Reprinted from:
AMS 21st Conference on Weather Analysis and Forecasting, August 1-5, 2005, Washington, D.C.

13B.1

USING GEOGRAPHIC INFORMATION SYSTEMS TO DEVELOP GRIDDED MODEL OUTPUT STATISTICS

Kari L. Sheets, Rachel A. Trimarco, and Kathryn K. Hughes
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) is developing a National Digital Guidance Database (NDGD) at a resolution of 5 km to complement the existing National Digital Forecast Database (NDFD, Glahn and Ruth 2003). To help accomplish this goal, MDL is creating a gridded forecast guidance system. Current 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. 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). The initial gridded MOS products were derived from NCEP's Global Forecast System (GFS) model and focused on the western third of the contiguous United States (CONUS). Traditional observing stations used to develop MOS for this region are sparsely located, leaving developers searching for additional observational datasets as well as better predictor variables to capture the meteorological effects of elevation, slope, aspect, land cover, and water. Efforts were made to gather, quality-control, and archive data from additional meteorological observing systems, but these data did not bring the observed data resolution to the desired NDGD resolution of 5 km. To supplement the meteorological data and tailor the MOS forecast guidance to terrain, we used a Geographic Information System (GIS) to generate additional geophysical variables at the proper NDGD grid resolution. For this purpose, grids of elevation, slope, aspect, land cover, and a land/water mask were created. Additionally, GIS was employed to generate the map specifications for the western third of the United States as well as a station dictionary including land/water designations for the observing stations. In this paper, we discuss the GIS efforts used so far in the development of the gridded MOS guidance, as well as some of the details of the GIS processes for developing the geophysical variables, including information about the parent data sets. Plans for the use of GIS to generate additional climatic and geophysical data sets for future gridded MOS development are also presented.

2. DATA COLLECTION

The Environmental Systems Research Institute (ESRI) produces the ArcInfo GIS package used for the tasks discussed in this paper*. The creation of geophysical and climatic data for NWS products was dependent on the collection of the following datasets: 1) United States base map, 2) current MOS forecast guidance site locations, 3) major water bodies of the United States, 4) 1-km elevation data, 5) 1-km land cover data, 6) monthly temperature and precipitation climatologies, 7) NWS grid specifications, and 8) 5-km gridded MOS specifications for the western third. Along with the software, ESRI packages data commonly used by their clients. The United States base map and major water bodies are two such files. The elevation terrain data were also part of the data media package provided with the ESRI ArcInfo package. The terrain data set is the tiled version of the United States Geologic Survey's (USGS) GTOPO30, Global 30 Arc Second

**Corresponding author address:* Kari L. Sheets, 1325 East-West Highway, Station 11330, 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.

Elevation Data. The land cover data were generated by the Global Land Cover Facility (GLCF) at the University of Maryland by using data collected between 1997 and 2004 (Hansen et al. 2000). Red, infrared, and thermal bands from satellite images, as well as the Normalized Difference Vegetation Index (NDVI), were used to provide the greatest discrimination between cover types. Fourteen classification were included in the resulting data ranging from water to forest to grassland to urban (Hansen et al. 2000).

The Parameter-elevation Regressions on Independent Slopes Model (PRISM) (Daly et al. 1997) temperature and precipitation climatology data were acquired from the Spatial Climate Analysis Service at Oregon State University. PRISM data's spatial coverage was limited to land areas of the CONUS and was representative of the years 1971-2000. Supplemental data for NDGD areas not covered by PRISM were obtained from the National Center for Atmospheric Research's (NCAR) International Comprehensive Ocean–Atmosphere Data Set (ICOADS). Precipitation data, valid over the Great Lakes, were obtained from NOAA's Great Lakes Environmental Research Laboratory (GLERL). Gridded MOS specifications were determined by MDL based on the NDFD and MOS archive specifications.

3. DATA CONFIGURATION

a. Grid-Based Data

Projections, geodetic datums, grid resolutions, and station dictionary formats were of primary focus before data analysis could begin. A geodetic datum is dependent on the assumed shape, ellipsoid or spheroid, and associated coordinate system of the Earth as well as a set of points and lines resulting from surveying (Bolstad 2002). Each of the data types collected for gridded MOS were configured differently and needed to be converted to a common coordinate system. MDL grids and coordinate systems were non-standard to the geographic community, so they also needed to be set within ArcGIS.

