

Time series analysis with Bilko Batch

C-RISe Lesson 1A

Using a monthly SST climatology to analyse seasonal variability

Author: Val Byfield, National Oceanography Centre, U.K.

Introduction

A climatology is the long-term average of a given variable, usually calculated over 30 years, provided a time series of this length is available. For example, a monthly climatology will produce a mean value for each of the twelve months; a weekly climatology will produce a mean value for each of the 52 weeks of the year. Climatologies show the typical conditions to be found in a region of interest at different times of year, and are useful for understanding seasonal changes.

This short tutorial demonstrates how to use Bilko in Batch mode by applying a formula document to a Bilko SET – a list of monthly sea surface temperature (SST) data from 2003 to 2017.

Learning objectives

By the end of the lesson you should be able to

- Identify the different components of a Bilko formula designed to calculate a monthly climatology and understand how they work.
- Use the Windows Command prompt and Bilko to create a SET to list the input data.
- Run the formula in Batch mode to create the climatology data set for a selected region.
- Use transects and data cores to analyse the seasonal variability seen in the output data.

The data set

The 1km resolution data set used in the lesson was provided by Plymouth Marine Laboratory as a contribution to the SOLSTICE project (www.solstice-wio.org). It covers the Western Indian Ocean, and was extracted from global Multi-scale Ultra-high Resolution (MUR) Sea Surface Temperature (SST) analysis data from NASA JPL PO.DAAC (<https://earthdata.nasa.gov/eosdis/daacs/podaac>) and covers a time period of 15 years from 2003 to 2017.

Content

1. Opening and examining the Bilko formula
2. Using the Windows Command prompt to rename multiple files
3. Creating a Bilko set of 2 files and examining it to understand the structure of set files
4. Creating a file list in the Command Window
5. Editing the Bilko set to include all the files to use in the calculations
6. Running the Bilko formula and set in Batch mode to carry out the calculations
7. Displaying the climatology to reveal seasonal and variability in the chosen area.

Disk space required

This lesson and the associated lesson on interannual variability requires a minimum of 20 GB of free space to complete.

The following data and tools should be downloaded and extracted to a single working directory:

- 180 NetCDF files '**Large_MURSST_YYYYMM.nc**' – each an SST data set of 86MB
- **mursst_monthly_climatology.frm** - Bilko formula used to calculate a climatology from this data
- **mursst_monthly_anomaly.frm** – Bilko formula used in the next lesson to calculate an anomaly time series
- **sst_rainbow_maxblack.pal** – colour palette used to display the climatology

The calculated climatology will consist of 12 files – requiring 1GB of free space.

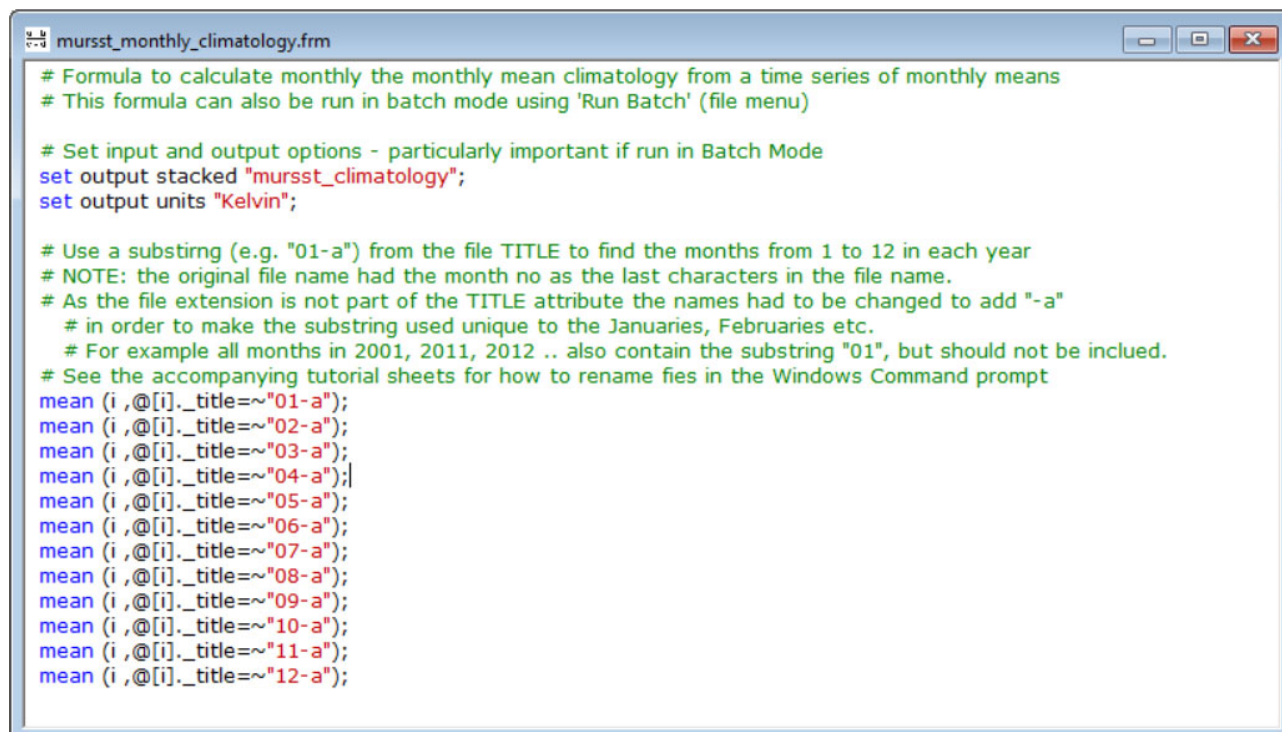
The anomaly data calculated for the Mozambique Channel in the following lesson on interannual variability requires a further 10 GB of free space.

