

Novel indicator geostatistics for water table mapping that incorporate elevation, land use, stream network and physical constraints to provide probabilistic estimation of heads and fluxes

T. J. Peterson^a, X. Cheng^b, A. W. Western^a, L. Siriwardena^a, S. R. Wealands^a

^a Department of Infrastructure Engineering, the University of Melbourne, Parkville, Vic. 3010, Australia

^b Department of Primary Industries, Parkville, Vic. 3010, Australia

Email: timjp@unimelb.edu.au

Abstract: Mapping of groundwater level observations often makes very little use of auxiliary data and is often undertaken simply by manual interpolation or ordinary kriging of the heads. Recently, a number of geostatistical methods have emerged that significantly improve estimates by incorporating the land surface elevation and groundwater flow or drawdown equations. However, at the regional scale heads are influenced by numerous other factors that cannot be considered by these methods. Such factors include the land cover type, aquifer basement elevation and upper limits to the heads (such as the land surface). Furthermore, all existing methods fail to include observation uncertainty; produce poor measures of prediction uncertainty; and assume the random field to be multi-Gaussian; that is, the spatial correlation in heads are independent of the head magnitude. To overcome these limitations and to make better use of the observation data, this paper presents a novel indicator geostatistical simulation method for mapping unconfined heads. The simulation method produces many equally probable maps and by post-processing produces quantitative uncertainty maps. Other post-processing could produce new products such as the probability of a stream having a gaining or losing hydraulic gradient and, if multiple time points are mapped, probabilistic changes in storage. To demonstrate the methodology, this paper presents an application for the Broken catchment, Victoria.

The method combines a multi-variate version of kriging with external drift (KED) and a modified Markov-Bayes indicator simulation algorithm to facilitate inclusion of physical constraints to groundwater head and soft data such as landuse. The KED facilitates inclusion of continuous variables that are linearly correlated with head and is used to produce a surface that results from these variables alone. As the difference between this surface and the observations were found to be spatially correlated, and approximately first and second order stationary, the head estimate was able to be refined using indicator kriging (IK) simulations. While IK was essential for inclusion of the land class data and the groundwater head constraints, it also allowed the spatial correlation to vary with the magnitude of the heads. In effect this means it relaxes an assumption required for multi-Gaussian methods such as sequential Gaussian simulations. The entire methodology was implemented within the *R* statistics package using the *Gstat* library and modified *GSLib* algorithms. The source code will be made publicly available with a forthcoming journal paper.

The study area comprised of both the Broken River and Broken Creek catchments with a 20 kilometre buffer to minimise boundary artefacts. Groundwater observations comprised of data from The Department of Primary Industries, Victoria; The Department of Sustainability and Environment, Victoria; and Department of Water and Energy, NSW. Importantly, the land class was found to be statistically important in estimating heads and the spatial correlation was not multi-Gaussian, thus challenging the use of standard kriging methods. All simulation head maps honoured all of the constraints. However, because of large nugget values and observations not located at the centre of grid cells, grid cells estimates can differ from observed water levels. Despite this implementation issue, overall the methodology produced highly plausible water table maps that increase the information extracted from observation bores by use of a very wide range of quantitative and qualitative spatial data.

Keywords: Geostatistics, Kriging, Indicator simulation, Groundwater, Water table, Potentiometry

1. INTRODUCTION

Mapping of groundwater level observations often makes very little use of auxiliary data and is often undertaken simply by manual interpolation or ordinary kriging of the heads (Kumar & Ahmed 2003). Recently, a number of geostatistical methods have emerged that significantly improve estimates by incorporating the land surface elevation (Desbarats et al. 2002, Peterson et al. 2003) and groundwater flow or drawdown equations (Peeters et al. 2010, Tonkin & Larson 2002). However, at the regional scale heads are influenced by numerous other factors that cannot be considered by these methods. Such factors include the land cover type, surface water bodies, geological structures and upper and lower limits to the heads such as the aquifer basement elevation and the land surface. Furthermore, all existing methods fail to include observation uncertainty; produce poor measures of prediction uncertainty; and assume the random field to be multi-Gaussian; that is, the spatial correlation in heads are independent of the head magnitude. To overcome these limitations and to make better use of the observation data, this paper presents a novel indicator geostatistical simulation method (simulation methods produces many realisations of equally probable maps) for mapping unconfined heads and presents an application for the Broken catchment, Victoria. While not the first simulation method for water table mapping (Peterson & Barnett. 2004), it is the first to: (i) utilise categorical predictors such as land cover or stratigraphy; (ii) allow non-Gaussian spatial correlations; and (iii) utilise spatially explicit hydrogeology knowledge by input or omission of constraints on the heads. By post-processing the simulation maps, new map can be derived such as the quantitative uncertainty, probability of a stream having a gaining or losing hydraulic gradient and, if multiple time points are mapped and the specific yield is estimated, the probabilistic changes in storage.

