GPS Signal Reception Classification Using Adaptive Neuro-Fuzzy Inference System

Rui Sun [1, 4], Li-Ta Hsu [2,*] Dabin Xue [1], Guohao Zhang [2] and Washington Yotto Ochieng [3, 1]

1 College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
2 Interdisciplinary Division of Aeronautical and Aviation Engineering, the Hong Kong Polytechnic University, Hong Kong;
3 Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK;
4 State Key Laboratory of Geo-Information Engineering, Xi'an Research Institute of Surveying and Mapping, Xi'an 710054, China
(E-mail: lt.hsu@polyu.edu.hk)

The multipath effect and non-line-of-sight (NLOS) reception of global positioning system (GPS) signals both serve to degrade performance, particularly in urban areas. Although, receiver design continues to evolve, residual multipath errors and NLOS signals remain a challenge in built-up areas. It is therefore desirable to identify the direct, multipath-affected and NLOS GPS measurements in order improve ranging based position solutions. The traditional signal strength-based methods to achieve this, however, use a single variable (e.g. C/N_0) as the classifier. Because the single variable does not completely represent the multipath and NLOS characteristics of the signals, the traditional methods are not robust in the classification of signal reception. This paper uses a set of variables derived from the raw GPS measurements together with an algorithm based on an adaptive neuro fuzzy inference system (ANFIS) to classify direct, multipath-affected and NLOS measurements from GPS. Results from real data show that the proposed method could achieve rates of correct classification of 100%, 91% and 84%, respectively, for the LOS, Multipath and NLOS based on a static test with special conditions results that are superior to the other three state-of-the-art signal reception classification methods.

KEYWORDS

1. NLOS. 2. Multipath. 3. Urban Canyon. 4. ANFIS

1. INTRODUCTION. Global Navigation Satellite Systems (GNSS) and, in particular, GPS, are used widely for positioning to support many applications and services. The multipath effect (resulting from the reception of both the direct and reflected signals) and non-line-of-sight (NLOS), or indirect, reception of GPS signals result in degraded performance, particularly in built environments such as urban areas. This in turn affects the provision of key services such as those designed to improve transport operations through intelligent transport systems (ITS). A number of ways have already been

developed to mitigate the effects of multipath and NLOS, including antenna design, signal processing, observable and measurement based modelling. Residual multipath effects remain a challenge, however, and, therefore, the detection of such signals would enable appropriate decisions to be made on whether or not to use the measurements for positioning.

Antenna design based methods include choke-ring antennas, dual-polarization and antenna array. Although effective in mitigating multipath effects from measurements at low elevation angles, such antennas are bulky, heavy and expensive, meaning that they are not appropriate for use in location based services. They are therefore mostly used in high precision and high accuracy geodetic applications (Tranquilla et al. 1994).

Signal processing based multipath mitigation was first proposed by Dierendonck et al. (1992) based on narrowing the spacing between early and late receiver code correlators. Subsequently, a popular auto-correlation function based technologies, multipath estimating delay lock loop (MEDLL), was proposed (van Nee et al. 1994). A review of these signal-processing-based multipath mitigation techniques can be found in Braasch (1996).In addition, numerical optimizations of correlator design, such as the Newton method-based fast iterative maximum-likelihood algorithm (FIMLA) (Sahmoudi and Amin 2008), and the space-alternating generalized expectation maximization algorithm (Fessler and Hero 1994), have also been developed to reduce the impact of multipath effects further. Advanced receiver-architecture vector tracking is also a promising technique to detect multipath effects (Bhattacharyya and Gebre-Egziabher 2014; Hsu et al. 2015). Ziedan (2012) has proposed a principal components analysis (PCA) and probabilistic neural network based method to deal with multipath effects in the adaptive tracking. Sokhandan et al. (2017), meanwhile, have proposed a support vector machine (SVM) based adaptive multipath compensation and tracking strategy to extract context information about the type of multipath environment and the state of motion of a GNSS receiver. Vector tracking, however, requires altering the traditional architecture of the GNSS receiver, which is not currently possible for low cost GNSS receivers. The signal processing based methods are therefore designed to mitigate only certain types of multipath effect, and no single method exists that accounts for all these effects.

Observable and measurement based methods use observables, measurements, satellite and signal information to mitigate multipath effects. Such information can be used to smooth the measurements. Smoothing techniques are based on the fact that carrier phase measurements are less sensitive to multipath signals than code measurements. Furthermore, satellite elevation and signal information (such as the carrier to noise power density ratio) can be used to apply weighting to the measurements in order to contribute to the position solution according to measurement quality in relation to multipath effects (Euler and Goad 1991; Kuusniemi 2005). Another way of dealing with multipath effects is to integrate GPS with complementary sensors not affected by such effects. A common approach is to integrate GPS observables or measurements with Inertial Measurement Units (IMU) (Cox 1978). The coupling, or integration, could be in the position domain (i.e. using positioning outputs of GPS), measurement domains (i.e. using carrier phase or code observables from GPS) or

tracking loop domain (i.e. using the code/carrier discriminator output of GPS). Although GNSS/IMU integration is effective if high grade IMU sensors are used, the expense associated with such sensors precludes their use in location based services.

In the case of NLOS reception, the ranging measurement errors that result from NLOS reception are different from those produced by multipath interference, and therefore cannot be corrected by most multipath mitigation techniques (Groves et al 2013). Although it has been shown that the dual-polarization antenna is able to detection NLOS reception, such an antenna is expensive and bulky compared with the patch antennas, again preventing their use for location based service. (Izadpanah et al. 2008; Jiang and Groves 2014; Palamartchouk et al. 2015). Another potential solution is to apply consistency check, which is based on multipath/NLOS contaminated measurement is not consist with other clean measurements (Groves and Jiang 2013). Consistency check between pseudorange measurements could detect and exclude multipath and NLOS effects when the number of clean measurement is sufficient (Hsu et al. 2017b).

More recently, research has focused on the use of spatial data (e.g. 3D city models) to assist in the detection of NLOS reception and to improve positioning, such as GNSS shadow matching in built environments (Groves 2011). In these approaches, the 3D buildings are used to detect satellite visibility, enabling the NLOS to be used actually to improve positioning accuracy, in contrast to the methods that detect and exclude NLOS that have the effect of significantly lowering satellite availability for positioning (Betaille et al. 2013; Peyraud et al. 2013). Adding NLOS by the prediction of satellites' azimuth and elevation angles, a cross-street accuracy of 5 m (54.3%) can be achieved by shadow matching (Wang et al. 2015). Subsequently, the NLOS measurement is further corrected using the method called ray-tracing simulation. It has been claimed that a positioning accuracy of 4.4 m (1σ) is achievable for pedestrian applications (Hsu et al. 2016). Research has been released illustrating the acceleration of ray tracing simulation with an associated enhancement of the 3D building model GNSS positioning method (Ziedan 2017). Recently, shadow matching has been integrated with range-based 3D mapping aided algorithms (Adjrad and Groves 2016; Suzuki 2016).

