

Estimation of Mean Fixed Frequency Biases: the significance of BFO Channels

BSM/GH December 2014, MH370 Independent Group

Background

The focus of this report is on the estimation of mean fixed frequency biases (FFBs), for applications to MH370 flight path analysis. This report investigates the extent to which the mean frequency bias is contingent on the number and type of channel associated with burst frequency offsets (BFOs). The approach adopted will be primarily via descriptive statistics; but a degree of hypothesis testing will be unavoidable. It is found that certain combinations of both channel type and number are significant covariates for the mean FFB, meriting consideration in improved modelling of MH370 flight paths. While investigation of the causes of factors of this kind are outside the scope of this paper, additional evidence on causation would be of interest, particularly if it can assist the air accident investigation.

This report represents a continuation of the analysis provided in [1], which was restricted to conditions when MH370 was on the ground. The introduction of channel-specific biases was identified and wider validation of their values was warranted. The current report includes BFO data for the early cruise phase of the flight and so represents a contribution to this effort. The sample size is increased from 131 ground observations to 263 total observations, which effectively doubles the sample size. Whilst the cruise data is somewhat more variable than the ground data, this increase in the sample size is sufficient to provide narrower confidence limits to our estimates.

The calculations reported here are given in [2]. One of the interesting findings is that a simple split by channel type (R, and T) is not the most effective way of identifying the inter-channel differences. A combination of channel type with channel number is needed for a more effective BFO analysis and in the assessment of competing flight path models.

Introduction

We estimate mean FFBs using ground and early cruise BFO data, and the PAR5 satellite model [3]

The sample values of the FFB terms were calculated using [4]:

$$FFB = BFO - \delta_{afc+sat} - \Delta F_{up} - \Delta F_{down} - \delta_{comp}$$

The FFBs are segmented by channel number (4, 8, 10, 11, and 12) and type (R, T). In the current data set channel 8 is the only channel number that appears as both type R and type T. Figure 1 shows the resulting FFBs against UTC time during the ground and early cruise, based on the PAR5 satellite positions and velocities and a model for the aeroplane position and velocity.

Inspection of figure 1 suggests that only three FFB value for blue channels R4 and R11 exceed the lowest value for the other channels (151 Hz). This is good *prima facie* evidence of channels R4 and R11 differing from the others. This topic will be subject to statistical testing, given below.

Table 1 shows, for each combination of channel number and type, the mean FFB, the standard deviation of the FFB and standard error of the estimated mean (the standard deviation divided by the square root of the count).

