ATLANTIC-WIDE RESEARCH PROGRAMME ON BLUEFIN TUNA (ICCAT GBYP – PHASE 5 - 2015)

ELABORATION OF DATA FROM THE AERIAL SURVEYS ON SPAWNING AGGREGATIONS

Report

26 February 2016

Ana Cañadas & José Antonio Vázquez

Alnilam Research and Conservation Ltd

Pradillos 29, 28491 Navacerrada, Madrid, Spain

Background

The objectives of the comprehensive ICCAT Atlantic-Wide Research Programme on Bluefin Tuna (GBYP) are to improve basic data collection and our understanding of key biological and ecological processes and to develop a robust scientific management framework.

An important element of this programme is to develop fisheries independent indices of population abundance. Therefore in 2010 and 2011 aerial surveys have been conducted in the Mediterranean on the selected spawning grounds. An extended survey was carried out in 2013 and 2015.

The purpose of this work is to elaborate the Aerial Survey data, collected under Phase 5 of the GBYP and to provide a comprehensive analysis of the results of all aerial surveys conducted so far under the framework of the GBYP.

In 2010 an analysis of the aerial survey was conducted and this included a power analysis that evaluated the ability of the survey to detect population trends in the East Atlantic and Mediterranean bluefin recovery plan. This original analysis was based on data from a single year and then it was repeated using 2011 data and then reassessed with a further analysis in GBYP Phase 3. However, inter-annual variation (e.g. due to environmental variation and changes in population distribution) in abundance levels within areas will result in uncertainty in abundance estimates to be underestimated and the power of the survey to detect recovery to be overestimated. Despite many operational and logistic difficulties and problems, data have been collected in 2013 and 2015 in much more extended areas.

Objectives for February 2016

Carry out an in-depth analyses of the collected data to assess the reliably and consistency with which the survey protocols have been implemented within years among the different companies and airplanes. Provide an evaluation of how any consistencies affect the robustness and reliability of the overall results. This should include analyses of the inclinometer versus the GPS derived estimates of the location of a school from the track-line, analyses of the GPS recorded flight flown compared to the survey designed track-line and further analyses of the GPS recorded flight path to see whether protocols for closing on schools were reliably followed.

Analyses of individual biases by spotter

- Re-analysis of all sighting data collected in all surveys in 2015 by each spotter, comparing two different number estimates of the same BFT school when available, trying to define possible individual different capacities or biases.
- Separate professional spotters & pilots from scientific spotters.
- Define individual biases and CVs, to be used as correcting factors if sufficient data are available.

- For the data collected on independent school size estimates by spotters on the same plane, provide estimates of the variance and biases between them. Evaluate how consistent these are among spotters on different planes.

Introducing correction factors into the estimates

- Assess the different number of sightings at each distance category between aircraft with bubble windows and aircraft without bubble windows, possibly identifying correction factors to be applied for previous surveys when bubble windows were not mandatory.
- Use data provided by mini-PATs for Bluefin tuna passing through the areas sampled by the aerial survey in the same period of time and year, in order to assess an average time at the surface (sighting corrector factor) for each internal area, if sufficient data are available.
- Assess the bias induced by some of the most relevant environmental factors (i.e.: wind <2 Beaufort, >2 Beaufort, glare).
- When average estimates are not available for one internal area/year, evaluate if estimates from other areas or years may provide a means to filling the gaps for missing years. This should take into account the information on the variability in across years and areas and the additional variance that such interpolation/extrapolations introduce into the estimate.
- Re-assess all estimates according to these correction factors.
- Possibly introduce the individual spotter correction factor for further re-assessing the estimates by area.

Provide a comprehensive analysis of the survey results across all years to provide an evaluation of the required areas to be surveyed and survey effort required to provide a useful index of relative abundance for stock assessment purposes[1] taking into account the additional variance (factors not able to be directly included within any single year estimate) due to variability in distribution of the animals among year across areas, the timing of the survey relative to variability in spawning, differences in g(0) among planes; accounted variance in estimates of school size and perpendicular distance measurements, and problems due to the survey being done in closing mode (i.e. issues of secondary sightings). Provide an evaluation of the likely achievable actual CV across years for surveys with different spatial coverage and overall survey effort.

Provide comments and suggestions on possible improvements to the survey design, protocols and implementation particularly with respect to issues of calibration of across planes and spotters with respect to g(0) and school size estimates.

[1] E.G. an index of relative abundance with a CV less than 30-40% across years and free from temporal biases

I. Assessment of the reliably and consistency with which the survey protocols have been implemented

In-depth analyses of the collected data to assess the reliably and consistency with which the survey protocols have been implemented within years among the different companies and airplanes. Evaluation of how any inconsistency affect the robustness and reliability of the overall results.

I.1 Analyses of the inclinometer versus the GPS derived estimates of the location of a school from the track-line

Data organization

It has been extremely complicated and time consuming process to obtain and "clean" the GPS data from different years and areas. In 2015 GPS data from all areas except B were available, whereas in 2013 and 2011 only GPS data from areas A and B and A-inside and C-inside were available respectively. It was not possible to obtain any GPS data from year 2010. The main problems found during data organization were:

- In some data sets it was not possible to be sure about the time used in the GPS, local or UTC, mainly in the first years.
- There were some GPS data collection errors in some areas. For an unknown reason[2] every random period of time the GPS signal seems to be lost and give latitude and longitude values very far away from previous ones. Because of this it was necessary to review each data set to detect these errors and delete them from the GPS data set.
- It was necessary to check the latitude and longitude format. While in some GPS data the latitude and longitude were collected as decimal degrees, others were recorded as degrees and decimal minutes. The later needed to be transformed into decimal degrees. It is very important that for futures surveys all airplanes use the same lat/lon format to collect position data in order to make the data organization easier.
- In many occasions it was difficult or even impossible to identify the circles corresponding to a specific BFT sighting.

Some examples of these difficulties are shown below:

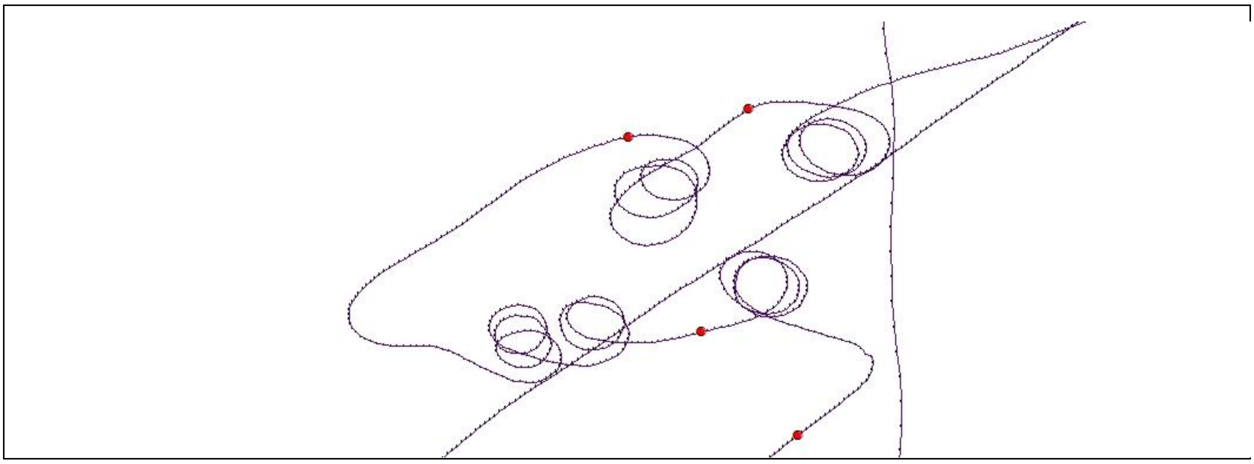


Figure 1. When various consecutive sightings were detected after a first one.

[2] It is well-known that in some cases and Mediterranean areas, during military activities or international crisis, the GPS signal may provide unreliable locations or even serious biases.

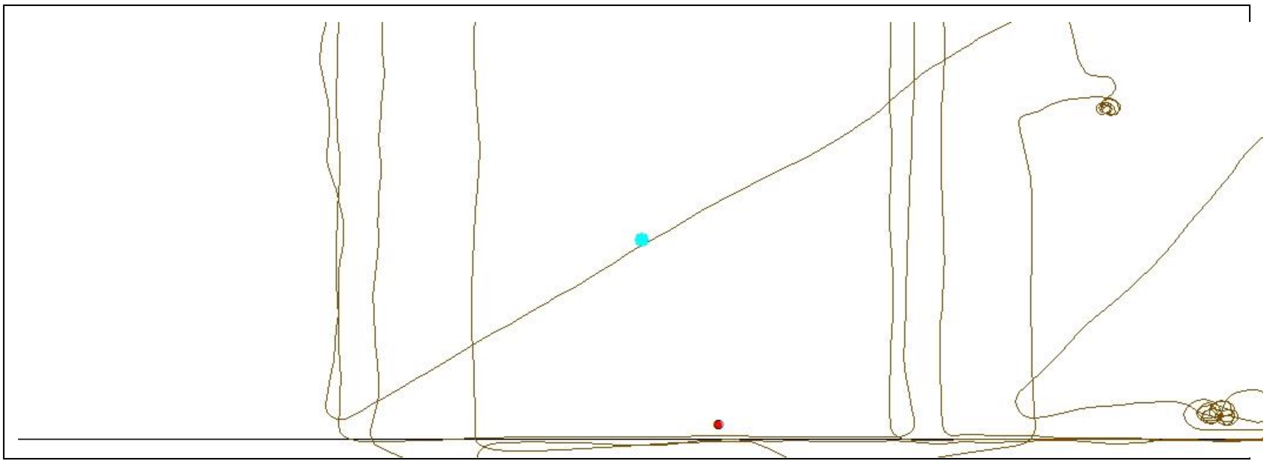


Figure 2. When there is no evidence of circling at all.

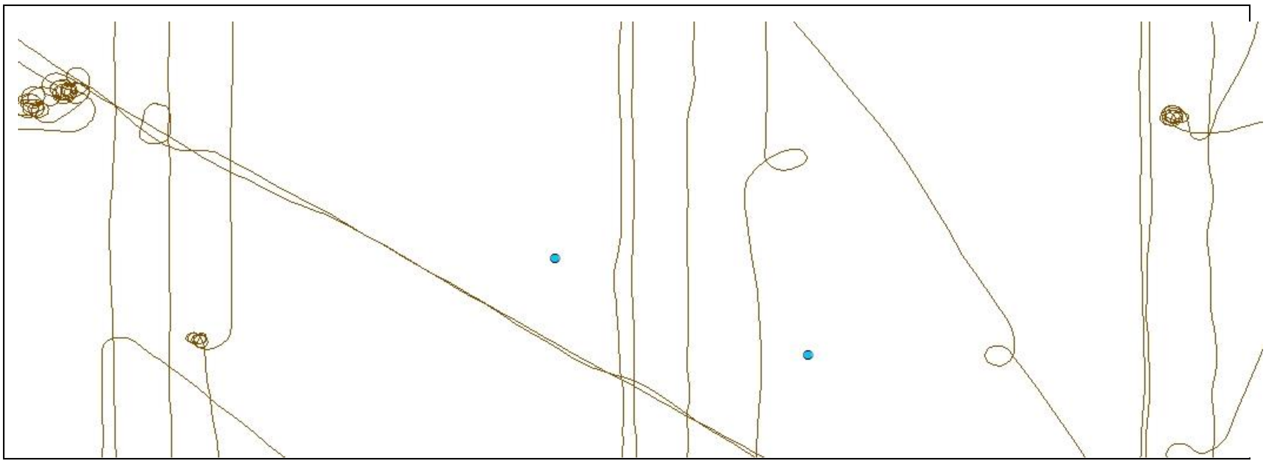
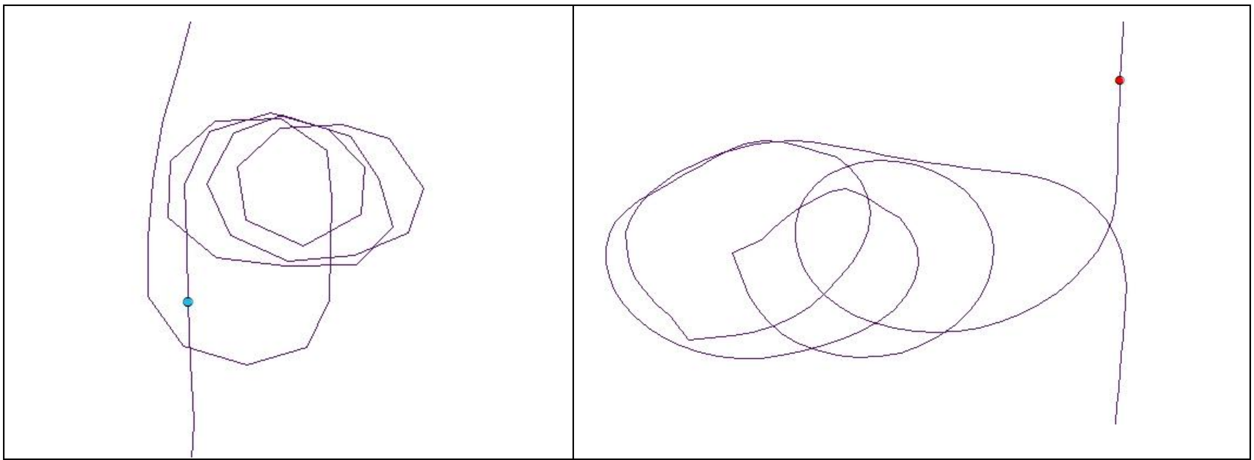


Figure 3. When the GPS data corresponding to the day when sightings were registered do not exist or were not provided.



Concentric circles. Animals stationary	No concentric circles. Animals moving

Figure 4. When animals are moving quickly so the circles are displaced spatially; this makes not easy to identify the right position.

Data analysis

ArcGIS software was used to estimate perpendicular distances based on spatial measurements. Each GPS data set was plotted on a map covering the study area together with the BFT sightings registered in that year/area. Perpendicular distances were estimated measuring the length between the centre of the contiguous circles made by the airplane while flying over the BFT to obtain school size and weight estimates, and the direction of flight in a straight line.

As explained before, not in all BFT sightings there were GPS data available. It was not possible to estimate perpendicular distances in 74 out of 91 observations in 2011, in 127 out of 161 in 2013 and in 17 out of 87 in 2015.

Table 1 shows the number of BFT sightings where it was possible to calculate the difference between the perpendicular distance estimates by clinometer and those calculated using GIS software, and the average differences by year and company (i.e. the average of all the differences per year and company), both considering the original values (negative or positive) or absolute values (not taking into account whether the differences are positive or negative). The negative values under "Original values" mean that the perpendicular distance estimated from the angle is smaller than that estimated with the GIS software.

Table 1. Number of BFT sightings where it was possible to calculate the difference between the perpendicular distance estimates by clinometer and those calculated using GIS software, and average differences (in meters) calculated by year and company. Both original values (negative or positive) and absolute values (not taking into account whether the differences are positive or negative) are shown.

	Original values						Absolute values			
	N	Mean	Min	Max	SD	CV (%)	N	Mean	SD	CV (%)
2011										
Action-Air										
Air-Med	5	-197	-1746	603	924	470	5	622	647	104
Perigord	1	-495	-495	-495			1	495		
Unimar	10	-522	-4043	588	1374	263	10	765	1241	162
Total	**16**	**-418**	**-4043**	**603**	**1176**	**281**	**16**	**703**	**1021**	**145**
2013										
Action-Air										
Air-Med	34	-476	-8016	3670	2021	425	34	1249	1647	132
Perigord										
Unimar										
Total	**34**	**-476**	**-8016**	**3670**	**2021**	**425**	**34**	**1249**	**1647**	**132**
2015										
Action-Air	20	-763	-7835	435	1830	240	20	831	1798	216
Air-Med	37	119	-4305	3078	1217	1020	37	767	945	123
Perigord										
Unimar	12	69	-2320	1198	893	1296	12	569	670	118
Total	**69**	**-145**	**-7835**	**3078**	**1417**	**976**	**69**	**751**	**1207**	**161**
TOTAL										
Action-Air	20	-763	-7835	435	1830	240	20	831	1798	216
Air-Med	76	-168	-8016	3670	1624	968	76	973	1307	134
Perigord	1	-495	-495	-495			1	495		
Unimar	22	-199	-4043	1198	1148	575	22	658	951	145
Total	**119**	**-276**	**-8016**	**3670**	**1581**	**572**	**119**	**887**	**1336**	**151**

Differences are very unequal between companies and years ranging from 763m larger, in average with GIS than with clinometer for Action-Air in 2015 to 119 m shorter, in average with GIS than with clinometer for Perigord in 2011; or between 495 m to 1249 m taking absolute values. This very large variability among years and companies, even among sightings of the same company for a certain year, makes it impossible to get a correction factor.

Figure 5 shows scattergrams, and their corresponding linear regression, of perpendicular distances estimated by GIS against perpendicular distances estimated by ANGLE (inclinometer).

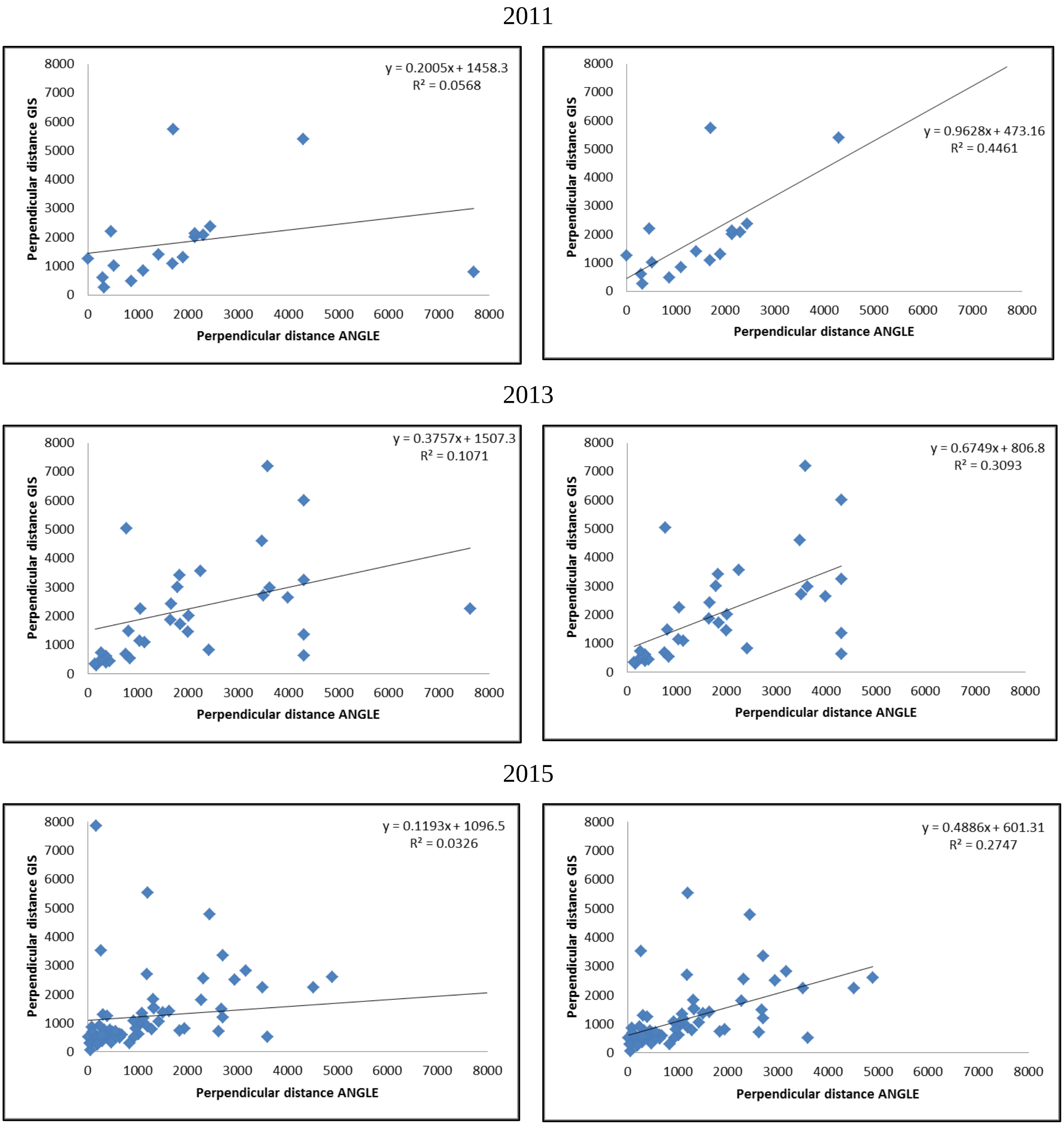


Figure 5. Scattergrams and linear regression of perpendicular distances estimated by GIS against perpendicular distances estimated by clinometer (ANGLE). Right figures show the outliers and left figures have removed the outliers. Upper figures correspond to 2011, medium figures correspond to 2013 and lower figures correspond to 2015.

In 2011, the outlier located in the right part of the graph corresponds to a BFT sighting where perpendicular distance estimated with clinometer was 7700m, but after identifying it in GIS there are no circles around it, except one at 796m (Figure 6). Therefore this observation was deleted from the comparison. By doing so, the relationship between GIS-ANGLE perpendicular distances in the scattergram changes drastically (Figure 5).

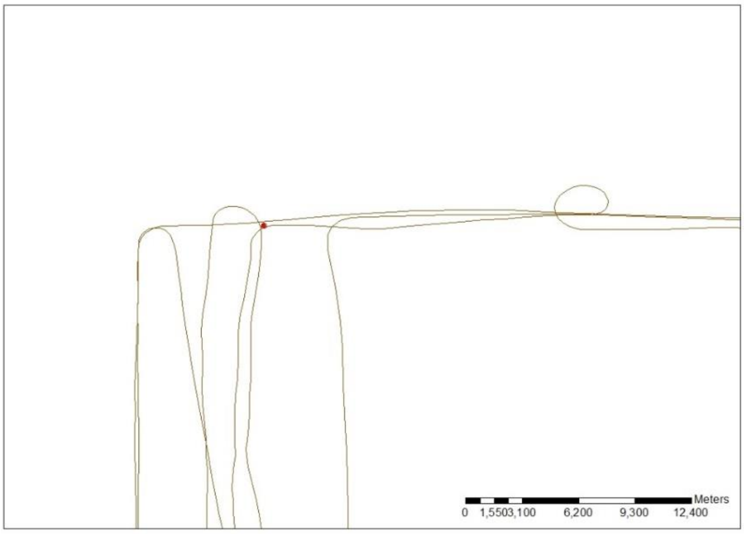


Figure 6. BFT sighting with no circles close to it; therefore it was not possible to estimate the perpendicular distance in GIS software.

The same effect is shown in 2013 when removing only two sightings with clearly wrong perpendicular distances estimated by angle measurement; 769 m (Figure 7a) and 7623 m (Figure 7b) produce changes in the scattergrams and linear regression (Figure 5).

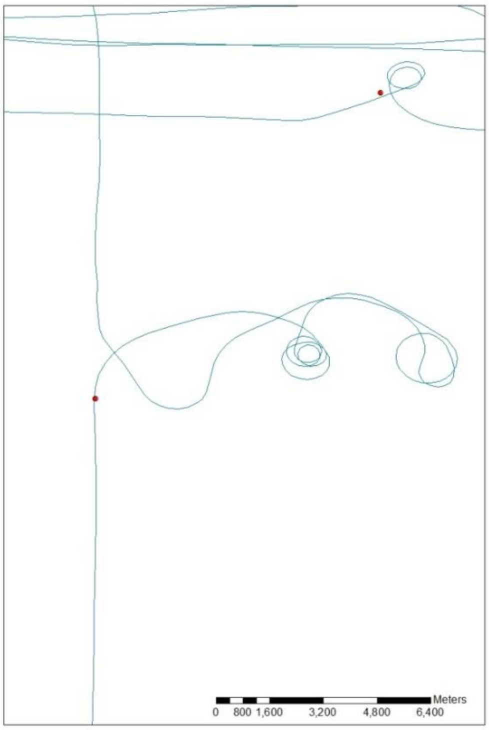


Figure 7a. BFT sighting with wrong perpendicular distance estimated by clinometer (769 m).

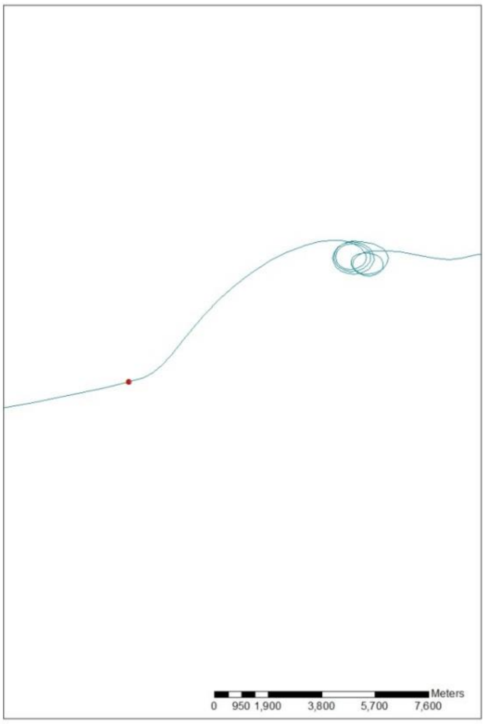


Figure 7b. BFT sighting with wrong perpendicular distance estimated by clinometer (7623 m).

The problem in Figure 7a is that two observations were reported with the same position and angle, but in fact the second one occurred 7 minutes later than the first one. Obviously the second one was a secondary sighting but reported as primary. The first circle in the figure corresponds to the first sighting, and the second one corresponds undoubtedly to the second, secondary sighting. Therefore the angle taken for this second one is wrong as it was recorded as the same as the first one. This distance from this wrong angle has been deleted. In the case of Figure 7b, the reported clinometer angle provided a distance value (7623m) was obviously not a perpendicular distance but a radial distance, as measuring in GIS the distance between the circle and the red point (point in the track when detection of the school) is 7420m, very similar to the distance estimated from the reported angle. Therefore, this is a wrong angle again.

Once again, in 2015 the extraction form the calculations of two BFT sightings where perpendicular distances estimated by clinometer were clearly wrong: 14,780m, when the GIS distance to the concentric circles is 1049m (Figure 8a) and 173m when the GIS distance to the concentric circles is 7,835m (Figure 8b), produces changes in the scattergram and linear regression (Figure 5).

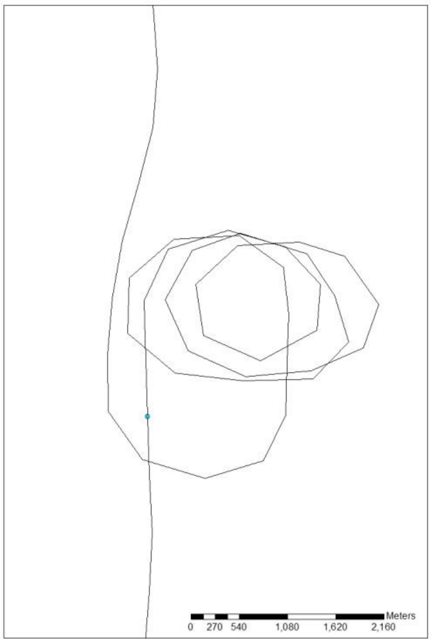


Figure 8a. BFT sighting with wrong perpendicular distance estimated by clinometer (14,780 m).

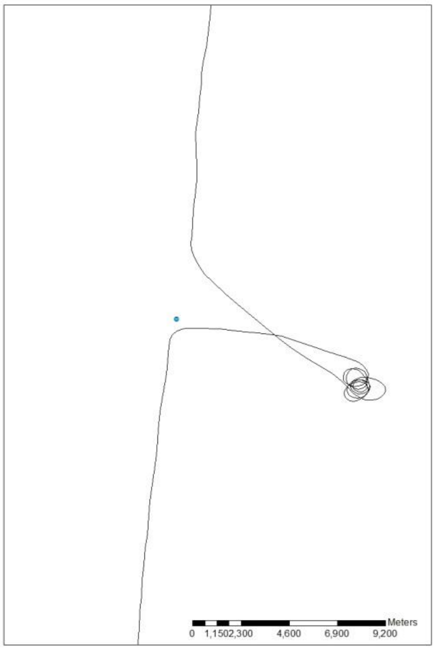


Figure 8b. BFT sighting with wrong perpendicular distance estimated by clinometer (173 m).

To test what difference would it make, using one estimate of perpendicular distance or the other, an experiment was done. A new Distance project was created, in which the GIS-derived perpendicular distances were used, wherever possible, substituting previously used clinometer (angle) derived. All models were run again with the same configuration as before. Table 2 shows the previous results, and the new ones with the “corrected” distances from GIS. The results are strikingly different in 2011 (much higher) and 2015 (much lower), although very similar in 2013. 2010 does not have data for GIS derived distances.

One of the effects of changing the perpendicular distances is that the amount of observations inside the truncation distance may vary. In this exercise, two observations from C in 2011 and two in A in 2013 got removed, one observation in A in 2015 got included. Additionally, the “corrected” distances derived from GIS are in some cases very different than the original ones derived from the angle. These differences can be of several thousand meters in some cases, which can produce an important effect on the detection function.

Table 2. Comparison of abundance estimates for overlapped sub-areas inside for each year, with and without GIS derived distances. Cells in grey show the changes.

Year	**2010**				**2011**			**2013**				**2015**			
Sub-area	**A inside**	**C inside**	**E inside**	**G inside**	**A inside**	**C inside**	**E inside**	**A inside**	**C inside**	**E inside**	**G inside**	**A inside**	**C inside**	**E inside**	**G inside**
Original estimates with only angle derived distances															
Number of schools ON effort	8	6	29	33	10	10	45	10	10	20	12	6	3	13	2
Density of schools (1000 km^{-2})	0.430	0.248	0.775	3.840	0.922	0.868	3.374	0.495	1.244	1.794	2.333	0.480	0.372	0.534	0.840
%CV density of schools	56.2	46.6	32.7	29.4	35.9	33.4	24.1	36.1	34.3	34.0	40.7	43.5	62.9	50.8	69.0
Mean cluster size (animals)		733	1,015		678	291	1,715	611	1,285	361	336	825	1,533	2,030	600
%CV abundance		36.5	19.0		27.9	30.7	21.5	26.0	17.0	67.3	36.7	11.0	19.0	56.8	66.7
Total abundance (animals)		**9,797**	**73,676**		**38,720**	**13,614**	**541,634**	**18,717**	**86,114**	**60,614**	**44,041**	**38,248**	**47,900**	**283,100**	**44,162**
%CV total abundance		59.2	37.8		45.5	45.3	32.3	44.5	38.3	75.4	54.8	44.7	65.5	64.1	95.9
L 95% CI total abundance		3,187	35,741		16,249	5,677	290,700	7,990	40,959	15,391	15,587	16,510	14,331	83,058	4,844
U 95% CI total abundance		30,016	151,880		92,266	32,649	1,009,200	43,845	181,040	238,710	124,440	88,610	160,100	964,970	402,600
Experimental estimates with GIS derived distances when available															
Number of schools ON effort					10	8	45	8	10	20	12	7	3	13	2
Density of schools (1000 km^{-2})					1.499	1.130	5.485	0.400	1.256	1.811	2.355	0.568	0.377	0.968	0.851
%CV density of schools					36.2	37.2	24.6	38.1	34.3	33.9	40.6	39.4	62.1	28.3	68.3
Mean cluster size (animals)					678	346	1,715	688	1,285	361	336	707	1,533	2,030	600
%CV abundance					27.9	29.7	21.5	27.5	17.0	67.3	36.7	19.9	19.0	56.8	66.7
Total abundance (animals)					**62,945**	**21,067**	**880,510**	**17,035**	**86,911**	**61,177**	**44,450**	**24,865**	**31,134**	**184,010**	**28,704**
%CV total abundance					45.7	47.6	32.6	47.0	38.2	75.4	54.7	44.2	65.0	63.5	95.5
L 95% CI total abundance					26,309	8,443	469,550	6,934	41,392	15,541	15,744	19,771	9,387	54,342	3,126
U 95% CI total abundance					150,060	52,565	1,651,100	41,849	182,490	240,820	125,500	57,402	103,260	623,080	263,570

Conclusion

Detecting BFT schools is a complex process that depends on the observer's experience, the sea state, and the number and activity of the animals in the school. Throughout the four years during which surveys have been conducted to estimate the number of spawners of this species in the Mediterranean, differences were detected in the way data collection protocols were applied: the location of the Professional spotter (PS) and Scientific Spotter (SS) in the airplane, the use of airplanes with bubble windows and inclinometer…etc. Additional unavoidable problems are created by the spotting platform (the aircraft itself), because of the speed and the instability, particularly in case of small turbulences, when detecting an angle is very difficult. According to the protocol, PS must occupy the co-pilot place and SS the place behind with bubble windows to see underneath the plane and to collect angles with inclinometers. Under this configuration, sometimes it occurs that the plane must leave the track when a BFT school is detected by the PS before the SS detects it and therefore SS is not able to read the angle when the plane is located perpendicular to the sighting. The reason for this is that otherwise the PS could lose the school and then approaching to estimate group size and weight would be impossible. On the other hand, the estimation of perpendicular distances based only on GPS data, measuring the length between the centre of contiguous circles where animals are supposed to be placed and the direction of flight in a straight line, is partially dependent on the analyst skills and not always possible due to several reasons (see Figures 1 to 4).

As shown in Figures 6, 7 and 8, there have been some examples where we found huge errors in measuring angles with clinometer by observers (SS), due to the complex detection process explained before and/or due to the inexperience of some observers that participated in the surveys. The observations with the largest differences or obvious problems in GIS were checked searching for possible errors in measuring angles. But there could be still other minor errors that went undetected. Even when checking, it is not possible to know exactly what the reason might be for the detected errors. The exploration made in this item points out the difficulties for the interpretation of the perpendicular distances data and, therefore, the quality of the data that has been used in Distance programme to estimate abundances, especially in 2010 and 2011 but also in 2013 and 2015. The examples shown here and that came up with this exploration, unfortunately mean that those wrong distances were used in the Distance analysis to estimate abundance, and the amount of other potential measurement errors like those ones is unknown unless all are checked, one by one, in GIS.

The comparison shows clear evidences about the problems to obtain accurate perpendicular distances through the use of clinometer and also in some cases by using GIS software. This is why it is needed to keep using both methods, enforcing SS to use clinometers in the right way and companies to provide accurate GPS data. Furthermore, the huge CVs in Table 1 evidence the very large variability in error measurements, in both directions, under and over estimations. In some cases, as stated above, the true distance can be reassessed with high certainty using GIS, but in some other cases it is not that clear which one would be the most accurate one.

Several observers reported practical problems in getting the right angle using clinometer, particularly when the school is small and close to the aircraft track. This difficulty was confirmed also by the GBYP staff when they had the opportunity to closely check the activity on board.

Our recommendation for futures surveys is that if the same PS/SS are maintained over time and that a protocol slightly changes in a way that a SS always annotates in the form if it was possible to get an accurate angle to the sighting when the plane was perpendicular to it before leaving the track or not. If this was possible, then the perpendicular distance obtained with the angle should be used. In the cases when the plane must leave the transect before the SS gets the angle, the perpendicular distance must be calculated in GIS measuring the length between the centre of the contiguous circles and the direction of flight in a straight line in the moment that the plane leaves the transect.

The huge differences on estimates in 2011 and 2015 show how important it is to get the most accurate possible perpendicular distances, as they influence enormously the probability of detection (detection function) and therefore the abundance estimates. Our recommendation is to continue with both methods simultaneously but making sure that both are done in a better way in the future, improving the performance as much as possible. In this way, we will be able not only to explore with more accuracy the differences and to attempt to make reliable corrections, but there will be more confidence in the data and therefore in the estimates.

I.2 Analyses of the GPS recorded tracks compared to the survey designed track-line

Data organization

To carry out this part of the work only 11 GPS data sets were available; areas A and C inside in 2011, areas A inside/outside and B outside in 2013, and areas A, C, E, G inside and A, B, D, E, F and G outside in 2015. There is no scientific rule or specification on how close to the designed tracks should the aircrafts fly to be considered "on track". In this instance 200m each side of the designed track has been arbitrary chosen as being close enough to be considered "on" the track.

All GPS data sets include trackings from the airport to the starting point and from the ending point to the airport. There are also some tracks over land, and also those corresponding to circles. All those track pieces[3] should not be included in the analysis to highlight how good the flights were carried out.

In order to select the GPS points corresponding to the moments when the airplane was following the designed tracks, the orientation of those designed tracks were calculated. All GPS points with headings in the range of -10º/+10º degrees or -170º/+170º (north or south directions parallel to the design tracks) were selected as "on track". For example, designed tracks in area C inside/outside 2015 were oriented 10º north /190º south (Figure 9a). By selecting points with headings between 0º-20º north and 180º-200º south from the original GPS points (Figure 9b) the result is shown in Figure 9c. To eliminate those points that still remain over land the "spatial selection" tool in GIS was used (Figure 9d). Once the GPS data was filtered, the distance between each point and the closest designed track was measured using "spatial join" tool in GIS. For the measurement of the distances a World Mercator projection was used.

Figure 9a. Designed transects for area C inside/outside 2015.

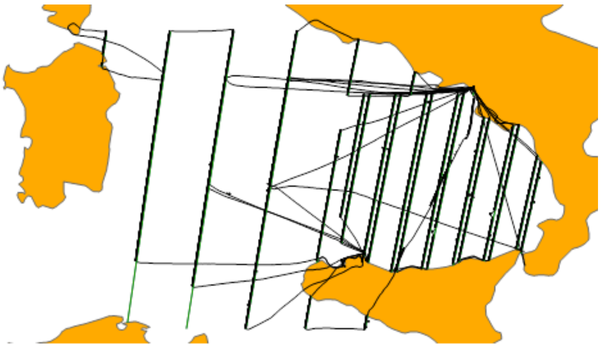


Figure 9b. Original points in the GPS data set for area C inside/outside 2015.

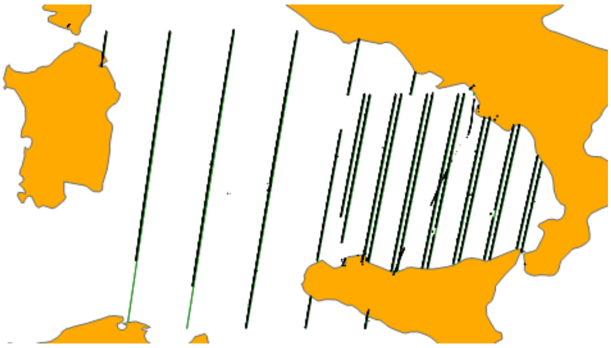


Figure 9c. GPS points with headings between 0º-20º north and 180º-200º south.

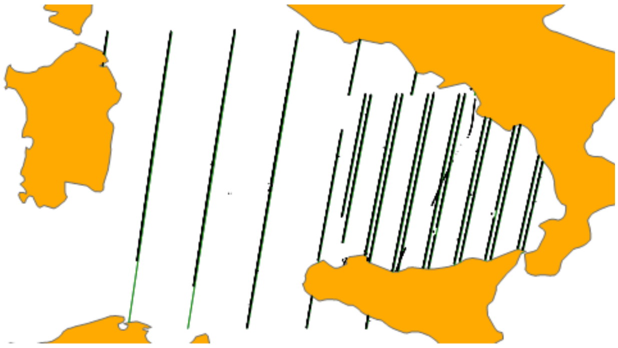


Figure 9d. GPS points after deleting those over land.

Figure 9. Example of data processing in area C 2015

[3] This info is anyway very useful for a better understanding of the logistic needs behind the survey.

Data analysis

For the analysis, the points with distances larger than 2000m were not included because it was considered that those points are remaining points corresponding to tracks form and to the airport that were not deleted during the "cleaning" process explained before. To compare how the fit is between the flights and the realized transect line, the frequency of the measured distances were plotted for each 50m bin between 0m and 2000m. Bins corresponding to 200m, 300m and 400m were marked in red as potential indicators of flights "on" track. Additionally, quartiles were calculated for each GPS data set. The third quartile (Q3) was selected as good indicator of fit when the value obtained was between 200-400m or below. Above 400m it was considered that the flight was not fitted to the designed line.

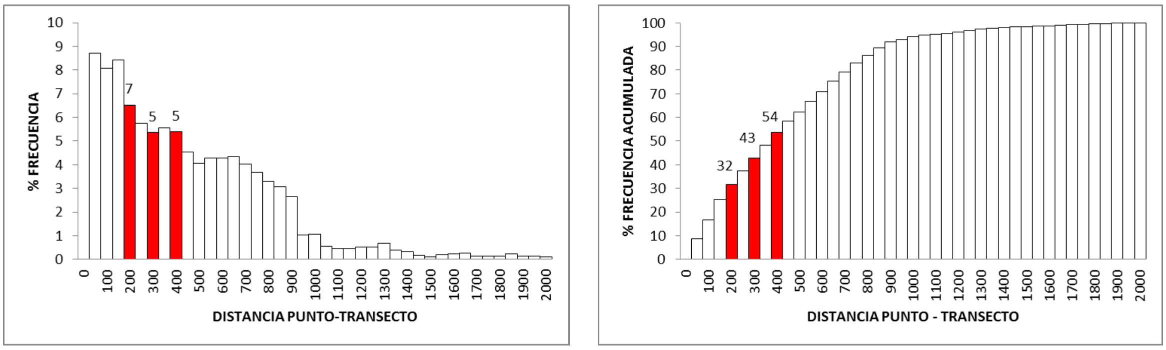


Figure 10. Frequency and cumulative frequency of the distances between each point of tracking and the closest transect. Area A 2011.

