Supplementary Information

A century of groundwater accumulation in Pakistan and northwest India

MacAllister et al.

Supplementary Figures

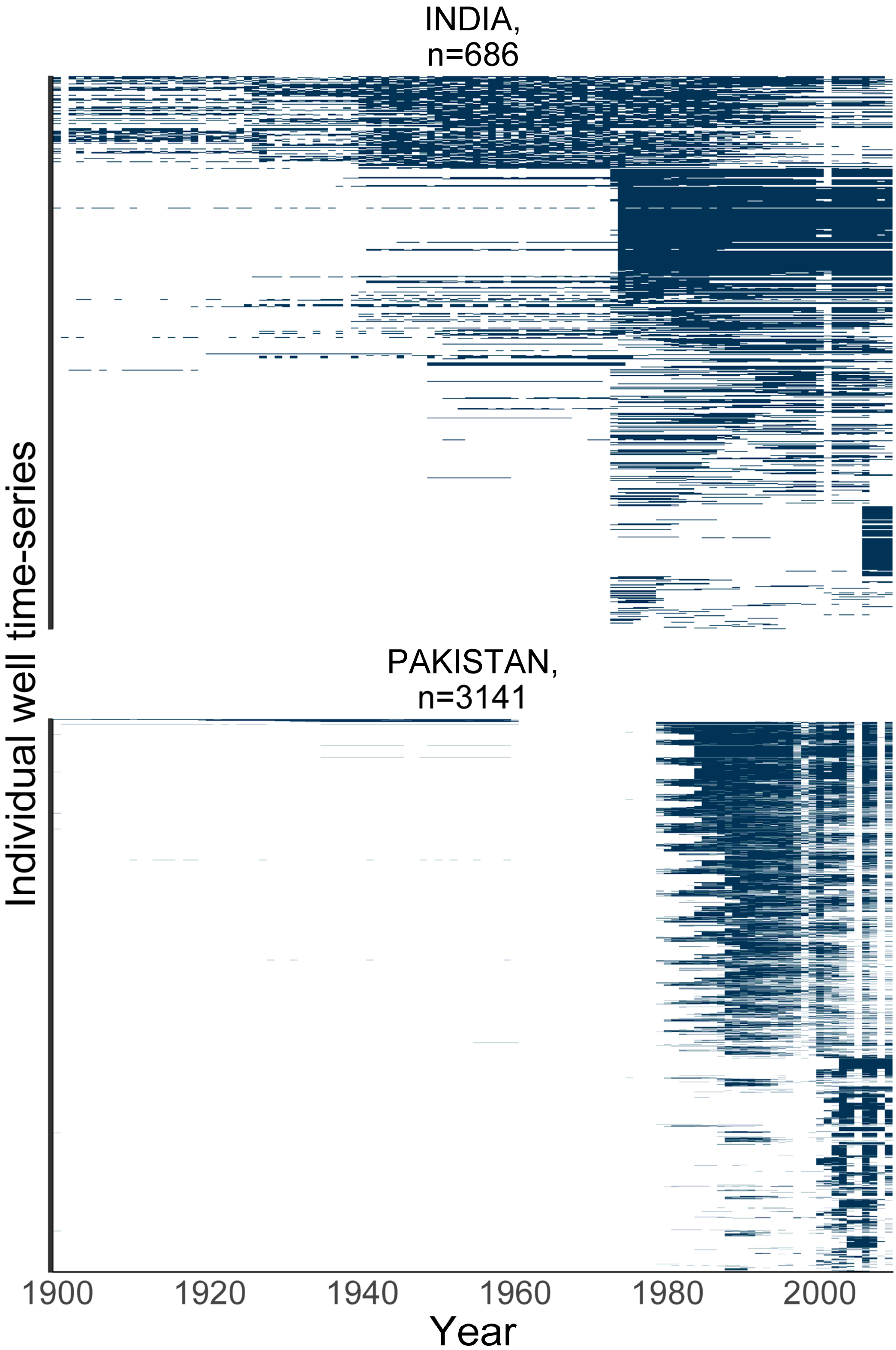

Figure 1 – *The full observation well time-series dataset for both India and Pakistan. Note that the large difference in sample size obscures the quantity of older data available for the analysis in Pakistan. We present all individual time-series in this format so that the nature of the dataset is clear and transparent. Supplementary fig 2 shows that the quantity of historic data available in India and Pakistan is comparable.*

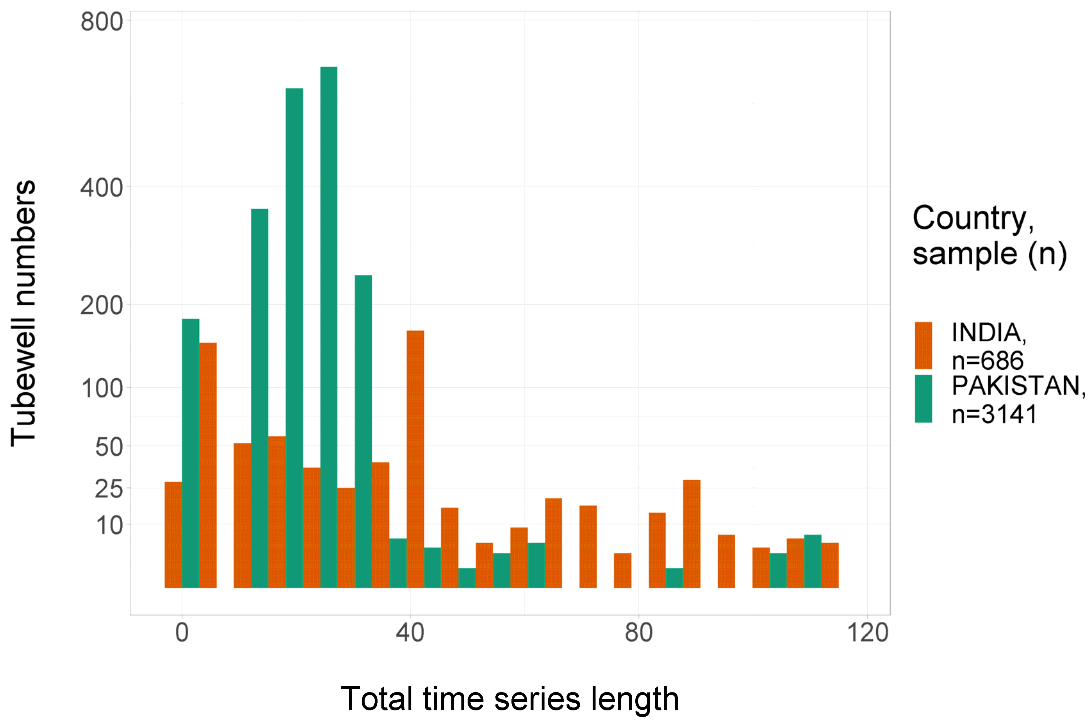

Figure 2 – Histogram showing distribution of tubewell time-series lengths.

