Evaluation of water-level trends using spatiotemporal kriging in the Mimbres Basin, southwest New Mexico

Geoffrey Rawling

Evaluation of water-level trends using spatiotemporal kriging in the Mimbres Basin, southwest New Mexico

Geoffrey Rawling

New Mexico Bureau of Geology and Mineral Resources

PROJECT FUNDING

Cover: The Mimbres River in flood 15 miles upstream of Deming, February 18, 2020.

CONTENTS

Figures

Tables

EXECUTIVE SUMMARY

Water-level measurements in the Mimbres Basin made in early 2020 were combined with data extending back to 1980 to characterize water level trends in the region. The geostatistical method of spatiotemporal kriging was used to create water-level maps every five years from 1980 until 2020. Changes in water-levels over these five-year intervals were calculated. Compared to traditional spatial kriging, the spatiotemporal approach offers improved precision, more realistic maps of water levels and water-level changes, fewer artifacts due to changing well networks over time, and overall less uncertainty in predictions.

Several notable patterns since 1980 are revealed in the sequential maps. From Deming to Columbus, water levels have declined up to 75 feet. Water-level declines and expansion of cones of depression appear to have slowed south of Deming and increased around Columbus in the past ten years. Water levels west of Red Mountain, east of the Florida Mountains, and northeast of Columbus have risen as much as 32 feet, presumably as a result of decline in pumping for irrigation in these areas that has resulted in flattening of cones of depression. The vicinity of Whitewater and Faywood shows net water-level rises over the 40 year period, but declines have occurred in the past ten years. Water levels have varied considerably along the reach of the Mimbres River south of Faywood, where most of the river's flow infiltrates.

The spatiotemporal kriging approach is more challenging than other interpolation methods traditionally used in hydrogeology. However, the cost in funds and staff-hours of field studies to gather water-level data has always been high and continues to rise. Therefore it is prudent to analyze the data collected at such great expense with methods that will extract the most useful information. The present work builds on other recent studies and demonstrates that spatiotemporal kriging of water-level data is superior to spatial kriging in this regard.
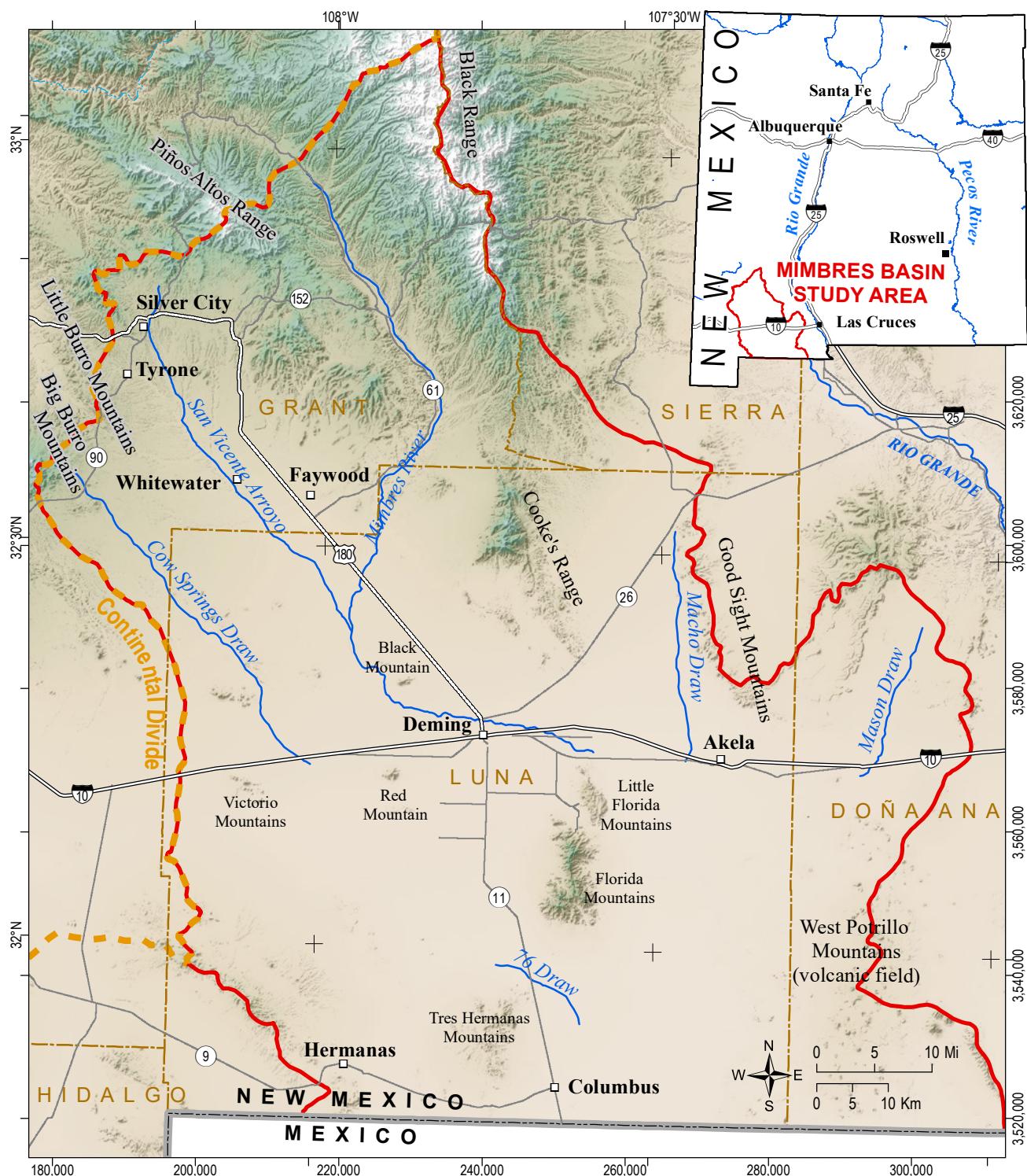
Figure 1. Location of the Mimbres Basin in southwest New Mexico, USA. Red line marks the boundary of the Mimbres Basin.

I. INTRODUCTION

The Mimbres Basin watershed and subjacent aquifers are located in Doña Ana, Grant, Luna, and Sierra Counties of southwest New Mexico (Figure 1). Major population centers in the basin include Silver City, Hurley, Tyrone, Deming and Columbus. The surface extent of the Mimbres Basin in the United States is shown in Figure 1, and covers about 11,590 km² (4,475 mi²). The topographic basin is closed and the watershed and aquifer extend south into Mexico. The northern boundary of the surface-water basin coincides with the Continental Divide north and west of Silver City. The Continental Divide also forms the western boundary of the basin and extends south from the Little Burro Mountains almost to the international border. On the east, the basin boundary extends south and east around the Good Sight Mountains and across the West Potrillo Mountains to the international border. The basin boundaries transect a great variety of topographic relief, geologic structures, and rock types (Figures 1 and 2).

Agriculture, mining, and ranching form the basis of the regional economy. Groundwater is the major water supply for agricultural, industrial, municipal and domestic uses. Luna County encompasses a large portion of the Mimbres Basin and almost all of the irrigated acreage. Magnuson et al. (2019) estimated that of the 29,880 acres irrigated in Luna County in 2015, 28,480 acres (95%) were irrigated solely with groundwater. Irrigated agriculture and public water supply together accounted for 83,100 acre-feet of groundwater withdrawals in 2015 (Magnuson et al., 2019).

Groundwater has been vital to the development of the Mimbres Basin since the early 20th century, and thus its importance has long been recognized. Groundwater development for agriculture began near Deming in 1908 and irrigated acreage and groundwater use expanded rapidly. Until that time, agriculture was limited to lands near the upper Mimbres River that could be irrigated with surface water (Hawley et al., 2000, and references therein). Water levels in wells across the basin have been measured by the U.S. Geological Survey (USGS), the New Mexico Office of the State Engineer (NMOSE), the New

Mexico Bureau of Geology and Mineral Resources (NMBGMR) and other entities since extensive groundwater development began. The database of water-level data maintained by the Aquifer Mapping Program at the NMBGMR contains measurements from the Mimbres Basin dating back to 1910.

Manual measurement of water levels in wells is a crucial task for understanding groundwater resources (Taylor and Alley, 2001), yet it is labor-intensive, time-consuming, and expensive. Figure 3 illustrates that the overall number and frequency of water-level measurements in the Mimbres Basin has declined greatly since 1980. Prior to 2020, the last basin-wide water-level measurement campaign was in 2012, and far fewer measurements were made at that time than in previous field studies. Developing water-table or water-level elevation maps is a major use of water-level measurements. Such maps convey vital information about an aquifer, such as groundwater flow directions, locations of recharge and discharge, groundwater-surface water interactions, and control of lithology and geologic structure on groundwater occurrence and flow. These maps can be used to infer aquifer hydraulic properties and their spatial variation, and assess the impacts of groundwater pumping on water resources. It is critical to use high-quality data and robust methods to interpolate between the locations of field measurements.

The geostatistical interpolation method of kriging was originally developed in the mining industry and has a strong theoretical background (Cressie, 2015). It is applied regularly across the gamut of earth and environmental sciences (Webster and Oliver, 2007) and groundwater hydrology in particular (Kitanidis, 1997), including recent studies in New Mexico (Rinehart et al., 2016, Rawling and Rinehart, 2018). Spatiotemporal kriging is an extension of spatial kriging to include the additional dimension of time, and it derives from an equally well-developed theory (Cressie and Wikle, 2011; Gräler et al., 2016; Wikle et al., 2019). It has also seen use in earth and environmental sciences (e.g., Huevelink and Griffith, 2010; Gasch et al., 2015) but only recently have there been applications to groundwater hydrology (Ruybal et al., 2019; Varouchakis and Hristopulos, 2019).
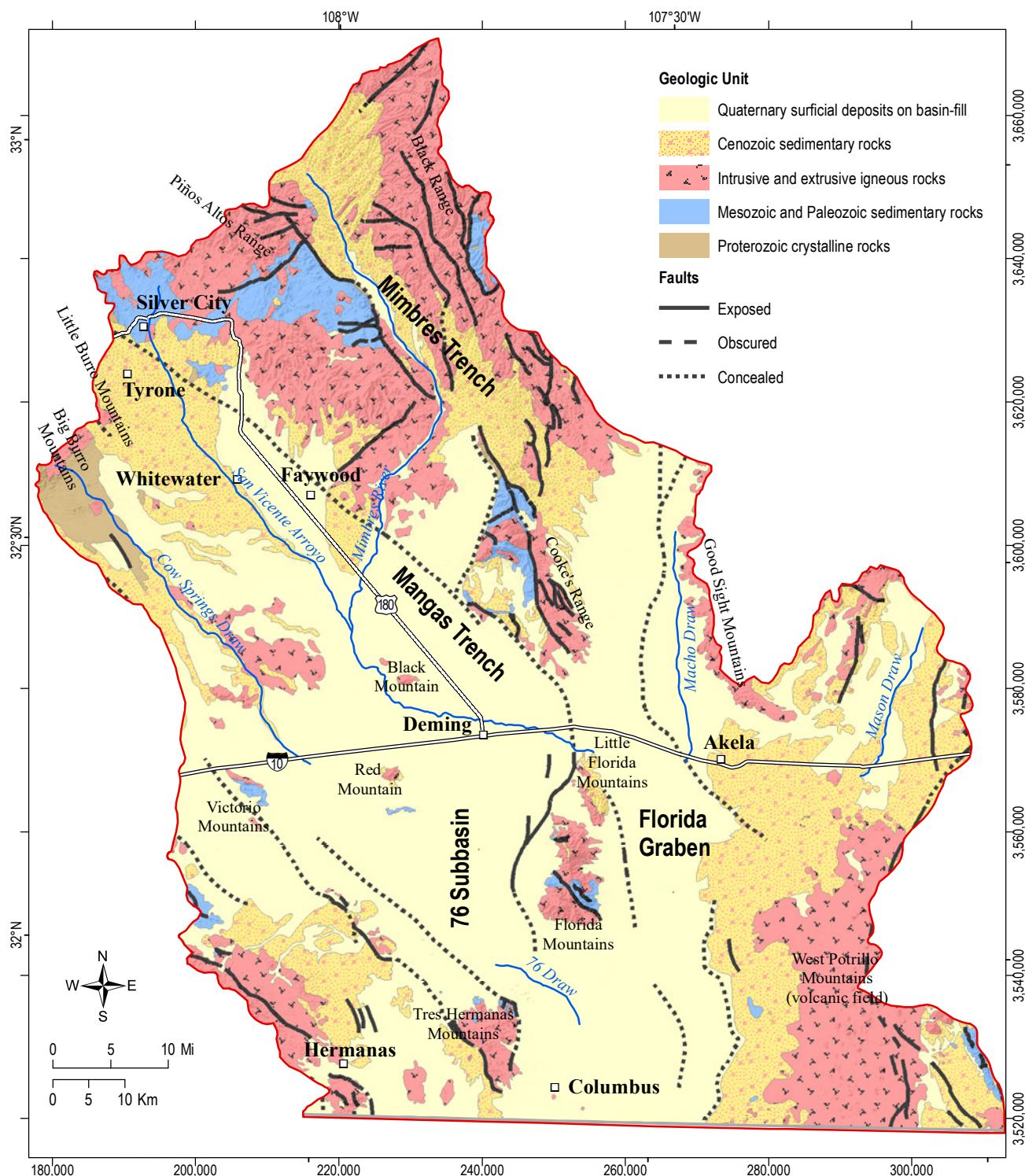
Figure 2. Geologic map of the Mimbres Basin. Geology simplified from New Mexico Bureau of Geology and Mineral Resources (2003).

The present work was initiated with two main objectives. The first was to characterize current groundwater conditions by again measuring water levels in wells across the Mimbres Basin. The goal was to revisit the wells measured in 2012, as these wells had been measured many times previously and thus new measurements would add to existing water-level time series. The second was to create a series of maps of water levels, and changes in water levels over time, using spatiotemporal geostatistical methods. The data manipulation and computations were performed using the R language (R Core Team, 2018) and the gstat, space, and spacetime packages (Pebesma, 2004; Pebesma, 2012, Bivand et al., 2013). The spatiotemporal approach maximizes the amount of information and insight that can be extracted from the new and existing water-level data by quantifying the correlation of the measurements in space and time. This increases the precision and accuracy of the derived water-level maps. The method allows estimates to be made of water levels at times when there are no data, and in places where there are no data, and provides estimates of the uncertainty in both cases.
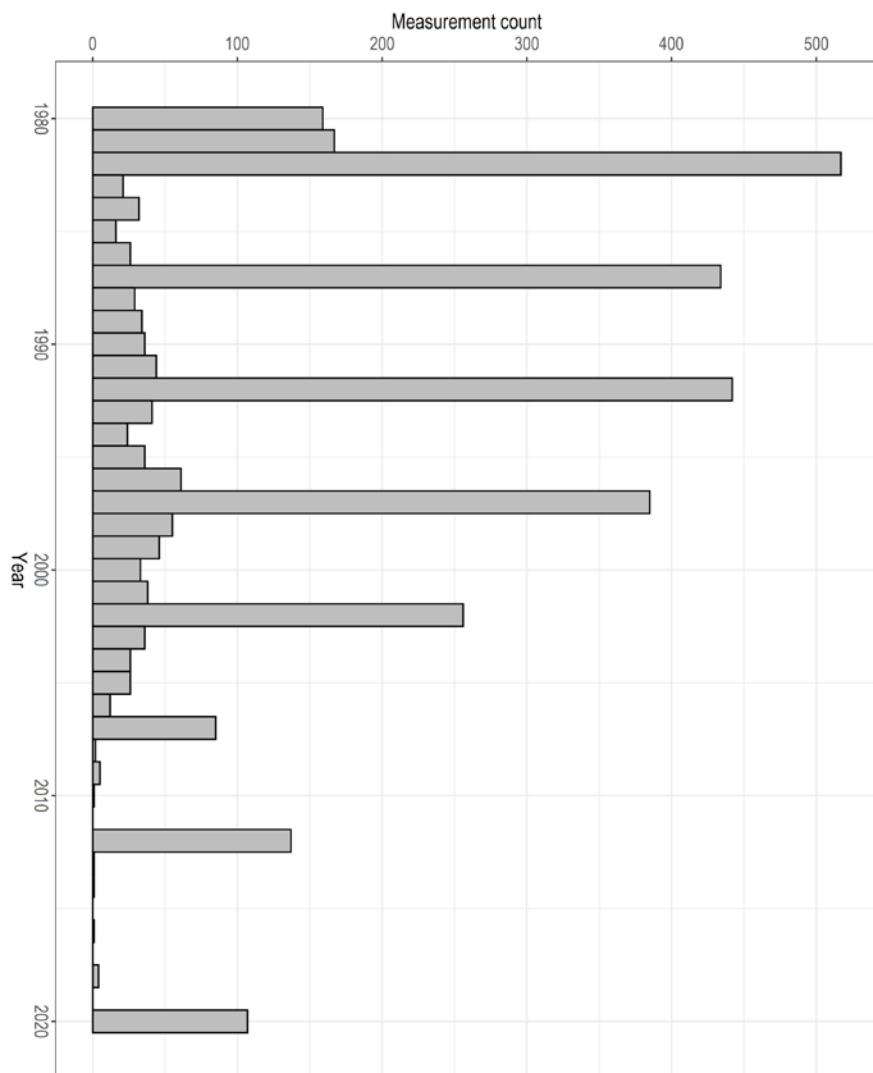
Figure 3. Histogram of water level measurements since 1980 in 663 wells in the Mimbres Basin illustrating the overall temporal variability in measurements and decline in frequency since the 1990s.

II. BACKGROUND

Previous Work

The geology of the Mimbres Basin is shown at 1:125,000 scale on regional geologic maps by Seager et al., (1982; northeast part of the basin), Seager (1995; southern part of the basin), and at 1:250,000 scale by Drewes et al. (1985; western part of basin). Clemons (1998) describes the geology of the Florida Mountains, the prominent range that rises near the center of the Mimbres Basin. Data from these maps and many other sources were compiled by Hawley et al. (2000) to create a geologic map of the entire Mimbres Basin at 1:500,000 scale, including the portions of the basin in Mexico. The aforementioned maps all contain geologic cross sections based on surface mapping, well data such as cuttings, core, and geophysical logs, and interpretation of gravity, magnetic, and seismic data. They also contain reference lists of older and/or larger scale mapping that was used in the geologic compilations. More recent work includes the important study of Heywood (2002), who interpreted the thickness of alluvial fill and depth to bedrock in the Mimbres Basin using isostatic residual gravity anomalies.

The earliest groundwater studies in the Mimbres Basin were the work of Darton (1914, 1916). Other historical regional-scale groundwater studies include White (1931), Theis (1939), Conover and Akin (1942), and Trauger (1972). McLean (1977) summarized existing hydrologic data and conditions in the mid-1970s and Hawley et al. (2000) contains a detailed and thorough summary of the geology and hydrogeology of southwestern New Mexico, including an exhaustive reference list. Kennedy et al. (2000) is an overview of the main conclusions of the Hawley et al. (2000) study.

Numerous unpublished reports by consultants have discussed issues such as regional water planning and estimates of current and future supply and demand (e.g., D.B. Stephens and Associates, 2009, for the Deming region), and numerical groundwater flow models developed to address particular hydrologic questions (e.g., Romero and Cook, 2009, managed aquifer recharge near Silver City). Hanson et al.

(1994) described the hydrogeologic framework of the basin and a preliminary two-dimensional numerical groundwater flow model. Finch et al. (2008) describe a more detailed three-dimensional flow model originally developed for the Chino Mines Company. The flow model currently used by the NMOSE to administer and manage water rights in the Mimbres Basin is described by Cuddy and Keyes (2011).

A notable consequence of ongoing groundwater pumping in the Mimbres Basin is land subsidence and the formation of earth fissures. Galloway et al. (1999) provide a review of the mechanisms and consequences of land subsidence due to extraction of groundwater. Contaldo and Mueller (1991) and Haneberg and Friesen (1995) described land subsidence and measured fissure formation in the Mimbres Basin.

Regional Geology

Simplified geology of the Mimbres Basin is shown on Figure 2. This overview is largely based on sources cited in the "Previous Work" section, Clemons and Mack (1988), and New Mexico Bureau of Geology and Mineral Resources (2003). The landscape is characterized by steep and rugged fault-block mountains at the basin margins that bound wide desert plains formed on the intervening structurally down-dropped valleys. This landscape has resulted from middle Miocene to late Pleistocene extensional block faulting (Clemons and Mack, 1988). The Florida, Tres Hermanas, and Victorio Mountains, and numerous smaller prominences, are intrabasin mountains surrounded by broad, low-relief pediments and bajadas (Figure 1). The regional geology is quite complex, as the Mimbres Basin lies at the intersection of three distinct geologic/tectonic provinces—the southern Colorado Plateau, the Basin and Range, and the Rio Grande rift (Hanson et al., 1994).

The bedrock in the mountains range from Proterozoic to Cenozoic in age. Proterozoic rocks are dominantly granite with minor syenite, gneiss, and schist, and are abundant in the Big Burro Mountains. Lower Paleozoic intrusive igneous rocks are present in

the Florida Mountains. Paleozoic sedimentary rocks range from Cambrian to Permian in age. They are dominated by limestone and dolomite, with subordinate sandstone, shale and siltstone, and claystone redbeds. These units are found in the Florida Mountains, Tres Hermanas Mountains, Cookes Range, and the Pinos Altos Range. Mesozoic rocks consist of Cretaceous sandstone and shale in scattered exposures in Cookes Range, the Tres Hermanas Mountains, and Pinos Altos Range. Lower Paleozoic intrusive igneous rocks are present in the Florida Mountains. Mesozoic rocks consist of Cretaceous sandstone and shale in scattered exposures in Cookes Range, the Tres Hermanas Mountains, and Pinos Altos Range. Tertiary sedimentary, volcaniclastic, and intrusive and extrusive igneous rocks are widespread in all of the mountains. Quaternary basalt flows comprise the volcanic field of the West Potrillo Mountains at the southeast margin of the area.

The sedimentary basins underlying the valleys are filled with Cenozoic sedimentary rocks, with lesser amounts of volcaniclastic and volcanic rocks. These units host the aquifer system that is the focus of this study. The depth to bedrock in the basins varies widely (see below), and the thickness of the sedimentary fill varies accordingly (Hanson et al., 1994; Seager, 1995; Hawley et al., 2000). As summarized by Hawley et al. (2000), regionally a distinction can be made between an upper, poorly consolidated basin-fill sedimentary unit with thickness ranging from 100 to 300 m (330 to 990 ft), and a lower, usually more indurated unit composed of conglomerate, sandstone, and mudstone as much as 1,000 m (3,280 ft) thick. The lower unit is often identified as, or correlated with, the well-indurated Gila Conglomerate and/or Gila Group, which is widely exposed at the northern margins of the Mimbres Basin area and in the mountains to the north. The majority of the wells in this study are completed in the upper unit, or the uppermost part of the lower unit. The near-surface deposits in the Mimbres Basin are Upper Quaternary in age and were deposited by the Mimbres River and San Vicente Arroyo drainage systems (Seager, 1995; Love and Seager, 1996). These unconsolidated deposits can have saturated thicknesses up to 30 m (100ft).

Hawley et al. (2000) proposed a conceptual framework of subdividing the basin fill into hydrostratigraphic units characterized by distinct lithofacies associations. This scheme relates depositional environment (e.g., alluvial fan, braided stream channel, basin-floor playa) to sediment textural characteristics, bedding thickness, spatial geometry, and connectivity, and ultimately hydraulic conductivity and groundwater production potential. Although geologically and hydrologically sound, in practice, these classifications cannot be mapped except in a very general way in the subsurface in the Mimbres Basin due in part to a lack of borehole geophysical data (Hawley et al., 2000; see plate 1 therein). The difficulty of correlating individual beds of gravel, sand, silt, and clay using lithologic logs or cuttings was noted even in the earliest studies in the basin (Darton, 1916; Contaldo and Meuller, 1991).

The subsurface geometry and geologic structure of the Mimbres Basin consists of numerous subbasins with great variations in thickness of the basin-fill units. Several workers prior to Heywood (2002) presented structural-tectonic maps and cross sections of the region with various names for the subbasins, their mapped or inferred extents, depth estimates from limited well, gravity and seismic data, and mapped or inferred geometries of bounding faults (e.g., Wilkins, 1986, Hanson et al., 1994, Seager, 1995; Hawley et al., 2000). The gravity study of Heywood (2002) clarified the subsurface geometry of the basins and clearly identified their shallow and deep portions (Heywood, 2002, Fig. 4; also reproduced as contours in Finch et al., 2008, Fig. 1). Great thicknesses of basin fill exist northeast of Deming, and along the northwest-southeast trend of San Vicente Arroyo and the lower Mimbres River (Figure 2). This area corresponds to the Mangas Trench of Hanson et al. (1994), also referred to as the San Vicente Subbasin by Hawley et al. (2000). The northwest-southeast region of thick basin fill west and southwest of the Florida Mountains corresponds to the Seventy-Six basin of Hanson et al. (1994), also referred to as the southern Deming Subbasin by Hawley et al. (2000). The Florida graben or subbasin east of Cookes Range and the Florida Mountains contains three areas of thick basin fill separated by saddles of shallower bedrock. Large areas of the Mimbres Basin have total basin-fill thickness of less than 100 m (330 ft).

The cross sections of Seager et al. (1982) and Seager (1995) show that almost everywhere in the central and southern Mimbres basin, the basin fill sedimentary units overlie middle to lower Tertiary volcanic rocks. This is also seen in the oil-test wells east and northwest of Deming studied by Clemons (1986). Hawley et al. (2000) notes that the upper Mimbres River drainage and a region east of the Tres Hermanas Mountains in the southern Florida subbasin/graben are the only areas in the Mimbres Basin where these pre-Gila Group volcanic rocks are important parts of the aquifer system.

Regional Hydrogeology

Hanson et al. (1994), Hawley et al. (2000) and Finch et al. (2008) discuss the surface and groundwater hydrology of the Mimbres Basin in detail. The Mimbres River usually has perennial flow from its headwaters to the vicinity of the Grant – Luna county line. Flood flows during spring snowmelt or exceptional rainfall events often extend further south to near Black Mountain northwest of Deming and very rarely to Deming itself (Figure 1). Infiltration of Mimbres River water is a main source of recharge to the Mimbres Basin aquifer (Hawley et al., 2000). Small amounts of perennial flow occur in upstream reaches of San Vicente Arroyo from urban irrigation and leakage in Silver City and treated effluent discharge from the sewage treatment plant. This water infiltrates upstream of the confluence with the Mimbres River. Flood flows in the normally dry drainages (e.g., Cow Springs Draw, Seventy-Six Draw, Macho Draw) extending from the basin bounding mountains contribute small amounts of recharge. Groundwater in the Mason Draw drainage in the northeast part of the Mimbres Basin has been shown to discharge to the Mesilla Basin to the east (Hawley et al., 2000).

Estimates of total recharge vary but all agree that it is only a few percent of the annual total precipitation over the basin. Potential evapotranspiration is very high and areal recharge across the basin floors is negligible. Most recharge comes from infiltration of Mimbres River water and mountain front recharge (Wilson and Guan, 2004) along the northern margins of the basin. Little or no recharge is derived from the highlands south of Deming, except perhaps the largest few watersheds. Hanson et al. (1994) developed a water budget and estimated total recharge to the United States portion of the Mimbres Basin at about 52,000 acre-feet per year. Their two-dimensional groundwater model suggested this was too high, and that a value of about 55% of this, or about 29,000 acre-feet was more reasonable. The smaller value is in agreement with results from calibration of the more elaborate model of Finch et al. (2008), who estimated total recharge at 29,050 acre-feet per year. Finch et al. (2008) did not specify whether their results were for the whole basin or just the portion in the United States. These numbers are about one third of the estimated groundwater withdrawals in Luna County in 2015 (83,100 acre-feet; Magnusson et al., 2019). Groundwater pumping and evapotranspiration from playas are the main discharges. The latter has largely been replaced by the former since large scale groundwater pumping began (Hawley et al., 2000).

Groundwater flow is generally south and southeast from the northern highlands to the international border, with progressively decreasing gradient. Gradients and flow directions are highly perturbed from pre-development conditions in areas with large water-level declines due to pumping. South and southeast of Columbus, flow directions have completely reversed, and groundwater flows north into the United States. Finch et al. (2008) subdivided the Mimbres Basin into four major hydrogeologic regions based on patterns of recharge and discharge, and large scale hydrogeologic properties determined by the geologic structure such as the depth to bedrock in the basins (Figure 4).

