

# chapter 2 - RF model

Dajun Wang

11/9/2019

## 1 Introduction

## 2 Material and methods

### 2.1 Random forest model construction

Random Forests (RF) is a relatively novel and powerful machine learning algorithm that has been reported to work well with complex ecological data that cannot be easily fitted with traditional methods such as generalized linear models (??). Through the use of tri-axial accelerometry measurements (i.e., acceleration values), RF models can also be used to predict unobservable behaviours in wild, free-ranging animals based on information collected from examining captive animals. To do this, the RF model must first be trained to recognize and predict labelled behaviours in a ‘training’ dataset with decision trees constructed with suitable predictor variables (Table 1) used for classifying and predicting behaviours (Table 2).

For the classification and labelling of known behaviours in accelerometry data, I collared 11 healthy adult dogs of three different dog breeds: German shepherds ( $n = 9$ , six males and three females, age range: 1–8 years, mean age: 5 years old, Rottweiler ( $n = 1$ , male, age: 8 years old) and Golden Retriever ( $n = 1$ , male, age: 1 year old) with an accelerometer-equipped wildlife tracking collar (Type 1C-heavy, E-obs GmbH; Grünwald, Germany). The wildlife tracking collar used in this chapter is also used for collaring and tracking free-roaming dogs in

the subsequent chapters. All eleven dogs were well-trained individuals that could perform the selected repertoire of movement behaviours (Table 2) solely with verbal instructions from their trainer while being off-leash.

Table 2: Description of the behavioural tasks performed by kennel club-trained domestic dogs.

Behaviour	Description
Stand	Individual under study is standing on all four limbs with no visible locomotion.  Resting posture of the dog's head will vary between individuals.
Lie	Individuals under study is lying prone on the ground with all limbs extended out. For some individuals, the dog may be lying fully on its side. In all recordings, there was no visible locomotion.
Sit	Individual under study is sitting on its haunches with no visible locomotion.  Similar to Stand, head posture may vary per individual but body posture remains relatively similar.
Eat	Individual under studying is eating or drinking from a bowl whilst standing. No visible locomotion but the posture of the head is angled approximately $45^\circ$ downwards.

Behaviour	Description
Walk	Individual under study is walking freely without leash at the same speed of the handler.
Forage	Similar to Walk but the position of the dog's head is angled downwards as it sniffs for hidden treats on the ground.
Run	Individual under study is sprinting after a tossed ball. Task is considered completed once the tossed ball has been received by the dog.

For the purpose of collecting data to train the RF models (i.e., training dataset), the wildlife tracking collar was programmed to sample continuously at 10 hz, and the collar was mounted such that the x-, y-, and z- axes of the tri-axial accelerometer were parallel to the median (surge), the dorsal (sway), and the dorsal (heave) planes of the animal, respectively. The design and alignment of the collar provided the tri-axial accelerometer the capacity to measure the surge (back and forth movement on the x-axis), sway (left and right motion on the y-axis) and heave (up and down on the z-axis) motion of the animal. Data was collected in October 2016 and May 2019 in a grassy field to a) simulate the environment where the intended free-roaming dogs for study are typically found, and b) provide sufficient ground for the dogs to move continuously. The terrain selected for this component was relatively flat to avoid compromising the gravitational forces acting on the heave axis.

All behavioural tasks (Table 2) were performed continuously for at least 2 min in a stipulated sequence, except for Forage, Eat and Run as these tasks are either too energetically taxing or behaviourally inconsistent. Each dog was tasked to perform the entire repertoire of behaviours at least twice (with consent from their trainer), and approximately 300 minutes of

video-recorded behavioural observations were obtained. The collected accelerometry data was viewed with an acceleration viewer (<http://www.movebank.org>, version 33) and the time-stamps of the collected accelerometer measurements were synchronized and labelled manually to the time-stamps of the video recordings of each performed behavioural task. This complete labelled dataset was subsequently used in the preparation and construction of the predictive RF model.

