SEA-EU Marine Data Literacy course
Practical session #3: from in situ observations to gridded data

Introduction

In this session we will process individual observations to create output on a regular grid of our choice. We will use the Ocean Data View software to do this today, but it is also possible to perform such gridding programatically (e.g. using python, MATLAB…).

Why learn how to grid observational data? We might want to do this for a number of reasons, but one of the most common is to make it easier to analyse changes that are occurring over time. For example, Argo floats drift freely with the ocean currents and report measurements every 10 days. If we want to use these data to quantify how the ocean temperature is changing over time at a certain point, we need to decide how to deal with the "random" spatial sampling. It is extremely unlikely that we will obtain regular measurements from exactly the same location when using a random sampling method. By using the measurements that we have to estimate the value at a different location, where this different location is fixed over time, we can overcome this problem.

Before beginning the practical:

You will need:
- a working copy of Ocean Data View installed on your computer
- the data files supplied for the practical: please unzip these into a folder of your choice. The folder containing these files will be referred to as **the data directory** in all of the instructions below.

In the instructions, where you see **red text like this**, this refers to menu items or options in the Ocean Data View program.

Instructions:

Part 1: making a data collection and loading an Argo profile

In Ocean Data View (ODV), we work with "collections" of data. These can include several different types of observations if you wish, but today we will be working with just Argo profiles. A collection of data is just a set of observations that you have loaded into Ocean Data View to analyse. It can include any number of observations, from zero up to thousands (depending on what your computer can cope with. A warning: if you try to load very many observations it can take a long time to do anything afterwards!)

We will begin by setting up a collection in ODV. When you open the program, you will have a screen that looks something like this:

To get started, we need to import some data. Click on **Import → Argo Formats → Profiles**
In the dialog box, navigate to the data directory where you extracted the files for this practical.
In the data directory, navigate to the folder:
ARGO_DATA/aoml/4902119/profiles, and select the file D4902119_027.nc. Click on **OK**, in the dialog boxes that pop up, and you should then see a dialog box telling you that the import was successful.
You should now see a map in the centre of your screen, which shows the ocean bathymetry and the profile that you have loaded, marked as a blue dot with a red crossnear Iceland. There will also be three new windows in a bar at the right hand side of your screen: **Station ID**, **Sample** and **Isosurface values**. These windows give you some information about the data that you have just imported: we can see some metadata, as well as some of the numerical values.

If we want to zoom in on the region surrounding our profile, we can right-click or ctrl-click on the map, and select the menu item **Zoom**. A red box will appear. You can adjust the position and size of the box using your mouse. Try reducing the size of the box to have a region of about 20° x 20° around the blue dot showing the station position. Press **Enter** when you are ready, and the map will be updated to show this new zoomed region.

Part 2: loading in many profiles

In step 1, we loaded in a single Argo profile. But to create our gridded outputs, we will need lots of data! So we need to load this into ODV. With the method that we used in step 1, we can only load one profile at a time. To use all of the data, we will instead create a list of files, which we will tell ODV to load. ODV has a tool to create this list for us: click on **Tools → List file generator**. A new dialog box will appear that looks like this:
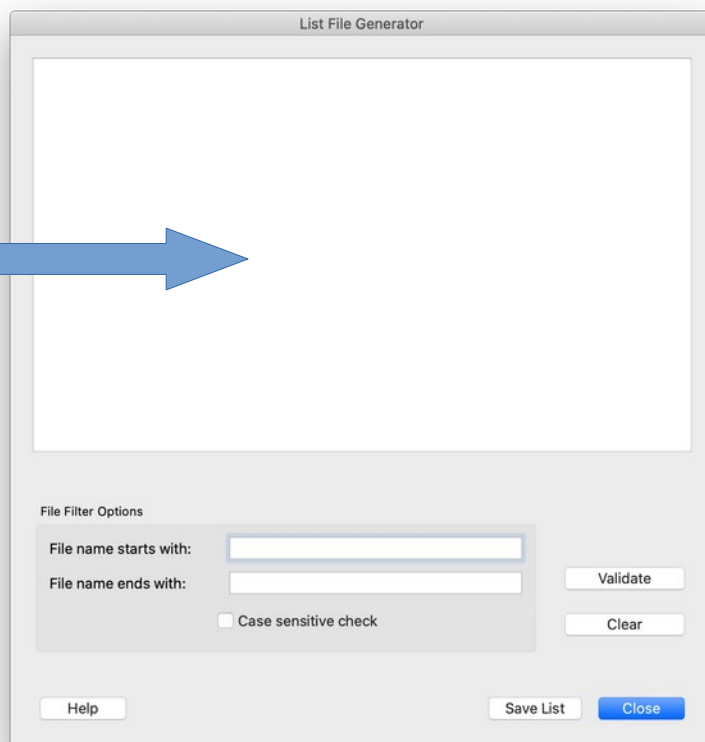
Open the file explorer for your operating system, and navigate to the data directory where you put the data files for this practical. Drag and drop the ARGO_DATA subfolder on to the main window here

A list of files should appear in the window.

You can then click on **Save List**, and choose a name and directory in which to save your list of files.

You can now click on **Close** to exit this tool.

List File Generator

File Filter Options

File name starts with: [   ]

File name ends with: [   ]   Validate

☐ Case sensitive check   Clear

Help   Save List   Close

We will now use this list to load all of our data files. As before, click on **Import → Argo Formats → Profiles** and then navigate to the directory where you stored the list of files. At the bottom of the window, change the option for Files of type from Argo Profile Files v3.1/v1.0 (*.nc *.NC) to the option File List Files (*.lst *.LST). You should now be able to select your file. Click on Open, and ODV will start preparing to import the files. ODV may ask you for a name to Save the collection file, if so, simply choose a directory and name and click Save. If Metavariable association and Import Options windows appear, click on **Ok**. ODV will then start reading in the files.

When the import has finished, many new dots will appear in the map window. You can adjust the zoom if necessary so that you can see all of the points. Follow the same steps

as in the step 1 to do this. Try to leave a bit of space around the points so that you can see the effects of the gridding process in the steps that follow.

To finish this import process, we will tell ODV to generate derived variables of longitude and latitude. These variables already exist in our data set, but some of the tools that we will use later will only work correctly if ODV has also generated these variables itself, so it is important not to skip this step!

Click on **View → Derived Variables** and a new window will pop up.
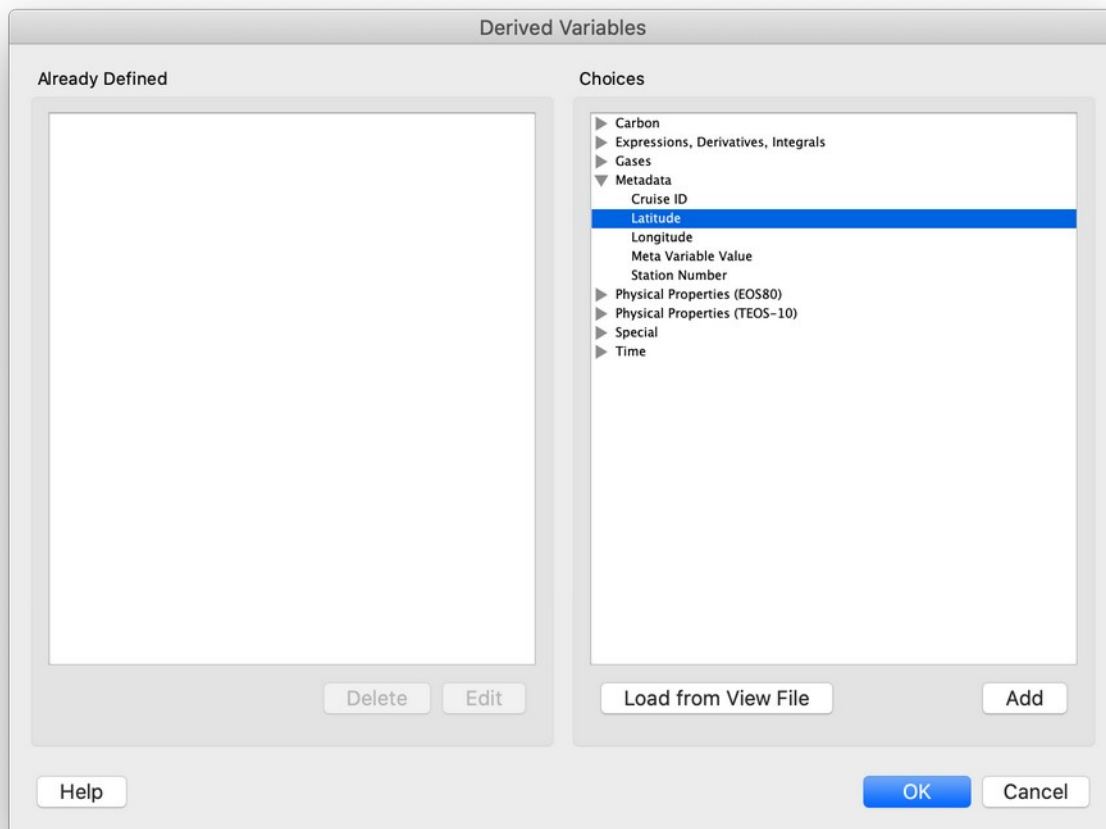
In the right hand side of the window, labelled Choices, expand the Metadata category. Click on Latitude in the list, then click the **Add** button, just below the window. Latitude should then appear in the Already Defined window on the left hand side. Repeat for Longitude, then click **OK**.

