Trend analysis tool

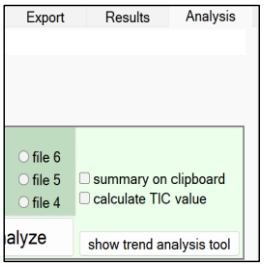
The tool determines trends in the detection rates from PODs at one or more fixed sites.

It uses '**Paired Year Ratio Assessment**' = PYRA.

> This method is described in 'Estimating cetacean population trends from static acoustic monitoring data using Paired Year Ratio Assessment (PYRA)' Grist *et al* https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0264289)

> The distinctive advantage of PYRA is that it avoids the need for estimation of both seasonal and diel patterns. Cetaceans often show big seasonal and diel patterns of activity and their estimation from gappy data is a major source of error in some other methods.

This tool is accessed via the Analysis page of the menu:



How it works:

Paired Year Ratio Assessment requires 2 or more years of data and handles gaps in the data by using only data from days that can be 'paired' or 'matched' with data from the same day in the following year. So the sum total of the detection metric in the second year is divided by the matching total in the first year to give a ratio that represents the growth or shrinkage – the 'change value' - of the data logged.

This two-year data window is moved forward one day at a time to produce a change value for each date. The date for this change value is the date between the two years. If the data window is moved further the number of matched days eventually falls as the second year becomes more incomplete. When the volume of detections becomes very low the change value becomes more and more wild! That's also true at any time within the body of the data when there are few detections, so to avoid wild results a small 'ballast' is added to both years. This has the effect of pushing the change value towards 1.0 = no growth.

> The process ends when there is no longer any paired data in the Y2 period. It can also start when there is only 1 paired day, but these 'thin ends' are of little value. So change values based on less than a mean of 1 DPM (detection positive minute) per day are not graphed … but all paired data is used in the assessment of overall trends.
>
> A measure of uncertainty is obtained by random resampling of whole days, with replacement, within a 21-day window centred on the paired day.

What does it tell us?

Sampling the trend in cetacean activity at a number of sites is often a very much cheaper way of getting a handle on what is happening to the population of a cetacean than doing repeated line transect surveys of its entire distribution from boats, planes or drones. Much larger numbers of detections can be made within the lower cost, and they give much greater precision. So what's not to like?

The sites chosen will usually cover only a small part of the distribution of the species (the monitoring of the Vaquita is a notable exception). So redistribution of the animals is potentially a major issue. Has the species declined or has it moved away? Has it increased or become more concentrated in the area logged? Habitat type will often be a factor in redistribution, and analysis of trends at logging sites with differing characteristics may be relevant (the app allows you to select and view subsets of the sites and to export data from all sites individually). These questions need to be considered and related to other biological or physical data, so that these trend results form a part of that larger picture.

The bottom line is: the trend values are simply the trend seen in the set of sites, and their predictive value for the 'local population' and the whole population needs to be determined using other data which might include prior data from such trend monitoring elsewhere, data on fish stock changes, physical oceanographic data, etc.

Limitations

Gappy data: when logged periods don't align across years the un-paired days are not used. A lot of data can be 'lost' this way, so if the whole year cannot be logged it's good to log the same parts of successive years.

Site stability: sites should be fixed, and any change in position of more than 200m should be treated as a new site. This is because we know that detection rates of porpoises vary significantly at this scale, even when the habitat looks superficially fairly uniform. (A fully randomised spatial design is theoretically possible and would use only the overall statistics for each year but would require many more sites).

