

# Verificación de taxímetros empleando datos imprecisos

José Ramón Villar, Adolfo Otero, José Otero, Luciano Sánchez,

**Resumen**—El único procedimiento fiable para la verificación de un taxímetro es la simulación del recorrido de un trayecto. Hasta hace poco era posible utilizar una bancada de rodillos con este fin, pero los recientes avances en ayudas a la conducción como ABS, ESP, TCS, etc. dificultan el proceso de verificación, ya que las ruedas tractoras giran mientras que las del otro eje no. Otra desventaja de este sistema es que la medida es afectada por la presión de inflado de las ruedas. Con el fin de solucionar los problemas mencionados se propone utilizar un GPS para realizar la verificación. Dada la inherente característica imprecisa de los datos generados por un GPS, esta verificación es impugnabile. Es preciso obtener una cota superior de la longitud de la trayectoria realizada por el taxi de forma que sea posible determinar con seguridad si un taxímetro es válido o no, y todo ello teniendo en cuenta la naturaleza imprecisa de los datos de partida. Mostraremos como, usando un algoritmo genético, restricciones basadas en la dinámica del vehículo y técnicas fuzzy proporcionamos una cota del error cometido por el taxímetro que permite rechazar de forma incontestable un taxímetro cuando la medición que proporciona es incorrecta.

**Palabras clave**—GPS, metrología, algoritmo genético, problema multicriterio, datos imprecisos.

## I. INTRODUCCIÓN

El servicio proporcionado por los Taxis tiene carácter concesional, con tarifas reguladas por algún organismo oficial. Anualmente las tarifas se actualizan, implicando ajustes en los taxímetros y la consiguiente verificación de los mismos en una ITV. La tarificación depende de la longitud del trayecto y de la velocidad con que se recorra. Si se desea simular el recorrido en un trayecto calibrado, un técnico de la ITV tendría que verificar constantemente la velocidad del vehículo para utilizar en cada momento la tarifa correspondiente. La forma de obtener la velocidad, si no se dispone de otro sistema, es realizando dicha lectura mediante el instrumental disponible en el vehículo, velocímetro y cuentakilómetros u ordenador de abordo. Esto no es adecuado, requiriéndose algún dispositivo que permita medir la velocidad y la distancia simultáneamente.

El grupo de Metrología y Modelos de la Universidad de Oviedo viene encargándose desde 1990 del diseño de los dispositivos que realizan esta tarea en Asturias. Hasta el momento, en la verificación de los taxímetros se utilizaba una bancada estática que dispone de unos rodillos sobre los que se sitúan las ruedas tractoras, con los que se mide satisfactoriamente la velocidad y la distancia, evidentemente bajo una cierta imprecisión derivada de la presión con la que puedan estar inflados los neumáticos, que puede influir en el radio de las ruedas y por lo tanto en la medida. Además, la sucesiva incorporación a los turismos de diversas ayudas a la conducción dificultan el proceso de verificación, pudiendo llegar a averiar la cen-

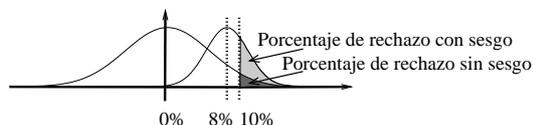


Fig. 1. Distribución del error de los taxímetros: a la izquierda la esperada, libre de sesgo; desplazada hacia la tolerancia máxima, la real.

tralita del vehículo debido a la diferencia de velocidad entre las ruedas tractoras y las del otro eje. En estas circunstancias, los sistemas de ayuda a la conducción constantemente intentan compensar esta diferencia, lo cual es imposible. En este trabajo se propone la utilización de un receptor GPS y una PDA con el fin de recoger los datos de velocidad y distancia durante la realización de un trayecto. De esta forma la verificación puede ser realizada sin la supervisión de un inspector de la ITV, abaratando los costes derivados de la mano de obra. Además es compatible con los sistemas de ayuda a la conducción, robusto frente a la variación de la presión de inflado de las ruedas y más barato que la bancada de rodillos. Existe un claro inconveniente relacionado con la imprecisión intrínseca de los datos obtenidos mediante GPS. El grado de ésta no es constante, variando en cada medida, por lo que es necesario proporcionar una cota superior del error cometido en la medida de la distancia en caso de que la verificación resulte negativa. En la sección II se explica la naturaleza de las medidas realizadas mediante GPS y se detalla como varía la precisión de estas.

El contexto legal de la verificación de los taxímetros tiene implicaciones en la precisión con la que es necesario medir distancia y velocidad: el margen legal del error en distancia es del 10%. Si un taxi presenta un error del 8% sólo se puede rechazar con seguridad si el proceso de verificación se produce con una precisión en la medida de la distancia, como mucho, del 2%. Durante nuestras investigaciones hemos encontrado que la distribución del error cometido por los taxímetros está sesgada hacia el límite legal. Por lo tanto, pequeños cambios en la tolerancia producen un gran aumento de la tasa de rechazo. En la figura 1 se ilustra esta situación, el dato de la media del error real es representativo de la realidad. Por este motivo es fundamental calcular la mencionada cota de la forma más precisa posible y además de modo que se beneficie al propietario del taxi. En este trabajo proponemos calcular la Menor Cota Superior (MCS) de todas las trayectorias compatibles con las medidas proporcionadas por el GPS.

Partiendo de la naturaleza imprecisa de los datos de partida, en este trabajo presentamos una nueva forma de calcular la MCS de una trayectoria. Mediante un algorit-

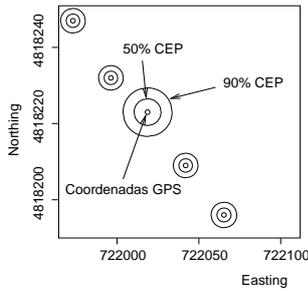


Fig. 2. Ejemplo real de medidas GPS, obsérvese como el CEP puede variar de una medida a otra.

mo genético multiobjetivo se filtran los datos de entrada de modo que se obtenga el subconjunto de los mismos más pequeño que defina una multipoligonal que cubra los datos de entrada. La MCS se calcula a partir de esa multipoligonal mediante un heurístico determinista.