In preparation of the training dataset, the labelled acceleration measurements were binned into two-second windows (or ‘bursts’) to accomodate at least two full strides for motion-based dog behaviours (i.e., walking, foraging) without influence from un-intentional behavioural transitions. Subsequently, series of summary statistics (mean, min, max, kurtosis, skewness, range, standard deviation; Table 1) for all three axes were applied onto each labelled burst of accelerometer measurements to characterize the accelerometry measurements of the different behavioural tasks and describe the predictor variables used in the RF model.

Table 1: Predictor variables and their statistic labels used for predicting dog behavioural tasks in the random forest models.

Statistics label	Predictor variables	Description
Mean	mean.x mean.y mean.z	The calculated mean of the acceleration measurements within each burst for axes x, y and z.
Min	min.x min.y min.z	The smallest acceleration measurement within each burst for axes x, y and z.
Max	max.x max.y max.z	The largest acceleration measurement within each burst for axes x, y and z.

Statistics label	Predictor variables	Description
Kurtosis	kurt.x kurt.y kurt.z	The relative flatness of the acceleration measurements within each burst for axes x, y and z.
Skew	skew.x skew.y skew.z	The relative skewness of the acceleration measurements within each burst for axes x, y and z.
Range	range.x range.y range.z	The calculated difference between the largest and smallest acceleration measurements of axes x, y, and z.
Standard deviation	sd.x sd.y sd.z	The calculated standard deviations of the acceleration measurements of axes x, y, and z.

The `randomForest` package (Liaw and Wiener 2002) in R was then used to prepare and construct the RF model used for predicting dog behaviours. The labelled dataset was subsampled at a 80% proportion as a training dataset to prepare the RF model, while the remaining 20% was used as a testing and validation dataset. The default number of trees ( $n = 500$ ) were used in the RF model construction as the Out-of-Bag (OOB) error estimates, a method of measuring prediction error rates, was found to be consistently stable (Figure 3). To refine and improve the RF model's predictive capacity, I cross validated the RF models by comparing the prediction accuracies of RF models constructed ( $n = 21$ ; Figure 1) with a range of `mTry` values mirroring the number of described predictor variables (Table 1). Following

which, the best predictive RF model was constructed and the classification error rate and mean prediction accuracy of each behavioural task in the model were tallied and tabled (Table 3).

## 2.2 Accelerometer threshold accelerometer-informed GPS sampling

The wildlife tracking collar is equipped with the capacity to sample GPS relocations dynamically based on the accelerometer measurements collected in real-time. With the use of a selected accelerometer threshold (ACT), researchers are able to selectively intensify the sampling of GPS re-locations only when the animal is moving (hence higher variance in accelerometer readings). This feature not only reduces the energy and memory consumption during the research period, it also allows the researcher to identify movement paths or patterns that are ecologically interesting (e.g., predation or harrassement events).

For this function to work realistically, I used the above constructed RF model to identify the variances of the predicted movement behaviours and selected a variance value (Table 5) for the axis that best represents movement or motion in a dog (i.e., Z axis; the acceleration in the heave axis typified by the rolling gait in dogs). Following which, I ground truth the selected variance by walking four dogs (Mongrel breeds, 4 females, age range 2-7 years old) with the wildlife tracking collar equipped. All dogs were walked continuously for approximately ten minutes, and with the studied dog resting (with the collar on) only in the beginning and end of each walk. The dogs were rested in a non-motion movement behaviour (e.g., lying, sitting or standing) to simulate the resting behaviour of free-roaming dogs in the wild. The GPS and ACC dataset were then retrieved from the collar, and the occurrence of quick- and long-burst gps relocations were plotted sequentially against the variance of the accelerometer measurements collected from all three axes (Figure 4).

