



El Periplo Sustentable

Universidad Autónoma del Estado
de México

<http://rperiplo.uaemex.mx/>

ISSN: 1870-9036

Publicación Semestral

Número: 36

Enero / Junio 2019

Artículo

Título

Búsquedas en Internet y su
influencia en los flujos de
turistas y visitantes.
El caso del avistamiento de
luciérnagas en
Nanacamilpa Tlaxcala.

Autor:

Ignacio Ibarra López
Adriana Montserrat Pérez Serrano
María del Pilar Cuecuecha Mendoza

Fecha Recepción:

09/01/2018

Fecha Reenvío:

26/06/2018

Fecha Aceptación:

27/11/2018

Páginas:

402- 431

Búsquedas en Internet y su influencia en los flujos de turistas y visitantes. El caso del avistamiento de luciérnagas en Nanacamilpa Tlaxcala.

Searches on the Internet and its influence on the flows of tourists and visitors. The case of the sighting of fireflies in Nanacamilpa Tlaxcala.

Resumen

En este trabajo el principal objetivo es el de identificar patrones de comportamiento de los turistas ocasionados por el uso de internet. Se analiza la forma en cómo las búsquedas de ciertas palabras en internet impulsan una respuesta en el flujo de turistas y visitantes domésticos que acuden al estado de Tlaxcala en México. La metodología empleada combina información en gran escala de internet (Big Data) y su análisis cuantitativo mediante un modelo de Vectores Autorregresivos (VAR). La principal conclusión es que existe un impacto diferenciado de acuerdo a la palabra que se busca en internet. El factor que atrae el turismo es la disponibilidad de cuartos, mientras que para el caso de las visitas es la existencia del avistamiento.

Palabras clave:

internet, flujo turístico, flujo de visitantes, información de gran escala.

Abstract

In this work the main objective is to identify patterns of behavior of tourists caused by the use of the internet. The way in which the searches of certain words on the internet drive a response in the flow of domestic tourists that come to the state of Tlaxcala in Mexico is analyzed. The methodology used is a quantitative analysis of large-scale information on the Internet (Big Data), using a Time Series autoregressive vector (VAR) model. The main conclusion is that there is a differentiated impact according to the word that is searched on the Internet. The factor that attracts tourism is the availability of rooms, while for the case of visits is the existence of sighting.

Keywords:

internet, tourism flow, Big Data

De los AUTORES

Ignacio Ibarra López

Doctor en Economía, Catedrático de El Colegio de Tlaxcala. Líneas de Investigación: Política Pública, Economía de las familias.

gnacio.ibarra@coltlax.edu.mx

Adriana Montserrat Pérez Serrano

Profesora - Investigadora de El Colegio de Tlaxcala., Tlaxcala.

María del Pilar Cuecuecha Mendoza

Profesora - Investigadora de El Colegio de Tlaxcala., Tlaxcala.

Introducción

Dentro de los problemas relacionados a la política pública en turismo, resulta fundamental conocer el flujo de turistas. Adicionalmente, es de interés conocer el comportamiento de las personas que deciden hacer turismo. Hasta hace algunos años, obtener información detallada tanto del flujo como del comportamiento implicaba un proceso complicado. Se han logrado avances mediante la utilización de datos de internet y datos de gran escala (Big Data) tal cual se ha presentado en diferentes trabajos (e.g. Yang *et al*, 2014; Gunter *et al*, 2016; Gawlik, *et al*, 2011; Bangwayo-Skeete y Skeete, 2015).

Para el caso de México son pocos los estudios en los que se utiliza el uso de información de internet (Lichtle *et al*, 2014) o Big Data (Domínguez *et al*, 2016; 2016a). En Domínguez *et al* (2016; 2016a) se analizan los flujos de turistas (y algunos patrones de consumo) con datos de las transacciones económicas realizadas con tarjetas de crédito débito. Dicha información por su carácter no es pública y este es el principal inconveniente de la propuesta. En cuanto al trabajo de Lichtle *et al* (2014), se estudian patrones de conducta de turistas hacia ciertos destinos del país (e.g. Querétaro, Puebla) y con datos de Twitter. En este caso, si bien se incluye información de gran escala que es parcialmente pública (Twitter restringe parte de su información), no se incorporan otros medios de interacción (e.g. Facebook, Youtube, Instagram), así como las búsquedas realizadas en internet¹.

En la mayoría de los trabajos generados hasta donde se conoce, se ha buscado pronosticar el flujo de turistas con datos de internet. Sin embargo, no se ha utilizado esta información para explicar la causalidad entre las búsquedas de internet y el flujo de turistas. En el presente trabajo se ha desarrollado una aplicación metodológica que explora este tipo de relaciones

1 El estudio utilizó una muestra de twits (7,955 usuarios que generan 827,424 twits) y en un período temporal de tres días (1, 2 y 3 de febrero de 2014)



causales, tomando como caso el flujo de turistas domésticos hacia el municipio de Nanacamilpa en el estado de Tlaxcala. Este lugar resulta de interés debido a que partir de 2012 se observa un importante crecimiento del flujo de turistas derivado de una mayor difusión del avistamiento de luciérnagas. En términos geográficos se identifica que la zona de avistamiento es una zona boscosa que comparte con el municipio de Tlahuapan (en Puebla). La particularidad de este ecosistema es que resulta propicio para la reproducción de una especie luminiscente de luciérnaga identificada como *Macrolampis palaciosi* (Zaragoza, 2015). Los avistamientos se realizan entre los meses de junio, julio y agosto generando un proceso estacional.

El principal objetivo del trabajo es el de explicar como la decisión de viajar a un destino en particular, se asocia con las búsquedas de ciertas palabras clave que identifican las personas. En términos metodológicos, se utiliza información de gran escala proveniente de un motor de búsqueda como “Google Trends”, datos del Instituto Nacional de Antropología e Historia, (INAH), datos del sistema de información Datatur y un modelo de Vectores Autorregresivos (VAR).

La premisa de la que se parte es que las personas que deciden viajar al avistamiento, buscaron información en internet para informarse sobre su viaje. Al mismo tiempo, las visitas al lugar de avistamiento, generan interacción en redes sociales y en la convivencia entre personas que motiva a su vez la decisión de otras personas de viajar hacia el avistamiento. Por lo tanto se plantea una doble causalidad entre las búsquedas de internet y el flujo de turistas.

1. Revisión de la literatura.

Los usos de internet trascienden la simple búsqueda de información y la generación de procesos de comunicación entre individuos. Una de las principales aplicaciones de internet es su capacidad para modelar mercados y generar pronósticos sobre precios. Según Engle (2000), la información que utilizan los modelos que estudian el comportamiento económico ha experimentado un cambio transcendental al pasar de datos anuales a datos mensuales, semanales, diarios y datos intradía (e.g. por hora, minuto o incluso segundo en un mismo día). Esta información se denomina de “alta frecuencia” y tiene como ventajas el aumentar la precisión de generar pronósticos y capacidad explicativa de las variables.



Con la penetración de internet en prácticamente todas las actividades humanas, comenzó a desarrollarse el concepto Big Data el cual puede ser ambiguo. Retomando a Ward *et al* (2013), a pesar de las diferencias en definiciones, el concepto se vincula con: 1) el tamaño de la información (bases de datos muy grandes); 2) la complejidad de la información (se presenta información muy detallada de todas y cada una de las observaciones) y 3) las herramientas y técnicas que se utilizan para analizar (se requiere de algoritmos y técnicas avanzadas de estadística para su procesamiento).

Por lo tanto, la integración de Big Data permite conocer con mayor precisión la conducta de las personas al analizar patrones agregados con una cantidad importante de observaciones (Da, Engelberg y Gao, 2011; Joseph, Wintoki y Zhang, 2011).

En el caso del turismo, se reconoce la importancia que tiene el internet, como un espacio donde los turistas obtienen información, pero al mismo tiempo influyen en los demás al compartir sus experiencias y recomendaciones de viajes (Machado, 2014). En un entorno con una alta interacción entre personas y con una rápida transmisión de información, el proceso de decisión del turista podría ser sensible a variables que cambian casi de forma instantánea (e.g. el tipo de cambio, los avisos sobre seguridad en las embajadas o el clima). Es por esta razón que han comenzado a aparecer con mayor frecuencia artículos donde se relaciona la actividad turística con la conducta de las personas en internet. A continuación se analizan algunos de los principales trabajos que se han podido identificar en la literatura reciente.