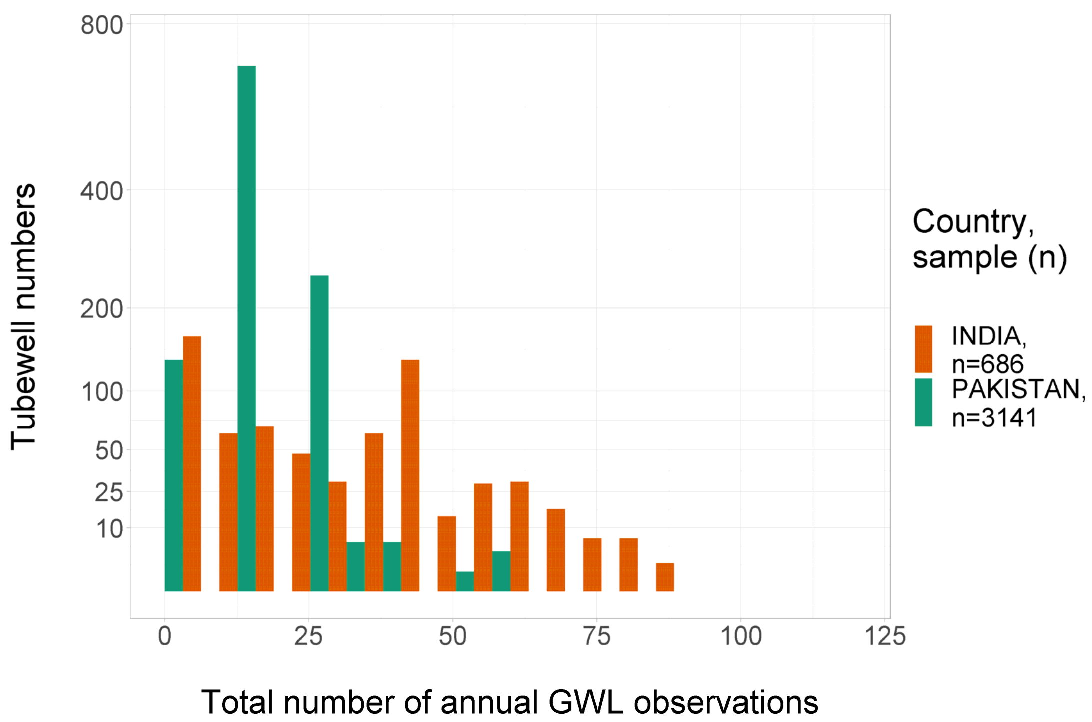

Figure 3 – Histogram showing distribution of total number of time-series annual groundwater level (GWL) observations.

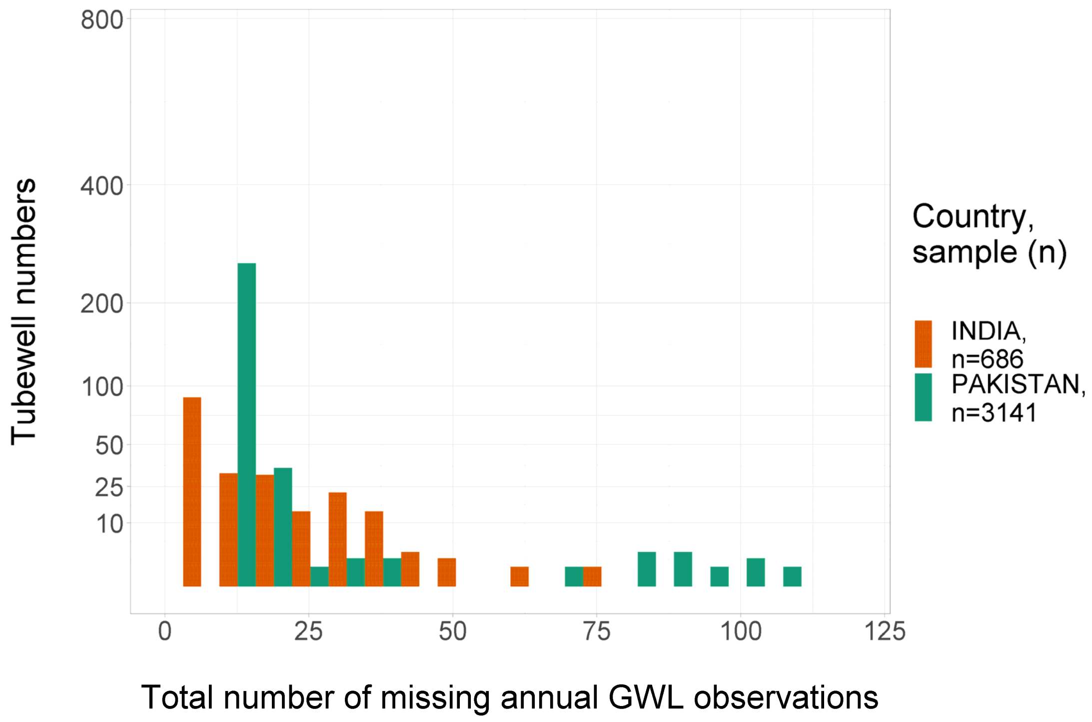

Figure 4 – Histogram showing distribution of total number of missing time-series annual groundwater level (GWL) observations.

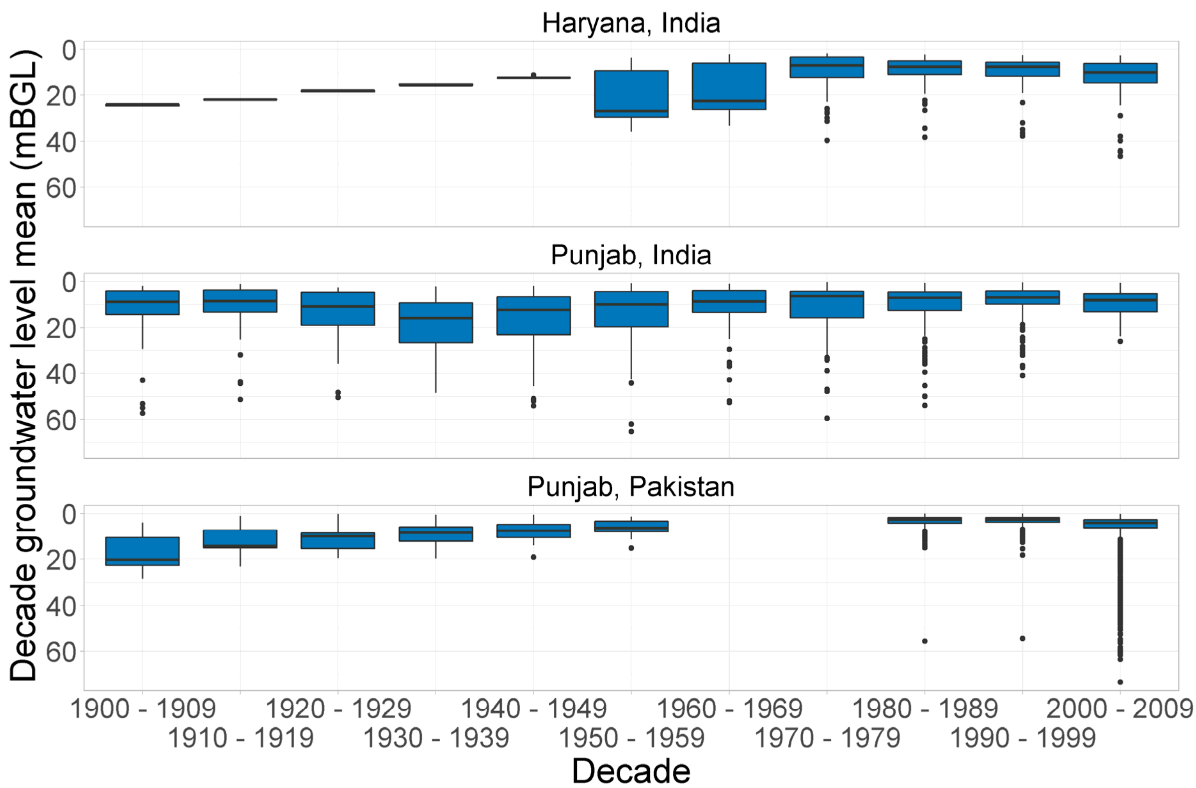

Figure 5 – Mean water level in each state and decade. The interquartile range is shown by the upper and lower bounds of the blue box, the median is represented by the black line and the mean by the black point in the blue box, the whiskers represent the interquartile range times 1.5 and the points represent outliers.