### 3 Results

For the purpose of building the training dataset of the RF model, I collected more than 400,000 raw acceleration measurements (per axis) and approximately 400 minutes of video footage from 11 kennel-trained domestic dogs. From the collected video footages, seven different dog behaviours were identified which included sitting, standing, lying, eating, walking, foraging, and running (Table 2). After identifying and cleaning the collected acceleration measurements, approximately 3000 bursts of labelled acceleration data comprising of nearly 1,260,000 transformed acceleration measurements (from the three axes) were used to construct and train the RF model.

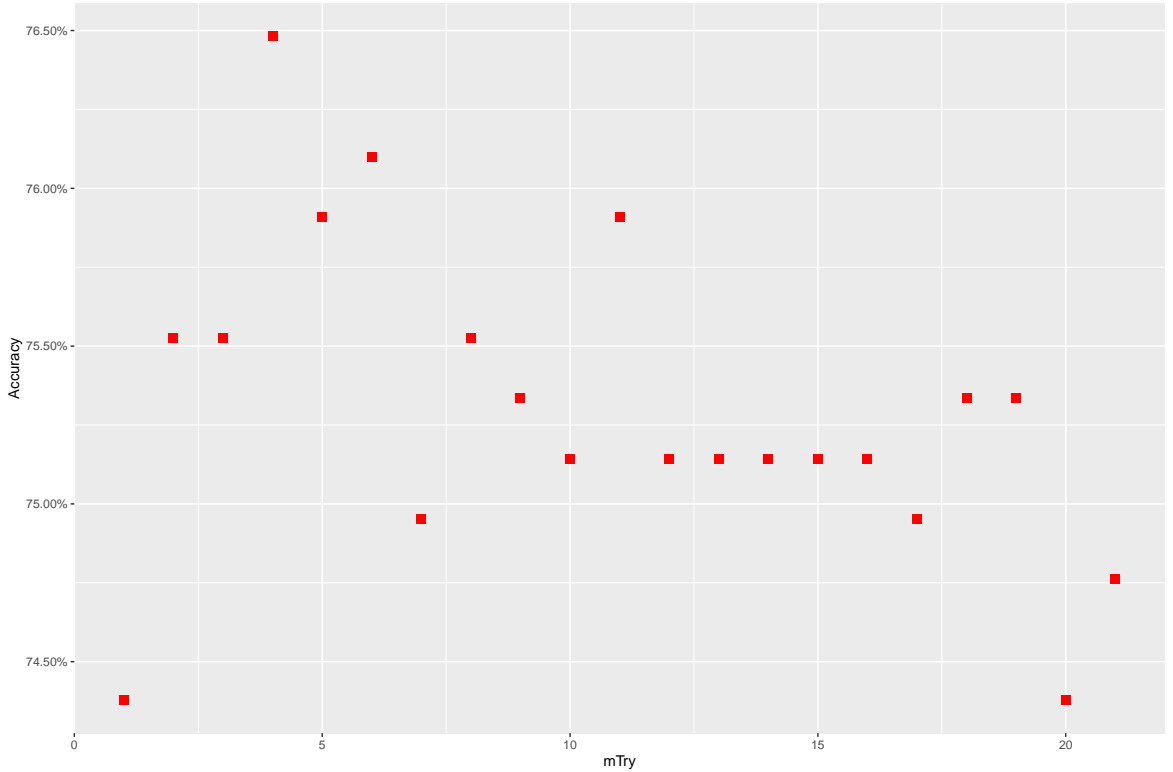


Figure 1: The prediction accuracies of the RF models constructed with a range of mTry values mirroring the number of predictor variables used.

From Figure 1, it was found that the constructed RF model had the highest prediction accuracy when only four predictor variables were available or used when describing the differences between each movement behaviour. At present, the results from Figure 1 found

that the difference in selecting between the best (i.e., 4) and worst (i.e., 1 or 20) performing mTry parameter resulted in an approximate 1% difference in mean prediction accuracy which suggests that the current RF model predictive capabilities, in spite of the large number of predictor variables used, is relatively consistent.