Dentro de los principales autores que comenzaron a explorar la aplicación de internet en materia turística se encuentran Choi *et al* (2009; 2009a; 2012) donde utilizando datos de Google, pronostican en el corto plazo valores de variables tales como ventas de automóviles, reclamos de seguros al desempleo, confianza del consumidor y planeación de viajes. El principio del que parten estos estudios es que cuando las personas realizan búsquedas en internet, revelan sus preferencias², así

2 Este problema es de particular interés para los economistas (en particular aquellos de corte neoclásico), debido a que los modelos que se plantean requieren de supuesto sobre la estabilidad de las preferencias. Anteriormente no era posible identificar estas preferencias de forma tan clara como ahora se puede hacer con los datos de internet.



como las interconexiones entre los diferentes conceptos que las delimitan. En lugar de hablar de predicciones a futuro (forecasting) indican que los datos de internet deben ser utilizados para realizar predicciones contemporáneas (nowcasting). En otras palabras: se deben utilizar los patrones de internet para analizar el comportamiento en tiempo real de las personas.

Dentro del trabajo existe una sección dedicada exclusivamente al análisis del internet y la planeación de viajes. Comparan los datos oficiales de visitas a Hong Kong recabados por la agencia oficial en temas de turismo (Hong Kong Tourism Board) con datos de Google Trends. Lo que encuentran es que el flujo de turistas de acuerdo a su nacionalidad (Estados Unidos, Canadá, Reino Unido, Alemania, Francia, Italia, Australia, Japón e India) coincide con el patrón del flujo de consultas a Google Trends en la mayoría de los casos (solo los turistas de Japón tienen un patrón diferente de comportamiento). En términos metodológicos, este trabajo recurre a modelos de series de tiempo que son muy comunes en el análisis de variables financieras y macroeconómicas.

El aporte de Choi *et al* ha impulsado el desarrollo de trabajos posteriores como el de Yang *et al* (2014) donde se analizan registros de hospedaje e información de internet para el caso de Charleston, South Carolina en Estados Unidos. La aportación adicional del trabajo es que busca identificar qué tipo de modelo es el mejor para trabajar con este tipo de información. En términos generales, compara diferentes modelos de series de tiempo tales como Autorregresivos con Media Móvil (ARMA), con modelo Autorregresivo con Media Móvil y Variable Exógena (ARMAX) donde se tiene información de internet. En el estudio, se concluye que los modelos ARMAX son mejores que los modelos ARMA, debido que permiten conocer y controlar estadísticamente las perturbaciones que existían en los primeros períodos del proceso.

En cuanto al caso de Europa, se pueden citar los trabajos de Artola *et al* (2015) y Claveria *et al* (2010) donde también utilizan datos de internet para la estimación de flujos de turistas retomando la idea de Choi *et al*, pero con datos de España. En estos trabajos se tienen modelos y supuestos muy parecidos a los expuestos anteriormente, demostrando también que las búsquedas de internet se relacionan estrechamente con el flujo turístico.



Por su parte, Gunter *et al* (2016) analizan pronósticos con datos de internet para diferentes países (Alemania, Italia, Japón, Reino Unido y Estados Unidos). En este trabajo en particular, se introduce una innovación al incluir una comparación de los pronósticos generados a partir de filtros como el Holt–Winters (HW). Este filtro es una herramienta estadística ampliamente utilizada para la generación de pronósticos en materia económica (e.g. la estimación del Producto Interno Bruto potencial). En el trabajo, se esperaba que la introducción de este filtro produjera mejores estimaciones de los flujos turísticos. Sin embargo, se encontró que la estimación generada por el filtro HW, es superada por la estimación de un modelo donde se incorpora información de internet para estimar el flujo de turistas. Con ello, se prueba nuevamente que el uso de datos de internet es una mejor alternativa.

En Gawlik, *et al* (2011) se introduce otra innovación: en lugar de modelar el flujo de turistas con información agregada oficial, se utilizan palabras específicas. Se retoma el caso de Hong Kong en los años 2005-2010 (datos mensuales) para desarrollar un algoritmo que demuestra tener un bajo nivel de error de pronóstico. La conclusión es que al utilizar información específica de búsquedas de internet, es posible reproducir el flujo de turistas.

En los artículos que se han comentado, una situación recurrente es que se usa información de Google. ¿Los hallazgos serán similares si se eligiera cualquier otra fuente de información? En el artículo de Yang *et al* (2015) se busca dar respuesta a esta interrogante. Utilizando datos de China (provincia de Hainan) y los modelos propuestos en un artículo anterior (Yang *et al*, 2014) se comparan las estimaciones del flujo de turistas realizadas con dos diferentes motores de búsqueda: Google y Baidu. Los modelos generados a partir de estas fuentes se comparan con las estimaciones de un Modelo Autorregresivo y de Media Móvil (ARMA). La principal conclusión es que se verifica que la utilización de motores de búsqueda reduce los errores de pronóstico y por tanto es mejor que la utilización de cualquier otro modelo. Sin embargo, se encuentra que existen diferencias de acuerdo a la fuente de los datos de internet. Por ejemplo, la información de Baidu permite generar mejores modelos en comparación con la información de Google. La explicación es que probablemente el turismo a Hainan es principalmente doméstico y por ello el motor de búsqueda Baidu (que concentra más información de China) se ajusta mejor a los datos.



En el caso de América Latina, se encuentra el trabajo de Bangwayo-Skeete y Skeete (2015) donde se analiza de forma particular la llegada de turistas a diferentes destinos en el Caribe. Para ello, se construye un indicador que pronostica la demanda a partir de datos de internet. En este caso utilizan un enfoque como el de Gawlik, *et al* (2011), donde primero se seleccionan palabras clave. El estudio se concentra en palabras como “hoteles y vuelos” para desarrollar pronósticos a partir de diferentes modelos estadísticos tales como un modelo Modelo Autorregresivo con Muestra de Información Combinada (AR-MIDAS), un Modelo Autorregresivo con Media Móvil Integrado y Estacional (SARIMA) o un Modelo Autorregresivo (AR) más simple. En el trabajo se demuestra que al utilizar como insumo las consultas de internet, se obtiene una alta capacidad de pronosticar correctamente el comportamiento turístico. Recomiendan que este tipo de técnicas sean integradas al conjunto de herramientas que se utilizan para planear la política en materia turística.

Adicionalmente, en Rivera (2016) se estima el flujo de turistas internacionales en Puerto Rico, tomando como insumo los registros en hoteles y la información de Google Trends. En el trabajo se indica que la incorporación de esta información les permite obtener mejores inferencias sobre la asociación entre el número de registros de no residentes en hoteles. Además, se generan predicciones más realistas para los intervalos de las predicciones. Una conclusión importante es que sus predicciones son mejores al considerar el corto plazo (6 meses), confirmando en parte lo comentado por Choi *et al* (2012) en relación a que los datos de internet son mejores para predicciones contemporáneas.

A manera de síntesis de los trabajos que relacionan el flujo de turistas (domésticos o extranjeros) con la información de internet, se pueden extraer algunas características importantes: 1) son trabajos que analizan la relación que tienen los datos de búsquedas con valores reales de registros de turistas (en su mayoría fuentes oficiales); 2) por lo regular se utilizan modelos de series de tiempo relacionados con procesos autorregresivos; 3) el principal objetivo es el de modelar las series de turistas para generar pronósticos; 4) independientemente del destino analizado, los modelos que incluyen datos de internet tienden a generar mejores pronósticos (i.e. con menos errores) en comparación con sus contrapartes que utilizan otro tipo de información (i.e. registros de visitantes); 5) existe evidencia de que al utilizar palabras específicas en lugar de búsquedas agregadas, los modelos son más precisos; 6) en los trabajos no se analizan propiamente relaciones causales o impactos de las búsquedas.



El caso de México

Si bien para México no se cuenta con muchos trabajos donde se relacionan el flujo de turistas y búsquedas de internet, es posible citar algunos documentos. En Lichtle *et al* (2014) se propone la utilización de información de Twitter para estimar flujos de turistas. En el caso particular de los datos de Twitter se tenía como objetivo "...conocer los movimientos de un perfil de usuario de Twitter, representativo de un turista promedio, con una prueba piloto en las entidades de Puebla y Guanajuato". Para ello, la SECTUR formó un equipo con el Instituto de Estadísticas y Geografía (INEGI) y se diseñó un ejercicio para determinar el lugar de origen de los turistas durante los días 1,2 y 3 de febrero de 2014 en los estados de Puebla y Guanajuato.