## Part 3: generating gridded output using all available data

The map that you can see in the ODV window displays only the positions of the observations. We would like to look at the values that have been measured. To do this, we first need to add some new windows to the ODV window layout.

Right click or ctrl-click within the ODV window, near your map, and select **Window Layout**. You can change the size of the map window if you want to, and/or move it by right-clicking / ctrl-clicking within the square labelled Map and selecting **Move / Resize window**. You can move and resize the window in the same way as for the zoom function, by adjusting the red square, and pressing Enter when you have finished. Make sure that

you have some space left in the area with squares on it, because we will be adding 2 more windows next to the map.

To add the new windows, right-/ctrl-click again and select **Create New Window**. You can then adjust the size and position as before. Arrange the window layout so that you are happy with it, and then press Enter again to exit the window layout mode.

You should have something that looks a bit like this (the position of the windows is not important, you just need to have a map, and two empty windows):
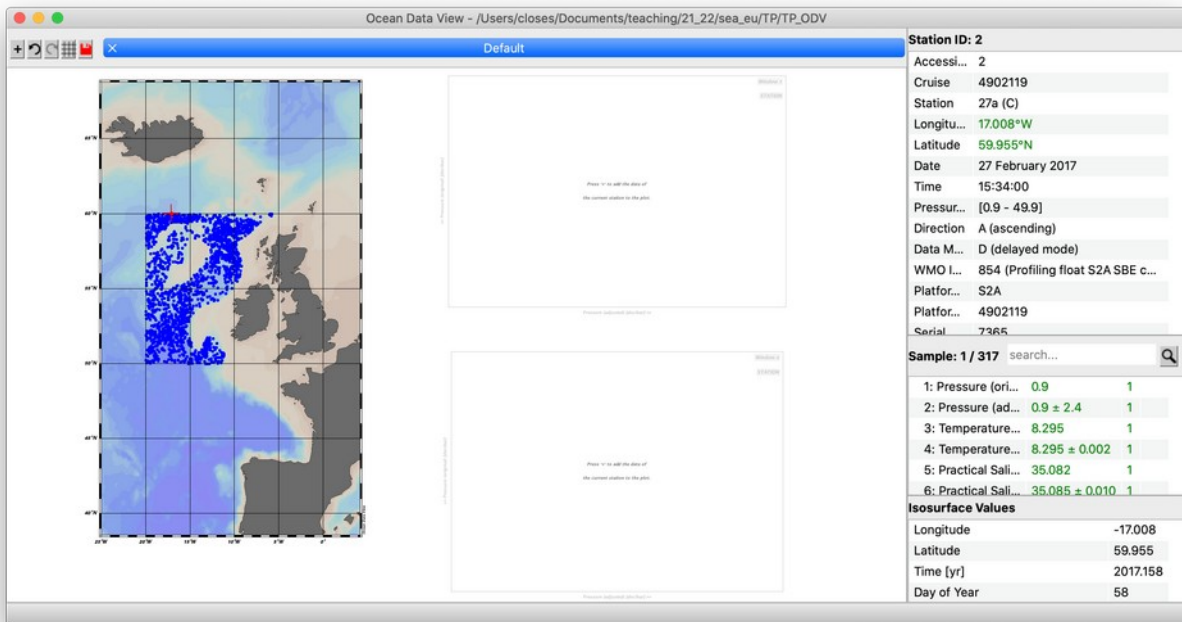
We are now ready to start displaying the temperature and salinity data.

Right-/ctrl-click in one of the empty boxes, and the following menu should pop up
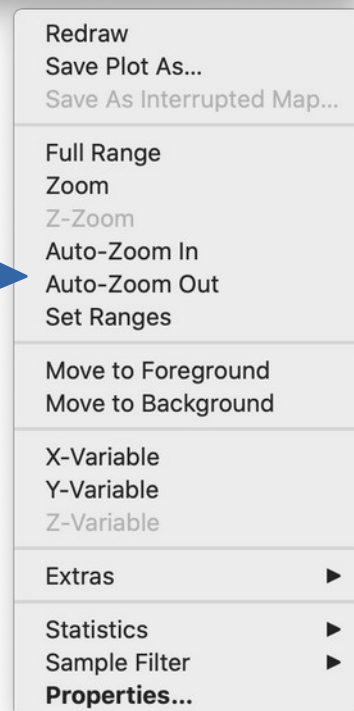
Select **Properties**, and a new dialog box will pop up.

Click on the **Data** tab near the top of the window. The first item on the page is an option named **Scope** which is currently set to STATION: Data of picked stations. In the drop down list, change this option to SURFACE: Isosurface data of all stations shown in the map.

The X-Axis, Y-Axis and Z-Axis options will now be activated in the bottom part of the window. This is where you choose what you would like to display in the figure. By default, X-Axis should be Longitude, and Y-Axis should be Latitude. In the Z-Axis drop down menu, change the choice from Pressure (adjusted) [decibar] @ Pressure (original) [decibar] = first to

Temperature (adjusted) [degrees_Celsius] @ Pressure (original) [decibar]=first. Don't click OK just yet, we'll make some more changes first!

Click on the Display Style tab at the top of the window. Make sure that the Original Data option is checked, with the option Colored Dots. We will use this to look at the individual point values of temperature before we create the gridded output. You can now click OK, and a new map should appear, showing the temperature values at the first depth level recorded in each profile as a coloured point.

In the second empty box, we will create the gridded output. Follow the same steps as above to select a surface plot, with the same variables, but in the Display Style tab, select **Gridded Field** instead of Original Data. We will leave all of the other options in this tab on their default values for the moment, you will try changing them very soon in the exercise below. Next, click on the Contours option, and click on the auto-create button. This tells ODV to add lines of constant temperature to our figure. Click on **OK** again, and ODV will create the gridded output.

Your 3 figures should now look something like this:
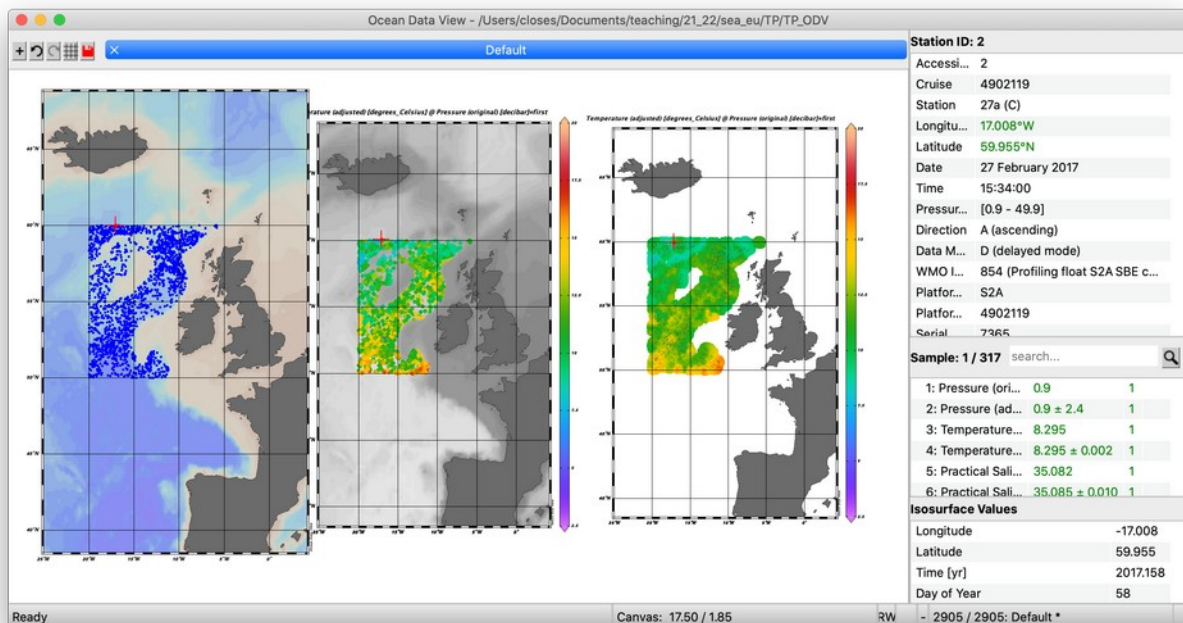
Note: you can rearrange your windows if necessary by right-/ctrl-clicking and selecting window layout again. In the window where the data have been gridded, you can see that the form of the output looks smoother: the gridding tool has tried to infer the temperature at points where there were no observations.

Exercise #1: We will now look at the effects of adjusting some of the gridding parameters. You will answer a multiple choice question based on this exercise for the evaluation. The question and possible responses are given after these instructions. Please be sure to note the response that you wish to give so that you can fill in your answer at the end of the session.

Exercise instructions: Reopen the Properties window for the gridded figure by right-/ctrl-clicking on the figure, as before. Click on the Display Style tab again. Uncheck the Automatic scale lengths box, and try changing the values for the X–

scale-length and Y-scale-length. Use the same number for both X and Y, and try first with small values (e.g. X-scale-length=5, Y-scale-length=5), then increase the values (e.g. 10, 20, 40, 80, 200). Note how the figure changes as you change these numbers: does it become smoother or less smooth? What happens to the maximum and minimum values as you change the numbers?