Gridded MOS developments use data on three different grids. The grids are derivations of grids outlined in NCEP Office Note 388 (National Weather Service 2002), so the geodetic datum is a spheroid with a radius of 6,372,100 meters. This geodetic datum will be hereafter referred to as the NCEP Sphere. MDL's archive grid is a 95-km grid based on NCEP's GFS grid specifications, so it is projected as a north polar stereographic map (National Weather Service 2002). The gridded MOS development grid, henceforth GMOS grid, is 1/16th the size of the MDL GFS archive grid yielding a grid resolution of 5953.125 meters. The NDGD/NDFD grid has a resolution of 5079.406 meters in Lambert Conformal projection as described by NCEP Grid 226 (National Weather Service 2002).

The USGS GTOPO30 and the 1-km GLCF land cover data were in a standard geographic projection. These land characteristic data sets as well as those derived from them needed to be converted to the MDL GFS archive grid, the GMOS grid, and the NDGD grid so they could be used as development predictors or provide influence for the gridded MOS analysis code.

Temperature and precipitation PRISM climatic data were obtained as ArcInfo ASCII grids with a 2.5-min (4-km) resolution. ICOADS temperature data and GLERL precipitation data were converted to PRISM's standard units and resolution. A separate GIS map "project" was created for each month (Trimarco et al. 2005). The compiled data sets were converted to the GMOS grid to be easily used as predictors.

b. Station-Based Data

Station metadata and the changes that occur at these reporting sites are maintained by MDL in a station dictionary (Allen 2001). The format for the station dictionary was established in TDL Office Note 00-1 (Glahn and Dallavalle 2000). The dictionary provides a history by documenting information about the past and present location, station type, and call letters for all stations in the MDL observational archive. Microsoft Access was used to convert the ASCII station dictionary to database format (.dbf) in order to add the data to GIS map layouts.

4. DATA ANALYSIS

a. Terrain Data

Terrain elevation analysis was the first GIS application for the gridded MOS effort. The raw elevation data needed to be converted to the NWS specified grids for use as both predictor data and for analysis routines. Since these functions require two separate coordinate systems, a map layout was created for each coordinate system.

ESRI's ArcInfo software is enabled with on-the-fly-projection, so the empty map layout's coordinate system was set to match the GMOS grid specifications for one map layout and NDFD specifications for the other map layout. Once the projected data were visible on the map, analysis could begin. For both layouts, ArcInfo's Spatial Analyst Map Calculator was used to resample the elevation data to the correct resolution. A nearest neighbor technique was used for the resampling. Since the original data were of finer resolution than the output data, the software chose the most common value occupying the same space as the resultant data cell. Ocean areas of the original USGS terrain dataset were set to missing, 9999. For use in MOS software, data are packed in an internal binary format, henceforth termed "tdlpack," so these no data values were converted to "0" by applying ESRI's Map Calculator (Fig. 1). In addition to the 5-km elevation files, the terrain elevation was analyzed with GIS to produce slope and aspect datasets (Figs. 2 and 3). Both of these computations are functions available in the Surface Analysis tool in ESRI's Spatial Analyst. Slope is determined by the greatest change in elevation between a cell and each of its eight neighbors. Aspect is the compass direction that a hill faces (McCoy and Johnston 2001). The elevation, slope, and, aspect on both the NDGD, GMOS, and MDL GFS archive grids were exported from ArcInfo to ASCII grids. A Fortran code was written to read these ASCII files and pack them into tdlpack files, which include their grid specifications.

b. Land Cover Data

Further surface analysis was completed to determine land characteristics for the NDGD and GMOS grids. Due to the absence of deciduous needleleaf forest cover over the whole CONUS, the land cover data were reclassified to create a continuous data field. Next, the data were resampled by nearest neighbor method to the 5-km grid resolutions. Finally, the null, or "no data," values were set equal to "0" (Fig. 4). The final surface dataset was the land water mask, which originated from the land cover dataset. The original 1-km land cover data were reclassified so that all non-water values were set to "2" and all water values were set to "1," creating a 1-km land water mask. The 1-km mask was then resampled to the 5-km grids and the null values converted to "0" (Fig. 5). As with the elevation datasets, the land characteristic datasets were exported from the GIS as ASCII files and converted to tdlpack with a Fortran routine.

c. Climate Data

To be used as predictor data for gridded MOS, the climate data needed to be converted to the GMOS grid specifications. GLERL and International Comprehensive Ocean – Atmosphere Data Set (ICOADS) were converted from point shapefile data to GIS vectors, then to raster data, and finally to GIS grids. All climate datasets were converted to the 5.953-km resolution by using the ArcInfo Map Calculator's nearest neighbor resample technique. The overall procedures for the temperature and precipitation data were the same, but due to differing water data sources, the intermediate steps were different.