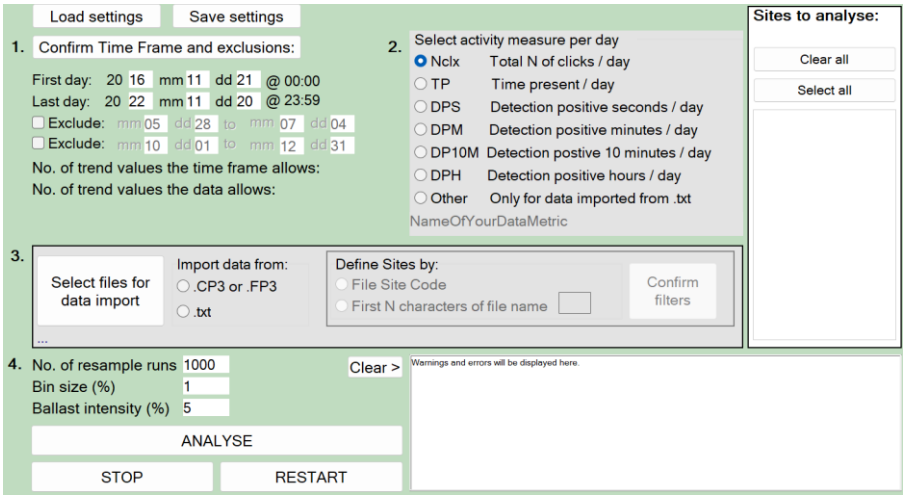
If a site changes – perhaps construction works starts nearby, or it is a maturing aquaculture site – then a local trend may be due to those changes, so it should be excluded from the overall trend analysis, but its own trend may still be of real interest.

Grist *et al.* (2022) discuss the limitations of PYRA.

Steps in using the trend analysis tool

In brief:

1. Give your data files site codes that identify each site correctly. If they have no site code the early part of the file name can be used and must uniquely identify each site in the study.
2. Crop all files to remove any time not in the water.
3. List all file times, (see 'Filers +files' page) then sort by dates of start and end to identify the dates that should be used in the trend analysis.
4. Enter the earliest and lates days ,of any file, in section 1 and also any exclusion period if part of each year was not normally logged or was subject to something that might interfere with the normal distribution of the cetaceans:



5. Select a detection metric. DPM – detection positive minutes is the default. Larger units may 'saturate' and it's a good idea to use only units that do not go above about 30% of the maximum value they can have in the

(which is 24 for DPH, 144 for DP10M, 1440 for DPM etc.) in the present data set or in the likely future of the project. Nclx may be conflate presence with feeding or social activity.

6. Click 'Confirm Time Frame and exclusions'
7. Set the filters for the input data e.g. Hi and Mod Quality, NBHF species, High species confidence. You could do assessments where you filtered out fast trains or slow trains etc. if you wished. If you are reading the data from a text file the filters will be those in force when the data was originally read.
8. In section 3 select where you will import data from.
9. Select 'Define sites by' unless you are reading from a text file.
10. Click Confirm filters.
11. Click 'Select files for data import'. The number of files and sites found will be shown. Later you can export a text file of the data and read that if you wish to save time.
12. Generally keep defaults for the resample runs, bin size and ballast …. more in Notes below.
13. Click 'save settings' at the top of the form and give it your project name.
14. Click Analyse. File reading and processing are followed by a 'View graphs…' popup message.
15. Now it gets interesting! But before you get into viewing the data click 'Put site list on clipboard'. The list will be on the clipboard so you can paste it directly into a spreadsheet and check that your files are correctly grouped into sites. If any have the wrong site code that can be changed via the File changes section on the Filters +files page – and you'll need to repeat the process above. If section 5 has disappeared, click 'Show PYRA controls'

Notes

The **resample runs** sets the number of times each paired day will be resampled – see 'How it works' above. If you set the resample runs to 1 there is actually no resampling - your original data is used. Increasing the number of resampling runs give a smoother distribution of change values. Larger numbers are a good idea.

Bin size: Each resampled PYRA change value will be added into a bin to produce percentile ranges and a histogram of results. The size of each bin can be set using the bin size value. Change values beyond the value of the most extreme bins will be added to that bin.

