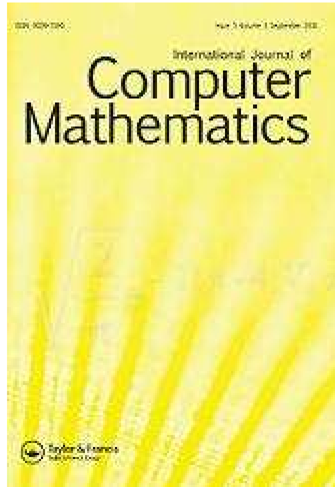
Comparison of GPS observations made in a forestry setting using functional data analysis - CMMSE 2010

Celestino Ordóñez Galán, Javier Martinez Torres, Francisco Javier de Cos, Fernando Sánchez Lasheras



Comparison of GPS observations made in a forestry setting using functional data analysis - CMMSE 2010

Journal:	*International Journal of Computer Mathematics*
Manuscript ID:	GCOM-2010-0661-B.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	10-Jan-2011
Complete List of Authors:	Ordóñez Galán, Celestino; VIGO UNIVERSITY Martinez Torres, Javier; University of Vigo, Natural Resources Engineering de Cos, Francisco; Universidad de Oviedo, Departamento de Prospección y Explotación Minera Lasheras, Fernando; Tecniproject
Keywords:	GPS, forest canopy, functional data, outlier, depth

Comparison of GPS observations made in a forestry setting using functional data analysis

C. Ordoñez[a,*], J. Martínez[b], J.F. de Cos Juez[c], F. Sánchez Lasheras[d]

[a]Department of Natural Resources and Environmental Engineering, University of Vigo, Vigo 36310, Spain.

[b]Centro Universitario de la Defensa. Academia General Militar, 50090 Zaragoza, Spain.

[c]Department of Mining Exploitation and Prospection. University of Oviedo. E.T.S.I. MINAS, 33007 Oviedo, Spain.

[d]Research Department, Tecniproject Ltd., 33004 Oviedo, Spain.

[*]Corresponding author: Celestino Ordóñez Galán. E.T.S.I.MINAS. Departamento de Ingeniería de los Recursos Naturales y Medio Ambiente. Universidad de Vigo. Rúa Maxwell s/n. 36310 Vigo (Spain). E-mail: cgalan@uvigo.es. Tel.: +34986814052. Fax: +34986812201.

Abstract

GPS receiver observations made in forestry settings are affected by data distortion and signal losses and this negatively affects precision and accuracy measurements. Using a technique for identifying functional outliers, we determine whether there are differences between errors for coordinates obtained at 10 different points of a forest characterized by a set of dasometric data. Our results indicate that the 2 points with highest error correspond to areas with dasometric values that would indicate these areas to have a more dense forest canopy than the remaining areas.

Keywords: GPS, forest canopy, functional data, outlier, depth.

1. Introduction

Geographic positioning system (GPS) receivers are frequently used in forestry settings for, among other applications [7], the monitoring of harvesting machinery [10] and cadastral forest surveys [3],[20]. In forestry settings, however, measurement precision and accuracy are affected by forest canopy interference in the satellite signals. Tree trunks, branches and leaves distort or break up satellite signals and negatively affect the accuracy and precision of receiver-measured coordinates in these conditions [6], [17]. For a forest of lodgepole pines (*Pinus contorta* Douglas, Gerlach (1989)) found that radio signal losses from a satellite could be attributed to tree trunks, branches and foliage 23%, 28% and 36% of the time, respectively.

The typical approach to studying the effect of the forest canopy on GPS positional accuracy is to seek associations between accuracy and dasometric variables characterizing the forest canopy, such as basal area, stand density and the Hart-Becking index [12], [13], [17], [19]. In these cases, tests for comparing statistical distributions (parametric or non-parametric) are generally used to determine whether or not precision values obtained in zones with different forestry variable values can be considered to come from the same population. Analysis of variance (ANOVA) is also frequently used to identify the variables that significantly affect GPS positional accuracy [1]. However, this kind of test is not appropriate when working with a dense set of data collected over time (such as GPS data), as such data is more suitably handled as observations made at discrete points of a smooth stochastic process.

