

Understanding natural hazards through space-time data mining with ArcGIS Pro

Overview

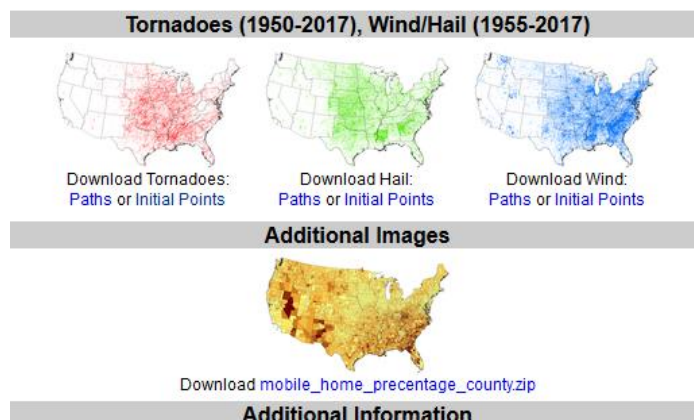
In this exercise, we will explore some of the space-time data mining functionality in ArcGIS Pro, looking at these in relation to natural hazards. Natural hazard risk assessment and management is a sub-discipline of environmental management, relevant both in managing new developments (e.g. flood risk for critical infra-structure such as hospitals, water treatment works, and power stations) but also natural resources too. For example, storm events affect the canopy structure of forests, with hurricane damage creating gaps where mature trees may be blown down, changing the mosaic of successional stages of forest. These tools can then thus be helpful in planning natural resource management as well as planning of the built environment by characterising changing patterns of natural hazards.

The data

The data for this exercise are taken from the US Storm Prediction Centre web site:

<https://www.spc.noaa.gov/gis/svrgis/>:

- Initial points for hail (i.e. where it was first recorded)
- Initial points for extreme wind events
- Initial points for tornadoes



There are also the paths of hailstorms, tornadoes etc too, as well as a polygon shapefile showing the distribution of mobile homes. For this exercise, choose one of the three sets of initial point data, and download it, then unzip it to a suitable folder. You may want to download and unzip some of the other map layers as well by way of extra information.

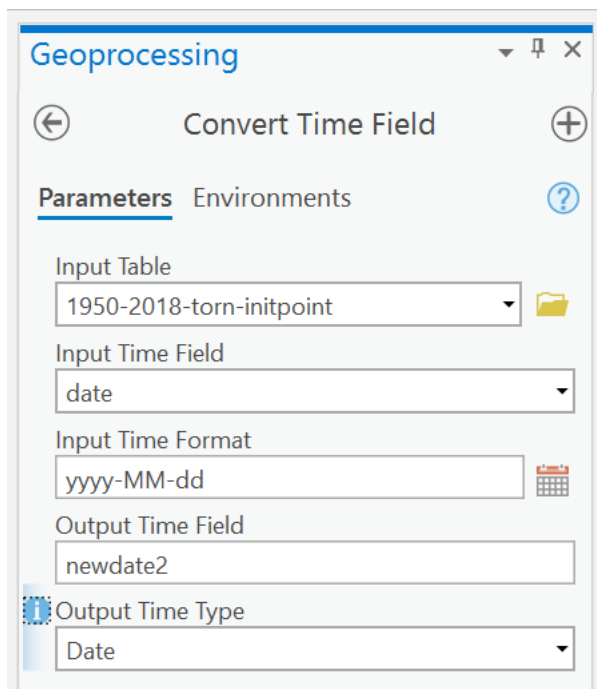
Instructions:

Begin by downloading the data, then exploring it using ArcGIS Pro (note: these instructions should also more or less work with ArcGIS Desktop, provided you have a copy of version 10.3 or later). For this exercise, we will work with the **initial points** for tornadoes (i.e. the locations where each tornado first formed). You may also be interested in downloading and viewing some of the other data sets (e.g. hailstorms or tornado tracks), but we will not be analysing these any further.

Take a look at the attributes of the initial points where tornadoes formed (**1950-2018-torn-initpoint**). One way of doing this is to open-up the attribute table (right-click on the layer, then choose *attribute table*). Notice that whilst there is information about the date and time when the tornadoes were observed forming, this is currently held in an attribute field of type string or text. You can see this by hovering your mouse over the header of the 'date' field and you should see information about the field, particularly that it is type 'text(80)'. A string field holds data as text, such as 'GIS123', so '12th November' would be stored as a set of letters and numbers, but not in a way where dates could be ordered chronologically. 'text(80)' means 80 characters – numbers, letters or punctuation – can be stored in total in each row. Alternatively, you can right-click on any field header and choose *fields* to bring up information about all fields. Make sure you close down the attribute field before going any further, since using tools on tables that are open sometimes causes problems.

