Electronic Supplementary Material ESM1 – Hydrogeology Journal

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

G. Rawling

New Mexico Bureau of Geology & Mineral Resources
New Mexico Institute of Mining & Technology
1015 Tijeras Avenue NW, Suite 200
Albuquerque, NM 87102-2909, USA
geoffrey.rawling@nmt.edu

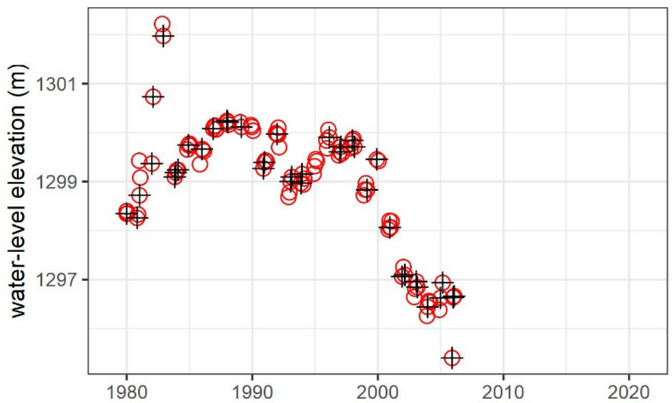
Section S1. Downsampling data at five wells

In the original development of this study, discontinuities in the kriging variance maps were evident around a cluster of four wells 12 km southeast of Silver City at the town of Bayard (Fig. 12b in Rawling, 2021). This appeared to be an artifact of the very large number of measurements at these wells combined with the use of a local spacetime neighborhood in the spatiotemporal kriging algorithm in the gstat package in R (Oleg Mahknin, New Mexico Tech Mathematics Department, personal communication, 2021). Using a local neighborhood is reasonable and practical in that wells outside of the spatial correlation range around a given well do not improve the predictions at that well, and also greatly reduces computation times to 3 – 4 hours. As noted by Wikle et al. (2019, p. 175), *"Spatio-temporal kriging … is relatively quick and easy to implement for small data sets, but it starts to become prohibitive as data sets grow in size, unless some approximation is used. For example, the function krigeST allows one to use the argument nmax to determine the maximum number of observations to use when doing prediction. The predictor is no longer optimal, but it is close enough to the optimal predictor in many cases of practical interest."* Attempts at spatiotemporal kriging using the entire 40-year dataset without setting a local neighborhood resulted in the R session crashing.

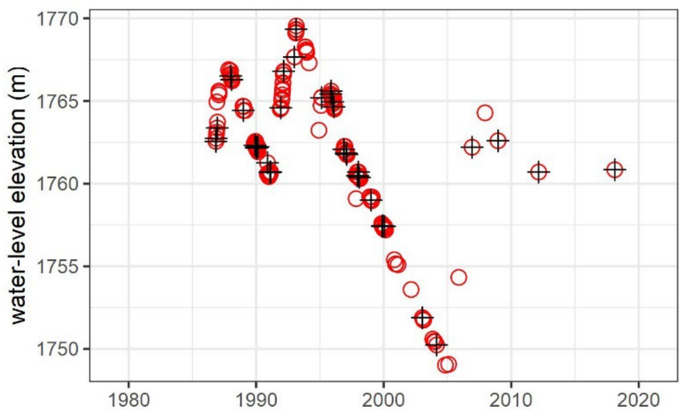
The data at these four wells (WL-0086, WL-0088, WL-0090, WL-0091) and a fifth well 8 km southwest of Deming (NM-02209) were downsampled to 40 measurements. The first and last measurement in the time series at each well was kept, and the intervening data were evenly sampled in sequence. Fig. S1 shows the original time series and the downsampled time series for each of the five wells. Spatiotemporal kriging using a local spacetime neighborhood of 200 data and the downsampled time series at these wells produced the smoothly varying kriging variance (and thus standard error) maps shown here.

