AREAL FORECASTING

Science specifications

Christophe Sarran
September 2011

Introduction

It was quickly identified that a definition of areal forecasts was required and the following working definition has been used:

Definition

By areal forecast is meant the characterising of a geographical area by a single meteorological parameter (or set of parameters) such that this parameter describes the area as a whole – and not some part of it. This means that an areal forecast is an average, or more generally a statistic, for a place defined by its area.

Areal forecasts include parameters such as the maximum, minimum or mean value of a parameter for the area, the cumulative sum over time, the probabilities of thresholds being reached within a period of time, etc. Any of these can be calculated using different weighting methods (e.g. population weighting) and kernel functions, both in time and in space. Each of these will be customer dependent, but the overall method of using an area to derive the customer-specified parameter for an areal forecast will be generic.

The capability development vision of the Met Office's areal forecasting products is to deliver areal forecasts for all areas and most parameters specified by a customer. This means that:

- Existing capabilities needed to be improved to allow for small areas and areas with multiple parts;
- Areal forecasts of extremes (minima, maxima) required exact polygon areal processing;
- The final areal forecast files need to be ASCII text files to be portable to a variety of delivery systems.

This document describes how these have been achieved and details the science specifications of the areal processing required for robust and generic areal forecasts.

Summary specifications

To process gridded weather forecasts into area forecasts it is necessary to use a definition of each area in terms of grid points. Therefore a static AFS (Areal Forecast Specifications) file would be used to generate areal forecasts from the gridded forecasts. Creating an accurate AFS file is paramount prior to processing gridded weather forecasts and an AFS utility has been created for this purpose. Figure 1 summarises the process.

Characteristics of the AFS utility:

- For each set of areas and grid specifications, creates an AFS file that contains exact areal specifications – necessary for the forecasting of extremes;

- The utility examines the proportion of the grid point cell (area closest to a grid point) that falls within each area polygon and returns that proportion as weight;

