Residual Current Uncertainty Ellipses.

Jonathan Tinker, Met Office Hadley Centre, 22/04/2022.

Based on the Appendix of Tinker et al. (2022).

Some analytical calculations have been undertaken with Wolfram Alpha (https://www.wolframalpha.com/).

Introduction

We introduce a python toolbox for the analysis of bivariate normally distributed residual currents. This is largely based on fitting uncertainty ellipses around the mean, and using the size and shape of these ellipses to help describe the underlying data. Note that tidal currents are not normally distributed (as they are sinusoidal), so typically, the tide must be removed before this analysis is undertaken.

Background

Many scalar properties (e.g. temperature and salinity) are normally distributed (they fit a Gaussian distribution) which allows the use of parametric statistical tests. When data is normally distributed, 95% of the data is within 1.96 standard deviations of the mean. Therefore, if zero is more than 1.96 standard deviations from the mean, the distribution is significantly different from zero. The eastward and northward (*u* and *v*) components of the tidal residual currents are typically normally distributed (Figure 1) and act as any other (normally distributed) scalar quantity.

However, as residual currents are vector quantities, they (generally) cannot be analysed separately. When considering two co-varying, normally distributed data sets in *u*-*v* space, their data distribution can be described as a bivariate normal distribution, which is a Gaussian surface. Like a Gaussian curve, most data lie around the mean. To illustrate this, we plot the *u* velocity component against the *v* component of an example data set (Figure 2a). Most data is within an ellipse centred on the mean, with a shape varying from a straight line (where *u* and *v* are 100% (or -100%) correlated, see the black and gold data in Figure 2a), to an ellipse (the red data in Figure 2a has a correlation of 50%), to a circle (where they are uncorrelated (and of equal variance), see green data Figure 2a).

In Figure 2b, we consider such a distribution, with the first 30 points shown in yellow. We can fit an ellipse around these data (Figure 2c) that is a constant number of standard deviations from the mean - this is equivalent to the mean ± x standard deviations for scalar distributions. The 2.45 standard deviations ellipse encloses 95% of the data (see section "Probability levels" below), and so in Figure 2c we would expect ~1.5 out of 30 points to be outside the ellipse. When the origin is outside the (2.45 standard deviation residual current variability) ellipse, we consider the current to be significantly different from zero. This is shown in Figure 2d, where the origin is 2.8 standard deviations from the mean (as shown by the green ellipse) – the origin is outside of the 2.45 standard deviations ellipse (red). When the mean residual current is between 1 and 2.45 standard deviations (the red and yellow ellipse in Figure 2e), it is not significant, but we describe is as being quasi-steady, as it tends to go in the same direction.

When the current is significantly different from zero (the origin is outside the 2.45 standard deviation ellipse), from the perspective of the origin, the ellipse has a finite (angular) width. By taking two tangents from the ellipse through the origin, we can quantify this angular width (Figure 2f). As the current vector is (largely) enclosed by the ellipse, the direction of the current is (largely) within this range. Depending on the size, shape and location of the ellipse (with respect to the origin), even when the residual current is significant (the origin is outside the ellipse), the current can have a relatively variable direction. For some applications, in addition to the current significance (or number of standard deviations the origin is from the mean), we also take this directional spread into account.

By fitting this ellipse to the data, we can use properties of the ellipse to help describe the underlying data.

This methodology allows you to:

- assess whether a residual current is significantly different from zero, given the (e.g. inter-annual) variability.
- compare the residual currents of two model runs
- compare whether a single year is significantly different from a climatology.
- consider the likely range of current directions.

The Python CurrUncertEllipses toolbox

Loading the toolbox

All the required functions are within CurrUncertEllipses.py.

CurrUncertEllipses_examples.py gives example analysis and figures.

CurrUncertEllipses_ChiSqProb.py provides tools to select the critical ellipse size (in standard deviations).

Example data sets are given in baroc_*.nc.

Once downloaded, and copied to the correct location, the toolbox can be imported with:

```
import CurrUncertEllipses
```

Data

The toolbox works on numpy arrays of the U and V component of the residual velocities. These should have three dimensions, with dimension [0] being time.

The github include example data sets (baroc_*.nc). CurrUncertEllipses_examples.py gives an example of how to load the example data set, with a loading function – the default data directory (`datadir`) is set as ‘`./`’ – this may need to be changed for your system.

Quick guide

```
#Data with time as Dimension[0], and latitude and longitude in dimensions
[1] and [2].
# U_mat_1, V_mat_1, U_mat_2, V_mat_2

#Extract ellipse coefficients
n_std = 2.45
ellipse_dict = {}
ellipse_dict ['UV_1'] = ellipse_params_add_to_dict(ellipse_params(U_mat_1,
V_mat_1, n_std=n_std))
ellipse_dict ['UV_2'] = ellipse_params_add_to_dict(ellipse_params(U_mat_2,
V_mat_2, n_std=n_std))

#Compare two datasets
overlap_dict = overlapping_ellipse_area_from_dict(
ellipse_dict['UV_1'],ellipse_dict['UV_2'])
OVL_dict = ellipse_overlap_coefficient_pdf_from_dict(
ellipse_dict['UV_1'],ellipse_dict['UV_2'])
```

Assessing normality of the data

To calculate whether the U and V components are normally distributed.

Based on scipy.stats.normaltest

```
U_isnorm,V_isnorm = ellipse_init_norm_test(U_mat,V_mat, min_time_samples =
8,alpha = 1e-3)
```

Where `min_time_samples = 8` is the minimum number of time samples required for the test and `alpha = 1e-3` is the probability threshold.

`U_isnorm,V_isnorm` are 1., when `U_mat,V_mat` are normally distributed (respectively), and 0. when they are not.

This is used in Figure 1.

