5A.14

EMPLOYING GEOGRAPHIC INFORMATION SYSTEMS FOR GRIDDED MODEL OUTPUT STATISTICS

Kari L. Sheets
Meteorological Development Laboratory
Office of Science and Technology
National Weather Service, NOAA
Silver Spring, MD

1. INTRODUCTION

The Meteorological Development Laboratory (MDL) of NOAA's National Weather Service (NWS) created the National Digital Guidance Database (NDGD) to provide additional tools for developing forecasts for the National Digital Forecast Database (NDFD, Glahn and Ruth 2003). Historically, forecast guidance is produced for the United States and its territories at approximately 1700 hourly observing sites and over 5000 cooperative observing sites by using the Model Output Statistics (MOS) technique (Glahn and Lowry 1972). In the MOS approach, observed predictand data are statistically related to predictors such as forecasts from dynamical models, surface observations, and geoclimatic information. Statistical model guidance is produced from this relationship. MOS guidance depends on a sufficiently long sample of high-quality observations to develop robust forecast equations for a variety of weather elements (Allen 2001).

MOS guidance is based on output from the National Centers for Environmental Prediction's (NCEP) numerical models (Dallavalle et al. 2004). Stations used in traditional MOS development are unevenly distributed, leaving developers searching for additional observational datasets as well as better predictor variables to capture terrain effects. Despite efforts to increase the resolution of the meteorological observation dataset, the network of quality-controlled observed data are shy of the desired NDFD resolution (Sheets 2007). To supplement the meteorological data and tailor the MOS forecast guidance to terrain and coastlines, we used a Geographic Information System (GIS) to generate additional geophysical variables at the agreed upon NDFD grid resolution (Glahn et al. 2008). As forecasters have begun to use gridded MOS, GIS has been used to compare forecast and guidance grids, to edit the aforementioned geophysical predictor variables, and to identify and remove misrepresentative stations to better represent the guidance area. GIS was utilized to manipulate final product coverage areas along the United States borders for the CONUS as well as to generate development areas for Alaska. The Environmental Systems Research Institute (ESRI) generated both ArcInfo and ArcView GIS packages used for the tasks discussed in this paper*. ArcMap is the primary interface for viewing and editing GIS data in both packages.

In this paper, we discuss the use of GIS in the development and evaluation of gridded MOS. We discuss some of the details of the GIS processes used to generate and assess geophysical components of gridded MOS, to create a station dictionary including land/water designations for the observing stations, and to analyze or troubleshoot problem areas in gridded MOS weather elements. Plans for the use of GIS to aid in future gridded MOS work will also be presented.

2. DEVELOPMENT TECHNIQUES

2.1 Station Data

MDL maintains current and historic MOS system station metadata including the changes occurring at reporting sites in a station dictionary (Allen 2001). This station dictionary is formatted according to guidelines in TDL Office Note 00-1 (Glahn and Dallavalle 2001). A GIS compatible station database was created by importing the ASCII station dictionary to database format (.dbf) in Microsoft Access. Current and past locations, type of observing site, elevation, and call letters are the key components of the station history documented in this database (Sheets et al. 2005).

**Corresponding author address:* Kari L. Sheets, 1325 East-West Highway, Station 11334, Silver Spring, MD 20910-3283; e-mail:Kari.Sheets@noaa.gov

**Disclaimer:* The use of this software package is in no way an endorsement of this company by the National Weather Service.

The addition of sites from unfamiliar sources presented the need for additional quality control measures, such as an elevation check. Stations were plotted for generic elevation and location checks. Gross station location checks were made by color coding stations according to the state in which they are located. Elevation values were extracted from the United States Geologic Survey (USGS) GTOPO30, Global 30 Arc Second Digital Elevation Model (DEM) to check for accuracy. As long as station elevation values matched within five hundred feet, they were viewed as close enough. Sites with discrepancies greater than 500 feet were crosschecked with other agencies maintaining metadata on the same sites.

Traditional MOS guidance has been available at specific locations, but changing forecaster requirements have led to the need to make the MOS guidance available on a grid. Forecasters need the highest quality guidance to populate the official forecast grids, ideally available at an equivalent resolution. Gridded MOS is produced because the existing MOS system is only available at specific sites. Gridded MOS is created using an extension of the Bergthorssen, Cressman, and Doos, or “BCD,” analysis technique used in the Local AFOS MOS Program (LAMP) implementation. This modified technique is now called “BCDG” to distinguish it from other “Cressman Analysis” techniques (Glahn and Dallavalle 2006). The major differences in BCD and BCDG analysis techniques are the effects of land, water, and elevation on the analysis. The triggers for the initialization of the analysis of station data to gridpoints are found in a station flag designation within the station dictionary. A station’s location over ocean water, inland water (lakes), or land is indicated by a value in this flag. GIS extraction techniques are used to set the land or water values for this flag (Figure 1).

2.2 Supplemental Data

In addition to the station based observation data, gridded MOS depends on geophysical data at the gridpoints in order to create realistic forecast guidance in areas lacking meteorological observations. The two geophysical constant fields used in gridded MOS are elevation and land/water coverage.

