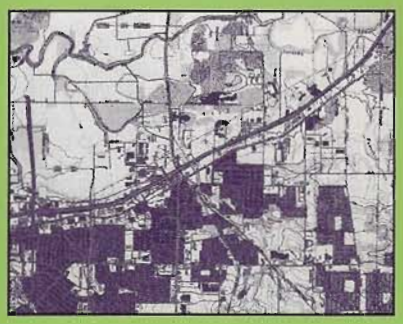
ASCE
AMERICAN SOCIETY OF CIVIL ENGINEERS

November 2011 • Volume 137, Number 4
ISSN 0733-9453 • CODEN: JSUED2



Journal of Surveying Engineering

Geomatics Division

Machine Learning Techniques Applied to the Assessment of GPS Accuracy under the Forest Canopy

Celestino Ordóñez[1]; José R. Rodríguez-Pérez[2]; Juan J. Moreira[3]; J. M. Matías[4]; and Enoc Sanz-Ablanedo[5]

Abstract: The geographic location of points using global positioning system (GPS) receivers is less accurate in forested environments than in open spaces because of signal loss and the multipath effect of tree trunks, branches, and leaves. This has been confirmed in studies that have concluded that a relationship exists between measurement accuracy and certain variables that characterize forest canopy, such as tree density, basal area, and biomass volume. However, the practical usefulness of many of these studies is limited because they are often limited to describing associations between the variables and mean errors in the measurement interval, when measurements should be made in real time and in intervals of seconds. In this work, machine learning techniques were applied to build mathematical models that would associate observation error and GPS signal and forest canopy variables. The results reveal that the excessive complexity of the signal prevents accurate measurement of observation error, especially in the *Z*-coordinate and in time intervals of a few seconds, but also reveal that forest cover variables have a significantly greater influence than GPS factors, such as position dilution of precision (PDOP), the number of satellites, or error for the base station. **DOI: 10.1061/(ASCE)SU.1943-5428.0000049.** *© 2011 American Society of Civil Engineers.*

CE Database subject headings: Global positioning systems; Neural networks; Satellites; Surveys; Forests.

Author keywords: Global positioning; Neural networks; Satellite; Surveys; Forests.

Introduction

Global navigation satellite system (GNSS) and particularly global positioning system (GPS) technologies are widely used in land surveying for engineering, mapmaking, and cadastral purposes. They are also used in forestry applications. GNSS receivers located in open spaces and operating in a differential positioning mode can measure the coordinates to a point with millimeter accuracy. Under a forest canopy, however, attenuation (Yoshimura and Hasegawa 2003) and temporary loss of lock of the GPS signal (Hasegawa and Yoshimura 2007) affect GPS positioning precision and accuracy. Nevertheless, it is possible to achieve an accuracy of better than 0.1 m under the forest canopy by using real-time kinematic (RTK) processing (Bakula et al. 2009), although the equipment needs to be reinitiated frequently to achieve redundancy in the observations.

[1]Associate Professor, Dept. of Environmental Engineering, Univ. of Vigo, Rúa Maxwell s/n, Campus Lagoas-Marcosende, 36310-Vigo, Pontevedra, Spain. E-mail: cgalan@uvigo.es

[2]Associate Professor, Geomatics Engineering Research Group, Univ. of León, Avda. Astorga s/n, 24400 Ponferrada, León, Spain (corresponding author). E-mail: jr.rodriguez@unileon.es

[3]Ph.D. Student, Dept. of Environmental Engineering, Univ. of Vigo, Rúa Maxwell s/n, Campus Lagoas-Marcosende, 36310-Vigo, Pontevedra, Spain. E-mail: jmoreira@uvigo.es

[4]Associate Professor, Dept. of Statistics, Univ. of Vigo, Campus Universitario As Lagoas-Marcosende s/n, 36310 Vigo, Pontevedra, Spain. E-mail: jmmatias@uvigo.es

[5]Assistant Professor, Geomatics Engineering Research Group, Univ. of León, Avda. Astorga s/n, 24400 Ponferrada, León, Spain. E-mail: esana@unileon.es

Note. This manuscript was submitted on June 22, 2010; approved on November 28, 2010; published online on December 9, 2010. Discussion period open until April 1, 2012; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Surveying Engineering***, Vol. 137, No. 4, November 1, 2011. ©ASCE, ISSN 0733-9453/2011/4-140–149/$25.00.

There exist many GNSS applications in forestry: GPS receivers are used to navigate forested areas (Mancebo and Chamberlain 2000), make maps to different scales, take forest inventory (Murphy et al. 2006), make area and perimeter estimates (Tachiki et al. 2005), manage forestry machinery (McDonald and Carter 2002; McDonald and Fulton 2005), characterize forest roads (Martin et al. 2001), and more. GPS is the main tool used for precision forestry, i.e., planning and implementing site-specific forest management activities and operations aimed at improving wood product quality and use, reducing waste, increasing profits, and maintaining the quality of the environment (Taylor et al. 2006). Precision and accuracy requirements depend on the forestry application; for example, precision in meters is sufficient to delimit large forests, whereas precision in decimeters and even centimeters is required for individual trees.

The receiver, thus, needs to be selected in accordance with the type of study performed. Code-phase differential GPS (DGPS) receivers (errors of 2–5 m) are widely used for most forestry applications, whereas carrier-phase DGPSs are required for precision forestry applications.

To assess GPS observation quality according to tree mass type, the factors that condition positioning precision and accuracy under a forest canopy need to be identified. Research in this field has focused essentially on identifying dasometric variables (tree diameter and height, treetop height, plantation density, and wood volume) and on comparative studies of measurement equipment and methods. Sawaguchi et al. (2003) demonstrated that forest type and antenna height condition positioning precision in differential mode. Basal area is one of the dasometric variables with the greatest bearing on precision for both pseudorange and carrier-phase positioning (Naesset 2001; Naesset and Jonmeister 2002). Yoshimura and Hasegawa (2003) concluded that the plantation structure also conditions positioning accuracy and precision. Other studies have focused on modeling error by using logistic regression to evaluate the probability of resolving ambiguity (Naesset et al. 2000; Hasegawa and Yoshimura 2003). Monte Carlo simulation has also

been used to estimate the relationship between satellite signal and forest structure (Sawaguchi et al. 2005). Other researchers have focused on evaluating accuracy and precision measurements obtained by using different types of GNSS receivers (Mancebo and Chamberlain 2000; Rodríguez-Pérez et al. 2007), comparing different kinds of positioning (Naesset and Joinmeister 2002; Gegout and Piedallu 2005), and comparing GPS and the GNSS (Naesset et al. 2000, 2001). Experiments indicate that the most accurate positioning is obtained in differential mode and by using dual-frequency receivers, although, there exist problems with the latter in resolving ambiguities because of cycle losses in the satellite signal (Hasegawa and Yoshimura 2007).

The results of these assessments cannot be compared because of the lack of standards, best practices, and systematic methods to assess GNSS measurements under forest canopy. To compare similar work, defining standards will have to address the assessment of the whole receiver, including traceability, methodology, terminology, acquisition, preprocessing (Beraldin et al. 2007), and the equations to calculate accuracy and precision.

The aim of our research was to study the usefulness of machine learning techniques for estimating GPS observation errors in forestry environments and to evaluate the effects of forest cover on GPS positioning accuracy and precision.

