Supplementary Material to:
"All Data are Wrong, but Some are Useful? Advocating the Need for Data Auditing"

GEO-REFERENCED TUNA CATCH DATA FROM GULF OF GUINEA

The total number of catch reports from 1992 to 2012 is 28,406 with the total catch for skipjack, yellowfin, and bigeye being 13,735, 11,702 and 2,969 metric tons, respectively. Circle plots of catch weight are available in Figures 1–4 below.

In the Gulf of Guinea, purse seine and baitboat are the main tuna fishing gears and these normally target yellowfin and skipjack, while bigeye is only by-catch thus resulting in a relatively low catch (Miyake, Miyabe, and Nakano, 2004). This is in line with what we observe in Table 1—where we give an account of the number of catch reports per vessel type—and in Table 2—where we provide the reported catch weight per vessel type.

Table 1: Number of catch reports per vessel type.

	Vessel type		
Species	Purse seine	Baitboat	Total
Skipjack	2,500	11,235	13,735
Yellowfin	2,151	9,551	11,702
Bigeye	2,472	497	2,969
Pooled	7,123	21,283	28,406

Species	Vessel type		Total
	Purse seine	Baitboat	
Skipjack	120,986.5	39,228.0	160,214.5
Yellowfin	67,233.9	16,050.6	83,284.5
Bigeye	6,887.3	3,657.8	10,545.1
Pooled	195,107.8	58,936.3	254,044.1

Table 2: Reported catch weight (in metric tons) per vessel type.

The geo-reference nature of our data is displayed in Figure 1, where we plot the locations of tuna catch per species, along with the demarcation of the relevant FAO fishing divisions, which consist of a geographic nomenclature by the Food and Agriculture Organization (FAO) of the United Nations (*www.fao.org*). The Gulf of Guinea falls covers a total of 14.2 million km^2 (Food and Agriculture Organization, 2011). The circle plots in Figure 2 complete the picture by also providing information on the size of reported catch.

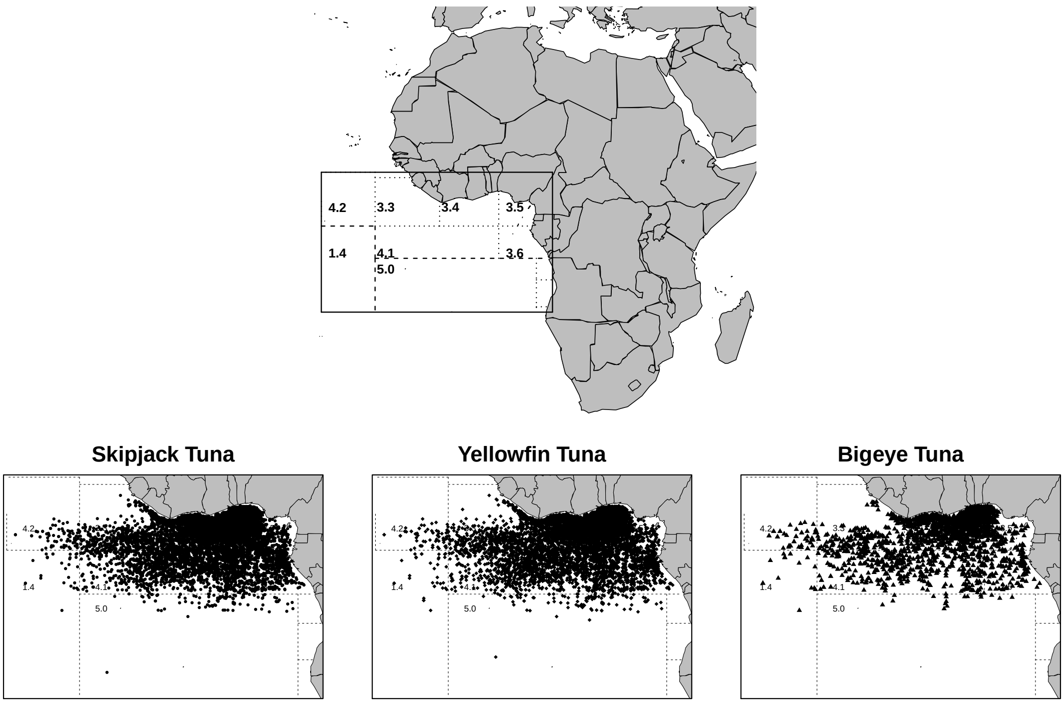

Figure 1: Map of locations of geo-referenced catch data with the demarcation of divisions of relevant FAO fishing areas. The locations of the catch for each species cuts across three fishing areas: Area 47 (Division 5), Area 41 (Division 1.4) and Area 34 (Divisions 3.3, 3.4, 3.5, 3.6, 4.1, 4.2), with the majority of catches being performed in the latter.