La estructura del trabajo es como sigue, en primer lugar se explica en la sección II como se obtienen las medidas GPS y la interpretación borrosa de las mismas. En la sección III se detalla el filtrado de los datos y el algoritmo determinista para el cálculo de la MCS. Posteriormente, en la sección IV se muestran los experimentos realizados y los resultados obtenidos.

## II. IMPRECISIÓN EN EL SISTEMA GPS

El término Global Positioning System (GPS) se refiere a un sistema formado por un conjunto de dispositivos (satélites y receptor) cuya misión es determinar la posición del receptor. El receptor recibe de los satélites ciertas señales que le permiten calcular su longitud, latitud, altitud, número de satélites usados en la medida, fecha, etc. Cada señal que llega al receptor contiene información acerca del momento en el que partió del satélite al receptor.

Dado que se producen perturbaciones debidas a la atmósfera, desviaciones de los satélites, errores en el reloj de los satélites, errores del receptor y rebotes, es necesario disponer de mayor número de satélites que los cuatro con los que en teoría bastarían para obtener una medida correcta.

Cuanto mayor sea el número de satélites, mejor será la precisión. Sin embargo, aún con un gran número de satélites disponibles, la configuración de la constelación de satélites es relevante a la hora de estimar la precisión de la medida. La influencia de este factor se mide con la magnitud denominada Dilution of Precision (DOP).

Los GPS comerciales proporcionan una indicación de su precisión utilizando una magnitud denominada Circular Error Probable (CEP), que informa de la probabilidad de que el receptor se encuentre en un entorno de cierto radio alrededor de la medida con determinada probabilidad. Así son comunes valores de CEP de 3 metros con un 50 % de probabilidad o de 7 metros con un 90 % de probabilidad.

En la figura 2 se muestra un ejemplo real de como el CEP puede variar de una medida a otra.

### A. Estimación del CEP a partir del número de satélites.

Tal y como hemos dicho, en la precisión de las medidas realizadas con un GPS no sólo influye el número de satélites disponible sino que la geometría de la constelación de satélites también tiene impacto en dicha precisión. De forma empírica se pueden estimar los parámetros de una fórmula que relaciona el número de satélites con el DOP para un modelo de receptor y antena determinado. Utilizaremos esta expresión en el proceso de simulación de la toma de medidas que se explica en la sección IV.

En el contexto de este artículo, la componente horizontal (HDOP) del DOP es la más importante, ya que está estrechamente relacionada con el CEP. A continuación se detalla como se relacionan estas magnitudes.

El RMS del error cometido en una serie de medidas GPS bajo un HDOP común se puede aproximar mediante la ecuación 1, donde  $A$  y  $B$  se pueden obtener de forma empírica para un receptor y antena específicos [30].

$$RMS\_Error(HDOP) = \sqrt{(A \cdot HDOP)^2 + B^2} \quad (1)$$

La componente HDOP se relaciona con el número de satélites disponible mediante la ecuación 2, donde  $C$  y  $D$  también son específicas.

$$RMS\_HDOP = \frac{C}{(\text{número de satélites})^2} + D \quad (2)$$

La probabilidad de que el error cometido en una medida sea menor que un CEP dado, condicionada a estar adquirida bajo un HDOP determinado sigue la distribución de Rayleigh, tal y como se muestra en la ecuación 3.

$$P(Err \leq CEP | HDOP) = 1 - e^{-\left(\frac{CEP}{RMS\_Error(HDOP)}\right)^2} \quad (3)$$

De las ecuaciones 1 y 3 y despejando CEP, se puede obtener este en función de HDOP y la probabilidad deseada. De forma análoga se puede obtener el CEP en función del número de satélites, tal y como se observa en la ecuación 4.

$$CEP = \left(-((A \cdot HDOP)^2 + B^2) \cdot \ln(1 - P(Err \leq CEP | HDOP))\right)^{0,5} \quad (4)$$

### B. Representación borrosa de datos GPS

En el contexto de las probabilidades imprecisas, tiene sentido entender un conjunto borroso como un conjunto de tolerancias, asignado a cada una de ellas un grado de confianza, de modo que la de menor confianza sea la mas pequeña [14]. En concreto, se dice que, dado un conjunto incompleto de intervalos de confianza para una variable aleatoria, podemos construir una variable aleatoria borrosa cuyos  $\alpha$ -cortes son intervalos de confianza de grado  $1 - \alpha$  [7], que contienen toda la información que conocemos sobre la variable aleatoria desconocida.

En el caso de las medidas obtenidas a través de un GPS el fabricante proporciona dos intervalos de confianza: al 50 % y al 90 %. Mediante la ecuación 4, para cada valor de probabilidad dado se puede obtener un valor CEP.

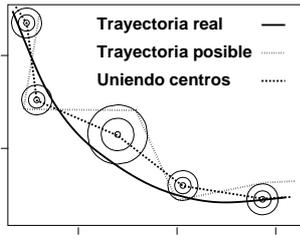


Fig. 3. Ejemplo sintético en el que se aprecia la diferencia entre la trayectoria real, la trayectoria obtenida uniendo centros y una trayectoria compatible con las medidas GPS.

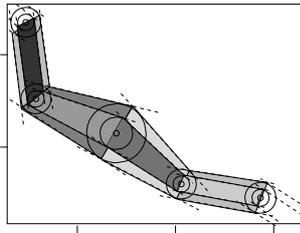


Fig. 4. Ejemplo sintético en el que se muestra como construir una poligonal que encierre a los puntos borrosos que representan las medidas GPS.

Además, se puede comprobar que cuanto mayor sea el valor de la probabilidad mayor será el valor del CEP. Esto tiene un significado físico claro en el caso de las medidas GPS: cuanto mayor confianza se desee tener en la inclusión en el CEP de la posición en la que realmente se encuentra el receptor, la magnitud del CEP aumenta.