Ellipse Parameters

To calculate the basic Ellipse parameters for a set of residual current data, a single function, `ellipse_params`, can be called:

```
n_std=2.45

U_mean,V_mean,UV_mean,U_std,V_std,U_var,V_var,UV_cov,UV_mat,ang_xy,X_elip_amp,Y_elip_amp,X_elip_phi,Y_elip_phi,X_elip_phi_cos,Y_elip_phi_cos,qmax,qmin, ecc, theta_max,
zero_ang,XY_std_dir_corr,XY_zero_num_std_from_mean,pX_dir,pY_dir,y_tang_1,y_tang_2,ang_wid,foci_max,foci_x_1,foci_y_1,foci_x_2,foci_y_2 =
ellipse_params(U_mat,V_mat, n_std= n_std)
```

Furthermore, for ease of use, these can all be saved into a dictionary:

```
ellipse_dict = ellipse_params_add_to_dict(ellipse_params(U_mat, V_mat,
n_std=n_std))
```

ellipse_params and ellipse_params_add_to_dict runs through a set of routines:

- `ellipse_init_proc`
- `confidence_ellipse_uv_mat_parametric_equation`
- `ellipse_parameters_from_parametric_equation`
- `find_num_std_to_point`
- `find_tangent_to_parametric_ellipse_at_a_point`
- `find_parameteric_ellipse_foci`

These are now described in turn.

ellipse_init_proc: the initial processing

```
U_mean,V_mean,UV_mean,U_std,V_std,U_var,V_var,UV_cov,UV_mat,ang_xy =
ellipse_init_proc(U_mat,V_mat)
```

Calculate the basic stats of the U and V data:

- U and V Mean: U_mean,V_mean
- Mean magnitude: UV_mean
- Standard deviations and variances: U_std, V_std, U_var, V_var
- Covariance: UV_cov
- Velocity Magnitude: UV_mat
- Current direction: ang_xy

confidence_ellipse_uv_mat_parametric_equation: Calculating the Ellipse equations

```
X_elip_amp,Y_elip_amp,X_elip_phi,Y_elip_phi,X_elip_phi_cos,Y_elip_phi_cos =
confidence_ellipse_uv_mat_parametric_equation(U_mat,V_mat, n_std=n_std)
```

Calculate the uncertainty ellipses:

Given the mean (μ_x, μ_y) and variability (σ_u^2, σ_v^2) of the u and v components of the residual velocity and their covariance ($cov_{u,v}$). We can now calculate the ellipse for a given number of standard deviation.

The uncertainty ellipses can be defined as:

$$X_{ellipse(\theta)} = \alpha_x \sin\left(\theta + \emptyset_x\right) + \mu_x \qquad (1)$$
$$Y_{ellipse(\theta)} = \alpha_y \sin\left(\theta + \emptyset_y\right) + \mu_x$$

For $\theta = 0{:}\,2\,\pi$

where:

$$\alpha_x = \sqrt{0.5\sigma_u^2}\sqrt{(\varepsilon_x)^2 + \left(-\varepsilon_y\right)^2} \qquad (2)$$
$$\alpha_y = \sqrt{0.5\sigma_v^2}\sqrt{(\varepsilon_x)^2 + \left(\varepsilon_y\sqrt{0.5}\right)^2}$$
$$\emptyset_x = \text{arctan2}\left(\varepsilon_x, -\varepsilon_y\right)$$
$$\emptyset_y = \text{arctan2}\left(\varepsilon_x, \varepsilon_y\right)$$

and ε_x and ε_y are the ellipse radius in the x and y direction:

$$\varepsilon_x = \sqrt{1+\rho} \qquad (3)$$
$$\varepsilon_y = \sqrt{1-\rho}$$

And ρ is the polarity

$$\rho = {cov_{u,v}}\Big/{\sqrt{\sigma_u^2\sigma_v^2}} = cov_{u,v}/\sigma_x\sigma_y \qquad (4)$$

The python function `confidence_ellipse_uv_mat_parametric_equation` calculates these terms as:

- X_elip_amp, Y_elip_amp, X_elip_phi, Y_elip_phi = α_x, α_y, $\emptyset_x$, $\emptyset_y$ in equation (1).

If (1) is written in terms of cos ($X_{ellipse(\theta)} = \alpha_x \cos\left(\theta + \emptyset_x\right) + \mu_x$)

- X_elip_phi_cos,Y_elip_phi_cos = $\emptyset_x$, $\emptyset_y$

ellipse_parameters_from_parametric_equation: Calculating the Ellipse parameters

```
qmax,qmin,ecc,theta_max,zero_ang =
ellipse_parameters_from_parametric_equation(X_elip_amp,Y_elip_amp,X_elip_ph
i,Y_elip_phi,U_mean,V_mean)
```

We can then calculate the parameters describing the shape of the ellipse.

The semi-major and semi-minor amplitudes q_{max} and q_{min} are defined as:

$$q_{max} = \sqrt{\left(\frac{{\alpha_x}^2 + {\alpha_y}^2 + \alpha^2}{2}\right)} \qquad (5)$$

$$q_{min} = \sqrt{\left(\frac{{\alpha_x}^2 + {\alpha_y}^2 - \alpha^2}{2}\right)} \qquad (6)$$

where:

$$\alpha^2 = \sqrt{\alpha_x{}^4 + \alpha_y{}^4 + \left(2\alpha_x{}^2\alpha_y{}^2\right)\cos\left(2\left((2\pi - \emptyset_x) - \left(2\pi - \emptyset_y\right)\right)\right)} \tag{7}$$

The semi-major and semi-minor amplitudes give the eccentricity ε:

$$\varepsilon = \left(\frac{q_{max} - q_{min}}{q_{max} + q_{min}}\right) \tag{8}$$

The semi major phase (θ_{max}) is calculated as:

$$\partial_{num} = \alpha_x{}^2 \sin\left(2\left((2\pi - \emptyset_x) - \left(2\pi - \emptyset_y\right)\right)\right) \tag{9}$$

$$\partial_{denom} = \alpha_x{}^2 + \alpha_y{}^2 \cos\left(2\left((2\pi - \emptyset_x) - \left(2\pi - \emptyset_y\right)\right)\right)$$

$$\partial = \arctan2(\partial_{num}, \partial_{denom})$$

$$\theta_{max_{num}} = \alpha_y \cos\left((2\pi - \emptyset_x) - \left(2\pi - \emptyset_y\right) - \partial\right) \tag{10}$$

$$\theta_{max_{denom}} = \alpha_x \cos(\partial)$$

$$\theta_{max} = \arctan2\left(\theta_{max_{num}}, \theta_{max_{denom}}\right)$$

The python function `ellipse_parameters_from_parametric_equation` calculates these terms as:

- qmax,qmin = q_{max}, q_{min} in (5) and (6)
- ecc = ε in (8)
- theta_max = θ_{max} in (10)
- zero_ang is the angle to the origin.

find_num_std_to_point: Finding the number of standard deviations to the origin

```
XY_std_dir_corr,XY_zero_num_std_from_mean,pX_dir,pY_dir =
find_num_std_to_point(U_mean,V_mean,X_elip_amp,Y_elip_amp,X_elip_phi,Y_elip
_phi)
```

To find the number of standard deviation from the mean to the origin, we first find the angle between the origin and the ellipse centre. This has to be converted to ellipse angles (i.e. θ in (1)). We then find the ellipse "radius" at this angle, and scale the distance from the origin to the ellipse centre by this point.

find_tangent_to_parametric_ellipse_at_a_point: Finding the tangents to the ellipse, and its angular width

```
y_tang_1,y_tang_2,ang_wid =
find_tangent_to_parametric_ellipse_at_a_point(U_mean,V_mean,
X_elip_amp,Y_elip_amp,X_elip_phi,Y_elip_phi,pnt_x=pnt_x,pnt_y=pnt_y)
```

We find the tangents between the ellipse and the origin as the maximum and minimum of all possible angles between the ellipse and the origin. Hence this is not defined if the origin is within the ellipse.