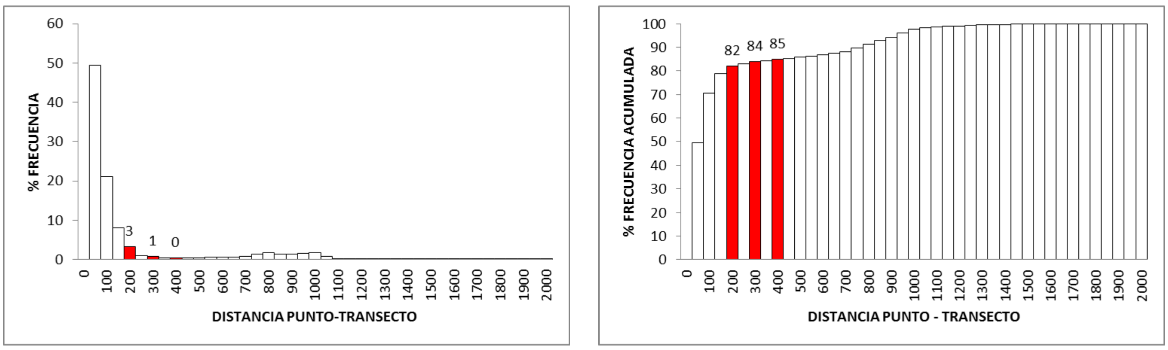


Figure 11. Frequency and cumulative frequency of the distances between each point of tracking and the closest transect. Area C 2011.

Table 3. Mean, standard deviation (sd), coefficient of variation (%CV) and quartiles 1, 2 and 3 of the distances in areas A and C 2011.

Area	n	mean	sd	CV	Q1	Q2	Q3
A	6,724	437	352.7	80.8	148	364	647
C	10,728	178	300.7	169.0	22	51	121

Figures 10 a. b and 11 a, b show the frequency and accumulated frequency of distances for area A and C in 2011 respectively. Whereas in area A only 32% of the distances were below 200m with a Q3 value of 647m, in area C 82% of the distances were below 200m with a Q3 value of 121 m (Table 3). This is a clear example where one pilot/team did not follow correctly the designed tracks (AirMed) while another did it much better (Unimar).

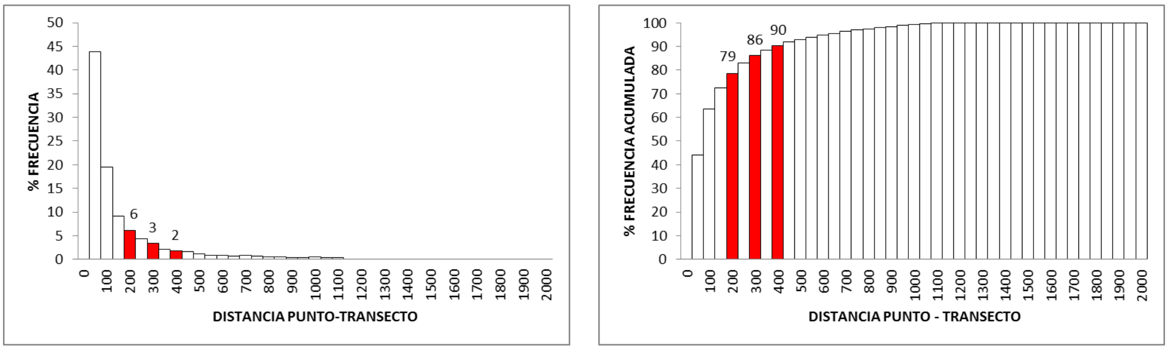


Figure 12. Frequency and cumulative frequency of the distances between each point of GPS and the closest designed track line. Area A 2013.

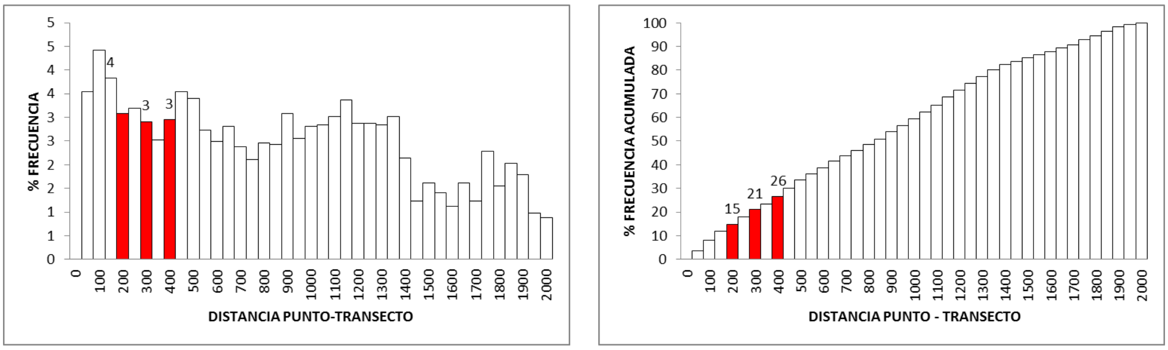


Figure 13. Frequency and cumulative frequency of the distances between each point of GPS and the closest designed track line. Area B 2013.

Table 4. Mean, standard deviation (sd), coefficient of variation (CV) and quartiles 1, 2 and 3 of the distances in areas A and B 2013.

Area	n	mean	sd	CV	Q1	Q2	Q3
A	11,182	143	200.2	140.1	26	61	168
B	2,850	851	549.8	64.6	373	836	1259

In 2013 (Figures 12 and 13) the pilot and team surveying area A improved their performance in comparison to 2011 because 79% of the distances were below 200m with a Q3 value of 168, whereas in area B another bad example was found; only 15% of the distances were below 200m with a Q3 value of 1259 m (Table 4). Therefore in 2013 AirMed followed the designed tracks correctly and Perigord did not.

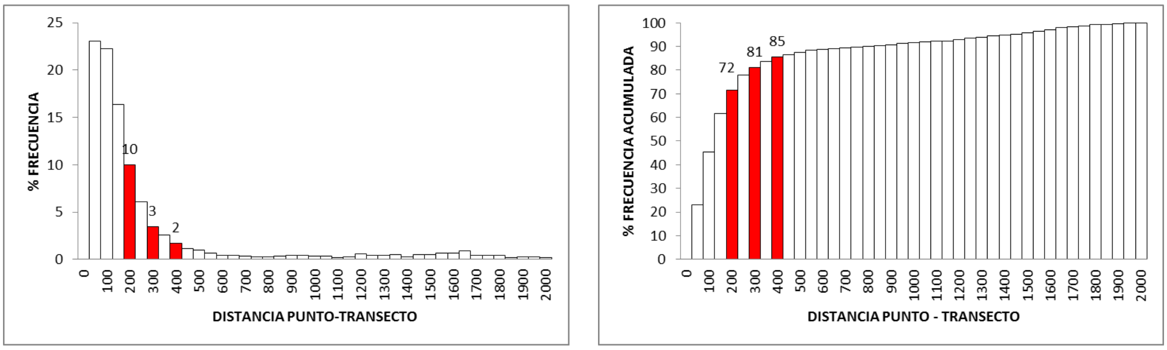


Figure 14. Frequency and cumulative frequency of the distances between each point of GPS and the closest designed track line. Area A 2015.

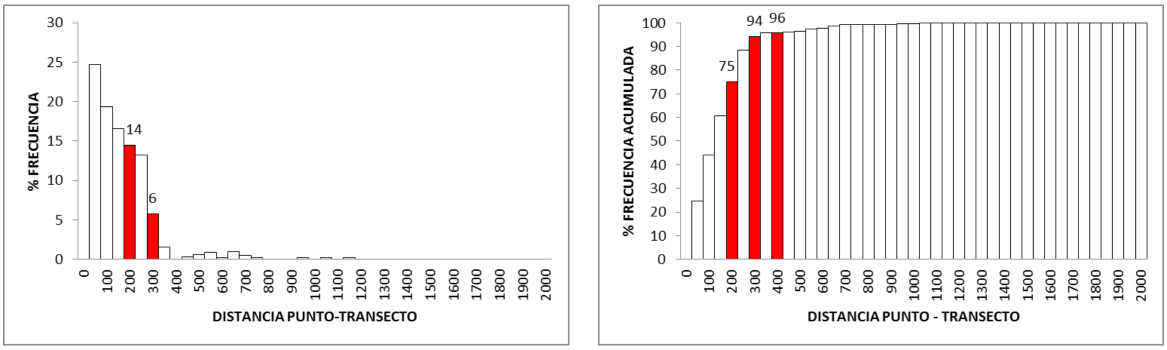


Figure 15. Frequency and cumulative frequency of the distances between each point of GPS and the closest designed track line. Area B 2015.

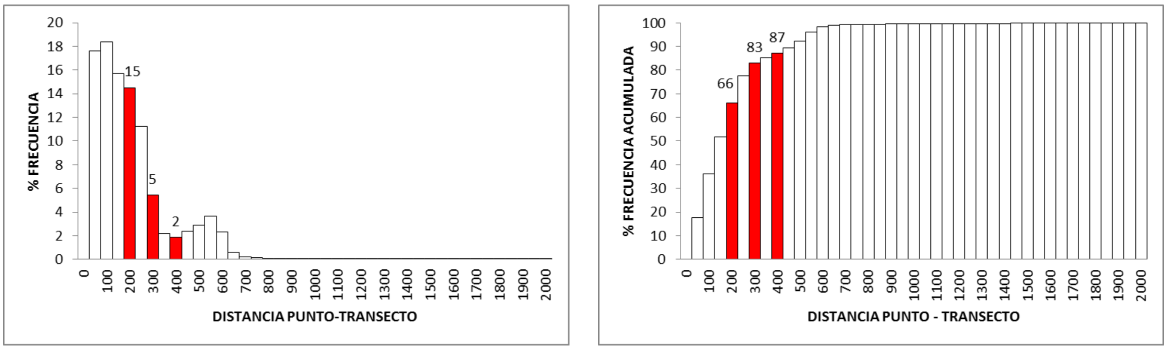


Figure 16. Frequency and cumulative frequency of the distances between each point of GPS and the closest designed track line. Area C 2015.

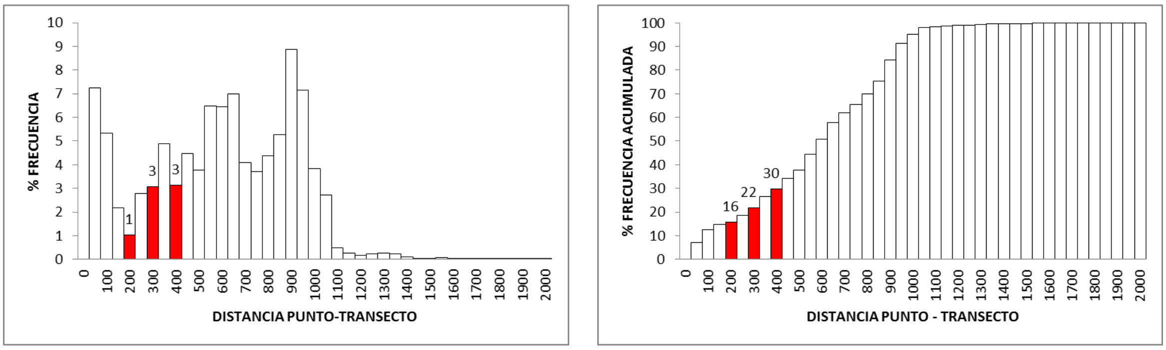


Figure 17. Frequency and cumulative frequency of the distances between each point of GPS and the closest designed track line. Area D 2015.

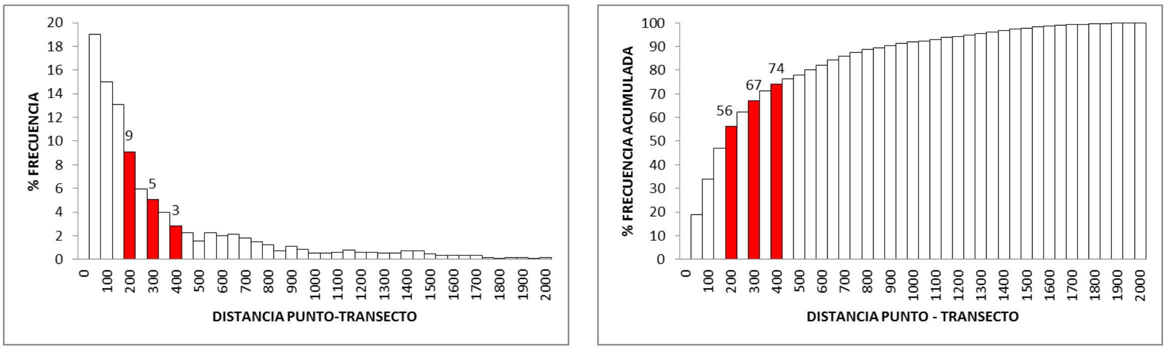


Figure 18. Frequency and cumulative frequency of the distances between each point of GPS and the closest designed track line. Area E 2015.

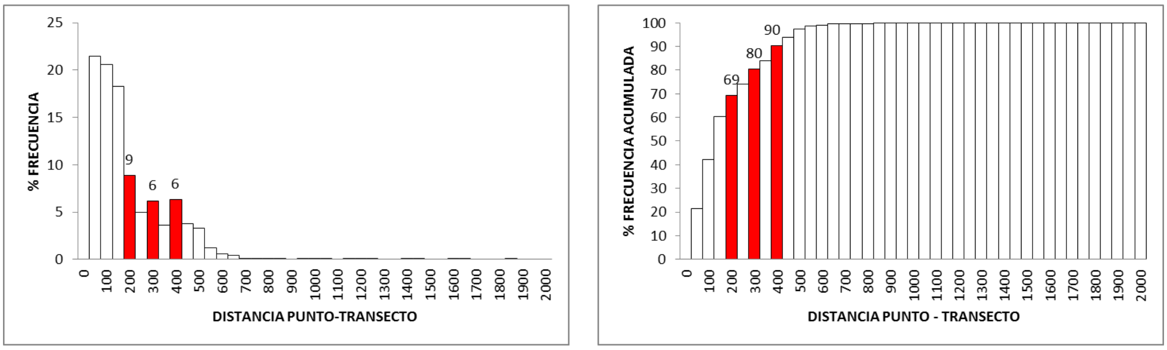


Figure 19. Frequency and cumulative frequency of the distances between each point of GPS and the closest designed track line. Area F 2015.

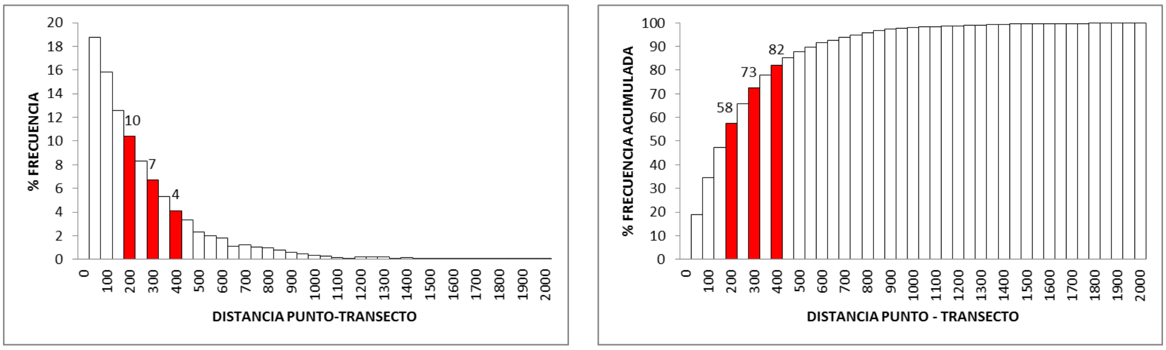


Figure 20. Frequency and cumulative frequency of the distances between each point of GPS and the closest designed track line. Area F 2015.

Table 5. Mean, standard deviation (sd), coefficient of variation (CV) and quartiles 1, 2 and 3 of the distances in all areas in 2015.

Area	n	mean	sd	CV	Q1	Q2	Q3
A	25,412	261	409.4	156.8	54	113	226
B	1,497	144	136.4	94.9	50	113	200
C	32,085	189	181.9	96.0	69	145	238
D	13,651	1367	3116.3	228.0	356	626	888
E	21,839	325	388.4	119.5	70	163	422
F	6,704	667	2239.4	335.5	62	127	290
G	6,120	243	260.8	107.4	69	164	321

In 2015 (Figures 14 to 20), except in areas D and E with 16% and 56% of the distances below 200m with Q3 values of 888 and 422 respectively (Table 5), the pilots/teams followed quite well the designed tracks. This year Unimar and ActionAir were the companies in charge of surveying areas D and E.

Conclusion

Due to the lack of GPS data sets in 2010 and some areas in 2011 and mainly in 2013, it was a problem to evaluate this issue those years in some areas. Fortunately, it seems that in 2015 the fit of the realized to the designed tracks has improved, but it is still necessary to remark and reinforce the requirement of this kind of data in the proper format and time in order to be analysed correctly in future surveys too.

According to these results, a high percentage of the pilots/teams seems to fly close enough to the designed track lines as to be considered "on" track. However, the fact that one company has very different performance in different years/areas, suggests that the flight procedure is highly dependent on pilot experience or on unknown external factors. Therefore, it is highly recommended for future surveys to ask the companies to work with the same pilots/teams.

I.3 Further analyses of the GPS recorded flight path to see whether protocols for encircling above schools were reliably followed

Analyses were done in point I.1. on the inclinometer *versus* the GPS derived estimates of the distance of a school from the track-line. For that reason, a detailed review of all BFT sightings was done in order to estimate perpendicular distances using GIS software. For those areas where GPS data were available, not all sightings were associated to circles. Most of the sightings have school size and weight estimates, so in theory the airplane observers would have needed to circle the animals to obtain the estimates, but that must not be the case in some instances. Table 6 shows a summary of the total BFT sightings per year/area, those for which perpendicular distance was possible to be estimated through GIS, and those in which it was no possible to find any circle associated to the sighting (NC: no circles).

Table 6. Number of BFT observations with distance estimation through inclinometer and GIS (GIS), and observations with no circling (NC). % indicates the proportion of observations with circling to estimate distances.

AREA	2010				2011				2013				2015			
	TOT	GIS	NC	%	TOT	GIS	NC	%	TOT	GIS	NC	%	TOT	GIS	NC	%
A inside	13				22	5	17	22.7	18	17	1	94.4	33	31	2	93.9
A off	13								21	19	2	90.5				
A outside									2	2		100.0	6	6	2	100.0
B													3	1		33.3
C inside	6				11	11		100.0	13				6	6		
C off	1								13							
C outside													4	2	2	50.0
D									1				4	4		
D off									1							
E inside	34				58				34				18	10	8	55.6
E off	10												8	7	1	87.5
E outside									2							
F									2				1	1		
F off									1							
G inside	33								37				2	1	1	50.0
G off																
G Outside									16				2	1	1	50.0
TOTAL	110				91	16	17	17.6	161	38	3	23.6	87	70	17	80.5

The percentage of number of sightings where GIS estimates was possible, i.e. when circles were available, with respect to the total number of BFT sightings recorded varies among years and areas. This effect could be related with the ability of the pilot/team to detect and follow the animals. Although there is a clear positive trend in percentage of BFT sightings where GIS estimates were possible, i.e. when circles were available, there are still areas in 2015 where this number is low.

The column % in the table represents the percentage of sightings with circles. But these circles sometimes are concentric and other times they are not. When they are not concentric, there could be two potential explanations. Either animals are moving and the plane follows them in subsequent circles, or the pilot/observers missed the animals and they are making circles while the airplane moves in one direction trying to find them again. It is not possible to know which was the reason unless it was specified it in the comments.

It would be highly recommended to remark the importance of these issues to the future teams and it would also be very important to maintain those teams that have been proved to work well in previous years.

II. Analyses of individual biases by spotter

II.1 Re-analysis of all sighting data collected in all surveys in 2015 by each spotter, comparing two different number estimates of the same BFT school when available, trying to define possible individual different capacities or biases.

Data organization

At any given moment, there were one professional spotter (PS from now on) and two scientific spotters (SS from now on) apart from the Pilot (P). The name of the observer making the observation was provided for all sightings of BFT. But there was some troubles when trying to deal with this item, so these different situations occurred:

- In many cases when the observation was made by the PS and he gave de estimated school size, there was no parallel estimation by the SS, so no comparison is possible
- When the observation was made by the PS and there was a parallel school size estimation from a SS, the name of the SS providing the parallel estimation was not provided, so no personal comparison is possible to a specific SS.
- When the observation was made by a SS, his/her name was provided, but in some cases no parallel estimation was given by the PS, so no comparison is possible.
- When the observation was made by a SS, his/her name was provided, and when there was a parallel estimation by the PS, as there was only one PS at any given time, his name was known. In this case comparison was possible.

Therefore, three types of comparison were possible. In all cases, the sample size is small (very small in the first case):

(a) Specific PS with specific SS when the observation was made by SS and both provided estimated school size (5 SS and a total of 16 observations)

(b) Specific PS with “general” SS, i.e. estimation by SS without specifying who, given than both provided estimation of school size (4 PS with general SS, and 5 SS with specific PS, total of 34 observations)

(c) Overall all PS against overall SS (76 school size estimations by PS and 35 by SS, and 34 parallel estimations).

Data analysis

There is too small sample size to perform a proper robust analysis of comparison, and especially to define possible individual different capacities or biases. Nevertheless, a rough comparison was made between overall estimations by PS and SS. Table 7 shows the observations where both PS and SS provided an estimate of school size, and the difference between both (comparisons a and b above). Table 8 shows the difference in the mean estimation of school sizes by PS and SS and the difference among them (comparison c above). Table 9 shows the same mean estimates for weight (in this case only 33 observations were available with parallel estimations).

Table 1 shows that, out of the 34 observations where comparison was possible, only 3 of them had the same estimate (difference =0), 12 had SS giving a smaller estimate than PS, and in the remaining 19 observations, the SS gave a larger estimate of school size. Only 3 PS provided parallel school size estimation to that given by the SS doing the observation, and the maximum one to one comparisons was 4 observations by SS 59 with PS number 26; all the rest pairs had even smaller sample size. Table 2 shows that the overall difference in the mean estimation of school size by PS and SS in observations with parallel estimations, is 22% larger estimates in SS than in PS. Table 3 shows that also in the estimation of weight, SS provide in average a 16% larger estimates than PS.

Conclusion

In average, SS give larger estimates than PS, by 22% in school size and by 16% in weight. But these differences are not constant and are just an average, with differences ranging from negative to positive in both cases, and with no observable pattern. No statistical analysis could be performed with such small sample size. Usually, SSs do not have enough experience in weight estimates.

Table 7. School size (Gsize) estimations by PS and SS in those observations were both estimates were provided. “Dif” is the difference of the estimate of the SS with respect to the estimate of the PS. “% dif” is the percentage of difference of the estimate by the SS with respect to the estimate of the PS. For observations by a known SS, the associated PS is given.

Observer ID	Obs. nº	Gsize PS	Gsize SS	Dif	% dif	
Observations by a PS						
26	196	800	1500	700	88	
	197	1500	2000	500	33	
	314	700	2000	1300	186	
	326	1500	700	-800	-53	
	370	2000	2000	0	0	
	402	700	1000	300	43	
	403	800	400	-400	-50	
	418	700	200	-500	-71	
	436	800	2000	1200	150	
	437	600	1300	700	117	
	439	500	1600	1100	220	
28	628	1600	1000	-600	-38	
	636	1000	1500	500	50	
	661	300	400	100	33	
	662	200	150	-50	-25	
	666	2500	2000	-500	-20	
31	675	200	200	0	0	
	680	1000	1500	500	50	
32	831	10	200	190	1900	
Observations by a SS						**PS**
38	647	2000	1500	-500	-25	28
	652	1500	1700	200	13	28
41	818	1500	2000	500	33	32
	835	3	1	-2	-67	32
	857	780	800	20	3	32
45	793	125	150	25	20	32
	837	4	1	-3	-75	32
	861	250	300	50	20	32
	882	300	300	0	0	32
48	313	600	3000	2400	400	26
	359	1000	500	-500	-50	26
59	238	2000	1500	-500	-25	26
	303	700	1500	800	114	26
	438	800	900	100	13	26
	522	1000	800	-200	-20	26

Table 8. Mean school size (Mean Gsize) estimations by PS and SS for observations with parallel estimations. “Dif” is the difference of the mean estimate of the SS with respect to the mean estimate of the PS. “% dif” is the percentage of difference of the mean estimate by the SS with respect to the mean estimate of the PS.

Observer ID	N PS	N SS	Mean Gsize PS	Mean GSize SS	Dif	% dif
Observations by a PS						
26	11	11	964	1336	373	39
28	5	5	1120	1010	-110	-10
31	2	2	600	850	250	42
32	1	1	10	200	190	1900
Observations by a SS						
38	2	2	1750	1600	-150	-9
41	3	3	761	934	173	23
45	4	4	170	188	18	11
48	2	2	800	1750	950	119
59	4	4	1125	1175	50	4
Total	34	34	882	1077	195	22

Table 9. Mean weight estimations by PS and SS for observations with parallel estimations. “Dif” is the difference of the mean estimate of the SS with respect to the mean estimate of the PS. “% dif” is the percentage of difference of the mean estimate by the SS with respect to the mean estimate of the PS.

Observer ID	N PS	N SS	Mean Weight PS	Mean Weight SS	Dif	% dif
Observations by a PS						
26	11	11	156818	197273	40455	26
28	5	5	144	318	174	121
31	2	2	101	102	1	0
Observations by a SS						
38	2	2	225	225	0	0
41	3	3	223348	241672	18323	8
45	4	4	21060	21515	455	2
48	2	2	100000	150000	50000	50
59	4	4	181250	165000	-16250	-9
Total	33	33	103202	119494	16293	16

II.2 Separate professional spotters & pilots from scientific spotters.

Data organization

School size

A new dataset was organized to run Distance software, in which a new column labelled "PS-SS" identified which observations had both school size estimations by the PS and the SS simultaneously. Only data from 2015 were available as it is the only year with simultaneous estimates of school size by PS and SS. Only 33 observations remained available thus for this analysis (the 34 observations with parallel school size estimations, minus one observation with no perpendicular distance). Of these 33 observations usable for a detection function, only 15 were on effort and therefore usable to estimate abundance from the detection function.

Encounter rates

Data were organized to perform a Chi-square test in order to compare the observed with the expected values by each type of observer, stratified by Team. As in each flight, theoretically, there is one PS and two SS, a third of the total amount of effort (in the blocks surveyed by each Team) was "allocated" to the PSs and two thirds to the SSs.

Distance

The same dataset used for obtaining the detection functions for the estimates of abundance was used to explore the effect of observer type, in each team, on the detection function.

Data analysis

School size

Two different detection functions were run for BFT data in 2015 to see what is the effect of the differences in school size estimation on the estimates of abundance (number of animals was used for school size):

a) Considering only the school size provided by PS

b) Considering only the school size provided by SS

Table 10 shows the mean estimated school size from the PS and SS, for observations on effort, with their corresponding CV and 95%CI. Due to the small sample size and the huge variability in the estimated school sizes, the CVs and 95%CI in some of the blocks are very large.

Table 10. Mean estimated school size from the PS and SS, for observations on effort, with their corresponding CV and 95%CI.

			PS				SS			
Sub-area	**n total**	**n on effort**	**Mean**	**% CV**	**95% Confidence Interval**		**Mean**	**% CV**	**95% Confidence Interval**	
A outside	4	2	1400	42.9	8	258300	1750	14.3	288	10651
A inside	13	4	725	6.6	588	894	1425	16.1	856	2367
C outside	3	1	2500				2000			
C inside	3	3	1533	18.9	683	3441	1333	12.5	780	2278
D	1	1	1000				1500			
E inside	8	4	441	81.2	46	4240	625	73.9	76	5117
F	1		200				200			
TOTAL	**33**	**15**								

Table 11 shows the estimated abundance and density based on those 15 on-effort observations and the detection function from the 33 observations on and off effort. The same detection function as in the previous report was used. Obviously, these are not abundance estimates to be considered for anything, except as an experiment to explore the differences in school size estimates from PS or from SS, and in any case it should be remembered that this exploration is based on a very small sample size. Table 12 shows the estimates for the sub-areas with observations with parallel school size estimates. When looking at Table 11, the overall abundance estimate for all sub-areas together is very similar in both cases, a bit larger when using the estimates from SS. This is in agreement with the results from II.1 where it was shown that, overall, the SS provide school size estimates 22% larger in average than the PS. However, these differences are not the same for all sub-areas (see Table 12), but given the extremely low sample size when stratifying by sub-area, these differences are not meaningful. The main purpose of this exploration, however, was to demonstrate how the differences in school size estimations yield differences in the abundance estimates from a survey.

Table 11. Overall abundance estimate for all sub-areas together

	PS				SS			
	Estimate	**%CV**	**95% CI**		**Estimate**	**%CV**	**95% CI**	
Density (schools)	0.00019	41.07	0.00009	0.00041	0.00019	41.07	0.00009	0.00041
Abundance	315,000	50.89	120,800	821,380	360,530	45.75	152,260	853,680

Table 12. Stratified abundance estimate for the sub-areas

	PS				SS			
	Estimate	**%CV**	**95% CI**		**Estimate**	**%CV**	**95% CI**	
A outside								
Density (schools)	0.00062	75.9	0.00016	0.00239	0.00062	75.9	0.00016	0.00239
Abundance	107,730	87.17	21,701	534,780	134,660	77.24	34,411	526,950
A inside								
Density (schools)	0.00045	55.52	0.00016	0.00126	0.00045	55.52	0.00016	0.00126
Abundance	20,457	55.92	7,326	57,127	40,210	57.8	13,946	115,930
C outside								
Density (schools)	0.00025	107.75	0.00004	0.00152	0.00025	107.75	0.00004	0.00152
Abundance	94,258	107.75	15,587	570,000	75,406	107.75	12,470	456,000
C inside								
Density (schools)	0.00044	64.88	0.00013	0.00144	0.00044	64.88	0.00013	0.00144
Abundance	43,181	67.59	12,578	148,240	37,549	66.07	11,207	125,800
D								
Density (schools)	0.00022	104.04	0.00004	0.00132	0.00022	104.04	0.00004	0.00132
Abundance	32,727	104.04	5,482	195,400	49,091	104.04	8,222	293,100
E inside								
Density (schools)	0.00032	55.88	0.00011	0.00090	0.00032	55.88	0.00011	0.00090
Abundance	16,649	98.54	2,299	120,540	23,618	92.7	3,734	149,390

Encounter rates

Table 13 shows the encounter rates by each type of observer, by Team. Table 14 shows the contingency tables for the Chi-square tests. Each Team is analysed separately, so each respective Team has only one degree of freedom (only PS and SS within each team is analysed, not across teams), and therefore Yates's correction is applied to the Chi-square calculation.

Table 13. Encounter rates (groups per 100 km searched) per observer type (PS = Professional Spotter; SS = Scientific Spotter) and Team

	Air-Med	Unimar	Perigord	Action-Air	Total
PS	0.347	0.107	0.327	0.050	0.185
SS	0.163	0.015	0.103	0.009	0.058
Total	**0.511**	**0.122**	**0.430**	**0.059**	**0.243**

Table 14. Contingency tables for the Chi-square tests.

	Obs. type	Action-Air	Air-Med	Perigord	Unimar	Total
Effort	PS	4,894	11,174	11,316	11,322	**38,707**
	SS	9,788	22,345	22,631	22,643	**77,407**
	Total	**14,683**	**33,521**	**33,950**	**33,968**	**116,122**
Observed	PS	51	36	111	17	**215**
	SS	24	5	35	3	**67**
	Total	**75**	**41**	**146**	**20**	**282**
Expected	PS	25	14	49	7	**94**
	SS	50	27	97	13	**188**
	Total	**75**	**41**	**146**	**20**	**282**
Chi-square	PS	26	35	79	15	154
	SS	14	19	41	9	79
	Total	**40**	**54**	**119**	**23**	**233**
p	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**	**<<0.001**

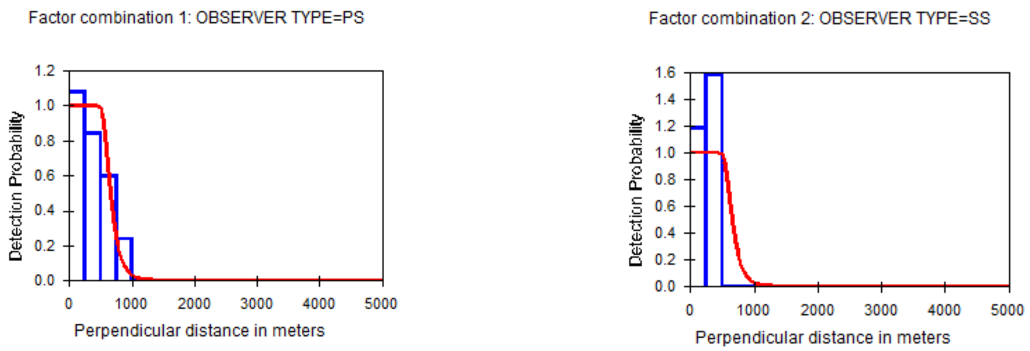
It is obvious from both tables that, in all teams, PS have a much higher encounter rate than SS, presumably due to the much larger experience searching for BFT from airplanes. The differences are highly significant in all cases.

Distance

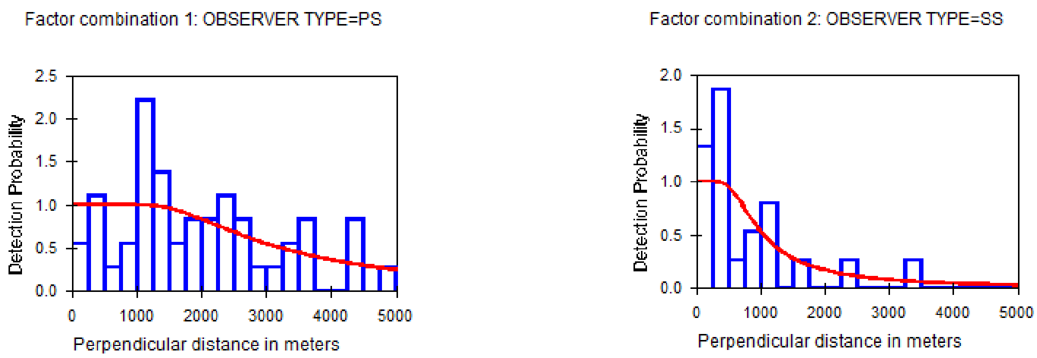
Detection functions were run for each team using observer type as covariate (Figure 21). A global detection function with all teams pooled together was run too (Figure 22). In all cases, year 2010 was removed as there were no bubble windows and therefore the lack of observations in the closest ranges that year could introduce noise in this exercise.

PS	SS

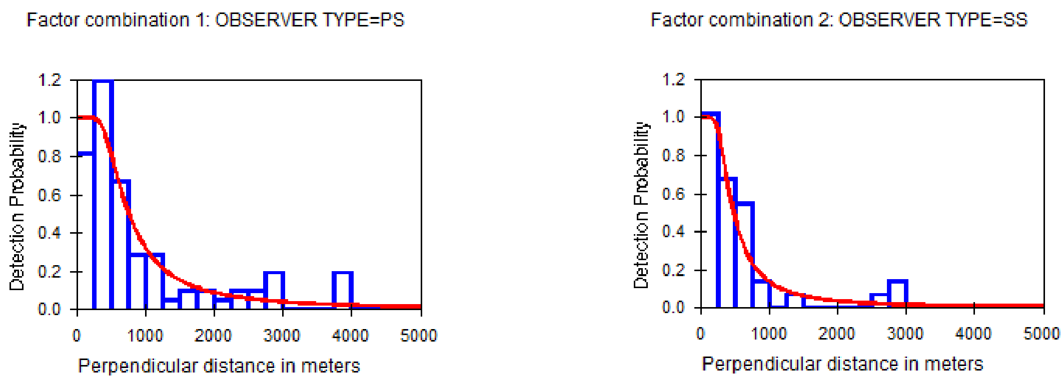
Action-Air



Air-Med



Perigord



Unimar

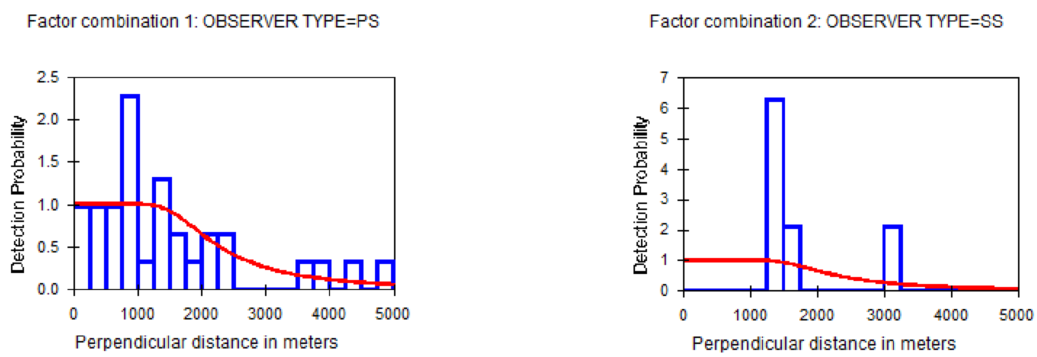
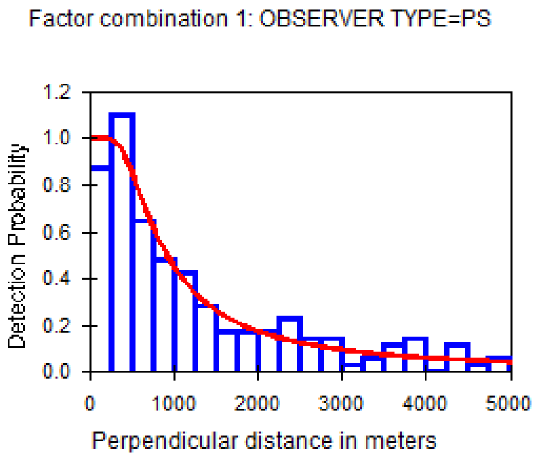


Figure 21. Detection functions for each team, with observer type as covariate.

PS	SS

All teams



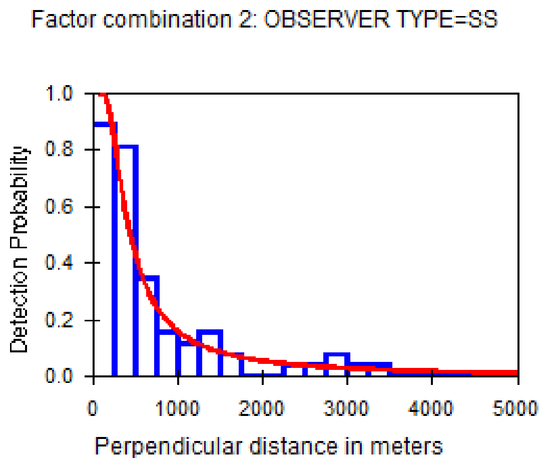


Figure 22. Detection functions for each team, with observer type as covariate.

Looking at all teams pooled together, the overall pattern is good for both types of observers, but with some differences. It is obvious that SS tend to search closer to the track line more often while PS tend to search all the way to long distances but lacking some effort in the shortest distances (first 250m), which is not a desirable effect.

When looking at each team individually, there are large differences in the search pattern between two types of observers, within each team and among teams, as already described in other sections of the report and in the previous report in October 2015. In Action-Air, there is a complete stop of searching effort after 1000m for PS and 500m for SS (even lower in the closest distances). In Air-Med there is not a defined pattern by the PS, who seem to be searching more or less equally at all distances with an important and undesirable lack in the shortest distances; while SS have a better searching pattern with more searching effort in the shortest distances. In Perigord, both PS and SS show a better pattern than Air-Med, although PS again lack effort in the closest distances. Finally Unimar has similar problems as Air-Med for the PS, but too few observations by SS to draw any conclusion.

Conclusion

School size

When sample size is very small (like in the stratified by sub-areas analysis), differences in the estimated school size between PS and SS yield differences in the abundance estimates which in some cases can be very large. See for example A inside, where the estimate from SS doubles that from PS. When all areas (and therefore all observers with their intrinsic differences) are pooled together, increasing also sample size (although still very small in this exploration), the differences become smaller.

The differences in school size estimation among observers is clearly an issue that can affect abundance estimates and create biases, which may be larger in some areas than in others, depending on the precision of the observers, but it is not possible to know if the biases are downward or upward as the ground truth of the actual school sizes is unknown.