Figure 2 demonstrates the mean decrease in Gini index in the predictor variables used for decision-making in the RF model. Node purity (as represented by the Gini index) refers to how well the trees split the data (i.e., decision-making) and predictor variables with a higher decrease in Gini index represents a purer node that makes has greater importance in the RF model predictive and decision making capacity. In any case, the removal of the top-most predictor variable (hence, highest decrease in Gini index) will result in a RF model with poorer prediction accuracies.

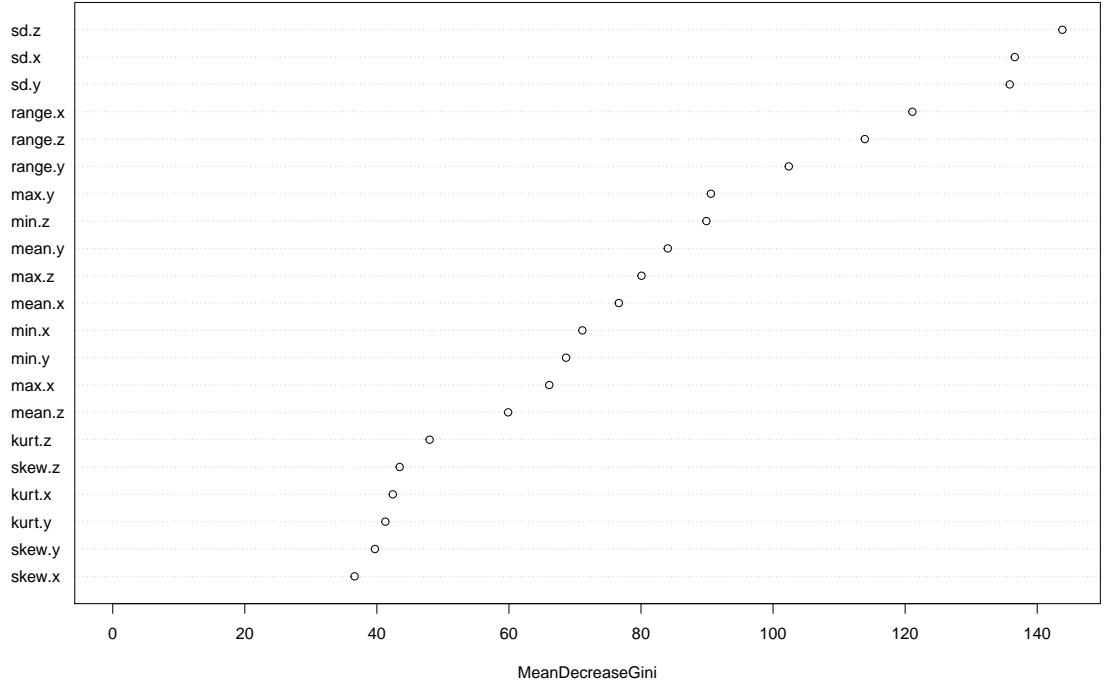


Figure 2: The variable importance of the predictor variables used in the best predictive RF model.

Using the selected mTry value as reference (Figure 1, mTry value of 4), it is suggestive



that the Y axis may have a larger role in differentiating between the different movement behaviours and postures as the important predictor variables used in the RF model included a higher number of statistical transformations (i.e., 2 out of 4) derived from the Y axis (i.e., standard deviation of Y;  $sd.y$ , and range of Y;  $range.y$ ). In addition, the ten most predictor variables in Figure 1 also found the statistical transformations of the Y axis were more prevalent than those from the X and Z axis (e.g., 40%, 30% and 30%; Y, X and Z respectively).

With regards to predictive stability in the constructed RF model, Figure 3: The Out-Of-Bag error estimates for the predictive RF model plotted against the number of trees ( $n$ ) used. found that the Out-of-Bag error rates remained relatively stable after 100 trees were grown. Despite using 500 trees to develop the RF model, the average run-time for predicting on the testing dataset (remaining 20% of the subsampled labelled acceleration dataset; approximately 600 bursts of acceleration measurements) took approximately two minutes. As such, increasing the number of trees used in the RF model will not significantly improve its predictive capacity, hence the trade-off between run-time and predictive power may not be feasible when larger datasets are used in the subsequent chapters.

