ISB 2013 BRAZIL
XXIV CONGRESS OF THE INTERNATIONAL SOCIETY OF BIOMECHANICS
XV BRAZILIAN CONGRESS OF BIOMECHANICS

THE EFFECT OF SAMPLE FREQUENCY, RUNNING VELOCITY AND CUT LENGTH ON ERRORS IN DISTANCE MEASUREMENTS FOR SIMULATED GPS TRACKING

Matthew T.G. Pain and John Morris
School of Sport, Exercise and Health Sciences, Loughborough University
email: m.t.g.pain@lboro.ac.uk, web: http://www.lboro.ac.uk/departments/ssehs/

SUMMARY
GPS systems are now commonly used to track player motion during training and games, but have limited sampling frequencies that can affect the accuracy of the required measures. Simulated GPS tracking data were created with a custom Matlab program. Number of cuts, range of velocities, and range of cut lengths were input as variables. Data were resampled at 1 Hz intervals, from 1 to 15 Hz. Regression equations were fitted to the percentage error values for line fits, surface fits and a 4D fit for all variables simultaneously. Error in distance covered was non-linearly dependent on cut length, sample frequency and velocity, with interactions between these three variables. A single regression equation relating error to the three variables $E= (((k1 \cdot x^{k2}) \cdot y)/z)+(((k3 \cdot y^{k4})/x)/z)+(((k5 \cdot y^{k6})/z) \cdot x$ gave an RMSE of 0.75%. A minimal level of error, from sampling limitations combined with running velocity and cut length, in distance travelled can be estimated from these regression equations for a wide range of parameter values.

INTRODUCTION
The use of GPS is becoming increasingly prevalent in team sports for the evaluation of the distance travelled, the speed of movement and even the acceleration of players during training and in competitive match play. Several research studies have assessed the validity and reliability of such equipment for these purposes [1,2,3] Early versions of GPS devices used in team sports had a measurement frequency of 1 Hz, whereas this has increased to 5 Hz and even 15 Hz in the latest models. The validity of a 1 Hz GPS system for the assessment of movements typical of those in team games has been established [3] but there is some debate regarding the adequacy of 1 Hz systems for measuring faster movements and movements that involve deviations from a straight line. The aim of this study was to use simulated GPS tracking type data to determine errors for distance travelled when sample frequency, running velocity and changes in direction were systematically altered.

METHODS
A Matlab script was developed to simulate GPS data for running at different velocities for different cut lengths (distance travelled in one direction before a change in direction) on a 100 m by 60 m field and allow resampling at different frequencies (Figure 1). Number of cuts, range of velocities, and range of cut lengths were input as variables. The velocities and cut lengths were normally distributed within the set ranges, and the cuts were constrained to remain on the field. The change in direction with each cut was randomly determined. This likely represents the worst case scenario for extreme changes in motion as a player's movements will tend to have some directional bias for extended periods of time relative to the time spent in a cut. Linear deceleration into, and acceleration out of, a change in direction was included so as not to overestimate changes in position during changes in direction. The transition velocity at the point of the cut was dependent on the mean of the incoming and outgoing velocity, and linearly dependent on the angle between the cuts, thus the greater the angular change between cuts the greater the deceleration and acceleration. The cumulative distance for all cuts covered with the set parameters was determined with an effective sample rate of 100 Hz (higher sample frequencies gave no improvement in accuracy to 5 decimal places). Once this had been done the data were resampled at other frequencies and the cumulative distance recalculated at the new sample frequencies. Finally a total error in distance relative to 100 Hz was calculated and expressed as a percentage for each new sample frequency.

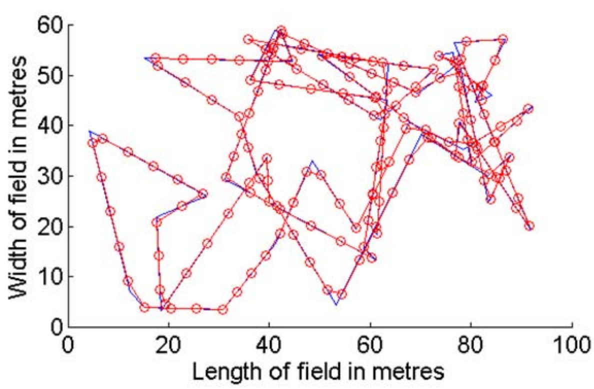

Figure 1: Simulated GPS data: 50 cuts between 2 and 40 m long at 4-7 ms^{-1}, in blue, and resampled at 1 Hz, in red.

The effect of velocity, frequency and cut length on cumulative distance error was systematically determined for a wide range of values. Cut lengths with mean values of 2 m, then 5-50 m in 5 m intervals were examined. For each cut length simulations with mean running velocities of 1-10 ms^{-1}, in 1 ms^{-1} steps were run. In both these cases instead of using a fixed value a range was used. The upper and lower bounds of the range were the desired value $\pm$ 0.5. A range of values was chosen instead of a single value to better represent real GPS data and to avoid any possible systematic errors from synchronous effects between variables. Each of these 110 simulations was resampled at 1 Hz intervals from 1 to 15 Hz. These values covered realistic ranges for, cut length and running velocity of players in training and games, and sample frequency of GPS units. However, some combinations of these three independent parameters will result in unrealistic situations, such as repeated 2 m cuts at a velocity of 10 ms^{-1}. This gave a total of 1650 individual trials covering all combinations of cut length, velocity, and sample frequency. For each cut length and velocity 5000 cuts were run in a single simulation in order to get a converged and stable error value when resampled. Finally regression equations were fitted to the percentage error values for line fits, surface fits and a 4D fit for all variables simultaneously, in order to allow interpolation and extrapolation of error values. This was done in Matlab using the sfit GUI and lsqcurvefit with multiple seeding.