For visualisation and further spatio-temporal analysis of these data, we need to have an attribute field that is held in a date or time format, not a string format. The most straightforward way to convert this is via the 'convert time field', which will create a new attribute field with dates held not as strings but in a date format (which can be ordered chronologically).

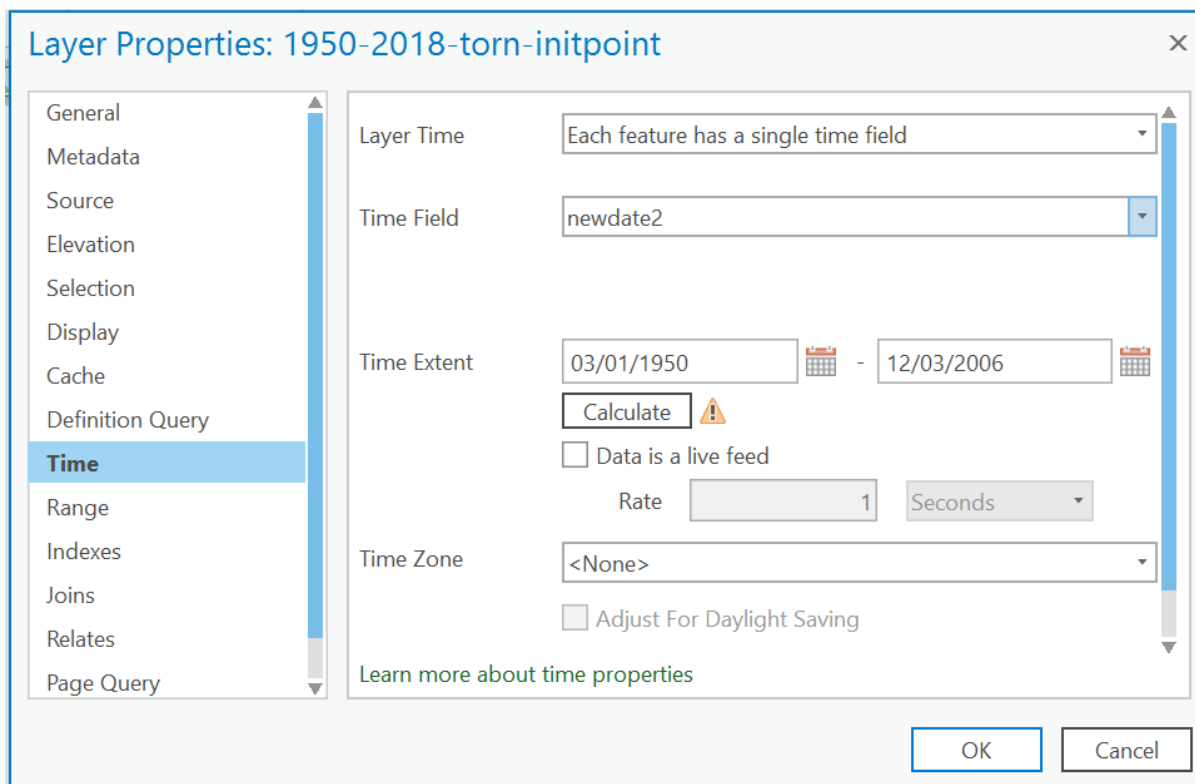
To run this tool, head for the *analysis* menu, then *tools*, and then search for 'time field':



Choose the string field **date** as the *input time field*. Choose *Date* as the *output time type* and give the new *output time field* a suitable name. As you may have noted earlier, in the input 'date' field, we had four digit years (e.g. 2010), two digit months, and then two digit dates (both with leading zeros, so January is '01' for example). Each is separated by '-'. We need to explain this format to the tool for it to process the data successfully. The 'i' next to *input time field* gives more help on this, but yyyy, MM, and dd describe these three formats to the tool, whilst '-' explains how each date component is separated. If you run the tool and then look at the tornado attribute field again, you should find that you now have a new date field added.

