Mapping geographic data

Our study of the Laguna's natural resources and anthropogenic effects has been aided by the use of a geographic information system, commonly referred to by the acronym "GIS". A geographic information system is a database of *spatially referenced* data that is stored and accessed using specialized software, which is nowadays commonly installed on a personal computer. This software is used as both an organizational filing system, where many disparate types of place-based data are co-located, and as an analysis tool, where the proximity of features enables the researcher to investigate correlations. The system employed during the development of this restoration and management plan is a software system called "ArcGIS version 9" from the Redlands, California company ESRI. Its use has proven to be very worthwhile.

The spatially referenced data used during the development of this report take four different forms. *Point data*, locates a place on earth using latitude north of the equator and longitude west of the prime meridian. Examples of point data include the location of mountain peaks; creek confluences and springs; weather, water quality and hydrology collection stations; restoration sites; rare species observations; and, invasive plant occurrences. Because a point does not occupy an area, some point data is applied in an approximate sense, that is, the referenced location is taken to be the center of the observed phenomena, with some approximate radii taken to be the range of the observation. An example of this might be the observation of a rare animal species, which the observer noted to be in one place, but which has a habitat larger than that single point.

Another type of data, *line data*, locates a feature using a series of points. Examples of this include the location of streams; geologic fault lines; streets, railroads and trails; and topographic contours. Isohyets, representing the contour of the average annual precipitation pattern, are another example of line data. Just as point data does not represent an area, line data has a certain fuzziness related to the width of the feature being represented. For example, the width of a creek is not precisely defined; indeed, its width varies throughout its course. Additional variability in line data is related to observational error, such as with geologic

fault lines, which are sometimes difficult to locate with any more than 50-foot accuracy. Accompanying all line data are assumptions about its width, such as the width of a road; or descriptions of width variability, such as stream classification systems which catalog different reaches of a watercourse using different symbols; or width error, such as isohyets that lie somewhere within a quarter-mile of the representative line. Researchers using line data for analysis need to be cognizant of this approximation, variability or error.

A third type of data are represented by ovals, rectangles, and other polygonal shapes. Each of these shapes covers an area in space that is represented by fixing its perimeter with a series of latitude/longitude pairs. Examples of *polygonal data* include the location of lakes and ponds; underground aquifers; surficial geologic deposits; soil types; land owner parcels; land uses; planning zones; and, public open spaces. Error and approximation enter into the precision of these perimeters. The degree of approximation in polygonal datasets is sometimes given in terms of scale, for example a layer representing the distribution of geologic alluvium may be accurate to no more than 1:62,500 (low resolution), or a soil boundary may be accurate to no more than 1:24,000 (medium resolution), or a landowner parcel boundary may be accurate to 1:1000 (high resolution).

The fourth type of spatial data used in the development of this restoration and management plan are data in the form of raster coverages. *Rasters* are pixilated representations of features, such as photographs, scanned images, and satellite acquired data. Such datasets consist of a very large number of pixels, organized into a regular arrangement of rows and columns, where each pixel represents a data point, and carries with it a value. Often the values are photographic grayscales, but they needn't be, pixels in a raster coverage frequently contain scaled ranges of values which represent densities, for example the coverage of impervious surfaces or tree canopy. To obtain full-color coverages, such as with color aerial photogrammetry, three raster datasets hold separate values for the red, green, and blue components of the dataset. The two most important raster datasets used by this plan are the aerial photogrammetric dataset of 2002 and the USGS 10-meter digital elevation model. Other rasters used in our analysis include: canopy cover; impervious surface cover; contemporary and historical USGS topographic maps; scanned maps from historical archives that we have georeferenced; and, early land grant *diseño's* from the 1850s.

Spatial data used for this report thus comes in four forms, point, line, polygon, and raster, providing a rich collection of information about the

watershed's natural resources and the patterns of human land use. But beyond this summary of data forms, it's necessary to document here another aspect, albeit an arcane one, of the data used in our work, that is, the projection of this data. Spatial data is most conveniently thought of as being a flat grid representing an area of the earth's surface, but in fact, this simplistic view of spatial data ignores the complex geometry associated with projecting the earth's not-quite-spherical shape onto a flat plane (or a piece of paper). Projections of many types have been created to reduce the distortion and error associated with this 3D to 2D conversion. For our document we have chosen the Lambert Conformal Conic geographic coordinate system using the "State Plane California II FIPS 0402 (Feet) North American Datum of 1983 (NAD83)" projection, for all of our data and maps. Data obtained from other sources has been re-projected to this coordinate system, prior to using it, with the tools provide in the ArcGIS 9.0 software.

