

Using fuzzy techniques for bounding the tolerance of GPS-based speed and distance measurements in taximeter verification

Adolfo Otero, José Otero, Luciano Sánchez
Computer Science Department. Oviedo University
Edificio departamental 1. Campus de Viesques s/n. 33204 Gijón - Asturias - Spain
{otero,jotero,luciano}@uniovi.es

Abstract

GPS sensors are a promising technique for verifying taximeters, because they do not require dedicated facilities and are compatible with a wide range of vehicles. The main drawback of this technology is based on legal issues: neither the absolute error of a GPS-based measurement nor the tolerance of the sensor can be known in advance, because they depend on environmental factors. In this paper we propose a technique that computes a dynamical tolerance for each measurement, using the Circular Error Probable at 50% and 95% levels. By combining the interpretation of a fuzzy set as a nested family of confidence intervals and a genetic algorithm-based interpolation, we have built an interval-valued estimation of the tolerance of a GPS-based verification of a taximeter.

1. Introduction

Taximeters are commonly verified by placing the drive wheels of the taxi on steel rollers with a known diameter. The angular speed of the rollers is regularly sampled. The test lasts a few minutes, and must be supervised by a technician. There are two variables of interest: the speed of the vehicle and the length of the trajectory, because a different fare is applied when the speed of the taxi surpasses a given threshold (changeover speed).

The use of a machine with rollers presents some drawbacks. First, the rollers have a relatively small radius, and the cab's tyres deform differently over the rollers than over a flat surface: tyres appear to be smaller for the system than they really are. Moreover, this error depends on the tyre conditions and the weight of the vehicle, making the whole test unreliable. Second, an employee can only verify one taximeter at a time, and this task lasts between 15 and 30 minutes. Third, there have been detected problems when verifying a taximeter in a car with electronic driving aids

(such as ESP, TCS, etc.)

In previous works [21, 22], we have introduced a new portable system, that uses a GPS sensor to sample the position and the speed of the taxi at regular intervals. Unfortunately, there are legal problems that raise difficulties in the use of GPS measurements, because the error of the measurements is not predictable; it depends on environmental factors. Unless we are able to bound the tolerance in our estimation of the length of the trajectory and the speed of the vehicle, we will not be able to legally reject a taximeter. In the same papers, we have also proposed to genetically filter the raw GPS data and remove redundant samples. This allowed us to obtain a tight upper bound on the tolerance, that fulfilled most of the legal requirements. However, that procedure was best suited for high-frequency GPS sensors, because if the frequency is low, there is a losing information risk. In this paper we go one step beyond trying to obtain bounds of length and speed without removing samples. The new technique is based on our own fuzzy interpretation of GPS data, and also on the genetic generation of a set of different trajectories that interpolate the data and are compatible with the probabilistic information encoded in the Circular Error Probable (CEP).

The structure of this paper is as follows: in the first section, we study the nature of a GPS measurement, and how it can be represented by a fuzzy set. In the second section we explain how to compute upper and lower bounds of the measurements that are compatible with the aforementioned representation. Sections 4 and 5 detail the experimental setup and numerical results of our empirical validation of the algorithm. Section 6 concludes the paper.

2. Uncertainty in GPS measurements

The term Global Positioning System (GPS) [10] refers to a set of devices (satellites, ground stations and receiver) working together to get a fix (the position) of the receiver. The receiver obtains some signals from the satellites and

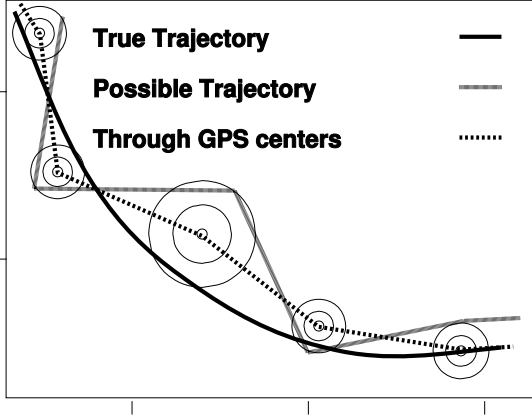


Figure 1. Simulated example where the differences between the true trajectory and the trajectory through the GPS coordinates are shown.

compute a set of measurements: longitude, latitude, altitude, number of satellites in use, time, etc. Each signal received from a satellite contains information about the time that the signal takes from the satellite to the receiver.