To use the NLOS, it must be detected correctly. Positioning accuracy is therefore highly dependent on the ability to classify the signals correctly. The satellite reception classification method is not robust enough, however, to achieve this when using the traditional single variable based method (usually C/N_0 based), which will always be suboptimal, since the status cannot be exactly known when it is determined only by one variable, especially in complex city environments. The concept of applying machine learning to improve GNSS positioning accuracy and classify measurement types has emerged in recent years, and has exhibited superior ability by considering various sources of information from the GNSS measurement to improve the successful rate of signal reception classification. The general idea is demonstrated in Figure 1.

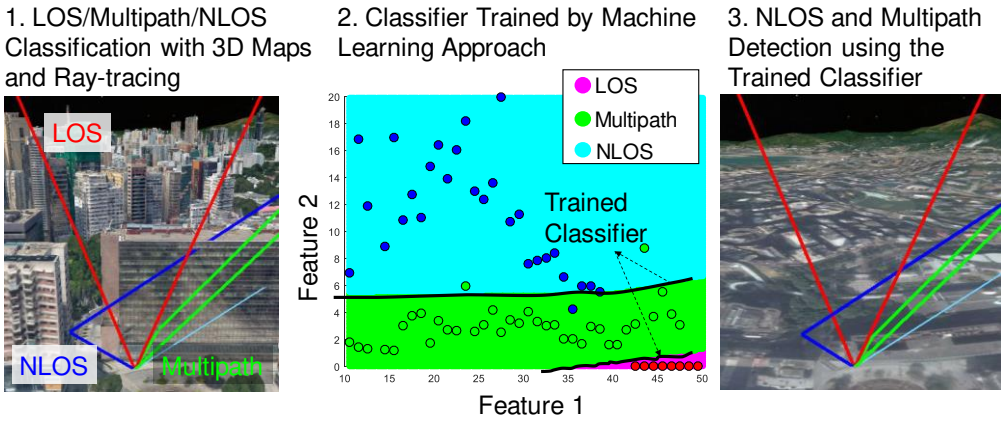

Figure. 1 Demonstration of the idea for training using a machine learning approach.

The related research can be illustrated as follows. Wang et al. (2013a) proposed the Wilcoxon-norm-regressor based on pseudorange residuals to detect biased pseudorange measurements. Phan et al. (2013) used elevation and azimuth angles as the key features for SVM to mitigate the multipath effect for static applications. Sokhandan et al. (2017) extracted variables from the correlators in the receiver signal processing stages to classify the different types of scenarios of GNSS receivers. A decision tree based approach has also been used to classify LOS and NLOS based on received signal strength and elevation angle (Yozevitch et al. 2016). It was indicated the accuracy of the classification prediction will be highly dependent on how the tree is designed according to the selection of features. It is difficult to obtain a trade-off balancing sufficiently high accuracy with sufficiently low computation load with this decision tree algorithm, since the features have to be selected manually. If we could design an algorithm which is able to choose the related features automatically, however, based on all the known representative variables and then use these related features for the classification algorithm that would improve the LOS/NLOS classification performance.

This paper develops a novel adaptive neuro fuzzy inference system (ANFIS) based algorithm to classify the received signals using the known representative variables from GPS raw measurements: i.e. the 1) received signal strength (RSS), 2) temporal difference of received signal strength (ΔRSS), 3) horizontal dilution of precision (HDOP), 4) vertical dilution of precision (VDOP), 5) satellite elevation angle (EA), 6) azimuth angle (AA), 7) pseudorange residual (η), 8) consistency between delta pseudorange and pseudorange rate (ζ) and 9) number of visible satellites (NS). The ANFIS is a machine learning method that uses a combination of neural network and fuzzy logic to produce a more reliable output by considering various input sources. The application of ANFIS varies from different disciplines, such as disease diagnosis, control engineering and share price forecasting etc. (Ubeyli 2009; Jilani et al. 2015; Wei 2016). It potentially has great advantages over the other machine learning methods for the GPS signal reception classification application. For example, as mentioned in Pradhan (2013), the ANFIS based models have exhibited superior predictive ability than the other machine learning methods (such as decision tree and SVM) in landslide susceptibility mapping. In the training phase, the algorithm is aided by ray-tracing

simulation and a 3D city model for offline data labelling. PCA is then used to extract the principal components, which are fed into the ANFIS based training algorithm to output the Fuzzy Inference System (FIS) rules. The variables captured in real-time are processed based on the extraction of PCA features, and are then fed into the FIS rules in order to output the classification results. The designed ANFIS based algorithm requires no additional hardware costs, and is therefore suitable for city based applications with low-cost GNSS receivers. The method is also easily extendable to multi-constellation GNSS for practical implementation. The multi-variable based classification, considering various sources of information from GPS measurements, has the potential to provide a robust signal reception classification, and therefore, address the limitations of the current single variable based classification approach.

2. ALGORITHM DESIGN.

2.1. *Process Overview.* Flowchart of the ANFIS based LOS/Multipath/NLOS classification is presented in Figure 2.

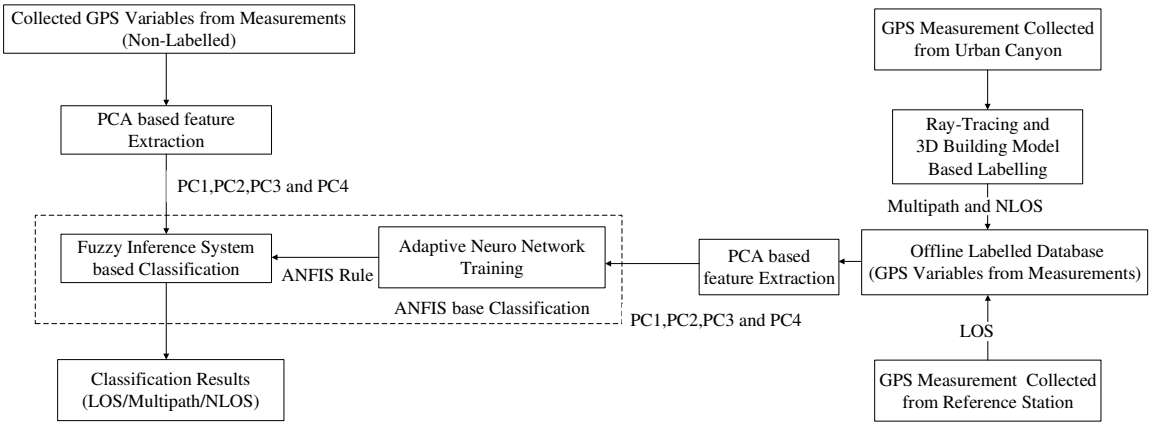

Figure 2. The Flow Diagram of the Algorithm

First, the offline dataset is created, including a large amount of LOS, multipath and NLOS data. The LOS data is collected and labelled for the GPS measurements from a GPS reference station, which is in an open area, therefore offering clean data. The multipath and NLOS data are collected in a densely built-up area in static mode, and labelled based on the ray-tracing and 3D building model (Section 3.1). PCA is then applied to extract the relevant combination of variables for dimension reduction (Section 2.2). Based on the dataset, the ANFIS based training is carried out to output the trained FIS rules (Section 2.3). These rules are then used together with the principal components extracted from the actual real-time measurements by FIS for classification of the signals in terms of LOS, multipath and NLOS.