Data mining techniques have advanced to the point where the exploitation of vectorial data has proven to be inadequate; this has led to the emergence of functional data analysis (FDA) [18]. FDA applications are very varied and include environmental research, [11], [16], sensors [22], industrial methods [8], [14] and medical research [2], [21].

In the functional approach, comparisons are made overall and take into account the time correlation structure of the data. This is the focus we give to our forest canopy problem. The method used to compare curves is based on the concept of functional depth, which is a measure of the centrality of a given curve within a group of curves [5].

The functional depth has been used by authors in a different problem in the environmental area, which was to compare the pollutant measurements made by different companies [9].

The article is structured on the basis of a description of our methodology, a description of our results and the most relevant conclusions to be drawn from the results.

2. Methodology

We identified outliers using a functional approach, in such a way that the sample of observations was considered to be composed of a series of curves rather than a discrete set of point observations. First the curves were fitted to the discrete data by means of a process called smoothing and then outliers were identified using the concept of functional depth. In this section we explain the basic concepts underlying the methodology for detecting outliers in the initial sample, namely, smoothing, functional depth and outlier identification.

2.1. Smoothing

One of the first studies in the FDA fields [18] considered functional data to be observations at discrete points of continuous random processes. Assume a set of observations $f\left(t_j\right)$ in a set of n_p points $t_j \in \mathbb{R}$, where t_j represents each instant of time.

These observations can be considered as discrete observations of the function, where is a functional space.

In order to estimate the function $f(t)$, it is considered that $F = span\left\{\phi_{1,\dots,}\phi_{n_b}\right\}$, where $\{\phi_k\}\, k = 1,\dots,n_b$ is a set of basis functions. In other words, for each function $f(t) \in \chi \subset F$, we have:

$$f(t) = \sum_{k=1}^{n_b} c_k \phi_k(t)$$

where $c_k \in \mathbb{R}; k = 1,\dots,n_b$ represent the coefficients of the function $f(t)$ respect to the chosen set of basic functions $\{\phi_k\}$ chosen.

The smoothing problem now consists of determining the solution to the following regularization problem [18]:

$$\min_{f \in F} \sum_{j=1}^{n_p} \left\{ z_j - f(t_j) \right\}^2 + \lambda \Gamma(f)$$

where $z_j = f(t_j) + \varepsilon_j$, ε_j is random noise with zero mean) represents each of the observations of the function f in the instant t_j, Γ is a differential operator that controls the complexity of the function and λ is a regularization parameter. In our case, we have used the operator $\Gamma(x) = \int_T \left[D^2 x(t) \right]^2 dt$ where $T = [t_{\min}, t_{\max}]$ and D^2 is the second-order differential operator.

Bearing in mind (1), the problem (2) may be written as:

$$\min_{\mathbf{c}} \left\{ (\mathbf{z} - \Phi\mathbf{c})^T (\mathbf{z} - \Phi\mathbf{c}) + \lambda \mathbf{c}^T \mathbf{R} \mathbf{c} \right\}$$

where $\mathbf{z} = (z_1, \ldots, z_{n_p})^T$ is the vector of observations, $\mathbf{c} = (c_1, \ldots, c_{n_b})^T$ is the vector of coefficients expressed in (1), $\mathbf{\Phi}$ is the $n_p \times n_b$ matrix with elements $\mathbf{\Phi}_{jk} = \phi_k(t_j)$ and $\mathbf{R}$ is the $n_b \times n_b$ matrix with elements:

$$R_{kl} = \left\langle D^2\phi_k, D^2\phi_l \right\rangle_{L_2(\mathbf{T})} = \int_{\mathbf{T}} D^2\phi_k(t) D^2\phi_l(t)dt$$

2.2 Functional depth

Depth measurement was originally introduced in the context of multivariate analysis to measure the centrality of a point with respect to a sample. Depth provides a way of ordering a sample from its centre in such a way that the points closest to the centre have greater depth. Like most vectorial concepts, the concept of depth has been generalized to the functional case [5]. Functional depth measures the centrality of a curve within a set of curves.