En el trabajo se expone que es posible realizar algún tipo de asociación entre los perfiles de Twitter y el tipo de turistas nacionales que llegan a destinos de ciudad y playa. Por ejemplo, en México, la mayoría (95%) de usuarios de Twitter se encuentra entre 18 y 49 años de edad, rango de edad que se asemeja al de los turistas nacionales en ciudad y playa, donde el 87% de los turistas tiene entre 18 y 55 años. Adicionalmente, se argumenta que los usuarios de Twitter se concentran en deciles de hogares con mayores ingresos y por ello pueden dedicar una mayor proporción de sus ingresos en esparcimiento y en particular para gastar en servicios turísticos. La principal conclusión de este ejercicio es que se observa una correlación entre la actividad en Twitter relacionada con el destino turístico y la tendencia mostrada en los datos oficiales. Sin embargo el ejercicio en realidad no constituye un análisis de flujos de turistas ya que se concentró principalmente en una muestra de twits (7,955 usuarios que generan 827,424) y en un período temporal de tres días (1, 2 y 3 de febrero de 2014).

Este primer ejercicio es importante porque permite fundamentar estudios posteriores de Domínguez *et al* (2016) y de Domínguez *et al* (2016a), donde se ha propuesto una metodología más precisa para que mediante el análisis de las transacciones bancarias, sea posible conocer la dinámica del turismo en diferentes destinos. En particular, en estos trabajos, se analiza el caso de Quintana-Roo (estado que contiene varios de los principales destinos de playa en el país) y los municipios donde se encuentran territorios beneficiados con el programa "Pueblos Mágicos"³.

3 Este programa desarrollado por el gobierno mexicano, busca contribuir a fortalecer las ventajas competitivas de la oferta turística en ciertos destinos turísticos (por lo regular Sitios o Ciudades Patrimonio y Áreas Naturales Protegidas) mediante la ejecución de obras, servicios relacionados y acciones para el desarrollo sustentable del turismo (DOF,2016).



En su metodología los autores han procesado una gran cantidad de información relacionada con las transacciones electrónicas que integran la actividad comercial en diferentes destinos turísticos. En particular, utilizan información de los sistemas de pago electrónico de BBVA Bancomer, S.A. (crédito y débito) de dos distintas fuentes: 1) Terminal Punto de Venta (TPV) de tarjetas emitidas por BBVA Bancomer; 2) la actividad de tarjetas emitidas por cualquier entidad en TPV operado por BBVA-Bancomer, aplicable al análisis de actividad de turistas extranjeros. La información comprende los meses de agosto, septiembre, octubre, noviembre y diciembre de 2015.

Con el análisis de estos datos, presentan algunas inferencias sobre patrones asociados al flujo, los gastos y la procedencia de los turistas nacionales (i.e. el estado donde radican) y extranjeros (i.e. país de origen). Algunas de las conclusiones tras la realización de estos estudios para el caso de Quintana Roo son: 1) los turistas internacionales tienden a concentrarse en Playa del Carmen, donde ocurre alrededor del 35% del gasto total registrado para este estado; 2) La concentración del gasto del turista nacional toma especial relevancia en dos destinos: Cancún y Playa del Carmen, representando respectivamente el 25% y el 27% del gasto total en destino; 3) La mayor concentración de gasto de turistas nacionales se realizó durante los meses de verano (Julio-Agosto) observándose también un repunte en las vacaciones decembrinas (Noviembre-Diciembre); 4) Existen diferencias en los patrones de gasto pues los turistas nacionales utilizan el pago con tarjeta principalmente para viajes dentro del centro turístico, mientras que los turistas internacionales utilizan el pago con tarjeta principalmente en actividades de entretenimiento; 5) La Ciudad de México, el Estado de México y Quintana Roo, son el principal origen del gasto en turismo a los destinos analizados (alrededor del 50% del gasto de turistas nacionales); 6) los resultados sobre el origen de los turistas internacionales muestran que el mayor porcentaje de gasto es generado por turistas de Estados Unidos, seguido por turistas de Argentina.

Para el caso de los municipios con territorios que pertenecen al programa “Pueblos Mágicos”, se tiene que: 1) existe una mayor actividad de los turistas nacionales con respecto a los internacionales en Pueblos Mágicos, lo cual implica, que el programa favorece el consumo doméstico de servicios turísticos; 2) el gasto tiende a concentrarse en destinos que tienen una mejor conexión vía terrestre con otros destinos turísticos que poseen una mayor oferta de atractivos turísticos y servicios



financieros, lo cual implica, una sinergia en la generación de servicios turísticos; 3) la actividad del turista nacional es regional y con comportamientos de gasto heterogéneos que dependen, tanto de la ubicación geográfica, como de la inclusión financiera que tenga el territorio beneficiado por el programa Pueblos Mágicos (Domínguez *et al*, 2016: 37).

De los trabajos realizados para México utilizando información de internet, se observa que en ninguno de ellos existe realmente un análisis del flujo de turistas utilizando la información de búsquedas de internet. En el caso del uso de Twitter se tiene una representación mediante mapas del número de twits que algunos usuarios reportan para dos estados (Puebla y Guanajuato) y solamente en tres días. En cuanto al análisis de datos en gran escala, se demuestra que es posible conocer perfiles de turistas (lugar de procedencia, género, edad), flujos y una taxonomía de gastos que se realizan. Sin embargo, las estimaciones solamente comprenden un período temporal específico (agosto a diciembre de 2015) y el origen de los datos es privado lo cual no soluciona el problema de un análisis oportuno de la información.

2. Metodología

En el presente trabajo se analiza la relación que tienen los flujos de turismo doméstico con las búsquedas de conceptos clave asociados al avistamiento de luciérnagas en el estado de Tlaxcala. En términos muy generales, la metodología a seguir es la siguiente: 1) se construye una base de datos con información de las búsquedas relacionadas con el avistamiento de luciérnagas y del flujo de turistas; 2) posteriormente, se analiza el comportamiento de las variables; 3) se utiliza un modelo de Vectores Atorregresivos, que es un sistema de dos o más series de tiempo y que sirve para conocer la interacción dinámica que pudiera existir entre ellas; 4) se realizan pruebas de robustez y calidad estadística del modelo propuesto; 5) se generan funciones de impulso respuesta y una descomposición de la varianza de dicho modelo para conocer el impacto de las búsquedas en el flujo de turistas. A continuación se comentan las principales características de la información utilizada.



Datos

Para construir la base de datos, se utilizó información de tres fuentes: 1) el Instituto Nacional de Antropología e Historia (INAH) la cual se encuentra en el Sistema Institucional Estadística de Visitantes (<http://www.estadisticas.inah.gob.mx>); 2) información de Google Trends (<https://trends.google.com/trends/?hl=es>) para las consultas realizadas a través de internet y; 3) la estimación del flujo de cuartos promedio ocupados de la Secretaría de Turismo (SECTUR) en el portal Datatur (<http://www.datatur.sectur.gob.mx>). Con esta información se genera una base de datos que comprende de enero de 2004 a diciembre de 2016. Las variables de flujo turístico (cuartos promedio ocupados y visitas registradas por el INAH) se han estandarizado en una escala de 0 a 100, donde el 100 representa el valor más alto en cada serie.

Integración de las búsquedas en internet.

Las variables relacionadas con las búsquedas en internet que realizan los turistas, se construye a partir de información de Google Trends, la cual se puede considerar como Big Data (Ward *et al*, 2013). El proceso que sigue Google para integrar la información es el siguiente: cada que se escribe un concepto o palabra en el buscador, la aplicación genera un extracto de datos utilizando un porcentaje de todas las búsquedas. Para los datos en tiempo real se genera una muestra aleatoria de las búsquedas de los últimos siete días. Los datos que no son en tiempo real, constituyen una muestra aleatoria de los datos de la búsqueda de Google que pueden obtenerse desde el año 2004 hasta 36 horas antes de que se realiza la búsqueda. Una vez recopilada la información, Google crea categorías (e.g. ocio, deportes, noticias) y se vinculan con un tema. Estos temas relacionados, son temas buscados con frecuencia junto a los conceptos que busca el usuario (e.g. Cancún con springbreaker). Finalmente, se elimina cualquier información personal para garantizar el anonimato y la protección de datos sensibles.

Para construir tendencias, Google Trends excluye datos de búsquedas realizadas por muy pocos usuarios. También elimina las búsquedas realizadas de forma reiterada por una misma persona durante un breve periodo de tiempo y filtra las consultas que contienen caracteres especiales (signos de admiración, interrogación, apóstrofes, entre otros).



Con los datos agregados a nivel mundial, Google elabora un índice (parecido a un índice de confianza al consumidor) con una escala de 0 a 100 donde valores más altos indican una mayor popularidad o frecuencia de aparición de un concepto. Este proceso es necesario para garantizar que los datos de búsquedas son proporcionales al periodo temporal y la ubicación de una consulta.

Selección de las palabras clave del estudio.

