# Guide to the PLANKTON INDEX method and software, v.3.992

Paul Tett

Scottish Association for Marine Science (SAMS)

October 11, 2016

## Abstract

This is a guide to the use of the Matlab software package that allows users to calculate and plot the Plankton Index and related indicators. It refers mainly to version PI2015_Oct_beta, which allows for spatial as well as temporal comparisons. It also provides a short introduction to the theory of the Plankton Index and Euclidian Distance, and some discussions of Lifeforms. It is moving towards version 4. A new section describes features in development, including assessment of significance of time-series. These features are available in version PI2016_Oct_beta.

# Contents

Author's email: **paul.tett@sams.ac.uk**. Please let me know of any errors or matters than need clarification.

*Suggested citation:* Tett, P. (2016). Guide to the Plankton Index method and software, v.3.992. Scottish Association for Marine Science, Oban. 33 pp. + appendices.

# List of Figures

# List of Tables

# 1 Quick Start

**What is the Plankton Index?**

The Plankton Index (PI, Gowen et al., 2011) is a measure of the state of a plankton community relative to a reference condition. When first described (Tett et al., 2008) it was called the Phytoplankton Community Index (PCI).

**Quick Guide**

You need a recent (2011 or later) version of the Matlab application.

1. Unzip the file `PI2015_Oct_beta.zip` and install the resulting folder where it will be accessible to the Matlab application. Rename it `PI2015`.

2. Within Matlab, edit the `cd` command in line 96 in `PI2A.m` to specify the path (on your computer) of the folder (directory) `PI2015`. Save.

3. Set the Matlab current directory to `PI2015` and run `PI2A.m`. Respond to options presented in the Command Window. More options become available as you proceed.

4. `PI2015`, by default, loads the textfile `ptHPLF2edata1on7at16.txt` from the subfolder `data` and saves a pdf copy of each plot in subfolder `diagrams`.The package includes example diagrams in this subfolder

5. The accessory script `cf4.m` (in subfolder `params`) contains control parameters, including the name of the data file, and can be edited. Some parameters can be reset when running PI2A.

**Licence**

The code and data in the zipped package are made available under the Creative Commons Attribution 3.0 Unported License. You are free:

to Share – to copy, distribute and transmit the work

to Remix – to adapt the work

to make commercial use of the work

Under the following conditions:

Attribution – You must attribute the program to Paul Tett and SAMS, and cite this report (see page 2). For any reuse or distribution, you must make clear to others the CCA3 licence terms of this work.

# 2    About the Plankton Index

**Purpose:** The Plankton Index (PI) is a computational tool for evaluating change in the condition of the 'pelagic habitat' component of 'Environmental Status' as defined by the Marine Strategy Framework Directive (MSFD) and COM (2010); it could also be used for monitoring against 'undesirable disturbance' associated with (marine) eutrophication and for quantifying the (transitional or coastal water) phytoplankton biological quality element of Annex V of the Water Framework Directive (WFD).

   **Data requirements:** The PI's main use is to analyse time-series of plankton data, sampled at monthly intervals or more frequently, but it is robust against missing data. It must be possible to aggregate taxon abundances or biomasses into abundances or biomasses of at least two *lifeforms*, which can be defined by the user. Three years of data are, ideally, required to define a *reference condition* with which previous or subsequent annual change can be compared (e.g. Figure 1). Data can be derived from water samples or the Continuous Plankton recorder, and may come from single sites or be aggregated over a region defined by its *ecohydrodynamics* (van Leeuwen et al., 2015). It is also possible to compare spatially distinct regions, so long as they are of similar ecohydrodynamic type.

   **Combining values of the PI:** PI values are between 1 (no change from the reference condition) and zero (complete change). If data are available for additional pairs of independent lifeforms (i.e. not overlapping in their species content with other lifeforms), then PI values may be averaged.[1]
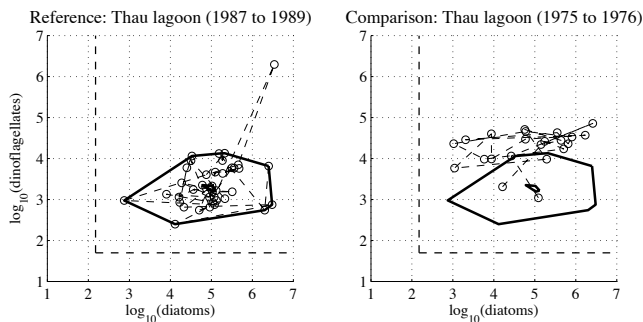


Figure 1: Example PI dagram for the lagoon of Thau in southern France (Gowen et al., 2015). Abundances in cells/L. The left side shows the reference envelope, which excludes the most extreme points. The right side shows new data, plotted over this envelope. PI = new points inside envelope ÷ total new points, and this case is 0.12. The 1975-76 condition was more eutrophic than that in 1987-89.

---

[1] Should they be arc-sine transformed first? Consult a statistician. See also section **??**.

# 3   Theory for the Plankton Index

The *Plankton Index* (PI) method draws on a number of key ideas. They are introduced here and are collectively referred to as the *Life-Form and State-Space Approach* (Gowen et al., 2011).

**Biomes** on land are characterised by the *lifeform* of their dominant primary producer. In the absence of human disturbance, a particular biome (such as 'deciduous woodland') would be expected to occur wherever climate and soils are suitable for this type. Applying this concept to the *pelagic component of marine ecosystems*, the characteristic primary producer lifeform, and its seasonal cycle, is considered to be a response to the prevailing *ecohydrodynamic* conditions (Tett et al., 2007; van Leeuwen et al., 2015). These are the physical conditions that select for species and communities in the sea, including water depth, stirring and stratifying tendencies, and light penetration.

**Ecosystems** are (in the theory presented here) seen as concrete things, instances of biomes. As defined by Lindeman (1942) they are *composed of the physical-chemical-biological processes active within a space-time unit of any magnitude*. Thus their limits can be arbitrary, as in the cases of the marine regions and subregions that are the management units of the MSFD. For PI purposes such larger and heterogenous units must be divided into ecohydrodynamically homogenous areas if sense is to be made of changes in PI values.[2]

**State variables and State Space.** *General Systems Theory* (GST) argues that an *internal description* of the state of an ecosystem - as of any system - can be given by values of "a set of n measures, called *state variables*" Furthermore, "change of the system is expressed by the trajectories that the state variables traverse in the *state space*, that is, the n-dimensional space of possible location of these variables" (Von Bertalanffy, 1972, p. 417). Given independent variables, states and trajectories can in principle be visualised by using the values of the state variables as the co-ordinates of points in an n-dimensional Euclidian space defined by n orthogonal axes.[3] In practice, visualising in more than 3 dimensions is difficult, and the PI method uses pairs of 2-dimensional graphs.

**Distance travelled in state space** In n-dimensional Euclidian space, distance travelled from a given starting point is a vector, made up of the

---

[2] For WFD purposes it would be acceptable to treat WFD *water-bodies* as the units.

[3] See Wikipedia page on Cartesian co-ordinates, and references cited there.

change of values ($\Delta y$) along each axis.[4] The vector can be converted to a scalar (the *Euclidian distance* $y_s$) by square-rooting the sum of squares of its components. That is, $y_s = \sqrt{\sum_{i=1}^{i=n}(\Delta y_i)^2}$. The assessment of change using Euclidian distance requires assumptions that each axis provides an *interval scale* and that the statistical distribution of errors is known. In contrast, the PI is calculated by counting points, which are categorised as either inside or outside the reference envelope (see e.g. figure 1), and thus it is a measure on a *nominal scale*.[5] It is a *non-parametric* statistic estimating distance travelled in state space, and its significance can be investigated with exact tests based on the binomial expansion or by approximations using $\chi^2$.

**Lifeforms.** The building blocks of ecosystems have typically been seen as (populations of) species, but recent ecological thought is tending towards the idea that the key units are functional groups, or *lifeforms*, made up of several or many species that play similar roles in ecosystem function (Folke et al., 2004; Hooper et al., 2005). Furthermore, the biome concept emphasises the lifeform (rather than the species) of the characteristic primary producer. The PI method uses the abundances of lifeforms (or *plankton functional groups*), instead of the abundances of species, as the variables describing the state of a pelagic ecosystem.

**Ecosystem organisation.** GST defines a system as "a set of elements standing in interrelation among themselves and with the environment" (Von Bertalanffy, 1972, p. 417). The pattern of relationships make up the *organisation* of the system; the intensity of interaction is the system's *vigour* (Mageau et al., 1995). Interactions amongst the biota of an ecosystem are often idealised in terms of a trophic network. The components are the nodes in this network; *organisation* describes the pattern of its connections; and *vigour* is what flows along these connections. Lifeforms may be equated with the nodes of the network, and changes in the relative abundances of lifeforms (i.e. in their relative positions in state space) seen as a proxy for their interactions.

**Seasonal succession.** A trophic network is a conceptual diagram, but many ecosystems (e.g. forest, coral reef) have a self-created physical organisation as well as a network. This is however less true of pelagic ecosystems, which have only the spatial structure that the inanimate environment imposes on them. Instead, the trophic interactions of temperate-zone planktonic ecosystems change with the seasons, and

---

[4] Notation: each state variable and axis is symbolised here by $y_i$, rather than by $x, y$ as is conventional for Cartesian co-ordinates, because the axes are supposed independent: they are like latitude and longitude on a map, in contrast to $x$-$y$ graphs in which $y$ is taken as a function of (and assumed to be causally dependent on) $x$.