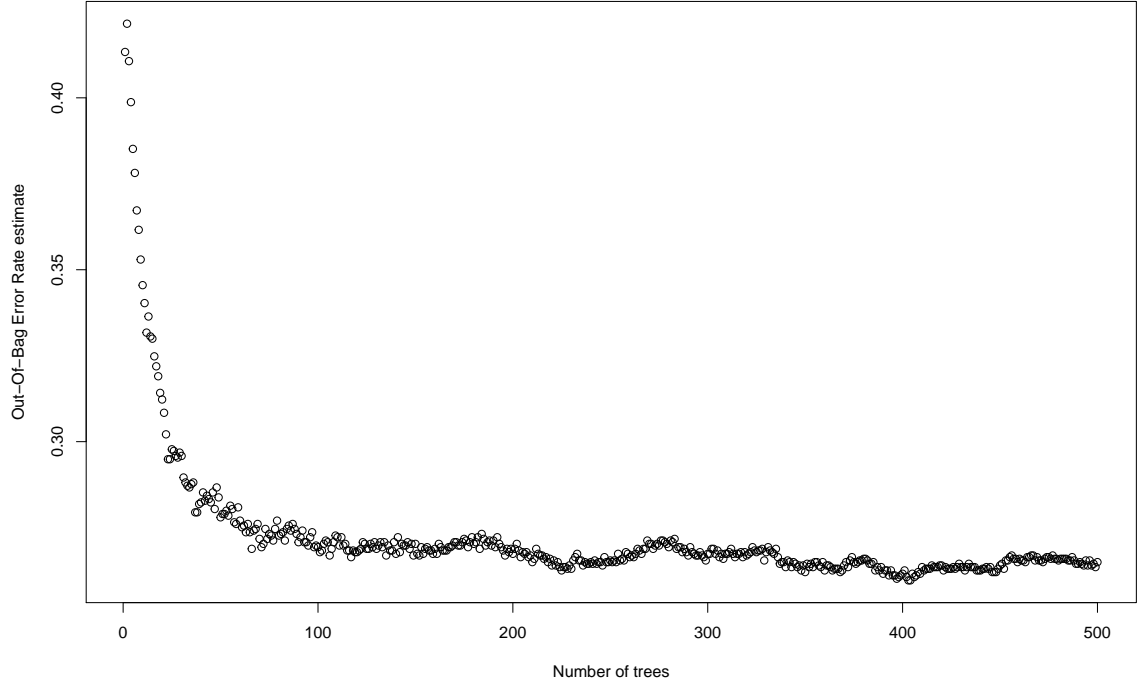


Figure 3: The Out-Of-Bag error estimates for the predictive RF model plotted against the number of trees ( $n$ ) used.

Table 3: The results of the RF model predicting on the training dataset.

	eat	forage	lie	run	sit	stand	walk	accuracy
eat	34	6	13	0	7	2	4	51.52%
forage	1	66	6	5	3	1	44	52.38%
lie	4	10	302	1	67	25	25	69.59%
run	0	2	1	353	0	0	52	86.52%
sit	3	11	87	0	231	14	10	64.89%
stand	2	4	23	2	20	113	17	62.43%
walk	0	9	4	54	7	3	424	84.63%

Looking at the RF model predictive capabilities on the training dataset (i.e., subsampled 80% of the labelled acceleration dataset), the RF model was found to predict movement-based

behaviours (e.g., ‘Run’, ‘Walk’ and ‘Forage’) more accurately than non-movement based behaviours (e.g., ‘Lie’, ‘Sit’, and ‘Stand’). For example, ‘Run’ and ‘Walk’ had comparatively higher prediction accuracies (86.52% and 84.63%, respectively) whereas ‘Sit’, ‘Stand’, and ‘Walk’ were mis-classified twice as much (approximately 34.37%). The poor predictive capabilities on non-movement behaviours was likely attributed to the the RF model inability to differentiate between the different postures held during non-movement behaviour tasks. For example, ‘Lie’ and ‘Sit’ often mis-classified interchangeably (e.g., ‘Sit’ as ‘Lie’, 15.44%; and ‘Lie’ as ‘Sit’, 24.44%) while the mis-classification for ‘Stand’ were evenly distributed between ‘Lie’, ‘Sit’ and ‘Walk’ (e.g., 12.71%, 11.05% and 9.39%, respectively).