Ante la gran cantidad de información que se puede encontrar en internet, es necesario definir criterios para la selección de las palabras clave. En el trabajo se utilizaron dos criterios. El primero consistió en aplicar una herramienta de correlación de datos (Google Correlate), la cual permite encontrar asociaciones entre diferentes palabras para todas las búsquedas realizadas en Google. Posteriormente, se seleccionaron las tres principales palabras que resultaron ser “Nanacamilpa”, “Luciérnagas” y “Piedra Canteada” (correlación por encima del 0.80). En los dos primeros casos, son palabras que tienen una asociación natural debido a que corresponden al lugar donde se verifica el avistamiento y a la forma coloquial de denominar a la *Macrolampis palaciosi*. En el tercer caso, la palabra identifica uno de los dos ejidos que actualmente gozan del permiso de explotación para realizar los recorridos.

Ubicando a estas tres palabras como las que mejor describen las búsquedas relacionadas con el avistamiento de luciérnagas en Tlaxcala, se utiliza Google Trends para generar las series una periodicidad mensual desde enero de 2004 a enero de 2016 (156 observaciones).

Finalmente, para garantizar que se tenían palabras relacionadas se calculó una matriz de correlación como la que se presenta en la tabla 1. Entre Nanacamilpa y Luciérnagas la correlación es de 0.60, entre Nancamilpa y Piedra Canteada es de 0.80 y entre Luciérnaga y Piedra Canteada de 0.64. En todos los casos son estadísticamente diferentes de cero con un p-value por debajo del 0.01. La conclusión es que los tres conceptos están muy relacionados.



Tabla 1. Correlación entre los conceptos relacionados con el avistamiento.

	Nanacamilpa	Luciérnaga	Piedra Canteada
Nanacamilpa	1		
Luciérnaga	0.6008***	1	
Piedra Canteada	0.7992***	0.6407***	1

Fuente: elaboración propia con datos de Sistema Institucional Estadística de Visitantes (para el flujo de turistas), Recuperado de <http://www.estadisticas.inah.gob.mx>. También se utiliza información de Google Trends de enero de 2004 a enero de 2017 mediante la aplicación: <https://trends.google.com.mx/trends/>.

Nota: *** p-value=0.01; ** p-value=0.05; * p-value=0.10.

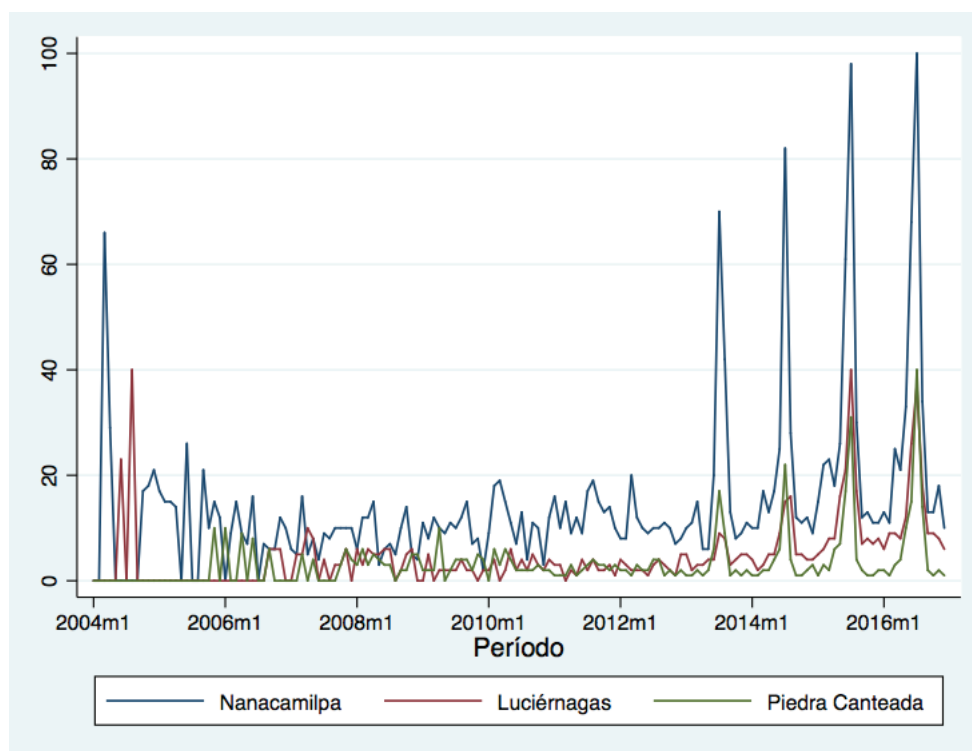
En la siguiente sección se presenta un análisis preliminar de los flujos de turistas domésticos y de las palabras relacionadas con el avistamiento de luciérnagas.

Análisis preliminar de la información.

En la gráfica 1 se presentan datos del comportamiento del turismo doméstico registrado en todos los posibles puntos de visita que identifica el INAH en Tlaxcala. También se incluye la información de los índices de Google Trends para cada una de las palabras relacionadas con el avistamiento de luciérnagas. Las consultas realizadas para la palabra “Nanacamilpa” desde enero de 2004. Es interesante observar un patrón de generación de ciclos en los últimos años (2014, 2015 y 2016), que coincide plenamente con la decisión de impulsar la explotación turística del avistamiento de luciérnagas. Hasta antes de esta fecha, realmente la palabra aparece con un volumen de búsquedas relativamente bajo (con excepción de las primeras búsquedas en 2004), el cual se mantiene prácticamente durante 10 años. A partir de 2014 justo en los meses donde ocurre el avistamiento (junio, julio y agosto) se observa un comportamiento cíclico acompañado de valores máximos cada vez más altos.



Grafica 1. Consultas de palabras clave en internet.



Fuente: elaboración propia a partir de datos de Google Trends (2017).

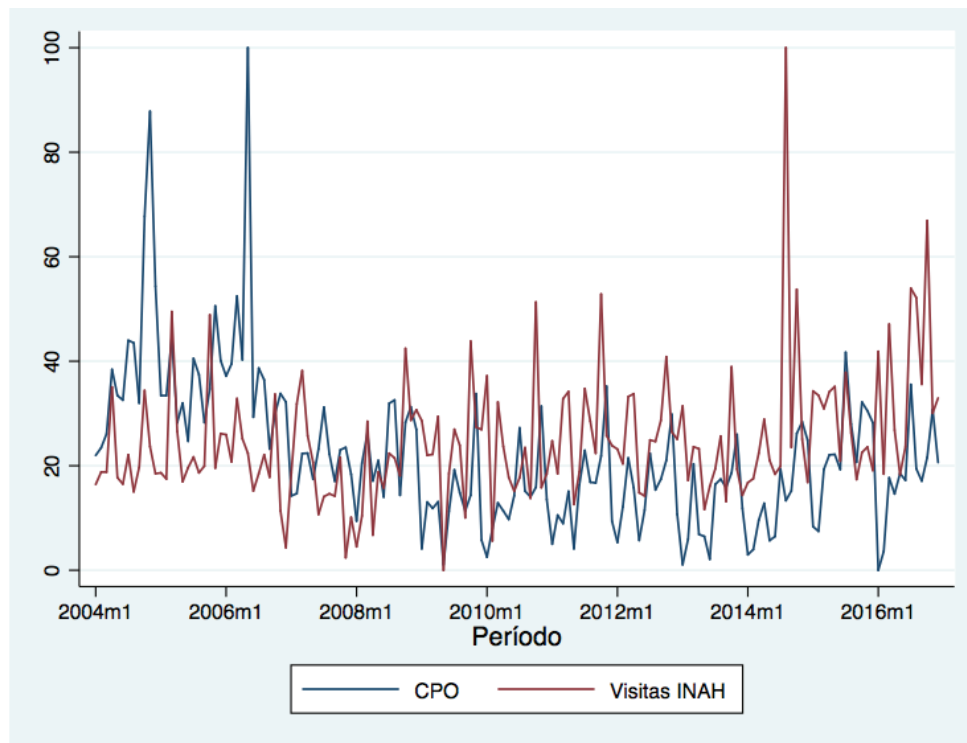
Para la palabra “Luciérnagas” en la gráfica 1, se identifica un patrón semejante al de Nanacamilpa. También es posible ver la aparición de un ciclo muy marcado en los meses donde ocurre el avistamiento (junio-agosto) a partir de 2004.