Ballast intensity

This is explained in 'How it works' above. The default is to add a ballast that is 5% of the overall sum of detections to both the year before and the year after the time point for the change value. It's effect on the overall trend is very small, but it heavily damps values where very few detections are present, pushing the change value towards no change.

Once your data has been analysed, the controls at the top of the trend analysis window and bottom of the window will be unlocked, allowing you to visualise your data and export values for further exploration in 3rd party software.

1. Visualise your data

"Graph trend statistics" shows the sequence of PYRA values for each site, with the overall trend sequence in a heavier red line.

"Graph spread of PYRA values" shows the distribution of trend values obtained in the resampling runs.

All plots will be created with a black background unless the "White Background" checkbox is selected. To view all of the trend analysis controls press the "Expand PYRA Controls" button.

To view the main FPOD window again, you can either close the trend analysis window (your settings will be maintained while in this session) or restore the window size as you would for any other application.

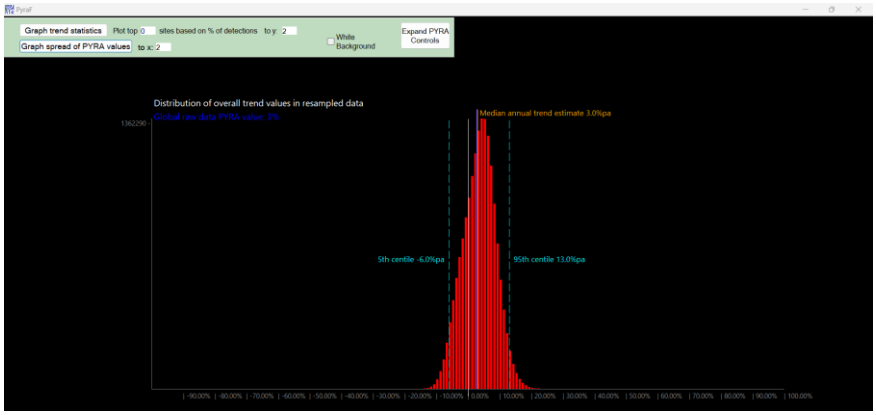
To view the PYRA values as trends, you can press the “Graph trend statistics” button at the top of the window. You can set the maximum y scale (PYRA value) by setting the “to y:” value. This defaults to a maximum PYRA value of 3. You can also specify whether you want to plot individual sites by setting the “plot top N sites based on the % of detections”. This will find that number of sites that contain the most detections and plot them alongside the overall trend. Here we set maximum Y value to 2 and plotted the top 3 sites (all sites in this case).

This shows three site trends with the overall / overall trend as the thicker red line.

In the legend on the right-hand side, the detection metric is displayed along with the overall trend and each site trend that is plotted. Each site also states the % of detections the site contains compared to the overall total. The overall trend line is calculated from all sites that had data, even if it is not plotted.

If you set the “Plot top N sites based on % of detections” to 0, only the overall trend will be displayed along with the overall 5th and 95th percentiles. The percentiles are only calculated with resampling runs > 100.

To view the spread of PYRA values calculated for the overall trend, you can select the “Graph spread of PYRA values” button. The x axis displays the % change which can be set in the “to x:” setting. A value of 3 will set the x axis maximum to 200%, a setting of 2 will be a maximum of 100% and so on.



This plot also displays the 5th and 95th centile for the overall resampled PYRA values, the median PYRA value and the PYRA value of the raw data (run 1, unsampled data). In our data we had 403 possible PYRA values within our time frame and a total of 10,000 resamples making a potential maximum of 4,030,000 PYRA values. On the Y axis you can see the maximum value for the bin that contained the most PYRA values (the highest peak).

2. Export your results

In the bottom left corner of the trend analysis window are five buttons that allow you to export the results of the trend analysis.

Put PYRA report on clipboard - Summary PYRA report

This gives the PYRA analysis version number, the detection metric, overall time frame, resampling runs and the smoothing intensity. It then states the median, 5th and 95th centile for the overall resamples. The final section shows the mean percentage change per annum for the overall sum and each site. It also states the start and end date, the % of detections out of the overall total and the number of PYRA values calculated for each site.