1. Opening and examining the Bilko formula

A Bilko formula (frm) document is a mini-program specifying the calculations that will be carried out on a set of two or more images.

Activity 1

In Bilko, open the formula document **mursst_monthly_climatology.frm** and examine it. (You may need to change the file type to FORMULAS to see this file.)



```
# Formula to calculate monthly the monthly mean climatology from a time series of monthly means
# This formula can also be run in batch mode using 'Run Batch' (file menu)

# Set input and output options - particularly important if run in Batch Mode
set output stacked "mursst_climatology";
set output units "Kelvin";

# Use a substring (e.g. "01-a") from the file TITLE to find the months from 1 to 12 in each year
# NOTE: the original file name had the month no as the last characters in the file name.
# As the file extension is not part of the TITLE attribute the names had to be changed to add "-a"
# in order to make the substring used unique to the Januaries, Februaries etc.
# For example all months in 2001, 2011, 2012 .. also contain the substring "01", but should not be included.
# See the accompanying tutorial sheets for how to rename files in the Windows Command prompt

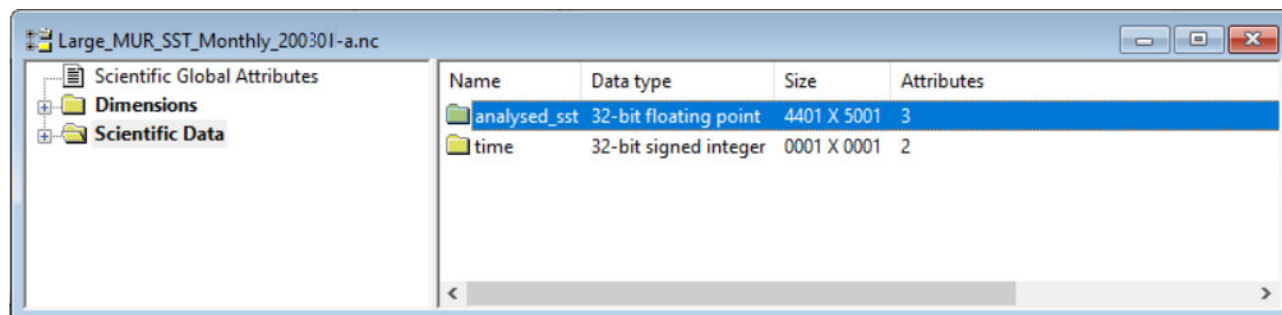
mean (i ,@[i]._title=~"01-a");
mean (i ,@[i]._title=~"02-a");
mean (i ,@[i]._title=~"03-a");
mean (i ,@[i]._title=~"04-a");
mean (i ,@[i]._title=~"05-a");
mean (i ,@[i]._title=~"06-a");
mean (i ,@[i]._title=~"07-a");
mean (i ,@[i]._title=~"08-a");
mean (i ,@[i]._title=~"09-a");
mean (i ,@[i]._title=~"10-a");
mean (i ,@[i]._title=~"11-a");
mean (i ,@[i]._title=~"12-a");
```

Text shown in green and preceded with **#** are **comments**. These are meant to explain to a human reader what the formula does, and are not read by the software.

Text in blue are **calls** to functions in the software; these lines are read because they do not start with a **#**. The black text specifies the data that will be affected by the function. Red text represents text **strings** (short bits of text that can be part of a file name, or related to the metadata included in a file).

The actual formula commands starts by setting the input and output options. In case there are no commands specifying the input because the software will read the necessary information from the metadata in the input files, which you can examine from the Bilko file structure window.

Open the file **WIO_MUR_Monthly_200301.nc**



The scientific global attributes usually gives general information about how the data was collected and processed. The **Dimensions** (latitude & longitude) tells you how to place the data in space (and time). **Scientific Data** contains the actual data values. Here there is only set of image data - **analysed_sst** which is provided using 32-bit floating point as the numerical format. The image size 4401 pixels in the X dimension and 5001 pixels in the Y dimension; these link back to the longitude (X) and latitude (Y) data under **Dimensions**. The SST data has 3 associated **attributes**. To see these, **right-click** on the **analysed_sst** folder in the right frame and select Open Properties.

The most important information here is that the Fill value (**Null** value) has been set to NaN (Not a Number). Since it is specified in the data attributes, file we do not need to specify this in the formula document*.

*Many files don't specify the Null value in the metadata, but rather in an accompanying README file or other data documentation. Where no null value is specified in the metadata, it is VERY IMPORTANT to set the nulls at the top of the formula document – otherwise the pixels with missing data will be included in the calculations. To do so, you use the phrase “**set nulls on**” followed by the number assigned to the missing data (e.g. due to land or cloud). Bilko defaults to using the same null values in the output as was given in the input, unless something else is specified. To specify output null values you use the phrase “**set nulls out**” followed by nan or the number you want to use (for example -999). Bilko Help gives more information about how to set input and output as part of the description of formulas – see *Help > Contents > Document types > Formulas* (near the bottom of the page).

Returning to the formula, you will see two lines that specify the **output**:

```
set output stacked "mursst_climatology";
set output units "Kelvin";
```

The first line makes sure the output is collected into a Bilko **set** named **mursst_climatology.set**, and that all the component images are given a common root name (mursst_climatology) followed by a 3 digit number – in this case corresponding to the month number. (Without this line the formula would produce 12 individual images called “Image” followed by a number.)