Data for our work has been collected from a number of different sources which are described here. We provide here basic bibliographic publishing information, a brief description of the data's accuracy when known, and the Uniform Resource Locators (URLs) which point to the owner's world wide web distribution portal for the GIS shapefiles containing the data; future researchers should use the most up-to-date versions of these datasets.

Area of interest

Watershed boundaries: The delineation of watersheds is a deterministic process. Given a digital elevation model and a pour point, GIS tools can be used to determine all water that flows to that point. This process has been done by several groups such as the California Dept. of Water Resources, California State Water Resources Control Board, California Resources Agency, California Dept. of Forestry and Fire Protection, and the US Geological Survey. In 1999, the California Interagency Watershed Mapping Committee worked to reconcile the different naming conventions and boundaries used by these groups. This work was updated, documented and republished in May 2004 as "California Interagency Watershed Map of 1999 (CalWater 2.2.1)". It was obtained from the California Spatial Information Library at http://gis.ca.gov.

For our definition of the Laguna de Santa Rosa watershed we combined these California Dept. of Forestry and Fire protection subwatersheds: Pool Creek, Upper Windsor Creek, Mark West Springs,

Porter Creek, Humbug Creek, Van Buren Creek, Monroe, Pine Creek Reservoir, Rincon Creek, Upper Matanzas Creek, South Fork Matanzas Creek, Santa Rosa Creek Reservoir, Santa Rosa Creek, Laguna de Santa Rosa, Crane Creek, and Blucher Creek. In addition, the eastern portion of the Mirabel Park watershed, at the confluence of the Mark West/Laguna complex with the Russian River, was added to these sixteen sub-watersheds to comprise the 253.7 square mile area used throughout the development of our work products.

Sonoma County: The Sonoma County outline was derived from the Sonoma County PRMD planning areas shapefile published by Sonoma County ISD in 2004 by appending the nine planning areas together and dissolving the result into a single polygon.

City incorporation boundaries: Five incorporated cities fall entirely or partially within the watershed: Santa Rosa, Rohnert Park, Cotati, Windsor and Sebastopol. The incorporation boundaries are published by the Sonoma County ISD and were obtained in 2004 from https://gis.sonoma-county.org.

Planning and human use

Planning zones: The Sonoma County Permit and Resource Management Department defines the unincorporated areas of the county as falling into one of several planning zones: diverse agriculture, land intensive agriculture, land extensive agriculture, resources & rural development, rural residential, urban residential, recreation/visitor serving commercial, general commercial, limited commercial, limited commercial traffic sensitive, general industrial, limited industrial, public & quasi-public, and river. The polygons representing these land use categories were obtained from Sonoma County ISD in 2004 from https://gis.sonoma-county.org.

Roadways: Sonoma County ISD publishes a shapefile containing the public roads, private lanes, railroads and paved trails of the county. This shapefile contains attributes which indicate street name, street class, public or private ownership, and address range. It was obtained in 2004 from https://gis.sonoma-county.org.

Parcels: The Sonoma County Assessors Office has developed a shapefile containing the approximate outlines of all owner parcels within the county. For each parcel several pieces of data are available: owner's name, owner's mailing address, assessor's parcel ID, and land use code. The land use code is used to document the principal human use of each parcel. It is finely categorized into approximately 400 codes falling into the broad

groupings of residential, commercial, industrial, irrigated and dry farming, recreational, institutional, and government. Note that this *use code* is not the same as the PRMD planning zones. There are more than 93,000 parcels within the Laguna's watershed. This dataset was obtained in 2004 from https://gis.sonoma-county.org with supplemental use code and owner name data obtained on CD from the Assessors office.

Additional datasets have been derived from the Parcels dataset for use in special purpose analyses for this plan. These derived products include: golf courses, car washes, industrial facilities, mineral processing places, gasoline service stations and service shops, and rural residences with septic systems.

