Improved Processing Prototype Based On New Data Streams

This task is to build a sequence of scripts that improve the processing of fire emissions and can serve as a prototype method for later automation. The goal of these scripts are to make particular use of enhance data availability in 2014 from various sources such as newer ground reporting systems that include polygon data as well as many more additional small fires, and newer satellites that include polygon data and enhanced geolocation and detection of small fires.

To accommodate these enhanced data sets, SmartFire will need to be significantly modified. Currently SmartFire follows three steps: within-data-set processing to clump and derive required data types, association between data sets, and finally reconciliation based on specific data set preferences into a unified data stream. Particularly with polygon data, this data processing can be short circuited. Instead a different methodology is needed.

Instead this task will develop scripts that do the following:

* Tranche the incoming data sets and their individual data reports into different groupings. Foremost in these will be data reports with polygon data. A second set will be those with point fire locations that are expected to be of good geolocation quality. A third set will be those with point fire locations that are expected to be of bad quality, but where a reasonably small bounding area can be derived (e.g. within a National Forest). A fourth set will be those with bad geolocation but no reasonably small bounding box. These data types can be called Tranches 1-4 as listed above.
* Process the data sets with polygon data first (Tranche 1).
* Create a best-guess polygon of the final (up-to-date current) overall fire perimeter by combining the polygon datasets based on their intrinsic errors.
* Assign (associate) the collection of earlier (not final/current) polygons to this final polygon.
* These polygons are to be considered “Primary Polygons”
* Then look at point based fire reporting datasets (Tranche 2) such as satellite hotspots.
* Assign (associate) as many hotspots/point fire locations/reports as possible to the known fire perimeters (“Primary Polygons”), without clumping the hotspots first.
* Only then create, via clumping or other methods, a set of final/current “Derived Polygons” for fires that do not have a “Primary Polygon”
* Assign (associate) the individual hotspots / point reports to these “Derived Polygons”
* Look for fire reports without a known good fire location but with a reasonable bounding box (Tranche 3) and try to assign them to the “Primary” or “Derived” polygons.
* Those in Tranche 3 that cannot be so assigned are to be created into a list based on their bounding box. This list can be called the “Not Found” list.
* Finally look at the Tranche 4 data and where possible assign to a “Primary” or “Derived” polygon.
* Those not assignable are to be put into a list called the “Questionable” list.

Deliverables are to include:

* A short technical writeup of what was done
* A set of slides (in Powerpoint) to explain the method
* All scripts including dependencies

The scripts are to use open-source software where possible.