Over-water temperature data were available only as average temperatures for each month, but we chose to use that average with both the PRISM maximum and minimum temperature datasets because of small diurnal variations in temperature over water. In order to merge these two datasets, the ICOADS data had to be converted to a vector polygon shapefile, so the ArcInfo erase tool could be used to remove PRISM's coverage area. ICOADS data were then converted back to a raster dataset (Trimarco et al. 2005). Over-water precipitation averages were only available for the Great Lakes, so the precipitation climate grids were created as a composite of what amounts to three datasets: GLERL data for the Great Lakes, extrapolated PRISM data for the oceans, and PRISM data for the CONUS. Both the GLERL data and the PRISM extrapolated data grids had to be converted to vector polygon shapefiles in order to remove the coverage of the other two datasets by using the ArcInfo erase tool. The resulting data were converted back to raster by using Spatial Analyst's shape to raster tool.

The remaining manipulations of monthly data were common to both types of climate datasets and were accomplished by ArcInfo's map calculator. In order to avoid quality data from being corrupted, the null and "no data" values for each grid were changed to "0." Finally, the water and land grids were added together and the composite grids were resampled to the full extent of the MDL GFS archive grid. The data were then output from ArcGIS as ASCII files. Maximum and minimum monthly temperature ASCII grids were ingested into a software package specifically designed to compute and evaluate cubic spline interpolation polynomials for a given set of points. The result of these computations was a maximum and minimum temperature normal valid every fifth day from day 5 (January 5th) until day 365 (December 31) (Trimarco et al. 2005).

d. Station Data

Station data attributes were also enhanced with the GIS. Preliminary gridded MOS experiments indicated the need for characterizing the individual station data according to its proximity to water. An attribute record of the station dictionary was modified to include station characteristic flags. The first field of this attribute indicates the origin of the station's data, that is, a METAR station, a Mesowest station, Cooperative Observing station (co-op), or River Forecast Center (RFC) site. The second field indicates the station's proximity to water, that is, in water, on land, or inland but influenced by water. GIS tools were used to select all observing stations located completely in water. From the remaining stations, the GIS buffer tool was used to select stations within 10 km of a major lake or sea in order to determine which inland stations should be flagged as land influenced by water. All remaining stations were flagged as land (Fig. 6).

5. FUTURE PROJECTS

As developers seek to improve and expand the gridded MOS system, the need for additional geophysical datasets will grow. Dew point and sky cover climate data sets, sky cover climatologies, radar data, and satellite images have all been discussed as being of interest to developers. GIS will be critical in converting this data to a format that can be used in the MOS system. Work has already begun to provide additional station characteristic flags such as proximity to major highways to assist in the quality control of wind data.

6. CONCLUSION

Prior to use of GIS, MDL's ability to ingest, manipulate, and analyze high-resolution data was very limited. Datasets created by using GIS have played a critical role in the development of the gridded MOS system. GIS tools will allow MOS developers to explore new analysis and predictor data, which will hopefully translate to better gridded MOS forecast guidance.

7. ACKNOWLEDGMENTS

GTOPO terrain elevation data are provided as a free public service by USGS at http://edcdaac.usgs.gov/gtopo30/gtopo30.asp. Land cover data were provided by the Global Land Cover Facility at the University of Maryland. Data are currently provided free to the public at http://glcf.umiacs.umd.edu/data/landcover/. The Spatial Climate Analysis Service and the Oregon Climate Service at Oregon State University provided the PRISM data as a free public service at http://www.ocs.orst.edu/prism. ICOADs data are provided as a free public service by NOAA-CIRES Climate Diagnostic Center at http://www.cdc.noaa.gov.coads/. The Great Lakes Environmental Research Lab provides the overlake precipitation data as free public data at ftp://ftp.glerl.noaa.gov/publications/tech_reports/glerl-083/ArchivedFiles/. The major lakes file is part of the media kit included with ESRI's off-the-shelf software, ArcGIS 9.0 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.

Daly, C., G. Taylor, and W. Gibson, 1997: The PRISM approach to mapping precipitation and temperature. Preprints, *10th Conference on Applied Climatology*, Reno, NV, Amer. Meteor. Soc., 10-12.

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.

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.

Hansen, M., R. DeFries, J. R. G. Townshend, and R. Sohlberg, 2000: Global land cover classification at 1-km resolution using a decision tree classifier. *International Journal of Remote Sensing*, **21**, 1331-1365.

McCoy, J., and K. Johnston, 2001: *Using ArcGIS Spatial Analyst*. ESRI, 232 pp.

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]

Trimarco, R. A., K. L. Sheets, and K. K. Hughes, 2005: Building a gridded climatological dataset for use in the statistical interpretation of numerical weather prediction models. Preprints, *15th Conf. on Applied Climatology*, Savannah, GA, Amer. Meteor. Soc., JP 1.6.