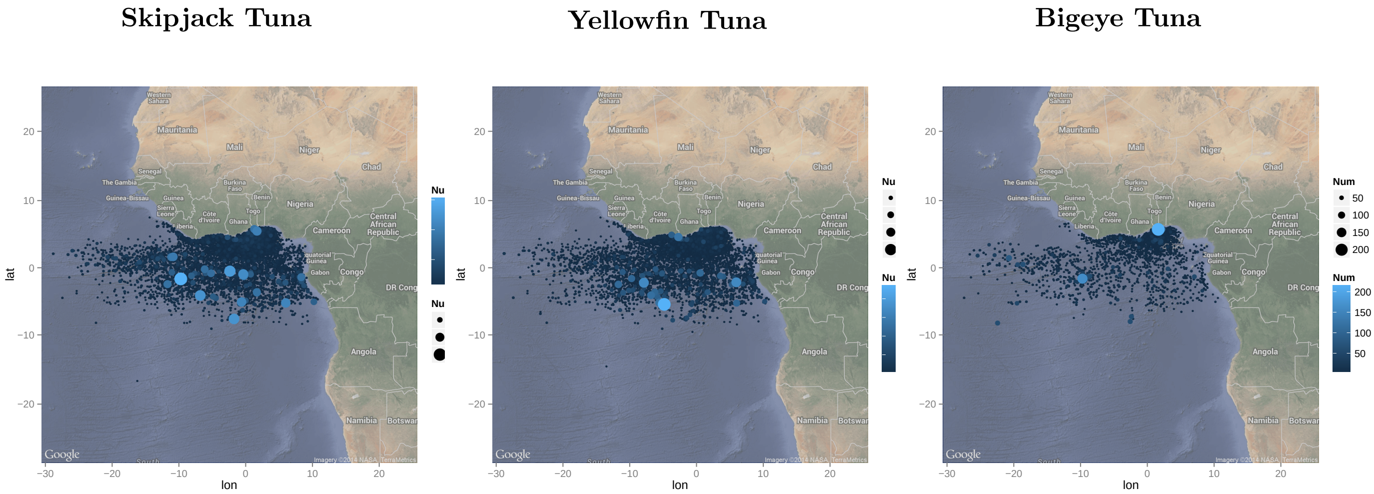

Figure 2: Circle plot type of representations with catch locations and catch weights. The circles are plotted with centres at the locations at which catch was reported, and the radii reflects size of reported the catch.

For each of the species, a further analysis carried out on each FAO fishing area depicts a similar story. Some comments regarding the motivation for this spatial analysis are in order. The fact that control and surveillance may be less common in certain regions, may provide a higher incentive to misreport in those regions. Though the locations of the catch were in eight different fishing areas, analysis was carried out on only five (divisions 3.3, 3.4, 3.5, 3.6, 4.1 all of fishing area 34). This was because, the amount of data in each of the other three were small for one to be sufficiently confident that the asymptotics underlying Pearson's test would apply. For skipjack, we reject the null hypothesis in FAO fishing divisions 3.3, 3.4, and 4.1, which accounts for a total of 83% of the 10,824 number of reports under analysis on this section for skipjack. In the neighboring fishing divisions 3.5 and 3.6, we cannot however reject the null hypothesis for skipjack. In the cases of bigeye and yellowfin tuna the null hypothesis is rejected on all five divisions. Pooling all skipjack, yellowfin and bigeye data together and carrying out the analysis by fishing areas basically leads to similar conclusions. In Figure 3 we plot the Pearson residuals corresponding to this analysis.

In Table 3 we present the number of catches per FAO fishing division of each species. As it can be observed division 3.4 contains about 62% of the total number of reports, while divisions 1.4, 4.2 and 5, only account for a total of 1.7% of the number of reports; see also Figure 1. In Table 4 we present the reported catch weight per FAO fishing division of each specie. Division 3.4 represents 46% of total catch weight.