Por lo tanto, se puede comprobar que las medidas realizadas a través de un GPS cumplen la condición dada en [14], [7], y podemos construir una variable aleatoria borrosa cuyos  $\alpha$ -cortes son círculos de confianza de grado  $1 - \alpha$  obtenidos a través de la ecuación 4, que contienen toda la información que conocemos sobre la variable aleatoria desconocida. En este trabajo se considerarán los datos como conjuntos borrosos obtenidos de la manera indicada. Para cada  $\alpha$ -corte se obtiene el radio CEP, que representa el intervalo de confianza para las coordenadas reales del taxi en ese momento.

### III. CÁLCULO DE LA MCS UTILIZANDO DATOS BORROSOS

Los datos GPS se toman a intervalos regulares. Cada muestra es un conjunto borroso, como se ha mencionado, cuyos  $\alpha$ -cortes son círculos. Por lo tanto cada círculo es un intervalo de confianza para las coordenadas del taxi en ese momento. Tomar los centros de esos círculos como coordenadas del taxi no es válido. En la figura 3 se muestra un ejemplo sintético en el que se pone de relieve esta cuestión.

Como se puede apreciar, el punto de la trayectoria (en trazo continuo) en donde se ha efectuado la medida puede estar contenido o no en el CEP correspondiente. En trazo discontinuo se representa la estimación de la trayectoria realizada utilizando las coordenadas GPS medidas. A puntos se representa una trayectoria compatible con las medidas GPS tal y como hemos explicado. Se puede comprobar que las longitudes de las trayecto-

rias son diferentes. Por este motivo necesitamos calcular la MCS cuyos extremos están contenidos en los círculos y esta longitud será siempre mayor que el valor obtenido uniendo los centros. Una aproximación a esa cota se puede conseguir construyendo una poligonal a partir de las tangentes a los CEP correspondientes a cada  $\alpha$ -corte. De esta forma la MCS en cada punto borroso pasaría por uno de los cortes de las tangentes a él. En la figura 4 se muestra un ejemplo de las poligonales construidas de la forma indicada, para dos  $\alpha$ -cortes.

Este cálculo no es sencillo. Si intentamos calcular la longitud máxima de todos los trayectos lineales a trozos cuyos extremos están en los círculos, es obvio que a menor tiempo de muestreo, más grande será la estimación de la MCS, coloquialmente podríamos decir que es posible zigzaguear más. Esto es inadecuado, la estimación debería de ser independiente del período de muestreo [21]. Hemos decidido preprocesar los datos borrosos y eliminar toda la información redundante utilizando un algoritmo genético, como se mostrará en la siguiente sección.

Si se utilizasen datos nítidos, la simplificación de una multipoligonal se ha estudiado en [10], [16]. La alternativa más parecida a la que se presenta en este trabajo, a la vista de la bibliografía consultada, usa datos borrosos de una base de datos geográficos [2] para reconstruir imágenes tridimensionales mediante B-splines borrosas [4]. El objetivo de este trabajo es obtener la MCS para un conjunto de datos borrosos obtenidos a través de un GPS, de forma que se pueda establecer una cota superior que no pueda ser cuestionada. La idea es trabajar con los datos borrosos sin defuzzificar, y generar una MCS que será, a su vez, borrosa. No se hará uso de ordenaciones totales de números borrosos, sino de ordenaciones parciales para establecer dominancia de uno frente a otro.

El proceso de cómputo de la MCS está dividido en tres fases: filtrado por colinealidad, filtrado de datos espúreos, y cálculo determinista de la MCS; todo ello para cada  $\alpha$ -corte. Dado un valor de  $\alpha$ , la representación de cada dato es un círculo de centro el de las coordenadas y radio el CEP. Por lo tanto, un conjunto de datos de entrada puede representarse como un conjunto de círculos para cada  $\alpha$ -corte. Si se trazan las dos tangentes exteriores comunes a dos círculos consecutivos, y se determinan los puntos de cruce para cada par de tangentes consecutivas, con las tangentes y dichos puntos de cruce se obtiene una superficie poligonal. La mayor trayectoria incluida en dicha poligonal representa la MCS. El resultado de todo el proceso es una MCS para cada valor de  $\alpha$ . En el proceso de filtrado no se pueden eliminar ni el primer ni el último puntos, dado que éstos marcan los límites del trayecto.

El filtrado por colinealidad es sencillo, puesto que para cada  $\alpha$ -corte, la representación de cada dato es un círculo de centro el de las coordenadas y radio el CEP. Trazando las dos tangentes exteriores comunes a dos puntos consecutivos, podremos determinar que tres puntos consecutivos son colineales si sus pares de tangentes exteriores comunes son paralelas.

El filtrado de datos espúreos es algo más complejo,

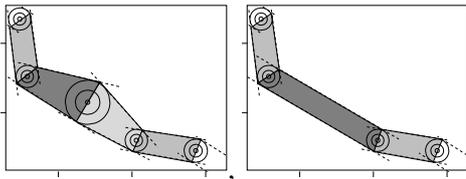


Fig. 5. Ejemplo sintético en el que se muestra el proceso de simplificación de la poligonal. La eliminación de un punto reduce la superficie, pero este puede ser representativo de la trayectoria seguida.

resultando un problema bicriterio como se verá a continuación. Se denomina en este trabajo dato espúreo aquél que de no filtrarlo induciría longitudes de trayectoria incorrectas, es decir, representa un punto cuyo error es cercano al máximo. Si se filtra un punto de máximo error se reduce el área de la poligonal definida por los restantes puntos para un determinado  $\alpha$ -corte. Es decir, que se pueden eliminar estos puntos buscando reducir el área definida por la poligonal de los datos para un  $\alpha$ -corte. Ahora bien, si se repite este paso un número de veces lo suficientemente grande, se podrá comprobar que el resultado del filtrado será un conjunto de muy pocos puntos, dado que eliminar un punto con casi la total seguridad genera una poligonal de menor área. Por lo tanto, es necesario disponer de otro criterio.