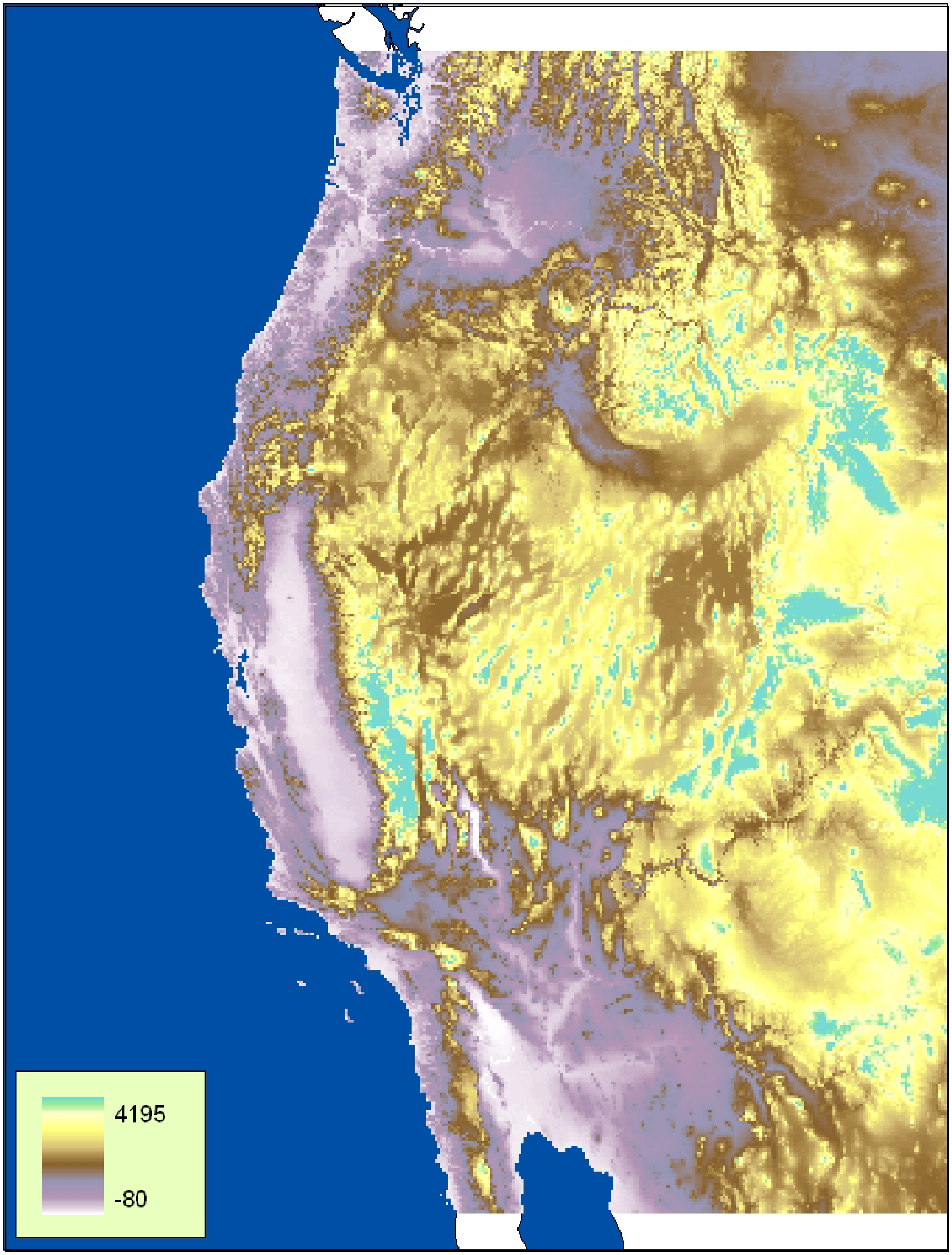

Figure 1. The 5-km elevation data, in meters, translated from the USGS1-km topographical data.

