

Nuevas formas de conteo de visitantes: estimación a través de telefonía móvil

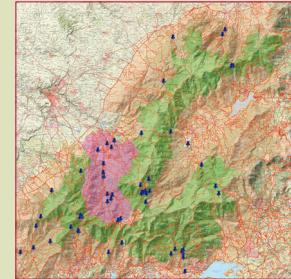
Parque Nacional de la Sierra de Guadarrama. Un escenario complejo

Límites, carreteras, tren, accesibilidad, aparcamientos, zonas tradicionales de recreo y esparcimiento, núcleos de población y grandes ciudades muy próximas, estaciones de esquí en su entorno, etc.

Necesidad de buscar nuevas formas de conteo

Métodos tradicionales no pueden abarcar todo el Parque: aforadores automáticos, muestreo de aparcamientos, registro de visitas en los Centros de Visitantes, etc.

Estos métodos precisan gran esfuerzo y presentan algunos inconvenientes



Sondas y ubicación de aforadores

Caracterización y estimación del número de visitantes a partir de datos de telefonía móvil

¿En qué consiste?

Estudio piloto para el conteo y la caracterización de visitantes a partir de datos de telefonía móvil como complemento a otras fuentes de datos.

¿Qué se obtiene?

- Estimación del nº de visitantes y caracterización sociodemográfica.
- Número de visitantes que entran por cada carretera
- Estimación del número de visitantes en áreas recreativas difusas

¿Cómo se hace?

A partir de los datos de telefonía móvil con posiciones geolocalizadas por triangulación (sin GPS)

Métodos estadísticos para extrapolar datos a partir de un número determinado de días analizados

Metodología

Sectorización del ámbito de estudio en función de las grandes áreas de uso público y de las coberturas

Determinación de visitantes: qué usuarios detectados son visitantes del Parque y cuáles no

- Determinación de patrones de actividad y movilidad de la población
- No se consideran los que están menos de 30 minutos por estar de paso
- Los que están más de 8 horas se excluyen por ser residentes y/o trabajadores
- Visitantes menores de 16 años no se tienen en cuenta por no ser titulares de línea.

Empleo de factores de corrección con encuestas de tipificación para no perder estos datos

Cálculo del flujo de visitantes en las entradas y salidas del Parque en vehículo

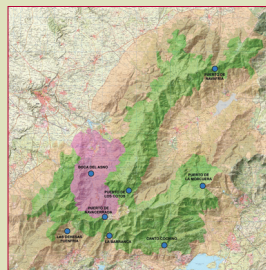
- Combinación de los datos de telefonía con los de la red de carreteras

Cálculo de afluencia a zonas recreativas difusas

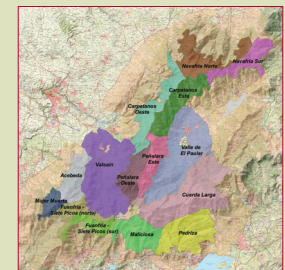
- Nº de visitantes que permanece al menos 4 horas. Estimación
- Nº de visitantes que permanece entre 30 minutos y 4 horas. Máximo razonable

Selección de los días analizados

- Selección de sólo algunos días para el análisis. Los complejos y costosos cálculos hacen necesario limitar los días de análisis.
- División del año en periodos siguiendo la distribución de visitantes al año dada por los contadores (% por meses y festivo/laborable).
- Selección de los días según la división del año para el análisis cuyos valores medios son los más próximos a los promedios de los aforadores para cada periodo.



