla 5.3 Modelos con efectos fijos ***(continuación)***

	Variable dependiente:		
	log(pibe_mun)		
	(1)	(2)	(3)
nombre_estadoBaja California		1.436***	1.233***
		(0.084)	(0.083)
nombre_estadoBaja California Sur		0.878***	0.678***
		(0.083)	(0.083)
nombre_estadoCampeche		2.984***	2.787***
		(0.059)	(0.059)
nombre_estadoChiapas		0.908***	0.718***
		(0.026)	(0.028)
nombre_estadoChihuahua		0.214***	0.021
		(0.031)	(0.033)
nombre_estadoCoahuila		0.603***	0.408***
		(0.038)	(0.039)
nombre_estadoColima		1.225***	1.031***
		(0.060)	(0.060)
nombre_estadoDistrito Federal		3.672***	3.475***
		(0.051)	(0.052)
nombre_estadoDurango		0.490***	0.298***
		(0.036)	(0.037)
nombre_estadoGuanajuato		1.052***	0.857***
		(0.036)	(0.037)
nombre_estadoGuerrero		0.918***	0.728***
		(0.029)	(0.030)
nombre_estadoHidalgo		0.926***	0.735***
		(0.029)	(0.030)
nombre_estadoJalisco		1.048***	0.856***
		(0.027)	(0.029)
nombre_estadoMéxico		0.999***	0.805***

Tabla 5.3 Modelos con efectos fijos ***(continuación)***

	Variable dependiente:		
	log(pibe_mun)		
	(1)	(2)	(3)
		(0.029)	(0.030)
nombre_estadoMichoacán		0.983***	0.792***
		(0.027)	(0.029)
nombre_estadoMorelos		0.767***	0.573***
		(0.039)	(0.040)
nombre_estadoNayarit		0.844***	0.651***
		(0.046)	(0.046)
nombre_estadoNuevo León		0.987***	0.793***
		(0.034)	(0.035)
nombre_estadoOaxaca		0.748***	0.566***
		(0.016)	(0.019)
nombre_estadoPuebla		0.864***	0.675***
		(0.023)	(0.025)
nombre_estadoQuerétaro		1.454***	1.259***
		(0.048)	(0.048)
nombre_estadoQuintana Roo		1.291***	1.094***
		(0.064)	(0.064)
nombre_estadoSan Luis Potosí		1.103***	0.912***
		(0.032)	(0.033)
nombre_estadoSinaloa		1.032***	0.835***
		(0.049)	(0.049)
nombre_estadoSonora		0.156***	−0.035
		(0.030)	(0.031)
nombre_estadoTabasco		1.639***	1.441***
		(0.050)	(0.051)
nombre_estadoTamaulipas		0.935***	0.742***
		(0.035)	(0.036)

Tabla 5.3 Modelos con efectos fijos ***(continuación)***

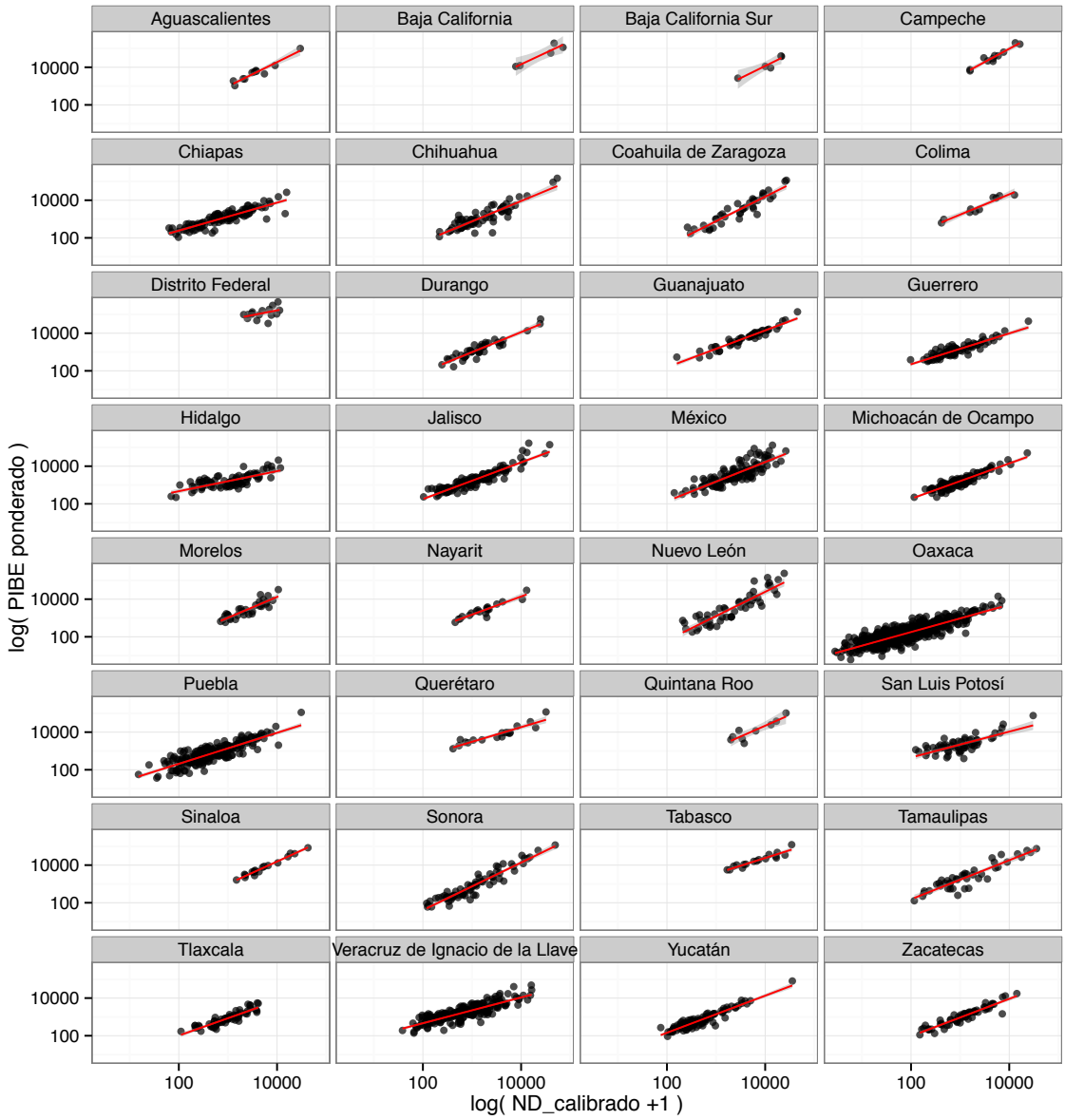
	Variable dependiente:		
	log(pibe_mun)		
	(1)	(2)	(3)
nombre_estadoTlaxcala		0.373***	0.183***
		(0.031)	(0.032)
nombre_estadoVeracruz de Ignacio de la Llave		1.388***	1.198***
		(0.024)	(0.026)
nombre_estadoYucatán		0.713***	0.524***
		(0.026)	(0.027)
nombre_estadoZacatecas		0.393***	0.201***
		(0.032)	(0.033)
N	27,016	27,016	27,016
R^2	0.990	0.993	0.993
Adjusted R^2	0.990	0.993	0.993
Error Std.	0.687 (df = 27004)	0.581 (df = 26983)	0.569 (df = 26973)

*p<0.1; **p<0.05; ***p<0.01

Fuente: Elaboración propia con datos de la NOAA [1], PIBE, Población Municipal, Población Estatal y superficie territorial del INEGI [29]. Tabla elaborada con el paquete `stargazer`, ver [26].