2.2. *Determination of Variables for GPS LOS, Multipath and NLOS Signal Classification.*

2.2.1 *Derivation of GPS Variables from Raw Data.* The raw data consist of the pseudorange and carrier phase measurements, the carrier to noise ratio, and the Doppler shift frequency in the RINEX format (Gurtner 1994). It is therefore assumed in this paper that the potential features extracted from the GPS raw data can be obtained from most new GNSS devices. The received GPS signal contains a variety of information that can be used to determine signal reception. These differences can be distinguished by the careful application of a number of variables as follows.

RSS: It is usually represented by C/N_0. The effect of signal reflection and additional travel time is to increase the signal propagation loss. C/N_0 is, therefore, commonly used in the mitigation of multipath effects (Hartinger and Brunner 1999).

$\Delta\boldsymbol{RSS}$: Due to the estimation of received signal strength in the receiver tracking loop, the received signal strength of multipath and NLOS could increase if the antenna stays static. The temporal difference in the received signal strength is calculated as:

$$\Delta RSS_k^{(i)} = RSS_k^{(i)} - RSS_{k-1}^{(i)} \tag{1}$$

where *i* is the satellite and *k* the epoch. As shown in previous research, the speed of the antenna is strongly related to the change rate of the multipath effect. (Kubo et al. 2017).

HDOP and VDOP: HDOP and VDOP indicate the strength of the geometric distribution of satellites in relation to the user's horizontal and vertical positioning dimensions, respectively. Conventionally, positioning error is closely related to the product of the strength of the geometric distribution (i.e. DOP) and the measurement error. In addition, areas with strong multipath and NLOS effects tend to have larger DOP values, mainly because of the surrounding physical features.

EA and AA: The measurements transmitted from higher elevation angles experience less multipath effects. It is therefore common to use the elevation angle as a weighting metric to reduce the multipath effect on the positioning accuracy (Euler and Goad 1991).

In addition to the variables above, the consistency of the measurements should be considered, including pseudorange residual and the consistency between delta pseudorange and pseudorange rate.

$\boldsymbol{\eta}$: Least squares estimation is a basic optimization method used for state estimation. The receiver state is estimated as:

$$r = (G^T G)^{-1} G^T \rho \tag{2}$$

where r is the receiver state, including three-dimensional position and the clock offset between the receiver and GPS system time. ρ denotes the pseudorange measurements. G denotes the design matrix consisting of the unit LOS vector between the satellite and receiver. The inconsistency between the pseudorange measurements represented by η is calculated as:

$$\eta = \rho - G \cdot r \tag{3}$$

Hsu et al. (2017b) have shown that the pseudorange residual could potential used as an indicator to exclude the multipath and NLOS signal if the number of measurements is sufficient.

$\boldsymbol{\zeta}$: The pseudorange and Doppler shift are estimated by the delay and frequency lock loops, respectively. This is the reason why they are independent if their minor cross-correlation is neglected. Delta pseudorange indicates the change of pseudorange between two epochs. It is calculated by:

$$\Delta\rho_k^{(i)} = \rho_k^{(i)} - \rho_{k-1}^{(i)} \tag{4}$$

where ρ is the pseudorange measurement and *k* the time in seconds. The pseudorange rate $\dot{\rho}$ is the change of pseudorange between two epochs. It is calculated from the Doppler shift as:

$$\dot{\rho}^{(i)} = \frac{c\left(f_{Doppler}^{(i)}\right)}{f_{L1}} + \left(v_x^{(i)} \cdot u_x^{(i)} + v_y^{(i)} \cdot u_y^{(i)} + v_z^{(i)} \cdot u_z^{(i)}\right) \tag{5}$$

where $f_{Doppler}^{(i)}$ is the Doppler shift in Hz, *c* the speed of light, f_{L1} the GPS L1 band carrier frequency of 1575.42 MHz, *v* is the speed of the satellite, *u* is the unit LOS vector between the satellite and receiver. Thus, their difference can be calculated as:

$$\zeta = |\Delta\rho - \dot{\rho}\,\Delta t| \tag{6}$$

where Δt is the time difference between two epochs.

NS: The number of the satellites tracked by the GPS receiver could be easily obtained from the National Marine Electronics Association (NMEA) GPGGA message.

The nine variables are the inputs to the feature extraction algorithm.

2.2.2 *Feature Extraction Using Principal Components Analysis.* PCA is applied to the GPS variables for pre-processing to extract the key features, reduce the dimensions and, therefore, simplify the rules for the ANFIS in the next step. The details of PCA are in Smith (2002). The first four Principal Components (PCs) extracted, which are linear combinations of the input GPS variables, represent 96.82% of the whole measurement information are shown in Figure 3. The blue line indicates the value of the accumulated percentage of the first m PCs. In this figure, we have obtained the first four PCs, which contribute to 96.82% of the whole features. Table 1 has shown the links between the GPS variables and the extracted PCs. The nine derived GPS Variables, mentioned in section 2.2.1 are denoted as x_1 ... x_9 respectively. The values within Table1 show the weights of the defined variable for the corresponding PC. Positive values indicate a positive correlation of the variables to the corresponding PC, while negative values

indicate a negative correlation. The larger the value of the corresponding variable, the more importance of the variable for the PC. Therefore, the linear combination equations could be derived for the first four PCs based on Table1. It is indicated that PC1 is mainly the linear combination of the number of satellites and C/N_0. PC2 is mainly formed by the linear combination of the azimuth angle. PC3 is mainly formed by the linear combination of the elevation. PC4 is mainly formed by the C/N_0 and the number of satellites.