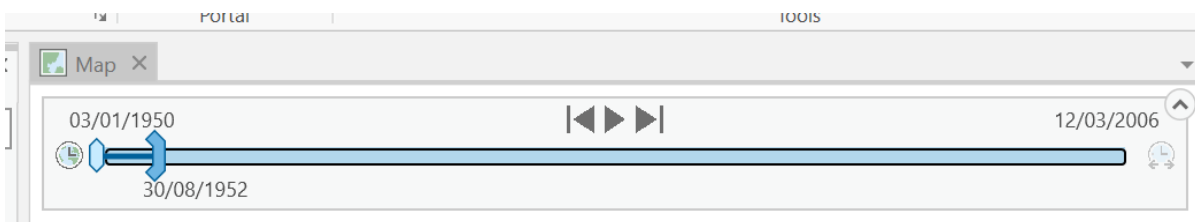
Visualising tornado patterns over time

We can now use this time field to visualise changing patterns of tornadoes over time. To do this, right click on the layer in the left-hand panel, choose *properties*, and then *time*:



If you choose *each layer has a single time field*, you should then be able to choose the new field that you just created as the *time field*. The *Time Extent* should then automatically be calculated from the values in this field.

Once you have done this, thereby enabling time on this point layer, you should now be able to animate the temporal aspects of this map layer via a time slider that appears at the top of your map display:

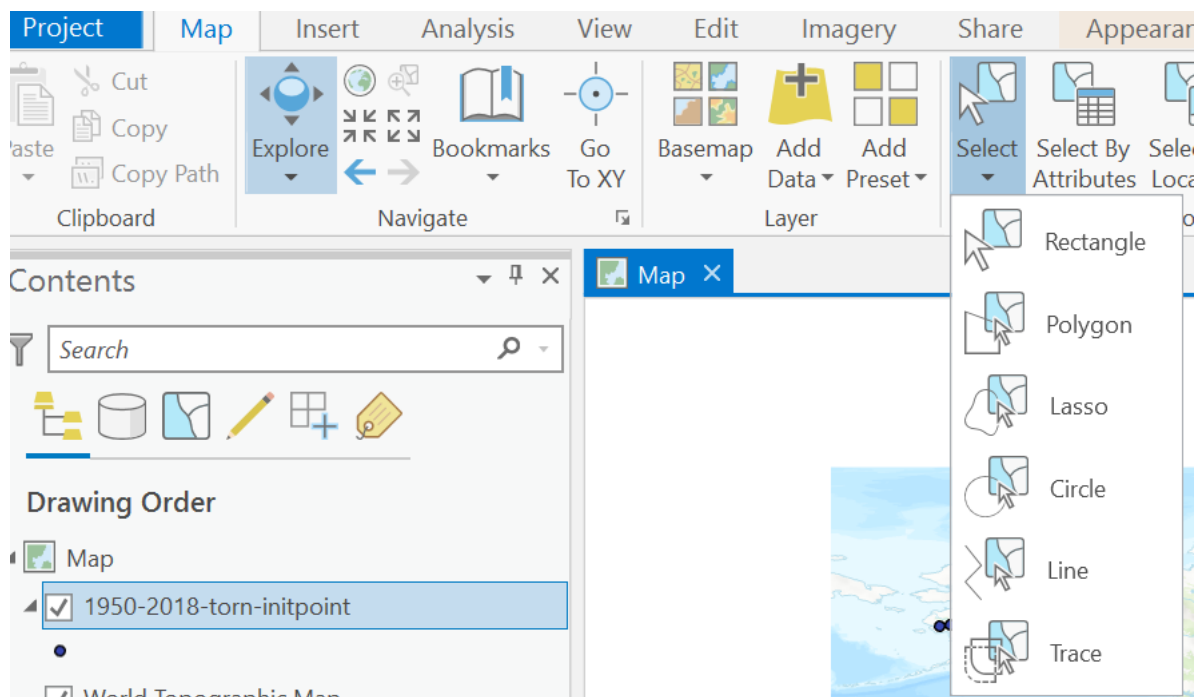


Move the time slider back and forth, and its position will represent a date between the earliest and most recent tornado event recorded. When you move this slider, only points that occurred on or before this date will be displayed. Note also that there is now a context-specific *Time* menu available for our tornadoes map layer.