Para complementar el análisis, de la imagen 5.5 a continuación se puede ver la relación del PIBE ponderado en el 2013 y su relación con la suma del número digital en cada municipio y dividido por Entidad Federativa. De la imagen se puede ver que la correlación que tienen estas variables incluso a este nivel es bastante alta. Se puede notar además que hay diferencias transversales en las distintas entidades de forma que hace sentido utilizar el modelo con efectos fijos por Entidad Federativa. En la imagen se muestran los datos para el 2013 por lo que no es posible comparar intertemporalmente, se pueden ver las variaciones anuales de las series para los estados más adelante en la imagen 5.7. Sin embargo, se decidió dejar los efectos anuales para seguir con el modelo de Henderson et. al. Es decir, se decidió trabajar con el modelo que incluye efectos fijos anuales y Estatales para la estimación de los parámetros del modelo teórico de la sección anterior.

Imagen 5.5: Relación PIBE y ND calibrado por Entidad Federativa, 2013



Fuente: Elaboración propia con datos de la NOAA [1], PIBE, Población Municipal, Población Estatal y superficie territorial del INEGI [29].

5.2.3. Estimación de los parámetros

En esta sección se estiman los parámetros de las ecuaciones (5.14) necesarios para poder definir del modelo teórico y así poder modelar el *verdadero* PIBE.

Como se explicó anteriormente, lo primero que se debe de hacer es elegir un valor de ϕ que representa qué tan bien mide el PIBE del INEGI al verdadero PIBE. En este caso, a diferencia del modelo de Henderson, et. al., se consideró un valor único de ϕ con la idea de que las medidas que tiene el INEGI tienen un error consistente en las distintas Entidades. Negar este supuesto implicaría hablar de temas escabrosos y no son tema de este trabajo porque supondría que el INEGI tiene genera distintos errores en la medición del PIBE según la Entidad Federativa.

Ahora, para elegir el parámetro correcto se consideró un criterio probabilístico. Siguiendo la idea de Mendoza en [39], se eligió ϕ de forma que la elección de ϕ genere un intervalo de confianza para la estimación del *verdadero* PIBE que sea tal que cumpla el Teorema de Tchebysheff, es decir, que se cumpla que

$$\mathbb{P}\Big[|\hat{y} - y| \geq k\sqrt{\mathbf{VAR}(\hat{y} - y)(\phi)}\Big] < \frac{1}{k^2} \tag{5.20}$$

donde k es una constante a elegir, y como se vio en las ecuaciones (5.11) y (5.16), $\mathbf{VAR}(\hat{y} - y)(\phi)$ es una función que depende del parámetro ϕ.

En la tabla 5.4 a continuación se puede ver el valor empírico de la probabilidad con distintas elecciones de ϕ tomando $k = 2$, con la idea de obtener un intervalo de confianza de dos desviaciones estándar. Es decir, se busca un valor de ϕ que sea tal que la $\mathbb{P}(\cdot)$ sea menor que $1/k^2 = 0.25$. Se puede ver que un valor de $\phi \in (0.93, 0.94)$ cumple razonablemente con este supuesto.

Tabla 5.4: Elección de ϕ

ϕ	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	1.00
$\mathbb{P}(\cdot)$	0.36	0.32	0.30	0.26	0.21	0.18	0.16	0.15	0.15	0.14	1.00

Fuente: Elaboración propia.

Ahora, de la tabla 5.5, se puede ver cómo cambia el valor de los parámetros estimados para los distintos valores de ϕ, recordando que λ es el parámetro correspondiente al PIBE

del INEGI por lo que, un menor valor de λ implica mayor peso a la variable *proxy* de iluminación nocturna. Hay que notar también que los valores de ϕ que se consideraron son tales que generan valores de $\lambda \in [0, 1]$ como se definió anteriormente.

Tabla 5.5: Parámetros para distintos valores de ϕ

ϕ	λ	β	$\sigma^2_{\tilde{y}}$
0.8400	0.0273	1.0420	0.0177
0.8500	0.0881	1.0297	0.0432
0.8600	0.1489	1.0177	0.0652
0.8700	0.2097	1.0060	0.0836
0.8800	0.2705	0.9946	0.0984
0.8900	0.3313	0.9834	0.1097
0.9000	0.3921	0.9725	0.1175
0.9100	0.4529	0.9618	0.1217
0.9200	0.5136	0.9513	0.1224
0.9300	0.5744	0.9411	0.1195
0.9400	0.6352	0.9311	0.1131
0.9500	0.6960	0.9213	0.1031
0.9600	0.7568	0.9117	0.0896
0.9700	0.8176	0.9023	0.0725
0.9800	0.8784	0.8931	0.0519
0.9900	0.9392	0.8841	0.0277
1.0000	1.0000	0.8752	0.0000

Fuente: Elaboración propia.

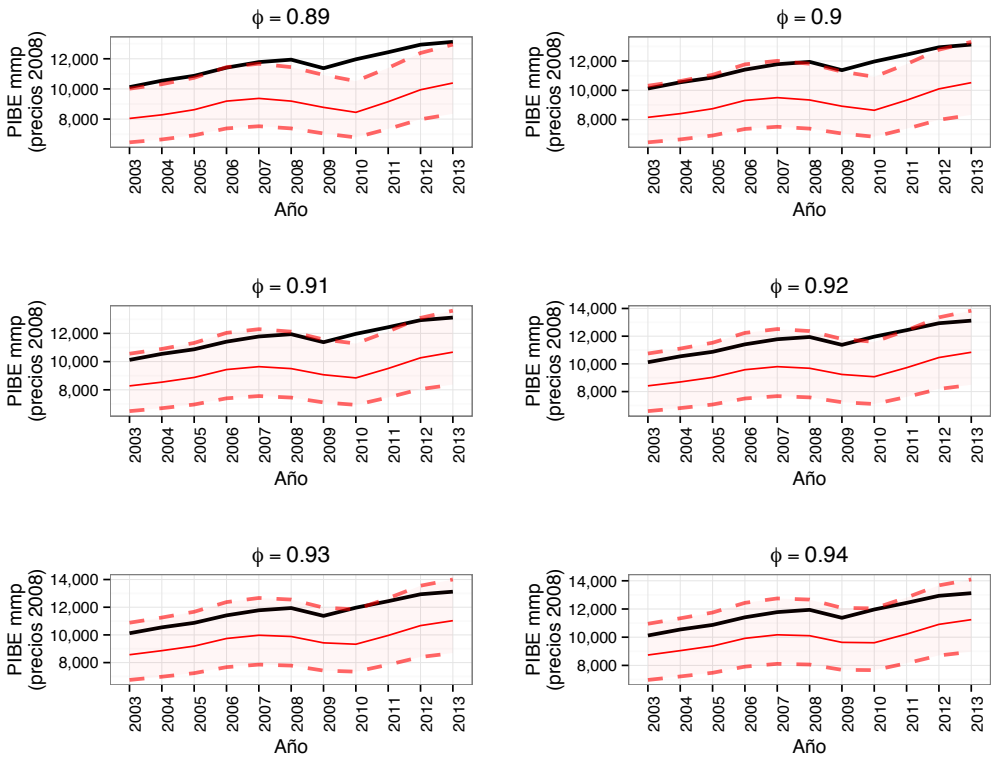
Nota: No se consideraron valores más pequeños de ϕ debido a que, dados los datos, la solución de los parámetros genera soluciones con varianzas negativas.

Para ilustrar el comportamiento de la estimación del PIB nacional con distintas elecciones de ϕ, se puede ver la imagen 5.6 en donde se puede ver el comportamiento de la serie del PIB medido según el INEGI (negro), la estimación del *verdadero* PIB (rojo) utilizando el modelo con distintas elecciones para valores de ϕ y su respectivo intervalo de confianza de $k = 2$ desviaciones estándar. De la imagen se puede notar que los valores más bajos

de ϕ, como el de 0.89 no generan un intervalo de confianza que contenga las suficientes observaciones para que se pueda considerar una buena elección del parámetro. Similarmente, elecciones más altas de ϕ como de 0.94 generan un intervalo que se acerca más a un intervalo de confianza congruente con el teorema de Tchebysheff. Es importante notar, que como la estimación se realizó a nivel Estatal, esta imagen no es tan representativa del verdadero intervalo de confianza, es sólo para ilustrar cómo es que depende del parámetro mencionado. Se puede ver la imagen 5.7 más adelante para ver el PIBE por Estado con elección de $\phi = 0.93$, que se eligió con la tabla 5.4.