To calculate the angles between the ellipse and the origin, we take the ratio between $Y_{ellipse(\theta)}$ and $X_{ellipse(\theta)}$ in (13) to give the tangent of the angle.

$$\tan(\emptyset_\theta) = \frac{Y_{ellipse(\theta)}}{X_{ellipse(\theta)}} \tag{11}$$

$$\theta = 0{:}\,2\,\pi$$

As the origin is (0,0) we do not need to subtract it – if interested in other points we use ($Y_{ellipse(\theta)}$-y_i)/ ($X_{ellipse(\theta)}$-x_i).

We then find the minimum and maximum angle to give the tangents. This step can be done numerically or analytically. By cycling through the θ in (26), we can calculate all the tangents, and then select the minimum and maximum value. Solving this analytically is possible using an algebraic solver (we used Wolfram Alpha), but the solution has more than 300 terms, so we have not included it in these appendices. Both methodologies are available in the python library.

The python function `find_tangent_to_parametric_ellipse_at_a_point` uses (11) to calculate $\tan(\emptyset_\theta)$, and then returns the maximum, minimum as the two tangents (y_tang_1,y_tang_2). The difference between them gives the angular width of the ellipse as ang_wid.

find_parameteric_ellipse_foci: Finding the Ellipse foci

```
foci_max,foci_x_1,foci_y_1,foci_x_2,foci_y_2 =
find_parameteric_ellipse_foci(qmax,qmin,theta_max,U_mean,V_mean,n_std)
```

The distance from the one focus to the curve of the ellipse and back to the other focus is constant, and equal to twice the semi-major amplitude (q_{max}). For an ellipse of a given number of standard deviations (n_{std}):

$$Foci_{hypot}(n_{std}) = n_{std}\sqrt{(q_{max})^2 - (q_{min})^2} \tag{12}$$

$$Foci_{x,1} = Foci_{hypot}\cos(\theta_{max}) + U_{mean}$$

$$Foci_{y,1} = Foci_{hypot}\sin(\theta_{max}) + V_{mean}$$

$$Foci_{x,2} = Foci_{hypot}\cos(\theta_{max} + \pi) + U_{mean}$$

$$Foci_{y,2} = Foci_{hypot}\sin(\theta_{max} + \pi) + V_{mean}$$

The python function `find_parameteric_ellipse_foci` calculates these terms as:

- foci_max = $Foci_{hypot}(n_{std})$ in (12)
- foci_x_1,foci_y_1,foci_x_2,foci_y_2 = $Foci_{x,1}$, $Foci_{y,1}$, $Foci_{x,2}$, $Foci_{y,2}$ in (12)

Comparing two sets of residual currents

We can compare two sets of residual current by comparing their uncertainty ellipses, and calculating their overlap:

```
overlap_dict = overlapping_ellipse_area_from_dict(dict_1, dict_2)
```

or we can compare the Gaussian distributions with the Overlap Coefficient:

```
OVL_dict = ellipse_overlap_coefficient_pdf_from_dict(dict1, dict2)
```

Background

<u>Area of intersection of ellipse</u>

Calculating the intersection area of two ellipses is complicated, so we take a numerical approach. For each ellipse, we create a Boolean mask of points within the ellipse. The masks extend beyond both ellipses in both the *x* and *y* directions. The higher the spatial resolution, the more accurate the result is. We then cycle through each point asking if it is within the ellipse (13, below), and repeat for the second ellipse. By comparing the two Boolean masks, we can ask which points are within both ellipses, and can use this to quantify the area.

You can ask if a point (p_x, p_y) inside an ellipse, you use the sum of the distance to the two foci. For a circle, a point is within the circle if the distance to the centre is less than the radius. For an ellipse the sum of the distance to both ellipse foci, *FociPntDist*, must be less than twice the semi major axis amplitude, q_{max}

$$FociPntDist = \sqrt{\left(Foci_{x,1} - p_x\right)^2 + \left(Foci_{y,1} - p_y\right)^2} + \sqrt{\left(Foci_{x,2} - p_x\right)^2 + \left(Foci_{y,2} - p_y\right)^2} \tag{13}$$

Gaussian (bivariate normal) Distribution Surface

To fit a bivariate Gaussian distribution surface to the residual current data, we use:

$$N\left(\bar{x}, \bar{y}, \sigma_x, \sigma_y, \rho\right) = \mathrm{A} \times \mathrm{e}^{\left(\frac{-Z}{(2(1-\rho^2))}\right)} \tag{14}$$

$$\rho = cov_{u,v}/\sigma_x\sigma_y$$

$$A = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

$$Z = \frac{(x-\bar{x})^2}{\sigma_u^2} + \frac{2\rho(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y}\frac{(y-\bar{y})^2}{\sigma_v^2}$$

Overlap coefficient

In order to compare two residual current distributions with the overlap coefficient, we take a similar, numerical, approach as when calculating the area of the intersection. For each residual current distribution we solve (14) on the same 2 dimensional array. We choose the limits of the array so that it captures most of the volume under the Gaussian surfaces. We then take the pointwise minimum of both distributions. This captures the volume under the surfaces, in a way analogous to the 1-dimensional case shown in Figure 2i. We then integrate this surface – this can be done by summing all the values and multiplying by the area of each array element.

overlapping_ellipse_area, overlapping_ellipse_area_from_dict

overlapping_ellipse_area and overlapping_ellipse_area_from_dict calculate the percentage overlap between two sets of ellipses.

```
overlap_dict = overlapping_ellipse_area_from_dict( dict_1, dict_2 )
```