This downsampling procedure is completely arbitrary but it is clear from Fig. S1 that very little information on the water level trends at these wells is lost in the process. In addition, the well NM-02209 southwest of Deming is in an area with abundant data in space and time, and the four wells southeast of Silver City are at the margins of the bedrock uplift east of the Silver City fault in an area of high relief, where the regional trend surface plus kriging model would not expected to perform well (Figs. 1 and 2, main paper). This is shown by the pink overlay on the maps in Figs. 7 and 8.

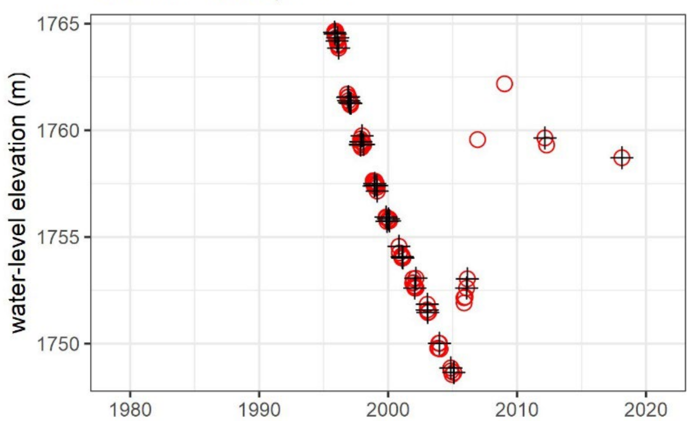
S1a NM-02209; n=99

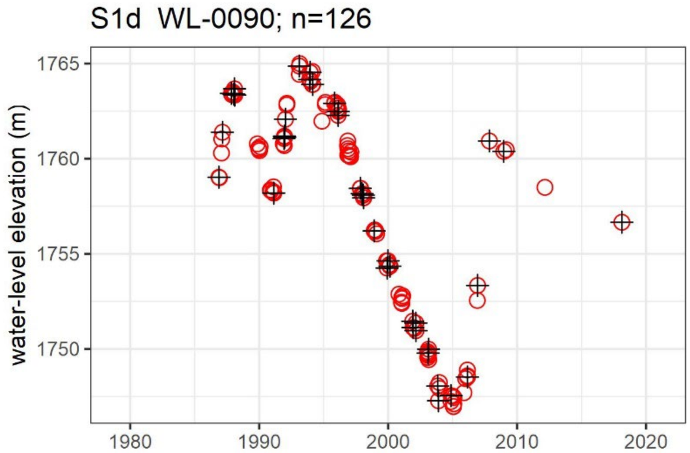


S1b WL-0086; n=187



S1c WL-0088; n=94





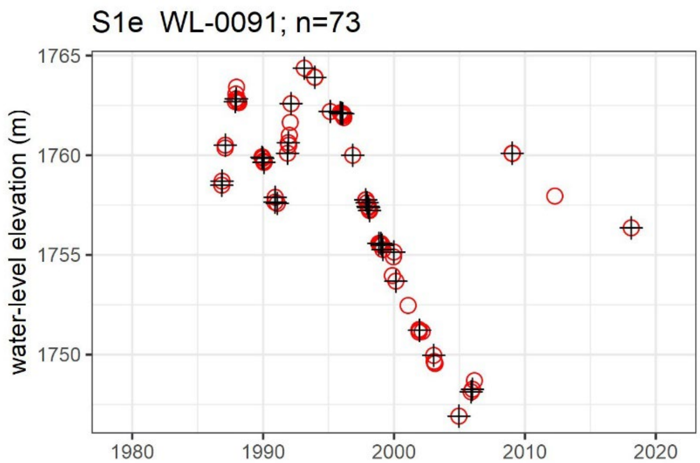


Fig. S1a-e. Hydrographs of wells with downsampled data (black crosses). *n* = number of original measurements.