Imagen 5.6: Serie del PIB y estimación para distintos valores de ϕ

Serie del PIB INEGI en negro. Estimación del PIB en rojo. Intervalo de confianza de 2 desviaciones estándar ($k = 2$), línea punteada roja.



Fuente: Elaboración propia con datos de la NOAA [1] y PIBE del INEGI [29].

Una vez que se analizó cómo cambia el intervalo de confianza para distintos valores de

ϕ, se decidió tomar un valor de $\phi = 0.93$. Con esto, de la tabla 5.5 se puede ver el valor correspondiente de los demás parámetros ajustados. Se tiene un valor de $1 - \lambda^* = 0.4256$, que es el parámetro del peso que tiene la *proxy* ajustada del modelo contra $\lambda^* = 0.5744$ que es el peso que tiene el PIBE del INEGI. El parámetro $\beta = 0.9411$ que es el representa la elasticidad del ingreso con respecto a la iluminación nocturna satelital y, finalmente, $\sigma^2_{\hat{y}} = 0.1195$, que representa la varianza que tiene el error de estimación de la estimación del PIBE *verdadero* ajustado.

Estos resultados no se pueden comparar directamente con lo que realizaron Henderson, et. al. en [24] por que ellos no reportaron el resultado de la estimación correspondiente para todos los países que ajustaron (alrededor de 120) y por lo tanto no se tiene los valores correspondientes para México.

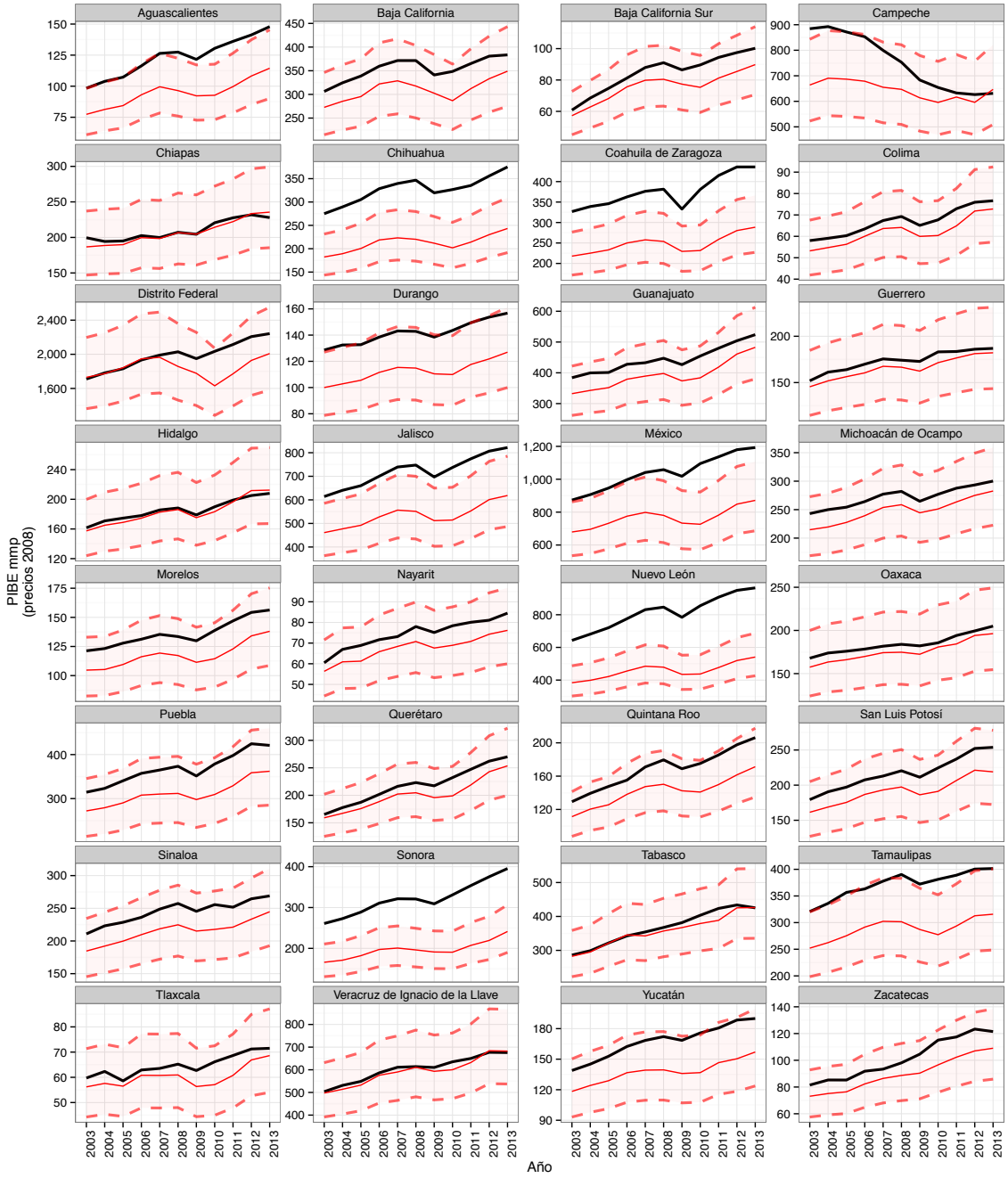
Adicionalmente de la imagen 5.5, se puede ver que el intervalo de confianza de la estimación del PIBE *verdadero* para cada uno de los Estados. Lo primero que salta a la vista es que en general la estimación del PIBE *verdadero*, con los parámetros dados por la tabla 5.5 correspondientes al valor de $\phi = 0.93$, se encuentra por debajo del nivel del PIBE medido por el INEGI en las cuentas nacionales. Esto sugiere que en general dos posibilidades. La primera, que la estimación esté sesgada a la baja o, la segunda, que realmente el PIBE se encuentra sobre estimado por el INEGI.

Para descartar la primera consideración, se comparó con los resultados nacionales obtenidos por Mendoza en [39], donde se estimó sólo el PIB a nivel nacional como se vio en la imagen 5.6. En su trabajo, siguiendo la metodología sugerida por Victor M. Guerrero, se realizó una estimación similar a la de este trabajo pero en la que se corrige el posible sesgo de la estimación generada por el parámetro ψ. En el artículo se compararon los resultados para México con toda las serie desde 1992 hasta 2008 a nivel nacional por que para estos años sí se cuenta con datos del PIB en el INEGI. Dado que en los años correspondientes a 2003 a 2008, los que se intersecan para los dos trabajos, su estimación del PIB *verdadero* se encuentra por encima del PIB del INEGI, descartando así la posibilidad de que la sobre estimación corresponda a un sesgo.

Es decir, se piensa que efectivamente el dato que reporta el INEGI para el PIBE y que se puede observar en negro en la imagen 5.5, se encuentra sobre estimado en general.

Imagen 5.7: Serie del PIBE y estimación por Entidad Federativa

Serie del PIB INEGI en negro. Estimación del PIB en rojo. Intervalo de confianza de 2 desviaciones estándar ($k = 2$), linea punteada roja.



Fuente: Elaboración propia con datos de la NOAA [1] y PIBE del INEGI [29].