Proposed public use: This dataset, developed through the cooperation of the RMP stakeholder council in 2005, shows the proposed location for bike trails, hiking trails, multi-use trails, kayaking sites, hunting sites, parking and access points, and picnicking sites. All of these are proposed sites, which have general support from community and advocacy groups, but do not have approval or funding to be created.

Public Land Survey: The township and range survey grid for the Laguna is referenced off the Mount Diablo baseline meridian, placing the watershed in townships 6, 7, 8, and 9 North and in ranges 7, 8, and 9 West. Each six-mile by six-mile township is divided into 36 one mile square sections. The townships with part or all of their surveyed area falling within the watershed are: 6N7W, 6N8W, 6N9W, 7N7W, 7N8W, 7N9W, 8N6W, 8N7W, 8N8W, 8N9W, 9N8W, 9N9W. This regular grid is squeezed between the Spanish and Mexican land grants that were honored when California became the 31st state of the United States of America. The grants that fall partially or wholly within the watershed are: a tiny part of Sotoyome, half of Molinos, all of San Miguel, all of Cabeza de Santa Rosa, half of Los Guilicos, all of Llano de Santa Rosa, most of Cotate, the corner of Blucher, a small part of Roblar de la Miseria, and a tiny part of Petaluma.

The PLS dataset contains the boundary lines for both the regular grid of townships and the confirmed land grants. This dataset is very useful in the process of georeferencing scanned versions of paper-based historic maps and *diseño's*. It was published in ArcGIS format in 1993 and revised in 1997 by Ray McDowell of CERES and was obtained in 2005 from http://gis.ca.gov

Public and protected lands: This dataset has been prepared by the Sonoma County Agricultural Preservation and Open Space District. It consists of polygonal shapes outlining SCAPOSD holdings, conservation easements,

parks, schools, utilities, etc. Every city, county, district, state, and federal land in Sonoma County is included. The 'Access' and 'Public' columns are used to symbolize these types of open space: public—open, private—open, public—closed or restricted, private—closed or restricted, utility, various county holdings—closed.

Land cover

Photogrammetric data: Several aerial photography datasets are available for Sonoma County. The first of these was flown in the early 1940s as part of the US Army effort to defend the West Coast during the onset of World War II. Subsequent flights in 1950 and 1960 were used to assist the Soil Conservation District in their publication of the soil survey CA097 for Sonoma County. Black and white aerial photos were taken in 1960, 1970, 1980, 1990, 2000, and 2002. Color aerial photographs were taken in 2004. The 1950 and 1960 photos are available in paper form at the USDA-NRCS offices in Petaluma and were examined for this plan. The other decadal datasets were not examined. For this report, the 2002 dataset, which we believe was taken in late spring of 2002, was used. It was obtained in MrSID data compression format and used directly from within the ESRI software. The resolution of this dataset clearly shows objects as small as automobiles; with some guesswork objects as small as cows and horses can be distinguished. This data was obtained in 2004 from https://gis.sonoma-county.org

Digital Ortho Quads: Photogrammetry from July 10, 1993 is available from the USGS website. It is divided into quadrangles which match their 7.5 minute series of topographic maps. These are not as clear as the more recent fly-overs.

Landsat Orthoimagery Mosaic: The June 10, 2001 and November 8, 2001 Landsat flyover have been tiled together to display the Laguna from space. Three bands have been selected from the eight spectral bands available for each frame. These are bands 4 (near-infrared), 3 (red), and 2 (green), typically displayed as red, green and blue, respectively. Resolution is based on a 1-arc-second (approximately 30-meter) sample interval. The image simulates color infrared film as a "false color composite". It is available from http://seamless.usgs.gov

Impervious surfaces: A raster dataset showing the extent of impervious surfaces within the Laguna has been derived from the National Landcover Dataset of 1992. Original GIS source data is available from http://landcover.usgs.gov/natllandcover.asp.

Canopy cover: A raster dataset showing the extent of canopy cover within the Laguna has been derived from the National Landcover Dataset of 1992. Original GIS source data is available from http://landcover.usgs.gov/natllandcover.asp.