[5] Siegel (1956) explains these distinctions between types of measurement and statistic.

it seems to the authors of the PI method that this seasonality is a key part of pelagic ecosystem *organisation*.[6] Much of the variability observed in the pelagic ecosystems of temperate coastal waters is seasonal.[7] In terms of state-space plots, this means that (i) a trajectory derived from weekly or monthly samples extends over a much greater area, volume or hypervolume than does a trajectory plotted from annual statistics; and (ii) taking an annual mean of lifeform abundance would lose important information about pelagic communities.

---

[6] For instance, the succession from spring diatoms to summer flagellates increases the efficiency with which the phytoplankton uses available resources of light and nutrients, in the same way that shrub layers increase the proportion of sunlight used by a wood.

[7] E.g., seasonal variation made up 74% of total variation in (log-transformed) chlorophyll concentrations observed 1972-1979 in the Scottish sea-loch Creran (Tett & Grantham (1980). The rest was measurement error, spatial variability, and inter-annual change.

# 4 History

The Plankton Index (PI) method began life as the *Phytoplankton Community Index* (PCI), a deliverable from a UK Defra-funded Cefas project led by David Mills for 'Development of a UK Phytoplankton Trophic Index' (2004-2006). Motivation was, in part, the need to detect disturbance to the balance of organisms associated with eutrophication. It was clear that seasonal changes needed to be taken into account, but it was also the case that we wanted to distinguish changes in 'balance' (using a state-space method) from changes in the timing of annually repeating events (the subject of *phenology*). Finally, we wanted a transparent, and easy to understand, method rather than for example one that involved multivariate statistical analyses and the use of *principal axes* to measure distance travelled in state space.

The method was implemented by a Matlab script and dependent functions. The program has been through numerous revisions to reach the present version comprising a main script `PI2A` and an augmented set of dependent functions. The guide to the use of the first 'published' script (Tett, 2006) could be used as an introduction to the present guide: the theoretical treatment has grown more rigorous as the method has developed.

A peer-reviewed description of the PCI method was published by Tett et al (2008). In adapting the method for proposed MSFD monitoring purposes, Gowen et al. (2011) revised its name to *Plankton Index.* The use of state-space to describe change in ecosystems was introduced in reports (Tett et al. 2004) and a paper (Tett et al., 2007) on *Undesirable Disturbance in the context of eutrophication.* A proposed *Framework for monitoring Ecosystem Health* (Tett et al., 2013) sets out the arguments for the use of functional groups and state-space. The UK MSFD methodology (see appendix A was further developed in the Defra-funded 'Lifeforms' project (2013-2014) led by Richard Gowen of AFBI (Scherer, Gowen et al., 2014). This project developed a 'master-list' for allocating plankton species to lifeforms and a procedure by which PI calculations would be carried out on the Defra-funded EMECO website.

Development of the Matlab software continued for research purposes (e.g. Gowen et al., 2015). However it began to seem that it would be convenient if the stand-alone program could be used by MSFD monitoring agencies (if desired) in order to explore local data more flexibly than possible using the stripped-down set of EMECO PI tools. A workshop in Aberdeen, convened for this purpose by Eileen Bresnan in September 2015, introduced users to version `PCI1H` of the program (documented by Tett, 2015), and led to the improvements included in this $\beta$ version of `PI2A`. The main change is that the program can read in two data files and so make spatial comparisons; in order to do this, the 'control' script has been updated to `cf4`.

# 5 Which Lifeforms?

A *lifeform* is *a set of species (not necessarily taxonomically related) that play similar roles in ecosystem function* (Tett et al., 2013). But an ecosystem is a complex network of functional relationships, and it is unclear into how many categories these should be analysed. Clearly it makes sense to distinguish the function of photosynthesis and primary production from that of consumption and recycling, and thus phytoplankters from zooplankters, but what further distinctions are needed to characterise all the key functions relevant to a pelagic ecosystem? They might include distinctions based on biogeochemical cycling (e.g. separating silicon users from non-users) and those relating to size and feeding mechanism, which help determine the structure of the food web. The designers of models of pelagic ecosystems have of necessity made judgements about the way in which they slice and dice the function of primary production (Follows & Dutkiewicz, 2011). To take an example, the 'Green Ocean Model' of Le Queré et al. (2005) distinguished 6 phytoplankton functional types: pico-autotrophs, $N_2$-fixers, calcifiers, DMS producers, silicifiers, and the remainder.
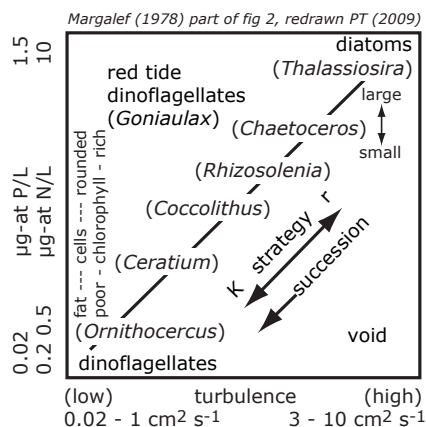


Figure 2: 'Margalef's mandala', redrawn from Margalef (1978).

The problem has been approached from consideration of phytoplankton functional traits, how these determine the ecological niches of species and, thus, where and when these species are abundant in phytoplanktonic communities (Smayda & Reynolds, 2001; Litchman & Klausmeier, 2008; Litchman et al., 2012). This approach is usually traced back to the seminal work by Margalef (1978), which was immediately influential (Pingree et al., 1978; Bowman et al., 1981; Jones et al, 1984).

Since Raunkiær, early in the $20^{th}$ century, botanists and botanical ecol-

ogists have categorized land (and related aquatic) plants according to their *life-forms*, including *phanerophytes* (trees and shrubs) and *chamaephytes* (woody ground cover).[8] Margalef applied the term *lifeform* to phytoplankters, arguing that the "different life-forms observed in phytoplankton [should be][9] functionally interpreted as adaptions to survival in an unstable and turbulent environment". These life-forms, he considered, lay along a continuum (figure 2) from r-selected species, able to grow fast when nutrient concentrations were high, to K-selected species, which could slowly accumulate biomass despite low nutrient concentrations. Such growth was opposed by losses due to vertical mixing or patch spreading,. This analysis shows how diatoms dominate the spring bloom (high nutrients, high losses) and dinoflagellates prevail under summer conditions (low nutrients, low losses). It reinforces phytoplanktonologists' intuitions that diatoms and dinoflagellates are different functional groups, as well as different taxa. In contrast, as figure 2 makes clear, Margalef's scheme dealt with species.[10]

Clearly, there is as yet no universally-agreed set of functional groups, or indeed agreed theory for them. In addition, there is the practical difficulty of lack of data relating to smaller or more fragile phytoplankters, which are not well sampled or easily identified and enumerated by standard methods using sedimentation microscopy. It is the responsibility of PI users to decide how they understand the concept 'lifeform' and how to aggregate their species-level data into lifeforms. The software works with time-series for pairs of lifeforms, so, as a minimum, two lifeforms must be used. For work in coastal waters, the diatom-dinoflagellate pair is often appropriate. These groups dominate primary producer biomass during much of the year, show seasonal succession, likely play different roles in pelagic food webs, and are almost always distinguished during microscopical analysis of phytoplankton samples. Depending on the amount of detail available from such analysis, it might be desirable to define:

**a pelagic diatom lifeform** excluding *tychopelagic diatoms* (Smetacek, 1986) which are better seen as phytobenthos;

**a photosynthetic dinoflagellate lifeform** excluding pure heterotrophs (those without chloroplasts).

Further suggestions can be found in Appendix B, and the UK has defined pairs of planktonic lifeforms for its MSFD monitoring (Appendix A).

---

[8] According to Wikipedia's page on Raunkiær's plant life-forms, the concept was first described in a paper in Danish (Raunkær, 1904) that was later translated into English as part of his collected works (Raunkær, 1934).

[9] I have replaced Margalef's 'are' with 'should be', because he was making a claim.

[10] Or, if not species, certainly low level taxa; not diatoms or dinoflagellates as a group. In general, reference to 'species' in this guide should be understood as 'the lowest-level taxa routinely indentified'.

# 6 Preparing the data

Data should be prepared in columns, using a spreadsheet, and then saved as a textfile. The file `ptHPLF2edata1on7at16.txt`, supplied in the 2015 package, provides an example. Here are a few rows:

```
%% year.day Si-users non-Si-users in biomass mm3/L
 1970.6740  1.54e+000  1.51e-001
 1971.2329  1.05e+000  2.39e-002
```

Rows starting with '%' are treated by Matlab as comments, and so ignored. They can be used to record information about the data.

The data themselves must be numeric. One column, in this case column 1, is the year, in decimal format (i.e. $<year> +(<day\ of\ year> /365)$). The other columns are the abundances of lifeforms, in this case in units of biomass (cell volume) concentration. Columns should be separated either by a fixed number of spaces, or by a tab character, and the row ended with an end-of-line character appropriate to the platform in use. Once saved, the file should be checked with a plain text editor for potential problems. If for example the spreadsheet application has saved a comment line as

```
'%% year.day Si-users non-Si-users in biomass mm3/L
```

the initial quote should be removed. Once checked, the file should be placed in the `PI2015` sufolder `data`. It can have any convenient name, so long as this is specified in the `cf4` script (see sections 10 & 11).

The lifeforms used in preparing the example were *Si-users* and *non Si-users*. Biomasses in the 'Si-user' column were calculated by summing over all silicon-using taxa the product of cell number concentration and typical cell volumes. The taxa included all diatoms and a few other micro-algae, such as silicoflagellates. The biomasses in the other column were calculated from similar sums over all other photosynthetic microplankters (i.e. including autotrophs and myotrophs but not pure heterotrophs). As already mentioned, the assignment of taxa to lifeforms is a user responsibility, as is providing cell volume (or other biomass) information.[11] The use of biomass data is recommended when one or more lifeforms includes a wide range of cell sizes. In other cases (e.g. the diatom-dinoflagellate pair), data on cells per unit volume will be adequate. Zero values may be entered into one or both lifeform columns: these will be interpreted in a special way (see section 13.