	A	B	C	D	E
1	PYRA version=	under development	Detection metric=	Nclx/day	
2	Trend from	21/11/2016 00:00	to	20/11/2022 00:00	
3	Resampling runs	10000	Smoothing Intensity	0	
4		Global			
5	Median %change pa	3.00%			
6	5thCentile	-6.00%			
7	95thCentile	13.00%			
8		Global	Site01	Site02	Site03
9	Mean %change pa	2.85	23	-12.6	-0.201
10	Start Date		21/11/2016	21/11/2016	21/11/2016
11	End Date		20/11/2022	20/11/2022	20/11/2022
12	% of detections	100	33.3	42.1	24.6
13	Nof PYRA vals	n/a	1825	1825	1825
14	Yr1sum	7.54496E+14			
15	Yr2sum	7.76015E+14			
16					

Put overall trend data on clipboard – Overall trend report

This is the data that is plotted when you press “Graph trend statistics” with Top N sites set to 0.

The year values are the midpoints between Y1 and Y2 and are in days from 01-01-1900. To convert that to a date in Excel, format it as a date. In R the values need to be multiplied by 86400 and formatted using a function such “as as.POSIXct”, specifying the origin as 1899-12-31.

In this report you will get the overall results for each PYRA period, including the PYRA year number, the year midpoint (days from 01-01-1900), sum of all resampled detections for year one, sum of all resamples for year 2, the PYRA value, the 5th centile and 95th centile (if resampling runs > 100).

YrN	Yr/Yr mid point	SumYr1	SumYr2	SumYr2/SumYr1	5thCentile	95thCentile	Data start =	2016 11 21	N of runs=	10000	Data: Sum of all resampliong runs
0	43060	462324282188	446133135246	0.964978809	0.919999957	1.039999962					
1	43061	462499373968	446157903510	0.964667043	0.919999957	1.039999962					
2	43062	462923004743	446082166788	0.96362065	0.909999967	1.039999962					
3	43063	461924315898	446094163568	0.965729987	0.919999957	1.039999962					
4	43064	461895023476	446097371580	0.965798177	0.919999957	1.039999962					
5	43065	462277037943	446189549264	0.965199464	0.919999957	1.039999962					
6	43066	462326840736	446257993676	0.965243534	0.919999957	1.039999962					
7	43067	461426386350	446337838441	0.967300206	0.919999957	1.039999962					

Put Site Trends on clipboard – Individual Site trend report

This is the data that is plotted when you press "Graph trend statistics" with top N sites set to the maximum number of sites. It also contains the sum of resamples for each year of each site.

Year	Site01	Site01 Yr + 1	Site01 Yr2/yr1	Site02	Site02 Yr + 1	Site02 Yr2/yr1	Site03	Site03 Yr + 1	Site03 Yr2/yr1	Data start =	2016 11 21	N of runs=	10000	Data: Sum of all resampliong runs
0	87223003216	121413684808	1.39199	253006965142	218738796890	0.86456	122094313829	105980653548	0.86802					
1	87429683625	121594675996	1.39077	252756804383	218568056457	0.86474	122312885959	105995171057	0.86659					
2	87628698632	121728453553	1.38914	252769345079	218338930468	0.86379	122524961031	106014782767	0.86525					
3	87298091954	121864427769	1.39596	252165877016	218159401044	0.86514	122460346927	106070334755	0.86616					
4	87468037856	122001826051	1.39482	251991171921	217941186196	0.86488	122435813698	106154359333	0.86702					
5	87319033419	122129940805	1.39866	252352525515	217819777950	0.86316	122605479008	106239830509	0.86652					

Put original data on clipboard – Raw data

This report can be useful to check whether the data you expected to be analysed was found. Days that were identified as having no data or less than a full day of data will be signified by -10.