Materials and Methods

Data Collection and Calculation of the Errors

The experimental points were located in Sancedo (a province of León) in the northwest of Spain (Fig. 1). Observations were made at 12 different points located in areas planted with *Pinus radiata* D. Don. Measurements were also made at a reference point located in a clearing so that the coordinates for the 12 experimental points could be calculated.

The data were collected by using two dual-frequency GPS receivers (Hiper Plus, Topcon Positioning Systems, Inc., Livermore, CA) that observed GPS pseudorange and carrier phase. The data for the 12 experimental points were collected over a 4-day period (August 20–23, 2007). Measurements were made at each experimental point for at least 1.5 h, and the data were reviewed to ensure continuity. The data were reduced to obtain 12 data sets that represented data collection over a period of 1 h. Antenna heights ranged from 1.45 to 1.60 m, and the logging rate was 1 s.

The position of the reference point was obtained at the same time as the position of the experimental point for each of the 12 experimental points. The coordinates for the reference point were 42°41′08.79872″N, 6°38′03.210587″W (latitude-longitude, WGS84), and ellipsoidal height was 933.829 m. These coordinates were calculated by differential correction by using data from the Ponferrada base station (the nearest reference station in the regional GNSS network: http://gnss.itacyl.es/). The reference point was projected by setting up the following (easting, northing) coordinates: (693,814.623, 4,728,635.531) m (Universal Transverse Mercator (UTM)—datum ETRS89; zone 29N). The reference point coordinates were used to calculate the coordinates assumed as true of the experimental points by differential correction (the coordinates are shown in Table 1).

Next, the accuracies obtained were determined for every second observation in each plot. The Joint Committee for Guides in Metrology (JCGM) defined accuracy as the closeness of agreement between a measured quantity value and a true quantity value of a measurement (JCGM 2008). In this work, horizontal accuracy and vertical accuracy were calculated for each sample by using the following equations:

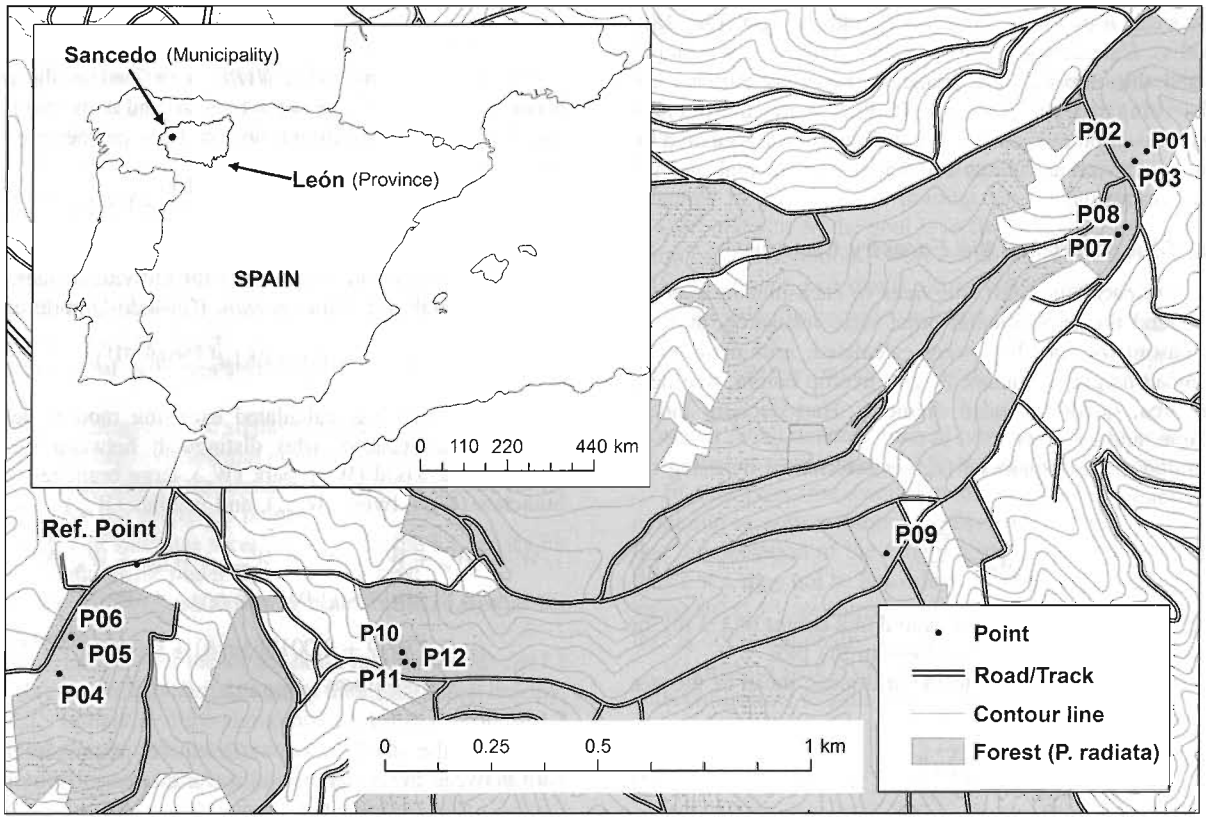


Fig. 1. Location of the experimental points and the reference point in Sancedo (León, Spain)

Table 1. Accurate Coordinates, Trueness, and Precision for the Experimental Points and the Reference Point (ETRS89, UTM29N)

Point	Easting (X)	Northing (Y)	Ellipsoidal height	Horizontal trueness	Vertical trueness	Horizontal precision	Vertical precision
P01	696,104.03	4,729,765.54	1,013.13	1.83	6.80	0.77	2.40
P02	696,058.33	4,729,778.21	1,013.72	4.21	3.68	1.79	2.29
P03	696,078.14	4,729,740.30	1,013.97	2.39	4.91	1.14	3.50
P04	693,652.72	4,728,368.28	913.64	1.74	4.21	0.91	2.60
P05	693,696.85	4,728,436.42	918.90	3.87	4.18	1.58	2.36
P06	693,673.89	4,728,454.87	921.66	2.12	7.13	0.99	3.64
P07	696,068.87	4,729,584.83	1,007.37	4.81	5.21	2.50	3.20
P08	696,052.45	4,729,566.48	1,008.89	3.78	4.42	1.95	3.33
P09	695,566.00	4,728,784.62	989.02	2.07	8.30	1.06	4.40
P10	694,450.03	4,728,474.45	954.21	2.85	5.49	1.45	2.76
P11	694,457.89	4,728,452.08	955.89	2.01	10.02	1.20	5.83
P12	694,478.74	4,728,446.77	954.93	3.12	5.71	1.52	3.04
RP	693,814.62	4,728,635.53	933.83	1.32	2.33	0.54	1.42

$$H_{\text{acc}} = \sqrt{(E_i - E_{\text{true}})^2 + (N_i - N_{\text{true}})^2} \quad (1)$$

$$V_{\text{acc}} = |Z_i - Z_{\text{true}}| \quad (2)$$

in which H_{acc} and V_{acc} indicate horizontal and vertical accuracy, respectively; E_i, N_i, and Z_i = positions measured at the ith second; and E_{true}, N_{true}, and Z_{true} = coordinates assumed as true in the easting, northing, and ellipsoidal height directions, respectively.

The measurement trueness is defined as the closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value (JCGM 2008), whereas the precision is the closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions (JCGM 2008). Precision calculations were made for different time intervals n measured in seconds (with $n = 60$, 120, 180, 240, or 300 s). Calculating precision values simplifies the output function, and this, theoretically, makes it easier to estimate the function by using machine learning. Table 1 shows the trueness (calculated by using mean values) and the precision (calculated by using standard deviation values) for both the horizontal and the vertical measurements.