Específicamente, de la imagen 5.7, se puede notar que hay Estados como Chihuahua, Choahila, México, Nuevo León y Sonora que se encuentran por fuera del intervalo de dos errores estándar de la estimación del PIBE *verdadero*. Más interesante aún es que, Entidades como Campeche y el Distrito Federal, se alejan de la estimación del PIBE *verdadero* sólo para algunos años y no para toda la serie. Esto sugiere por ejemplo en Campeche, que el PIBE para los primeros años de la serie, que coinciden con la caída de la producción petrolera de Cantarell, como entre 2003 y 2007 se encontraba sobre estimado y que, más tarde, esta tendencia se regularizó. Contrariamente en el Distrito Federal, la serie parece estar sobre estimada para los años posteriores a la crisis económica del 2008 y 2009. Esto indica que el INEGI, de una manera optimista, sobre estimó la recuperación ante la caída del PIBE para los años posteriores a la crisis en el Distrito Federal.

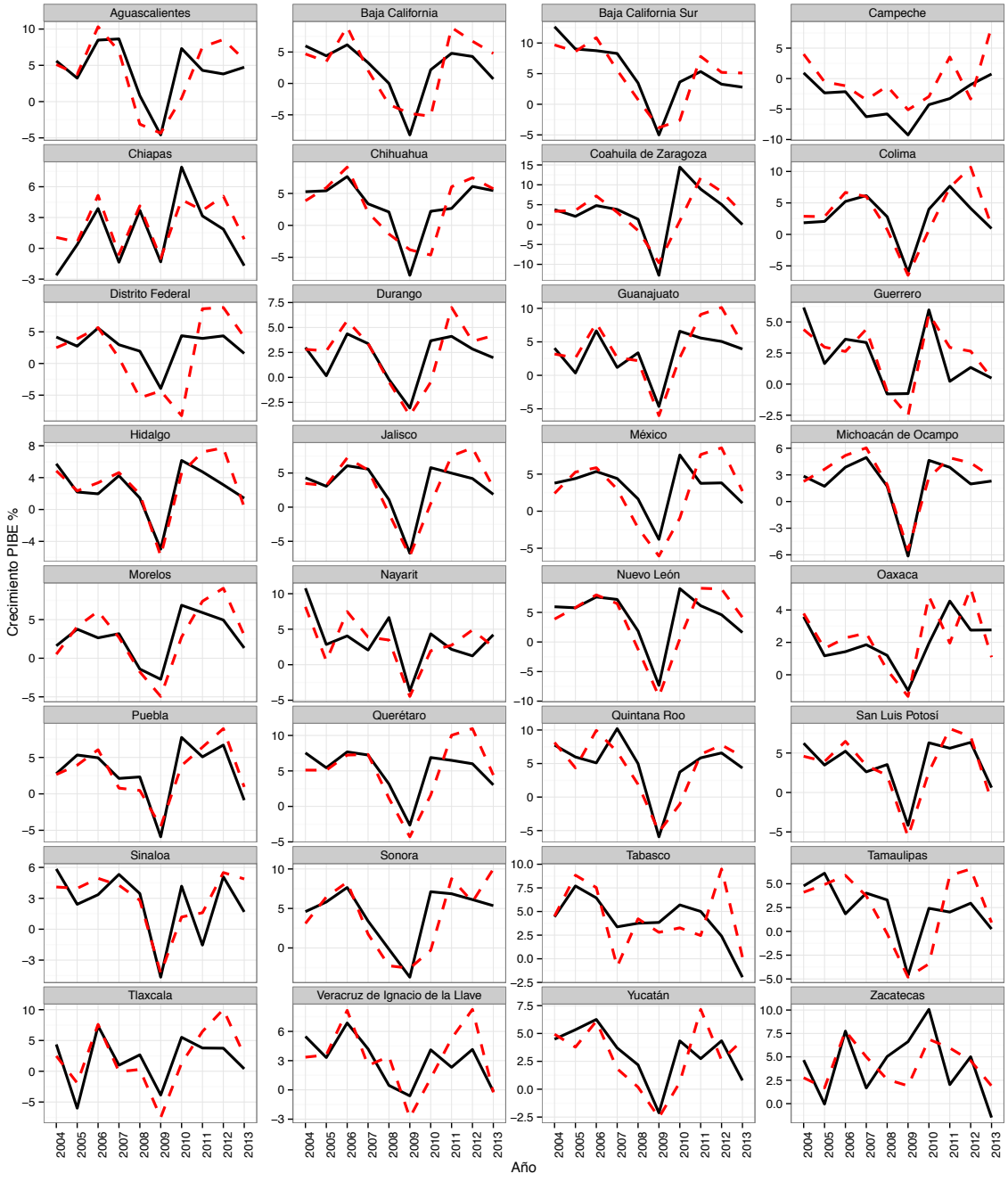
Sin embargo, a pesar de que en general el PIBE se encuentra sobre estimado por el INEGI, la tendencia que sigue la serie estimada del PIBE *verdadero* sigue un patrón similar de crecimiento que el de la serie generada por el INEGI lo cual es deseado dado que el objetivo final de este trabajo es estimar el *verdadero* crecimiento económico de las Entidades Federativas.

5.3. Crecimiento económico *verdadero* en los Estados

Una vez que se estimaron los parámetros adecuados para estimar el *verdadero* PIBE se puede analizar el crecimiento económico. Para esto, se tomaron los crecimientos reales anuales de ambas series, la del INEGI y la estimación del *verdadero* PIBE. Con esto, en la imagen 5.8 a continuación se puede ver el crecimiento anual en cada Entidad Federativa según el INEGI (negro) y según la estimación del *verdadero* PIBE (línea punteada roja). De la imagen se puede apreciar que en general el *verdadero* PIBE presenta crecimientos muy similares a los del PIBE medido por el INEGI. Se puede apreciar la caída correspondiente a la crisis del 2008 – 2009 en las Entidades. Un detalle notable es que para la mayoría de los Estados, el crecimiento del PIBE sobre estima la recuperación en el 2010. Es decir, parece que el INEGI fue algo optimista posiblemente para incentivar la actividad económica. También se puede notar que, en general, la sobre y sub estimación del crecimiento varia mucho con el tiempo. Para esto, se tomaron promedios anuales para todo el periodo 2003 – 2013. Se pueden ver en la tabla 5.6 más adelante.

Imagen 5.8: Crecimiento del PIBE y estimación del PIBE *verdadero* por Entidades

Crecimiento del PIBE INEGI en negro. Estimación del PIBE *verdadero*, linea punteada roja



Fuente: Elaboración propia con datos de la NOAA [1] y PIBE del INEGI [29].

De la tabla 5.6 a continuación se puede apreciar el crecimiento anual del INEGI, el *verdadero* y la diferencia para todos los Estados. La tabla se encuentra ordenada con respecto a la diferencia. De entrada resalta que Campeche es el primero en la lista y que, a diferencia de los demás Estados, es el único que presenta crecimiento negativo promedio para los periodos entre 2003 y 2013. Es decir, la caída del crecimiento promedio *verdadero* en Campeche está sobre estimada con respecto a la caída estimada por el INEGI. Esto siguiere que la caída debería ser un poco menos dramática de lo que siguieren las cuentas nacionales.

Para los demás Estados, los que presentan una diferencia negativa significa que el crecimiento del PIBE del INEGI se encuentra sub estimada. Es decir, debería de ser más alto de lo que presenta el INEGI. Por otro lado, los que tienen una diferencia positiva, tienen un crecimiento que se encuentra sobre estimado. Es decir, el crecimiento reportado debería ser menor.

De primera instancia si se analiza sólo el promedio anual, no es tan alarmante pensar que haya una diferencia de medio punto base (0.5). Sin embargo, para esto se incluye la última columna, en donde se puede ver la implicación que tienen las diferencias anuales en el crecimiento de largo plazo. Es decir, ya no es cualquier cosa que el crecimiento del PIBE, por ejemplo, en el Distrito Federal, se encuentre sobre estimado por el INEGI en 15 % ni, en el otro extremo, que por ejemplo Chiapas, parezca que haya crecido 12 % según los datos cuando en realidad no es así.