Encounter rates

PS prove to be much more efficient in finding schools of BFT than SS, presumably due to their larger experience in this task.

Distance

It is clear that in most cases PS tend to look further away than SS, despite having repeatedly insisted that the most important observations for the analysis are those at shortest distances. The longest ones will be discarded from the analysis through truncation, and it is the shortest ones, in the first bins close to the track, the ones that define and shape the detection function and therefore those which will influence the most the resulting estimates of abundance. This searching pattern by the PS needs to be corrected in a way that they search more at closer distances.

II.3 Define individual biases and CVs, to be used as correcting factors if sufficient data are available.

No sufficient data are available for this analysis, because there are too many different observers, from different teams, and with different degrees of experience. Table 15 shows the list of observers from all years, classified by Team and type of observer, and providing the total observations of BFT and of all species pooled together made by each observer. Figure A1-1 (in Annex 1 for space issues in the main report) shows the plot of histogram of perpendicular distance at detections for each observer and the scaled overall detection function with observer as the unique covariate, both for BFT and for all species.

The number of observations of BFT made by observer is very variable and generally very low in most cases. The small sample size gives much weight to the random factor and makes difficult the observation of a real pattern. Therefore an appropriate exploration of the searching patterns of individual observers would only be possible for a few of them (see Table 15 and Figure A1-1). The number of observations of all species is much larger, but also extremely variable among observers, with only a few with good sample size. Hence, in practice it is impossible to make reliable comparisons.

Even having much larger sample size of observation by pooling all species together, it is obvious that there is a very different observing pattern between Professional and Scientific Spotters. Professional spotters tend to concentrate on BFT and do not seem to put much attention to other species, while Scientific spotters tend to pay attention to other species too. BFT corresponds, in average, to 56% of the total observations by Professional Spotters, while they are only 27% of all the observations by Scientific spotters, indicating that SS look and record all species they observe while PS seem to ignore in many cases the non-BFT species. Therefore, using all species for trying to make comparisons among PS or between PS and SP could lead to important biases.

Furthermore, no correction factor can be estimated, partly because the small and highly heterogeneous sample size per individual observer, but also because no ground truth information is available in terms of true school size and true distance or any other parameters.

The only way to attempt to estimate a correction factor would be with a calibration experiment in which the ground truth is known and compared with the estimations of the different observers. In terms of distances, for example, the ground truth could be the real distance from the airplane position (GPS) to known buoys or other objects at sea with known coordinates. The problem would be the resources to do such experiment with so many observers as they are. In terms of group sizes, a possibility would be to calibrate the observers' estimates through testing their estimates of the group sizes in pictures (seen during the same approximate time as a tuna school can be observed from air during surveys). But even if these calibrations can be made, they would need many replicates to obtain enough sample size to obtain such an estimated calibration. The second step would be to apply such calibration to the actual estimations obtained at sea (where in many cases the sample size is very small too). Much uncertainty would be accumulating during the process as to ensure the reliability of the corrections, so the cost-efficiency of such process should be evaluated and seems very poor anyway. So far, no one in the world has never attempted the calibration of so many observers at the same time and the many different difficulties are very clear. A first attempt to provide a comprehensive SWOT analysis of this problem was provided by Di Natale in 2015 (in press), while an extensive discussion took place at the SCRS BFT Species Group in 2015, where the most expert and experienced scientists supported the results of the SWOT analysis, excluding any realistic calibration with so many observers and taking into account the many nationalities, recommending to possibly use the same spotters in the same area over the years, for keeping the bias as close as possible to a constant level.

Table 15. List of observers from all years, classified by Team and type of observer. The total observations of BFT and of all species pooled together made by each observer is provided.

Observers	ID Obs	2010	2011	2013	2015	Total	BFT	All species	Proport.
ActionAir				**3**	**12**	**15**			
Pilot				**1**	**3**	**4**			
Bertrand Badie	4				x	1		2	0
Daniel Walther	5			x	x	2			
Patrick Feron	16				x	1			
Professional Spotter				**1**	**3**	**4**			
Khalifa Zariohi	24			x	x	2	4	13	31
Rosario Migliore	30				x	1	3	17	18
Silvere Fontanet	32				x	1	17	38	45
Scientific spotter				**1**	**6**	**7**			
Aurélie Monthy	41				x	1	3	88	3
Devis Monthy	45				x	1	4	98	4
Josselin Juan	51			x		1			
Manoelle Chauveau	53				x	1		1	0
Manon Joguet	54				x	1			
Pierre Yves Le Guern	60				x	1		3	0
Rémi Allard	62				x	1		4	0
Air-Med		**9**	**5**	**5**	**4**	**23**			
Pilot		**3**	**1**	**2**	**1**	**7**			
Alain Echevarría Aguirre	2				x	1	1	2	50
Francisco J. Hevia Busoño	9	x		x		2			
Ignacio Cubedo	10		x			1			
Iñigo Ortiz de Zarate	11			x		1	1	15	7
Miguel Salguero	13	x				1	1	1	100
Roberto Corral	18	x				1			
Professional Spotter		**2**	**1**	**1**	**1**	**5**			
Carlos dos Santos	21	x				1	21	28	75
Luis Navarro Martínez	26	x	x	x	x	4	73	186	39
Scientific spotter		**4**	**3**	**2**	**2**	**11**			
Ariadna Purroy Albert	40			x		1	4	116	3
Edwin Zammit	46	x				1	2	9	22
Joan M. Sorrell Baron	47	x	x			2	18	20	90
José A. Vázquez Bonales	48				x	1	7	310	2
Jose Luis Cort	49	x				1			
Joseph Borg	50	x				1			
M'Hamed Idrissi	57		x			1			
Miguel Cojan Burgos	58		x			1	1	1	100
Ohiana Revuelta Avín	59				x	1	6	229	3
Samantha Bowgen	63			x		1	6	107	6
Périgord		**6**	**5**	**8**		**19**			
Pilot		**1**	**2**	**2**		**5**			
Alexandre Kratz	3	x	x	x		3	7	10	70
Fabien Roque	6			x		1	1	3	33

Olivier Modene	14		x			1			
Professional Spotter		**3**	**1**	**3**		**7**			
Alessio Maglio	20			x		1	13	174	7
Francisco Bilora Valv.	22	x				1	82	96	85
José Molina	23	x	x			2	5	5	100
Lino Vassalo	25	x				1	5	5	100
Olivier Rigal	36			x		1			
Yoan Aucante	35			x		1	34	47	72
Scientific spotter		**2**	**2**	**3**		**7**			
Benjamin Charreyre	42			x		1	3	76	4
Daniel Grima	43		x			1	6	6	100
Deborah Belleney	44	x				1			
Kevin Cuevas	52			x		1	12	20	60
Ramonet Morgane	61	x				1	2	2	100
Sébastian Stradal	64		x			1	7	7	100
Vincent Bretille	67			x		1	13	218	6
Unimar		**5**	**5**	**9**	**12**	**31**			
Pilot		**1**	**1**	**2**	**4**	**8**			
Francesco Orrico	7				x	1			
Francesco Ruggiero	8		x	x	x	3			
Marco Fitz Seung	12				x	1			
Paolo Di Gruccio	15				x	1			
Pietro Fattiroso	17	x				1			
Stefano Kirchmayr	19			x		1			
Professional Spotter		**1**	**1**	**3**	**3**	**8**			
Michelangelo Cirillo	27			x		1	6	12	50
Pietro Paolillo	28				x	1	8	12	67
Raffaele Giordano	29			x		1	18	26	69
Salvatore De Martino	31	x		x	x	3	14	38	37
Vincenzo Consiglio	33		x			1	1	1	100
Vincenzo Severino	34				x	1		3	0
Scientific spotter		**3**	**3**	**4**	**5**	**15**			
Adriano Mariani	37	x	x		x	3			
Andrea Fusari	38			x	x	2	5	33	15
Andrea Poggi	39			x		1		3	0
Marco Dell'Aquila	55	x				1			
Massimiliano Valastro	56	x	x		x	3		3	0
Sergio Lombardo	65			x	x	2	1	44	2
Simone Serra	66		x	x	x	3	1	13	8
TOTAL		**20**	**15**	**25**	**29**	**89**	416	2145	50
	n						38	45	45
	Min						1	1	0
	max						82	310	100

III. Introducing correction factors into the estimates

III.1 Assess the different number of sightings at each distance category between aircraft with bubble windows and aircraft without bubble windows, possibly identifying correction factors to be applied for previous surveys when bubble windows were not mandatory.

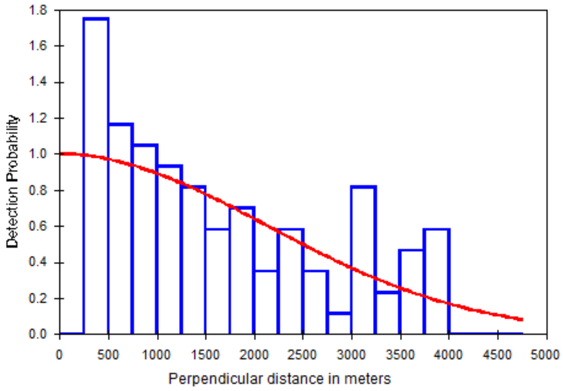
In practice, the absence of bubble window precludes the observation of approximately 86m from the track line, below the airplane (depending on the height of the airplane), even it is not excluded that some sightings along the main line of the track can be done by the spotters (pilot and professional observer) sitting on the front, when looking forward the nose of the aircraft.

In 2010 there were no bubble windows in any of the airplanes, but the minimum distance recorded this year was 254m, indicating a not very good searching pattern by the observers, not concentrating on the shorter distances. The rest of the years there are observations with and without bubble windows, as only the rear windows had bubble while the pilot and co-pilot in front had flat windows. The frequency of observations in the closest range (the range allowed by bubble windows in comparison with normal windows, 0-86m) varies among years and between the two types of airplanes, with higher frequency in Cessna and in 2013. However, the sample size is too small for proper comparison (see Table 16).

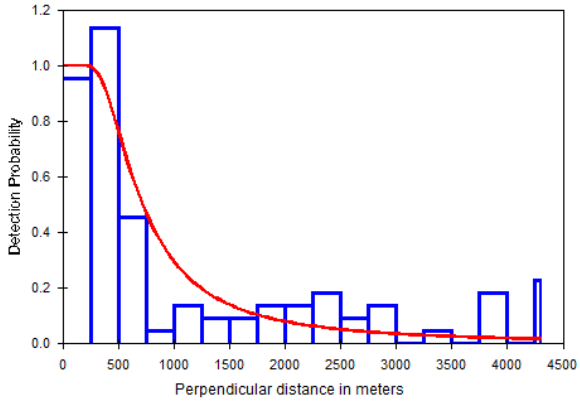
Table 16. Distribution of perpendicular distances at observations with (Y) and without (N) bubble windows.

Year	Perpendicular distance	Cessna		Partenavia		Total
		N	Y	N	Y	
2010	0-86					
	87-250					
	251-400	2		6		8
	401-550	4		5		9
	551-700	4		2		6
	>700	41		36		77
2011	0-86		3		1	4
	87-250	1	15			16
	251-400		23	2		25
	401-550		3	2		5
	551-700		8			8
	>700	5	10	9	9	33
2013	0-86		9		1	10
	87-250	1	3		4	8
	251-400	1	9	1	2	13
	401-550	2	2	3		7
	551-700	1	6	1		8
	>700	11	20	11	45	87
2015	0-86		1		1	2
	87-250	9	2	1	3	15
	251-400	6	1	2	3	12
	401-550	2	3	4	1	10
	551-700	4		2	1	7
	>700	2		30	5	37
	Total	**96**	**118**	**117**	**76**	**407**

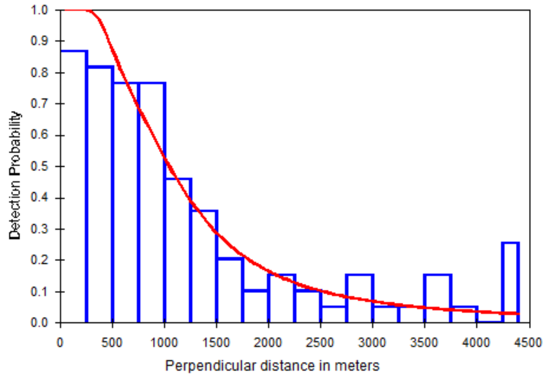
In general, it is expected that the detection functions would have a shoulder in the short distances so that probability of detecting animals in the short distances is fairly stable up to a certain distance when it starts falling down. When looking at the detection functions and perpendicular distances frequencies for each year (all areas and team pooled together; Figure 23), for all years, except 2013, this “falling down” seems to be occurring quite sharply beyond 500m of perpendicular distance from the track line. In 2013 the sharp drop seems to occur beyond 1000m. In the figure for 2010 the gap in the first 250m is visible.



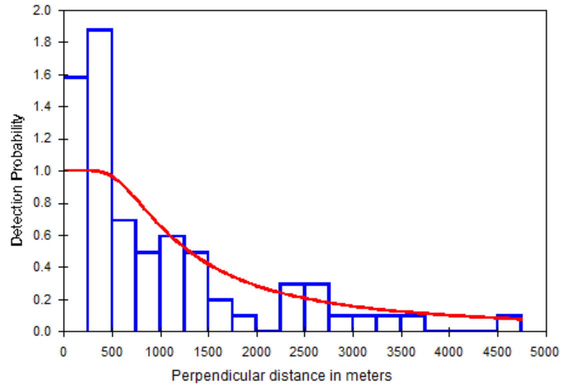
2010



2011



2013



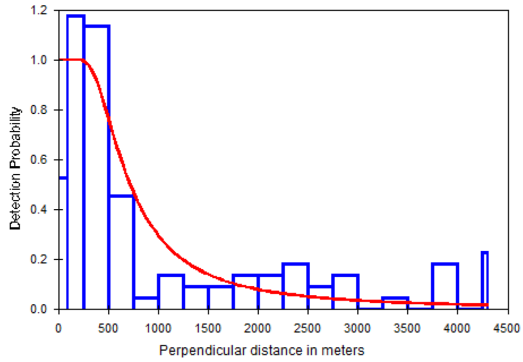
2015

Figure 23. Detection functions and perpendicular distances frequencies for each year (with and without bubble windows, and all teams and sub-areas pooled together).

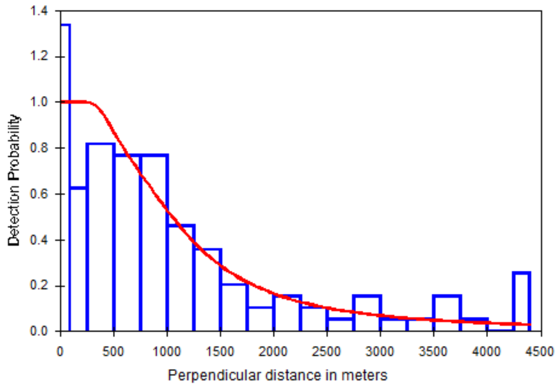
The relative frequency of observations in the closest range (86m), corresponding to the blind sector without bubble windows, changes with year, being 2013 the year with highest frequency (see Figure 24). In 2013 the frequency of observations is much higher in the first 86m than in the rest of the first bin of 250m, while in 2011 and 2015 the frequency is much lower than in the rest of the bin until 250m. Reasons for this difference are unknown.

In this Figure it is obvious that there is not a consistent pattern of relative frequency of observations in the range visible only from bubble windows (86m). Therefore, it would be speculative to make inferences on what such frequency could be in 2010 for the first 86m. However, the first 250m of perpendicular distances are relatively similar to the 250-500m when detection usually drops as mentioned before, therefore it may be relatively safe to assume that also in 2010 the missing 0-250m might be similar to the 250-500m bin.

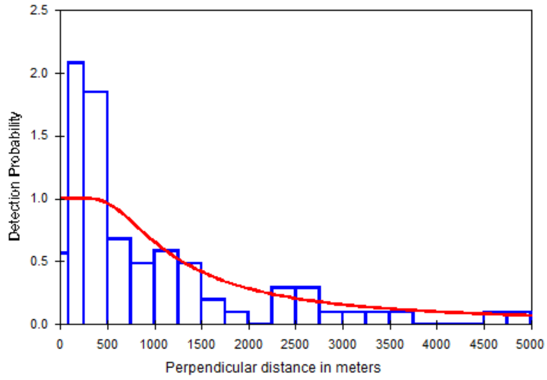
For the previous analysis (see report October 2015), left truncation was applied at 250m in the analysis of 2010 to account for those missing data. Left truncation eliminates the area that has not been searched from analysis, i.e. only data beyond the left truncation distance is used, so the detection function is fit only to these data and is extrapolated back to distance zero, with a hazard-rate function that creates a shoulder at sort distances. It is not possible to know whether this left truncation may be creating a bias upwards or downwards but it was considered the safest approach, usually used in aerial surveys with no bubble windows.



2011



2013



2015

Figure 24. Detection functions and perpendicular distances frequencies for each year (with and without bubble windows, and all teams and sub-areas pooled together), separating the first 86m of blind sector for no bubble windows. 2010 is not represented as it would be the same figure as in Figure 23.

Conclusion

Given the variability of relative frequencies in the blind sector for bubble windows (86m) and the use of left-truncation distance in 2010 accounting for those missing data, we consider that no further correction factor is possible to obtain reliably to improve the estimates of 2010.

III.2 Use data provided by mini-PATs for Bluefin tuna passing through the areas sampled by the aerial survey in the same period of time and year, in order to assess an average time at the surface (sighting corrector factor) for each internal area, if sufficient data are available.

Density estimates from line transect surveys are usually subject to “availability bias” due to animals not always being available for detection while within detectable range (Buckland *et al*. 2004), and to “perception bias” due to observers failing to detect animals even though they are available (Buckland *et al.*, 1993), causing both a negative bias.

What will be explored in this point is the availability bias, through the estimates of the average time BFT spends at or near surface and therefore available to be seen by the spotters. However, there are no means to explore the perception bias given the amount and diversity of experience of spotters, and the small amount of observations by most of them (see point III about individual biases by spotters).

Data organization

Electronic tagging data for the period 2011-2015 (ICCAT GBYP Phase 2, 3, 4 and 5) were provided by ICCAT. Data were prepared by ICCAT beforehand indicating the percentage of time, in periods of 6 hours in average, that the observed fish spent in different depth bands, during the observed time period. For this purpose, sea water column was divided in layers (bands) i.e. depth categories, like category 0 (0 m), category 2 (0.1-2m), category 10 (2.1-10m) and so on. The sum of time percentage values for all depth categories in one observed time period is always 100 (%).

For the scope of this exercise, only depths up to 10m were considered, given that, according to inquiries to some observers, animals at greater depths are much more difficult to spot. According to them, also, when tunas are at 10m or less, they are actually occupying the whole column from surface to 10m or more, therefore, the total sum of percentage of time spent at 0m, from 0 to 2m and from 2 to 10m was used.

All data points in 2011, 2013 and 2015 were associated to their corresponding survey sub-areas during those years. All data points were also associated to the overlapped sub-areas.

Data was filtered for the months of June and July, the survey months, and further filtered in different ways for exploration. Table 17 shows these different filtering options and the sample size remaining for each of them.

Table 17. Different filterings applied and sample size remaining for day, night, or both.

Months	Years	Sub-areas	Only day	Only night	Day and night
June-July	2011-2015	All areas	1,770	1,152	2,922
		Overlap	154	166	320
		Original inside	190	349	539
	2011, 2013, 2015	Overlap	144	162	306
		Original inside	190	349	539

Data analysis

Exploration of pooling and stratification

As a first step, the average, and associated CVs, percentage time spent from surface to 10m depth was calculated for each period of 6 hours of each tag, year and sub-area, as well as for all the tags within each overlap sub-area/year pooled together, each year (across sub-areas) pooled together and each overlap sub-area (across years) pooled together, and all tags, years and sub-areas together, for day time and for night time, to explore if there were differences among them or whether they could be pooled together in order to increase sample size and reduce CV. The same procedure was followed using the original sub-areas before overlapping. Table 18 shows the available data and the averages and CVs of the percentage time spent at 0-10m depth for June and July all years without considering the sub-areas; Table 19 shows the same results for the overlap areas for day time and Table 20 for night time; and Table 21 shows the same results for the previous areas for day time. The sampling unit is each measure of a 6 hours period for each tag.

Table 18. Number of sampling units (n) per year in June-July, mean percentage of time spent at 0-10m depth (Mean %) and Coefficient of Variation (CV). Day time, night time and whole period are presented.

	Day			Night			Total		
	n	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**
2011	81	68.8	29.7	61	60.2	36.4	142	65.1	32.9
2012	216	52.9	58.6	73	49.1	63.2	289	51.9	59.7
2013	1219	47.2	69.0	563	57.5	56.9	1782	50.5	65.3
2014				6	61.8	16.8	6	61.8	16.8
2015	254	58.8	37.2	449	59.4	32.6	703	59.2	34.3
Total	**1770**	**50.6**	**61.6**	**1152**	**57.9**	**47.6**	**2922**	**53.5**	**56.1**

Table 19. Number of sampling units (n) per year in June-July during day time for the overlap sub-areas, mean percentage of time spent at 0-10m depth (Mean %) and Coefficient of Variation (CV). No tags were available for sub-area G during day light. Tags in 2011 were outside the overlap sub-areas during day-light.

	A			C			E			TOTAL		
	n	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**
2013	46	62.4	47.3	9	60.8	30.3	41	29.3	86.9	96	48.1	65.1
2015	13	60.6	22.1	8	63.2	27.0	27	55.3	32.8	48	58.0	28.9
Total	**59**	**62.0**	**43.1**	**17**	**61.9**	**27.9**	**68**	**39.6**	**65.8**	**144**	**51.4**	**53.8**

Table 20. Number of sampling units (n) per year in June-July during night time for the overlap sub-areas, mean percentage of time spent at 0-10m depth (Mean %) and Coefficient of Variation (CV).

	A			C			E			G			TOTAL		
	n	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**
2011	4	59.4	9.8										4	59.4	9.8
2013	16	58.8	41.0	3	65.6	4.1	14	32.8	72.2				33	48.4	54.2
2015	13	49.6	31.3	8	59.6	37.3	25	51.7	28.5	79	70.4	23.7	125	63.8	29.2
Total	**33**	**55.3**	**35.6**	**11**	**61.2**	**30.8**	**39**	**44.9**	**45.3**	**79**	**70.4**	**23.7**	**162**	**60.6**	**34.8**

Table 21. Number of sampling units (n) per year in June-July during day time for the original inside sub-areas, mean percentage of time spent at 0-10m depth (Mean %) and Coefficient of Variation (CV). No tags were available for sub-area G during day light in any year. Tags in 2011 were outside the overlap sub-areas during day light.

	A			**C**			**E**			**TOTAL**		
	n	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**
2013	46	62.4	47.3	9	60.8	30.3	41	29.3	86.9	96	48.1	65.1
2015	13	60.6	22.1	8	63.2	27.0	29	54.1	34.8	50	57.2	30.4
Total	59	62.0	43.1	17	61.2	27.9	70	39.6	65.5	146	51.2	53.9

In a first approach, looking at Table 19 and 21, the only difference is 2 sampling units more in the second one, i.e. 2 observations in 2015 in sub-area E inside, but outside the overlap areas. Therefore, and as the two extra observations do not change the results but increase slightly the CV, the first option was chosen, i.e. using the sub-areas inside with overlapping.

When comparing sub-areas by year (Table 19), A and C are very similar in 2013 and 2015 for day time, and are also very similar between them, therefore, it was decided to pool together A and C 2013 and 2015 to increase sample size and decrease (substantially) the CV, resulting in 76 sampling units with a mean percentage of time spent between 0 and 10m depth of 62% and a CV of 40.0%. At night time, A and C 2013 and 2015 had 44 sampling units, with a mean percentage of time spent between 0 and 10m depth of 56.8% and a CV of 34.3%. The mean at night is lower than during day time by around 8%, and the CV is also a bit smaller. Pooling together A-C 2013-2015 day and night time gives 120 sampling units (including 4 from 2011 with similar mean), with a mean percentage of time spent between 0 and 10m depth of 60.1% and a CV of 38.3%. A non-parametric U Mann-Whitney test was used to test for significant differences between day time and night time in A-C 2013-2015, and the null hypothesis was not rejected, i.e., there are no significant differences between both sets of data, day and night. This lack of difference during June and July may be due to the spawning behaviour during those months, with little or no feeding behaviour. However, given the paper by Aranda et al (2013) in the Balearic Islands showing some more deep diving behaviour during the day in July, the same test was done just for sub-area A, and still no significant differences were found. However, given the very little gain in CV going from using only day to using day and night, the fact that surveys were done only during day time, and that the mean time spent at 0-10m depth is lower at night (despite not being significant differences), led us to consider that the best and precautionary option was to use A-C 2013-2015 only at day time.

Sub-area E during day time seems to be more variable and in general with a lower percentage of time spent between 0 and 10m depth than the other sub-areas, and the same happened at night time, therefore this sub-area was kept separated. At the same time, 2013 and 2015 yielded very different mean percentages of 0-10m depth time both during day and during night. On the other hand, there are no significant differences between day and night for E 2013 and for E 2015. However, the CV in 2015 remains basically the same when pooling together day and night time (32.3%) compared with only day (32.8%), so there is no gain in pooling them together, and following the same reasoning as for A-C, only day was kept. In the case of E in 2013, there is a decrease in CV from 86.9% using only day to 82.4% when pooling together day and night, being the difference in mean time spent at 0-10m depth very small. Unfortunately, the CV is very large in any case, making this estimate rather useless (although still the only available correction estimate). Therefore, to keep consistency with the previous reasoning and choices, it was decided to keep day time too.

Sub-area G only has data for 2015 during night (Table 20). Give the lack of significance between day and night in the other areas, this information is taken as the best available, considering the original G inside sub-area before overlapping to increase sample size.

Table 22 shows the final pooling and stratification done (4 values: A-C day-night 2013-215; E day-night 2013; E day-night 2015; G night 2015) in order to estimate availability bias, based on the existence or lack of significant differences.

Table 22. Number of sampling units (n) per year in June-July during night time for the overlap sub-areas, mean percentage of time spent at 0-10m depth (Mean %) and Coefficient of Variation (CV).

	A-C day			E day			G night		
	n	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**	**n**	**Mean %**	**CV (%)**
2013				41	29.3	86.9			
2015				27	55.3	32.8	79	70.4	23.7
Total	**76**	**62.0**	**40.0**						

Conclusion

Given the speed of the observation platform on an aerial survey, it is assumed that the survey is virtually instantaneous and therefore the forward distances that would affect availability bias (Borchers et al. 2013) in a ship-based survey do not affect, or does it very little, the availability bias from the aerial survey. In other words, if the observers can see animals when they are more than a short time ahead of the plane, then g(0) estimation (at any perpendicular distance) is more complicated than just the proportion of time the animals are available for detection. Here "short time" means a short time relative to the tunas' average time diving below 10m depth (estimated depth for detection from the aircraft), which is the case as observers are not able to detect the animals more than very few seconds ahead of the aircraft due to its speed. This is because (for any perpendicular distance) the probability of seeing them at the forward distance that they were seen is a combination of the probability of missing them until that forward distance, and the probability of seeing them there, given they were available and as yet unseen at that forward distance. And the probability of missing them depends on their availability over the whole period until they were seen, not just their availability at the point they were seen. However, in our case the time the observers are able to detect the BFT schools ahead of the plane is very short (a few seconds at maximum) compared with the available-unavailable cycle of the BFT, and therefore it is considered practically instantaneous.

Table 23 shows the g(0) estimates and their respective CVs.

Table 23. Estimates of g(0)

Day / Night	Sub-area	Years	G(0)	CV (%)
Day	A-C	2013-2015	0.380	40.0
Day	E	2013	0.707	86.9
Day	E	2015	0.459	34.8
Night	G	2015	0.302	23.7

It is important to highlight that the g(0) estimated here is, as stated at the beginning of this point, only based on the availability bias. Perception bias is still unknown and therefore there is still a negative bias unaccounted for in the estimation of g(0).

III.3 Assess the bias induced by some of the most relevant environmental factors (i.e.: wind <2 Beaufort, >2 Beaufort, glare).

Data organization

To explore the effect of environmental (but also survey) factors on the probability for detection and therefore the estimates of abundance, two data sets were prepared:

a) observations of all species (BFT, other fish, cetaceans, turtles, sharks): total of 687 to 1725 observations (some years not all the variables were recorded)

b) observations of only BFT: total of 73 to 353 observations available (some years not all the variables were recorded)

The first dataset has a much larger sample size, so it could be useful to detect more differences, but the dataset with only BFT is important by being the target species, with its particularities[4]. In both cases all data available from 2010 to 2015 were used.

Data analysis

Two types of approach were done to explore this issue. No modelling with GAM or GLM could be done, as would be ideal (response variable encounter rate and covariates the factors to be explored – Beaufort, glare, etc.), given that the covariates are all factors and no continuous variables. The two approaches were:

a) a simple Chi-square analysis to evaluate whether the difference between the observed and expected frequencies of observations in the different levels of the variables were significant. This is to explore the effect of this variables in the encounter rates (number of observations per km of searching effort).

b) detection functions for each of the levels of the variables to evaluate if they make a difference in the global detection functions. This is to explore the effect of the variables in the probability of detection

<u>Chi-square analysis</u>

The test was done for Beaufort, Glare Intensity, Turbidity, Clouds, Subjective, Airplane and Team, both for all species together and only for BFT. The expected frequencies in all cases were based on the proportion of on-effort (km) in each level of each variable. Yates corrections was applied to contingency tables with only one degree of freedom. Tables A2-1 to A2-8 in Annex 2 show these contingency tables for all variables. An arbitrary degree of "Effect" was assigned to each level of each variable depending on the difference between the expected and the observed values in each category; if the difference in the category is more than ±50% of the expected value, a 'big effect' was assigned (either positive or negative); if the difference in the category is between ±50% and ±25% of the expected value, a 'medium effect' was assigned; if the difference in the category is between ±25% and ±10% of the expected value, a 'small effect' was assigned; if the difference in the category is less than ±10%, no effect is assumed.

For most variables, the results for all species and for BFT are in agreement, although in some there are some variations.

- *Beaufort*: this is the variable with more disagreement between all species and BFT. For all species the result is somehow puzzling as it shows a 'medium' negative effect in the best Beaufort sea state 0 while a strong positive effect of sea state 2. However, as expected, there is a negative effect of the worse Beaufort sea state (4). The result, overall, is highly significant. When considering only BFT, however, the best sea state from 0 to 2 have a small positive effect (or null), while the worse (3 and 4) have a medium and strong negative effect respectively, more accordingly to what would be expected. Nevertheless, in the case of BFT the result is less significant than for all species, but still significant.

- *Glare*: in this case results are similar for all species and for only BFT. Surprisingly, there is a strong positive effect of medium glare on the encounter rate, both for all species and for BFT, with a negative effect of stronger and lighter glare intensity. It is be understandable a negative effect of strong glare, but the reasons for a strong positive effect of medium glare and negative effect of none or little glare are unknown. In both datasets the result is highly significant.

[4] Of course, the different surfacing behaviour between a fish species and other marine species (such as marine mammals or turtles) may affect the sightings.

- *Turbidity*: For all species, the clearest waters did not have an effect (or slightly positive), but the waters with some turbidity, from little to strong, had a strong negative effect, with high significance. For BFT, the waters with none or little turbidity had a null or medium positive effect respectively, but the result is not significant.

- *Haze*: For all species, the only strong effect was negative for thickest haze. Only none or little haze had small positive or no effect, respectively. For BFT, no haze had a medium positive effect, while having haze had small or medium negative effect. In both cases results are highly significant.

- *Clouds*: This factor is not significant in the encounter rate of BFT, although it is for all species together, with a small positive effect of mostly clear and very cloudy skies and medium negative effect of moderately cloudy skies.

- *Subjective*: The subjective category is highly significant for all species, with positive effect of good conditions and negative effect of moderate to poor conditions. This factor is not significant for BFT.

- *Airplane*: Both in the case of BFT and all species, Cessna has a medium positive effect and Partenavia a small or medium negative effect.

- *Team*: Both for all species and for only BFT, the result is highly significant, with medium to strong positive effect for Air-Med and Perigord, and medium to strong negative effect for Unimar and Action-Air.

<u>Detection functions</u>

The detection functions included both sightings On and Off effort. Ranking the detection functions by their AIC (Akaike Information Criterion), from best fit to the worse, we got:

- For all species: Subjective > Clouds > Glare > Turbidity > Haze > Beaufort > Airplane > Team
- For BFT: Subjective > Clouds > Glare > Haze > Turbidity > Beaufort > Team > Airplane

Looking at the p-values of the Cramer von Misses GOF (Goodness of Fit), which indicate whether the differences between the fitted function and the observed data at the short perpendicular distances close to the track line are significant or no, they yield a relatively similar ranking for all species, but a different one for BFT:

- For all species: Subjective = Clouds (p=0.50) > Turbidity = Haze = Team (p=0.05) > Beaufort (p=0.025) > Glare (p=0.005) > Airplane (p=0.001)
- For BFT: Team (p=0.9) > Subjective (p=0.8) > Glare = Turbidity (p=0.6) > Haze (p=0.5) > Beaufort (p=0.3) > Airplane (p=0.1) > Clouds (p=0.005)

In the case of BFT, all the GOFs show a good fit, with no significant differences, except for Clouds. In the case of all species, only Subjective and Clouds had no significant differences between fitted and observed data. The pattern is the same when considering the Kolmogorov-Smirnoff GOF, which measures the difference between the fitted function and the observed data all over the whole range of perpendicular distances. But we need to be careful with these comparisons as the variables Subjective and Clouds were only recorded in 2015, while the others for all the years.

The covariates in a detection function affect the scale parameter (not the shape), so the coefficient for each level of the covariate gives an indication on how close or far from the trackline are the detections under the conditions of such level of the covariate.

Figures A2-1 to A2-8 in Annex 2 show the detection functions for the different levels of covariates tested, both for all species and for BFT. The coloured lines in the last plots of each figure show the effect on the scale parameter of each of the levels of the covariate, and the global detection function (in blue) for comparison.

Conclusion

- *Beaufort*: This is not one of the most influencing variables, and results don't show a consistent pattern. This is probably due to the fact that effort was constrained to moderately good sea state conditions up to

Beaufort 4 max[5]. The detection function incorporating only Beaufort was not very significant (and the green line in Figure 7 for BFT corresponds only to 1 observation off effort, so not worth considering it). It seems that, although small and as expected, better sea states have a better effect on the observation rates than worse sea states (Table X), although its effect of the detection function is very small, with slightly longer perpendicular distances in average with better sea states (Figure 7). Looking at these results, it may be advisable to reduce the on effort searching time up to Beaufort 3 (as stated by the contract) instead of 4 giving the negative effect of this worse sea state.

- *Glare*: The results for glare intensity are also counter instinctive. There were less observations than expected with no or little glare intensity and many more with moderate glare (glare 2). On the other hand, with this moderate glare the observations were done in average at slightly closer perpendicular distances than the no or little glare which were done in average at larger distances, not because at moderate glare there were more observations further away, but because with none or little glare there were many more at shorter distances. Intense glare diminishes the encounter rate as expected. Maybe the moderate glare has less effect at short distances (depending on the angle of the sun, presumably) so observers tend to look closer to the track line where the glare is less intense, increasing the chances of finding groups which in theory should be easier to detected at shorter distances. When there is none or little glare, possibly the observers tend to look further away from the track line and also loosing concentration in the closer areas where the probability of detection is greater. Therefore, if this theory is true, the glare intensity actually changes the searching behaviour of the observers, demonstrating at the same time that when forced to look closer to the track line the encounter rate increases considerably compared to when they look further away.

- *Turbidity*: With moderate and strong turbidity there are practically no observations, although these conditions were recorded only during less than 2% of the searching effort. Therefore only null or little turbidity had significant amount of searching effort. When looking only at BFT, there are no significant differences between these categories, but when looking at all species even small turbidity has a very negative effect on encounter rates compared to no turbidity at all. In any case, for BFT this variable does not seem to have an effect on detection.

- *Haze*: In this case, it is clear, as expected, that the encounter rate both of BFT and of all species is better with no haze and worse with increasing thickness of fog. In terms of distance of detection, it does not seem to have an effect when considering all species together, but it does when looking at BFT only. However, the two more extreme scale effects of the different levels of haze, 3 and 4 (the thickest), have only 3 and 7 observations respectively, and therefore those lines might be so extreme just due to the small sample size. With moderate haze (2) the observations are in average closer to the track line, maybe due to the same effect as with glare intensity (being easier to search at closer distances), while with none or little haze searching effort is more spread-out in terms of distance from the track line. In this case, even if with moderate haze there is a tendency to look more at closer distances, the haze itself has a negative effect on the encounter rate at all distances, while no haze at all facilitates detection everywhere.

- *Clouds*: the coverage of clouds does not have an effect on the detection of BFT, neither on the encounter rate nor on the distance of detection from the track line. But this lack of observable effect could be due to the small sample size as this variable was recorded only in 2015 and there were only 39 observations of BFT associated with a given cloud coverage.

- *Subjective*: As in the previous case, this variable was only recorded in 2015, with only 39 observations of BFT associated, which maybe the reason for the lack of significant differences in encounter rates. When looking at all species, it is clear that, as expected, "good" conditions increase the encounter rates above expected values, while the "moderate" and "poor" conditions decreases it. This should be further explored in future aerial surveys, thus having bigger sample size, given that it produces an effect when looking at all species, and because at least in terms of detection function there is an observable effect. There are more detections closer to the track line with "good" conditions while, in average, detections expand further away with "moderate" and "poor" conditions. But, again, this is based on very small sample size so this apparent effect should be taken cautiously.

- *Airplane*: Cessna has a positive effect on the encounter rate compared to Partenavia. In the case of BFT it could be argued that as the different airplanes were used in different areas, these effect may be due mainly

[5] The technical requirements, by contract and protocol, excluded the possibility to carry out the survey with Beaufort over 3. In some cases, sightings were made even with Beaufort 4 when the aircraft was already on effort.

to the different densities of BFT in the different areas. But the fact that the same effect is observed for all species pooled together suggests that it is not only a matter of density, but maybe the type of airplane has indeed some real effect for some reason. On the other hand, it does not have any effect on the distance of detection when considering all species together, but it does when looking only at BFT. With Cessna there is a tendency for detection at closer distances in average, and much larger with Partenavia in average. Whether this is an effect of different searching behaviours by the different teams using each airplane (see below) or due to the configuration of the airplane itself, it is unclear and unknown.

- *Team*: The effect observed for team (described above: positive effect on encounter rate for Air-Med and Perigord, and negative for Unimar and Action-Air) does not seem to be related with the type of airplane used by each team, as Air-Med (positive) and Unimar (negative) used Partenavia, while Perigord (positive) and Action-Air (negative) used Cessna. Some potential reasons for these differences could be: a difference in effectiveness among observers from each team; or a difference in density (BFT) of the areas surveyed by each team; or overall differences in searching conditions (environment) among the areas surveyed by each team. Overall, comparing the 4 years of survey, density in sub-area E is the largest and has been surveyed by Perigord (positive), Air-Med (positive) and Action-Air (negative); in sub-area C, density is lower, but higher than in A and G overall and it was surveyed by Unimar (negative); A and G had in average lower densities and they were surveyed by Perigord (positive), Air-Med (positive) and Action-Air (negative). Therefore, no clear pattern is observed here. Moreover, the same result is obtained when looking at all species together, in which case density is not an issue. When looking at the detection functions, the pattern is quite consistent when looking at all species or only at BFT. Action-Air has a tendency of searching closer to the track line, while Unimar further away. Perigord has in average intermediate distances. The case of Air-Med is different when looking at all species or only at BFT; for all species being a good curve in agreement with the global average, but for BFT with more observations further away in average than the other teams and not very good shape for the detections against perpendicular distances. Again, it is unclear how different casualties in presence or distribution of the various species could affect the analysis.

III.4 When average estimates are not available for one internal area/year, evaluate if estimates from other areas or years may provide a means to filling the gaps for missing years. This should take into account the information on the variability in across years and areas and the additional variance that such interpolation/extrapolations introduce into the estimate.

Table 24 shows the estimates of abundance for the overlapped areas every year, as reported in the report in September.