2. HYDROGEOLOGY OF THE BROKEN CATCHMENT

Three major aquifer systems dominate the Broken Catchment– the bedrock, Calivil/Renmark and Shepparton Formation aquifer systems. There are also some localised aquifer systems occurring in colluvium associated bedrock in uplands. All of these are interconnected to varying degrees and may be in direct contact or they may be separated by units of low hydraulic conductivity. While perched water tables are likely to exist, in this study the water table has been conceptualized as existing within the bedrock and the Shepparton Formation. The Shepparton Formation covers most of the Riverine Plain and the sand beds in the formation form important aquifers in many areas. They tend to be highly irregular in their shape and distribution and, as a consequence, it is difficult to predict their occurrence. The size and number of aquifers varies from one area to another. Although sand beds tend to occur more frequently at certain depths, they can be considered as being randomly distributed vertically and horizontally. The bedrock outcrops in the upland area and is buried by the unconsolidated alluvial sediment on the riverine plain. In the west of the catchment, particularly gently undulating hills close to the Riverine Plain, the bedrock is highly weathered, often to depths of 50m or more. The non-weathered sedimentary and volcanic rocks form the major aquifers.

3. METHODS

3.1. Geostatistical Algorithms

The water table surface is the product of a wide range of deterministic (e.g. topography and geology) and stochastic drivers (e.g. climate and land use). This geostatistical algorithm aims to use this information to significantly improve the regional scale mapping of the water table and to quantify its spatial uncertainty. It uses a multi-variate version of kriging with external drift (KED) to incorporate external drivers that are linearly correlated with head, for example land surface elevation. The component of the head not resulting from these external drivers was then interpolated using indicator kriging methods (IK) (Journel & Alabert 1989). Standard IK methods contribute to the above aims as they allow inclusion of categorical land classes, such as land cover or stratigraphy units, using a Markov-Bayes methodology, but also allow the spatial correlation to change with the head (Goovaerts 1997). In effect the latter relaxes statistical assumptions required for multi-Gaussian methods such as sequential Gaussian simulations. More importantly, IK allowed development of a novel algorithm to incorporate physical constraints (or lack of) on the head, such as the land surface elevation or aquifer basement elevation. Overall the new methodology facilitates inclusion of a considerable amount of prior hydrogeological knowledge into the mapping. The entire methodology was implemented within the *R* statistics package using the *Gstat* (Pebesma 2004) and *GSLib* (Deutsch & Journel 1998) modified to include the constraints. Below is a summary of the major steps of the methodology.

Kriging with External Drift (KED)

The first step was to approximate the component of the head not resulting from the external drivers in order to develop the model variograms for the KED step. This was undertaken by developing a weighted linear regression model between external drivers and the head. While multiple drivers can be included, herein only land surface elevation was used. The regression weights were included in the model to account for the location of the observation. The residual heads from this relationship were then calculated and found to be first order globally stationary. These residuals approximated the head not resulting from the external drivers.

The second step was to map the component of the head resulting from only the external drivers and to calculate the difference to the observed water levels. Firstly, an experimental anisotropic variogram was derived for the above regression residuals and a model was fit. In fitting the model, the nugget was estimated from the observation and location errors and the temporal averaging undertaken to derive a water level data point. Next, using this model, KED mapping was undertaken to produce a water table map. Herein, only the DTM was used. Lastly, a second KED map was produced that estimates the DTM related component to the heads. Importantly, because the difference between this surface and the observed heads were found to be spatially correlated and near stationary, the mapping could be refined using indicator kriging simulations.