El tercer concepto que se analiza es “Piedra Canteada”, un ejido en Nanacamilpa el cual es el punto de referencia para el avistamiento. En este caso, se observa también con claridad la aparición de ciclos (solo que a partir de 2013) pero no la tendencia positiva que aparecía en las otras búsquedas. En el movimiento cíclico se observa que los picos son progresivamente más altos, lo cual implica una mayor variación en las búsquedas para “Piedra Canteada” año con año a partir de 2014



En cuanto a la variable de flujo de visitas registras por el INAH y los cuartos promedio ocupados (CPO), en Gráfica 2 se observa en general un comportamiento estacionario (i.e. no existe una tendencia).

Gráfica 2. Flujo de visitas registradas por el INAH y cuartos promedio ocupados.



Fuente: elaboración propia a partir de datos de INAH (2017) y Datatur (2018)

En todos los años se identifica un comportamiento cíclico donde se alcanza un pico en el mes de octubre con la celebración de la feria (cerca de 29 mil personas en promedio). Sin embargo, a lo largo del año, también se observan momentos importantes para la atracción de turistas. En particular, en marzo de cada año con la llegada del carnaval (en promedio cerca de 23 mil personas y en agosto con el desarrollo de las fiestas en honor a la Virgen de la Caridad (poco más de 22 mil personas en promedio). Recientemente y a partir de 2014 se observa otro pico en junio, el cual podría corresponder a la temporada de avistamiento de luciérnagas.



Las conclusiones que se pueden extraer de este análisis preliminar de la información son las siguientes: 1) se confirma la existencia de ciclos sincrónicos en los flujos de turistas y de las búsquedas en internet con mayor claridad a partir del año de 2014; 2) las búsquedas de palabras como Nana-camilpa, Luciérnagas y Piedra Canteada generan ciclo a partir de junio; 3) se presenta una tendencia positiva en la búsqueda de la palabra Luciérnagas y en el flujo de turistas también a partir de 2014.

En la siguiente sección se presentan los aspectos técnicos del modelo que permite estimar el impacto de las palabras relacionadas con el avistamiento en el flujo de turistas. Se busca encontrar el signo (positivo o negativo) y magnitud de este impacto para cada variable. Estas estimaciones permitirán conocer cuál de las tres búsquedas tiene el mayor impacto en el turismo que genera el avistamiento de luciérnagas en Tlaxcala.

Modelo para analizar el flujo de turistas en Tlaxcala a partir de consultas en internet.

Como se ha comentado, la principal motivación de este trabajo es la de conocer si es posible utilizar las búsquedas de internet como una variable que aproxima el flujo de turistas. Como se ha planteado, existe sospecha de que las personas que deciden viajar al avistamiento, buscaron información en internet para informarse sobre su viaje. Al mismo tiempo, las visitas al lugar de avistamiento, generan interacción en redes sociales y en la convivencia entre personas que motiva a su vez la decisión de otras personas de viajar hacia el avistamiento.

A partir de la metodología planteada en Johansen (1991) se sabe que un Modelo de Vectores Autorregresivos (VAR) tiene como ventaja el no imponer ninguna restricción, no se generan errores de especificación al asumir relaciones de causalidad entre las variables. Tampoco se requiere especificar ciertas variables como exógenas de la misma forma en que se hace para cualquier modelo econométrico de ecuaciones simultáneas.

Esta metodología consiste en plantear un sistema de ecuaciones donde las variables explicativas en cada ecuación, serán el conjunto de rezagos (i.e. valores en t-1). Las ecuaciones que se utilizan, aparecen en forma reducida y sin restringir. Cuando las ecuaciones son de forma reducida, los



valores contemporáneos de las variables en el modelo no serán variables explicativas en ninguna de las ecuaciones. Las variables explicativas en cada ecuación, serán el conjunto de retardos de las otras variables del modelo (como en un proceso autorregresivo, AR). Por ello, si las ecuaciones no están restringidas, cada variable del modelo aparece en cada ecuación.

En este tipo de modelos, usualmente se incluyen otras variables deterministas tales como la tendencia temporal, variables de respuesta binaria estacionales, o variables que identifican una coyuntura (i.e. cambio estructural). Este tipo de modelos resultan ser muy útiles cuando se asume simultaneidad entre un grupo de variables, pero además cuando la relación entre estas variables prevalece durante un cierto número de períodos.

A partir de la estimación de los parámetros de este sistema de ecuaciones, se calcula el impacto de cada variable. El modelo VAR usado en este estudio es de la siguiente forma:

$$(1) \quad Y_{it} = \Gamma_0 + \sum_{j=1}^k \Gamma Y_{it-j} + \varepsilon_{it}$$

Donde Y_{it} es un vector de cinco variables: flujo de visitantes a Tlaxcala registrados por el INAH, cuartos promedio ocupados, búsquedas para “Nanacamilpa”, búsquedas para “luciérnagas” y búsquedas para “Piedra Canteada” en el momento t . El vector Y_{it} se encuentra en función del vector Y_{it-j} donde se tienen las mismas variables que en Y_{it} solo que con $it-j$ rezagos. El último término ε_{it} se asume como un error con una distribución Normal. Se utiliza la descomposición de Cholesky para identificar el modelo.

Para determinar los impactos de un coeficiente, se utilizan Funciones Impulso Respuesta (FIR) que provienen de la matriz de funciones impulso respuesta a partir de los coeficientes estimados de cada VAR, así como sus errores estándar. Precisamente estos errores estándar permiten definir bandas de error al 5% y 95% mediante simulaciones Monte Carlo.

A continuación se comentan estos resultados.



4. Resultados.

Relación entre búsquedas de internet y el flujo de turistas.

En esta sección se comentan los resultados de las estimaciones del VAR (tabla 2). Por las características de las variables y pruebas de robustez realizadas al modelo, se ha determinado que se debe elegir un máximo de tres rezagos en cada variable. En la sección de Anexos están las pruebas que se realizaron al modelo estimado. El principal objetivo del trabajo es el de explicar cómo la decisión de visitar y hospedarse a un destino en particular, se asocia con la búsqueda de ciertas palabras clave en internet. Existe un énfasis en comentar los casos donde la respuesta es el flujo de turistas domésticos medido como los cuartos promedio ocupados y el número de visitas registradas por el INAH.

Tabla 2: Resultados de las Estimaciones del VAR.

	a) Cuartos promedio ocupados		b) Visitantes INAH	
	Coeficiente	p-value	Coeficiente	p-value
Cuartos promedio t_{-1}	0.42	*** 0.00	-0.05	0.55
Cuartos promedio t_{-2}	0.09	0.27	-0.12	0.16
Cuartos promedio t_{-3}	0.25	*** 0.00	0.14	* 0.06
Nanacamilpa t_{-1}	0.13	0.16	0.44	*** 0.00
Nanacamilpa t_{-2}	-0.09	0.42	-0.23	** 0.03
Nanacamilpa t_{-3}	0.01	0.91	0.16	0.13
Luciérnagas t_{-1}	-0.20	0.30	-0.07	0.73
Luciérnagas t_{-2}	0.45	** 0.02	0.13	0.46
Luciérnagas t_{-3}	0.29	0.14	-0.13	0.52
Piedra Canteada t_{-1}	-0.04	0.89	-0.11	0.73
Piedra Canteada t_{-2}	-0.56	* 0.06	-0.04	0.90
Piedra Canteada t_{-3}	-0.05	0.88	0.28	0.36
Visitantes INAH t_{-1}	0.13	* 0.10	0.05	0.50
Visitantes INAH t_{-2}	-0.06	0.45	0.14	* 0.07
Visitantes INAH t_{-3}	-0.14	* 0.07	0.06	0.41

Notas: Un asterisco (*) denota significancia al 10%, doble asterisco (**) denota significancia al 5%, y triple asterisco (***) denota significancia al 1%.



Cuando la variable dependiente son los cuartos promedio (inciso a de la tabla 2), se identifica un efecto tendencial del uso de las habitaciones (el tercer y primer rezago son positivos y diferentes de cero); es decir, si incrementa la ocupación en un periodo, a futuro se esperaría una mayor ocupación. Resulta interesante que la búsqueda del lugar destino de viaje (i.e. Nanacamilpa) no genere ningún impacto en la cantidad de cuartos promedio ocupados. Los tres coeficientes asociados a esta palabra no son estadísticamente diferentes de cero. Sin embargo, en el caso de las búsquedas de la palabra Luciérnagas si se observa un impacto positivo en la ocupación. Al menos una de las tres variables relacionadas con la palabra resultan ser estadísticamente diferentes de cero (el segundo rezago de la variable). También resulta interesante que la búsqueda del principal lugar de hospedaje en Nanacamilpa (i.e. Piedra Canteada), genera un impacto negativo en la ocupación. Este resultado puede conducir a una interpretación interesante sobre el comportamiento de los turistas y visitantes: 1) la persona busca un lugar donde quedarse en Nanacamilpa y la principal referencia es Piedra Canteada; 2) dado que Piedra Canteada tiene pocas habitaciones y se saturan rápido, al confirmar que no hay espacio, las personas deciden no ir; 3) se puede argumentar que a medida que más personas buscan Piedra Canteada y no encuentran hospedaje, se incide negativamente en la ocupación de cuartos promedio en el estado lo cual evidencia que el interés de los posibles turistas es hospedarse en el lugar del avistamiento.