Under certain conditions, GPS measurement errors follow a bidimensional Gaussian distribution. When many satellites are available, that distribution can be regarded as circular [20]. Because of this, consumer grade GPS give an indication of their precision through a magnitude called Circular Error Probable (CEP). Given a probability threshold, the CEP indicates the radius of a circle. This circle is approximately centred on the position where the receiver was when it registered the measurement.

Consumer grade GPS, that will be used in this paper, do not send information related to the standard errors, but it is possible to carry out an empirical estimation of the CEP, whose details can be found in [4, 8, 13, 17].

3. Multilevel calculation of upper and lower bounds with fuzzy techniques

In certain contexts, fuzzy sets can be interpreted as a family of confidence intervals of a random variable [3]. In particular, we will consider that each measurement obtained by the GPS sensor is a fuzzy set, whose α -cuts are circles centered on the GPS coordinates. Each of these circles will be a confidence interval of the position of the taxi at the time when the measurement was taken.

In Figure 1, some simulated GPS measurements and trajectories are shown. The position where the measurement was taken is on the real trajectory -continuous line-, and it

can be inside or outside the respective circle of radius CEP. The trajectory using the GPS coordinates is drawn using a dashed line. A trajectory totally compatible with the GPS measurements is also drawn as a dotted line. Notice that the lengths of these trajectories are different, but all of them are compatible with the measurements of the GPS. Thus, to know the accuracy of the measure, we want to compute the upper and lower upper bounds of the lengths compatible with the CEP (for each confidence level).

Observe that, if CEP information is discarded, GPS measurements are crisp, and a spline model can be fitted to the data [5][6][2]. We will use instead an interval-valued approach [1]. Apart from this, we can also assume certain prior knowledge about the path: there are regularity conditions in the trajectory, as a straight line is unrealistic for most of the vehicles, because of inertia, and also because cars follow a path over a road, that has a smooth shape. Finally, we found a way to add heading and speed information, that is usually available from consumer grade GPS receivers. This measurements are more precise than coordinates measurements [19] and thus, improve the overall accuracy. This approach bears some relationship with the so-called Fuzzy Information Fusion Techniques [7]. In the next paragraphs we will explain how to interpolate physically realizable trajectories from GPS data, and how to bound their length.

3.1. Interpolation of a trajectory with fuzzy data and information fusion

A planar trajectory is represented by the equation

$$c(t) = x(t) \vec{i} + y(t) \vec{j} \quad (1)$$

Let us assume that each component of the trajectory is a polynomial,

$$a_n t^n + a_{n-1} t^{n-1} + \dots + a_1 t + a_0 = x(t) \quad (2)$$

(equations of the “y” component are similar) and therefore

$$n a_n t^{n-1} + (n-1) a_{n-1} t^{n-2} + \dots + a_1 = \dot{x}(t) \quad (3)$$

In words, at each GPS sample centered in $(x(t), y(t))$, the velocity is $(\dot{x}(t), \dot{y}(t))$ and Equation 3 holds, where t stands for time and a_i, b_i are constants to be estimated. This means that four equations are obtained for each sample. The number of unknowns is $2(n+1)$. A solution can be estimated if the number of measurements m satisfy $m \geq 2(n+1)$, thus a (possibly) over constrained system is obtained.

If that system is strictly constrained, the fitted polynomial

$$\hat{x}(t) = \hat{a}_{1,n} t^n + \hat{a}_{1,n-1} t^{n-1} + \dots + \hat{a}_{1,1} t + \hat{a}_{1,0} \quad (4)$$

contains the measured coordinates. If the system is over-constrained, the obtained model does not pass exactly over the measured GPS coordinates.