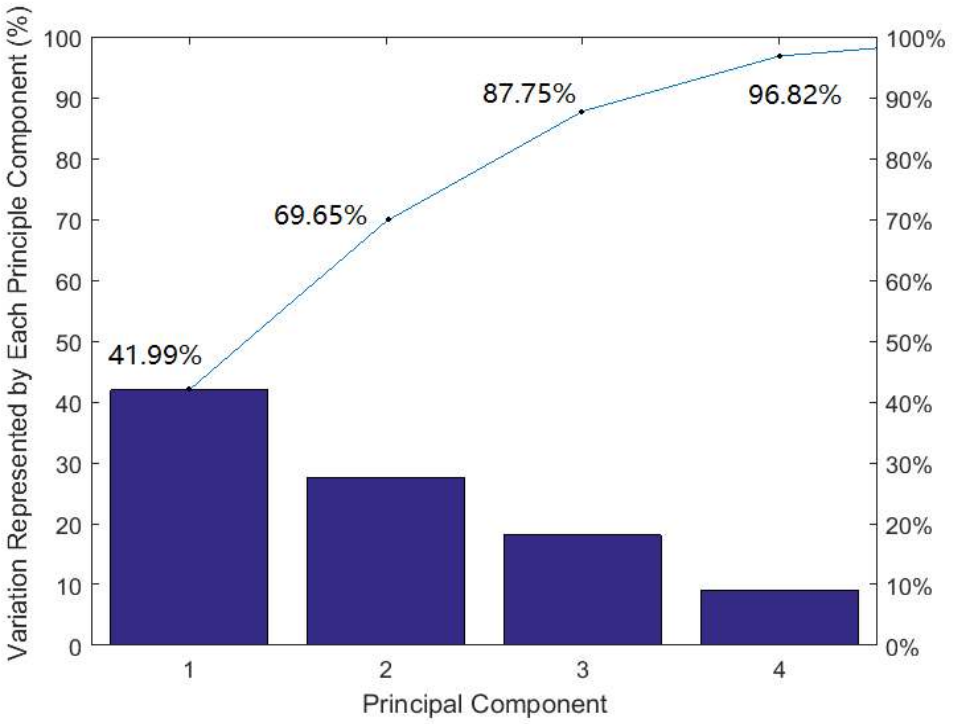

Figure 3. The first four PCs extracted and the accumulated variation represented by each PC

Table 1 The relationship between the GPS variables and the extracted PCs

Variable Name	Variable	Principal Component			
		PC1	PC2	PC3	PC4
x_1	RSS	0.517277	0.247048	0.457374	0.56531
x_2	ΔRSS	-0.01006	-0.08638	-0.23553	-0.35748
x_3	HDOP	-0.1284	-0.00479	0.025696	0.111117
x_4	VDOP	-0.06398	-0.00234	0.025046	0.085101
x_5	EA	-0.24635	-0.03352	0.8483	-0.46645
x_6	AA	-0.19563	0.962334	-0.10485	-0.15625
x_7	η	-0.00347	0.00926	-0.00656	-0.02479
x_8	ζ	0.000425	0.000393	0.000871	-0.00204
x_9	NS	0.782795	0.064657	-0.05827	-0.53893

2.3. *ANFIS based Classification Model.* Generally, ANFIS is the integration of neural network (NN) architectures with FIS. It is able not only to take linguistic rules from human experts, but also to adapt itself using input-output data to achieve better performance. Mamdani and Sugeno are two types of basic fuzzy systems. In these systems, the first two parts of the fuzzy inference process, fuzzifying the inputs and applying the fuzzy operator, are the same. The main difference between Mamdani and Sugeno is that the Sugeno output membership functions are either linear or constant. In ANFIS, therefore, Sugeno type is used since it can output linear or constant membership functions (Jang 1993).

From our initial analysis, although the classification accuracy could be improved if we increase the use of PCs, the similar accuracy results are obtained by using 4PCs and 5PCs. However, the computation load of 5PCs is much higher than that of 4PCs. Therefore, 4 PCs are adopted for our ANFIS algorithm. The five layers for the generated structure of the ANFIS based classification are presented in Figure 4.

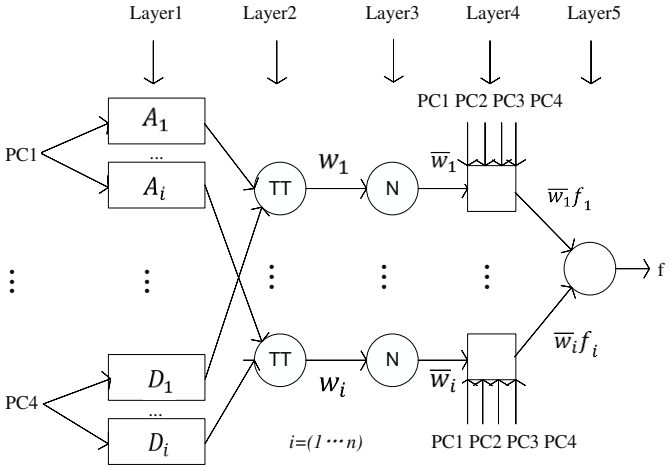

Figure 4. Structure of the ANFIS based GPS LOS, multipath and NLOS signal classification.

Layer 1 assumes that every node i in this layer is shown as a square with a node function.

$$O_i^1 = \mu_{A_i}(\mathrm{PC1}) \tag{8}$$

where, PC1 is the input of node i, and A_i is the linguistic lable (e.g. small, medium, large, very large) with node i. O_i^1 is the membership function of A_i and it specifies the degree to which the given PC1 satisfies the quantifier A_i. In our case, the initial membership functions $\mu_{A_i}(\mathrm{PC1})$ are set as Gaussian function with maximum value equal to 1 and minimum value equal to 0 based on the characteristics of the input information.

$$\mu_{A_i}(\mathrm{PC1}) = Gaussian\,(\mathrm{PC1}; \sigma, c) = \mathrm{e}^{e^{-\frac{(\mathrm{PC1}-c)^2}{2\sigma^2}}} \tag{9}$$

where c is the parameter to determine the centre of the membership function and σ determines the width of the curve. As the values of these parameters change, the Guassian membership functions vary accordingly, thus exhibiting various forms of membership functions on linguistic label A_i.

PC2, PC3, PC4 and their corresponding initial membership functions $\mu_{B_i}(\text{PC2})$, $\mu_{C_i}(\text{PC3})$ and $\mu_{D_i}(\text{PC4})$ are defined based on the same method. The initial rule could then be extracted based on the first-order Sugeno fuzzy model (Takagi and Sugeno, 1983).

$$f_i = p_i * \text{PC1} + q_i * \text{PC2} + r_i * \text{PC3} + s_i * \text{PC4} + t_i \quad (10)$$

where, the membership functions will be modified along with the parameters p_1, q_1, r_1, s_1 and t_1 during the following NN training. f_i is the initial i-th rule. The parameters in this layer are considered as the premise parameters.