Variables Characterizing the Forestry Canopy

To characterize each tree mass, the trees in each plot in a radius of 10 m around the experimental point were measured, and the following dasometric variables were calculated: arithmetic mean diameter, mean height, dominant height, treetop height, tree density, basal area, quadratic mean diameter, Hart-Becking index, wood volume, and biomass.

The arithmetic mean diameter (d_m) is determined by using the following equation:

$$d_m = \frac{\Sigma d_i}{n} \quad (3)$$

in which d_i = normal diameter (measured at a height of 1.3 m) for each tree; and n = total number of trees in the plot.

Mean height h_a corresponds to the arithmetic mean of the tree heights

$$h_a = \frac{\sum h_i}{n} \quad (4)$$

in which h_i = tree height; and n = number of trees measured in each plot.

Dominant height (H_0) was calculated as the mean height of the four largest trees in each plot.

Treetop height (h_c) was determined by calculating the difference between total height and height to the first branch.

Tree density (N) was calculated as the number of trees per hectare.

Basal area (G) reflects the cross-section area at the normal height of the trees in the plot

$$G = \sum \frac{\pi}{4} d_i \quad (5)$$

The quadratic mean diameter (d_g) measures tree diameter compared to mean basal area

$$d_g = \sqrt{\frac{4G}{\pi n}} \quad (6)$$

The Hart-Becking index (HBI) is defined as the relationship between mean spacing between trees (a) and dominant height (H_0), and HBI is related to the number of trees per hectare (N)

$$HBI = \frac{a}{H_0} 100 = \frac{100}{H_0 \sqrt{N}} 100 \downarrow\downarrow \quad (7)$$

To calculate wood volume (V) for individual trees, we used a specific formula for *Pinus radiata* (Castedo-Dorado et al. 2009)

$$V = (0.000081 d_i^{1.7746} h^{0.9711}) N \quad (8)$$

Biomass (W) was calculated by using models developed by Balboa et al. (2006), who distinguish between the following components: wood (W_m), bark (W_c), large branches (W_{rg}), small branches (W_{rf}), twigs (W_{ram}), and needles (W_{ac})

$$\begin{aligned} W &= (W_m + W_c + W_{rg} + W_{\text{mrf}} + W_{\text{ram}} + W_{ac}) N \\ &= (0.0123 d_i^{1.6042} h^{1.4131} + 0.0036 d_i^{1.2.6564} \\ &\quad + (1.37699 + 0.001065 d_i^2 h) + 0.0363 d_i^{2.6091} h^{-0.9417} \\ &\quad + 0.0078 d_i^{1.9606} + 0.0423 d_i^{1.7141}) N \end{aligned} \quad (9)$$

Finally, the slenderness coefficient (C) expresses the relationship between mean height and mean diameter of the mass

$$C = \frac{h_a}{d_m} \quad (10)$$

Table 2. Dasometric Variables for Each Plot

Plot	d_m (m)	h_m (m)	H_0 (m)	h_c (m)	N	G (m^2/ha)	d_g (m)	HBI (%)	V (m^3/ha)	W (kg/ha)	C
P01	0.18	14.48	16.05	9.58	764	18.73	0.18	22.54	124.85	57,717	82.51
P02	0.20	16.42	17.05	10.85	732	23.65	0.20	21.68	182.46	83,696	81.64
P03	0.20	14.86	15.70	9.94	764	24.68	0.20	23.04	190.21	87,207	73.86
P04	0.28	24.58	26.48	17.08	605	38.28	0.28	15.36	385.36	179,483	88.24
P05	0.29	27.40	27.65	19.14	668	45.94	0.30	13.99	473.19	220,913	94.30
P06	0.28	25.55	25.93	16.55	509	31.59	0.28	17.10	315.00	146,441	92.27
P07	0.14	15.92	15.00	9.12	2,037	37.00	0.15	14.77	236.55	114,336	111.69
P08	0.16	18.50	18.95	10.10	1,751	39.83	0.17	12.61	278.51	131,151	114.48
P09	0.15	15.15	15.28	9.81	1,814	32.49	0.15	15.37	191.13	91,907	103.58
P10	0.14	21.78	21.95	11.49	3,056	59.87	0.16	8.24	442.67	216,048	155.76
P11	0.13	20.55	22.05	11.79	2,960	53.27	0.15	8.34	379.54	188,170	153.78
P12	0.12	22.48	22.28	12.06	2,992	46.22	0.14	8.21	312.87	157,546	182.16

Measurement points were selected to have groups with internally similar dasometric variables that were considerably different from the other groups. Hence, Points 1, 2, and 3 referred to areas with substantially less forest cover and Points 10, 11, and 12 referred to areas with substantially more forest cover (see Table 2), bearing in mind variables such as N, W, G, and HBI.

GPS Signal Variables

In addition, we recorded a series of variables for each measurement second that condition observation accuracy independently of tree canopy. These variables are listed as follows:

- $PDOP_p$: Position dilution of precision (PDOP) for each experimental point under forest canopy;
- $PDOP_r$: PDOP for the reference point;
- $\sigma_{X_{r\_acc}}$: Error X for the reference point;
- $\sigma_{Y_{r\_acc}}$: Error Y for the reference point;
- $\sigma_{XY_{r\_acc}}$: Error XY for the reference point;
- $\sigma_{Z_{r\_acc}}$: Error Z for the reference point;
- E_p: Mean elevation angle for the satellites transmitting the signal for the experimental point under the forest canopy;
- E_r: Mean elevation angle for the satellites transmitting the signal received by the reference point;
- $DLLSNCA_p$: Indicator of the signal-noise ratio in coarse/acquisition (CA) code (in dB • Hz) for the experimental point under the forest canopy;
- $DLLSNCA_r$: Indicator of the signal-noise ratio in CA code (in dB • Hz) for the reference point;
- nCA_p: Number of satellites receiving CA code for the experimental point under the forest canopy;
- nCA_r: Number of satellites receiving CA code for the reference point;
- $DLLSNL1_p$: Indicator of the signal-noise ratio in P code for L1 (in dB • Hz) for the experimental point under the forest canopy;
- $DLLSNL1_r$: Indicator of the signal-noise ratio in P code for L1 (in dB • Hz) for the reference point;
- $nL1_p$: Number of satellites receiving code in the L1 carrier for the experimental point under the forest canopy;
- $nL1_r$: Number of satellites receiving code in the L1 carrier for the reference point;
- $DLLSNL2_p$: Indicator of the signal-noise ratio in P code for L2 (in dB • Hz) for the experimental point under the forest canopy;
- $DLLSNL2_r$: Indicator of the signal-noise ratio in P code for L2 (in dB • Hz) for the reference point;
- $nL2_p$: Number of satellites receiving code in the L2 carrier for the experimental point under the forest canopy; and
- $nL2_r$: Number of satellites receiving code in the L2 carrier for the reference point.

Regression Models

To calculate GPS observation accuracy according to dasometric variables and GPS variables and to evaluate the influence of each, we used three different machine learning techniques, each briefly described subsequently. The models obtained by using these techniques are all nonlinear models. Other simpler models tested (such as univariate and multivariate regression) produced unsatisfactory results because they were incapable of establishing relationships among accuracy and the different explanatory variables. This is the main reason why they are not mentioned in this section; in addition, they are well-known methods that have been discussed in similar research.