Originally, GIS was used to perform a nearest neighbor resample of the most recent GTOPO30, to the NDFD grid specifications (Sheets et al. 2005). During the development phase of the gridded MOS analysis technique, developers realized this elevation data differed from the elevation data available to forecasters in the Graphical Forecast Editor (GFE) within the Advanced Weather Interactive Processing System (AWIPS) environment at NWS weather forecast offices (Jones et al. 2003). GFE elevation data are also derived from a USGS GTOPO30 DEM, but an earlier version. Therefore, the GFE elevation was ingested into the GIS to quality control the grid specification and exported as an ASCII grid.

The technique for developing the land water mask also evolved since the prototype gridded MOS was developed for the western third of the CONUS. For this prototype, a land cover dataset was reclassified to indicate water or land and resampled to the NDFD grid specifications (Sheets et al. 2005). This dataset included too many small water bodies when compared to the availability of meteorological observations over water. Therefore, the development team decided to include only those lakes with observing sites and the oceans as water. The best method available to achieve this dataset was to use shapefiles of the desired lakes and shapefiles of landmasses within the NDFD domain. Shapefiles for the ocean were not readily available, so land and lake files were removed from a shapefile just larger than the NDFD extent in order to create an ocean file. Each shapefile was given a value to indicate its designation and merged with the other shapefiles to form a land water shapefile. This shapefile was converted to a raster file, a GIS grid file, matching the NDFD grid specifications (Figure 2). The final step was to export an ASCII grid of the land water mask.

The geophysical ASCII grids were moved to a UNIX development environment where they were processed by FORTRAN codes developed to pack ASCII constant data into a MOS binary data format. The MOS constant file includes grid specifications as well as identifying data information needed for use in additional MOS FORTRAN codes.

2.3 Coverage Areas

Creating weather forecast guidance on grids based on station guidance presents several quality control issues with respect to the area for which the guidance is valid. Areas needing quality control include the reference system for the data as well as the extent of the data itself.

A geodetic datum is dependent on the assumed shape of the earth, ellipsoidal or spherical, and associated coordinate system, as well as a set of points and lines resulting from surveying (Bolstad 2002). If data are not properly transformed to the geodetic reference system on which the product is based, erroneous values will result. A primary data source for MOS guidance is NCEP model output (Dallavalle et al. 2004), which references a spherical datum unique to meteorological applications (Sheets 2007). The NDFD grid specifications are based on a grid defined in NCEP Office Note 388 (National Weather Service 2002) which also describes the associated geodetic datum, referred to as the NCEP Sphere (Sheets et al. 2005) as a spheroid with a radius of 6,371,200 meters. Improper transformations or projections result in a skewing of the data area, often times resulting in data being misplaced in coastal areas and/or rectangles appearing as rounded polygons (Figure 3). These erroneous representations of the data create inconsistent valid areas for gridded products.

Once the coordinate systems are set, the next question of coverage relates to the analysis technique. The first step was to finalize the output area, or the area for which gridded MOS would be valid. This area is slightly larger than NDFD because it is an input in the forecasting process and forecasters need the ability to advect systems into their area. After consulting with NWS field representatives, the gridded MOS development team decided the CONUS, Coastal Marine Zones (including the Great Lakes), and a 50-km buffer beyond these areas was a reasonable valid area. This area was created by combining AWIPS shapefiles for the aforementioned areas into a single shapefile.

A search radius is used when computing gridpoint values, so data are needed for an area larger than the data area disseminated as official guidance. Therefore, we created a computational area by extending the boundaries of the output area 200 km. Analysis input data are trimmed to this buffered area to eliminate processing unnecessary points. This was accomplished using buffering GIS geoprocessing tools. The polygon for the valid area was given a value of “1” and the polygon area outside of the valid area was given a value of “0.” The coverage area shapefiles were converted to an ASCII grid in the same method as previously described for the land/water mask.

3. Evaluation Techniques

3.1 Observation Data

The analysis of station data to surrounding gridpoints may reveal problems with the quality of a station observation. Areas with large differences from surrounding areas appear as bull’s eyes on a guidance grid. GIS tools are used to plot the grids and stations on one surface. In areas of densely populated observation sites, this technique makes identifying suspect stations easier (Figure 4). Stations identified as being suspect of poor data quality can be removed from the analysis for the problematic element.

3.2 Constant Data

Forecasters were eager to use high quality guidance grids to initialize their forecast grids. The idea is the availability of better guidance will lessen a forecaster’s need to hand edit forecast grids, especially for the extended forecast periods. Inconsistencies between the MOS development constant data and similar tools available in GFE became evident. One of the biggest areas was in the land water designations at gridpoints. The original philosophy for generating the land water mask was to use a coarse coastline because of the 5-km grid resolution. The forecasters, however, use a high resolution coastline in the forecast process, so gridded MOS values did not agree with forecaster expectations based on GFE coastlines available in AWIPS. AWIPS provides shapefiles on the web, so a new land/water mask was generated using the high resolution coastal and lake shapefiles. Even the use of the same dataset did not address all of the forecaster concerns. Therefore, strategic hand edits were made for areas where the gridpoint fell over land/water, but the area surrounding it was more representative of the opposite designation. GIS tools were used to modify the coastline to force these designations (Figure 5). Grids were differenced to show changes (Figure 6) and shared with concerned forecasters before implementing a new mask.