This calls:

```
overlap_1,overlap_2,overlap_and,overlap_or,area_1,area_2,overlap_1not2,over
lap_2not1,overlap_xor, perc_overlap, perc_ratio_of_overlap, perc_area_rat =
overlapping_ellipse_area(
X_elip_amp_1,X_elip_phi_1,U_mean_1,Y_elip_amp_1,Y_elip_phi_1,V_mean_1,foci_
x_1_1,foci_y_1_1,foci_x_2_1,foci_y_2_1,qmax_1,qmin_1,
X_elip_amp_2,X_elip_phi_2,U_mean_2,Y_elip_amp_2,Y_elip_phi_2,V_mean_2,foci_
x_1_2,foci_y_1_2,foci_x_2_2,foci_y_2_2,qmax_2,qmin_2, npnt_counting = 100)
```

ellipse_overlap_coefficient_pdf, ellipse_overlap_coefficient_pdf_from_dict

ellipse_overlap_coefficient_pdf and ellipse_overlap_coefficient_pdf_from_dict calculate the overlap coefficient (OVL) between two sets of residual currents.

```
OVL_dict = ellipse_overlap_coefficient_pdf_from_dict( dict_1, dict_2 )
```

This calls:

```
gauss_1_2_overlapping_coef,gauss_1_int,gauss_2_int =
ellipse_overlap_coefficient_pdf(
```

```
X_elip_amp_1,X_elip_phi_1,U_mean_1,U_var_1,UV_cov_1,
Y_elip_amp_1,Y_elip_phi_1,V_mean_1,V_var_1,
X_elip_amp_2,X_elip_phi_2,U_mean_2,U_var_2,UV_cov_2,
Y_elip_amp_2,Y_elip_phi_2,V_mean_2,V_var_2, npnt_counting = 100)
```

Description of process

Figure 2c shows hypothetical *u* and *v* residual current data (from a pair of correlated normal distributions). A (red) 2.45 standard deviation ellipse can be fit around the mean (black cross) in Figure 2c. The distance from the mean to the origin can be worked out as a number of standard deviations (Figure 2d).

By fitting tangents between the ellipse and the origin, we can calculate the angular width of the ellipse (assuming the origin is outside the ellipse) (Figure 2e). Note that the angular width is calculated for an ellipse of a given number of standard deviations – the angle will be different for a 1 standard deviation ellipse and a 2.45 standard deviation ellipse. Furthermore, if the origin is within the ellipse, the angle is undefined.

There are two method that we can use to quantify the similarity of two climatologies. The most intuitive of these is to consider the overlap of the two ellipses, but we can also consider the overlap of the two bivariate normal distributions. To calculate the percentage ellipse overlap, we first calculate the area of the 2.45 standard deviation ellipse from one climatology Figure 2g). We then repeat this for a second ellipse, and work out the intersection area (Figure 2h). We can compare the ratio of the uncertainty of the two climatologies (area of one/area of other) and how consistent they are (area of intersection/total area in either ellipse). To compare the overlap of the distributions, we calculate the volume under the distribution surfaces, as illustrated in Figure 2i.

Climatology

The residual current ellipse shows how variable the circulation pattern is. By plotting the number of standard deviations between the mean and the origin (Figure 3) you can show how the variability compares to the mean. When the mean of the residual current is greater than 2.45, the origin is outside the 2.45 standard deviation ellipse, and we consider the current significant (within the white contour on Figure 3). When the current is greater than 1 standard deviation, we consider it to be “quasi-steady”, and this is shown with a grey contour in Figure 3.

When the origin is outside the residual current ellipse, the angular width can be calculated (Figure 4). This is specific to an ellipse of a given number of standard deviations. When the ellipse passes through the origin, this is 180°, and it decreases with distance for the ellipse. In many regions, it is less than 90°, but in the English Channel it is often much higher.

Figure 2a and b make use of both these techniques to highlight the differences in the circulation of two model runs. Firstly, currents that are not quasi-steady are not shown (white). Currents that are significant (c.f. Figure 3) and have an angle less than 90° (c.f. Figure 4) are shown in bold, and currents that are quasi-steady, or significant with an angle greater than 90° are shown with a transparent mask (much lighter colouring).

Comparison of two climatologies

We can use this approach to compare the residual current distributions between two model simulations. There are two related methodologies that we can use: the ellipse overlap, and the Overlap Coefficient. For a given point, we can compare the uncertainty ellipses of the two simulations and calculate their overlap. When the currents agree (in terms of mean, variance, and covariance), the ellipses perfectly align, and the overlap is 100%. We can also compare the similarity of the residual current distributions directly, using the OVL (Overlap Coefficient, Inman and Bradley, 1989). This is less intuitive than the ellipse percentage overlap, but more robust, so we use it in this paper. When comparing to two (scalar) Gaussian curves (e.g. Figure 2i), the overlap area under both curves is a measure of their similarity. When the two curves are identical the OVL is at its maximum of 1 (as the area integrated under a Gaussian curves is one). As the two curves become increasingly different the overlap areas decrease. As Gaussian curves are infinite, the OVL tends to 0, but is always >0. We use a two-dimensional analogue of the OVL to compare our two bivariate normal distributions and rely on the fact that the volume under the Gaussian curve integrates to 1. We note that there is a near linear relationship between the ellipse overlap and the overlap coefficient.

We first describe the percentage ellipse overlap methodology. Figure 2h shows two ellipses, and their overlap. Comparing the area of the intersection of the two ellipses to the total (union) area of the two ellipses (i.e. the sum of the area of the two ellipses minus the intersection area) gives a percentage of how similar the ellipses are. If the two ellipses are identical (in shape, size, and location) this value is 100%, difference in their size, shape or location leads to a lower value, and if they do not overlap, the value is 0%.

Calculating the overlap between two ellipses is not a trivial problem, as there are 10 possible orientations between two ellipses (Hughes & Chraibi, 2012). Instead, we do this numerically. For each ellipse, we take a grid of regularly spaced *u* and *v* values (the same grid for each ellipse) and use (13) to ask if each point is within the ellipse to give a Boolean mask (Figure 2g). We can then compare these arrays of masks to find the intersection.

We also calculate OVL numerically. Again, we take a grid of regularly spaced *u* and *v* values and calculate the Gaussian surfaces for these point with equation (14). We then take the minimum of these two distributions and integrate the volume. We can approximately integrate this simply by summing up all the values and multiplying by the grid box area. This is the approach taken in this study.

We can also compare the area of one ellipse to the other to quantifies the difference in the current variability. The area of the ellipse is easy to calculate using (15).

$$Area(n_{std}) = {n_{std}}^2 q_{max} q_{min} \pi \tag{15}$$

Assessment of an individual year within a climatology.