Table 3: Number of catch reports per FAO fishing division.

	FAO fishing division								
Species	1.4	3.3	3.4	3.5	3.6	4.1	4.2	5	Total
Skipjack	29	795	8,516	1,291	1,017	1,901	90	96	13,735
Yellowfin	35	689	7,159	1,078	786	1,781	72	102	11,702
Bigeye	17	216	1,832	232	148	459	49	16	2,969
Pooled	81	1,700	17,507	2,601	1,951	4,141	211	214	28,406

Table 4: Reported catch weight (in metric tons) per FAO fishing division.

	FAO fishing division							
Species	1.4	3.3	3.4	3.5	3.6	4.1	4.2	5
Skipjack	386.5	15,381.8	75,846.0	10,689.1	19,177.4	34,289.9	2,459.9	1,983.9
Yellowfin	462.3	5,404.9	37,069.3	5,691.8	8,745.7	23,301.7	971.3	1,637.5
Bigeye	140.5	1,107.3	4,596.3	408.4	589.1	3,001.2	548.8	153.5
Pooled	989.3	21,894.0	117,511.6	16,789.3	28,512.2	60,592.8	3,980.0	3,774.9

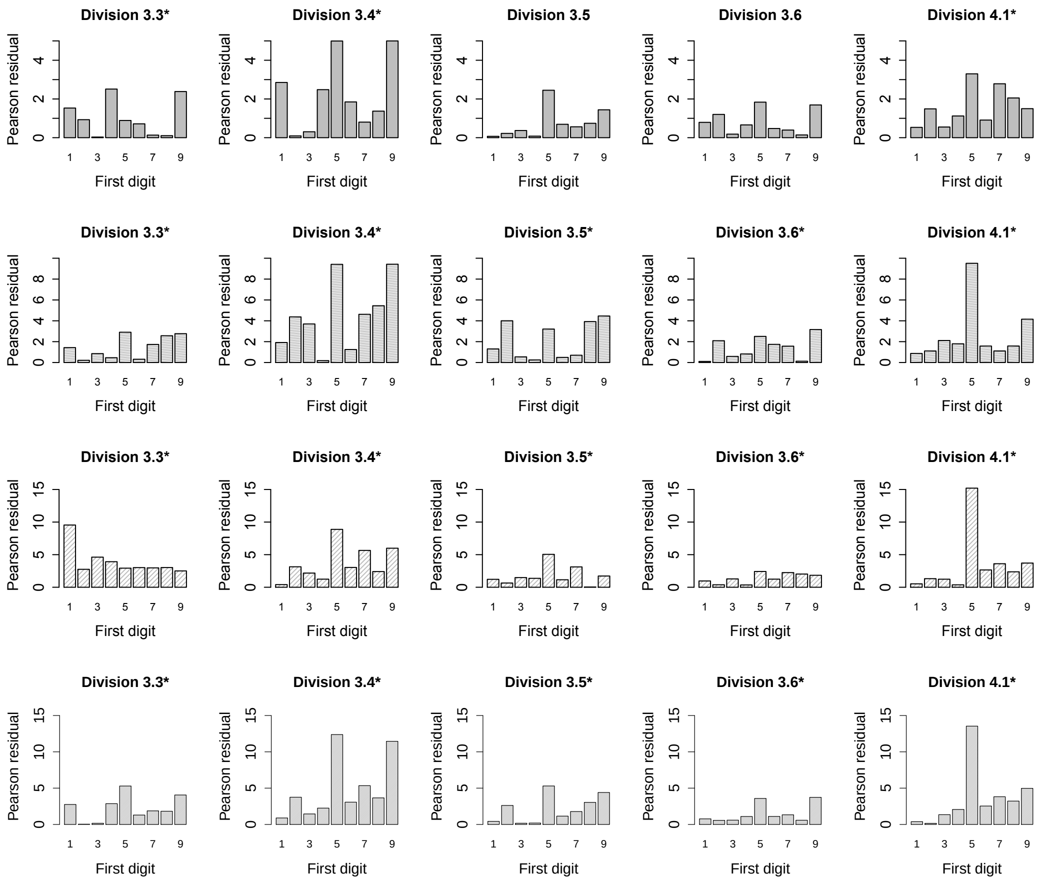

Figure 3: Benford's law-based spatial analysis over FAO fishing areas. Pearson residuals of per FAO fishing area; the asterisks ($\star$) in the titles of the graphs represent rejections at 5% significance level.