Darton (1916) noted water-bearing zones of 1.5 to 12 m (5 to 40 ft) in thickness and highly variable lateral continuity. Darton (1916) and White (1931) described the Mimbres Basin aquifer as overall phreatic or unconfined, with local zones of partial confinement and subartesian conditions, the latter indicating water levels that rise in wells above the water-bearing zone, but not flowing at the surface. Hawley et al. (2000) noted that the upper 100 m (330 ft) of the Mimbres basin aquifer "at many localities" contains fine-grained material such as playa-lake beds and/or is partly indurated, suggesting potential for local confined or partially confined conditions. Almost all of the wells in this study are located in outcrop areas of basin-fill sediment or Gila Group (Figure 4) and are assumed to be completed in these units. Many irrigation wells have multiple screens and likely tap more than one water bearing zone. Water levels do not show distinct trends with depth of wells. It is assumed that the water levels analyzed here are derived from one complex but continuous, regional, unconfined aquifer system.
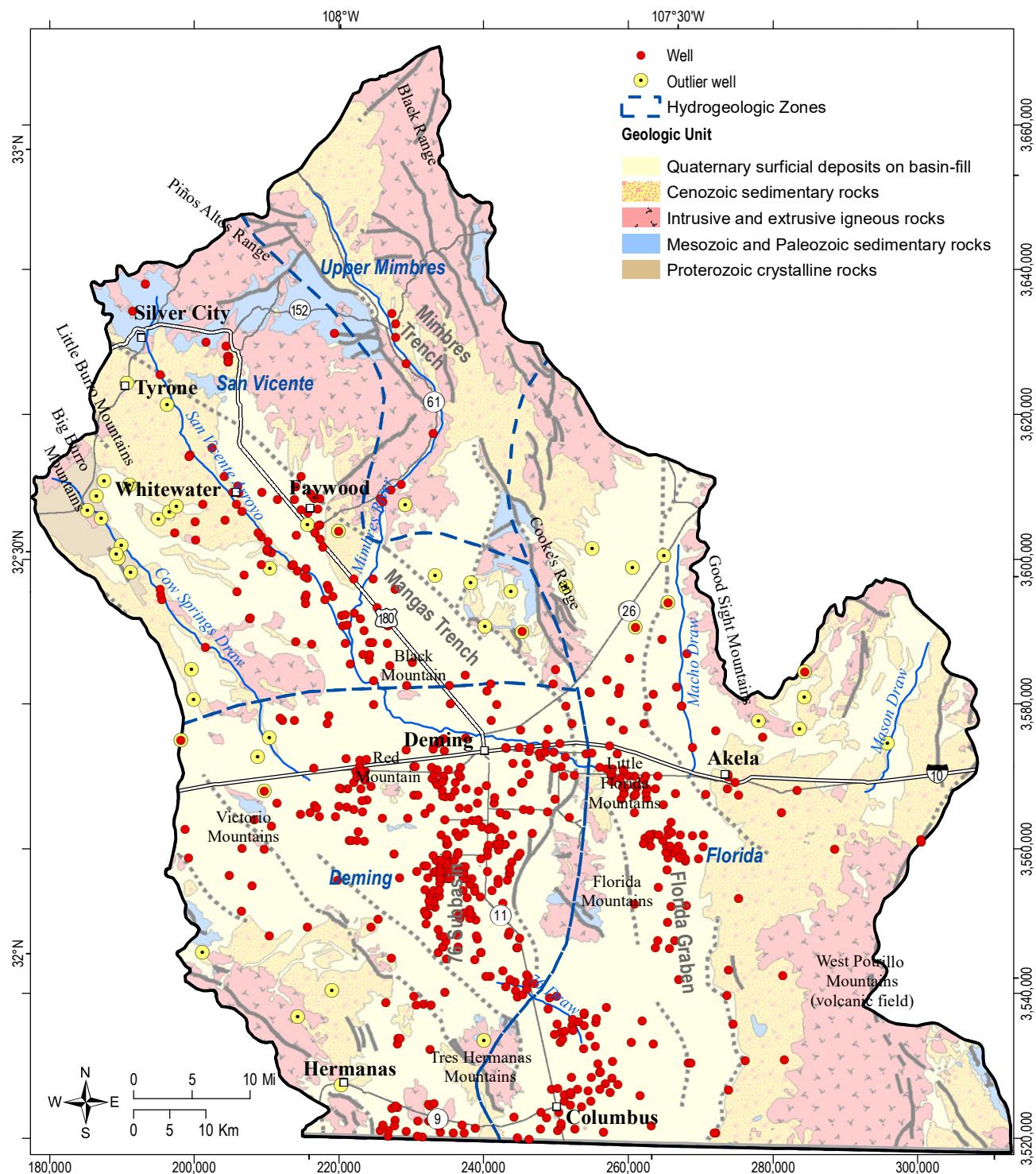
Figure 4. Map showing locations of wells with water-level measurements since 1980 used in this study. Outliers are explained in the text. Note that some wells have both outlier measurements and "good" data. Hydrogeologic zones are defined by Finch et al. (2008). Geology as in Figure 2.

III. FIELD METHODS AND DATA REVIEW

The last basin-wide set of water-level measurements was conducted in late winter and early spring 2012 by a private consultant under contract to the NMOSE. The field notes and records of this work were obtained from the NMOSE with the goal of revisiting all of the wells for additional measurements in late winter 2020. In 2012, 158 wells were measured, in January and February 2020, 107 wells were measured. We could not measure numerous wells for various reasons, including collapse of the well, denial of permission by the owner, locked gates, muddy and impassable roads, and flooded crossings of the Mimbres River. Water levels were measured with a steel tape or an electric water-level sounder if the well was unequipped. At least two depth-to-water measurements were made at each well to attempt repeatability within 0.02 ft (0.6 cm). Well locations were recorded with a handheld GPS unit.

In addition to the 2012 and 2020 data, well locations and water levels for all wells in the Mimbres Basin were obtained from the NMBGMR Aquifer Mapping program database. This database contains all readily available water-level data from the USGS, NMOSE, and NMBGMR for the Mimbres Basin since 1910. These data were filtered and processed in the following ways:

1. Only wells with water-level measurements since 1980 were considered, and only data since 1980 were included in this study.

2. Hydrographs for each well were viewed in an interactive script, where all measurements during the normal irrigation season, March through October inclusive, were removed. The exception was year 2012, in which there were many measurements in March. These were kept unless there was an obvious irrigation signal.

3. Measurements with a USGS data quality flag were removed. A data flag indicates measurement issues, such as the water-level recovering from pumping, or adjacent wells were being pumped at the time of the measurement.

4. Any other measurements that appeared to be significant outliers were removed, based on interactive inspection of the hydrographs.

5. All well locations were projected from UTM coordinates to Lambert Conformal Conic projection. This was done because the boundary of UTM zones 12 and 13 passes though the study area, resulting in well locations in two different projections, and in some cases wells in UTM zone 12 were recorded with UTM zone 13 coordinates. These were converted to UTM zone 12 before projection to Lambert Conformal Conic.

6. The land surface elevation at each well was determined from a 4.5 meter (15 ft) lateral resolution digital elevation model, and the water-level elevation was then determined from the depth-to-water measurements.

Water-level data are recorded and used by regulatory and management agencies in New Mexico in English units (feet, for depth to water and water-level elevation), whereas location and elevation data for wells may be recorded in English units, metric units, or both. In this study all data were converted to metric units prior to the analyses. Intermediate results of analyses are presented here in metric units. Final results such as water-level elevations and their changes, and estimates of uncertainty, were converted to feet after the analyses were complete. Water-level data and well information retained after the filtering and processing described above are presented in Appendix 4.

Water-level Analysis

Elevations in the Mimbres Basin range from 3,099 m (10,170 feet) in the Black Range to 1,207 m (3,959 ft) along the international border southeast of Columbus, with a strong north to south regional topographic gradient (Figure 1). Thus, there is similarly strong regional gradient in water-level elevations. This is a regional water-level trend that must be accounted for when modelling the water-level surface.

Mathematically, the model used here is spatiotemporal regression kriging (Kyriakidis and Journel, 1999; Hengl et al., 2007; Heuvelink and Griffith, 2010; Ruybal et al., 2019) where the water-level elevation at any point $Z(s,t)$ is a function of space *(s)* and time *(t)*, and consists of two parts:

$$Z(s,t) = m(s,t) + \varepsilon(s,t). \qquad (1)$$

where $m(s,t)$ is a deterministic trend and $\varepsilon(s,t)$ is the stochastic, or random, residual with a normal distribution and zero mean. The trend may be a function of space and time as shown, of space or time alone, or a constant value. Here it is assumed to be constant in time; i.e., $m(s)$. This is the "regression" in spatiotemporal regression kriging. It can be thought of as the long-term baseline pattern of water levels and their variation across the region that is controlled by factors that are essentially constant in time but variable in space on a large scale; e.g., land-surface elevation, broad patterns of subsurface geology, long-term natural patterns of recharge and discharge, etc. Deviations from this trend are the residuals and they account for the local geometry of water levels and how they change with smaller scale spatial patterns of discharge, recharge, and variations of these processes in time. Quantifying and modelling the correlation of the residuals in space and time is what is investigated in the "spatiotemporal kriging" part of spatiotemporal regression kriging.

Regional Trend

In many geologic situations, the water-table elevation of an unconfined aquifer is a subdued replica of the land surface topography (King, 1899; Tóth, 1963; Desbarats et al., 2002). This suggests a model of the regional water-level trend in the Mimbres Basin as a function of geographic coordinates and elevation. The mathematical form of such a model is subjective and nonunique (Huevelink and Griffith, 2010), but ideally it is physically reasonable and explains the large scale variation. Importantly, the residuals from the model (the differences of the data values from the trend model predictions) should be normally distributed, have zero mean, and exhibit stationarity (Webster and Oliver, 2007; Hengl, 2009). The first two conditions indicate that, when all of the data are considered, the trend model does not tend to over- or under-predict the water levels. Stationarity means that the statistics of the residuals do not show spatial trends.

Specifically the mean should be constant and the covariance or correlation of the residuals at any pair of points should only depend on their separation, and not their absolute position (Webster and Oliver, 2007).

Rawling and Rinehart (2018) modelled the regional trend of water levels in the High Plains Aquifer in eastern New Mexico using a third-order polynomial function of the easting and northing coordinates of the wells. This approach was not successful in the Mimbres Basin, as residuals from polynomial functions of the geographic well coordinates are distinctly bimodal, have non-zero mean and are skewed. Experimental spatial variograms calculated from these residuals did not reach a constant sill, but rather showed continuous increase with increasing spatial lag, indicating non-stationarity (Webster and Oliver, 2007; variograms are discussed in the next section).

Principal Component Analysis (PCA) was used to find a better predictor for the regional water-level trend. PCA is a technique for reducing the dimensionality of datasets while minimizing loss of information (Abdi and Williams, 2010; Jolliffe and Cadima, 2016). It creates new, uncorrelated, variables from a dataset under the constraint of maximizing the variance explained by the new variables. The new variables are called the principal components and are linear combinations of the original data, here, the easting, northing, and land-surface elevation coordinates for each well location (in meters). The principal components are completely defined by the structure of the original data, and the solution is unique apart from the sign, which is arbitrary (Jolliffe and Cadima, 2016).

The first two principal components together explain 96% of the variation in the three-dimensional coordinates of the wells (Figure 5; Table 1 and Appendix 1), effectively reducing the well locations to a two-dimensional space. All three spatial coordinates contribute to all three of the principal components (Table 1; each coordinate is multiplied by the loading and these are summed to create the principal component). The regional water-level trend was then

Table 1. Principal component analysis (PCA) loadings for the three principal components.

|	PC1	PC2	PC3
Easting	0.53	-0.78	-0.34
Northing	0.57	-0.62	0.54
Land Surface Elevation	-0.63	-0.09	-0.77
Standard Deviation	1.53	0.73	0.33
Proportion of Variance	0.78	0.18	0.04
Cumulative Proportion	0.78	0.96	1.00
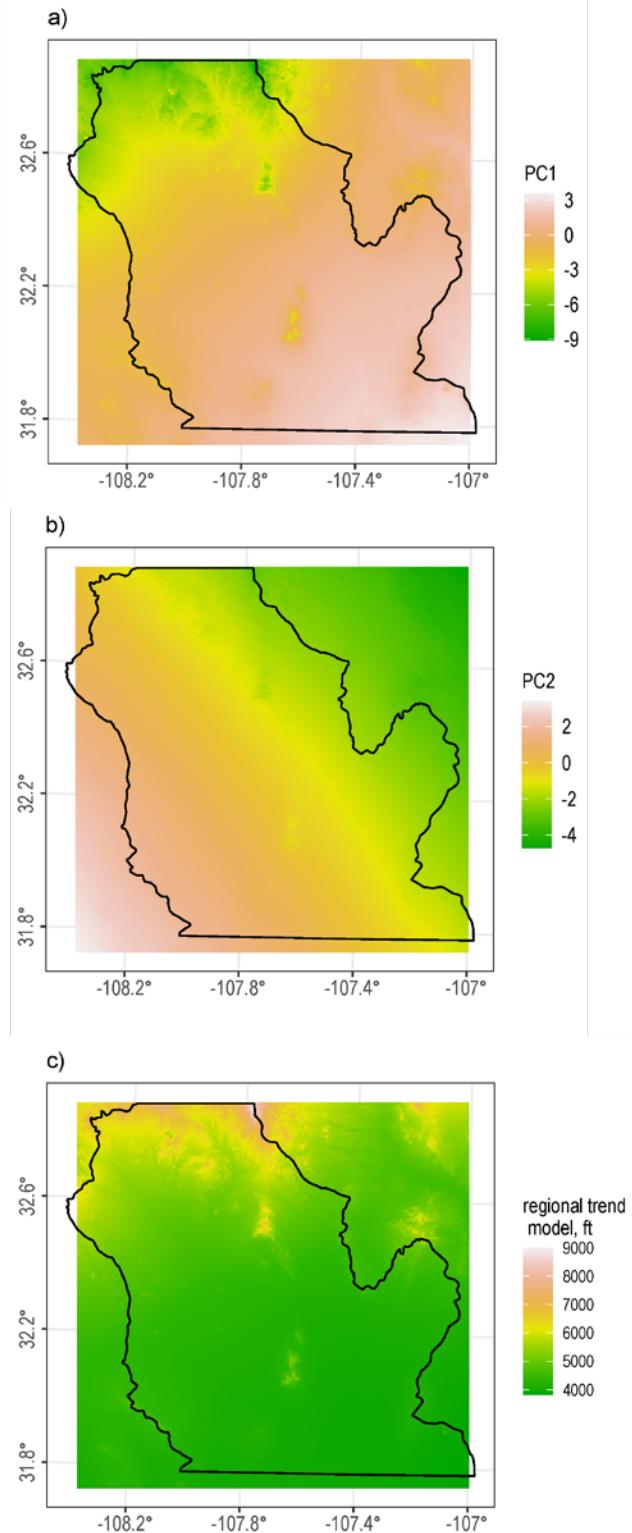
Figure 5. Maps of a) principal component 1 (PC1), b) principal component 2 (PC2), and c) the third-order polynomial trend model of water-level elevation in feet above sea level. The principal components are derived from the east coordinate, north coordinate, and land surface elevation of each well. Only the first two of the three principal components are used in the polynomial trend model, as the first two principal components account for 96% of the variance of the original data set.

modelled with a third-order polynomial function of the first two principal components (Figure 5). This trend is constant in time; for most of the wells there are multiple water-level measurements, so the trend is fit to all of the measurements at each well, regardless of the date of measurement. In Figure 5c, the loadings of the well coordinates (Table 1) were used to calculate the water-level trend prediction at each point of a 1,000 m square grid spanning the study area, where elevations for each grid point were determined from a 4.5 m digital elevation model. Note that the regional trend and subsequent spatiotemporal kriging analysis was not extended to the northern terminus of the basin due to the lack of data in this region. More details of the regional trend are presented in Appendix 1.

Figures 6 and 7 show the residuals from the regional trend fit to the data. Two criteria were combined to identify data points that were not fit, or predicted, adequately by the polynomial trend function. Data points with standardized residuals that are greater than two standard deviations from zero are often considered outliers (Schuenemeyer and Drew, 2011). One can also calculate Cook's Distance, which is an iterative process wherein the parameters of the fitted model are recalculated by dropping each data point in turn and using only the remaining data to fit the model. One common criteria for a high value of Cook's Distance is any value greater than $4/n$, 0.0012 in this study, where n is the number of data (Glen, 2020). Any water-level measurements that satisfied *both* of these criteria were flagged as outliers. These water-level measurements were then dropped from further analysis, and were not used in the spatiotemporal kriging. Well NM-05236 (one measurement) was also dropped. The standardized residual for this measurement was 3.08 and Cook's distance was 0.011. However, it clearly plots amongst the other outlier measurements in Figure 6b. Wells with outlier measurements are shown as yellow circles in Figure 4. Outlier measurements are shown in red in Figures 6b and c. Figure 7 shows the statistics and the spatial distribution of the residuals with the outlier measurements removed.

One could repeat the above procedure of identifying outliers indefinitely, by repeatedly refitting the regional trend model and identifying and removing outliers using the same combined criteria as above. However, there is no definite end to this process and no reason to think that a deterministic trend will ever adequately represent *all* of the data. Most of the outlier data points are on the margins of the study region, in geologically and topographically distinct

areas, where it is plausible that a regional model may not represent extreme, local conditions. In addition, a polynomial trend model will tend to perform worse on the margins of the region, where data is often sparse, and it will be overly influenced by regions of extreme variation (Webster and Oliver, 2007). The outliers are addressed further in the discussion.

Both with and without the outlier points, the residuals from the polynomial trend model form a minimally skewed, approximately normal distribution with near zero mean (Figures 6a and 7a respectively). Spatially, they tend to cluster in groups of mostly positive and negative values, but many areas have a mix of both, and there is no regional trend (e.g., all positive values in the south, all negative values in the north) (Figure 6d and 7d). The positive and negative clusters result in the double peak in the histogram

(Figure 7a) but the Q-Q plot (Figure 7b) shows that the distribution of residuals after removal of outliers is very nearly normal.

Sample and Fitted Variograms

The residuals from the regional trend model are the differences of each water-level elevation measurement from the time-invariant trend model prediction at that point. The spatial and temporal correlation of these residuals are investigated by calculating the sample spatiotemporal variogram. The variogram is then modelled with a mathematical function. The variogram model is then used in spatiotemporal kriging to predict water levels at locations and times for which there is no data.
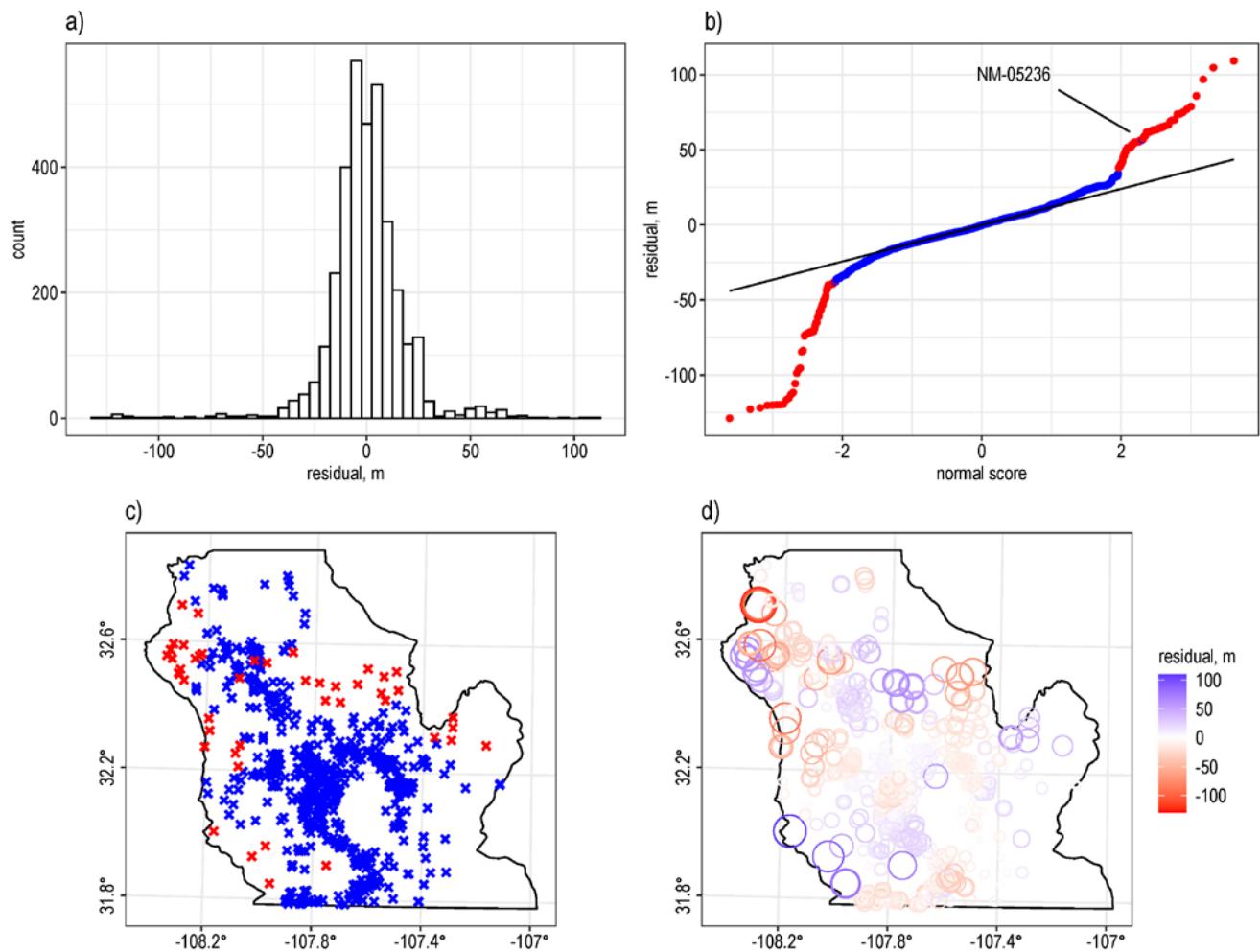
Figure 6. All residuals from the regional trend model, including outliers as described in the text. a) Histogram of all residuals. b) Q-Q plot of all residuals, compared with the normal distribution (black line). Outliers are shown in red. Well NM-05236 is discussed in the text. c) Map showing locations of outliers (red). d) Map showing magnitude (in meters) of all residuals. Trend-surface predictions above the water level are positive (blue) and below the water level are negative (red).
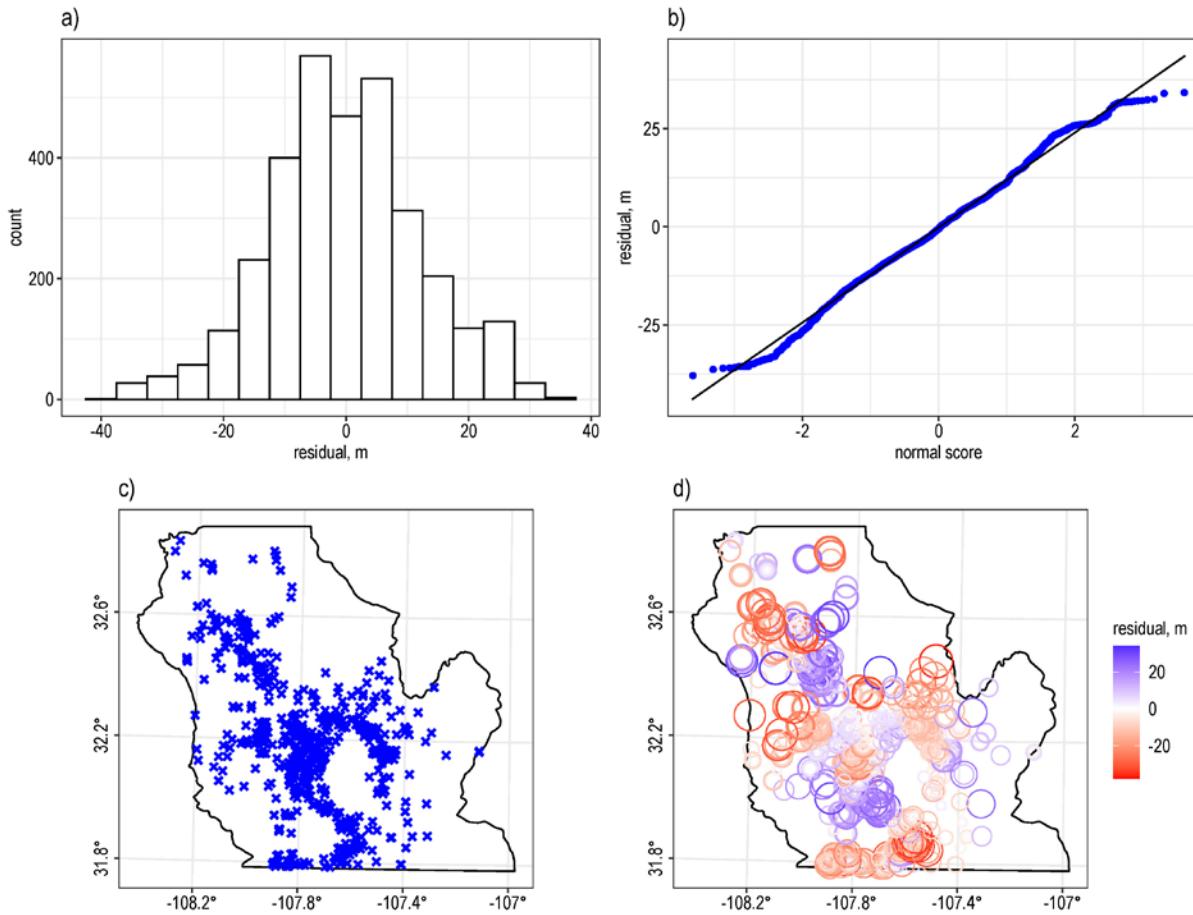
Figure 7. Residuals from the regional trend model, with outliers removed. a) Histogram of residuals. Note the smaller spread than in Figure 6a. b) Q-Q plot of residuals, compared with the normal distribution (black line). c) Map showing wells without outlier measurements. d) Map showing magnitude (in meters) of residuals without outliers. Trend surface predictions above the water level are positive (blue) and below the water level are negative (red). Note the much smaller range compared to Figure 6d.

The empirical or sample spatiotemporal variogram is calculated as:

$$\gamma(h,u) = \frac{1}{2 \cdot N(h,u)} \sum_{i=1}^{N(h,u)} [\varepsilon(s,t) - \varepsilon(s+h, t+u)]^2, \qquad (2)$$

where h is the spatial separation, or lag, u is the time separation, or lag, and $N(u,h)$ is the number of paired residuals separated by spatiotemporal lag (h,u) (Heuvelink and Griffith, 2010; Ruybal et al., 2019). The variogram value, gamma, is thus the average squared difference in residuals at each combination of spatial and temporal lag. Because the data points are not regularly arranged in space or time, the data are grouped into bins of spatial and temporal lags. The result is a three-dimensional variogram surface describing the correlation of the residuals in space and time (Figure 8).

The sample spatiotemporal variogram was fitted with a sum-metric variogram model (Snepvangers et al., 2003; Gräler et al., 2016). The entire suite of available spatiotemporal variograms models available

in the gstat package (Gräler et al., 2016) were tested for the best fit to the sample spatiotemporal variogram. The sum-metric model was chosen based on 1) its superior fit based on the optim and MSE parameters generated by the numerical optimization scheme used by R to fit the model parameters to the sample variogram, and 2) comparison of "leave-one-out" cross-validation statistics and spatial patterns of cross-validation residuals generated for each model in turn (Cressie and Wikle, 2011; see Appendix 2 for details). Cross-validation involves removing each data point in turn and predicting the value at that location (and time in the spatiotemporal case) using the remaining data. It is a measure of the strength of the model and how well it can reproduce the actual data.

The general formula for the sum-metric spatiotemporal variogram model is

$$\gamma_{ST}(s,t) = \gamma_S(h) + \gamma_T(u) + \gamma_J\left(\sqrt{h^2 + (\kappa \cdot u)^2}\right). \qquad (3)$$

The sum-metric model is the sum of separate spatial, temporal, and joint spatio-temporal variogram components. The joint component is a metric variogram, where κ is the space-time anisotropy ratio (space/time) that relates spatial and temporal distances into a single space-time distance (Kyriakidis and Journal, 1999; Snepvangers et al., 2003). κ less than one means that, for example, the correlation at 1 m distance equals the correlation at 5 days time, and vice-versa if the anisotropy is greater than one (Gasch et al., 2015).