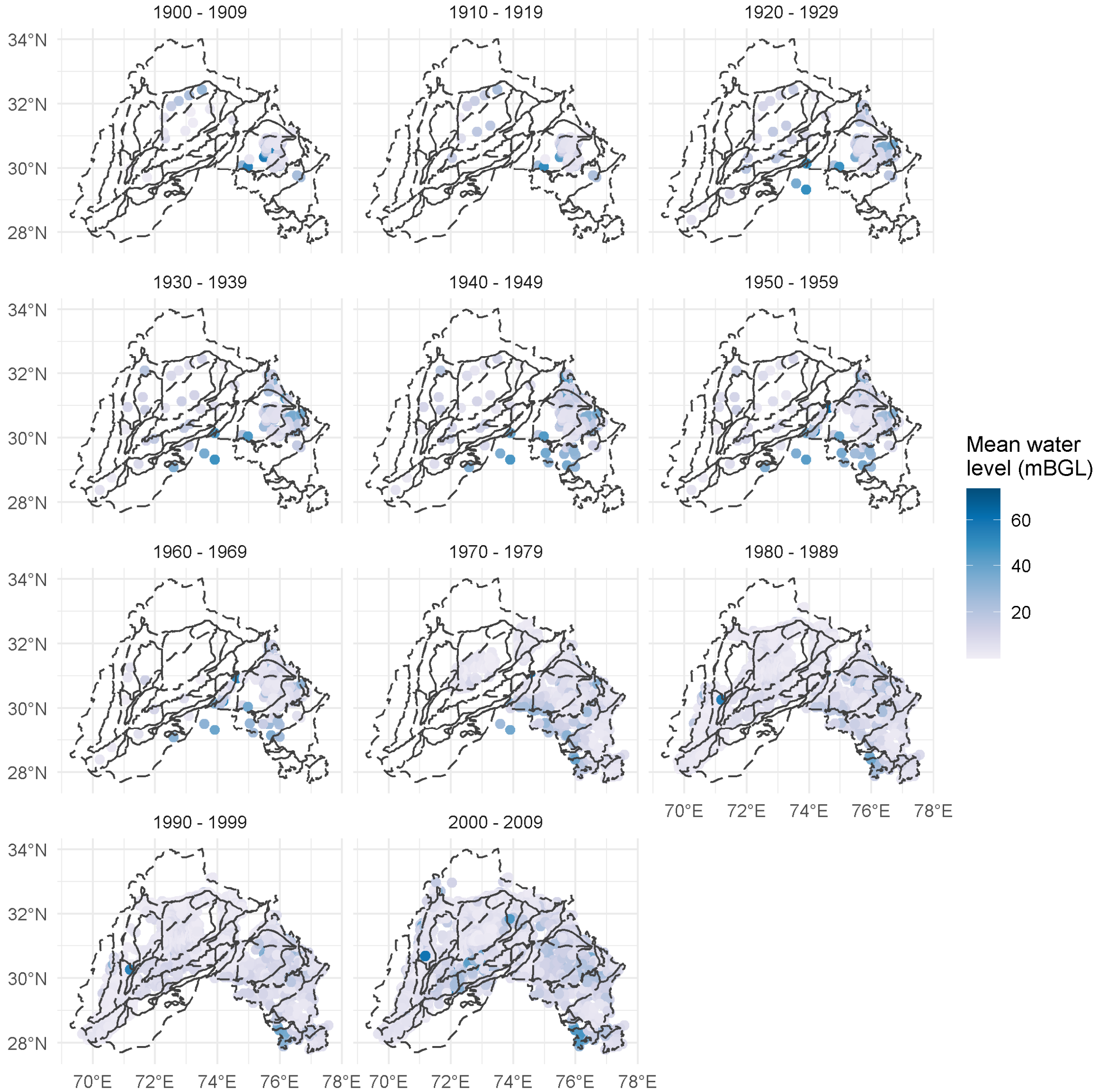

Figure 6 – Map of mean water level depth for individual observation wells in each decade.

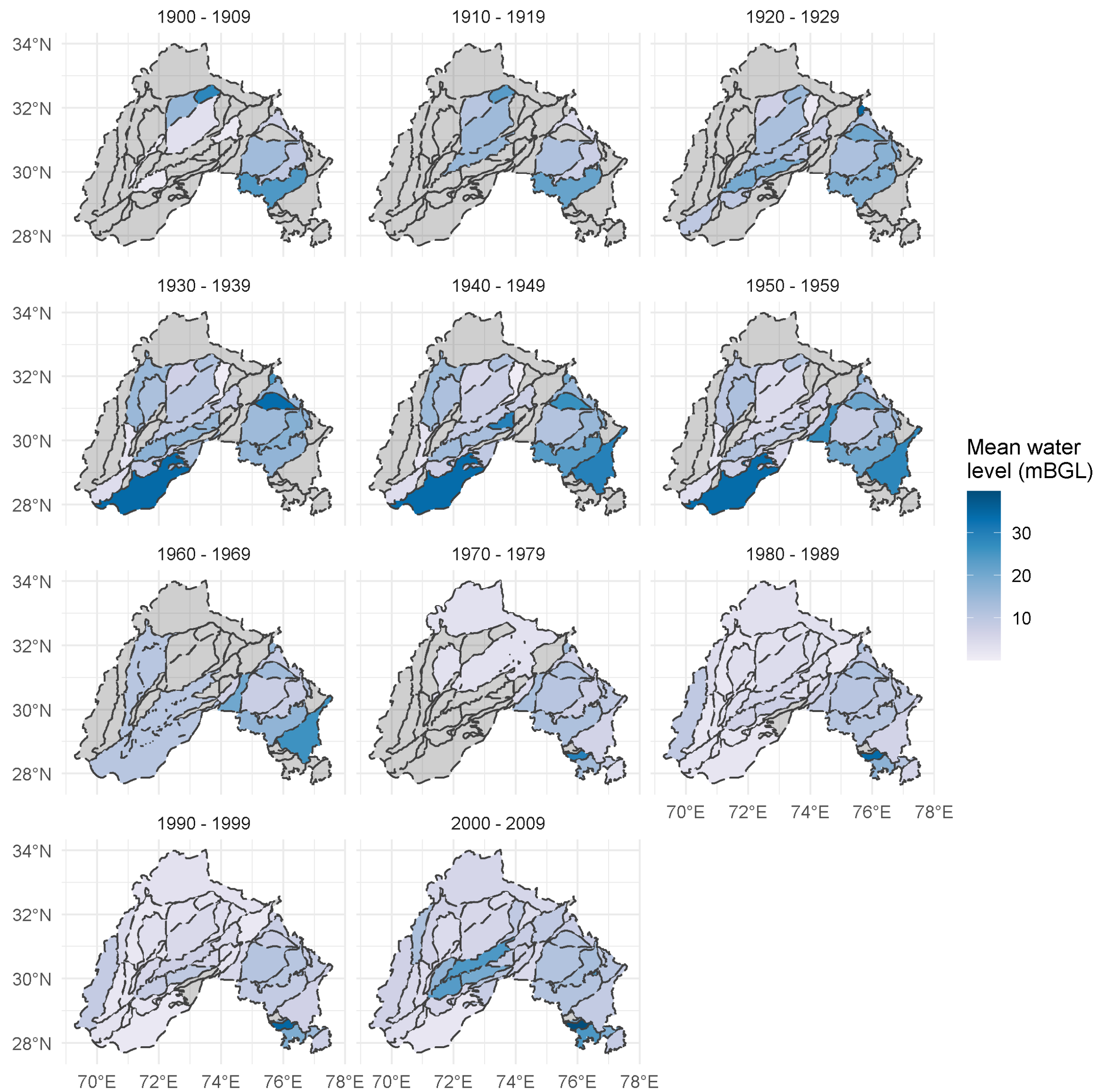

Figure 7 – Map of mean water level depth for canal commands in each decade.