You can also try using different numbers for the X-scale-length and the Y-scale-length. See how this affects the output.

Finally, in the drop-down menu where Weighted-average gridding is selected, change this option to use DIVA gridding. First of all, check the Automatic scale lengths box again, and compare the output of the DIVA gridding method with the Weighted-average gridding method using the default parameters. You can then uncheck the Automatic scale lengths box and look at the effects of changing the numbers for X- and Y-scale-lengths again.

Multiple choice question: (choose only one answer)
Select the answer which describes your results from exercise #1:

a) When changing the gridding parameters, as we increase the X-scale-length and Y-scale-length we find that the output field becomes smoother and the difference between the maximum and minimum output values increases. When using the default scale values, the DIVA gridding method fills in more of the gaps in the spatial coverage than the Weighted-average gridding method.

b) When changing the gridding parameters, as we increase the X-scale-length and Y-scale-length we find that the output field becomes smoother and the difference between the maximum and minimum output values decreases. When using the default scale values, the DIVA gridding method fills in more of the gaps in the spatial coverage than the Weighted-average gridding method.

c) When changing the gridding parameters, as we increase the X-scale-length and Y-scale-length we find that the output field becomes smoother and the difference between the maximum and minimum output values decreases. When using the default scale values, the DIVA gridding method fills in fewer of the gaps in the spatial coverage than the Weighted-average gridding method.

d) When changing the gridding parameters, as we increase the X-scale-length and Y-scale-length we find that the output field becomes less smooth and the difference between the maximum and minimum output values increases. When using the default scale values, the DIVA gridding method fills in fewer of the gaps in the spatial coverage than the Weighted-average gridding method.

# Part 4: setting the depth level and generating new variables

The variable that we have plotted so far was called: Temperature (adjusted) [degrees_Celsius] @ Pressure (original) [decibar]=first

What does this mean? Let's interpret the different parts of the name. The first part is fairly easy: Temperature (adjusted) [degrees_Celsius] tells us that we are looking at the temperature in °C, and "adjusted" indicates that the data have been quality controlled. When we download Argo data, we normally have the choice of two options for the temperature and salinity variables: Temperature (original) and Temperature (adjusted) / Salinity (original) and Salinity (adjusted). If we choose Temperature

(original) [degrees_Celsius], we would get the data that had not undergone quality control, and there are likely to be some anomalous values in the data.

The second part of the name of the variable is @ Pressure (original) [decibar]=first. This is telling us about the surface on which the data have been calculated. Here the quantity that we are considering is pressure, which is related to the depth of the measurement, so if we have Pressure = 0 decibar, we are looking at measurements on the sea surface, whereas if we have Pressure = 4000 decibar, this corresponds to a depth near the bottom of the ocean. Here, the variable that we have chosen has a value of first. This means that ODV is taking the shallowest available measurement for each profile. Note that this could be different for each profile: for example, one profile might include a measurement at 10 dbar, near to the ocean surface, whilst another profile might not have any available measurements until 350 dbar, if it had some measurement problems. In this case, the first available measurements will be 10 dbar for the first profile, and 350 dbar for the second, and these will be selected and plotted together in our figure. The advantage of this approach is that each profile is represented in our figure, but the (very important!) disadvantage is that we are mixing together two completely different situations: the temperature several hundred metres below the ocean surface is very, very different to the temperature 10m below the ocean surface! In this step, we will learn how to define variables at different, fixed depth levels, so that we can create coherent estimates.

In the ODV menu bar, select **View → Isosurface variables**. A window like this will appear:
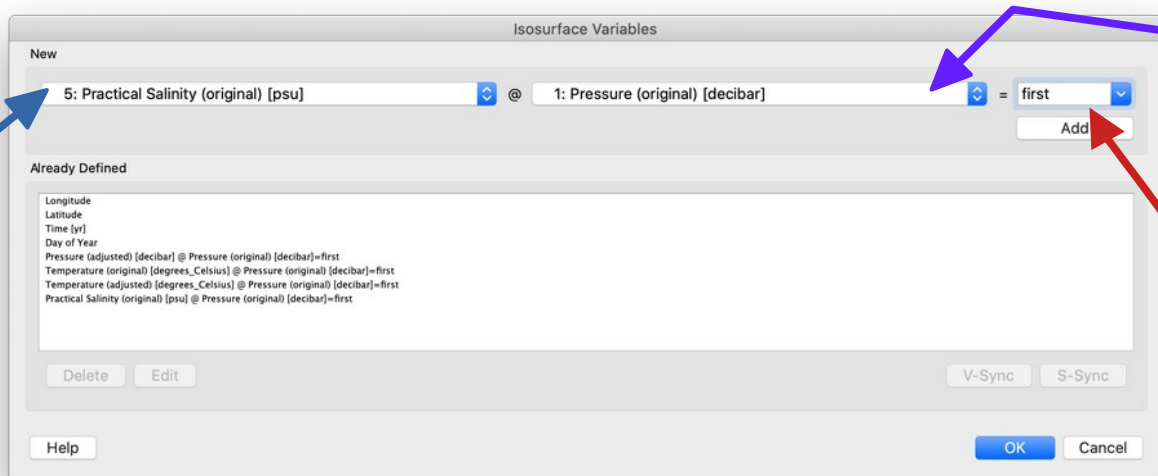
We will define a variable that will allow us to look at the salinity at a depth of 100 dbar.

Click on the first drop-down menu and select Practical Salinity (adjusted) [psu]

Click on the second drop-down menu and select Pressure (adjusted) [decibar]

Finally, in the third drop-down menu, select the empty option (1st item in the list) and type the number 100 in to the window. Then click the **Add** button.

You should see a new entry in the Already Defined window:

Practical Salinity (adjusted) [psu] @ Pressure (adjusted) [decibar] = 100.0

Click on **OK** to exit the window.

In the main ODV window, you can now plot this new variable by opening the `Properties` window, selecting the `Data` tab, and then selecting this new variable as the Z variable. In the `Display Style` tab, select the `DIVA gridding` option and make sure that the `Automatic scale-length` option is checked. Click **OK**, and ODV will calculate the new variable, and then display it.

You should find that the new field looks quite strange! We have a lot of values of salinity around 35, and then a series of very low values to the north west of Scotland. We can click in this region on the map to look at the data values in the `Sample` window, located at the right side of your screen in the centre panel. If you click here, you will see that the Practical Salinity values are indicated in red, and have the number 4 in the right hand column. This tells us that the data are bad quality: 4 corresponds to the quality control flat, and any value other than 1 indicates a potential problem with the data. Let's try excluding them.

Right-/ctrl-click on the map, then select `Sample Filter → Customize settings`. In the `Quality` tab, in the `Acceptable Quality Flags` box, you will see that all values are currently selected. Click in the box so that only `1: good data` is selected, and then click **Apply to all variables**. Click on **Ok** to close. ODV will redraw the map without these data — it should look much more uniform now!

Exercise #2: In this exercise you will practise producing depth-varying fields by repeating the above procedure. The associated multiple choice question is given below, and asks you to describe the changes that you see as you change the pressure level. Again, make sure to note the response that you wish to give so that you can fill in your answer at the end of the session.

Exercise instructions: create gridded temperature outputs at 10, 25, 50, 100, 200, 500, 1000 and 1500 decibars. Make a grid of 8 figure windows by altering the window layout, and plot each of these gridded outputs in these 8 windows. Observe how the outputs change as the pressure level increases.
To make it easier to compare the properties at different pressure levels, you might want to fix the colourbar scale, so that the colours always correspond to the same values when you change the level. To do this, once you have made the figures, right-/ctrl-click on each figure, and select **Set Ranges**. You can then change the Z-Axis values to always have the same minimum and maximum in each figure.

Multiple choice question: (choose only one answer)
Select the answer which describes your results from exercise #2.
As the pressure increases:

a) The output fields become less smooth, with more small patches of variability. Temperature increases uniformly over most of the region as the pressure increases.

b) The output fields become smoother, with fewer small patches of variability. Temperature decreases uniformly over most of the region as the pressure increases.

c) The output fields become smoother, with fewer small patches of variability. Temperature first increases, and then decreases as the pressure increases.

d) The output fields become less smooth, with more small patches of variability. Temperature first decreases, and then increases as the pressure increases.


Part 6: obtaining data values at chosen grid points

So far we have made figures where the input data have been gridded automatically by Ocean Data View. We might also want to choose a grid ourselves, and to obtain the data values at those particular grid points. Ocean Data View has tools that can help us to do this.