In this study we focus on 2 of the most widely used depth measurements:

- Fraiman-Muniz depth (FMD): Let the empirical distribution function $F_{n,t}(f_i(t))$ be [5] for the functional sample $\{f_i(t)\}_{i=1}^n$, $t \in [a,b]$ be:

  $$F_{n,t}(f_i(t)) = \frac{1}{n}\sum_{k=1}^{n} I(f_k(t) \le f_i(t))$$

  where $I(\cdot)$ is the indicator function. The FMD for a curve f_i is given by:

  $$FMD_n(f_i(t)) = \int_a^b D_n(f_i(t))dt$$

  where $D_n(f_i(t))$ is the point depth of $f_i(t), \forall t \in [a,b]$ given by:

  $$D_n(f_i(t)) = 1 - \left|\frac{1}{2} - F_{n,t}(f_i(t))\right|$$

- H-modal depth (HMD): The functional mode (based on the mode concept) is defined as the curve most densely surrounded by the other curves in a sample. HMD is expressed as:

$$MD_n(f_i,h)=\sum_{k=1}^{n}K\left(\frac{\|f_i-f_k\|}{h}\right)$$

where $K:\mathrm{R}^+\to\mathrm{R}^+$ is a kernel function, $\|\cdot\|$ is a norm in a functional space and h is the bandwidth parameter. One of the most widely used norms for a functional space is L^2, expressed as:

$$\|f_i(t)-f_j(t)\|_2=\left(\int_a^b(f_i(t)-f_j(t))^2\,dt\right)^{1/2}$$

The infinite norm L^∞ is sometimes used:

$$\|f_i(t)-f_j(t)\|_\infty=\sup_{t\in(a,b)}|f_i(t)-f_j(t)|$$

Different kernel functions $K(\cdot)$ can also be defined, among them the truncated Gaussian kernel:

$$K(t)=\frac{2}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right),t>0$$

2.3 The functional outlier concept

A functional sample may include elements that, although they do not constitute error in themselves, may feature patterns different from the rest of the sample. Depth measurement, as described above, is used to identify outliers in functional samples. Depth and outlier are inverse concepts: outliers in functional samples have considerably less depth than non-outliers.

For this study we used the HMD to measure depth. The cutoff C was obtained so that type 1 error—the percentage of correct observations wrongly identified as outliers—was approximately 1%:

$$\Pr(MD_n(x_i(t)) \le C) = 0{,}01, \quad i = 1, \ldots, n$$

In practice, the distribution of the functional depths for which the value for C needs to be estimated is unknown. Of the different estimation methods available [4], we selected the bootstrapping approach as the method used by many authors to solve this problem, the ease to implement the algorithm and to present good results in several comparative studies [15].

3. Application. Forest canopy impact on GPS accuracy

3.1 GPS measurements

The sample used in this research $\left\{(t_1, t_2, \ldots, t_{3600})_j\right\}_{j=1}^{10}$ consists of a set of GPS observations measured, for a set of 10 points, second by second over 1 hour, where t_{ij} represents measurement in instant *i* (in seconds) over the period of 1 hour at point *j*.

The data were collected using a double-frequency GPS receiver (HiperPlus, Topcon Positioning Systems, Inc., Livermore, CA, USA) while observing GPS pseudorange and carrier phase.

The GPS experimental data were collected during 2 days over periods of 5-6 hours between 18 and 21 July 2008. Antenna heights ranged from 1.35 to 1.70 m and the logging rate was 1 second. The collection of observations lasted for at least 1.5 hours and the process was repeated 3 times a day. GPS data were revised to ensure continuity and were cut to obtain 10 datasets representing 1 hour.

The observation point was located at latitude 42°41'08.79872"N and longitude 6°38'03.210587" W (WGS84) and at an ellipsoidal height of 933.829 metres (considered the Z_{true} coordinate). These coordinates were calculated by differential correction in static surveying. Post-processing correction was carried out using the PONF base station as the nearest reference station in the regional GNSS network (http://gnss.itacyl.es/). This point was projected and the position set up as the 'true' position for calculating horizontal and vertical accuracy. The UTM coordinates were X_{true}=693814.623 and Y_{true}=4728635.531 (Datum ETRS89; zone 29N).

