Methods for spatial clustering IPDB

etreml. Update on 3 June 2015 (/Spatial_3June2015/)

Technical details:

- Projection: World Mollweide (equal area) projection, Central Meridian at 155 East (centered between 20°E and 70°W).
- Extent: use full global extent to allow for spatial filtering downstream
- Use default origin of this projection for all fishnet generation
- Downloaded **ipdb.txt** on 3June2015 from GitHub.
- Extracted columns [IPDB_ID, Genus_species_locus, locality, decimalLatitude, decimalLongitude, country] to create **ipdb_sub.txt** (**35,897** records)
- Use <tab> delimited file format throughout as many special characters are in fields (: , _ -)
- Add 'GISid' as an integer unique identifier for spatial manipulation -> **ipdb_sub.shp**
- Projected ipdb_sub.shp to Moll-155 -> ipdb_sub_m155.shp (Fig 1)
- Spatial intersect with fn500_m155.shp (Fig 2), fn100_m155.shp (Fig 3), meow_ecos_m155 (Fig 4), World_EEZ_v8_2014_HR_m155 (Fig 5), and PellissierRichness_m155 (Fig 6) created in **Pellissier_DataWork.mxd** from **RichnessFish.csv** received from Peter Cowman on 13Apr2015 at NESCent .
- Exported to ipdb_sp.tsv (ipdb with spatial data) via .dbf

- **Note:** PellissierRichness_m155 created from richnessFish.csv with columns [longitude, latitude, current.iso.km, current.area.km2, past.iso.km, past.area.km2, total.rich] with no metadata. Lat/long represents centres of 5° bins, assumed WGS84. Discovered origin of grid is at -179 for fishnet reproduces their Fig1c after removing all NoData cells in original fishnet. 275 rows of original data created 273 cells of spatial data due to some wonky points around Central America (Fig 7).
- **Known issues:**
    - EEZ data does not cover entire land mass, so some points are missing EEZ information, for example where points fall on islands (e.g., Big Island, Hawaii and Cook Islands, New Zealand)
    - Join with PellissierRichness_m155 duplicates 40 points as they fall *exactly* on the boundary of two cells. Using 'Find Identical' confirms this. 30 duplicates near Thailand (Fig 8) and 10 duplicate points in GBR (Fig 9). Used Delete Identical tool to remove duplicate rows, resulting in the original 35,897 points (ipdb_join5noDups.shp).
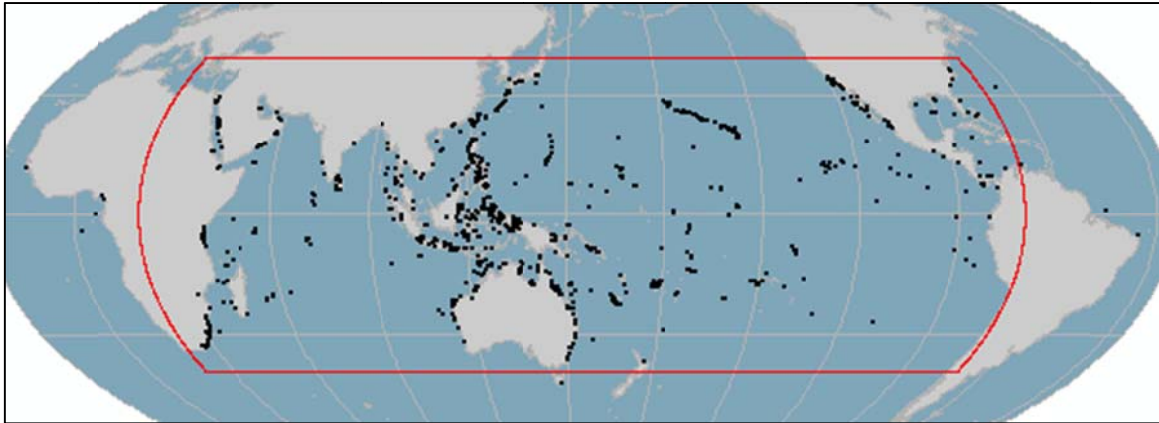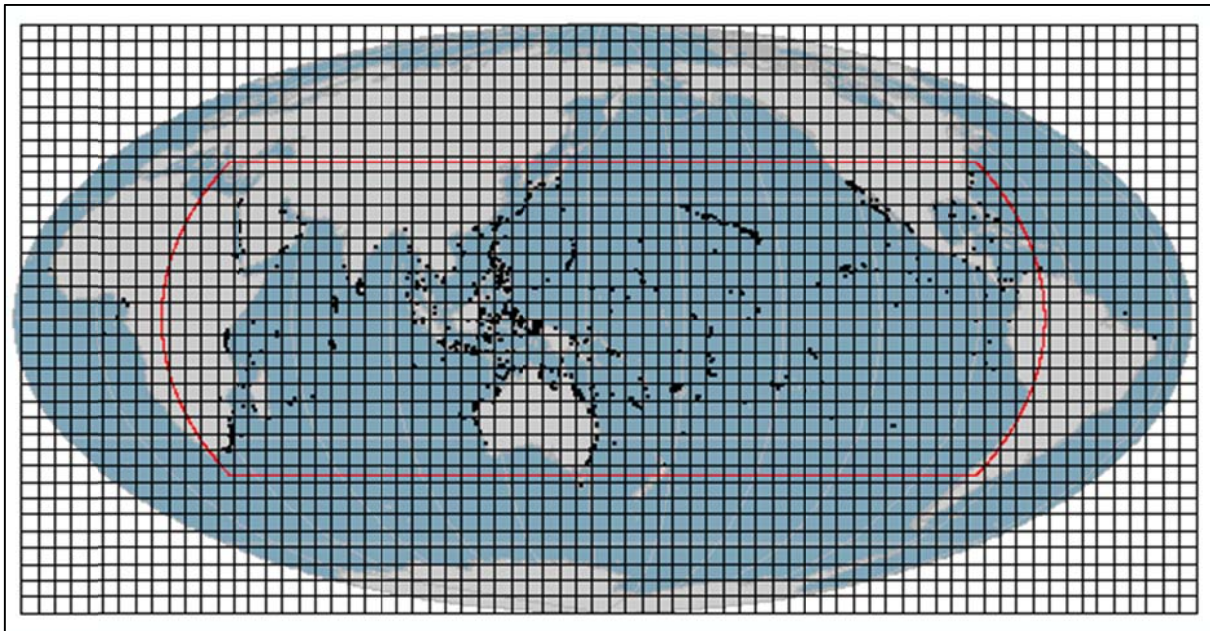
Figure 1. All data as of 3 June 2015

