PNMS BRUV steps

Loaded data into R – for initial data checks and prep.

Outliers – two huge biomass outliers were detected:

1. Tiger shark in Sample PAL22\_059 – W2
2. Mola Mola in PAL22\_137 – S1

Using the Tukey Interquartile Range (IQR) method:

Calculate quartiles (Q1 and Q3) of biomass data by determining the IQR as Q3-Q1 and then establishing outlier threshold rules using 1.5 x IQR rule:

* Lower threshold: Q1 - (1.5 × IQR)
* Upper threshold: Q3 + (1.5 × IQR)

**Biomass\_kg:**

Using this approach there are 21 outliers.

Focusing on extreme outliers using multiplying with 3, there are 7 outliers.

**MaxN:**

Using this approach there are 28 outliers.

Focusing on extreme outliers using multiplying with 3, there are 23 outliers.

Many of these are ecological important and real observations though.

\*\*So – will probably not be using this approach to determine outliers.

**Final decision with regards to outliers:**

Considering that large schools of decapterus and mahi mahi were observed across multiple strings and are characteristic of the pelagic environment, as well as a few observations of large billfish – it was decided to keep all of these observations.

Only two outliers were removed:

The single observation of a tiger shark and mola mola.

All further analyses etc. were done with these two outliers removed.

**Sample aggregations**

We have 28 strings with 5 non-independent samples on each string.

So – we will focus on the following metrics:

**Abundance as MaxN:**

Species/group specific Max N per string (This uses the maximum count of each species across the samples within a string, recognizing that the 5 samples on a string are not independent observations). We have 2-3 strings per site – so we can get an average MaxN for a species/group in each site based on 2-3 data points.

**Biomass:**

Here we will follow a similar approach as with MaxN – considering the issue of double counting fish across samples from the same string.

So, we will use the maximum biomass for a species/group found across the 5 samples – and use as the biomass estimate for that species/group in that string. We then again have 2-3 biomass estimates per species/group for each site.

**Diversity**

This is straightforward enough – and will be the maximum number of different species/group across all samples in a string. We will then get diversity per string and can average it across 2-3 strings in each site.

**Observation summary**

Of the 28 strings deployed, fish were not observed in 3 strings:

* PAL22\_3 N1 PNMS North
* PAL22\_8 N4 PNMS North
* PAL22\_21 S3 PNMS south

Successful observations on 89% of deployments

**Plankton data**

Went back to the Plankton data – and it seems the following samples are missing:

String 19, 22, 25, 28

Used the Plankton RAW data to start again.

First – linked the plankton data with a BRUV string (common variable of String). Here each sample (which is Zone, Site and then rep combined) relates to a single string. In each site we deployed 3 strings and did 3 tows.

Saved as:

Then – checked the data and removed:

* Encrusting algae
* Insects
* Rubber
* Spirorbid shells
* Sponge chunks
* Sponge spicules
* Terrestrial flotsam
* Trash

Then also removed:

* Bryozoans
* 13
* Ciliates
* Foliculina
* Forams
* Hydroids
* Oblong eggs
* Orbs
* Pummus

There was a huge bunch that was unknown – kept these and just added in the term Unknown for Plankton type etc.