Rev.R.Acad.Cienc.Exact.Fís.Nat. (Esp)
Vol. 102, Nº. 1, pp 251-263, 2008
IX Programa de Promoción de la Cultura Científica y Tecnológica

¿DE DÓNDE ERA PROBABLEMENTE D. QUIJOTE? UN ENFOQUE ESTADÍSTICO

FCO. JAVIER GIRÓN GONZÁLEZ-TORRE* Y M. JESÚS RÍOS INSUA**

* Real Academia de Ciencias Exactas, Físicas y Naturales. Valverde 22, 28004 Madrid.
** Universidad Complutense de Madrid.

"En un lugar de la Mancha de cuyo nombre no quiero acordarme …"

Desde la primera frase, Cervantes establece el principio que regula toda la redacción de 'El Quijote': la incertidumbre. Incertidumbre autoral, incertidumbre nominativa, incertidumbre genérica y, en consecuencia, el ojo crítico de la imaginación, dándole "una segunda realidad" a lo que pasa por ser "realidad".

Carlos Fuentes (2005)

RESUMEN

Cervantes —aunque no quiere decir en dónde vivía Don Quijote al principio del libro—, va dejando caer a lo largo del texto, con motivo de sus viajes, referencias acerca de las distancias entre ese lugar de la Mancha y otros pueblos, como El Toboso, Puerto Lápice, Sierra Morena, destinos de sus tres salidas e, incluso, algún otro. En este artículo pretendemos demostrar que esas distancias son consistentes —dentro de unos ciertos límites— y permiten determinar con bastante fiabilidad dicho lugar, teniendo en cuenta que de la lectura del libro se tiene idea del tiempo empleado por Don Quijote y Sancho en llegar a los distintos pueblos mencionados.

Comentamos algunas de las soluciones posibles, como la geométrica, así como la basada en la *Teoría de la Decisión Multicriterio,* señalando sus limitaciones y deficiencias. Ambas soluciones suponen que se conoce la velocidad con la que Rocinante y Rucio cabalgaban y que además ésta es constante en todas las salidas de D. Quijote. Ambos enfoques están íntimamente relacionados, como se demuestra en el desarrollo del artículo.

La solución estadística del problema se revela como la mejor opción puesto que no solamente ordena los pueblos candidatos a ser el lugar de la Mancha sin necesidad de conocer la velocidad ni suponerla constante, sino que, además, permite calcular la probabilidad a posteriori de cada uno de ellos —objetivo fundamental de este estudio—. También, como subproducto del modelo estadístico, se puede estimar una cota inferior de las velocidades a los diferentes destinos.

1. INTRODUCCIÓN

Muchos —curiosos y estudiosos— han sido los que desde la publicación de la primera parte del *Quijote* por Miguel de Cervantes en 1605, han intentado situar el *lugar de la Mancha* dentro del territorio geográfico de esa zona. Algunos teóricos de la Literatura han opinado que ese lugar era puramente imaginario, por lo que no tendría sentido buscarlo. No obstante, está claro que Cervantes —desde la primera frase *«En un lugar de la Mancha de cuyo nombre no quiero acordarme […]»*, hasta uno de los últimos párrafos del último capítulo de la segunda parte de la novela *«Este fin tuvo el Ingenioso Hidalgo de la Mancha, cuyo* *lugar no quiso poner Cide Hamete puntualmente, por dejar que todas las villas y lugares de la Mancha contendiesen entre sí por ahijársele y tenérsele por suyo, como contendieron las siete ciudades de Grecia por Homero*»—, va dejando caer pistas que, aunque no exentas de inconsistencias en algunos casos, llevan al convencimiento de que hay realismo geográfico en la novela.

De una lectura cuidadosa del *Quijote* se pueden obtener los siguientes datos:

a) El *lugar de la Mancha* está dentro del Campo de Montiel (prólogo; Cap. 1; Cap. VII; Cap. LII, Primera parte y Cap. VII, Segunda Parte).

b) El *lugar de la Mancha* está a dos días de Sierra Morena (SM), (Cap. XXIX, Primera parte).

c) El *lugar de la Mancha* está a una noche y dos días de El Toboso (ET), (Cap. XXXVII, Primera parte).

d) El *lugar de la Mancha* está a dos días y algunas horas de Puerto Lápice (PL), (Caps. XII a XXIV, Primera parte).

e) El *lugar de la Mancha* está entre un día y medio y dos días del *Punto Tarfe* (Munera) (Cap. LXXII, Segunda parte).

f) La velocidad de Rocinante/Rucio está comprendida entre los 30 y 35 kms por jornada (Caps. XI, XXVII, Primera parte).

Al cumplirse, en el año 2005, el cuarto centenario de la publicación del *Quijote*, un trabajo llevado a cabo por un equipo multidisciplinar de la UCM (Parra et al., 2005) *resuelve*, desde distintos enfoques (literarios, sociológicos, topológicos, …), el problema de situar geográficamente el *lugar de la Mancha*. En un segundo trabajo presentado en las XII JAEM (Ríos, Montero y Parra, 2005) se aborda directamente el problema desde el punto de vista de la *Teoría de la Decisión Multicriterio* asociando a cada uno de los pueblos candidatos a ser el lugar de la Mancha un vector de discrepancias, construyendo distintas funciones de valor que se basan en las distancias o métricas usuales, lo que conduce a varias posibles soluciones que proporcionan distintos órdenes de preferencia de los pueblos candidatos, teniendo además en cuenta los diversos valores de la velocidad. En los comentarios y críticas sobre estos enfoques se pone de manifiesto que, dada la incertidumbre inherente a los modelos matemáticos y los datos, se hace necesario emplear un modelo estadístico para valorar dicha incertidumbre que, además, permita calcular las probabilidades a posteriori de cada uno de los pueblos candidatos a ser el lugar de la Mancha. La cita introductoria del gran escritor mexicano Carlos Fuentes sobre el papel de la incertidumbre en el Quijote se nos antoja como una premonición para enfocar el problema de determinar el *lugar de la Mancha* desde una disciplina como es la Estadística, que tiene como objetivo fundamental el estudio de aquellos problemas en los que hay presente incertidumbre en cualquiera de sus formas o manifestaciones.

El artículo se estructura de la siguiente manera: en la sección 2 se describen los datos que se obtienen de la lectura del libro, haciendo la salvedad de que los correspondientes a las jornadas recorridas son solamente aproximaciones y que además estos datos contienen incertidumbre, a diferencia de lo que ocurre con las distancias entre los pueblos de origen y los cuatro destinos, que son precisas. Por razones de espacio y de complejidad en los modelos estadísticos, en este artículo no se considera el problema de incluir incertidumbre dicional en el modelo cuando hay datos inciertos. En la sección 3 se resumen los enfoques previos encaminados a resolver el problema y se hace una breve revisión del *método geométrico* y del basado en la *Teoría de la Decisión Multicriterio*, señalando sus puntos débiles. En la sección 4 se proponen tres posibles modelos estadísticos para determinar el lugar de la Mancha. Usando técnicas de selección de modelos se elige el mejor de ellos, se estiman sus parámetros y, finalmente, se calculan las probabilidades a posteriori de los pueblos candidatos. Por último, en la sección 5, se plantean brevemente las posibles extensiones del modelo básico.