3.3 Guidance Area

As forecasters began using guidance grids in the northeast, they realized the 50-km buffer into Canada, the standard buffer set in the development stage of gridded MOS, did not provide enough data to effectively advect grids for their hourly forecasts. MOS developers used GIS tools to generate options for the forecasters to consider and ultimately expand the gridded MOS guidance area available in AWIPS (Figure 7).

3.4 Product Comparisons

Thus far we have discussed evaluating modifications to gridded MOS analysis input data based on the effects of the changes on the constant data; however, GIS is also used to compare the output gridded MOS guidance for specific elements. GIS tools, namely ArcMap's raster calculator, made it easy to demonstrate the change in the guidance grids. Guidance grids excluding previously discussed suspect stations were subtracted from the existing grids including those stations. The resulting difference grids showed the minor adjustments of the gridpoints less affected by the problematic stations as well as the elimination of the bull's eye areas used to identify station issues (Figure 8). These images displaying these alterations to analysis input fields have proven to be an invaluable tool for communication when addressing customer concerns.

4. AVAILABLE PRODUCTS

Information concerning availability of GMOS, based on both 0000 and 1200 UTC cycles of NCEP's GFS model, is available at http://www.nws.noaa.gov/mdl/synop/gmos.html. Gridded MOS guidance is available every 3 to 12 hours, depending on the element, out to 192 hours. Available gridded guidance elements include: maximum and minimum temperature, temperature, dew point temperature, relative humidity, wind direction, wind speed, wind gusts, sky cover, precipitation type, probability of precipitation, probability of thunderstorms, quantitative precipitation, and snowfall amount.

5. FUTURE PROJECTS

As developers seek to improve and expand the gridded MOS system, the need for additional geophysical datasets will grow. Work is in progress to identify and better relate stations and gridpoints with similar characteristics. Development of gridded MOS for Alaska is also advancing with the goal of issuing temperature guidance grids in early 2008. Gridded wind guidance is proving problematic along the coasts, so developers are investigating methods to improve gridded MOS wind products. Developers are proceeding with the creation of gridded MOS guidance at a 2.5-km resolution with a goal of having prototype grids available by summer 2008.

6. CONCLUSION

GIS has increased the speed and ability of MDL to process high resolution data. These data have been integral in the gridded MOS development and evaluation process. GIS tools are aiding MOS developers in thinking outside the box for troubleshooting existing gridded MOS products as well as incorporating a diverse set of data available to the analysis process. These efforts will continue to facilitate collaboration with NWS field offices ultimately improving gridded MOS forecast guidance.

7. ACKNOWLEDGMENTS

The statistical modeling branch works as a team to generate MOS products. Special thanks are given to Rebecca Cosgrove for assistance in learning and running the gridded MOS analysis codes. NOAA's NWS assembles and propagates spatial geographic information issued by the NWS for their user community in a map database found at http://www.weather.gov/geodata. The coarse shapefiles used are part of the media kit included with ESRI's off-the-shelf software, ArcGIS 9.2 Desktop.

8. REFERENCES

Allen, R. L., 2001: Observational data and MOS: The challenges in creating high-quality guidance. Preprints, *18th Conference on Weather Analysis and Forecasting*, Ft. Lauderdale, FL, Amer. Meteor. Soc., 322-326.

Bolstad, P., 2002: *GIS Fundamentals: A first test on geographic information systems*. Eider Press, 412 pp.

Dallavalle, J. P., M. C. Erickson, and J. C. Maloney, 2004: Model Output Statistics (MOS) guidance for short-range projections. Preprints, *20th Conference on Weather Analysis and Forecasting,* Seattle, WA, Amer. Meteor. Soc., CD-ROM, P6.1.

_____, J. P., and B. Glahn, 2005: Toward a gridded MOS system. *Preprints, 21st Conference on Weather Analysis and Forecasting*, Washington, D.C., Amer. Meteor. Soc., 13B.2.

Glahn, H. R., and D.A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203-1211.

____, and J. P. Dallavalle, 2000: MOS-2000. TDL Office Note 00-1, National Weather Service, NOAA, U.S. Department of Commerce, 131 pp.

____, and D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor., Soc.*, **84**, 195-201.

____, and J. P. Dallavalle, 2006: Gridded MOS – Techniques, Status, and Plans. Preprints, *180th Conference on Probability and Statistics in the Atmospheric Sciences,* Atlanta, GA, Amer. Meteor. Soc., CD-ROM, P2.1

_____, K. K. Gilbert, R. L. Cosgrove, D. P. Ruth, and K. L. Sheets, 2008: Gridded MOS guidance in the national digital guidance database. *Preprints, 19th Conference on Probability and Statistics*, New Orleans, Amer. Meteor. Soc., 11.3

Jones, D. R., C. Bullock, W. Carrigg, M. desJardins, E. Mandel, D. Rhine, J. Roe, and M. T. Young, 2003: AWIPS Build 5 in review. *Preprints, 19th Conference on Interactive Information Processing Systems*. Long Beach, Amer. Meteor. Soc., 4.1

National Weather Service, 2002: ON 388 GRIB (Edition 1) The WMO Format for the Storage of Weather Product Information and the Exchange of Weather Product Messages in Gridded Binary Form as Used by NCEP Central Operations. [Available online at http://www.nco.ncep.noaa.gov/pmb/docs/on388/introduction.html]

Sheets, K. L., 2007: Supporting Gridded Model Output Statistics forecast guidance system. Conference Proceedings, *27th Annual ESRI International User Conference,* San Diego, CA, ESRI, CD-ROM, p1737.