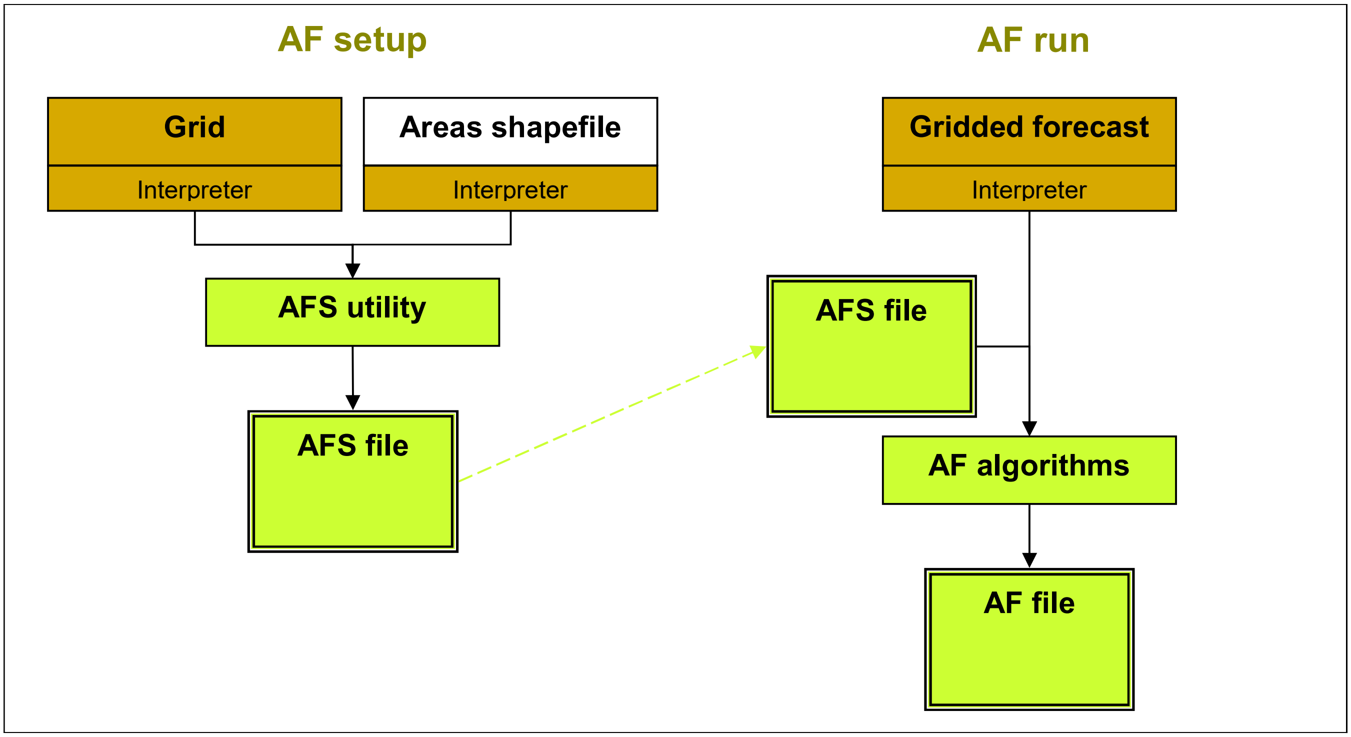


Figure 1. Schematic diagram of proposed Areal Forecasting (AF) specifications, showing (a) the AF setup requiring grid specifications and a customer-provided shapefile defining the areas and producing and Areal Forecast Specifications (AFS) file using an AFS utility, and (b) the AF run which processes the gridded forecast with a set of AF algorithms using the AFS file and producing a final AF file containing the areal forecast; interpreters are used to convert the data in the grid files and shape files to the appropriate inputs to the AFS utility and AF algorithms.

- The AFS utility takes a standard input from interpreted grid and area information – this means that only the interpreters need to be adapted for new formats;

- User-defined masks and gridded weights can be used to change the utility's default weighting.

Characteristics of the AF algorithms:

- Efficient because all the areal processing has been done by the AFS utility;

- Coordinate system independent – as the AFS file specifies areas in terms of grid point indices;

- Modular so as to only run those algorithms that produce areal forecasts that are used, and so that new customer requirements can be easily added.

Example

Upon receipt of customer specifications (area definitions and required parameters), the setting-up of a new Areal Forecast (AF) would consist of two stages (figure 1):

Christophe Sarran
September 2011

- AF setup:

The recommended setup is to use the customer-provided area shape file and the grid specifications of the required gridded forecast to generate an Areal Forecast Specifications (AFS) file. This requires an AFS utility which is now available[1,2] as a compiled Fortran program. The AFS file written by the utility would be held by Locations Management or equivalent information management system.

- AF run:

The specifications held in the AFS file are used to process gridded forecast data to produce an areal forecast that can subsequently be further processed and written to the Areal Forecast (AF) file.

Detailed specifications

The development of the AFS utility has required the setting of specifications for the standard input and output of the utility program. Figure 2 illustrates the structures of the standard input and output. The data that forms the standard input is obtained from the grid specifications and the areas shapefile, each through file format specific interpreters to convert the relevant file information into the required utility input. At least one interpreter is needed for each grid file format. The interpreter for the areas shapefile does not only depend on the shapefile format but also on the structure given to the file by the person who wrote it – there is therefore no standard format of the ESRI shapefile in terms of the data required by the AFS utility.

Parameters:

gtype	Horizontal grid type (0=NG, 1=lat/lon, 2=space view, 3=polar stereographic, 4=x/y grid)	udw	User-defined weights (0=no, 1=yes)
ny	Number of rows in field	nr	Number of Records
nx	Number of columns in field	rn	Record Number
yi	Northing or latitude or start line of first row of data	cl	Content Length
dy	Interval between rows	x	X coordinate
xi	Easting or longitude or start pixel of first point of first row of data	y	Y coordinate
		ngp	Number of grid points
dx	Interval between columns	gpid	Grid point index
order	Order of the grid points (0=contiguous x, 1=contiguous y)	w	Weight

Standard input:

The first 8 parameters, gtype to order, are the grid specifications; for example a UKPP grid specification may be: 0 704 547 1222000 -2000 -238000 2000 0