In Layer 2, every node is a circle node labelled $\prod$, which multiplies the incoming signals and sends the product out. Each node in this layer calculates the firing strength of each rule via multiplication, noted as w_i. In our case, we use the AND T-norm operator here, given by,

$$O_i^2 = w_i = \mu_{A_i}(\text{PC1}) * \mu_{B_i}(\text{PC2}) * \mu_{C_i}(\text{PC3}) * \mu_{D_i}(\text{PC4}), i = 1,2,3,4 \quad (11)$$

In Layer 3, every node is a circle node labelled N. The ith node of this layer calculates the ratio of the ith rule's firing strength to the sum of all the rules' firing strengths, which is represented by $\overline{w}_i$.

$$O_i^3 = \overline{w}_i = \frac{w_i}{w_1+w_2+w_3+w_4}, i = 1,2,3,4 \quad (12)$$

In Layer 4, the multiplication of the input from Layers 3 and 1 is implemented, given by:

$$O_i^4 = \overline{w}_i f_i' = \overline{w}_i(p_i' * \text{PC1} + q_i' * \text{PC2} + r_i' * \text{PC3} + s_i' * \text{PC4} + t_i') \quad (13)$$

where $\overline{w}_i$ is the output of layer 3 and $\{p_i'\ q_i'\ r_i'\ s_i'\ t_i'\}$ is the parameter set. Parameters in this layer are called consequent parameters, which have been modified after NN training. f_i' here is the output of the i-th rule.

Layer 5 computes the overall outputs as the summation of all incoming signals.

$$\sum_i \overline{w}_i f_i = \frac{\sum_i w_i * f_i}{\sum_i w_i} \quad (14)$$

Subtractive clustering is applied for the initial FIS design in order to reduce the computation complexity (Chiu, 1994). In addition, during the learning process, the premise parameters in Layer 1, and the consequent parameters in the Layer 4, are tuned until the desired response of the FIS is achieved.

3. FIELD TEST AND RESULTS ANALYSIS.

3.1. *Experiment Setup – Data Collection, Labelling and Processing.*

We created five datasets collected from four different locations. The relationships between these datasets are depicted in Figure 5.

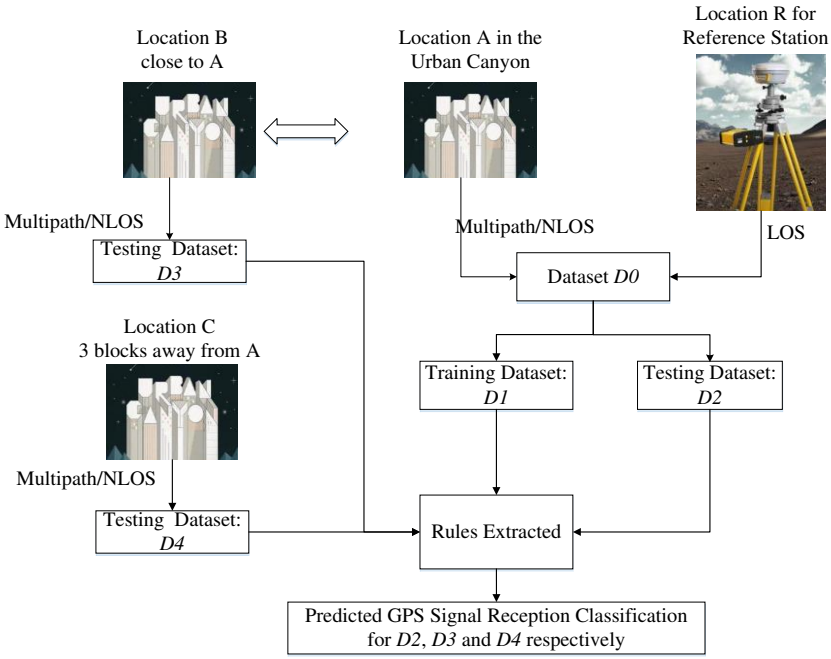

Figure 5 The relationship of the datasets in the field test

To create the combined dataset *D0*, two types of data are collected. One is collected in location A from an urban canyon (NLOS and multipath) and the other is collected from location R in the SatRef HKSC station (LOS). To record a large amount of multipath and NLOS data, a static experiment was carried out in location A, a densely built-up area in Hung Hom, Hong Kong. The left and middle panels of Figure 6 show the environment in which the data for location A were collected. The antenna was attached to a pole out of a window. A commercial GPS receiver, u-blox NEO-M8T, was deployed to collect multipath and NLOS data. 24 hours of raw GPS measurements were collected. In this dataset, and because of the environment, most of the measurements were affected by the buildings in the vicinity. In other words, the urban dataset should consist predominantly of multipath and NLOS measurements.

A ray-tracing method to identify the signal reception types is applied for labelling the multipath and NLOS signals in the urban canyons. The principle of ray-tracing in GPS is to use known satellites, reflector and receiver geometry to trace the direct and reflecting paths (Lau and Cross, 2007). Satellite positions can be estimated from the broadcast ephemeris. A 3D building model was used to search for reflectors. The ground truth was provided by the topographic map from the Land Department of the HK government with the 2D accuracy of 20cm. The height was determined from Google map plus the height of the equipment.

Once the positions of the satellite, reflector and receiver were known, ray-tracing was performed. The right panel of Figure 6 shows the skyplot with building sites. This

skyplot was generated using ray-tracing simulation and a 3D building model. The grey area indicates where the direct transmission was blocked according to the building models. If the elevation and azimuth angles of a satellite were blocked according the tailored skyplot, its measurements were labelled as NLOS. Otherwise, they were labelled as multipath, i.e. we assume that all non-NLOS measurements are multipath. Regarding LOS (clean) GPS data, the HK Land Department has established a GPS network called SatRef to provide differential corrections for HK users. The archived RINEX data of the location R, which is the SatRef HKSC station, were used as the LOS data because it is located in a clear environment. 24 hours of clean data were collected with a measurement update interval of 30 seconds.

The training dataset *D1* is randomly selected from the dataset *D0* at first and the testing dataset *D2* is randomly selected from the rest of the dataset *D0* (exclude *D1*). Although some of the variables will be correlated over time, yet the time dependency of the data does not affect the ANFIS performance, since the algorithm treats each epoch as an independent event. The training dataset, *D1*, contained 24,000 measurement samples, with nine variables for each sample, of which a third each were labelled as LOS/multipath/NLOS data (labelled by the 3D map and ray tracing), were processed with PCA to extract the four principal components to feed the ANFIS for offline training. Then the ANFIS rules were extracted from this training.

The testing dataset, *D2*, contained 24,000 measurement samples, with nine variables for each sample, of which a third each were LOS, multipath and NLOS (reference label - labelled by the 3D map and ray tracing). To determine the performance of the input testing dataset, we first created a reference label called 'unknown', manually deleting the labelled signal reception in advance. We then marked the signal reception as 'unknown' for the test set and processed this with PCA, and then fed the extracted features into the ANFIS rules trained from the dataset *D0*. We then compared the results from the ANFIS predicted signal reception with the reference label to compute the accuracy. In order further to verify the validity of the extracted rules, two more testing datasets collected from other locations were also used to feed the rules. One testing dataset, *D3*, was collected from location B, which is close to location A in the urban canyon, while the other testing dataset, *D4*, was collected from location C, which is about three blocks away from location A in the urban canyon, see Figure 7. The summary of the datasets is shown in Table2.