---

[11] In the case of UK and the monitoring of plankton for the Marine Strategy Framework Directive, there is a national list assigning taxa to MSFD lifeforms: see Appendix A.

# 7 Running the PI script

## 7.1 Introduction

`PI2A`[12] is the Matlab *script* contained in the file `PI2A.m`,. It calls a number of child functions (see section 12 and figure 7).[13]

Unzipping the current package gives `PI2015_Oct_beta`, which is a folder or directory containing: `PI2A.m`; various child functions; and three sub-directories: `data`, `figures` and `params`. The main folder or directory (but not the sub-directories) may be renamed. It will be assumed, here, that the directory has been installed somewhere that Matlab can find it, and that it has been renamed `PI2015`. You may call it anything you like.

To run the script: make `PI2015` the current directory, and type `PI2A` at the prompt in the Matlab Command Window (CW). Alternatively, open `PI2A.m` for editing – which you may want to do if you change the name of `cf4` – and invoke `debug` from the `Run` pull-down menu. Accept the option to `add PI2015 to current path`, and then switch to the CW.

## 7.2 Example output

The package includes an example of graphical output in the file `ptPCI1H-ptHP2005-12-20-41.pdf`, displayed as figure 3.
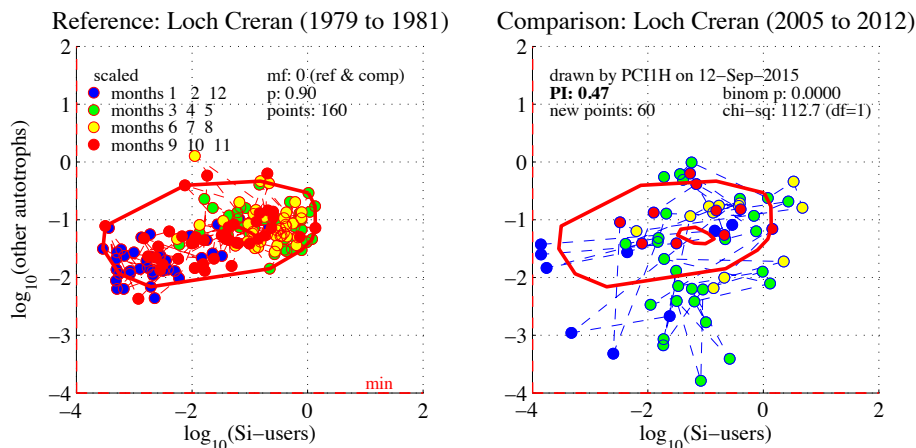


Figure 3: Example PI diagram for loch Creran. Made with program version `PCI1H`; `PI2A` should give almost identical results. Parameter `cp.COL = 1`, resulting in a colour plot and verbose commenting: cf. fig. 1. Abundance units might have been shown by assigning the value 'Si-users, `mm3/L`' to `cp.dsv1` in `cf4`.

---

[12] Earlier versions of the Matlab program for the Plankton Index (up to `PCI1H`) started with the letters `PCI`, referring to the 'Phytoplankton Community Index'.

[13] Consult Matlab `Help` for the difference between *scripts* and *functions*.

The left hand diagram shows reference data and an envelope drawn to include 90% of these data. The right hand side shows the value of the Plankton Index (PI), calculated by plotting new points and finding the proportion that fall within the reference envelope. In this case there were 60 new points, but only 28 of them fell within the envelope; thus the PI value was 0.47.

The diagram includes `binom p`, the probability, calculated by an exact binomial expanansion, of finding this proportion of points inside the envelope, given chance variation and an expectation of 0.9 (or whatever has been set in the parameter $p$). A probability less than the conventional value of 0.05 would imply that there has been a significant shift in community organisation: i.e. that the PI value is significantly different from the range between 0.9 and 1 that would imply no change. The diagram also contains the $\chi^2$ value associated with this PI value, for use as an approximate test of significance when there are more than about 200 points (when the binomial expansion becomes tedious to calculate).

The state-space axes are logarithmic: this is mainly for visualisation purposes, although the transformation does have beneficial effects when points are excluded from the envelopes (see section 13). The dashed lines labelled `min` are drawn at the values of the $z_i$ parameters (see section 13). Data entered as `0` will plot here. Plots exhibiting many values on one or both `min` lines point to defects in data acquisition: very likely, insufficient cells have been counted in the species assigned to that lifeform, and this should be addressed, perhaps by sedimenting a greater volume of sample.

The version of the PI diagram shown in figure 3 also colour-codes points by the season (the group of months) in which the plankton was sampled. Some scientific journals require black-&-white diagrams with minimal text, and setting the option `cp.COL` to `0` in `cf3` will output in such a format. `cp.COL =1` gives the 'verbose' and coloured output (see section **??**).

### 7.3   Running the script with test data

Running `PI2A`, without changing the default parameter values or file names given in the supplied version of `cf4`, should result in (something like) the following in the Matlab Command Window. A few things will be different: the program name (`PI2A` not `PCI1H`); the current directory (whatever you've set it to); and the control file name (`cf4` not `cf3`) and details.

```
********************************************************
PCI1H started: 12-Sep-2015 20:39:47
abort with ctrl-C, ctrl-break or <apple>-<.>
Current directory is C:\Users\sa02pt\Documents\MATLAB\PCI2015

Inserted the control file cf3, which was updated 7 Sep 2015 by PT,
including new plot limits
```

```
Expecting 1 data file(s)

Loaded data file ptHPLF2edata1on7at16.txt for Loch Creran
 -- Maximum biovolume, log10(mm3/L) =  0.74

Enter your 2-letter ID, e.g. [XX]: pt
```

→ At the prompt, type your initials; these will be prefixed to output files.Then press the 'enter' key (⟨enter⟩). The main menu will appear:

```
------------- MAIN MENU FOR PI ----------------
End PCI1H (0)
    Plot data time-series (1)
        Make reference envelope (2)
        Change values of control parameters (5)
    Make state space plot of annual medians (6)
Enter a number [0]: 2
----------------------------------
```

→ Option 2 has been selected by typing 2⟨enter⟩ at the prompt in the last line. The value in square brackets is the default, and can be selected with ⟨enter⟩.

As data are analysed, the available menu expands, until it is:

```
------------- MAIN MENU FOR PI ----------------
End PCI1H (0)
    Plot data time-series (1)
        Make reference envelope (2)
            single PI for comparison data (3)
            annual PI time-series (7)
                Plot PI diagram (4)
        Change values of control parameters (5)
    Make state space plot of annual medians (6)
Enter a number [0]:
----------------------------------
```

→ Choosing option 3 will lead to a request for a year range for calculation of a single PI value. The range may be a single year (e.g. 2012 to 2012, or a longer period as exemplified below:

```
Estimating PI with mf = 0
  using data for Loch Creran from file ptHPLF2edata1on7at16.txt
Select range of years for PCI:
Enter (integer) start year [1970]: 2005
```

```
Enter (integer) end year [2012]: 2012
--- using the year range: 2005 to 2012.999

   year range          PI    N-total N-out  binom p  chi-sq
 2005.00   2013.00     0.47      60      32   0.0000   112.67
```

The final year has been rounded up, to confirm that data used were those from the period finishing at midnight on 31 December 2012.

→ Option 7 is independent of option 3 but requires a reference envelope to have been calculated using option 2. The PI time-series option calculates a value of the PI for each year of the time series, outputting both to the CW and to a PI time-series diagram; it reports NaN if there are insufficient data available for a particular year.

→ Option 4 will not appear until option 3 has been selected at least once. Option 4 should result in the diagram shown in figure 3 (but with the current date). The diagram's new (and almost unique) file name will be printed in the CW, for reference. The last part of its name will differ from that given in subsection 7.2.

→ Option 5 allows some (numerical) parameter values set in cf4 (section 10) to be changed dynamically (i.e. whilst PI2A is running). The option steps through the following – press ⟨enter⟩ to accept defaults at each step.

**mf** – controlling selection of data from time-series: mf = 0 selects all data; mf = 1 calculates and plots monthly means; mf <0 randomly selects 1 in -mf of the data

**p** – sets fraction of plotted data included in envelope: usual value is 0.9; the function findenv2 corrals values within 0.5 − 1.0; see section 13.

**sc** – controls axis scaling: scaling (cp.sc=1) is used in the function findenv2 to standardize variability along the two axes of the plot; it should improve the fit of the envelope, but does not otherwise change the plot.

**reference period** : the start (cp.refs) and end (cp.refe) years can be changed. Use whole years only; the example period (set initially in cf3) 1979 − 1981 includes 3 whole years of data.

**COL** – switches between verbose colour (cp.COL=1) and laconic B&W (0).

**nmax or bmax** – maximum axes values in a PI diagram: the appropriate value (of biomass for cp.c1svt1 = 2, of cell numbers for = 1) will be presented for confirmation or alternation.