The planimetric error in each instant of time *i* was calculated using the expression:

$$E_{XY} = \sqrt{\left(X_i - X_{true}\right)^2 + \left(Y_i - Y_{true}\right)^2}$$

The altimetric error in each instant was calculated using the formula:

$$E_Z = \left|Z_i - Z_{true}\right|$$

3.2 Forest environment characterization

With a view to characterizing the forest lots studied, we calculated the parameters associated with the forestry characteristics of each. The parameters were determined by measuring the trees in a radius of 10 metres around the point where the GPS observations were made. The dominant specie in all the lots was *Pinus radiata.*

The parameters studied were the following:

- DBH: Diameter at Breast Height (measured at a height of 1.37 metres).
- H_m: Mean height.
- H_0: Dominant height (mean height of the 4 thickest trees).
- H_{tt}: Treetop height (total height less the height to the first branch).
- N: Number of feet/hectares.
- G: Basal area (cross-section at normal tree height).
- D_g: Mean squared diameter ($D_g = \sqrt{\frac{4G}{\pi n}}$; n = number of trees).
- HBI: Hart-Becking index (relationship between mean spacing between trees *a* and dominant height H_0 ($IH = \frac{a}{H_0} 100$)

- V: Wood volume.
- W: Biomass.
- SC: slenderness coefficient (relationship between mean height and mean diameter of the biomass)

Table 1 shows the parameter values for the 10 studied plots and also minimum, maximum and mean values and standard deviation for the entire set.

4. Results

The first step in identifying possible outliers in the data was to fit curves to the set of values for the planimetric and altimetric errors. The smoothing method described in Section 2.1 was used for this purpose using a set of Fourier basic functions because of their orthonormal character, resulting in a set of 10 curves for each error type. Figure 1 shows the 10 planimetric error curves. The great functional complexity of the sample is evident in the irregularity of the functions.

Figure 1. Functional sample for planimetric errors, with 2 outliers depicted in black.

Given the complexity of the sample, it was necessary to select a set of 1000 basis functions in order to obtain the most information possible. The outcome was a fit defined by an R-squared value (RSQ) of 0.99.

In order to study the functional outliers the HMD function was selected considering the norm L^2 of a functional space. The analysis identified 2 functional outliers (depicted in black in Figure 1) in the sample for the specific case of the planimetric errors; these outliers corresponded to Lots 5 and 8 in Table 1. No functional outliers for the error in Z were detected. As it can be seen in Figure 1, functions detected as outliers have large values for planimetric error and it is easy to distinguish some sharp points.

This result differs from that obtained using the Kruskal-Wallis test for all the positions except numbers 5 and 8 (the fact that the errors do not follow a normal distribution was first checked and was the reason a non-parametric test was used). This test rejected the null hypothesis, for a 99% significance level, that the 8 error observations (all except numbers 5 and 8) came from the same population. Recall that, since the Kruskal-Wallis test compares the medians of the groups, it is possible to have a tiny *p* value—clear evidence that the population medians are different— even if the distributions overlap considerably.

The 2 points corresponding to the outliers are characterized by high values for tree density (N), volume (V) and biomass (W), low values for the Hart-Becking index and the largest basal areas. Our result corroborates the results obtained by other authors, such as Naesset (1999, 2001)—who found that the basal area is a parameter that has a significant bearing on GPS measurement accuracy—and Rodríguez-Pérez et al. (2007)—who detected an association between the Hart-Becking index and precision.

5. Conclusions

We have analysed the application potential of functional data analysis to the evaluation of the impact of the forest canopy on the accuracy of GPS receiver measurements. Adopting a functional approach means that error measurements over the period of an hour can be considered as a continuous function. In other words, we can work with all the information rather than just with mean values, which is the basis for the traditional statistical approach.

Outliers were detected using functional depth measurement (a generalization of the vectorial case), which provides information on the distance of a function from the centre of a sample. In contrast with conventional vectorial techniques, this methodology does not require a normality hypothesis for the sample (whether this hypothesis would be valid was, nonetheless, checked) and also takes into account the time correlation structure of the data.

The 2 outliers detected in our study can be explained by the fact that they correspond to points located in lots with the highest basal areas and also high tree density, volume and biomass values and a low Hart-Becking index.

References

[1] C.H. Deckert and P. Bosltadr, *Forest Canopy, Terrain and Distance Effects on Global Positioning System Point Accuracy*. Photogrammetric Engineering & Remote Sensing, **62(3)** (1996), 317-321.