Constrained Sequential Indicator Simulations

The third step was to derive realisations of head that incorporate the KED external drivers plus categorical land classes and the upper and lower physical constraints. The first of four sub-steps was the assignment of indicator thresholds to adequately characterise the empirical cumulative density function (CDF). For this application, seven indicators were adopted at the following percentiles: 0.05, 0.1, 0.25, 0.5, 0.75, 0.9 and 0.95.

The second sub-step was the fitting of model indicator variograms at each of the seven thresholds. For each threshold, a variogram map was first derived to identify anisotropic spatial correlations or a direction of minimal non-stationarity. Next, an experimental anisotropic variogram was derived and then a model variogram was fit using a least squares Newtonian fitting algorithm with well considered initial estimates.

The third sub-step was the inclusion of landscape class as a prior predictor of head using the Markov-Bayes algorithm (Goovaerts 1997). An appealing aspect of this algorithm over others (e.g. collocated indicator cokriging) is that additional indicator variograms are not derived for each landscape class nor for the cross correlations with the residuals. It was implemented by: (i) calculating the CDF of the residuals for each land use; (ii) estimating prior probabilities for each land type using the previous CDFs and extracting probabilities at the indicator thresholds; (iii) assessing the significance of the prior probabilities by calculating the change in the mean prior probabilities when partitioned into above and below a given indicator threshold.

The fourth sub-step was the indicator simulation. A feature of many geostatistical simulation methods (excluding annealing) is that grid cells for estimation are randomly chosen and input point data is sourced from both local observations and local prior cell estimates. A conditional distribution is then derived using the prior estimates and is randomly sampled to obtain a grid cell estimate. Repeating the mapping with a different sequence of cells produces different realisations, each conditioned by the observation data. A key contribution of this paper is the inclusion of upper and lower constraints or no constraint on the head. Figure 1 summarises the three scenarios of: interpolating the CDF with an upper constraint; extrapolating the CDF with an upper constraint at the cell; and extrapolating the CDF with no upper constraint at the cell. The first scenario is achieved by shifting the upper CDF probability to 1.0 during kriging of the CDF and, when sampling the CDF, the interpolation is between the lower value and the upper constraint. The constrained extrapolation scenario is achieved simply by linearly extrapolating between the upper probability and the constraint set to a probability of 1.0. The unconstrained extrapolation scenario is achieved by estimating the slope between the two upper most probabilities and extrapolating to a probability of 1.0.

3.2. Input Data

Groundwater observation data was compiled using a Victorian groundwater database built for this project. It aggregates databases provided by The Department of Sustainability and Environment (DSE), Victoria; The Department of Primary Industries (DPI), Victoria; and The Department of Water and Energy (DWE), NSW.

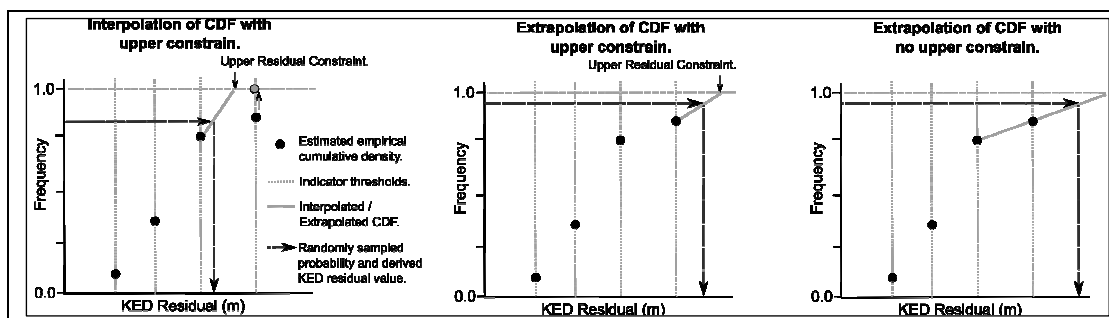


Figure 1. Diagrams how an upper constraint was incorporated into the indicator simulations.