Table 2 The summary of the datasets

Dataset ID	*D0*	*D1*	*D2*	*D3*	*D4*
Dataset Type	Combined	Training	Testing	Testing	Testing
Total samples	96992	24000	24000	25033	11572
LOS (labelled as 1)	25987	8000	8000	0	0
Multipath (labelled as 0)	18164	8000	8000	8830	3087
NLOS (labelled as -1)	52841	8000	8000	16203	8485

Figure 6. The left panel indicates the installation of the patch antenna, the middle panel illustrates the data collection environment and the right panel indicates the skyplot with building sites used to label NLOS and multipath measurements

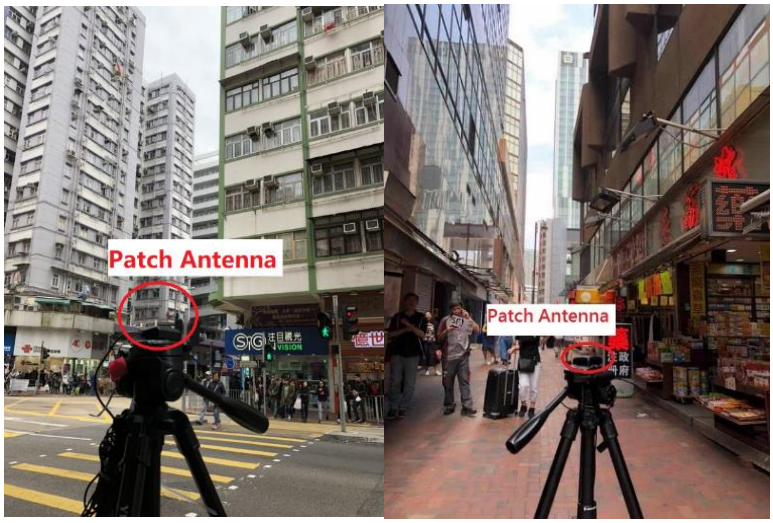

Figure 7. GPS data collection environment for location B (left) and location C (right)

3.2 *Results Comparison and Analysis.* Figure 8 compares the results from the ANFIS prediction with the reference label (signal reception labelled by the 3D map and ray-tracing) based on dataset *D2*. It is clear that the LOS signals can be identified with a high accuracy of 100%, while the errors of classification are entirely distributed in the multipath and NLOS areas. The predicted values are further rounded to the closest value of 1, 0 or -1. For example, if the ANFIS predicted value is within the interval of [-0.5, 0.5], it will be rounded to the value 0 (multipath). Some NLOS measurements were misclassified as multipath in the experimental results. The reason for the errors distributed in the multipath and NLOS areas are attributable to 3D map border errors and scattered reflection in the labelling stage. In this paper, the shapes of the buildings are just considered to be cuboid, whereas in reality the roofs of some buildings are likely to have different shapes, errors may be induced during the labelling stage when we are using the ray-tracing method. In addition, we don't consider the scattered reflection when using ray-tracking, which means we assume the signal could only be reflected once during the propagation.

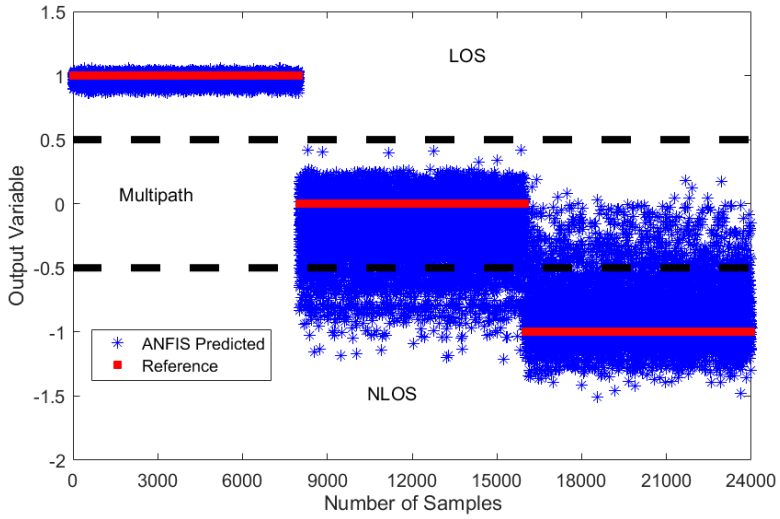

Figure 8. Comparison of the ANFIS predicted results for testing dataset *D2* and the labelled reference. 1, 0 and -1 of y-axis denote LOS, multipath and NLOS for the output variable, respectively.

The following paragraphs compare the proposed ANFIS based algorithm with other state-of-the-art algorithms for signal reception identification based on different testing datasets. Specifically, Decision Tree and SVM, two typical machine learning methods that could be used for the identification of satellite visibility (Yozevitch et al., 2016, Hsu, 2017a), both for multi-variable based classification and traditional single C/N0 based classification, are compared with the ANFIS-based algorithm designed here.

The confusion matrix of the LOS, multipath and NLOS (1, 0 and -1) classification results for different algorithms based on the testing datasets *D2* are listed and compared in Table 3. The rows labelled Accuracy is calculated as the ratio (as a percentage) of the number of correctly detected activities to the total number of known activities (Blasch et al 2011). It is obvious that the proposed ANFIS algorithm outperforms Decision Tree, SVM and traditional C/N_0 based classification in the GPS signal reception classification, with a total classification accuracy of 91.8%. NLOS detection accuracy is calculated as the ratio (as in percentage) of the number of instances correctly classified as NLOS to the total number of known NLOS instances, which is labelled NLOS. In particular, the ANFIS based algorithm significantly improves the accuracy of the NLOS detection with an accuracy of 91.1%, while it is 49.5%, 65.1% and 64.9% for the decision tree, SVM and single C/N_0 based algorithms.

The multipath detection accuracy is calculated as the ratio (as a percentage) of the number of instances correctly classified as multipath to the number of total known instances of multipath, which is labelled multipath. For multipath detection accuracy, the ANFIS also has a high detection accuracy of 84.1%. Although the decision tree based method performs slightly better with a detection accuracy of 89.5%, it also has a higher number of misdetections, where NLOS signals are incorrectly identified as multipath. It is also noted that the overall performance of the classification accuracy for the multi-variable based algorithms (Proposed ANFIS, Decision Tree and SVM) is superior to the single C/N_0 based algorithms.