2. DESCRIPCIÓN DE LOS DATOS

Tal como se ha explicado en la introducción, se tiene una idea aproximada de las jornadas que necesitaron las cabalgaduras en las tres salidas a Puerto Lápice, Sierra Morena y El Toboso, más la correspondiente al denominado *punto Tarfe* que Parra *et al.* (2005) sitúan finalmente en Munera. Estas jornadas, medidas en días, las representamos por el vector $d = (d_1, d_2, d_3, d_4)$, donde $d_1 = 2.25$, $d_2 = 2.00$, $d_3 = 2.50$ y $d_4 = 1.75$.

Pueblos	PL	SM	ET	MU	Pueblos	PL	SM	ET	MU
Albadalejo	96.8	63.2	99.6	54.4	Ossa	74.4	88.0	64.0	24.0
Alcubillas	70.0	48.4	84.4	64.4	Puebla P.	96.0	51.6	104.0	64.8
Alhambra	59.6	65.2	67.6	51.2	Ruidera	64.8	80.8	59.6	35.6
Almedina	89.2	52.0	97.6	61.2	S. C. Caña.	92.4	59.6	96.8	55.2
Argamasilla	40.0	84.8	43.2	53.2	Solana	47.2	61.6	66.0	65.6
Cañamares	89.6	73.6	87.2	40.0	Terrinches	95.6	60.4	100.0	56.8
Carrizosa	67.6	63.6	73.6	49.2	T. J. Abad	89.2	45.2	102.0	70.8
Castellar	88.4	22.8	109.6	88.0	Torrenueva	76.0	26.8	100.8	87.6
Cozar	81.2	44.8	94.0	66.0	Villahermosa	82.0	64.8	84.4	46.4
Fuenllana	77.6	59.6	83.6	51.6	Villamanrique	95.2	45.2	106.4	70.4
Membrilla	40.8	62.0	66.8	74.8	Villa Fuente	96.4	76.0	93.2	41.6
Montiel	86.8	62.8	90.0	49.6	Villa Infantes	76.4	54.4	85.6	56.8

Tabla 1. Distancias en kms, medidas en línea recta, de los veinticuatro pueblos candidatos a los cuatro destinos. Fuente: Mapa de carreteras Michelin 576 Regional España.

Por otra parte, no se conocen con precisión las distancias reales a los cuatro destinos desde los veinticuatro pueblos candidatos a ser el *lugar de la Mancha* al no saber con seguridad los caminos y veredas de la época que conducían a estos destinos. Por ello, en esta primera aproximación al problema, hemos tomado como datos las distancias en línea recta desde el pueblo de origen al destino, ya que son fáciles de medir y constituyen una cota inferior de las distancias reales recorridas. Los datos, en kilómetros, se han obtenido de la guía Michelin y aparecen en la Tabla 1.

Otra alternativa a las distancias entre pueblos y destinos es considerar los datos de las coordenadas geográficas de los veinticuatro pueblos candidatos y de los cuatro destinos. Esta información, expresada en coordenadas cartesianas tomando como origen de coordenadas Sierra Morena, se recoge en la Tabla 2, aunque las coordenadas de los destinos no aparecen en ésta. Obviamente de la Tabla 2 se deducen las distancias de la Tabla 1. Recíprocamente, a partir de esas distancias también se pueden calcular las coordenadas geográficas, salvo errores de redondeo.

El por qué considerar ambas fuentes de información se explica por la formulación estadística del problema que se expone en la sección 5.

Todos los pueblos que aparecen en la Tabla 1, exceptuando Argamasilla de Alba, estaban situados en la época de Cervantes en el Campo de Montiel. La razón de que hayamos incluido Argamasilla de Alba en el estudio es sobre todo histórica y tiene como fin confirmar, mediante el análisis estadístico que se describe en la sección 2, la imposibilidad práctica de que Argamasilla fuera el lugar de la Mancha. No olvidemos que Rubén Darío y Azorín, con motivo del tercer centenario de la publicación de la primera parte del Quijote, *determinaron* que la cuna de don Alonso Quijano era Argamasilla de Alba, pero sin dar explicaciones objetivas del por qué.

Acerca de la velocidad de las cabalgaduras, aunque existen datos dentro del propio texto del Quijote que dan una idea de cual podría ser ésta, tal como se señala en el apartado f) de la introducción y en Parra *et al.* (2005), no se han utilizado en este trabajo por varios

Pueblos	x	y	Pueblos	x	y
Albadalejo	59.50	21.56	Ossa	65.01	59.49
Alcubillas	31.88	36.20	Puebla P.	49.20	15.94
Alhambra	38.78	52.41	Ruidera	53.25	60.58
Almedina	47.00	22.35	S. C. Caña.	54.83	23.68
Argamasilla	35.14	77.21	Solana	22.88	57.13
Cañamares	64.69	35.31	Terrinches	56.60	20.76
Carrizosa	43.65	46.18	T. J. Abad	37.50	17.97
Castellar	19.07	12.84	Torrenueva	11.90	24.09
Cozar	36.66	26.04	Villahermosa	54.02	36.04
Fuenllana	47.17	36.29	Villamanrique	43.17	13.53
Membrilla	13.18	60.37	Villa Fuente	69.65	30.56
Montiel	54.99	30.63	Villa Infantes	42.10	34.30

Tabla 2. Coordenadas de los 24 pueblos: Origen de coordenadas Sierra Morena (Venta de Cárdenas). Fuente: Mapa de carreteras Michelin 576 Regional España.

Figura 1. Mapa del Campo de Montiel que se guarda en la *Real Academia de Historia de Madrid* y corrige un error de orientación del original. Cortesía de la *Real Academia de Ciencias Exactas, Físicas y Naturales*.

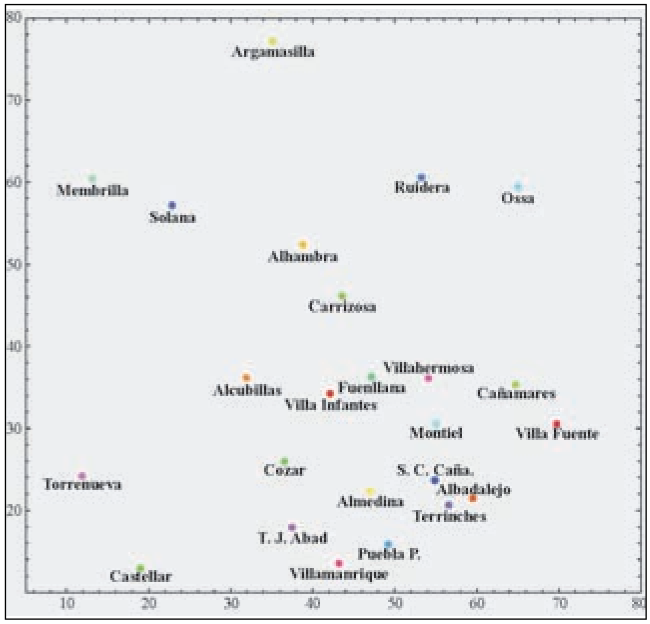


Figura 2. Campo de Montiel, en la actualidad, en el que además se muestra Argamasilla de Alba. Las coordenadas de los pueblos, en kms, se toman considerando como origen de coordenadas Sierra Morena (Venta de Cárdenas), que no aparece en la Figura.

motivos. En primer lugar, la velocidad depende no solamente de los destinos sino también, al menos en principio, de los pueblos de origen y puede ser bastante variable. En segundo lugar, y este es el motivo principal para no incluirla como *parámetro conocido*, es que en el modelo estadístico que proponemos en la sección siguiente para determinar el lugar de la Mancha —que incluye como parámetros subsidiarios las diferentes velocidades— no es necesario conocer éstas.