The best fitting sum-metric model is shown in Figure 8 and is given by:

$$\gamma_S(h)(\text{Circular}) = \begin{cases} \sigma_S^2\left(1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{a_S}\right) + \frac{2h}{\pi a}\sqrt{1 - \frac{h^2}{a_S^2}}\right) + \sigma_{nS}^2, \ 0 < h \leq a_S \\ \sigma_S^2 + \sigma_{nS}^2, \ h > a_S \\ 0, \ h = 0 \end{cases}, \qquad (4)$$

$$\gamma_T(u)(\text{Nugget}) = \sigma_{nT}^2(1 - \delta(h)), \qquad (5)$$

$$\gamma_J(h,u)(\text{Exponential}) = \begin{cases} \sigma_J^2\left(1 - \exp\left(-\frac{\sqrt{h^2+(\kappa \cdot u)^2}}{a_J}\right)\right) + \sigma_{nJ}^2, \ 0 < \sqrt{h^2 + (\kappa \cdot u)^2} \\ 0, \ \sqrt{h^2 + (\kappa \cdot u)^2} = 0 \end{cases}. \qquad (6)$$

where $\sigma_{S,T}^2$ and $a_{S,T}$ are the variance and range of the spatial and joint components of the sum-metric variogram, $\sigma_{nS,nT,nJ}^2$ is the nugget variance, and $\delta(h)$ is the Kronecker delta, which equals 1 when $h = 0$ and 0 otherwise. The nugget variance is present in all three components, but has different magnitudes for each, and is the sum of spatio-temporal variation in the data at ranges smaller than the smallest lag distance and measurement errors (Cressie and Wikle, 2011).

Circular and exponential (exp) variogram models are two of a limited group of functions used in geostatistics to model variograms because they possess mathematical properties necessary for the kriging equations to be solvable (Webster and Oliver, 2007). Note that the temporal component $\gamma_T(u)$ is a pure nugget variogram; all of the spatiotemporal correlation is accounted for in the joint component.

The structure of the sample and fitted spatio-temporal variograms in Figure 8 suggests why the space-time approach is appealing for prediction of water levels in the Mimbres Basin. In the spatial domain, the variogram rises rapidly to a well-defined sill value ($\sigma_S^2 + \sigma_{nS}^2$) of about 145 m², which is the "global" variance of the residuals of the water level data. This sill is reached at a range of 15,380 m. This

means that pairs of data points (wells) separated by a distance, or lag, larger than this are essentially uncorrelated, and a measurement at one provides no information about the water level at the other. At lags less than 15,380 m, the residuals of the water-level data are progressively more strongly correlated as the lags decrease. In the time domain, the variogram rises very slowly, indicating strong temporal correlation of water levels to the largest temporal lags of nearly 22 years. The fluctuations, or cyclicity, about the sill value of the sample variogram reflect the distribution of positive and negative trend residuals in space. Several patterns visible in the maps of residuals (Figures 6 and 7) show up in the sample variogram (Prycz and Deutsch, 2003).

1. The lag distance of the first peak reflects the average spatial extent of the positive and negative groups, about 15,380 m.
2. The lag distance at the first trough is the sum of the average extent of the groups and the average distance between the groups, about 20,000 m.
3. The distance from the peaks to the troughs is the average spacing of the groups.

In simple kriging the mean is known and constant and is an input to the kriging algorithm. In ordinary kriging the mean of the data is estimated from the data, rather than assumed constant and/or known beforehand. Prycz and Deutsch (2003) note that the type of correlation structure described above violates the assumption of strong global stationarity inherent in simple kriging and that ordinary kriging is more appropriate to use. Ordinary spatiotemporal kriging was used in this study. The penalty for estimating the mean (ordinary kriging) rather than specifying a known mean (simple kriging) is higher kriging variance (Cressie, 2015).

The Mimbres Basin covers over 10,000 km², and is geologically and topographically diverse, with a variety of land uses. These facts also bear on the issue of stationarity: is it appropriate to use a single sample variogram and fitted model, or "global" model, to describe the correlation structure of the water-level data across the entire region? Even with the fluctuations just described, the sample variogram does have a well-defined sill that reaches a plateau value and does not continually increase. Beyond the range value of 15,380 m, the residuals have essentially constant variance upon which the local fluctuations in the residuals are superimposed. (Figure 8 and Appendix 2).

To further explore this issue, the data were subsetted into the San Vicente, Deming, and Florida hydrogeologic regions identified by Finch et al. (2008, Figure 4), with overlap of 5 km along the boundaries.

The Upper Mimbres region was not investigated due to the paucity of data. Sample spatiotemporal variograms were calculated for each region and a best fitting model was chosen using the same optimization criteria described above. Cross-validation statistics and spatial patterns of cross-validation residuals were compared for the wells in each region using the region-specific model, and the "global" model (Appendix 3). In all cases the "global" model performed as well or better than any of the region-specific models, and so only it was used in further analyses.

Spatiotemporal Kriging

Once the spatiotemporal variogram model has been chosen, it is straightforward to implement spatiotemporal kriging at the time periods of interest. An important user-specified parameter in spatiotemporal kriging is nmax, the maximum number of nearest (in the sense of most highly correlated) data points to use in calculating the prediction at any given spacetime point. This value was set to a neighborhood of 50. Increasing this number greatly increases processing time and resulted in poorer, not better predictions. Using a local neighborhood in kriging also relaxes strict stationarity assumptions, and smooth variations in the mean, as are seen here, can be accommodated (Gräler et al, 2016).
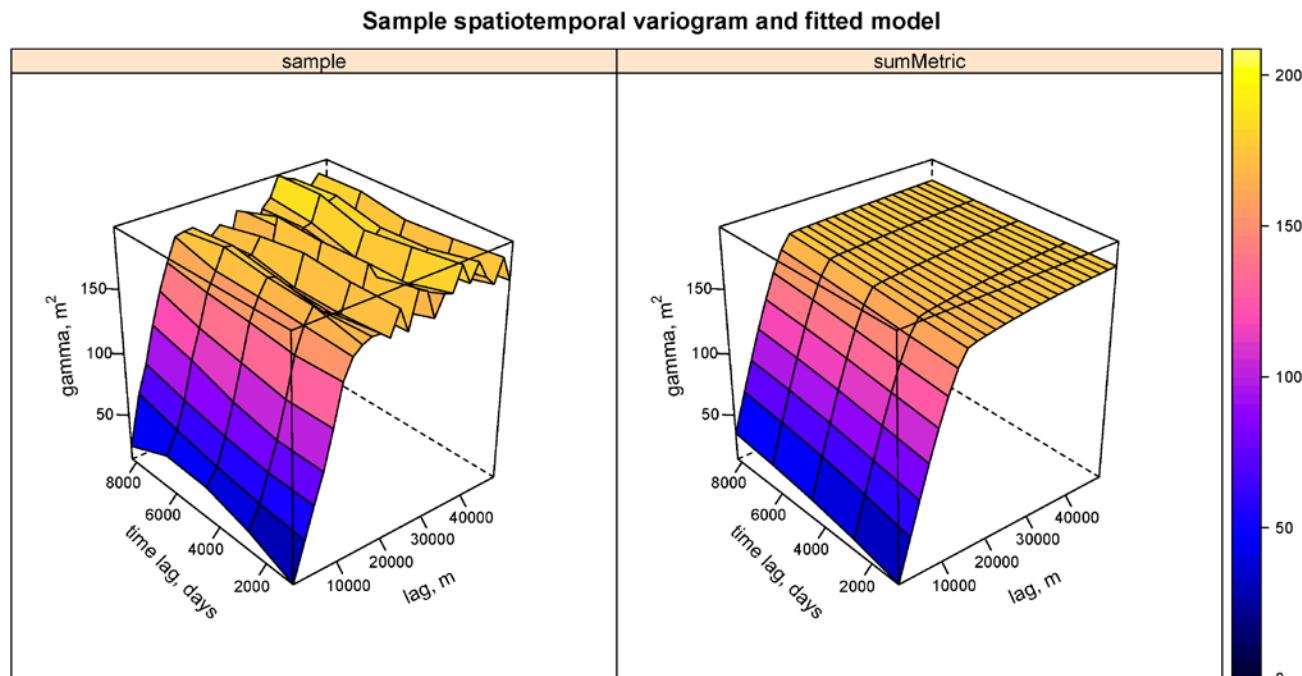

Figure 8. Sample spatiotemporal variogram of regional trend residuals and fitted sum-metric model.

Results were calculated for late winter at five-year intervals from 1980 to 2020. The water level elevation is the sum of the trend function prediction at each point (which is constant through time) and the predicted residual from the spatiotemporal kriging calculation, which is unique to each prediction time, or instance (Figure 9). Flow direction vectors were calculated from the predicted water-level surface for each time instance. (Figure 10). Higher resolution maps were prepared for the Deming – Columbus region, where the majority of the irrigated agriculture is located. (Figure 11). These maps utilize the same data with the same 1 km spatial resolution as those in Figure 9. Kriging, whether spatial or spatiotemporal, provides an estimate of the variance of the predictions, and thus the uncertainty in the results is quantified. Figure 12a shows the standard error (the square root of the variance; this has the same units as the original measurements) of the regional trend prediction, and the standard error from the spatiotemporal kriging for each instance (Figure 12b). The total variance is the sum of the variance of the trend surface prediction and kriging variance (Cressie, 2015; Cressie and Wikle 2011; Hengl et al., 2007, Ruybal et al., 2019). The former is constant and the latter changes for each instance. Both vary in space. The total prediction standard error is the square root of the total prediction variance (Cressie and Wikle, 2011). The changes in water levels over each five-year time period are shown in Figure 13, with the net water-level change over the 1980–2020 period shown in Figure 14. Hydrographs from example wells are shown in Figure 15 with field measurements and predicted values, and the 95% confidence intervals about the predictions. Several of the aforementioned

figures have an overlay of gray polygons to indicate the areas of lithified Cenozoic sedimentary rocks and older bedrock. Very few wells are completed in these regions so the interpolations and predictions there are deemphasized.

Comparison of Spatial and Spatiotemporal Kriging

The improvement in predictive ability gained by the use of spatio-temporal kriging versus spatial kriging can be demonstrated with the prediction standard error and via cross-validation. Figures 16 through 20 compare the results of the two methods for the years 2012 and 2020, the last two years with abundant water-level measurements. Statistics of the cross-validation results for the two methods are presented in Table 2. The predicted water-level elevation maps are very similar in gross appearance at the scale of the whole basin and it is difficult to see differences in the results from the two methods (Figure 16). Differences are more obvious in the maps of water-level change between 2012 and 2020 (Figure 17). Spatial kriging predicts much larger water-level changes, of up to 150 feet. These values are unreasonably large for an eight-year interval, even in areas of intensive irrigation. The largest changes are very localized, creating an irregular yet very smooth map appearance of closed highs and lows, and some changes are poorly constrained by data, such as between Whitewater and Faywood and southeast of Akela. The different well networks for the two time-periods may be contributing to this effect. Ruybal et al. (2019) showed that changing well networks though time is a major source of uncertainty when comparing spatial kriging maps made at

Table 2. Cross-validation statistics for comparison of spatial and spatiotemporal kriging. Better values are in bold.

	spatial 2012	spatiotemporal 2012	spatial 2020	spatiotemporal 2020
Mean	**-0.059**	0.45	0.37	-0.019
Mean squared error	110.92	**18.49**	119.84	**40.28**
Mean squared normalized error	1.16	**0.99**	1.57	**0.99**
Correlation between observed and predicted	0.55	**0.94**	0.61	**0.88**
Correlation between predicted and residual	-0.034	**-0.011**	-0.13	**0.021**
Variance of residuals	111.86	**18.44**	120.99	**40.71**
Standard deviation of residuals	10.58	**4.29**	10.99	**6.38**
Skewness	0.51	**-0.47**	0.79	**-3.06**
Minimum of residuals	-30.16	**-21.81**	-24.73	-40.10
Maximum of residuals	42.12	15.40	47.09	**15.26**

different times. The water-level changes predicted by spatiotemporal kriging are more reflective of natural variability, of much lower magnitude, and the areas of largest change coincide with wells, which are the locations with the highest reliability of prediction (lowest standard error, Figures 12 and 18).

The prediction standard error for the spatial-only kriging case rises rapidly away from the wells, which form "bulls-eyes" of accuracy, to the maximum value of about 40 to 50 feet, while overall the uncertainty is large (Figure 18a and c). The prediction standard error for the spatiotemporal kriging case shows large regions of low values (high precision), 20 to 30 feet or less (Figure 18b and d). This occurs in many areas where there were no measurements in 2012 and 2020, showing the beneficial influence of measurements at other times on the predictions. The high correlation of water-levels in time improves the precision at times when there are no data – overall the areas of relatively low standard error are much larger when using spatiotemporal kriging.

The standard error contributed by the regional trend model is not shown on the maps in Figure 18. It is the same for both the spatial and spatiotemporal kriging cases and is added to the standard error from kriging to result in the total standard error of prediction (see Figures 12a and 15). Recall that the trend surface was calculated using all of the data from 1980 – 2020, and was assumed to be constant in time. However, one could calculate a trend surface for 2012 and 2020 each, using only the data available for those two years. The fewer available data would result in more uncertainty in the trend surface, and a larger total variance of the prediction for either year. The use of data through time improves both aspects of the water level prediction, the trend surface and the kriging results.