Not surprisingly, the RF model had the lowest prediction rates for the ‘Forage’ and ‘Eat’ behaviours (51.52% and 52.38%, respectively). As discussed previously, it is likely that the RF model was unable to accurately predict ‘Forage’ and ‘Eat’ accurately as these behaviours bear strong postural and movement resemblance to ‘Walk’ and ‘Lie’ (e.g., 19.70% and 34.92%, respectively). For example, the ‘Eat’ behaviour represents the dog in-study consuming or drinking from a bowl and the these dogs often assume a lying position in the duration of this behavioural task. Likewise, the ‘Forage’ behaviour can be difficult for the RF model to differentiate from the ‘Walk’ behaviour as the behavioural task for ‘Forage’ was represented by the dog in-study walking with its head angled downwards (Figure 5).

Table 4: The results of the RF model predicting on the testing dataset.

	eat	forage	lie	run	sit	stand	walk	accuracy
eat	4	2	8	0	1	0	2	23.53%
forage	0	15	2	4	1	1	9	46.88%
lie	1	1	79	1	16	7	4	72.48%
run	0	0	0	89	0	0	14	86.41%
sit	0	0	10	2	68	3	7	75.56%
stand	0	0	4	0	6	28	8	60.87%
walk	0	5	1	5	2	1	112	88.89%

As with most RF models, overfitting (i.e., is a common occurrence when the constructed RF model is tasked to predict on the remaining subsampled testing dataset. Even though overall prediction accuracy between the training and testing datasets differed slightly (73.75% and 75.53%, respectively), the prediction accuracies of individual behaviours in the testing dataset may be significantly worse (Table 4). For example, the mis-classification rates for ‘Eat’ and ‘Forage’ increased by 27.99% and 5.51% respectively yet the prediction accuracies for ‘Lie’ and ‘Sit’ improved slightly ( $< 10\%$ ). Nevertheless, it is likely that these discrepancies in the prediction accuracies of individual behaviours were caused by the reduced sampling size of the testing dataset as the overall prediction accuracy for both datasets were still found to be comparable.

Table 5: The mean variance of acceleromenter measurements for predicted movement behaviours in dogs.

behaviour	mean.X	mean.Y	mean.Z
lie	4382.160	5236.047	3097.256
sit	5966.968	3639.992	7489.651
stand	10536.696	15618.678	7963.085
forage	26128.293	23022.594	22260.306
walk	56863.595	34096.077	42351.569
run	335325.959	152629.015	326456.067

Using the variances calculated from the labelled acceleration measurements (Table 5), I sampled a range of variance values between ‘Stand’ and ‘Forage’ across the three axes to identify a variance threshold value (VTV) that could consistently initiate the collection of short-bursts GPS re-locations in the wildlife tracking collar. To do this, I used the `rle()` function from base R to examine and count the frequency of which a feasible VTV (from the range of feasible variance values) is met or exceeded by the variance of movement-based labelled acceleration measurements across the three axes. The VTV that determines the highest frequency of met or exceeded variance values was then used to program the wildlife

tracking collar for the dog walking experiment.