Áreas difusas de alta ocupación



Sectorización para el conteo por telefonía

Resultados

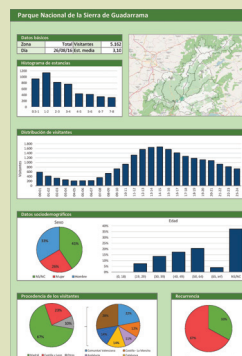
Resultados finales

- Se calculan los promedios de los resultados de los indicadores tanto para cada sector como para los totales por provincia y del total del Parque.
- Estos promedios se multiplican por el número de días reales según la división del año, sacando los resultados finales

Estimación total de visitantes en 2017					
	INVIERNO	PRIMAVERA	VERANO	OTOÑO	TOTAL
Madrid	357.669 (13,3%)	495.881 (18,4%)	405.352 (15,1%)	272.491 (10,1%)	1.531.393 (56,9%)
Segovia	191.682 (7,1%)	413.536 (15,4%)	325.178 (12,1%)	230.191 (8,6%)	1.160.587 (43,1%)
Total	549.351 (20,4%)	909.417 (33,8%)	730.530 (27,1%)	502.682 (18,7%)	2.691.980

Valores medios diarios del Parque Nacional en 2017				
	INVIERNO	PRIMAVERA	VERANO	OTOÑO
Día laborable medio	5.542 (x 63)	6.479 (x 60)	6.230 (x 64)	3.934 (x 59)
Día no laborable medio	7.415 (x 27)	16.797 (x 31)	11.852 (x 28)	8.199 (x 33)

* entre paréntesis el número de días de cada periodo por el que hay que multiplicar el total

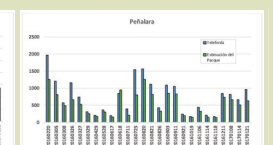
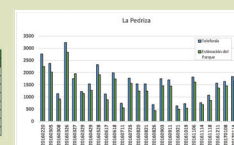
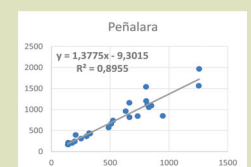
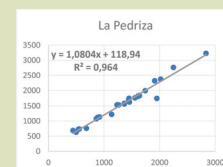


Ejemplos de los indicadores entregados

Indicador	Valor
Visitantes	1.531.393
Visitantes por día	4.223
Visitantes por hora	176
Visitantes por sector	...

Proyección: elaboración del modelo matemático extrapolable

- Los datos relativos por sectores (porcentajes) y las fórmulas que correlacionan aforadores y telefonía servirán para sacar los totales en años venideros utilizando las series de datos de los contadores



El número sí es importante. Nuevas formas para saber cuántos somos

¿Es realmente importante saber cuántos visitantes somos? La respuesta es fácil: sí que es importante y es necesario dedicar recursos a conocer las cifras. Pero, ¿cómo somos capaces de ofrecer cifras realistas? Esta entrada al blog científico explica a grandes rasgos cómo se hace en el Parque Nacional de la Sierra de Guadarrama y algunos resultados para 2017.

Queramos o no, la presencia del ser humano en el medio natural provoca impactos, eso sí, no igual de importantes. Hay zonas que son muy sensibles en las que determinado número de visitantes pueden causar impactos muy serios y a veces irreversibles. Otras, sin embargo, soportan mejor la presencia de visitantes sin que la flora y la fauna se vean afectadas seriamente.

Además, saber con la mayor precisión posible cuántos somos, cómo se reparten las visitas, en qué zonas y con qué frecuencia, conocer cuáles son los meses o días de mayor presencia y otros datos relacionados con los visitantes sirven para tomar las decisiones oportunas que ayuden a la conservación de este espacio natural protegido y que mejoren la calidad de la visita, de las infraestructuras y de los equipamientos.

Parque Nacional de la Sierra de Guadarrama. Un escenario complejo

Esta necesidad de conocimiento se convierte en un reto, pues es muy difícil dar cifras reales en un Parque como el de Guadarrama, en el que innumerables sendas lo recorren, con carreteras que lo cruzan y grandes aparcamientos. Además, una tradición montañera y de recreo histórica hace que sea el destino preferido para muchos, sobre todo cuando lo separan de grandes núcleos de población unos pocos kilómetros.