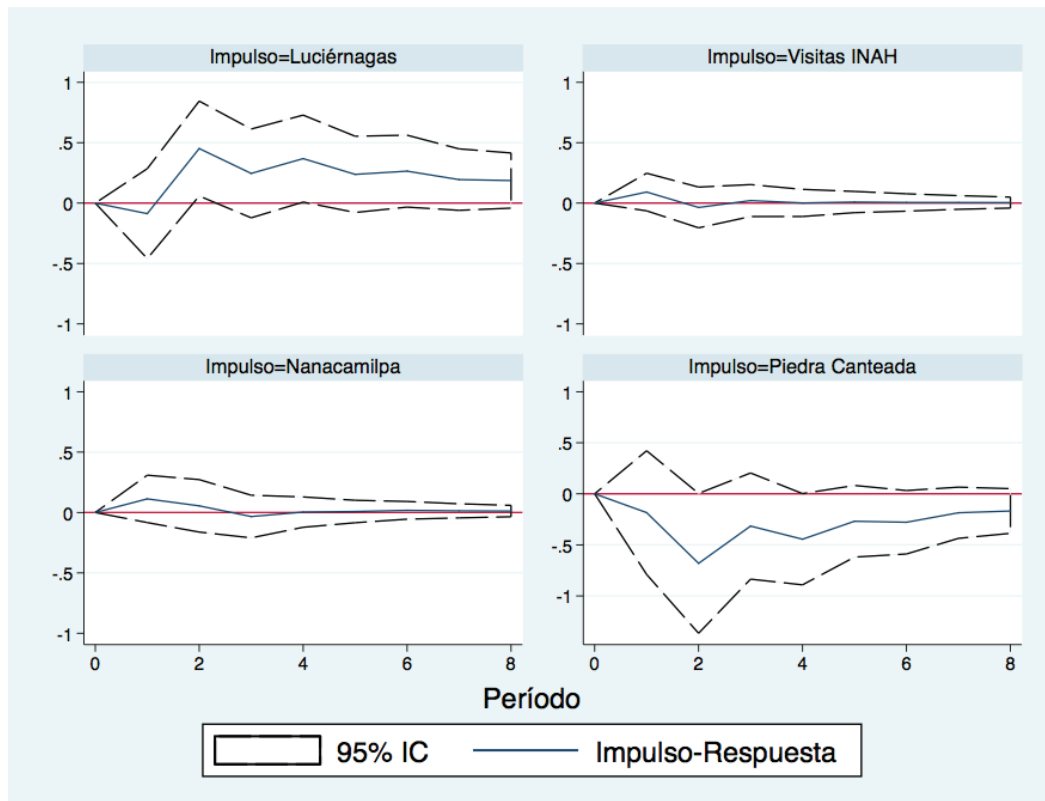
Para la variable de flujo de visitas registradas por el INAH, se observa que a medida que incrementan la cantidad de cuartos promedio ocupados, también incrementan las visitas (en particular el rezago de tercer orden es positivo). Este resultado es consistente con el comportamiento de un turista e implica que se buscan realizar más actividades en adición al avistamiento. Se nota que las búsquedas de la palabra Nanacamilpa generan un impacto mixto. Para el rezago de primer orden es positivo, mientras que para el rezago de segundo orden es negativo. Este resultado puede interpretarse de la siguiente manera: las personas buscan ir a Nanacamilpa y al igual que en el caso de Piedra Canteada, si no encuentran lugar donde hospedarse, deciden ya no asistir al avistamiento. Esto explicaría el coeficiente con valor negativo en el segundo período. Sin embargo, existen personas que deciden en muy poco tiempo (i.e. con un rezago) acudir al avistamiento (o bien son personas que del período $t-2$ al $t-1$ continuaron buscando hospedaje). De paso a Nanacamilpa, podrían visitar alguno de los sitios que custodia el INAH. Esto explica el coeficiente positivo de la variable con un rezago. La última variable que muestra un impacto diferente de cero (en este caso positivo) es la de visitantes registrados por el INAH con un rezago de segundo orden. La explicación es que existe una inercia en ciertos meses del año que puede durar al menos dos períodos. La más clara de estas inercias es la de los meses de junio, julio y agosto donde coinciden el avistamiento y la Huamantlada.



La conducta de los turistas domésticos y las búsquedas de internet en el flujo de turistas.

A partir del modelo estimado en la sección anterior se puede conocer el efecto que tienen las búsquedas en Google Trends en el flujo de turistas buscando explicar la conducta de las personas. Para ello, se debe calcular una función impulso-respuesta donde se estima qué ocurrirá en el tiempo con el flujo de turistas ante un impulso o cambio inesperado (i.e. un shock) de las otras variables. Cuando la variable que genera el impulso es la misma que la de la respuesta se habla de una autorrespuesta.

Gráfica 3. Respuesta en los cuartos promedio promedio ocupados



Fuente: elaboración propia a partir de datos de Google Trends (2017), INAH (2017) y Datatur (2018).



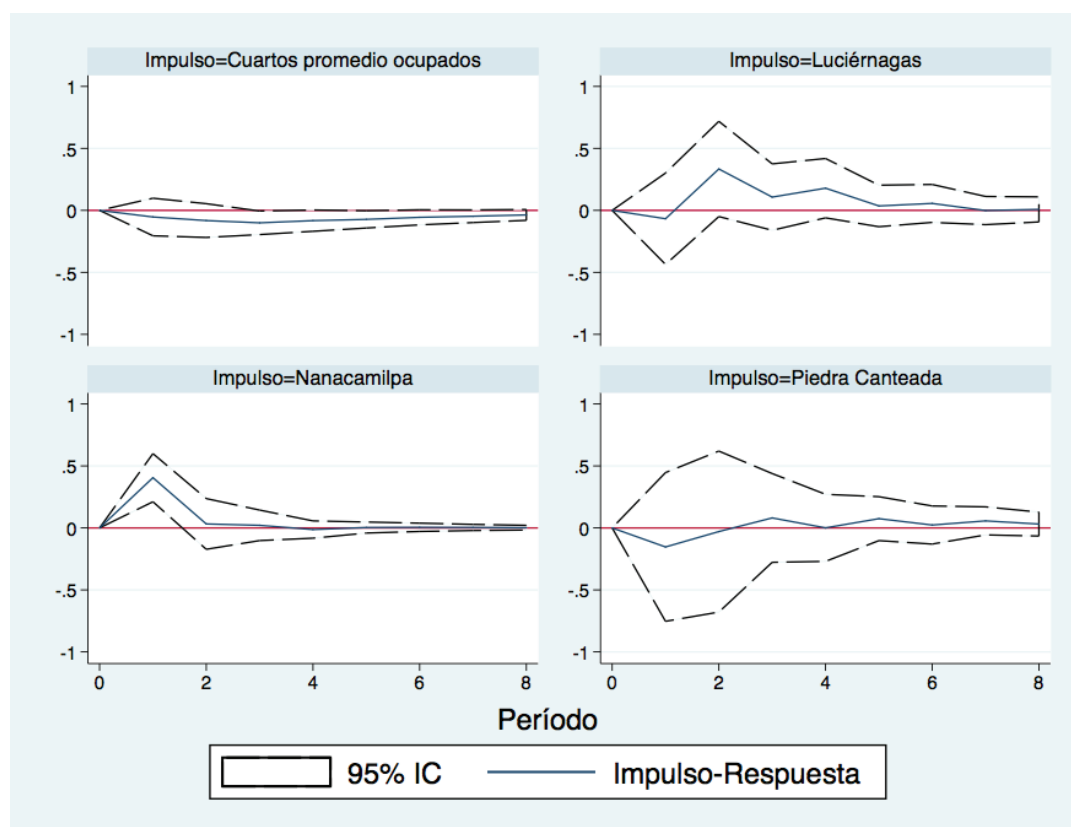
Se puede considerar a las funciones impulso-respuesta como una simulación de los efectos que tendrían shocks aleatorios de cada variable en aquella que se selecciona como variable de respuesta.