Questionable water level observations were removed using three iterations of an exponential moving average filter. Importantly, the bores used in the mapping should only be monitoring the watertable, and not the deeper aquifers. This is difficult to estimate but was approximated using GIS stratigraphy layers and bore depth. Following detailed hydrogeological review of the data, the mean head for summer 2008 was derived for each bore. All heads were relative to the elevation from the VicMap 20m DTM and the vertical flow was assumed sufficiently small that the head equalled the water table elevation. At locations not denoted as bedrock, the lower constraint to the heads was the Shepparton Formation basement. The upper constraint was the land surface elevation except at major water bodies (form VicMap-HYDRO reaches classed as ‘medium’ or ‘high’). Australian Bureau of Rural Sciences (BRS) land class was adopted as the categorical land type.

3.3. Implementation

The study area comprised of both the Broken River and Broken Creek catchments with a 20 kilometre buffer to minimise boundary artefacts. Twenty indicator simulation realisations were derived using the closest 100 observations points and 50 prior cell estimates within a maximum search radius of 120 km. The realisations were then post processed to the mean depth to water table and standard deviation in the head estimate. The number of realisations was limited to only 20 because of computational constraints. This will produce some uncertainty in the mean and standard deviation and future applications aim to derive 100 realisations.

4. RESULTS

The regression model between head and land elevation produced a coefficient of determination of 0.98. The residuals from this model were then used to derive the KED anisotropic residual variogram model in Figure 2. Two correlation scales were observed (less than 2 km and less than 40 km). The fine scale correlation was isotropic and incorporated into the larger scale variogram by rescaling the total sill to the experimental nugget of the large scale variogram.

Using the KED variogram, the KED residuals were estimated and are summarised in Figure 3. It also details the seven thresholds values and maps of the indicator values at four thresholds. For each of these thresholds, indicator variogram maps and variograms were derived (Figure 4). Importantly, the variogram range appears to decline as the absolute threshold value declined. To illustrate the basis for the inclusion of land class, Figure 5 shows the CDFs for dryland and irrigated regions. They differ most significantly at the 5th, 10th, 90th, and 95th threshold percentiles.

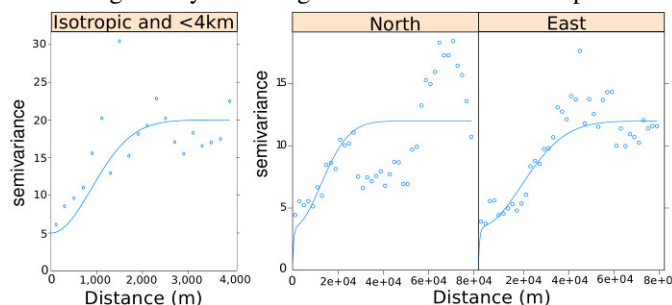


Figure 2. KED fine scale (left) and regional scale (right) experimental and model residual variograms.

Figure 6 presents maps of the mean depth to water table and standard deviation from the 20 realisations. It also details the input constraint layers, the BRS land use and a cross section of heads. The map of mean depth to water table clearly shows the influence of the lower constraint layer. Near to Lake Mokoan the water table was estimated as significantly deeper at regions without a lower head boundary. With respect to the upper constraint, the depth to water table was below the land surface at all locations with an upper constraint (which equalled the land surface elevation). This could not have been achieved using kriging methods. At