Table 3: Confusion matrix of LOS (noted as 1) , multipath (noted as 0) and NLOS (noted as -1) classification results for testing dataset *D2*

		Algorithms Predicted Results					
		Proposed ANFIS			**Decision Tree**		
		-1	0	1	-1	0	1
Label Results	-1	7291	709	0	3963	4036	1
	0	1269	6731	0	837	7163	0
	1	0	0	8000	0	0	8000
Accuracy		91.8%			79.7%		
Accuracy for each class		-1	0	1	-1	0	1
		91.1%	84.1%	100%	49.5%	89.5%	100%
		SVM			**C/No based Classification**		
		-1	0	1	-1	0	1
Label Results	-1	5206	2793	1	13499	6603	708
	0	1995	6005	0	3152	3339	1902
	1	0	0	8000	0	779	10018
Accuracy		80.1%			67.1%		
Accuracy for each class		-1	0	1	-1	0	1
		65.1%	75.1%	100%	64.9%	39.8%	92.8%

Although the proposed ANFIS based algorithm has superior performance based on the testing dataset *D2*, more testing databases from other locations were used to verify the validity of the proposed algorithm. As we have mention in section 3.1, dataset *D3* and dataset *D4* were collected to feed the extracted rules from training dataset *D1* to identify the GPS signal reception classification. Since locations B and C are both in the urban canyon, we can only obtain the NLOS and multipath signals. The labelling of NLOS and multipath is also based on the 3D city model and ray-tracing introduced in section 3.1. It should be noted that it is difficult for us to get NLOS, LOS and multipath data in one environment. Even if this kind of environment exists, it would still not be possible to label the data based on the current 3D city model and ray-tracing based method as one cannot determine the 'true' LOS existing in the area where multipath and NLOS exist.

The classification accuracy based on different algorithms for the testing datasets *D3* and *D4* are listed and compared in Table 4. The ANFIS has a similar performance as SVM but with a higher calculation speed. ANFIS will therefore be more suitable than the SVM for real-time applications. The classification accuracy of the ANFIS based results for datasets *D3* and *D4* is much worse than it is in the case of dataset *D2*, which indicates the data sensitivity of the proposed algorithm. In addition, we have found that when conducting a PCA process on the datasets *D3* and *D4*, although the main components in the PCs were the same as they were in *D0*, *D1* and *D2*, the percentages

for the main PCs vary across the different collecting locations. Furthermore, we have also found that if we use parts of the data from *D3* or *D4* for training and the rest for testing, respectively, the proposed ANFIS could still perform with very high accuracy. This means that for the proposed ANFIS algorithm, if the training and testing data are from the same dataset, highly accurate classification results can be obtained. That bring us to the idea that the rules extracted from the ANFIS could not be suitable for different environments due to changing the built environment factors, such as the materials of the building surface. This suggests that real-time on-line training algorithms may be a logical next step in this line of research.

Table 4: The Multipath (noted as 0) and NLOS (noted as -1) classification results for testing dataset *D3* and *D4*

Algorithms for Comparison		**Proposed ANFIS**		**Decision Tree**		**SVM**		**C/No based Classification**	
Class Type		-1	0	-1	0	-1	0	-1	0
D3	Accuracy	73.0%		56.2%		75.4%		64.8%	
	Accuracy for each class	76.6 %	66.4 %	50.8 %	66.2 %	80.6 %	66.2 %	100 %	0.3 %
D4	Accuracy	71.5%		53.2%		71.2%		71.9%	
	Accuracy for each class	64.2 %	91.5 %	39.3 %	91.4 %	65.7 %	86.5 %	96.8 %	3.8 %

4. CONCLUSION AND FUTURE WORK. This paper has presented an ANFIS based multi-variable based GPS measurement classification algorithm for the identification of LOS, multipath and NLOS measurements by considering the known representative variables from GPS raw measurements. The experimental results show that the proposed algorithm delivers an overall detection accuracy of 91.76% using static test based on the testing datasets from the same locations as the training datasets, with 91.1% and 84.1% for the NLOS and multipath classification accuracy respectively. This is significantly higher than the other three approaches (i.e. decision tree, SVM and C/N_0 based classification), whose overall performances range from 67.1% to 80.1%. For the testing results from datasets *D3* and *D4*, however, which are not collected from the same location as the training data *D1*, the changing of the environment degraded the algorithm performance, pointing to the sensitivity of the algorithm to such issues.

In the future, an online data training mechanism will be combined with the ANFIS based decision system to build a real-time applicable system to achieve high classification accuracy. In addition, the ANFIS classifier will be collaborated with consistency-check to exclude the multipath/NLOS measurements from the GPS single point positioning.

ACKNOWLEDGMENTS

The authors are grateful for the sponsorship of the National Natural Science Foundation of China (Grant No.41704022), National Natural Science Foundation of Jiangsu Province (Grant No. BK20170780), Shenzhen Municipal Science and Technology Innovation Committee (Grant No. JCYJ20170818103653507).

REFERENCES

Adjrad, M. and Groves, P. D. (2016). Intelligent Urban Positioning using Shadow Matching and GNSS Ranging Aided by 3D Mapping, *ION GNSS+ 2016*, Portland, Oregon, September 2016, pp. 534-553.

Betaille, D., Peyret, F., Ortiz, M., Miquel, S. and Fontenay, L. (2013). A new modeling based on urban trenches to improve GNSS positioning quality of service in cities. *IEEE Intelligent transportation systems magazine*, **5(3)**, 59-70.

Bhattacharyya, S. and Gebre-Egziabher, D. (2014). Vector loop RAIM in nominal and GNSS-stressed environments. *IEEE Transactions on Aerospace and Electronic Systems*, **50(2)**, 1249-1268.

Blasch, E. P., Salerno, J. J., Tadda, G. P. (2011). Measuring the worthiness of situation assessment. *IEEE Aerospace & Electronics Conference*, 87-94.

Braasch, M. S. (1996) Multipath effects. In: *Global Positioning System: Theory and Applications. 1*: 547-568.

Chiu, S (1994). Fuzzy Model Identification Based on Cluster Estimation. *Journal of Intelligent & Fuzzy Systems*, **2(3)**.

Cox, D. B. (1978). Integration of GPS with Inertial Navigation Systems. *Navigation,* **25(2)**, 236-245.

Dierendonck, V. A., Fenton, P. and Ford, T. (1992). Theory and performance of narrow correlator spacing in a GPS receiver. *Navigation*, **39(3)**, 265-283.

Euler, H. J. and Goad, C.C. (1991). On optimal filtering of GPS dual frequency observations without using orbit information. *Journal of Geodesy*, **65(2)**, 130-143.

Fessler, J. A. and Hero, A. O. (1994). Space-alternating generalized expectation-maximization algorithm. *IEEE Transactions on Signal Processing*, **42(10)**, 2664-2677.

Groves, P. D. (2011). Shadow matching: A new GNSS positioning technique for urban canyons. *The journal of Navigation*, **64(3)**, 417-430.

Groves, P. D., Jiang, Z., Rudi, M.and Strode, P. (2013) A Portfolio Approach to NLOS and Multipath Mitigation in Dense Urban Areas. In: Proceedings of the 26th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2013). (pp. 3231 - 3247). The Institute of Navigation: Manassas, US.