The second line sets the output units to degrees Kelvin. The units could have been specified in the NetCDF data attributes, but since it was not, we take the opportunity to do it here.

The final part of the formula document contains the climatology calculations, one line per month.

```
mean (i ,@[i]._title=~"01-a");
mean (i ,@[i]._title=~"02-a");
mean (i ,@[i]._title=~"03-a");
mean (i ,@[i]._title=~"04-a");
mean (i ,@[i]._title=~"05-a");
mean (i ,@[i]._title=~"06-a");
mean (i ,@[i]._title=~"07-a");
mean (i ,@[i]._title=~"08-a");
mean (i ,@[i]._title=~"09-a");
mean (i ,@[i]._title=~"10-a");
mean (i ,@[i]._title=~"11-a");
mean (i ,@[i]._title=~"12-a");
```

The **mean** function calculates the statistical mean at each pixel location based on the input images specified in the text in brackets. As you can see there is a line for each month – each acts as a command to the software, and ends in a semi-colon.

The semicolon signals the end of a command. Missing a semi-colon leads to a “Syntax error” which Bilko will point out when you apply the formula. Missing semi-colons is one of the most common reasons why a formula is not working.

The letter **i** inside the square brackets is a counter, which is also used to refer to any image **@[i]** in the stack.

@[i]._title

refers to the **title attribute** of the image **@[i]** which has a location in the stack corresponding to the number **i**. This is used to decide if an image should be included in the calculations or not.

=~"01-a"

refers to all the files that contains the substring **“01-a”** in the file name (or title). The output of this command is a single image containing the mean temperature for all the Januaries.

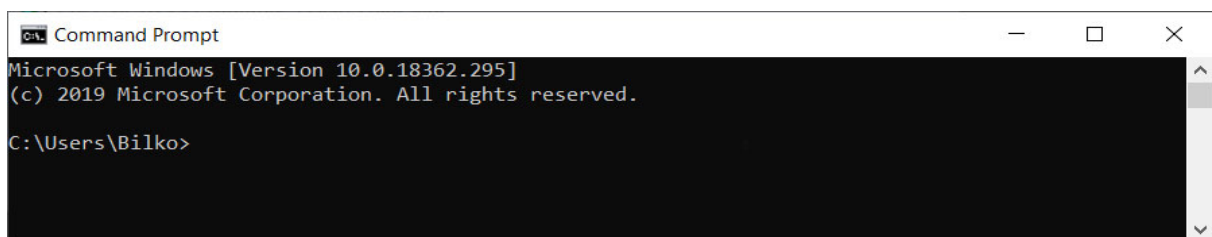
Selecting files based on the **title** is what allows you to select all files with month number 01 to calculate the climatology for January, 02 for the February and so on. We can do this because all the files in the time series follow a common naming convention – the string “Large_MUR_Monthly_” where the wildcard * represents the year and month number in the format YYYYMM.

Question 1

Why was it necessary to add -a to the file names in this case? (Note that “.nc” is the file extension for NetCDF file, not part of the file name. Therefore, it is not included in the *title attribute*.)

2. Using the Windows Command prompt to rename multiple files

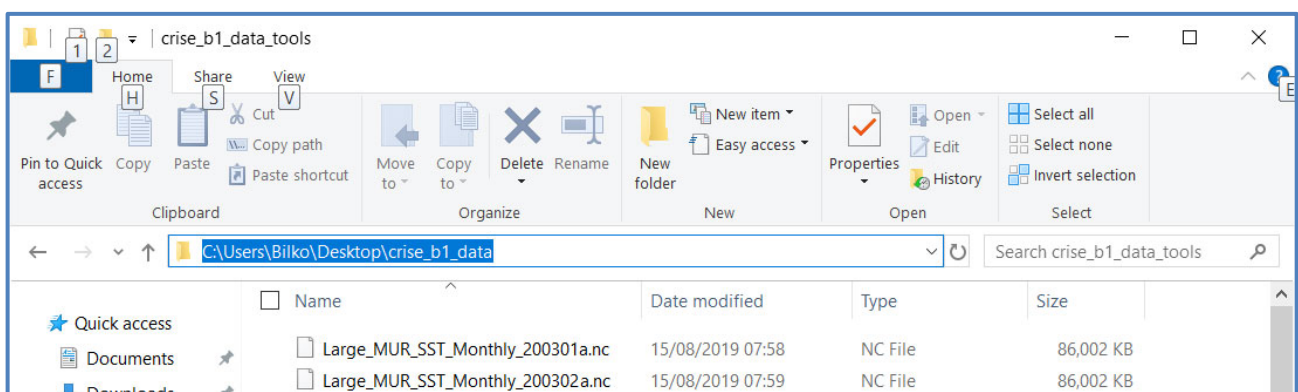
As the original file names do not have “-a” at the end, we need to append this by renaming all the files we use in the calculations. This can be done most effectively using the Windows **Command** prompt (below) - found by typing Command into Windows search.



The Command Prompt will open in your user directory, but you can navigate to the directory that contains the data you want to list by following the steps below.

Activity 2

Find the directory you in File Explorer. To copy the path, simply click in the top address bar and click copy (keystroke [CTRL+C]) – as shown below.



Return to the Command Prompt window. Type **cd**, then paste the path using keystroke [CTRL+V] and press return to move to the selected directory.