No obstante, como subproducto del modelo estadístico considerado, cuya finalidad primordial —no lo olvidemos— es *estimar probabilísticamente* el lugar de la Mancha, también se puede dar una estimación de la cota inferior de las velocidades de las cabalgaduras que es consistente con las estimaciones de los expertos. En Parra *et al.* se puede leer que los estudiosos de este tema han establecido que la velocidad de Rocinante/Rucio estaría comprendida entre los 30 y 35 kms por jornada.

Las Figuras 1 y 2 representan el mapa histórico y la situación actual del Campo de Montiel, respectivamente; en la Figura 2 se ha incluído, como hemos comentado, el pueblo de Argamasilla de Alba, situado al norte del Campo de Montiel.

3. NUEVOS ENFOQUES AL PROBLEMA DE DETERMINAR EL LUGAR DE LA MANCHA

Desde hace 400 años se viene intentando encontrar el lugar de la Mancha de cuyo nombre no quiso acordarse Cervantes. Las soluciones al problema han ido desde las cabalísticas, como por ejemplo *el lugar debe tener tantas letras como años la esposa de Cervantes al casarse*, hasta las sociológicas derivadas, por ejemplo, del estudio de las condiciones y características que se describen en el texto cervantino. (Parra *et al*., 2005).

Soluciones más consistentes se basan en considerar la determinación de *El lugar de la Mancha* como el enunciado de un problema geométrico de determinación de un punto o lugar geométrico (de cuyo nombre Cervantes no quiso acordarse), dadas las distancias desde dicho punto a tres/cuatro lugares (de cuyos nombres Cervantes sí se acordó) o, como hemos explicado anteriormente, conociendo las coordenadas geográficas de los pueblos candidatos y los destinos de los viajes.

Otra posibilidad, es considerar el problema desde el punto vista de la *Teoría de la Decisión Multicriterio* que es, como comentaremos, complementario del enfoque geométrico.

El principal inconveniente de estos dos procedimientos es que consideran la misma velocidad de marcha a cada uno de los cuatro destinos y, además, ésta debe suponerse conocida.

Otro inconveniente —aunque quizás menor—, es que estas soluciones solamente proporcionan una ordenación de los pueblos, encaminada a determinar cuál de ellos podría ser el lugar de la Mancha pero, al ser la solución puramente ordinal, ésta no proporciona una medida objetiva fácilmente interpretable de la misma.

La solución que proponemos como más satisfactoria, que se estudia en la sección 5, es considerar el problema como uno en el que hay *incertidumbre* en el modelo matemático que relaciona los pueblos candidatos con los cuatro destinos y, por consiguiente, se hace necesario adoptar un procedimiento estadístico.

Como veremos, las ventajas del enfoque estadístico son: primera, que no es necesaria la hipótesis de velocidad constante a todos los destinos, ni siquiera que ésta sea conocida, pues el modelo estadístico más adecuado nos permitirá comprobar que, en efecto, las velocidades dependen solamente del destino; segunda, que permite eliminar del estudio siete de los veinticuatro pueblos candidatos a ser el lugar de la Mancha y, tercera, nos da la medida de la incertidumbre acerca del lugar de la Mancha en términos de probabilidades (una escala de medir la incertidumbre mucho más precisa que una simple escala ordinal, que es la que proporcionan los otros dos enfoques).

4. LA SOLUCIÓN GEOMÉTRICA Y LA BASADA EN LA TEORÍA DE LA DECISIÓN MULTICRITERIO

La solución geométrica del problema consiste, tal como puede apreciarse en la Figura 3, en trazar circunferencias con centro en cada uno de los cuatro destinos y radios proporcionales a la velocidad de las cabalgaduras v hasta que éstas —las circunferencias— se intersequen, si acaso esto ocurriese, aunque sea de forma aproximada. Esta solución supone que la velocidad a los cuatro destinos es la misma. En la Figura 3 se aprecia que para una velocidad v de 30 kms por jornada aproximadamente, las cuatro circunferencias confluyen en un punto muy próximo a Carrizosa, con lo cual, en principio, la solución geométrica del problema de localizar el lugar de la Mancha sería el pueblo de Carrizosa y la velocidad de las cabalgaduras se situaría en unos 30 kms por jornada aproximadamente, suponiendo que fuesen en línea recta.

Sin embargo, si examinamos con más detalle el procedimiento geométrico, se puede comprobar que para velocidades mayores que 30 kms por jornada, las cuatro circunferencias no se intersecan en un mismo punto, sino solamente tres de ellas e incluso en este caso, de modo aproximado. La que tiene su centro en Sierra Morena (de color verde), se aleja del punto de confluencia de las otras tres a medida que la velocidad aumenta. Como acabamos de comentar, estas últimas circunferencias solamente se intersecan de modo aproximado en un punto, dibujando una trayectoria, tal como se representa en la Figura 4, que va desde Carrizosa para una velocidad de 30 kms por jornada, hasta un punto próximo a Villanueva de los Infantes para una velocidad de 35–36 kms por jornada aproximadamente. Esto pone de manifiesto que el considerar una velocidad constante para los cuatro destinos no parece una hipótesis razonable. Intuitivamente, la velocidad en el camino hacia Sierra Morena debería ser menor que la de los otros destinos.

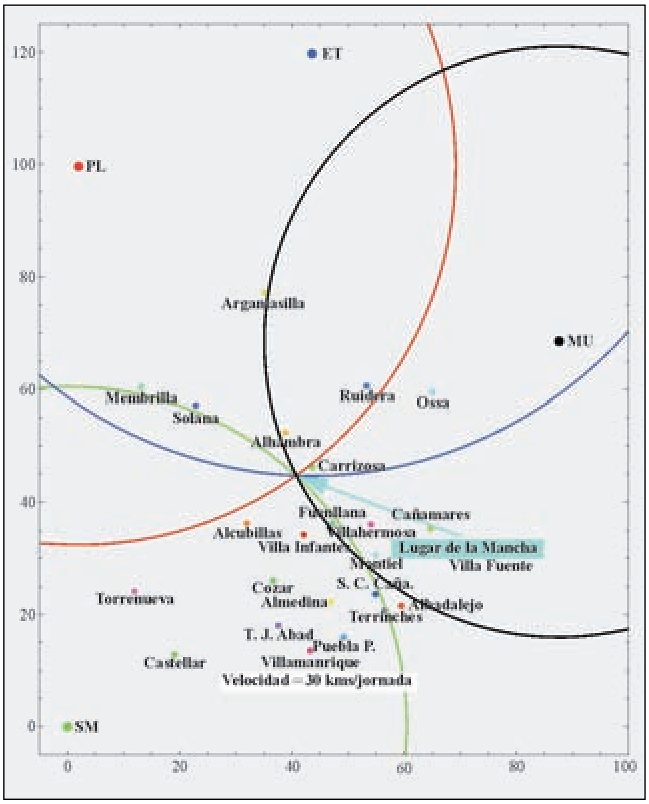


Figura 3. Ilustración de la solución geométrica. El punto de intersección aproximado de las cuatro circunferencias se obtiene para una velocidad $v \approx 30$ kms por jornada.