SUPPLEMENTARY EMPIRICAL REPORTS

Table 5: Benford law-based analysis by species and by pooled species. Pearson chi-square (X^2), Kolmogorov–Smirnov ($\mathbb{D}$), Euclidean distance (ρ), and Chebyshev distance (m) statistics per species and for pooled species.

Species	X^2	$\mathbb{D}$	ρ	m
Skipjack	118.3131*	2.7654*	3.1865*	1.9663*
Yellowfin	447.8661*	5.9170*	5.7353*	3.8033*
Bigeye	494.5211*	4.6922*	6.0503*	5.1004*
Pooled	722.0985*	7.2376*	7.0953*	5.4573*

NOTE: ** and * represent rejections at 10 and 5% significance levels, respectively.

Table 6: Benford law-based analysis by species and by pooled species. Pearson chi-square (X^2), Kolmogorov–Smirnov ($\mathbb{D}$), Euclidean distance (ρ), and Chebyshev distance (m) statistics per vessel type

Vessel	Species	X^2	$\mathbb{D}$	ρ	m
Purse seine	Skipjack	38.4980⋆	1.0124⋆	1.7674⋆	1.3809⋆
	Yellowfin	251.1814⋆	3.4325⋆	4.3889⋆	3.7880⋆
	Bigeye	716.5923⋆	3.6508⋆	7.6225⋆	6.9817⋆
	Pooled	485.7883⋆	3.9593⋆	6.0333⋆	2.0700⋆
Baitboat	Skipjack	88.9205⋆	2.5800⋆	2.8236⋆	1.6032⋆
	Yellowfin	287.7371⋆	4.9205⋆	4.6339⋆	2.4122⋆
	Bigeye	200.7758⋆	3.6487⋆	3.8601⋆	2.4591⋆
	Pooled	423.4504⋆	6.1359⋆	5.5111⋆	5.5111⋆

NOTE: ⋆⋆ and ⋆ represent rejections at 10 and 5% significance levels, respectively.

Table 7: Benford law-based analysis by species and by pooled species. Pearson chi-square (X^2), Kolmogorov–Smirnov ($\mathbb{D}$), Euclidean distance (ρ), and Chebyshev distance (m) statistics per species and per FAO fishing division.

FAO Area	Species	X^2	$\mathbb{D}$	ρ	m
3.3	Skipjack	16.5297*	1.2303*	1.3519*	0.8399
	Yellowfin	28.8017*	1.6707*	1.5067*	0.8170
	Bigeye	91.6407*	1.4914*	2.9182*	2.2383*
	Pooled	68.7483*	2.0650*	2.5558*	1.5098*
3.4	Skipjack	70.2656*	2.0509*	2.5573*	1.5651*
	Yellowfin	266.5399*	4.6828*	4.4868*	2.6491*
	Bigeye	177.9344*	2.8837*	3.6296*	2.4985*
	Pooled	357.6326*	5.3578*	4.9563*	2.9323*
3.5	Skipjack	9.6249	0.7906	0.8227	0.6896
	Yellowfin	64.4154*	2.1367*	2.4366*	1.6804*
	Bigeye	45.4831*	1.4095*	1.9447*	1.4201*
	Pooled	68.2540*	2.3535*	2.2630*	1.4912*
3.6	Skipjack	9.1704	0.4344	0.9951	0.5165
	Yellowfin	27.1739*	1.4757*	1.4770*	0.8772**
	Bigeye	22.7483*	1.7125*	1.3215**	0.6807
	Pooled	32.5525*	1.5356*	1.5000*	1.0079*
4.1	Skipjack	29.9949*	1.6911*	1.5236*	0.9283**
	Yellowfin	123.6528*	1.9220*	3.1041*	2.6771*
	Bigeye	274.2721*	2.8859*	4.5868*	4.2781*
	Pooled	245.5649*	3.3671*	4.2583*	43.8090*

NOTE: ** and * represent rejections at 10 and 5% significance levels, respectively.

REFERENCES

Food and Agriculture Organization (2011), "Review of the State of World Marine Fisheries Resources," *FAO Fisheries and Aquaculture Technical Paper*, 569.

Miyake, M. P., Miyabe N., and Nakano, H. (2004), "Historical Trends of Tuna Catches in the World," *FAO Fisheries and Aquaculture Technical Paper*, 467.