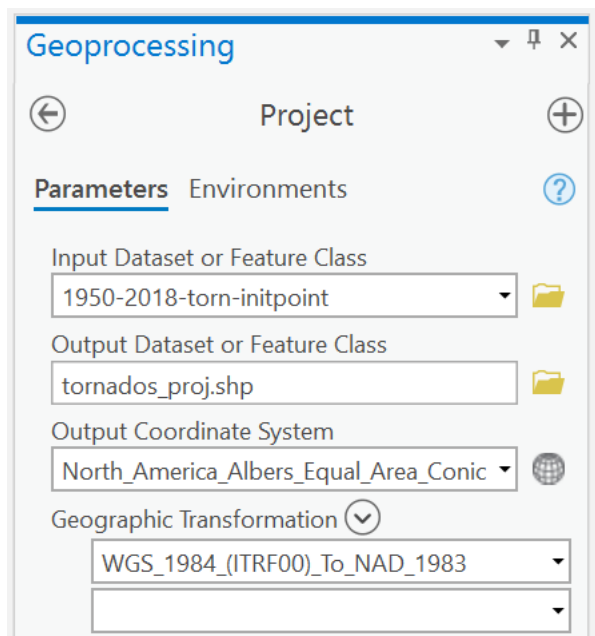
Analysing tornado patterns over time

Now that we have a time-enabled map layer, we are in a position to analyse it using space-time data mining tools. To do this, we first need to convert its coordinate reference system from unprojected latitude and longitude (geographic coordinates) to projected coordinates in metres.

Before doing so, as there are a few tornadoes in places like Puerto Rico, which lie outside the mainland USA, you may wish to first select just the points that lie in the mainland USA, for example by using the *rectangle* selection tool:



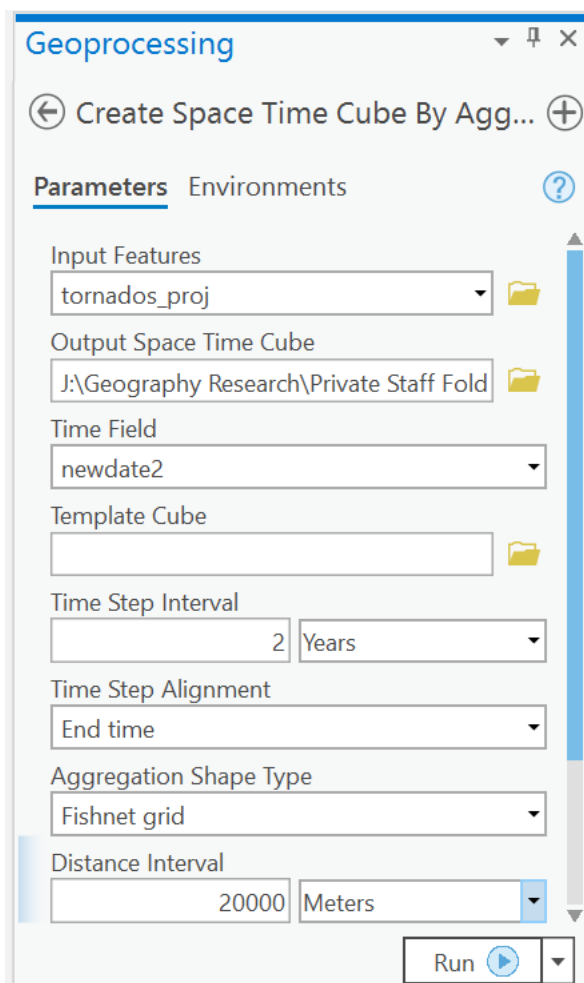
Only these selected points will be processed when you reproject the data. To reproject the map layer, head back to the geoprocessing panel once more and search for the 'project' tool, then run it:



Choose your tornadoes layer as the input, and give a suitable output file name. As the *output coordinate system*, choose one of the projections under *Projected / continental / North America*. I have used Albers equal area conic above. Note that this uses the North American Datum 1983, so a transformation is automatically selected for converting to this from the World Geodetic Survey 1984 typically used with unprojected (geographic) coordinates.

Our next step is to take our newly projected tornado points, then convert them to a space-time cube. In ArcGIS, a space-time cube is basically a NetCDF file with a time dimension. The NetCDF format is flexible enough to work with both raster and vector data. In this case, a fishnet (a vector format grid) is superimposed on the points, and then the points in each period in each grid cell are counted. Alongside counting features, it is possible also to summarise the characteristics of each point (e.g. as mean proportion of damaged buildings), but we will not do so here.