El segundo criterio viene dado porque la poligonal resultante debe contener la trayectoria, debe ser representativa de la trayectoria realizada. En otras palabras, el porcentaje de puntos de los datos iniciales contenidos en la poligonal debe ser máximo. De esta forma, la poligonal generada tras el filtrado continúa siendo representativa de la trayectoria cuya MCS es buscada. En la figura 5 se ilustra este proceso para uno de los dos  $\alpha$ -cortes. Como se aprecia, al eliminar uno de los puntos, el área de la poligonal ha decrecido. Como consecuencia de esto la MCS será menor, se ha reducido el espacio bidimensional en donde por donde se considera que ha podido pasar el vehículo. Sin embargo el punto podría ser representativo, por lo que es evidente que los objetivos son contrapuestos.

De todo esto se concluye que se buscará el mejor filtrado en un problema multicriterio. En la solución de este problema se usará un algoritmo genético multicriterio.

Asímismo, en el filtrado se considerará el efecto de la dinámica del vehículo en durante la conducción, como se verá más adelante. Finalmente, una vez realizado el filtrado, se aplicará un algoritmo determinista para el cálculo de la longitud de la mayor trayectoria contenida en una poligonal definida como se ha dicho en las fases anteriores.

#### A. Datos de partida

Los datos de entrada al algoritmo serán directamente las medidas en el plano bidimensional obtenidas por el GPS. Se trabaja en el plano bidimensional puesto que la pérdida de información de esta representación no es significativa: trayectorias de entre 3 y 5 kilómetros frente a diferencias de altitud de pocos metros. Se toma una medida de GPS cada segundo, y se almacena en una PDA.

De esta manera, se envía el taxi con el taxímetro a contrastar, junto con el GPS y la PDA, a realizar un trayecto. Una vez de vuelta el taxi, las medidas son volcadas a un computador, donde se realizarán los cálculos. Cada medida incluye las coordenadas del GPS (centro de la circunferencia), y el radio CEP calculado en función del número de satélites, etc., para cada  $\alpha$ -corte. Por lo tanto, la entrada al modelo es una lista de puntos borrosos según se explicó en el apartado II-B.

#### B. Evaluación de la dominancia de Pareto

Un aporte interesante en este trabajo es que, al igual que los datos de entrada son borrosos, las funciones de evaluación de los diferentes criterios también lo son. El área de la superficie poligonal y el porcentaje de datos contenidos en la poligonal varían con el valor de  $\alpha$  utilizado. Por lo tanto, cada individuo está valorado con dos números borrosos, y el problema es buscar un individuo que minimice el primero mientras maximice el segundo. Esto implica ordenar números borrosos.

La ordenación de números borrosos puede ser absoluta, y hay muchos trabajos en la literatura que permiten obtener ordenes totales [1], [22], [27], [24], [26]. Desafortunadamente, todos los ordenes parciales tienen el inconveniente de necesitar un criterio extra, que será dependiente del problema; el uso general implica inconsistencias en la ordenación [31], [29]. De igual manera, el coste computacional de estas ordenaciones es importante [5], [31].

Sin embargo, no todos los algoritmos de optimización requieren de relaciones de orden total, sino que muchas veces es suficiente una relación de orden parcial. En concreto, las decisiones de dominancia en problemas multicriterio según el criterio de optimalidad de Pareto requiere únicamente de relaciones de orden parcial que no presenten inconsistencias [12], [6].

Para establecer un relación de orden parcial para números borrosos es posible el uso de una representación intervalar del número borroso para cada  $\alpha$ -corte. Esta idea ha sido aplicada con éxito en [17], [29] introduciendo nuevas restricciones al problema. En [13] se propone cómo establecer ordenes parciales restringido a funciones trapezoidales. En nuestro caso, no es una aplicación válida, dado que se desconocen las funciones de pertenencia.

Se han presentado diversos trabajos para evaluar la dominancia de Pareto con funciones fitness borrosas. En [32] se utilizan reglas borrosas para inferir si un individuo domina a otro. Trabajos similares se disponen en [28], [18].

Otras propuestas extienden el concepto de dominancia de Pareto para funciones fitness borrosas. Un ejemplo claro es el presentado en [33], conocido como  $\alpha$ -dominancia, donde para cada valor de  $\alpha$  se evalúa la dominancia. En [19] se propone determinar la dominancia mediante la agregación de las reglas borrosas *grado de dominancia de x sobre z* y *grado de dominado de x por z*. Hapke y otros proponen el uso de una medida de orden parcial entre criterios borrosos para establecer el grado

de dominancia de un individuo sobre otro [15]. En [11] se propone una generalización del concepto de dominancia de Pareto sobre fitness borrosos, generando una superficie de decisión como resultado de la evaluación de la dominancia. Finalmente, en [20] se considera la dominancia una variable borrosa, evaluada para cada par de individuos. Cuando el grado de dominancia  $\mu(a, b) = 0$  se dice que  $a$  domina a  $b$ , mientras que si  $\mu(a, b) = 1$  se dice que  $b$  domina a  $a$  (para minimización). En este trabajo se utiliza el concepto de  $\alpha$ -dominancia, evaluando cada individuo para cada valor de  $\alpha$ . Además, para establecer el orden parcial entre dos individuos, necesario para establecer la relación de dominancia, se utiliza el criterio de intersección nula para especificar un individuo dominante tal y como se expone en [23].

### C. Algoritmo genético multicriterio

Como algoritmo de búsqueda se plantea el uso de un algoritmo genético, dado que es conocido el buen comportamiento de este tipo de algoritmos en problemas con bajo número de generaciones e individuos [8]. En concreto, el algoritmo genético multicriterio utilizado es NSGA-II [8], [9].