First, we need to define the grid of latitude and longitude points at which we want to create output values. Click on **Tools → Grid file generator** and a new window will appear which looks like this:
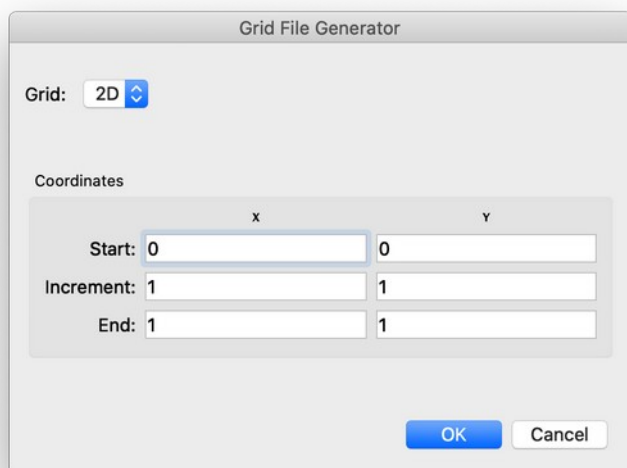
We will specify values for x and y, and the tool will then create a text file for us, containing a grid of values corresponding to the required x points and y points. To see how it works, we'll start with a simple example, changing the value of Increment to 0.25 for x and 0.5 for y, and then clicking on **OK**. A dialogue box will pop up asking you to select a file name and location for the output file. Choose where you want to store the file and click **Save**. Then, open this file using your computer's text editor. You will see that the output looks something like this:
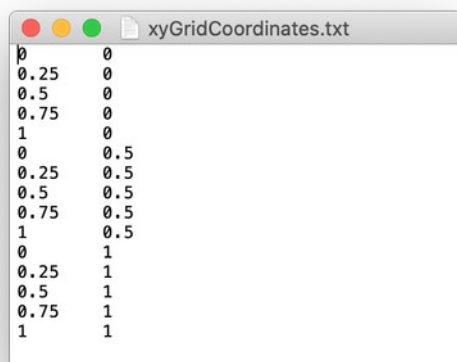
The first column contains the x-values of the grid, which go from the Start value (0) to the End value (1) in steps (increments) of 0.25.

The second column contains the y-values of the grid, which go from the Start value (0) to the End value (1) in steps (increments) of 0.5.

For each specified y-value, we have a set of 5 associated x-values that specify all possible values of x. Similarly, for each specified x-value, we have a set of 3 associated y-values that specify all possible values of y. We have thus specified the full grid of x/y points.

We will now repeat this to create a grid of latitude and longitude points, which we will use to generate gridded values for our data.

Open the grid generator again: **Tools → Grid file generator**. We will now specify longitude values for the x variable, and latitude values for the y variable, so that we generate a grid situated in the region of our data.

If we look at the map of our data, we can see that we have data contained within the region 20°W to 5°W and 50°N to 60°N. We will generate a grid with a spacing of 2° in longitude and 1° in latitude. To do this, for the x coordinate, set the Start value to -20, the Increment to 2 and the End value to -5. Next, for the y coordinate, set the Start value to 50, the Increment to 1 and the End value to 60. Click **OK**, and **Save** the file.

We are now ready to produce our gridded output. Right-/ctrl-click on the figure containing gridded temperature data at a depth of 10 dbar, and select **Extras → 2D Estimation**. A new window will pop up with the title Select File with X/Y Coordinates. In this window, select the grid file that you just created and click **Open**. A message should then pop up telling you that the gridded estimates have been written to a text file.

If you open the estimate file that ODV has produced in a text editor, you will see:

In order, the columns contain Longitude, Latitude, a "Dummy" variable, Temperature estimates at the chosen grid points, and temperature gradients in the x-/y-directions.

We can look up values directly in the file. However, these outputs can also be imported into ODV so that we can see the gridded structure. We will load these into a new collection so that we can easily see the grid pattern. Before we create the new file, let's save the figure of the Temperature at the 10 dbar pressure level that we created so that we can compare with the gridded output afterwards. Right-/ctrl-click on the figure, select **Save Plot as…** and select a file name, location and format in the window that pops up.

Now we will load the gridded file. Click **File → Open** and at the bottom of the window, change Files of type to Data files (*.txt, *.csv,*.o4x). Now select the gridded estimate file that we just created (it's default name will be xyGridCoordinates_est.txt). The following window will appear:

In the right hand panel, we can see that ODV has correctly identified all the column names. We can click **OK** without changing anything here. In the next window, ODV will try
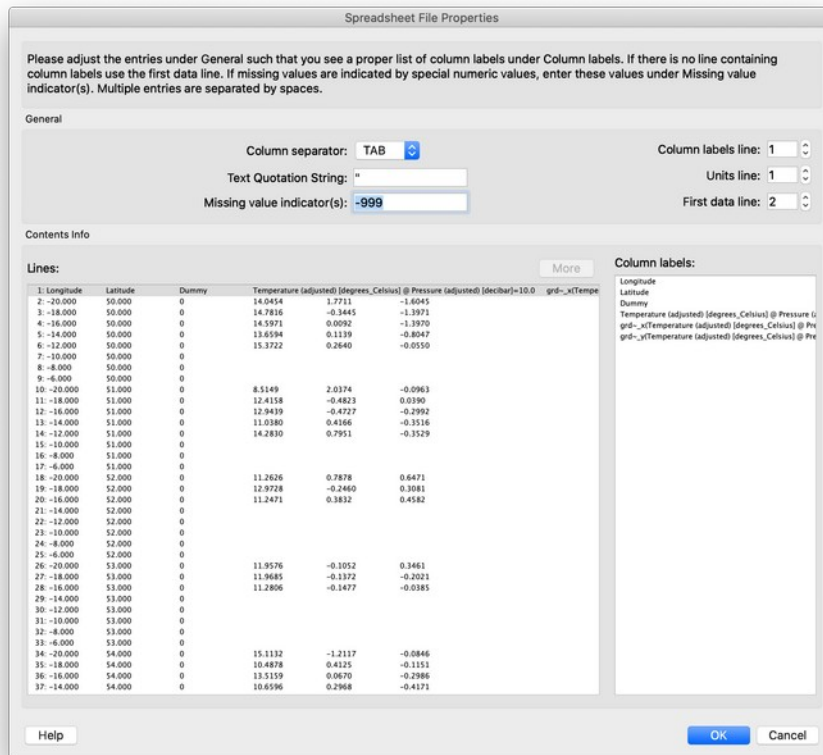
to identify metadata information automatically. Our file does not contain any, and we can just click **OK** again.

In the next window, ODV will try to identify the data and metadata variables. You should see a window that looks like this:
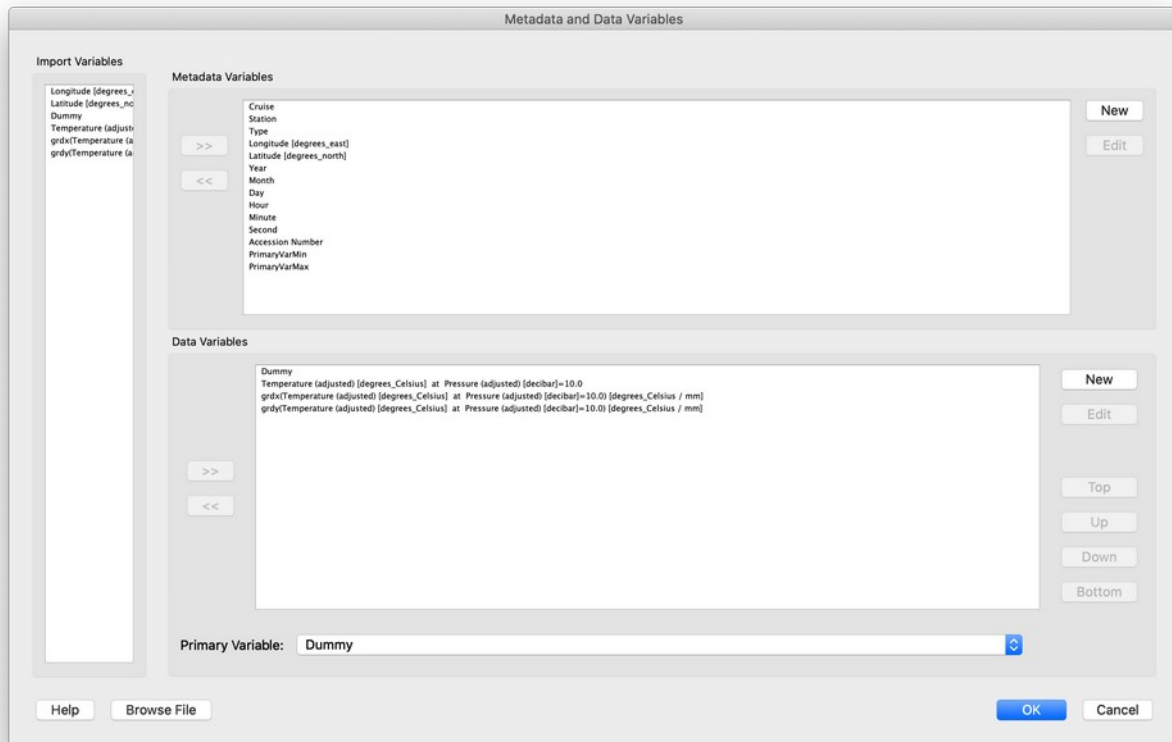
ODV has correctly identified the Longitude and Latitude, as well as the 4 data variables. Click **OK** again.

The next screen asks for information about the type of data: we don't need to fill this in, just click **OK** again.