Type **dir** to list all the file in the directory

As you can see from the list in the command window, the files do not yet have the appended “-a” needed for turning the month numbers into unambiguous identifiers. To rename the file we use the following command (see bottom line in the Command Prompt window above.)

```
ren *.nc Large_MUR_SST_Monthly_?????-a.nc
```

Here the **ren** command is followed by an expression identifying the files to be renamed:

```
*.nc
```

where ***** is a wildcard standing in for all the filenames in the directory and **.nc** ensures that only files with that extension are renamed.

The next expression gives the new name of the files:

```
Large_MUR_SST_Monthly_?????-a.nc
```

The **root name** of all the data files is “**Large_MUR_SST_Monthly_**”. This is followed by 6 wildcard characters (**??????**) which represent the year (4 digits) and the month (2 digits). These are unique to a single file, and will remain unchanged. Finally **-a** is the short string that will be appended to the end of each file name.

Type this command into the Command Prompt window and press return.

After the renaming is completed, you can use the **dir** command to check that the files now have a new name – ending in the month number and “-a”.

3. Creating a Bilko set and examining it to understand its structure

A Bilko **set** is a plain text file that tells the software where to find the data you want to access for processing. All images in a set must have the same geographical projection, spatial window and resolution. When combined with a **formula** document a **set** allow users to carry out image processing operations specified in the formula on the data sets specified in the **set**.

The processing may be run interactively, with Bilko holding all the data in memory. However, for large data sets it is better to use Batch mode; this avoids the software crashing because you run out of memory. Batch mode is also the best solution for long time series because you can process up to 4096 images. Interactive processing can only deal with 256 images, and these cannot be large because that would cause memory problems.

The text in a set has a specific structure, and has to be saved as a text file in MSDOS format, with the file extension **.set**. If these conditions are not met, Bilko will not recognize the SET file. The best way to ensure you get this right is to create a small set interactively in Bilko and edit it later to add the full file list.

Activity 3

Return to the Bilko window. If you have not already done so close any files and windows you have opened earlier using **Window > Close All** from the top menu bar.

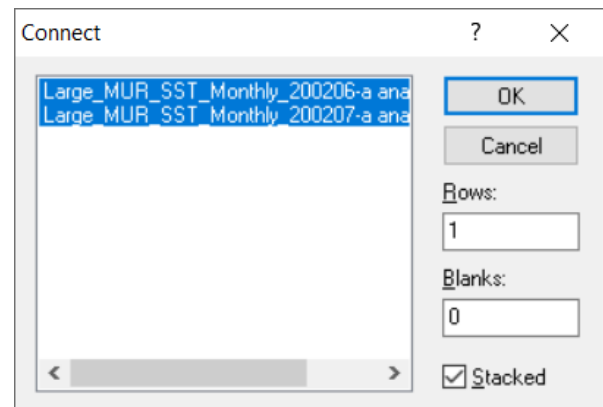
Open the files for January and February 2003:

- From the file structure Window select Scientific Data in the left frame, and double-click on the **analysed_sst** in the right frame to open the first image.
- Set the **Extract box** (below) to open the file at full resolution (Sample 1, 1) and click OK.
- Accept the defaults suggested by the Redisplay dialog to open the first image.
- Repeat for the second image.

Once the two images are open, connect them into a Stack:

- Select [Image > Connect] from the top menu bar will open the Connect dialog (below).
- Select both images,
- Make sure the Stacked box is ticked, and there are 0 blanks, and
- click OK to create the set

From the **File** menu select **Save Set**, and give the set a suitable name, for example **mursst_month_2003_2017.set**.



This new set includes only the first two images in the time 19-year time series. The next step will therefore be to add the remaining 226 images – from Mar'98 to Dec'16. This can be easily done in two steps:

1. Use the Command Prompt to create a file list
2. Edit the Bilko set so it includes the analysed_sst data extracted from all the files in the list

4. Using the Command Prompt to create a file list

Typing all the file names into the set document manually would take a long time. Luckily there is a command that will list all the files of a given type contained in the current directory and write them to a text file that may later be copied and pasted into the set document.

Activity 4

Close Bilko and the set you just created, and return to Command Prompt window.

To list all files in the directory you use the `dir` command, but this time the list should be written to a text file, not to the computer screen. To achieve this type the following:

```
dir *.nc /a /b /-p /o:gen >images.txt
```

Here ***.nc** specifies that you want to list all files with the file extension **nc**, but not any of the other files that may be in directory.

A, B, P, and O are **switches** that modify the output of the **dir** command - preceded by **/**. **/a** specifies all such files ready to be archived. **/b** strips away excess information such as time stamps, leaving only the file names. **/-p** tells the command **not (-)** to pause, but continue to the end of the file list. It is not strictly needed, as the default is to continue without pausing. **/p** is often used with **dir** when you list to the screen, to allow you to see the files before pressing return to continue.

/o:gen orders the output, the letters after the **:** gives the sort order: g – group folders first, e – sort by file extension (file type), n – sort by name. Strictly speaking you don't need g and e here, since *.nc specifies that you only list files with the nc extension (i.e. no folders, no files that are not nc-files).

> tells the software to send the output to somewhere other than the computer screen – in this case to a text file called **images.txt**, which is then saved.

Both the Bilko set you created earlier and **filelist.txt** may be edited in a text editor such as Wordpad. Open these two files in separate windows so you can compare them and copy text from the file list to the set document.