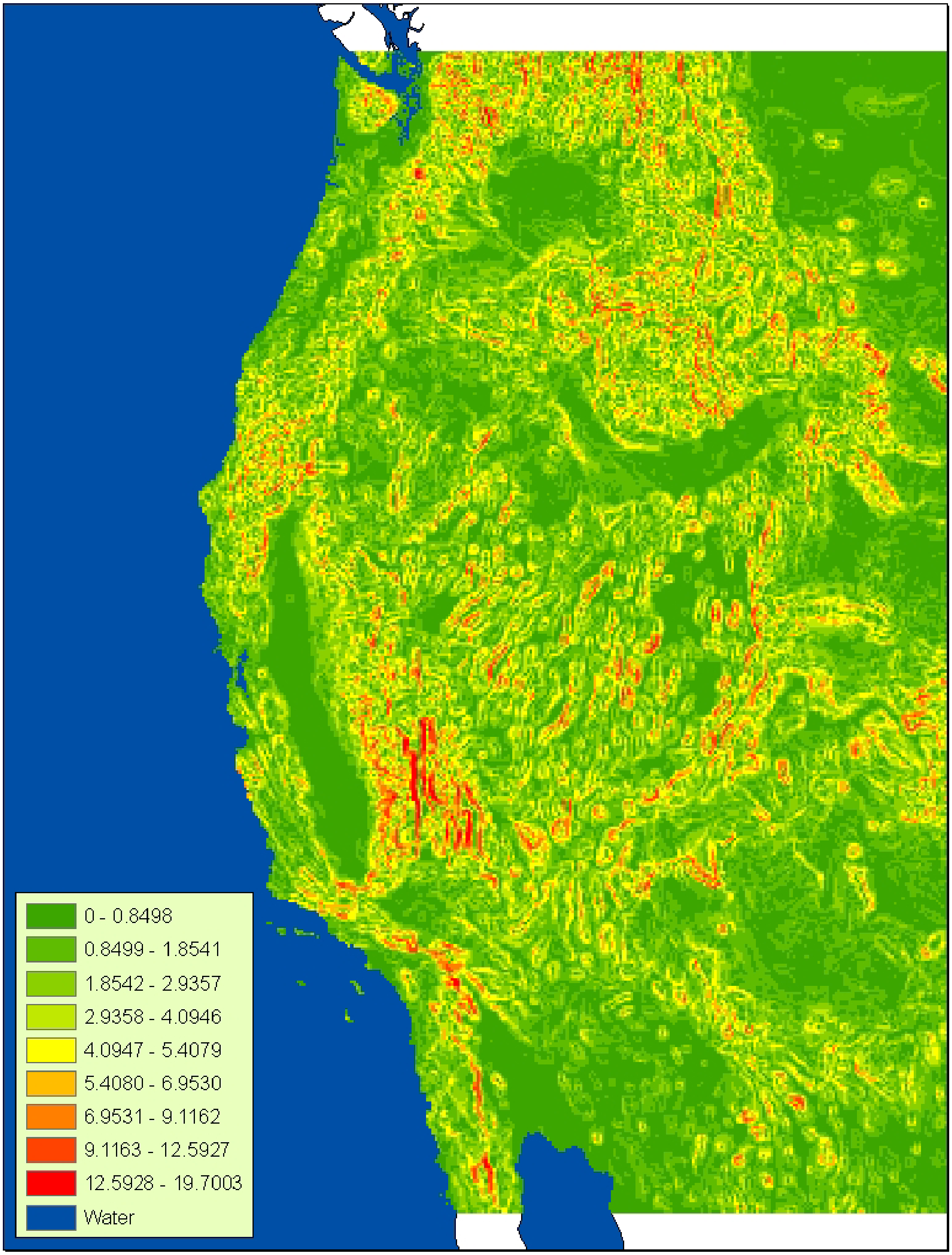

Figure 2. The 5-km slope, representing change in elevation for each cell as a percent, created from the 5-km elevation.