Métodos tradicionales y la necesidad de buscar nuevas formas

Hasta ahora las formas de contabilizar a los visitantes se basan en varios sistemas que se han combinado entre sí. Los más tradicionales consisten en contar los vehículos en los principales aparcamientos y, con la estimación de los ocupantes por vehículo, dar cifras que se aproximan bastante a la realidad. El principal inconveniente de este sistema es que requiere de un conteo continuo de vehículos en muchos aparcamientos y del muestreo periódico de ocupación. Además los innumerables accesos y los medios de transporte complican estas estimaciones.

Otras formas empleadas se basan en la estimación a partir de los registros en los Centros de Visitantes y las encuestas de caracterización. Esta cifra también se queda corta, pues muchos de los visitantes que a ellos se acercan visitan el Parque por primera vez.

Desde hace unos años se están empleando contadores o aforadores automáticos, sistemas muy fiables que permiten saber cuántos visitantes pasan por determinadas sendas y en qué sentido de paso. Se ubican estratégicamente en puntos por los que se concentra el paso, pero lógicamente se necesitarían muchísimos dispositivos de este tipo para poder abarcar todo el Parque. Además estos sistemas requieren de un esfuerzo importante de mantenimiento y gestión.

Aunque la combinación de estos sistemas con otras fuentes de datos ha permitido estimar el número de visitantes, desde el área de uso público del Parque Nacional se ha decidido investigar nuevas formas de conteo que mejoren las estimaciones ofrecidas en los últimos años.

Proyecto de caracterización y estimación del número de visitantes a partir de datos de telefonía móvil

¿En qué consiste?

En 2016 y 2017 se ha realizado un estudio piloto para el conteo y la caracterización de los visitantes a partir de datos anonimizados de telefonía móvil. El proyecto, ideado desde el área de uso público del Parque, nace como complemento al resto de las fuentes de información para ofrecer una visión más amplia de la realidad de uso público en el Parque Nacional y en el Área de Especial Protección de los Montes de Valsaín.

¿Qué se obtiene?

Este proyecto tiene varios objetivos que de forma resumida se agrupan en la estimación del número de visitantes y su caracterización. Se trata de conocer la cifra de visitantes clasificados por edad, por género, su residencia o el tiempo de estancia en el Parque y la recurrencia de la visita, es decir, si han estado en el Parque en más de una ocasión en el periodo de un mes alrededor del día de estudio.

También se obtienen datos de los accesos por carretera utilizados por los visitantes. No son simplemente las cifras dadas por los aforadores de las carreteras, sino cuántos de esos registros son visitantes en cada uno de los puntos de entrada y salida del Parque por carretera.

Otro de los aspectos más interesantes de este proyecto es conocer la concentración de visitantes en puntos de interés o áreas difusas con importante uso público como son las Dehesas de Cercedilla, Puertos de Los Cotos, Navacerrada, Navafría y Morcuera, Canto Cochino y La Barranca.

¿Cómo se hace?

Para abordar esta idea, el Parque Nacional buscó entre los distintos proveedores de telefonía móvil aquellos que se ajustaban mejor a las necesidades de conteo y a las dificultades de un proyecto de esta envergadura. El equipo técnico del área de uso público y el Centro de Investigación, Seguimiento y Evaluación del Parque Nacional plantearon distintas opciones metodológicas contando con las formas disponibles para la obtención de datos, otras cifras que apoyaran y fortalecieran los cálculos, y las fórmulas de validación precisas para verificar los resultados.

Finalmente se contó con Orange como suministrador de datos, que con su 27,5% de cuota de mercado, abarca también a los usuarios de Jazztel, Amena y Simyo. La limpieza, el cálculo y los análisis han sido realizados por Kineo Mobility Analytics trabajando de la mano con los técnicos

del Parque Nacional hasta afinar al máximo la precisión del estudio. Para ello se han desarrollado nuevas encuestas y muestreos, ideando formas de confirmación y corrección de los algoritmos. Mediante cálculos estadísticos complejos y modelos matemáticos se proyectan los datos al resto de la población.