[1] Currently limited to orthogonal grids.
[2] *[removed]*

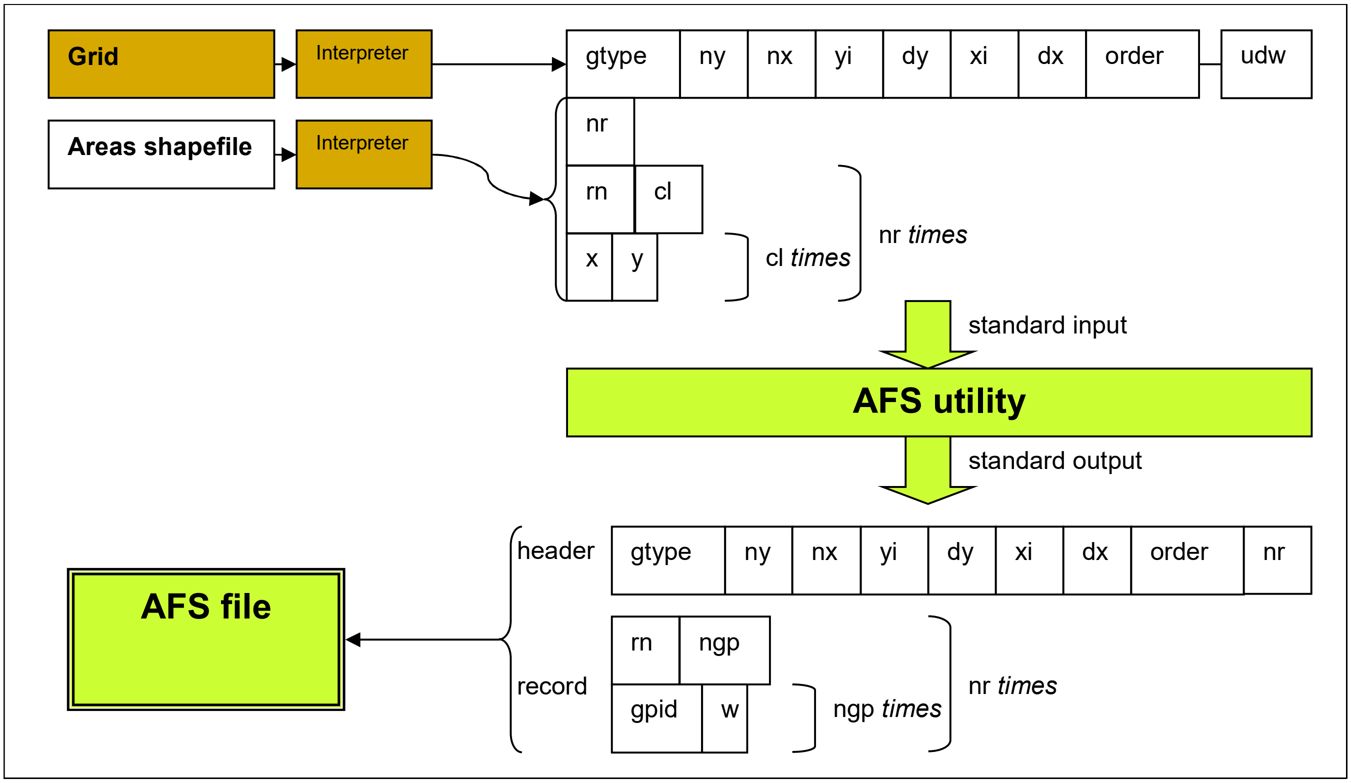


Figure 2. Schematic diagram of the standard input and output required by the AFS utility; the standard input includes the grid specifications that may be obtained from a grid file, and the area polygon identifiers and vertex coordinates from a shapefile; the standard output consists of a header and a set of records that are saved to an AFS file.

If udw is set to 1, then the utility expects, immediately after udw and before nr, the gridded weights for the grid points and in the order set by the grid specifications. For the example UKPP grid specification above, 704 × 547 = 385088 weight values would be expected before specifying the number of records nr.

The remaining parameters are normally obtained from the areas shapefile. First the number of records nr is specified (nr is required on a separate line at standard input). Then for each record, the record number rn and content length cl is obtained from the shapefile. The content length specifies the number of vertices that the polygon is defined by, and the number of x and y vertex coordinates expected by the utility.

For example, the first five lines of the standard input may be:

```
0 704 547 1222000 -2000 -238000 2000 0 0
380
1 4292
84011 5378
84009 5374
…
```

Line 1 specifies a National Grid (NG) system with 704 rows and 547 columns, starting at -238000 east and 1222000 north and with -2000 m intervals between rows and 2000 m intervals between columns. The grid points are ordered by contiguous x and there are no user-defined weights. The second line specifies that there are 380 area polygons. The next line has the first polygon record number (1) and the number of vertices (4292) it is defined by. The next two lines are the x and y coordinates of the first two vertices.

Standard output:

The standard output is used to write the AFS file. It consists of a header that contains the grid specifications (gtype to order) and the number nr of areas. Following the same example as above, the header would be: 0 704 547 1222000 -2000 -238000 2000 0 380

Following the header are the AFS records, each consisting of a record header and a list of grid point indices and associated weights. The record header specifies an area identifier rn and the number ngp of grid points associated with the area – there are hence that number of grid point indices gpid and associated weights w following the record header.