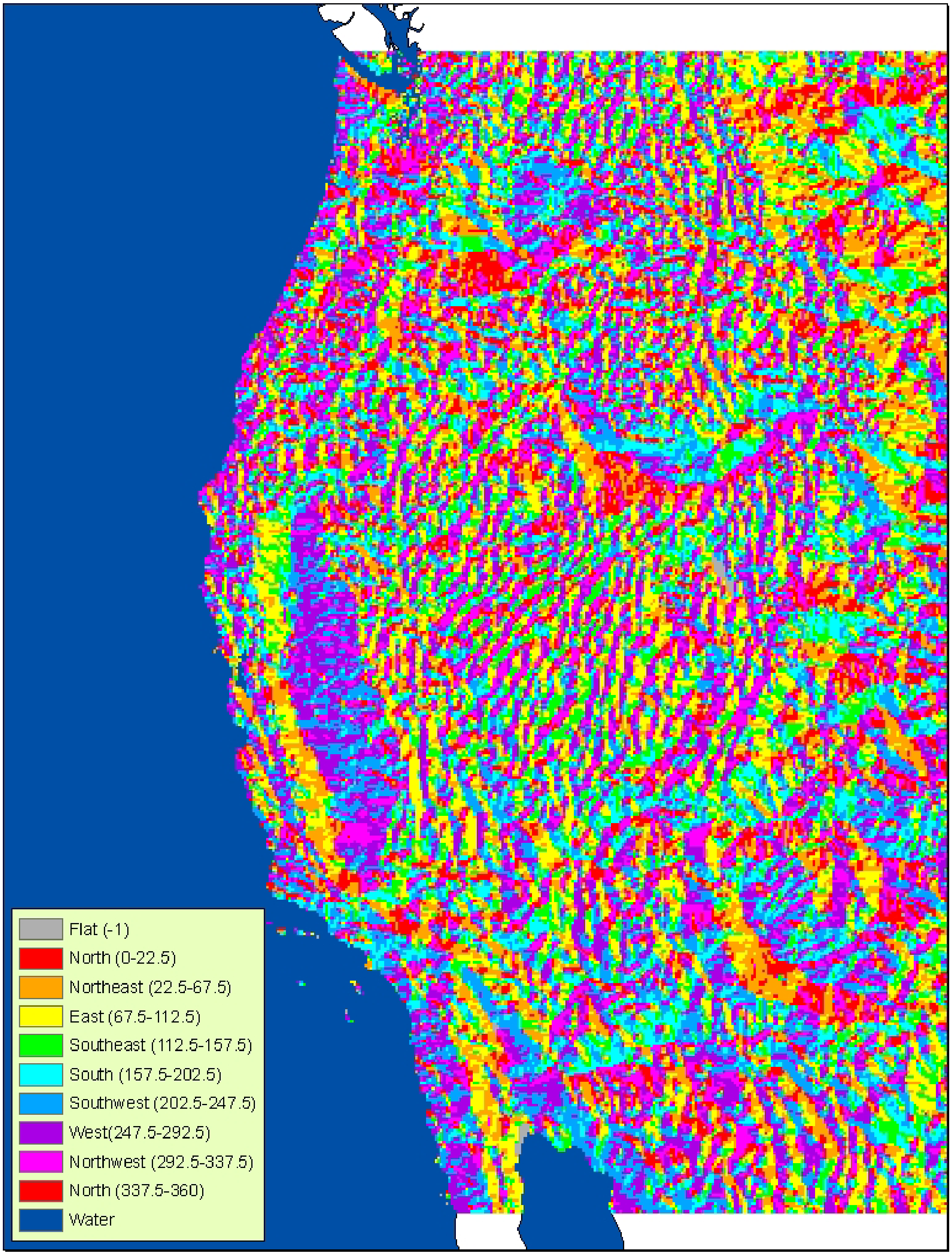

Figure 3. The 5-km aspect data, indicating the compass direction of the downward facing slope, created from the 5-km elevation.