#### C.1 Codificación de cada individuo:

En este trabajo, un individuo es representado como un vector booleano de puntos borrosos incluidos -no filtrados- en el individuo. Cuando un valor del vector es *true* el punto correspondiente está incluido por el individuo para determinar la poligonal. Cuando un valor del vector es *false* el punto correspondiente ha sido filtrado. Para generar un individuo en la población inicial, cada punto no extremo se incluye en aquél con probabilidad  $p$ . Los puntos extremos del camino siempre estarán incluidos en los individuos.

#### C.2 Fitness borroso multiobjetivo:

Cada candidato a solución se evalúa de la forma siguiente: en primer lugar se construye la cadena de poligonales para cada  $\alpha$ -corte de los datos seleccionados, usando las superficies tangentes a los puntos borrosos seleccionados. El objetivo es que la cadena contenga el máximo número de puntos y que ocupe el menor área posible. Por lo tanto, se calcula el área de la poligonal para dicho  $\alpha$ -corte, así como el porcentaje de puntos cubiertos. Ambos objetivos son números borrosos y definen las funciones fitness del problema multicriterio. En concreto se utilizan los criterios área de la poligonal y  $1 - \tau$ , donde  $\tau$  es el porcentaje de puntos cubiertos por la poligonal, ambos números borrosos. Ambos criterios se minimizarán. Para evaluar la dominancia de Pareto se utiliza el criterio de la  $\alpha$ -dominancia explicado anteriormente.

#### C.3 Operadores Genéticos:

Como operadores genéticos se utilizan el cruce y la mutación. Las definiciones de las operaciones de cruce y mutación deben reducir el número de vértices. Dados dos progenitores  $A$  y  $B$ , los descendientes son dos cadenas  $C$  y  $D$  tales que  $A \cap B \subseteq C$  y  $A \cap B \subseteq D$ ; un vértice

$v \in A - B$  tiene una probabilidad  $p^+$  de estar en  $C$ , un vértice  $v \in B - A$  tiene una probabilidad  $p^-$  de estar en  $C$ , donde  $p^-$  es mucho más pequeña que  $p^+$ . El conjunto  $D$  se construye de la misma forma. La mutación se define como la eliminación aleatoria de un punto no extremo de la cadena. Todo individuo incluirá el primer y último puntos, de forma que siempre incluirá el inicio y fin de trayecto.

### D. Evaluación de la dinámica del vehículo

Considérese un vehículo que circula por una vía a una determinada velocidad. Existe un ángulo máximo de giro pasado el cuál el coche vuelque. De igual manera, cuando se conduce un vehículo existe un ángulo de giro a partir del cuál la conducción no es cómoda para los pasajeros.

Si se tiene en cuenta esta reflexión se puede introducir un nuevo heurístico en el proceso filtrado a través del algoritmo genético multicriterio. El máximo ángulo de giro admitido depende de la velocidad del vehículo, a mayor velocidad, menor ángulo de giro máximo. Cuando se tenga en cuenta la dinámica del vehículo en el filtrado, cada vez que se generen nuevos individuos tras una operación genética se analizará el ángulo de giro en cada punto, de forma que si es mayor a un valor preestablecido se considerará que el individuo no es válido, del mismo modo que cualquier otra restricción de las empleadas en NSGA-II [9].

Para determinar este ángulo en cada punto se utilizan el punto anterior y el siguiente, por supuesto, asumiendo que son datos borrosos. Como se ha trabajado para un determinado  $\alpha$ -corte, se dispone de la poligonal para dicho  $\alpha$ . Entonces, se busca la mayor trayectoria incluida en la poligonal de tres puntos. Esta trayectoria está formada por dos segmentos que se cortan en uno de los vértices de la poligonal para el punto en cuestión. Si ambos segmentos forman un ángulo mayor que el máximo ángulo de giro para la velocidad actual, entonces el individuo tiene puntos no válidos físicamente.

### E. Calculo determinista de la longitud de la trayectoria más larga

Una vez que los datos se preprocesan mediante el algoritmo genético, obteniendo el subconjunto más pequeño de vértices que contiene el recorrido real, necesitamos calcular su MCS.

Para cada  $\alpha$ -corte de la  $B$ -spline borrosa que contiene la trayectoria del taxi, tenemos una cadena poligonal construida a base de trapecoides, como se ve en la figura 6. El sentido de la marcha se indica mediante una flecha. Cada vértice de un trapecoide se identifica mediante una pareja de enteros. Los que están a la izquierda de la flecha, según el sentido de la misma, tienen un cero como primer valor de ese par, los que están a la derecha un uno. El otro número es el índice del tramo. En cada tramo, la MCS pasa por  $(0, i)$  o por  $(1, i)$ . El conjunto de vértices que forman parte de la MCS puede ser determinado por exploración exhaustiva, pero este cálculo es muy costoso e impracticable para cualquier trayectoria real del orden de 500 puntos. Este problema ha sido estudiado previa-

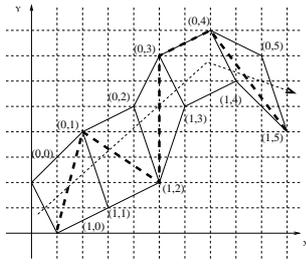


Fig. 6. Ejemplo del cálculo de la MCS (línea gruesa a trazos) mediante el heurístico determinista.

mente, relacionado con el conocido como “Longest Path With Forbidden Pairs” [3] el cual es NPO-PB completo.

Debido a estas cuestiones y a que, en la práctica, los cambios en dirección y en velocidad (que implica una cierta distancia entre vértices) están limitados por la dinámica del taxi, la geometría del recorrido y el comportamiento del GPS, usamos un heurístico que es lineal con el número de vértices. El heurístico se basa en la selección de los vértices convexos: cuando un vehículo gira, el trayecto más largo que puede recorrer pasa por el exterior de las curvas que tome. La convexidad de los vértices se analiza en base a las rectas que pasan por los vértices anterior y siguiente.