The descriptions of the machine learning techniques used in this research are, out of necessity, brief. Readers interested in more in-depth explanations can look up the references provided.

Support Vector Regression

Support vector regression (SVR), which was proposed by Vladimir Vapnik and others to solve regression problems (Drucker et al. 1997), is based on concepts developed previously for classification problems and gives rise to support vector machines (SVMs) (Vapnik 1995). Given a training set of l samples $(\mathbf{x}_i, z_i)$, with $x_i \in R^n$ and $z_i \in R$, the goal is to find a function $f(\mathbf{x})$ with a maximum of one deviation ε with respect to the values z_i observed for the entire training set that is also as minimally complex as possible.

If we initially consider a linear function

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \quad \text{with} \quad w \in R^n, \qquad b \in R \tag{11}$$

in which $\langle w, x \rangle$ denotes internal product in R, the problem consists of finding $\mathbf{w}$ and b that satisfy the previously mentioned goal.

In this case, the condition of minimum complexity implies obtaining w with a minimal norm $\|\mathbf{w}\|^2 = \langle \mathbf{w}, \mathbf{w} \rangle$. The condition that the deviation of the function with respect to z_i should be less than $\varepsilon \; \forall \; i$, may not be feasible in some cases. For this reason, slack variables ξ, ξ^* are introduced to obtain an optimization problem that is computationally solvable.

Regression with SVR can be posed as an optimization problem with constraints (Smola and Schölkopf 2004):

$$\tau(\mathbf{w}, \xi, \xi^*) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}(\xi + \xi^*) \tag{12}$$

subject to

$$z_i - (\mathbf{w}, \mathbf{x}_i) - b \le \varepsilon + \xi_i \tag{13}$$

$$(\mathbf{w}, \mathbf{x}_i) + b - z_i \le \varepsilon + \xi_i^* \qquad \xi, \xi^* \ge 0 \tag{14}$$

The constant $C > 0$ determines the compromise between the complexity of f and model accuracy.

To generalize to a nonlinear regression, we used a method based on transforming the original space to a larger dimension space in such a way that linear solutions in the new space give rise to nonlinear solutions in the original space. Defining a positive definite function k (kernel) such that the chosen transformation is $\phi : R^n \to R^r$, and letting

$$\phi(\mathbf{x}) \cdot \phi(\mathbf{x}') = \sum_i \phi_i(\mathbf{x})\phi_i(\mathbf{x}') = k(\mathbf{x}, \mathbf{x}') \tag{15}$$

it is demonstrated that this function k is a well-defined scalar product (Boser et al. 1992).

Therefore, introducing the Lagrange multipliers $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$, the function to be minimized can be expressed as

$$W(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*) = \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T Q(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) + \varepsilon \sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} z_i(\alpha_i - \alpha_i^*) \tag{16}$$

subject to

$$\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0, \qquad 0 \le \alpha_i, \qquad \alpha_i^* \le C, \qquad i = 1, \ldots, l \tag{17}$$

in which

$$Q_{ij} = K(x_i, x_j) \equiv \varphi(x_i)^T \cdot \varphi(x_j) \tag{18}$$

The estimated regression function at a given point is, therefore,

$$f(\mathbf{x}) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)k(\mathbf{x}_i, \mathbf{x}) + b \tag{19}$$

in which b is obtained from the fact that Eq. (13) is converted into equality with $\xi_i = 0$ if $0 \le \alpha_i \le C$, and Eq. (14) is converted

Table 3. Means and Standard Deviations for R_2 Obtained for Horizontal Accuracy with Just One Explanatory Variable

	SVR		MLP		RBF	
Variable	R^2	σ_{R^2}	R^2	σ_{R^2}	R^2	σ_{R^2}
CE	0.7572	0.1399	0.6745	0.2230	0.5756	0.2689
HBI	0.7524	0.1385	0.6919	0.1946	0.5756	0.2689
V	0.7521	0.1345	0.6857	0.2082	0.5756	0.2689
H_0	0.7510	0.1482	0.6478	0.2424	0.5756	0.2689
N	0.7501	0.1291	0.6781	0.2280	0.5707	0.2612
G	0.7483	0.1278	0.6790	0.2386	0.5756	0.2689
h_a	0.7467	0.1505	0.6653	0.2420	0.5756	0.2689
W	0.7453	0.1277	0.6712	0.2364	0.5756	0.2689
d_g	0.7448	0.1277	0.5304	0.2903	0.5756	0.2689
h_c	0.7446	0.1274	0.6450	0.2568	0.5756	0.2689
d_m	0.5828	0.2446	0.5122	0.3098	0.4384	0.2967
DLLSNCA_P	0.5057	0.2860	0.4811	0.2591	0.0038	0.0169
DLLSNCA_R	0.4050	0.2908	0.2788	0.2362	0.1873	0.2201
DLLSNL1_P	0.3808	0.2784	0.2237	0.2568	0.1929	0.2630
nCA_r	0.3774	0.2226	0.2828	0.1825	0.2017	0.1926
nL1_r	0.3694	0.2185	0.3182	0.2087	0.0734	0.1551
DLLSNL2_p	0.3558	0.2966	0.2001	0.2487	0.1388	0.2234
PDOP_p	0.3426	0.2308	0.0740	0.1340	0.0815	0.1656
nL2_p	0.3309	0.2566	0.1467	0.2259	0.1072	0.2075
nL1_p	0.3309	0.2566	0.1204	0.1803	0.1072	0.2075
E_r	0.3278	0.2710	0.2724	0.2027	0.1769	0.1717
$\sigma_{XY_{r\_acc}}$	0.3066	0.2607	0.2032	0.1926	0.1233	0.1919
DLLSNL2_r	0.2905	0.2217	0.2081	0.2054	0.1548	0.2071
DLLSNL1_r	0.2862	0.2200	0.1871	0.1898	0.1918	0.2352
$\sigma_{Y_{r\_acc}}$	0.2827	0.2685	0.1581	0.1607	0.1088	0.1562
PDOP_r	0.2704	0.2260	0.1703	0.1534	0.0112	0.0500
nCA_p	0.2681	0.2592	0.0974	0.1251	0.0331	0.0894
$\sigma_{Z_{r\_acc}}$	0.2641	0.2068	0.1716	0.1786	0.0858	0.1798
E_p	0.2457	0.2185	0.2121	0.2379	0.1112	0.1817
$\sigma_{X_{r\_acc}}$	0.2304	0.2181	0.0992	0.1419	0.0570	0.0930

Table 4. Means and Standard Deviations for R^2 Obtained for Vertical Accuracy with Just One Explanatory Variable