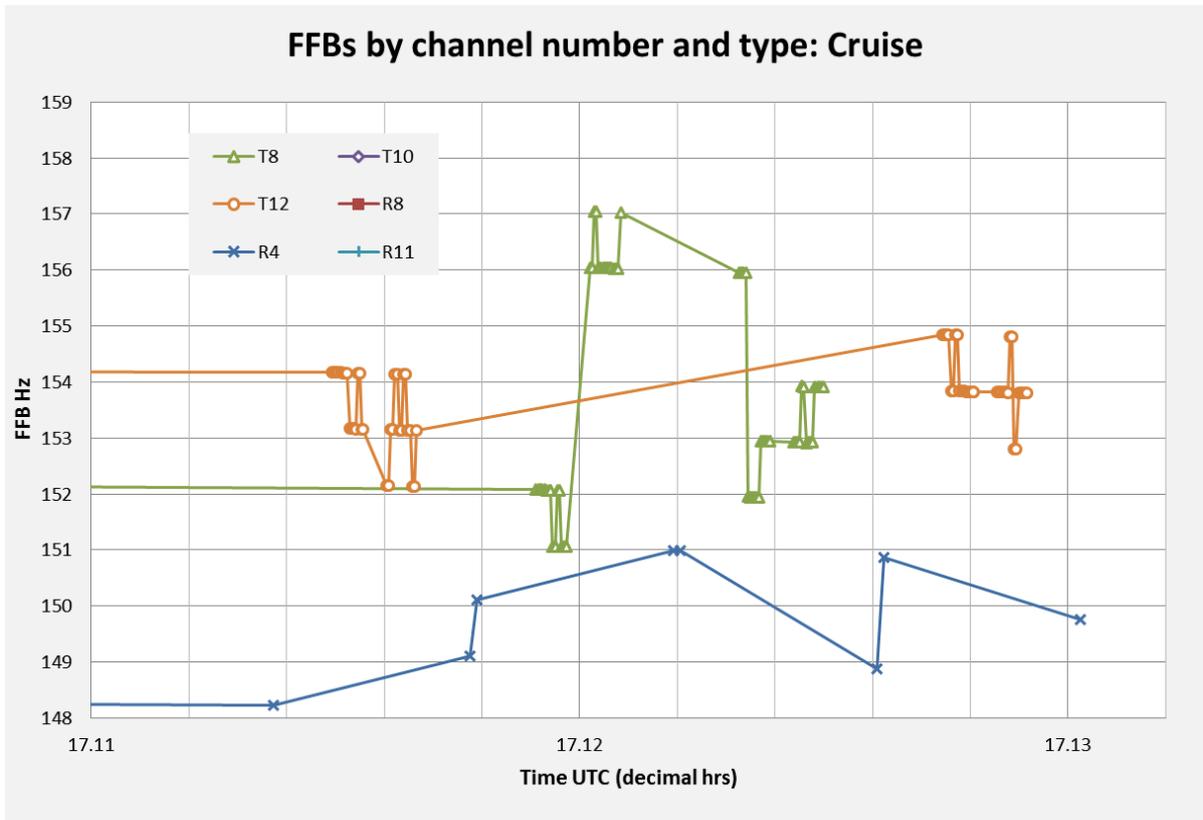
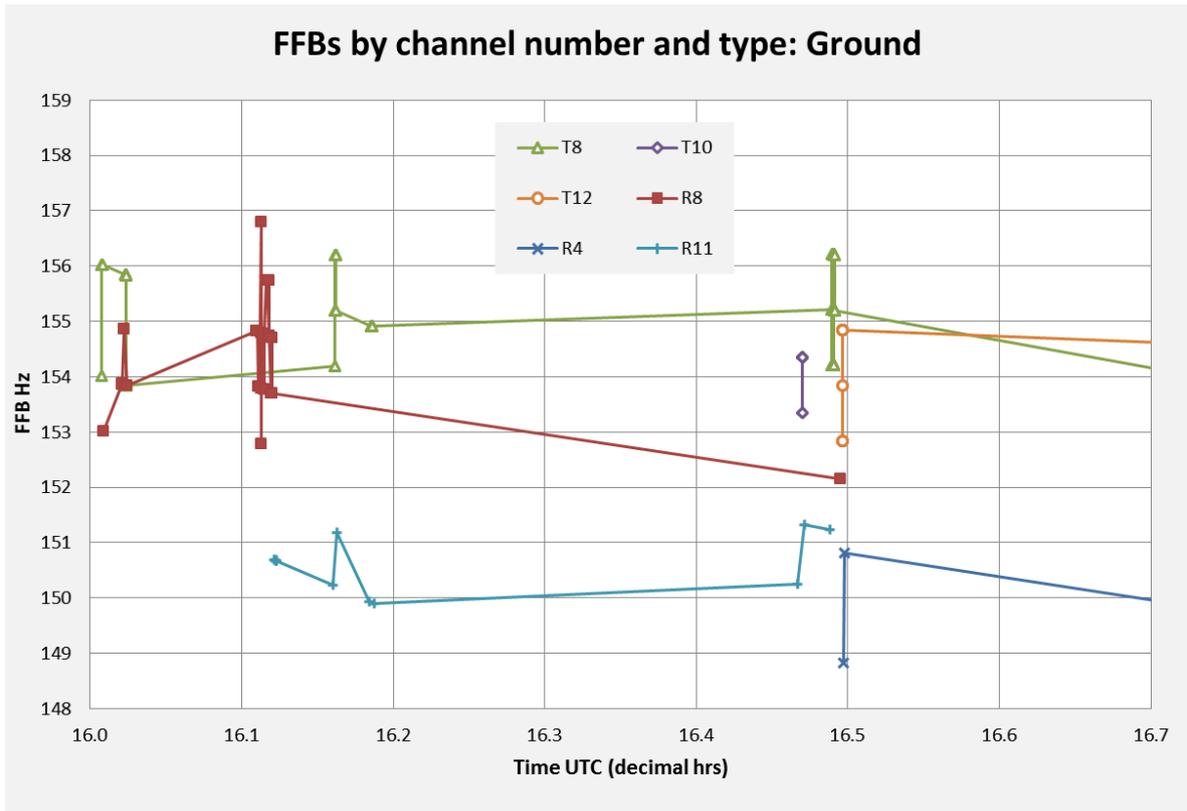