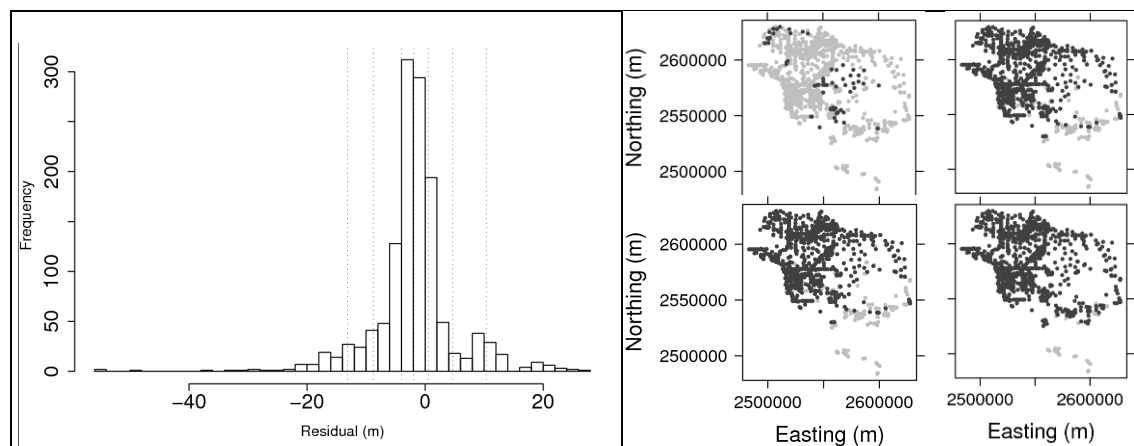


Figure 3. Indicator thresholds and the spatial distribution from four thresholds where ● depicts an indicator value of zero and ● of one (top left: 5th percentile (residual<-13.156); top right: 50th percentile (residual<-1.902); bottom left: 75th percentile (residual<0.541); bottom right: 95th percentile (residual<10.373).

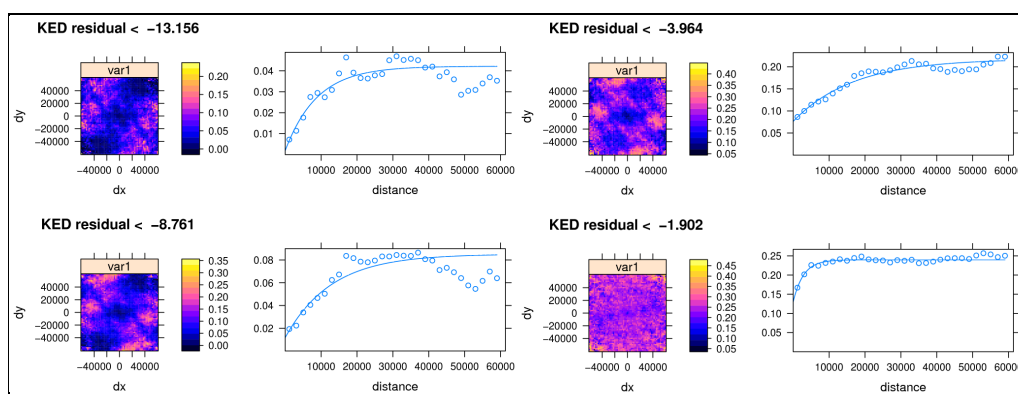
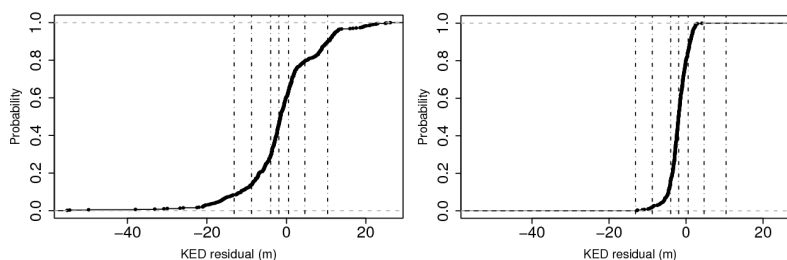


Figure 4. Isotropic variograms at the 5th, 10th, 25th and 50th percentile indicator thresholds.

locations without an upper constraint, such as along the Goulburn River, the influence on the mean water table depth is less apparent than for omission of the lower constraint. One of the few clear influences is the large artesian area near to Yambuna. The omission of the upper constraint, and to a lesser degree the lower constraint, does however significantly increase the local uncertainty in the head. This is most apparent near to Honeysuckle Creek. By induction, inclusion of either constraint reduces the estimated local uncertainty.

To assess the realisations in more detail, Figure 6f presents cross-sections containing all 20 simulations. It shows the distribution of realisations at a given location to be negatively skewed toward a lower elevation, most likely as a result of the upper constraint limiting the positive tail of the distribution. It also shows that across the cross-section the variance is fairly uniform. Even at the Broken Creek (depression on lower inset plot) the variance is comparable. Figure 6f also presents a comparison of the realisations against head observations. It shows observed heads are within the envelope of realisations but many realisations do not honour the observations. However, observations that are located at the centre of a grid cell (those in Figure 6f are not centred) are honoured by all realisations. When an observation is not centred within the grid cell, a large indicator variogram nugget (and possible the very short range of some variograms) can cause the cell estimate to quickly deviate from the observation.