Un aspecto relevante que conviene destacar es que los datos suministrados han sido *anonimizados*, es decir, se les ha quitado toda la información que pudiera permitir identificar los datos personales de los visitantes, hasta el punto de que si en una zona se registran menos de 10 usuarios, no se suministran los datos para asegurar el anonimato.

Los datos con los que se trabaja contienen las posiciones geolocalizadas de los dispositivos móviles, tanto para eventos activos (llamadas, SMS, tráfico en Internet, etc.) como para algunos eventos pasivos (transferencia de comunicaciones entre antenas, actualización de las conexiones a Internet, etc.) por triangulación entre antenas, sin necesidad de tener activado el GPS.

Metodología

Sectorización del ámbito de estudio

El primer paso consistió en sectorizar el ámbito de estudio en 15 zonas. El criterio de zonificación se basó en la situación de las grandes áreas de uso y disfrute de los visitantes, de la ubicación de los aforadores automáticos y de las coberturas prestadas por el operador en el área de estudio. Esta división permite conocer con mayor detalle los patrones de comportamiento de los visitantes en cada zona, que en ocasiones son bastante dispares.

Determinación de patrones de actividad y movilidad de la población a partir del análisis de los datos de telefonía móvil

Para la determinación de los patrones de actividad y movilidad de la población a partir de los datos de telefonía móvil se han aplicado los algoritmos desarrollados por Kineo Mobility Analytics. En primer lugar, mediante el análisis de los patrones de comportamiento de varias semanas se determinó para los distintos usuarios su lugar de residencia, su lugar de trabajo y otros lugares frecuentemente visitados. Del mismo modo, para los días concretos de análisis se realizó un estudio en detalle de la movilidad de dichos usuarios, así como de las actividades realizadas y los lugares visitados a lo largo del día. Aplicando modelos estadísticos sobre esta información se obtuvieron los patrones de movilidad elevados al total de la población. A partir de esta información y la fusión con otras fuentes de datos (información de la red de carreteras, contadores de pisada, etc.) se calcularon los indicadores para la caracterización de los visitantes al Parque

Determinación de visitantes

Un factor clave en este estudio es distinguir qué usuarios detectados por el sistema son realmente visitantes al Parque, discriminando los que atraviesan el Parque por carretera o los que son detectados por ser residentes de la zona aunque no entren al Parque. En este sentido

hay que tener en cuenta que el relieve del Parque crea sombras de cobertura que introducen incertidumbre respecto a la localización, con lo que no es posible saber si los usuarios son visitantes o se encuentran en poblaciones cercanas. Para hacer esta distinción se definieron una serie de criterios que determinarán qué personas son visitantes y cuáles no. Los criterios utilizados fueron:

- Usuarios identificados como residentes y/o trabajadores no se consideran visitantes. Como muchos de los visitantes sí son residentes, se utilizó posteriormente un factor corrector a partir de las encuestas de tipificación para meter a los residentes que visitan el Parque, y así no perder estos datos.
- Los que están menos de 30 minutos no se consideran visitantes sino que están de paso.
- Los que sin ser residentes, pernoctan en el parque o en su entorno y pasan más de 8 horas de estancia se excluyen del estudio en un principio. Para no perderlos en las estimaciones también se utilizan factores de corrección obtenidos a partir de las encuestas para conocer cuántos de los que se alojan en el entorno (hotel, segunda residencia, camping, etc.) entran en el Parque.
- Visitantes menores de 16 años. Dado que las muestras de usuarios de telefonía móvil para ese segmento de la población son muy bajas, se excluyeron inicialmente. Para incluir a este segmento de población, se ha utilizado un factor corrector a partir de las encuestas de tipificación que proporcionan información sobre el porcentaje de visitantes que son menores.