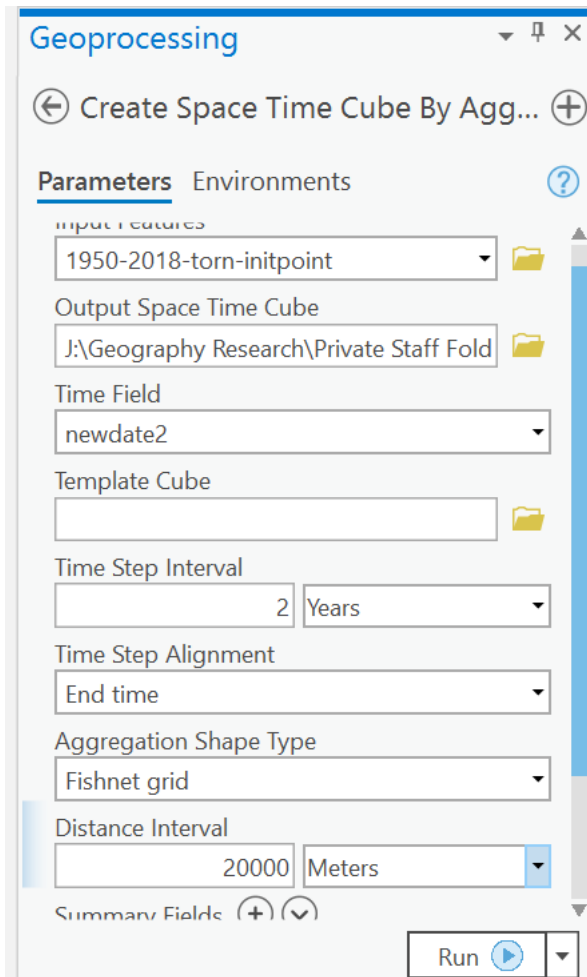
To create this space-time cube, head back to the geoprocessing panel again, then search for 'space-time cube' and run 'create space-time cube by aggregating points':



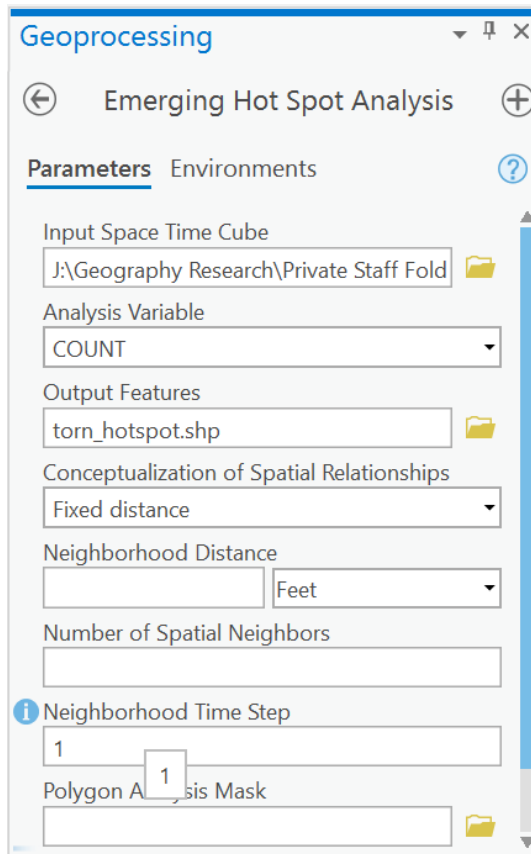
The *input features* should be your projected tornado points, with the *time field* being the dates you created, whilst the *output space time cube* will be a new NetCDF file. The other options are more subjective and require judgement:

- **Aggregation shape type:** To see how the tool works, stick to the default of *fishnet grid*. However, note that in spatial analysis, some have suggested that hexagons are preferable, given their more compact shapes. For more socio-economic or health applications, a really useful option here is *defined locations*. This enables polygons from a given map layer to be used instead of a grid or hexagons. For example, if you had points from postcodes of visitors to a nature reserve, it might be appropriate to use census boundaries (e.g. Lower-level Super Output Areas, which have similar resident population sizes) to aggregate up the points, rather than a grid.

- *Time step interval, distance interval:* This is the interval of time and size of grid cells within which tornados will be counted. Ideally, you should choose an interval where the ‘bins’ (time and space containers) are not largely full of zeros, but you should also make sure there is a minimum of ten time intervals in your series (decades would be too large in our case for example; if you decide to set a distance interval, my one in the screenshot above is probably too small). If you leave these settings blank, ArcGIS will choose values for you, but ideally, they should be chosen with the phenomenon being studied in mind. It can sometimes be worth experimenting with different values for these settings.