El heurístico es como sigue: el primer segmento de la MCS une los vértices  $(x, 0)$  e  $(y, 1)$  siendo  $x$  e  $y$  los valores escogidos de entre 1, 0 que maximizan la distancia entre esos vértices. Desde el vértice del paso 1 hasta al penúltimo, si sólo hay un vértice convexo, la MCS pasa por ese vértice. Si hay dos vértices convexos, la MCS pasa por el vértice más alejado del anterior. Si no hay vértices convexos, la MCS pasa por el vértice más alejado del anterior.

El último tramo de la MCS termina en el vértice más alejado del vértice seleccionado en la etapa anterior. En la figura 6 se muestra un ejemplo de una MCS calculada mediante este heurístico, indicada por una línea gruesa a trazos.

#### IV. EXPERIMENTOS Y RESULTADOS

Se ha decidido evaluar el algoritmo propuesto en un recorrido realista que cubra las situaciones encontradas habitualmente en la verificación de un taxímetro, con varias curvas, aceleraciones y deceleraciones, cambios en HDOP, CEP y número de satélites disponibles. Las longitudes y latitudes obtenidas se convirtieron en un par Northing/Easting del sistema Universal Transverse Mercator [25]. De este modo se puede utilizar la norma Euclídea para calcular distancias entre mediciones, al proyectarse la superficie terrestre en unos ejes cartesianos.

La trayectoria se ha muestreado cada segundo, en cada punto el número de satélites disponibles es una variable aleatoria uniforme que varía entre 4 y 9. A partir de estos datos, construimos un conjunto de mediciones GPS realizadas cada una de ellas en el mismo instante en el que se ha muestreado la trayectoria. El proceso de simulación se realiza de la siguiente forma: con una probabilidad igual

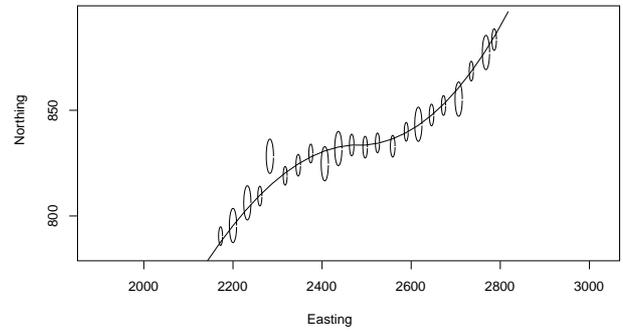


Fig. 7. Parte de los datos usados en los experimentos. Los círculos (elipses por el escalado) representan las mediciones GPS. La línea continua la trayectoria real. La mayor parte de los puntos de la trayectoria están en el CEP al 95 %.

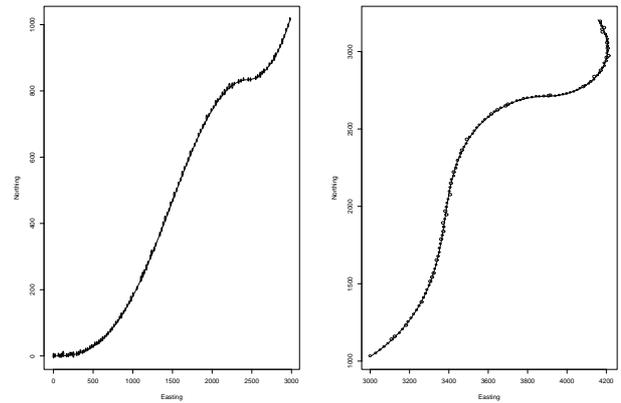


Fig. 8. Trayectorias utilizadas en los experimentos.

a 0.9, se selecciona un punto que dista del correspondiente de la trayectoria menos que el CEP correspondiente al número de satélites adjudicado y a la mencionada probabilidad. Con probabilidad 0.1 el punto se toma más allá del correspondiente CEP. Este procedimiento refleja la incertidumbre que ocurre cuando se usa un GPS y los datos obtenidos sirven para averiguar como de ajustadas son las MCS obtenidas con nuestro algoritmo. En la figura 7 se muestra parte de los datos generados. Las mediciones GPS se representan como círculos (elipses en realidad, debido al escalado) con radio igual al CEP al 90 % y la trayectoria original como una línea continua. Como se puede observar, la mayor parte de los círculos (de hecho el 90 %) cortan a la trayectoria, es decir, el punto correspondiente de la trayectoria está en el círculo de radio CEP centrado en la medida.

Se realizaron dos experimentos con dos trayectorias generadas como se ha comentado y de tamaño 120 cada una, que se pueden ver en la parte izquierda y derecha de la figura 8. La primera mide 3228.57 m., la segunda 2741.31 m. Utilizando las coordenadas de las mediciones GPS para determinar la longitud se obtiene 3238.52 m. para la primera y 2696.49 m. para la segunda. Como se puede ver, es cierto que usando las mediciones de GPS no se obtiene una cota superior, puesto que para el segundo caso la longitud a través de las mediciones GPS es menor que la real.

Parámetro	Valor			
	No. Generaciones	4000	6000	8000
No. Individuos	15 or 25			
No. Poblaciones	1			
Mínimo % de puntos fuzzy cubiertos	0.85			
Prob. $p^+$ en cruce	0.5			
Prob. $p^-$ en cruce	0.1			
Prob. de cruce	0.7			
Prob. de Mutación	0.1			

TABLA I  
PARÁMETROS UTILIZADOS PARA EL ALGORITMO NSGA-II.

Nº de generaciones	Nº de individuos	NSGA-II	
		Mejor	Medio
4000	15	3456.61	3499.48
6000	15	3462.01	3493.53
8000	15	3480.07	3501.01
10000	15	3483.20	3502.04
4000	25	3462.09	3494.41
6000	25	3429.14	3485.15
8000	25	3433.16	3490.54
10000	25	3417.43	3466.98

TABLA II  
RESULTADOS EN LA MCS PARA 10 EJECUCIONES DE NSGA-II SIN ANÁLISIS DE DINÁMICA DE VEHÍCULOS, PRIMERA TRAYECTORIA.