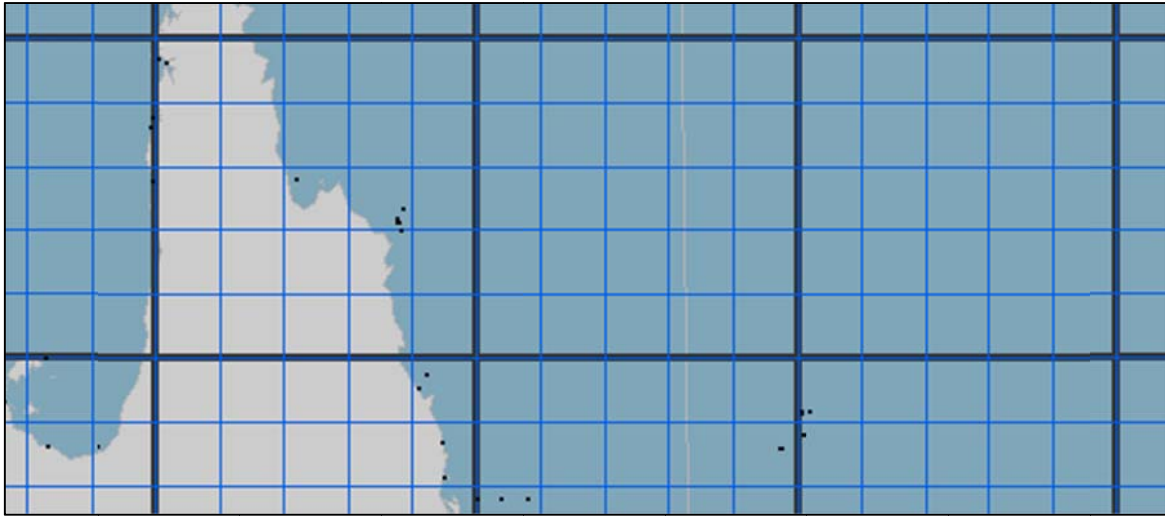Figure 2. fn500_m155 shapefile. fn100 nests within these cells