Agriculture: The California Dept. of Conservation publishes county-by-county datasets which show the pattern of land use as categorized into five types: 1) prime farmland, 2) important or unique farmland, 3) grazing land, 4) forested, mining, rural, or other land, and 5) urban areas. These datasets are published every two years since 1984, showing the trends over time. The most recent dataset for Sonoma County, at the time of this report is 2002, but for most other counties is 2004. These datasets were obtained in 2005 from ftp://ftp.consrv.ca.gov/pub/dlrp/FMMP/.

Natural resources

USGS National Elevation Dataset (NED): This digital elevation model (DEM) was obtained from the USGS's ⅓ arc-second seamless dataset. Each pixel represents 10 meters in an X/Y direction. Absolute vertical accuracy is poor, with the metadata for the data stating "A vertical RMSE of 7 meters is the desired accuracy standard. A RMSE of 15 meters is the maximum permitted." Nevertheless, the relative accuracy is reasonably good and has been used, with proper precautions, to make hydrologic interpretations for this plan. When LIDAR data becomes available, or when the USGS 1/9 arc-second (3-meter resolution) data becomes available, a review of this document should be undertaken to correct any elevation and hydrologic deficiencies. The NED used for this report was obtained in 2005 from http://seamless.usgs.gov.

Geology: The US Geological Survey published, in the year 2000, Open-File Report 00-444, "Preliminary maps of Quaternary Deposits and Liquefaction Susceptibility, nine-county San Francisco Bay Region, California : A digital database". The datasets within this report where compiled, in part, from 1:24,000 quadrangle maps carefully added to an ArcGIS system. The resolution of this dataset should not be interpreted beyond that limit. Three important data layers are present in this dataset. The first is the location and type of Quaternary deposits; these are surficial deposits of the Holocene and late Pleistocene epochs, which in the Laguna are principally alluvial fan, basin, terrace, and undifferentiated deposits. This first layer comes in two parts, with the area covered by the deposit represented separately from the observed edge of the deposit, and whether this observed edge is approximate, concealed or well-located.

The second data layer contains the location of geologic faults with an accompanying classification of the fault's age. The third layer classifies the study area's susceptibility to liquefaction.

Soils: The Soil Survey Geographic (SSURGO) Database is published by NRCS and made available online via the Soil Data Mart at http://soildatamart.nrcs.usda.gov/report.aspx?Survey=CA097&State=CA.

Sonoma County's soils are also reported under the electronic Field Office Technical Guide (eFOTG) report 097 which can be accessed at http://efotg.nrcs.usda.gov/treemenuFS.aspx?Fips=06097.

Also available for use with this dataset are two spreadsheets of data, "Soil characteristics", which lists the saturated conductivity (kSat) values taken from USDA-NRCS "Sewage Disposal" report for Sonoma County soils [CA097], and "RUSLE2 related attributes", which documents hydrologic group (A, B, C or D), erosion loss tolerance (Kf with coefficients ranging from 0.17 to 0.43), erodability loss tolerance (T values from 1 to 5), and sand-silt-clay percentages (collectively summing to 100%). Many other soil-related attributes are available for Sonoma County soils.

Isohyets: Average annual precipitation within Sonoma County is mapped using isohyetal lines, which represent equal values of rainfall for the county. These values were plotted by the Sonoma County Water Agency, and published most recently in 2005. It is believed that these values cover the period from 1900—1983, although verification of this has not been possible. Weather stations within Sonoma, Marin, Napa, and Mendocino are sparse and data collection efforts vary, with some weathers stations only recording a few years of precipitation data. For the purposes of this plan however, the isohyets give an accurate enough representation to make good restoration and management decisions. For the most up-to-date version of this dataset contact the Sonoma County Water Agency at http://scwa.ca.gov.

24K Hydrography: The National Hydrography Dataset is a comprehensive collection of points, lines, and polygons, representing the wells and springs, the creeks and rivers, and the lakes and reservoirs of the United States. It was developed, in electronic form, by digitizing the 1:100,000 and 1:24,000 maps of the USGS and other agencies. The dataset is published in regional subsets, with the Laguna de Santa Rosa being covered within the "1801" basin. The dataset is published in medium resolution as NHDM1801, and in high resolution as NHDH1801. It was obtained in 2004 from http://nhd.usgs.gov/.