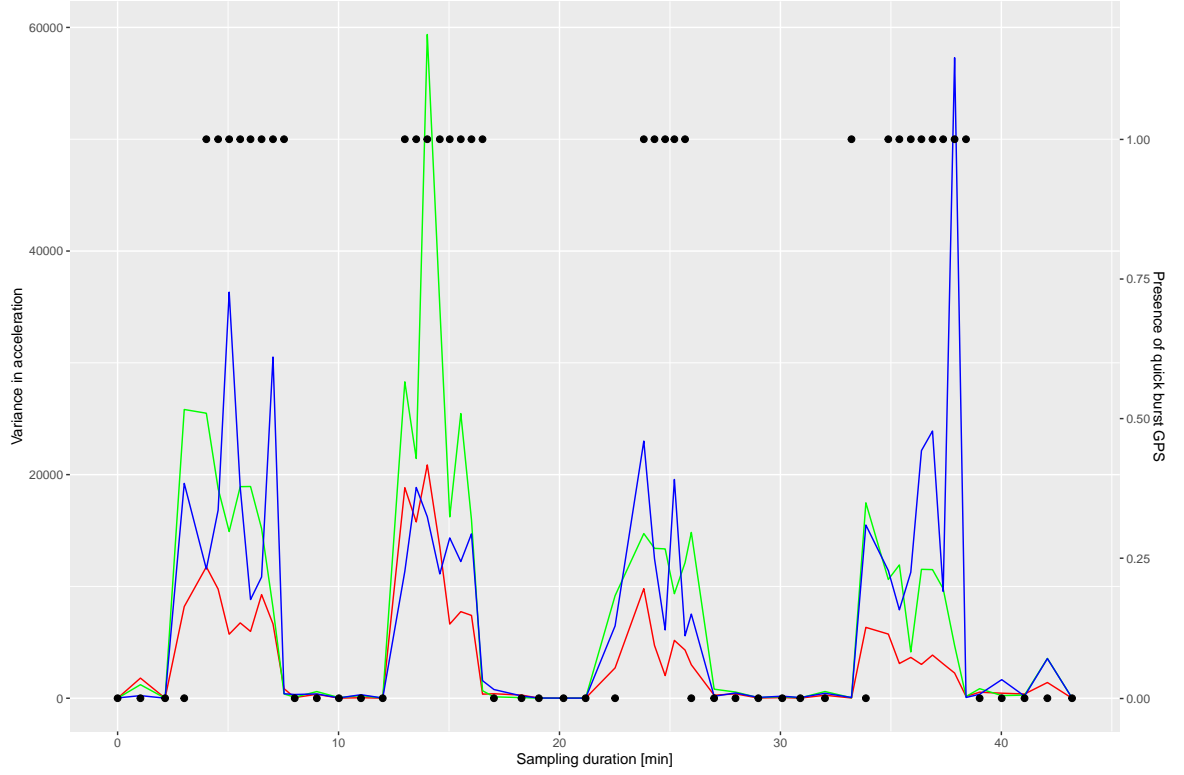


Figure 4: The sampling of GPS re-locations based with an accelerometer threshold of 10,000 in the Z axis. The red, green and blue lines represents variance in acceleration measurements collected from the X, Y and Z axis.

With the VTV identified and programmed into the collar, Figure 4 found that the collection short-burst GPS re-locations was most consistent when a VTV of 10,000 is programmed onto the Z-axis. For example, the collection of short-term GPS re-location is only triggered if the variance of collected acceleration measurements from the Z axis exceeds 10,000 repeatedly. That being said, movement is typically associated with acceleration in the surge axis (X-axis; red line) yet the heave axis (Z axis) appears to be more sensitive to the up-and-down acceleration forces exerted by rolling gait of a walking dog. Seemingly, having access to the ground truthed information on the acceleration forces of a moving dog allows one to program a suitable VTV to examine biologically and ecologically interesting movement patterns in the subsequent chapters.

## 4 Discussion

- no real need to prune the RF model even with 21 p.v as analysis ran relatively quick and n trees used also showed consistency.
- average run time was less than a minute

## 5 Conclusion

## 6 References

### Supplementary codes

Liaw, Andy, and Matthew Wiener. 2002. “Classification and Regression by randomForest.”  
*R News* 2 (3): 6.