[2] D.A. Dombeck, M.S. Graziano and D.W. Tank, *Functional clustering of neurons in motor cortex determined by cellular resolution imaging in awake behaving mice*, Journal of Neuroscience, **29(44)** (2009) 13751-13760.

[3] D. Evans, R. Carraway and G. Simmons, *Use of global positioning system (GPS) for forest plot location*. Southern J. Applied Forestry, **16(2)** (1992) 67–70.

[4] M. Febrero, P. Galeano and W. González-Manteiga, *Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels*. Environmetrics (2008) 19:331-345. doi: 10.1002/env.878.

[5] R. Fraiman and G. Muniz, *Trimmed means for functional data*. Test **10** (2001) 419-440.

[6] H. Hasegawa and T. Yoshimura, *Application of dual-frequency GPS receivers for static surveying under tree canopies*. J. For. Res. **8** (2003) 103-110.

[7] L.R. Kruczynski and A. Jasumback, *Forestry Management Applications: Forest Service Experiences with GPS*, Journal of Forestry, **91(8)** (1993) 20-24.

[8] M. López, J.M. Matías, J.A. Vilán and J. Taboada, *Functional Pattern Recognition of 3D Laser Scanned Images of Wood-Pulp Chips*, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **4477(1)** (2007) 298-305.

[9] J. Martínez, P.J. García, L. Alejano and A.N. Reyes, *Detection of outliers in gas emissions from urban areas using functional data analysis*, Journal of Hazardous Materials, **186 (1)** (2011), 144-149.

[10] T.P. McDonald, E.A. Carter and S.E. Taylor, *Using the global positioning system to map disturbance patterns of forest harvesting machinery*. Canadian J. Forest Research, **32(2)** (2002) 310–319.

[11] J.M. Matías, C. Ordóñez, J. Taboada and T. Rivas, *Functional support vector machines and generalized linear models for glacier geomorphology analysis*, International Journal of Computer Mathematics, **86(2)** (2009) 275-285.

[12] E. Næsset, *Point accuracy of combined pseudorange and carrier phase differential GPS under forest canopy*. Canadian J. Forest Research, **29(5)** (1999) 547–553.

[13] E. Næsset, and T. Jonmeister, *Assessing point accuracy of DGPS under forest canopy before data acquisition, in the field and after postprocessing*. Scandinavian J. Forest Research, **17(4)** (2002) 351–358.

[14] J.I. Park, S.H. Baek, M.K. Jeong and S.J. Bae, *Dual features functional support vector machines for fault detection of rechargeable batteries*, IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, **39(4)** (2009) 480-485.

[15] L. Peng and Y. Qi, *Bootstrap approximation of tail dependence function*, Journal of Multivariate Analysis **99 (8)** (2008) :1807-1824. doi: 10.1016/j.jmva.2008.01.018.

[16] J. M. Paruelo and F. Tomasel, *Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models*, Ecological Modelling, **98** (1997) 173-186.

[17] J.R. Rodríguez-Pérez, M.F. Álvarez and E. Sanz-Ablanedo, *Assessment of low-cost GPS receiver accuracy and precision in forest environments*. J. Surv. Eng. **133** (2007) 159-167.