This data is in the format required for the data memo, so you can save it in a spreadsheet and paste it back into the trend memo instead of rereading the files again. This can save a lot of time!

DayN	Site01	Site02	Site03
42695	44547	93782	60633
42696	0	68067	4949
42697	30076	109099	49524
42698	36600	88022	0
42699	0	67833	32056
42700	73247	40833	91761
42701	0	109105	34091

Put site list on clipboard – File report

Check this list to confirm that all files have been attributed to the sites you expect with a realistic number of days.

Site name	Site valid Days	File valid days	read from file
C:\Users\Chelonia\Site01 2020 11 14 FPOD_6288 file0 PART 79d 23h 19m.FP3	618	80	TRUE
C:\Users\Chelonia\Site01 2021 02 02 FPOD_6288 file0 PART 55d 19h 22m.FP3	618	56	TRUE
C:\Users\Chelonia\Site01 2021 04 13 FPOD_6288 file0 PART 126d 9h 43m.FP3	618	126	TRUE
C:\Users\Chelonia\Site01 2021 08 18 FPOD_6288 file0 PART 123d 1h 1m.FP3	618	123	TRUE
C:\Users\Chelonia\Site01 2021 12 19 FPOD_6288 file0 PART 103d 21h 1m.FP3	618	104	TRUE
C:\Users\Chelonia\Site01 2022 04 11 FPOD_6288 file0 PART 128d 9h 1m.FP3	618	129	TRUE
end of data from site 1			
C:\Users\Chelonia\Site02 2020 10 01 FPOD_6287 file0 PART 23d 23h 39m.FP3	566	0	TRUE
C:\Users\Chelonia\Site02 2020 10 26 FPOD_6287 file0 PART 125d 18h 51m.FP3	566	107	TRUE
C:\Users\Chelonia\Site02 2021 03 01 FPOD_6287 file0 PART 125d 23h 6m.FP3	566	126	TRUE
C:\Users\Chelonia\Site02 2021 07 05 FPOD_6287 file0 PART 127d 15h 41m.FP3	566	127	TRUE
C:\Users\Chelonia\Site02 2021 11 17 FPOD_6287 file0 PART 106d 21h 23m.FP3	566	107	TRUE
C:\Users\Chelonia\Site02 2022 03 04 FPOD_6287 file0 PART 98d 14h 6m.FP3	566	99	TRUE
Excluded	C:\Users\Chelonia\Site02 2020 10 01 FPOD_6287 file0 PART 23d 23h 39m.FP3		
end of data from site 2			
C:\Users\Chelonia\Site03 2020 10 06 FPOD_6286 file0 PART 39d 21h 45m.FP3	604	2	TRUE
C:\Users\Chelonia\Site03 2020 11 16 FPOD_6286 file0 PART 77d 20h 52m.FP3	604	78	TRUE
C:\Users\Chelonia\Site03 2021 02 22 FPOD_6286 file0 PART 141d 1h .FP3	604	141	TRUE
C:\Users\Chelonia\Site03 2021 07 14 FPOD_6286 file0 PART 92d 2h 46m.FP3	604	92	TRUE
C:\Users\Chelonia\Site03 2021 10 14 FPOD_6286 file1 PART 34d 11h 44m.FP3	604	35	TRUE
C:\Users\Chelonia\Site03 2021 11 19 FPOD_6286 file0 PART 125d 1h 14m.FP3	604	125	TRUE
C:\Users\Chelonia\Site03 2022 03 24 FPOD_6286 file0 PART 130d 12h 51m.FP3	604	131	TRUE
end of data from site 3			

When the data memo is used to enter the data for analysis, this will list the number of valid days in each site.