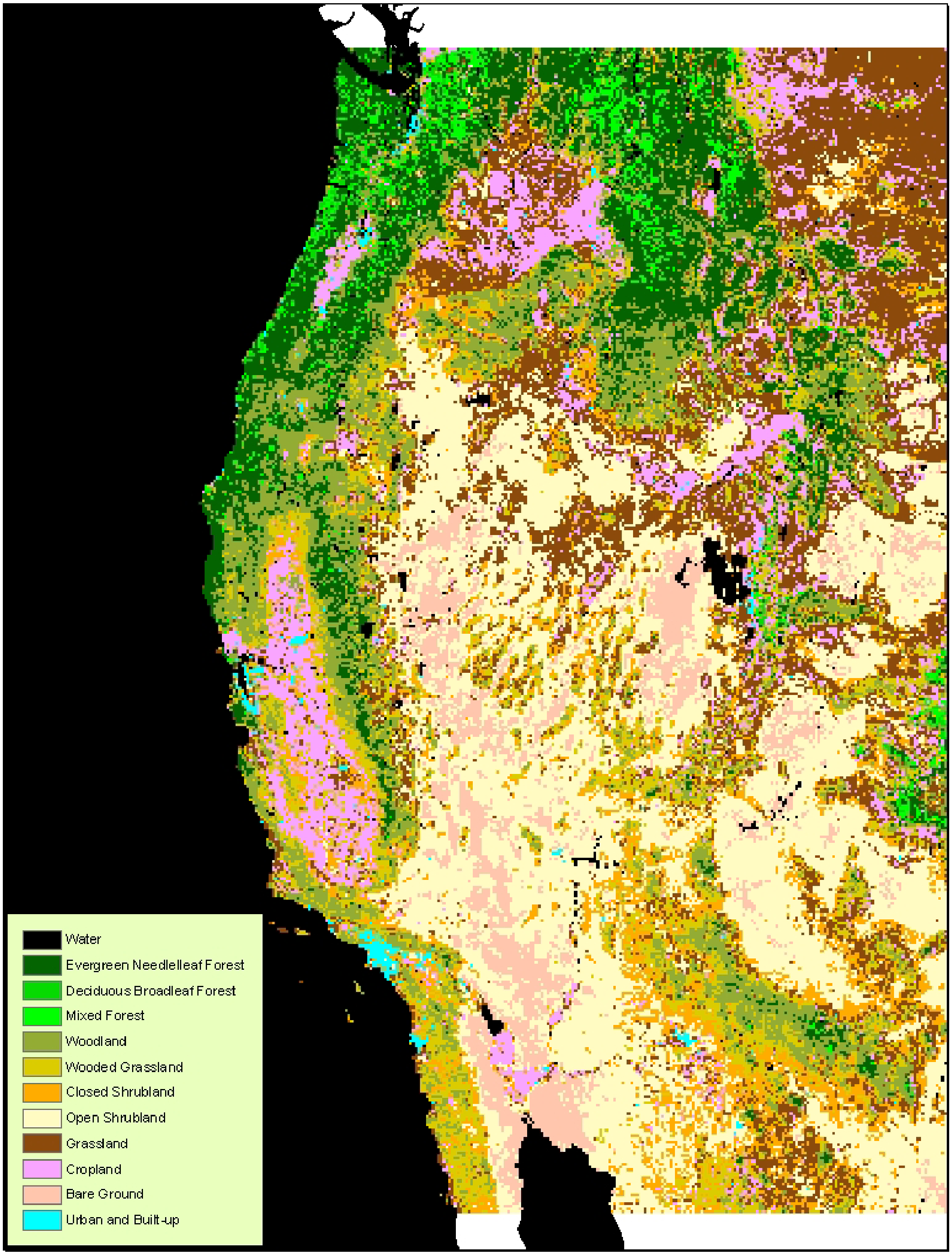

Figure 4. The 5-km land cover maps created from the 1-km GLCF land cover data.