For example, the first four lines of the AFS file may be:

```
0 704 547 1222000 -2000 -238000 2000 0 380
1 31
329459 0.041129436
329460 0.045815587
…
```

Line 1 is the header which parameters have already been described above. The next line has the first polygon record number (1) and the number of grid points (31) associated with that polygon. The next two lines are the indices and associated weights of the first two grid points. In this example, no user-defined weights were used: the weight values mean that 4.11% of grid point 329459 and 4.58% of grid point 329460 are associated with polygon 1. (Specifically, in this example, 4.11% of the area closest to grid point 329459 covers Round Island, Men-a-vaur and part of Saint Helen’s, and 4.58% of 329460 covers White Island, Pernagie Isle, Plumb Island and part of Saint Martin’s, where polygon 1 is the local authority district of the Isles of Scilly.)

The AFS file can then be used directly by the Areal Forecast (AF) algorithms to process gridded data as illustrated in figure 3. Again, the gridded forecasts require interpreters to transform the data into the format specific to the standard input of the AF algorithms. There will be as many AFS files as there are grid specifications and customer specifications of sets of area definitions: it is therefore essential that the gridded data match the grid specifications of the AFS file header.

Specifications of geometry

Each area polygon that is specified at the AFS utility standard input is processed as follows to generate the weighted list of grid points at the standard output. How this has been achieved computationally is described in the fully commented AFS utility program code available at *[removed]*.

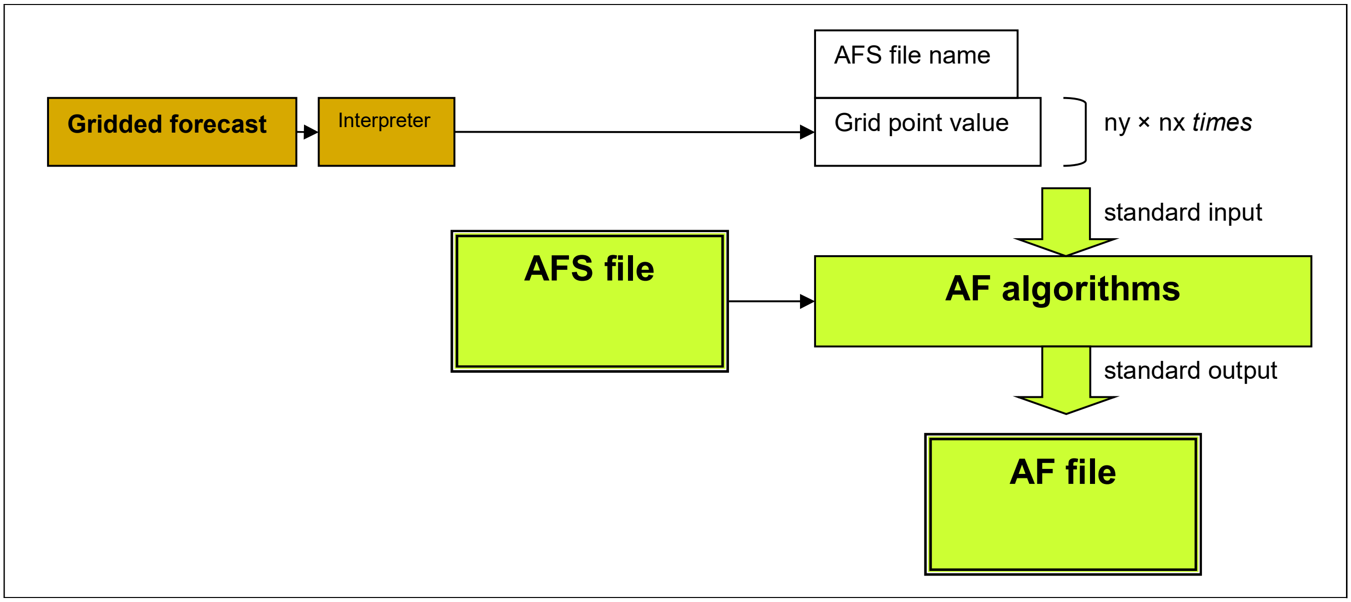
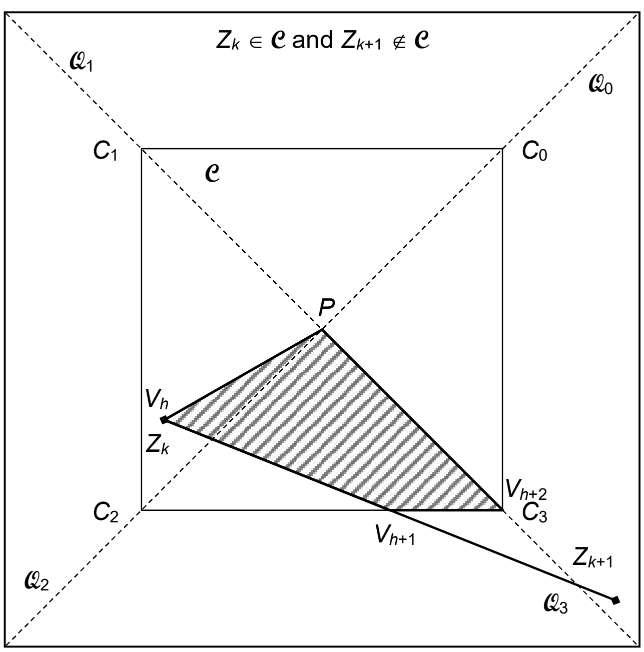


