**Allocation comparison: actual sites surveyed vs allocation targets**

**Introduction**

Site allocation targets from past cruises where coral reef fish were surveyed were either compiled from existing spreadsheet files or calculated from a combination of variance of past data at the trophic level and the area of each strata (hard bottom reef area for each island, sector, reef zone, and depth bin).

**Methods**

1. Locate and compile existing spreadsheets from past cruise planning
   1. For these cruises, Ivor calculated 3 allocation percentages for each sector: 1) the weighted percentage based on a combination of variance from past surveys and area of each strata (variance\*area = AREA\_VAR\_PCT), 2) the hard bottom reef area percentage (AREA\_PCT), and 3) a manually adjusted percentage based on IVOR’S BRAIN and I don’t know what the logic was (MANUAL\_PCT). These areas are:
      * 2019 MHI
      * 2018 Samoa
      * 2018 PRIAs
      * 2017 Marianas
      * 2016 MHI
      * 2016 NWHI: Kure, Lisianski, ONLY FOREREEF for Pearl & Hermes and French Frigate
      * 2016 Samoa
      * 2016 PRIAs
      * 2015 PRIAs: Jarvis, Palmyra, Baker, Howland
   2. Cleaned up spreadsheets to keep the 3 percentage values used to calculate allocation based on number of sites we think we can visit per island.
2. Back-calculate allocation for areas/years that did not have a spreadsheet or for islands that had to be recalculated due to sections of islands not surveyed (for various reasons, i.e. weather: either sectors or reef zones)
   1. These areas are:
      * 2017 NWHI
      * 2016 NWHI: French Frigate and Pearl & Hermes
      * 2015 NWHI
      * 2015 PRIAs: Kingman, Johnston – Kingman WEST sites are changed to protected slope, as that designation was added later. Johnston protected slope sites were allocated as lagoon
      * 2015 MHI
      * 2015 Samoa
   2. 2 of the 3 allocation percentages were calculated in R - the missing piece for this is when Ivor would manually adjust. To calculate allocation:
      * Pull site level mean biomass estimates from the last 3 visits, include strata information: sector, reef zone, and depth bin (Can use site summary file from R scripts = TMPwsd.Rdata: this is generated from scripts 01\_gitFishREACleanData and 02\_gitFishREACalcWD and starts with file ALL\_REA\_FISH\_RAW.rdata – all files on GitHub-FishPaste).
      * Calculate standard deviation of mean biomass estimates at the higher trophic level (pisc, plank, prim, sec, tot fish) for each strata.
      * Calculate proportional SD for each trophic group at the strata level = SD value divided by the island level sum of all SD values for that trophic group.
      * Calculate proportional variance for each strata = average of proportional SD for all trophic groups at that strata.
      * Calculate proportional area = Get area of each strata (from fish-paste/data/Sectors-Strata-Areas.csv) = area for each strata divided by the total area for an island.
      * Multiply proportional variance by proportional area to get a value = mean\_v
      * Calculate strata proportion by summing mean\_v for each island and dividing each strata’s mean\_v value by the island sum. Value = AREA\_VAR\_PCT
      * Save as csv file with AREA\_VAR\_PCT value and AREA\_PCT for each strata for that year.

3. In R, load clean spreadsheets from past cruises and R files from allocations described above.

4. Data cleaning: Change values so files would match the clean data as it is now to reflect changes that were made along the way:

* 1. group Guam MP's into sector = GUA\_MP
  2. pool Rose lagoon and backreef depth bins to all
  3. for Tutuila in 2015, change analysis sectors- anything in Fagalua gets pooled in SW open, and Aunuu A gets pooled in SE open.
  4. Johnston sites in 2015 named protected slope now in master file- was originally lagoon when it was allocated, so change back to lagoon

5. Summarize number of sites surveyed each year at the sector level and add it to the data frame. Calculate allocations for each of the 3 percentage values, and add number of sites actually surveyed at each strata. Save summary file = allocation\_comparison\_2015\_2019.csv

**Summary file spreadsheet fields:**

* AREA\_PCT = percent used to calculate N\_AREA (by multiplying by SEC\_sites)
* AREA\_VAR\_PCT = percent used to calculate N\_AREA\_VAR
* MANUAL\_PCT = percent used to calculate N\_MANUAL – zero for areas where we didn’t have the number
* SEC\_sites = total number of sites that were surveyed in sector
* N\_AREA = number of sites we should have surveyed based solely on the proportion of reef area
* N\_AREA\_VAR = number of sites we should have surveyed based on a combination of the proportion of variance plus the proportion of reef area
* N\_MANUAL = Ivor’s magical manual adjustment to allocation
* N\_SITES = number of sites that were actually surveyed in that strata

**Uses**

1. Compare N\_SITES with any of the other N sites that seem appropriate (N\_AREA, N\_AREA\_VAR, N\_MANUAL) using file: **allocation\_comparison\_2015\_2019.csv**
2. Moving forward: we have scripts that can calculate allocation based on past variance, and we can use Ivor’s manual adjustment proportions for the areas where they exist – if we are surveying the same combination of sectors and reef zones at each island.

**Notes**

1. Folder structure:
   1. Metatdata: T:\Fish\Site allocation\allocation\_comparison\_2015\_2019\Documents
   2. Past data for variation and site numbers: T:\Fish\Site allocation\Data
   3. Original allocation spreadsheets: T:\Fish\Site allocation\allocation worksheets from past cruises
   4. Ivor’s manual adjustments (“fiddled”) and clean allocation spreadsheets: T:\Fish\Site allocation\allocation\_comparison\_2015\_2019\clean\_past\_allocation\_for\_R
   5. R scripts: T:\Fish\Site allocation\allocation\_comparison\_2015\_2019\R Scripts
      1. past\_allocation\_generation: produces files in T:\Fish\Site allocation\allocation\_comparison\_2015\_2019\Summary files
      2. allocation\_vs\_sites\_surveyed: produces summary comparison file
   6. Comparison file: T:\Fish\Site allocation\allocation\_comparison\_2015\_2019\allocation\_comparison\_2015\_2019.csv
2. NA values in final spreadsheet:
   1. MAI\_SE sector in 2019: hasn’t been surveyed before, so no variance to calculate