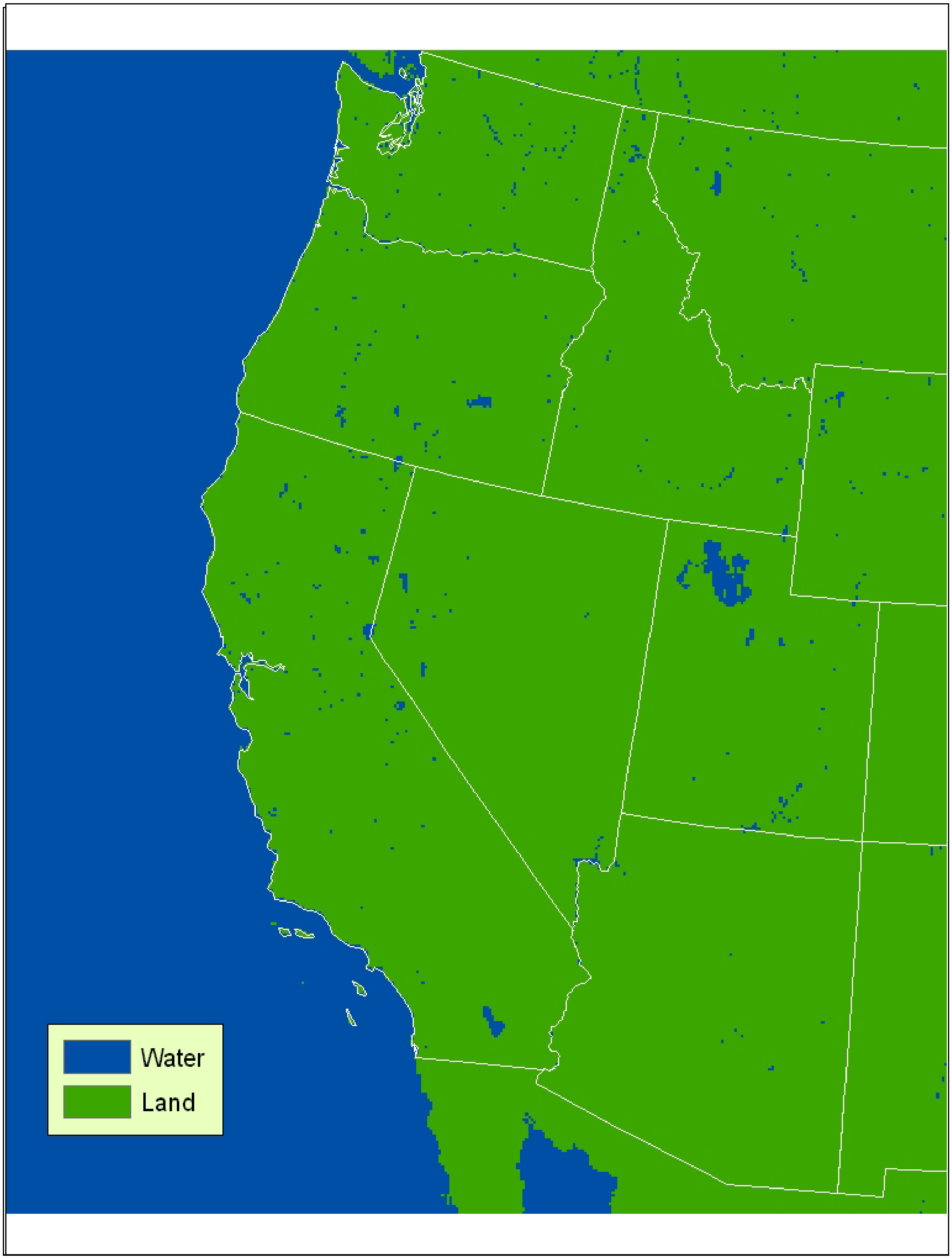

Figure 5. The 5-km land water masks developed from the 1-km GLCF land cover data. These data are the foundation of the station designations discussed later in the paper.

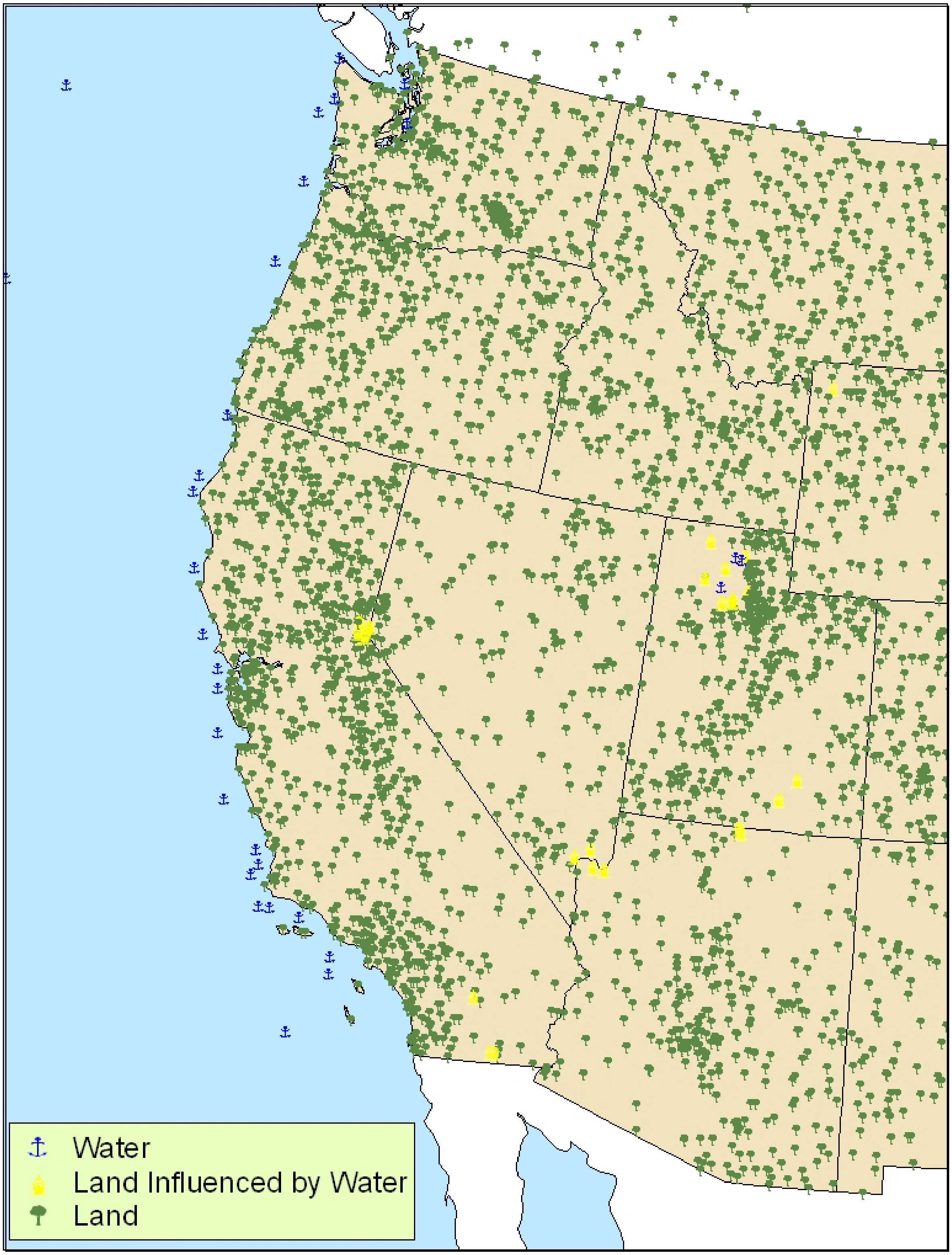

Figure 6. Sites designated in the gridded MOS system as land stations, inland stations influenced by water, and water stations.