De lo anterior podemos entresacar algunos comentarios y posibles deficiencias de la solución geométrica:

1º. El suponer que los cuatro recorridos se realizan con velocidad constante no parece razonable. La velocidad para llegar a Sierra Morena (SM) debería ser menor que las demás.

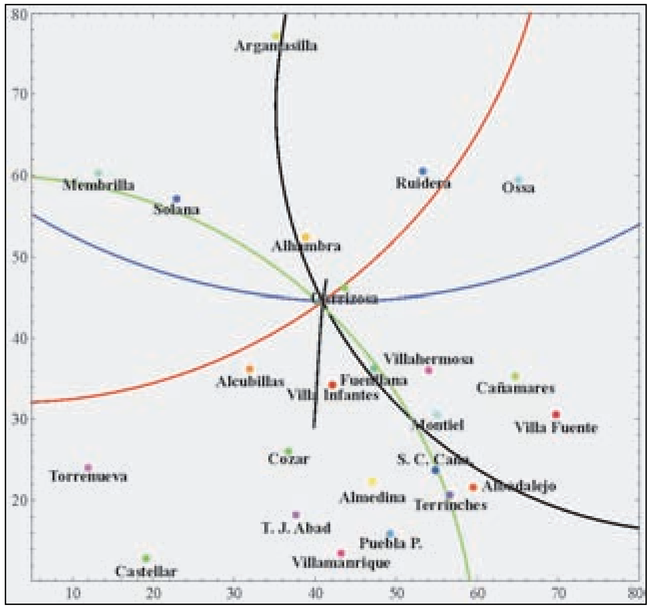


Figura 4. Trayectoria (en negro) de la intersección de las tres circunferencias con origen en Puerto Lápice, El Toboso y Munera, al variar la velocidad de 30 a 36 kms por jornada. La circunferencia con origen en Sierra Morena no pasa por el punto de intersección de las otras tres para velocidades superiores a 30 kms por jornada.

2º. El suponer que los recorridos se realizan en línea recta —lo que justificaría utilizar la distancia euclídea en la solución geométrica— tampoco parece razonable. Hay destinos a los que puede ser más complicado llegar, como por ejemplo Sierra Morena.

3º. En cualquier circunstancia, sí parece razonable que las velocidades podrían estar comprendidas entre 28 y 36 kms/jornada.

4º. Los pueblos situados aproximadamente en el eje Norte-Sur de la Figura 4, desde Alhambra hasta Cózar, pasando por Alcubillas, Fuenllana y Villanueva de los Infantes, serían candidatos a ser el lugar de la Mancha desde la perspectiva geométrica.

El otro enfoque al problema, basado en la Teoría de la Decisión Multicriterio, lo podemos resumir de la siguiente manera, de modo que no sea necesario detallar dicha teoría, sino simplemente adaptarla a nuestro problema.

Para ello, a cada uno de los veinticuatro pueblos candidatos a ser el lugar de la Mancha P, y a cada valor de la velocidad en línea recta v, se le asigna un vector de discrepancias $\delta(P, v) = (\delta_1(v), \delta_2(v), \delta_3(v), \delta_4(v))$ —o de *criterios*, en el lenguaje de la Teoría de la Decisión Multicriterio—, que representa las distancias euclídeas de un punto genérico del Campo de Montiel P a cada una de las cuatro circunferencias consideradas en la solución geométrica, que en cierto modo mide las discrepancias de cada pueblo candidato a cada destino, para cada velocidad v, tal como se puede ver en la Figura 5.

La solución del problema se reduce ahora a comparar u ordenar estos vectores, lo que no es inmediato, salvo que haya dominancia coordenada a coordenada y, en este caso, elegir aquel pueblo que minimice todas las coordenadas lo que, en general, no es factible ni siquiera en nuestro caso.

Si para algún pueblo P y una cierta velocidad v el vector de discrepancias fuese el vector nulo (0, 0, 0, 0), ese pueblo sería la solución del problema. En lenguaje de la Teoría de la Decisión Multicriterio, esa sería la solución que domina a cualquier otra, al ser las discrepancias cantidades positivas.

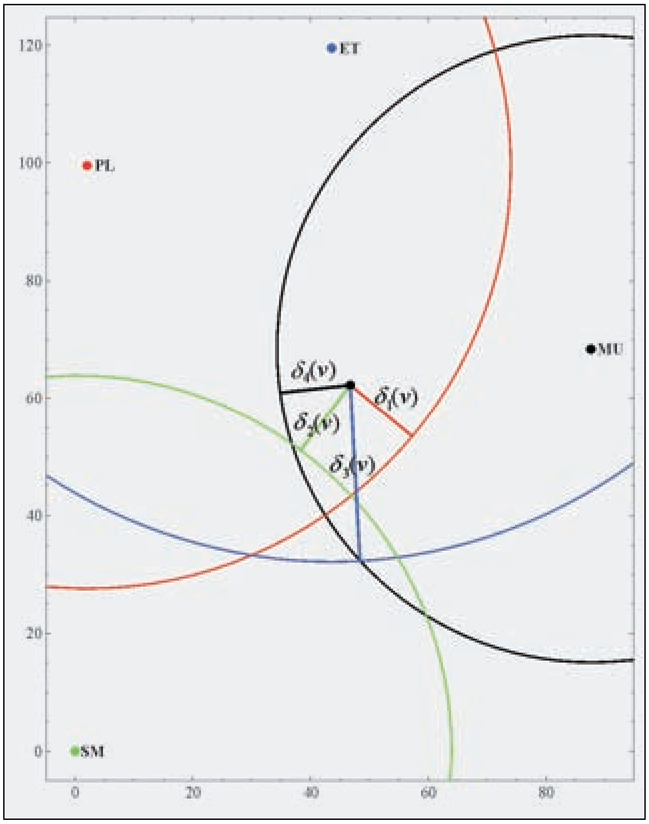


Figura 5. Discrepancias asociadas a cada lugar del Campo de Montiel y a cada uno de los cuatro destinos, en función de la velocidad v de las cabalgaduras.

Pero, en general, la dominancia no ocurre para ningún pueblo P ni para ningún valor de la velocidad v. Ni siquiera para ningún punto del Campo de Montiel.

Cuando no hay dominancia —el caso más frecuente—, las posibles soluciones se basan en construir o considerar una *función de valor* $Z(\delta(P, v))$ que asocia a cada pueblo P y a cada velocidad v un número positivo que permita comparar, en nuestro caso, los veinticuatro pueblos.