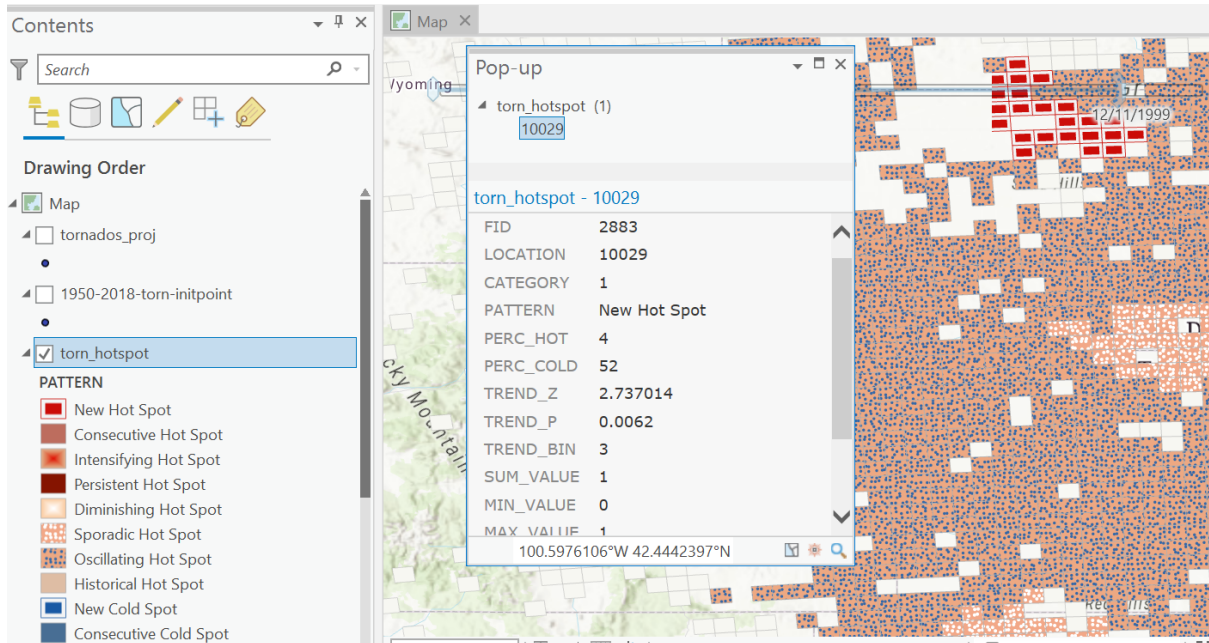


Nothing will appear on screen, but we do now have our space-time cube. There are multiple tools available for analysing such a cube, and we will explore just one here – emerging hotspot analysis. In the geoprocessing panel, search for ‘emerging hotspot’ and then run the *emerging hotspot analysis* tool:



- As inputs, you should be able to choose your *input space time cube* and as the *analysis variable*, you should be able to choose *count* of tornado events.
- Choose a suitable output file name. The tool will create a fishnet (or hexagon) vector file, with each feature classified in terms of its spatio-temporal characteristics.
- To see how the tool works, set the *neighbourhood time step* to **1** and leave the other settings to their default values (note: if you decide to use these tools ‘for real’, such as in a dissertation, there is a lot more thinking to be done about these settings. See <https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/modeling-spatial-relationships.htm> for more).

Your output should look something like this:



An overview of how the tool works, together with how to interpret these different output classes, is provided here:

<https://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/learnmoreemerging.htm>

Notice some of the attribute fields created too. Aside from the overall type of pattern, there are fields such as:

- Perc_hot, perc_cold: % of time periods classified as hot or cold
- Min_value, max_value, sum_value: The minimum and maximum tornado counts in any period for the grid square, together with their sum.
- Trend_Z, Trend_P: A statistical test for a significant trend (known as a Z statistic – see <https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/what-is-a-z-score-what-is-a-p-value.htm>), together with the probability of the trend being significant (<0.01 would suggest significance).

Other tools for spatio-temporal analysis:

Emerging hotspot analysis is just one of several ways of exploring spatio-temporal data. Other tools enable forecasting and identification of trends or outliers within a spatio-temporal data set. You can read more about these other tools here: <https://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/an-overview-of-the-space-time-pattern-mining-toolbox.htm>