En la gráfica 3 se presenta el caso donde la respuesta es los cuartos promedio ocupados, asumiendo cambios positivos e intervalos al 5% y 95% de confianza y un horizonte temporal de 8 períodos (i.e. meses). El eje de las ordenadas indica el impacto en la variable de respuesta y el de las abscisas el tiempo que transcurre después del shock aleatorio en la variable de impulso. Se ha mantenido para todas las gráficas la misma escala $[-1,1]$ con la finalidad de hacer las comparaciones.

Los resultados permiten concluir que la búsqueda de la palabra luciérnagas genera la mayor respuesta en la ocupación. Este resultado, prueba que efectivamente los turistas que se hospedan en Tlaxcala lo hacen por el evento del avistamiento. Las visitas registradas por el INAH y la búsqueda de la palabra Nanacamilpa también generan una respuesta positiva en los cuartos promedio ocupados; sin embargo, las respuestas son de menor magnitud en relación a la que genera la búsqueda en internet de la palabra luciérnagas. Lo anterior, implica que las personas no le otorgan importancia a la búsqueda de la palabra Nanacamilpa como el lugar de origen del avistamiento. Por otro lado, sorprende que los sitios arqueológicos y museos del INAH no generen una mayor respuesta en la ocupación promedio de cuartos en Tlaxcala. En cuanto a la variable de búsquedas en internet de Piedra Canteada, sorprende que genere una respuesta negativa y persistente en los cuartos promedio ocupados. Una posible explicación de este comportamiento como ya se ha comentado, es que ante una oferta de cuartos limitada en Nanacamilpa, si las personas buscan Piedra Canteada y no encuentran lugar para pasar la noche, no buscarán otros lugares para quedarse dentro del estado de Tlaxcala, sino en otros estados (e.g. Hidalgo). Otra posible explicación es que una vez finalizado el avistamiento, las personas no deciden quedarse y regresan a su lugar de origen, lo cual implica que mucho del turismo que acude a Nanacamilpa debe provenir de lugares muy cercanos (e.g. Puebla, Hidalgo, CDMX y Estado de México). Esto último implicaría que en realidad el avistamiento genera visitantes y no turistas



Gráfica 4. Respuesta en las vistas registradas por el INAH.



Fuente: elaboración propia a partir de datos de Google Trends (2017), INAH (2017) y Datatur (2018).

El análisis que se realiza ahora es sobre las funciones de impulso respuesta cuando la variable de respuesta es el número de visitas registradas por el INAH (Gráfica 4). En este caso, sorprende que la ocupación hotelera genere una respuesta negativa y persistente (aunque de baja magnitud). En el caso de la búsqueda de la palabra luciérnagas, en el corto plazo (un período) se tiene una respuesta negativa en las visitas registradas por el INAH pero a partir del segundo período se genera una respuesta positiva que dura al menos otros cinco periodos. En cuanto a la búsqueda de la



palabra Nanacamilpa en internet, se observa que genera una respuesta positiva en las visitas del INAH. En este caso, prácticamente la respuesta es de la misma magnitud que la de la búsqueda de la palabra luciérnagas, pero no es persistente en el tiempo pues su efecto dura un periodo.

Finalmente, para la variable de las búsquedas en internet de la variable Piedra Canteada, se observa que genera una respuesta negativa de corto plazo (primer y segundo período), seguida de una respuesta positiva y persistente hasta el octavo período. Desafortunadamente, la magnitud de la respuesta positiva es muy pequeña. Una posible interpretación de este resultado es que las personas que buscan la palabra Nanacamilpa lo hacen con cierto tiempo, diseñando itinerarios que podrían implicar visitas a los lugares que tiene bajo su custodia en INAH. Por esta razón la respuesta positiva aparece hasta el tercer período (o mes, de acuerdo a la temporalidad de los datos).

Con el análisis de las graficas 3 y 4 se puede inferir la conducta de los turistas domésticos: las búsquedas relacionadas con la palabra luciérnagas generan una respuesta positiva de corto plazo tanto en la ocupación hotelera como en las visitas que registra el INAH. Si la visita es planeada con anticipación (e.g. 4 meses al menos) y se busca como lugar de hospedaje Piedra Canteada, es posible que las personas decidan visitar lugares que tiene bajo su resguardo el INAH. La cuestión es que si ya no hay hospedaje en este lugar, las personas deciden ya no quedarse en el estado de Tlaxcala produciendo una respuesta negativa en las visitas que registra el INAH. La respuesta que genera la búsqueda de la palabra Nanacamilpa es positiva en ambas variables flujo aunque mucho menor en el caso de los cuartos promedio ocupados. Esto último exhibe el hecho de que las personas podrían estar buscando lugares adicionales a Piedra Canteada para poder hospedarse. Si llegaran a encontrar hospedaje, es muy probable que aumenten las también las visitas que el INAH registra.

5. Conclusiones y discusión.

En el presente trabajo se ha demostrado que la incorporación de búsquedas en internet permite explicar el comportamiento de turistas y visitantes para el turismo en Tlaxcala. Se observa que los análisis generados son consistentes con los hallazgos realizados a nivel internacional. En particular, la relación entre la búsqueda de palabras que identifican la oferta de hospedaje y el flujo de turistas.



Hasta donde se revisó, este podría ser uno de los primeros trabajos generados para México partiendo de un enfoque de estudio sobre una muestra de información a nivel estatal.

Se tienen dos aportaciones a la literatura que estudia el uso de internet y el flujo de turistas. Por un lado, propone una adecuación de la metodología cuantitativa que vincula los datos generados en consultas de internet con datos oficiales para conocer el comportamiento de los turistas (e.g. Gawlik, *et al*, 2011; Choi *et al*, 20012; Yang *et al*, 2014; Bangwayo-Skeete y Skeete, 2015; Gunter *et al*, 2016). Por otro lado, constituye una innovación importante al profundizar en el entendimiento de la conducta de los turistas, a partir de uso de modelos de VAR. En este sentido, el uso de este tipo de modelos ha permitido demostrar que búsquedas específicas en internet tienen un impacto en el flujo de turistas.

En cuanto a las conclusiones, se descubrió que la palabra Nanacamilpa, Piedra Canteada (qué es el principal lugar de hospedaje) y Luciérnagas tienen un impacto importante. En relación al impacto de cada palabra buscada se ha encontrado que el lugar donde se hospedan las personas (i.e. Piedra Canteada) incide en mayor medida en el flujo de hospedajes. Un resultado interesante es que la búsqueda del lugar destino de viaje (i.e. Nanacamilpa) no genera ningún impacto en la cantidad de cuartos promedio ocupados. También resulta interesante que la búsqueda del lugar más importante para el hospedaje en Nanacamilpa (i.e. Piedra Canteada), genere un impacto negativo en la ocupación.

Para la variable de flujo de visitas registradas por el INAH, se observa que a medida que incrementan la cantidad de cuartos promedio ocupados, también incrementan las visitas. Sin embargo, para el caso de las búsquedas de la palabra Nanacamilpa se encuentra que si bien en el corto plazo hay un impulso positivo ocasionado por las búsquedas en internet, en periodos más rezagados el impacto es negativo. La interpretación que se dio de este resultado es que si las personas buscan Nanacamilpa lo hacen para hospedarse. Si no encuentran disponibilidad de cuartos, deciden ya no asistir al avistamiento. Finalmente, se identifica que los visitantes registrados por el INAH en un período previo, impulsa dichas visitas confirmando la inercia que se verifica en el flujo de visitas, particularmente en los meses de junio, julio y agosto donde coinciden el avistamiento y la Huamantlada.



Tras el análisis de los datos se identifica que las personas que acuden al avistamiento buscan un lugar donde quedarse en Nanacamilpa y la principal referencia es Piedra Canteada. Como Piedra Canteada tiene pocas habitaciones y se saturan rápido, al confirmar que no hay espacio, las personas deciden no ir al avistamiento.

Al relacionar los hallazgos se pudo identificar una recomendación de política pública:

- 1) No existen elementos para posicionar la palabra Nanacamilpa como una marca turística destinada al hospedaje debido a que su efecto es nulo en la ocupación de cuartos. Sin embargo, dado que se encontró una relación de corto plazo positiva en el flujo de visitas, se puede explorar la posibilidad de relacionar la palabra con otros destinos dentro del estado.
- 2) Derivado de lo anterior, no se recomienda incrementar la oferta de cuartos en Nanacamilpa ya que no es sustentable. Como se demuestra en las gráfica 1, existe un punto en el tiempo (la temporada de avistamiento) donde se genera una alta demanda, pero también que contemple que la demanda por hospedaje será baja (o inexistente) en el resto del año.
- 3) Una opción sería la de desarrollar otros servicios turísticos a lo largo del año en Nanacamilpa, de tal forma que sea más atractiva la llegada al municipio.
- 4) Se debe integrar la información generada de las consultas de internet a las variables de flujo de turistas a nivel local para planear estrategias de promoción y políticas relacionadas con el turismo en Tlaxcala. Es un recurso que se puede obtener prácticamente a un costo cero y que arroja información importante de las personas que deciden viajar al estado.