Section S2. Regional trend surface 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 (all in meters) was transformed into three principal components (Table S1):

Table S1. 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.

$$\mathrm{PC1} = -0.53 \times \text{easting} + 0.57 \times \text{northing} + 0.63 \times \text{land surface elevation} \qquad \text{(A2.1)}$$

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:

$$a_1 + a_2x + a_3y + a_4x^2 + a_5xy + a_6y^2 + a_7x^3 + a_8x^2y + a_9xy^2 + a_{10}y^3 \qquad \text{(A2.2)}$$

where PC1 = *x* and PC2 = y. The coefficients are shown in Table S2 and maps of PC1, PC2 the regional trend model, and its standard error are shown in Figure S2.

Table S2. Coefficients and statistics of the regional trend model.

coefficient	Value	Standard error	t-value	Pr(>|t|)
a_1	1282.47	0.57	2253.00	< 2e-16
a_2	63.73	0.43	148.92	< 2e-16
a_3	18.36	0.96	19.22	< 2e-16
a_4	10.70	0.23	46.83	< 2e-16
a_5	9.17	0.55	16.67	< 2e-16
a_6	6.66	0.58	11.55	< 2e-16
a_7	-0.51	0.056	-8.99	< 2e-16
a_8	-1.32	0.19	-6.97	3.77e-12
a_9	3.10	0.36	8.55	< 2e-16
a_{10}	2.33	0.40	5.85	5.47e-09
Multiple r^2	0.99		Adjusted r^2	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 deviation. 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. 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. 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. 2007).