De la imagen 5.8 se puede ver que el INEGI hace una buena labor al estimar las cuentas nacionales a nivel Estatal; sin embargo, es un hecho que hay economía no cuantificable de forma que se debe tener cuidado al utilizar las cifras de crecimiento económico. Es por esto que se considera que la tabla 5.6 es un resultado muy importante de esta tesis ya que permite ver dónde están los focos rojos en cuanto al crecimiento del PIB en las Entidades de forma que se puedan tomar mejores decisiones que consideren la actividad económica no observable por el INEGI.

Este trabajo representa una metodología concreta para que se pueda mejorar el cálculo del crecimiento del PIBE específico de cada Entidad Federativa para cuando no se cuenta con datos. Es decir, las imágenes satelitales de iluminación nocturna desde el espacio permiten realizar cálculos incluso a niveles más desagregados, lo que permitiría ampliar el horizonte de exploración y dotar de mayor certeza a las estadísticas nacionales.

Tabla 5.6: Crecimientos anuales promedio en %

Estado	PIBE INEGI	PIBE *verdadero*	Diferencia	Diferencia 03-13
Campeche	-3.27	-0.18	-3.09	-26.05
Chiapas	1.39	2.38	-0.99	-11.97
Guanajuato	3.21	3.90	-0.69	-8.86
Michoacán de Ocampo	2.18	2.83	-0.66	-8.19
Hidalgo	2.60	3.11	-0.51	-6.26
Durango	2.02	2.46	-0.44	-5.07
Colima	2.88	3.28	-0.40	-4.72
Sinaloa	2.51	2.90	-0.38	-5.01
Morelos	2.61	2.88	-0.27	-2.96
Baja California	2.36	2.63	-0.27	-2.78
Veracruz de Ignacio de la Llave	3.02	3.27	-0.25	-2.98
Tlaxcala	1.89	2.13	-0.24	-2.36
Oaxaca	2.03	2.24	-0.22	-2.53
Guerrero	2.12	2.30	-0.18	-2.19
Tabasco	4.07	4.25	-0.18	-2.23
Jalisco	3.02	3.08	-0.06	-0.32
Tamaulipas	2.31	2.35	-0.04	-0.12
Zacatecas	4.13	4.10	0.03	-0.04
Puebla	3.03	2.96	0.06	0.71
Aguascalientes	4.23	4.10	0.13	2.43
Coahuila de Zaragoza	3.14	3.00	0.14	0.98
Chihuahua	3.24	3.03	0.21	2.93
Querétaro	5.09	4.86	0.22	4.10
Yucatán	3.20	2.91	0.30	4.04
Nayarit	3.47	3.12	0.35	4.58
Quintana Roo	4.86	4.50	0.36	5.64
Sonora	4.31	3.91	0.39	6.15
San Luis Potosí	3.58	3.17	0.41	5.85
Baja California Sur	5.22	4.71	0.51	7.94
México	3.21	2.62	0.59	8.24
Nuevo León	4.24	3.64	0.60	9.00
Distrito Federal	2.78	1.66	1.12	14.99

Fuente: Elaboración propia.

Capítulo 6.

Conclusiones

En este capítulo se plantean las conclusiones generales de este trabajo así como algunas consideraciones finales.

Para empezar, se puede concluir de este trabajo que los datos iluminación nocturna satelital visible desde el espacio es una proxy muy útil desde un punto de vista geográfico y temporal. Del capítulo 2 se puede ver que este tipo de datos de Estadística Espacial y su relación con los Sistemas de Información Geográfica se pueden aplicar en distintas metodologías y con mucho alcance. Este trabajo es prueba de ello.

Adicionalmente en el capítulo 2 se presenta una revisión extensa de la literatura relacionada al uso específico de la serie de iluminación nocturna satelital visible del espacio del cual queda claro que la serie se ha utilizado desde hace tiempo y que, sin embargo, no se ha explotado del todo su potencial. Por lo que se piensa que este trabajo representa uno de los primeros que comienzan a alcanzarlo.

Del capítulo 3 se pudo ver que el problema de intercalibración es problema serio que presenta la serie de iluminación nocturna satelital ya que los datos de la serie no son directamente comparables entre distintos años ni entre distintos satélites. Afortunadamente se logró plantear una solución óptima, específica para México y generalizable a cualquier país para resolver este problema. Es por esto que se planteó una metodología completa para poder atacarlo. El impacto de corregir por intercalibración fue muy evidente al momento del modelado en el capítulo 5.

Por otra parte y como complemento al trabajo realizado en este proyecto, en el capítulo

3, se generó un índice de Gini de desigualdad de iluminación nocturna satelital para todos los años de la serie de iluminación correspondiente a cada una de las Entidades Federativas en México. Esta medida puede ser muy útil debido a que no se cuenta con muchos índices de este estilo desagregados a este nivel por lo que se considera un valor agregado importante.

Finalmente se demostró la capacidad que puede tener la serie de iluminación nocturna satelital en el capítulo 5 en donde se explicó cómo se puede utilizar la serie como variable *proxy* para poder estimar el crecimiento económico en las Entidades Federativas de México. Este capítulo desarrolla la metodología completa para poder mezclar la serie de iluminación con el crecimiento económico de forma que se pueda generalizar a cualquier país y cualquier nivel de agregación en las cuentas nacionales. Aunque se piensa que puede haber algunas fuentes de error en el uso de la serie de iluminación, ninguno sugiere que el error pueda estar coorelacionado con el error de medición de las cuentas nacionales por que se piensa que es un modelo apropiado.

Para todos los Estados, la serie de iluminación nocturna satelital puede representar una medida fundamental para estimar el crecimiento a niveles incluso más desagregados o agregados, en donde no se cuenta con datos espaciales detallados.

Evidentemente, la contabilidad nacional es un proceso altamente complejo de cuantificar y una parte de esta complejidad se incrementa en mayor medida al contar con una gran proporción de economía no observada. El análisis que se presentó en este trabajo permite tener una estimación comparable para todos los Estados. Los resultados que se obtuvieron muestran que para el periodo que va de 2003 – 2013, el país tuvo un crecimiento menor al reportado por el INEGI. La sobre estimación del crecimiento promedio anual mexicano para este periodo alcanzó 15 por ciento y de 1.2 por ciento en promedio anual. La sub estimación del crecimiento promedio anual mexicano para este periodo alcanzó 12 % y de 1.0 por ciento en promedio anual. Lo cual tiene fuertes implicaciones económicas.

Es fundamental poder medir de la manera más certera posible los indicadores económicos, ya la mayoría de decisiones de política pública e incluso muchas de carácter privado dependen de ello. Es por esto que se propone el la aplicación de este trabajo para que por medio de la serie de iluminación se pueda aportar a tener mejores estimaciones del crecimiento económico Nacional y Estatal.

Bibliografía

[1] Defense meteorological satellite program (DMSP). `http://ngdc.noaa.gov/eog/index.html`, 2014.

[2] Richard A. Becker and Allan R. Wilks. Maps in s, 1993.

[3] L. Bhandari and K. Roychowdhury. Night lights and economic activity in india: A study using dmsp - ols night time images. *Hanoi Asia - Pacific Advanced Network*, 2011.

[4] Roger S. Bivand, Edzer J. Pebesma, and Virgilio Gómez-Rubio. *Applied Spatial Data Analysis with R*. Springer, Baltimore, MD, USA, 2011.

[5] Victor A. Bloomfield. *Using R for Numerical Analysis in Science and Engineering*. Chapman & Hall/CRC, 2014.

[6] Xi Chen and William D. Nordhaus. Using luminosity data as a proxy for economic statistics. *DOI:1017031108*, Vol. 108(núm. 21):págs. 507–14, May 2011.