We can assess how an individual year (exemplar value) compares to the climatological residual current ellipse (Figure 5). This is not used in this study but is included for completeness. If a particular year is outside the uncertainty ellipse we can ask whether a simple change in direction or, magnitude can explain the difference from the climatological values, or whether both are different. Here we use the 1 standard deviation ellipse, as most points are not significantly different from the climatology when using a 2.45 standard deviation ellipse).

We can divide the *u-v* current space into regions based on properties of the ellipse (Figure 5) and ask where the exemplar value fits. If the point fits within the ellipse (white), there is no significant difference between it and the climatological values. It can have a direction and magnitude that fall within the range of the ellipse, but still fall outside the ellipse – this region is shown in purple in Figure 5. In this region, changing either the angle or magnitude of the exemplar current to the climatological mean will make it fall within the ellipse. When the magnitude is correct, but the angle is incorrect, it falls within the green region. When the angle is correct, but the magnitude is incorrect, it falls into the red or blue regions (if the exemplar magnitude is greater than or less than the climatology respectively). If the exemplar current is either in the wrong direction OR magnitude it falls into a coloured (green, blue, red or purple) section. If the current is of the wrong magnitude AND direction, it falls into the grey regions.

When the origin is within the ellipse this approach is not directly applicable, and so using a 1 standard deviation ellipse may be appropriate.

We can use this classification scheme to show how the residual current field of a particular year compares to the climatology. Figure 6 shows how the annual mean current field of 2020 compares to the 1990-2020 CMEMS climatology, using the same colour schemes as in Figure 5.

Probability Levels

For data that fits a Gaussian distributions, there is a related distribution of the distance (squared) between the data points and the mean – the chi-squared distribution. For scalar data (with one degree of freedom), 95% of the data (p=0.05) are less than 1.96 standard deviations of the mean. This is reflected in the chi-squared distribution table, where for p=0.05, 1 d.f., the critical chi-squared value is 3.84, which is the distance squared ($\sqrt{3.84}$ = 1.96 standard deviations).

We can transform bivariate Gaussian data to have zero mean and covariance, and unitary variance. For this transformed data, the distance between a data point and the mean can be described as:

$$u^2 + \mathrm{v}^2 = s^2 \qquad (16)$$

Again, we note that s^2 fits the chi-squared distribution with 2 degrees of freedom. For this transformed data, the uncertainty ellipse become a circle. We can choose the size of this circle (in terms of standard deviations from the mean) to encapsulate a given proportion of the data. Again, the chi-squared distribution table (with 2 degrees of freedom) can describe the distribution of the distance (squared, s^2) to the mean. To encapsulate 95% of the data (p=0.05, 2 d.f.), the chi-squared critical value of 5.99 suggest that the distance to the mean must be ($\sqrt{5.99}$ =) 2.45 standard deviations. This value is unaltered when we transform the data back.

For 90%, 95% and 99% of the data, the ellipse must be 2.15, 2.45 and 3.03 standard deviations (see Table 1 and Figure 7).

We can confirm this numerically with two independent methods. For the first method, we create a large (n=1000) bivariate normally distributed dataset of pseudo *u* and *v* currents (for a given mean, variance and covariance), fit uncertainty ellipses for a range of standard deviations (between 0 and 3 in steps of 0.1) and ask what proportion of the dataset fit within the ellipse. We then cycle through ~1000 combinations of a range of means, variances and covariances. We present the mean of these in Figure 7, with the uncertainty range. For the second approach, for a given mean, variance and covariance, we calculate a discretised bivariate Gaussian surface. We fit a series of uncertainty ellipse (for the same range of standard deviations), and numerically integrate the values within these ellipses. We cycled through the same combinations of means, variances and covariances to confirm that the result wasn't dependent on the underlying details of the ellipse. Both methods confirm the chi-squared approach, and are presented in Figure 7.

Figures and Tables

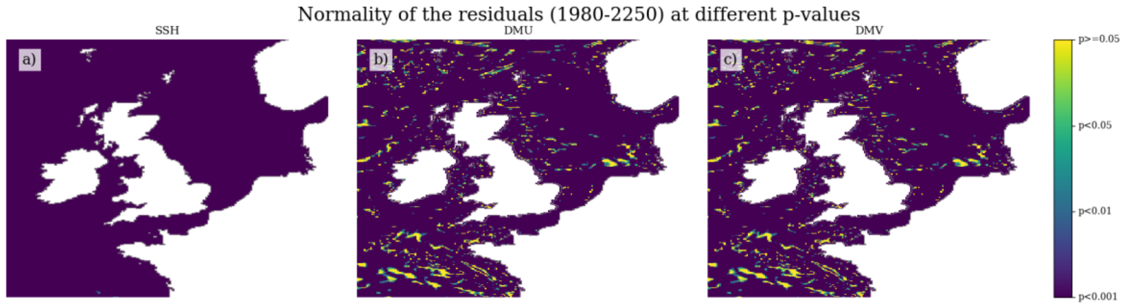

Figure 1 The significance of a test for Normality of tidal residual sea surface height (SSH, a), *u*- and *v*-components of depth-mean velocity (DMU, DMV, b and c respectively), from a 270-year present day control simulation (Tinker et al., 2020) as calculated with monthly mean. The p-value shows where the normality is significant at the 0.1%, 1% and 5% levels, and where it is not significant at the 5% level (shown in yellow).