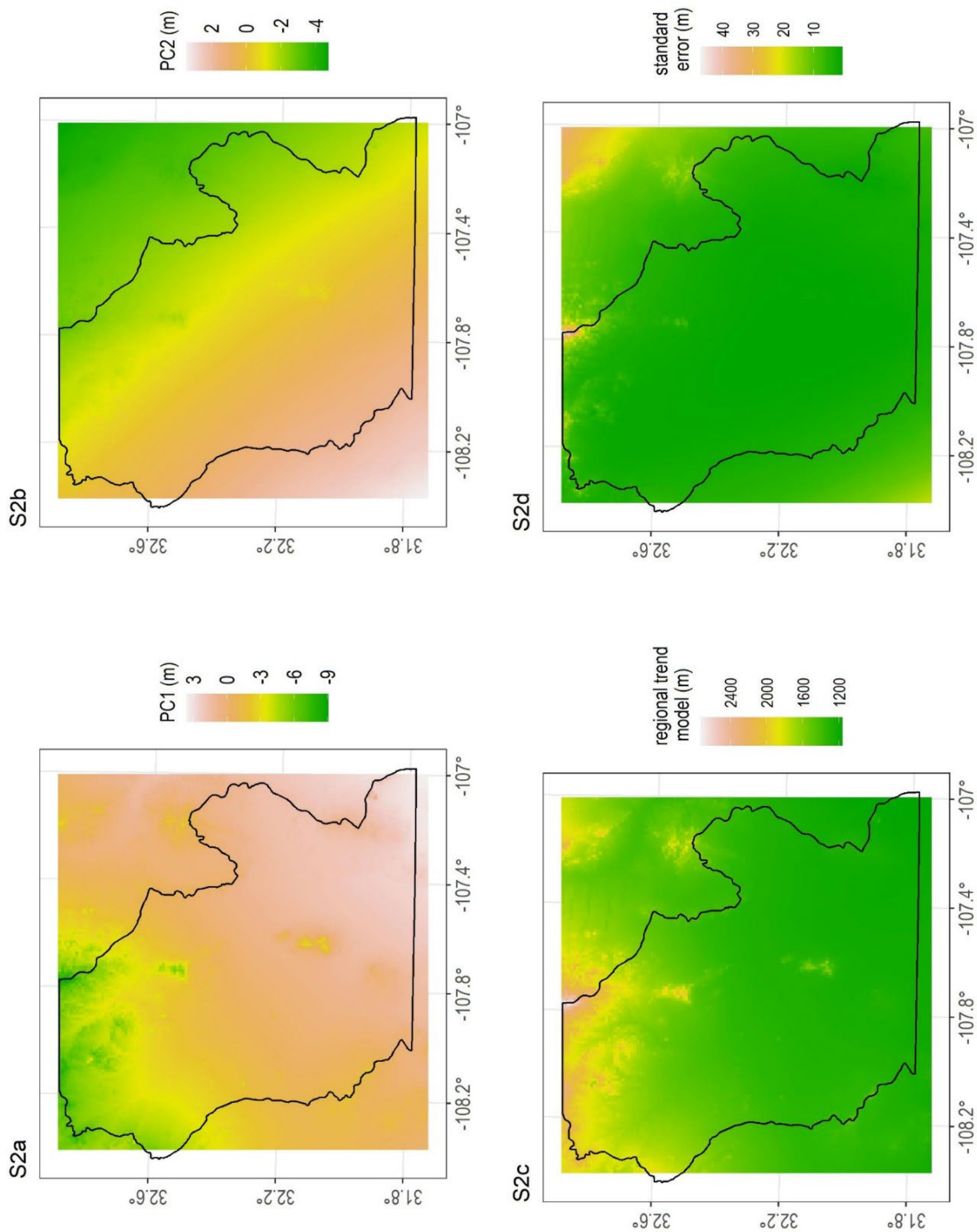


Fig. S2 a-d Maps of the first (PC1) and second (PC2) principal components, the regional water-level trend surface model, and the standard error of the trend surface.

Section S3. Trend surface residuals

Fig. S3 shows the residuals from the regional water-level trend surface before removal of outlier measurements using the criteria described in Section *'Regional Trend Model'* of the main article. Well NM-05236 southeast of Deming (one measurement) was also dropped. The standardized residual for this measurement was 2.94 and Cook's Distance was 0.010, however, it clearly plots with the other outlier measurements. Outliers are shown in red in parts b and c of the figure.

Fig. S4 shows the residuals from the regional water-level trend surface after removal of outlier measurements.The residuals from the regional water-level trend surface form a minimally skewed, approximately normal distribution with near zero mean after removal of the outliers (Fig. S4a and b). 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) (Fig. S4c). The positive and negative clusters result in the double peak in the histogram (Fig. S4a) but the Q-Q plot (Figure Fig. S4b) shows that the distribution of residuals after removal of outliers is very nearly normal.