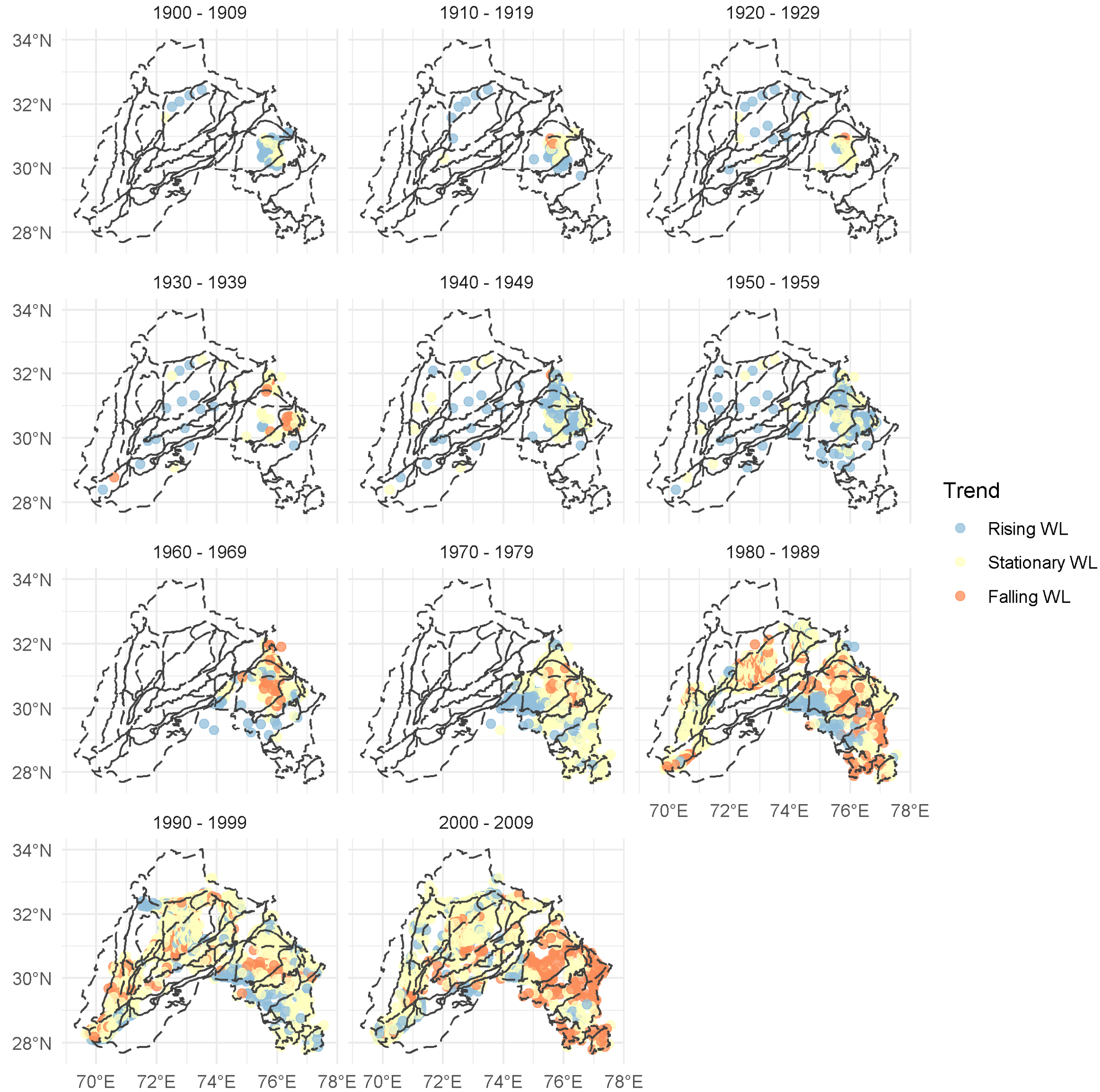

Figure 8 – Tubewell Mann-Kendall trend results using a minimum of six data points in a decade, as shown in the main article but presented here for easy comparison with Supplementary Fig 5 which shows the Mann-Kendall trend results using a minimum of four data points in a decade and Supplementary Fig 6 which shows the Mann-Kendall trend results using a minimum of eight data points in a decade.

.

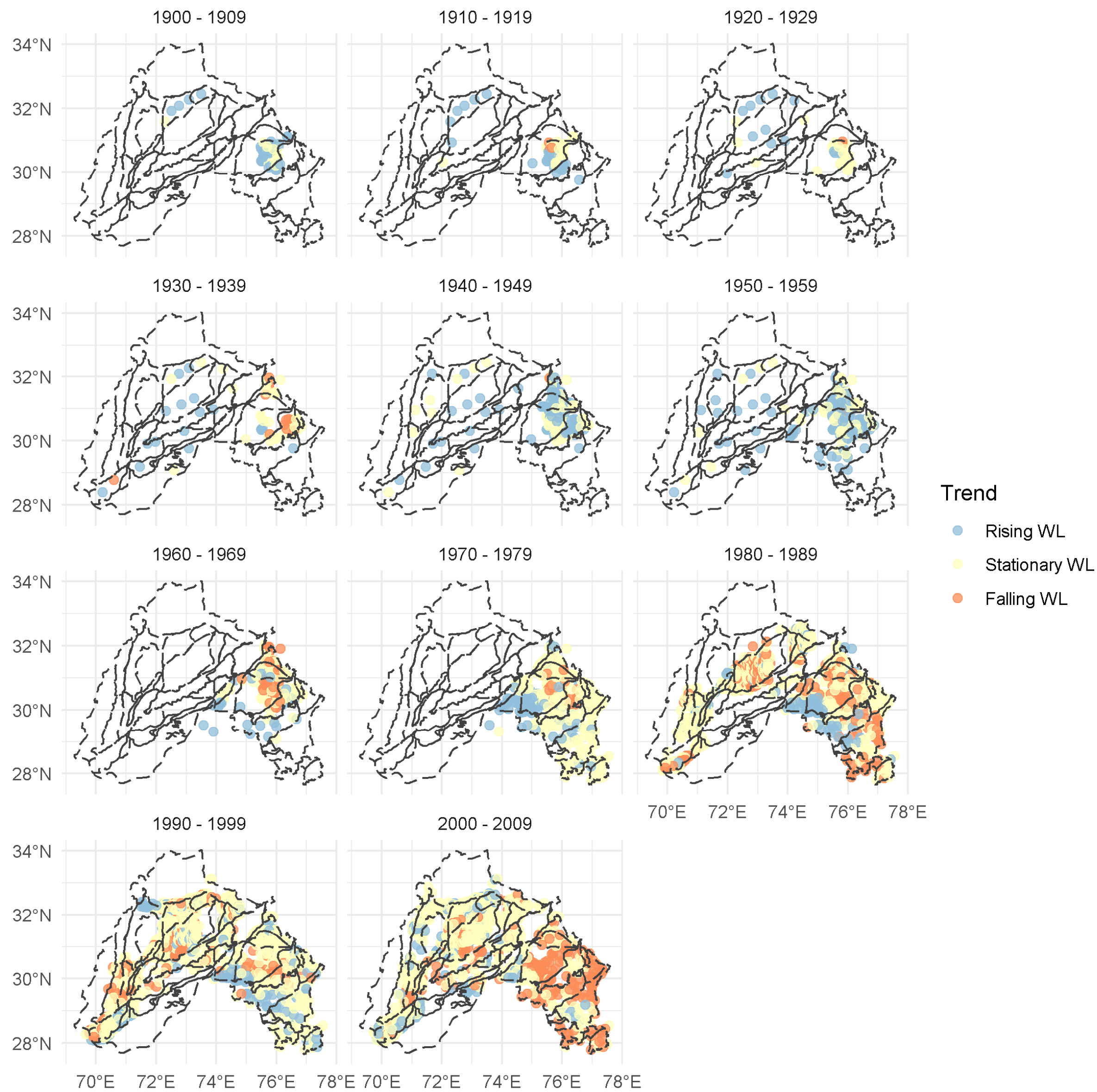

Figure 9 – Tubewell Mann-Kendall trend results using a minimum of four data points in a decade.