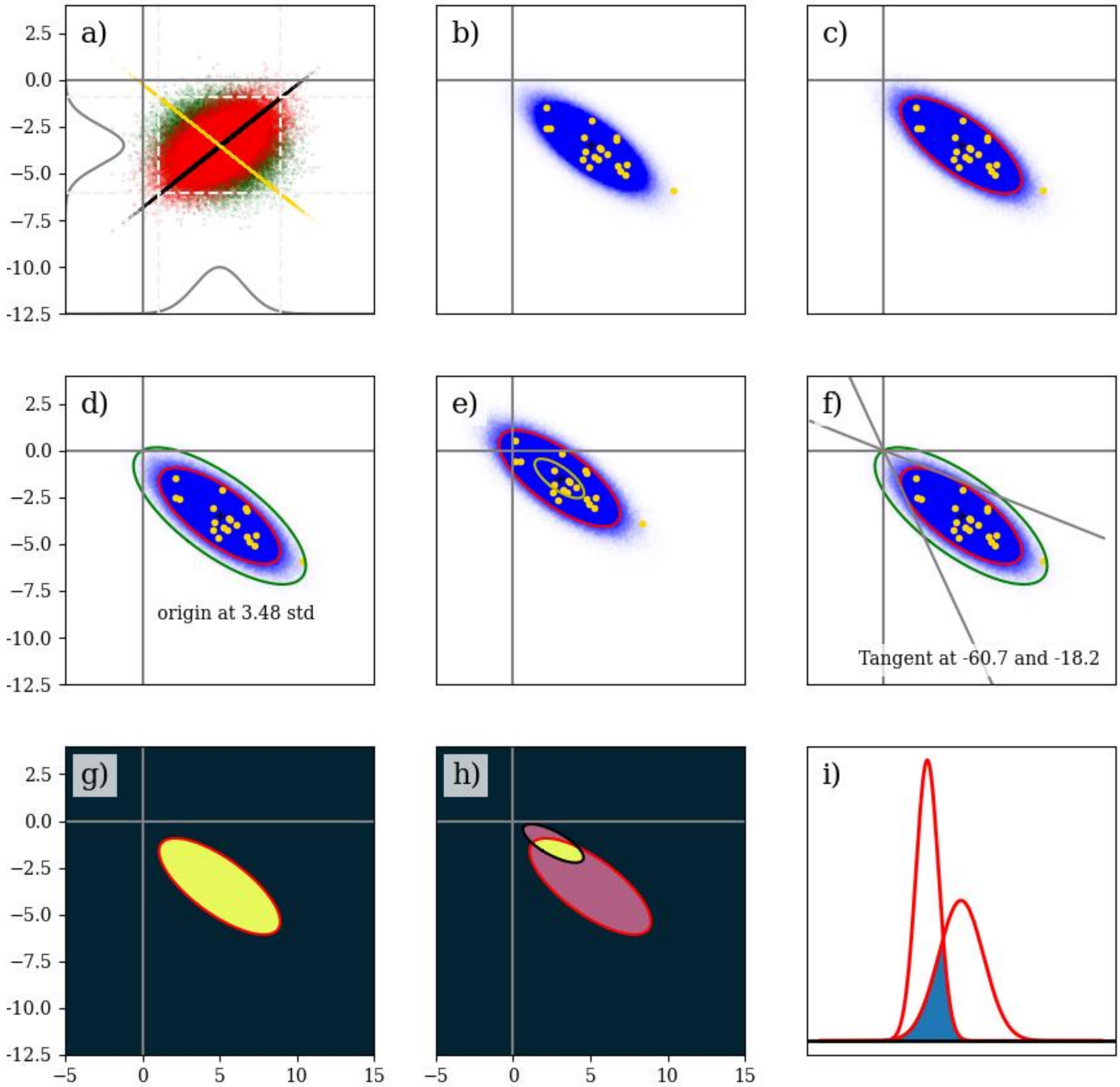

Figure 2 Examples of how to build up the uncertainty ellipse approach. See text for details.

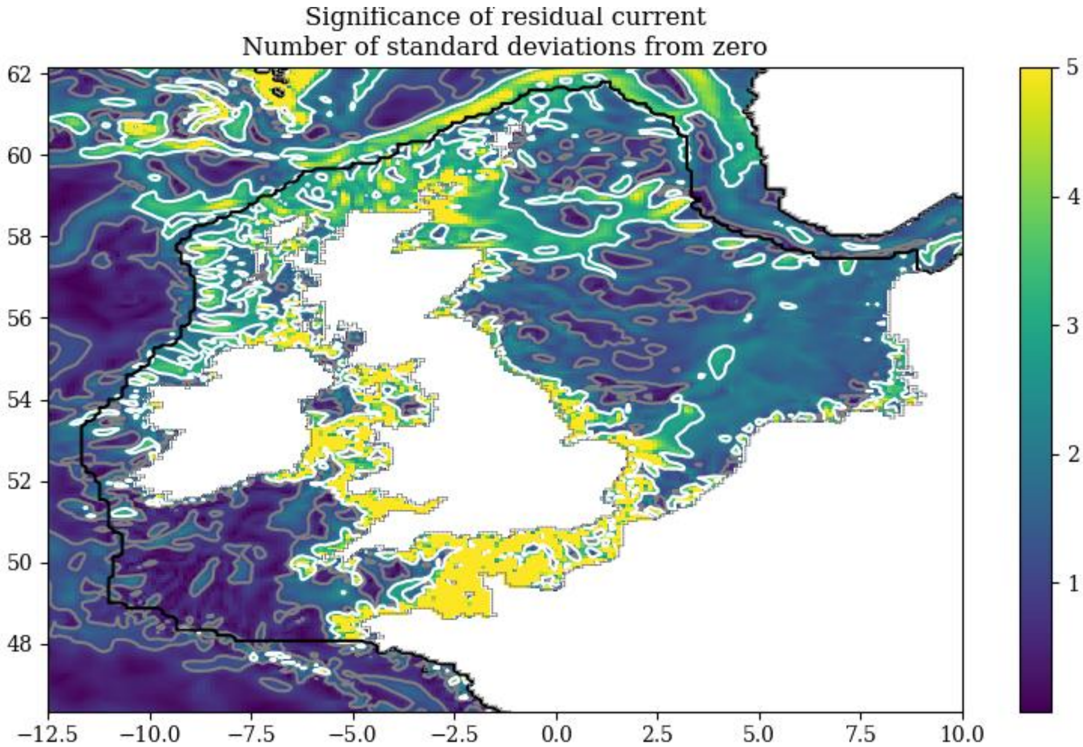

Figure 3 Number of standard deviations between the origin and the mean residual current. When it is greater than 2.45 (white contour), the current is significant.