(a) Dryland agricultural lands. (b) Irrigated agricultural lands.
Figure 5. Empirical cumulative probability functions of KED residuals disaggregated by land class.

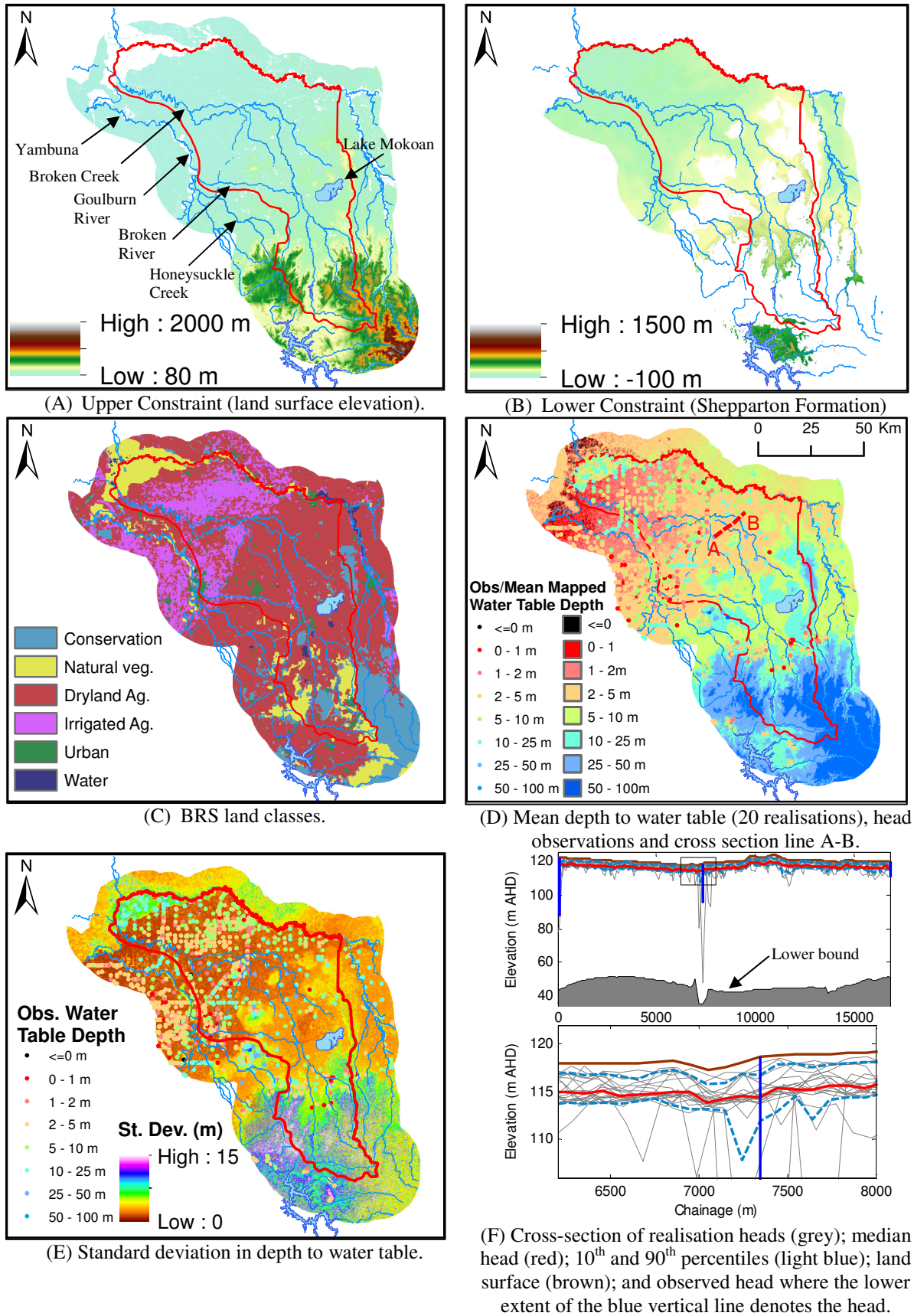


Figure 6. Spatial inputs (Figs. A to C); post-processed results (Figs. D and E); and cross sections (Fig. F);

5. DISCUSSION AND CONCLUSION

This paper outlined a novel approach that advances the mapping of unconfined heads by incorporating a wide range of secondary data, specifically: surface elevation, land class and upper and lower head constraints. The head constraints were found to eliminate unrealistic artesian values, inform the influence of bedrock regions and major water bodies and reduce the local error estimates. However, as the distribution of realisations at a location appears to be non-normal it is concluded that non-parametric measures of uncertainty, such as the interquartile range, should be adopted instead of the standard deviation. With respect to the land class, it resulted in clear statistical differences in the departure from the topographic estimate of head. However, its influence should be more thoroughly quantified by undertaking simulations without its inclusion.