Se decidió realizar dos tandas de experimentos. En la primera no se incorporó la restricción por la dinámica del vehículo y se utilizaron distintas combinaciones de número de individuos y generaciones, repitiendo 10 veces la ejecución del algoritmo con cada combinación. Los parámetros del algoritmo NSGA-II se muestran en la tabla I. Los resultados de los experimentos (para el  $\alpha$ -corte al 90 %) se visualizan en la tabla II para la primera trayectoria y en la tabla III para la segunda. A la vista de los resultados no parece existir una mejora del rendimiento que justifique utilizar más de 4000 generaciones y 15 individuos. Por esa razón, los resultados que se muestran en la tabla IV se corresponden con esa combinación de número de generaciones e individuos (para el  $\alpha$ -corte al 90 %).

Como se puede extraer de estos resultados, el uso de la dinámica de los vehículos reduce enormemente la cota superior que se había obtenido en la primera tanda de ensayos. Los valores resultantes son efectivamente una cota superior y las longitudes obtenidas son similares a las longitudes reales de las trayectorias, siendo además incontestable esta estimación por haber tenido en cuenta la imprecisión de las medidas GPS. Recuérdese además que los puntos muestreados de la trayectoria sintética no se proporcionan al algoritmo sino que se proporcionan medidas GPS simuladas, incluyendo por lo tanto dicha imprecisión.

Nº de generaciones	Nº de individuos	NSGA-II	
		Mejor	Medio
4000	15	3126.99	3192.51
6000	15	3078.44	3216.72
8000	15	3156.23	3243.73
10000	15	3088.41	3211.55
4000	25	3168.87	3238.76
6000	25	3038.61	3210.98
8000	25	3146.52	3249.75
10000	25	3064.06	3181.56

TABLA III  
RESULTADOS EN LA MCS PARA 10 EJECUCIONES DE NSGA-II SIN ANÁLISIS DE DINÁMICA DE VEHÍCULOS, SEGUNDA TRAYECTORIA.

Dataset	Longitud real	Con los GPS fixes
		1
2	2741.31	2696.49

Dataset	NSGA-II	
	Mejor	Medio
1	3242.64	3243.45
2	2765.63	2798.01

TABLA IV  
RESULTADOS PARA 10 EJECUCIONES DE NSGA-II USANDO ANÁLISIS DE LA DINÁMICA DE VEHÍCULOS.

## V. CONCLUSIONES Y TRABAJO FUTURO

Durante el desarrollo de este trabajo hemos visto que existen problemas legales si usamos directamente los datos obtenidos mediante el GPS. Los propietarios de los taxis podrían impugnar fácilmente cualquier verificación que resultara negativa esgrimiendo como argumento la imprecisión de las medidas utilizadas. Sin embargo, debido a los problemas comentados en la sección I, la única alternativa factible para verificar los taxímetros hoy en día es el uso de GPS. Por ello hemos decidido modificar la forma en la que una ITV presenta la información, dando una estimación de la MCS de modo que si el taxímetro está un 10 % por encima de ella, entonces no hay duda de que el taxímetro está mal ajustado. Además esta alternativa es menos restrictiva, lo cual se ajusta mejor a la realidad de la distribución del error cometido por los taxímetros, sesgado hacia la tolerancia máxima.

Con este trabajo se ha demostrado que el uso de las medidas de GPS directamente para establecer una cota superior de la longitud de la trayectoria seguida por un taxi no es válido. Se ha propuesto el uso de un algoritmo multicriterio borroso que pueda gestionar la inherente imprecisión de los datos generados obtenidos a través de dispositivos GPS. De igual manera se ha utilizado una de las técnicas conocidas para realizar la evaluación borrosa de la dominancia de Pareto. Además, se han incluido diversas heurísticas en sucesivas etapas del proceso derivadas de la naturaleza del problema. Los resultados

muestran la importancia de resolver problemas imprecisos manejando los datos como tal, con algoritmos que gestionen apropiadamente dicha imprecisión. En este caso, la imprecisión ha sido manejada como conjuntos borrosos con un resultado muy satisfactorio. El proceso, en su globalidad, genera una MCS apropiada para establecer un límite legal para el rechazo de taxímetros.

El trabajo futuro incluye el uso de distintos algoritmos evolutivos como recocido simulado con operadores genéticos, algoritmo más rápido y produce mejores resultados en problemas multiobjetivo [23]. En el mismo sentido se deben probar distintos conceptos de dominancia borrosa, con el fin de estimar de forma más ajustada la MCS.

#### AGRADECIMIENTOS

Este trabajo ha sido subvencionado por el Ministerio de Educación y Ciencia Español, en el marco del proyecto TIN2005-08386-C05.