	SVR		MLP		RBF	
Variable	R^2	σ_{R^2}	R^2	σ_{R^2}	R^2	σ_{R^2}
CE	0.6410	0.2223	0.5795	0.2292	0.4682	0.2850
HBI	0.6427	0.2257	0.6268	0.2055	0.4682	0.2850
V	0.6411	0.2224	0.5664	0.2181	0.4682	0.2850
H_0	0.6410	0.2223	0.6339	0.1944	0.4682	0.2850
N	0.6112	0.2501	0.5137	0.2584	0.4351	0.2654
G	0.6455	0.1858	0.6056	0.2255	0.4685	0.2850
h_a	0.6433	0.1651	0.6351	0.1637	0.4682	0.2850
W	0.6449	0.2239	0.5637	0.2762	0.4682	0.2850
d_g	0.6215	0.2252	0.5947	0.2066	0.4682	0.2850
h_c	0.6410	0.2223	0.6338	0.2005	0.4682	0.2850
d_m	0.6413	0.1629	0.5888	0.2415	0.4394	0.2893
DLLSNCA_P	0.2851	0.2775	0.0806	0.0819	0.0600	0.1485
DLLSNCAr	0.3431	0.2673	0.2130	0.1886	0.1119	0.1452
DLLSNL1_P	0.3242	0.3107	0.2017	0.1926	0.1245	0.1823
nCA_r	0.2672	0.2688	0.1922	0.1707	0.0403	0.1103
nL1_r	0.2696	0.2675	0.2310	0.2037	0.1585	0.2554
DLLSNL2_p	0.2418	0.1958	0.0881	0.1360	0.0744	0.0873
PDOP_p	0.1989	0.2030	0.1342	0.1495	0.0562	0.1296
nL2_p	0.2905	0.2836	0.2525	0.2029	0.0080	0.0358
nL1_p	0.2905	0.2836	0.2025	0.2004	0.0080	0.0357
E_r	0.2589	0.2936	0.1016	0.1204	0.0164	0.0732
$\sigma_{XY_{r\_acc}}$	0.3328	0.2806	0.1237	0.1507	0.1331	0.2171
DLLSNL2_r	0.3155	0.2826	0.1864	0.1600	0.0766	0.1392
DLLSNL1_r	0.1996	0.2162	0.1396	0.1509	0.0798	0.1084
$\sigma_{Y_{r\_acc}}$	0.2536	0.2685	0.0780	0.1046	0.0797	0.1319
PDOP_r	0.2803	0.3036	0.0368	0.1198	0.0704	0.1719
nCA_p	0.2673	0.1681	0.1450	0.1804	0.0816	0.1464
$\sigma_{Z_{r\_acc}}$	0.3501	0.2544	0.1433	0.1904	0.1628	0.1878
E_p	0.3501	0.2391	0.2202	0.2188	0.1090	0.1655
$\sigma_{X_{r\_acc}}$	0.3252	0.2804	0.1153	0.1564	0.1067	0.1534

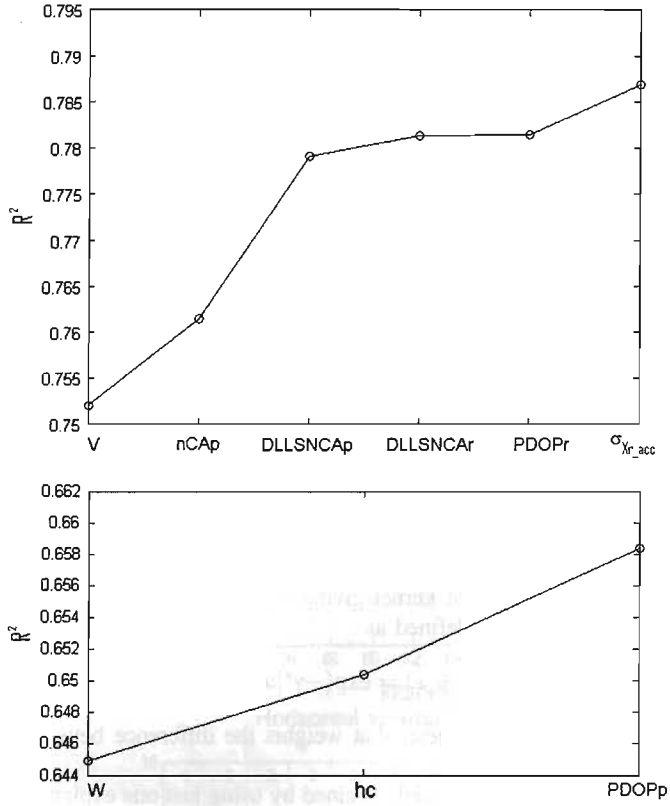


Fig. 2. Variation in R^2 for horizontal accuracy (top) and vertical accuracy (bottom) as explanatory variables are added to the model

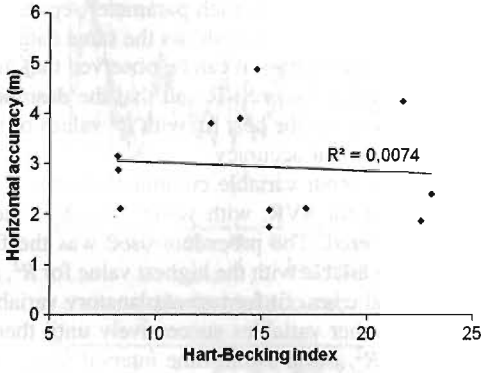


Fig. 3. Scatter plot of the mean horizontal errors versus the Hart-Becking index for each of the 12 experimental points

into equality with $\xi_i^* = 0$ if $0 \leq \alpha_i^* \leq C$. This function $f(\mathbf{x})$ depends solely on the training subsets for which $\alpha_i, \alpha_i^* \neq 0$. The elements of this subset are known as support vectors.

Multilayer Perceptron

The first modern neural model, proposed by McCulloch and Pitts (1943), has served as the basis for many current models. The two architectures considered are multilayer perceptron (MLP) and the radial basis function network (RBFN), both of which are based on a more general multilayer feedforward model (Haykin 1999).

MLP is a neural network formed by an input layer, an output layer, and one or more hidden layers (Pinkus 1999). Each neuron in layer j can be connected to all the neurons in layer $j+1$ or to just a specific number of neurons. Its output is given by the expression

$$T\left(\beta_0 + \sum_{i=1}^{d} \beta_i x_i\right) \tag{20}$$

in which $x_i = i$th input; T = nonlinear activation function for R in R; and $\beta_0, \beta_1, \ldots, \beta_d$ = numeric parameters (the neuron weights).

The MLP output proposed in this case takes the following general form (activation function T is a sigmoid with d neurons in the input layer, h neurons in the hidden layer, and one neuron in the output layer):

$$f(x) = \sum_{j=1}^{h} c_j u_j + c_0 \tag{21}$$

in which $c_0, c_1, \ldots, c_h$ = weights associated with the neurons in the hidden layer; and $u_j, j = 1, 2, \ldots, h$ = outputs for each neuron in this layer expressed in Eq. (20).

To train the network, we used the Bayesian regularization algorithm, which adjusts the weights and values of the constants according to Levenberg-Marquardt optimization (MacKay 1992; Foresee and Hagan 1997). This minimizes the combination of quadratic errors and weights and determines the correct combination produced by a network that generalizes well.