Figure 3. Schematic diagram of the proposed standard input required by the AF algorithms; the standard input includes the name of the AFS file to be used and the grid point values obtained from a grid file; the standard output consists of the Areal Forecasts for each area as specified by the AFS file and are saved to an AF file.

Let a polygon $\mathscr{P}$ be defined by a series of n points Z_k with coordinates $z_k \in \mathbb{C}$ in the complex plane such that $z_{k+n} = z_k$ with $(k;n) \in \mathbb{Z}^2$.

Let a grid point P of coordinate $g_0 \in \mathbb{C}$ have an associated square unit cell $\mathcal{C}$ defined by four corners C_x with coordinates $c_x \in \mathbb{C}$ such that $c_{x+4} = c_x$ with $x \in \mathbb{Z}$ and given by:



$$c_x = g_0 + \tfrac{1}{\sqrt{2}} e^{\frac{i\pi}{4}(2x+1)}$$

Furthermore, let four quadrants $\mathcal{Q}_x$ delimited by $[PC_{x-1})$ and $[PC_x)$ consist of all points of coordinate $q \in \mathbb{C}$ given by $q = g_0 + |q - g_0| e^{i\theta}$ with $\theta \in \left[\tfrac{\pi}{4}(2x-1); \tfrac{\pi}{4}(2x+1)\right]$.

Let the m points V_h with coordinates $v_h \in \mathbb{C}$ such that $v_{h+m} = v_h$ with $(h;m) \in \mathbb{Z}^2$ define the intersection $\mathcal{C} \cap \mathscr{P}$ of the cell with the polygon.

For every two points Z_k and Z_{k+1} of the polygon $\mathscr{P}$ the coordinates v_h of the points V_h are given as follows.

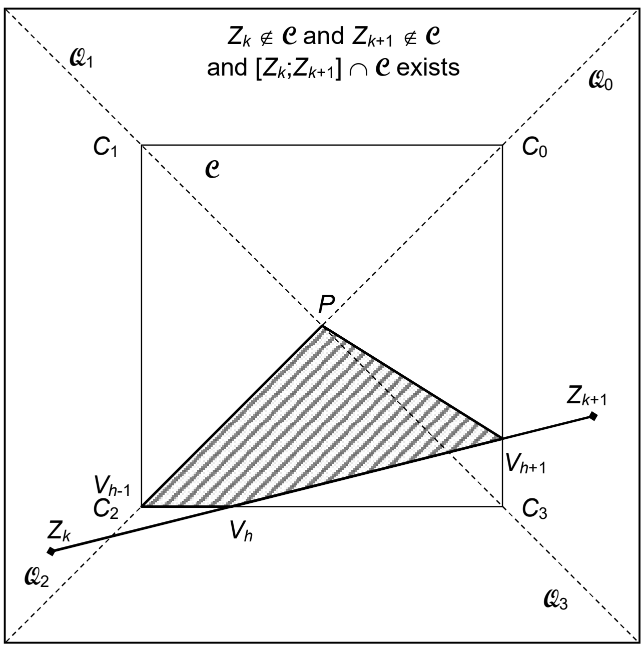
Christophe Sarran
September 2011

If $Z_k \in \mathcal{C}$, the point Z_k is in the cell, then $v_h = z_k$.