RESULTS AND DISCUSSION

Error in total distance covered was non-linearly dependent on cut length, sample frequency and velocity. Obviously as sample frequency increased error decreased, similarly for cut length, and as velocity increased error increased (Figure 2). Also of interest were the interactions between the three variables (Figure 2)

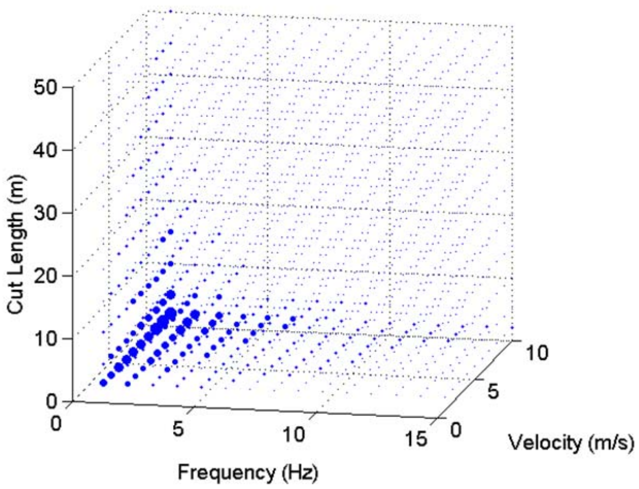

Figure 2: Percentage error versus sample frequency, velocity, and cut length. Size of dot represents error, min = 0.036% max = 59.6%.

Regression equations fitted to error versus sample frequency were best represented by families of power relationships of the form $E = a \cdot x^b + c$, and had R^2 values of >0.95 for their respective optimized constants. Regression equations fitted to error versus velocity were best represented by families of second order polynomials (R^2 >0.95), but for ~85% of them there was no notable improvement from using a first order polynomial. Regression equations fitted to error versus cut length could also be well represented by families of power relationships of the form $Y = a \cdot x^b + c$, and generally had R^2 values of >0.95 for their respective optimized constants. For all 2D fits RMS error was between 0.03% and 4% in terms of absolute percentage error and less than 10% of relative percentage error, to the maximum error in a given curve. The largest relative errors occurred with the lowest absolute errors and these are towards the top right of Figure 2.

Surface fits of error versus:

sample frequency (x) and velocity (y)

$$\mathbf{E=((a1 \cdot x^{a2}) \cdot y) + ((a3 \cdot y + a4)/x)} \qquad \text{eq.1}$$

velocity (x) and cut length (y)

$$\mathbf{E=((b1 \cdot x^{b2}) \cdot y)+b3+((b4 \cdot y\text{^}2\ b5 \cdot y+b6) \cdot x)} \qquad \text{eq.2}$$

sample frequency (x) and cut length (y);

$$\mathbf{E=((a3 \cdot x^{b3}) \cdot y) +c2 + ((p3 \cdot y + q3) \cdot x)} \qquad \text{eq.3}$$

all gave fits with $R^2 \geq 0.99$, RMSE of < 0.5 and a mostly uniform looking residuals, with only one or two outlying residuals per fit. There are a number of effectively equivalent solutions, for R^2 values and RMSE to 3 decimal places but a set of constants determined from least squares fitting are in Table 1

Using the results from the three surface fits a single fit to the data shown in Figure 2 was performed that gave an RMSE fit of 0.75. Using equation 4

$$\mathbf{E= (((k1 \cdot x^{k2}) \cdot y)/z)+(((k3 \cdot y^{k4})/x)/z)+(((k5 \cdot y^{k6})/z) \cdot x} \qquad \text{eq.4}$$

Where:

x = sample frequency; y = velocity; z = cut length

One set of constants that give this level of fit is:

k1= 0.0002	k2= 3.2039	k3= 10.5838
k4= 37.3578	k5= -0.0492	k6= -0.3614

CONCLUSIONS

Error due to sample rate limitations in GPS tracking of motion are inter-related to running velocity and distance run in a straight line, the cut distance. A minimal level of error, due to sampling limitations, combined with running velocity and cut length, in distance travelled can be estimated from the regression equations provided here. This could aid with planning future experiments or assessing the accuracy of already collected data sets.

REFERENCES

1. Coutts, A.J. & Duffield, R. *Journal of Science and Medicine in Sport.* **13**, 133–135, 2010.
2. Gray, A.J., et al. *Journal of Sports Sciences.* **28**, 1319–1325, 2010.
3. MacLeod, H., et al.(2009). *Journal of Sports Sciences.* **27**, 121-128, 2009.

Table 1: One set of constants for each surface fit equation, equations 1 to 3.

Constant	a1	a2	a3	a4	b1	b2	b3	b4	b5	b6	c1	c2	c3	c4	c5
Value	-351	-0.999	352	1.44	16.7	0.753	-0.18	-0.0003	0.023	-0.58	-99.3	-1.16	-0.59	0.001	0.002