	A	B	C	D
1	Site name	Site valid Days	File valid days	read from file
2	Site01	2191		FALSE
3	end of data from site 1			
4	Site02	2191		FALSE
5	end of data from site 2			
6	Site03	2191		FALSE
7	end of data from site 3			
8				

Run through with real data

The site names in this example indicate what the trend of that site should be. So here the orange line should show a negative population growth and thus be below 1.0, while the positive trend should be above 1.0. In both cases we can see this to be the case with the negative trend ranging between 0.7 and 1.0 (a population decrease of 0 to 30 %) while the positive trend ranges from 1.1 to 1.7 (a population increase of 10 – 70%). The green line represents a seasonal trend where the population increases to a peak around May-June time and drops to a low around November-December time. Since the sum of the 365 days from year to the next doesn't change, the trend remains around 1.0 but does range from 0.8 to 1.3 (we will see later that this is an artefact of something else). The blue site shows a Biannual trend, where in one year a population decreases and then the following year the population increases. This trend will be less common in real world data but is shown here to help with understanding the mechanism of how the trend analysis works.

One thing to note is that the trends are not indicative of actual population sizes, only their change in size. In this data the site with the largest population is the negative population which contains around 40% of the individuals. The site with the smallest population is the biannual population with 16% of the individuals. You could try using a smoothing factor of 0.1 on the data and rerun the analysis and see how it effects the biannual trend compared to the others.

Since this data has been generated for the purpose of this guide, it is fairly well populated and so have very little yearly variation, but it becomes evident that as the trend analysis comes towards the end of the time frame the variability starts to increase. This can be reduced by increasing the number of days of data that are required in a year and rerunning the analysis. When you allow smaller numbers of data days within a year, a single day can have a much larger impact on the PYRA value and increase yearly variability.

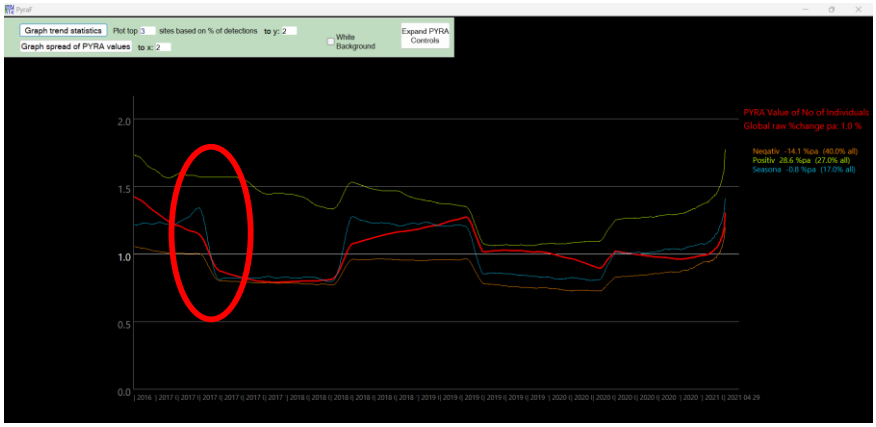
In this next plot we only plotted the top three sites based on their % of detections and set the y value to 2. You can see that all sites seem to have sudden drops and increases at roughly the same time (one instance is circled in red below). This is because in each site there are periods of 0 detections every 2 years between 28th May and the 4th July (maybe all of the animals were excluded from an area somehow?). For the positive site, this coincides with when no data was collected in 2018 which is why there is no initial drop as with other sites. To fix this we can set an exclusion period from 28th May – 4th July and rerun the analysis. If your data set includes several leap years, and you have an exclusion period, consider running PYRA on, say, 10-year sections to reduce artefacts created by drift. To remove this entirely, run PYRA on 4-year sections and set the exclusion period to finish a day after the intended period.