[7] Andrew D. Cliff and J. K. Ord. *Spatial Autocorrelation*. Pion Limited, 1973.

[8] Andrew D. Cliff and J. K. Ord. *Spatial Processes: Models & Applications*. Pion Limited, 1981.

[9] Secretaría de Gobernación CONAPO. México en cifras. `http://www.conapo.gob.mx/es/CONAPO/Mexico_en_cifras`, 2015.

[10] Christopher N. H. Doll. Thematic guide to night-time light remote sensing and its applications. *CIESIN Thematic Guide*, 2008.

[11] Christopher N.H. Doll, Jan-Peter Muller, and Christopher D. Elvidge. Night-time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions. *AMBIO*, Vol. 29 núm. 3:págs. 157 a 162, mayo 2000.

[12] Christopher N.H. Doll, Jan Peter Muller, and Jeremy G. Morley. Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, Vol. 57, núm. 1:págs. 75 a 92, 2005.

[13] Steeve Ebener, Christopher Murray, Ajay Tandon, and Christopher C. Elvidge. From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. *International Journal of Health Geographics*, Vol. 4, 2005.

[14] Christopher D. Elvidge, K. E. Baugh, S. J. Anderson, P. C. Sutton, and T. Ghosh. The night light development index (NLDI): a spatially explicit measure of human development from satellite data. *Social Geography*, Vol.7:págs. 23 a 35, 2012.

[15] Christopher D. Elvidge, Kimberly E. Baugh, H. Korel, E. R. Davis, and C. W. Davis. Relation between satellite observed visible near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*, Vol. 18:págs. 1373 a 1379, 1997.

[16] Christopher D. Elvidge, Daniel Ziskin, Kimberly E. Baugh, Benjamin T. Tuttle, Tilottama Ghosh, Dee W. Pack, Edward H. Erwin, and Mikhail Zhizhin. A fifteen year record of global natural gas flaring derived from satellite data. *Energies*, Vol. 2(núm. 3):págs. 595 a 622, 2009.

[17] Tilottama Ghosh, Sharolyn Anderson, Rebecca L. Powell, Paul C. Sutton, and Christopher D. Elvidge. Estimation of Mexico's informal economy and remittances using nighttime imagery. *Remote Sensing*, Vol. 1:págs. 418 a 444, agosto 2009.

[18] Tilottama Ghosh, Rebecca L. Powell, Christopher D. Elvidge, Kimberly E. Baugh, Paul C. Sutton, and Sharolyn Anderson. Shedding light on the global distribution of economic activity. *The Open Geography Journal*, Vol. 3:págs 147 a 160, 2010.

[19] Tilmann Gneiting, Marc G. Genton, and Peter Guttorp. Geostatistical space-time models, stationarity, separability and full symmetry. *Department of Statistics University of Washington*, Technical Report no. 475.

[20] David F. Good. The economic lag of central and eastern europe: Income estimates for the habsburg successor states, 1870 - 1910. *Journal of Economic history*, Vol. 54 núm. 4:págs. 869 a 91, 1994.

[21] P. Goovaerts. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, 1997.

[22] Victor M. Guerrero Guzmán. *Análisis estadístico y pronóstico de series de tiempo económicas*. Jit Press, México, D.F., 3º edition, 2009.

[23] Vernon Henderson, Adam Storeygard, and David N. Weil. A bright idea for measuring economic growth. *American Economic Review*, Vol. 101(Núm. 3):págs. 194 a 199, 2011.

[24] Vernon Henderson, Adam Storeygard, and David N. Weil. Measuring economic growth from outer space. *American Economic Review*, Vol. 102(Núm. 2):págs. 994 a 1028, 2012.

[25] T. Hengl. *A Practical Guide to Geostatistical Mapping of Environmental Variables*. Office for Official Publications of the European Communities, Luxemburgo, septiembre 2007.

[26] Marek Hlavac. stargazer: LaTeX/HTML code and ASCII text for well-formatted regression and summary statistics tables. R package version 5.1. `http://CRAN.R-project.org/package=stargazer`, 2014.

[27] Qingxu Huang, Xi Yang, Bin Gao, Yang Yang, and Yuanyuan Zhao. Application of DMSP/OLS nighttime light images: A meta-analysis and a systematic literature review. *Remote Sensing*, Vol. 6:págs. 6844 a 6866, 2014.

[28] Instituto Nacional de Estadística y Geografía INEGI. Marco geoestadístico nacional. `http://www.inegi.org.mx/geo/contenidos/geoestadistica/m_geoestadistico.aspx`, 2010.

[29] Instituto Nacional de Estadística y Geografía INEGI. PIB - entidad federativa, anual. `http://www.inegi.org.mx/est/contenidos/proyectos/cn/pibe/`, 2015.

[30] R. Kulkarni, K. Haynes, R. Stough, and J. Riggle. Revisiting night-lights as proxy for economic growth: A multi-year light based growth indicator (LGBI) for China, India and the U.S. *GMU School of Public Policy - Research Paper*, 2011.

[31] P. C. Kyriakidis and A. G. Journel. Geostatistical space-time models: A review. *Mathematical Geology*, Vol. 31(núm. 6), 1999.

[32] X. Li, X. Chen, Y. Zhao, J. Xu, F. Chen, and H. Li. Automatic intercalibration of night-time light imagery using robust regression. *International Journal of Remote Sensing*, (núm. 4):págs. 45 a 54, 2013.

[33] Z. Liu, C. He, Q. Zhang, Q. Huang, and Y. Yang. Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008. *Landsc. Urban Plan*, (106):págs. 62 a 72, 2012.

[34] Paul A. Longley, Michael F. Goodchild, David J. Maguire, and Rhind David W. *Geographical Information Systems: Principles, Techniques, Applications and Management*. Abridged, 2 edition, 2005.

[35] Matt Lowe. Night lights and ArcGIS: A brief guide. *Economic Journal*, 2014.

[36] Banco Mundial. Datos. `http://datos.bancomundial.org`, 2015.

[37] National Geophysical Data Center NGDC. National oceanic and atmospheric administration (NOAA). `http://ngdc.noaa.gov/ngdc.html`, 2014.

[38] Petróleos Mexicanos PEMEX. Base de datos institucional. `http://ebdi.pemex.com/bdi/bdiController.do?action=temas`, 2015.

[39] Juan Antonio Mendoza Perea. ¿crecemos mÁs de lo que creemos? anÁlisis sobre el crecimiento econÓmico de mÉxico visto desde el espacio. *Tesis de licenciatura en economía, ITAM*, 2014. Asesor: Victor M. Guerrero Guzmán.

[40] Nicola Pestalozzi. Nighttime lights as proxy for the spatial growth of dense urbanized areas. *Department of Management, Technology and Economics (D-MTEC) Chair of Entrepreneurial Risks, Chair of Land Use Engineering ETH Zürich*, Noviembre 2012.

[41] GNU project. LaTeX - a document preparation system. http://www.latex-project.org.

[42] GNU project. The R project for statistical computing. http://www.r-project.org/about.html.

[43] RStudio. http://www.rstudio.com.

[44] Björn Sandvik. World borders dataset. http://thematicmapping.org/downloads/world_borders.php.

[45] Paul. C. Sutton, Christopher D. Elvidge, and Tiltottama Ghosh. Estimation of gross domestic product at sub national scales using nighttime satellite imagery. *International Journal of Ecological Economics and Statistics (IJEES)*, Vol. 8(núm. 07):págs. 5–21, 2007.

[46] O. Tange. GNU Parallel - the command line power tool. *The USENIX Magazine*, Vol. 36(núm.1):págs. 42 a 47, Feb 2011.

[47] W. R. Tobler. A computer model simulation of urban growth in the detroit region. *Economic Geography*, 46(2):págs. 234 a 240, 1970.