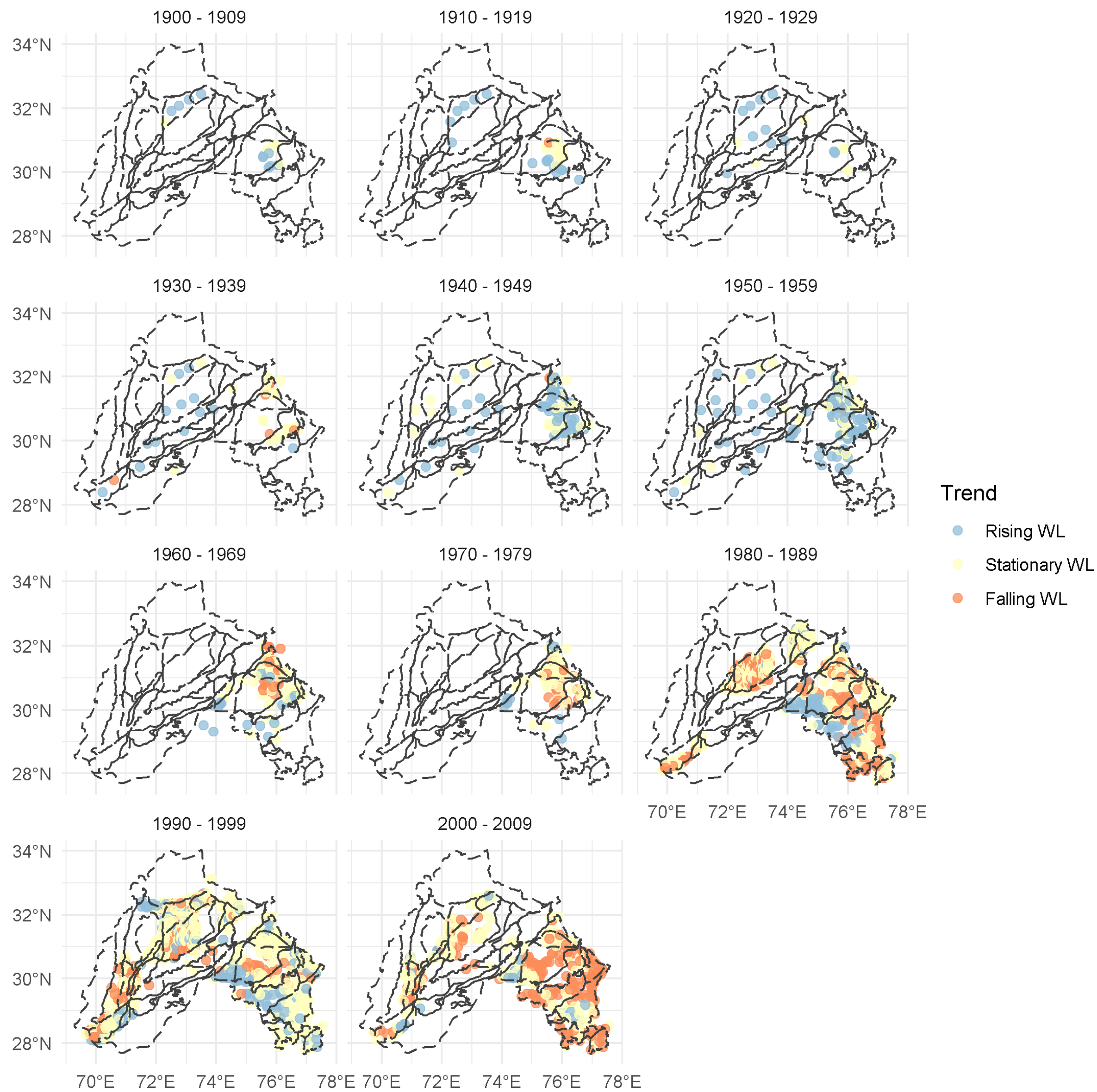

Figure 10 – Tubewell Mann-Kendall trend results using a minimum of eight data points in a decade.

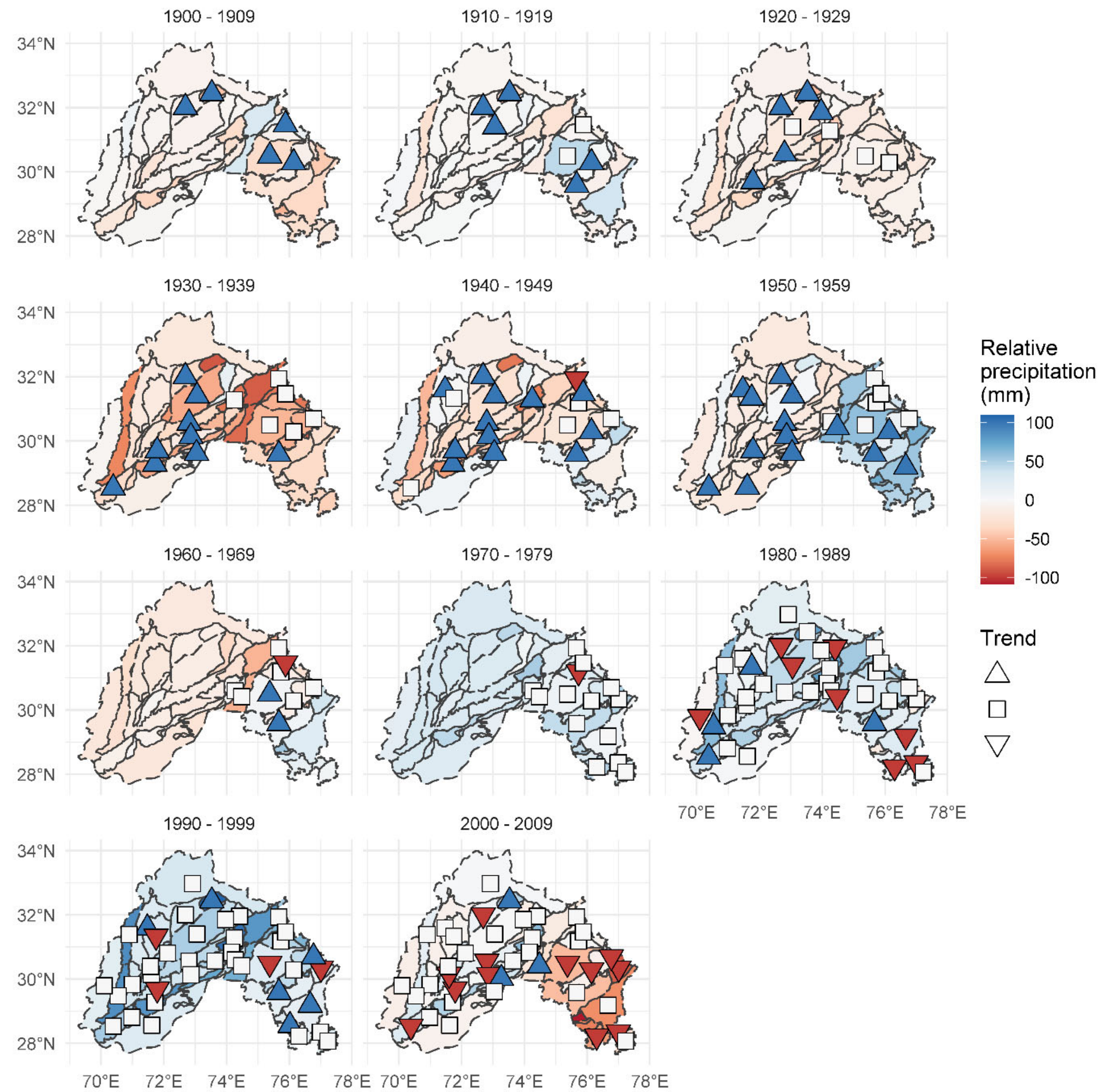

Figure 11 – Deviation from mean precipitation and mode (> 50% of tubewells) Mann-Kendall water level trend in each canal command.