Given this fitted polynomial, the length of the trajectory can be estimated by integrating a length differential along the fitted polynomial between any two t values. Observe that the effect of assuming a straight trajectory between measurements is mitigated: the length measured over the polynomial will always be greater than the distance between coordinates.

In order to reformulate length estimation as a problem of optimization constrained to the CEP, we choose adding to $x(t)$ and $y(t)$ the appropriate increments or decrements that maximise or minimise the length of the trajectory. The easiest way to do this is to represent the increments or decrements using polar coordinates. Thus we will use two variables $\alpha \in [0, 2\pi]$ and $\rho \in [0, CEP]$. The system is transformed as follows:

$$\begin{aligned} a_{1,n}t^n + a_{1,n-1}t^{n-1} + \dots + a_{1,0} &= x_1(t) + \rho \cos \alpha \\ na_{1,n}t^{n-1} + (n-1)a_{1,n-1}t^{n-2} + \dots + a_{1,1} &= \dot{x}_1(t) \\ &\dots \\ a_{m,n}t^n + a_{m,n-1}t^{n-1} + \dots + a_{m,0} &= x_m(t) + \rho \cos \alpha \\ na_{m,n}t^{n-1} + (n-1)a_{m,n-1}t^{n-2} + \dots + a_{m,1} &= \dot{x}_m(t) \end{aligned} \quad (5)$$

With the addition of the deltas to the GPS coordinates, the length of the trajectory between to instants of time becomes a function of α and ρ because the constants that represent the polynomials are, also, a function of α and ρ . The model of the trajectory is now

$$\hat{x}(t, \alpha, \rho) = \hat{a}(\alpha, \rho)_{1,n}t^n + \hat{a}(\alpha, \rho)_{1,n-1}t^{n-1} + \dots + \hat{a}(\alpha, \rho)_{1,1}t + \hat{a}(\alpha, \rho)_{1,0} \quad (6)$$

and the length of the trajectory is, therefore, given by

$$L(t_i, t_j, \alpha, \rho) = \int_{t_i}^{t_j} \sqrt{\dot{\hat{x}}(t, \alpha, \rho)^2 + \dot{\hat{y}}(t, \alpha, \rho)^2} dt \quad (7)$$

The proposed schema is illustrated in Figure 2. Note that in Equation 5 the speed is crisp, in words, we *trust* these measurements. If the manufacturer of the GPS receiver provides some tolerance for speed and heading (similar to CEP), they could be used too.

For a trajectory with p points, a polynomial can be used to model each portion of the trajectory between two instants of time, and the length of the whole trajectory can be estimated by the addition of the lengths of each polynomial that fits the last two points. The trajectory then comprises $p - 1$ fragments, each one modeled by the polynomials of Equation 4 obtained solving the mentioned system. The total

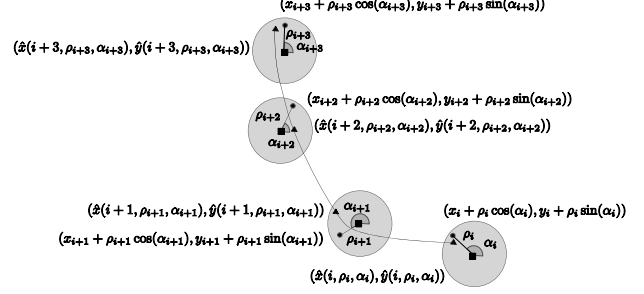


Figure 2. Polynomial fitted to the GPS measurements with the deltas added, compatible with a given CEP. The square marks are the original GPS measurements, the dots the coordinates plus the deltas and the triangles the estimation of the model.