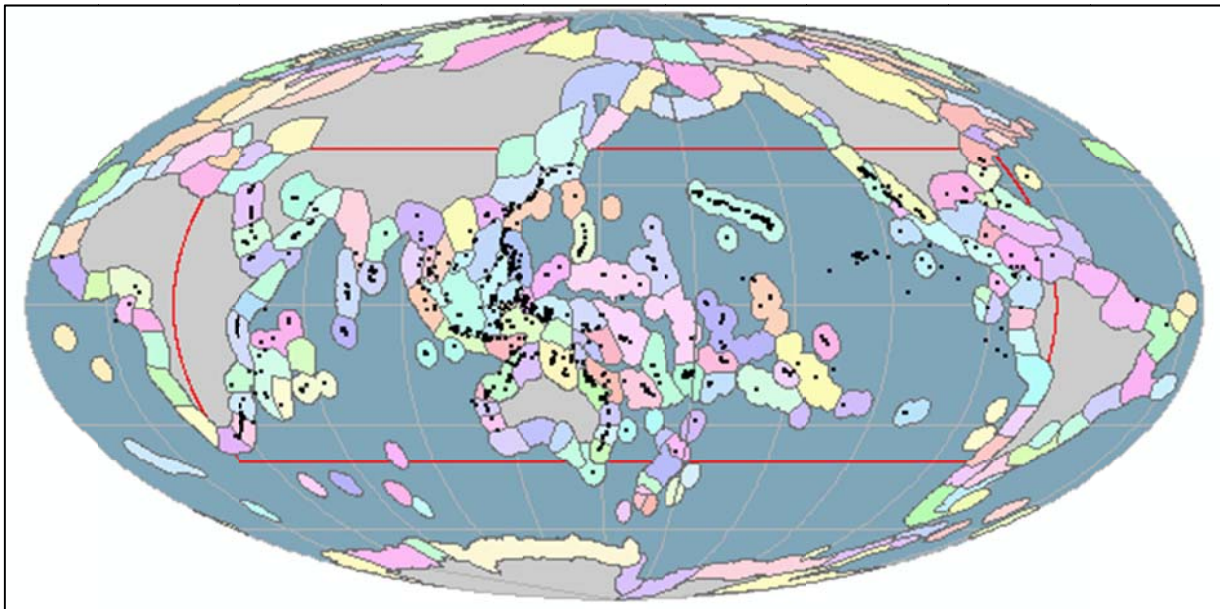**Figure 3. fn100_m155 cells (blue) within fn500 cells**

Figure 4. Marine Ecoregions of the World (Spalding et al. 2007)
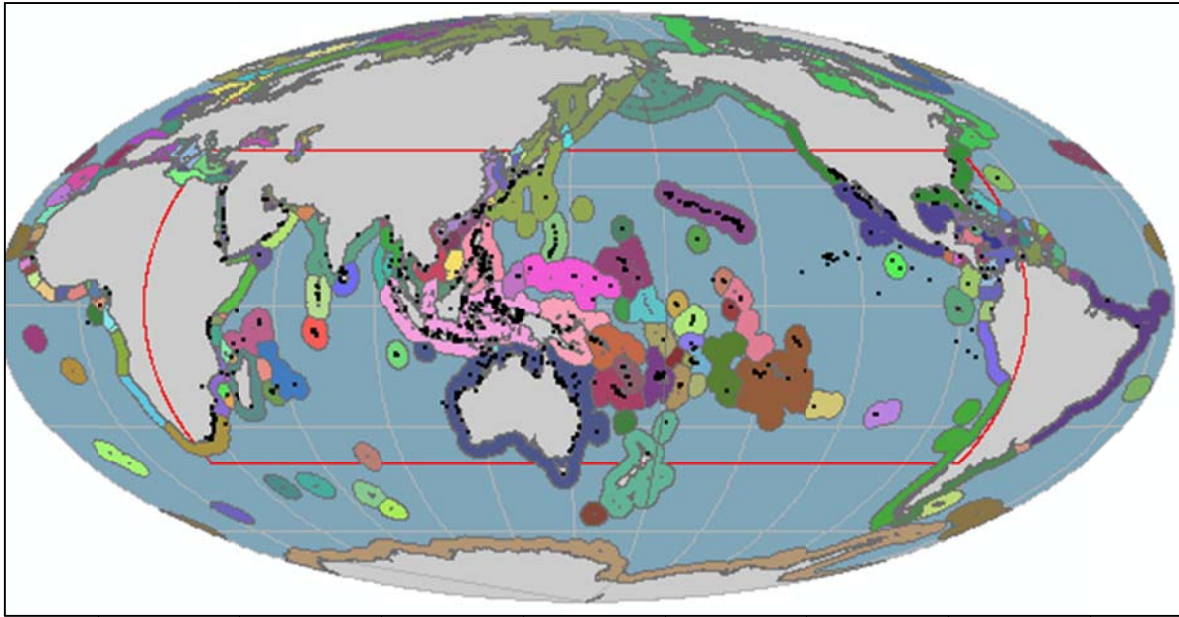
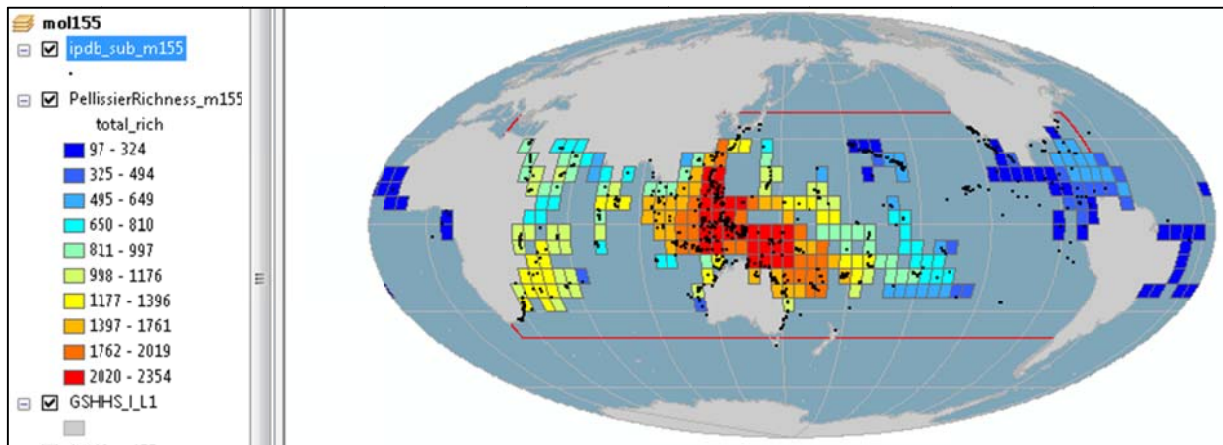**Figure 5. Exclusive Economic Zone data (v8, Marineregions.org)**