16

# 8 Timeseries of Plankton index values

Invoking option 7 from the main menu leads to the calculation of one PI value for each year between limits that are prompted for in the CW. Results are tabulated in the CW and plotted in a diagram such as that in figure 4.
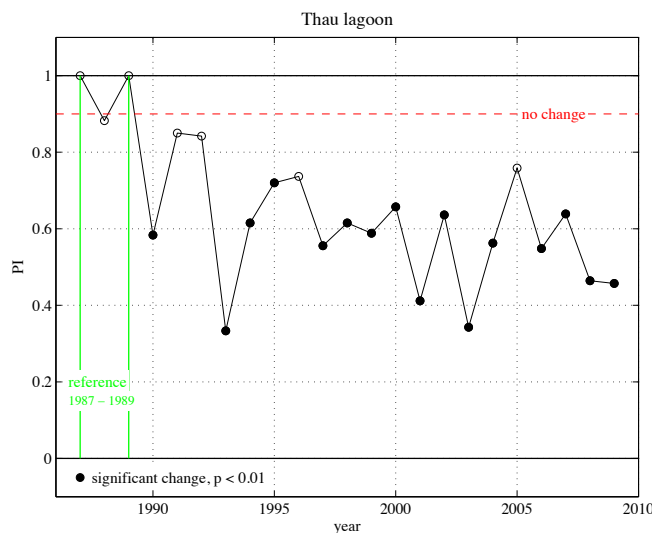


Figure 4: Time-series of PI values for the diatom-dinoflagellate pair from the lagoon of Thau. Filled circles show a PI value significantly different from the expectation in the range $p$ to 1.

The PI value is, by definition, $p$ during the reference period; distance travelled in (some transformation of) state-space is given by $(1 - p)$. The option 7 calculation routine will (if the prompted limits are set to include them) calculate individual PI values for each year or a multi-year reference period. These cannot exceed 1 but may be less.[14]

It should be kept in mind that the PI simply measures change from a reference condition: it does not indicate whether the change is for better or for worse. See also section 14.

---

[14] I have not explored the possibility of a PI greater than $p$, or whether such could be significant (perhaps for low values of $p$). Such a state of affairs would imply less system variability than that in the reference condition.

# 9 Other ways to present change

## 9.1 Ratio time-series

The time-series graphing option 1 ... ... used to check the $y_i$ time-series loaded by the program. The ... ... a time-series of the ratio $\frac{y_1 -}{y_1 + y_2}$, providing a way to scalarise position in a 2-D $(y_1, y_2)$ state space.
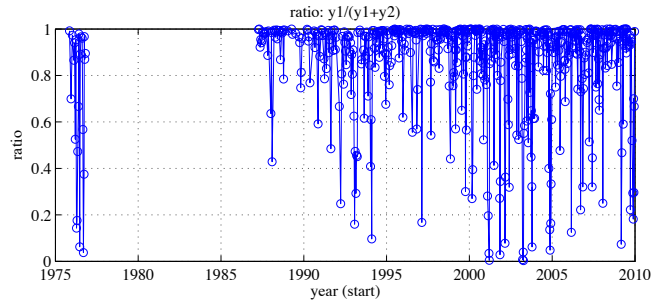


Figure 5: Time-series of ratios from the lagoon of Thau. $y_1$ is diatom abundance and $y_2$ that of dinoflagellates. Data source: `Thau-TSDD.txt`.

## 9.2 Euclidian distance

Menu option 6 calculates annual medians of each state variable and use these as co-ordinates in $y_1, y_2$ state-space. Euclidian distance travelled is also plotted, calculated from $y_s = \sqrt{\sum_{i=1}^{i=n}(\Delta y_i)^2}$ where $\Delta y_i[t] = y_i[t] - \overline{y_i}$. The final symbol refers to the median $y_i$ during the reference period.
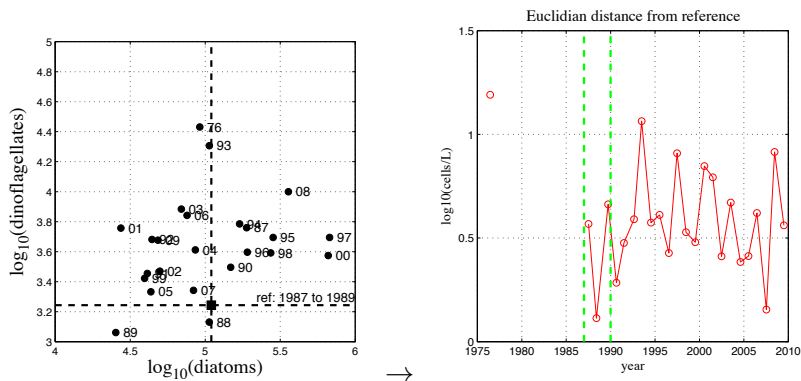


Figure 6: Lagoon of Thau: annual medians in state-space, and Euclidian distance travelled from a reference condition in 1986-88.

18

# 10    Parameters that can be changed

`cf4` is the name of a short script stored in the `params` subfolder and loaded by `PI2A` when this runs. It provides parameter values via the structure `cp` (known by other names, including `cd`, in functions called by `PI2A`). It may be convenient to keep several versions of cf4, with different names, and to edit the following line in `PI2A` for the required name:

```
cf4; % contains run parameters in a structure called cp
```

Example code from `cf4` follows; values given to fields of `cp` can be edited to relate to data to be loaded or analyses to be performed. It is assumed that the following default is set in `cf4`, instructing the script to load one data file only.

```
% number of data files to be loaded (1 or 2)
    cp.ndf = 1; % if 1, information about file 2 is ignored
```

See section **??** concerning the use of two data files.

## 10.1    Parameters relating to data

```
% name of data file (should include extension)
    cp.df1n = 'ptHPLF2edata1on7at16.txt';
```

→ this is the file that will be read, from subfolder `data`, when `PCI1H` runs. Edit the string 'ptHPLF2edata1on7at16.txt' to change the file-name.

```
% description of data
    cp.d1des = 'Loch Creran';
% number of column to use for date values
    cp.c1dat=1;
% column number and description of first state variable
    cp.c1sv1=2;
    cp.dsv1='Si-users';
% column number and description of second state variable
    cp.c1sv2=3;
    cp.dsv2='other autotrophs';
```

→ the numeric parameters identify columns in the data file, and the text strings provide labels for diagrams and printed output.

```
% state variables are numbers (1) or biomass (2)
    cp.c1svt1 = 2;
% file state variable data will be transformed, log10(y+z)
% in next 4 rows' comments, 'col 1' = column specified by c1sv1 (etc)
```

```
    cp.z1 = 150; % min cells/L in col 1 - half obsv min of 300 cells/L dia
    cp.bz1 = 0.0001; % min value for biovolume (mm3/L) in column 1
    cp.z2 = 50; % min cells/L in col 2- half obs min of 100 cells/L dino
    cp.bz2 = 0.0001; % min value for biovolume (mm3/L) in column 2
```

→ parameters `z1` etc estimate minimum cell concs. $z_i$ for each state variable $y_i$. If $y_i = 0$, the plotted value is $\log_{10}(z_i)$. Biomass minima are held in fields `.bz`. Which are applied depends on the value in `.c1svt1`.

## 10.2   Parameters relating to analysis and plotting

```
% first year of reference period (an integer)
   cp.refs=1979; % Creran
% last year of reference period (an integer; ---
%       all data from this year will be included)
    cp.refe=1981; % Creran
% data aggregation or subsampling
% if mf is > 0.5, data are averaged over month
% if mf is < 0, data are chosen at random with probability of 1/-mf
% if mf = 0, all data are used
    cp.mf=0;
% set switch to apply mf to reference data as well as sample data
    cp.mfref = 1; % 1 = apply to reference also; 0 = don't apply
% set number of iterations for random subsampling when mf < 0
    cp.mfn=10;
% set fraction of data to be excluded from envelope (outliers are
%   1 - (p/2) at top and bottom of distribution)
    cp.p = 0.9;
% set scaling of axes: 0 not scaling, 1 scaling on (used in FINDENV2)
    cp.sc = 1;
% set minimum number of values required for calculation of PI
    cp.mindata = 6;
```

→ some of these can be reset whilst the program is running.

```
% output information
   cp.COL = 0; % COL=0 sets black & white output with minimal text;
   % COL > 0.5 uses colour and provides more text
   cp.pt='pdf';% saved graphics type; alternatives are 'ai', 'ps', 'pdf'
                % and others - see function PIPRINT
   cp.nmax = 7; % default upper limit (log10(cells/L)) for cell no. plot
                % should not include space needed for text inserts
   cp.bmax = 0; % default upper limit (log10(mm3/L)) for biovolume plot
```

→ some of these can be reset whilst the program is running.

```
% seasonal selection
%       edit to allocate months to seasons for display in PI diagram
    cp.Winter = [1 2 12];
    cp.Spring = [3 4 5];
    cp.Summer = [6 7 8];
    cp.Autumn = [9 10 11];
```

$\rightarrow$ omitting a month number will result in no points plotted for that month;
setting a selection to the empty set [] will omit that season from the plot.

# 11 Spatial comparisons: editing `cf4`

Depending on the settings in the control file `cf4` (which replaces `cf3` of previous versions of the program) the PI program can read either one or two data files. The two-file case is useful for spatial comparisons, enabling a reference envelope to be made from data in one file and the comparisons carried out using data from the other file. Of course, the comparison must only be made between (1) waters of the same ecohydrodynamic type, and (2) abundance data collected by similar methods.

## 11.1 Selecting one or two files

Edit `cp.mdf` in the following line in `cf4`:

```
% number of data files to be loaded (1 or 2)
    cp.ndf = 2; % if 1, information about file 2 is ignored
```

When only one file is loaded, the reference and comparison data are both taken from it. When two files are loaded, `cp.reff` and `cp.compf` must be edited in `cf4` to specify from which file these data are to be taken.

```
% selection information --
% set switch to select file for reference data
    cp.reff = 1; % either 1 or 2
    if cp.ndf < 1.5, cp.reff=1; end % ensure it's 1 if only one file
```

and

```
% set switch to select file for comparison data
    cp.compf = 2; % either 1 or 2
    if cp.ndf < 1.5, cp.compf=1; end % ensure it's 1 if only one file
```

It is possible to load 2 files and use only one for comparison and reference!