length is:

$$L(\alpha_1, \rho_1, \alpha_2, \rho_2, \dots, \alpha_p, \rho_p) = \sum_{i=2}^{i=p} \int_{t_i}^{t_j} \sqrt{\dot{\hat{x}}(t, \alpha_i, \rho_i)^2 + \dot{\hat{y}}(t, \alpha_i, \rho_i)^2} dt \quad (8)$$

Using Equation 8, the problem of finding the maximum and minimum lengths compatible with GPS measurements is easily formulated. Note that the whole summation has to be maximised or minimised, not just each portion of the trajectory. Moreover, this can be done at different levels of probability, constraining ρ to the corresponding CEP at a given probability and a set of, nested, length intervals can be obtained from these estimations, using Equation 9. In this equation, P_1, P_2, \dots, P_q are CEP probabilities, and CEP_{P_q} is the corresponding CEP at that probability. Lastly,

$$\begin{aligned} L_{\min P_1} &= \min\{L(\alpha_1, \rho_1, \dots, \alpha_p, \rho_p)\} \rho_i \in [0, CEP_{P_1}] \\ L_{\max P_1} &= \max\{L(\alpha_1, \rho_1, \dots, \alpha_p, \rho_p)\} \rho_i \in [0, CEP_{P_1}] \\ &\dots \\ L_{\min P_q} &= \min\{L(\alpha_1, \rho_1, \dots, \alpha_p, \rho_p)\} \rho_i \in [0, CEP_{P_q}] \\ L_{\max P_q} &= \max\{L(\alpha_1, \rho_1, \dots, \alpha_p, \rho_p)\} \rho_i \in [0, CEP_{P_q}] \\ P_1 &< P_2 < \dots < P_q \end{aligned} \quad (9)$$

and $[L_{\min P_i}, L_{\max P_i}] \subset [L_{\min P_j}, L_{\max P_j}]$ for $P_i > P_j$.

4. Experimental Setup

The approaches presented in this paper have been tested with both realistic (synthetic) and real data. The purpose of the realistic data is to test the algorithm with a dataset of a-priori known properties: length of the trajectory, distribution of errors and so on. The purpose of the real data is

testing the approach in real world situations. In the following, both datasets are presented.

All data was translated to Universal Transversal Mercator (UTM) coordinate system for speeding calculations [16]. From each dataset we compute the upper and lower bounds of the trajectory length using: straight lines between GPS measurements with no speed and heading data, second degree polynomials with no speed and heading data and second degree polynomials *with* speed and heading data.

The optimization has been carried with Genetic Algorithms. We have chosen the Genoud optimization algorithm [11]. In particular, we have used the “rgnoud” implementation, available as a R package. We have chosen this algorithm because it combines evolutionary algorithm methods with a derivative-based (quasi-Newton) method and it can solve problems with many local minima and a large number of variables.

4.1. Synthetic Data

The synthetic dataset is obtained by simulation of 10 laps to a closed loop circuit with turns in all possible orientations, and with a speed that varies from a maximum in straight portions to a minimum in the hardest turns:

$$\begin{aligned} x(t) &= r \left(\cos \left(\frac{t}{t_f} n 2\pi \right) + e \right) * \sin \left(\frac{t}{t_f} 2\pi \right) \\ y(t) &= r \left(\cos \left(\frac{t}{t_f} n 2\pi \right) + e \right) * \cos \left(\frac{t}{t_f} 2\pi \right) \end{aligned} \quad (10)$$

where r stands for radius, t for time, t_f stands for the time to complete one loop (the longer this time, the less medium velocity), n stands for the number of convexities in the trajectory and e for eccentricity (the lesser this value, the rounder the trajectory is). The speed over the trajectory and the length between any to instants of time can be computed as explained in section 3.1. GPS measurements are simulated sampling that equation at one second intervals.

In order to simulate the behaviour of a typical GPS receiver, we have added gaussian noise to the original, such that the real position of the receiver at time t must lie inside of the CEP with a given probability. The velocity is also modified with random gaussian noise in order to simulate the measurement errors that affect this magnitude.