To open the SET document in Wordpad, right-click on the files and **select Open with > Wordpad** from the pop-up menu. Otherwise it will default to Notepad, which is more cumbersome when it comes to finding and replacing text multiple times. When you open the **set**, you are likely to be asked what application to use – Wordpad will be near the bottom of the list of options. Wordpad will most likely wrap the lines of the set, but your edits become easier if the text is not wrapped. Select the **View** tab and uncheck the **Wrap Text** check box.

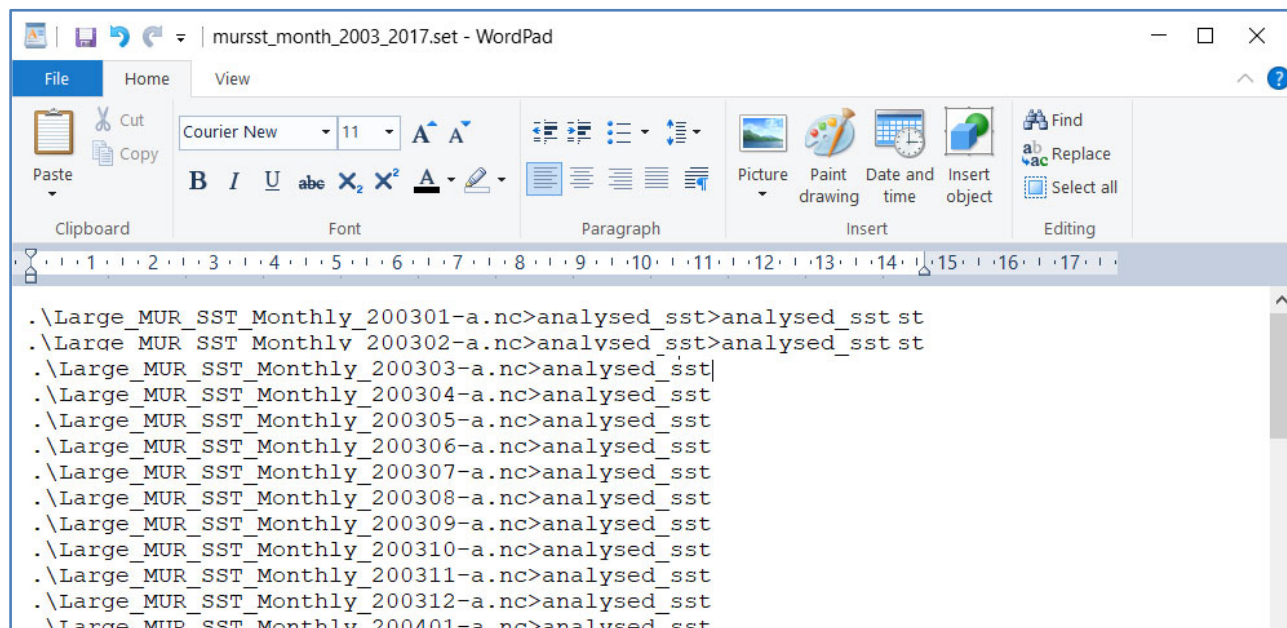
Question 2

What are the main differences between the original Bilko set and the file list you just created?

5. Editing the Bilko set to add all the files from the file list

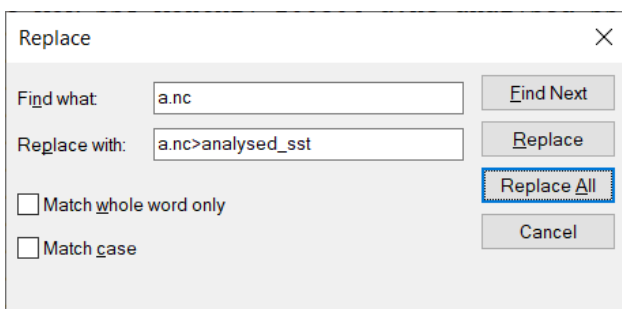
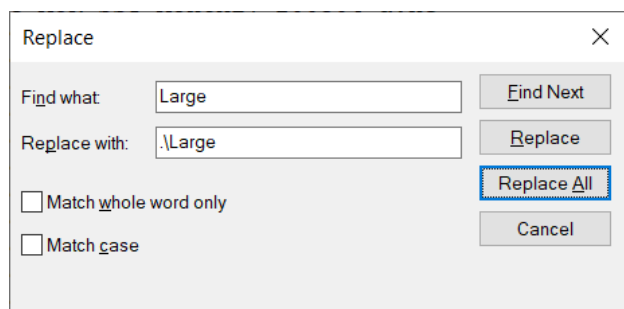
Activity 5

Start by copying the lines from the file list to the SET file.



You can now edit the set using the **Replace** function in Wordpad (menu bar, top right):

- First add `./` to the start of each line (below, left).
- Next, add `>analysed_sst` to the end of each line (above, right).



As you can see Replace all has also added characters at the beginning and end of the first two lines. However, as these are no longer needed, you can delete both lines, making sure you leave no empty lines at the top of the document. (Empty lines act as commands to open a blank image window).

Save the set by clicking Save on the toolbar (keystroke [CTRL+S]).

Windows will ask if you really want to save the file using this format – the answer is yes.

You can use Save As, but then you have to make sure you select MSDOS as the text format. Wordpad will add the extension `.txt` to the end of the set file, and you have to find and remove the extra extension in File Explorer. This is not straightforward because the new `.txt` extension will not be visible (only the first `.set` which has now become part of the name rather than an extension. Confusing? Definitely. You can avoid all that by editing the set (rather than the file list) as shown above, and then just saving the file in the normal way (i.e. don't use Save As). If you want to rename the set, you can do that later in File Explorer.