Figure 6. Data recreated from RichnessFish.csv from Pellissier et al. 2014, showing Richness values (as in Fig1C in Pellissier)
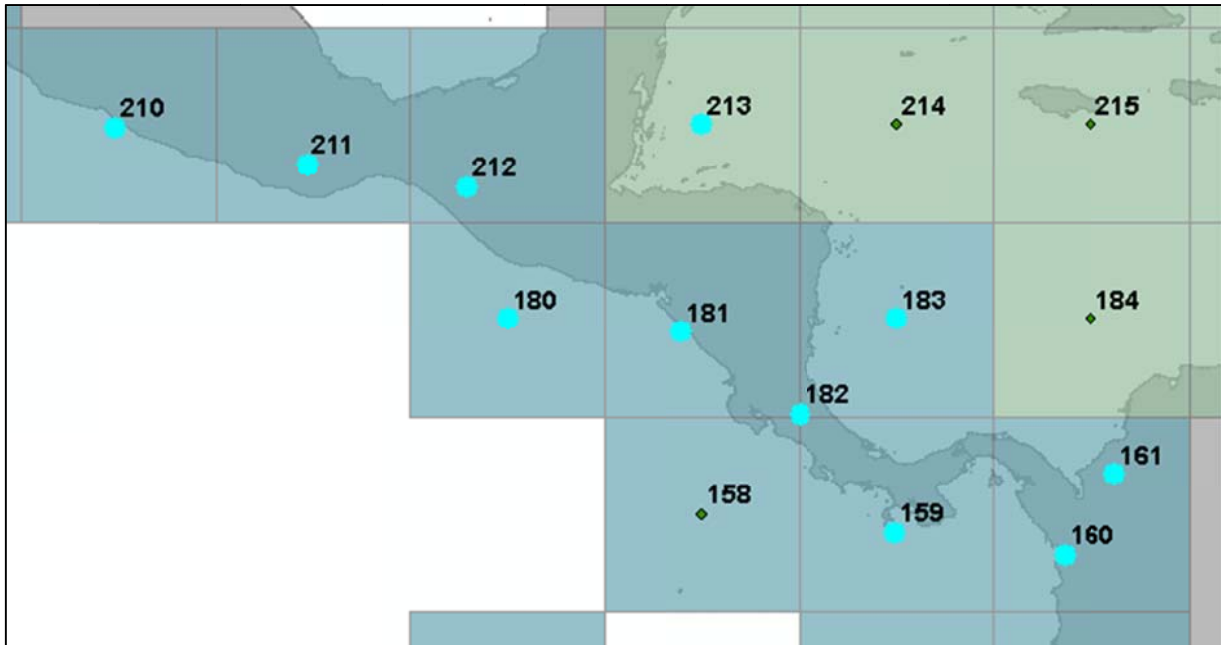
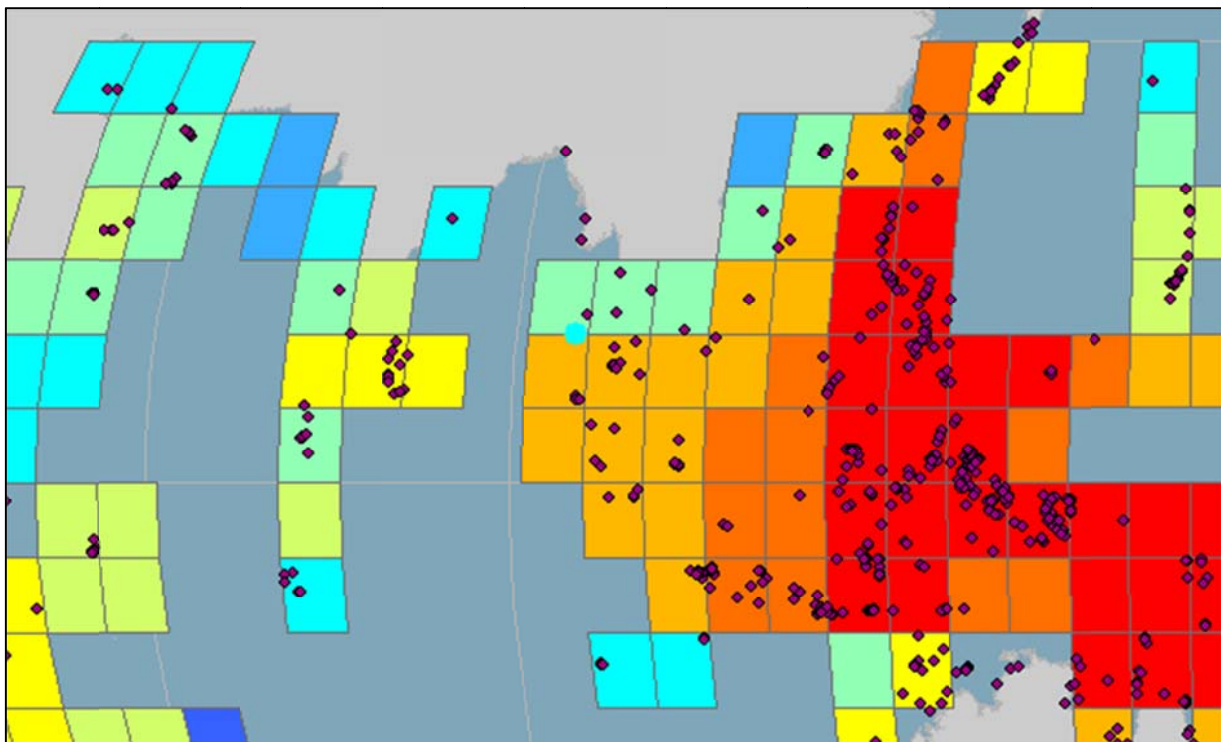Figure 7. Wonky points in original .csv file. Points in Coasta Rica (pnt 182) and SW Colombia (pnt 160) were deleted.

Figure 8. 30 Duplicate points (FID >=8523 & <= 8582 in ipdb_join5)
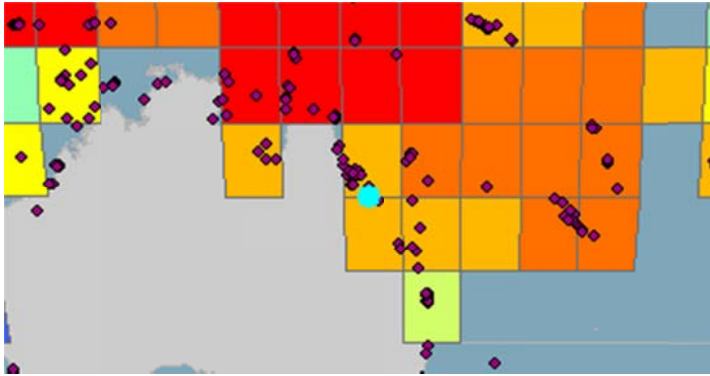
Figure 9. Duplicate points (FID >=19704 & <= 19723 in ipdb_join5)