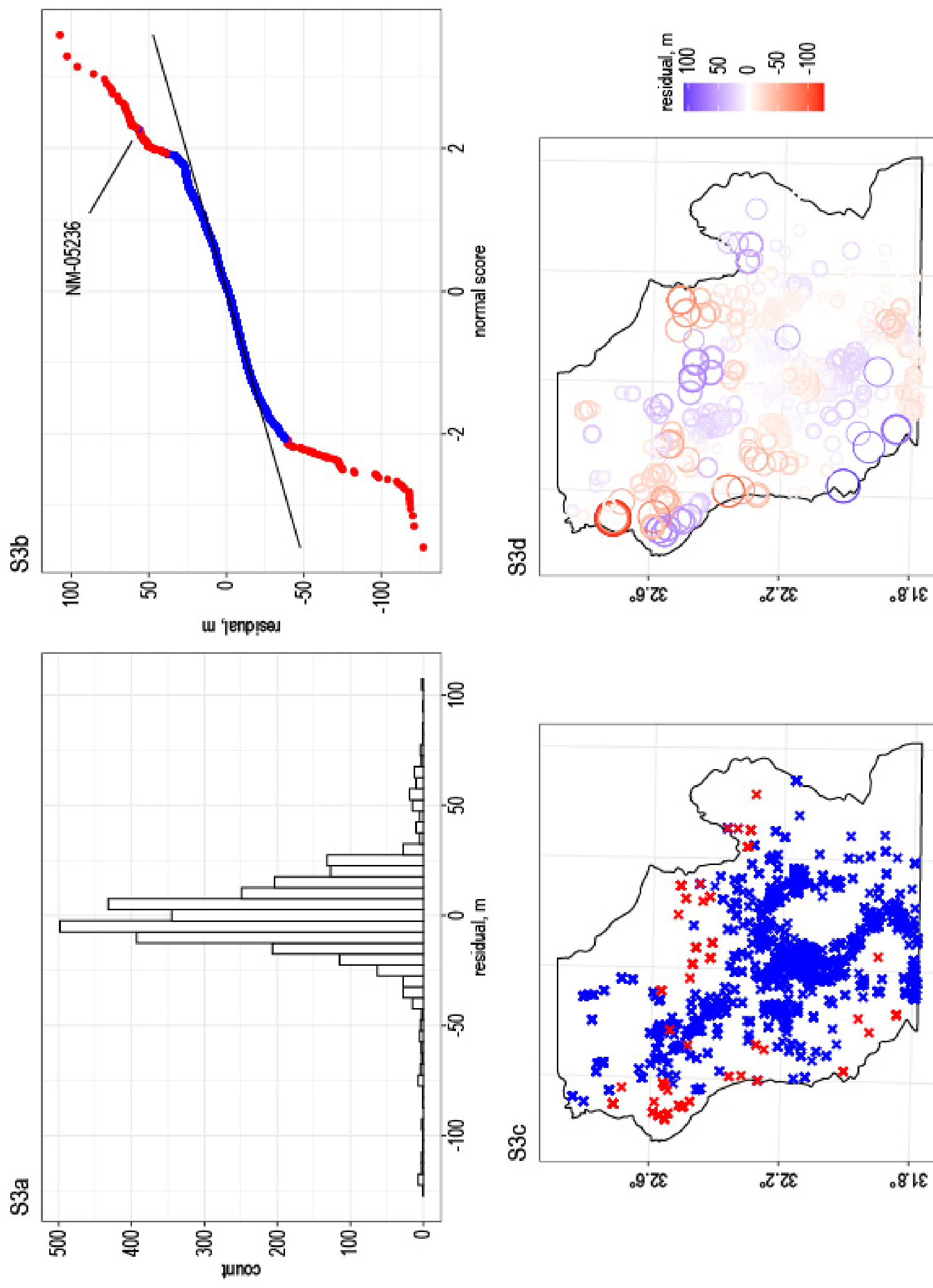


Fig. S3 **a** histogram of trend surface residuals; **b** QQ plot of trend surface residuals (outliers in red); **c** wells with at least one outlier measurement (red); **d** Spatial distribution of trend surface residuals. Predictions above the water level are positive (blue) and below the water level are negative (red).