Cross-validation histograms and scatterplots show the clear improvement of spatiotemporal kriging over spatial kriging alone (Figures 19 and 20). The latter shows significant bias, under-predicting the high values and over predicting the low values (Figures 19a and 20a). The spatiotemporal cross-validation results more closely follow the 1-to-1 line and show less spread about that line (Figures 19c and 20c). Similar results were observed by Ruybal et al. (2019). Histograms of the cross-validation residuals show a narrower peak and less overall spread (Figures 19b and d, Figures 20b and d). The cross-validation statistics also confirm the improvement in prediction using spatiotemporal methods (Table 2).
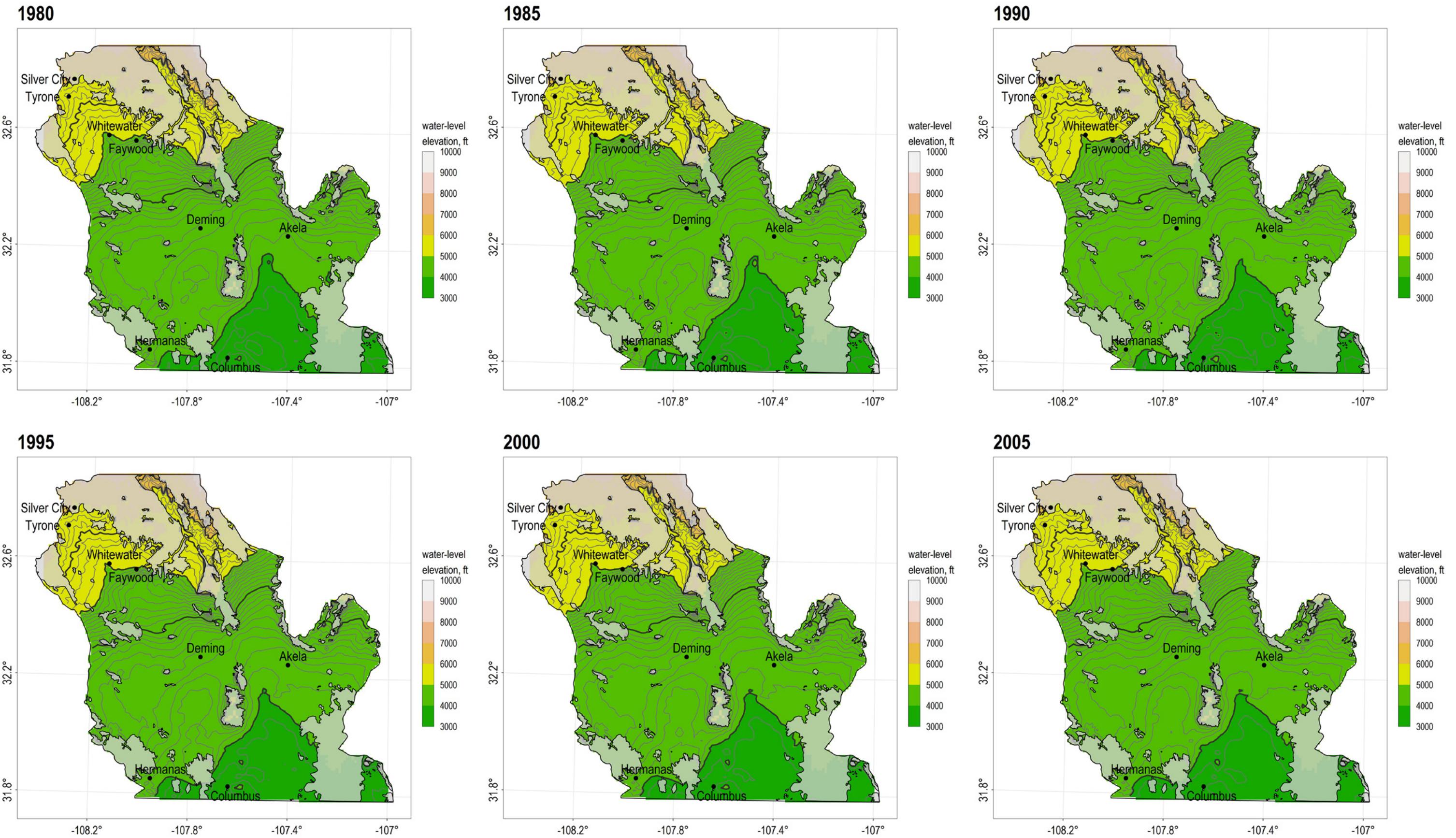
Figure 9. Maps of water-level elevation for nine time instances from 1980 to 2020. Contour interval is 50 ft from 3,800 ft to 4,800 ft, and 200 ft from 5,000 to 9,000 ft. Gray shaded areas are lithified Cenozoic sedimentary rocks and older sedimentary, igneous, and metamorphic rocks, undivided.
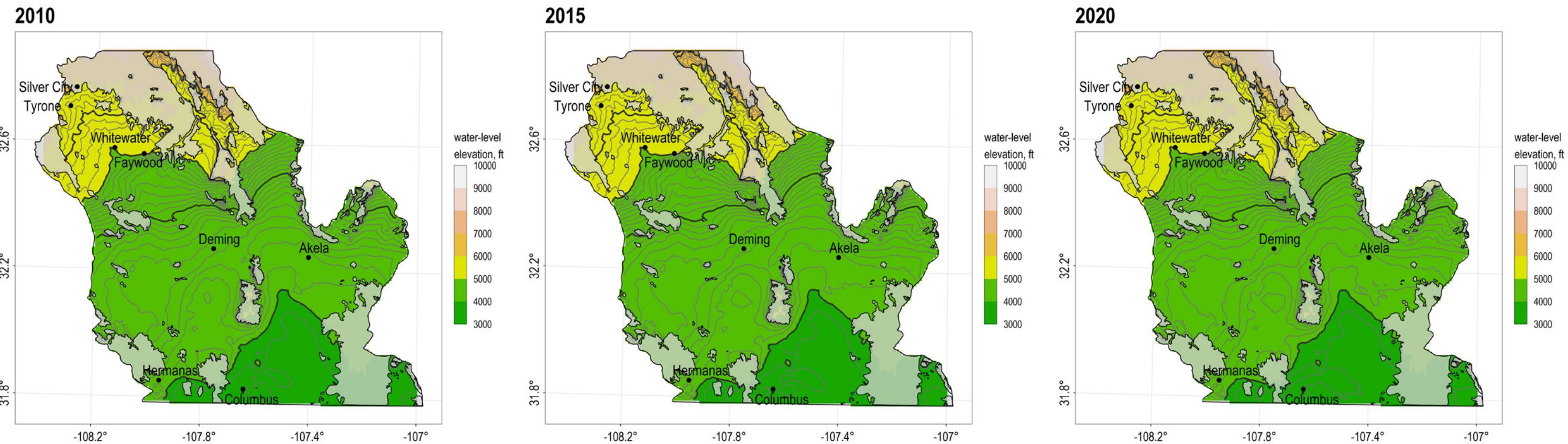
Figure 9, continued. Maps of water-level elevation for nine time instances from 1980 to 2020. Contour interval is 50 ft from 3,800 ft to 4,800 ft, and 200 ft from 5,000 to 9,000 ft. Gray shaded areas are lithified Cenozoic sedimentary rocks and older sedimentary, igneous, and metamorphic rocks, undivided.
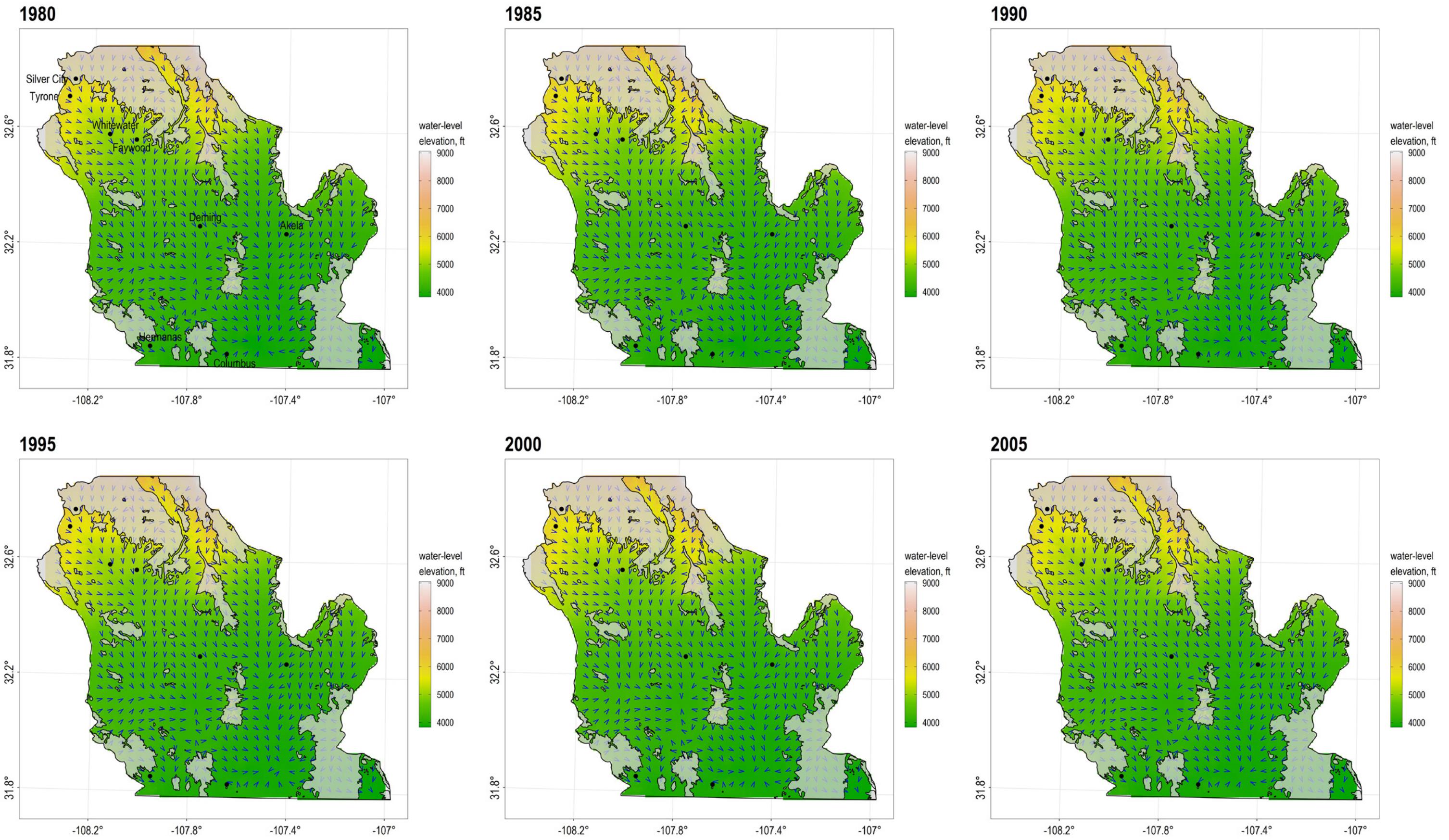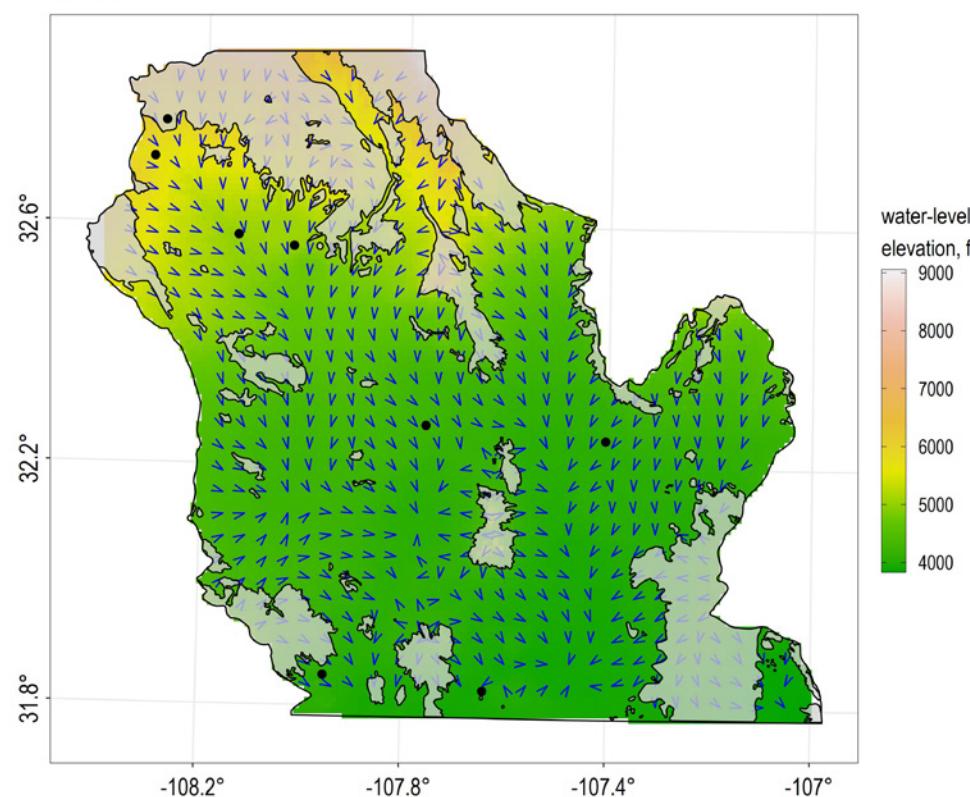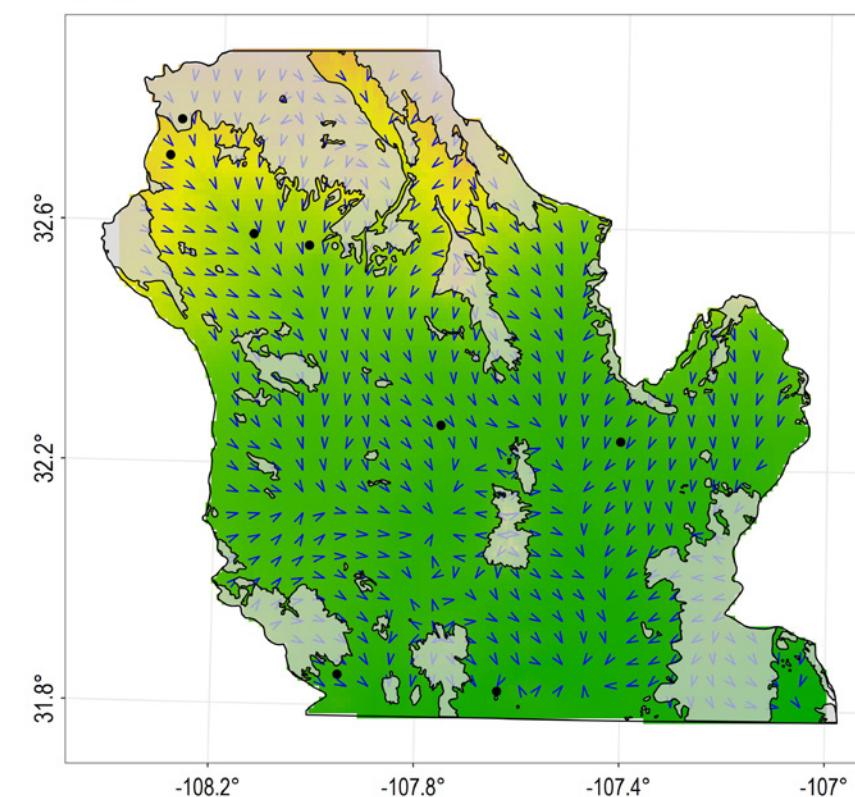
Figure 10. Maps of water-level elevation for nine time instances from 1980 to 2020 with flow direction vectors shown in blue. Gray shaded areas are lithified Cenozoic sedimentary rocks and older sedimentary, igneous, and metamorphic rocks , undivided. Labels of towns are only shown in the first map for clarity
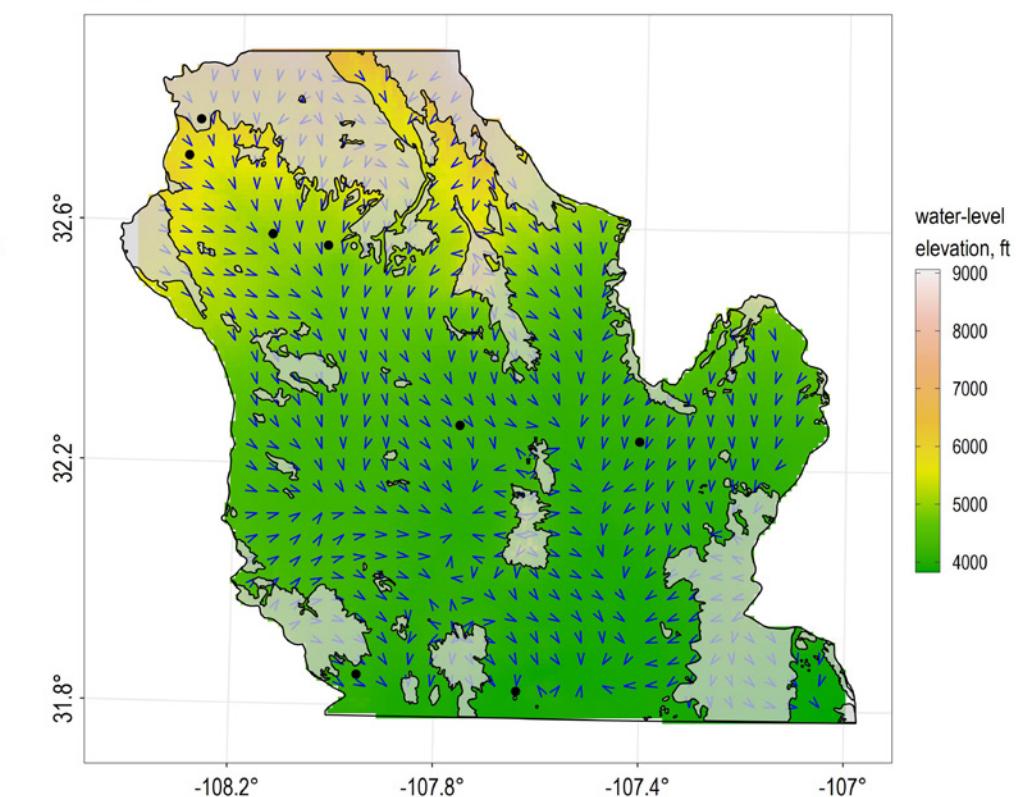
Figure 10, continued. Maps of water-level elevation for nine time instances from 1980 to 2020 with flow direction vectors shown in blue. Gray shaded areas are lithified Cenozoic sedimentary rocks and older sedimentary, igneous, and metamorphic rocks , undivided. Labels of towns are only shown in the first map for clarity
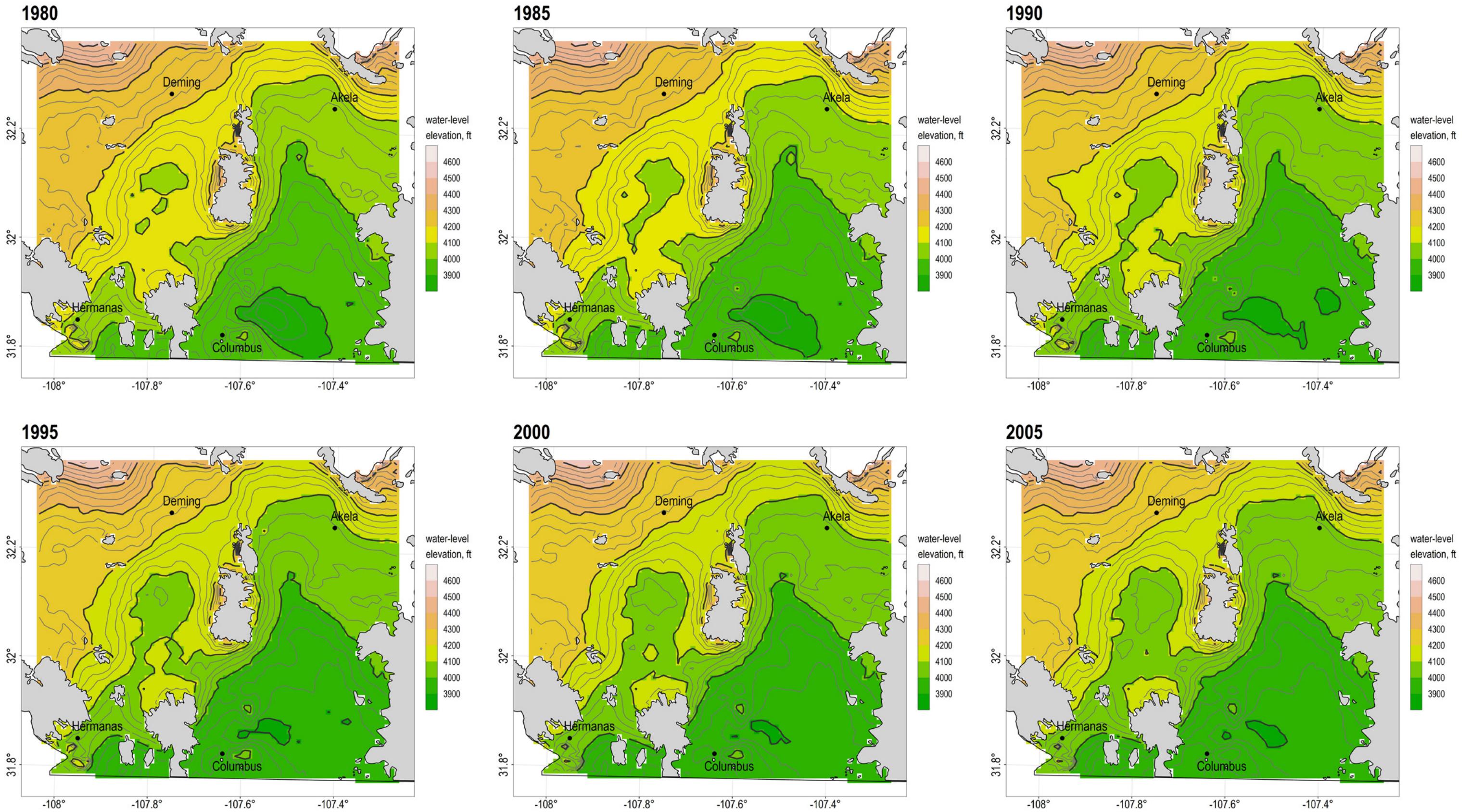
Figure 11. Maps of water-level elevation for nine time instances from 1980 to 2020 focused on the Deming–Columbus region. Major contours are 100 foot intervals. Minor contours are 25 foot intervals. Gray shaded areas are lithified Cenozoic sedimentary rocks and older sedimentary, igneous, and metamorphic rocks, undivided.
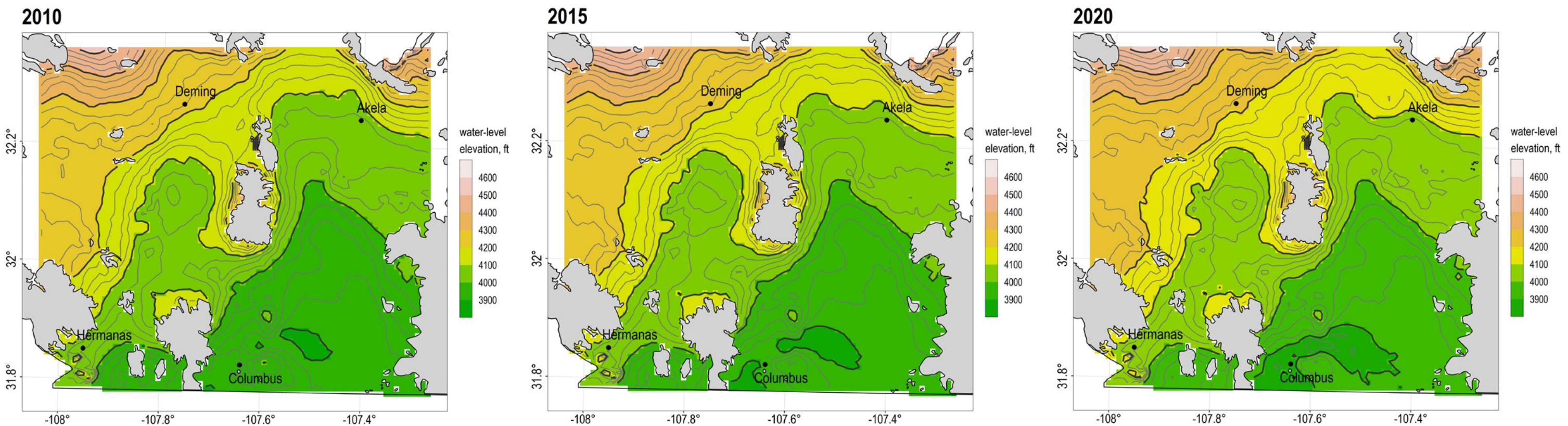
Figure 11, continued. Maps of water–level elevation for nine time instances from 1980 to 2020 focused on the Deming–Columbus region. Major contours are 100 foot intervals. Minor contours are 25 foot intervals. Gray shaded areas are lithified Cenozoic sedimentary rocks and older sedimentary, igneous, and metamorphic rocks , undivided.
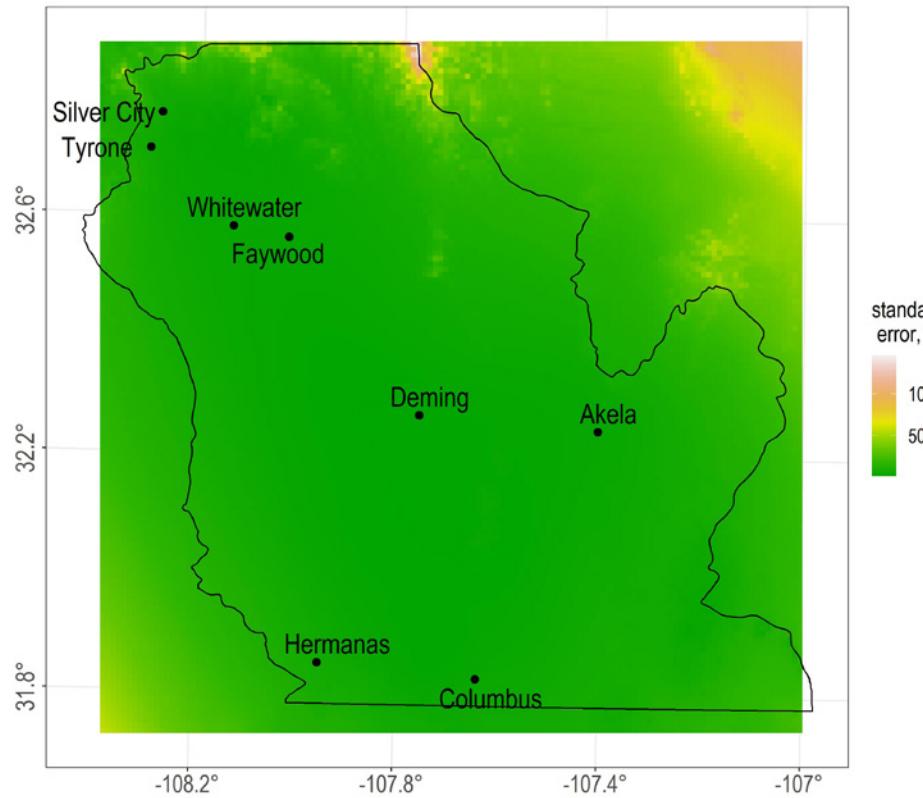
Figure 12a. Map of the standard error from the regional trend model. This is constant for each time instance of prediction.
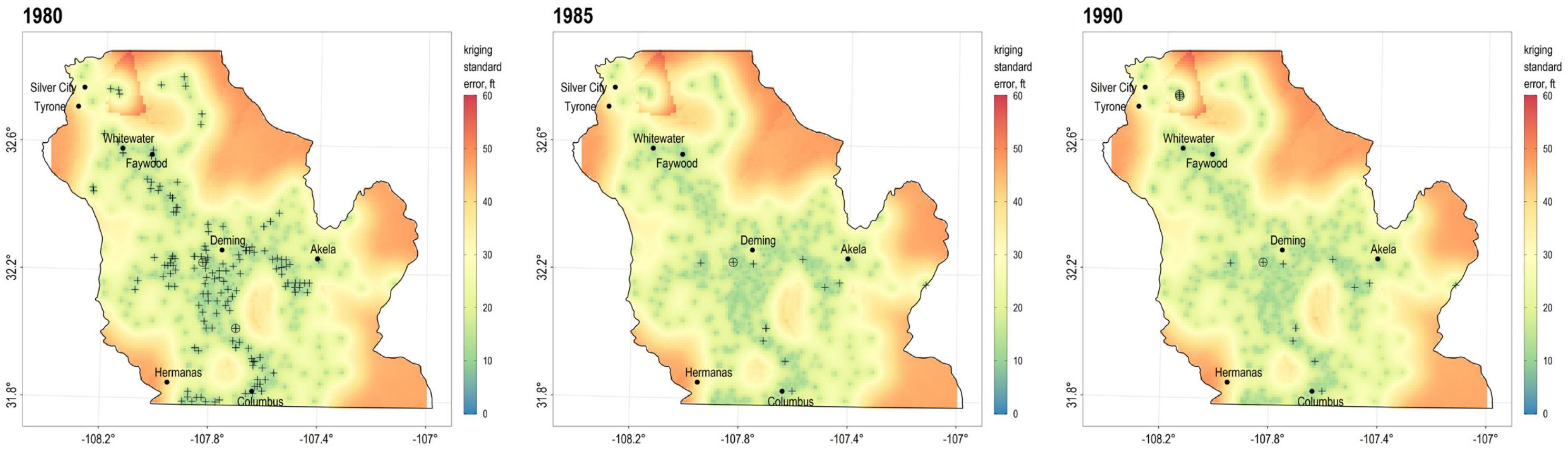
Figure 12b. Maps of the spatiotemporal kriging standard error, which varies for each time instance of prediction. Wells with measurement in the year of each map are shown as crosses. Wells with more than one measurement are shown as a cross in a circle. The same color scale applies to each map.
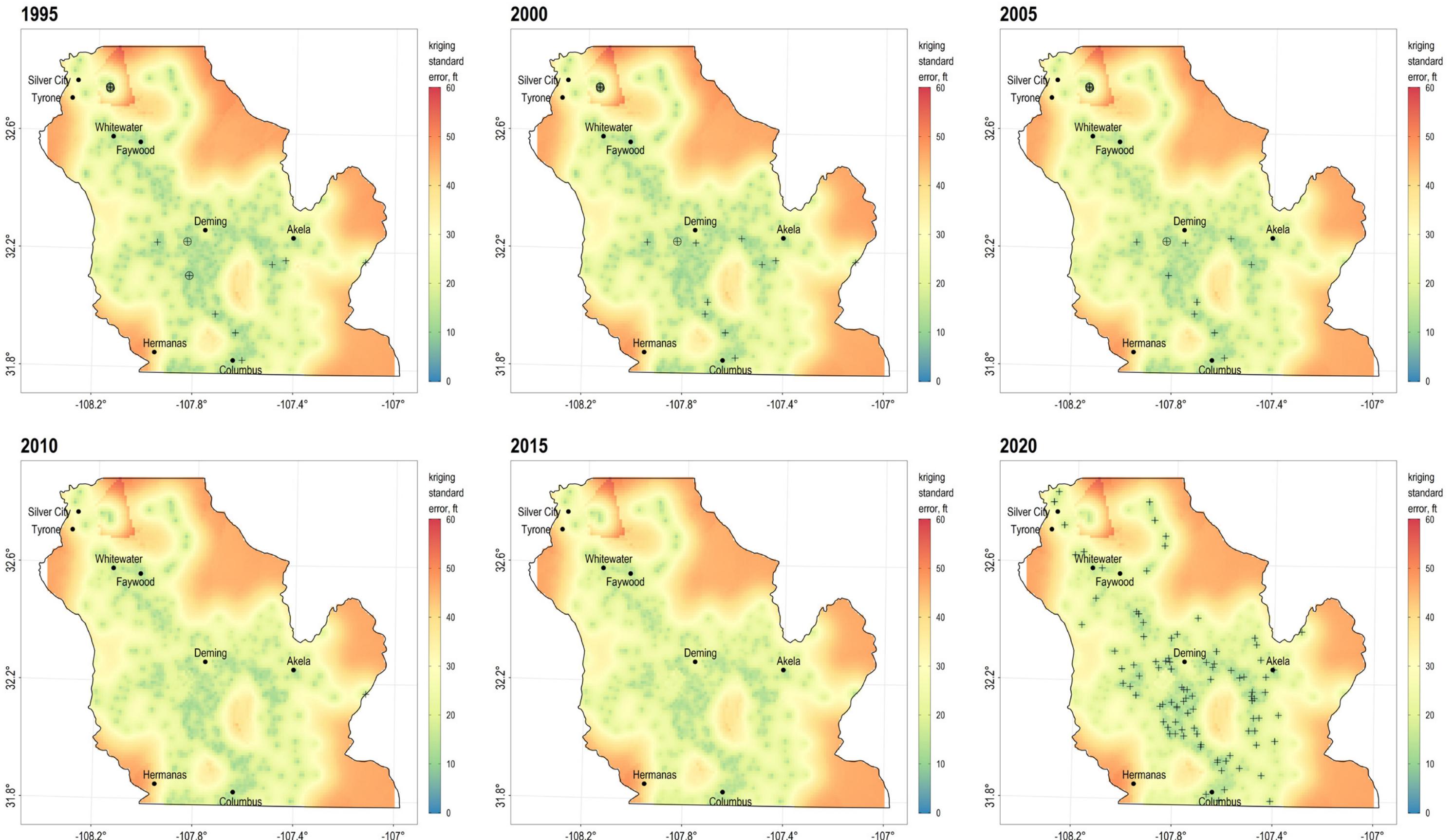
**Figure 12b, continued.** Maps of the spatiotemporal kriging standard error, which varies for each time instance of prediction. Wells with measurement in the year of each map are shown as crosses. Wells with more than one measurement are shown as a cross in a circle. The same color scale applies to each map.
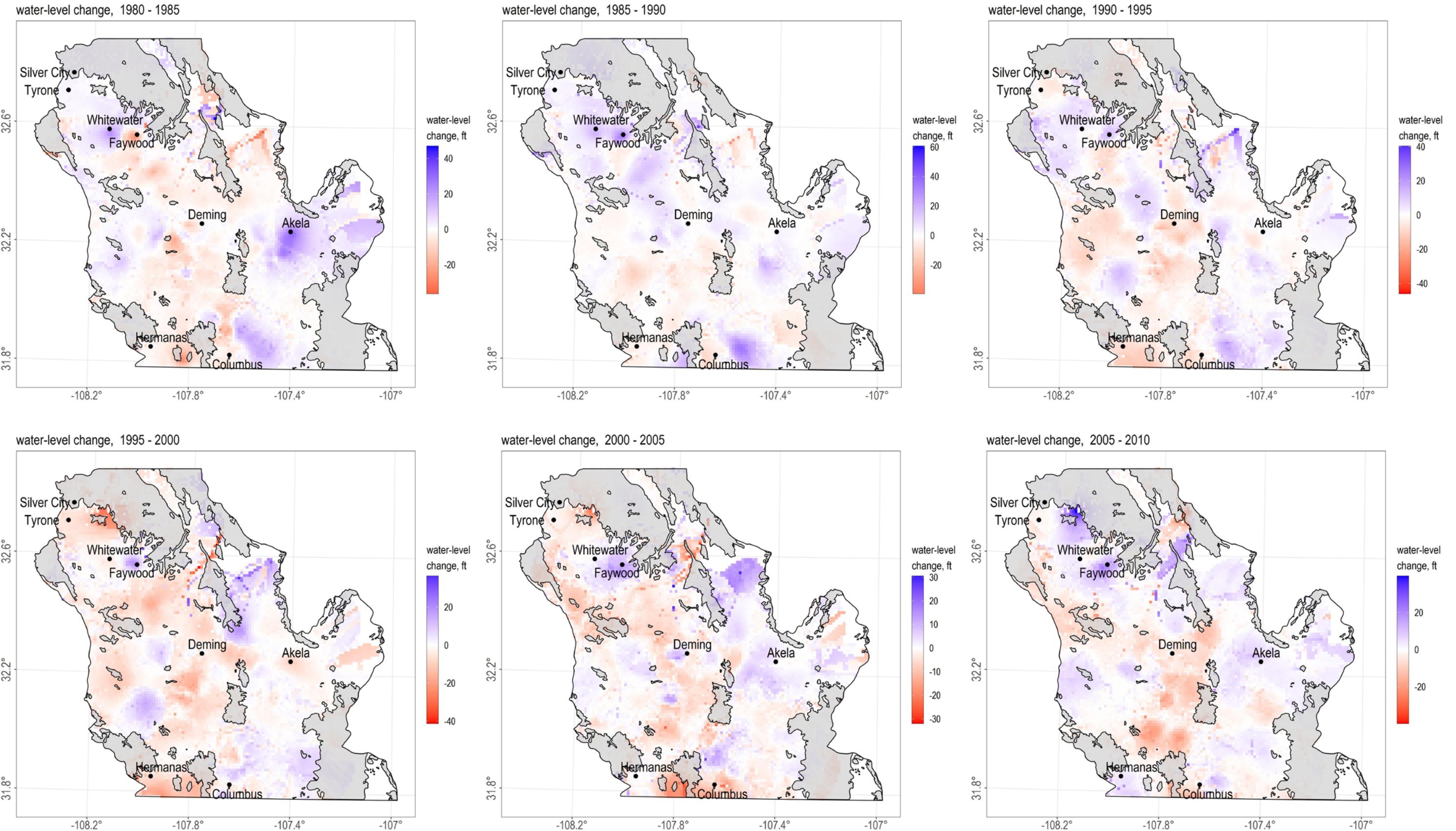
Figure 13. Change in water levels over five-year time periods. Note different color scale on each map. Gray shaded areas are lithified Cenozoic sedimentary rocks and older sedimentary, igneous, and metamorphic rocks, undivided.
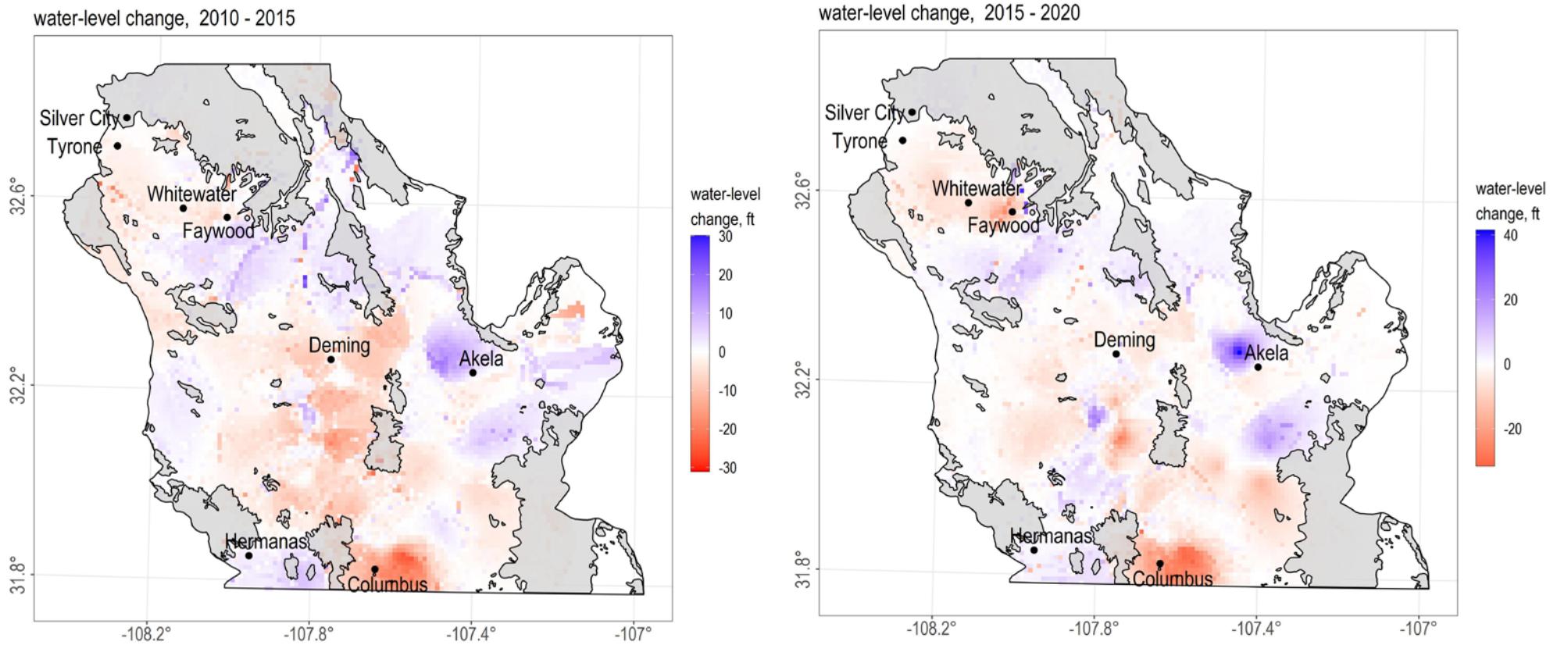
**Figure 13, continued.** Change in water levels over five-year time periods. Note different color scale on each map. Gray shaded areas are lithified Cenozoic sedimentary rocks and older sedimentary, igneous, and meta-morphic rocks, undivided.
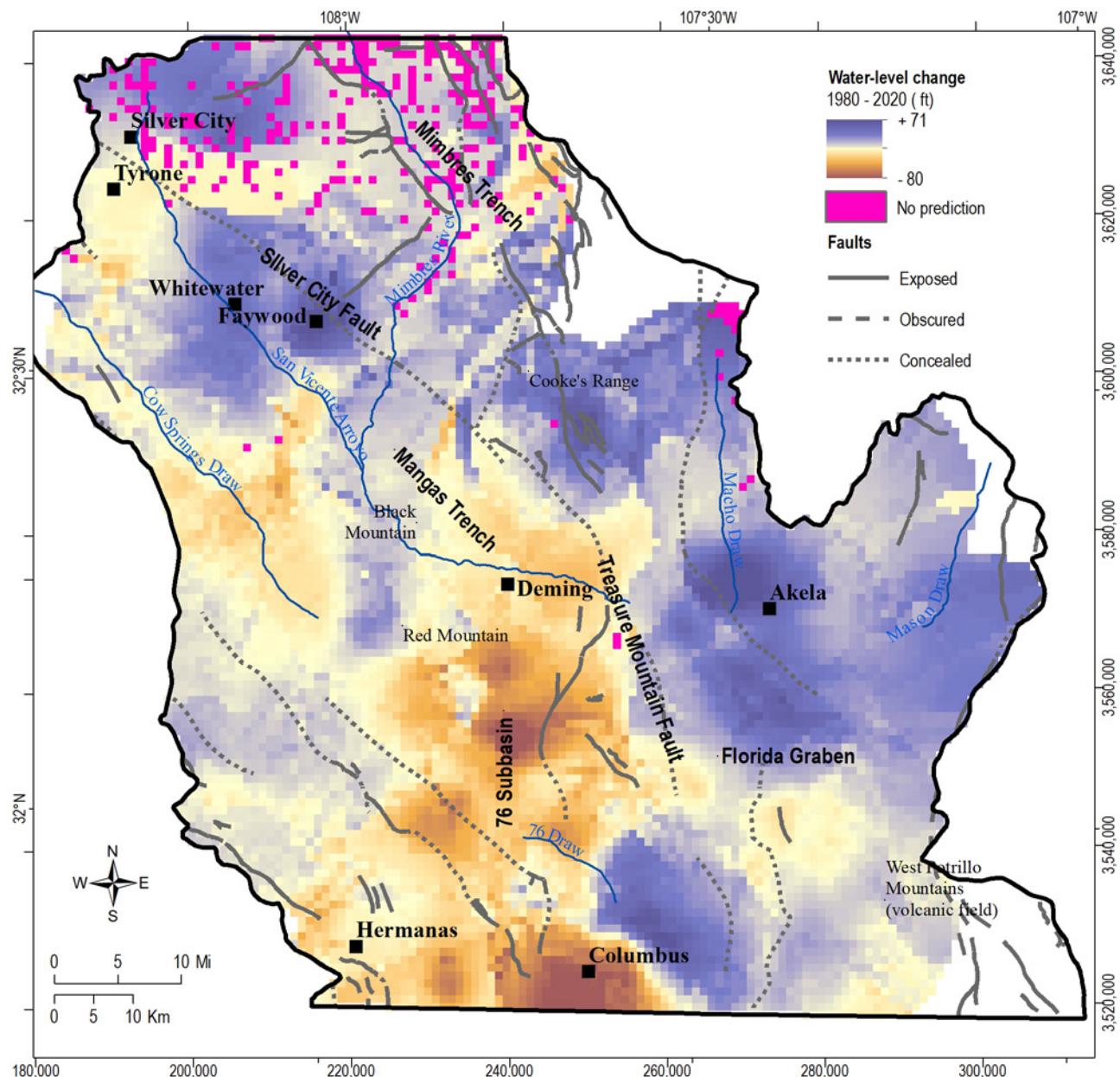
Figure 14. Net water-level change over the time period 1980–2020. Pink areas are where the regional trend model is not valid, resulting in predicted water-levels above ground level.
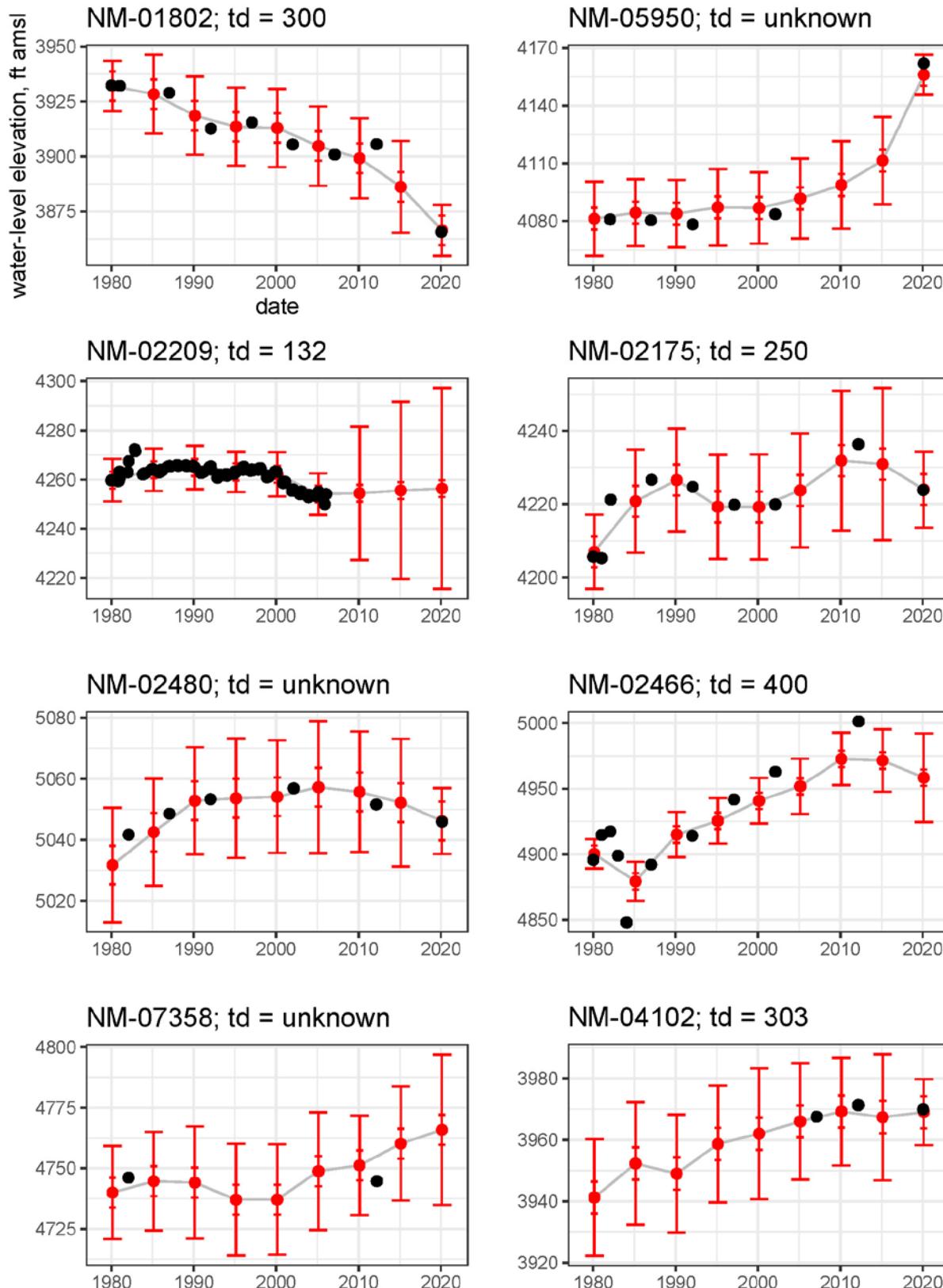
Figure 15. Representative hydrographs with measured (black) and predicted (red) water table elevations. "td" is total depth of well in feet. Error bars indicate 95% confidence intervals from regional trend prediction (small bar) and regional trend prediction plus spatiotemporal kriging interpolation (large bar). Note the different vertical scales on the hydrographs. Locations shown on map.
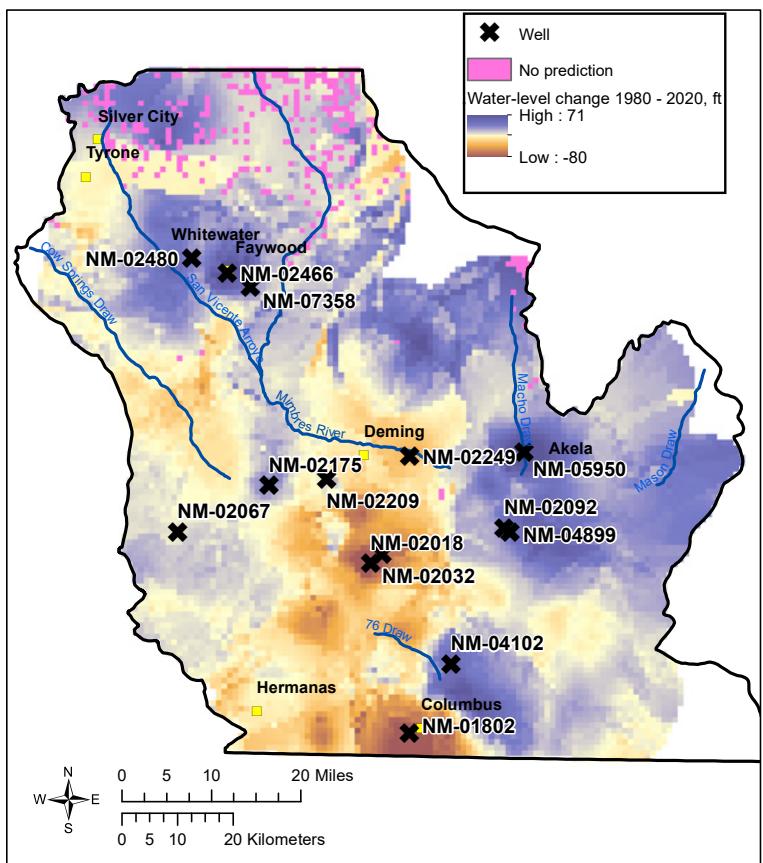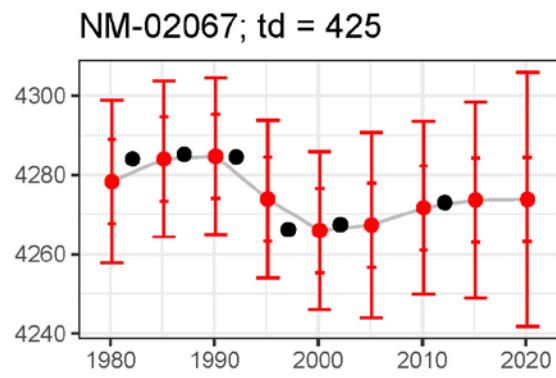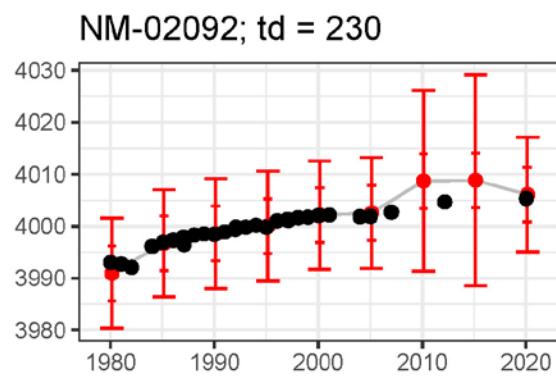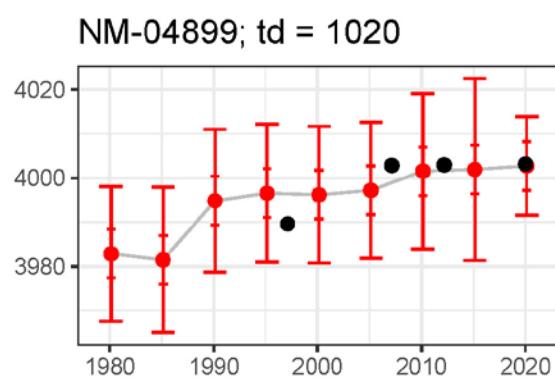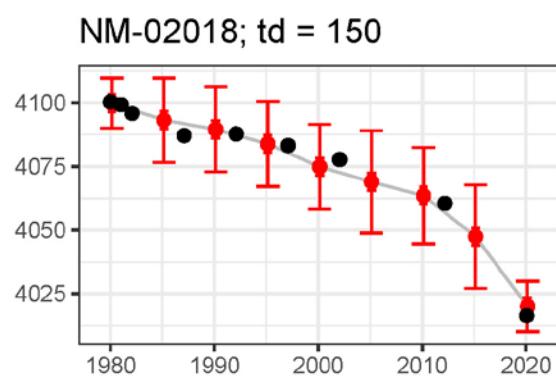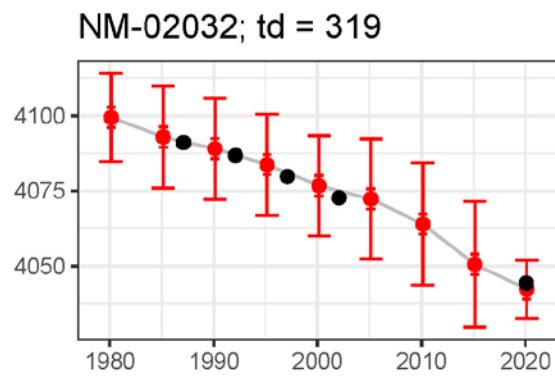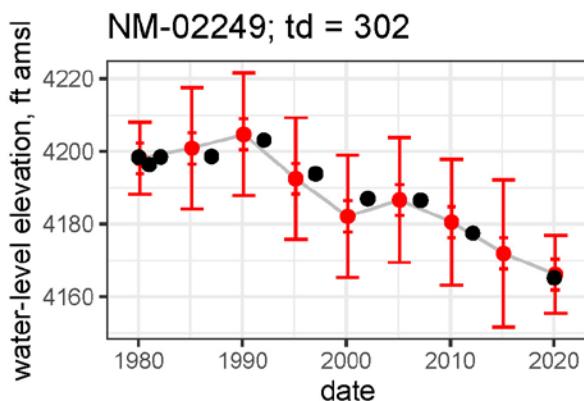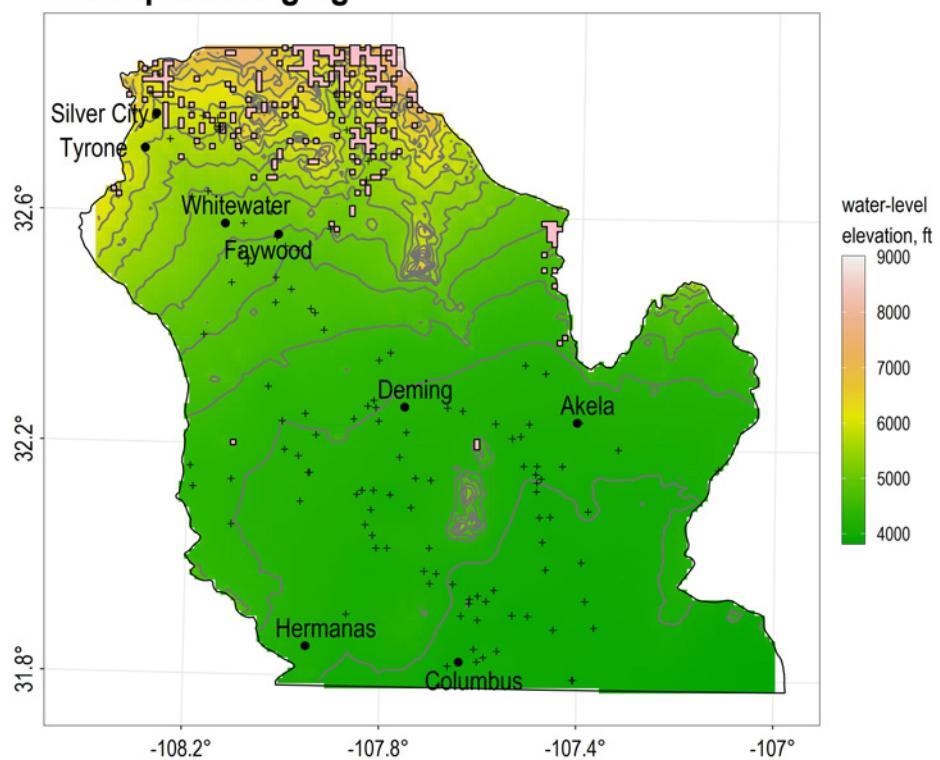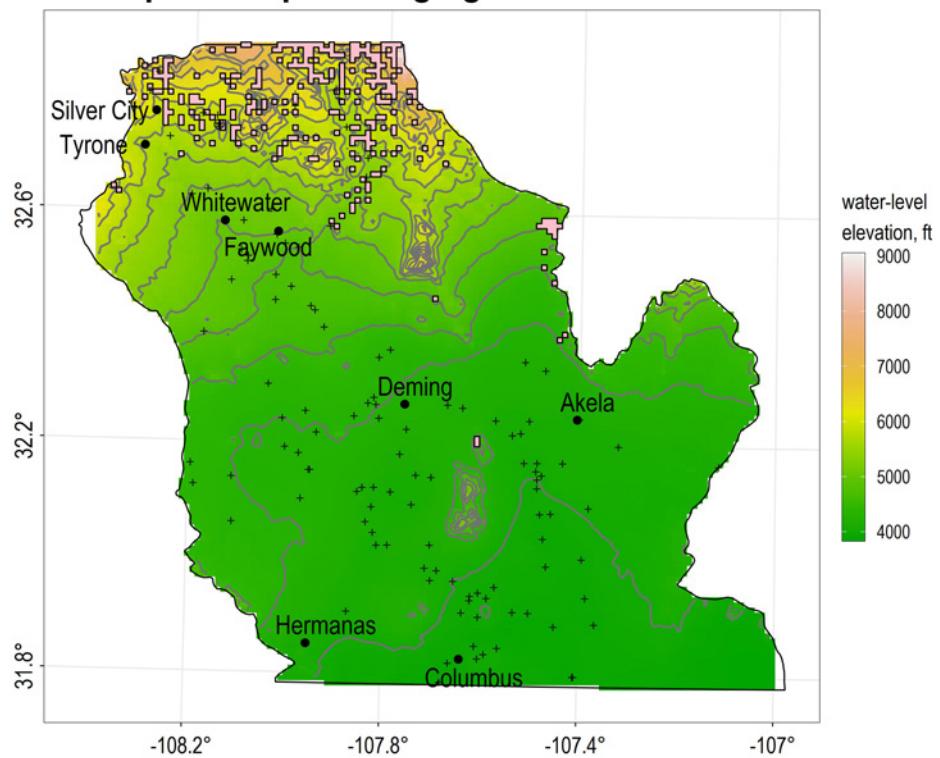
Figure 15, continued.