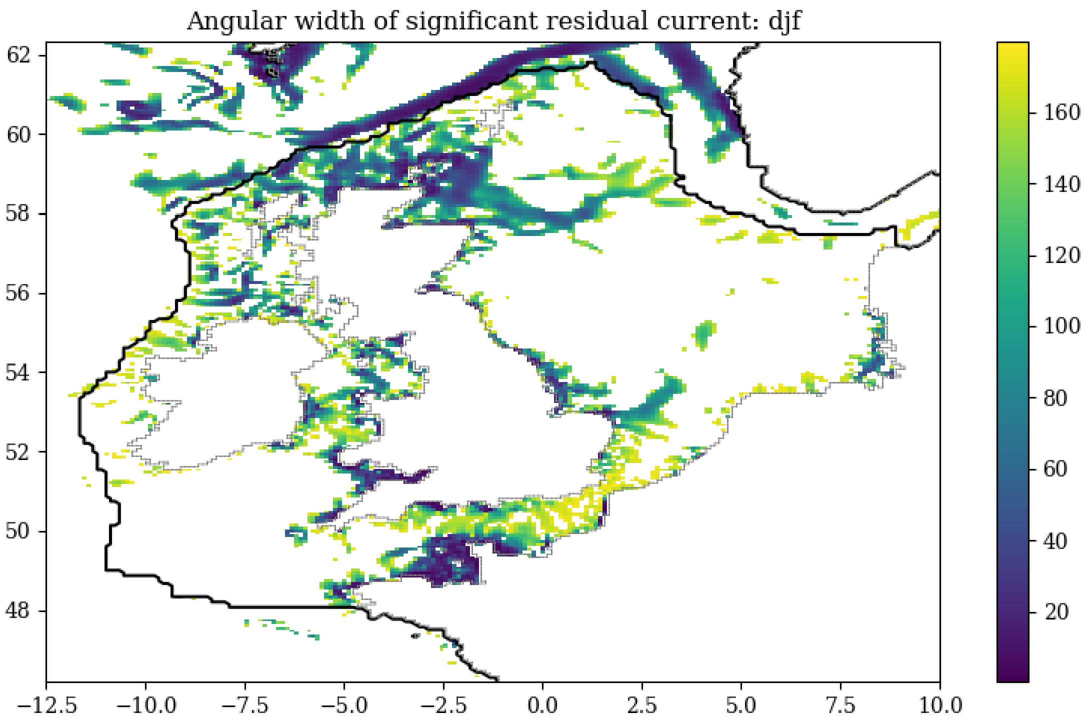

Figure 4 The significant residual circulation (at 2.45 standard deviations), showing the angular width of the 2.45 standard deviation ellipse (the angle between the two tangents of the ellipse that pass through the origin. Note that when the origin is within the ellipse (i.e. the residual circulation is not considered significant), there are no tangents that pass through the origin, and so the angle is undefined).

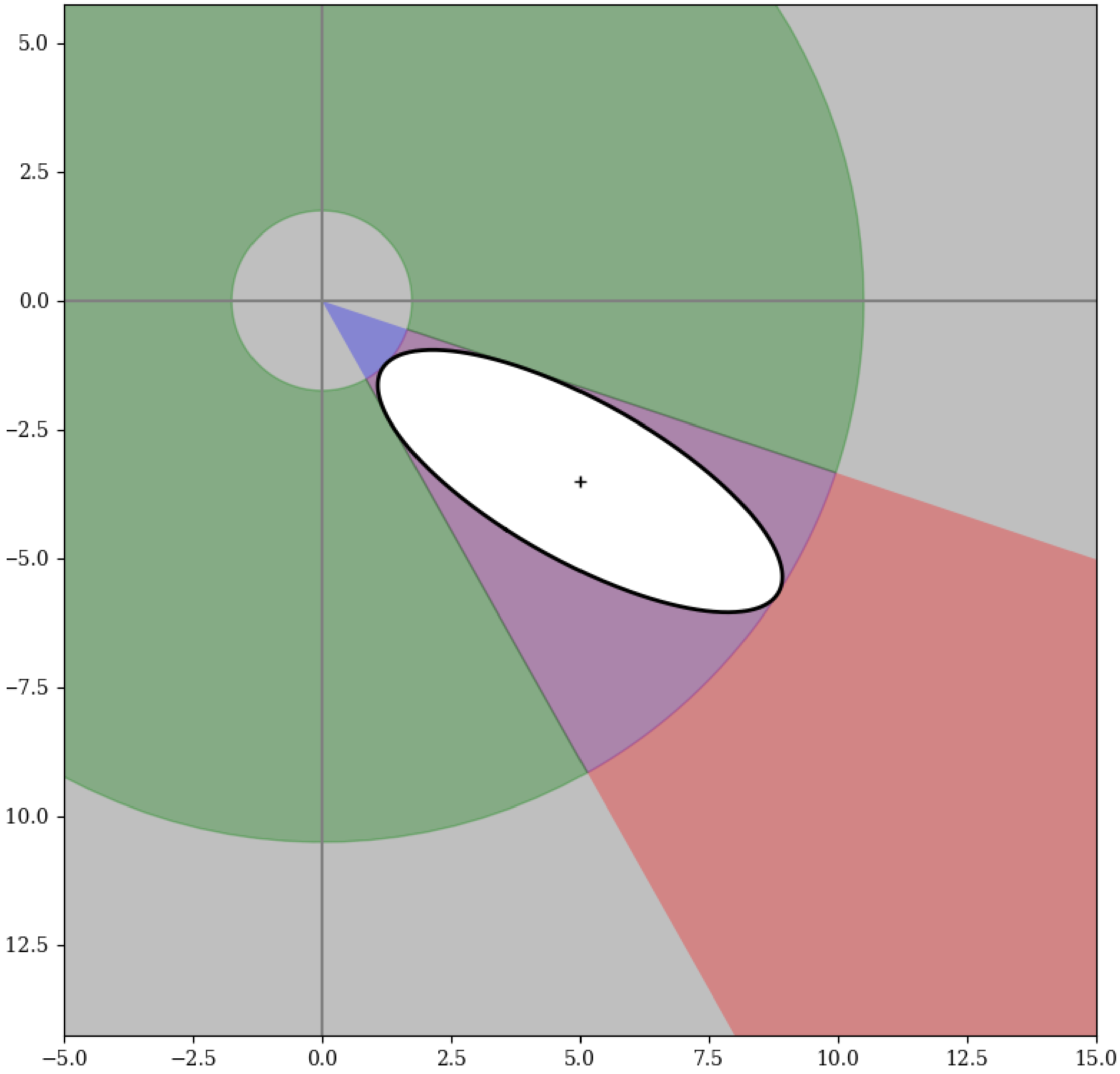

Figure 5 Regions of current space, with respect to the current ellipse. See text for details.