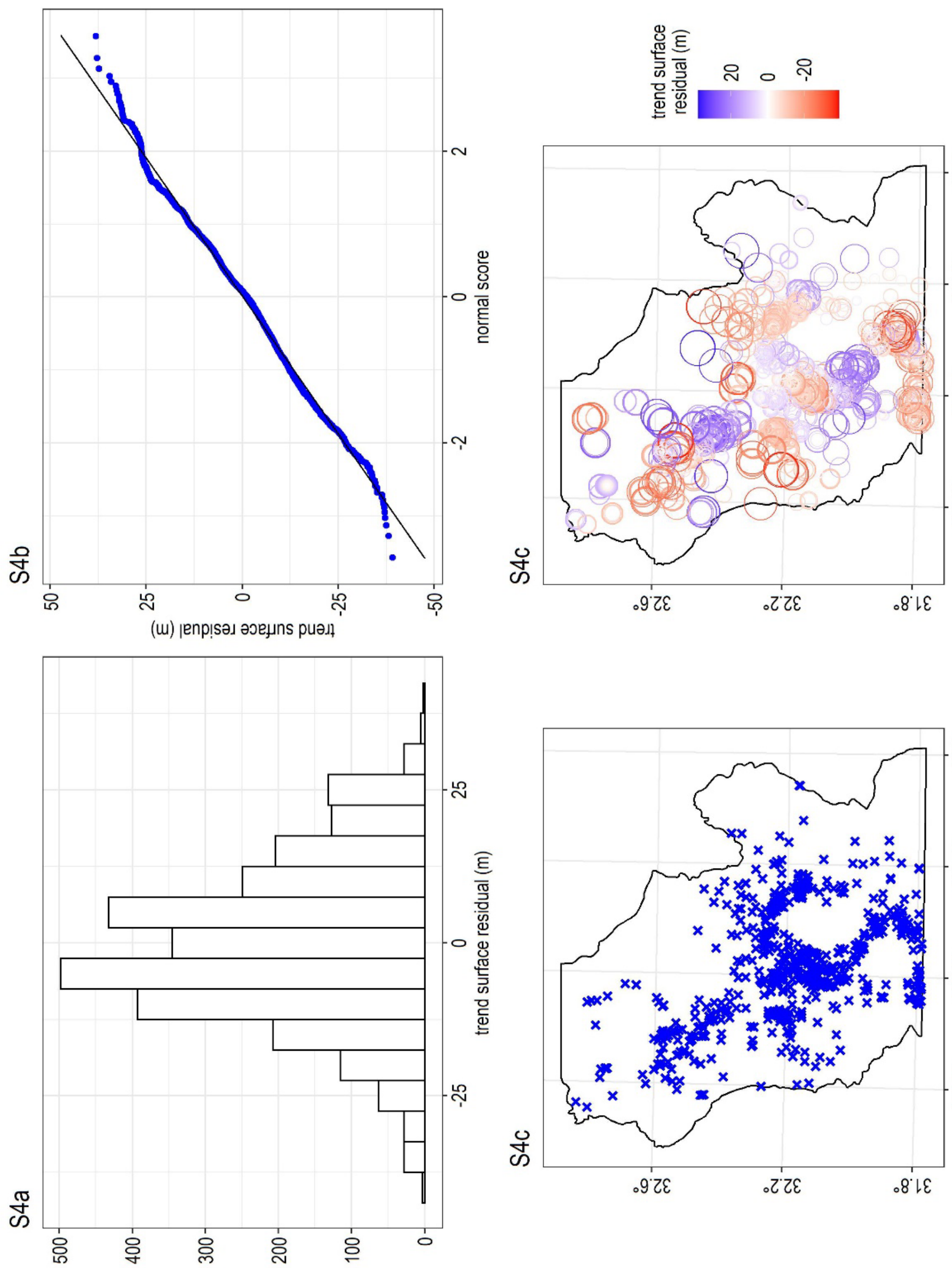


Fig. S4 a histogram of trend surface residuals after removal of outliers; **b** QQ plot of trend surface residuals after removal of outliers; **c** wells with only outlier measurements have been removed; **d** Spatial distribution of trend surface residuals. Predictions above the water level are positive (blue) and below the water level are negative (red). Note different scale than S3d.

Section S4. Spatiotemporal variogram model-fitting

The Sum-metric model was chosen for the spatiotemporal variogram based on the combination of the model-fitting parameters and the slightly better cross-validation statistics when compared to the Simple sum-metric model.

Table S3. Model-fitting parameters for spatiotemporal variogram models. Better values are in italics.

	Metric	Product sum	Sum-metric	Separable	Simple sum metric
optim	0.034	0.0062	*0.0016*	0.0046	0.0020
mse	342.45	177.34	*74.73*	96.63	74.74

Table S4. Leave-one-out cross-validation statistics of spatiotemporal variogram models. Better values are in italics.