En la literatura sobre la Teoría de la Decisión Multicriterio se pueden encontrar diversas funciones de valor. Entre las más utilizadas, tenemos las siguientes:

$$Z_1(\delta_1(v),\delta_2(v),\delta_3(v),\delta_4(v)) = \delta_1(v)+\delta_2(v)+\delta_3(v)+\delta_4(v),$$
$$Z_2(\delta_1(v),\delta_2(v),\delta_3(v),\delta_4(v)) = \delta_1(v)^2+\delta_2(v)^2+\delta_3(v)^2+\delta_4(v)^2,$$
$$Z_\infty(\delta_1(v),\delta_2(v),\delta_3(v),\delta_4(v)) = \max\{\delta_1(v),\delta_2(v),\delta_3(v),\delta_4(v)\}.$$

En la Figura 6 se presenta una solución del problema cuando la velocidad se supone constante e igual a 34 kms por jornada y la función de valor utilizada es la Z_1. Para cada pueblo, se calcula el vector de discrepancias y en la última columna se calcula la función de valor para cada uno de los pueblos, que en este caso es la suma de las discrepancias. La solución sería el pueblo para el cual esa función fuese mínima. En nuestro ejemplo, con los datos anteriores, éste sería *Villanueva de los Infantes* ya que hace mínima la función de valor Z_1 para esa velocidad.

Los inconvenientes de la solución basada en la Teoría de la Decisión Multicriterio son similares a los de la solución geométrica:

1º. El suponer que los recorridos se realizan con velocidad constante no parece razonable.

2º. El suponer que los recorridos se realizan en línea recta tampoco parece razonable.

3º. La solución depende de la velocidad que, por otra parte como ya hemos comentado, se supone igual para los cuatro destinos, que no se conoce.

4º. La solución depende, *a priori*, de la función de valor que se elija para ordenar, aunque afortunadamente hay cierta robustez en la solución, es decir, no depende excesivamente de la función de valor elegida.

La figura 7 ilustra la íntima relación que existe entre la solución geométrica y la basada en la Teoría de la Decisión Multicriterio, para una velocidad v de 33 kms por jornada y función de valor la Z_1.

5. EL ENFOQUE ESTADÍSTICO

En los dos enfoques de la sección anterior para localizar el lugar de la Mancha, los datos que se utilizan son exclusivamente los de la Tabla 1, pues ambos dependen de las distancias de los veinticuatro pueblos a los cuatro destinos y, obviamente, de la duración de las jornadas para desplazarse a estos.

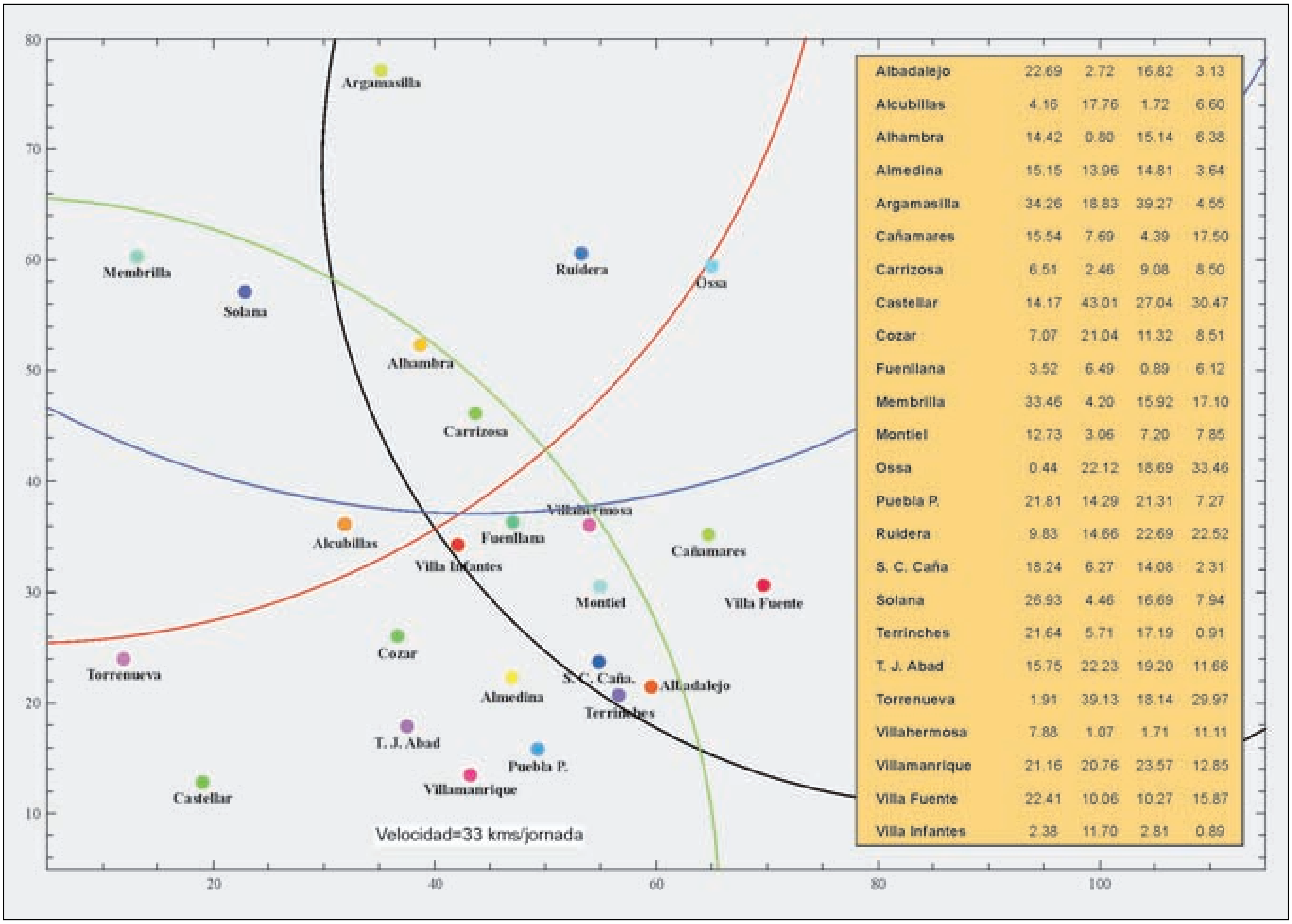
Sin embargo, la perspectiva de optar por un enfoque estadístico, nos ofrece las dos posibilidades complementarias:

1ª. Considerar como datos la duración de las jornadas y las distancias de los pueblos a los destinos dados en la Tabla 1, que serán importantes para estimar los parámetros del modelo estadístico, que es el paso previo y necesario para dar solución probabilística a la pregunta básica de cúal es el pueblo más probable de ser el lugar de la Mancha. Este punto de vista proporciona, además, estimaciones de la cota inferior de las velocidades a los cuatro destinos.