Cálculo del flujo de visitantes en las entradas y salidas del Parque en vehículo

Para saber por qué carreteras acceden los visitantes, se han combinado los datos de telefonía con los de la red de carreteras. Mediante el análisis de los datos de telefonía se determinaron el origen, el destino y las posiciones intermedias de todos los viajes hacia o desde el Parque realizados por los visitantes. Fusionando esta información con datos de la red de carreteras y utilizando un modelo de asignación de ruta, se obtuvo la ruta más probable para cada viaje y se contabilizó el número de visitantes por cada uno de los accesos al Parque.

Cálculo de afluencia a los puntos de interés

En el Parque hay ciertos puntos de especial interés para los visitantes por ser áreas recreativas difusas con un importante e histórico uso público, donde es difícil el conteo con los métodos tradicionales. Los visitantes de estas áreas, en términos generales, no tienen el mismo comportamiento que los senderistas, pues mayoritariamente realizan actividades de ocio y esparcimiento en torno a estas zonas, sin penetrar en el Parque. Por este motivo se presentan dos estimaciones:

- Estimación: nº de visitantes que permanece al menos 4 horas. Este indicador proporciona la mejor estimación del número de visitantes mediante los criterios definidos.

- Máximo: nº de visitantes que permanece entre 30 minutos y 4 horas. Este indicador proporciona un máximo razonable de visitantes que puede haber en el punto.

Selección de los días analizados

Como se ha explicado, estos cálculos son muy complejos y, por tanto, costosos. Dado el gran volumen de información que hay que procesar es imprescindible restringir los análisis sólo una serie de días para extrapolar los resultados a todo el año, siempre con la significancia estadística suficiente. La elección de los días a analizar se realiza de la siguiente manera:

- Los datos de los contadores automáticos, aunque sean sólo de ciertas zonas del Parque, ofrecen la distribución de visitantes al año, conociendo qué meses tienen más visitantes, cuál es la proporción entre festivos y laborables, etc. Estos datos permiten elaborar los promedios diarios, por estaciones, por festivos, etc.
- Se hace la selección de los días concretos para el análisis con los datos de los contadores, eligiendo los días cuyos valores medios son más próximos a los promedios para cada periodo. Así se piden al proveedor el análisis de ciertos días concretos.

Resultados finales

Con los datos recibidos se calculan los promedios para cada resultado en fin de semana y en día laborables en cada época del año, tanto por sectores, como por provincia y del total del Parque. Estos promedios se multiplican por el número de días reales de cada periodo, obteniendo los resultados finales.

Estimación total de visitantes en 2017 en datos absolutos										
	INVIERNO		PRIMAVERA		VERANO		OTOÑO		TOTAL	
Madrid	357.669	13,3%	495.881	18,4%	405.352	15,1%	272.491	10,1%	1.531.393	56,9%
Segovia	191.682	7,1%	413.536	15,4%	325.178	12,1%	230.191	8,6%	1.160.587	43,1%
Total	549.351	20,4%	909.417	33,8%	730.530	27,1%	502.682	18,7%	2.691.980	100,0%

Proyección: elaboración del modelo matemático extrapolable

Estos datos han sido validados con las series dadas por los contadores automáticos, consiguiendo una significancia estadística más que suficiente para poder elaborar los modelos predictivos de años venideros.

Al contar con la sectorización mencionada anteriormente, se conocen los datos relativos de cada sector respecto del total del Parque. Estos porcentajes y la correlación con las series de datos de los contadores automáticos, posibilitan la elaboración de los modelos matemáticos necesarios para poder seguir estimando el número de visitantes.

No obstante, el estudio no se queda aquí, sino que se sigue trabajando con Orange y Kineo para mejorar el sistema, viendo cómo ser aún más precisos, proponiendo la instalación de aforadores en zonas poco cubiertas con estos sistemas y así mejorar los coeficientes de

correlación, y planteando nuevos análisis o cruces de datos para conseguir el objetivo con el que arrancaba esta entrada al blog: conocer realmente cuántos somos y cómo nos distribuimos.