Referencias bibliográficas

- Artola, C., Pinto, F., y de Pedraza García, P. (2015). Can internet searches forecast tourism inflows?. *International Journal of Manpower*, 36(1)[103-116].
- Bangwayo-Skeete, P. F., y Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, 46[454-464].
- Claveria, O., y Datzira, J. (2010). Forecasting tourism demand using consumer expectations. *Tourism Review*, 65(1)[18-36].
- Choi, H., y Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88(s1)[2-9].
- Da, Z., Engelberg, J., y Gao, P. (2011). In search of attention. *The Journal of Finance*, 66(5)[1461-1499].
- Diario Oficial de la Federación (DOF) (2016), Acuerdo por el que se emiten las Reglas de Operación del Programa de Desarrollo Regional Turístico Sustentable y Pueblos Mágicos (PRODERMAGICO), para el ejercicio fiscal 2017, México, (DOF: 29/12/2016). Recuperado de http://www.dof.gob.mx/nota_detalle.php?codigo=5468076&fecha=29/12/2016. Consultado el: 22/09/ 2017.
- Domínguez Torres, M. L., Sánchez Salinas, J.C. Murillo Arias, J., Valero Lapaz, H. y Romero Palop, J.D. (2016). Big Data y Turismo en México: Pueblos Mágicos. *Documentos de Investigación Estadística y Económica No. 2016-5*. Subsecretaría de Planeación y Política Turística. Recuperado de http://www.datatur.sectur.gob.mx:81/Reportes/bigdata/pdf/2016_5.pdf. Consultado el 22/09/2017.
- Domínguez Torres, M. L., Sánchez Salinas, J.C. Murillo Arias, J., Valero Lapaz, H. y Romero Palop, J.D. (2016a). Big Data y Turismo en México: Quintana Roo. *Documentos de Investigación Estadística y Económica No. 2016-4*. Subsecretaría de Planeación y Política Turística. Recuperado de http://www.datatur.sectur.gob.mx:81/Reportes/bigdata/pdf/2016_4.pdf. Consultado el 22/09/ 2017.
- Engle, Robert (2000). The econometrics of ultra-high-frequency data. *Econometrica*, 68(1), 1-22.
- Gawlik, E., Kabaria, H., y Kaur, S. (2011). Predicting tourism trends with Google Insights. Recuperado de <https://pdfs.semanticscholar.org/f943/6afa83b3194cbf3e82676a6d2b1cf40afba3.pdf> Consultado el 01/01/2018.



- Gunter, U., y Önder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni-and multivariate models employing monthly data. *Tourism Management*, 46[123-135].
- Johansen S. (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models" *Econometrica*, 59, pp. 1551-1580.
- Joseph, K., Wintoki, M. B., y Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting*, 27(4)[1116-1127].
- Lichtle Fragoso, P.M. y Sánchez Salinas, J.C. (2014). Uso Productivo de Big Data y Redes Sociales en el Sector Turismo Octubre, 2014. Recuperado de http://www.datatur.sectur.gob.mx/Documentos%20Publicaciones/2014_1_DocInvs.pdf. Consultado el 22/09/2017.
- Machado Mendes Filho, L. A. (2014). Empowerment in the context of User-Generated Content in the Travel Industry: A research model proposal. *El Periplo Sustentable* [en línea] 2014, (Julio-Diciembre). Recuperado de <http://www.redalyc.org/articulo.oa?id=193432842002>. Consultado el 5/01/ 2018.
- Rivera, R. (2016). A dynamic linear model to forecast hotel registrations in Puerto Rico using Google Trends data. *Tourism Management*, 57[12-20].
- Ward, J. S., & Barker, A. (2013). Undefined by data: a survey of big data definitions. Recuperado de: arXiv preprint arXiv:1309.5821. Consultado el: 15 de agosto de 2018.
- Yang, Y., Pan, B., y Song, H. (2014). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research*, 53(4)[433-447].
- Yang, X., Pan, B., Evans, J. A., y Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management*, 46[386-397].
- Zaragoza Caballero, S. (2015). Nuevas especies de Photinus (Coleoptera: Lampyridae: Photinini) del bosque tropical caducifolio del Pacífico mexicano. *Revista mexicana de biodiversidad*, 86(3) [638-651].



Anexos.

En esta sección se presenta un análisis complementario con diferentes pruebas estadísticas que permiten el conocimiento de su estabilidad estructural.

Pruebas para la identificación del VAR. Estacionariedad

Tabla A1. Prueba de estacionariedad de las series que conforman el VAR

	<i>Estadístico t</i>				
	p<=0.10	Nanacamilpa	Luciérnagas	Piedra Canteada	Flujo de Turistas
Sin tendencia y sin deriva	-3.13	-2.3969	-2.0073	-3.3496	-4.5359
Con deriva	4.07	2.6084	2.7884	4.0919	7.0667
Con deriva y tendencia	5.47	3.5272	3.854	5.6737	10.4811

Fuente: elaboración propia a partir de datos de Google Trends (2017), Datatur (2018) e INAH (2017)

La conclusión de la tabla A1 es que de los cuatro casos solo las series de búsquedas en Google Trends para Piedra Canteada y de los datos del INAH sobre el Flujo de Turistas presentan un estadístico t que permite rechazar la hipótesis nula (i.e se puede rechazar la hipótesis nula de no estacionariedad). Los resultados de la prueba son congruentes con lo que se comentó en el análisis.



Rezagos

Siguiendo a Lutkepohl (2005) se recomienda elegir el número de rezagos en función de los criterios BIC y HQ con valores más bajos debido a que el criterio AIC sobreestima el orden del rezago. En este caso, se plantea que la mejor especificación con los datos de la tabla A2 se obtiene al elegir 2 rezagos.

Tabla A2. Resumen de estadísticos para identificar el número de rezagos en el VAR

Rezago(p)	AIC	BIC	HQ	M(p)	p-value
0	27.539	27.539	27.539	0.000	0.000
1	26.724	27.034	26.850	144.027	0.000
2	26.661	27.282	26.913	36.414	0.003
3	26.626	27.556	27.004	31.795	0.011
4	26.752	27.992	27.256	9.897	0.872
5	26.789	28.340	27.419	20.747	0.188
6	26.888	28.749	27.644	12.532	0.707
7	27.003	29.174	27.884	10.367	0.847
8	26.993	29.474	28.000	24.130	0.087
9	27.015	29.806	28.149	19.738	0.232
10	26.950	30.052	28.210	28.190	0.030

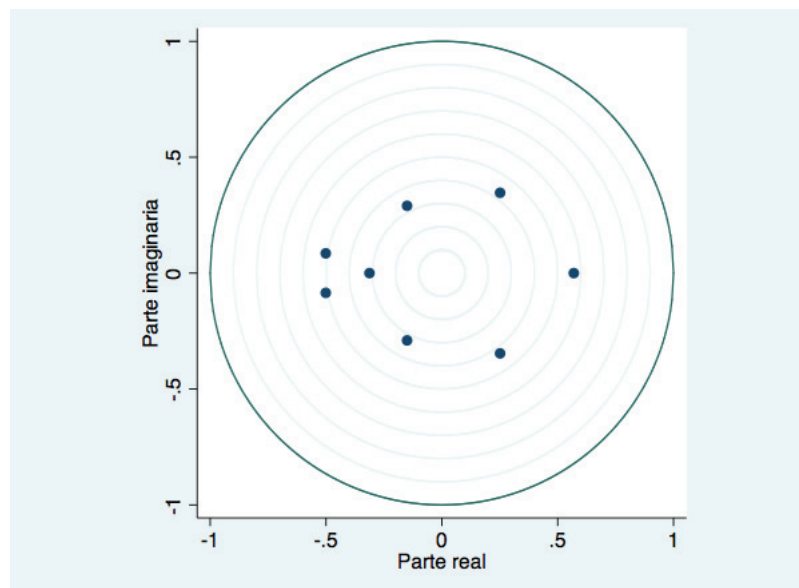
Fuente: elaboración propia a partir de datos de Google Trends e INAH (2017)

Prueba de estabilidad

En la gráfica A3 se presenta como prueba de estabilidad, la relación entre las raíces inversas del polinomio autorregresivo del VAR. Se concluye que existe una relación de tendencia entre las variables y por ende una buena relación de cointegración entre las variables.



Gráfica A3. Estabilidad del VAR



Fuente: elaboración propia a partir de datos de Google Trends e INAH (2017)