Table 24. Results for abundance of animals and weight for overlapped sub-areas inside for each year

Year	**2010**				**2011**			**2013**				**2015**			
Sub-area	**A inside**	**C inside**	**E inside**	**G inside**	**A inside**	**C inside**	**E inside**	**A inside**	**C inside**	**E inside**	**G inside**	**A inside**	**C inside**	**E inside**	**G inside**
Survey area (km^2)	61,933	53,868	93,614	56,211	61,933	53,868	93,614	61,933	53,868	93,614	56,211	61,933	53,868	93,614	56,211
Transect length (km)	6,277	8,168	12,621	2,900	7,975	8,466	9,806	6,743	2,682	3,720	1,716	4,119	2,658	4,484	785
Number of schools ON effort	8	6	29	33	10	10	45	10	10	20	12	6	3	13	2
Abundance of schools	27	13	73	216	57	47	316	31	67	168	131	46	31	139	74
%CV abundance of schools	56.2	46.6	32.7	29.4	35.9	33.4	24.1	36.1	34.3	34.0	40.7	43.3	62.7	29.6	68.9
Encounter rate of schools	0.0013	0.0007	0.0023	0.0114	0.0013	0.0012	0.0046	0.0015	0.0037	0.0054	0.0070	0.0015	0.0011	0.0029	0.0025
%CV encounter rate	54.6	44.6	29.9	26.3	33.8	31.2	21.0	35.0	33.2	32.9	38.7	41.1	61.2	26.2	67.5
Density of schools (1000 km^{-2})	0.430	0.248	0.775	3.840	0.922	0.868	3.374	0.495	1.244	1.794	2.333	0.749	0.580	1.490	1.309
%CV density of schools	56.2	46.6	32.7	29.4	35.9	33.4	24.1	36.1	34.3	34.0	40.7	43.3	62.7	29.6	68.9
Mean weight (t)	131.25	124.17	110.14	63.62	122.43	38.87	118.05	194.1	173.5	11.0	4.0	160.7	190.0	391.6	9.0
%CV weight	6.2	5.6	33.9	12.7	19.2	44.4	19.2	23.8	22.1	66.0	40.2	11.7	19.9	54.8	66.7
Mean cluster size (animals)		733	1,015		678.1	291	1,715	611	1,285	361	336	825	1,533	2,030	600
%CV abundance		36.5	19.0		27.9	30.7	21.5	26.0	17.0	67.3	36.7	11.0	19.0	56.8	66.7
Density of animals (km^{-2})		0.182	0.787		0.625	0.253	5.786	0.302	1.599	0.647	0.783	0.618	0.889	3.024	0.786
%CV density of animals		59.2	37.8		45.5	45.3	32.3	44.5	38.3	75.4	54.8	44.7	65.5	64.1	95.9
Total weight (t)	**3,496**	**1,658**	**7,995**	**13,733**	**4,296**	**1,999**	**39,344**	**3,572**	**11,830**	**1,882**	**534**	**8,736**	**5,965**	**54,889**	**666**
%CV total weight	56.6	46.9	47.1	32.1	46.2	54.9	32.2	40.6	40.9	74.3	57.2	41.9	65.8	62.2	95.8
L 95% CI total weight	1,218	678	3,284	7,387	1,775	689	21,147	1,640	5,365	486	181	3,956	1,776	16,632	73
U 95% CI total weight	10,037	4,056	19,464	25,532	10,398	5,794	73,198	7,780	26,081	7,284	1,574	19,296	20,034	181,140	6,070
Total abundance (animals)		**9,797**	**73,676**		**38,720**	**13,614**	**541,634**	**18,717**	**86,114**	**60,614**	**44,041**	**38,248**	**47,900**	**283,100**	**44,162**
%CV total abundance		59.2	37.8		45.5	45.3	32.3	44.5	38.3	75.4	54.8	44.7	65.5	64.1	95.9
L 95% CI total abundance		3,187	35,741		16,249	5,677	290,700	7,990	40,959	15,391	15,587	16,510	14,331	83,058	4,844
U 95% CI total abundance		30,016	151,880		92,266	32,649	1,009,200	43,845	181,040	238,710	124,440	88,610	160,100	964,970	402,600

The only missing sub-area/year was G inside in 2011, as well as the abundance of animals in A and E in 2010 (when only weight was provided).

Abundance of animals missing in 2010

In the case of abundance of animals missing in 2010, there exist the possibility of assuming an estimate of abundance of animals based on the other available information that year (estimated weight, encounter rate, density of groups) and the comparison of available results for the same area for all years. For example, for sub-area A inside (overlap area), see Table 25, and for Year 2010 comparing all sub-areas, see Table 26.

Table 25. Results for abundance of animals and weight for overlapped sub-area A inside for each year (from report September 2015).

A inside				
Year	**2010**	**2011**	**2013**	**2015**
Survey area (km^2)	61,933	61,933	61,933	61,933
Transect length (km)	6,277	7,975	6,743	4,119
Area searched (km^2)	18,602	10,846	20,207	12,499
Number of schools ON effort	8	10	10	6
Abundance of schools	27	57	31	46
%CV abundance of schools	56.2	35.9	36.1	43.3
Encounter rate of schools	0.0013	0.0013	0.0015	0.0015
%CV encounter rate	54.6	33.8	35.0	41.1
Density of schools (1000 km^{-2})	0.430	0.922	0.495	0.749
%CV density of schools	56.2	35.9	36.1	43.3
Mean weight (t)	131.25	122.43	194.1	160.7
%CV weight	6.2	19.2	23.8	11.7
Mean cluster size (animals)		678.1	611	825
%CV abundance		27.9	26.0	11.0
Density of animals (km^{-2})		0.625	0.302	0.618
%CV density of animals		45.5	44.5	44.7
Total weight (t)	**3,496**	**4,296**	**3,572**	**8,736**
%CV total weight	56.6	46.2	40.6	41.9
L 95% CI total weight	1,218	1,775	1,640	3,956
U 95% CI total weight	10,037	10,398	7,780	19,296
Total abundance (animals)		**38,720**	**18,717**	**38,248**
%CV total abundance		45.5	44.5	44.7
L 95% CI total abundance		16,249	7,990	16,510
U 95% CI total abundance		92,266	43,845	88,610

When looking at Sub-area A (overlap version, Table 25) to explore any temporal issue, it is fairly stable over time in terms of encounter rate of schools. However density of schools was double in 2011 compared with 2010, 2013 and 2015, while total weight was double in 2015 compared with the previous years, despite having a similar mean weight. At the same time abundance of animals was half in 2013 compared with 2011 and 2015, probably due to a much lower density of groups although similar mean school size than 2013, and similar density of schools and smaller mean school size than 2015.

When looking at year 2010 (overlap version, Table 26) to explore any spatial issue, no pattern is found when comparing the different sub-areas. Encounter rate of groups is highly variable among sub-areas, as is the density of schools and the available mean school size (C and E). Mean weight seems more homogeneous with some decrease from W to E, being in G half of the rest of the sub-areas.

In general, not enough clues can be found when looking at sub-areas in 2010 or at sub-area A across years, to extrapolate an estimate for the missing estimates of abundance of animals in 2010, independently of the degree of uncertainty associated to such extrapolation due to all the variances, additional variance included.

Table 26. Results for abundance of animals and weight for overlapped sub-areas in 2010 (from report September 2015).

2010				
Year	**A inside**	**C inside**	**E inside**	**G inside**
Survey area (km^2)	61,933	53,868	93,614	56,211
Transect length (km)	6,277	8,168	12,621	2,900
Area searched (km^2)	18,602	24,205	37,401	8,594
Number of schools ON effort	8	6	29	33
Abundance of schools	27	13	73	216
%CV abundance of schools	56.2	46.6	32.7	29.4
Encounter rate of schools	0.0013	0.0007	0.0023	0.0114
%CV encounter rate	54.6	44.6	29.9	26.3
Density of schools (1000 km^{-2})	0.430	0.248	0.775	3.840
%CV density of schools	56.2	46.6	32.7	29.4
Mean weight (t)	131.25	124.17	110.14	63.62
%CV weight	6.2	5.6	33.9	12.7
Mean cluster size (animals)		733	1,015	
%CV abundance		36.5	19.0	
Density of animals (km^{-2})		0.182	0.787	
%CV density of animals		59.2	37.8	
Total weight (t)	**3,496**	**1,658**	**7,995**	**13,733**
%CV total weight	56.6	46.9	47.1	32.1
L 95% CI total weight	1,218	678	3,284	7,387
U 95% CI total weight	10,037	4,056	19,464	25,532
Total abundance (animals)		**9,797**	**73,676**	
%CV total abundance		59.2	37.8	
L 95% CI total abundance		3,187	35,741	
U 95% CI total abundance		30,016	151,880	

Lack of survey in sub-area G in 2011

Table 24 above shows the results for abundance of animals and weight for overlapped sub-areas for each year. Just looking at sub-area G, the one not surveyed in 2011, the conclusion is the same as in the previous case: too much variability. Encounter rate and density of schools decreases progressively from 2010 to 2015. At the same time, mean weight is much higher in 2010 than in 2013 and 2015. There is no information on mean school size and abundance of animals in 2010, so the only information available is in 2013 and 2015, when abundance of animals is almost identical whilst mean school size is almost double in 2015 than in 2013. Table 27 summarizes these changes in terms of point estimates, and Table 28 shows the associated CVs (without considering additional variance). Figure 25 shows them graphically.

The very large CVs in 2015 are due to the extremely low sample size, with only 2 observations of BFT in sub-area G. In 2013 the CVs are also larger than in 2010 as the number of observations decreases from 33 to 12 (partly because of a reduction in effort from 2,900 km to 1,716km due to the time spent on surveying the outside areas).

Table 27. Summary of variability in point estimates in sub-area G across years.

Year	ER	Ds	Mean weight	Total weight	Mean school size	Abundance of animals
2010	0.0114	3.840	63.62	13733		
2011						
2012						
2013	0.0070	2.333	5.00	534	336	44,041
2015	0.0025	0.840	9.00	666	600	44,162

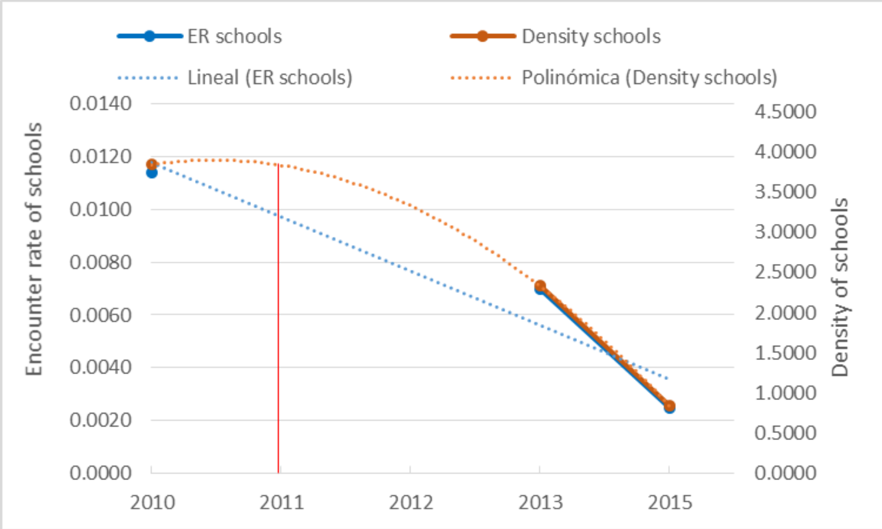
Table 28. Summary of the CVs (%) of the point estimates in sub-area G across years.

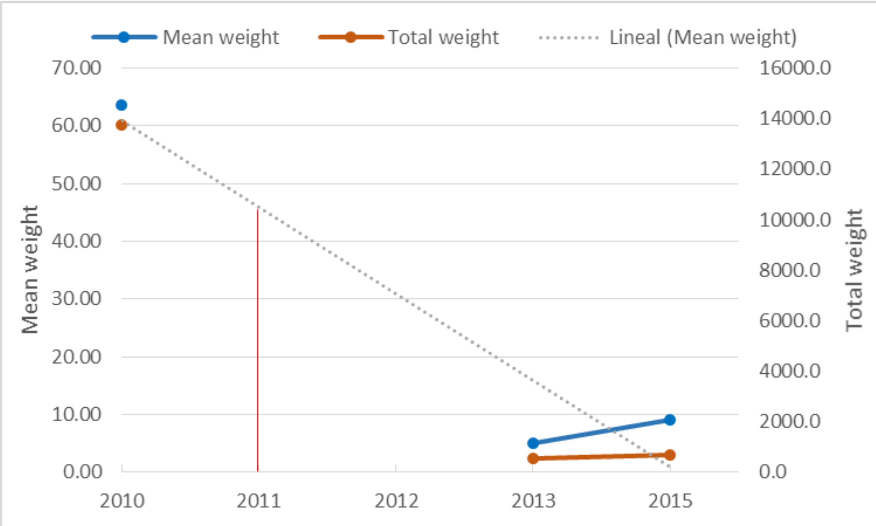
Year	ER	Ds	Mean weight	Total weight	Mean school size	Abundance of animals
2010	26.3	29.4	12.7	32.1		
2011						
2012						
2013	38.7	40.7	40.2	57.2	36.7	54.8
2015	67.5	69.0	66.7	95.8	66.7	95.9

Attempting to extrapolate an estimate for sub-area G in 2011 when no survey was carried out, and only three other years with information available, two years with full information and one year with partial information (no mean school size not abundance of animals) is a very risky and a non-reliable exercise.

However, if the attempt needs to be made, it would be useful to look at Figure 25. In the upper graph, the trend for encounter rates and density of groups is the same, as expected. Two possible trend lines have been plotted (always taking into account that they are built based on only 3 points), a lineal trend (blue dotted line) and a polynomial, curved, trend line (orange dotted line). The red vertical line shows where the hypothetical points for 2011 would be in both cases. When considering a lineal trend, density of schools could be in the order or 3.250 schools per km^2 and when considering a curved trend, density of schools could be in the order of almost 4.00 schools per km^2. If we look at weight (graph in the centre), being 2013 and 2015 very similar and both much smaller than 2010, the only logical option, without any more information, would be a lineal trend line, in which case a cut in 2011 would yield a potential total weight of between 10,000 and 11,000 tons. Not even a guess can be made in terms of mean school size and abundance of animals (see lower graph).

Another way of looking at it is through the estimates of abundance of animals for each sub-area, each year. Figure 26 shows the point estimates for all sub-areas (above) and with a zoom to A, C and G (removing the noise created by the extreme estimated for E) in the plot below. Sub-area E has very extreme interannual variability, making it useless for comparison or attempt of extrapolation for another sub-area. In the second plot, the behaviour of each area is completely different in terms of interannual variability. There is no clear trend upwards or downwards by all sub-areas. This fact precludes any attempt to extrapolate or even guess at what level the abundance of animals could be in sub-area G in 2011 (especially since that information is not available for 2010 either).





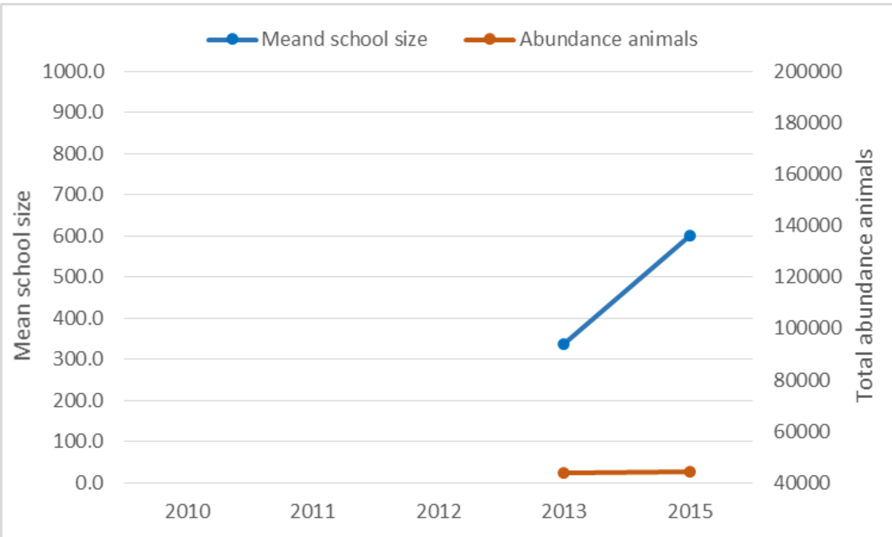
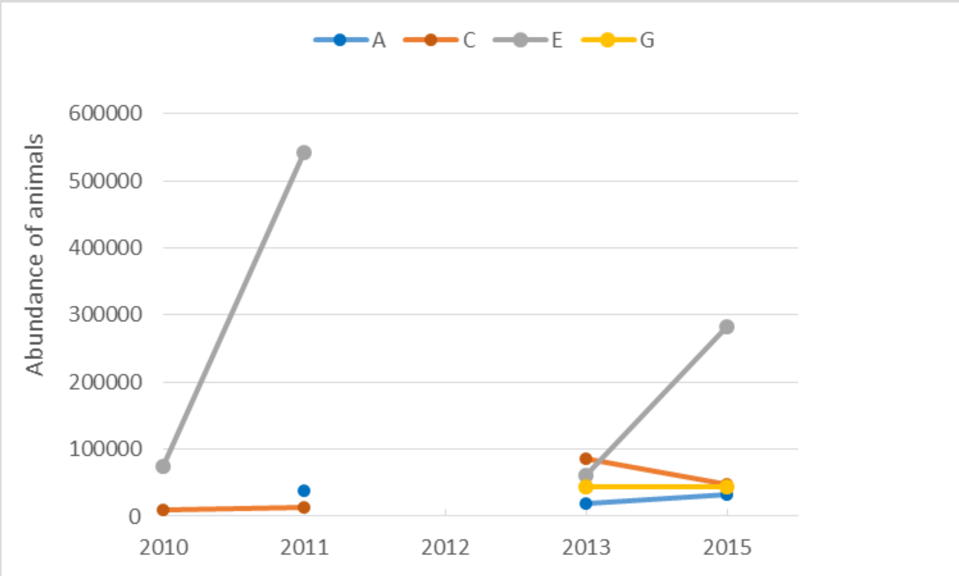


Figure 25. Plots of point estimates of encounter rate and density of schools (above), mean and total weight (centre) and mean school size and abundance of animals (below). Dotted lines are trend lines and red vertical lines hypothetical cut of the trend lines in 2011.



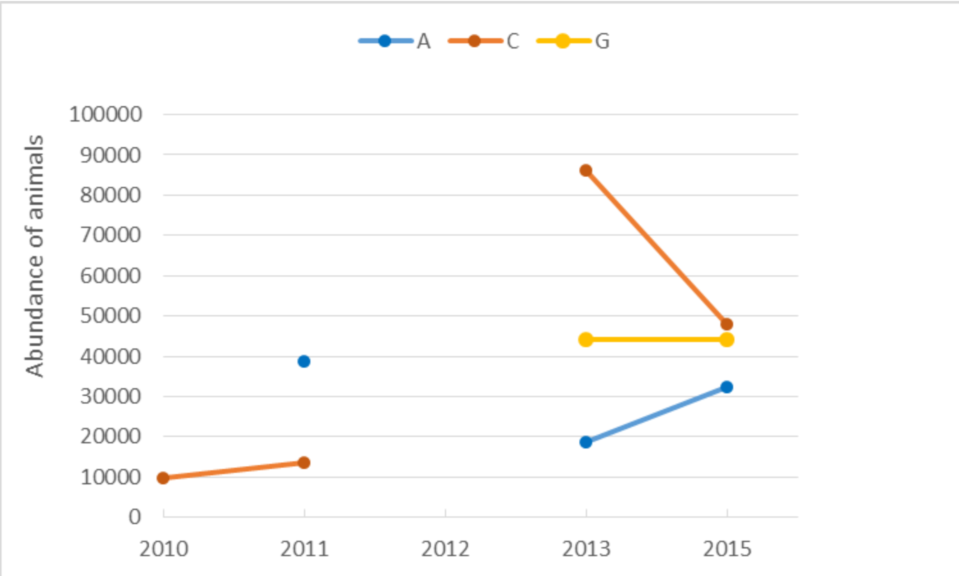


Figure 26. Plots of point estimates of abundance of animals for each sub-area in each year.

III.5 Re-assess all estimates according to these correction factors.

The only correction factor to be applied with some reliability is the g(0) (the availability bias component). No reliable correction factors can be applied for the lack of bubble windows, for missing surveys in areas/years or for environmental factors.

In the case of lack of bubble windows, left truncation, as already applied, is probably the best option to account for the missing sector. The analysis of missing surveys in areas/years has shown extreme variability which precludes any attempt of extrapolation. The environmental factors can be accounted for in the detection functions when they are significant and improve the fit. However, the covariates selected in the final models for each year (see report of September 2015) were effort-related variables (such as team, airplane or sub-area) and no environmental factors were selected, as they influenced the probability of detection to a lesser extent.

The standard approach in this case of instantaneous survey to obtain corrected estimates by the availability bias is to divide conventional line transect estimates by the proportion of time animals at zero perpendicular distance are available for detection. This is, the complement to the time spent between 0 and 10m depth (i.e. the time spent below 10m depth), which would be the time they are generally not available to the spotters to be detected. Therefore, to correct for the availability bias, abundance estimates, both for animals and for weight, are divided by the corresponding g(0) according to year and sub-area (from Table 23). Table 29 shows the pre-corrected and corrected estimates. It is important to highlight that the correction factor is only available in a direct way (from information on the diving patterns in the sub-area and year of the survey) for A, C and E in 2013 and 2015, and for G (with the caveat of being only from night data) in 2015. Therefore, the corrections in Table X for these sub-areas/years are direct. However, an assumption needs to be made if correction has to be applied also for the other years shaded in grey in the table): for A and C 2010 and 2011, the same correction was applied as for 2013-2015; for E 2010 and 2011, the correction of E 2015 was applied given that this correction factor is more similar to those from the other areas, and has also a smaller CV, while 2013 has a very large CV; and for G 2010, 2011 and 2013 the correction for G2015 was applied. All these extrapolations should be taken with caution.

The delta method (Seber 1982) was used to combine the CV from the estimates with the CV from the correction factor, giving a final CV. This final CV is, obviously, larger than the previous one with no correction factor, as the variability inherent to the correction factor is now incorporated and added to the variability in the abundance estimates.

Table 29. Abundance of animals and weight for overlapped sub-areas inside for each year, with and without correction for availability bias. In light grey those corrections extrapolated from other years when information on availability bias exists.

Year	Sub-area	Without correction				Corrected for availability bias			
		A inside	C inside	E inside	G inside	A inside	C inside	E inside	G inside
2010	**Total weight (t)**	**3,496**	**1,658**	**7,995**	**13,733**	**9,200**	**4,363**	**17,418**	**45,474**
	%CV total weight	56.6	46.9	47.1	32.1	69.3	61.6	58.6	39.9
	Total abundance (animals)		**9,797**	**73,676**			**25,782**	**160,514**	
	%CV total abundance		59.2	37.8			71.4	51.4	
2011	**Total weight (t)**	**4,296**	**1,999**	**39,344**	**4,296**	**11,305**	**5,261**	**85,717**	**14,225**
	%CV total weight	46.2	54.9	32.2	46.2	61.1	67.9	47.4	51.9
	Total abundance (animals)	**38,720**	**13,614**	**541,634**	**38,720**	**101,895**	**35,826**	**1,180,031**	**128,212**
	%CV total abundance	45.5	45.3	32.3	45.5	60.6	60.4	47.5	51.3
2013	**Total weight (t)**	**3,572**	**11,830**	**1,882**	**534**	**9,400**	**31,132**	**2,662**	**1,768**
	%CV total weight	40.6	40.9	74.3	57.2	57.0	57.2	114.3	61.9
	Total abundance (animals)	**18,717**	**86,114**	**60,614**	**44,041**	**49,255**	**226,616**	**85,734**	**145,831**
	%CV total abundance	44.5	38.3	75.4	54.8	59.8	55.4	115.1	59.7
2015	**Total weight (t)**	**8,736**	**5,965**	**54,889**	**666**	**22,989**	**15,697**	**119,584**	**2,205**
	%CV total weight	41.9	65.8	62.2	95.8	57.9	77.0	71.3	98.7
	Total abundance (animals)	**38,248**	**47,900**	**283,100**	**44,162**	**100,653**	**126,053**	**616,776**	**146,232**
	%CV total abundance	44.7	65.5	64.1	95.9	60.0	76.7	72.9	98.8

As mentioned before, it is important to highlight that this correction is only based on the availability bias. Perception bias is still unknown and therefore there is still a negative bias unaccounted for and the estimates could be a bit larger to an unknown extent. However, it is important to highlight also the large uncertainty of these corrected estimates (large CVs).

One issue to be mentioned here also, is that it is not possible to estimate a particular g(0) to each airplane or team, as the source of the correction factor comes from an external source (mini-PATs) (also known as additional variance factor) and not from experiments within each airplane or team.

III.6 Possibly introduce the individual spotter correction factor for further re-assessing the estimates by area.

It is not possible to assess any individual spotter correction factor.

IV. Evaluation of the required areas to be surveyed and survey effort required to provide a useful index of relative abundance for stock assessment purposes taking into account the additional variance. Evaluation of the likely achievable actual CV across years for surveys with different spatial coverage and overall survey effort.

Data organization

For this exercise we worked with estimates including school size (number of animals), as all the parameters and the CVs are very similar to the ones including weight, so duplication of the calculations which would provide only similar results was avoided.

As has been showed in previous reports and in this one, the variability of estimates and CVs among years and areas is very large. Principle parameters that affect abundance estimates are year, area, amount of effort, density of animals, variability in school sizes and searching patterns and efficiency of the observers (and therefore the detection function and its derived effective strip width).

The large variability in density, and the difference in surface area, among the 4 subareas suggests that it is best to do the power analysis separately for each area. Nevertheless, the power analysis for the 4 areas pooled together is also provided, to have a better overview of the total effort for each scenario.

Trying to build scenarios with all the possible combinations of those variables would yield an unmanageable amount of potential scenarios. Therefore, to simplify this exercise, four types of scenarios have been built for each area (Table 30). For each of them, 6 potential scenarios of percentage coverage of the area were built (from 10% to 60%, considering that the coverage of the areas for each year ranges from 3% to 40%, see Tables 31 and 32):

- Average: the average density of schools, the average CV of the detection functions and the average CV of the school sizes observed over the years in each area, were used as basis for the scenarios
- Average-H: the highest density of schools, the average CV of the detection functions and the average CV of the school sizes observed over the years in each area, were used as basis for the scenarios
- Best: the highest density of schools, the smallest CV of the detection functions and the smallest CV of the school sizes observed over the years in each area, were used as basis for the scenarios
- Worse: the lowest density of schools, the highest CV of the detection functions and the highest CV of the school sizes observed over the years in each area, were used as basis for the scenarios

All potential variations of these parameters can be built if necessary.

Table 30. Variables included in the Scenarios: density of schools in each area, CV of de detection function (CV df) and CV of school sizes.

	Density schools (1000 km^2)						CV school sizes				
Scenario type	**A**	**C**	**E**	**G**	**Total**	**CV df**	**A**	**C**	**E**	**G**	**Total**
Average	0.65	0.74	1.86	2.50	1.46	0.12	0.22	0.26	0.41	0.52	0.34
Average-H	1.00	1.25	3.35	3.85	2.00	0.12	0.22	0.26	0.41	0.52	0.34
Best	1.00	1.25	3.35	3.85	2.00	0.09	0.11	0.17	0.19	0.37	0.21
Worse	0.40	0.25	0.75	1.30	1.09	0.14	0.28	0.37	0.67	0.67	0.50

Data analysis

Calculation of coefficients of variation

To be able to calculate coefficients of variation, a hypothetical density had to be applied to each block. These densities are those shown in Table 30 for the four types of scenarios. For each type of scenario, six levels of coverage are shown, from 10% to 60% (Tables 33 to 36). The amount of effort (transect length) needed to reach such coverage in each area was calculated as:

$$L = \frac{A \times C}{esw \times 2}$$

Where *L* is the transect length needed on effort, *A* is the surface area of the block, *C* is the coverage, and *esw* is the effective strip width. An *esw* of 2 was chosen as an intermediate value with respect of those obtained in the 4 years of survey (Table 31).

The derived expected number of schools per block/scenario are also shown in Tables 33 to 36. These are calculated as:

$$Exp(n) = A \times \frac{D}{1000} \times C$$

Where *A* is the surface area of the block, *C* is the coverage, and *D* is the density (divided by 1000 as the density is expresses as number of animals per 1000 km^2).

The coefficients of variation for the density of schools, for the detection functions and for the school size are shown in Tables 33 to 36. As mentioned above, the CVs for the detection function and for the school size are estimates based on the values obtained in the four years of survey in each area (Table 30). For each scenario, the expected coefficient of variation of $E(n_{ij})$ was estimated for each block and scenario as follows:

$$CV_{E(n_{ij})} = 100 * \frac{\sqrt{E(n_{ij}) * 2}}{E(N_{ij})}$$

Where $CV_{E(n_{ij})}$ is the coefficient of variation of $E(n_{ij})$; and $E(n_{ij})$ is the expected number of schools in block *i* in scenario *j* given the coverage and density specified for each block and scenario. Some overdispersion is assumed (variance of E[n] is twice E[n]). The total CV for each block in each scenario was estimated using the Delta method (Seber, 1982) combining the CVs of $E(n_{ij})$, detection function ($CV_{df_{ij}}$) and cluster size ($CV_{s_{ij}}$):

$$CV_{T_{ij}} = \sqrt{{CV_{df_{ij}}}^2 + {CV_{s_{ij}}}^2 + {CV_{E(n_{ij})}}^2}$$

Additional variance

Two sets of additional variance have been used, according to the work done in Cañadas and Ben Mhamed (2016). One comes from evaluating spatial and vertical differences between spawning seasons using electronic tagging data (additional variance 1 = 28.2%), and the other one from the results of Distance using a joint model between the density and the school size (additional variance 2 = 80%). The two different additional variances have been incorporated to obtain a final CV through the delta method above.

Power analysis

Power analyses can be undertaken to evaluate the survey CV and frequency required to be reasonably certain to detect a given change in population size (Gerrodette 1987). Here we chose to examine a significance level of 5% and a detection power of 60% based on conventional practice (Fortuna *et al*. 2014). Survey CVs considered were those obtained for each scenario in Tables 33 to 36.

Table 31. Observed data per year. Obs n = number of schools detected; CVn = CV of density of groups; CVdf = CV of the detection function; CVs = CV of school size; CVabun = CV of abundance of animals

	Area (km²)	Density schools (1000 km²)	Transect length	*esw* x2	% coverage	Obs n	CV n	CV df	CV s	CV abun	Additional variance 1	Final CV 1	Additional variance 2	Final CV 2
2010	265,627	1.236	29,967	2.96	0.33	76	0.23	0.13	0.19	0.30	0.28	0.41	0.80	0.85
2011	209,416	2.004	26,247	1.36	0.17	65	0.21	0.12	0.37	0.30	0.28	0.41	0.80	0.85
2013	265,627	1.494	14,862	2.95	0.17	52	0.22	0.09	0.19	0.30	0.28	0.41	0.80	0.86
2015	265,627	1.094	12,046	1.94	0.09	24	0.27	0.14	0.41	0.47	0.28	0.55	0.80	0.93

Table 32. Observed data per year and block. Obs n = number of schools detected; CVn = CV of density of groups; CVdf = CV of the detection function; CVs = CV of school size; CVabun = CV of abundance of animals

Block	Year	Area (km²)	Density schools (1000 km²)	Transect length	*esw* x2	% coverage	Obs n	CV n	CV df	CV s	CV abun	Additional variance 1	Final CV 1	Additional variance 2	Final CV 2
A	**2010**	61,933	0.43	6,277	2.96	0.30	8	0.56	0.13						
	2011	61,933	0.922	7,975	1.36	0.18	10	0.36	0.12	0.28	0.46	0.28	0.54	0.80	0.92
	2013	61,933	0.495	6,743	2.95	0.32	10	0.36	0.09	0.26	0.45	0.28	0.53	0.80	0.92
	2015	61,933	0.749	4,119	1.94	0.13	6	0.43	0.14	0.11	0.45	0.28	0.53	0.80	0.92
C	**2010**	53,868	0.248	8,168	2.96	0.45	6	0.47	0.13	0.37	0.59	0.28	0.66	0.80	1.0000
	2011	53,868	0.868	8,466	1.36	0.21	10	0.33	0.12	0.31	0.45	0.28	0.53	0.80	0.92
	2013	53,868	1.244	2,682	2.95	0.15	10	0.34	0.09	0.17	0.38	0.28	0.48	0.80	0.89
	2015	53,868	0.58	2,658	1.94	0.10	3	0.63	0.14	0.19	0.66	0.28	0.71	0.80	11.03
E	**2010**	93,614	0.775	12,621	2.96	0.40	29	0.33	0.13	0.19	0.38	0.28	0.47	0.80	0.88
	2011	93,614	3.374	9,806	1.36	0.14	45	0.24	0.12	0.22	0.32	0.28	0.43	0.80	0.86
	2013	93,614	1.794	3,720	2.95	0.12	20	0.34	0.09	0.67	0.75	0.28	0.81	0.80	1.1010
	2015	93,614	1.49	4,484	1.94	0.09	13	0.30	0.14	0.57	0.64	0.28	0.70	0.80	11.03
G	**2010**	56,211	3.84	2,900	2.96	0.15	33	0.29	0.13						
	2011														
	2013	56,211	2.333	1,716	2.95	0.09	12	0.41	0.09	0.37	0.55	0.28	0.62	0.80	0.97
	2015	56,211	1.309	785	1.94	0.03	2	0.69	0.14	0.67	0.96	0.28	1.00	0.80	1.25

Table 33. Scenarios Average 1 to Average 6. Exp n = expected number of schools for the given coverage; CVn = CV of density of groups; CVdf = CV of the detection function; CVs = CV of school size; CVabun = CV of abundance of animals

Block	Area (km^2)	Density schools (1000 km^2)	Transect length	% coverage	Exp. n	CV n	CV df	CV s	CV abun	Add. Var. 1	Final CV 1	Add. Var. 2	Final CV 2
	61,933	0.65	3,097	0.10	4	0.70	0.12	0.22	0.75	0.28	0.80	0.80	1.10
	61,933	0.65	6,193	0.20	8	0.50	0.12	0.22	0.56	0.28	0.63	0.80	0.98
A	61,933	0.65	9,290	0.30	12	0.41	0.12	0.22	0.48	0.28	0.55	0.80	0.93
	61,933	0.65	12,387	0.40	16	0.35	0.12	0.22	0.43	0.28	0.52	0.80	0.91
	61,933	0.65	15,483	0.50	20	0.32	0.12	0.22	0.40	0.28	0.49	0.80	0.90
	61,933	0.65	18,580	0.60	24	0.29	0.12	0.22	0.38	0.28	0.47	0.80	0.89
	53,868	0.74	2,693	0.10	4	0.71	0.12	0.26	0.76	0.28	0.81	0.80	1.11
	53,868	0.74	5,387	0.20	8	0.50	0.12	0.26	0.58	0.28	0.64	0.80	0.99
C	53,868	0.74	8,080	0.30	12	0.41	0.12	0.26	0.50	0.28	0.57	0.80	0.94
	53,868	0.74	10,774	0.40	16	0.35	0.12	0.26	0.46	0.28	0.54	0.80	0.92
	53,868	0.74	13,467	0.50	20	0.32	0.12	0.26	0.43	0.28	0.51	0.80	0.91
	53,868	0.74	16,160	0.60	24	0.29	0.12	0.26	0.41	0.28	0.50	0.80	0.90
	93,614	1.86	4,681	0.10	17	0.34	0.12	0.41	0.55	0.28	0.61	0.80	0.97
	93,614	1.86	9,361	0.20	35	0.24	0.12	0.41	0.49	0.28	0.57	0.80	0.94
E	93,614	1.86	14,042	0.30	52	0.20	0.12	0.41	0.47	0.28	0.55	0.80	0.93
	93,614	1.86	18,723	0.40	70	0.17	0.12	0.41	0.46	0.28	0.54	0.80	0.92
	93,614	1.86	23,404	0.50	87	0.15	0.12	0.41	0.45	0.28	0.53	0.80	0.92
	93,614	1.86	28,084	0.60	104	0.14	0.12	0.41	0.45	0.28	0.53	0.80	0.92
	56,211	2.5	2,811	0.10	14	0.38	0.12	0.52	0.65	0.28	0.71	0.80	1.03
	56,211	2.5	5,621	0.20	28	0.27	0.12	0.52	0.60	0.28	0.66	0.80	1.00
G	56,211	2.5	8,432	0.30	42	0.22	0.12	0.52	0.58	0.28	0.64	0.80	0.99
	56,211	2.5	11,242	0.40	56	0.19	0.12	0.52	0.57	0.28	0.63	0.80	0.98
	56,211	2.5	14,053	0.50	70	0.17	0.12	0.52	0.56	0.28	0.63	0.80	0.98
	56,211	2.5	16,863	0.60	84	0.15	0.12	0.52	0.56	0.28	0.62	0.80	0.97
	265,626	1.46	13,281	0.10	39	0.23	0.12	0.34	0.43	0.28	0.51	0.80	0.91
	265,626	1.46	26,563	0.20	78	0.16	0.12	0.34	0.39	0.28	0.49	0.80	0.89
Total	265,626	1.46	39,844	0.30	116	0.13	0.12	0.34	0.38	0.28	0.48	0.80	0.89
	265,626	1.46	53,125	0.40	155	0.11	0.12	0.34	0.38	0.28	0.47	0.80	0.88
	265,626	1.46	66,407	0.50	194	0.10	0.12	0.34	0.37	0.28	0.47	0.80	0.88
	265,626	1.46	79,688	0.60	233	0.09	0.12	0.34	0.37	0.28	0.47	0.80	0.88

Table 34. Scenarios Average 1-H to Average 6-H with highest observed densities of school per area. Exp n = expected number of schools for the given coverage; CVn = CV of density of groups; CVdf = CV of the detection function; CVs = CV of school size; CVabun = CV of abundance of animals

Block	Area (km^2)	Density schools (1000 km^2)	Transect length	% coverage	Exp. n	CV n	CV df	CV s	CV abun	Add. Var. 1	Final CV 1	Add. Var. 2	Final CV 2
A	61,933	1	3,097	0.10	6	0.57	0.12	0.22	0.62	0.28	0.68	0.80	1.01
	61,933	1	6,193	0.20	12	0.40	0.12	0.22	0.47	0.28	0.55	0.80	0.93
	61,933	1	9,290	0.30	19	0.33	0.12	0.22	0.41	0.28	0.50	0.80	0.90
	61,933	1	12,387	0.40	25	0.28	0.12	0.22	0.38	0.28	0.47	0.80	0.89
	61,933	1	15,483	0.50	31	0.25	0.12	0.22	0.36	0.28	0.45	0.80	0.88
	61,933	1	18,580	0.60	37	0.23	0.12	0.22	0.34	0.28	0.44	0.80	0.87
C	53,868	1.25	2,693	0.10	7	0.54	0.12	0.26	0.62	0.28	0.68	0.80	1.01
	53,868	1.25	5,387	0.20	13	0.39	0.12	0.26	0.48	0.28	0.56	0.80	0.93
	53,868	1.25	8,080	0.30	20	0.31	0.12	0.26	0.43	0.28	0.51	0.80	0.91
	53,868	1.25	10,774	0.40	27	0.27	0.12	0.26	0.40	0.28	0.49	0.80	0.89
	53,868	1.25	13,467	0.50	34	0.24	0.12	0.26	0.38	0.28	0.47	0.80	0.88
	53,868	1.25	16,160	0.60	40	0.22	0.12	0.26	0.36	0.28	0.46	0.80	0.88
E	93,614	3.35	4,681	0.10	31	0.25	0.12	0.41	0.50	0.28	0.57	0.80	0.94
	93,614	3.35	9,361	0.20	63	0.18	0.12	0.41	0.46	0.28	0.54	0.80	0.92
	93,614	3.35	14,042	0.30	94	0.15	0.12	0.41	0.45	0.28	0.53	0.80	0.92
	93,614	3.35	18,723	0.40	125	0.13	0.12	0.41	0.45	0.28	0.53	0.80	0.92
	93,614	3.35	23,404	0.50	157	0.11	0.12	0.41	0.44	0.28	0.52	0.80	0.91
	93,614	3.35	28,084	0.60	188	0.10	0.12	0.41	0.44	0.28	0.52	0.80	0.91
G	56,211	3.85	2,811	0.10	22	0.30	0.12	0.52	0.61	0.28	0.68	0.80	1.01
	56,211	3.85	5,621	0.20	43	0.21	0.12	0.52	0.58	0.28	0.64	0.80	0.99
	56,211	3.85	8,432	0.30	65	0.18	0.12	0.52	0.56	0.28	0.63	0.80	0.98
	56,211	3.85	11,242	0.40	87	0.15	0.12	0.52	0.55	0.28	0.62	0.80	0.97
	56,211	3.85	14,053	0.50	108	0.14	0.12	0.52	0.55	0.28	0.62	0.80	0.97
	56,211	3.85	16,863	0.60	130	0.12	0.12	0.52	0.55	0.28	0.62	0.80	0.97
Total	265,626	2	13,281	0.10	53	0.19	0.12	0.34	0.39	0.28	0.48	0.80	0.89
	265,626	2	26,563	0.20	106	0.14	0.12	0.34	0.38	0.28	0.47	0.80	0.88
	265,626	2	39,844	0.30	159	0.11	0.12	0.34	0.37	0.28	0.47	0.80	0.88
	265,626	2	53,125	0.40	213	0.10	0.12	0.34	0.37	0.28	0.46	0.80	0.88
	265,626	2	66,407	0.50	266	0.09	0.12	0.34	0.37	0.28	0.46	0.80	0.88
	265,626	2	79,688	0.60	319	0.08	0.12	0.34	0.37	0.28	0.46	0.80	0.88

Table 35. Scenarios Best 1 to Best 6. Exp n = expected number of schools for the given coverage; CVn = CV of density of groups; CVdf = CV of the detection function; CVs = CV of school size; CVabun = CV of abundance of animals