_____, K. L., R. A. Trimarco, and K. K. Hughes, 2005: Using geographic information systems to develop gridded model output statistics. *Preprints, 21st Conference on Weather Analysis and Forecasting*, Washington, D.C., Amer. Meteor. Soc., 13B.1.

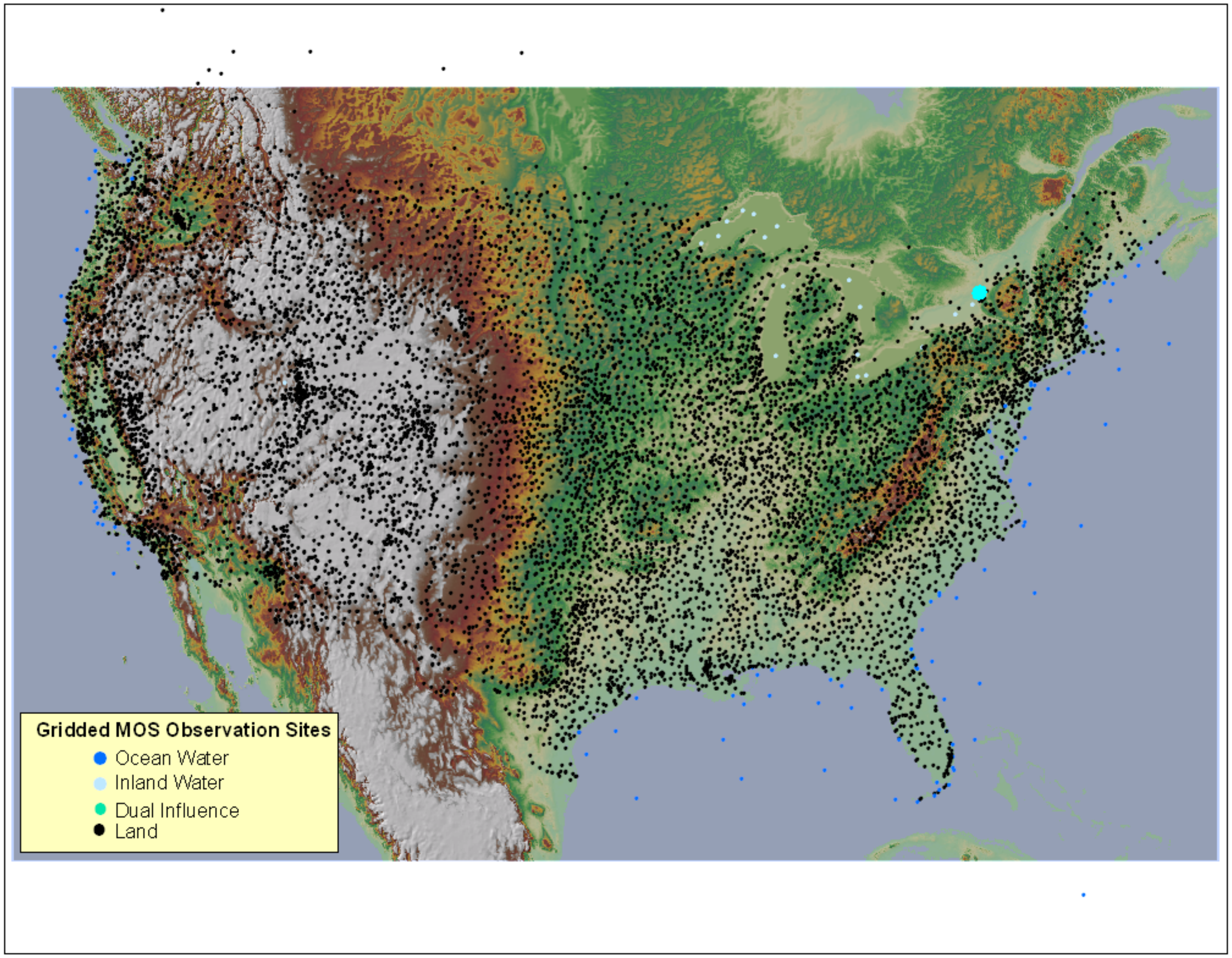

Figure 1 – Stations are designated as having observations capable of influencing land, water, or both depending on the characteristic of the surface underneath the site.

Figure 2 – Land water mask created from AWIPS high resolution shapefiles.