4.2. Real Data

We have tested our approach in a open street surrounding our campus in order to verify the quality of the results. In figure 4.2 an aerial image of the street can be seen, with the trajectory overimposed. The length of the circuit was measured using an ISO-9002 certified odometer, and the measured length was 1093 meters. In order to obtain some

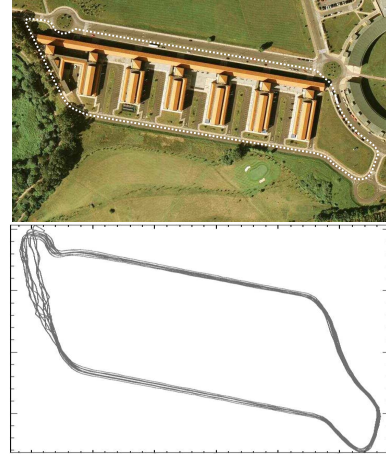


Figure 3. Top: Aerial picture of the real circuit used to test the approach presented in this paper. The test vehicle follows clockwise the circuit marked with a dashed line. Bottom: Plot of ten laps to the circuit.

statistics of the measured length and also for testing the behaviour of the stochastic nature of the approach, the test vehicle completed ten laps to the circuit. The collected data is shown in the same figure. Observe that the measured trajectories are not identical, due to GPS errors, mostly in the left part of the image, where the trees and the building jam GPS signals.

5. Numerical results

In this section we discuss the results obtained with the proposed approach on the datasets mentioned in the preceding section.

5.1. Synthetic data

In Table 1 the results (mean and standard deviation) obtained with the synthetic trajectory, using the three tested approaches, are shown. We have performed two batches of ten laps to the trajectory with 3 and 5 meters CEP. In Fig. 4 the length sample data is represented as boxplots, along with the results of the other approaches. From left to right, each group of four boxplots correspond to one approach. In each group of four, the first two (from left to right) are obtained with a CEP of 3 meters and the other two with a CEP of 5 meters. It is remarked that the tightest bounds are obtained adding the speed and heading measurements to the model.

Note also that with the proposed approach the minima are comparable but the maxima are much lower. This is be-

Straight Lines				
CEP	Min Mean	Min SD	Max Mean	Max SD
3	1041.878	8.754915	1553.377	24.56959
5	991.7894	8.175372	1975.620	28.81959
Second degree polynomials				
3	1033.29	6.047035	1427.402	13.70028
5	987.7802	5.593647	1698.075	19.02959
Second degree polynomials plus speed & heading				
CEP	Min Mean	Min SD	Max Mean	Max SD
3	1035.776	8.99255	1203.178	10.24138
5	990.5467	6.237333	1270.932	8.539436

Table 1. Mean and standard deviation obtained with synthetic data and the three tested approaches with 3 and 5 meters CEP.

cause the speed and heading measurements are much more precise and then the accuracy of the overall model is increased.

5.2. Real data

The aim of this section is to validate our approach with real data obtained with a consumer grade GPS in an open traffic environment. We choose to use the data obtained after ten turns of the circuit in Fig. 4.2. We perform the same experiments that we did with synthetic data: straight lines, second order polynomial and second order polynomial with heading and speed information added.

In Table 2, the results (mean and standard deviation) obtained with the trajectory shown in Fig. 4.2 are shown. Observe that the same situation that we found while dealing with synthetic data also happens here: the lesser the CEP, the tighter the obtained bounds are.

In Fig. 4, the dispersion of the lengths of the sample data is represented by boxplots, along with the results of the other approaches. From left to right, each group of four boxplots corresponds to one approach. In each group of four, the first two (from left to right) are obtained with a CEP of 3 meters and the other two with a CEP of 5 meters. The tightest bounds are found, again, when we use the speed and heading information in order to improve the accuracy of the results.