6. Running the Bilko formula and set in Batch mode

Now that you have both a set document and an associated formula, running the processing in Batch mode is very straight forward.

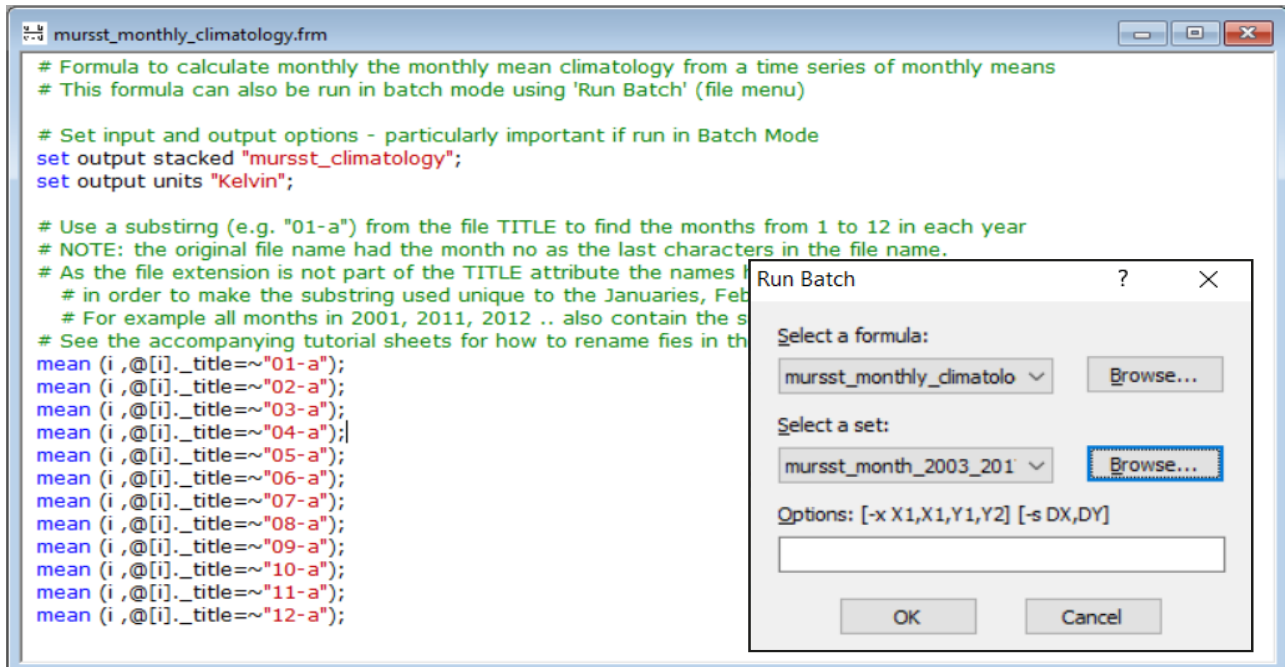
Activity 6

Before you start, it is best to close all but the formula you want to use:

- Click on the formula window to make this the active window
- From the Windows menu select **Close all except**

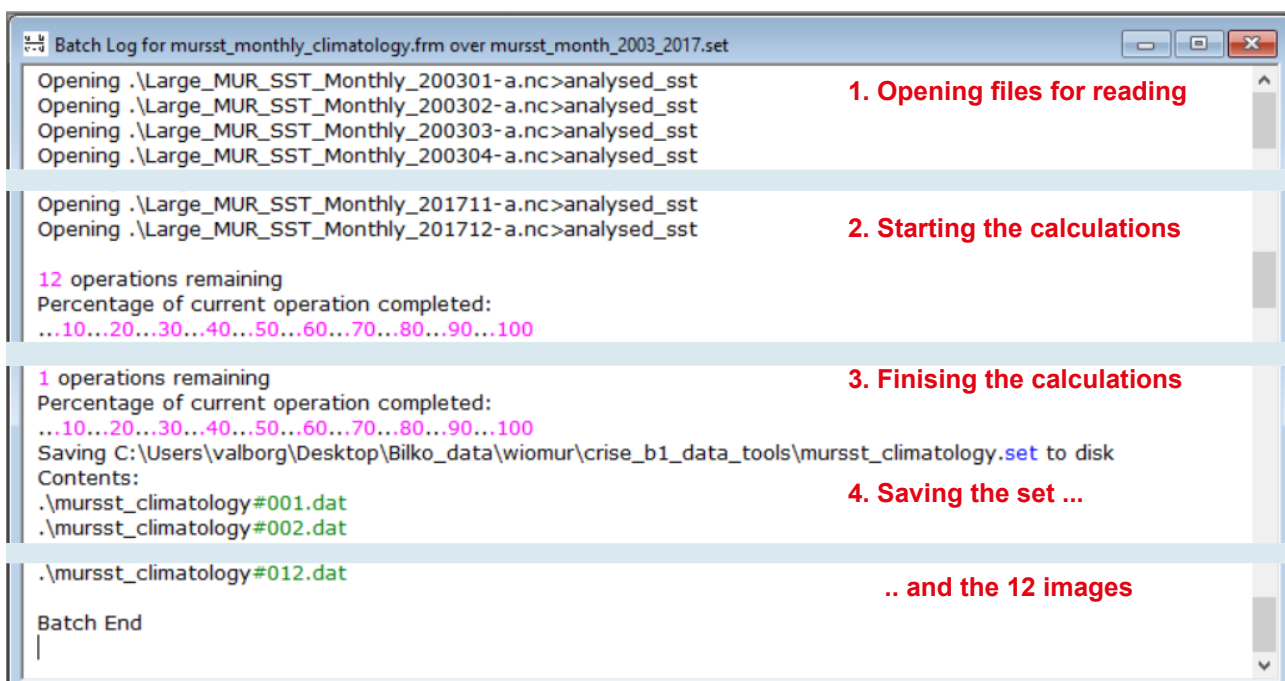
Select **File > Run Batch** from the menu bar to open the Batch dialog (below):