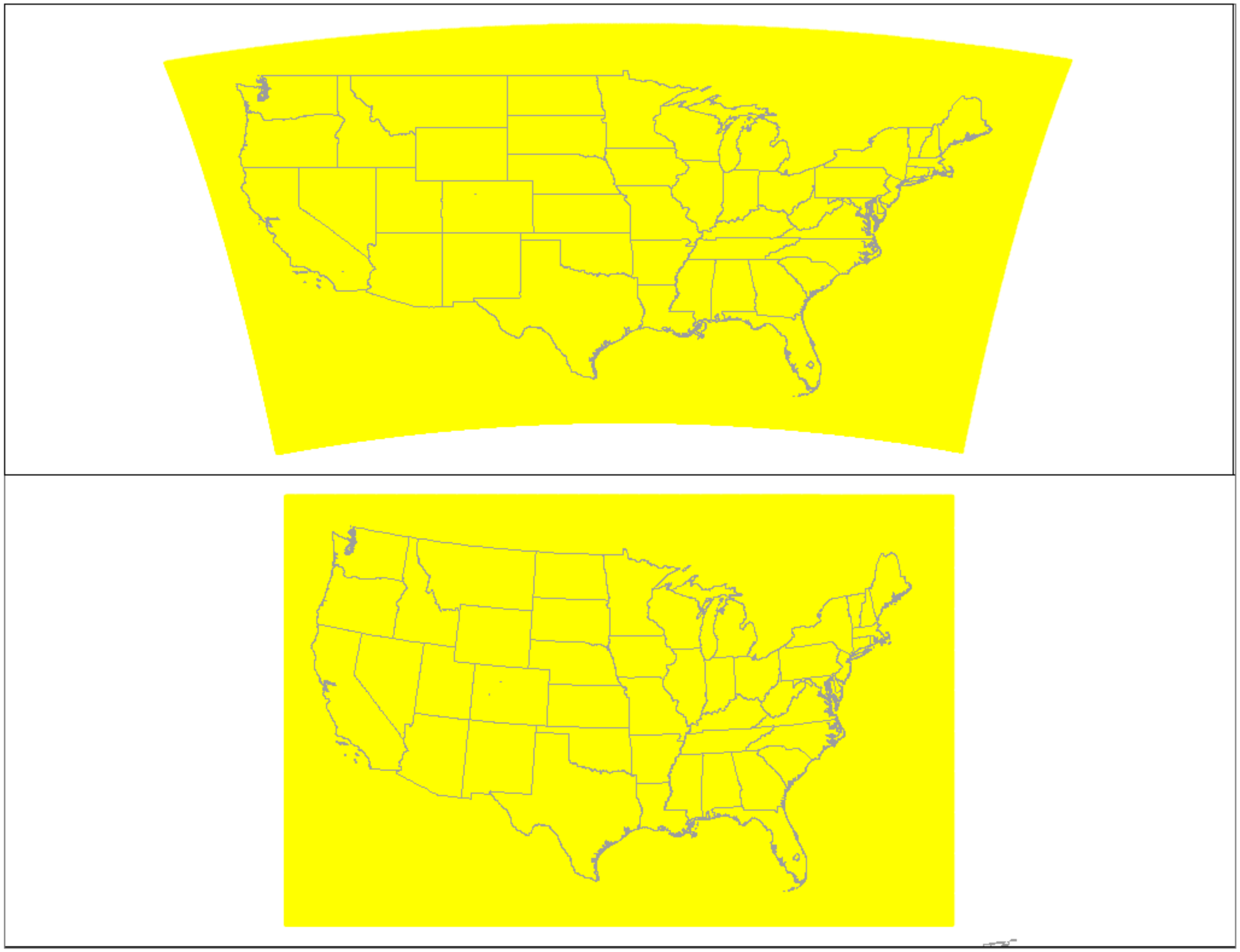

Figure 3 – These images depict the importance of correctly projecting data. The yellow areas are the NDFD/NDGD CONUS grid. The NCEP sphere is the datum in both images. The top image neglects choosing a projection, while the lower image is projected according the GRIB table B specifications for Lambert Conformal valid over the CONUS.