If $Z_k \in \mathcal{C}$ and $Z_{k+1} \notin \mathcal{C}$, the point Z_k is in the cell and the point Z_{k+1} is outside the cell, then there is one additional point $V_{h+1} \in [Z_k Z_{k+1}] \cap [C_x C_{x+1}]$ on the cell boundary with coordinate v_{h+1} given by:

$$\begin{cases} v_{h+1} = z_k + \dfrac{z_{k+1} - z_k}{\mathrm{Re}(z_{k+1}) - \mathrm{Re}(z_k)} \left[\mathrm{Re}(g_0) - \mathrm{Re}(z_k) + \dfrac{\mathrm{Re}(z_{k+1}) - \mathrm{Re}(z_k)}{2\left|\mathrm{Re}(z_{k+1}) - \mathrm{Re}(z_k)\right|} \right] \\ \text{or} \\ v_{h+1} = z_k + \dfrac{z_{k+1} - z_k}{\mathrm{Im}(z_{k+1}) - \mathrm{Im}(z_k)} \left[\mathrm{Im}(g_0) - \mathrm{Im}(z_k) + \dfrac{\mathrm{Im}(z_{k+1}) - \mathrm{Im}(z_k)}{2\left|\mathrm{Im}(z_{k+1}) - \mathrm{Im}(z_k)\right|} \right] \end{cases} \qquad \left| v_{h+1} - g_0 \right| \le \tfrac{1}{\sqrt{2}}$$

And furthermore, if the points V_{h+1} and Z_{k+1} belong to different quadrants $\mathcal{Q}_x$ and $\mathcal{Q}_{x+1}$ then there is one more point V_{h+2} at the cell corner C_x with coordinate $v_{h+2} = c_x$ given by:



$$v_{h+2} = g_0 + \tfrac{1}{\sqrt{2}} e^{\frac{i\pi}{4}(y+y')}$$
$$|y - y'| = 1 \pmod 8$$

y and y' correspond to the quadrants that V_{h+1} and Z_{k+1} belong to, respectively, $(y;y') \in \mathbb{Z}^2$, $V_{h+1} \in \mathcal{Q}_y$ and $Z_{k+1} \in \mathcal{Q}_{y'}$, such that:

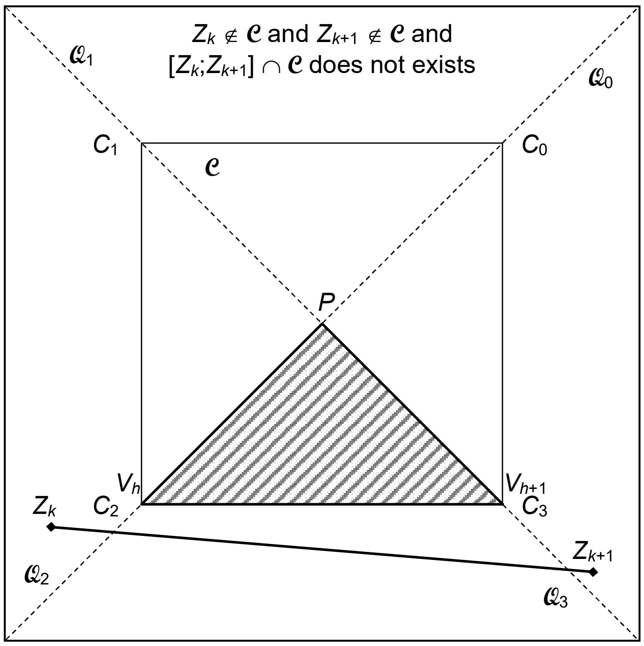
$$\begin{cases} \arg(v_{h+1} - g_0) \in \left[\tfrac{\pi}{4}(y-1); \tfrac{\pi}{4}(y+1) \right] \\ \arg(z_{k+1} - g_0) \in \left[\tfrac{\pi}{4}(y'-1); \tfrac{\pi}{4}(y'+1) \right] \end{cases}$$

If $Z_k \notin \mathcal{C}$ and $Z_{k+1} \notin \mathcal{C}$, both the points Z_k and Z_{k+1} are outside the cell, and if the segment $[Z_k Z_{k+1}]$ crosses the cell $\mathcal{C}$, there exists points V_h and V_{h+1} on the cell boundary such that $[V_h V_{h+1}] = [Z_k Z_{k+1}] \cap \mathcal{C}$ with coordinates v_h and v_{h+1} given by:

$$\begin{cases} v_h = z_k + \dfrac{z_{k+1} - z_k}{\mathrm{Re}(z_{k+1}) - \mathrm{Re}(z_k)} \left[\mathrm{Re}(g_0) - \mathrm{Re}(z_k) - \dfrac{\mathrm{Re}(z_{k+1}) - \mathrm{Re}(z_k)}{2\left|\mathrm{Re}(z_{k+1}) - \mathrm{Re}(z_k)\right|} \right] \\ \text{or} \\ v_h = z_k + \dfrac{z_{k+1} - z_k}{\mathrm{Im}(z_{k+1}) - \mathrm{Im}(z_k)} \left[\mathrm{Im}(g_0) - \mathrm{Im}(z_k) - \dfrac{\mathrm{Im}(z_{k+1}) - \mathrm{Im}(z_k)}{2\left|\mathrm{Im}(z_{k+1}) - \mathrm{Im}(z_k)\right|} \right] \end{cases} \qquad \left| v_h - g_0 \right| \le \tfrac{1}{\sqrt{2}}$$