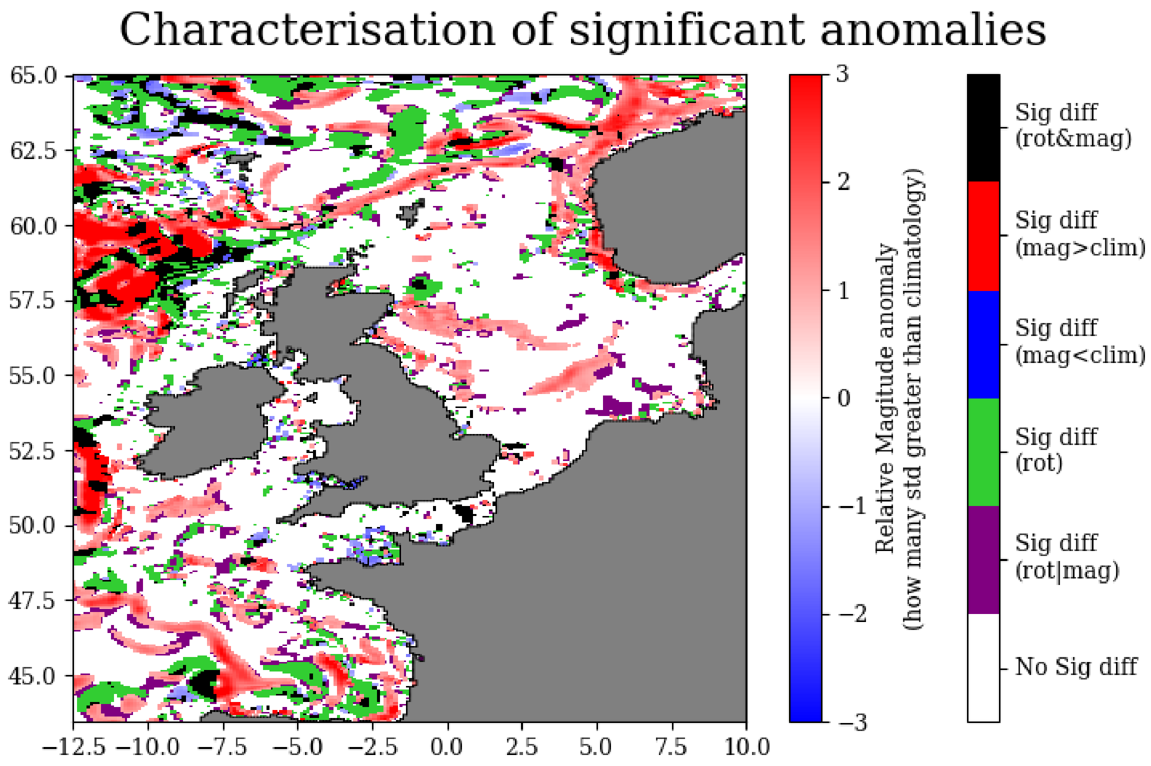

Figure 6 Characterisation of the anomaly significance relative to the climatology.

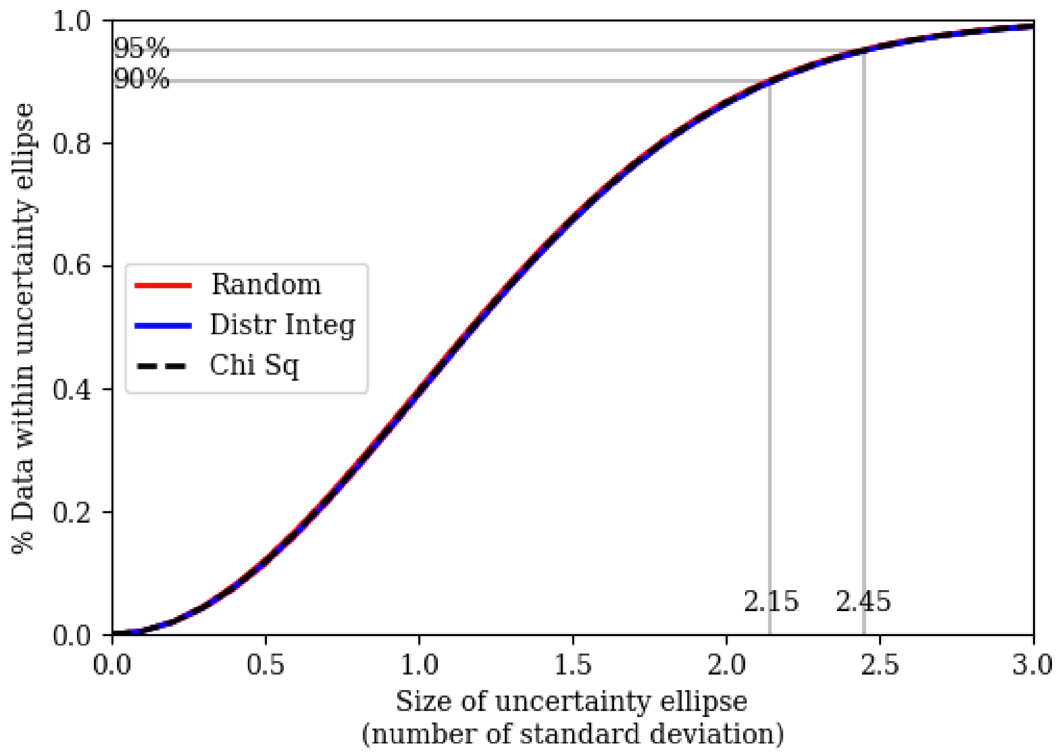

Figure 7 Percentage of data encapsulated by uncertainty ellipses of a given size (in standard deviations). The black dashed line is calculated from the chi-squared distribution table. The red lines are calculated by interrogating a bivariate normally distributed dataset, and asking what proportion is within a series of ellipses. The blue lines are calculated by integrating the area of a bivariate Gaussian distribution surface within a series of ellipses. The values are tabulated in Appendix Table 1.

Table 1 Size of uncertainty ellipse (in terms of standard deviations) required to a given percentage of the data. These values are derived from the chi-squared distribution, which are given in the 3rd and 4th column. Note that the fourth column is second column squared.

Uncertainty Ellipse size (in standard deviations) and data coverage (%)		**Chi-Squared Distribution Table (with 2 degrees of freedom)**	
Percentage of data within Uncertainty Ellipse	**Size of uncertainty ellipse (# standard deviations)**	**Critical value**	**Probability of exceeding the critical value**
50.0%	**1.1774**	**0.500**	**1.386**
75.0%	**1.6651**	**0.250**	**2.773**
90.0%	**2.1460**	**0.100**	**4.605**
95.0%	**2.4477**	**0.050**	**5.991**
97.5%	**2.7162**	**0.025**	**7.378**
99.0%	**3.0349**	**0.010**	**9.210**
99.5%	**3.2552**	**0.005**	**10.597**