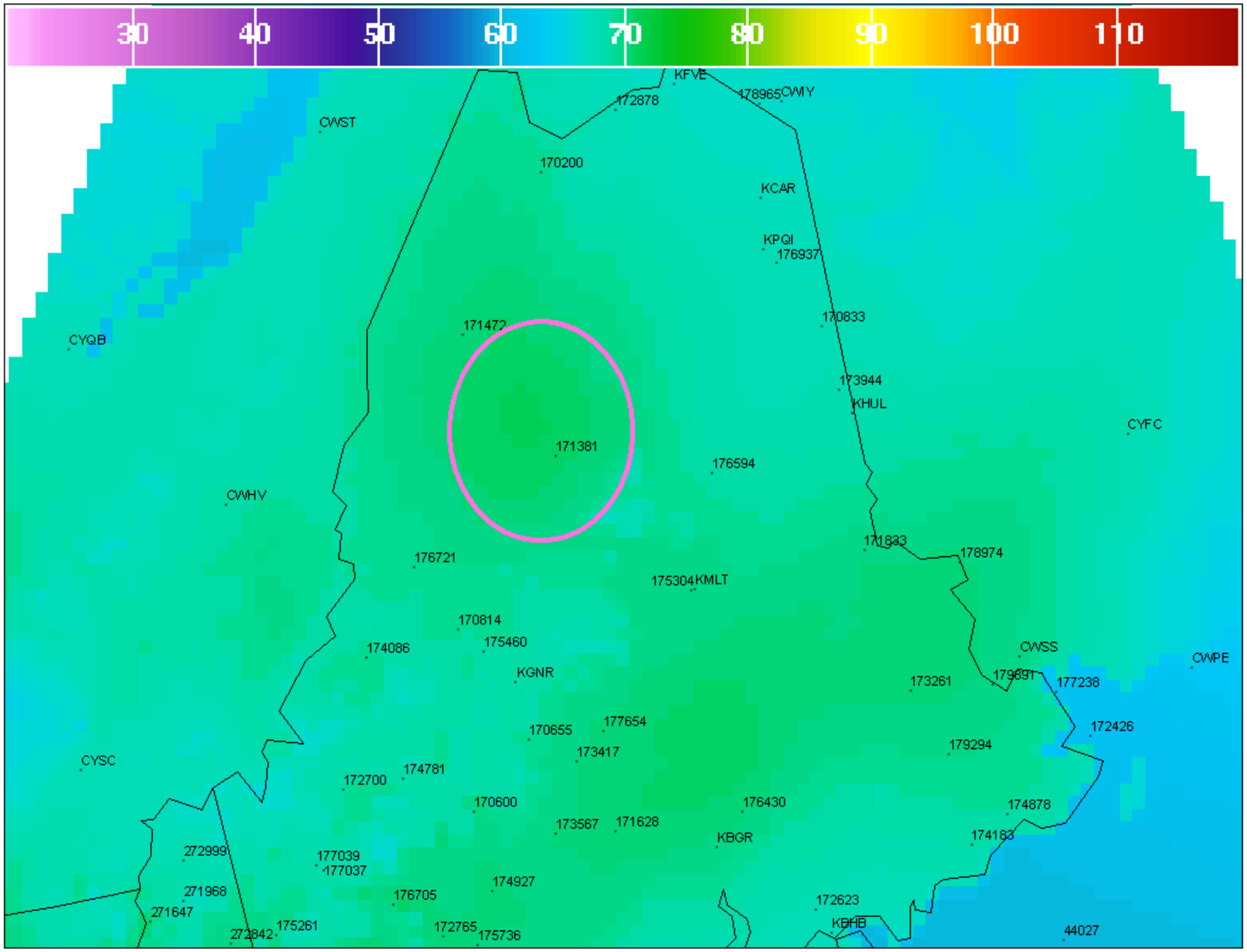

Figure 4 – The data from the cooperative observing site near the center of the circled area caused the entire circled area to be a few degrees Fahrenheit warmer than the surrounding areas in this GMOS 30-hr Maximum Temperature guidance product. This station has since been removed.