[48] Diego Valle-Jones. Descargar shapefiles de manazanas, AGEB, municipios, estados y mucho más del inegi. http://blog.diegovalle.net/2013/06/shapefiles-of-mexico-agebs-manzanas-etc.html.

[49] Hadley Wickham. *ggplot2: Elegant graphics for data analysis*. Springer, 2009.

[50] J. Wu, S. He, J. Peng, W. Li, and X. Zhong. Intercalibration of DMSP-OLS night-time light data by the invariant region method. *International Journal of Remote Sensing*, (34):págs. 7356 a 7368, 2013.

[51] Naizhuo Zhao, Nate Currit, and Eric Samson. Net primary production and gross domestic product in China derived from satellite imagery. *Ecological Economics*, Vol. 70(núm. 5):págs. 921 a 928, 2011.

Apéndice A.

Software

En este trabajo se utiliza el lenguaje `R Project for Statistical Computing`[1] [2]

Para resumir la información de los datos de forma gráfica se usa el paquete `ggplot2` [49] en R creado por Wickham para mostrar las imágenes y evaluaciones de los resultados de los análisis usando una herramienta sencilla y flexible.

Para la exploración de las imágenes satelitales se puede utilizar QGIS. Adicionalmente, dado que las imágenes requieren de mucho poder de cómputo, se utilizó el paquete Parallel [46] para poder distribuir procesos.

En el desarrollo y escritura de esta tesis se empleó el software libre RStudio [43] y LaTeX- A document preparation system [41].

[1] "R es un software libre para un ambiente de computo estadístico y gráficas. Compila y ejecuta en un amplia variedad de plataformas de UNIX, Windows y MacOS." Cita disponible en [42]

[2] Ver [5] para una guía de R.

Apéndice B.

Descarga de la base

La base de datos se puede descargar en la página de la NOAA, ver [1]. En este trabajo se utilizó la base de DMPS-OLS porque es la base que contiene más años, de 1992 a 2012. Para algunos años contiene información de más de un satélite. Como se mencionó, cada archivo contiene dos sets de archivos comprimidos para cada satélite y año "Average Visible, Stable Lights", "Cloud Free Coverages" y "Average Lights x Pct".

B.1. Descarga en paralelo

En este proyecto se programó un script de descarga. Se probó en el sistema operativo *Mac OS X - Mavericks (10.10).* Antes de ejecutar cualquier script en este proyecto es necesario instalar algunos paquetes básicos.

Con los paquetes básicos instalados se puede proceder a la descarga (esto descarga toda la base de datos de la NOAA de DMPS-OLS, cerca de 100GB).

Código B.1: Código de descarga

```
#!/bin/bash

rm -rf base-cruda-v4/
mkdir base-cruda-v4
cd base-cruda-v4/

rm satelites.urls

for satelite in "F101992" "F101993" "F101994" "F121994" "F121995"\
   "F121996" "F121997" "F121998" "F121999" "F141997"\
   "F141998" "F141999" "F142000" "F142001" "F142002"\
   "F142003" "F152000" "F152001" "F152002" "F152003"\
   "F152004" "F152005" "F152006" "F152007" "F152008"\
   "F162004" "F162005" "F162006" "F162007" "F162008"\
   "F162009" "F182010" "F182011" "F182012" "F182013"
do

echo "http://ngdc.noaa.gov/eog/data/web_data/v4composites/$satelite.v4.tar"\
 >> satelites.urls
done

cat satelites.urls | parallel curl -O
```

B.2. Descomprimir en paralelo

El siguiente código descomprime toda la base de DMSP-OLS.

Código B.2: Descompresión

```
#!/bin/bash

# Creo carpeta de salida para gzips y tifs
mkdir base-cruda-v4-tifs

# Inicio el proceso
cd base-cruda-v4

# Descomprimo en paralello
ls *.tar | parallel -j +0 --eta 'tar -xvzf {} -C ../base-cruda-v4-tifs/'

cd ../base-cruda-v4-tifs

ls *.gz | parallel -j+0 'gzip -dv {} '

# Ordeno un poco por el tipo de raster

mkdir average
mkdir cloud-free
mkdir stable

mv ls *web.avg*  average/
mv ls *web.cf*  cloud-free/
mv ls *web.stable*  stable/

# Es importante checar que diferencia hay en v4b con c y d.
# c y d solo sucede en 2010 - 2013
rename 's/.v4[b|c|d]_web.stable_lights.avg_vis././' stable/*
rename 's/.v4[b|c|d]_web.cf_cvg././' cloud-free/*
rename 's/.v4[b|c|d]_web.avg_vis././' average/*
```

B.3. Descarga del censo económico, INEGI

EL código a continuación (ver Valle-Jones [48]) sirve para descargar shapefiles de manazanas, AGEB, municipios, estados y mucho más del INEGI [28].

Código B.3: Código de descarga INEGI

```
#! /bin/bash

# Para hacer compatibles las coordenadas
PROJECCION="+proj=longlat +ellps=WGS84 +no_defs +towgs84=0,0,0"
WGET="wget -w 5 --random-wait --tries=10 "

declare -a files=("ageb_urb" "eje_vial"
                  "estatal" "loc_rur" "loc_urb"
                  "manzanas" "municipal" "servicios_a"
                  "servicios_l" "servicios_p")
 declare -a estados=("ags" "bc" "bcs" "camp" "coah"
                     "col" "chis" "chih" "df" "dgo"
                     "gto" "gro" "hgo" "jal" "mex"
                     "mich" "mor" "nay" "nl" "oax"
                     "pue" "qro" "qroo" "slp" "sin"
                     "son" "tab" "tamps" "tlax"
                     "ver" "yuc" "zac");
function reproject {
  for i in "${files[@]}"
  do
    ./ogr2ogr.py $1/$2_$i.shp "$1"/$i.shp -t_srs "$PROJECCION"
    rm "$1"/$i*
  done
  cd $1/tablas
  rename "s/^cpv2010/$2_cpv2010/" cpv2010*
  rm -rf cpv2010*
  cd ../../..
}
for i in {1..32}
do
   if [ $i -lt 10 ]
   then
     FILENUM="0$i"
   else
     FILENUM="$i"
   fi
   $WGET "http://www.inegi.org.mx/sistemas/consulta_resultados/
   scince2010.aspx?_file=/sistemas/consulta_resultados/scince2010/
   Scince2010_$FILENUM.exe&idusr=80085" -O ${estados[$i-1]}_scince.exe
   innoextract --lowercase --silent ${estados[$i-1]}_scince.exe
   mkdir -p shps/${estados[$i-1]}
   cp -r app/$FILENUM/* shps/${estados[$i-1]}
   rm -rf app
   rm -rf tmp
   rm -rf ${estdos[$i - 1]}_scince.exe
   reproject shps/${estados[$i - 1]} ${estados[$i - 1]}
   sleep 20
done
```

Apéndice C.
Cortar imágenes satelitales

El siguiente código corta los satelites utilizando como base para recortar un archivo tipo Shape de todos los municipios de México.