- The name of the formula is already there (if not, use the browse function to select it).
- Select the set you just saved as **occci_month_1998_2016.set**
- Click **OK** to carry out the calculations out the calculations



Do NOT shut down your computer or Bilko. (If it goes to sleep while you are away, the processing will stop, but continue when your computer wakes up again.)

The processor first opens the files for reading. It then carries out the commands of the main calculations – 12 operations, each one line in the formula. Finally it saves the new set and the component images. The batch log (below) will tell you when this process has been completed.



What if the batch does not run?

The main reasons failing to run in batch mode are:

The set has been saved in the wrong format; it MUST be in MS DOS text format. (This is the reason why you copied the file list to the SET file and ignored Wordpad's suggestion to save in another format).

Bilko cannot find the files listed in the SET Document (this could happen if you have made an editing error, or if you have moved any of the images to a different directory from the SET Document).

Error in the formula syntax (you can check for this by activating the formula and using keystroke [CTL+C]; if there is a syntax error, Bilko will alert you, and indicate where it might be).

The batch log can give you an idea of what went wrong. If the set has the wrong format, the batch will not start at all. If one of the files is missing or there is an editor, the first step – opening the files for reading – will halt at the missing file. If formula syntax error in one of the commands will stop the processing at that point.

7. Displaying the climatology to reveal seasonal variability in different locations.

At the end of the batch run, the output set of 12 images will have been saved to the same directory as the formula and given the name specified in the formula output – “**mursst_climatology**”. If all

Activity 7

Close the batch log and open the new set to examine it:

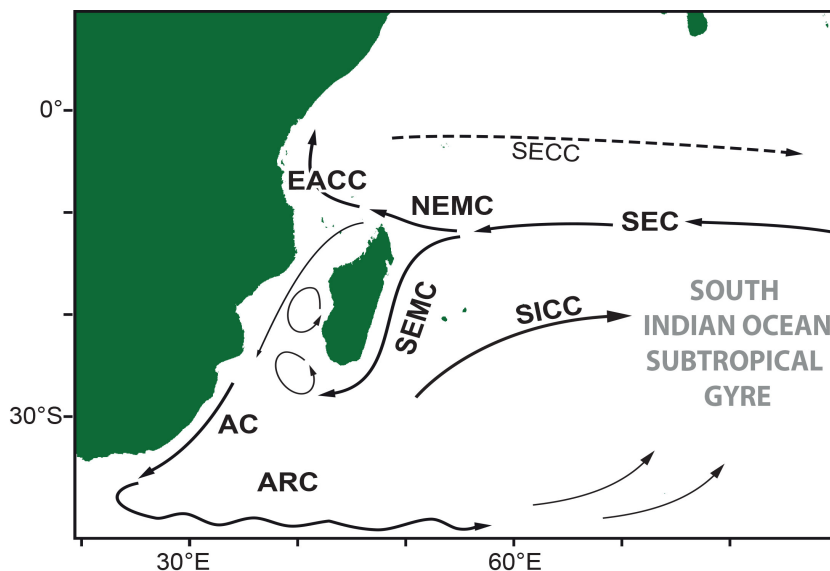
- Select file type SET to see the set to open it.
- Accept the **Extract** default (whole image at full resolution)
- In the **Redisplay** dialog choose a **Linear** stretch with **Min 288.15** and **Max 304.15**
- Zoom out to see the whole image: **View > Zoom > Preserve Shape**.

Apply a colour palette:

- with the image stack active select **File > Open > file type PALETTES**
- open **sst_maxblack.pal**

TAB through the images to explore seasonal variability in the Western Indian Ocean (WIO) region or click the film strip icon on the top tool bar to run the images as an animation.

Use the current map below to help you try to identify different features of the South West Indian Ocean circulation, and note how they change with time



South West Indian Ocean Currents

SEC: South Equatorial Current (SEC),
SECC: South Equatorial Counter-Current (absent during northern summer due to the
SICC: South Indian Ocean Counter-Current,
NEMC: Northeast Madagascar Current
SEMC: Southeast Madagascar Current
EACC: East African Coastal Current
AG: Agulhas Current
ARC: Agulhas Return Current

Thin southward arrow in the Mozambique Channel shows the effects of eddy propagation.

Question 3

- In what month does the southern end of the Agulhas Current (AC) reach its highest temperature? (Taking a **data core** through the set at a selected location may help with this.)
- In what month does the East African Coastal Current and the northern Mozambique Channel reach peak temperatures? (A few data cores may help with this. Alternatively you may take a histograms for a sub-selection of the image containing only this region.)

DATA CORES, TRANSECTS AND HISTOGRAMS IN BILKO

Data cores are cuts through the whole time series taken at a selected locations. They are a common tool for exploring how a parameter changes with time.

1. Select Point Selection (+) from the tool bar.
2. Place the cursor at the point where you want the core to be extracted. (You can also use the GoTo dialog to select a specified location).
3. From the file menu or tool bar open a **new** TRANSECT document (File >New, new icon or keystroke [CTRL+N]), then select TRANSECT
4. In the new file dialog make sure the apply stretches check box is **not** ticked.