Christophe Sarran
September 2011

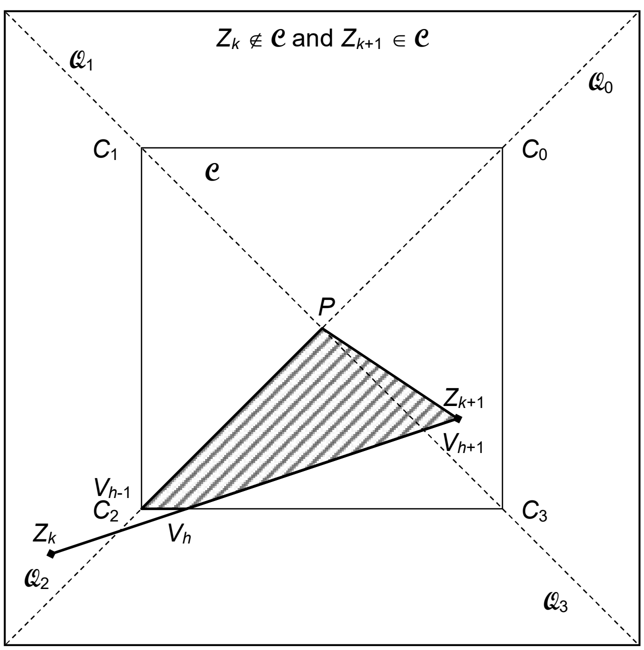
$$\begin{cases} v_{h+1} = z_k + \dfrac{z_{k+1}-z_k}{\mathrm{Re}(z_{k+1})-\mathrm{Re}(z_k)}\left[\mathrm{Re}(g_0)-\mathrm{Re}(z_k)+\dfrac{\mathrm{Re}(z_{k+1})-\mathrm{Re}(z_k)}{2\left|\mathrm{Re}(z_{k+1})-\mathrm{Re}(z_k)\right|}\right] \\ \text{or} \\ v_{h+1} = z_k + \dfrac{z_{k+1}-z_k}{\mathrm{Im}(z_{k+1})-\mathrm{Im}(z_k)}\left[\mathrm{Im}(g_0)-\mathrm{Im}(z_k)+\dfrac{\mathrm{Im}(z_{k+1})-\mathrm{Im}(z_k)}{2\left|\mathrm{Im}(z_{k+1})-\mathrm{Im}(z_k)\right|}\right] \end{cases} \qquad \left|v_{h+1}-g_0\right| \le \tfrac{1}{\sqrt{2}}$$



Furthermore, if the points Z_k and V_h belong to different quadrants $\mathcal{Q}_x$ and $\mathcal{Q}_{x+1}$ then there is one more point V_{h-1} at the cell corner C_x with coordinate $v_{h-1} = c_x$ given by:

$$v_{h-1} = g_0 + \tfrac{1}{\sqrt{2}} e^{\frac{i\pi}{4}(y+y')}$$
$$\left|y-y'\right| = 1 (\mathrm{mod}\, 8)$$

y and y' correspond to the quadrants that Z_k and V_h belong to. Similarly, if the points V_{h+1} and Z_{k+1} belong to different quadrants then there is one more point V_{h+2} at a cell corner, with coordinate v_{h+2} given in the same way as for v_{h-1}, with y and y' corresponding to the quadrants that V_{h+1} and Z_{k+1} belong to.



If $Z_k \notin \mathcal{C}$ and $Z_{k+1} \notin \mathcal{C}$, both the points Z_k and Z_{k+1} are outside the cell, and if the segment $[Z_k Z_{k+1}]$ does not cross the cell, $[Z_k Z_{k+1}] \cap \mathcal{C}$ does not exist, then if the points Z_k and Z_{k+1} belong to different quadrants there are one point, V_h, or two points, V_h and V_{h+1}, at cell corners. With y and y'' corresponding to the quadrants that Z_k and Z_{k+1} belong to, respectively, $(y;y'') \in \mathbb{Z}^2$, $Z_k \in \mathcal{Q}_y$ and $Z_{k+1} \in \mathcal{Q}_{y''}$, if $\left|y-y''\right| = 1 (\mathrm{mod}\, 4)$ then there is one point V_h with coordinate v_h given by:

$$v_h = g_0 + \tfrac{1}{\sqrt{2}} e^{\frac{i\pi}{4}(y+y'')} \qquad \left|y-y''\right| = 1 (\mathrm{mod}\, 8)$$