1K Hydrography: Responding to the need to create a finer resolution to the hydrography dataset for the watershed, the Laguna Foundation

developed a new dataset to represent the location of ponds and creeks. The protocol for the development of this dataset was multi-step: 1) using the 2002 aerial photogrammetry of Sonoma County within ArcGIS, the route of all visible waterways and water bodies was traced using a 1:1000 scale, 2) using the National Elevation Dataset and its byproducts (100ft, 50ft, 20ft, 10ft, 5ft, 2ft, and 1ft contours), obscured waterways were traced at a 1:1000 scale; 3) layered examination of georeferenced USGS topographic quadrangles was performed to double check the work for completeness (but not for accuracy) and to add feature names; 4) layered examination of soils and surficial geology was undertaken to spot-check ambiguous areas; 5) field reconnaissance was conducted using a hand-held GPS unit to verify the position of stream crossings at roadways. Note that the National Elevation Dataset, which strives for 7-meter positional accuracy, was adjusted on the fly and heavier reliance was place on the aerial photography. We believe this 1K Hydrography dataset to be an accurate depiction of the creeks, channels and ponds of the watershed as it looked in 2005 to a positional accuracy of 10 feet.

Vernal pools: The Llano and Piner regions are home to a concentrated collection of vernal pools. The locations of these have been carefully mapped by the California Dept. of Fish and Game as part of the work to create a California Tiger Salamander (CTS) Conservation Strategy for the Santa Rosa Plain. Creation of this dataset entailed the use and interpretation of georeferenced color aerial photographs taken in 2004, analysis of CTS occurrences documented within the California Natural Diversity Database, and consultation with knowledgeable professionals. The dataset consist of 1225 features categorized into possible and verified CTS sites. This is a credible and well documented dataset. For more information contact the Wildlife and Habitat Data Analysis Branch at CDFG http://www.dfg.ca.gov/.

Vernal pool plants: In addition to the geographic data for CTS, locations of three endangered Santa Rosa Plain vernal pool plants: Sonoma Sunshine (*Blennosperma bakeri*), Sebastopol meadowfoam (*Limnanthes vinculans)*, and Burke's goldfields (*Lasthenia burkei*) were mapped by Dr. Christina Sloop during spring of 2006. This work was done as part of a CDFG-funded survey of the genetic variability of the accessible extant populations in the Santa Rosa Plain. The demographic and geographic data of this effort spans forty-three logged population sites, and is retained in the Laguna Foundation's GIS repository and reported to the California Natural Diversity Database.

Groundwater protection areas: The California Dept. of Pesticide Regulation (DPR) has identified 3718 sections of land throughout the state which are vulnerable to pesticide movement to ground water. Each section is about one square mile in size and its condition indicates whether the section is primarily sensitive to pesticide movement to ground water via runoff, leaching, or both runoff and leaching. This dataset was obtained in 2005 from http://gis.ca.gov

Groundwater basins: The California Dept. of Water Resources has published Bulletin 118, revised most recently in 2003, which delineates groundwater basin boundaries within the state. DWR defines a groundwater basin as "an area underlain by permeable materials capable of furnishing a significant supply of groundwater to wells or storing a significant amount of water." The delineation of these basins were initially based on the presence of unconsolidated alluvial soils and then further evaluated through review of geologic and hydrogeologic reports, well completion reports, court-determined adjudicated basin boundaries, and contact with local agencies. This dataset was obtained in 2005 from http://www.groundwater.water.ca.gov.

Data stations: There are many concurrent efforts to collect data about natural phenomena that occur within the watershed. Agencies that collect this data use field locations with special on-site measuring devices. The location of these field sites may change over time, but for the most part sites are chosen strategically with an eye toward balancing cost and statistical significance. Some of the agencies collecting data include the California Irrigation Management Information System (CIMIS), the US Geological Survey (USGS), the California Dept. of Water Resources (DWR), City of Santa Rosa Utilities & Public Works Departments, Sonoma County Water Agency (SCWA), National Climate Data Center, and the California Dept. of Forestry and Fire Protection. Data collected at these stations is highly specific and frequency of collection varies from continuous (every 15 minutes) to monthly. Some data is automatically collected and published to the world wide web, other data is manually entered. The range of data collection spans the realms of water quality, hydrology and weather. Some additional volunteer data collection sites, for such things as water quality and well water depths, are also collected, but are poorly represented in this dataset. Note that this dataset, developed by the Laguna Foundation, only documents the location of these data stations, it does not attempt to republish the actual data.