In the next step, you should see that ODV has successfully identified the Longitude and Latitude variables (with a little star next to their names):
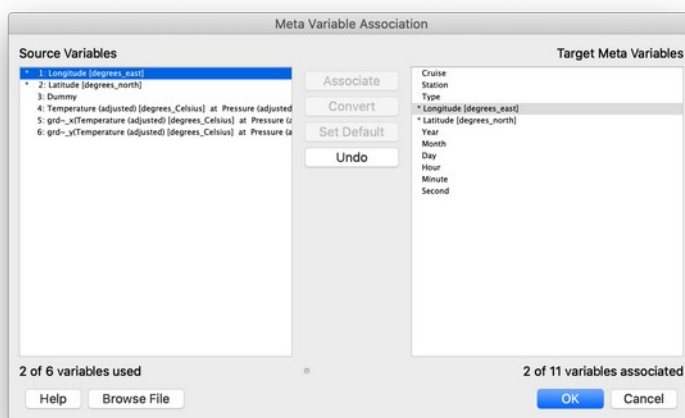
Click **OK** again without changing anything.

In the next window, ODV will associate the variables in the file with data variables: again, there is nothing to change, you can just click **OK**.

Finally, you should see a message saying that ODV has successfully imported 88 stations. Click **OK** to finish the import.

On the map, zoom into the study region, as previously. You should see that we have a regular grid of values, with the spacing that we specified when creating the grid. We can again add a new window, and create a surface plot to look at the original data values:
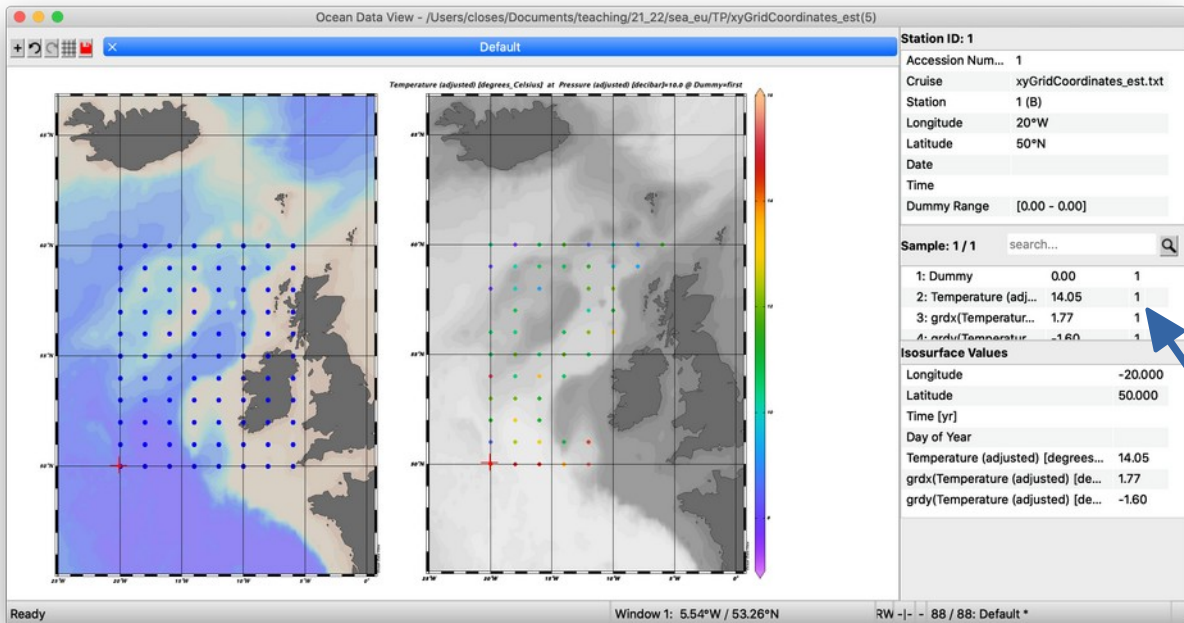
We can read the values at a certain grid point by clicking on it: the data values will appear in the middle panel of the right hand side window here

These values can also be regridded, as for the original data:
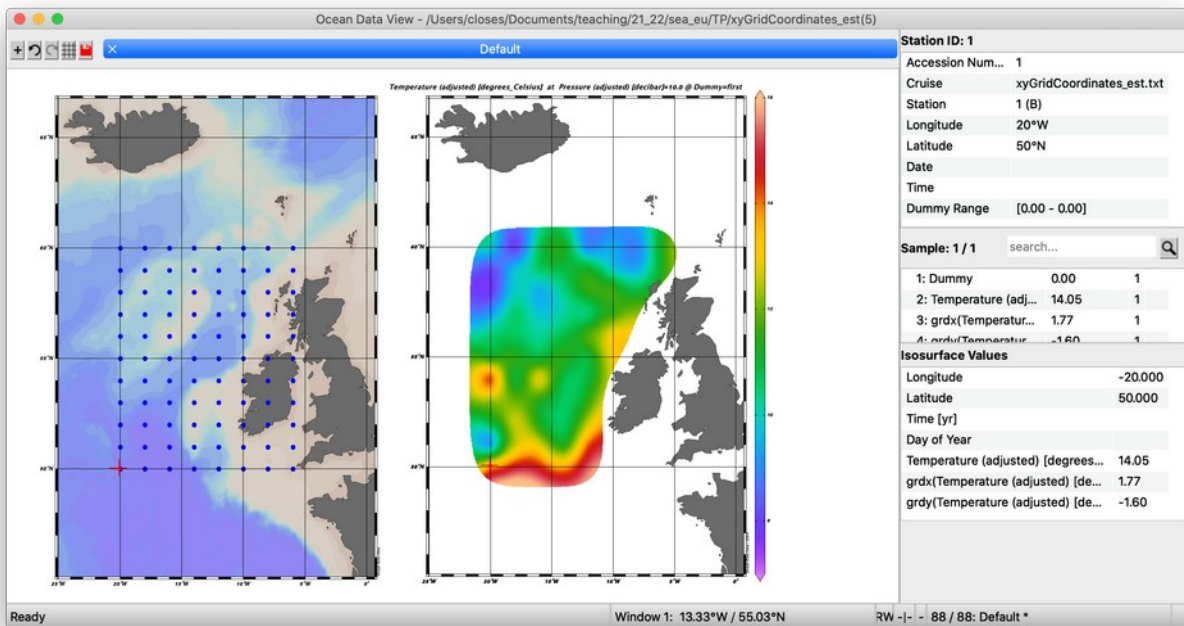
Open the figure file that you saved earlier and compare the gridded data with the original data. What has happened to all the small-scale detail?

Exercise #3: you will now practice generating gridded output. Reload the original data set (**File → Recent Files →** *name of your Argo data set collection*)

1. define a new grid covering the region 20°W to 5°W and 50°N to 60°N with a spacing of 0.25° in longitude (x-direction) and 0.25° in latitude (y-direction).

2. For the Temperature (adjusted) at Pressure level = 10 dbar, create a gridded figure using the DIVA gridding method, with X-scale = 15 and Y-scale = 15. Check that you have selected only data with a quality flag of 1 in the `Sample Filter → Customize settings` menu. Once you have created the figure, use this grid that you created in step (1) with the 2D estimation tool to create an output file containing the gridded values.

3. Open this gridded file in ODV to look at the output. Find the gridded estimate of the temperature at 18.25°W, 53.75°N and make a note of the value at this location so that you can answer the multiple choice question. Create a new gridded figure using this data. Save an image file of the figure so that you can refer to it later.

4. Re-open the original data collection, and repeat step (2), again using the DIVA gridding method, but with X-scale = 50 and Y-scale = 50. Again, perform the 2D estimation and save the output file containing the gridded values.

5. Repeat step 3 for the new output file that you created in step (4). Again, make a note of the gridded estimate of the temperature at 18.25°W, 53.75°N and save the figure as an image file.

6. Compare the image files for your two estimates. How do the patterns compare between the two figures? How do the patterns compare to the original figure that you saved earlier?

Multiple choice question: (choose only one answer; the numbers reported have been rounded to 2 decimal places)

    a) With X-scale=15, Y-scale=15, the gridded temperature estimate at 18.25°W, 53.75°N is 12.91°C. With X-scale=50, Y-scale=50, the gridded temperature estimate at 18.25°W, 53.75°N is 12.32°C.

    b) With X-scale=15, Y-scale=15, the gridded temperature estimate at 18.25°W, 53.75°N is 11.38°C. With X-scale=50, Y-scale=50, the gridded temperature estimate at 18.25°W, 53.75°N is 10.01°C.

    c) With X-scale=15, Y-scale=15, the gridded temperature estimate at 18.25°W, 53.75°N is 12.31°C. With X-scale=50, Y-scale=50, the gridded temperature estimate at 18.25°W, 53.75°N is 12.91°C.

    d) With X-scale=15, Y-scale=15, the gridded temperature estimate at 18.25°W, 53.75°N is 10.01°C. With X-scale=50, Y-scale=50, the gridded temperature estimate at 18.25°W, 53.75°N is 11.38°C.

Part 7: how accurate are our estimates?

Our gridded output files contain estimates of the values of our chosen variable at the points that we have chosen, but how accurate are our estimates when we make these gridded products?

We can imagine several important possible sources of error in our estimate. Let's look at some of the most likely ones.