Velocidad=34 kms/jornada

Pueblos	PL	SM	ET	MU	L_1
Albadalejo	20.44	4.72	14.32	4.88	44.35
Alcubillas	6.41	19.76	0.78	4.85	31.80
Alhambra	16.67	2.80	17.64	8.13	45.24
Almedina	12.90	15.96	12.31	1.89	43.05
Argamasilla	36.51	16.83	41.77	6.30	101.40
Cañamares	13.29	5.69	1.89	19.25	40.13
Carrizosa	8.76	4.46	11.58	10.25	35.04
Castellar	11.92	45.01	24.54	28.72	110.20
Cozar	4.82	23.04	8.82	6.76	43.43
Fuenllana	1.27	8.49	1.61	7.87	19.24
Membrilla	35.71	6.20	18.42	15.35	75.69
Montiel	10.48	5.06	4.70	9.60	29.83
Ossa	1.81	20.12	21.19	35.21	78.33
Puebla P.	19.56	16.29	18.81	5.52	60.17
Ruidera	12.08	12.66	25.19	24.27	74.20
S. C. Caña	15.99	8.27	11.58	4.06	39.90
Solana	29.18	6.46	19.19	6.19	61.01
Terrinches	19.39	7.71	14.69	2.66	44.46
T. J. Abad	13.50	24.23	16.70	9.91	64.34
Torrenueva	0.34	41.13	15.64	28.22	85.33
Villahermosa	5.63	3.07	0.79	12.86	22.35
Villamanrique	18.91	22.76	21.07	11.10	73.84
Villa Fuente	20.16	8.06	7.77	17.62	53.62
Villa Infantes	0.13	13.70	0.31	2.64	16.77

Figura 6. Ejemplo de aplicación de la Teoría de la Decisión Multicriterio cuando la velocidad v es de 34 kms por jornada y la función de valor es la Z_1, que da como solución óptima el pueblo de Villanueva de los Infantes.



Albadalejo	22.69	2.72	16.82	3.13
Alcubillas	4.16	17.76	1.72	6.60
Alhambra	14.42	0.80	15.14	6.38
Almedina	15.15	13.96	14.81	3.64
Argamasilla	34.26	18.83	39.27	4.55
Cañamares	15.54	7.69	4.39	17.50
Carrizosa	6.51	2.46	9.08	8.50
Castellar	14.17	43.01	27.04	30.47
Cozar	7.07	21.04	11.32	8.51
Fuenllana	3.52	6.49	0.89	6.12
Membrilla	33.46	4.20	15.92	17.10
Montiel	12.73	3.06	7.20	7.85
Ossa	0.44	22.12	18.69	33.46
Puebla P.	21.81	14.29	21.31	7.27
Ruidera	9.83	14.66	22.69	22.52
S. C. Caña	18.24	6.27	14.08	2.31
Solana	26.93	4.46	16.69	7.94
Terrinches	21.64	5.71	17.19	0.91
T. J. Abad	15.75	22.23	19.20	11.66
Torrenueva	1.91	39.13	18.14	29.97
Villahermosa	7.88	1.07	1.71	11.11
Villamanrique	21.16	20.76	23.57	12.85
Villa Fuente	22.41	10.06	10.27	15.87
Villa Infantes	2.38	11.70	2.81	0.89

Figura 7. Relación entre el enfoque geométrico y el basado en la Teoría de la Decisión Multicriterio. Comparación cuando la velocidad v es de 33 kms por jornada y la función de valor es la Z_1

2ª. Considerar como datos básicos la duración de las jornadas y las coordenadas de los pueblos respecto de un cierto origen de coordenadas, dados en la Tabla 2. Para ello, hemos elegido como origen de coordenadas la Venta de Cárdenas situada en Sierra Morena por ser el lugar geográfico situado más al sur y al oeste respecto del Campo de Montiel. Los resultados que se obtienen son, obviamente, independientes de la elección del origen. Este punto de vista alternativo es útil para calcular las probabilidades a posteriori de cada pueblo de ser el lugar de la Mancha, objetivo principal de cualquier estudio sobre el tema.

El enfoque que describimos en esta sección es totalmente general en lo que se refiere al número de pueblos y de destinos.

Supongamos, pues, que hay m pueblos de origen a ser candidatos al lugar de la Mancha y k destinos. Los datos que se conocen son:

1. $d = (d_1, \ldots, d_j, \ldots, d_k)$, donde d_j representa la jornada empleada por las cabalgaduras desde el posible lugar de la Mancha hasta el destino j-ésimo.
2. x_{ij} representa la distancia euclídea entre el i-ésimo pueblo de origen y el destino j- ésimo.

Los parámetros que interviene en el modelo son:

3. $\theta = (\theta_1, \ldots, \theta_j, \ldots, \theta_k)$ punto del espacio paramétrico $\Omega = R^{+k}$, donde θ_j representa la distancia euclídea desde un punto genérico —no necesariamente un pueblo— cualquiera del mapa del Campo de Montiel y sus entornos al destino j-ésimo.

4. v_{θ_j} representa la *velocidad* de las cabalgaduras en kms/jornada, desde un pueblo genérico arbitrario θ al destino j-ésimo.

5. γ_{θ_j} representa el *factor de inflación de la distancia* desde un pueblo genérico arbitrario θ al destino j-ésimo. Representa el factor (mayor que 1) que hay que aplicar a la distancia euclídea θ_j para obtener la distancia real entre el pueblo θ y el destino j-ésimo.

6. σ^2 es un factor de variabilidad.

La ecuación básica que han de cumplir los datos y los parámetros es la siguiente, que nos dice que la distancia real recorrida entre el pueblo i-ésimo y el destino j-ésimo es aproximadamente igual al producto de la velocidad por la jornada correspondiente, es decir:

$$\gamma_{\theta_j} x_{ij} \approx v_{\theta_j} d_j \text{ y } \gamma_{\theta_j}\theta_j \approx v_{\theta_j} d_j \text{ para todo } i=1,\ldots,m;\ j=1,\ldots,k \text{ y } \theta \in \Omega. \tag{5.1}$$

Tenemos, por consiguiente, una única observación d que depende de los parámetros θ, v_{θ_j}, γ_{θ_j} y σ^2, que mediría el grado de aproximación de la fórmula (5.1). El modelo estadístico que proponemos, basado en esta fórmula, es el siguiente: Para $j = 1,\ldots, k$ se tiene

$$d_j \left| \theta_j, v_{\theta_j}, \gamma_{\theta_j}, \sigma^2 \sim N\left(d_j \left| \gamma_{\theta_j}/v_{\theta_j}\theta_j, \sigma^2\theta_j^2\right.\right)\right.. \tag{5.2}$$

La justificación del modelo se basa en que la hipótesis sobre la esperanza de d_j se corresponde con la ecuación (5.1), mientras que la varianza de la observación d_j depende del factor de varianza σ^2 y del parámetro θ_j, de tal forma que el coeficiente de variación no dependa de éste último. Es decir, estamos suponiendo que a las distancias más largas corresponde mayor variabilidad que a las más cortas.

La última hipótesis sobre el modelo, que representa que las salidas son independientes, es que para j= 1,…, k las dj son condicionalmente independientes, con lo cual el modelo (5.2) se puede escribir como una normal k variante

$$d \left| \theta, M_\theta, \sigma^2 \sim N_k\left(d \left| M_\theta \theta, \sigma^2 D(\theta)\right.\right)\right., \tag{5.3}$$

donde M_θ es una matriz diagonal de dimensiones $k\times k$, en la que cada elemento de la diagonal principal es $\gamma_{\theta_j}/v_{\theta_j}$ y $D(\theta)$ es una matriz diagonal de dimensiones $k\times k$, cuyos elementos son θ_j^2.