From the results above, we encourage to model the trajectory between consecutive points with second order polynomials, because the model is closer to the real behaviour of vehicle's trajectory. Moreover, if an overconstrained estimation is done, then a filtering process is implicit in the estimation that makes the length computation less susceptible of perturbation from erroneous GPS measurements. Finally, the last batch of experiments lead us to recommend the use

Straight Lines				
CEP	Min Mean	Min SD	Max Mean	Max SD
3	971.4588	18.32016	1391.11	71.80738
5	940.204	19.19713	1785.405	122.5640
Second degree polynomials				
CEP	Min Mean	Min SD	Max Mean	Max SD
3	965.7175	17.82051	1304.862	55.59313
5	931.8834	19.83234	1578.455	95.01664
Second degree polynomials plus speed & heading				
CEP	Min Mean	Min SD	Max Mean	Max SD
3	1032.780	6.586798	1206.878	8.433708
5	989.7233	9.371552	1277.49	11.48268

Table 2. Mean and standard deviation obtained with real data and the three tested approaches with 3 and 5 meters CEP.

of speed and heading information from GPS measurements too. This information has less errors and thus improves the precision of the whole process.

6. Conclusions and future work

In this work we have presented a novel approach to the measurement of lengths with a GPS based system, based on a fuzzy representation of GPS data. This allows the computation of nested upper and lower bounds of the length at different levels of confidence. The estimation is done by a search algorithm that fits those models that, being compatible with the GPS measured data, have maximum and minimum length. Moreover, we have proposed the fusion of speed and heading measurements with GPS coordinates measurement, in order to improve the overall precision.

This procedure can be expanded to any kind of device that provides speed, heading and even acceleration, and thus devices like IMU's, digital compasses or combinations of more than one GPS receivers can be added to the system to improve the precision of the computed bounds.

Acknowledgement

This work was funded by Spanish M. of Education, under the grant TIN2008-06681-C06-04.

References

- [1] A. M. Anile, B. Falcidieno, G. Gallo, M. Spagnuolo, and S. Spinello. Modeling uncertain data with fuzzy b-splines. *Fuzzy Sets Syst.*, 113(3):397–410, 2000.

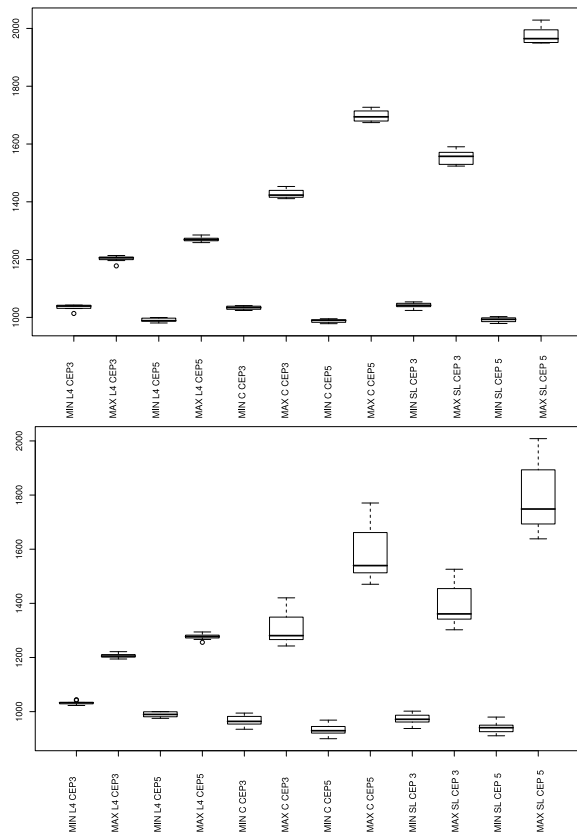


Figure 4. Top: Synthetic data. Bottom: Real data. From left to right: boxplots of minimum and maximum length for each approach at 3 meters CEP and 5 Meters CEP. 'L4' is fuzzy information fusion approach, 'C' uses second order polynomials fitted to the GPS measurements, 'SL' uses straight lines from one GPS measurements to the next. The tightest bounds are obtained with L4 approach, then 'C' and then 'SL'.