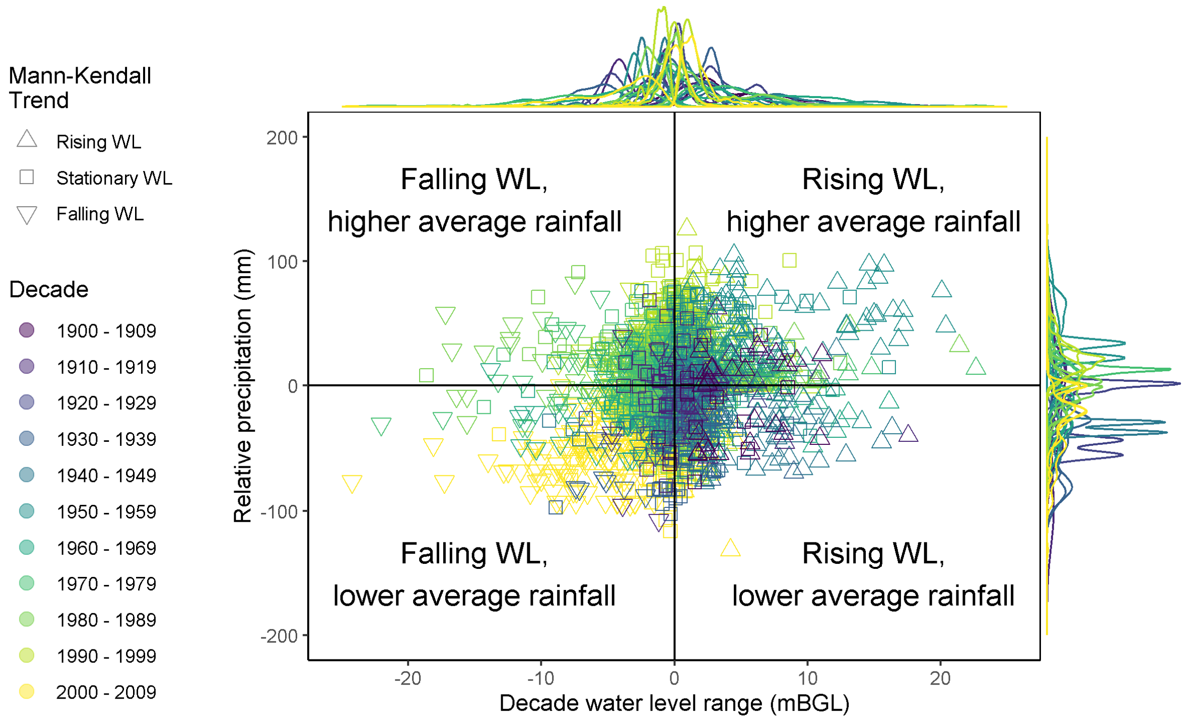

Figure 12 – Relative precipitation versus water level change at the start and end of each decade. Individual tubewell observations are coloured by the decade the canal command they are located in was constructed. The histograms on the top and the left of the figure panel show the underlying distribution. Shapes indicate the Mann-Kendall results for individual tubewells in each decade and provide confidence in our use of water level change at the start and end of each decade in the analysis. Mann-Kendall rising water level trends are generally found in the right hand site of the plot which relates to sites with shallower water levels at the end of the decade than at the start and vice versa.

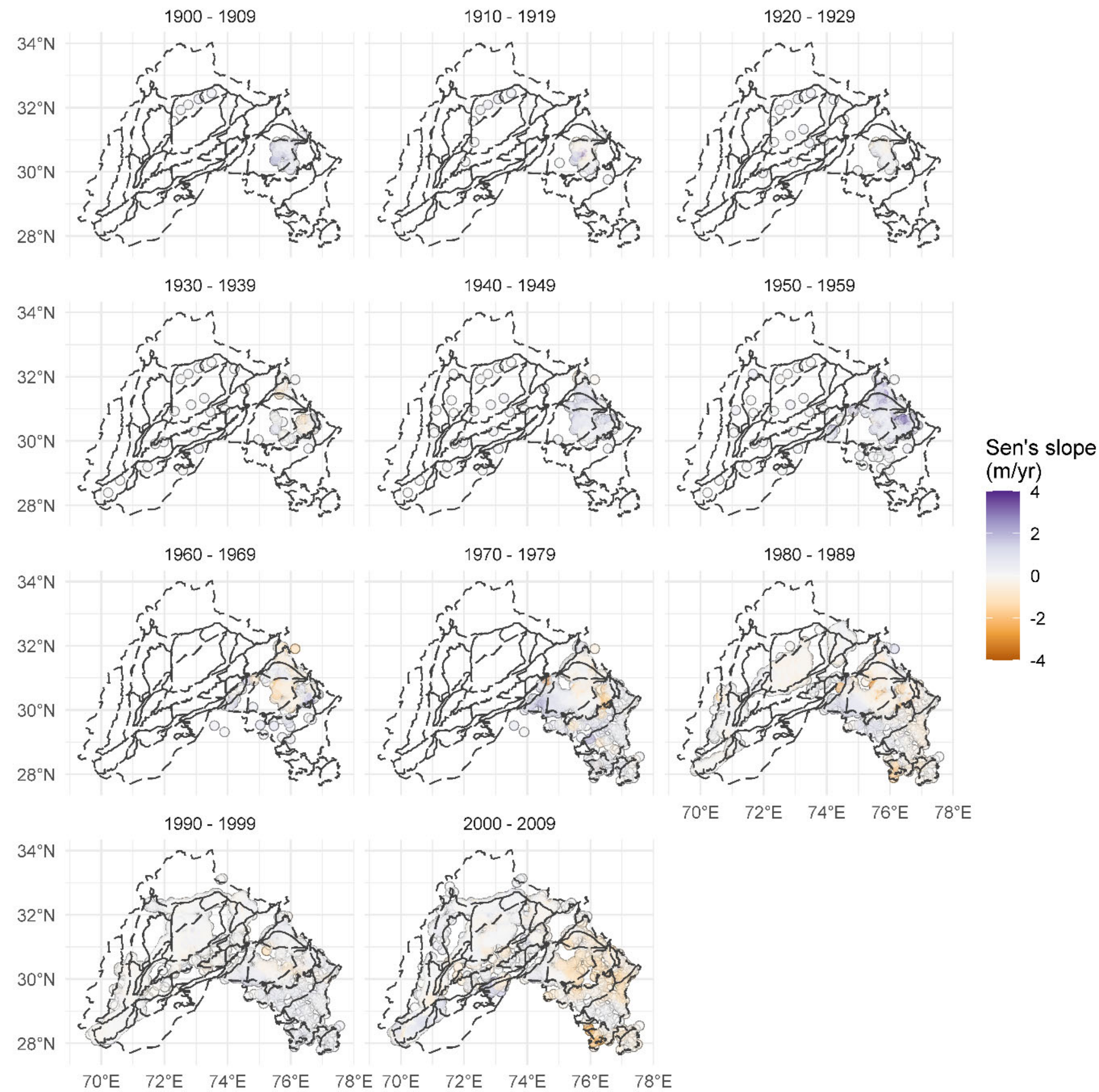

Figure 13 – Sen's slope estimator results using six data points in a decade.

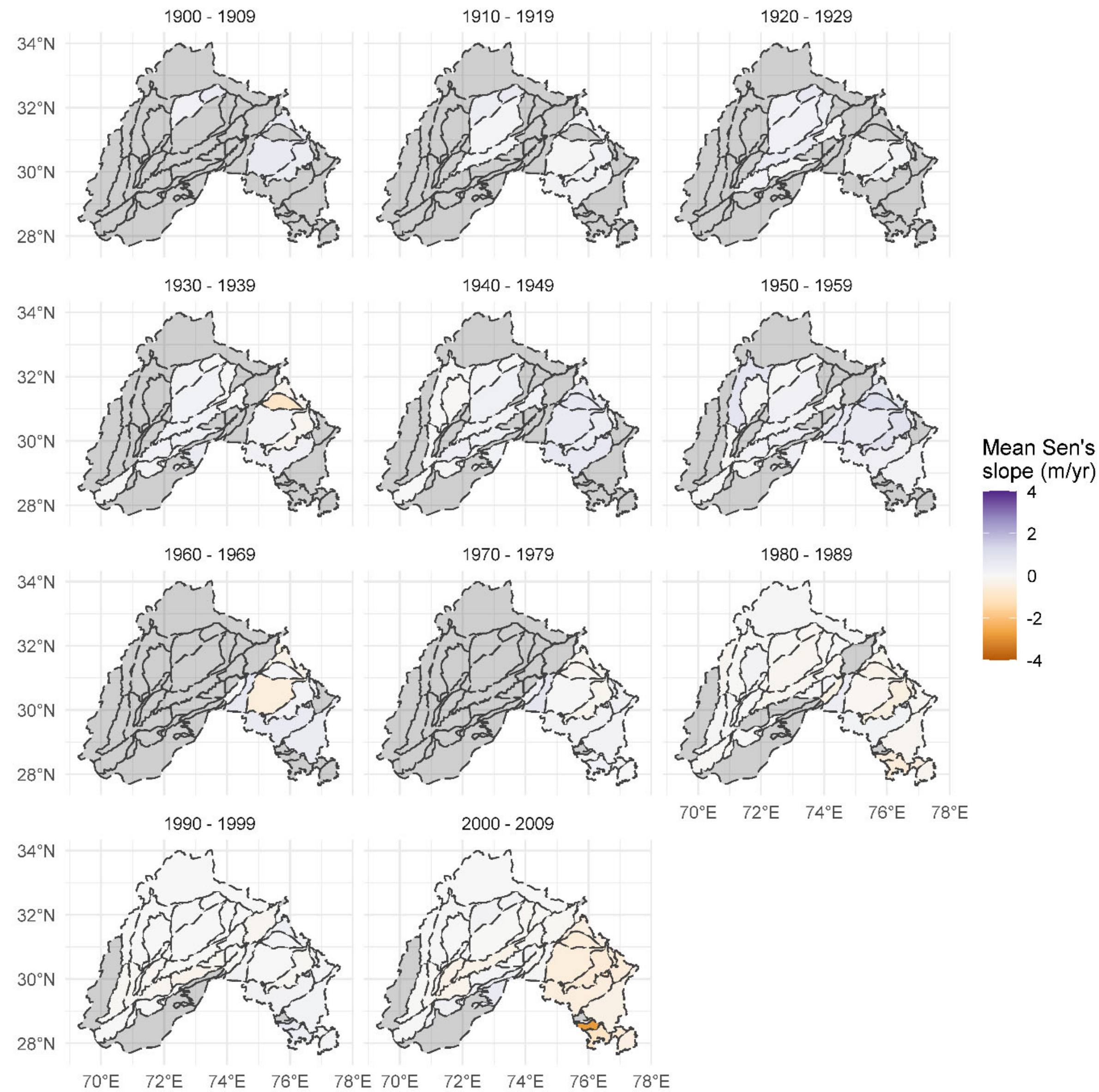

Figure 14 – Mean canal command Sen's slope estimator results based on six data points in a decade and used to calculate groundwater level accumulation in the study area.