Radial Basis Function Network

The RBFN differs from MLP is several ways (Bishop 1995). First of all, it generally has a fixed architecture composed of one input

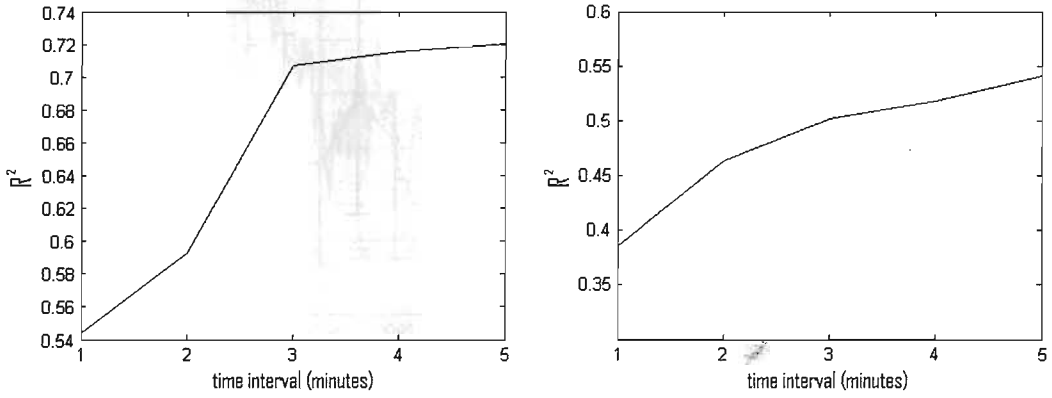


Fig. 4. Variation in R^2 for the SVR model as the time interval increases: horizontal accuracy (left) and vertical accuracy (right)

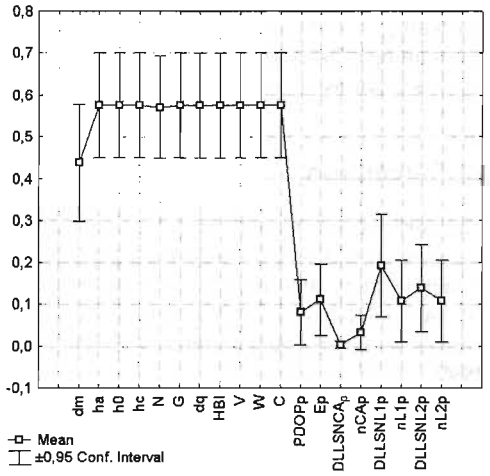


Fig. 5. Means of the R^2 values and the confidence intervals (95%) for the horizontal error, for each dasometric variable, and for each GPS signal variable, for all the experimental points

layer, one output layer, and one hidden layer. Second, the neurons in the hidden layer have a radial basis function (RBF) as the activation function. RBFs also have the peculiarity that they only depend on the distance from the argument to a specific center. In this case, we chose a Gaussian function as giving the best results

$$h_j(\mathbf{x}) = e^{-\|\mathbf{x}-c_j\|^2/r_j^2} \tag{22}$$

in which $\mathbf{x}$ = input vector; h_j = output of the jth neuron in the hidden layer; and c_j and r_j = center and the width of the corresponding radial function, respectively.

The operation in the output layer is represented by the expression

$$y_k(\mathbf{x}) = \sum_{j=1}^{n} w_{kj} h_j(\mathbf{x}) + b_k \tag{23}$$

in which y_k = response of the kth neuron in the output layer; w_{kj} = weight corresponding to the connection between the kth neuron in the output layer and the jth neuron in the hidden layer; and b_k = a constant term.

Results

Each of the machine learning techniques previously described was used to make an estimate of the horizontal and vertical accuracy of the observations. In this work, the sample was composed of a total of 43,200 data observations. The goodness of fit for each model was measured by using the coefficient of determination R^2, whose values range between 0 and 1. R^2, which indicates the fraction of variance in a GPS error explained by the model under consideration (1 represents a perfect fit), is calculated as follows:

$$R^2 = \frac{(\mathrm{SSY} - \mathrm{SSE})}{\mathrm{SSY}} \tag{24}$$

in which

$$\mathrm{SSY} \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad \text{and} \quad \mathrm{SSE} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{25}$$

in which y_i = ith value for the variable to be predicted; $\bar{y}$ = variable measurement; and $\hat{y}_i$ = prediction of the model for y_i.

A cross-validation procedure was used for the SVR method to determine the type of kernel giving the best results. This kernel, based on RBFs, is defined as

$$k(u, v) = \exp(-\gamma^* |u - v|^2) \tag{26}$$

in which γ = a parameter that weights the difference between u and v.

The network was initially trained by using just one explanatory variable each time, with a view to determine which variable yielded most information on the error. The network parameters were determined in each case by using 20-fold cross-validation. Table 3 shows the means and standard deviations for R^2 obtained on predicting horizontal accuracy with each parameter separately and by using each of the methods. Table 4 shows the same data for error in the Z coordinate. In both tables, it can be observed that the method producing the best results was SVR and that the dasometric variables for each plot yielded the best fit, with R^2 values of more than 0.7, mostly for horizontal accuracy.

By using various input variable combinations, the best results were again obtained for SVR, with values that depended on the time interval considered. The procedure used was the following: we started with the variable with the highest value for R^2, added the variable producing the best fit for two explanatory variables to the model, and added other variables successively until there was no further increase in R^2. For a 5-min time interval (i.e., considering

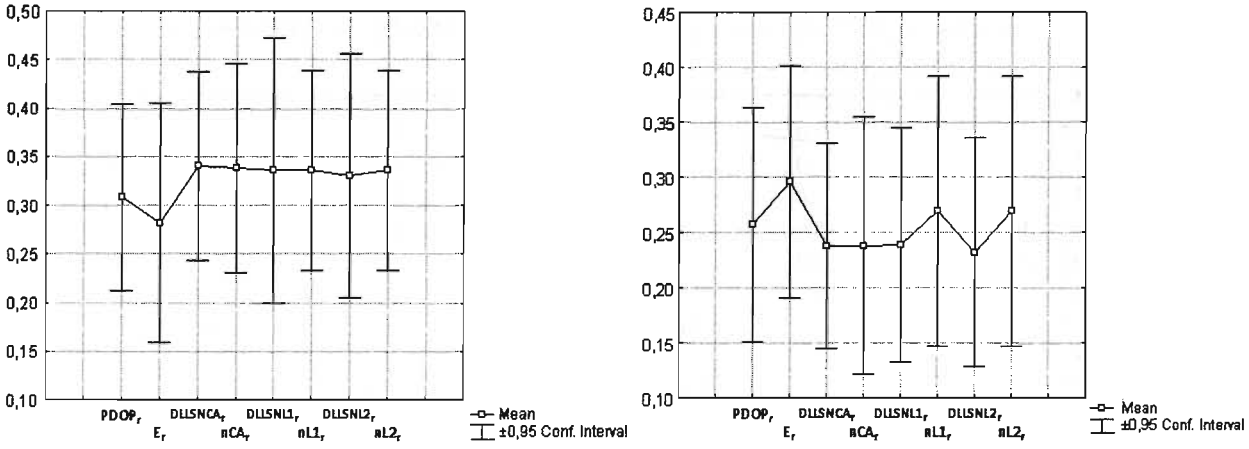


Fig. 6. Means of the R^2 values and the confidence intervals (95%) for the horizontal accuracy (left) and the vertical accuracy (right) for the reference point, for the GPS signal variables

mean error every 5 min as the dependent variable), the optimal parameters were $C = 1$, $\varepsilon = 0.05$, and $\gamma = 0.1$ for the XY error (horizontal accuracy) and $C = 1000$, $\varepsilon = 0.001$, and $\gamma = 2$ for the Z error (vertical accuracy).

Fig. 2 shows the contribution of each variable to the total precision of the model. It can be observed that the dasometric values best explained most of the error. The contribution of the remaining parameters and variables was quite small. This was confirmed by applying the same techniques to the errors at the reference point to locate a model that related these errors with the GPS variables (the forest canopy variables, obviously, had no bearing on reference point error).