## 11.2 Parameters for the second file

These must be supplied in `cf4`, as shown. Using this example, the same file would be loaded twice; normally, a different name would be given for second file.

```
% ===== information about data file 2 (which may not be needed)
% data in file 2 must be comparable with data in file 1
% for example, file 1 cp.z and .bz will be used for both data sets
% name of data file 2 (should include extension)
    cp.df2n='ptHPLF2edata1on7at16.txt';
% description of data
    cp.d2des='Creran loch (2)';
```

```
% number of column to use for date values
    cp.c2dat=1;
% column number of first state variable
    cp.c2sv1=2; % must be same as first SV from file 1
% column number and description of second state variable
    cp.c2sv2=3; % must be same as second SV from file 1
%
```

As the commenting emphasises, the variables in file 2 must be comparable
with those in file 1 in units and minimum values. COLUMN NUMBERS?

# 12   Structure of the PI program

Figure 7 shows the main components of the program. Each function m-file contains header comments giving further details.
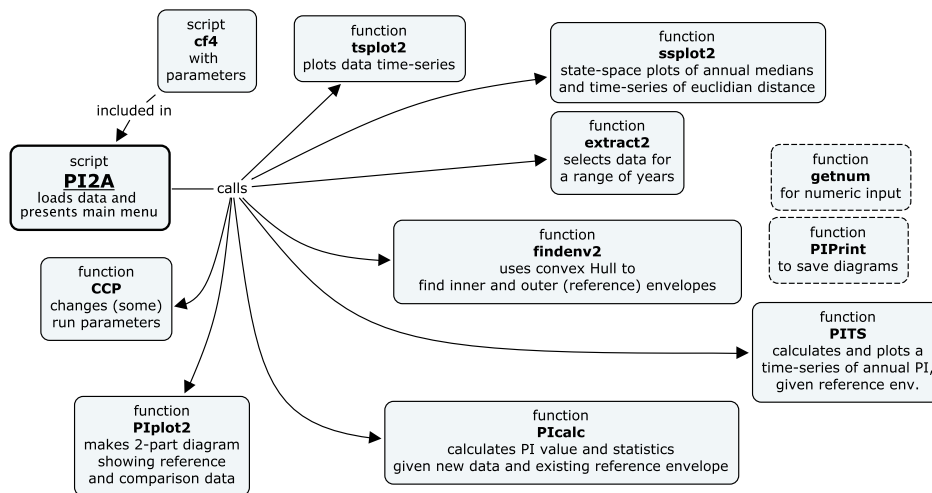


Figure 7: Dependencies of the Matlab script `PI2A`, which can load one or two data files

In addition,

**getnum** is used whenever numeric input is required with a default as a prompt.

**findenv2** uses the Matlab cart2pol, pol2cart and convhull functions, and the *private* function `trimmedwith`

**PIcalc** uses the Matlab inpolygon and nchoosek functions

**PIPlot2** calls `PIPrint`

**PIPrint** is called whenever diagrams are to be saved; check it for allowable formats.

**PITS** calls `extract2`, `PIcalc` and `PIPrint`

**ssplot2** calls `extract2` and `PIPrint`

**tsplot2** calls `extract2` and `PIPrint`

# 13   Numerical Methods

**Logarithmic transformation** When plankton data are analysed by parametric statistical methods, they are usually logarithmically transformed in order to normalize error distribution and render it homoscedastic. This is not necessary for the PI method, as it relies on categorizing points as falling inside or outside the reference envelope, and the Convex hull procedure will draw the envelope around the same points whether, or not, variables are transformed. However, transformation is useful for visualising data which encompasses several orders of magnitude. Furthermore, logarithmic transformation gives more equal weight to variables when they are combined in calculating Euclidian distance, because it is proportionate change that is, then, evaluated.[15] Finally, the procedure used to exclude $(1 - p)$ of points is influenced (beneficially) by transformation. In essence, points are ranked according to their Euclidian distance from the centre of the distribution of points; logarithmic transformation helps exclude equal proportions of high-valued and low-valued points.

**Dealing with zero values** The time-series data read into the PCI program may include `0` or `NaN` values in either lifeform. The function `extract2` takes data for specified years from the input data, deleting rows with `NaN`, and transforming all numeric data according to $\log_{10}(y_i + z_i)$, where $y_i$ is an abundance and $z_i$ is a parameter given a value (as `cp` fields `.z1`, `.z2`, `.bz1` or `.bz2` ) in `cf3` (see section **??**). All subsequent calculations are performed on these log-transformed values. A suitable value for a $z_i$ is half the minimum observable abundance for the corresponding $y_i$. This is simple in the case of cell abundances: if 10 mL has been sedimented, the minimum observable abundance (calculated when 1 cell has been seen) is 100 cells/L, and thus $z = 50$. In the case of biomasses, all available values should be ranked. If there are zero values at the bottom of the ranking, then the first non-zero value is the minimum observed biomass.[16]

**Aggregating or subsampling the data** Depending on the value of *mf*, `extract2` can average sample data over 12 equal periods ('months') within a year (for `mf = 1`) or (for `mf <0`) randomly subsample the

---

[15] For example: diatom abundance changes from 10,000 to 100,000 and dinoflagellate abundance from 1,000 to 10,000; the untransformed Euclidian distance is 90,449 – i.e. biased towards the larger change. After log-transformation, diatoms increase from 4.0 to 5.0 and dinoflagellates from 3.0 to 4.0; the resulting Euclidian distance is 1.41, equally influenced by both variables.

[16] These remarks are, mainly, based on experience with data derived from microscopic analysis of water-sampled microplankton. They may need to be expanded to deal with experience of net-sampled plankton or CPR data.

data, taking a proportion $\frac{1}{-mf}$. The latter is useful for investigating statistical properties using a Monte Carlo method. The default value, $\mathtt{mf} = 0$, uses all data without aggregating or subsampling.

**Making the reference envelope** in `findenv2.m`. This is done by fitting a *Convex Hull* around the cloud of points, after excluding outliers. For points plotted in 2 dimensions, the Hull is the shape of a rubber band stretched around these points, and the Matlab `convhull` function returns the co-ordinates of the points touched by the band. Then the cloud is turned inside out and a second set of Hull co-ordinates obtained, corresponding to the inner envelope.[17] The processes of excluding fraction $1-p$ outliers, and turning the cloud inside out, requires the cloud's centre to be found. `findenv` places this at co-ordinates determined by the median value of each variable. The points' Cartesian co-ordinates are transformed to polar co-ordinates about this centre, specified by an angle and a radial distance from the centre. The $N$ points are then ranked on their radial distances, and the bottom and top $(1-\frac{p}{2})N$ are excluded from the Hull fitting process. 'Turning inside out' simply requires taking the reciprocal of the radial co-ordinate.

**The value of $p$** This is the proportion $p$ of points included in the reference envelope and the expectation for subsequently-calculated values of the PI. $p = 1$ retains all points including extreme outliers and results in a large reference envelope, so that the method becomes insensitive to change. Excluding a large fraction of points results in a smaller envelope, so that the method becomes more sensitive to change; however, the values of the PI that are generated might be less likely to be significant. $p = 0.9$ seems to give the best outcome. Excluding 10% of points seems reasonable and is in accord with practice in other phytoplankton tools.[18]

**Scaling** in `findenv2.m` (with $\mathtt{sc} = 1$) uses for each lifeform the transformation $\frac{y_i - \overline{y_i}}{\sigma_{y_i}}$, where $\overline{y_i}$ is the median value and $\sigma_{y_i}$ is the standard deviation (of the log-transformed variable). When applied, scaling has the effect of fitting the envelope more tightly about a cloud of points, especially if the cloud is markedly elliptical in shape. Scaling completes the equalising of the contribution of each variable.

---

[17] Why two envelopes? See Appendix C. However, the argument given here might not apply to all lifeform pairs used for UK MSFD monitoring, and needs to be considered further. In the next version of the program I will include an option to disable the inner envelope.

[18] The proportion $1 - p$ of points is the total excluded from the 2-component reference envelope or 'doughnut'; $\frac{1-p}{2}$ are excluded from outside the doughnut and the same proportion from inside the doughnut.

**Calculating the PI and the statistics** in `PIcalc.m`. The value of the plankton index, for a particular set of comparison data, is the proportion of new points plotting between the reference envelopes:

$$PI = \frac{n}{N} \tag{1}$$

where $N$ refers to total new points and $n$ is the number between the inner and outer envelopes. The identification of 'inside' and 'outside' points uses the Matlab function `inpolygon`.

Absent change, the expected outcome is $PI = p$, with $pN$ points between the envelopes. The probability of the actual result (on the null hypothesis of no change) can be calculated by the exact binomial expansion,

$$\sum_{k=N-n}^{k=N} \frac{N!}{j!k!} \cdot p^j \cdot q^k \tag{2}$$

where $q = 1 - p$, $j = N - k$, and the sum includes the probability of all more unlikely events (those with fewer than $n$ points between the envelopes). The factorial calculations use the Matlab function `Nchoosek`.

The $\chi^2$ (with 1 degree of freedom),

$$\frac{((N-n) - qN)^2}{qN} + \frac{(n - pN)^2}{pN} \tag{3}$$

provides an alternative way to estimate the significance of the PI value.

**Finding medians** in `findenv2.m` and `ssplot.m`. These use the Matlab `median` function to find a median value on the lifeform 1 axis (ranking all values on this axis) and on the lifeform 2 axis; together, these medians give the co-ordinates of the cloud centre, required in `findenv2` for data manipulation, and in `ssplot` as state-space co-ordinates to be plotted. These medians are unaffected by any of the data transformations used.[19]

---

[19] The formal requirement is that the transformation does not re-order magnitudes.