	Metric	Product sum	Sum-metric	Separable	Simple sum metric
Mean error	*-0.027*	-0.055	-0.059	-0.073	-0.060
Mean squared error	9.62	12.97	9.36	9.38	*9.30*
Correlation between observed and predicted	*0.97*	0.96	*0.97*	*0.97*	*0.97*
Correlation between predicted and residual	-0.248	*-0.14*	-0.21	-0.162	-0.23
Variance of residuals	9.62	12.97	9.36	9.37	*9.30*
Skewness	-0.79	*-0.49*	-0.91	-0.93	-0.97
Maximum	25.68	24.74	*18.07*	27.10	19.06
Minimum	-29.08	-33.53	-29.63	*-18.03*	-30.46

Section S5. Comparison of global and regional variogram models

The global sum-metric model was chosen to represent the spatiotemporal correlation structure over the entire region, as opposed to using local models for each hydrgeologic region shown in Fig. 2. The global model fits the global experimental variogram better than any local model fits the local experimental variograms (Table S5). Cross validation statistics offer no compelling argument to choose the added complexity of a local model over the global model in any of the hydrogeologic zones (Table S6).

Table S5. Model-fitting parameters for best-fitting global model and models specific to each and hydrogeologic zone. Better values are in italics.

	Global, sum-metric	Deming, sum-metric	San Vicente, sum-metric	Florida, metric
optim	*0.0016*	0.0054	0.0037	0.0079
MSE	*74.73*	620	6828	702

Table S6. Cross-validation statistics for comparison of performance of local and global variogram models in each hydrogeologic zone. Better values are in italics.

	Deming		San Vicente		Florida	
	Global	Local	Global	Local	Global	Local
Mean	-0.018	*-0.007*	-0.061	*-0.055*	-0.11	*-0.079*
Mean squared error	**7.77**	8.54	*10.30*	17.94	12.10	*11.97*
Mean squared normalized error	0.999	0.999	0.998	0.998	0.999	0.999
Correlation between observed and predicted	0.97	0.97	*0.98*	0.96	0.96	0.96
Correlation between predicted and residual	*-0.18*	-0.24	*-0.23*	-0.25	*-0.26*	-0.29
Variance of residuals	*7.77*	8.54	*10.31*	17.97	12.10	*11.97*
Skewness	*0.26*	0.30	-1.62	*-0.84*	-1.33	-1.33
Minimum of residuals	-17.57	*-15.84*	*-29.72*	-30.12	*-24.51*	-27.51
Maximum of residuals	*17.19*	17.52	*16.82*	22.13	*18.07*	26.38

Section S6. Comparison of spatial and spatiotemporal kriging for years 2012 and 2020

Table S7. Cross-validation statistics for comparison of spatial and spatiotemporal kriging. Better values are in italics.

	spatial 2012	spatiotemporal 2012	spatial 2020	spatiotemporal 2020
Mean	*0.032*	0.11	0.294	*-0.22*
Mean squared error	133.88	*16.31*	106.57	*32.88*
Mean squared normalized error	1.24	*0.99*	1.42	*0.92*
Correlation between observed and predicted	0.522	*0.96*	0.71	*0.92*
Correlation between predicted and residual	-0.027	*0.004*	*0.0090*	-0.036
Variance of residuals	135.03	*16.44*	107.61	*33.17*
Standard deviation of residuals	11.62	*4.06*	10.37	*5.76*
Skewness	*0.62*	-0.63	*0.29*	-1.92
Minimum of residuals	-33.69	*-22.31*	*-22.84*	-29.62
Maximum of residuals	43.03	*15.43*	31.27	*15.39*

7. Data tables

See Table S8 in file ESM2.

8. ESM References

Hengl T, Heuvelink GB, Rossiter DG (2007) About regression kriging: From equations to case studies, Computers and Geosci. 33:1301–1315, doi 10.1016/j.cageo.2007.05.001

Rawling GC (2021) Evaluation of water-level trends using spatiotemporal kriging in the Mimbres Basin southwest New Mexico. NM Bur Geol Min Resour Open-File Rep 616

Schuenemeyer JH, Drew LJ (2011) Statistics for Earth and Environmental Scientists. John Wiley & Sons

Wikle CK, Zammit-Mangion A, Cressie N (2019) Spatio-Temporal Statistics with R. Chapman & Hall/CRC