- [2] R.-S. R.-S. J. Castro M, Iglesias L. Geometric modelling of highways using global positioning system (gps) data and spline approximation. *Transportation Research Part C: Emerging Technologies*, 14(4):233–243, August 2006.
- [3] G. P. Couso I, Montes S. 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.
- [4] N. Cressie. *Statistics for spatial data*. Wiley, New York, USA, 1993.
- [5] M. R. David Ben-Arieh, Shing Chang and G. Zhang. Geometric modeling of highways using global positioning system data and b-spline approximation. *J. Transp. Engrg.*, 130(5):632–636, September/October 2004.
- [6] de Boor C. On calculating with b-splines. *Journal of Approximation Theory* 6, 50-62, 1972., (6):50–62, January 1972.
- [7] W. Dou, S. Ruan, Y. Chen, D. Bloyet, and J.-M. Constans. A framework of fuzzy information fusion for the segmentation of brain tumor tissues on mr images. *Image Vision Comput.*, 25(2):164–171, 2007.
- [8] O. J. P.-H. J. Dussault C., Courtois R. Influence of satellite geometry and differential correction on gps location accuracy. *Wildlife Society Bulletin*, 1(29):171–179, 2001.
- [9] I. R. Goodman and H. T. Nguyen. *Uncertainty Models for Knowledge-Based Systems; A Unified Approach to the Measurement of Uncertainty*. Elsevier Science Inc., New York, NY, USA, 1985.
- [10] C. Hofmann-Wellenhof, Lichtenegger. *GPS Theory and Practice*. Springer Verlag, 2004.
- [11] W. R. Mebane, Jr., and J. S. Sekhon. The rgenoud package, 2002.
- [12] d. B. R. Meratnia N. Trajectory representation in location-based services : problems and solutions. *Proceedings of the 3rd IEEE Workshop on Web and Wireless Geographical Systems*, 2003.
- [13] L. R.B. The mathematics of gps. *GPS World*, 2(7):45–49, 1991.
- [14] L. R.B. Dilution of precision. *GPS World*, 10(5):52–59, 1999.
- [15] I. Sanchez, L. Couso. Advocating the use of imprecisely observed data in genetic fuzzy systems. *Fuzzy Systems, IEEE Transactions on*, 15(4):551–562, Aug 2007.
- [16] J. P. Snyder. *Map Projections - A Working Manual*. U.S. Geological Survey Professional Paper 1395. U.S. Geological Survey, 1987.
- [17] G. Strang and K. K. Borre. *Linear algebra, geodesy, and GPS*. Wellesley, MA, USA, 1997.
- [18] C. I. C. J. Sánchez, L. Modeling vague data with genetic fuzzy systems under a combination of crisp and imprecise criteria. *First IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making*, 2007.
- [19] A. W. T. Witte. Interval efficiency assessment using data envelopment analysis. *Accuracy of non-differential GPS for the determination of speed over ground*, 37(12):1891–1898, 2005.
- [20] van Diggelen F. Gns accuracy: Lies, damn lies, and statistics. *GPS World*, 9(1):41–45, 2007.
- [21] J. Villar, A. Otero, J. Otero, and L. Sánchez. Taximeter verification using imprecise data from gps. *Eng. Appl. Artif. Intell.*, 22(2):250–260, 2009.
- [22] J. Villar, A. Otero, J. Otero, and L. Sánchez. Taximeter verification with gps and soft computing techniques. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 22, 2009.
- [23] Y. J.-B. Wang Y.-M., Greatbanks R. Interval efficiency assessment using data envelopment analysis. *Fuzzy Sets and Systems*, 153(3):347–370, January 2005.
- [24] C. P. Zhang J., Pham B. Fuzzy genetic algorithms based on level interval algorithm. *Proceedings The 10th IEEE International Conference on Fuzzy Systems*, pages 1424–1427, 2001.