2012 spatial kriging

a) Water-level elevation predicted for 2012 using spatial kriging.
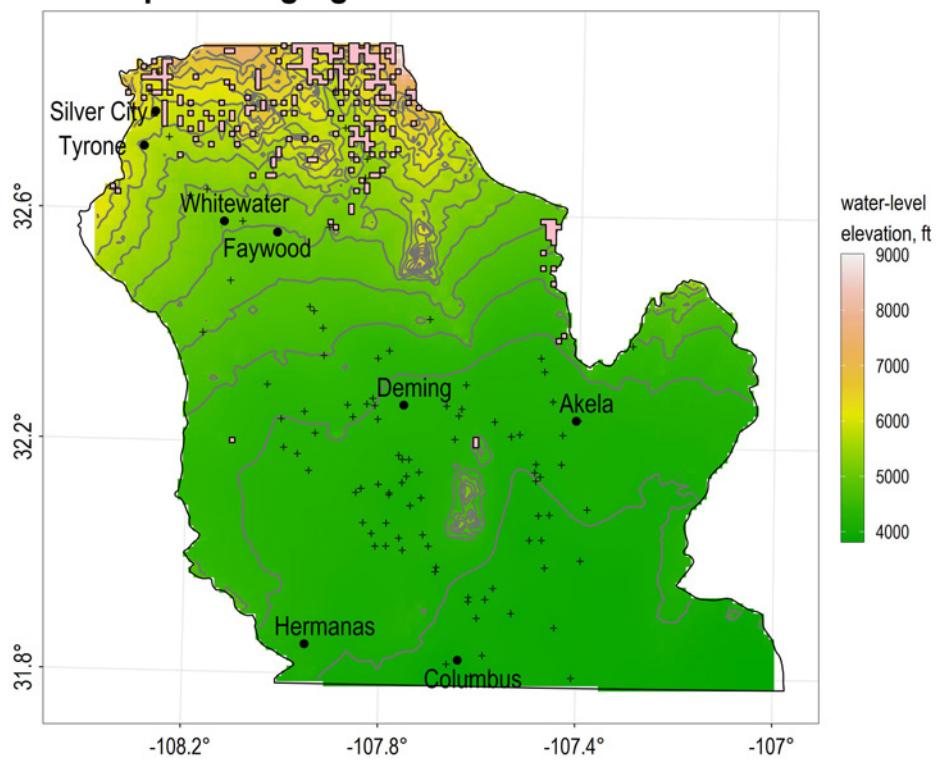
2012 spatiotemporal kriging

b) Water-level elevation predicted for 2012 using spatiotemporal kriging.
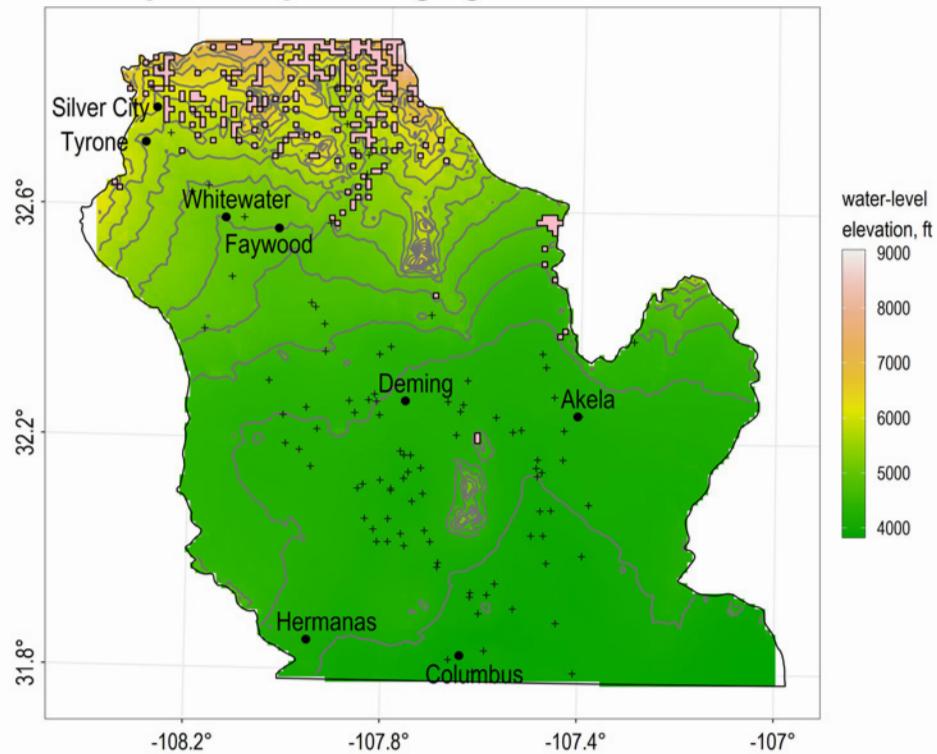
Figure 16 a-b. Pink areas in all four maps are where the regional trend model is not valid, resulting in predicted water-levels above ground level. Contour interval is 250 ft, starting at 4000 ft.

2020 spatial kriging

c) Water-level elevation predicted for 2020 using spatial kriging.
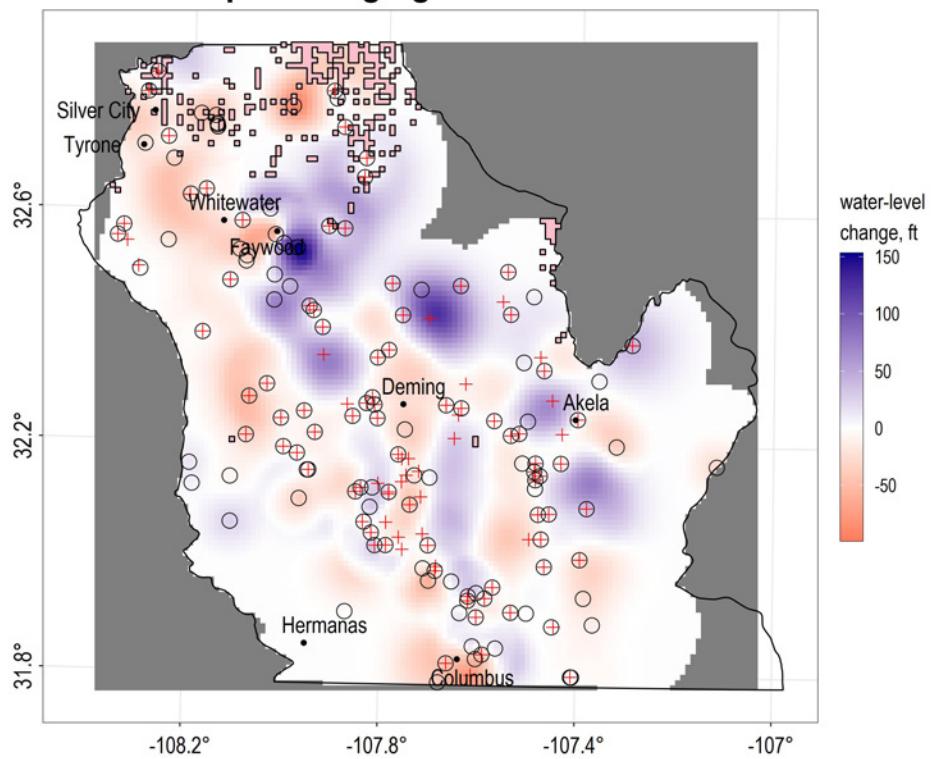
2020 spatiotemporal kriging

d) Water-level elevation predicted for 2020 using spatiotemporal kriging.
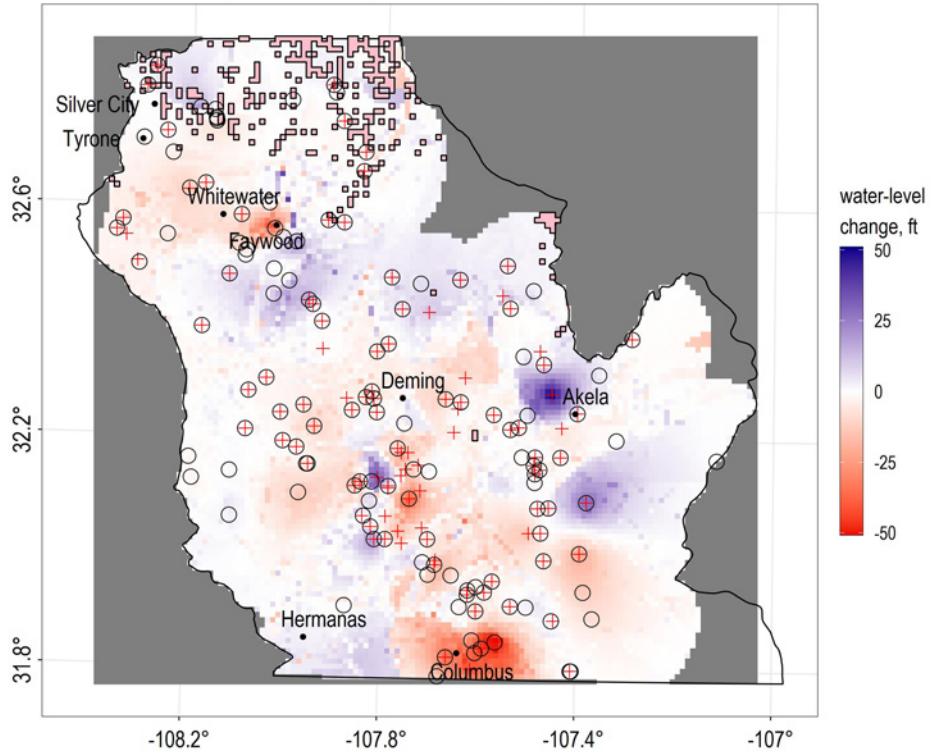
Figure 16 c-d. Pink areas in all four maps are where the regional trend model is not valid, resulting in predicted water-levels above ground level. Contour interval is 250 ft, starting at 4000 ft.

2012–2020 spatial kriging

a) Predicted water-level change from 2012 to 2020 using spatial kriging.
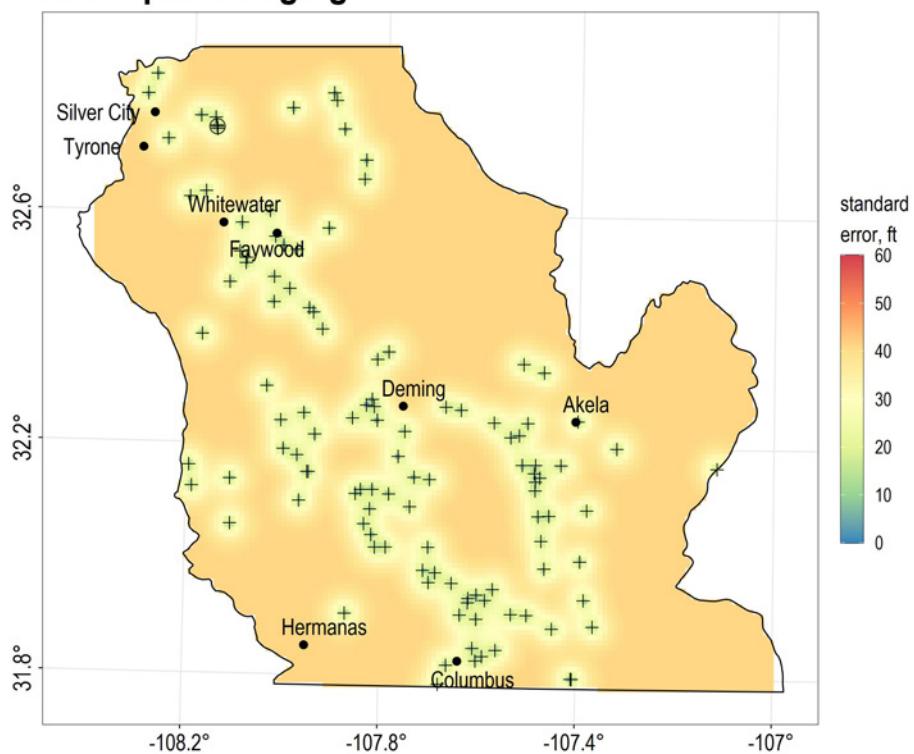
2012–2020 spatiotemporal kriging

b) Predicted water-level change from 2012 to 2020 using spatiotemporal kriging.
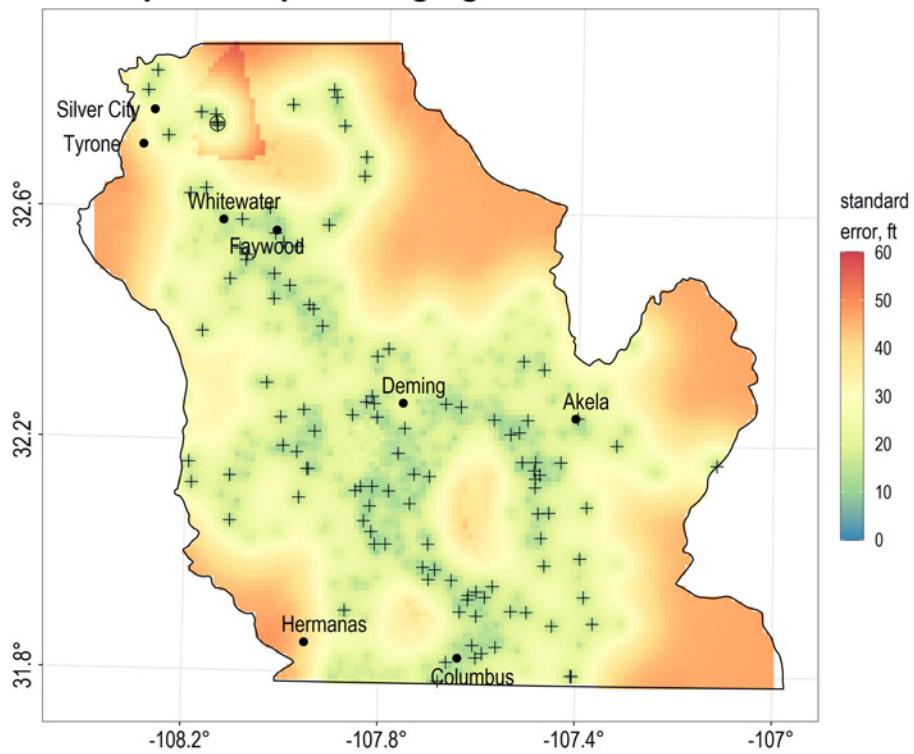
Figure 17 a–b. Note the different scales on the two figures. Pink areas are where the regional trend model is not valid, resulting in predicted water-levels above ground level. 2012 measurements are shown as blcack circles; 2020 measurements are shown as red crosses.