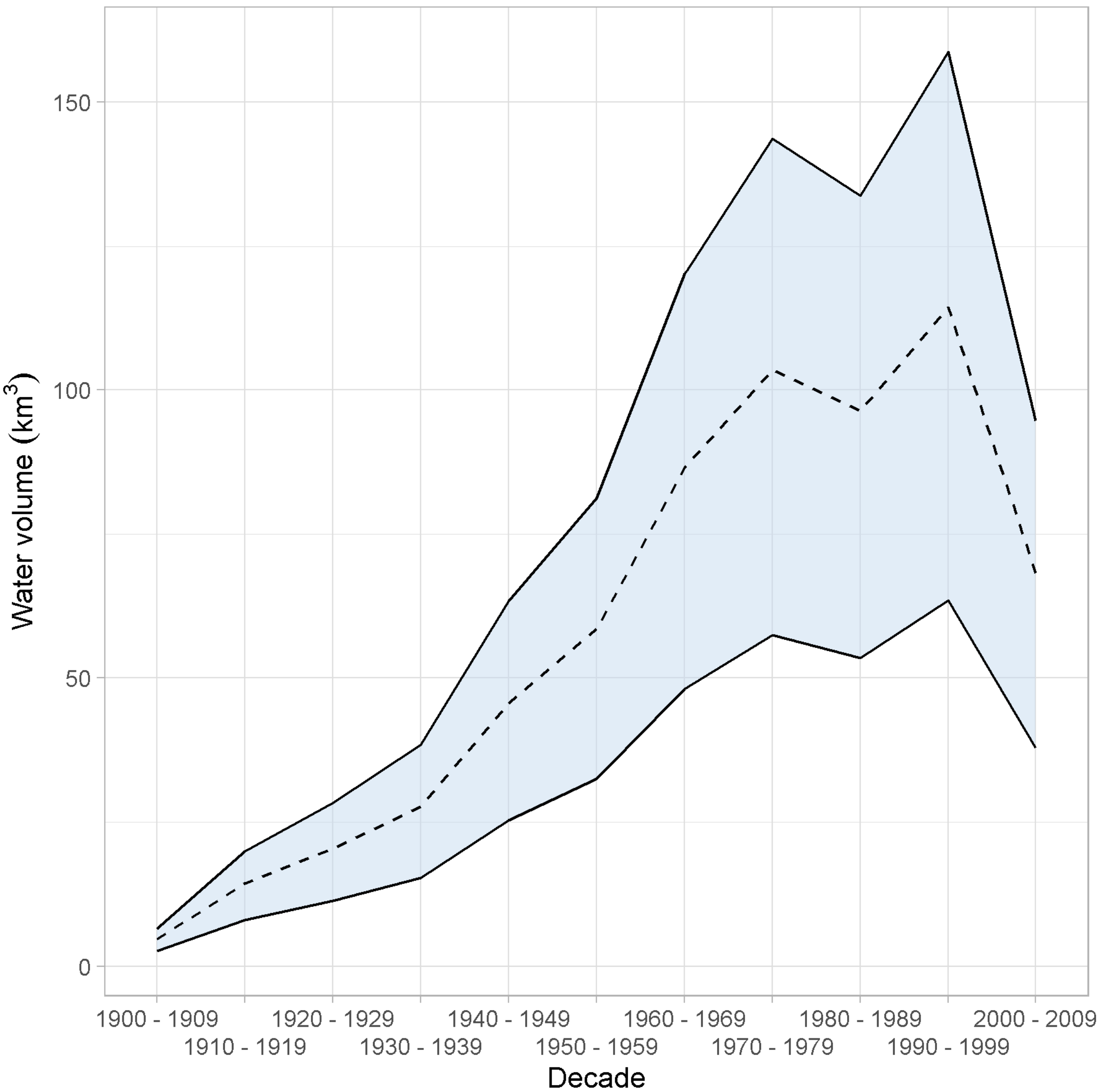

Figure 15 – *Estimated groundwater accumulation in Haryana, India. Note only two sites, which were in close proximity, were available to estimate accumulation in Haryana between 1900 – 1939, so the estimates in this period only reflect localised groundwater accumulation in a limited area of north-east Haryana (see fig 6 for location of these sites) rather than a average of accumulation across Haryana.The blue envelope shows the estimated range of accumulation, the dashed line shows the best estimate, the porosity values used to calculate groundwater accumulation were based on previous work in the study area.*

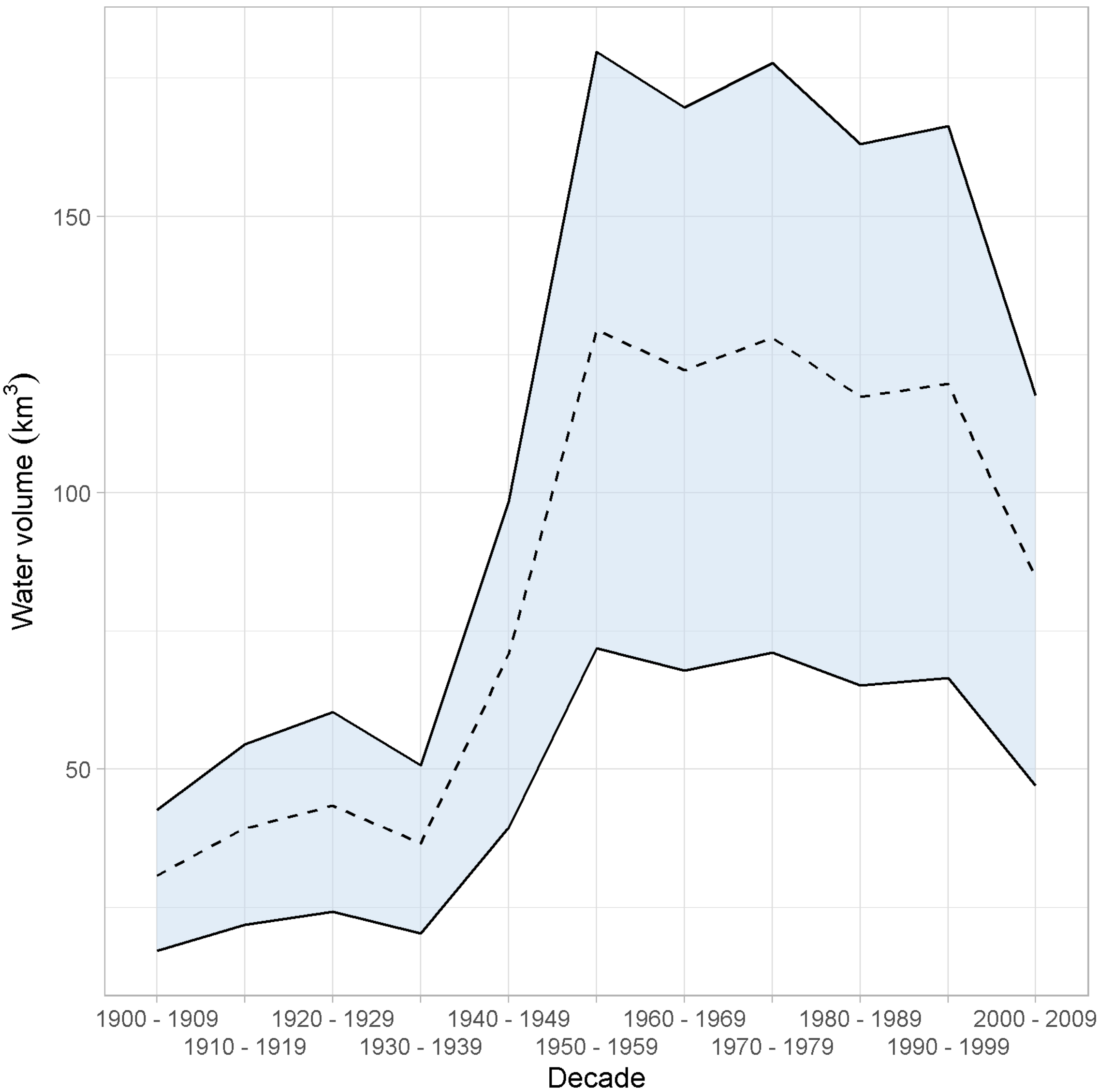

Figure 16 – Estimated groundwater accumulation in Punjab, India. The blue envelope shows the estimated range of accumulation, the dashed line shows the best estimate, the porosity values used to calculate groundwater accumulation were based on previous work in the study area.