History

Topographic quadrangles: The Laguna has been mapped by the USGS and its predecessors since 1919. The most recently published series of 7.5 minute (1:24,000) quadrangles were published in 1954 and photo-revised between 1971 and 1980. They are available in two file formats, the first being TIF files, where each file contains one digitized quadrangle sheet. These are available in the folder 38122 at ftp://casil-mirror1.ceres.ca.gov/casil/gis.ca.gov/drg/. The quadrangles for the Laguna are: Calistoga, Camp Meeker, Cotati, Glen Ellen, Healdsburg, Kenwood, Mark West Springs, Santa Rosa, Sebastopol, and Two Rock (Files C5, C6, C7, D5, D6, D7, D8, E5, E6, E7)

The second file format is MrSID, where multiple quadrangles are combined to form a mosaic. These are also available at ftp://casil-mirror1.ceres.ca.gov/casil/gis.ca.gov/drg under the folder "7.5_minute_series_albers_nad83_mosaic/MrSID". The Laguna (and Sonoma County) is covered in files o_nw0301 and o_nw0401.

Historic waterways: Examination of historic maps has revealed changes to the watershed's hydrography. Some of these historic waterways were traced using georeferenced versions of the "New Historical Atlas of Sonoma County, California. Illustrated.", published in 1877 by Thomas H Thompson and the "Illustrated atlas of Sonoma County, California" published in 1898 by Reynolds & Proctor, and the 1:62,500 and 1:24:000 USGS topographic quadrangles from 1914-16, 1942-44, and 1951-54.

Historic topographic maps: Historic topographic maps are available with partial coverage for the Laguna as follows: Santa Rosa 7.5' quadrangle for 1916, 1944, and 1954; Sebastopol 7.5" quadrangle for 1942 and 1954; Healdsburg 7.5" quadrangle for 1940 and 1955.

Two 30x60 topographic maps cover the entire Laguna at a scale of 1:100,000—Healdsburg (1972) and Napa (1983).

A US Army Corps of Engineers tactical map is available for Calistoga at the scale of 1:62,500. It was published in 1927 and reprinted 1936 using survey work of 1915, which interestingly does not yet include contours for all of the mountainous areas.

Historical maps of Sonoma County: Spanish and Mexican land grants were formally defined by survey work done after 1848, which was submitted to the U.S. Northern California District Court. Prior to the formal survey and court case, the applicant was required to submit a sketch (*diseño*) of the area that he believed was granted to him. These sketches usually showed important features of the landscape including features of the sur-

rounding area. Although there are only a few such land grants wholly or partially within the Laguna's watershed, because there were such contentious boundary disputes, many land cases which appeared before the court include sketches which cover areas within the Laguna. The major grants within the Laguna are: half of Molinos, all of San Miguel, all of Cabeza de Santa Rosa, half of Los Guilicos, all of Llano de Santa Rosa, and most of Cotate. The land cases which are of interest to the Laguna watershed are:

A-197 and B-196, Rancho los Guillicos
B-107 and B-108, Cañada de Jonive
B-128 Llano de Santa Rosa,
B-267, B-269, B-270, B-271, B-272 Cabeza de Santa Ros
B-351 Cotate

Land case *diseño's* were obtained through the courtesy of the University of California Bancroft Library via the Internet at http://www.oac.cdlib.org/findaid/ark:/13030/hb8489p15p

The county's first assessor map was published in 1865 by Bowers. Only four copies are known to exist, one at the State Library in Sacramento, two owned by the California Dept. of Parks and Recreation, and one at the Sonoma County Museum. An effort, spearheaded by volunteers of the Sonoma County Heritage Network, is underway to capture this very large map (it is approximately 5' by 6' in size) and make it available to researchers.

Two historical atlases for the county are of significance to the study of land use patterns in the county. The "New Historical Atlas of Sonoma County, California. Illustrated.", published in 1877 by Thomas H. Thompson is available in scanned format courtesy of the David Rumsey Map Collection.

The "Illustrated atlas of Sonoma County, California" published in 1898 by Reynolds & Proctor, is also available in scanned format courtesy of the David Rumsey Map Collection. An original is available at the Sonoma County Archives in Los Guillicos.