2012 spatial kriging

a) Standard error of 2012 spatial kriging prediction.
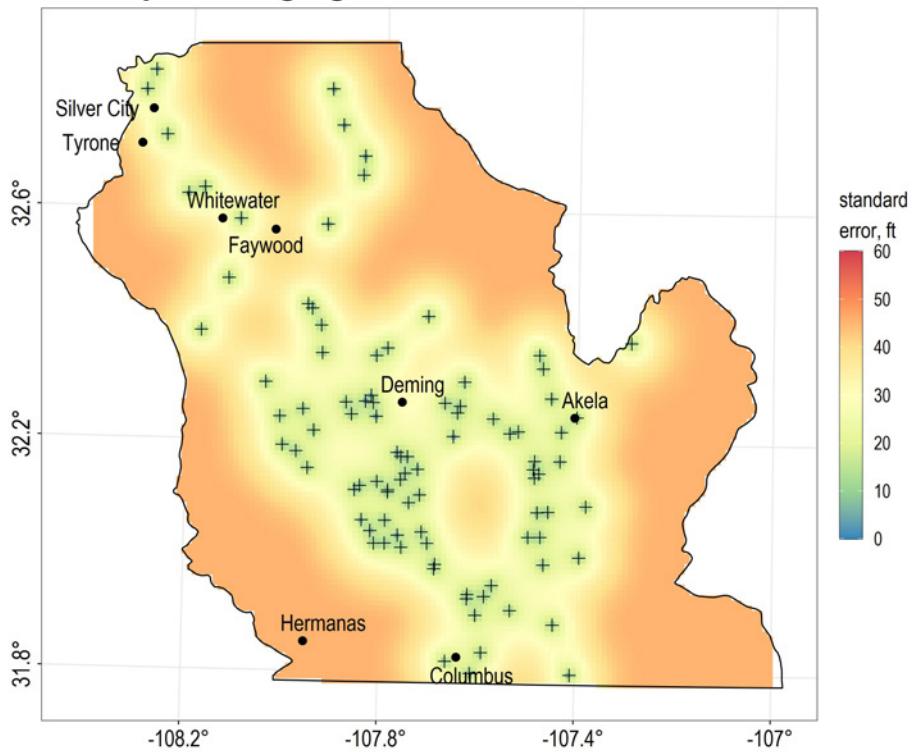
2012 spatiotemporal kriging

b) Standard error of 2020 spatial kriging prediction.
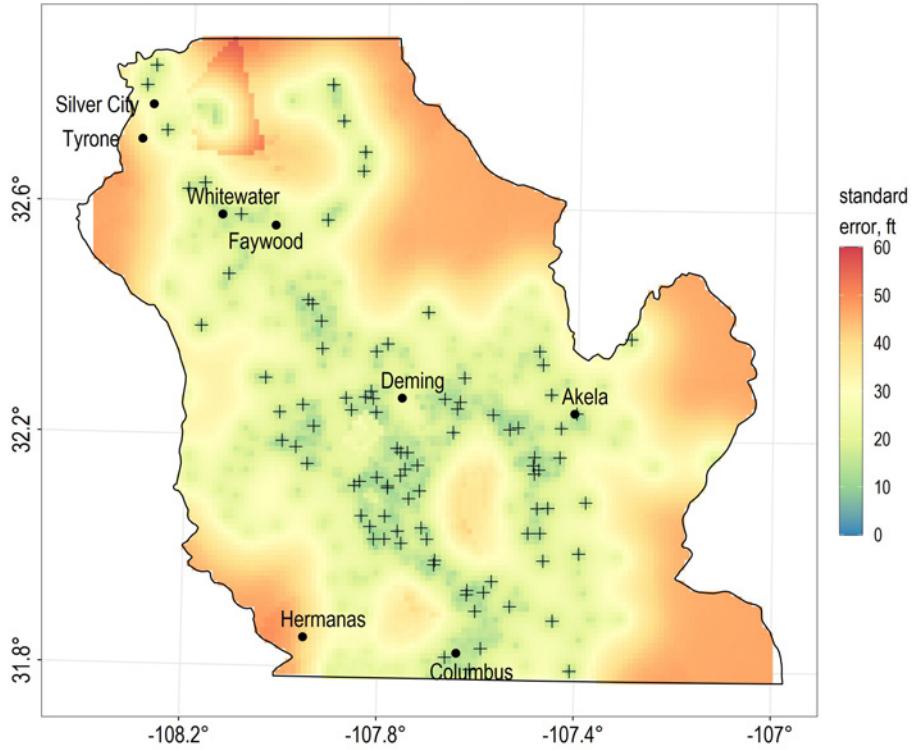
Figure 18 a–b. The same scale is used for the four maps. Wells for the time instance of prediction are shown as crosses; wells with more than one measurement are shown as crosses in a circle.

2020 spatial kriging

c) Standard error of 2012 spatiotemporal kriging prediction.

2020 spatiotemporal kriging

d) Standard error of 2020 spatiotemporal kriging prediction.

Figure 18 c–d. The same scale is used for the four maps. Wells for the time instance of prediction are shown as crosses; wells with more than one measurement are shown as crosses in a circle.
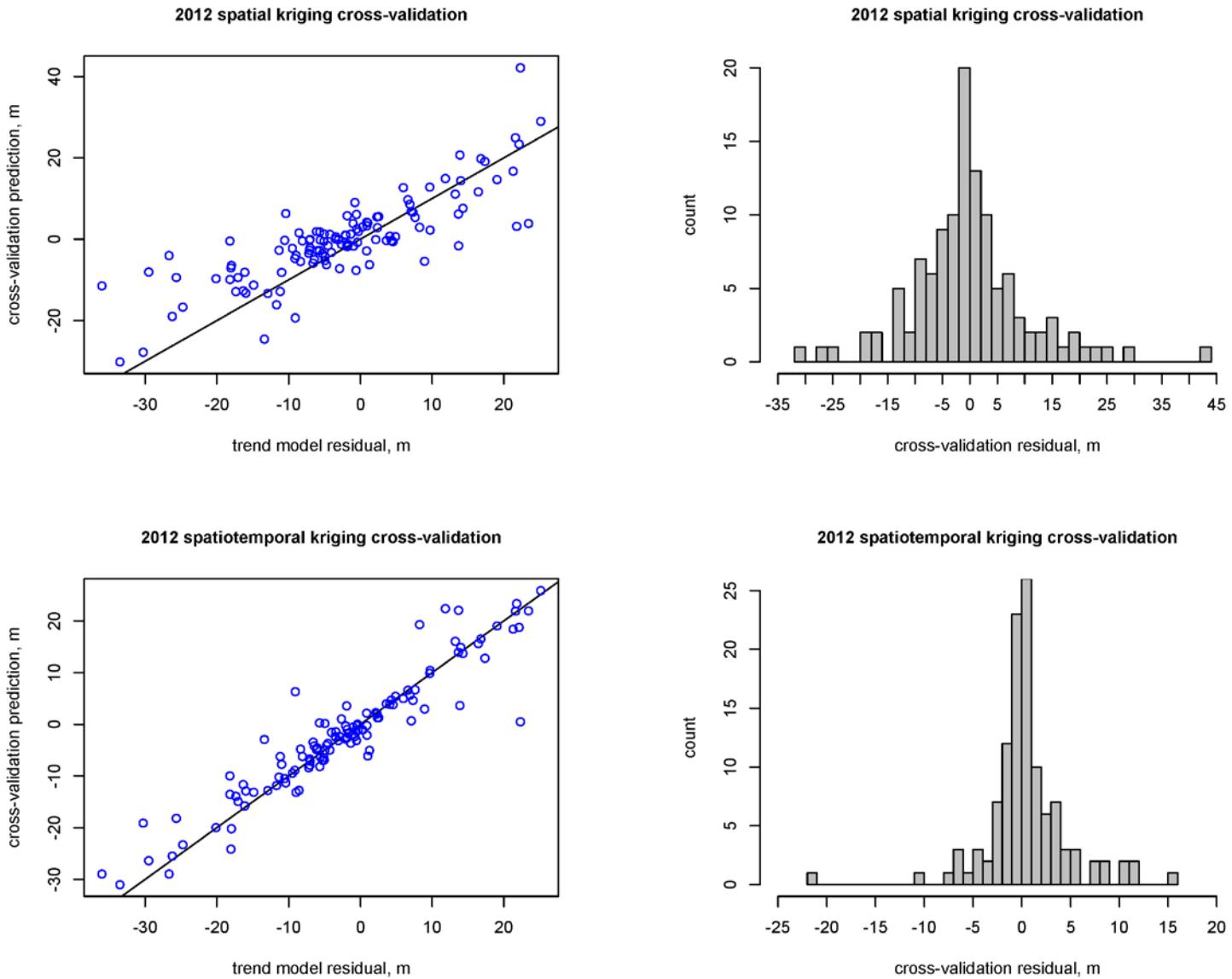
Figure 19. Scatterplots and histograms comparing cross-validation results for spatial and spatiotemporal kriging for 2012. 1-to-1 line shown in black on scatter plots.
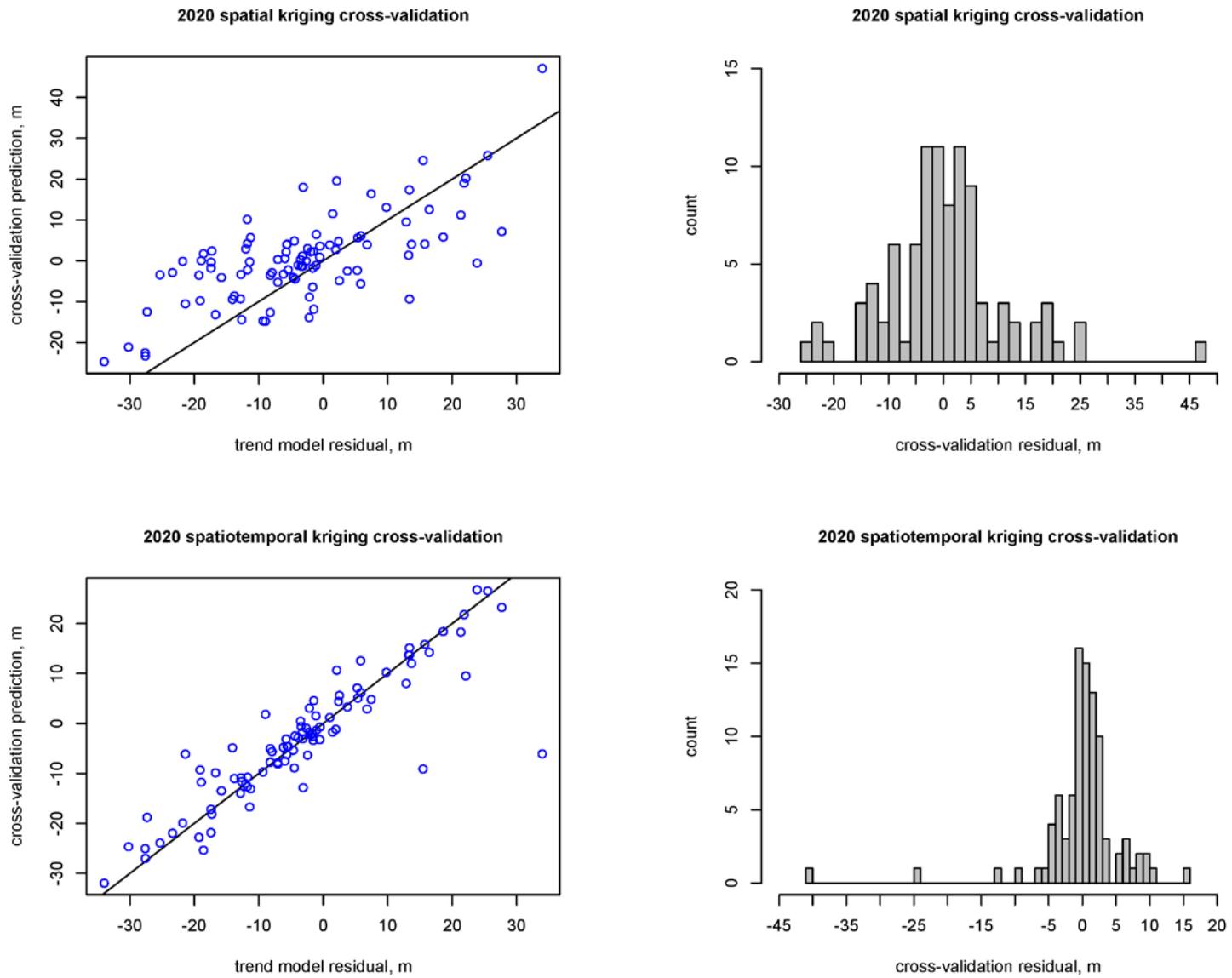
Figure 20. Scatterplots and histograms comparing cross-validation results for spatial and spatiotemporal kriging for 2020. 1-to-1 line shown in black on scatter plots.
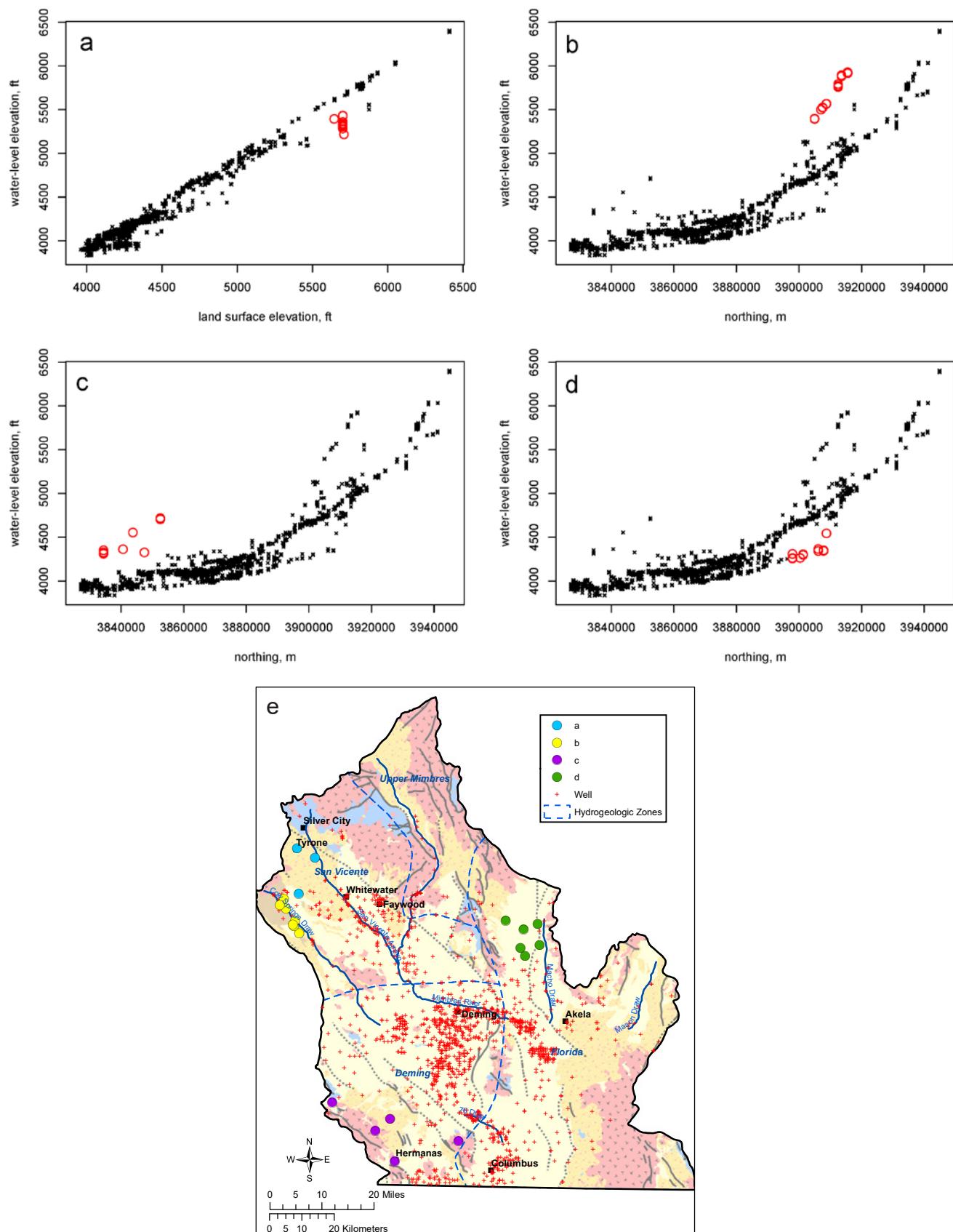
Figure 21. Comparison of outlier water-levels to regional trends, see text for discussion. Water levels from: a) Three deep wells south of Tyrone; b) Seven shallow wells south of Tyrone; c) Four wells north of Hermanas; and d) Six wells between Cooke's Range and the Goodsight Mountains. e) Colored circles show the locations of the wells in a-d. See Figure 2 for explanation of geology.
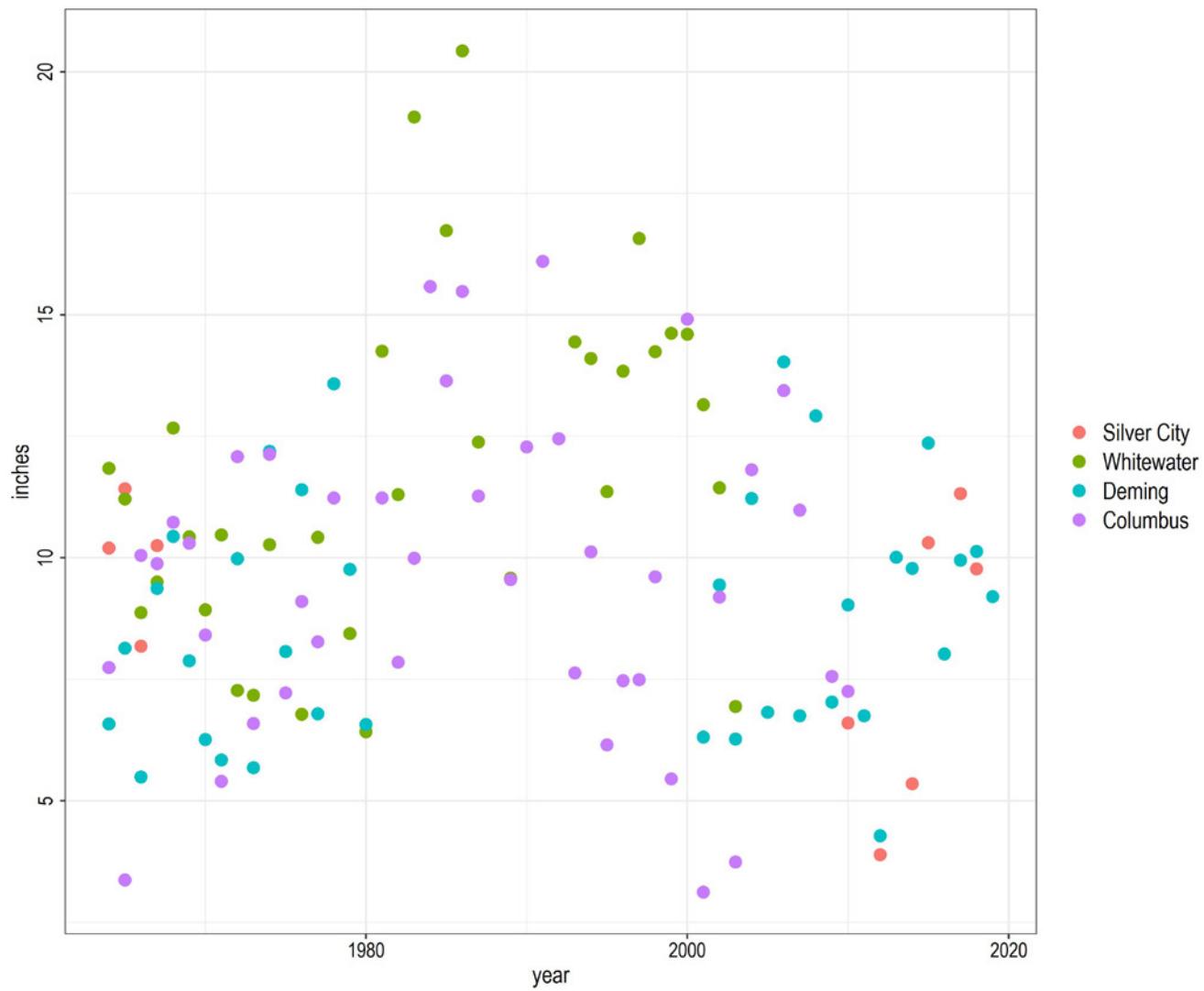
Figure 22. Annual precipitation in inches from 1964 to 2019 at four stations in the Mimbres Basin. Missing data points are years with incomplete records.
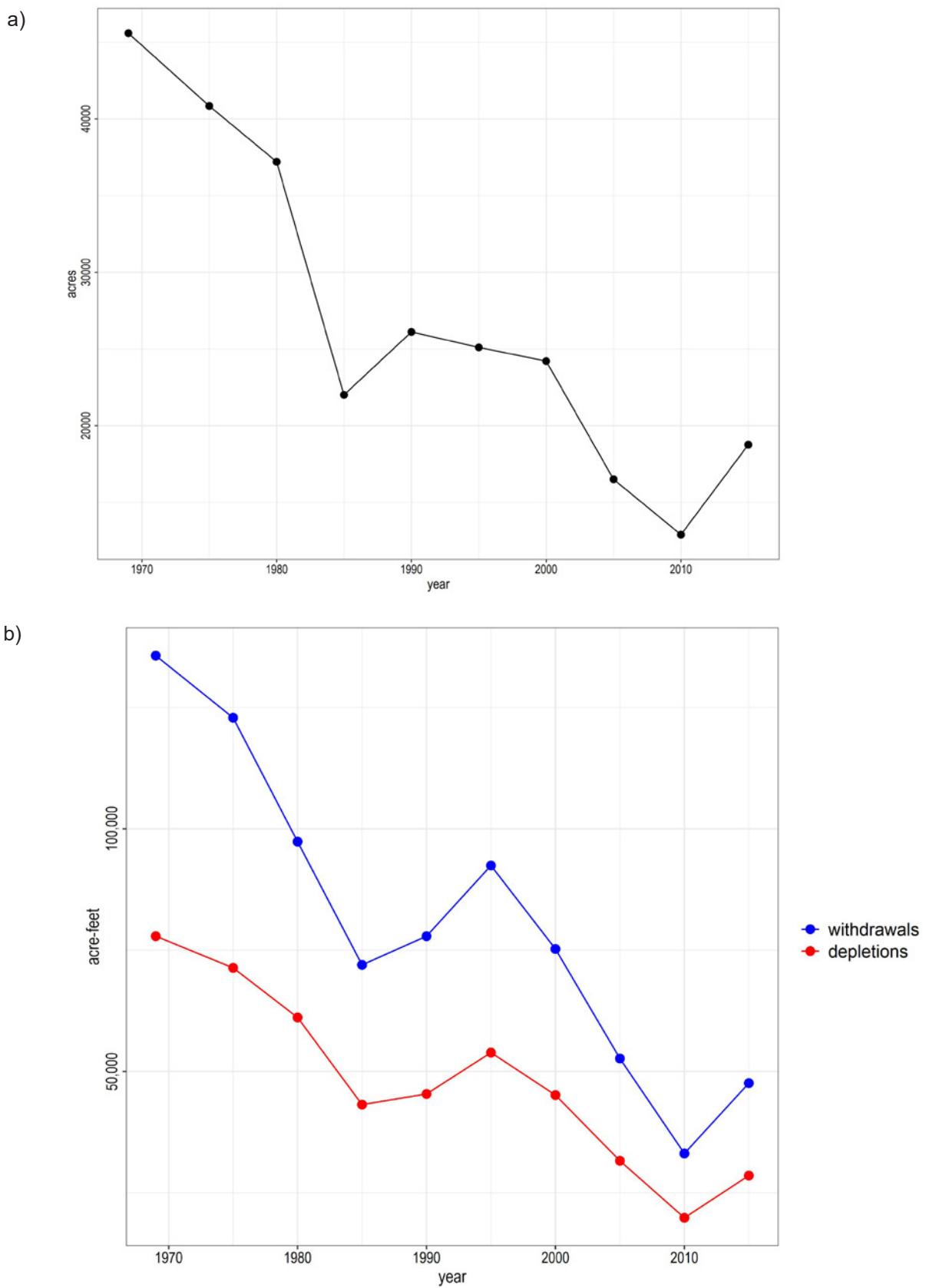
Figure 23. a) Estimates of irrigated acreage in the Mimbres Basin b) Estimates of groundwater withdrawals and depletions for irrigation in the Mimbres Basin. See text for sources of these data.

DISCUSSION

Outlier Water-levels and Wells

The paired criteria of high standardized residual (> 2) and Cook's distance > 4/n, where n is the number of data, was used to identify outlier measurements not adequately represented by the regional trend model. This criteria resulted in 144 of 3,376 water-level measurements being classified as outliers, which are shown in Figures 4 and 6. Most of the outlier data are from wells near the perimeter of the study area, and 21 of the 46 wells are located in older lithified Cenozoic sedimentary units, or older bedrock, rather than basin-fill sediments. It is likely these well are not completed in basin-fill sediments.

Many of the outlier data are clearly off of the regional trends of water-level elevation vs. land surface elevation and northing (Figure 21). Three wells south of Tyrone are located in bedrock, are 340 to 555 feet deep and plot below the regional water-level versus land-surface trend. They are likely completed in a fractured bedrock aquifer (Figure 21a). Seven wells south of Tyrone in the southeastern Big Burro Mountains are less than 100 feet deep and have water levels above the regional water-level vs. northing trend. They are also likely completed in fractured bedrock and/or a local perched aquifer (Figure 21b). Four wells northwest of Hermanas and one in the Tres Hermanas Mountains are located in or near bedrock uplands and plot above the regional water-level elevation vs. northing trend. These are likely completed in fractured bedrock or local perched aquifers (Figure 21c). Six wells between Cooke's Range and the Goodsight Mountains are completed in basin fill but have water-levels below the regional water-level elevation vs. northing trend (Figure 21d). However, overall, the third-order polynomial function of the first two principal components of easting, northing, and elevation adequately describes the regional water-level trend for the majority of the wells completed in basin fill.

Water-level Maps and Patterns of Water-level Change

It is very important when interpreting the water-level maps and changes in water-levels over time to consider the patterns of kriging variance (or standard error as shown here), as lower values indicates increased reliability of the predictions (Figure 12). The spatial patterns of standard error reflect the spatial well density and the number of measurements at each well; values are lowest (precision is highest) in areas with many wells and many measurements over time, and vice- versa.