Figure 1: FFB variation over time by channel

Channel unit and type	Mean FFB (Hz)	StDev (Hz)	Count	StErr (Hz)
R4	149.86	1.05	10	0.33
R8	154.45	0.96	34	0.17
T8	154.35	1.63	120	0.15
T10	154.17	0.40	11	0.12
R11	150.62	0.48	12	0.14
T12	153.82	0.74	76	0.09
All	153.86	1.67	263	0.10

Table 1: Initial FFB Descriptive Statistics by Channel

Stage 1 Tests

We now turn to an issue requiring some basic testing of hypotheses. The first question to be addressed will be how do estimated FFB means for the BFOs associated with R4 and R11 differ from each other, and is this difference statistically significant? From the numbers given in table 1 we would clearly expect to obtain a negative result. We will however proceed, for the purpose of illustrating the tests. We apply both a non-parametric and a parametric test. The non-parametric is the Mann-Whitney U -test [5], [6] and the parametric is the Welch-Aspin test [7]. The Welch-Aspin test accommodates both unequal group sizes and unequal group variances. The Mann-Whitney test only requires information about rank and is distribution-free.

Both tests yield a test statistic that is a ratio of a measure of difference to a measure of its standard error. In both tests we will refer to this statistic as its Z score. We may consider a score in excess of 3, in both tests, to be significant. A score in excess of 5, (the 'five sigma level') in both tests, is taken to be highly significant.

Table 2 contains the results for R4 vs R11. The non-parametric Z score is 1.45 and the parametric score is 2.13. To reject the null hypothesis both scores were required to exceed 3, so we have not met this condition.

Non-parametric		Parametric	
N_1	10	R4 statistics	
N_2	12	Count	10
R_1	93	Mean	149.86 Hz
U_1	82	Stdev.s	1.05 Hz
U_2	38	St error	0.33 Hz
U	38	R11 statistics	
m_U	60	Count	12
σ_U	15.2	Mean	150.62 Hz
		Stdev.s	0.48 Hz
		St error	0.14 Hz
		Offset	0.76 Hz
Z (with ties)	1.45	Z	2.13

Table 2: Testing for differences between estimated means: Channel R4 v channel R11

For the purpose of our current estimation task we treat R4 and R11 as equivalent channels. The distributions of the estimated means are presented stylistically below, as 'spike plots'. These plots have been designed to give a visual indication of how well determined the FFB means are, given the input data being used for estimation. The x-axis denotes the FFB scale in Hz, with the peaks set at the estimated mean values. The width of each spike is two standard errors, one on each side of the estimated mean. The areas under the spikes are designed to be

approximately equal, with the height of each spike set at the reciprocal of the corresponding standard error. Spike plots are given both before and after merging R4 with R11.