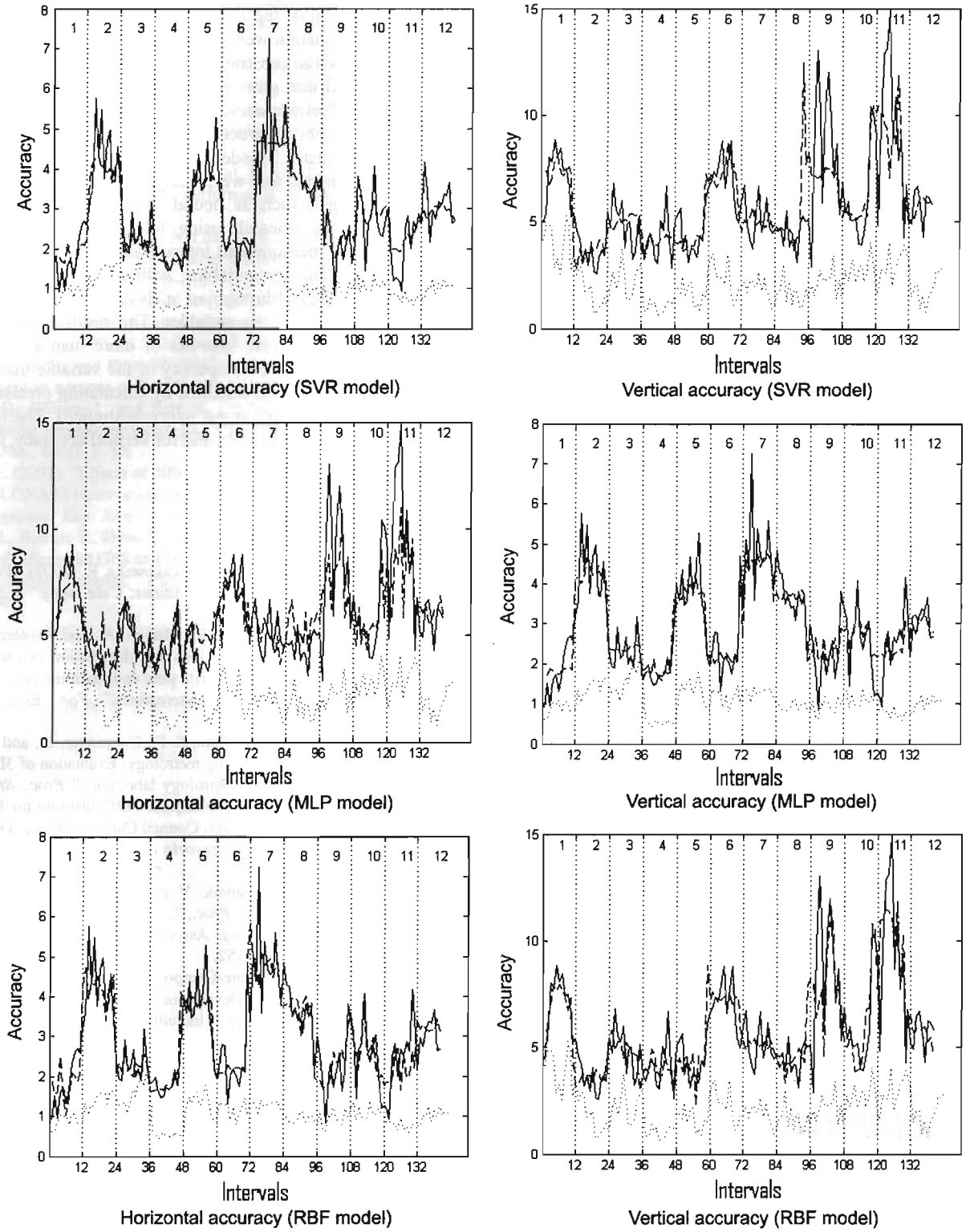


Fig. 7. Horizontal and vertical accuracies for the 12 experimental points and the reference point; the values for the 12 experimental points are represented consecutively; the solid line is assigned to the calculated accuracy and the dashed line to the estimated one, for each of the three regression models; the calculated accuracy for the reference point is shown with a dotted line

Table 5. Maximum R^2 Obtained for Accuracy for Each of the 12 Points under the Forest Canopy by Technique Used

	SVR	MLP	RBF
H_{acc}	0.7869	0.7393	0.6109
V_{acc}	0.6584	0.6422	0.5646

This error dependence on the forest canopy was not detected by simpler models, such as linear regression or polynomial models. Fig. 3 shows a mean value for horizontal accuracy for each experimental point with reference to the *HBI* variable. There exists no observable direct relationship between them. The same conclusion can be drawn regarding the remaining forest cover and GPS signal parameters and variables.

It can be concluded from the results that the influence on accuracy of the parameters associated directly with the GPS signal is low.

Error variability is important when trying to obtain a good fit. The fit improved considerably with longer time intervals because there exist fewer fluctuations in the output variable (Fig. 4).

To determine whether significant differences exist for the values for R_2 obtained by using 20-fold cross-validation, means were compared for each of the variables by using the t-test, with the results showing that for the dasometric variables (which provide most information on the model), only d_m differed from the rest. Fig. 5 shows the R_2 means and confidence intervals for horizontal error, for the dasometric variables and the GPS signal variables, and for each experimental point under the forest canopy. It confirms the t-test, and the R_2 obtained for d_m is significantly lower than the other variables. Therefore, with the exception of d_m, the dasometric variables explain the accuracy of the measures equally.

None of the analyzed dasometric variables directly measure foliage density, and Sigrist et al. (1999) concluded that, although the existence of the canopy overhead may degrade positional precision, it is the foliage that plays the greatest role in signal reception. For the same kind of analysis made for accuracy at the reference point, it was concluded that there were no significant differences between the different GPS signal variables (the dasometric variables do not, obviously, affect accuracy at this point). The results of the t-test are shown in Fig. 6. The best fit was for *DLLSNCAr*, in which $R_2 = 0.34$, so the signal variables were insufficient to determine error at this point. To corroborate this result, larger data sets are needed; however, Yeh et al. (2010), who studied the quality and precision of GPS positioning in reference stations, found no significant clock offset and multipath effects.

In Fig. 7, which shows the observed (solid line) and estimated (dashed line) precision values for 5-min time intervals, jumps corresponding to the 12 experimental points are evident, although they have been marked by using vertical dotted lines to facilitate their recognition. The dotted line corresponds to the precision of the reference point. As expected, the error for the reference point (dotted line) is considerably smaller than on points under the forest canopy; therefore, as the statistic analysis has confirmed, most of the error is because of the effect of vegetation on the GPS signal. The corresponding values for R_2 are shown in Table 5. SVR was the model with the best fit, although it was similar to the fit for the MLP network.

Conclusions

We evaluated the influence of dasometric variables and the variables associated with the GPS signal for the accuracy of observations made with a GPS receiver under the forest canopy. The results obtained for the three machine learning techniques used pointed to the same conclusion: accuracy and precision depend greatly on the presence of forest cover and far less on GPS signal variables, such as PDOP, the number of satellites, or the signal-noise ratio.

A comparison of the coefficients of determination for each dasometric variable that was analyzed revealed no significant differences between the variability explained by each variable, with the sole exception of the mean diameter, for which lower coefficients of determination were obtained. Therefore, any of these variables can be used to construct models to estimate error, although it should be noted that other factors that were not included in the model could affect accuracy.

The great variability in accuracy over time makes it complicated to locate mathematical models to estimate measurement error for the kind of variables that were analyzed here, including complex nonlinear models, such as neural networks. This may explain why, in previous research using linear models, accuracy and precision of the measurement interval are associated with a small number of explanatory variables, with no information provided about the fit, merely [through an analysis of variance (ANOVA)] on the significance of the variables. The results improved when mean values for the time intervals of more than 1 s were used because this reduced the complexity of the variable that was estimated. The better fit was obtained by calculating precision values for 5-min intervals (the longest intervals studied). The best fit for horizontal accuracy was 80%; and for vertical accuracy, the best fit was only 66%.