Groves, P. D. and Jiang, Z. (2013). Height Aiding, C/N 0 Weighting and Consistency Checking for GNSS NLOS and Multipath Mitigation in Urban Areas. *Journal of Navigation*, 66(05), 653-669.

Gurtner, W. (1994). Innovation: Rinex--The Receiver Independent Exchange Format. *GPS world*, **5(7)**, 48-53.

Hartinger, H. and Brunner, F. K. (1999). Variances of GPS phase observations: the SIGMA-ε model. *GPS solutions*, **2(4)**, 35-43.

Hsu, L. T., Jan, S. S., Groves, P. D. and Kubo, N. (2015). Multipath mitigation and NLOS detection using vector tracking in urban environments. *GPS Solutions*, **19(2)**, 249-262.

Hsu, L. T., Gu, Y. and Kamijo, S. (2016). 3D building model-based pedestrian positioning method using GPS/GLONASS/QZSS and its reliability calculation. *GPS solutions*, **20(3)**, 413-428.

Hsu, L.T. (2017a). GNSS multipath detection using a machine learning approach. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Yokohama, Japan, 2017, pp. 1-6.

Hsu, L-T, Tokura H, Kubo N, Gu Y, Kamijo S (2017b). Multiple Faulty GNSS Measurement Exclusion based on Consistency Check in Urban Canyons IEEE Sensors Journal, **17(6)**, 1909-1917

Izadpanah, A., O'Driscoll, C. and Lachapelle, G. (2008, September). GPS multipath parameterization using the extended kalman filter and a dual LHCP/RHCP antenna. In *Proceedings of the 21st International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2008), Savannah, GA, USA* (pp. 16-19).

Jang, J. S. (1993). ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, **23(3)**, 665-685.

Jiang, Z. and Groves, P. D. (2014). NLOS GPS signal detection using a dual-polarisation antenna. *GPS solutions*, **18(1)**, 15-26.

Jilani, A., Murawwat, S. and Jilani, S. O. (2015). Controlling Speed of DC Motor with Fuzzy Controller in Comparison with ANFIS Controller. Intelligent Control and Automation, **06(01)**, 64-74.

Kubo, N., Kobayashi, K., Hsu, L. T. and Amai, O. (2017). Multipath Mitigation Technique under Strong Multipath Environment Using Multiple Antennas. *Journal of Aeronautics, Astronautics and Aviation*, **49(1)**, 75-82.

Kuusniemi, H., (2005). *User-Level Reliability and Quality Monitoring in Satellite-Based Personal Navigation*. PhD thesis, Tampere University of Technology, Finland, 2005. Publication 544

Lau, L. and Cross, P. (2007). Development and testing of a new ray-tracing approach to GNSS carrier-phase multipath modelling. *Journal of Geodesy*, **81(11)**, 713.

R. D. J. v. Nee, J. Siereveld, P. C. Fenton and B. R. Townsend (1994). The multipath estimating delay lock loop: approaching theoretical accuracy limits, in *Position Location and Navigation Symposium, IEEE*, (pp. 246-251).

Palamartchouk, K., Clarke, P. J., Edwards, S. J. and Tiwari, R. (2015). Dual-polarization GNSS observations for multipath mitigation and better high precision positioning. In *Proceedings of the 28th International Technical Meeting of the ION Satellite Division, ION GNSS+ 2015, Tampa, Florida* (pp. 2772-2779).

Peyraud, S., Bétaille, D., Renault, S., Ortiz, M., Mougel, F., Meizel, D. and Peyret, F. (2013). About non-line-of-sight satellite detection and exclusion in a 3D map-aided localization algorithm. *Sensors*, **13(1)**, 829-847.

Phan, Q., Tan, S., McLoughlin, I. V., & Vu, D. V. (2013). A Unified Framework for GPS Code and Carrier-Phase Multipath Mitigation Using Support Vector Regression. *Advances in Artificial Neural Systems*, 1-14.

Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences,* **51(2)**, 350-365.

Sahmoudi, M. and Amin, M. G. (2008). Fast iterative maximum-likelihood algorithm (FIMLA) for multipath mitigation in the next generation of GNSS receivers. *IEEE Transactions on Wireless Communications*, *7*(11).

Smith, L. I. (2002). A tutorial on principal components analysis. *Cornell University, USA*, **51(52)**, 65.

Sokhandan, N., Ziedan, N., Broumandan, A., & Lachapelle, G. (2017). Context-aware adaptive multipath compensation based on channel pattern recognition for gnss receivers. *Journal of Navigation*, 1-19.

Suzuki, T. (2016). Integration of GNSS Positioning and 3D Map using Particle Filter, *Proceedings of the 29th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2016)*, Portland, Oregon, September 2016, pp. 1296-1304.

Takagi, T. and Sugeno, M (1983). Derivation of fuzzy control rules from human operator's control actions. *Proc. Of the IFAC Symp. on Fuzzy Information, Knowledge Representation and Decision Analysis*, p. 55–60.

Tranquilla, J. M., Carr, J. P. and Al-Rizzo, H. M. (1994). Analysis of a choke ring groundplane for multipath control in global positioning system (GPS) applications. *IEEE Transactions on antennas and propagation*, **42(7)**, 905-911.

Ubeyli, E. D. (2009). Adaptive Neuro-Fuzzy Inference Systems for Automatic Detection of Breast Cancer. *Journal of Medical Systems*, **33(5)**, 353-358.

Wang, H. S., Kao, C. Y. and Chen, J. F. (2013a). Sequential quadratic method for GPS NLOS positioning in urban canyon envi-ronments. *International Journal of Automation & Smart Technology, 3*(1), 37-46.

Wang, L., Groves, P. D. and Ziebart, M. K. (2013b). Urban positioning on a smartphone: Real-time shadow matching using GNSS and 3D city models. The Institute of Navigation.

Wang, L., Groves, P. D. and Ziebart, M. K. (2015). Smartphone shadow matching for better cross-street GNSS positioning in urban environments. *The Journal of Navigation*, **68(3)**, 411-433.

Wei, L. (2016). A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting, Applied Soft Computing, **42**, 368-376.

Yozevitch, R., Moshe, B. B. and Weissman, A. (2016). A robust GNSS los/nlos signal classifier. *Navigation*, **63(4)**, 429-442.

Ziedan, N. I. (2012). Multipath Channel Estimation and Pattern Recognition for Environment-Based Adaptive Tracking Conference: ION GNSS 2012, 25th International Technical Meeting of the Satellite Division of the Institute of Navigation 2012, At Nashville, Tennessee

Ziedan, N. I. (2017). Urban Positioning Accuracy Enhancement Utilizing 3D Buildings Model and Accelerated Ray Tracing Algorithm. *Proceedings of the 30th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2017)*, Portland, Oregon, September 2017, pp. 3253-3268.