#### REFERENCIAS

- [1] Abbasbandy S., Asady B., *Ranking fuzzy numbers by sign distance*, Information Sciences 176, 2405-2416, 2006.
- [2] Anile A.M., Falcidieno B., Gallo G., Spagnuolo M., Spinello S., *Modeling uncertain data with fuzzy B-splines*, Fuzzy Sets and Systems 113, 397-410, 2000.
- [3] Berman, P., and Schnitger, G., *On the complexity of approximating the independent set problem*, Inform. and Comput. 96, 77-94, 1992
- [4] de Boor C., *On Calculating with B-splines*, Journal of Approximation Theory 6, 50-62, 1972.
- [5] Chen L-H., Lu H-W., *The preference order of fuzzy numbers*, Computer and Mathematics with Applications 44, 1455-1465, 2002.
- [6] Coello C.A., *A Comprehensive Survey of Evolutionary-Based Multiobjective Optimization Techniques*, Knowledge and Information Systems. An International Journal 1(3), 269-308, 1999.
- [7] Couso I., Montes S., Gil P., *The necessity of the strong alpha-cuts of a fuzzy set*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9-2, 249-262, 2001
- [8] Deb K., Agrawal S., Pratap A., T. Meyarivan T., *A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II*, En Marc Schoenauer, Kalyanmoy Deb, Günter Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo, and Hans-Paul Schwefel, editors, Proceedings of the Parallel Problem Solving from Nature VI Conference, 849-858, Springer. LNCS, 2000.
- [9] Deb K., Goel T., *Controlled Elitist Non-dominated Sorting Genetic Algorithms for Better Convergence*, In Eckart Zitzler, Kalyanmoy Deb, Lothar Thiele, Carlos A. Coello Coello, and David Corne, editors, First International Conference on Evolutionary Multi-Criterion Optimization, 67-81. Springer-Verlag. LNCS No. 1993, 2001.
- [10] Estkowski R., Mitchell J. S. B., *Simplifying a polygonal subdivision while keeping it simple*, SCG '01: Proceedings of the seventeenth annual symposium on Computational geometry, ISBN 1-58113-357-X, 40-49, ACM Press, New York, NY, USA, 2002
- [11] Farina M., Amato P., *Fuzzy Optimality and Evolutionary Multiobjective Optimization*, Peter J. Fleming, Eckart Zitzler, Kalyanmoy Deb and Lothar Thiele eds. Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003, pp. 58-72, Springer. LNCS. Volume 2632, Faro, Portugal, April 2003
- [12] Carlos M. Fonseca and Peter J. Fleming, *Multiobjective Optimization and Multiple Constraint Handling with Evolutionary Algorithms—Part I: A Unified Formulation*, IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 28 (1), 26-37, 1998.
- [13] Ganesan K., Veeramani P., *Fuzzy linear programming with trapezoid fuzzy numbers*, Annals of Operation Research 143, 305-315, 2006.
- [14] Goodman, Nguyen, *Uncertainty Models for Knowledge-based Systems*, North-Holland. 1985
- [15] Hapke M., Jaskiewicz A., Slowinski R., *Pareto Simulated Annealing for Fuzzy Multi-Objective Combinatorial Optimization*, Journal of Heuristics, 6(3), 329-345, 2000.
- [16] Hershberger J., Snoeyink J., *Speeding up the douglas-peucker line simplification algorithm*, in Proceedings of the 5th International Symposium on Spatial Data Handling, 134-143, Charleston, United States, 1992.
- [17] Jahanshahloo G. R., Soleimani-damaneh M., Nasbaradi E., *Measurement of efficiency in DEA with fuzzy input-output levels: a methodology for assessing, ranking and imposing of weight restrictions*, Applied Mathematics and Computation 156, 175-187, 2004.
- [18] Kiyota T., Tsuji Y., Kondo E., *New Multiobjective Fuzzy Optimization Method and Its Application*, American Control Conference ACS'2000 6, 4224-4228. Chicago, Illinois, USA, 2000.
- [19] Köppen M., Franke K., Nickolay B., *Fuzzy-Pareto Dominance Driven Multiobjective Genetic Algorithm*, In Proceedings of the 10th IFSA World Congress (IFSA 2003), pages 450-453, Istanbul, Turkey, June, 2003.
- [20] Koduru P., Das S., Welch S., Roe J.L., *Fuzzy Dominance Based Multi-objective GA-Simplex Hybrid Algorithms Applied to Gene Network Models*, Genetic and Evolutionary Computation—GECCO 2004. Proceedings of the Genetic and Evolutionary Computation Conference. Part I Springer-Verlag, Lecture Notes in Computer Science Vol. 3102.
- [21] Meratnia N., de By R.A., *Trajectory representation in location-based services: problems and solutions*, in Proceedings of the 3rd IEEE Workshop on Web and Wireless Geographical Systems (W2GIS 2003) in conjunction with the Fourth International Conference on Web Information Systems Engineering (WISE), Rome, Italy, 2003.
- [22] Mitchell H.B., *Ranking type-2 fuzzy numbers*, IEEE Transactions on Fuzzy Systems 14(2), 287-294, 2006.
- [23] Sánchez L., Otero J., Villar J.R., *Boosting of fuzzy models for high-dimensional imprecise datasets*, in Proceedings of the Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU06, Paris, 2006.
- [24] Sheen J. N., *Fuzzy evaluation of Cogeneration Alternatives on a Petrochemical Industry*, Computers and Mathematics with Applications 49, 741-755, 2006.
- [25] Snyder, J. P., *Map Projections Used by the U. S. Geological Survey*, 2nd edition, Geol. Survey Bulletin 1532, 313 p., U. S. Government Printing Office, Washington, D. C., 1982.
- [26] Sun H., Wu J., *A new approach for ranking fuzzy numbers based on fuzzy simulation analysis method*, Applied Mathematics and Computation 174, 755-767, 2006.
- [27] Tran L., Duckstein L., *Comparisons of fuzzy numbers using a fuzzy distance measure*, Fuzzy Sets and Systems 130, 321-341, 2002.
- [28] Trebi-Ollennu A., White B. A., *Multiobjective Fuzzy Genetic Algorithm Optimization Approach to Nonlinear Control System Design*, IEEE Proceedings on Control Theory and Applications 2(144), 137-142, 1997.
- [29] Wang Y-M., Greatbanks R., Yang J-B., *Interval efficiency assessment using data envelopment analysis*, Fuzzy Sets and Systems 153, 347-370, 2005.
- [30] Wilson, D., David L., *Wilson's GPS Accuracy Web Page*, <http://users.erols.com/dlwilson/gps.html>.
- [31] Yeh C-H., Deng H., *A practical Approach to fuzzy utilities comparison in fuzzy multicriteria analysis*, International Journal of Approximate Reasoning 35, 179-194, 2004.
- [32] Youssef H., Sait S.M., Khan A., *Fuzzy Simulated Evolution Algorithm for Topology Design on Campus Networks*, 2000 Congress on Evolutionary Computation 1, 180-187. Piscataway, New Jersey, USA.
- [33] Zhang J., Pham B., Chen P., *Fuzzy Genetic Algorithms Based on Level Interval Algorithm*, In Kazmierczak, E, Eds. Proceedings The 10th IEEE International Conference on Fuzzy Systems, pages pp. 1424-1427, Melbourne, Australia, 2001.