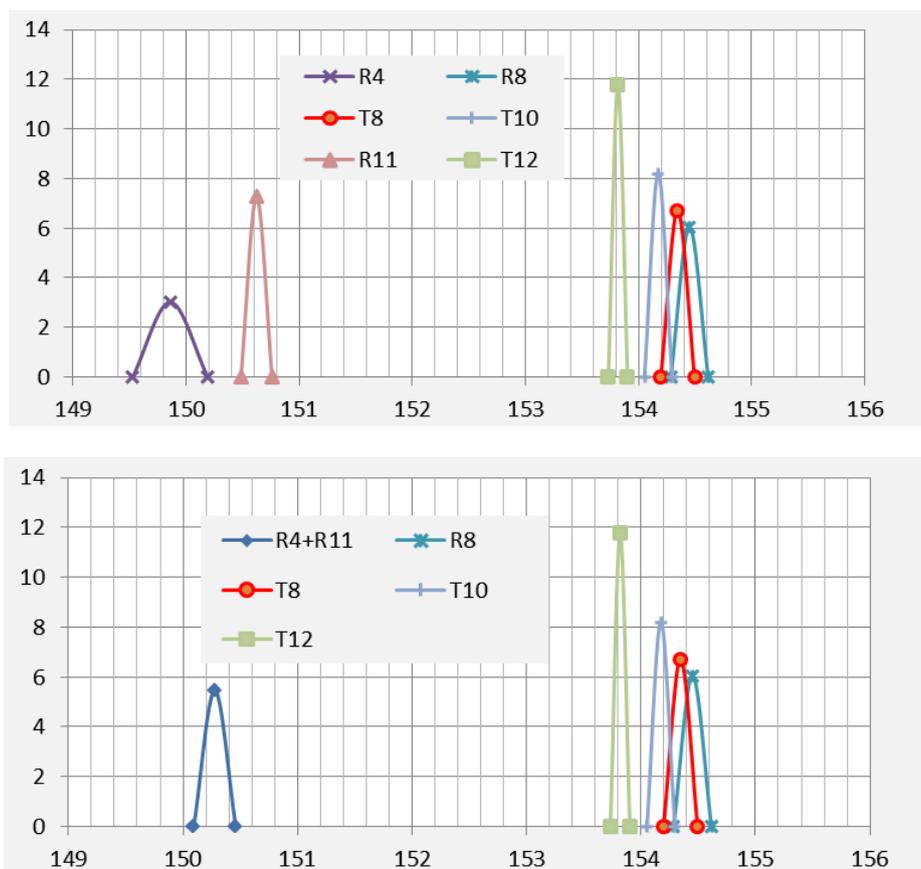


Figure 2: Spike plots by channel, before and after merging R4 with R11

The next test of interest is the statistical significance of the differences between the estimated means for the R4+R11 combination and the combinations of the other channels: T10+T12+R8+T8.

The results are given in table 3. The non-parametric Z value is in excess of 7 and the parametric Z value exceeds 19. This is taken as a highly significant difference between the estimated FFB means for the R4+R11 combination and the combination of the remaining channels. The relative offset has a value of -3.9 Hz, i.e. the mean FFB for R4+R11 is nearly 4 Hz lower than the mean FFB for the other channels combined. The Z values for both tests exceeded the five sigma threshold, leading to a strong rejection of the null hypothesis so that, in the absence of a true systematic difference, the reported results should typically appear no more than once in (at least) 3.5 million trials.

Table 4 provides the descriptive statistics for the resulting two-way split between channels, corresponding to the greatest two-way differentiation available from the data being used. While this may be incorporated in the analysis of flight paths, it is of interest to investigate whether the data would support a further level of discrimination.

Non-parametric		Parametric	
N_1	241	T10+T12+R8+T8 statistics	
N_2	22	Count	241
R_1	34445	Mean	154.19 Hz
U_1	18	Stdev.s	1.30 Hz
U_2	5284	St error	0.08 Hz
U	18	R4+R11 statistics	
		Count	22
m_U	2651	Mean	150.28 Hz
σ_U	341.5	Stdev.s	0.86 Hz
		St error	0.18 Hz
		Offset	-3.91 Hz
Z (with ties)	7.71	Z	-19.39

Table 3: Testing for differences between channel R4+R11 vs all other channels combined

Channel unit and type	Mean FFB (Hz)	StDev (Hz)	Count	StErr (Hz)
R4+R11	150.28	0.86	22	0.18
T10+T12+R8+T8	154.19	1.30	241	0.08

Table 4: FFB Descriptive Statistics after Stage 1

Now suppose that instead of splitting the channels in this way we had simply split them according to type (R, T). What would we have obtained? Table 3 would have been replaced by table 3a.