The regional pattern of water-level elevations across the whole Mimbres Basin appears to change little over the time period of the study, 1980 to 2020 (Figure 9). The flow direction vectors indicate regional flow to the south and southeast away from recharge areas in the mountains in the northern part of the basin (Figure 10). Closed depressions in the water table surface are evident as areas of converging vectors representing anthropogenic discharge though groundwater pumping south of Deming and east of Columbus. Water-level declines are notable in the region from Deming south to Columbus when viewed at larger scale (Figure 11). The maps of water-level change over the eight five-year periods reveal large spatial and temporal variations (Figure 13), which are summed in the net water-level change map (Figure 14). Many of these changes can be related to patterns of land use and groundwater recharge and discharge. Net water-level declines of up to 80 feet dominate the region from Deming south to Columbus and Hermanas (Figure 14). These are due to groundwater pumping for irrigated agriculture.

The focused water-level maps (Figure 11) and individual change maps (Figure 13) show how the loci in space and time of water-level declines have varied in this agricultural region. Possible influences on these trends are changes in precipitation, land use, and onset and cessation of groundwater pumping. Precipitation records from four stations in the Mimbres Basin since 1964 are shown in Figure 22 (Western Regional Climate Center, 2021). Estimates of irrigated acreage and groundwater withdrawals and depletions in the Mimbres Basin are shown in Figure 23. These data were compiled from Sorenson (1977), Sorenson (1982), Wilson (1986), Wilson (1992), Wilson and Lucero (1997), Wilson et al. (2003), Longworth et al. (2008), Longworth et al. (2013), and Magnuson et al. (2019).

Withdrawal is the quantity of water taken from a ground or surface water source. Depletion is that part of a withdrawal that has been evaporated, transpired, incorporated into crops or products, consumed by man or livestock, or otherwise removed from the water environment. It includes that portion of ground water recharge resulting from seepage or deep percolation (in connection with a water use) that is not economically recoverable in a reasonable number of years, or is not usable (Wilson, 1992). The NMOSE stopped calculating depletions after 2003. Depletions for years after that shown in Figure 23 were estimated by using the average ratio of depletions to withdrawals from 1975 to 2000 of 60%. The NMOSE reported depletions but not withdrawals for 1969 and 1975. Withdrawals for these two years were estimated from average ratios of irrigated acreage to withdrawals, and depletions to withdrawals, for years in which all three types of data were available.

From 1980 to 2005 the closed depression in the water table south of Deming defined by the 4100 foot contour expanded in area and depth, spreading to the southwest and southeast. After 2005 it became deeper but not much larger. Conversely, the closed depression northeast of Columbus defined by the 3900 foot contour decreased in size from 1980 to 2005, but then expanded greatly after 2005. The areas of net water-level rise west of Red Mountain, south of Akela, and northeast of Columbus all likely represent (partial) recovery of water levels and flattening of the water table after abandonment of extensive irrigation (Figure 14). Figure 23 shows that irrigated acreage, and groundwater withdrawals and depletions have declined since the 1970s across the Mimbres Basin as a whole, but have begun to increase in the last ten years. Delineating spatial and temporal trends of irrigated acreage within the basin using remote sensing imagery is a topic of ongoing research at the NMBGMR, so future work may be able to more accurately relate the patterns of water-level change seen in these maps with irrigation patterns.

The areas of net water-level rise include two centers at Whitewater and Faywood (Figure 14). These are due to infiltration of streamflow in San Vicente Arroyo and emergence of deep- sourced water at springs around Faywood, including Faywood Hot Springs (Hawley et al., 2000). The springs at Faywood and associated net water-level rise are adjacent to the Silver City fault zone on the southwest side of the Pinos Altos Range that separates the range from the Mangas Trench or graben to the southwest (Figures 2 and 14; Hawley et al., 2000; Heywood, 2002). Although the net water-level change since 1980 is positive (Figure 14), the individual

water-level change maps show that both of these areas have experienced water-level declines from 2010 – 2015 and 2015 - 2020 (Figure 13). The precipitation trend in the Mimbres Basin (Figure 22) shows a steady decline since the 1980s. This long-term precipitation trend may be negatively affecting stream flow and spring discharge, causing the water-level declines.

The reach of the Mimbres River south from where it emerges from the bedrock constriction east of Faywood to where it turns to the east, south of Black Mountain, is a region of net water-level rise (Figures 2 and Figure 14). This is the reach where flows in the river usually completely infiltrates; flow rarely extends to Deming. Groundwater-levels have both risen and fallen in this reach since 1980. This is probably due to variation in flow of the Mimbres River.

There appears to be minimal structural control of the patterns of water-level change (Figure 14). Faults between Columbus and the West Potrillo Mountains approximately bound areas of water-level rise and decline. The Treasure Mountain Fault between Deming and Cooke's Range bounds the Mangas Trench on the northeast, separating basin-fill sediments from bedrock of the mountains, and also approximately bounds areas of net water-level rise and decline. The other mapped faults show little relation to the patterns of water level change, suggesting there is little structural compartmentalization of the basin fill aquifer at the depths investigated by the wells in this study.

Water-level predictions and patterns of water-level change are not reliable northeast of Faywood, Whitewater, and Silver City, due to the poor performance of the regional trend surface there because of the large topographic relief, and the general paucity of data. The most obvious manifestation are the areas shown in pink in the water level surfaces and change maps in this region, which are grid cells where the water level is predicted to be above the ground surface (Figures 14, 16, and 17). These areas are not indicated on the earlier figures for clarity, but the pattern is present for all of the data. The net water-level rise north of Silver City is not considered reliable. Net water-level rises in Cooke's Range and to the northeast are of somewhat questionable validity, as this is an area of generally high standard error (Figure 12). The same can be said for water-level changes along the eastern margin of the study area; although the regional trend model fits well here, the overall standard error is high.

The area of net water-level rise northwest of Akela appears to be largely due to one measurement in well NM-05950, on 2/17/2020 (Figures 13, 14, and 15). This was a high-quality measurement in that it was repeatable, but the last measurement at this well was in

2002. Overall, there are few wells and data in this area. The sequential water level change maps (Figure 13) faithfully reproduce the small water-level changes in this well from 1982 to 2002 (Figure 15), yet they also reflect that fact that kriging is a smooth interpolator and that one data point can be highly influential in regions where data is sparse, regardless of the interpolation method used.

The hydrographs in Figure 15 illustrates the variety of water level trends across the study area, with rises up to 83 feet (well NM-05950) and declines up to 84 feet (well NM-02018). Measured water levels are plotted with predicted values produced by spatio-temporal kriging. The latter includes two error bars, the smaller indicating the standard error from the regional trend model and the larger indicating the total standard error (regional trend plus spatiotemporal kriging). Note that the former is constant for each well while the latter varies though time. As in purely spatial kriging, the standard error estimates are independent of the actual data values, and are purely a function of the pattern of the measurements in space-time (or space alone in purely spatial kriging; Isaaks and Srivastava, 1989).

The total standard error around each predicted water-level elevation on the hydrographs is the 95% confidence interval about that value, or the 95% confidence about the mean. The meaning of this is rather subtle. The entire water-level dataset used here may be viewed as a sample from a (normal or Gaussian) distribution of water levels spread over space and time. If the measurements in space and time that constitute the dataset could be repeated 100 times, and the trend model and spatiotemporal kriging were recalculated for each dataset repetition, then we would expect the mean value of all 100 predictions at each well point at the chosen prediction times to fall within the error bars 95 times, or 95% of the time (Ross, 2009, p. 377).

The changes in water levels seen in the hydrographs match the changes in water levels seen on the maps through time (Figures 13 and 15). For wells with no recent measurements, the width of the confidence intervals increases as one moves further away in time from the last measurement (Figure 15; wells NM-02209, NM-02466, NM-02067), and is lower in intervals with abundant data (Figure 15; the early periods of wells NM-02209, NM-02246, NM-02018). This is a useful illustration of the advantage of the spatiotemporal method and can help clarify understanding about the reliability of patterns of water-level change through time. For wells with very few measurements (Figure 15; well NM-04102), the spatiotemporal kriging method allows a systematic "filling out" of the time series using data from other wells, and with a

lower standard error than can be achieved using only spatial kriging at any given time instance. A possible application of this is to generate realistic, data-based, time series for water levels at wells, which could then be used for history-matching in calibrating groundwater flow models.

Spatial vs. Spatiotemporal Kriging

Rawling and Rinehart (2018) and Rinehart et al. (2016) used extensive space-time datasets of water-level data to assess usable aquifer lifetimes and calculate storage changes over time periods of fifty years. In both studies, median water levels were calculated for each well for each decade, then spatial kriging was used to interpolate median water-level surfaces. The approach is straightforward and produced reliable and useful results in both studies, yet the information contained in the temporal correlation of water-levels is lost by using median water-levels at each well. In addition, at any given decade for which spatial kriging was performed, the extent of the interpolation was limited to regions within the correlation length (the range) around each well, which in some cases results in discontinuous prediction maps (Rinehart et al., 2016). This is significant as groundwater measurement networks are rarely spatially consistent through time. One can then only present a lower bound on some derivative quantity, for example storage changes, as a reliable value cannot be calculated in areas where predictions were not made.

In the spatiotemporal kriging method, predictions can be made at times for which there are no data, "filling out" the time series as described above (Figure 15, well NM-4102), and allowing the spatial extent of predictions to be larger at any given time. In essence, every well with a measurement in the study area can contribute some information to prediction of a groundwater surface at any time the user chooses. Evenly spaced five-year intervals were chosen for this study, and Figure 12 shows that for most of the time instances very few data were available. The standard error is larger in areas and at times with sparse data, yet a prediction can be rigorously be made, and one can take the standard error into account in interpretation or further analysis. Figure 12 shows that the increase in standard error is not significant in the time instances with few data—this is because of the high correlation in time of the water-level data in this study. Spurious patterns of change that are an artifact of changing well networks are greatly reduced using the spatiotemporal method (Figure 17, Ruybal et al., 2019).

Abandoned stock well south of the Florida Mountains

CONCLUSIONS

This study has extended the long history of water-level measurements in the Mimbres Basin and used the geostatistical method of spatiotemporal kriging to create water-level maps every five years from 1980 until 2020. Changes in water-levels over these five-year intervals were calculated. Compared to spatial kriging, the spatiotemporal approach offers improved precision, more detailed maps of water levels and water-level changes, predictions at times with no data as well as at locations with no data, fewer artifacts due to changing well networks over time, and overall less uncertainty in predictions.

Several important water-level trends since 1980 are revealed in the maps. In the areas of most abundant groundwater pumping for irrigation, from Deming to Columbus, water levels have declined up to 84 feet. Water-level declines and expansion of cones of depression appear to have slowed south of Deming and increased around Columbus in the past ten years. Water levels west of Red Mountain, east of the Florida Mountains, and northeast of Columbus have risen up to 42 feet, presumably as a result of declining irrigation in these areas that has resulted in flattening of cones of depression. The vicinity of Whitewater and Faywood shows net water-level rises over the 40 year period, but declines have occurred in the past ten years. Water-levels have varied considerably along the reach of the Mimbres River south of Faywood, where most of the river's flow infiltrates.

The spatiotemporal kriging approach is more mathematically complex than spatial kriging, very demanding of computational resources, and multidimensional space-time datasets with thousands of data must be manipulated to implement it. The user must be familiar enough with the theory, methodology, and potential pitfalls to make informed decisions at several points during the analysis that will affect the final results. However, the cost in funds and staff-hours of field studies to gather water-level data has always been high and continues to rise. Therefore it is prudent to analyze the data collected at such great expense with methods that will extract the most useful information. The present work builds on the recent studies of Ruybal et al. (2019) and Varouchakis and Hristopulos (2019) to demonstrate that spatio-temporal kriging of water- level data is superior to spatial kriging in this regard.

Any hydrogeologic analysis that is dependent on water-level maps should benefit from the improvements offered by spatiotemporal kriging. This includes assessment of water-level changes (as demonstrated here), assessment of storage changes, calibration of numerical groundwater-flow models, and design and updating of water-level monitoring networks.

ACKNOWLEDGEMENTS

REFERENCES

Abdi, H. and Williams, L.J., 2010, Principal component analysis: WIREs Computational Statistics, v. 2, p. 433–459.

Bivand, R.S., Pebesma, E., and Gomez-Rubio, V., 2013, Applied spatial data analysis with R, Second edition: Springer, NY, https://asdar-book.org/.

Clemons, R.E., 1986, Petrography and stratigraphy of Seville-Trident exploration wells near Deming, New Mexico: New Mexico Geology, v.8, p. 5–11.

Clemons, R.E., 1998, Geology of the Florida Mountains, southwestern New Mexico: New Mexico Bureau of Geology and Mineral Resources Memoir 43, 113 p.

Clemons, R.E. and Mack, G.H., 1988, Geology of southwestern New Mexico in Mack, G.H., Lawton, T.F., and Lucas, S.G., eds., Cretaceous and Laramide Tectonic Evolution of Southwestern New Mexico: New Mexico Geological Society Guidebook 39, p. 45–57.

Conover, C.S., and Akin, P.D., 1942 Progress report on the ground-water supply of the Mimbres Valley, New Mexico, 1938–1941. 14th and 15th Biennial Reports of the State Engineer of New Mexico, 45 p.

Contaldo, G.J., and Mueller, J.E., 1991, Earth fissures of the Mimbres Basin, southwestern New Mexico: New Mexico Geology, v.13, p. 69–74.

Cressie, N., 2015, Statistics for spatial data, Hoboken, NJ, John Wiley & Sons, 900 p.

Cressie, N., and Wikle, C.K., 2011, Statistics for spatio-temporal data, Hoboken, NJ, John Wiley & Sons, 588 p.

Cuddy, A.S., and Keyes, E., 2011, Groundwater model of the Mimbres Basin, Luna, Grant, Sierra, and Doña Ana Counties, New Mexico: New Mexico Office of the State Engineer Hydrology Bureau Technical Report 11-1, 43 p.

Darton, N.H., 1914, Underground water of Luna County, New Mexico: U.S. Geological Survey Water-Supply Paper 345C, 40 p.

Darton, N.H. 1916, Geology and underground water of Luna County, New Mexico: U.S. Geological Survey Bulletin 618, 188 p.

D.B. Stephens and Associates, 2009, Estimates of region-wide and Deming area water supplies: unpublished report prepared for the New Mexico Interstate Stream Commission, 28 p.

Desbarats, A.J., Logan, C.E., Hinton, M.J., and Sharpe, D.R., 2002, On the kriging of water table elevations using collateral information from a digital elevation model: Journal of Hydrology, v. 255, p. 25–38.

Drewes, H., Houser, B.B., Hedlund, D.C., Richter, D.H., Thorman, C. H., and Finnell, T.L. 1985, Geologic map of the Silver City 1x2 degree quadrangle, New Mexico and Arizona: U.S. Geological Survey, Miscellaneous Investigations Series Map I-1310-C, scale 1:250,000.

Finch, S.T., McCoy, A., and Melis, E., 2008, Geologic controls on ground-water flow in the Mimbres Basin, southwestern New Mexico; in Mack, G., Witcher, J., and Leuth, V.W., eds., Geology of the Gila Wilderness-Silver City area: New Mexico Geological Society Guidebook 59, p. 189–198.

Galloway, D.L., Jones, D.R., and Ingebritsen, S.E., eds., 1999, Land subsidence in the United States: U.S. Geological Survey Circular 1182, 165 p.

Gasch, C.K., Hengl, T., Gräler, B, Meyer, H., Magney, T.S., and Brown, D.J., 2015, Spatiotemporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set: Spatial Statistics, v.14, p. 70–90.

Glen, S., 2020, Cook's Distance / Cook's D: Definition, Interpretation: https://www.statisticshowto.com/cooks-distance/, accessed 6/3/2020.

Gräler, B., Pebesma, E., and Heuvelink, G., 2016, Spatio-temporal interpolation using gstat: The R Journal, v. 8, p. 204–218.

Haneberg, W.C., and Friesen, R.L., 1995, Tilts, strains, and ground-water levels near an earth fissure in the Mimbres Basin, New Mexico: Geological Society of America Bulletin, v.107, p. 316-326.

Hanson, R.T., McLean, J.S., and Miller, R.S., 1994, Hydrogeologic framework and preliminary simulation of ground-water flow in the Mimbres Basin, southwestern New Mexico: U.S. Geological Survey Water-Resources Investigations Report 94-4011, 118 p.

Hawley, J.W., Hibbs, B.J., Kennedy, J.F., Creel, B.J., Remmenga, M.D., Johnson, M., Lee, M.M., and Dinterman, P., 2000, Trans-international boundary aquifers in southwestern New Mexico: New Mexico Water Resources Research Institute Technical Completion Report, 126 p.

Hengl, T., Heuvelink, G.B., and Stein, A., 2003, Comparison of kriging with external drift nad regression-kriging: Technical Note, ITC, http://www.itc.nl/library/Academic output/

Hengl, T., Heuvelink, G.B., and Stein, A., 2004, A generic framework for spatial prediction of soil variables based on regression-kriging: Geoderma, v. 120, p. 75–93.

Hengl, T., Heuvelink, G.B., and Rossiter, D.G., 2007, About regression kriging: From equations to case studies: Computers and Geosciences, v.33, p. 1301–1315.

Hengl, T., 2009, A Practical Guide to Geostatistical Mapping, 2nd Edition: Luxembourg, Office for Official Publications of the European Communities, 270 p.

Heywood, C.E., 2002, Estimation of alluvial-fill thickness in the Mimbres ground-water basin, New Mexico, from interpretation of isostatic residual gravity anomalies: U.S. Geological Survey Water-Resources Investigations Report 02-4007, 16 p.

Huevelink, G.B., and Griffith, D.A., 2010, Space-time geostatistics for geography: A case study of radiation monitoring across parts of Germany: Geographical Analysis, v. 42, p. 161–179.

Isaaks, E.H., and Srivastava, R.M., 1989, An introduction to applied geostatistics, Oxford, Oxford University Press, 561 p.

Jolliffe, I.T., and Cadima, J., 2016, Principal component analysis: a review and recent developments, Philosophical Transactions of the Royal Society A, v. 374, 201502020, https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202.

Kennedy, J.F., Hawley, J.W., and Johnson, M., 2000, The hydrogeologic framework of basin-fill aquifers and associated ground-water-flow systems in southwestern New Mexico—An overview; in Lawton, T.F., McMillan, N.J., and McLemore, V.T., eds., Southwest Passage-A Trip through the Phanerozoic: New Mexico Geological Society Guidebook 51, p. 235 – 244.

King, F.H., 1899, Principles and conditions of the movements of groundwater: U.S. Geological Survey 19th Annual Report Part 2, 59-294.

Kitanidis, P.K., 1997, Introduction to geostatistics: applications in hydrogeology, Cambridge, Cambridge University Press, 249 p.

Kyriakidis, P.C., and Journel, A.C., 1999, Geostatistical space-time models: A review: Mathematical Geology, v. 31, p. 651–684.

Lark, R.M., and Webster, R., 2006, Geostatistical mapping of geomorphic variables in the presence of trend: Earth Surface Processes and Landforms, v. 31, p. 862–874.

Love, D.W., and Seager, W.R., 1996. Fluvial fans and related basin deposits of the Mimbres drainage: New Mexico Geology, v.18, p. 81–92.

Longworth, J.W., Valdez, J.M., Magnuson, M.L., Albury, E.S., and Keller, J., 2008, New Mexico water use by categories 2005: New Mexico Office of the State Engineer Technical Report 52, 126 p.

Longworth, J.W., Valdez, J.M., Magnuson, M.L., and Richard, K., 2013, New Mexico water use by categories 2010: New Mexico Office of the State Engineer Technical Report 54, 144 p.

Magnuson, M.L., Valdez, J.M., Lawler, C.R., Nelson, M., and Petronis, L., 2019, New Mexico water use by categories 2015: New Mexico Office of the State Engineer Technical Report 55, 126 p.

McLean, J.S., 1977, Hydrologic maps and data in the Mimbres Basin, New Mexico: U.S. Geological Survey Open-File Report 77-317, 531 p.

New Mexico Bureau of Geology and Mineral Resources, 2003, Geologic Map of New Mexico, 1:500,000: New Mexico Bureau of Geology and Mineral Resources.

Pebesma, E., 2004, Multivariable geostatistics in S: The gstat package: Computers & Geosciences, v. 30, p. 683–691.

Pebesma, E., 2012, Spacetime: Spatio-temporal data in R: Journal of Statistical Software, v. 51, p. 1–30.

Pyrcz, M.J. and Deutsch, C.V., 2003, The whole story on the hole effect: Geostatistical Association of Australasia, Newsletter, v. 18.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at https://www.R-project.org/.

Rawling, G.C., and Rinehart, A.J., 2018, Lifetime projections for the High Plains aquifer in east-central New Mexico: New Mexico Bureau of Geology and Mineral Resources Bulletin 162, 44 p.

Rinehart, A.J., Mamer, E., Kludt, T., Felix, B., Pokorny, C., and Timmons, S., 2016, Groundwater level and storage changes in alluvial basins in the Rio Grande Basin, New Mexico: New Mexico Water Resources Research Institute Technical Completion Report, 41 p.

Romero, D.M., and Cook, C.W., 2009, Groundwater recharge analysis and estimate of recharge option costs: unpublished technical memorandum prepared for the Town of Silver City, Balleau Groundwater, Inc., 55 p.

Ross, S.M., 2009, Introduction to probability and statistics for scientists and engineers, Elsevier Academic Press, 664 p.

Ruybal, C.J., Hogue, T.S., and McCray, J.E., 2019, Evaluation of groundwater levels in the Arapahoe aquifer using spatiotemporal regression kriging: Water Resources Research, v. 55, p. 2820–2837, https://doi.org/10.1029/2018WR023437

Schuenemeyer, J.H., and Drew, L.J., 2011, Statistics for Earth and Environmental Scientists, Hoboken, NJ, John Wiley & Sons, 397 p.

Seager, W.R., Clemons, R.E., Hawley, J.W., and Kelley, R. E., 1982, Geology of northwest part of Las Cruces 1° x 2° sheet, New Mexico: New Mexico Bureau of Geology and Mineral Resources Geologic Map 53, scale 1:125,000, 3 sheets.

Seager, W.R., 1995, Geology of southwest quarter of Las Cruces and northwest El Paso 1° x 2° sheets, New Mexico: New Mexico Bureau of Geology and Mineral Resources Geologic Map 60, scale 1:125,000, 5 sheets.

Snepvangers, J.J.J.C., Heuvelink, G.B.M., and Huisman, 2003, Soil water content interpolation using spatio-temporal kriging with external drift: Geoderma, v. 112, p. 253–271.

Sorenson, E.F., 1977, Water use by categories in New Mexico counties and river basins, and irrigated and dry cropland acreage in 1975: New Mexico Office of the State Engineer Technical Report 41, 44 p.

Sorenson, E.F., 1982, Water use by categories in New Mexico counties and river basins, and irrigated acreage in 1980: New Mexico Office of the State Engineer Technical Report 44, 60 p.

Taylor, C.J., and Alley, W.M., 2001, Ground-water-level monitoring and the importance of long- term water-level data: U.S. Geological Survey Circular 1217, 68 p.

Theis, C., 1939. Progress report on the ground-water supply of the Mimbres Valley, New Mexico. 12th and 13th Biennial Reports of the State Engineer of New Mexico 1934–1938, p. 135–153.

Tóth, J., 1963, A theoretical analysis of groundwater flow in small drainage basins: Journal of Geophysical Research, v. 68, p. 4795–4812.

Trauger, F.D., 1972, Water resources and general geology of Grant County, New Mexico: New Mexico Bureau of Geology and Mineral Resources Hydrologic Report 2, 118 p.

Varouchakis, E. A. and Hristopulos, D.T., 2019, Comparison of spatiotemporal variogram functions based on a sparse dataset of groundwater level variations: Spatial Statistics, v. 34, http://dx.doi.org/10.1016/j.spasta.2017.07.003

Webster, R., and Oliver, M.A., 2007, Geostatistics for Environmental Scientists, Hoboken, NJ, John Wiley & Sons, 315 p.

Western Regional Climate Center, 2018: Reno, Nevada: https://wrcc.dri.edu/ (accessed January 2021).

White, W.N., 1931, Preliminary report on the ground-water supply of Mimbres Valley, New Mexico: U.S. Geological Survey, Water-Supply Paper 637-B, p. 69–90.

Wikle, C. K., Zammit-Mangion, A., and Cressie, N., 2019, Spatio-Temporal Statistics with R: Boca Raton, FL: Chapman & Hall/CRC, 380 p.

Wilkins, D.W. 1986, Geohydrology of the southwest alluvial basins, regional aquifer-systems analysis in parts of Colorado, New Mexico, and Texas: U.S. Geological Survey Water Resources Investigations Report 84 No.4224. 61 p., 7 plates.

Wilson, B.C., 1986, Water use in New Mexico in 1985: New Mexico Office of the State Engineer Technical Report 46, 87 p.

Wilson, B.C., 1992, Water use by categories in New Mexico counties and river basins, and irrigated acreage in 1990: New Mexico Office of the State Engineer Technical Report 47, 149 p.

Wilson, B.C. and Lucero, A.A., 1997, Water use by categories in New Mexico counties and river basins, and irrigated acreage in 1995: New Mexico Office of the State Engineer Technical Report 49, 145 p.

Wilson, B.C., Lucero, A.A., Romero, J.T., and Romero, P.J., 2003, Water use by categories in New Mexico counties and river basins, and irrigated acreage in 2000: New Mexico Office of the State Engineer Technical Report 51, 173 p.

Wilson, J.L, and Guan, H., 2004, Mountain-block hydrology and mountain front recharge, *in* Hogan, J.F. ,Phillips, F.M., and Scanlon, B.R., eds., Groundwater recharge in a desert environment: Washington, D.C., American Geophysical Union, p. 113–137.

APPENDIX 1 — PRINCIPAL COMPONENT ANALYSIS AND REGIONAL TREND MODEL

Principal component analysis was used to derive a regional water-level trend model. The easting, northing, and land surface elevation coordinate of each well (in meters) was transformed into three principal components as shown in Table 1, and repeated here:

Table A1.1. Principal component analysis loadings.

	PC1	PC2	PC3
Easting	0.53	-0.78	-0.34
Northing	0.57	-0.62	0.54
Land Surface Elevation	-0.63	-0.09	-0.77
Standard Deviation	1.53	0.73	0.33
Proportion of Variance	0.78	0.18	0.04
Cumulative Proportion	0.78	0.96	1.00

The three coordinates were centered and standardized prior to the analysis. The three principal components are uncorrelated (geometrically orthogonal) and are linear combinations of the original data, where each coordinate is weighted by its loading on each component, e.g.

$$PC1 = 0.53 \times easting + 0.57 \times northing - 0.63 \times land\ surface\ elevation. \quad (A.1)$$

The first two components explain 96% of the variation in the original data. These two components were combined in a third order polynomial function to derive a regional trend model. The function was fit to all of the water-level elevation data for each well, regardless of time. Ordinary least-squares fitting was implemented in R (R Core Team, 2018) to optimize the model. The resulting regional trend model is of the form:

$$a_1 + a_2x + a_3y + a_4x^2 + a_5xy + a_6y^2 + a_7x^3 + a_8x^2y + a_9xy^2 + a_{10}y^3, \quad (A.2)$$

Where PC1 = x and PC2 = y. The coefficients are shown in Table A1.2.

Table A1.2. Coefficients and statistics of the regional trend model.

coefficient	value	standard error	t-value	Pr(>|t|)
a_1	1,282.97	0.53	2431.93	~ 0
a_2	-63.88	0.41	-157.58	~ 0
a_3	-18.30	0.91	-20.18	1.47e-85
a_4	10.64	0.22	49.45	~ 0
a_5	9.27	0.52	17.75	1.73e-67
a_6	628	0.54	11.55	2.69e-30
a_7	0.49	0.053	9.07	1.95e-19
a_8	1.20	0.18	6.78	1.43e-11
a_9	-2.82	0.33	-8.42	5.51e-17
a_{10}	-2.32	0	-6.15	8.47e-10
Multitude r² 0.99			**Adjusted r²** 0.99	

The standard errors are small relative to the coefficient values. The t-values are a measure of how far the coefficient estimates are from zero, in units of standard deviations. Larger numbers here are better. The last column (Pr(>|t|)) indicates the probability that the relationship between the term of the model and the predicted variable could occur by chance. All of the probabilities are vanishingly small, indicating that all of the terms are statistically significant. The multiple r^2 indicates the proportion of variance in the data explained by the model. The adjusted r^2 does the same, but accounts for the number of variables in the model. It is preferred when there are many variables in a multiple regression.