Block	Area (km^2)	Density schools (1000 km^2)	Transect length	% coverage	Exp. n	CV n	CV df	CV s	CV abun	Add. Var. 1	Final CV 1	Add. Var. 2	Final CV 2
	61,933	1	3,097	0.10	6	0.57	0.09	0.11	0.59	0.28	0.65	0.80	0.99
	61,933	1	6,193	0.20	12	0.40	0.09	0.11	0.43	0.28	0.51	0.80	0.91
A	61,933	1	9,290	0.30	19	0.33	0.09	0.11	0.36	0.28	0.46	0.80	0.88
	61,933	1	12,387	0.40	25	0.28	0.09	0.11	0.32	0.28	0.42	0.80	0.86
	61,933	1	15,483	0.50	31	0.25	0.09	0.11	0.29	0.28	0.41	0.80	0.85
	61,933	1	18,580	0.60	37	0.23	0.09	0.11	0.27	0.28	0.39	0.80	0.84
	53,868	1.25	2,693	0.10	7	0.54	0.09	0.17	0.58	0.28	0.64	0.80	0.99
	53,868	1.25	5,387	0.20	13	0.39	0.09	0.17	0.43	0.28	0.51	0.80	0.91
C	53,868	1.25	8,080	0.30	20	0.31	0.09	0.17	0.37	0.28	0.46	0.80	0.88
	53,868	1.25	10,774	0.40	27	0.27	0.09	0.17	0.33	0.28	0.44	0.80	0.87
	53,868	1.25	13,467	0.50	34	0.24	0.09	0.17	0.31	0.28	0.42	0.80	0.86
	53,868	1.25	16,160	0.60	40	0.22	0.09	0.17	0.29	0.28	0.41	0.80	0.85
	93,614	3.35	4,681	0.10	31	0.25	0.09	0.19	0.33	0.28	0.43	0.80	0.86
	93,614	3.35	9,361	0.20	63	0.18	0.09	0.19	0.28	0.28	0.39	0.80	0.85
E	93,614	3.35	14,042	0.30	94	0.15	0.09	0.19	0.26	0.28	0.38	0.80	0.84
	93,614	3.35	18,723	0.40	125	0.13	0.09	0.19	0.25	0.28	0.37	0.80	0.84
	93,614	3.35	23,404	0.50	157	0.11	0.09	0.19	0.24	0.28	0.37	0.80	0.83
	93,614	3.35	28,084	0.60	188	0.10	0.09	0.19	0.23	0.28	0.37	0.80	0.83
	56,211	3.85	2,811	0.10	22	0.30	0.09	0.37	0.49	0.28	0.56	0.80	0.94
	56,211	3.85	5,621	0.20	43	0.21	0.09	0.37	0.44	0.28	0.52	0.80	0.91
G	56,211	3.85	8,432	0.30	65	0.18	0.09	0.37	0.42	0.28	0.51	0.80	0.90
	56,211	3.85	11,242	0.40	87	0.15	0.09	0.37	0.41	0.28	0.50	0.80	0.90
	56,211	3.85	14,053	0.50	108	0.14	0.09	0.37	0.40	0.28	0.49	0.80	0.90
	56,211	3.85	16,863	0.60	130	0.12	0.09	0.37	0.40	0.28	0.49	0.80	0.89
	265,626	2	13,281	0.10	53	0.19	0.09	0.21	0.30	0.28	0.41	0.80	0.85
	265,626	2	26,563	0.20	106	0.14	0.09	0.21	0.27	0.28	0.39	0.80	0.84
Total	265,626	2	39,844	0.30	159	0.11	0.09	0.21	0.25	0.28	0.38	0.80	0.84
	265,626	2	53,125	0.40	213	0.10	0.09	0.21	0.25	0.28	0.38	0.80	0.84
	265,626	2	66,407	0.50	266	0.09	0.09	0.21	0.24	0.28	0.37	0.80	0.84
	265,626	2	79,688	0.60	319	0.08	0.09	0.21	0.24	0.28	0.37	0.80	0.84

Table 36. Scenarios Worse 1 to Worse 6. Exp n = expected number of schools for the given coverage; CVn = CV of density of groups; CVdf = CV of the detection function; CVs = CV of school size; CVabun = CV of abundance of animals

Block	Area (km^2)	Density schools (1000 km^2)	Transect length	% coverage	Exp. n	CV n	CV df	CV s	CV abun	Add. Var. 1	Final CV 1	Add. Var. 2	Final CV 2
A	61,933	0.4	3,097	0.10	2	0.90	0.14	0.28	0.95	0.28	0.99	0.80	1.24
	61,933	0.4	6,193	0.20	5	0.64	0.14	0.28	0.71	0.28	0.76	0.80	1.07
	61,933	0.4	9,290	0.30	7	0.52	0.14	0.28	0.61	0.28	0.67	0.80	1.00
	61,933	0.4	12,387	0.40	10	0.45	0.14	0.28	0.55	0.28	0.62	0.80	0.97
	61,933	0.4	15,483	0.50	12	0.40	0.14	0.28	0.51	0.28	0.58	0.80	0.95
	61,933	0.4	18,580	0.60	15	0.37	0.14	0.28	0.48	0.28	0.56	0.80	0.93
C	53,868	0.25	2,693	0.10	1	1.22	0.14	0.37	1.28	0.28	1.31	0.80	1.51
	53,868	0.25	5,387	0.20	3	0.86	0.14	0.37	0.95	0.28	0.99	0.80	1.24
	53,868	0.25	8,080	0.30	4	0.70	0.14	0.37	0.81	0.28	0.86	0.80	1.14
	53,868	0.25	10,774	0.40	5	0.61	0.14	0.37	0.73	0.28	0.78	0.80	1.08
	53,868	0.25	13,467	0.50	7	0.54	0.14	0.37	0.67	0.28	0.73	0.80	1.05
	53,868	0.25	16,160	0.60	8	0.50	0.14	0.37	0.64	0.28	0.70	0.80	1.02
E	93,614	0.75	4,681	0.10	7	0.53	0.14	0.67	0.87	0.28	0.91	0.80	1.18
	93,614	0.75	9,361	0.20	14	0.38	0.14	0.67	0.78	0.28	0.83	0.80	1.12
	93,614	0.75	14,042	0.30	21	0.31	0.14	0.67	0.75	0.28	0.80	0.80	1.10
	93,614	0.75	18,723	0.40	28	0.27	0.14	0.67	0.73	0.28	0.79	0.80	1.09
	93,614	0.75	23,404	0.50	35	0.24	0.14	0.67	0.72	0.28	0.78	0.80	1.08
	93,614	0.75	28,084	0.60	42	0.22	0.14	0.67	0.72	0.28	0.77	0.80	1.08
G	56,211	1.3	2,811	0.10	7	0.52	0.14	0.67	0.86	0.28	0.91	0.80	1.18
	56,211	1.3	5,621	0.20	15	0.37	0.14	0.67	0.78	0.28	0.83	0.80	1.12
	56,211	1.3	8,432	0.30	22	0.30	0.14	0.67	0.75	0.28	0.80	0.80	1.10
	56,211	1.3	11,242	0.40	29	0.26	0.14	0.67	0.73	0.28	0.79	0.80	1.08
	56,211	1.3	14,053	0.50	37	0.23	0.14	0.67	0.72	0.28	0.78	0.80	1.08
	56,211	1.3	16,863	0.60	44	0.21	0.14	0.67	0.72	0.28	0.77	0.80	1.07
Total	265,626	1.09	13,281	0.10	29	0.26	0.14	0.50	0.58	0.28	0.65	0.80	0.99
	265,626	1.09	26,563	0.20	58	0.19	0.14	0.50	0.55	0.28	0.62	0.80	0.97
	265,626	1.09	39,844	0.30	87	0.15	0.14	0.50	0.54	0.28	0.61	0.80	0.97
	265,626	1.09	53,125	0.40	116	0.13	0.14	0.50	0.54	0.28	0.61	0.80	0.96
	265,626	1.09	66,407	0.50	145	0.12	0.14	0.50	0.53	0.28	0.60	0.80	0.96
	265,626	1.09	79,688	0.60	174	0.11	0.14	0.50	0.53	0.28	0.60	0.80	0.96

Results

Coefficients of variation

Tables 33 to 36 show the expected CV of abundance of animals for each scenario/block. As expected, CVs decrease with increased coverage and with increased density of schools. Two sets of Final CVs are provided, Final CV 1 incorporating the additional variance from evaluating spatial and vertical differences between spawning seasons using electronic tagging data; and Final CV 2 incorporating the additional variance from the results of Distance using a joint model between the density and the school size. The second one is much smaller than the first one, allowing for a more useful power analysis.

Power analysis

Figure 27 shows the results of the power analysis for six levels of survey CV (20% to 120%) and two survey frequencies (1 and 2 years). CVs of this level are plausible if uncorrected estimates are considered as relative abundance estimates, and if it can be assumed that availability and perception bias remains constant over the survey periods investigated. The curved lines represent a power of 0.6 to detect a trend in the population for each combination of survey CV and frequency. For a given *r* (y axis), the time taken to detect a significant change in the population can be found from the intercept of the corresponding horizontal line with the power curves.

For example, for a CV of 20% and an *r* of 0.1 with surveys every 1 or 2 years it would take 6 and 7 years, respectively, to detect a change in the population. In contrast, detection of a growth rate of –0.1 would take longer, i.e. 7 and 10 years, respectively. With growth rates of between 0.05 and –0.05 it appears difficult to detect any change in the population within a reasonable timeframe.

Looking at the final CVs obtained at the different scenarios (Tables 33 to 36), their range from 37% to 131% in the case of additional variance from tagging data (CV Final 1), with a minimum CV between 37% and 39% depending on the block and under the Best type of scenario. In the case of additional variance calculated from the results from Distance, the final CVs (Final CV 2) range from 83% to 151%, with a minimum CV between 83% and 89% depending on the block and under the Best type of scenario.

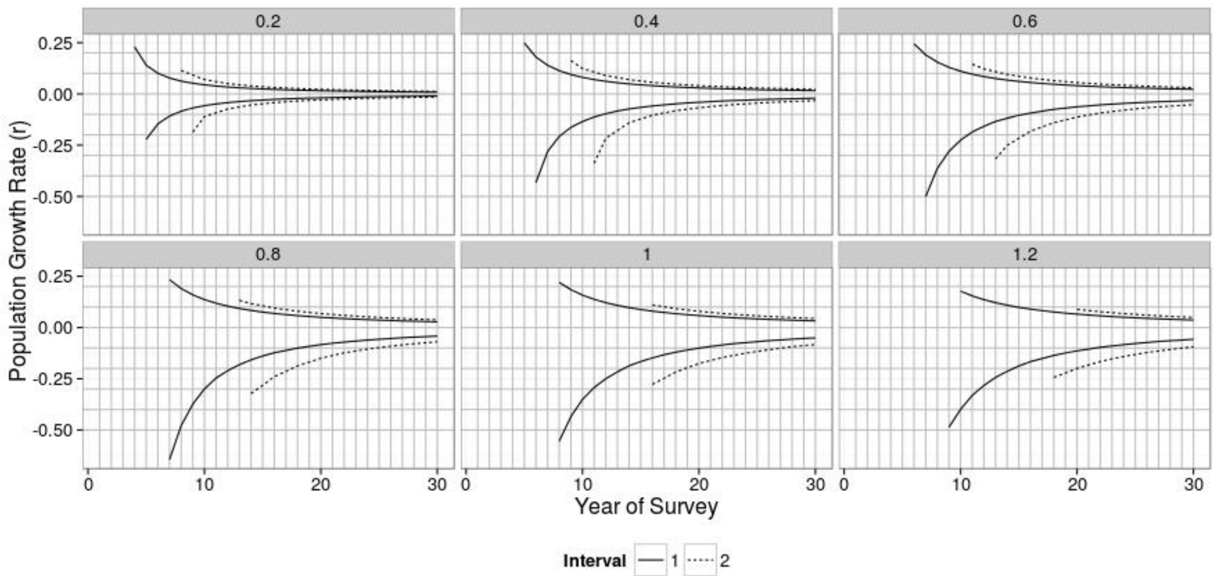


Figure 27. Power analysis: contours correspond to a probability of 0.6 that the null hypothesis (i.e. no change in the population) will be rejected when the null hypothesis is false. Panels correspond to the range of assumed CV of the survey abundance estimate (0.2 to 1.2) and lines to annual and biennial survey cycles. Horizontal lines correspond to a given population growth rate and where this intercepts a power curve the number of years required before a change in the population is detectable can be read off the x axis.

No calculations of the power have been done for the additional variance 2, as it is so high that it would not be useful at all. Therefore the next examples are given for the Additional Variance 1 (from the tagging data).

According to those results:

<u>Medium scenarios</u>

- For a CV of 60% (roundedFinal CV 1 for 30% coverage in Aall blocks and C60% in block G) and an *r* of 0.1 with surveys every 1 or 2 years it would take 11 and 14 years, respectively, to detect a change in the population. In contrast, detection of a growth rate of –0.1 would take longer, i.e. 15 and 22 years, respectively.
- For a CV of 60% (roundedFinal CV 1 for 30% coverage in Aall blocks and C60% in block G) and an *r* of 0.2 with surveys every 1 or 2 years it would take 7 and around 10 years, respectively, to detect a change in the population. In contrast, detection of a growth rate of –0.2 would take longer, i.e. 11 and 16 years, respectively.
- For a CV of 50% (1 for 30% coverage in E and G, and approximaterounded Final CV 1 for 60% coverage of all areasin areas A, C and E) and an *r* of 0.1 with surveys every 1 or 2 years it would take 99 and 12 years, respectively, to detect a change in the population. In contrast, detection of a growth rate of –0.1 would take longer, i.e. 13 and 19 years, respectively. These were estimated as intermediate between a CV of 40% and 60%.% from the panels in Figure 27.
- For a CV of 50% (1 for 30% coverage in E and G, and approximaterounded Final CV 1 for 60% coverage of all areasin areas A, C and E) and an *r* of 0.2 with surveys every 1 or 2 years it would take 6 and around 99 years, respectively, to detect a change in the population. In contrast, detection of a growth rate of –0.2 would take longer, i.e. 99 and 14 years, respectively. These were estimated as intermediate between a CV of 40% and 60%.% from the panels in Figure 27.

If more detailed examples are required, they can be provided.

Discussion

Main sources of uncertainty

The main sources of uncertainty in the abundance estimates of BFT from the aerial surveys are:

- Year (interannual variability)

- Season (month)

- Area (geographical variability)

- Amount of effort (percentage coverage of the sampling area)

- Density of animals (which can vary geographically and interannually)

- Variability in school sizes

- Searching patterns and efficiency of the observers (and therefore the detection function and its derived effective strip width).

- Behaviour of the animals (diving patterns)

Of these, the only "controllable" source of uncertainty is the amount of effort. All the others are extrinsic factors that we cannot control. Even year, month and area are sources of uncertainty in the sense of the variability of the density of the animals, which can vary temporally and spatially without our control. These sources are also potentially related with temporal and spatial variability in the diving behaviour of the animals, and therefore their detectability. The detection function is controllable up to a certain level only, depending on the efficiency and searching pattern of the observers, which should be improved when necessary and kept as constant as possible. However, other variables affect the detection function like the searching conditions (described above in this report), the detectability of the animals (e.g. diving behaviour) and some randomness in the distribution of the observations especially when the sample size is small.

Additionally, there are interrelations or synergy among many of these sources. For example, looking at Tables 33 to 36, in a given type of scenario and area, changing only the % coverage, the Final CV (for example CV 2) can be 30% lower going from minimum 10% coverage to maximum 60% coverage in an area of low density (see Table 33, Block A), but only 10% lower in an area of high density (see Table 33, Block G). If the density is higher in each block (see Table 34), the difference on Final CV between low and high coverage is a bit smaller, only 20% in the lower density area (A) and 7% in the high density area (G). This seems to indicate that a higher coverage has larger effect on areas with lower density than higher density of animals.

It can also be observed in Table 32 that the range of CVs of the detection functions is small, and with small values, ranging from 9% to 14%. Therefore this parameter had relatively small influence on the final CV, and little variability. However, it would be important to keep it small and quite constant by ensuring an improved and effective searching pattern of the observers, and their continuity over time.

The uncertainty in the diving patterns, which is base for the additional variance 1, might be further decreased by increasing the sample size to obtain a more precise diving/surfacing pattern. However, there is potentially a large variability intrinsic to the behaviour of the animals. This affects too to the probability of detecting the animals.

Special case of uncertainty in school sizes

The CVs of the school sizes (and the same happens when looking at estimates of weight), are highly variable (Table 32), ranging from 11% to 67% depending on year and area. This variability up to certain point out of control as it is an extrinsic factor.

There could be two ways of potentially reduce the uncertainty in the estimates of school size. On one hand, if the estimates of group sizes are improved though training (as suggested above in this report) and are provided always by the same observers it would reduce the potential inter-observer variability. On the other hand, an increase in number of observations with increased coverage would theoretically reduce the uncertainty around the school sizes estimates. If a regression line is built in a plot of number of observations against school size CVs, a negative trend is found with decreasing CVs for increasing number of observations. For all areas combined, the slope of the regression line would be around 0.2% decrease in CV for each increment of one observation (Figure 28). If only areas E and G are considered (those with higher variability in school size CVs, the trend is even clearer, with a slope of around 1% decrease in CV for each increment of one observation (Figure 29).

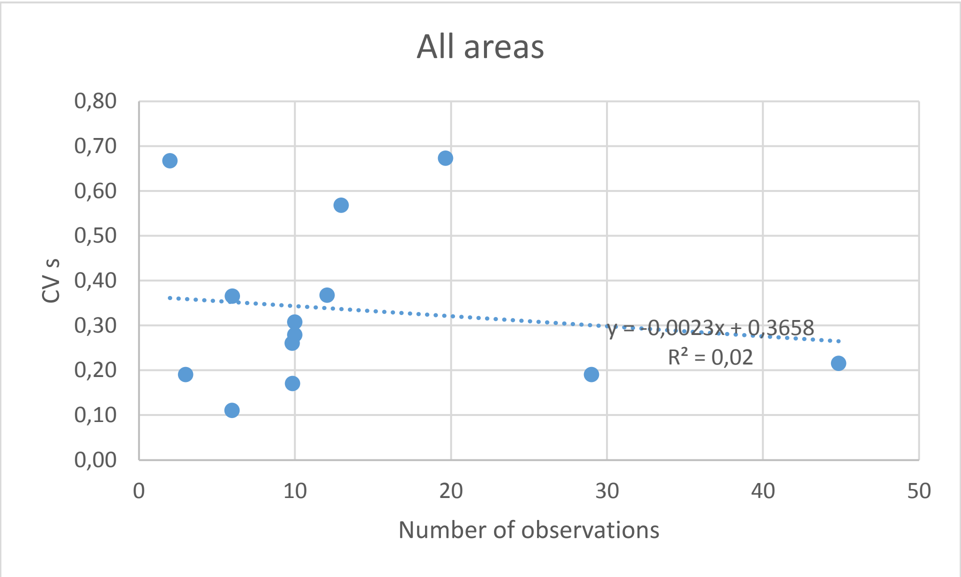


Figure 28. Number of observations of BFT against school size CVs, and trend line, pooling together all areas and years

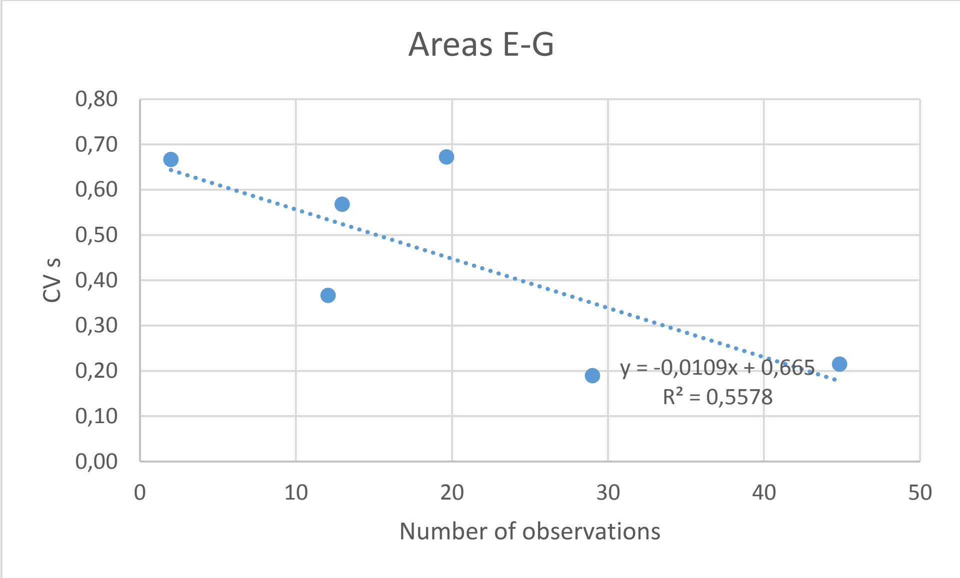


Figure 29. Number of observations of BFT against school size CVs, and trend line, pooling together areas E and G, all years

As example, Table 36 shows the scenarios Average 1 to 6 but including the potential variation of school size CVs with slope 0.2%. Obviously, the decrease would not be lineal and there would be a point of increased number of observations where the CV of school size would not decrease any more or very slightly because the true intrinsic variability of school sizes would still remain no matter how large the sample size is.

The main source of variability in these CVs are precisely the presumably true large variability of school sizes of BFT at sea.

Power analysis

The power analysis shows that the additional variance has a big impact on the final CV. Depending on the method to obtain it, 28% or 80%, so the true additional variance probably lays somewhere in between those values. It certainly seems to have more effect that the CV of the detection function and in most cases than the CV of the density and similar to that of the school sizes.

The additional variance calculated from the spatial (inter-area) and temporal (inter-annual) variability is very high. Ideally, an additional variance area-specific should be estimated, considering only the inter-annual variability and therefore decreasing the final CV considerably. However when this was attempt the models did not converge so not reliable area-specific additional variance could be obtained, probably due to the small number of data points per area (4 years in A, C and E, and 3 years in G). It is possible that after a new year survey these could be estimated.

With the level of final CVs, quite large, due to all the accumulated uncertainties, it would need many years to detect a trend in population growth, usually between a bit more than half and two decades depending on the scenario chosen. The main way to reducing the final CV and therefore increasing the power in detecting trends, which is controllable by us, is to preferably increase or at least maintain the maximum level of coverage per survey as possible over time.

Table 36. Scenarios Average 1 to Average 6. Exp n = expected number of schools for the given coverage; CVn = CV of density of groups; CVdf = CV of the detection function; CVs = CV of school size; CVabun = CV of abundance of animals

Block	Area (km^2)	Density schools (1000 km^2)	Transect length	% coverage	Exp. n	CV n	CV df	CV s	CV abun	Add. Var. 1	Final CV 1	Add. Var. 2	Final CV 2
A	61,933	0.65	3,097	0.10	4	0.70	0.12	0.24	0.75	0.28	0.80	0.80	1.10
	61,933	0.65	6,193	0.20	8	0.50	0.12	0.23	0.56	0.28	0.63	0.80	0.98
	61,933	0.65	9,290	0.30	12	0.41	0.12	0.22	0.48	0.28	0.55	0.80	0.93
	61,933	0.65	12,387	0.40	16	0.35	0.12	0.21	0.43	0.28	0.51	0.80	0.91
	61,933	0.65	15,483	0.50	20	0.32	0.12	0.20	0.39	0.28	0.48	0.80	0.89
	61,933	0.65	18,580	0.60	24	0.29	0.12	0.19	0.37	0.28	0.46	0.80	0.88
C	53,868	0.74	2,693	0.10	4	0.71	0.12	0.28	0.77	0.28	0.82	0.80	1.11
	53,868	0.74	5,387	0.20	8	0.50	0.12	0.27	0.58	0.28	0.65	0.80	0.99
	53,868	0.74	8,080	0.30	12	0.41	0.12	0.26	0.50	0.28	0.57	0.80	0.94
	53,868	0.74	10,774	0.40	16	0.35	0.12	0.25	0.45	0.28	0.53	0.80	0.92
	53,868	0.74	13,467	0.50	20	0.32	0.12	0.24	0.42	0.28	0.50	0.80	0.90
	53,868	0.74	16,160	0.60	24	0.29	0.12	0.23	0.39	0.28	0.48	0.80	0.89
E	93,614	1.86	4,681	0.10	17	0.34	0.12	0.49	0.61	0.28	0.67	0.80	1.00
	93,614	1.86	9,361	0.20	35	0.24	0.12	0.45	0.52	0.28	0.59	0.80	0.96
	93,614	1.86	14,042	0.30	52	0.20	0.12	0.41	0.47	0.28	0.55	0.80	0.93
	93,614	1.86	18,723	0.40	70	0.17	0.12	0.37	0.42	0.28	0.51	0.80	0.91
	93,614	1.86	23,404	0.50	87	0.15	0.12	0.33	0.38	0.28	0.48	0.80	0.89
	93,614	1.86	28,084	0.60	104	0.14	0.12	0.29	0.34	0.28	0.44	0.80	0.87
G	56,211	2.5	2,811	0.10	14	0.38	0.12	0.58	0.71	0.28	0.76	0.80	1.07
	56,211	2.5	5,621	0.20	28	0.27	0.12	0.55	0.62	0.28	0.69	0.80	1.02
	56,211	2.5	8,432	0.30	42	0.22	0.12	0.52	0.58	0.28	0.64	0.80	0.99
	56,211	2.5	11,242	0.40	56	0.19	0.12	0.49	0.54	0.28	0.61	0.80	0.96
	56,211	2.5	14,053	0.50	70	0.17	0.12	0.46	0.50	0.28	0.57	0.80	0.94
	56,211	2.5	16,863	0.60	84	0.15	0.12	0.42	0.47	0.28	0.54	0.80	0.93
Total	265,626	1.46	13,281	0.10	39	0.23	0.12	0.52	0.58	0.28	0.64	0.80	0.99
	265,626	1.46	26,563	0.20	78	0.16	0.12	0.43	0.47	0.28	0.55	0.80	0.93
	265,626	1.46	39,844	0.30	116	0.13	0.12	0.34	0.38	0.28	0.48	0.80	0.89
	265,626	1.46	53,125	0.40	155	0.11	0.12	0.25	0.30	0.28	0.41	0.80	0.85
	265,626	1.46	66,407	0.50	194	0.10	0.12	0.16	0.23	0.28	0.36	0.80	0.83
	265,626	1.46	79,688	0.60	233	0.09	0.12	0.07	0.17	0.28	0.33	0.80	0.82

V. Suggestions on possible improvements to the survey design, protocols and implementation particularly with respect to issues of calibration of across planes and spotters with respect to g(0) and school size estimates.

V.1 Design

Due to the encounter rates recorded in previous years, it would be highly recommend to concentrate the survey effort in the high density areas (called "inside"), also known as the main spawning areas for bluefin tuna in the Mediterranean Sea. This would improve both the number of replicates and the number of expected recorded sightings and therefore the estimates precision would be much higher. Equally, for the mean of comparisons, we highly recommend to concentrate the survey only on the overlap areas.

If in a given year an extension would be made for surveying also outside areas or extra adjacent areas to the main overlap inside ones, a different independent design should be done for those "extra", preferably with extra resources too, in order to avoid the reduction of the effort allocated to the main areas.

Keeping the same areas over the years will improve also the budget assessment and the survey logistic, reducing many administrative problems linked to flight permits in too many FIRs.

V.2 Protocols

It is fundamental to ensure that the procedures included in the ICCAT BFT survey protocol are properly understood and enforced by all teams and observers before carrying out a new survey. It has been shown in this report how errors during data collection can produce essential problems for the analysis to estimate abundances, and the very large potential biases that can be produced, especially due to the searching pattern and the recording of the necessary data for estimating perpendicular distances.

These are some recommendations that would improve the quality of data collection.

1. GPS functioning

The use of at least 2 GPS at the same time should be a requirement. The GPS data set analysis has shown malfunction of the signal and lost data producing lack or wrong lat/lon values. After each flight the SS should check the GPS data in order to identify and solve, when possible, such problems.

2. Recoding of times and positions – use of software

The ICCAT BFT survey protocol implies getting precise times and positions (lat/lon) of different events that must be recorded:

- F: first sighting
- LE: leaving the transect
- A: animals abeam
- C: start circles
- RE: re-join transect

In many occasions these events are very close one from the other or even are almost simultaneous, so this requires the SS to be quick and precise. Recording these data by hand writing in a notebook is not an easy task and can lead (as already seen in many occasions when reviewing the data) to typos and errors in the numbers, and also an unavoidable delay between checking the time and GPS locations and writing them down, with not enough time to do it precisely and instantaneously for all events. Additionally, the stress of doing it fast enough is a potential source of errors in the writing, of which many have been found.

Nowadays there are laptops with batteries long enough up to 6 hours or even more. There are many softwares to collect time and GPS data automatically. With an easy training, SS would be able to collect the event data in a much quicker and automatic way and with much more quality, and therefore the data supplied to the analyst would be much better. It would be highly recommended to implement such an automatic collection data system in future surveys. In order to get a good implementation of this system a training session of minimum 4 hours would be necessary. This training session could be done during the training course in ICCAT installations. It would also be desirable to get an additional field training with the teams selected to carry out the surveys in each study area, even if this is currently complex due to the short survey season and the many teams and areas engaged in it, without considering that an additional person on board will limit the flight time available.

We recommend the software LOGGER, a free software developed by IFAW (International Fund for Animal Welfare) and widely used by many researchers in the field of marine mammals surveys. The analysts for ICCAT surveys, Cañadas and Vazquez, have experience of 20 years using such software.

Logger is connected directly to the GPS and records automatically the time and position (lat/lon) every xx seconds, showing it in the map screen, which in turn can show different colours for different types of effort (on, off, circling, etc.) depending on how it is configured (Logger is completely customizable), see Figure 30. In this map the recorded sightings of BFT and/or any species can be shown in real time too (Figures 31 and 32). Additionally, the designed tracks can be plotted, so important deviations from them could be detected in real time and therefore correct the course on transects.

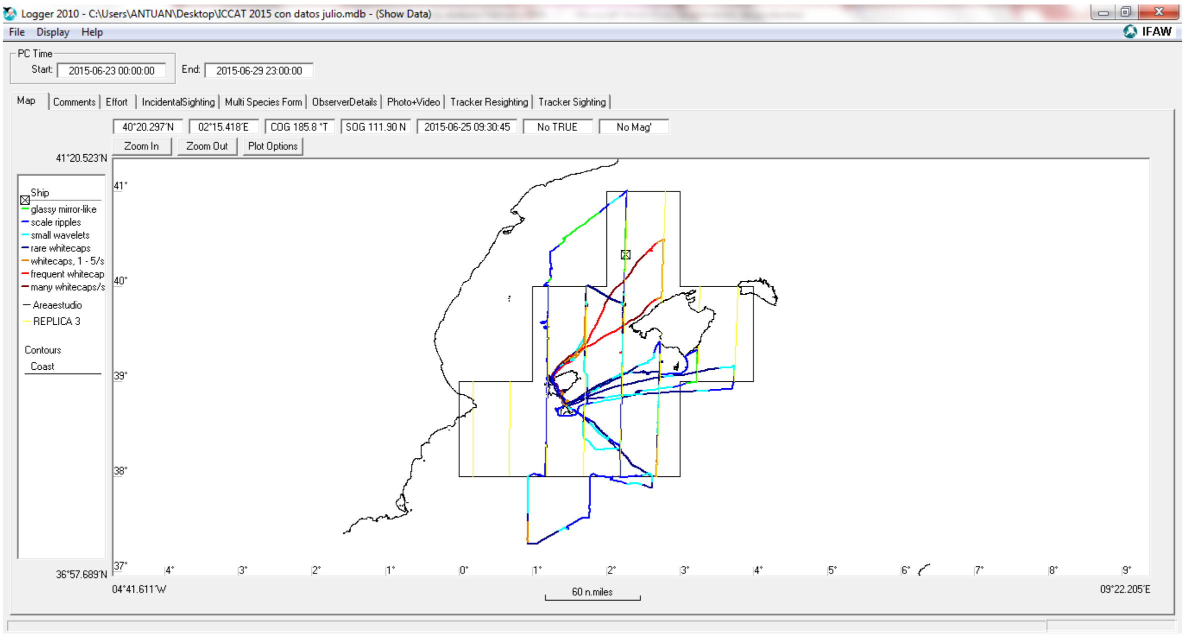


Figure 30. Example of map screen of Logger with effort (used in area A in 2015 as a test)

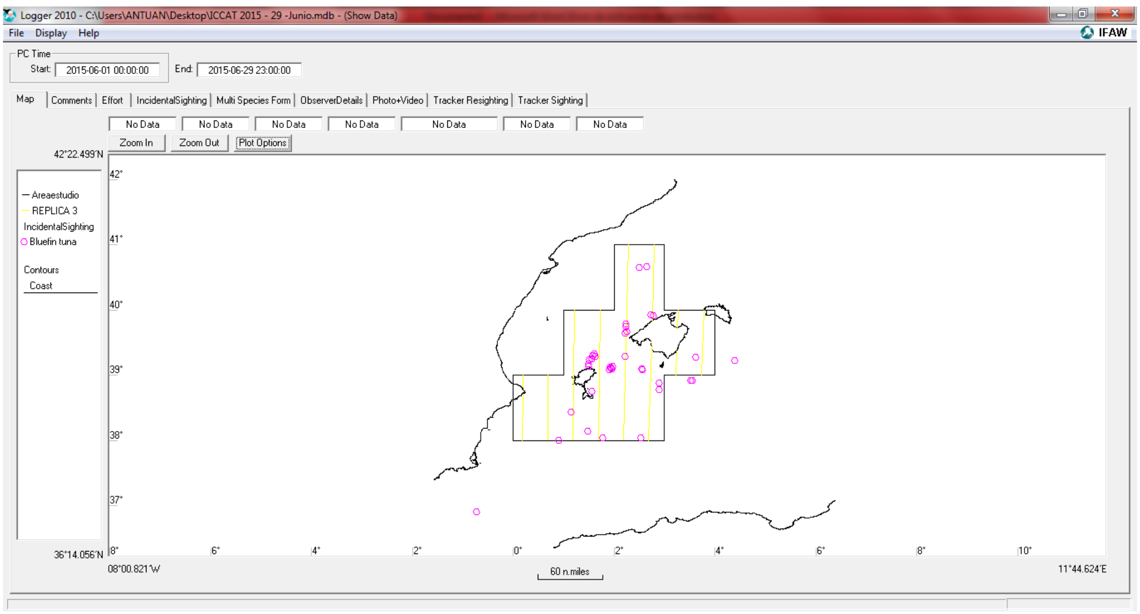


Figure 31. Example of map screen of Logger with sightings (used in area A in 2015 as a test)

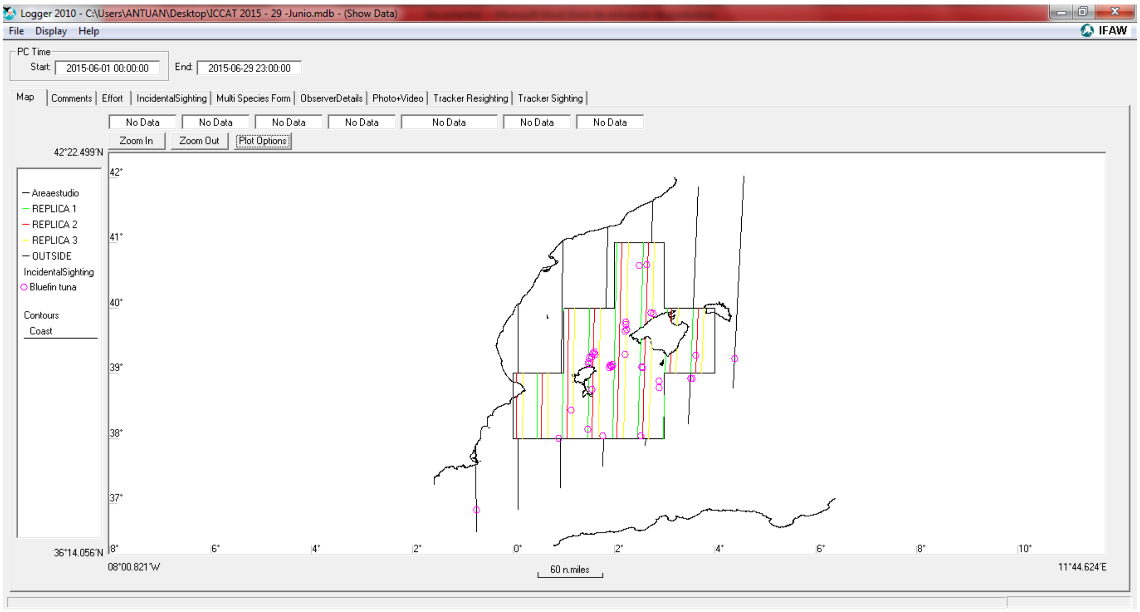


Figure 32. Example of map screen of Logger with effort and sightings (used in area A in 2015 as a test)

Logger has other screens to introduce data, for example a change in effort or an observation of BFT (or whatever species). When introducing data on the taps for effort or observations, a timestamp is recorded automatically (by pressing F1) obtaining the exact time and position when the record is done, so there is no need to record those data manually. The tabs for effort and observations can be fully customized to have the required fields for the survey. Figures 33 and 34 show examples of these screens for data entry from a cetacean survey (but as said, these can be completely customized for the BFT surveys).

We strongly recommend using this software for next surveys as it would simplify enormously the data collection by the observers, and the data checking and analysis by the analysts, reducing the amount of errors considerably, increasing therefore the reliability of the data.

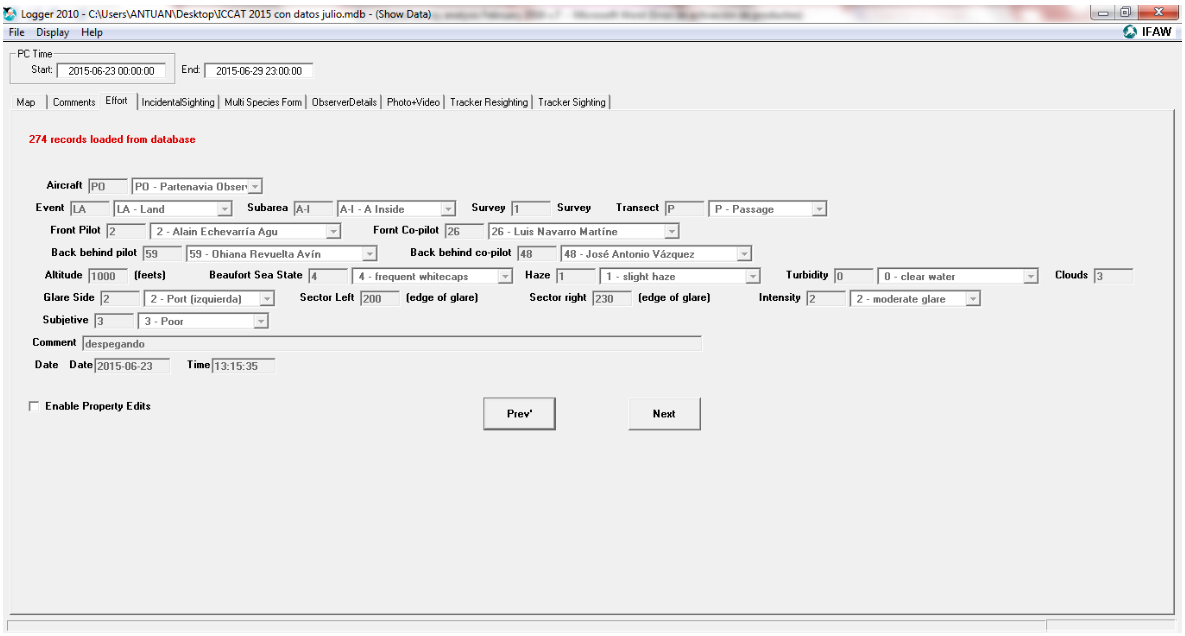


Figure 33. Example of effort data entry screen in Logger (used in area A in 2015 as a test)

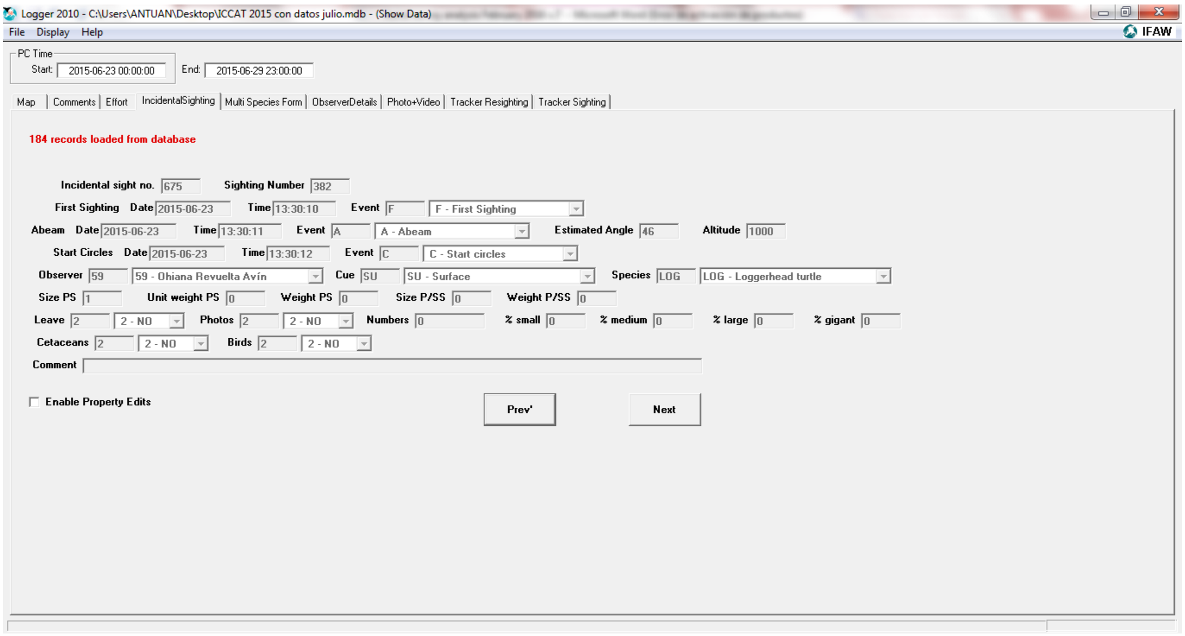


Figure 34. Example of observations data entry screen in Logger (used in area A in 2015 as a test)

3. Columns for data quality for perpendicular distances

Due to the problems identified during perpendicular distance estimation using both the angle from the clinometer and the one calculated using GIS software, it would be highly recommended to include two new columns in the sighting form. The first one to indicate whether the angle from the inclinometer has been collected in the right way or not (animals abeam and clear view of the group by the SS), and the second one to inform in which circle animals were visible. This is very important mainly when observers need more than one circle to detect animals. It would make it easier to know which circles need to be chosen to estimate perpendicular distance by GIS.