Non-parametric		Parametric	
N_1	56	R statistics	
N_2	207	Count	56
R_1	5364	Mean	152.81 Hz
U_1	7824	Stdev.s	2.25 Hz
U_2	3768	St error	0.30 Hz
U	3768	T statistics	
		Count	207
m_U	5796	Mean	154.14 Hz
σ_U	505.0	Stdev.s	1.34 Hz
		St error	0.09 Hz
		Offset	1.33 Hz
Z (with ties)	-4.02	Z	4.23

Table 3a: Testing for differences between channel R and T channels

It can be noted that the result of this test produces a significant result, but it is not highly significant (at the 5 sigma level). Further, the relative offset between the T and R channels is only 1.3 Hz. The standard deviation for the grouping of R channels has increased substantially, resulting in wider confidence bands for the mean FFB. On the current data set, the R/T segmentation would be inferior to that given in tables 3 and 4 and was not retained for the next stage in the analysis.

Stage 2 Tests

We now take a closer look at the UTC time profile of the FFBs for the four unmerged channels: T8, T10, T12 and R8.

From table 1 it looks like T12 might have a slightly lower mean FFB than the three remaining groups, T8, T10 and R8. If we were to merge the latter three we would arrive at the grouping of channels depicted in figure 4.

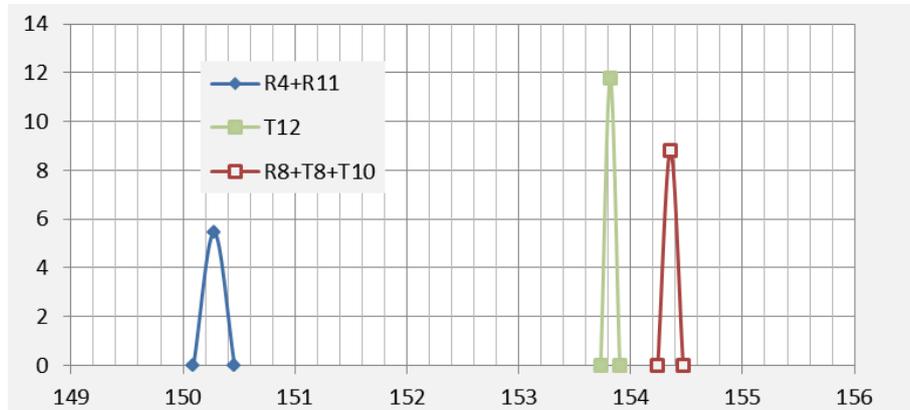


Figure 4: Spike plots by channel combination, after merging R8, T8 and T10.

How does the estimated mean FFB for T12 differ from the estimated mean FFB for the R8+T8+T10 combination? The results of the two tests are given in table 5. The Z values for both tests exceeded the three sigma threshold, indicating that, in the absence of a systematic difference, the null hypothesis would be incorrectly rejected with an (estimated) expected frequency of less than one in 300. The result is considered significant but not highly significant. The relative offset is 0.54 Hz, so that the mean FFB for T12 is approximately one half hertz smaller than the others.

Non-parametric		Parametric	
N_1	165	T12 statistics	
N_2	76	Count	76
R_1	21916	Mean	153.82 Hz
U_1	4319	Stdev.s	0.74 Hz
U_2	8221	St error	0.09 Hz
U	4319	R8+T8+T10 statistics	
m_U	6270	Count	165
σ_U	502.9	Mean	154.36 Hz
		Stdev.s	1.46 Hz
		St error	0.11 Hz
		Offset	0.54 Hz
Z (with ties)	3.88	Z	3.79

Table 5: Testing for differences between mean FFBs for channel T12 vs channel combination R8+T8+T10

The descriptive statistics for the resulting channel combinations are given table 6.