References

Bakula, M., Oszczak, S., and Pelc-Mieczkowska, R. (2009). "Performance of RTK positioning in forest conditions: Case study." *J. Surv. Eng.*, 135(3), 125–130.

Balboa, M., Rodríguez-Soalleiro, R., Merino, A., and Álvarez-González, J. G. (2006). "Temporal variations and distribution of carbon stocks in aboveground biomass of radiata pine and maritime pine pure stands under different silvicultural alternatives." *For. Ecol. Manage.*, 237(1–3), 29–38.

Beraldin, J. A., Blais, F., El-Hakim, S. F., Cournoyer, L., and Picard, M. (2007). "Traceable 3D imaging metrology: Evaluation of 3D digitizing techniques in a dedicated metrology laboratory." *Proc., 8th Conf. on Optical 3-D Measurement Techniques*, NRC Institute for Information Technology; National Research Council Canada, Ottawa, ON, Canada.

Bishop, C. M. (1995). *Neural networks and pattern recognition*, Oxford Univ., New York.

Boser, B., Guyon, I., and Vapnik, V. (1992). "A training algorithm for optimal margin classifiers." *Proc., 5th Annual ACM Workshop on Computational Learning Theory*, Association for Computing Machinery (ACM), New York, 144–152.

Castedo-Dorado, F., Crecente-Campo, F., Álvarez-Álvarez, P., and Barrio-Anta, M. (2009). "Development of a stand density management diagram for radiata pine stands including assessment of stand stability." *Forestry*, 82(1), 1–16.

Drucker, H., Burges, J. C., Kaufman, L., Smola, A., and Vapnik, V. (1997). "Support vector regression machines." *Proc., 1996 Conf. NIPS: Advances in Neural Information Processing Systems 9*, MIT, Cambridge, MA, 155–161.

Foresee, F. D., and Hagan, M. (1997). "Gauss-Newton approximation to Bayesian learning." *Proc., Int. Joint Conf. on Neural Networks*, IEEE, New York, 1930–1935.

Gegout, J. C., and Piedallu, C. (2005). "Effects of forest environment and survey protocol on GPS accuracy." *Photogramm. Eng. Remote Sens.*, 71(9), 1071–1078.

Hasegawa, H., and Yoshimura, T. (2003). "Application of dual-frequency GPS receivers to static surveying under tree canopies." *J. For. Res.*, 8(2), 103–110.

Hasegawa, H., and Yoshimura, T. (2007). "Estimation of GPS positional accuracy under different forest conditions using signal interruption probability." *J. For. Res.*, 12(1), 1–7.

Haykin, S. (1999). *Neural networks. A comprehensive foundation*, 2nd Ed., Prentice Hall, Upper Saddle River, NJ.

Joint Committee for Guides in Metrology (JCGM). (2008). *International vocabulary of metrology (VIM)—Basic and general concepts and associated terms*, 3rd Ed., ⟨http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2008.pdf⟩ (Sep. 5, 2010).

MacKay, D. J. C. (1992). "Bayesian interpolation." *Neural Comput.*, 4(3), 415–447.

Mancebo, S., and Chamberlain, K. (2000). "Performance testing of the Garmin eTrex, Garmin GPSIII plus, Magellan GPS 2000XL, and Magellan Blazer recreation type global positioning system receivers." ⟨http://www.fs.fed.us/database/gps/mtdc/gps2000/Nav_3-2001.htm⟩ (Feb. 5, 2010).

Martin, A. A., Owende, P. M., and Ward, S. M. (2001). "The effects of peripheral canopy on DGPS performance on forest roads." *Int. J. For. Eng.*, 12(1), 71–79.

McDonald, T. P., and Carter, E. A. (2002). "Using the global positioning system to map disturbance patterns of forest harvesting machinery." *Can. J. For. Res.*, 32(2), 310–319.

McDonald, T. P., and Fulton, J. P. (2005). "Automated time study of skidders using global positioning system data." *Comput. Electron. Agric.*, 48(1), 19–37.

McCulloch, W. S., and Pitts, W. (1943). "A logical calculus of the ideas immanent in nervous activity." *Bull. Math. Biophys.*, 5, 115–133.

Murphy, G., Wilson, I., and Barr, B. (2006). "Developing methods for pre-harvest inventories which use a harvester as the sampling tool." *Aust. For.*, 69(1), 9–15.

Næsset, E. (2001). "Effects of differential single- and dual-frequency GPS and GLONASS observations on point accuracy under forest canopies." *Photogramm. Eng. Remote Sens.*, 67(9), 1021–1026.

Næsset, E., Bjerke, T., Øvstedal, O., and Ryan, L. H. (2000). "Contributions of differential GPS and GLONASS observations to point accuracy under forest canopies." *Photogramm. Eng. Remote Sens.*, 66(4), 403–407.

Næsset, E., and Jonmeister, T. (2002). "Assessing point accuracy of DGPS under forest canopy before data acquisition, in the field and after postprocessing." *Scand. J. For. Res.*, 17(4), 351–358.

Pinkus, A. (2008). "Approximation theory of the MLP model in neural networks." *Acta Numer.*, 8, 143–195.

Rodríguez-Pérez, J. R., Álvarez, M. F., and Sanz-Ablanedo, E. (2007). "Assessment of low-cost GPS receiver accuracy and precision in forest environments." *J. Surv. Eng.*, 133(4), 159–167.

Sawaguchi, I., Nishida, K., Shishiuchi, M., and Tatsukawa, S. (2003). "Positioning precision and sampling number of DGPS under forest canopies." *J. For. Res.*, 8(2), 133–137.

Sawaguchi, I., Saitoh, Y., and Tatsukawa, S. (2005). "A study of the effects of stems and canopies on the signal to noise ratio of GPS signals." *J. For. Res.*, 10(5), 395–401.

Sigrist, P., Coppin, P., and Hermy, M. (1999). "Impact of forest canopy on quality and accuracy of GPS measurements." *Int. J. Remote Sens.*, 20(18), 3595–3610.

Smola, A. J., and Schölkopf, B. (2004). "A tutorial on support vector regression." *Stat. Comput.*, 14(3), 199–222.

Tachiki, Y., Yoshimura, T., Hasegawa, H., Mita, T., Sakai, T., and Nakamura, F. (2005). "Effects of polyline simplification of dynamic GPS data under forest canopy on area and perimeter estimations." *J. For. Res.*, 10(4), 19–427.

Taylor, S. E., McDonald, P., Fulton, J., Shaw, J., Corley, F. W., and Brodbeck, C. J. (2006). "Precision forestry in the southeast U.S." *Proc., 1st Int. Precision Forestry Symp.*, University of Washington, Seattle, 397–414.

Vapnik, V. (1995). *The nature of statistical learning theory*, Springer-Verlag, New York.

Yeh, T. K., Chen, Y. J., Chung, Y. D., Feng, C. W., and Xu, G. (2010). "Clarifying the relationship between quality of global positioning system data and precision of positioning." *J. Surv. Eng.*, 136(1), 41–45.

Yoshimura, T., and Hasegawa, H. (2003). "Comparing the precision and accuracy of GPS positioning in forested areas." *J. For. Res.*, 8(3), 147–152.