4. Glare

In ICCAT BFT 2015 protocol glare intensity and coverage (left and right angle) were recorded, but in general glare only affects one side of the plane. It would be worth investigating whether it would be necessary to consider only one side in sample area calculations in the cases when glare is intense on one side of the plane.

Another point to be considered is that there is no way to distinguish among situations where glare angles are similar but the area affected by it is not. This situation is shown in Figure 35 a, b, c and d; angles would be similar (around 90-100) but whereas in Figure 35a only a small part of the area is affected (from 20 to 40 degrees), and the closest area to the transect is clear without glare, in Figure 35d, the whole area is affected from 0 to 90 degrees. Therefore, for the same angle and intensity recorded, the visibility is still very different. This shows how complex the issue of glare is. Maybe it would be worth to record also the vertical angle of the glare, although it makes data recording more complex.

5. Bubble windows

In the light of the analysis in this report, bubble windows should always be used in all surveys.

6. School size estimation

As mentioned in section II of this report, we make two recommendations in this issue. The first one is that it should be insisted to the observers that both SS and PS should always try to give an estimate of school size and weight independently for each school detected. The second one is described below under "Calibration".

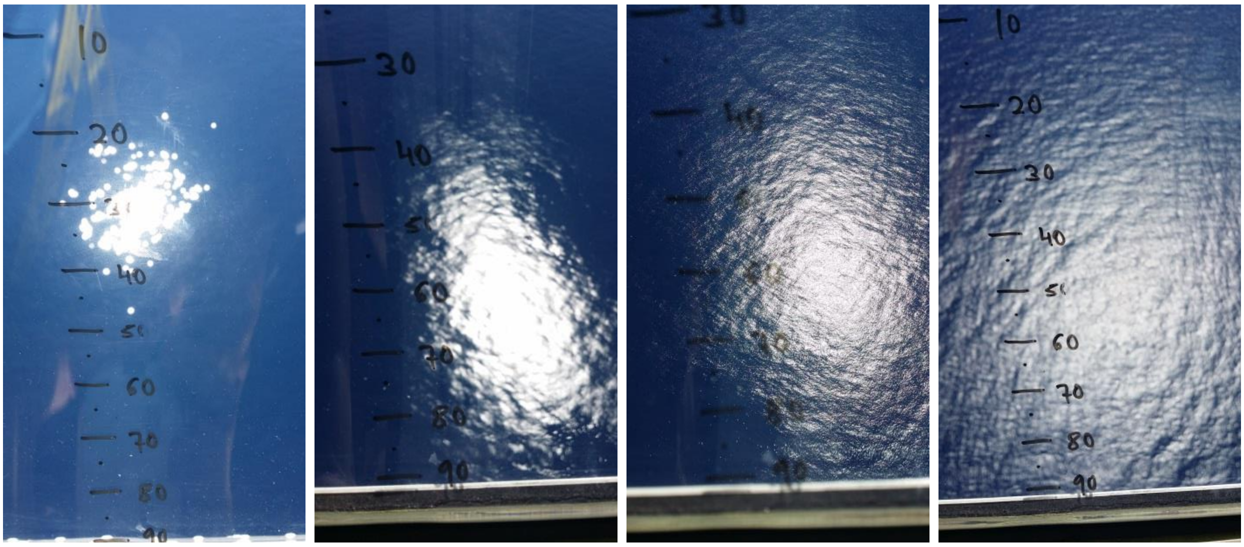


Figure 35 a, b, c and d. Effect of glare.

V.3 Calibrations

The possible calibration among all spotters engaged in the ICCAT GBYP aerial survey has been analysed in a recent paper presented to SCRS in 2015 (Di Natale, in press). This issue was deeply discussed also by the SCRS BFT Species Group, which shared the conclusions of the paper and mostly recommended to keep the same spotters in the same area over the years. Calibration is anyway an issue to be further discussed, at least in order to find operational solutions for smoothing some problems.

1. School size estimation

The only way to attempt to estimate a correction factor for school size is through an experiment with the available photographs, which are the only means to obtain a ground truth of the size of a school (the most accurate one, as there might be animals not visible on the picture). With a good collection of pictures where the number of fish have been `previously counted, show them to the observers (for example during a training session at the ICCAT office) during the time (xx seconds) that groups are usually available when spotted from the airplane, and asking the observers to write down the numbers estimated. It can be done both in almost instantaneous images (emulating those detection when it is not possible to circling) and with the average time they are usually seen while circling. With enough sample size per observer, it can be investigated whether there are trends to under or overestimate group sizes by each observer and if so estimate potential correction factors. The limit of this exercise is that human vision has much more detection capabilities than any available photo methodology (except those used by military institutions, which are not available for civil uses, like the old “Water penetration” film by Kodak). This exercise is therefore useful for assessing the individual capability for assessing the school as in the photo and for detecting overestimation or underestimation individual trends.

2. PS-SS same airplane

In general, SS should concentrate their searching effort on the closest area to the track line, especially having the bubble windows, although in practice they still search towards longer distances too. PS do not have bubble windows and therefore they miss the schools closer to the track line, but in turn they have much more experiencing detecting BFT and therefore have, in general, more detections probability than SS. On the other hand, even in those cases when BFT are detected first by SS, in most cases the PS is able to see the school too while circling the animals. Hence, the PS almost always have the opportunity of providing school size and weight estimates.

It is very important that both PS and SS provide an estimate of school size, independently of who detects it first. If enough sample size of parallel estimates are available for each spotter, and the school size estimate of PS is considered the reference for the analysis (for means of consistency and based on the usually larger experience of PS), then potentially a correction factor of the SS estimates could be attempt to be obtained.

This would be useful for the very few cases when there is only school size estimate from SS (e.g. school lost before circling, so no estimate from PS is available), to try to correct it if a pattern in the comparisons were found.

3. PS-PS different airplanes and areas

In the light of all what has been explained above, the only way that a comparative study would make sense, would be a relative comparison between all the PS working in different areas through a pilot study. Such pilot study might be carried out during the purse seiner fishing period before the ICCAT BFT survey. The 4 PS could go in one airplane covering a high density area during a period with enough days to get a statistically sufficient number of observations, taking into account the fact that the observers are sitting on two different sides of the aircraft. In this way, the estimates from the four experienced PS could be compared and see the level of discrepancies or otherwise. Nevertheless, we are aware of the enormous difficulties of doing so from the logistical and economical point of view; a dedicated time and budget should be made available in case this exercise would be decided. The bias induced by the comparative sonar estimates is another factor to be taken into account in this exercise.

Alternatively, the same experiment for group size estimation described above, with photographs, could be used to test the accuracy of the PS and their agreements or discrepancies. This would be a much less expensive option (although maybe not as close to real scenarios).

V.4 Stability of the same team over the years and areas

This is possibly the main recommendation, because the stability of teams over the years and the areas is one of the main factor for reducing both variance and bias. Even if we are aware of the many constrains and administrative limits in ICCAT GBYP, we must point out this factor and this need. It is very clear from all the analyses that moving the spotter and teams in different areas over the following years will induce a lot of additional biases, while a different composition of the team within the same company in different years add further biases.

Furthermore, a stable team will contribute to better internal coordination and a possible professional improvement of all members.

V.5 The problem of the Professional Spotters

It is important to note that the Professional Spotters became as such during many years of spotting activity for the fishing fleets. This happened while aerial spotting was permitted for the eastern Atlantic bluefin tuna fishery, up to 2006. After the enforcement of ICCAT Rec. 06-05, aerial spotting was not possible anymore. This happened 10 years ago. As a matter of fact, many professional spotters available at that time, with different experience (some of them had already more than 25 years of spotting experience) moved to other jobs, sometimes in the fishery sector. This situation results in the fact that only some of them are still available for the ICCAT GBYP aerial survey, while their availability is absolutely essential, both because of their capacities and the fact that they are the ones able to train also the scientific spotters, particularly for assessing the weight.

This is a limiting factor that should be considered for future surveys and the recommendation provided in the previous point would possibly smooth this problem.

VI. General conclusions

- The initial choice of the four main spawning areas, at that time based on GPS data of the main fleets over three years and more than two centuries of scientific knowledge, confirmed its validity and robustness;
- The aerial survey on spawning aggregations, besides limits and constrains, is one of the very few fishery independent methodologies for getting trends over the years;
- The behaviour of bluefin tuna is quite different from any other non-fish species (marine mammals or turtles), for which this survey technique was mostly developed; therefore, even spotting this species is more difficult than spotting others, and this needs to be taken into account;
- The best strategy for getting more solid results should be carrying out the survey without gaps between years, preferably always with the same team in each area;
- Extended surveys could be planned from time to time, in order to catch any possible change in spawning behaviour, but only if the extended survey can ensure the same previous coverage for the main area and set additional coverage for the new areas;
- The extremely complex geography of the Mediterranean induces a high variability of different environmental conditions even in very short spaces within the same day; this further complicates the efforts of the teams; how this biases the results of the many analyses, it is difficult to properly assess;
- The logistic behind the extended survey was extremely complex, due to the many constraints, including the limited availability of the right type of fuel in most of the airports and the recent security controls;
- The complex situations of both FIRs and national regulations made and makes the ICCAT GBYP surveys a real challenge and the results obtained so far, with all limits included in this report, must be evaluated accordingly.
- The amount of effort per area needs to be, at least maintained as in the first years, but it would be better to increase it if any useful index of abundance with appropriate CVs are to be obtained, in order to assess trends over time at manageable time frames.

References

Aranda, G., Abascal, F.J., Varela, J.L. and Medina, A. 2013. Spawning Behaviour and Post-Spawning Migration Patterns of Atlantic Bluefin Tuna (*Thunnus thynnus*) Ascertained from Satellite Archival Tags. PLoS ONE 8(10): e76445. doi:10.1371/journal.pone.0076445

Borchers, D.L., Zucchini, W., Heide-Jorgensson, M.P., Cañadas, A. and Langrock, R. 2013. Using hidden Markov models to deal with availability bias on line transect surveys. Biometrics. DOI: 10.1111/biom.12049

Buckland, S.T., Anderson, R., Burnham, K.P., Laake, J.L. 1993. *Distance sampling: estimating abundance of biological populations.* London; Chapman and Hall.

Buckland, S.T., Anderson, R., Burnham, K.P., Laake, J.L., Borchers, D.L. and Thomas, L. 2004. Advance Distance Sampling. Examining abundance of biological populations. Oxford University Press.

Di Natale A., in press. Tentative SWOT analysis for the calibration of ICCAT GBYP aerial survey for Bluefin tuna spawning aggregations. SCRS/2015/143: 19 p.

Fortuna, C., Kell, L, Holcer, D. Canese, S., Filidei, E.Jr., Mackelworth, P. and Donovan, G. 2014. Summer distribution and abundance of the giant devil ray (*Mobula mobular*) in the Adriatic Sea: Baseline data for an iterative management framework. Sci. Mar. 78(2): 227237. doi: http://dx.doi.org/10.3989/scimar.03920.30D

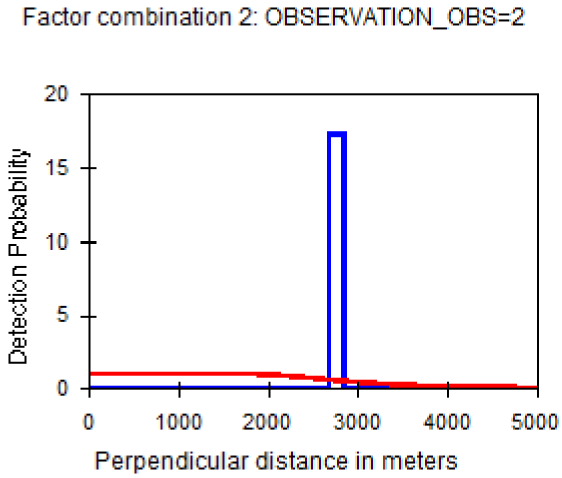
Gerrodette T. 1987. A power analysis for detecting trends. Ecology 68(5): 13641372. http://dx.doi.org/10.2307/1939220

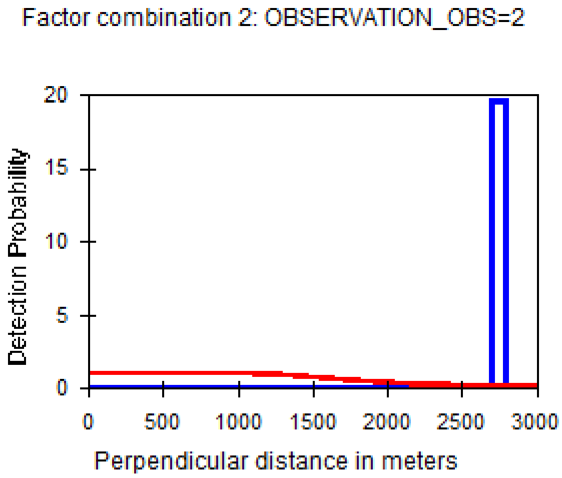
Seber, G.A.F. 1982. The Estimation of Animal Abundance and Related Parameters. MacMillan, New York.

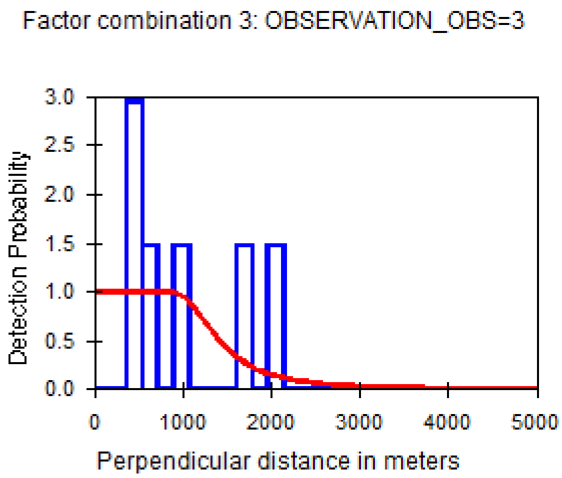
Annex 1

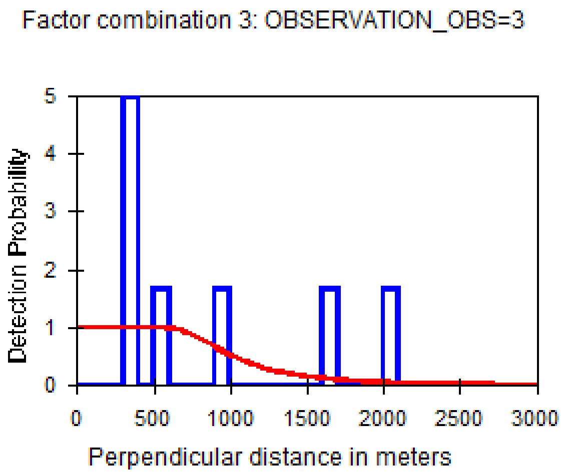
The "OBSERVATION_OBS='n'" corresponds to the Observer ID as reflected in Table 7 of the report.

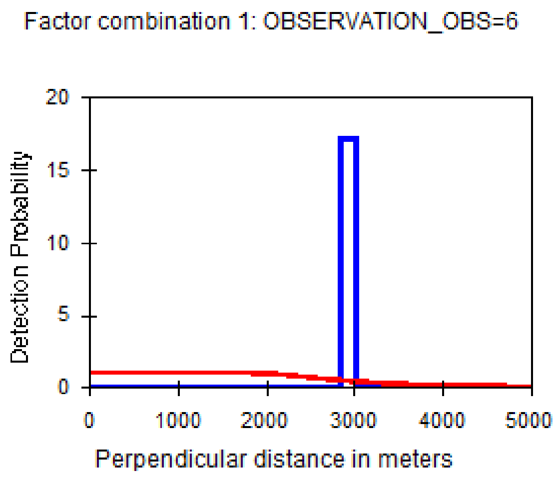
BFT	All species
Pilots	
Perigod	

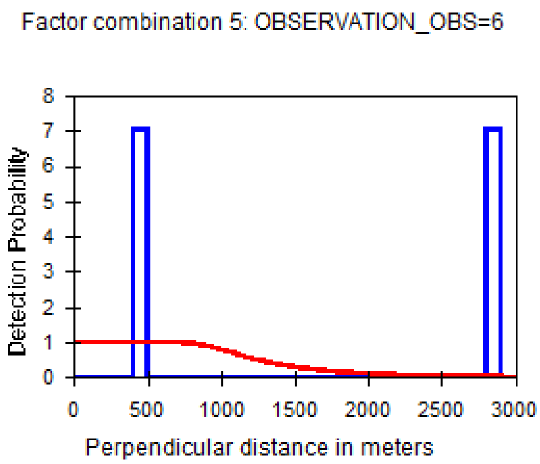




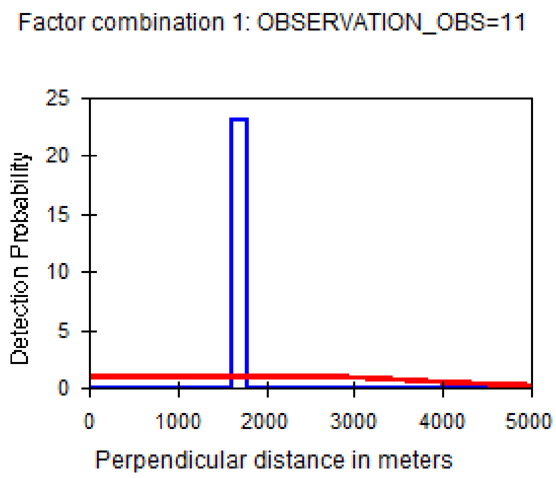


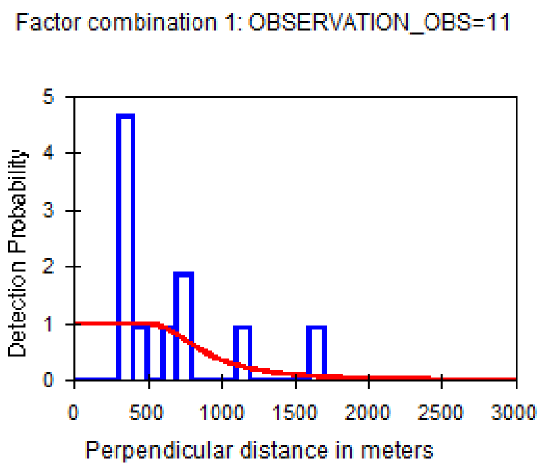






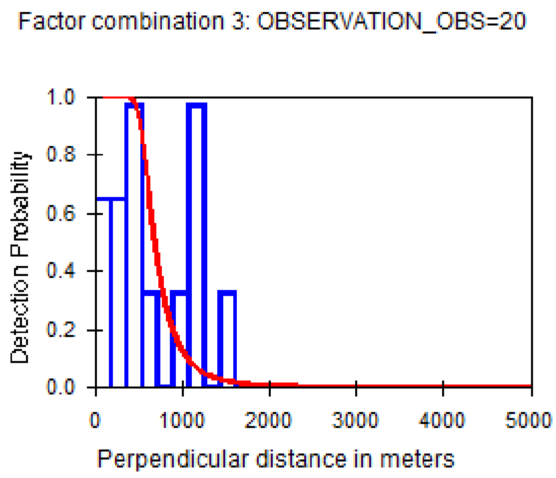
Air-Med

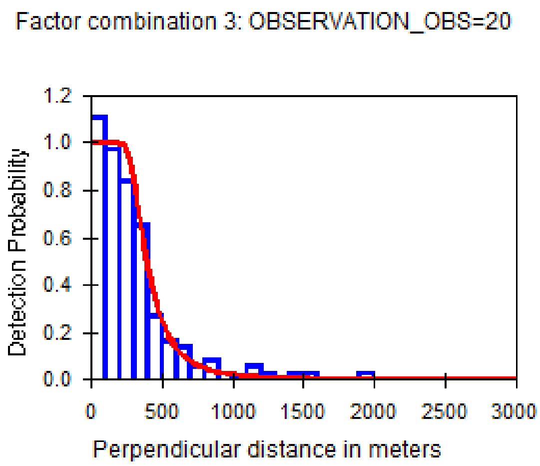




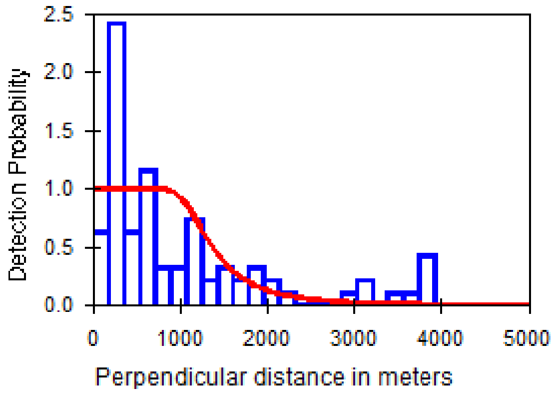
Professional Spotters

Perigod

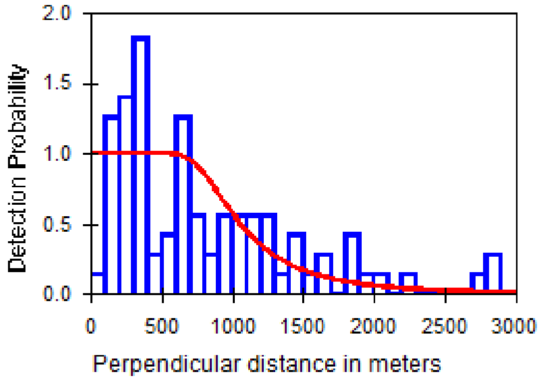




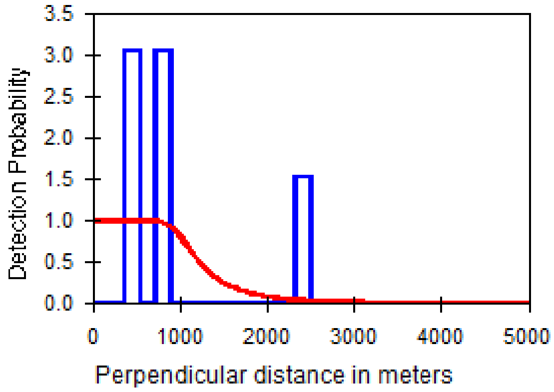
Factor combination 5: OBSERVATION_OBS=22



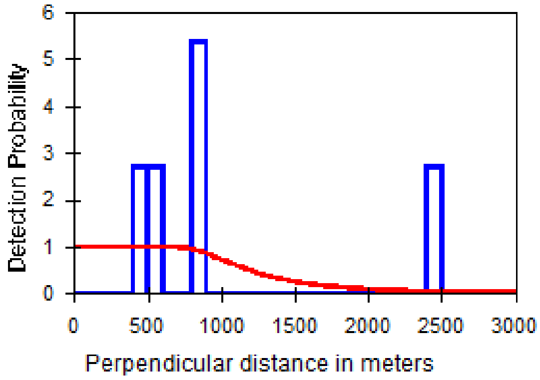
Factor combination 5: OBSERVATION_OBS=22



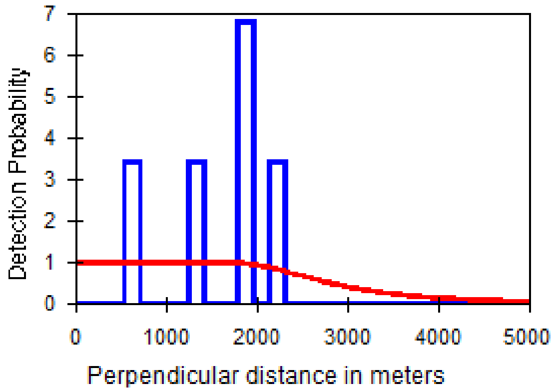
Factor combination 6: OBSERVATION_OBS=23



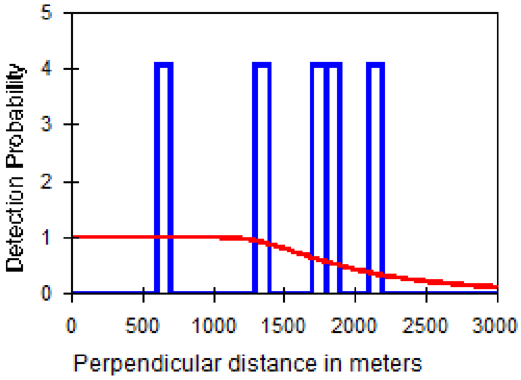
Factor combination 6: OBSERVATION_OBS=23



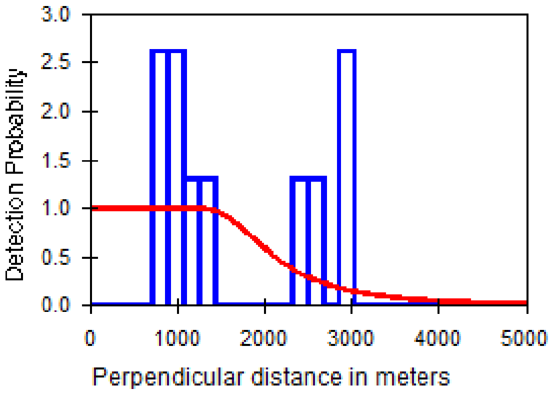
Factor combination 8: OBSERVATION_OBS=25



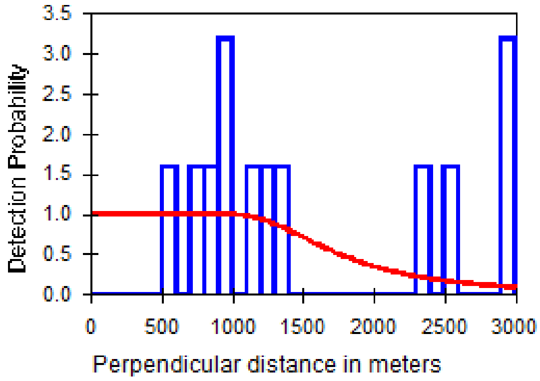
Factor combination 8: OBSERVATION_OBS=25



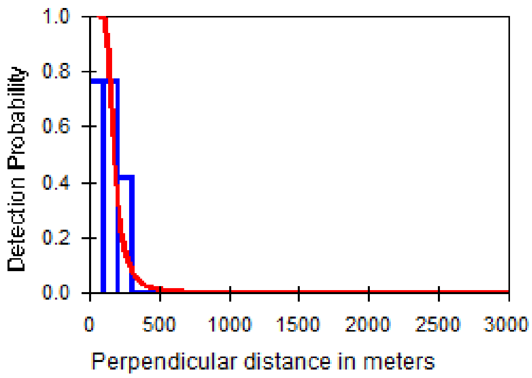
Factor combination 8: OBSERVATION_OBS=35



Factor combination 9: OBSERVATION_OBS=35

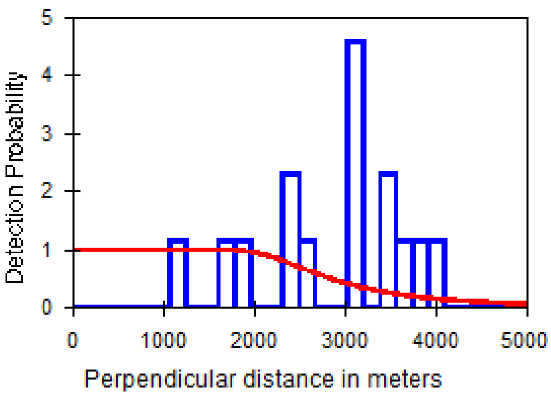


Factor combination 3: OBSERVATION_OBS=68

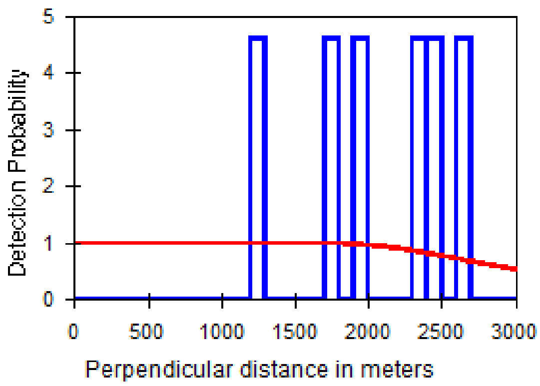


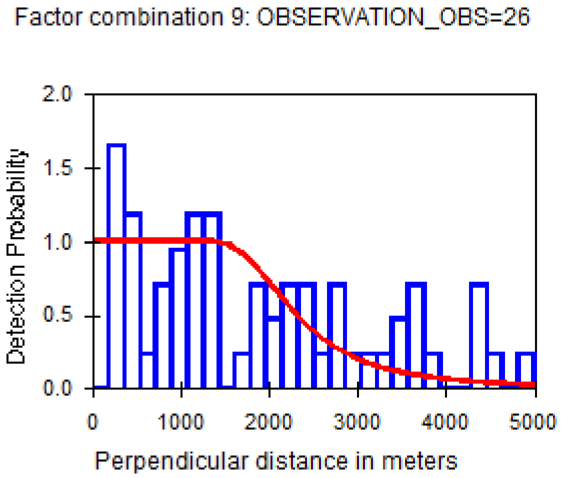
Air-Med

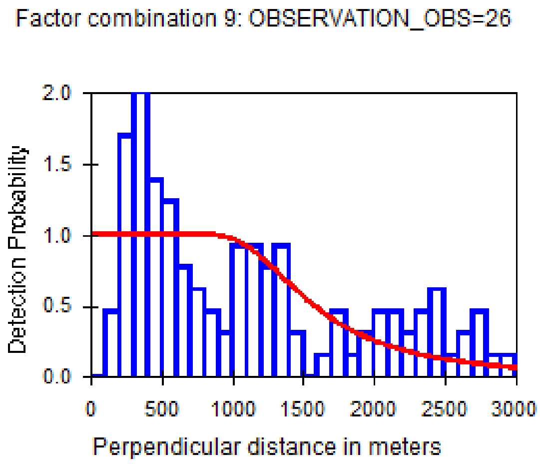
Factor combination 4: OBSERVATION_OBS=21



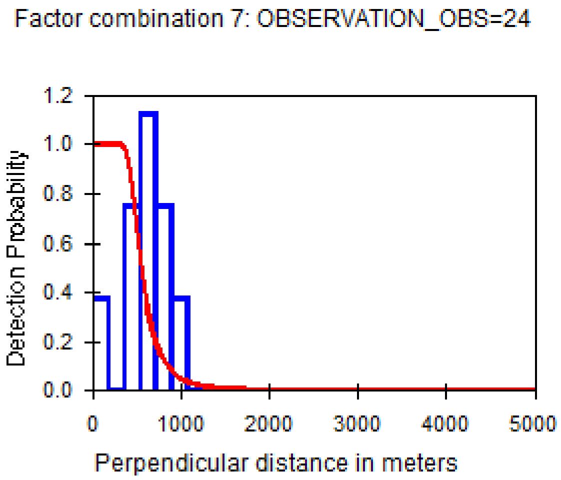
Factor combination 4: OBSERVATION_OBS=21

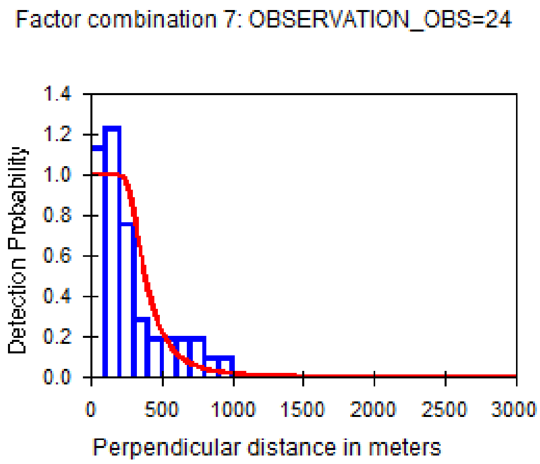


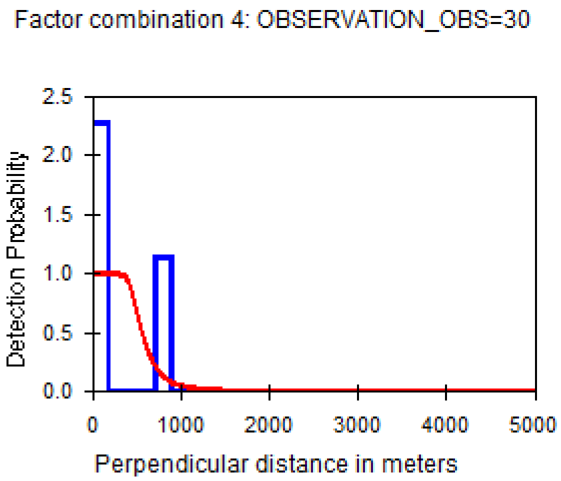

Factor combination 9: OBSERVATION_OBS=26
Detection Probability
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters

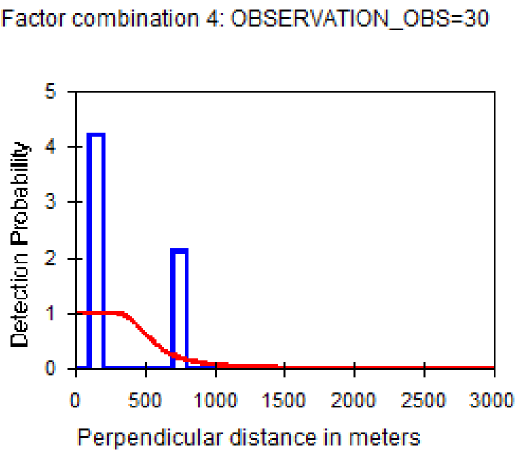

Factor combination 9: OBSERVATION_OBS=26
Detection Probability
2.0
1.5
1.0
0.5
0.0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters

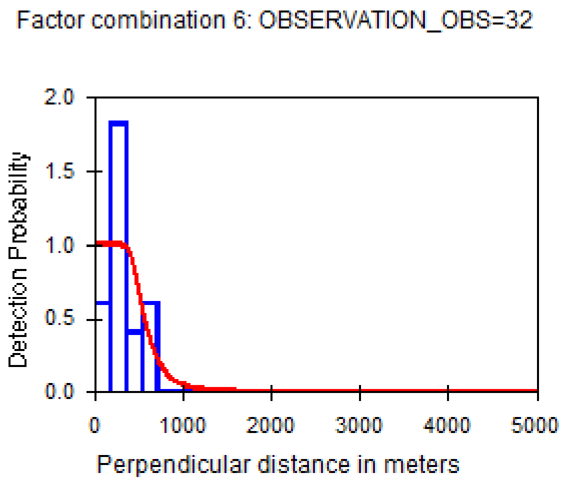
ActionAir

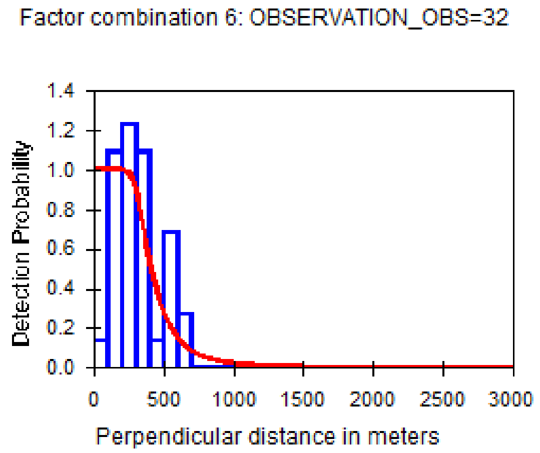

Factor combination 7: OBSERVATION_OBS=24
Detection Probability
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters


Factor combination 7: OBSERVATION_OBS=24
Detection Probability
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters

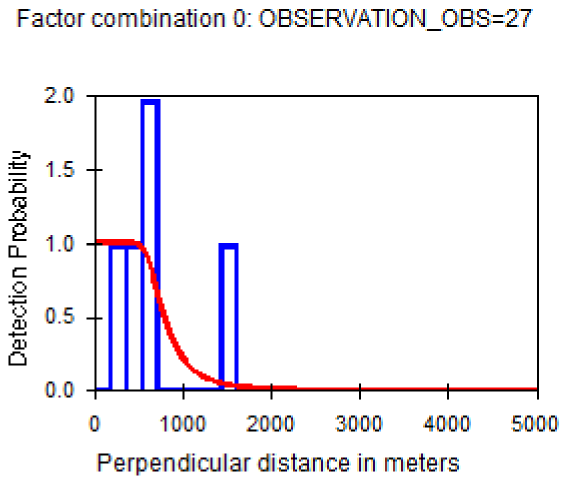

Factor combination 4: OBSERVATION_OBS=30
Detection Probability
2.5
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters

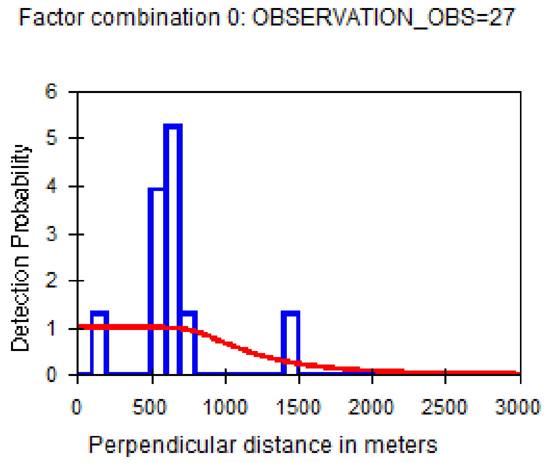

Factor combination 4: OBSERVATION_OBS=30
Detection Probability
5
4
3
2
1
0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters

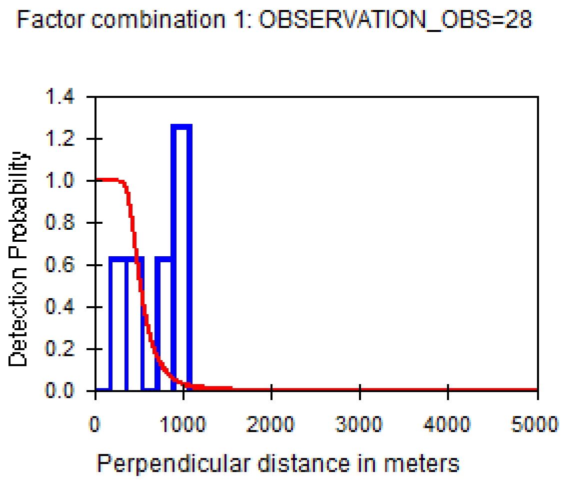

Factor combination 6: OBSERVATION_OBS=32
Detection Probability
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters

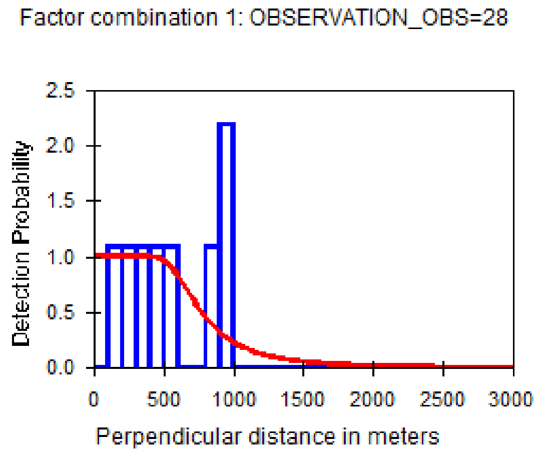

Factor combination 6: OBSERVATION_OBS=32
Detection Probability
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters