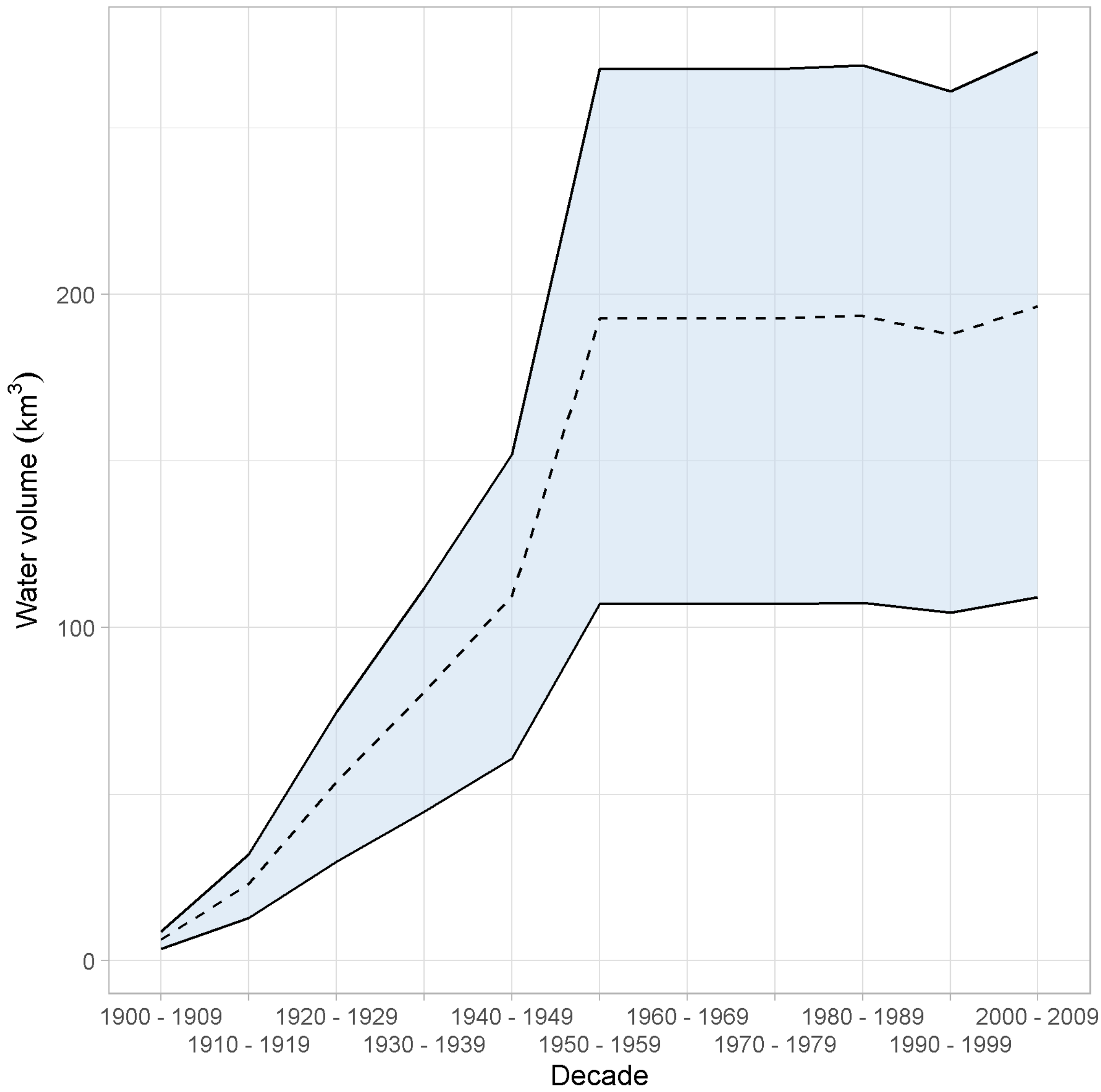

Figure 17 – *Estimated groundwater accumulation in Punjab, Pakistan. Note that there were no data with which to estimate accumulation in 1960 – 1969 or 1970 – 1979. So these decades do not contribute to our accumulation estimates. The blue envelope shows the estimated range of accumulation, the dashed line shows the best estimate, the porosity values used to calculate groundwater accumulation were based on previous work in the study area.*

Supplementary Notes

Note 1: Mann – Kendall Trend Test for Autocorrelated Data

The Mann-Kendall test is a non-parametric hypothesis test based on ranked data that aims to identify a monotonic trend, linear or non-linear, in a time series. The test is based on the S statistic defined as:

$$S = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \operatorname{sgn}(Y_j - Y_i),$$

Here N is the total number of observations in the time series; Y_i and Y_j are sequential observations (ranks) of the data Y, and:

$$\operatorname{sgn}(\theta) = 1 \text{ if } \theta > 0;\ = 0 \text{ if } \theta = 0; = -1 \text{ if } \theta < 0.$$

A positive value of S is indicative of a positive trend; conversely, a negative value of S indicates a negative trend in the data. In the absence of trend and when $N \geq 8$, the statistic is approximately normally distributed[1,2] with mean:

$$E[S] = 0$$

and variance:

$$\operatorname{var}(S) = \frac{N(N-1)(2N+5)}{18}.$$

To perform the hypothesis test, the Z standardised test statistic is used, which follows the standard normal distribution:

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{\operatorname{Var}(S)}} & S > 0 \\ 0 & S = 0 \\ \dfrac{S+1}{\sqrt{\operatorname{Var}(S)}} & S < 0 \end{cases}$$

The null hypothesis of no trend is rejected at α_L, where L is the chosen significance level, if the absolute value of Z is greater than the theoretical value $Z_{1-\alpha_L/2}$.

To account for serial autocorrelation, a modified Mann-Kendall test was applied called block-bootstrap Mann-Kendall[3,4]. The modified method is most useful in the presence of autocorrelation[3,5], the method resamples the data in predetermined blocks a large number of times to estimate the significance of the observed test statistic, i.e. the S test statistic. The block length was set to the number of contiguous significant serial correlations plus one[5]. The process of assessing the significance of a trend in the presence of serial autocorrelation was therefore (1) estimate the S test statistic from the original dataset, (2) estimate the number of significant contiguous serial correlations k and add one, (3) resample the original time series in blocks of k+1 approximately 2000 times[3], compute the S test statistic, estimate the simulated distribution, and (4) assess significance of the trend. If the test statistic calculated in step 1 lay outside the confidence interval of the simulated distribution from step 3, the null hypothesis was rejected and the trend was considered significant.

References

1 Mann, H. B. Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, 245-259 (1945).

2 Kendall, M. G. Rank correlation methods. (1948).

3 Önöz, B. & Bayazit, M. Block bootstrap for mann–kendall trend test of serially dependent data. *Hydrol. Process.* **26**, 3552-3560 (2012).

4 Kundzewicz, Z. *et al. Detecting trend and other changes in hydrological data*. (World Meteorological Organization, 2000).

5 Khaliq, M. N., Ouarda, T. B. M. J., Gachon, P., Sushama, L. & St-Hilaire, A. Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of canadian rivers. *J. Hydrol.* **368**, 117-130 (2009).