The indicator variograms suggest an increasing spatial correlation with increasing departure from the topographic estimate of head. If the observed residual heads were sampled from a multi-Gaussian random field, then the indicator variogram range would be at a maximum for the 50th percentile threshold and decrease toward the upper and lower percentiles (Gómez-Hernández & Wen 1998, Western et al. 1998). It is therefore concluded that unconfined heads are not multi-Gaussian and that indicator simulation methods should be adopted. Adoption of multi-Gaussian methods, such as sequential Gaussian simulation (Deutsch & Journel 1998), would result in poor estimation of extreme heads such as groundwater mounds and sinks. Additionally, only indicator methods would facilitate the inclusion of uncertainty in the observations.

Finally, in applying these methods consideration should be given as to whether a grid cell head estimate should honour an observation not at the centre of a cell. If so, then the nugget for each indicator variogram should be set to a near-zero fraction of the total sill and, if the variogram range is small, consideration should be given to shifting observations to the centre of cells.

6. ACKNOWLEDGMENTS

The authors are very grateful to the funding and support from Australian Research Council (LP0991281) and the following research partners: Dr. Frost and Dr. Daamen of The Bureau of Meteorology, The Department of Primary Industries, Victoria and Mr. McAuley of The Department of Sustainability, Victoria.

REFERENCES

- Desbarats, A. J., Logan, C. E., Hinton, M. & Sharpe, D. R. (2002), 'On the kriging of water table elevations using collateral information from a digital elevation model', *Journal of Hydrology* **255**(1-4), 25–38.
- Deutsch, C. V. & Journel, A. G. (1998), *GSLIB Geostatistical Software Library and User's Guide*, 2nd edn, Oxford University Press, New York.
- Gómez-Hernández, J. J. & Wen, X.-H. (1998), 'To be or not to be multi-gaussian? a reflection on stochastic hydrogeology', *Advances in Water Resources* **21**(1), 47 – 61.
- Goovaerts, P. (1997), *Geostatistics for natural resource evaluation*, Applied Geostatistics Series, Oxford University Press, New York.
- Journel, A. & Alabert, F. (1989), 'Non-gaussian data expansion in the earth sciences', *Terra Nova* **1**(2), 123–134.
- Kumar, D. & Ahmed, S. (2003), 'Seasonal behaviour of spatial variability of groundwater level in a granitic aquifer in monsoon climate', *Current Science* **84**(2), 188–196.
- Pebesma, E. (2004), 'Multivariable geostatistics in s: The gstat package', *Computers and Geosciences* **30**(7), 683–691.
- Peeters, L., Fasbender, D., Batelaan, O. & Dassargues, A. (2010), 'Bayesian data fusion for water table interpolation: Incorporating a hydrogeological conceptual model in kriging.', *WRR* **46**.
- Peterson, T. J. & Barnett, B. (2004), Spatially quantifying the uncertainty of salinity risk assessments, in S. Dogramaci & A. Waterhouse, eds, 'Engineering Salinity Solutions - 1st National Salinity Engineering Conference', Engineers Australia, Perth, Australia.
- Peterson, T. J., Clifton, C. & Chaplin, H. (2003), Kriging and the drift of salinity risk mapping, in M. Boyd, ed., 'About water : 28th International Hydrology and Water Resources Symposium', Engineers Australia, Wollongong, Australia.
- Tonkin, M. J. & Larson, S. P. (2002), 'Kriging water levels with a regional-linear and point-logarithmic drift.', *Ground Water* **40**(2), 185–193.
- Western, A. W., Blöschl, G. & Grayson, R. B. (1998), 'How well do indicator variograms capture the spatial connectivity of soil moisture?.', *Hydrological Processes* **12**(12), 1851.