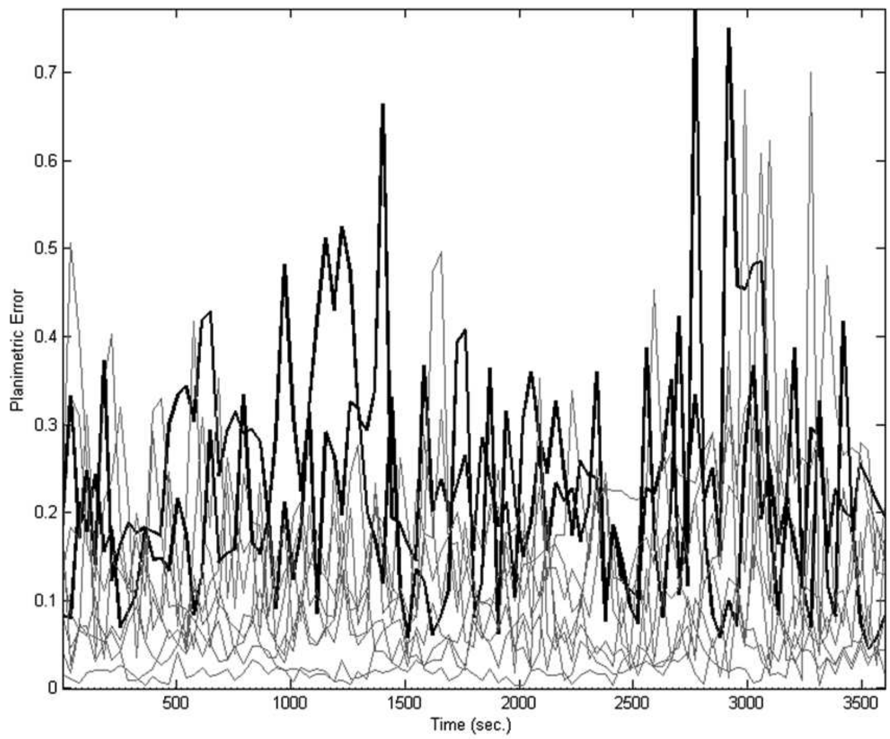
[18] J.O. Ramsay, and B.W. Silverman, *Functional Data Analysis*. New Yor, Springer, 1997.

[19] I. Sawaguchi, K. Nishida, M. Shishiuchi and S. Tatsukawa, *Positioning precision and sampling number of DGPS under forest canopies*. J. Forest Research, **8(2)** (2003) 133–137.

[20] T. Soler, G. Álvarez-García, A. Hernández-Navarro and R.H. Foote, *GPS high accuracy geodetic networks in Mexico*. J. Surv. Eng., **122(2)** (1996), 80–94.

[21] R. Viviani, G. Grön and M. Spitzer, *Functional principal component analysis of FMRI data*, Human Brain Mapping, **24(2)** (2005) 109-129.

[22] D. Wu, S. Huang and J. Xin, *Dynamic compensation for an infrared thermometer sensor using least-squares support vector regression (LSSVR) based functional link artificial neural networks (FLANN)*, Measurement Science and Technology, **19(10)** (2008) n°105202.



254x190mm (96 x 96 DPI)

Table 1. Forestry paramaters for 10 forest lots

Lot	DBH (cm)	H_m (m)	H_0 (m)	H_{tt} (m)	N (ft/ha)	G (m^2/ha)	D_g (cm)	HBI (%)	V (m^3/ha)	W (kg/ha)	SC
1	20.11	17.02	17.85	11.72	732.48	24.47	20.41	20.69	182.89	83696.33	82.42
2	20.28	15.34	16.40	10.36	764.13	25.61	20.31	22.06	191.20	87207.64	74.29
3	28.38	25.32	26.73	17.46	605.82	38.53	29.13	15.20	385.41	179483.25	88.50
4	29.68	27.69	28.01	19.21	668.30	46.84	29.79	13.81	474.05	220912.96	94.46
5	28.66	25.73	26.07	17.30	2509.59	54.38	28.17	7.66	415.45	146441.65	192.57
6	14.92	16.01	15.02	9.85	2037.15	37.20	15.92	14.75	237.06	114335.66	111.71
7	16.41	19.19	19.11	10.56	1751.19	40.50	17.22	12.51	279.05	131151.88	115.13
8	14.42	22.63	22.61	12.43	3056.36	60.18	16.08	8.00	442.99	216048.37	156.75
9	13.46	21.41	22.23	12.48	2960.08	53.96	16.01	8.27	380.39	188170.60	153.83
10	12.73	23.26	22.85	13.06	2992.34	46.35	14.23	8.00	312.88	157547.27	182.95
Min	12.73	15.34	15.02	9.85	605.82	24.47	14.23	7.66	182.89	83696.33	74.29
Max	29.68	27.69	28.01	19.21	3056.36	60.18	29.79	22.06	474.05	220912.96	192.57
Mean	20.12	21.38	21.66	13.62	1811.63	42.72	20.94	13.39	329.86	152467.07	126.62
STD	6.70	4.33	4.48	3.33	1044.93	11.91	6.05	5.28	104.81	49092.53	43.06