However, if $\left|y-y''\right| = 2 (\mathrm{mod}\, 4)$ then there are two possibilities. If the measure of the angle

$\left(\overrightarrow{PZ_k};\overrightarrow{PZ_{k+1}}\right)$ is positive, then there are two points V_h and V_{h+1} with coordinates v_h and v_{h+1} given by:

$$v_h = g_0 + \tfrac{1}{\sqrt{2}} e^{\frac{i\pi}{4}(y+y')} \qquad y' - y = 1 (\bmod 8)$$

$$v_{h+1} = g_0 + \tfrac{1}{\sqrt{2}} e^{\frac{i\pi}{4}(y'+y'')} \qquad y'' - y' = 1 (\bmod 8)$$

Alternatively, if the measure of $\left(\overrightarrow{PZ_k};\overrightarrow{PZ_{k+1}}\right)$ is negative, then there are two points V_h and V_{h+1} with coordinates v_h and v_{h+1} given by:

$$v_h = g_0 + \tfrac{1}{\sqrt{2}} e^{\frac{i\pi}{4}(y+y')} \qquad y - y' = 1 (\bmod 8)$$

$$v_{h+1} = g_0 + \tfrac{1}{\sqrt{2}} e^{\frac{i\pi}{4}(y'+y'')} \qquad y' - y'' = 1 (\bmod 8)$$

If $Z_k \notin \mathcal{C}$ and $Z_{k+1} \in \mathcal{C}$, the point Z_k is outside the cell and the point Z_{k+1} is in the cell, then there is a point $V_h \in [Z_k Z_{k+1}] \cap [C_x C_{x+1}]$ on the cell boundary with coordinate v_h given by:

$$\begin{cases} v_h = z_k + \dfrac{z_{k+1} - z_k}{\mathrm{Re}(z_{k+1}) - \mathrm{Re}(z_k)} \left[\mathrm{Re}(g_0) - \mathrm{Re}(z_k) - \dfrac{\mathrm{Re}(z_{k+1}) - \mathrm{Re}(z_k)}{2\left|\mathrm{Re}(z_{k+1}) - \mathrm{Re}(z_k)\right|} \right] \\ \text{or} \\ v_h = z_k + \dfrac{z_{k+1} - z_k}{\mathrm{Im}(z_{k+1}) - \mathrm{Im}(z_k)} \left[\mathrm{Im}(g_0) - \mathrm{Im}(z_k) - \dfrac{\mathrm{Im}(z_{k+1}) - \mathrm{Im}(z_k)}{2\left|\mathrm{Im}(z_{k+1}) - \mathrm{Im}(z_k)\right|} \right] \end{cases} \qquad \left|v_h - g_0\right| \le \tfrac{1}{\sqrt{2}}$$

And furthermore, if the points Z_k and V_h belong to different quadrants $\mathcal{Q}_x$ and $\mathcal{Q}_{x+1}$ then there is one more point V_{h-1} at the cell corner C_x with coordinate $v_{h-1} = c_x$ given by:

$$v_{h-1} = g_0 + \tfrac{1}{\sqrt{2}} e^{\frac{i\pi}{4}(y+y')} \qquad \left|y - y'\right| = 1 (\bmod 8)$$

y and y' correspond to the quadrants that Z_k and V_h belong to, respectively, $Z_k \in \mathcal{Q}_y$ and $V_h \in \mathcal{Q}_{y'}$.

Having defined the intersection $\mathcal{C} \cap \mathcal{P}$ of the cell with the polygon with the m points V_h, the surface area a of the intersection $\mathcal{C} \cap \mathcal{P}$ is given by:

$$a = \sum_{h=1}^{m} \frac{\left|v_h - g_0\right| \times \left|v_{h+1} - g_0\right|}{2} \sin\left(\arg\left(v_h - g_0\right) - \arg\left(v_{h+1} - g_0\right)\right)$$

The surface areas can be adjusted with the user-defined weights if required to produce the weights associated with each grid point for each area polygon.

Christophe Sarran
September 2011

Further Science developments

It is intended to further develop the Science capabilities for areal forecasting by:

- Extending the types of grids processed by the AFS utility to grids aligned with spherical coordinates (latitude-longitude);
- Create a Fortran algorithm to generate areal means from UKPP gridded data (temperature and precipitation rate);
- Create a Fortran algorithm to generate 24 h temporal means (0 h to 0 h) from the areal means.

Other developments will be dependent on precise stakeholder requirements.