The r^2 values indicate that globally, 99% of the variance in the water-level elevation data is explained by the model. This might suggest that the model is sufficient to explain the water level variations across the study area and that spatiotemporal kriging of the trend model residuals is superfluous. This is not the case as illustrated in Figure A1.1. The regional trend prediction at any well is constant through time, and the spatiotemporal kriging accounts for the temporal fluctuations in the water level.

Ordinary least squares optimization of the regional trend model was used in this study. Ordinary least squares assumes that the residuals are independent and normally distributed (Schuenemeyer and Drew, 2011). The regional trend predictions were subtracted from the data values and the residuals were then interpolated with spatiotemporal kriging. This procedure is straightforward and commonly used with purely spatial kriging under the general name "regression kriging" (Hengl et al., 2003, 2007) The total variance of the resulting prediction, which is the sum of the separately calculated variance of the trend model prediction plus the kriging variance (spatial or spatiotemporal) is equal to the universal kriging variance, in which the regional trend model and the kriging weights are calculated simultaneously (Hengl et al., 2003, 2004).
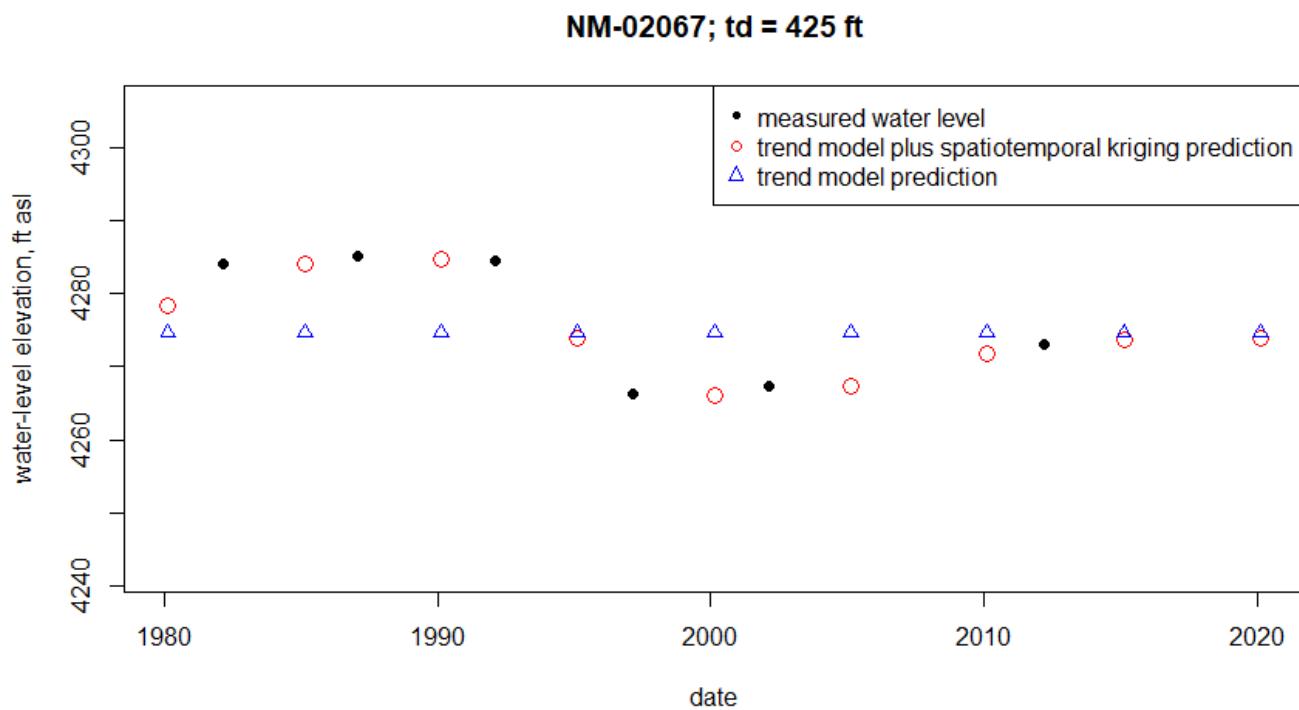

Figure A1.1. Example hydrograph showing regional trend model prediction at the well point, measured water levels, and spatiotemporal kriging prediction.

It is known that using ordinary least squares optimization to derive the regional trend model is not ideal for several reasons (Webster and Oliver, 2007). The resulting model does not yield minimum variance estimates unless the sample sites are selected randomly (although one could argue that the well locations in this study are essentially random samples of water levels in the aquifer, at least in space). The variogram calculated from the trend model residuals is biased because the residuals are spatially correlated and are nonlinearly dependent on the trend model parameters, which are themselves estimated from the same data and have errors. However, somewhat fortunately, the bias increases with lag distance in the variogram and the shorter spatial (and temporal in this study) lags are most important in the kriging interpolation.

Dealing with this issue rigorously is difficult and entails a significant increase in mathematical and computational complexity (Hengl et al., 2007; Lark and Webster, 2006; Webster and Oliver, 2007). Ordinary and generalized least squares optimization are used to estimate the regional trend, with the modelled covariance structure of the residuals (i.e., the variogram) from the former used as an input in the latter procedure. Iteration is usually required until the generalized least squares model parameters stabilize. Lark and Webster (2006) used the residual maximum likelihood method to estimate the regional trend of a geomorphic surface and it's variogram of residuals and showed variograms that were "substantially different" from previous work that employed regression kriging using ordinary least squares. However, in the end, their estimation of the regional trend and residual variation did not differ notably from previous work, probably because of the importance of short lags in the variogram (Lark and Webster,2006). Hengl et al. (2007, p. 1305) commented that the covariance structure of residuals from a regional trend determined by generalized least squares "in practice…will hardly differ from the covariance structure of the ordinary least square residuals".

Thus it is not clear that, in general, the extra effort required to address the shortcomings of ordinary least squares estimation of the regional trend surface is warranted. The issue was briefly explored here by using the R package caret (https://cran.r-project.org/web/packages/caret/vignettes/caret.html) to calculate a generalized linear model of the third order polynomial function of the first two principal components. This model only relaxes the assumption of independent, normally distributed residuals in fitting the model to the data; it does not address spatial correlation of the residuals. The resulting trend model coefficients were identical to those generated by ordinary least squares. Further calculations and the spatiotemporal kriging used the regional trend model generated by ordinary least squares. One may conservatively assume that the results presented in the report contain some additional measure of bias and uncertainty beyond that shown, though the effect is likely to be small.

APPENDIX 2 — COMPARISON OF SPATIOTEMPORAL VARIOGRAM MODELS TO THE EXPERIMENTAL VARIOGRAM

As was observed by Ruybal et al. (2019), all classes of spatiotemporal variogram models seem to fit well and visually appear to capture basic structure of the experimental variogram. The sum-metric is the best-fitting model, based on the optim value of 0.0028 and MSE value of 93 (Table A2.1). optim is the mean of the weighted squared deviations between the sample spatiotemporal variogram and the variogram model; ideally it should be zero. The weights are based on how many data points fall in each space-time lag bin. MSE is the unweighted sum of the mean-squared errors between the sample spatiotemporal variogram and the variogram model.

Table A2.1 Model-fitting parameters for variogram models. Better values are in bold.

parameter	metric	product-sum	sum-metric	separable	simple sum-metric
optim	0.0114	0.0034	**0.0028**	0.029	0.0030
MSE	115	112	**93**	128	94.32

"Leave-one-out" cross-validation was performed for each variogram model fit to the experimental spatiotemporal variogram. The nmax parameter was set to 50 to minimize computation time; only the nearest 50 datapoints in terms of spatiotemporal correlation were used in the cross-validation (this was also done for the kriging interpolations). The statistics also tended to degrade of the value were set higher. The statistics of the cross-validation residuals are shown in Table A2.2. All of the models except the simple sum-metric achieve a best value of one of the statistics.

Table A.2.2. Leave-one-out cross-validation statistics of spatiotemporal variogram models. Better values are in bold.

| parameter	metric	product-sum	sum-metric	separable	simple sum-metric
Mean	**-0.025**	-0.05	-0.033	-0.056	-0.033
Mean squared error	11.17	8.35	8.41	**8.11**	8.27
Correlation between observed and predicted	**0.96**	0.97	0.97	0.97	0.97
Correlation between predicted and residual	-0.31	**-0.18**	-0.21	-0.168	-0.224
Variance of residuals	11.17	8.35	8.41	**8.11**	8.27
Skewness	**-0.78**	-1.3	-1.24	-1.37	-1.26
Maximum of residuals	23.3	18.29	**17.65**	17.86	18.32
Minimum of residuals	**-38.66**	-39.4	-40.1	-40.2	-40.17
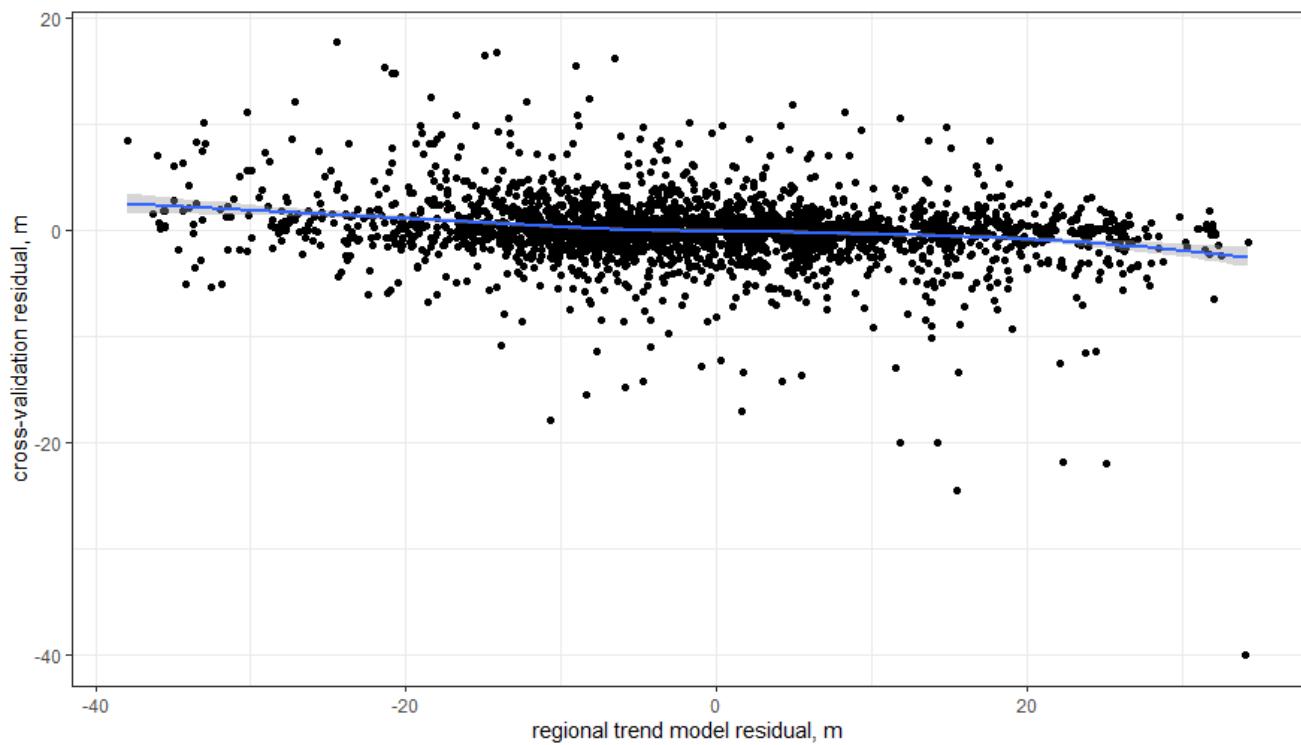
Figure A2.1. Cross-validation residuals (cross validation prediction – trend model residual) plotted against the regional trend model residuals. The data cluster about zero across nearly the full range of trend model residuals, illustrating accurate predictions across the range of the data. The curve is a local smoothing fit, showing a small bias at the lowest and highest values of the regional trend residual.
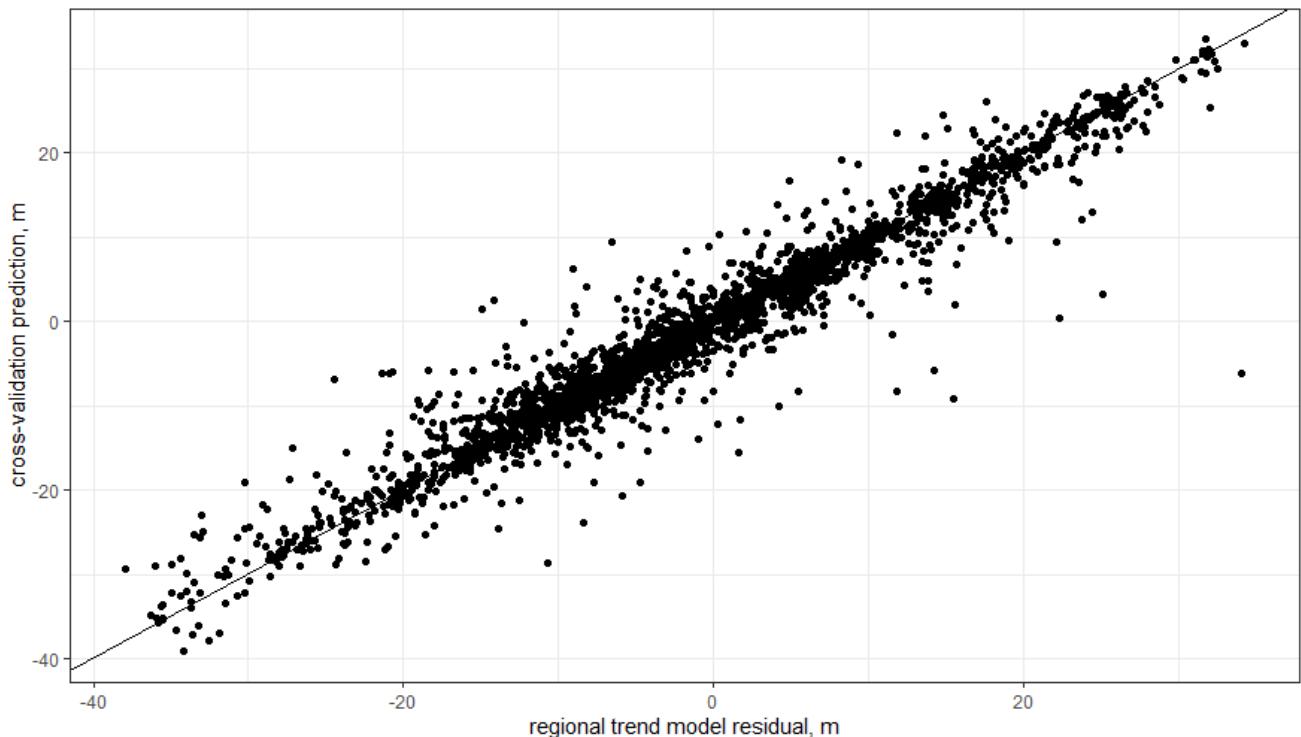

Figure A2.2. Cross-validation predictions plotted against the regional trend model residuals, with the line x=y for comparison. The data follow the line closely, showing that the sum-metric spatiotemporal model provides accurate predictions across the range of data values.

WATER-LEVEL TRENDS IN THE MIMBRES BASIN, SOUTHWEST NEW MEXICO

APPENDIX 3 — COMPARISON OF GLOBAL AND REGIONAL VARIOGRAM MODELS

Results are presented here for comparison of the global spatiotemporal variogram model as applied to the Deming, San Vicente, and Florida hydrogeologic zones (Figure 4), and local models fit to the subset of water-level data from wells in those zones. Table A3.1 compares the model fitting parameters for the best-fitting model to the global data, and the best-fitting model to the data in each zone. Figures A3.1–A3.4 illustrate global experimental variogram and best-fitting model, and the local region experimental variograms and their best-fitting models. The global model fits the global sample variogram better than any of the regional models fit the regional data. Put another way, the regional models are poorer representations of the regional spatiotemporal correlation structures, although all models appear to fit the smallest spatiotemporal lags reasonably well, and these are the most important for the kriging interpolation.

The cross validation statistics in Table A3.2 indicate that the global model, when applied to the data from the individual regions, performs better in the Deming and San Vicente zone, and about as well in the Florida zone. There is no strong justification for choosing the added complexity of individual regional models over the global model, and so the global model was used in the subsequent analysis.

Table A3.1. Model fitting parameters; smaller is better.

	Global, sum-metric	Deming, sum-metric	San Vicente, sum-metric	Florida, metric
optim	**0.0028**	0.0049	0.0065	0.0078
MSE	**93**	566	5549	683

Table A3.2. Cross-validation statistics for comparison of performance of local and global variogram models in each hydrogeologic zone.

	Deming		San Vicente		Florida	
	Global	Local	Global	Local	Global	Local
Mean	**-0.0038**	0.038	**-0.026**	-0.038	-0.10	**-0.074**
Mean squared error	**7.47**	7.72	**7.67**	9.09	12.68	**11.97**
Mean squared normalized error	0.99	0.99	0.99	0.99	0.99	0.99
Correlation between observed and predicted	0.974	0.974	**0.975**	0.971	0.958	**0.961**
Correlation between predicted and residual	**-0.195**	-0.242	-0.220	**-0.155**	**-0.245**	-0.267
Variance of residuals	**7.47**	7.72	**7.68**	9.10	12.68	**11.97**
Skewness	**0.32**	0.40	-2.34	**-1.39**	-1.93	**-1.56**
Minimum of residuals	-17.57	**-15.68**	-35.76	**-33.58**	**-35.11**	-36.12
Maximum of residuals	**16.71**	17.53	**17.65**	18.73	**13.09**	26.94

## Sample global spatiotemporal variogram and fitted model
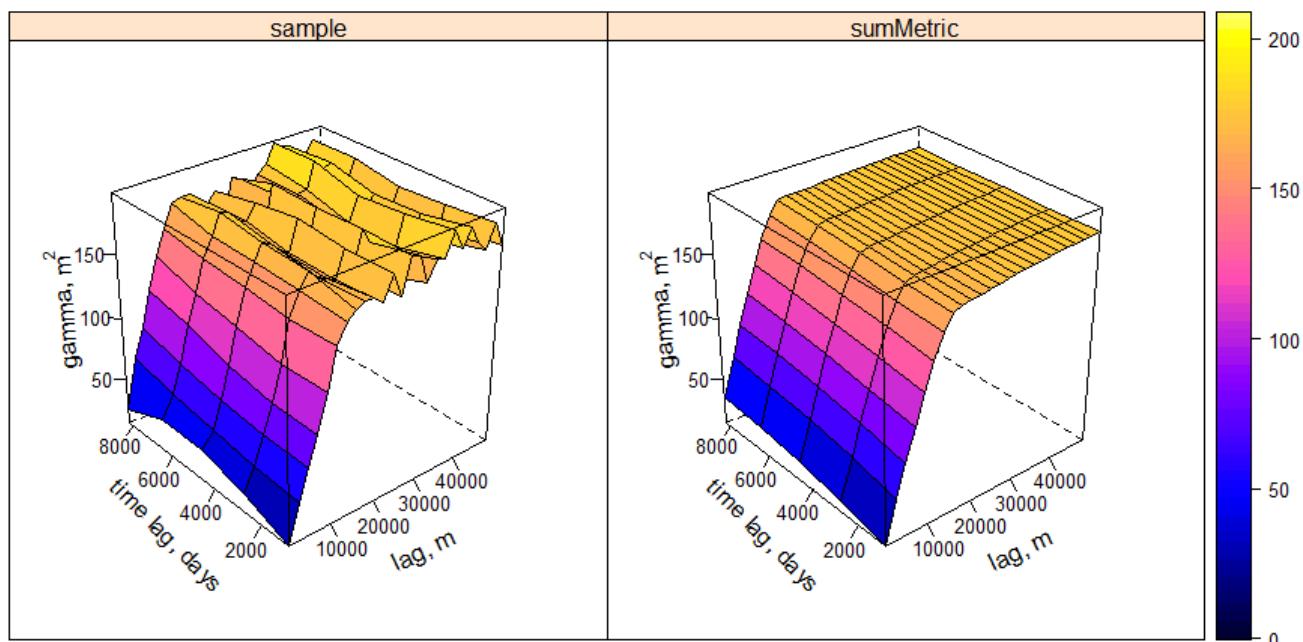
Figure A3.1. Global sample spatiotemporal variogram and best-fitting model.

## Deming hydrogeologic zone, sample spatiotemporal variogram and fitted model
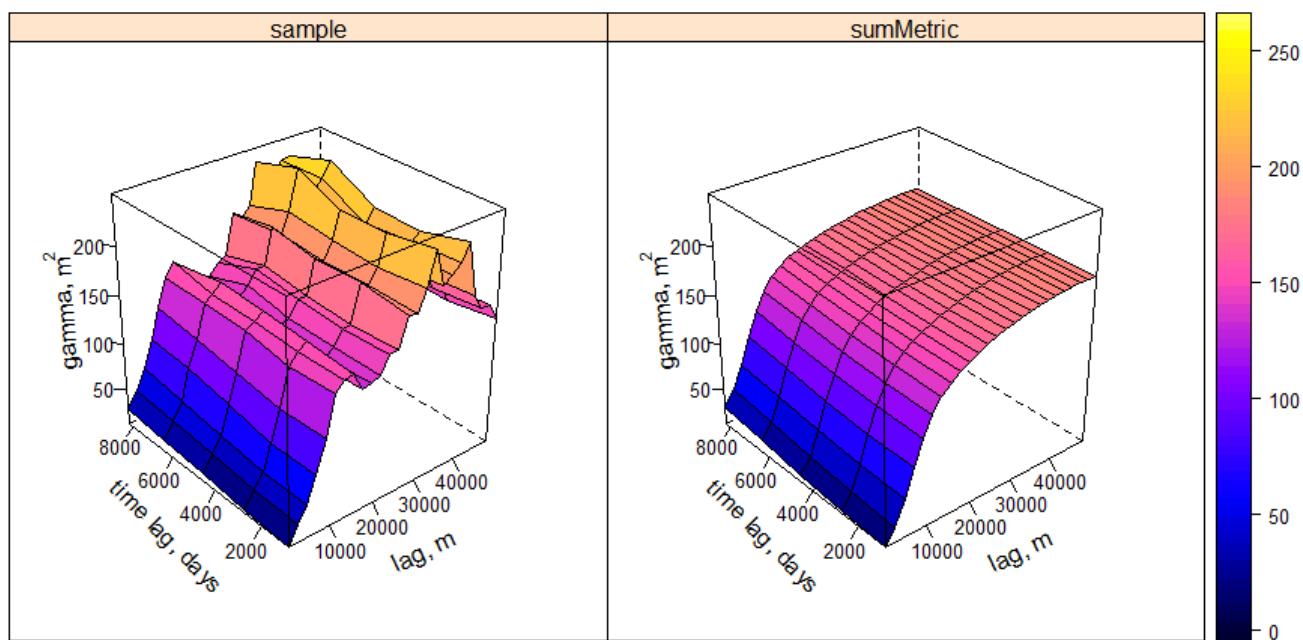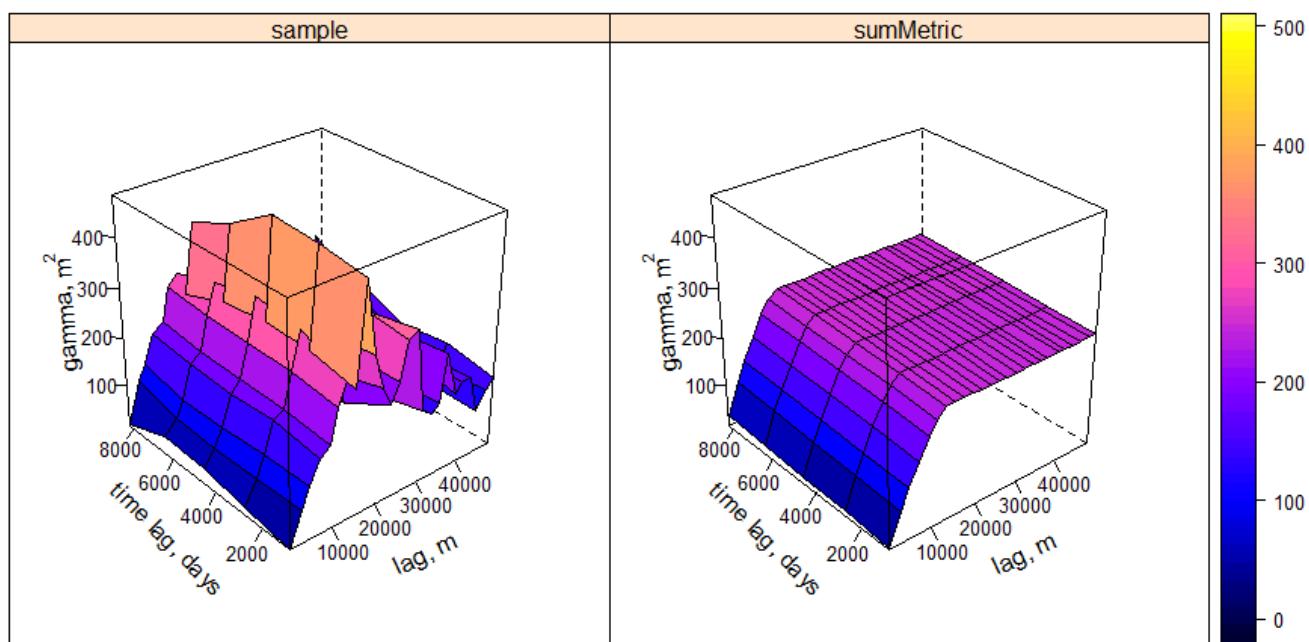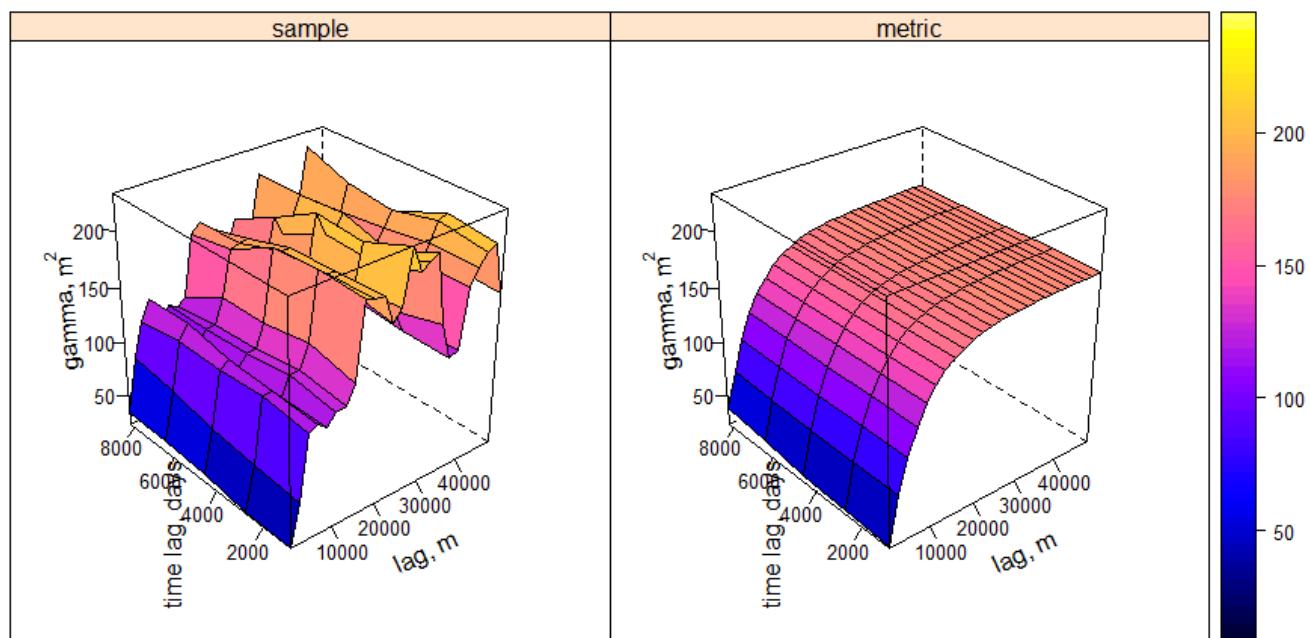
**Figure A3.2.** Sample spatiotemporal variogram and best-fitting model for the Deming hydrogeologic zone.

Figure A3.3. Sample spatiotemporal variogram and best-fitting model for the San Vicente hydrogeologic zone.


Figure A3.4. Sample spatiotemporal variogram and best-fitting model for the Florida hydrogeologic zone.

New Mexico Bureau of Geology and Mineral Resources

A Research Division of New Mexico Institute of Mining and Technology

Socorro, NM 87801
(575) 835-5490
geoinfo.nmt.edu