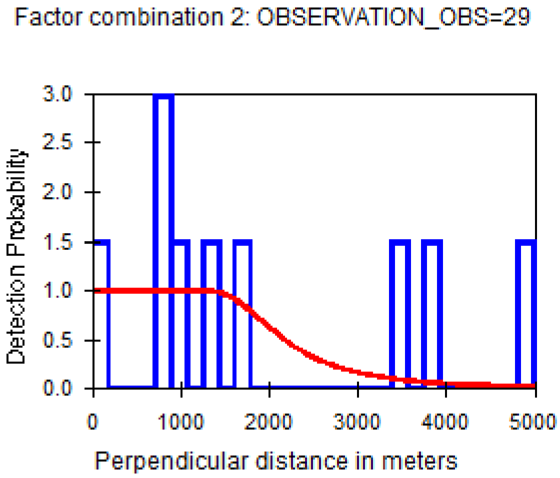
Unimar

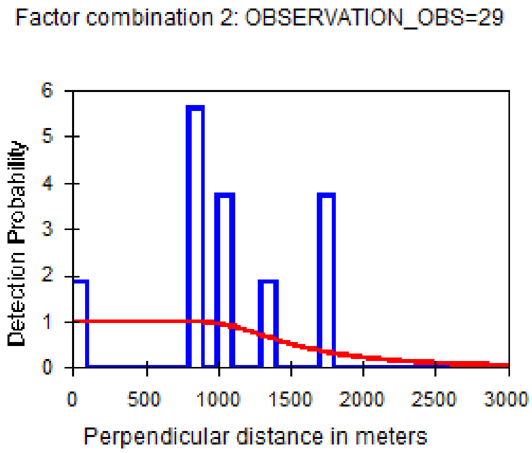

Factor combination 0: OBSERVATION_OBS=27
Detection Probability
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters

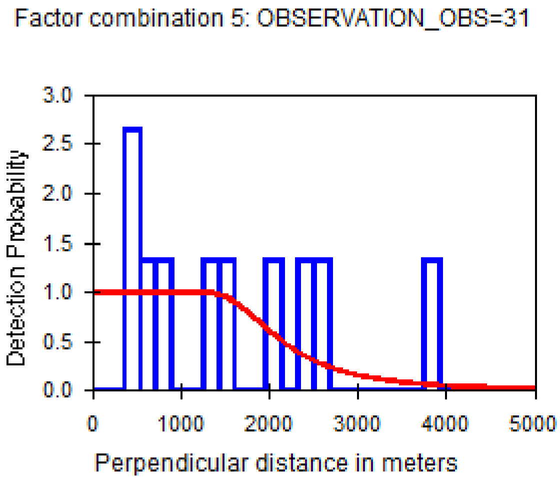

Factor combination 0: OBSERVATION_OBS=27
Detection Probability
6
5
4
3
2
1
0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters

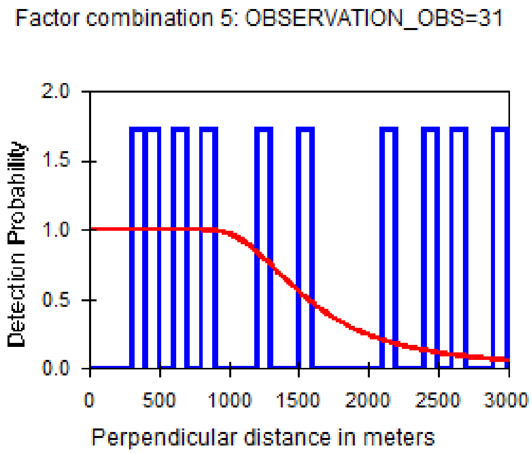

Factor combination 1: OBSERVATION_OBS=28
Detection Probability
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters

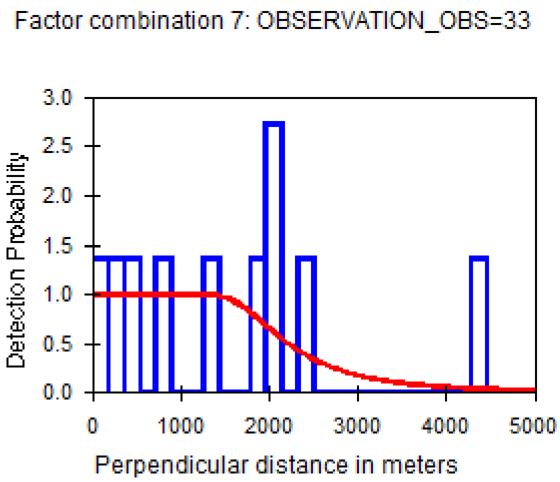

Factor combination 1: OBSERVATION_OBS=28
Detection Probability
2.5
2.0
1.5
1.0
0.5
0.0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters

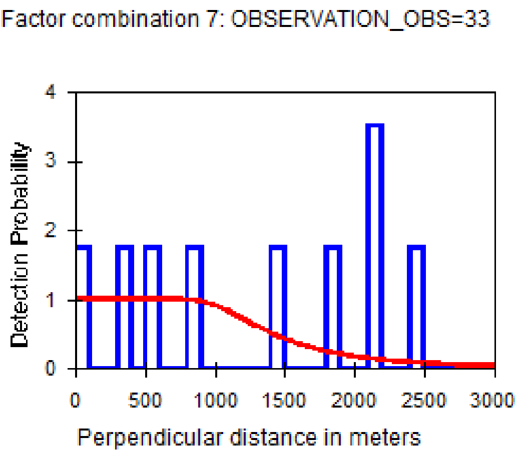

Factor combination 2: OBSERVATION_OBS=29
Detection Probability
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters


Factor combination 2: OBSERVATION_OBS=29
Detection Probability
6
5
4
3
2
1
0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters

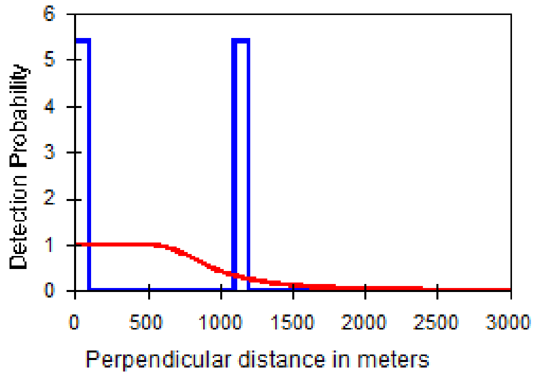

Factor combination 5: OBSERVATION_OBS=31
Detection Probability
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters


Factor combination 5: OBSERVATION_OBS=31
Detection Probability
2.0
1.5
1.0
0.5
0.0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters


Factor combination 7: OBSERVATION_OBS=33
Detection Probability
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters


Factor combination 7: OBSERVATION_OBS=33
Detection Probability
4
3
2
1
0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters

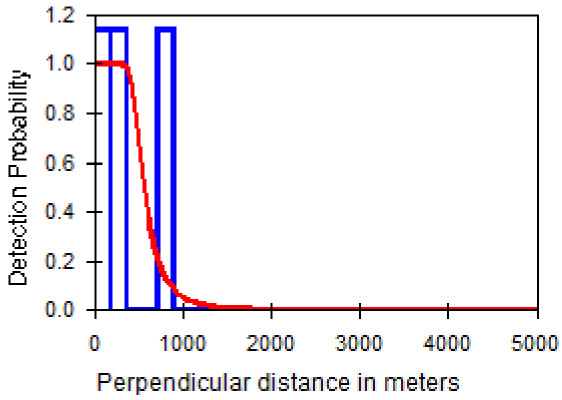
Factor combination 8: OBSERVATION_OBS=34



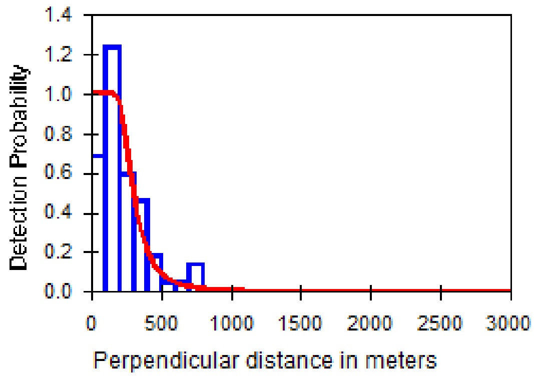
Scientific Spotters

Perigod

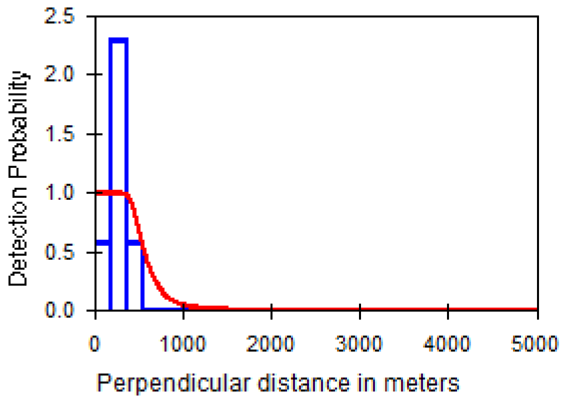
Factor combination 2: OBSERVATION_OBS=42



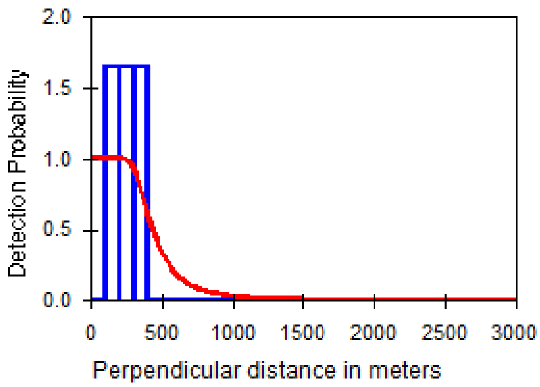
Factor combination 4: OBSERVATION_OBS=42



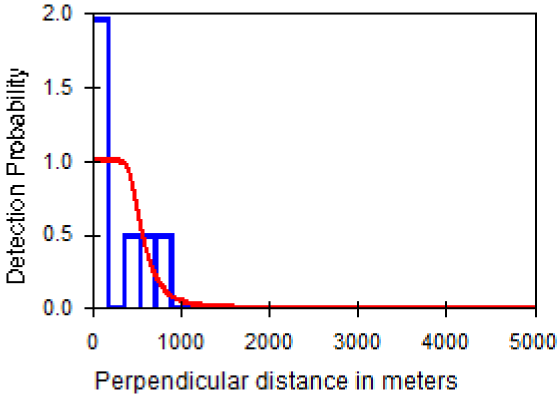
Factor combination 3: OBSERVATION_OBS=43



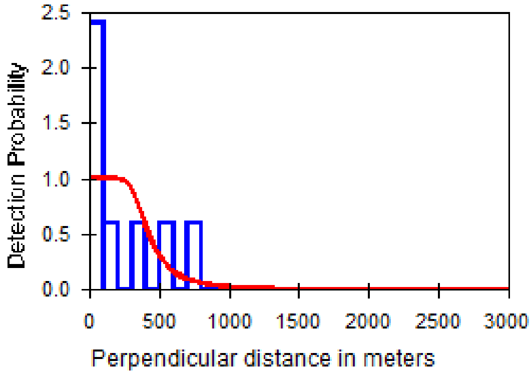
Factor combination 6: OBSERVATION_OBS=43



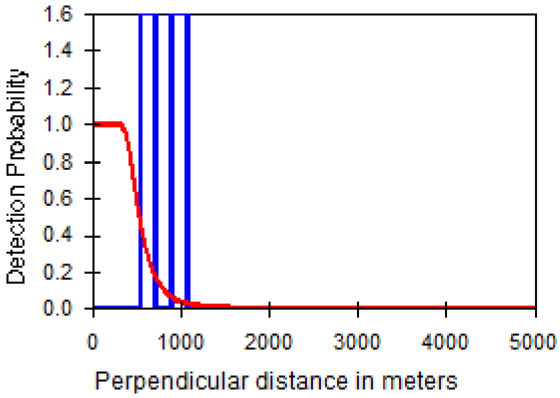
Factor combination 8: OBSERVATION_OBS=52



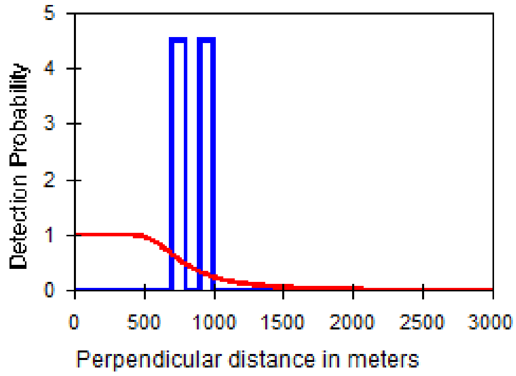
Factor combination 0: OBSERVATION_OBS=52



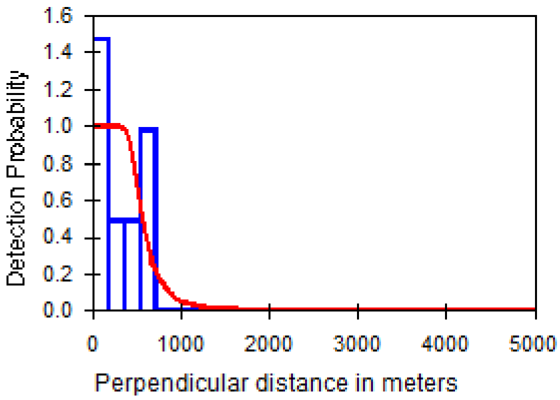
Factor combination 2: OBSERVATION_OBS=61



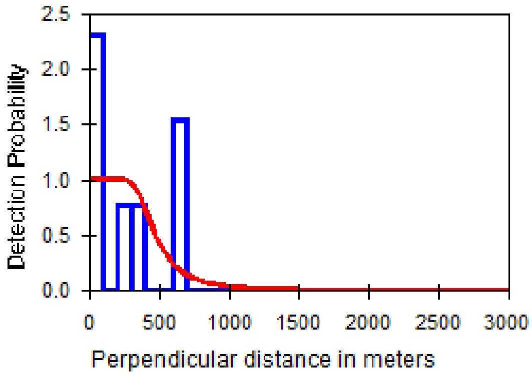
Factor combination 6: OBSERVATION_OBS=61

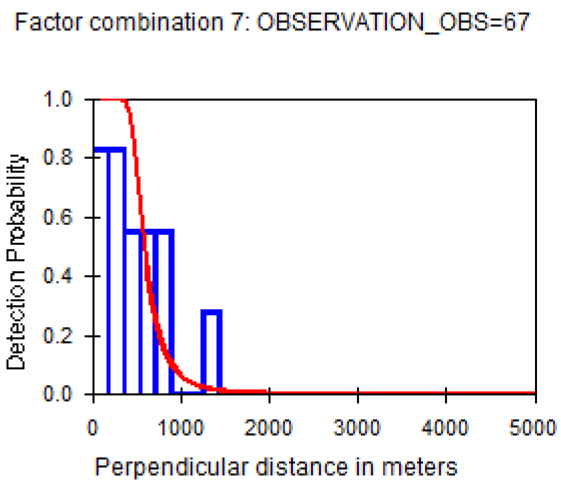


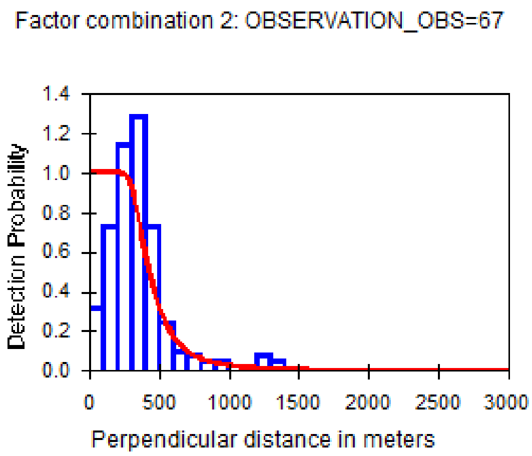
Factor combination 4: OBSERVATION_OBS=64

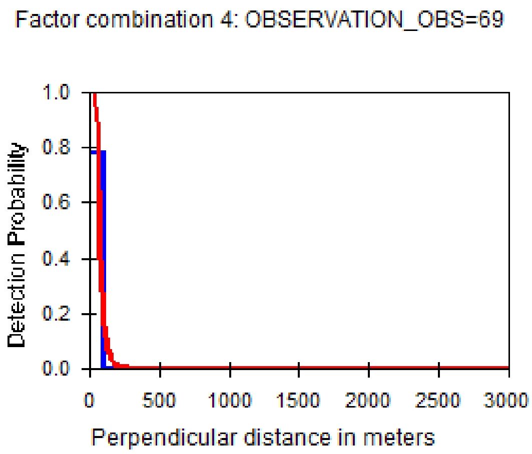


Factor combination 9: OBSERVATION_OBS=64

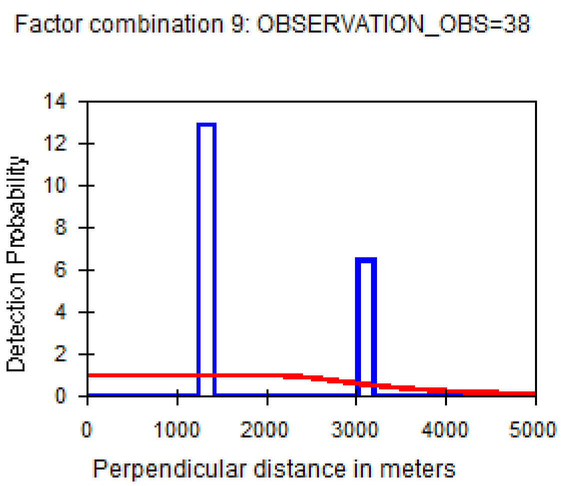


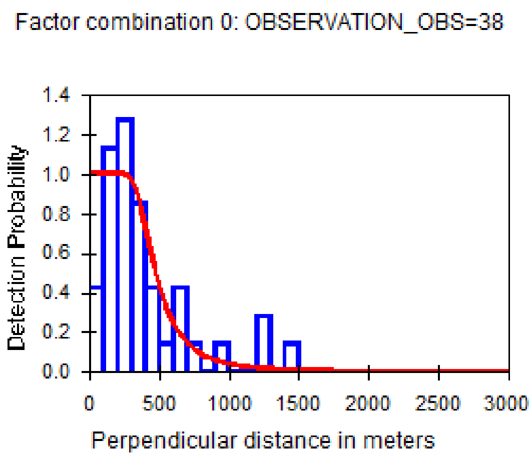

Factor combination 7: OBSERVATION_OBS=67
Detection Probability
Perpendicular distance in meters


Factor combination 2: OBSERVATION_OBS=67
Detection Probability
Perpendicular distance in meters

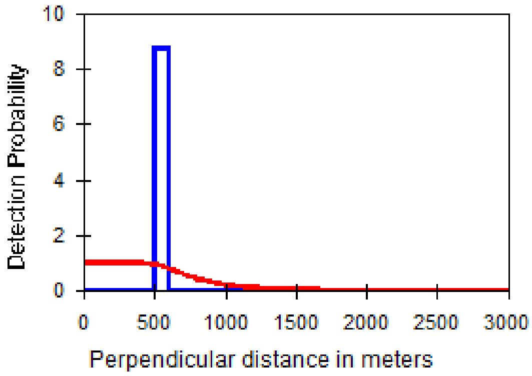

Factor combination 4: OBSERVATION_OBS=69
Detection Probability
Perpendicular distance in meters

Unimar

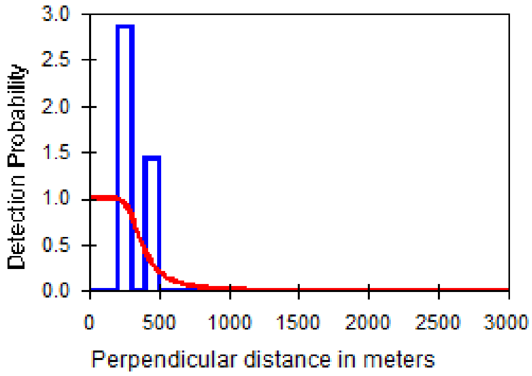

Factor combination 9: OBSERVATION_OBS=38
Detection Probability
Perpendicular distance in meters


Factor combination 0: OBSERVATION_OBS=38
Detection Probability
Perpendicular distance in meters

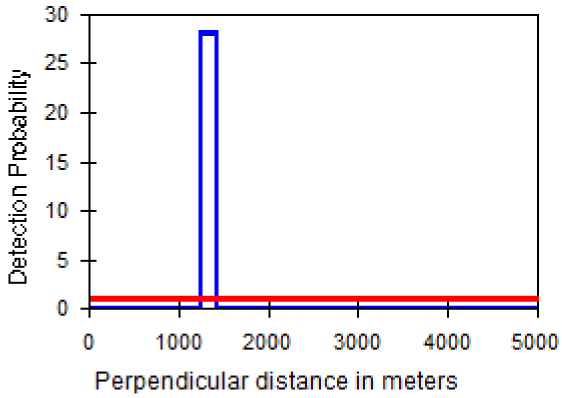
Factor combination 1: OBSERVATION_OBS=39



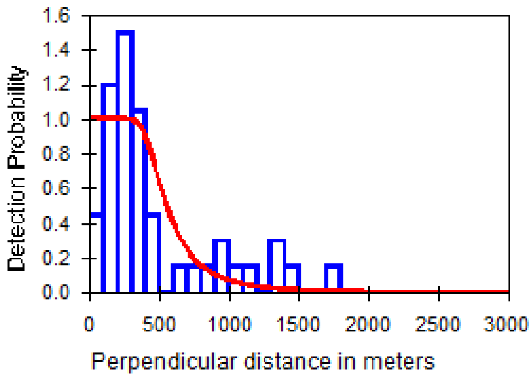
Factor combination 2: OBSERVATION_OBS=56



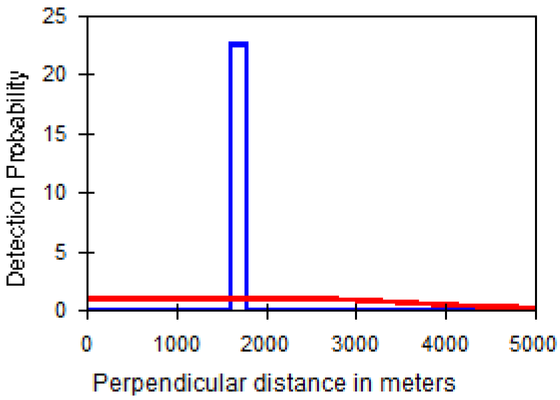
Factor combination 5: OBSERVATION_OBS=65



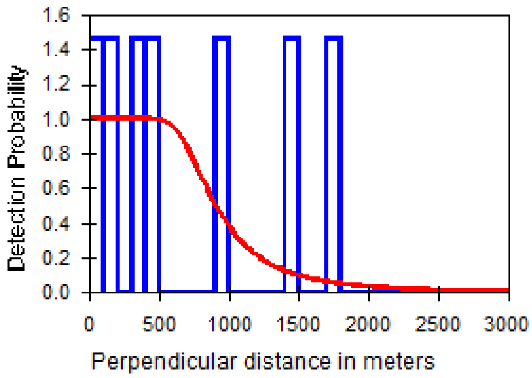
Factor combination 0: OBSERVATION_OBS=65



Factor combination 6: OBSERVATION_OBS=66

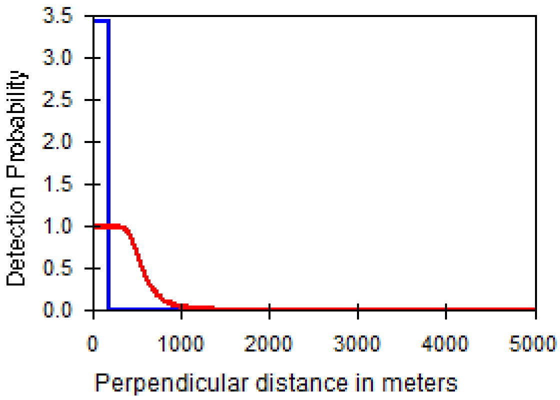


Factor combination 1: OBSERVATION_OBS=66

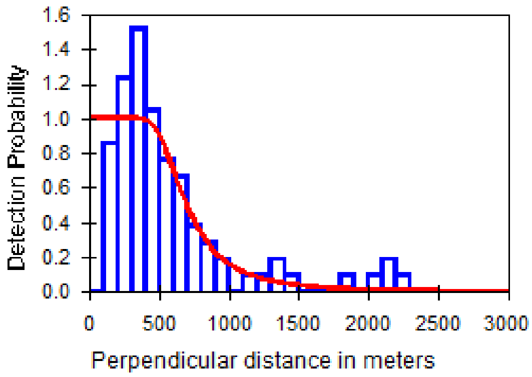


Air-Med

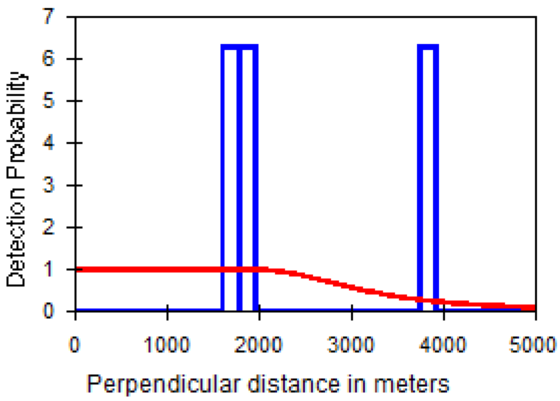
Factor combination 0: OBSERVATION_OBS=40



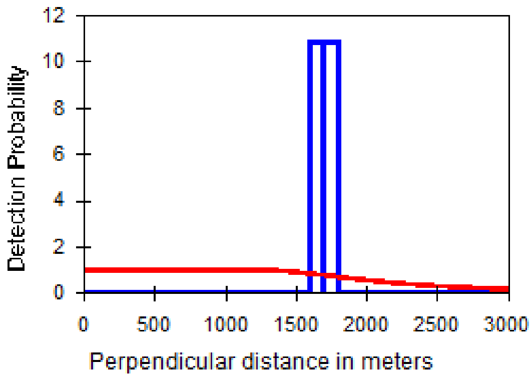
Factor combination 2: OBSERVATION_OBS=40

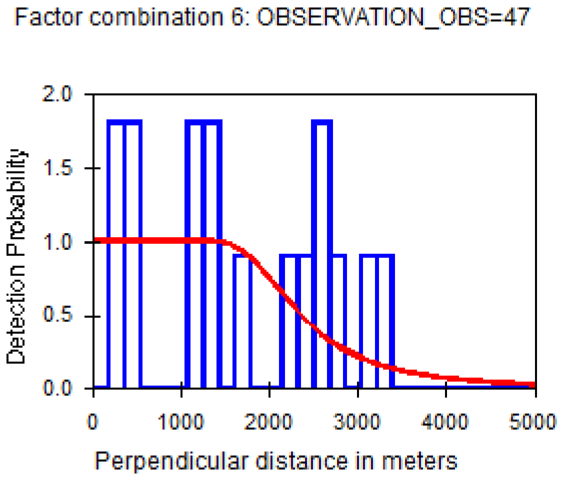


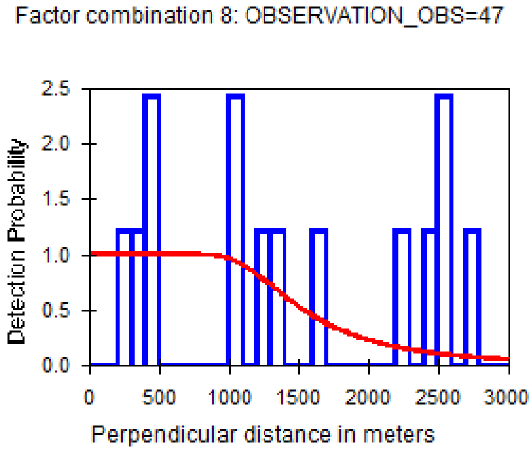
Factor combination 5: OBSERVATION_OBS=46

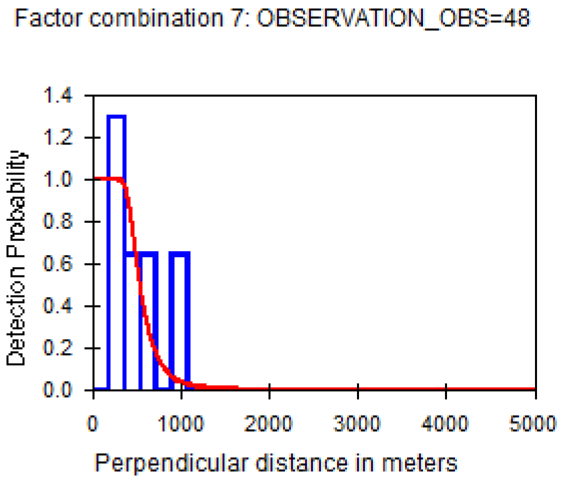


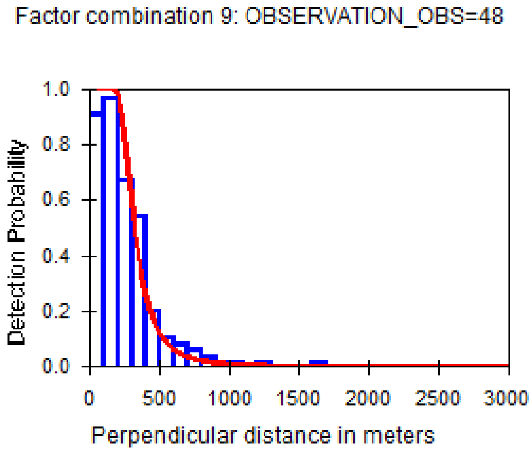
Factor combination 7: OBSERVATION_OBS=46

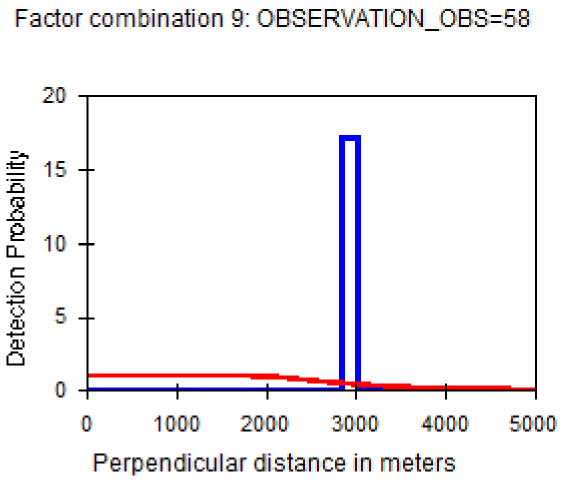


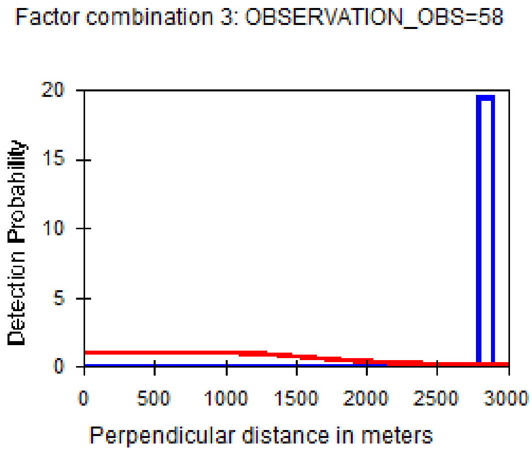

Factor combination 6: OBSERVATION_OBS=47
Detection Probability
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters

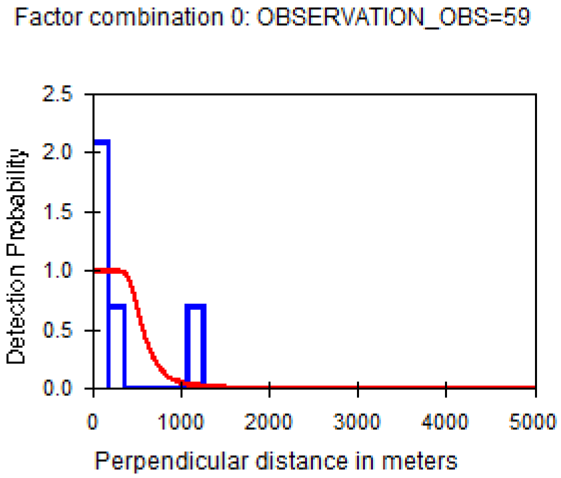

Factor combination 8: OBSERVATION_OBS=47
Detection Probability
2.5
2.0
1.5
1.0
0.5
0.0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters

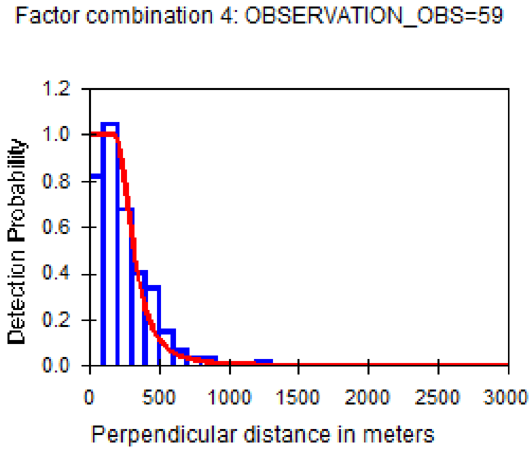

Factor combination 7: OBSERVATION_OBS=48
Detection Probability
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters

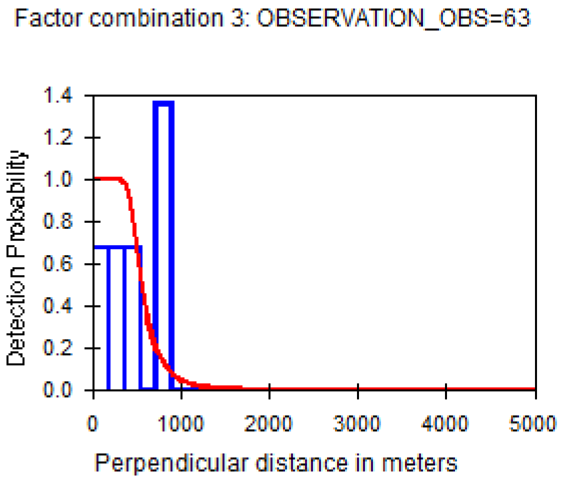

Factor combination 9: OBSERVATION_OBS=48
Detection Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters

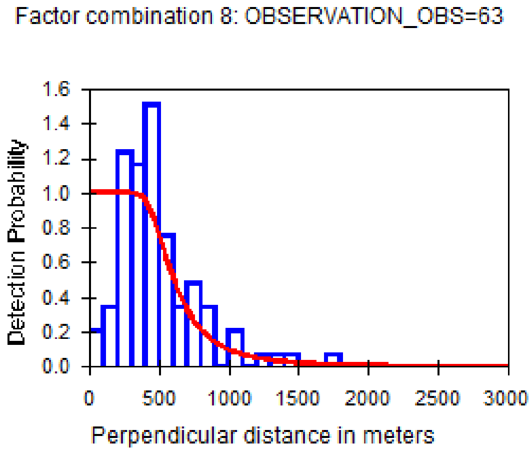

Factor combination 9: OBSERVATION_OBS=58
Detection Probability
20
15
10
5
0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters


Factor combination 3: OBSERVATION_OBS=58
Detection Probability
20
15
10
5
0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters

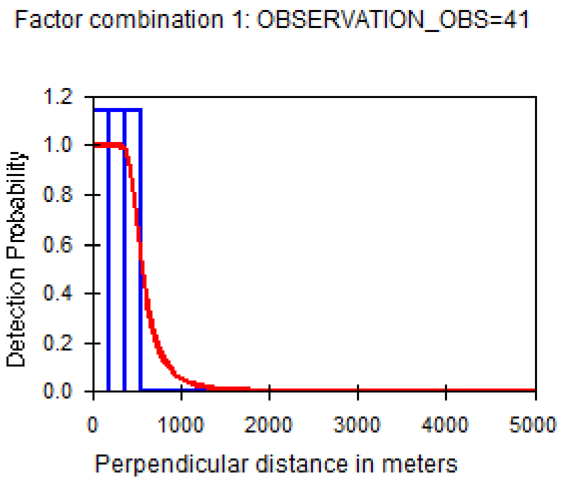

Factor combination 0: OBSERVATION_OBS=59
Detection Probability
Perpendicular distance in meters

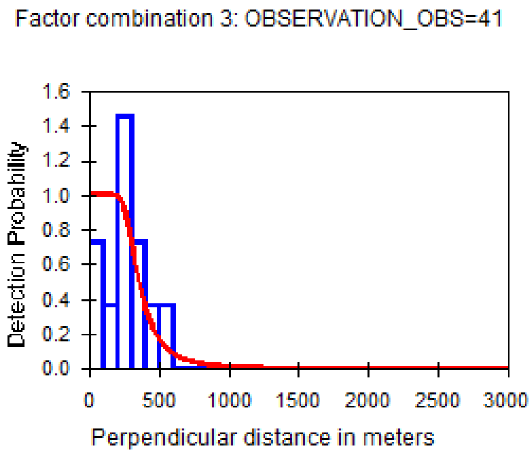

Factor combination 4: OBSERVATION_OBS=59
Detection Probability
Perpendicular distance in meters

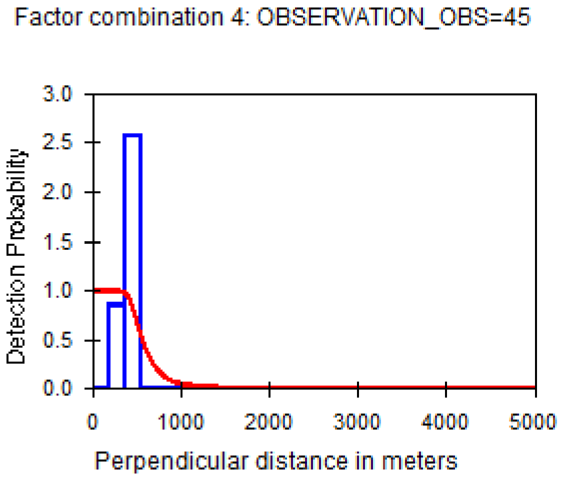

Factor combination 3: OBSERVATION_OBS=63
Detection Probability
Perpendicular distance in meters

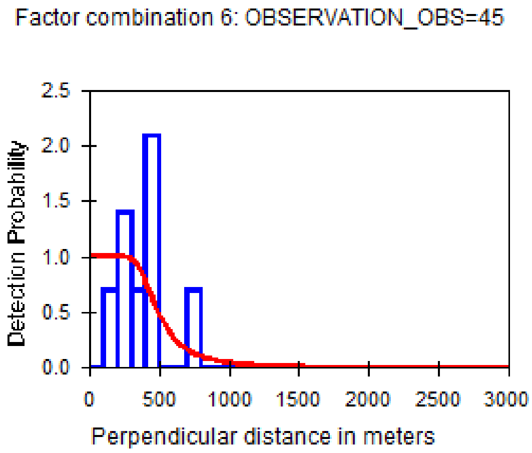

Factor combination 8: OBSERVATION_OBS=63
Detection Probability
Perpendicular distance in meters

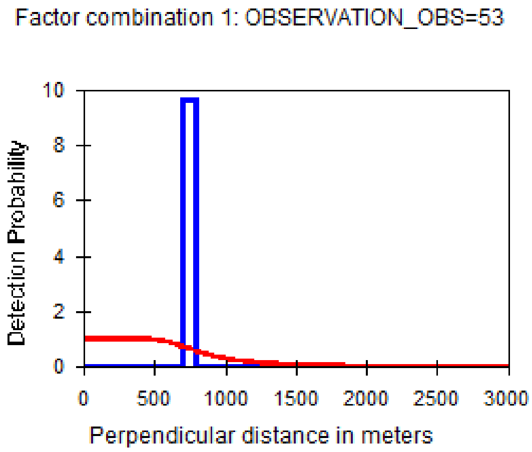
ActionAir

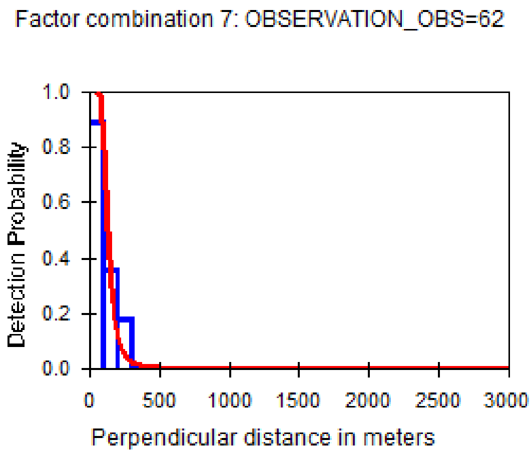

Factor combination 1: OBSERVATION_OBS=41
Detection Probability
Perpendicular distance in meters


Factor combination 3: OBSERVATION_OBS=41
Detection Probability
Perpendicular distance in meters


Factor combination 4: OBSERVATION_OBS=45
Detection Probability
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
1000
2000
3000
4000
5000
Perpendicular distance in meters


Factor combination 6: OBSERVATION_OBS=45
Detection Probability
2.5
2.0
1.5
1.0
0.5
0.0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters


Factor combination 1: OBSERVATION_OBS=53
Detection Probability
10
8
6
4
2
0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters


Factor combination 7: OBSERVATION_OBS=62
Detection Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
3000
Perpendicular distance in meters

ANNEX II

Table 1. Chi-square test for Beaufort. '+/-' = small positive/negative effect, '+ +/- -' = medium positive/negative effect, '+ + + / - - -' big positive/negative effect.