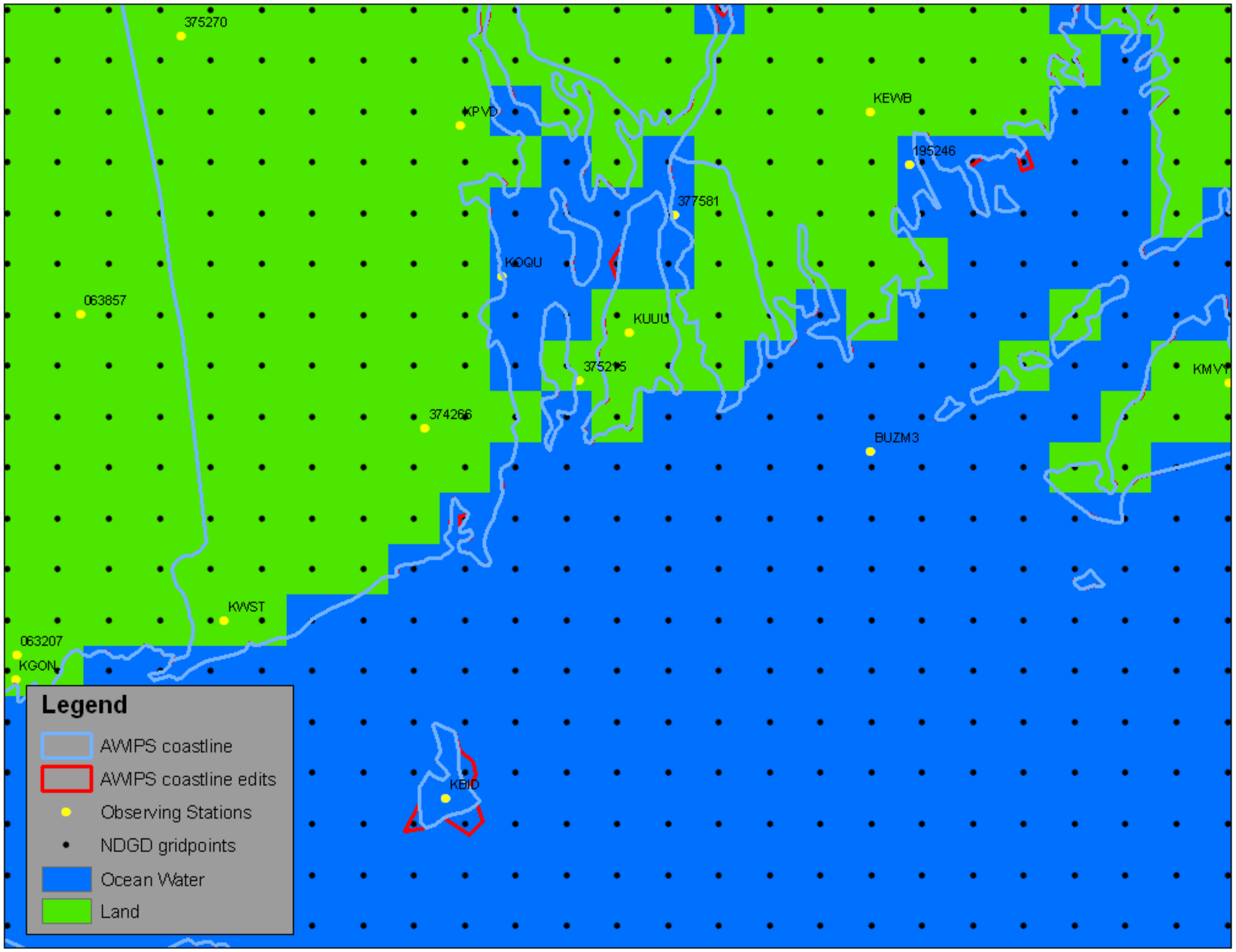

Figure 5 – The AWIPS coastline, light blue lines, for Block Island fit neatly in between NDFD/NDGD gridpoints, so the island was missing from the resulting land water mask. In order to include Block Island as land in the mask, the coastline was edited so the island covered gridpoints nearest the KBID observing site.

Figure 6 - This image depicts the grid cell changes resulting from a change in gridpoint designation. This is what we expect to see because the cell actually represents the gridpoint at its center.