The transect shows the image number (month number) along the x-axis, and the temperature along the y-axis.

Transects are lines drawn across a selected feature from one point to another. They are a commonly used to explore features such as fronts, upwelling and eddies.

1. Select Line Selection (\) from the tool bar.
2. Place the cursor at the point where you want the core to be extracted.
3. From the file menu or tool bar open a **new** TRANSECT document.
4. In the new file dialog make sure the apply stretches check box is **not** ticked.

The displayed transect will show the image numbers distance (in pixels) from the starting point along the x-axis and the temperature along the y axis. Each series represent the spatial changes in one of the image of the stack. It is difficult to interpret the transects if you have more than about 4-5 images in the stack, however. For this reason it is often a good idea to export the data to a spreadsheet, and plot the series as and X-Y scatterplot there.

To **export the data series**, simply make the transect the active document. Click copy on the tool bar (or use keystroke [CTL+C]), then paste the data into the spreadsheet ([CTRL+V]).

Histograms can be taken for the whole image or selected regions of interest.

Select region of interest (□) from the tool bar.

Use your mouse to draw a square around the selected region.

Open a **new** file, selecting HISTOGRAM as the document type

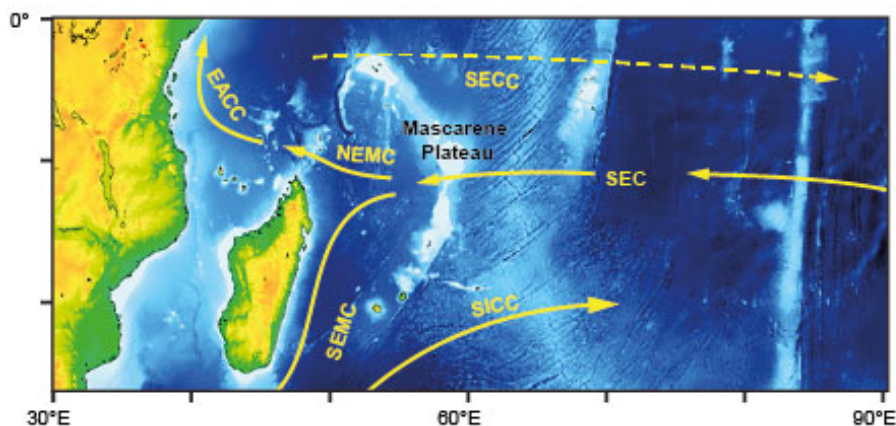
Make sure the Apply Stretches tickbox is **not** checked, and click OK to open the histogram

You can now tab through the stack of histograms to see how the temperature distribution in the region changes from month to month.

Note that Bilko auto-scales the temperatures displayed along the x-axis, so you need to remember this when you compare each month to the one before.

Question 4

- a) As the South Equatorial Current (SEC) approaches northern Madagascar it diverges to form two currents – the Southeast Madagascar Current (SEMC) which flows southward along the east coast of Madagascar and the Northeast Madagascar Current (NEMC). How does the temperature in the NEMC compare with the temperature west of Madagascar at the northern entrance to the Mozambique Channel?
- b) In what month is the front between the NEMC and the Mozambique Channel water most prominent? What is the temperature difference across the front? (Taking a transect across the front, copying this to a spreadsheet to plot the temperature variations may help you with this question).
- c) To the north of the NEMC an area of cooler water is apparent, particularly during the cooler months (June-Aug). How might you explain this phenomenon? (Hints: vertical mixing, current interaction with bathymetry). What may be the reason for the seasonal variation in this feature?



Question 5

- a) How does temperature along the south coast of Madagascar compare with the temperature of the South-East Madagascar Current (SEMC) offshore?
- b) Take a meridional (along a meridian = y-direction) transect from the south coast across the SEMC towards the cooler water to the south. Based on the transect, approximately how wide is the SEMC at this point? How far offshore is the core of the SEMC? (Assume that 1 pixel equals about 1km).
- c) Take a zonal (West to East = x-direction) across the cooler water along the Madagascar south coast. Approximately how wide is this coastal feature? (1 pixel is approximately 1km). What are the main seasonal temperature changes in this water? What is the approximate temperature difference between this coastal water and the warmer water of the SEMC and Mozambique Channel?

References and further reading

Western Indian Ocean circulation

- [1] Manyilizu, M., Penven, P., Reason, C.J.C (2016): Annual cycle of the upper-ocean circulation and properties in the tropical western Indian Ocean, African Journal of Marine Science 38(1), 81-99. DOI: 10.2989/1814232X.2016.1158123. [Fulltext available: https://www.researchgate.net/publication/301705955_Annual_cycle_of_the_upper-ocean_circulation_and_properties_in_the_tropical_western_Indian_Ocean#fullTextFileContent]

MUR-SST data access and documentation

- [2] NASA-JPL PODAAC: GHRSSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis (v4.1) – Data access and information at <https://podaac.jpl.nasa.gov/dataset/MUR-JPL-L4-GLOB-v4.1>

Acknowledgements

This lesson was created for the C-RISe Project (<http://www.satoc.eu/projects/c-rise/>) with support from the UK Space Agency International Partnerships Programme.

The data used in the lesson were provided by Plymouth Marine Laboratory (PML) for the SOLSTICE-WIO project (www.solstice-wio). It is a subset of data available from the PODAAC data service as NASA-JPL.