# 14   Interpretations

In the preceeding text I have commented in several places about what certain features in PI diagrams or values imply about the PI method or about the methods used to collect data to input to the PI tool. For example, a PI diagram in which many points have plotted onto the min lines implies that insufficient organisms have been counted during sample analysis.[20] The volume of water sedimented for inverted microscopy, or the length of a plankton tow, may be too small. Or, the lifeform might be a poor choice, containing only rare species – or, at least, species that appear infreqently in samples,

However, the core purpose of the PI (as a number and as a method) is to illuminate the behaviour of the target ecosystem. For example,

**PI diagram** can often be interpreted in terms of change. Figure 3 shows significant change with a PI of 0.47. The abundance of Si-users (mainly diatoms) has decreased during late winter and Spring, and there is some evidence of increase in the summer biomass of non-Si-users (flagellates and dinoflagellates). The latter might have been expected from the development, subsequent to the reference period, of a nutrient-enriching salmon farm in the loch.

**Change** in PI may be related to time-series of pressure. The time-series in figure 4 shows movement of the state of the lagoon of Thau from an arbitrary reference condition; this change has been interpreted (Gowen et al., 2015) as the lagoon returning to better condition as nutrient loading, and consequent eutrophication, has reduced. The current UK MSFD monitoring procedure (Scherer et al, 2015) uses this approach: if a trend in PI correlates with a trend in a manageable pressure, such as nutrient loading, then further investigation will be carried out, leading to a *programme of measures* to deal with the pressure if the relationship is shown to be causal.

**GES** (Good Environmental Status) is defined by the MSFD as that of a fully-functioning marine ecosystem; we should like to know what the reference envelopes for pelagic habitat lifeforms looked like for GES - i.e. where in state-space they were situated. Given such information for each ecohydrodynamic regime, PI values could be interpreted as showing, or not showing, GES. However, we don't yet have such information.[21]

---

[20] Include an example of such a diagram in the next edition.

[21] But work is in hand to find coastal waters in GES and establish reference envelopes for them.

# 15   Under development

This section lists features that are currently being added to the program but are not available in the current 'published' version PI2015_Oct_beta. They are available in version PI2016_Oct_beta.

### Assessing time-series

Function ssplot2 has been modifed to ssplot2b to assess the Euclidian distance time-series using Kendall's tau; if the series is significant, a polynomial is fitted and plotted.

Function PITS has been modifed to PITSb to assess the PI time-series using Kendall's tau; if the series is significant, a polynomial is fitted and plotted.

Both call the new function TStrend to fit the polynomial, and this in turn calls Kendl1 (based on the Press et al. 1989 function) to calculate tau.

This behaviour can be switched as follows in cf4

```
% select order of polynomial to fit time-series and plot
% cp.poly = 0 does not fit or plot
% cp.poly = 1 fits linear trend
cp.poly = 3; % order 3 polynomial is usually best
%
```

# 16 References

- Anon (2004). *Understanding of Undesirable Disturbance in the Context of Eutrophication, and Development of UK Assessment Methodology for Coastal and Marine Waters: Stage 1 Report: What is Undesirable Disturbance? Stage 2 Report: Measuring Undesirable Disturbance.* Napier University, Edinburgh.

- Bowman, M. J., W. E. Esias & M. B. Schnitzer (1981). Tidal stirring and the distribution of phytoplankton in Long Island and Block Island Sounds. Journal of Marine Research, 39, 587-603.

- COM (2010). COMMISSION DECISION of 1 September 2010 on criteria and methodological standards on good environmental status of marine waters (2010/477/EU). Official Journal of the European Union, L 232(2.9.2010), 14-24.

- Folke, C., S. Carpenter, B. Walker, M. Scheffer, T. Elmqvist, Lance Gunderson & C. S. Holling (2004). Regime shifts, resilience, and biodiversity in ecosystem management. Annual Review of Ecology Evolution and Systematics, 35, 557-581.

- Follows, M. J. & S. Dutkiewicz (2011). Modeling Diverse Communities of Marine Microbes. Annual Review of Marine Science, 3, 427-451.

- Gowen, R. J., A. McQuatters-Gollop, P. Tett, M. Best, E. Bresnan, C. Castellani, K. Cook, R. Forster, C. Scherer & A. Mckinney (2011). *The Development of UK Pelagic (Plankton) Indicators and Targets for the MSFD. A report of a Defra workshop held at AFBI 2nd - 3rd June 2011.* 41 pp. Belfast, AFBI.

- Hooper, D. U., F. S. Chapin, et al. (2005). Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. Ecological Monographs, 75, 3-35.

- Jones, K. J., R. J. Gowen & P. Tett (1984). Water-column structure and summer phytoplankton distribution in the Sound of Jura. Journal of Experimental Marine Biology and Ecology, 78, 269-289.

- Le Quéré, C., S. P. Harrison, et al. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. Global Change Biology, 11, 2016-2040.

- Litchman, E. & C. A. Klausmeier (2008). Trait-Based Community Ecology of Phytoplankton. Annual Review of Ecology, Evolution, and Systematics, 39, 615-639.

- Litchman, E., K. F. Edwards, C. A. Klausmeier & M. K. Thomas (2012). Phytoplankton niches, traits and eco-evolutionary responses to global environmental change. Marine Ecology Progress Series, 470, 235-248.

- Lindeman, R. A. (1942). The trophic-dynamic aspect of ecology. Ecology, 23, 399-417.

- Mageau, M. T., R. Costanza & R. E. Ulanowicz (1995). The development and initial testing of a quantitative assessment of ecosystem health. Ecosystem Health, 1, 201-213.

- Margalef, R. (1978). Life forms of phytoplankton as survival alternatives in an unstable environment. Oceanologica Acta, 1, 493-509.

- Pingree, R. D., P. M. Holligan & G. T. Mardell (1978). The effects of vertical stability on phytoplankton distributions in the summer on the northwest European Shelf. Deep Sea Research, 25, 1011-1028.

- Raunkær, C. (1904) Om biologiske Typer, med Hensyn til Planternes Tilpasninger til at overleve ugunstige Aarstider. Botanisk Tidsskrift 26, p. XIV.

- Raunkiær, C. (1934) *The Life Forms of Plants and Statistical Plant Geography, being the collected papers of C. Raunkiær.* Oxford University Press, Oxford. Reprinted 1978 (ed. by Frank N. Egerton), Ayer Co Pub., in the *History of Ecology* Series.

- Scherer, C., R. J. Gowen, P.Tett, A. Atkinson, M. Baptie, M. Best, E. Bresnan, K. Cook, R. Forster, S. Keeble & A. McQuatters-Gollop (2014). Development of a UK Integrated Plankton Monitoring Programme: A final report of the Lifeform and State Space project. London, The Department of Environment, Food and Rural Affairs.

- Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. New York, McGraw-Hill Book Company, Inc.

- Smayda, T. J. & C. S. Reynolds (2001). Community assembly in marine phytoplankton: application of recent models to harmful dinoflagellate blooms. Journal of Plankton Research, 23, 447-461.

- Smetacek, V. S. (1986). Impact of freshwater discharges on production and transfer of materials in the marine environment. *The role of freshwater outflow in coastal marine ecosystems.* S. Skreslet (ed). Berlin, Springer-Verlag. vol.: 85-106.

- Tett, P. & Grantham, B. (1980). Variability in sea-loch phytoplankton. Fjord Oceanography. D. H. Freeland, D. M. Farmer and C. D. Levings (ed). New York, Plenum. 435-438.

- Tett, P. (2006). *Using the PCI-LF - a draft user guide.* Report. Napier University, Edinburgh.

- Tett, P., R. Gowen, D. Mills, T. Fernandes, L. Gilpin, M. Huxham, K. Kennington, P. Read, M. Service, M. Wilkinson & S. Malcolm (2007). Defining and detecting Undesirable Disturbance in the context of Eutrophication. Marine Pollution Bulletin, 55, 282-297.

- Tett, P., C. Carreira, D. K. Mills, S. van Leeuwen, J. Foden, E. Bresnan & R. J. Gowen (2008). Use of a Phytoplankton Community Index to assess the health of coastal waters. ICES Journal of Marine Science, 65, 1475-1482.

- Tett, P., R. Gowen, et al. (2013). A framework for understanding marine ecosystem health. Marine Ecology Progress Series, 494, 1-27.

- Tett, P. (2015). Guide to the Plankton Index method and software, v.3.01. Scottish Association for Marine Science, Oban. 27 pp + appendices.

- van Leeuwen, S. M., P. Tett, D. K. Mills & J. van der Molen (2015). Stratified areas in the North Sea: long-term variability and biological and policy implications. Journal of Geophysical Research, 120, do1: 10.1002/2014JC010485.

- von Bertalanffy, L. (1972). The history and status of General Systems Theory. The Academy of Management Journal, 15(4), 407-426.

APPENDICES

# A UK lifeforms for the MSFD 'pelagic habitat'

The UK *Lifeforms project*, funded by Defra and led by Richard Gowen, AFBI, brought together an expert group to agree sets of lifeform pairs for monitoring plankton in relation to MSFD *Qualitative Descriptors* (QDs) relevant to GES in the *pelagic habitat*. The tables in this section are modified from Table 3.2 in Scherer, Gowen et al. (2014).