El primer paso para obtener las probabilidades a posteriori de los m pueblos candidatos a ser el lugar de la Mancha es *estimar* los parámetros v_{θ_j}, γ_{θ_j} y σ^2 del modelo con los datos d_j y x_{ij}. La hipótesis de independencia, para todo i = 1,…,m y j = 1…, k, es del todo plausible en nuestro caso, de modo que tenemos el modelo

$$d_j \left| x_{ij}, v_{ij}, \gamma_{ij}, \sigma^2 \sim N\left(d_j \left| \mu_{ij} x_{ij}, \sigma^2 x_{ij}^2\right.\right)\right., \tag{5.4}$$

donde $\mu_{ij} = \gamma_{ij}/v_{ij}$. Definiendo los nuevos datos $y_{ij} = d_j/x_{ij}$, el modelo anterior lo podemos escribir como

$$y_{ij} \left| \mu_{ij}, \sigma^2 \sim N\left(y_{ij} \left| \mu_{ij}, \sigma^2\right.\right)\right.. \tag{5.5}$$

La primera conclusión que se extrae del modelo (5.5) es que los parámetros v_{ij} y γ_{ij} *no son identificables* —solamente lo son los μ_{ij}—, de modo que no se pueden estimar por separado, a no ser que se tenga información a priori sobre ellos.

Como el número de parámetros de (5.5) es muy elevado, $m\times k$ —casi tanto como el de observaciones—, establecemos las siguientes hipótesis simplificadoras sobre los μ_{ij} que dan lugar a tres modelos posibles.

i) $\mu_{ij} = \mu$, es decir, el cociente entre el factor de inflación y la velocidad es constante y no depende ni del origen ni del destino.

ii) $\mu_{ij} = \mu_j$, es decir, el cociente entre el factor de inflación y la velocidad solamente depende del destino j-ésimo.

iii) $\mu_{ij} = \alpha_i + \beta_j$, es decir, la contribución del origen i-ésimo y del destino j-ésimo al cociente entre el factor de inflación y la velocidad se supone separable y aditiva.

Los tres modelos para los datos y_{ij}, de menos a más complejo, se corresponden con un modelo normal ordinario, un ADEVA de una vía y un ADEVA de dos vías, respectivamente.

Los datos se analizaron para los tres modelos y tras eliminar las siete observaciones anómalas por pueblos, básicamente coindidentes para los tres modelos, y comprobar la homogeneidad del factor de varianza σ^2 para los destinos, se aplicó la técnica bayesiana descrita en Girón, Moreno y Martínez (2005) para seleccionar el mejor de los tres modelos, que resultó ser el segundo, es decir

$$y_{ij} \left| \mu_j, \sigma^2 \sim N\left(y_{ij} \left| \mu_j, \sigma^2\right.\right)\right. \text{ para } i = 1, \ldots, m'; \; j = 1, k, \tag{5.6}$$

donde m' es el número de pueblos una vez eliminados los anómalos que resultaron ser, por orden lexicográfico: Argamasilla de Alba, Castellar, Membrilla, Ossa, Ruidera, Solana y Torrenueva.

La selección del segundo modelo como el que mejor describe los datos es muy ilustrativa ya que básicamente implica que *la velocidad de las cabalgaduras y/o el factor de inflación solamente dependen del destino y no del pueblo de origen*; además, al rechazarse también el primer modelo, *la velocidad no se puede suponer constante para todos los destinos*, como reflejan las estimaciones.

El análisis bayesiano del problema arroja los siguientes resultados sobre las cotas inferiores de las velocidades a los cuatro destinos, representada por $\underline{v}_j = 1/\mu_j$, para $j = 1, 2, 3, 4$, que se obtienen haciendo $\gamma_j = 1$ para todo j.

Estas estimaciones, basadas en la moda $\tilde{\underline{v}}_j$, y en la media armónica $\hat{\underline{v}}_j$ de sus correspondientes densidades a posteriori, expresadas en kilómetros por jornada, son, respectivamente:

$$\tilde{\underline{v}}_1 = 36.83, \; \tilde{\underline{v}}_2 = 28.33, \; \tilde{\underline{v}}_3 = 35.83, \; \tilde{\underline{v}}_4 = 31.01.$$
$$\hat{\underline{v}}_1 = 36.97, \; \hat{\underline{v}}_2 = 28.40, \; \hat{\underline{v}}_3 = 35.96, \; \hat{\underline{v}}_4 = 31.09.$$

Aunque no debe sorprendernos, llama particularmente la atención el hecho de que la cota inferior de la velocidad a Sierra Morena sea menor que al resto de los destinos, probablemente debido a la mayor dificultad del camino. Además, estas estimaciones son consistentes con la velocidad que Cervantes sugiere implícitamente en la primera parte del Quijote.

La información sobre el parámetro θ vendría dada por su distribución a posteriori que, por cierto, no es una distribución estándar a la que hemos denominado *t*-multivariante invertida. La moda $\tilde{\theta}$ y la media armónica $\hat{\theta}_h$ de esta distribución son, respectivamente

$$\tilde{\theta} = (78.18, 54.68, 84.75, 52.01) \quad y$$
$$\hat{\theta}_h = (83.19, 56.80, 89.90, 54.40).$$

Estas estimaciones, que representarían la distancia desde el *punto o lugar ideal* a los cuatro destinos, sirven para localizar geométricamente el lugar de la Mancha, tal como se ilustra en la Figura 8.

Finalmente, a partir de esta distribución también se pueden calcular las probabilidades a posteriori de cada uno de los m pueblos candidatos a ser el lugar de la Mancha, que se presentan en la Tabla 3, en la que observamos que Villanueva de los Infantes es el pueblo más probable seguido, muy de cerca, por Fuenllana.

Como comentábamos al principio de esta sección, la otra alternativa para determinar el lugar de la Mancha era utilizar el enfoque de coordenadas y calcular la densidad a posteriori de éstas. La Figura 9 representa las curvas de nivel de esta densidad a posteriori

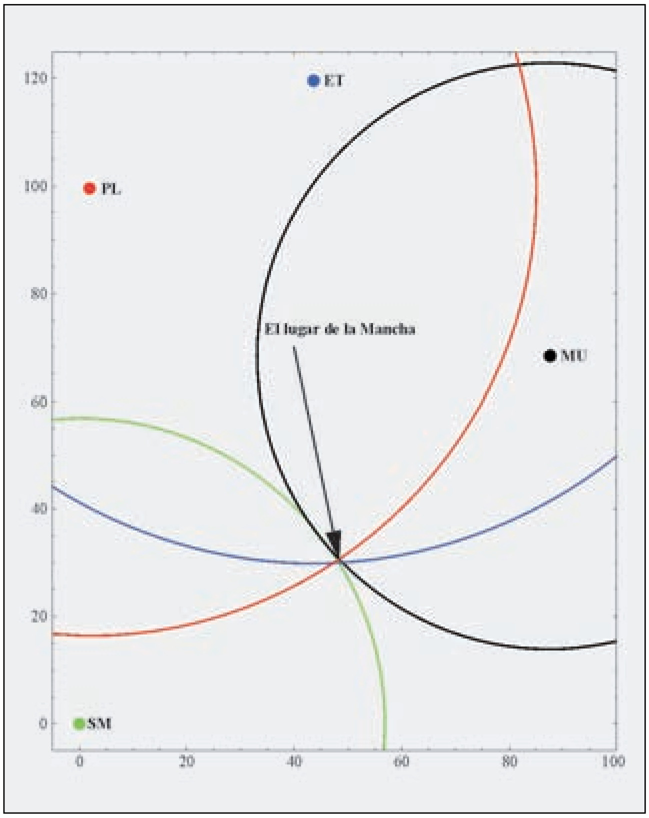


Figura 8. Representación geométrica de la estimación del ***lugar de la Mancha*** usando la media armónica como estimador de los parámetros θ_j, que son los radios de sus respectivos círculos. Las coordenadas del lugar, de acuerdo con el convenio de tomar como origen de coordenadas ***Sierra Morena*** y las distancias expresadas en kms, son aproximadamente (48.6,30.4).