Channel unit and type	Mean FFB (Hz)	StDev (Hz)	Count	StErr (Hz)
R4+R11	150.28	0.86	22	0.18
T12	153.82	0.74	76	0.09
R8+T8+T10	154.36	1.46	165	0.11

Table 6: FFB Descriptive Statistics after Stage 2.

In flight path analysis the principal statistics that are used are the means and standard deviations for the channel number and type associated with the BFO applicable to the phase of flight under investigation. These statistics are depicted in figure 6, illustrating the detailed distinctions obtained from both the stage 1 and 2 analyses.

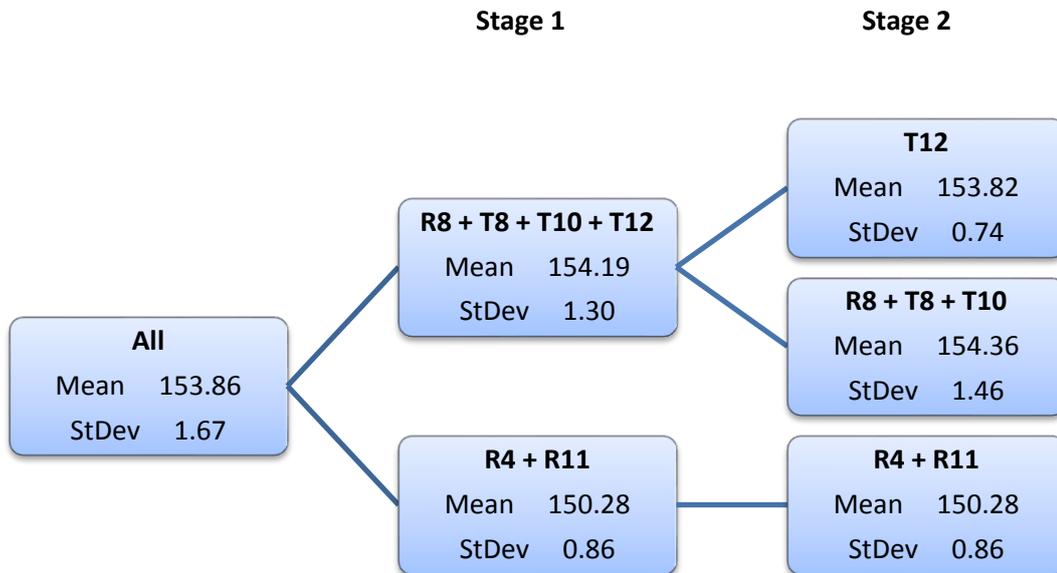


Figure 6: FFB means and standard deviations for identified channel groupings

Figure 6 shows how an all-channel mean of 153.86 Hz first becomes split into two groups, with a -3.9 Hz difference between R4 + R11 and the remaining channels. The remainder then are split into two further groups, with a -0.54 Hz difference between T12 and the other channels.

Discussion and Conclusion

The findings from the stage 1 analysis appear to be particularly robust and support the broad findings of the ground based analysis, at a high level of statistical significance. Alternative forms of stage 1 segmentation, such as a simple split by channel type, did not provide the same level of discrimination. The results from the stage 2 analysis are currently tentative, and the level of discrimination that it provides is modest. Pending a deeper understanding of these issues, it is recommended that only the results from stage 1 should be used in flight path analysis.

References

- [1] FFB_Estimates_Ground_V3b.docx
- [2] BFOFFB_GHBSM05_MWrev4.xlsx
- [3] Rydberg H (2014) <http://bitmath.org/mh370/satellite-par5-ecef.txt.gz> MH370 Independent Group.
- [4] <http://www.atsb.gov.au/mh370/mh370-definition-of-underwater-search-areas.aspx>
- [5] http://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test
- [6] <http://projecteuclid.org/euclid.aoms/1177730491>
- [7] http://en.wikipedia.org/wiki/Welch%27s_t_test