Table 1: Lifeform pairs for **descriptor 1, Biodiversity**.

| | *Lifeform pair* | | *Reasoning* | *Pressures* |
|---|---|---|---|---|
| 1 | **Diatoms** (all) | **Dinoflagellates** (all) | evolutionarily distinct groups with different attributes and biology | nutrient enrichment; changes in EHDC |
| 2 | **'jellies'** (pelagic ctenophora and cnidaria) | **fish larvae** (including eggs) | indicators of alternative ecosystem states (and potential services in food provision) | fishing |
| 3 | **holoplanktonic crustacea** (not eggs) | **other holoplankton** (not 'jellies', crustacea, eggs) | evolutionarily distinct groupings that capture holoplankton features not included in pair 2 | fishing; nutrient enrichment |

Table 2: Lifeform pairs for **descriptor 4, Food-Webs**.

| | *Lifeform pair* | | *Reasoning* | *Pressures* |
|---|---|---|---|---|
| 1 | **Phytoplankton** (as chlorophyll or PCI) | **Zooplankton** (all) | shows the magnitude of energy flows and the balance between adjacent trophic levels | fishing |
| 2 | **large phytoplankters** ($\geq 20\mu m$) | **small phytoplankters** ($< 20\mu m$) | relates to the efficiency of energy transfer to higher trophic levels (smaller phytoplankters $\rightarrow$ microplankters) | fishing |
| 3 | **large copepods** ($\geq 2mm$) | **small copepods** ($< 2mm$) | food web structure and energy transfer (smaller less efficient) | fishing; nutrient enrichment; change in stratification |

Table 3: Lifeform pairs for **descriptor 5, Eutrophication**.

| | Lifeform pair | | Reasoning | Pressures |
|---|---|---|---|---|
| 1 | **Diatoms** (all) | **Dino-flagellates** (autotrophs and myx-otrophs only) | shift in community composition ('balance of organisms') | nutrient enrichment |
| 2 | **Ciliates** (including tintinnids) | **Micro-flagellates** ($< 20\mu m$) | shift from dominantly primary production to more heterotrophic ecosystem | nutrient enrichment |
| 3 | **Toxin-producing diatoms** (i.e. *Psudo-nitzscia* spp.) | **Toxin-producing dinoflag-ellates** (all species on FSA list) | shift in algal community structure towards dinoflagellate HABs | nutrient enrichment |

Table 4: Lifeform pairs for **descriptor 5, Sea floor integrity**.

| | Lifeform pair | | Reasoning | Pressures |
|---|---|---|---|---|
| 1 | **Holoplankton** (all) | **Meroplankton** (excluding fish larvae) | strength of benthic-pelegic coupling | bottom trawling, dredging |
| 2 | **Pelagic diatoms** | **Tychopelagic diatoms** | effects of seabed disturbances and resuspension events | climate change |
| 3 | - | - | - | - |

There is a UK master list for assigning species to these lifeforms. Finally, an overall evaluation of the state of the pelagic habitat (QD1.7) can be made by combining all PI values.[22]

---

[22] At present we're proposing that this should be done by simple averaging of PIs generated from all available lifeform pairs. As noted earlier, however, it might be desirable to arc-sine transform the PI values first. I'm planning to investigate this by fitting a Convex hull to a 4-D group of points in order to estimate PI values for tetraplets of lifeforms.

# B   Suggestions for lifeforms

As stated in the main text, the choice of lifeforms is currently a matter for expert judgement rather than application of a single widely-accepted theory. Thus, more than one set of choices is possible, depending on the expert and the problem addressed. The UK *Lifeforms* project has proposed a set of planktonic lifeforms for the monitoring of the status of the *pelagic environment* in the context of the EU Marine Strategy Framework Directive (Gowen et al., 2011). Those lifeforms were listed in appendix A.

The set proposed in this Appendix grew out of discussions in that project, but differs in several ways. It focuses on **microplankton** and hence does not include multicellular zooplankton. It is a personal suggestion, not a consensus list. It is intended to provide an comprehensive list for the pelagic environment, irrespective of locality and practicality (although excepting waters containing resuspended microphytobenthos). Users could chose a sub-set appropriate to their available data and the issues to be addressed in a lifeforms & state-space approach.

## General key

**Sizes** - important in trophic relationships; ESD = 'equivalent spherical diameter'

*Large* (greater than 20 $\mu$m ESD, perhaps in chains)

*Medium*: 5 - 20 $\mu$m ESD: the smallest size group that can be reliably identified and counted using conventional light microscopy

*Small*: 2 - 5 $\mu$m ESD

*Very small*: less than 2 $\mu$M ESD - the formal definition of picoplankton

**Trophy**

*(Photo)autotrophy* - light driven primary production from inorganic raw materials (bicarbonate, water, inorganic nutrients)

*Myxotrophy* - use, in combination with photosynthesis, of organic sources of C and nutrients; many myxotrophs use phagotrophy to get N and P

*Heterotrophy* - reliance on organic sources of C and nutrients, including particle-feeding (phagotrophy), uptake of dissolved organics, and other methods

**Acronyms**

*SFVT* - shellfish vectored toxins - produced by micro-algae and retained by filter-feeding shellfish

## The lifeforms

These are described in tables 5 to 14. The groupings might be subdivided, especially those that are polyphyletic.

Table 5: Pelagic Diatoms

| Major taxa | Bacillariophyceae |
|---|---|
| Description: | all pelagic centric and pennate diatoms except for (thick-walled) tychopelagic species; most are chain-forming |
| Trophic (etc) role: | medium & large photo-autotrophs that are part of silicon cycle |
| When, where: | most abundant under low-light high-nutrient conditions; their spring bloom in subpolar and temperate waters typically provides the main annual input of new primary production |
| Notes: | typically identified and counted by visible light microscopy |
| Impact: | a few species (e.g. *Pseudo-nitzchia* spp.) contain SFVT |

Table 6: Autotrophic/mixotrophic Dinoflagellates

| Major taxa | Dinophyceae |
|---|---|
| Description: | all armoured and naked dinoflagellates containing chloroplasts |
| Trophic (etc) role: | medium and large photo-autotrophs and myxotrophs |
| When, where: | more abundant in warm, nutrient-depleted waters (i.e. during summer in temperate latitudes); complex life-cycles and migratory behaviour adapt to stratification and local circulation |
| Notes: | typically identified and counted by visible light microscopy |
| Impact: | many species give rise to red tides and to mildly toxic deoxygenating biomass, others contain SVFT |

Table 7: Autotrophic/mixotrophic Microflagellates

| Major taxa | Cryptophyceae, Euglenophyceae, Haptophyta, Raphidophyceae |
|---|---|
| Description: | all unicells ($> 5\mu$m) with flagell-um/a and chloroplast(s), from a variety of micro-algal groups, except calcified forms |
| Trophic (etc) role: | medium photo-autotrophs and myxotrophs |
| When, where: | more abundant in warm, nutrient-depleted waters but may also contribute semi-constant background |
| Notes: | pigment analysis is alternative to identification by light microscopy |
| Impact: | a few species are toxic and give rise to harmful blooms |

Table 8: Coccolithophores

| Major taxa | Haptophyta - Coccolithophyceae |
|---|---|
| Description: | flagellated/non-flagellated unicells with calcarous plates |
| Trophic (etc) role: | medium photo-autotrophs and mixotrophs; their coccoliths are an oceanic C sink |
| When, where: | most abundant in clear ocean or outer shelf waters, where they may form late spring blooms |
| Notes: | coccoliths need neutral or alkaline preservatives |
| Impact: | highly reflective (harmless) blooms |

Table 9: Colonial Cyanobacteria

| Major taxa | Cyanobacteria |
|---|---|
| Description: | all chain-forming 'blue-green algae' |
| Trophic (etc) role: | medium and large photo-autotrophs, some being N-fixers; because of size and toxicity are grazed by specialists or controlled by phages |
| When, where: | adapted to stratified waters; abundant in Baltic in summer, and in some warm-water upwelling regions |
| Notes: | - |
| Impact: | form reflective near-surface rafts under calm conditions; also 'icecream-coloured water' when cells decay; some species are toxic |

Table 10: Picophytoplankton

| Major taxa | Cyanobacteria, Prasinophyceae |
|---|---|
| Description: | all unicells ($< 5\mu$m) with chlorophyll; includes green and blue-green cyanobacteria ($< 2\mu$m) and coccoid and flagellated eukaryotic algae ($1 - 5\mu$m) |
| Trophic (etc) role: | small and very small photo-autotrophs (some of the eukaryotes may be myoxtrophic) |
| When, where: | picophytoplanktonic cyanobacteria are a semi-constant background in waters $> 6°$C and are an important component of production under oligotrophic conditions |
| Notes: | best analysed by flow cytometry, or by pigment analysis on fraction passing $5\mu$m filter pores |
| Impact: | some eukaryotic picoplankters have formed toxic blooms inshore |

Table 11: *Myrionecta*

| Major taxa | Ciliata - cyclotrichida |
|---|---|
| Description: | cyclotrich ciliate with symbiotic cryptomonads |
| Trophic (etc) role: | medium-large myxotroph and photo-autotroph; used as a source of chloroplasts by the dinoflagellate *Dinophysis* |
| When, where: | - |
| Notes: | easily identified by light microscopy in fresh Lugol-preserved samples |
| Impact: | can form (harmless) red tides |

Table 12: Oligotrich Ciliates

| Major taxa | Ciliata - oligotrichida |
|---|---|
| Description: | all oligotrich ciliates including tintinnids |
| Trophic (etc) role: | medium-large phagotrophs (as a group, feeding on most other protist groups and on eubacteria) |
| When, where: | - |
| Notes: | recognisable collectively by light microcopy in lugol-preserved samples; lower level taxa easy for tintinnids, difficult for non-cased oligotrichs |
| Impact: | - |

Table 13: Heterotrophic Dinoflagellates

| Major taxa | Dinophyceae |
|---|---|
| Description: | all dinoflagellates without chloroplasts |
| Trophic (etc) role: | medium-large predators on most other protist groups, using a range of feeding methods |
| When, where: | roughly the same as Dinoflagellates AM |
| Notes: | typically identified and counted by light microscopy |
| Impact: | might carry toxins from prey |

Table 14: Heterotrophic Microflagellates

| | |
|---|---|
| *Major taxa* | Choanoflagellidea, Kinetoplastea, Bicosoecophyceae, class containing *Solenicola* |
| *Description:* | all flagellated unicells without chloroplasts (some giving rise to hyphae) |
| *Trophic (etc) role:* | small-medium phagotrophs (on smaller algae and bacteria) and parasites (on diatoms especially) |
| *When, where:* | ? |
| *Notes:* | poorly known, with many $< 5\mu$m, generally difficult to identify reliably except in a few cases (eg. Choanoflagellates); this life-form should perhaps be split between parasitical and free-living microheterotrophs |
| *Impact:* | - |

# C   Why a doughnut-shaped envelope?

*This is a re-arrangement of:* Tett, P. & D. K. Mills (2009). Patterns in phytoplankton life-form space: ring doughnuts? Poster: ASLO Aquatic Sciences Meeting 2009, Nice, France. *The text uses the old acronym of PCI; state-space plots were made with software that did not exclude outlying points from the envelope.*

---

**1.** Phytoplankton Community Indices (or PCIs) belong to a set of 'tools' for quantifying the phytoplankton biological quality element of Annex V of the Water Framework Directive and for monitoring against 'undesirable disturbance' of UK marine waters (Tett et al., 2007). The 'LifeForm' PCI rests on two concepts. The first is the idea that the states of an ecosystem can be shown by plotting their co-ordinates in a space defined by a number of state variables.