Código C.1: Cortar satelites

```
#!/bin/bash

rm -R sat_cortados_mexico

mkdir sat_cortados_mexico

for satelite in "F101992" "F101993" "F101994" "F121994" "F121995"\
   "F121996" "F121997" "F121998" "F121999" "F141997"\
   "F141998" "F141999" "F142000" "F142001" "F142002"\
   "F142003" "F152000" "F152001" "F152002" "F152003"\
   "F152004" "F152005" "F152006" "F152007" "F152008"\
   "F162004" "F162005" "F162006" "F162007" "F162008"\
   "F162009" "F182010" "F182011" "F182012" "F182013"
do

        mun='inegi_mun_agregado/agregado_municipio_claves.shp'
        #echo $ageb

        imagen='base-cruda-v4-tifs/stable/'$satelite'.tif'
        #echo $imagen

        archivo='sat_cortados_mexico/'$satelite'_mexico.tif'

        echo $satelite
        gdalwarp -q -cutline $mun -crop_to_cutline $imagen $archivo
done
```

Código C.2: Filtrar satelites

```
rm(list=ls());gc();
source('utils.R')
carga_paquetes(); library(doMC); registerDoMC(4)

muns <- readOGR(dsn = '../inegi_mun_agregado',
                layer='agregado_municipio_claves')
crs_muns <- CRS(proj4string(muns))
muns_f <- fortify(muns, region='CVEGEO')
sats <- dir('../sat_cortados_mexico/',
            pattern = '*.tif', full.names = T )
names(sats)  <- substr(dir('../sat_cortados_mexico/',
                           pattern = '*.tif'),
                       start = 1,  stop = 7)

lista_mun <- unique(as.character(muns$CVEGEO))
names(lista_mun) <- unique(as.character(muns$CVEGEO))

df_rasters <- ldply(.data = sats, .id = 'satelite',
                    .fun = function(f){
                      x <- raster( f )
                      print(substr(f, start=25, stop=31))
                      x[x==255]=NA
                      sat <- data.frame(rasterToPoints(x))
                      colnames(sat) <- c('long', 'lat', 'dn')
                      saveRDS( sal <- ldply(.data = lista_mun, .parallel=T,
                            .id = 'CVEGEO', function(mun){
                              print(mun)
                              df <- filter(muns_f, id==mun)
                              sat_filtro <- filter(sat,long >=min(df$long) &
                                                   long <=max(df$long) &
                                                   lat >=min(df$lat) &
                                                   lat <=max(df$lat) )
                              sat_filtro_sp <-
                                SpatialPointsDataFrame(coords =
                                                       sat_filtro[,c(1,2)],
                                                       proj4string = crs_muns,
                                                       data = sat_filtro %>%
                                                       dplyr::select(dn) )
                              int <- gIntersects( sat_filtro_sp,
                                                  muns[muns$CVEGEO==mun,],
                                                  byid = T)
                              ind <- which(apply(int==T, MARGIN = 2, any))
                              salida <-na.omit(data.frame(sat_filtro_sp[ind,]))
                              salida
                            })
    , file=paste0('sat_join_mun/',substr(f, start=25, stop=31),'.rds') )
  sal
})
saveRDS(df_rasters, file='dn_municipio_raster.rds')
```

Apéndice D.
Modelo de intercalibración

Tabla D.1: Modelo de intercalibración

	Variable dependiente:
	log(dn_base + 1)
	(1)
sateliteF101992	-1.640^{***}
	(0.022)
sateliteF101993	-1.713^{***}
	(0.017)
sateliteF101994	-1.850^{***}
	(0.016)
sateliteF121994	-1.864^{***}
	(0.021)
sateliteF121995	-2.178^{***}
	(0.018)
sateliteF121996	-1.732^{***}
	(0.017)
sateliteF121997	-1.597^{***}
	(0.019)

Tabla D.1 Modelo de intercalibración ***(continuación)***

	Variable dependiente:
	log(dn_base + 1)
	(1)
sateliteF121998	-1.996^{***}
	(0.018)
sateliteF121999	-2.080^{***}
	(0.018)
sateliteF141997	-1.385^{***}
	(0.016)
sateliteF141998	-1.343^{***}
	(0.014)
sateliteF141999	-1.516^{***}
	(0.015)
sateliteF142000	-1.649^{***}
	(0.014)
sateliteF142001	-1.531^{***}
	(0.014)
sateliteF142002	-1.439^{***}
	(0.016)
sateliteF142003	-1.549^{***}
	(0.015)
sateliteF152000	-2.263^{***}
	(0.017)
sateliteF152001	-2.099^{***}
	(0.017)
sateliteF152002	-2.148^{***}
	(0.017)
sateliteF152003	-1.470^{***}
	(0.015)
sateliteF152004	-1.388^{***}

Tabla D.1 Modelo de intercalibración ***(continuación)***

	Variable dependiente:
	log(dn_base + 1)
	(1)
	(0.012)
sateliteF152005	−1.333***
	(0.012)
sateliteF152006	−1.286***
	(0.012)
sateliteF152007	−1.331***
	(0.013)
sateliteF162004	−1.785***
	(0.016)
sateliteF162005	−1.563***
	(0.018)
sateliteF162006	−1.748***
	(0.015)
sateliteF162007	−1.930***
	(0.015)
sateliteF162008	−1.749***
	(0.015)
sateliteF162009	−1.462***
	(0.015)
sateliteF182010	−2.044***
	(0.016)
sateliteF182011	−1.789***
	(0.016)
sateliteF182012	−2.091***
	(0.018)
sateliteF182013	−1.877***
	(0.018)

Tabla D.1 Modelo de intercalibración ***(continuación)***

	Variable dependiente:
	log(dn_base + 1)
	(1)
sateliteF101992:log(dn + 1)	1.562***
	(0.010)
sateliteF101993:log(dn + 1)	1.524***
	(0.009)
sateliteF101994:log(dn + 1)	1.573***
	(0.008)
sateliteF121994:log(dn + 1)	1.567***
	(0.010)
sateliteF121995:log(dn + 1)	1.643***
	(0.009)
sateliteF121996:log(dn + 1)	1.508***
	(0.008)
sateliteF121997:log(dn + 1)	1.472***
	(0.009)
sateliteF121998:log(dn + 1)	1.576***
	(0.009)
sateliteF121999:log(dn + 1)	1.582***
	(0.008)
sateliteF141997:log(dn + 1)	1.494***
	(0.008)
sateliteF141998:log(dn + 1)	1.471***
	(0.008)
sateliteF141999:log(dn + 1)	1.493***
	(0.008)
sateliteF142000:log(dn + 1)	1.532***
	(0.008)

Tabla D.1 Modelo de intercalibración ***(continuación)***

	Variable dependiente:
	log(dn_base + 1)
	(1)
sateliteF142001:log(dn + 1)	1.485***
	(0.008)
sateliteF142002:log(dn + 1)	1.484***
	(0.008)
sateliteF142003:log(dn + 1)	1.469***
	(0.008)
sateliteF152000:log(dn + 1)	1.656***
	(0.008)
sateliteF152001:log(dn + 1)	1.597***
	(0.008)
sateliteF152002:log(dn + 1)	1.628***
	(0.008)
sateliteF152003:log(dn + 1)	1.537***
	(0.008)
sateliteF152004:log(dn + 1)	1.498***
	(0.007)
sateliteF152005:log(dn + 1)	1.461***
	(0.007)
sateliteF152006:log(dn + 1)	1.445***
	(0.007)
sateliteF152007:log(dn + 1)	1.479***
	(0.007)
sateliteF162004:log(dn + 1)	1.537***
	(0.008)
sateliteF162005:log(dn + 1)	1.539***
	(0.009)
sateliteF162006:log(dn + 1)	1.576***

Tabla D.1 Modelo de intercalibración ***(continuación)***

	Variable dependiente:
	log(dn_base + 1)
	(1)
	(0.008)
sateliteF162007:log(dn + 1)	1.550***
	(0.007)
sateliteF162008:log(dn + 1)	1.484***
	(0.007)
sateliteF162009:log(dn + 1)	1.400***
	(0.008)
sateliteF182010:log(dn + 1)	1.464***
	(0.007)
sateliteF182011:log(dn + 1)	1.428***
	(0.008)
sateliteF182012:log(dn + 1)	1.530***
	(0.008)
sateliteF182013:log(dn + 1)	1.488***
	(0.008)
N	280,371
R^2	0.927
Adjusted R^2	0.927
Error Std.	0.406 (df = 280303)

*p<0.1; **p<0.05; ***p<0.01