			All species					BFT				
Levels	**Effort (km)**	**% Effort**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**
0	29342	25	282	18	398	34	- -	86	30	72	3	+
1	43971	38	549	35	596	4		102	36	107	0	
2	26935	23	556	35	365	100	+ + +	73	26	66	1	+
3	12166	10	162	10	165	0		22	8	30	2	- -
4	3578	3	23	1	48	13	- - -	0	30	9	9	- - -
Total	**115992**	**100**	**1572**	**100**	**1572**	**151**		**283**	**100**	**283**	**15**	
Df						**4**					**4**	
P						**<<0.001**					**<0.005**	

Table 2. Chi-square test for Glare. '+/-' = small positive/negative effect, '+ +/- -' = medium positive/negative effect, '+ + + / - - -' big positive/negative effect.

			All species					BFT				
Levels	**Effort (km)**	**% Effort**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**
0	15583	13	143	9	211	22	- -	21	7	38	8	- -
1	31855	27	346	22	431	17	-	81	29	78	0	
2	27780	24	685	44	376	254	+ + +	94	33	68	10	+ +
3	40904	35	398	25	554	44	- -	87	31	100	2	-
Total	**116122**	**100**	**1572**	**100**	**1572**	**336**		**283**	**100**	**283**	**20**	**-**
Df						**3**					**3**	
P						**<<0.001**					**<0.001**	

Table 3. Chi-square test for Turbidity. '+/-' = small positive/negative effect, '+ +/- -' = medium positive/negative effect, '+ + + / - - -' big positive/negative effect.

			All species					BFT				
Levels	**Effort (km)**	**% Effort**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**
0	102756	88	1500	95	1391	9		243	86	250	0	
1	11868	10	69	4	161	52	- - -	38	13	29	3	+ +
2	1217	1	2	0	16	13	- - -	1	0	3	1	- - -
3	281	0	1	0	4	2	- - -	1	0	1	0	
Total	**116122**	**100**	**1572**	**100**	**1572**	**76**		**283**	**100**	**283**	**20**	
Df						**3**					**3**	
P						**<<0.001**					**>0.1**	

Table 4. Chi-square test for Haze. '+/-' = small positive/negative effect, '+ +/- -' = medium positive/negative effect, '+ + + / - - -' big positive/negative effect.

			All species						BFT				
Levels	**Effort (km)**	**% Effort**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**
0	49433	43	743	47	669	8	+	161	57	120	14	+ +
1	48351	42	664	42	655	0		92	33	118	6	-
2	12489	11	143	9	169	4	-	20	7	30	4	- -
3	1947	2	7	0	26	14	- - -	3	1	5	1	- -
4	3902	3	15	1	53	27	- - -	7	2	10	1	- -
Total	**116122**	**100**	**1572**	**100**	**1572**	**54**		**283**	**100**	**283**	**24**	
Df						**4**					**4**	
P						**<<0.001**					**<0.001**	

Table 5. Chi-square test for Clouds. '+/-' = small positive/negative effect, '+ +/- -' = medium positive/negative effect, '+ + + / - - -' big positive/negative effect.

			All species						BFT				
Levels	**Effort (km)**	**% Effort**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**
0-2	18212	67	389	74	353	4	+	27	69	26	0	
3-5	7481	28	105	20	145	11	- -	12	31	11	0	
6-8	1408	5	31	6	27	1	+	0	0	2	2	- - -
Total	**27100**	**100**	**525**	**100**	**525**	**15**		**39**	**100**	**39**	**2**	
Df						**2**					**2**	
P						**<0.001**					**>0.1**	

Table 6. Chi-square test for Subjective. 'G'= Good; 'M'= Moderate; 'P'= Poor. '+/-' = small positive/negative effect, '+ +/- -' = medium positive/negative effect, '+ + + / - - -' big positive/negative effect.

			All species						BFT				
Levels	**Effort (km)**	**% Effort**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**
G	18212	67	414	79	340	16	+	27	69	25	0	
M	7481	28	95	18	154	23	- -	9	23	11	1	
P	1408	5	16	3	31	7	- -	3	8	2	0	
Total	**27100**	**100**	**525**	**100**	**525**	**46**		**39**	**100**	**39**	**1**	
Df						**2**					**2**	
P						**<<0.001**					**>0.1**	

Table 7. Chi-square test for Airplane. 'C'= Cessna; 'P'= Partenavia. '+/-' = small positive/negative effect, '+ +/- -' = medium positive/negative effect, '+ + + / - - -' big positive/negative effect.

			All species						BFT				
Levels	**Effort (km)**	**% Effort**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**
C	48268	42	874	56	653	74	+ +	166	59	118	19	+ +
P	67854	58	698	44	919	53	-	117	41	165	14	- -
Total	**116122**	**100**	**1572**	**100**	**1572**	**127**		**283**	**100**	**283**	**34**	
Df						**1**					**1**	
P						**<<0.001**					**<<0.001**	

Table 8. Chi-square test for Team. '+/-' = small positive/negative effect, '+ +/- -' = medium positive/negative effect, '+ + + / - - -' big positive/negative effect.

			All species					BFT				
Levels	**Effort (km)**	**% Effort**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**	**N Observ.**	**% Observ.**	**Expected**	**Chi-square**	**Effect**
Air-Med	14683	13	530	34	199	552	+ + +	76	27	36	45	+ + +
Unimar	33521	29	167	11	454	181	- - -	41	14	82	20	- -
Perigod	33950	29	637	41	460	68	+ +	146	52	83	48	+ + +
Action-Air	33968	29	238	15	460	107	- -	20	7	83	48	- - -
Total	**116122**	**100**	**1572**	**100**	**1572**	**909**		**283**	**100**	**283**	**161**	
Df						**3**					**3**	
P						**<<0.001**					**<<0.001**	

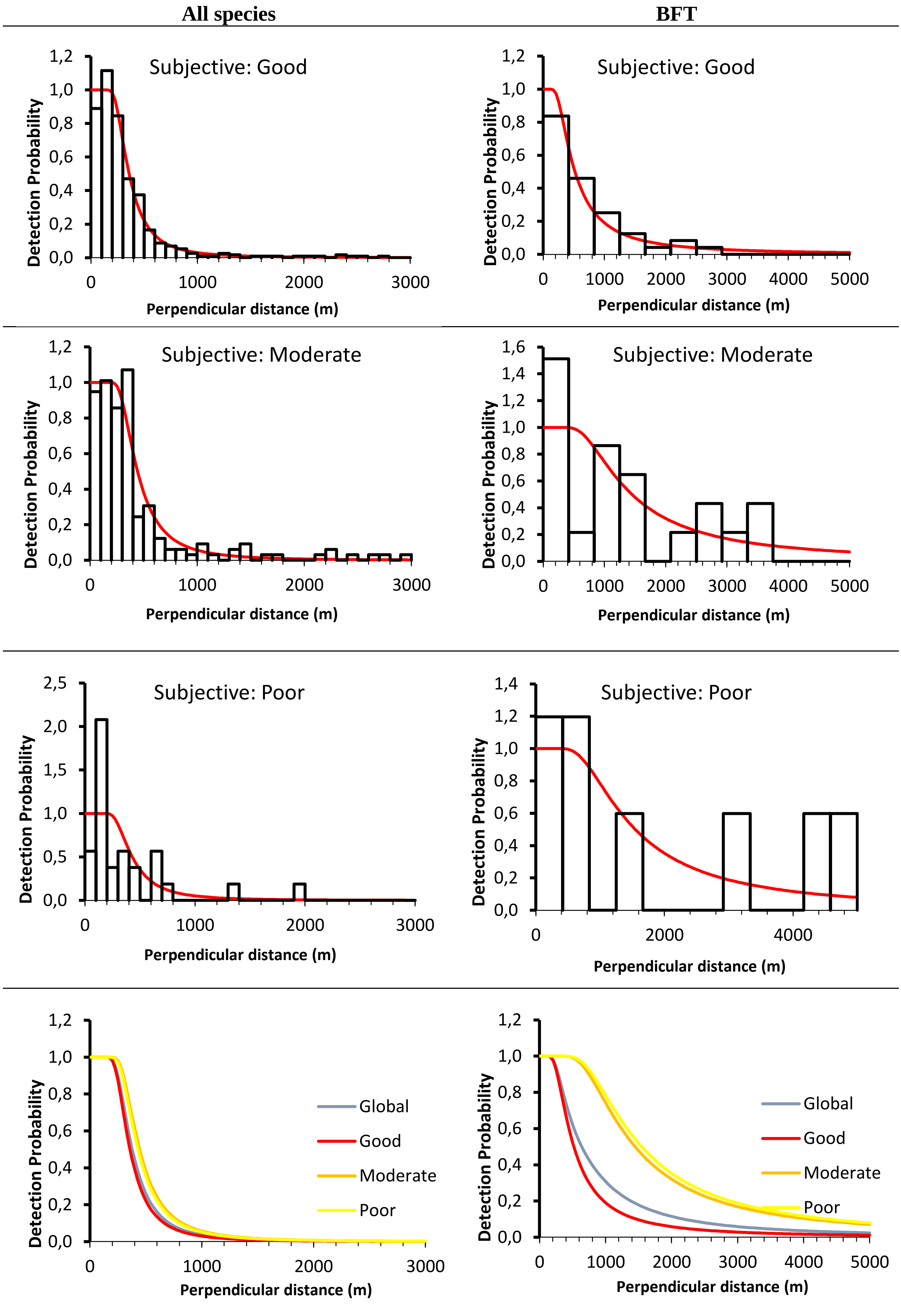


Figure 1. Effect of the factor "Subjective", with three levels key to describe 'good', 'moderate' or 'poor' searching conditions, as defined by the observers.

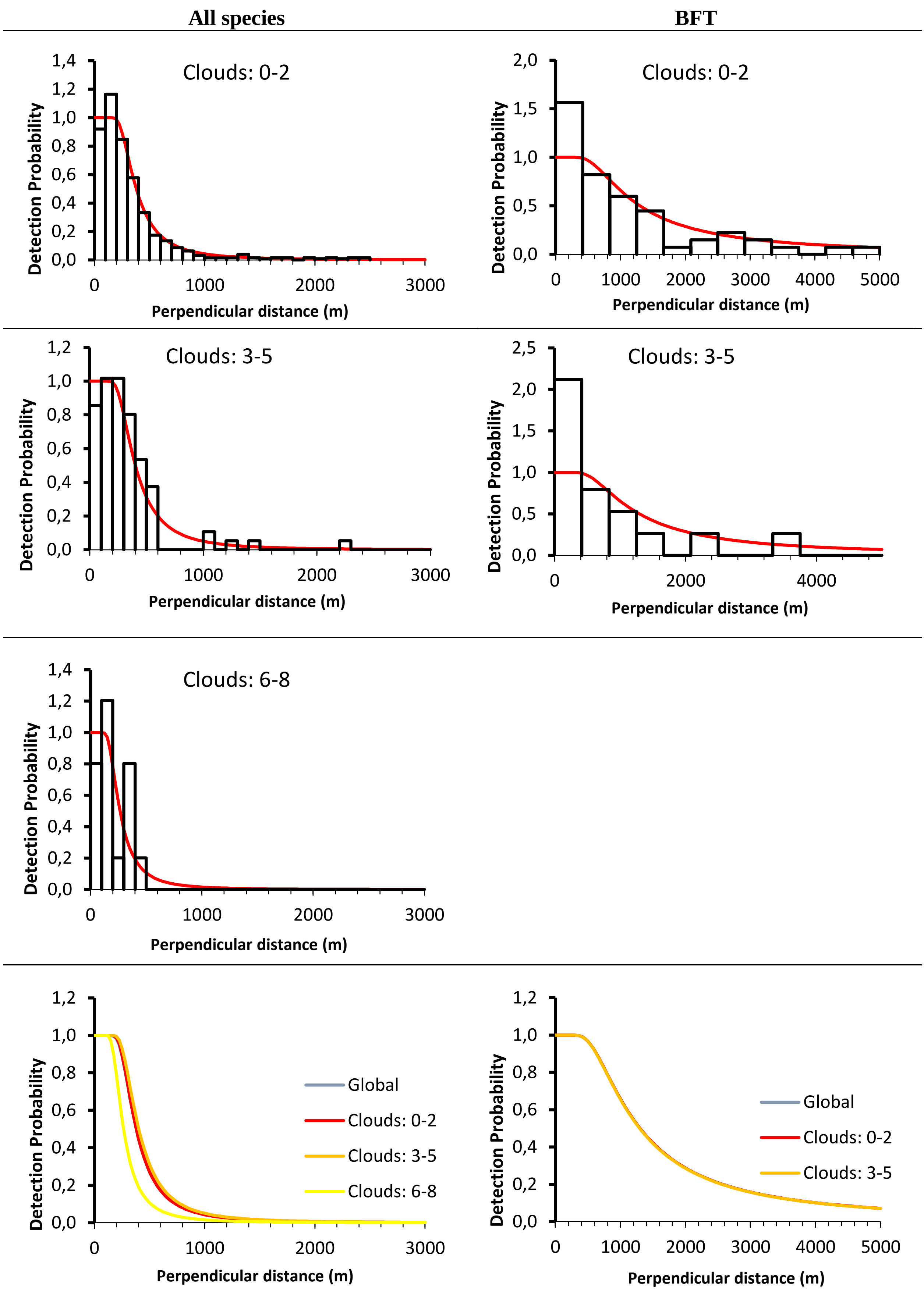


Figure 2. Effect of the factor "Clouds", with three levels: 0-2 (clear or slightly cloudy), 3-5 (moderately cloudy) and 6-8 (mostly cloudy).

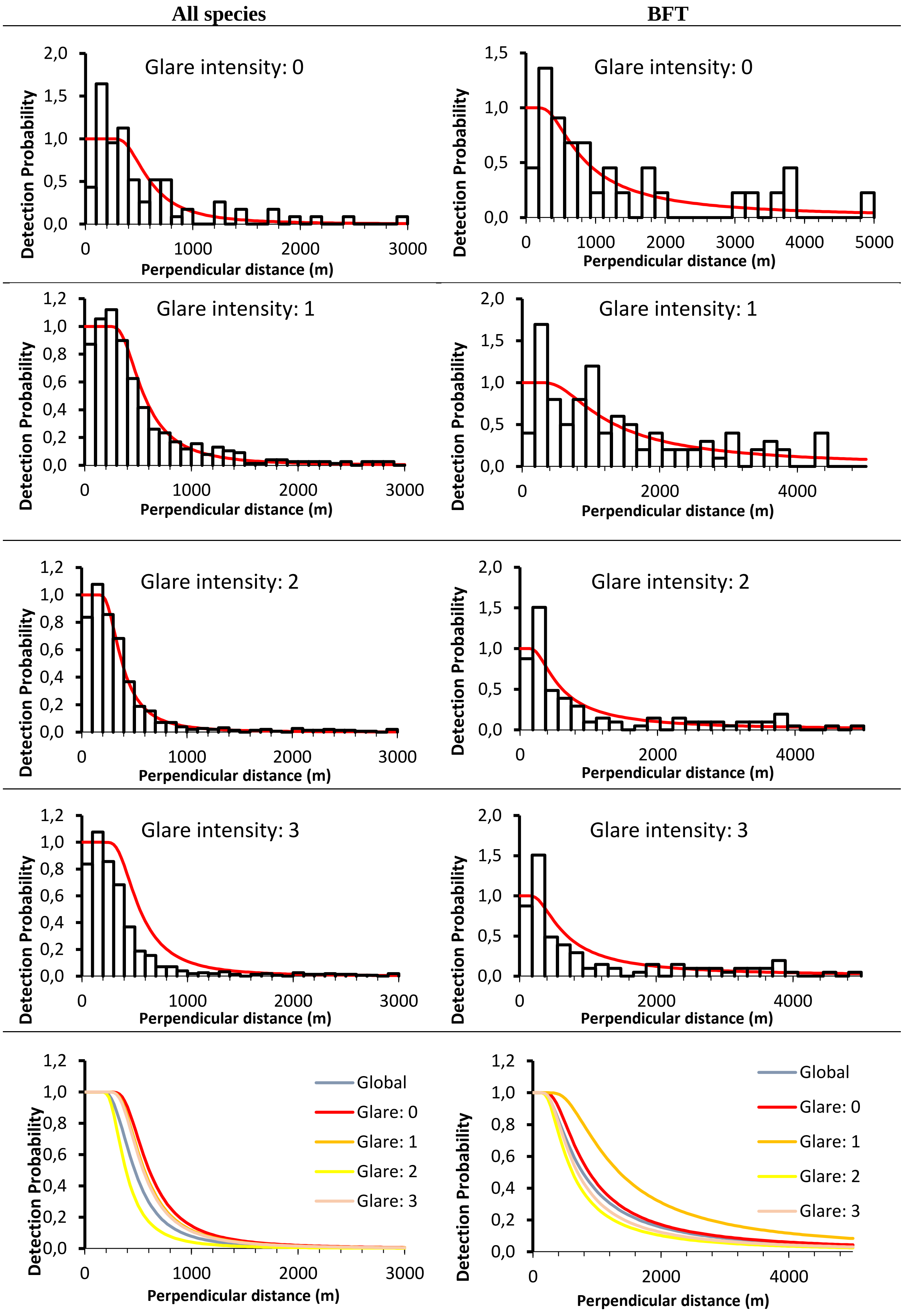
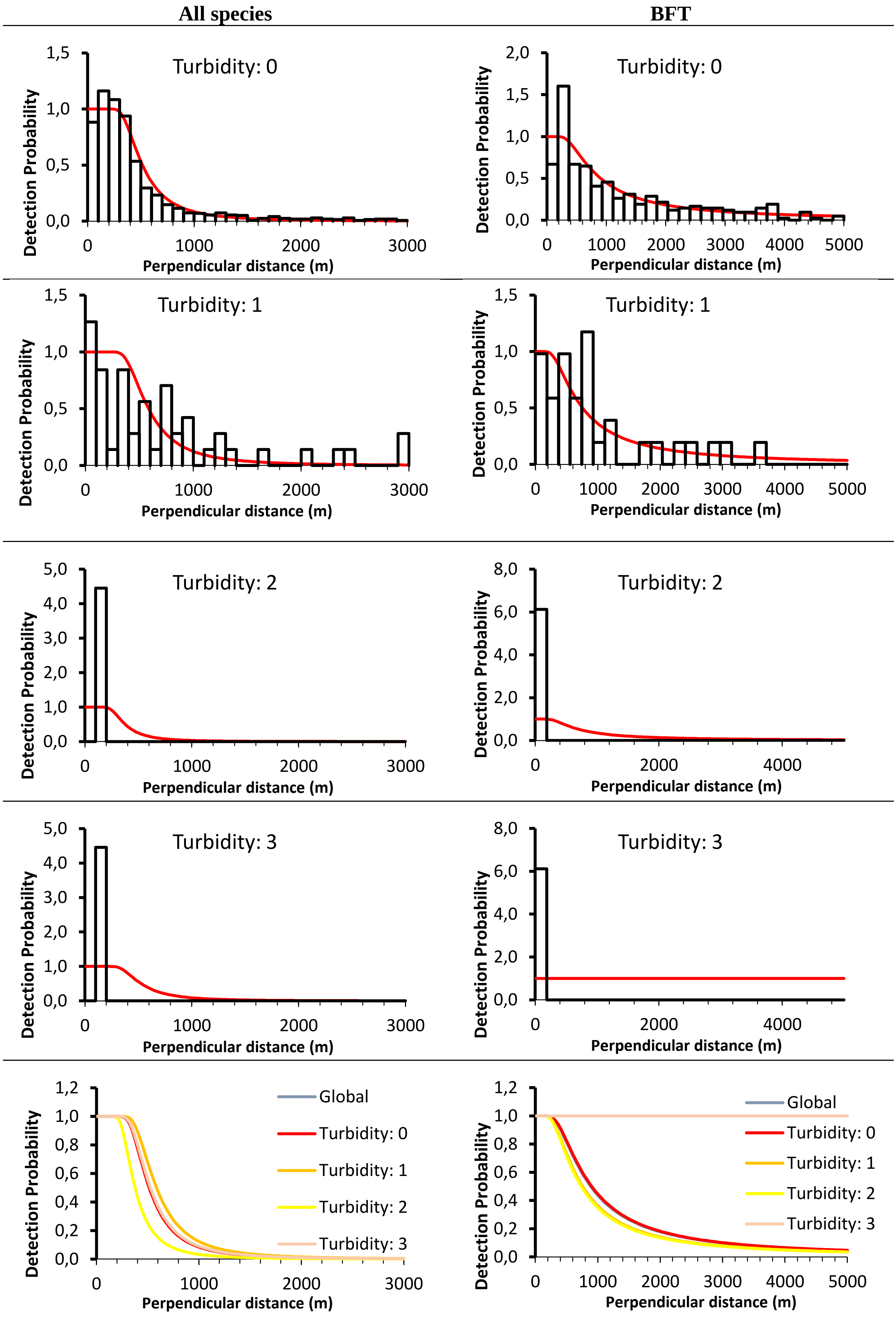


Figure 3. Effect of the factor “Glare”, a four levels from null (0) to intense (3).



All species
BFT
Turbidity: 0
Turbidity: 1
Turbidity: 2
Turbidity: 3
Detection Probability
Perpendicular distance (m)
Global
Turbidity: 0
Turbidity: 1
Turbidity: 2
Turbidity: 3

Figure 4. Effect of the factor "Turbidity", a four levels from null (0) to intense (3).

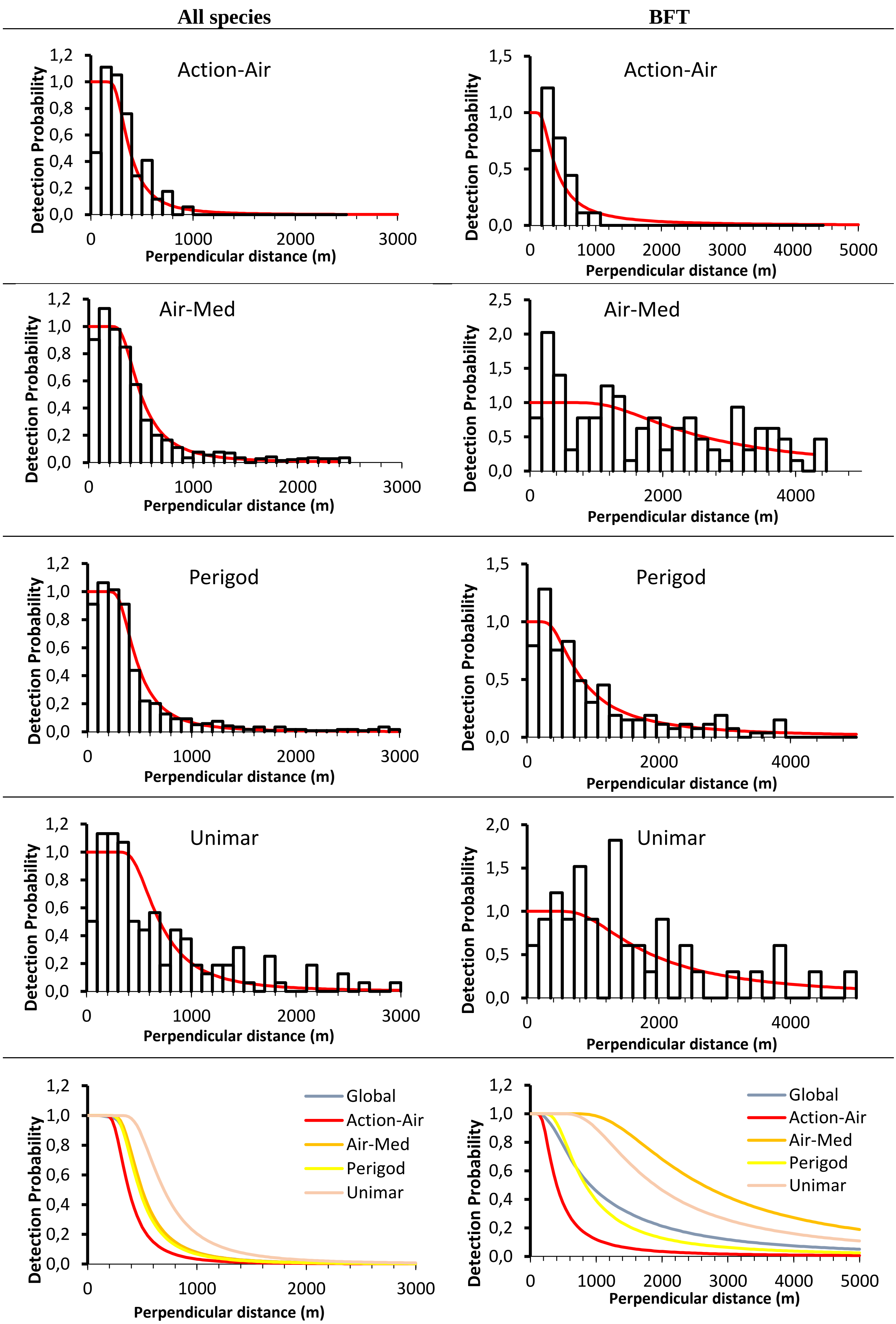


Figure 5. Effect of the factor “Team” with a four levels.

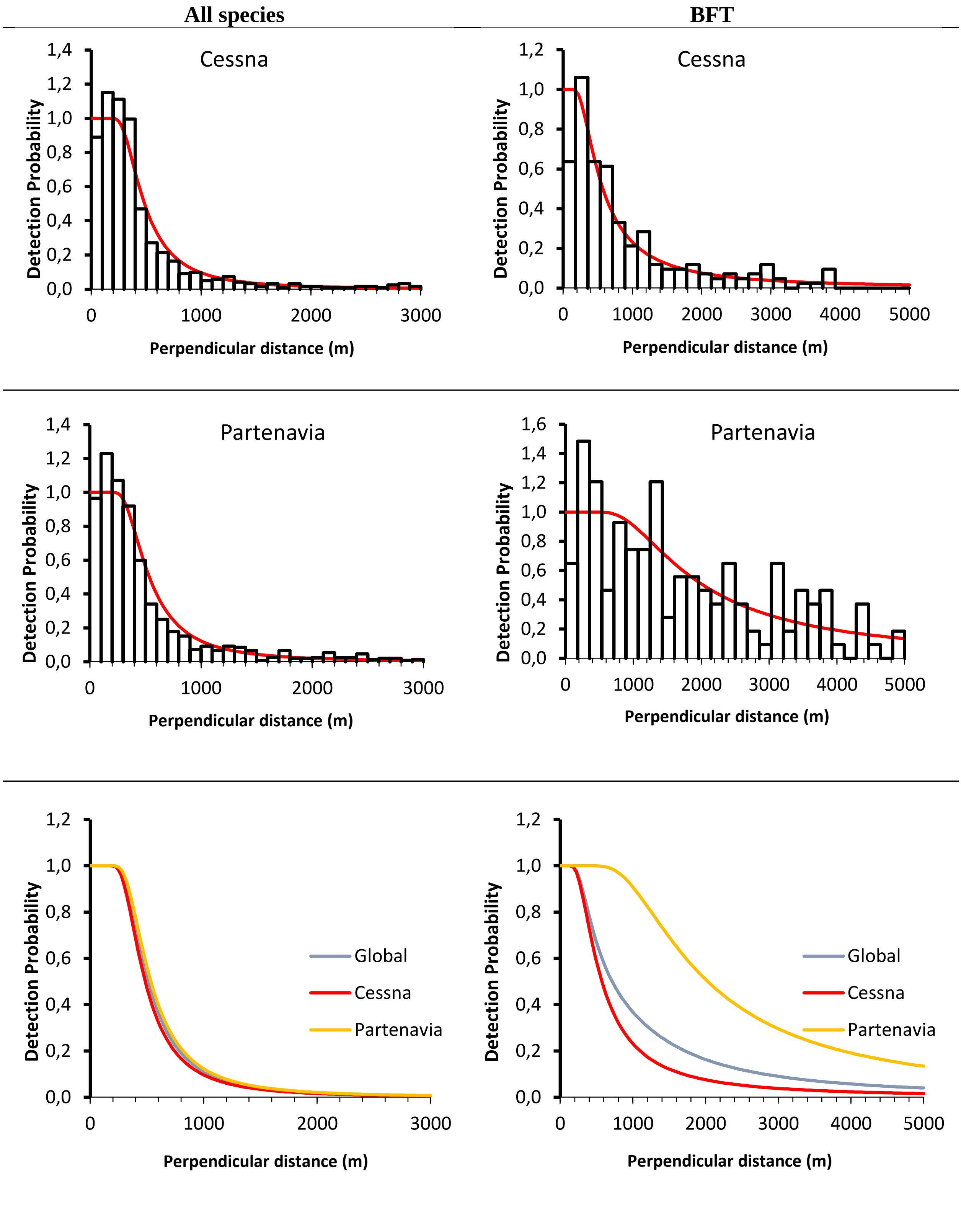
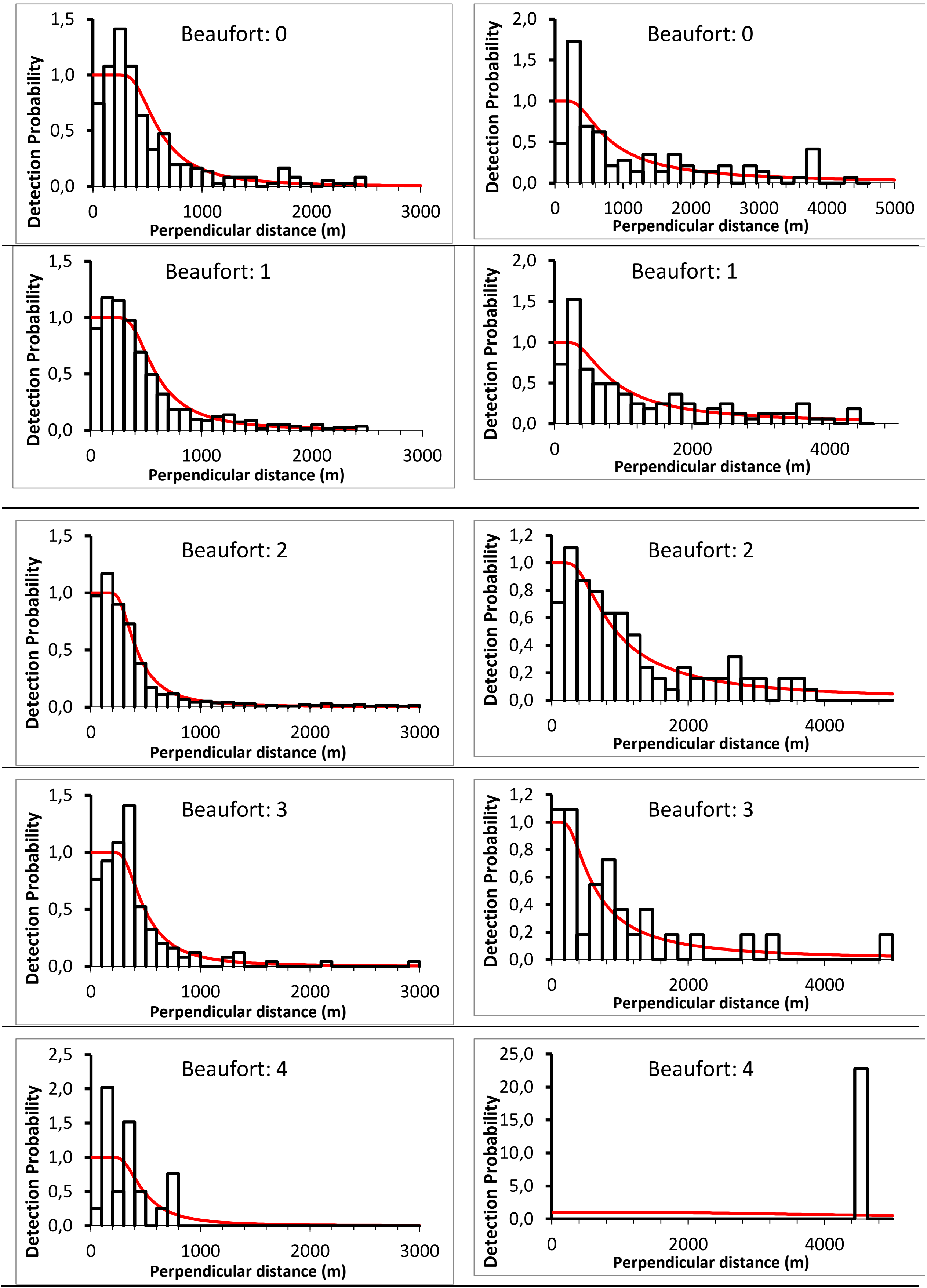


Figure 6. Effect of the factor “Airplane”, with two levels

All species **BFT**



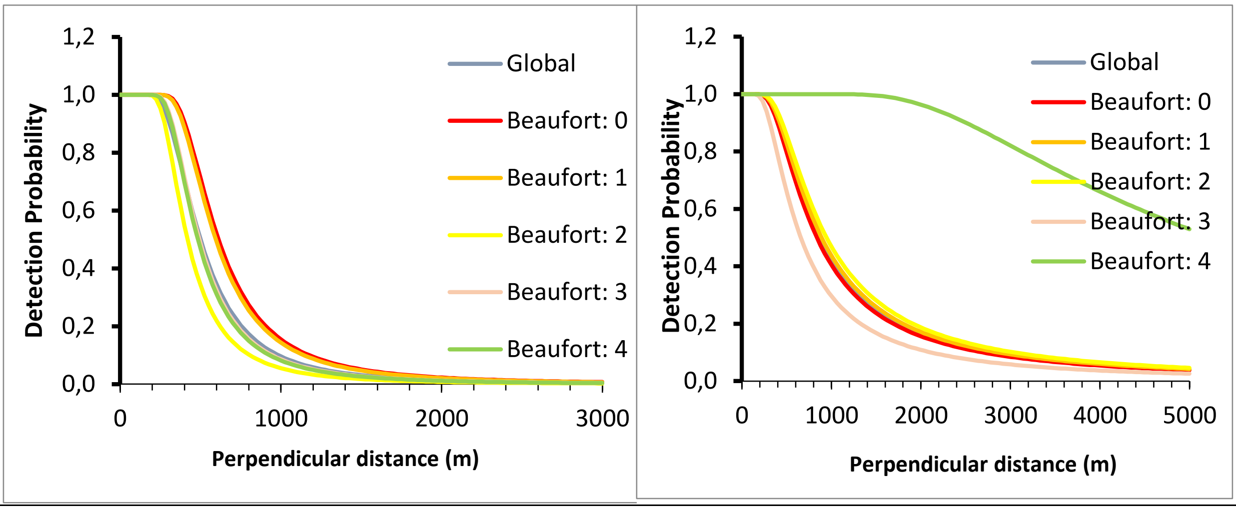
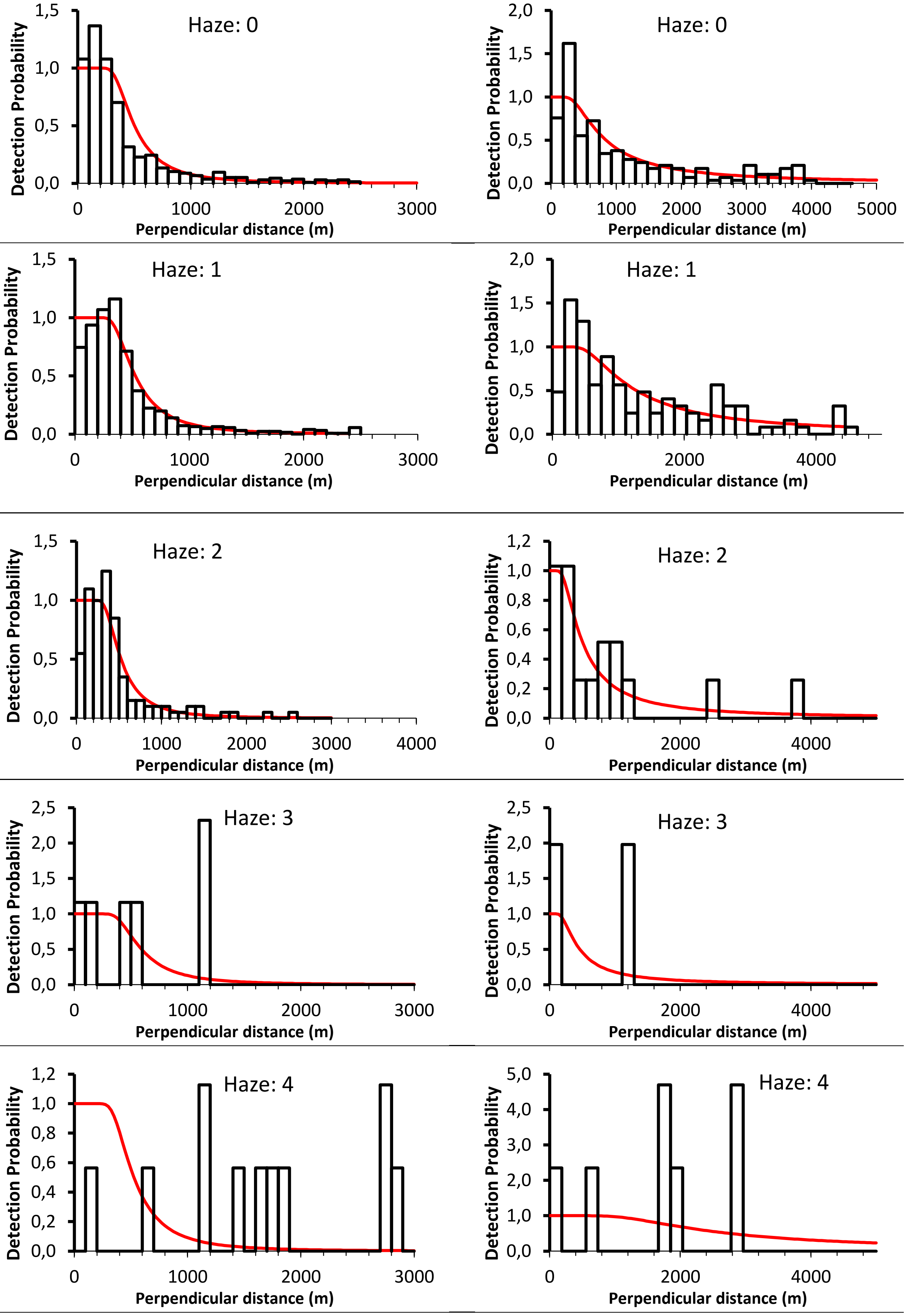


Figure 7. Effect of the factor "Beaufort", a five levels from calm (0) to medium-heavy (4).

All species | **BFT**



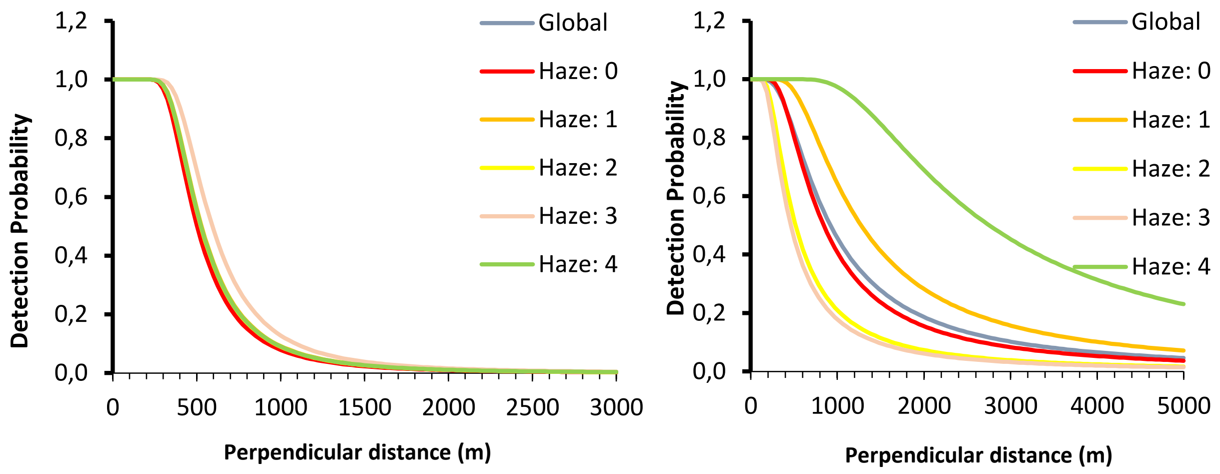


Figure 8. Effect of the factor "Haze", a five levels from null (0) to intense (4).