---
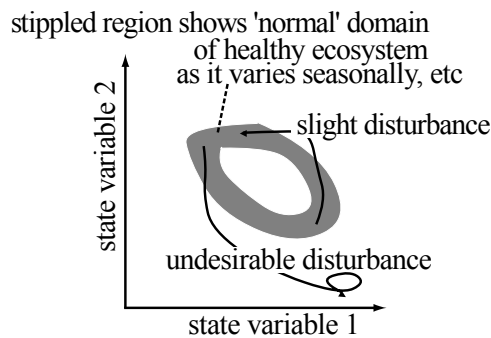
**2.** The second idea is that suitable state variables are the abundances of groups of species of similar lifeform (Margalef, 1978). Seasonal variation in lifeform abundance is characteristic of phytoplankton in temperate waters, and should be taken into account in judging ecosystem state.



---
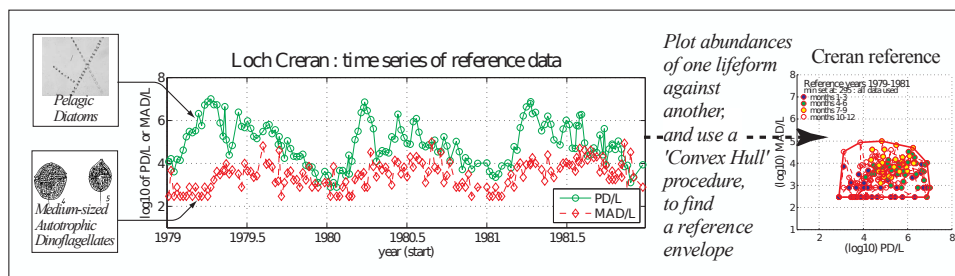
3. In order to assess change from a reference condition, new data are plotted over the reference envelope.

$$PCI = \frac{\text{new points inside envelope}}{\text{total new points}}$$

A value of: 1 = no change; 0 = complete change. (Tett et al., 2008) *[This example uses data from Loch Creran in western Scotland]*
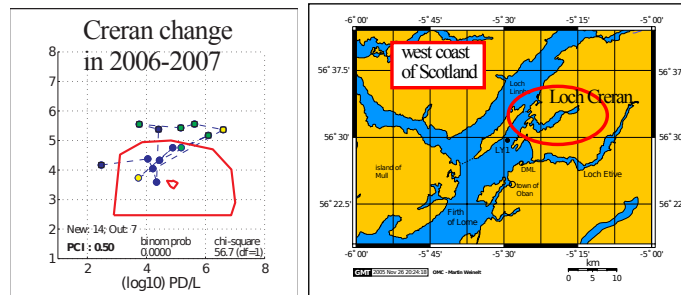


A second example uses Continuous Plankton Recorder data from SAHFOS. Notice that there is a bigger 'hole' in the 'doughnut', which leads to a more sensitive measure of change.



_____ .
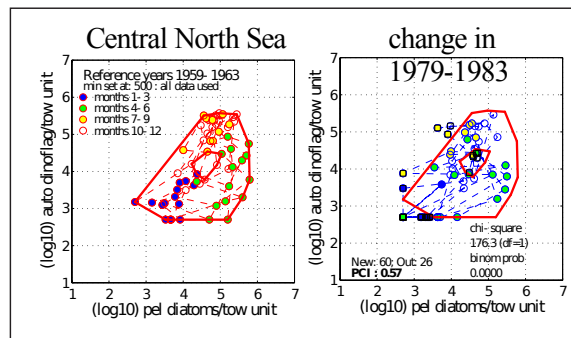
4. Should we expect such clouds of points to have an empty centre? – should there always be a 'hole' in the dougnut ? Here is a theory involving mapping from an 'environmental' to a 'community' space.
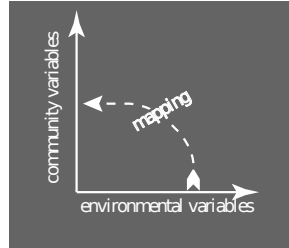


(i) If a variable $y$ is a function of time with a partly-regular seasonal component, it can be seen as a truncated Fourier Series plus an error term:

$$y = f(t) = a_0 + \sum_n [a_n \cdot \sin(nt) + b_n \cot \cos(nt)] + \epsilon$$

where:

$a_0$ is the mean value of the series;
$a_n$ and $b_n$ are coefficients;
$\epsilon$ is a stochastic term corresponding to the non-seasonal part of $y$;
$n = 1$ to $m$; $m = 3$ in the present case;
$t = 2\pi \cdot day/365$ (thus, the wave with $n = 1$ has a period of one year);

(ii) Environmental conditions are defined by a set $\{y_e\}$ of environmental variables and community state is defined by a set $\{y_b\}$ of biological variables. Specific growth rate is an example of a function that transforms an environmental set (irradiance $E$ and nutrient concentration $S$) into a community set (the biomasses $y_i$ of species $i$):

$$\{y_e\} \rightarrow \{y_b\} : \mu_i = f_i(E, S) \text{ and } y_{i[t+\Delta t]} = y_{i[t]} \cos e^{\mu_i \Delta t}$$

(iii) Integrating a sine-cosine function causes a phase shift; the integration of growth rate results in (potentially) exponential growth; and hence the transformational function ought to result, for each lifeform, in a phase shift of the seasonal cycle that is specific to that lifeform, and hence a graph that is somewhat circular when plotted on a logarithmic axis.

―――――――――――――――――――――――――――――――――――――――――――.

5. This is demonstrated for loch Creran. Truncated Fourier Series (TFS) have been fitted to these environmental time-series, and the TFS plotted against each other in 'environmental state space'.
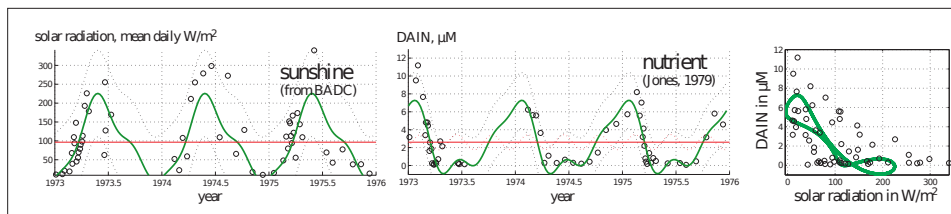


6. Community state in loch Creran: TFS fitted to life-form time-series, and life-form state-space plots
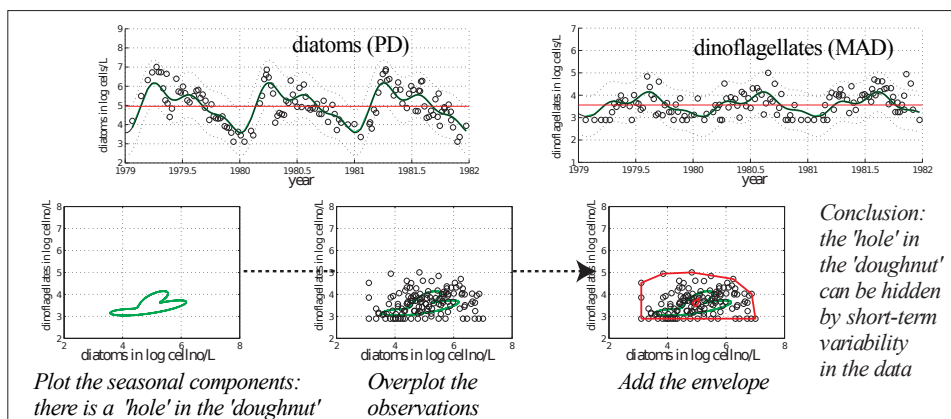


## References

**Jones, K.J. (1979)** Studies on nutrient levels and phytoplankton growth in a Scottish Sea Loch. PhD. University of Strathclyde.

**Margalef, R. (1978)** Life forms of phytoplankton as survival alternatives in an unstable environment. Oceanologica Acta, 1, 493-509.

**Tett, P.,** Carreira, C., Mills, D.K., van Leeuwen, S., Foden, J., Bresnan, E., & Gowen, R.J. **(2008)** Use of a Phytoplankton Community Index to assess the health of coastal waters. ICES Journal of Marine Science, 65, 1475-1482.

**Tett, P.,** Gowen, R., Mills, D., Fernandes, T., Gilpin, L., Huxham, M., Kennington, K., Read, P., Service, M., Wilkinson, M., & Malcolm, S. **(2007)** Defining and detecting Undesirable Disturbance in the context of Eutrophication. Marine Pollution Bulletin, 55, 282-297.