Commented [NT1]: I'm lost here … what is this odd data? … OK I get it, but I'm not sure it's a good educational tool, it is hard to read, although it did work as a test. We should discuss this

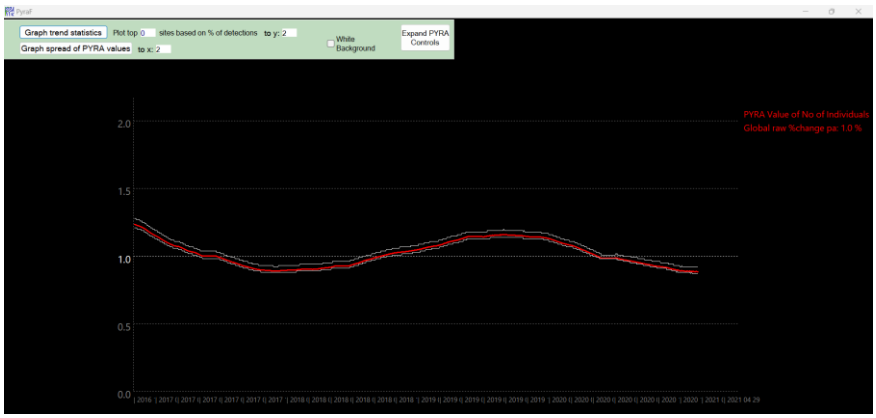
I'll stop at this point – I'm thinking a graph of data volumes by site would be good – I did code that at some stage.

Also a real example – maybe the Pesut data with a commentary on how to interpret it.

Commented [CD2R1]: Use real data in a run through at the end of the guide.

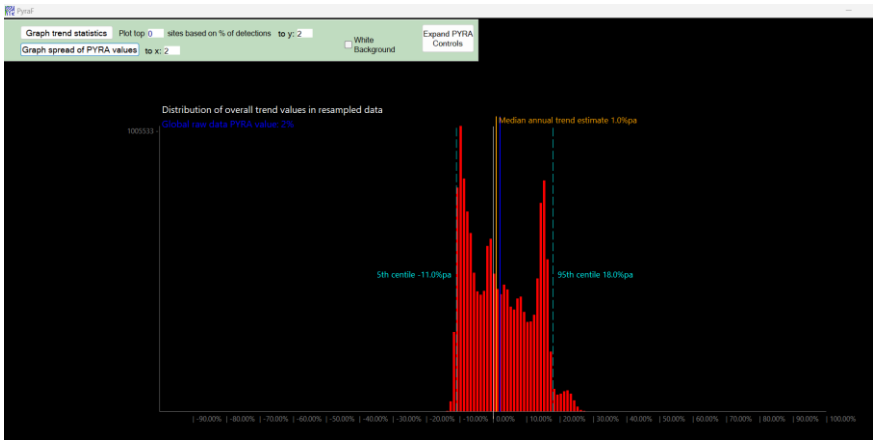


The next two images show the same data with the exclusion period removed and a minimum of 100 days in a year with data. In the site trends, the exclusion periods are plotted with dashed lines to indicate that data was removed from those years. In the positive and negative populations, you can still see two periods that have a very flat PYRA value (circled red). This is unlikely to be natural as populations are very rarely that stable. Instead, this could be periods where no data was collected, meaning that as the two-year PYRA window moves over them, the yearly PYRA values do not change as no new data is being added or removed compared to the previous year.



In this plot you can see that there are three distinct peaks. This is largely due to the two year exclusion period being included in the data when it shouldn't be. These periods of 0 values are driving the PYRA values to be much higher or lower than they should be for all sites, irrespective of their actual trend (as seen by the peaks and troughs in the PYRA trend graph).

After the data was analysed with exclusion periods and a minimum of 100 days of data in a year, the spread of PYRA values were centred much more closely around 1 with much narrower centiles. There does still appear to be three peaks, likely owing to the biannual trend increasing or decreasing the overall year values as it reaches its peaks and troughs. Try running the analysis again without the biannual site to see how the PYRA spread changes.



References

Grist, E.P.M., Mckinley, T.J., Das, S., Tregenza, T., Jeffries, A. & Tregenza, N. 2022. Estimating cetacean population trends from static acoustic monitoring data using paired year ratio assessment (pyra). *PLoS One,* 17**,** e0264289.