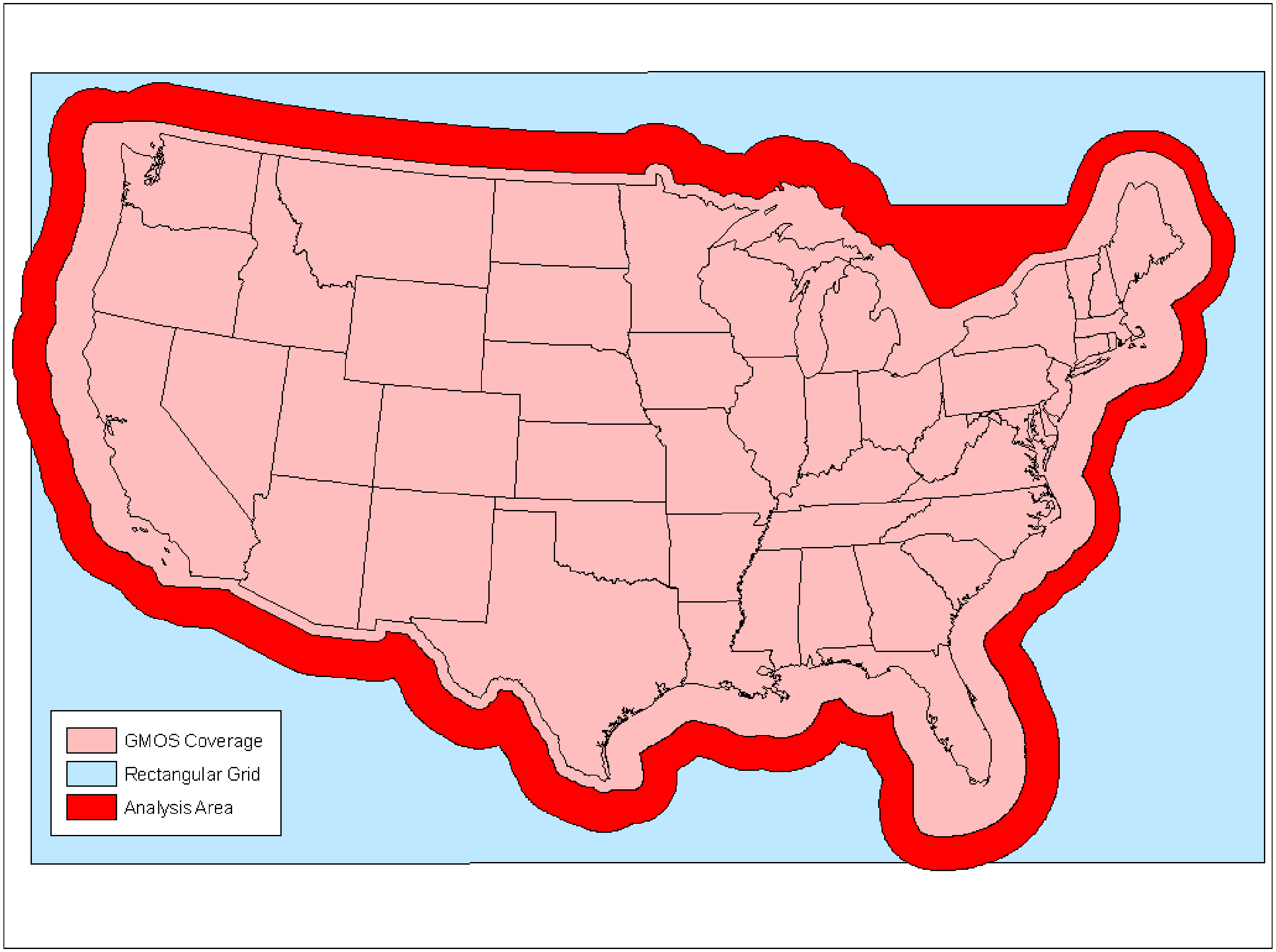

Figure 7 – Values are computed outside of the official guidance area in order to give the best information within the coverage area, so developers work with data in the red area while forecasters only have access to the areas in pink.

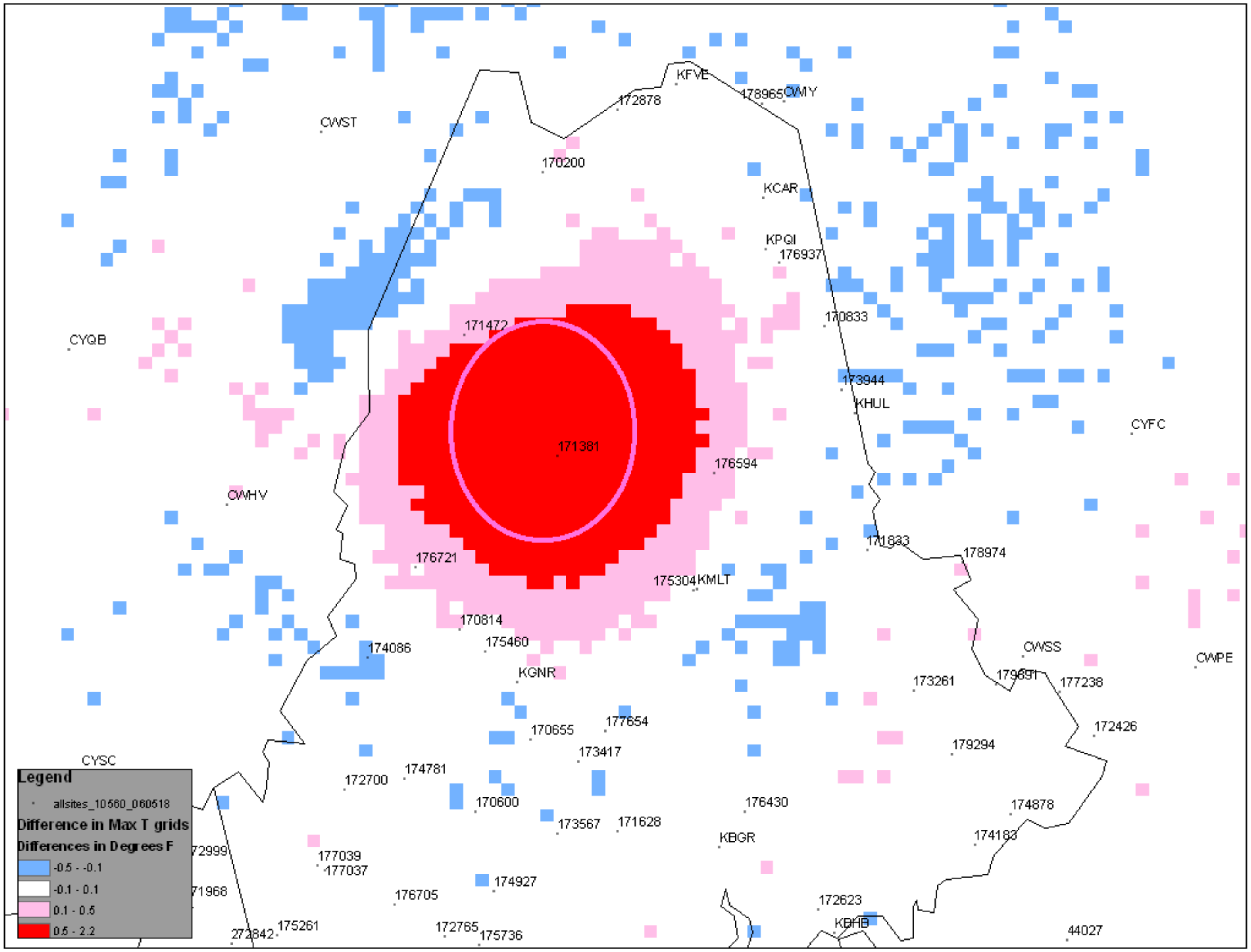

Figure 8 – Stations consistently causing suspect data in GMOS guidance were removed from the offending element's data pool, resulting in the maximum temperature differences shown above. The circled area corresponds to the circled area in Figure 4.