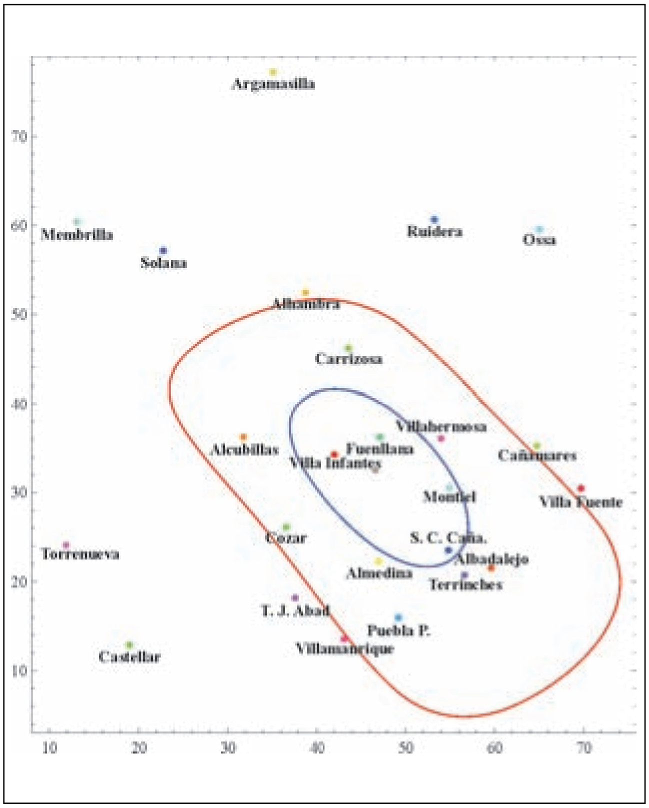


Figura 9. Representación geométrica de las curvas de nivel de la densidad a posteriori de las coordenadas del *lugar de la Mancha*. El lugar más probable es la moda de la densidad a posteriori (representada por un punto gris). Las curvas de nivel se corresponden con los contornos de las regiones de máxima densidad a posteriori de contenido probabilístico 0.50 (en azul) y 0.95 (en rojo).

—regiones de máxima densidad a posteriori de contenido probabilístico 0.50 (en azul) y 0.95 (en rojo)— que permiten identificar *el lugar más probable*, representado por un punto gris, así como determinar las probablidades a posteriori de los veinticuatro pueblos, a partir de esta función de densidad. Como puede deducirse de la Figura 9, los pueblos más cercanos a la moda —el hipotético lugar de la Mancha más probable— son Villanueva de los Infantes, Fuenllana, Montiel y Santa Cruz de Cañameros. Estos cuatro pueblos suman más del 57 % de la probabilidad de que alguno de ellos sea el lugar de la Mancha. Nótese que están contenidos en la región de máxima densidad a posteriori de contenido probabilístico 0.50, cuyo contorno está señalado en azul. Tambien cabe destacar que los siete pueblos clasificados como *anómalos* por el modelo estadístico están muy alejados de la región de máxima probabilidad a posteriori de contenido probabilístico 0.95, cuyo contorno está señalado en rojo.

5. CONCLUSIONES Y POSIBLES EXTENSIONES DEL MODELO ESTADÍSTICO

Aunque el análisis bayesiano de datos positivos con modelos normales, como hemos hecho con los datos del Quijote, tiene la ventaja de permitir obtener resultados analíticos, un análisis más ajustado a la realidad emplearía modelos estadísticos para los datos con soporte positivo como, p. ej., los modelos gamma o gamma-invertido. Hemos realizado un análisis análogo al aquí presentado con estos modelos, que no incluímos por razones de espacio y por su mayor complejidad —ya que este análisis requiere técnicas bayesianas basadas en métodos MCMC—, que conduce a resultados similares a los obtenidos con los modelos normales.

Otro aspecto de gran interés que ofrecen estos datos es el elevado número de observaciones anómalas —siete pueblos de un total de veinticuatro que están

Pueblos	Probabilidades	Pueblos	Probabilidades
Villa Infantes	0.17444	Villamanrique	0.00432
Fuenllana	0.17022	Alhambra	0.00336
Montiel	0.11654	Villa Fuente	0.00317
S. C. Caña	0.11314	Cañamares	0.00315
Almedina	0.08237	T. J. Abad	0.00148
Terrinches	0.07866	Solana	0.00000
Villahermosa	0.07073	Ruidera	0.00000
Albaladejo	0.06446	Membrilla	0.00000
Alcubillas	0.03383	Torrenueva	0.00000
Carrizosa	0.03265	Argamasilla	0.00000
Puebla P.	0.03223	Ossa	0.00000
Cozar	0.01525	Castellar	0.00000

Tabla 3. Probabilidades a posteriori de cada uno de los 24 pueblos candidatos a ser el lugar de la Mancha, ordenadas en orden decreciente. Las probabilidades menores que 0.00001 se han redondeado a 0.

concentrados en una área geográfica relativamente pequeña— que se debe fundamentalmente a que *los datos* no se han obtenido a partir de un diseño estadístico previamente planificado. Obviamente, como revela la Tabla 3, estos datos se identifican con los pueblos más improbables de ser el lugar de la Mancha.

Por último, la inclusión de incertidumbre en la duración de las jornadas d y en las distancias *reales* entre los pueblos de origen y los destinos (mayores, por definición, que las x_{ij}), así como la incorporación de información a priori sobre los factores de inflación y las velocidades se deja, como en el Quijote, para una *segunda parte*…, aunque esperamos no tardar diez años en escribirla, como le sucedió a Cervantes.

REFERENCIAS

1. Girón, F. J., Moreno, E. and Martínez, M. L. (2005). *An objective Bayesian procedure for variable selection in regression*. In *Advances on distribution theory, order statistics and inference*. pp. 393–408. Eds. N. Balakrishnan et al., Birkhauser Boston.
2. Parra, F. et al. (2005). *El lugar de la Mancha es…: El Quijote como un sistema de distancias/tiempos*. Editorial Complutense: Madrid.
3. Ríos, M. J., Montero, F. J. y Parra, F. (2005). *Encontrando el «lugar de la Mancha» con las Matemáticas*. XII, JAEM: Albacete.
4. Girón, F. J. y Ríos, M. J. (2006). *La determinación del "lugar de la Mancha" como problema estadístico*. Boletín de la Sociedad de Estadística e Investigación Operativa. Vol. 22, nº. 1, pp. 23-29.