You have already seen that changing the X- and Y-scales of the gridded product affects the values that we obtain as output. With bigger scales the gridded product is much smoother: the small-scale patterns disappear. Conversely, if we have small values for the X- and Y-scales, our gridded product retains almost all of the patterns that we see when we plot all of the data together. This leads us to our first question: **what are the "right" scales to use? Do we want our gridded product to precisely reproduce all of the patterns of our input data?** Our choices with respect to what scales to use are a first potential source of error in our estimate. We will return to these questions in the following sections.

Next, let's consider the effects of data availability. Consider the following 1-hour average "snapshot" of the ocean SST from a high-resolution numerical model (taken from https://www.youtube.com/watch?v=HzalL01j2bQ, using the NATL60 simulation by Ocean Next and the MEOM group in Grenoble):
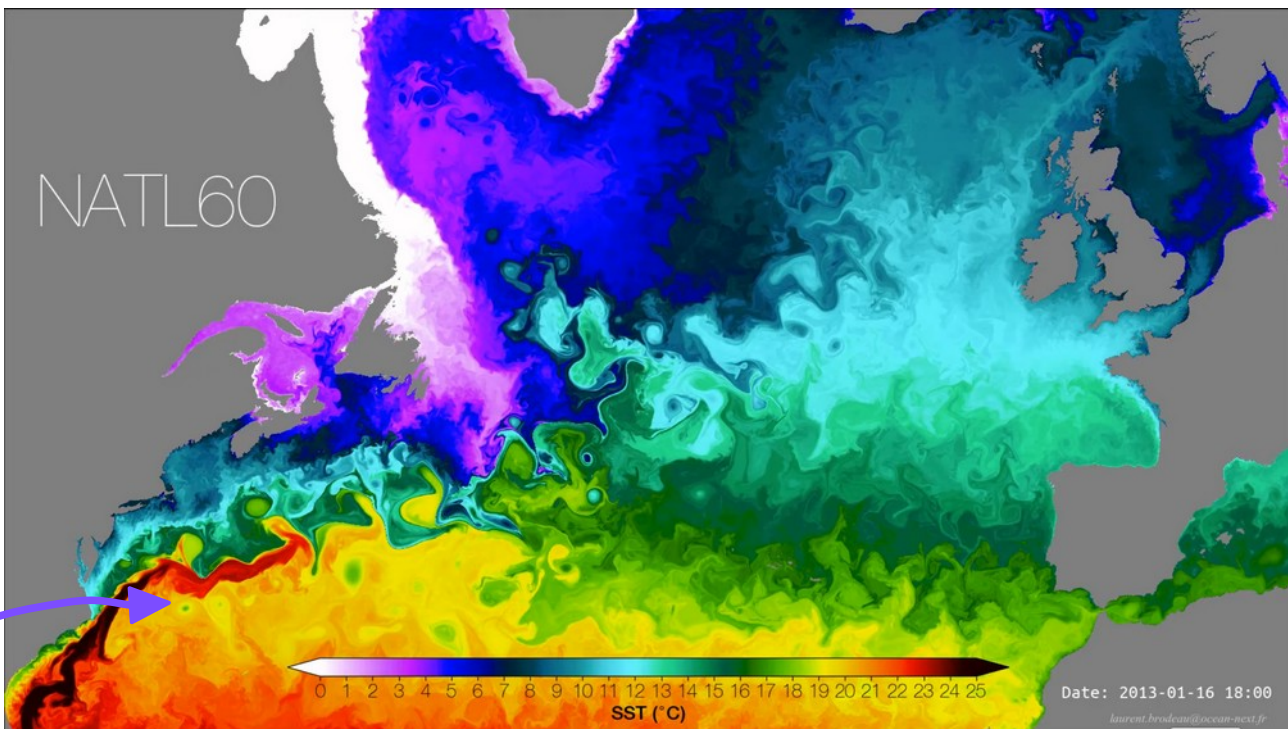

Imagine that this is the "real" state of the ocean. Consider one of the small-scale patterns in this image, such as one of these small green regions of cooler sea surface temperature. This feature is not very big. However, unless we have made a measurement within this small region, our gridded product will not represent this small-scale pattern.

Conversely, if, by chance, our measurements had been taken within these small-scale green circular patterns, rather than in the more uniform yellow-orange regions surrounding them, our gridding tool would try to "expand" these regions outwards, and we would again have errors due to the random nature of the sampling. **The data availability is thus an important source of potential error: we might not manage to sample a pattern that we would like to represent, or, alternatively, we might sample patterns that we don't really want to represent.** We have already seen that, depending

on the X- and Y-scales that we choose, the small-scale patterns can be well-represented, or smoothed out, and so the question of which scales to use is linked to this question of data availability. But this idea of not wanting to represent certain patterns might seem strange: why would we not want to represent all of the features that we can see in the image? Let's look at another image taken from the same video above, but 3 days later...
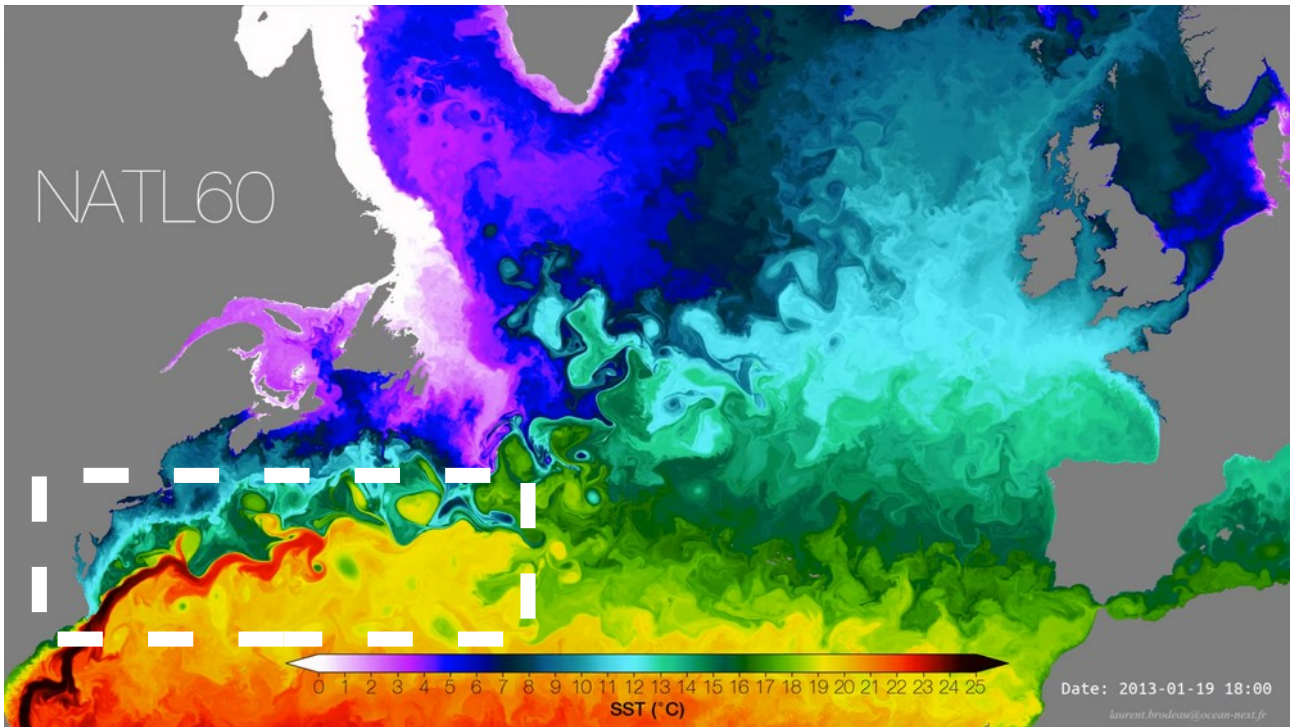
In only 3 days, there have been quite a few changes in the SST. If we focus particularly on the region surrounded by the white dashed rectangle, we can see that the warm, red current has started to deform relative to it's earlier state. Some other small features have also changed. And yet we have considered only a very short time interval! Argo floats measure the state of the ocean every 10 days, so let's compare a snapshot 10 days after the original image with the original image:

We can see changes in many different places!
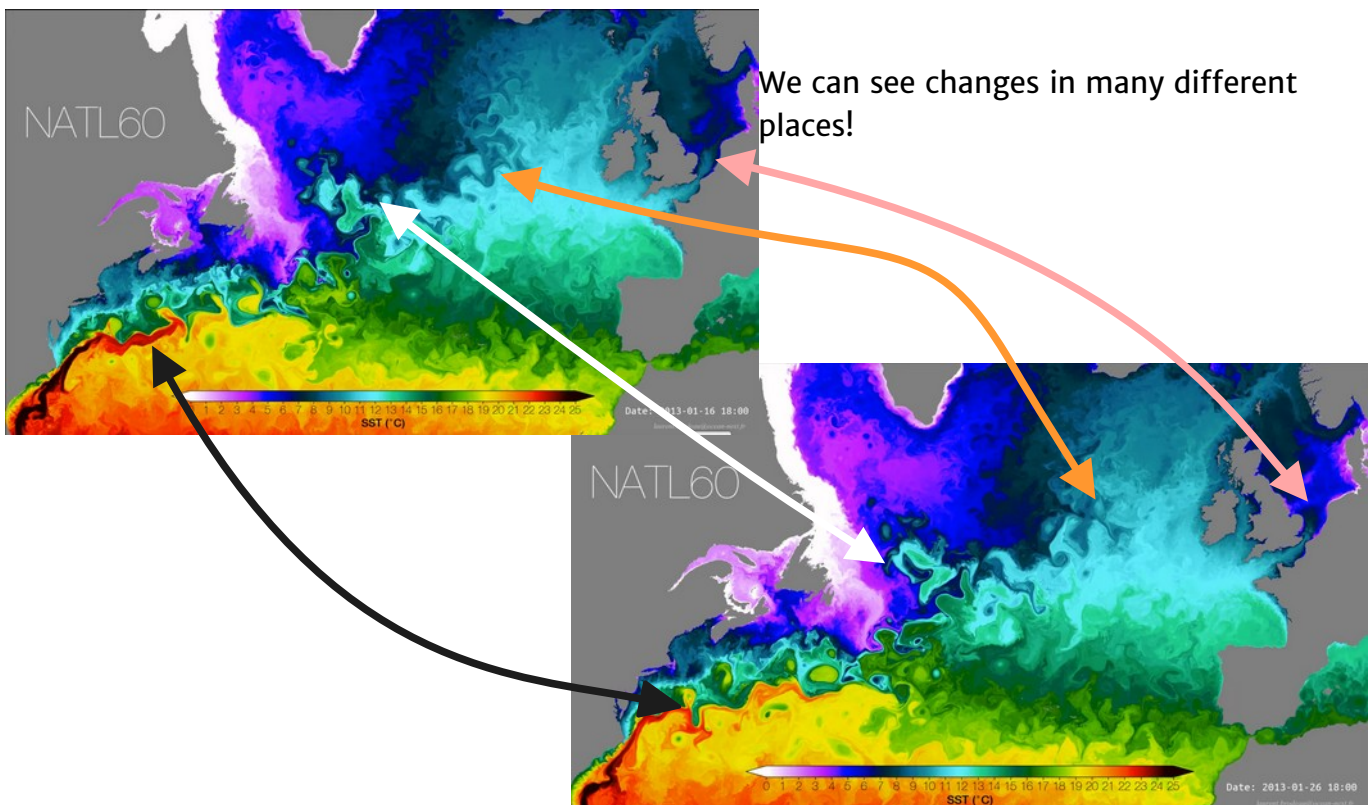
If we sample small patterns of variability, we might be representing patterns that no longer exist 10 days later: they evolve very quickly. This is a problem because of the error due to the problem that we considered above: data availability. We do not collect enough data within a 3-day or 5-day or even 10-day period to resolve all of the tiny patterns that we saw in the previous images. We need to collect data over a reasonable time period in order to have enough data to be able to make our estimates. However, if, by chance, we have sampled the small-scale, quickly-evolving patterns that we have seen in the previous images, they might not provide a relevant description of the average state of the ocean calculated over the time interval of our data collection. A third source of error is thus **the representativity of the data for the time scales that we are trying to resolve**. Again, this is linked to the X- and Y-scales that we began by talking about. There is little physical sense in having very small X- and Y-scales when we are making a gridded product using data collected over several months (or longer), because small-scale patterns evolve very quickly. Moreover, we cannot have enough in situ measurements to make a true average of this small-scale variability. This problem is most relevant at the surface of the ocean, where the ocean state evolves very quickly and seasonal variations can be very large, but it can also still be relevant at deeper levels of the ocean. To summarize this introduction to the problem: the X- and Y-scales that we choose need to be adapted to the length of the time interval over which our data have been collected, and also the data availability. In the rest of this practical, we will use the skills that you have learnt so far to see the effects of this problem, and to look at how our results change as a function of the time-scale that we are averaging over, and as a function of data availability.

Firstly, we might like to know whether a particular profile has a strong influence on our gridded output. We can do this by excluding that profile from our data set, and using the gridding tool to estimate the value at the location of that profile. If our gridded estimate is very different from the measured value, this might indicate any of the following:

1. that we have sampled a pattern that is not captured by the surrounding measurements (perhaps a small-scale pattern?)

2. that we have sampled a pattern that only persisted for part of the time period

3. that the instrument that provided that measurement had a technical problem and was returning bad data

Let's try to look at this in the region of the point that you looked at in the last exercise. Firstly, if you don't already have it open, re-open the original data set that we were working on.

If we zoom in to the region around 18.25°W, 53.75°N, the point where we estimated the temperature in the last exercise, we can see that there is no measurement directly at this point, but that we have a number of surrounding measurements. We will try excluding one of these, and estimating the gridded value at that point. By clicking on the surrounding points, I can see that there is a measurement at 17.98°W, 53.645°N, with

Cruise identifier 6900561 and station number 164 (C). This information will allow us to select, and exclude this profile.

Click on **View → Station Filter**. In the tab Name/Range enter the value 6900561 into the box labelled Cruise, and 164 − 164 into the 2 boxes labelled Station. Next, click on the **Invert** tab, and click Invert selection. Then click **OK**. This will select all profiles except for the one that we have identified. It will disappear from the map once you click OK.

Now we want to estimate the value at the location of this profile. To do this, we can use the gridding tool again. We will use a text editor to create a new file that contains just the coordinates of this one profile, which we will then use with the 2D estimation tool to obtain the gridded estimate at the profile location.

Open the text editor on your computer, and enter −17.98, then press the tab key, then enter 53.645 (all on the same line). Save this file somewhere where you can find it easily: we will use it as the grid values file for the 2D estimation input.

Next, in ODV, for your gridded surface estimation at 10 decibar, redraw the figure using DIVA gridding with X-scale = 25, Y-scale = 15. Then right-/ctrl-click on the gridded temperature estimate at the 10 decibar pressure level, and choose **Extras → 2D estimation**, and select the text file containing the station location that you just created as the grid location file in the Select file with X/Y coordinates window. ODV will create a file containing the estimated temperature value at that location, which we can check by opening the text file.

If we now click **View → Station Filter** again, uncheck the **Invert selection** option in the **Invert** tab and then click **OK**, ODV will now show only the profile that we just excluded, so we can easily check the measured value of the temperature.

The true value is 10.154°C, and the estimated value is 12.54°C. We can see that the quality flags for this profile indicate that it is considered to be "good" quality and so given this fairly big temperature difference between the expected and the obtained value we can conclude that with this profile we might have measured some small-scale or short time-scale variability.

To reselect all profiles, click **View → Station Filter** again, select the **Name** tab, click on the **Relax** button and then click **OK** again.

For the next step of our investigation of potential errors, we will compare our gridded temperature field on the 10 decibar surface with a satellite sea surface temperature (SST) product. A 10 decibar pressure corresponds to approximately 10m depth, whilst the SST gives the value at the surface = 0m depth. There is thus a slight depth offset between the two fields, but we still should expect the values to be close.

The field that we will compare with is a multi-year mean, calculated over the period 2018-2020. This corresponds to the time range of the profiles included in our data set.

In the SST subdirectory of your data directory, there is a file named `sst_2018_2020_mean.nc`. This is the mean of all daily fields over 2018-2020, calculated

from a product that combines satellite data and in situ measurements (NOAA Optimum Interpolation (OI) Sea Surface Temperature (SST), https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.html). We will load these data into ODV and use them to make a "reference field" that we will then compare with our gridded estimate.

Click **File → Close** to close the original data set that we have been working on, and then **File → Open** and navigate to the SST subdirectory of your data directory, and select the file `sst_2018_2020_mean.nc`. Click **Open**. ODV will now begin to import the data via the NetCDF Setup Wizard: a number of windows will appear as ODV asks you whether it has correctly identified the data and metadata in this file. Normally all the settings should be correct, and you can just keep clicking **Next >** until the process finishes (click **Finish** on the final screen).

First we will create the Latitude and Longitude variables within ODV: click **View → Derived Variables**, and under the Metadata category, click on Latitude, and the **Add** button, then Longitude and the **Add** button. Click **OK**. ODV will now create these two variables in a form that it will recognise when we reuse this data later.

We will now make a surface plot of this SST data. First we will look at the original data points, then we will make a gridded field again. Set up your window layout so that you have a spare window in which you can create the figure, then, as before, right-/ctrl-click, click **Properties →** select the **Data** tab → select SURFACE as the Scope → select the Display Style tab and make sure that Original Data is selected. Click **OK**. A plot of the SST data values should appear. You can see that this data is on a grid with 1° spacing between the points. You might also notice that it seems to be more uniform than our original data. Create the gridded field using the DIVA gridding tool and the default scales. How does this figure compare to our original data set? You can compare with the image file that you saved earlier if you don't remember how it looked.

We will now export this data as a reference field, so that we can compare it to our original data set. Click on **Export → Window data as reference**. Enter SST in the Exporting Window Data... dialogue box that asks you to enter a string ID, and press **OK**. A message will tell you that the data has been written to a directory somewhere on your computer.

Close this data set now, and reopen the gridded output field of temperature at the 10 decibar level using X-scale=15, Y-scale=15 that you created during the last exercise. We will now create a surface plot showing the difference between this estimate and the SST data.

First, we need to create some new variables. To begin, we need to create special Latitude and Longitude variables again. click **View → Derived Variables**, and under the Metadata category, click on Latitude, and the **Add** button, then Longitude and the **Add** button. Click **OK**.

Next, click on **View → Derived variables** again, and in the Special category select Difference from Reference. Click **Add**. A window will appear, where you will need to select the folder named SST, and then the file ExportID.txt. Click **Open**, then **OK**. In the next two windows, ODV should successfully identify the longitude and latitude variables.

In the 3rd window, it will ask you to identify the variable "`sst @ Dummy = first`": we will be comparing with the temperature field rather than the SST here, so you should select `Temperature (adjusted) [degrees_Celsius] at Pressure (adjusted) [decibar] = 10.0`. Click **OK** again twice to finish.

As a final step, before we can make our figure, we need to turn this new variable showing the difference between our estimate and the mean SST field into a quantity that we are able to make a surface plot of. Click on **View → Isosurface variables** and define the variable `drvd: Temperature (adjusted) [degrees_Celsius] at Pressure (adjusted) [decibar] = 10.0 – (Ref=SST) @ 1: Dummy = first`. Click **Add**, then **OK**.

We can now make a surface plot of this new variable showing the difference between our estimate and the mean SST field. Create a new surface plot in the usual way, and select this new variable as the Z-variable for your figure. You should see that the colourbar has a range of approximately −5°C to 5°C. If the SST field is the same as our estimate, the value will be 0°C. Positive or negative values indicate that the fields are different.

Right-/ctrl-click on the figure, and select **Statistics → Summary**. A new window will pop up, showing some values that characterise the field. In the third line, we can see that the temperature is cooler on average in our estimate than in the SST field, by -0.03°C, and that the range of values goes from -4.08°C to +4.67°C (note: these are quite important differences!). If you click on the Z-histogram button, a window showing the data distribution will appear. This confirms that our estimate is in general colder than the SST estimate. Click **Close** twice to return to the original figure.

Exercise #4: you will now calculate the variable representing the difference between the SST and the estimate of temperature that you made previously using X-scale=50, Y-scale=50 to see whether the smoother gridded field fits the SST data better.

Exercise instructions:

Repeat the steps above, but select the gridded estimate with X-scale=50, Y-scale=50 that you created previously in place of the gridded estimate with X-scale=15, Y-scale=15.

Multiple choice question: (choose only one answer; the numbers reported have been rounded to 2 decimal places)
   a) With X-scale=50, Y-scale=50, the gridded temperature estimate is slightly warmer than the SST, by 0.18°C. The difference between the two estimates has a minimum value of -2.36°C and a maximum value of +4.10°C.

   b) With X-scale=50, Y-scale=50, the gridded temperature estimate is slightly warmer than the SST, by 0.18°C. The difference between the two estimates has a minimum value of -2.36°C and a maximum value of +4.10°C.

   c) With X-scale=50, Y-scale=50, the gridded temperature estimate is slightly cooler than the SST, by 1.19°C. The difference between the two estimates has a minimum value of -4.10°C and a maximum value of +2.36°C.

d) With X-scale=50, Y-scale=50, the gridded temperature estimate is slightly warmer than the SST, by 1.19°C. The difference between the two estimates has a minimum value of -4.10°C and a maximum value of +2.36°C.

Part 8: selecting subsets of the data to look at the effects of variation over time

At the sea surface, large changes in temperature can occur from season to season: the SST in winter will be much colder than in summer. This can cause problems when we are trying to create our gridded estimates. If we combine measurements taken from the winter season with measurements taken from the summer season, we will create a field that does not have physical meaning: it will be representing the warm summer season at certain locations, and the cold winter season at other locations. This is thus a problem with the estimates that we have created so far, where we have used all available data!

To address this problem, we need to separate the data into different subsets, one for each season (perhaps even one for each month, if we have enough data), so that we are not mixing together physically-different situations. However, when we do this, we will reduce the total number of stations that we are using to make our estimate. As discussed previously, this can affect the accuracy of our results.

In ODV, we can use the Station Filter tool to select these subsets of the data.

First, reopen the original data set again.

Now click **View → Station Filter**, and the tool will open. We will select only the winter season: we will keep only measurements that were taken in January, February or March. Click on the **Date/Time** tab, and in the Season boxes, select **From: Jan / 01 To: Mar / 31**. ODV will redraw your figures. You should see that there is now much less data being used. In the figure properties window, you might like to click on **Auto-Adjust** in the Color mapping tab, so that you can see the detail clearly in your new gridded estimates. Save the figure showing temperature on the 10 decibar surface so that you can compare it to your estimates for the other seasons.
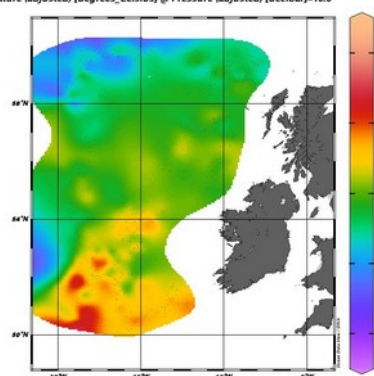
We will now change the station filter settings to look at spring. Reopen the Station Filter tool, and in the Season boxes, select **From: Apr / 01 To: Jun / 30**. Save your figure again. You can then repeat this for the summer and autumn seasons.

For summer and winter, your figures should look something like this:
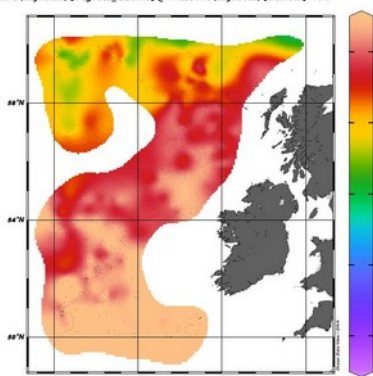
If we want to combine all of our data at the surface, one possible solution is to use these seasonal estimates to create reference fields, and then subtract these reference fields from the observed values to create fields of temperature anomalies, which can then be combined. Because we have only a limited amount of time, we will not look at this further here. Instead, for the final exercise, you will look at the feasibility of creating daily, weekly and monthly estimates of the gridded fields by comparing your estimated gridded outputs with three corresponding SST fields.

Exercise #5: you will create three different estimates in this exercise, and compare them with SST fields. Your estimates will be calculated using one month of data, one week of data, and finally one day of data. You will combine all of the techniques that you have seen so far in the practical for this final part!

Exercise instructions:

1. Using the station filter, in the Period section of the Date/Time tab, choose only data from 1st July 2018 to 31st July 2018. Grid the data using the DIVA gridding method and automatic length scales. Create a grid covering the region 20°W to 5°W and 50°N to 60°N with 1° grid spacing in both the latitude and longitude directions. Use the 2D estimation tool to save gridded output values of the month's temperature on the 10 decibar pressure surface using this 1° grid.

2. Close the original data set, open this 1° gridded temperature output from step (1), and use it to create a reference data set.

3. Close the data set, and open the SST monthly average for July 2018. This file is located in the SST subfolder of your data directory, and is named: `july_2018_SST.nc`. Following the same procedure as in Part 7, calculate the difference between the SST and the gridded output that you just saved. How well do the two estimates correspond? What do the statistics tell you?

4. Close the SST data, and reload the original data set again. Using the station filter again, in the Period section of the Date/Time tab, choose only data from 1st July 2018 to 8th July 2018. As in step (1), use the 2D estimation tool to save gridded output values of the week's temperature on the 10 decibar pressure surface using your 1° grid. Repeat steps (2) and (3), comparing with the weekly SST file: `01072018_08072018_SST.nc`. How close are the two estimates when comparing weekly averages? Are the results better or worse than for the monthly average?

5. Finally, repeat the process again, selecting only data from 1st July 2018 00h00 – 23h59. Create a new grid covering the region 20°W to 5°W and 50°N to 60°N with 0.25° grid spacing in both the latitude and longitude directions for your 2D estimation. We will use a slightly higher output resolution in this step to correspond to the higher spatial resolution of the daily SST data product that you will look at. Repeat the same steps as above to compare with the daily SST file: `01072018_SSTday.nc`. How close is your estimate to the SST data when comparing daily averages? Are the results better or worse

than for the monthly and weekly averages? How do you think the gridding scales and the data availability are affecting the accuracy of the output?

Bravo! You have finished the practical.

Optional extra exercises if you still have some time left:

- Try removing the seasonal cycle from your 10 decibar pressure level temperature data to see if you can create a better estimate of the long-term mean. You will need to: make an estimate for each season, adjusting the gridding scales so that you get an output that does not show too much small-scale variability. Save this gridded output as a reference field, and then define a new variable that gives the difference between your original data and the reference field. Grid the data and export it using Export → Window data. Repeat this for each season. Finally, create a new data collection, into which you import all 4 of these difference fields. You should now be able to create the gridded output.

- Use the Coriolis data server (http://www.ifremer.fr/co-dataSelection/?theme=argo) to download Argo data for the year 2018 in the Drake Passage region of the Southern Ocean. Choose Start date: 01/01/2018, End data: 31/12/2018, for the 4 geographical limits of the map, choose 55°S − 75°S and 80°W to 60°W. When you click "Refresh" at the top of the page, you should see that 514 profiles are available. You can click Download (NetCDF Argo) to download these data, then enter your e-mail to start the data extraction process. You will receive an e-mail once your data are ready. You can then try gridding the data in this region to see how the physical scales of variability and data availability affect the outputs in this region.