Engineers Australia
29th Hydrology and Water Resources Symposium
21–23 February 2005, Canberra

Quantitative measures for the local similarity of hydrological spatial patterns

Stephen R. Wealands
PhD Student, Department of Civil and Environmental Engineering,
The University of Melbourne, Victoria, Australia
Email: srweal@civenv.unimelb.edu.au

Rodger B. Grayson
Director, Cooperative Research Centre for Catchment Hydrology,
The University of Melbourne, Victoria, Australia
Email: rodger@civenv.unimelb.edu.au

Jeffrey P. Walker
Senior Lecturer, Department of Civil and Environmental Engineering,
The University of Melbourne, Victoria, Australia
Email: j.walker@unimelb.edu.au

ABSTRACT: The task of assessing similarity between data sets is common in hydrological modelling. While this has been widely researched for temporal data sets, the similarity between spatial patterns has been largely ignored. This has been due to a lack of spatial pattern data. Today there is widespread use of distributed hydrological models and increasing availability of observed spatial patterns. These observed spatial patterns are useful for model calibration and optimisation, though at present there is limited use of the spatial information contained in them, other than through visual comparison. This is mostly due to a lack of understanding of methods to make optimal use of this information-rich data. The work in this paper investigates quantitative measures for judging the similarity between observed and simulated spatial patterns, with a particular emphasis on local similarity techniques. Three methods — fuzzy comparison, importance maps and image segmentation — are introduced, with a detailed demonstration using fuzzy comparison. Fuzzy comparison allows users to specify their tolerance for errors in value and location when comparing spatial patterns. The different measures presented here can be used to assess many aspects of similarity, which is important for automated model calibration and/or evaluation.

INTRODUCTION

In hydrological modelling, assessing the similarity between data sets is an everyday task, regardless of whether the data is temporal or spatial. Many methods exist for doing this, but most were <u>not</u> developed specifically for, or applied to, hydrological data sets. As such, it is necessary to understand the methods, their application, and what their resulting measures actually represent.

Research into similarity measures for comparing temporal data has made some progress in hydrological modelling. Legates and McCabe (1999) provide an evaluation of many methods used for assessing similarity between temporal data sets. Some methods are sensitive to matching extreme values, while others provide a test of fit but ignore absolute differences. It is concluded that relative, absolute, local and global measures should <u>all</u> be used when assessing the similarity between data sets. Additionally, the use of specialised methods for particular types of data can provide more informative similarity measures. Boyle et al. (2000) present a method for temporal data in which the hydrograph is divided into 'process-related' components. Each component is then compared, providing a measure of similarity that can be directly related to the quality of process representation in the model. This requires prior knowledge about the phenomenon being compared, which is non-trivial for spatial data sets (herein referred to as spatial patterns).

There are many methods available for assessing similarity between spatial patterns data, but few have been applied in hydrological modelling. Together, these global and local methods can describe the similarity between the values in the spatial patterns. But as with many of the temporal measures, they ignore the specific arrangement of the values (especially the global methods). As a result, most hydrologists rely on visual comparison for assessing similarity (Grayson et al. 2002). Visual comparison can be thought of as a specialised method, as it incorporates knowledge about the hydrological phenomenon and other ancillary information. However, its weaknesses are that it is neither automated, objective, repeatable nor quantitative — all things that are important when assessing similarity between many data sets. This research aims to address some of these weaknesses by emulating parts of the visual comparison process computationally. There is no expectation that a computer algorithm will be able to emulate the human brain. However, the steps undertaken during visual comparison suggest many new avenues to pursue for developing specialised methods for assessing similarity in hydrological comparisons.

This paper discusses the background to similarity assessment for spatial patterns and describes three different methods for assessing local similarity, including an example of their uses. A detailed demonstration of using fuzzy comparison is then given, using observed and simulated spatial patterns of soil moisture.

BACKGROUND

A review of the literature on computer vision, image processing and pattern recognition has identified the major processes undertaken during visual comparison and methods for their emulation (Wealands et al. submitted). A visual comparison involves both global and local similarity assessment (Hagen 2003, Hay et al. 2003). During local comparison, the image is viewed as a set of homogeneous regions, rather than individual pixels (Hay et al. 2003). The visual comparison also focuses on particular features or parts of the spatial pattern, rather than treating every location equally (Tompa et al. 2000). During this process, observations such as the similarity of shape, location and intensity are noted. Finally, the observed similarities and differences observed are explained and/or interpreted using extensive background knowledge (Grayson et al. 2002). Thus, the procedure used during visual comparison can be described as 'global similarity assessment, followed by local similarity assessment of regions in the spatial pattern using various measures, with a tolerance for minor differences, and additional focus on more important parts of the spatial pattern'.

Global methods for assessing similarity of spatial patterns are plentiful (see Scheibe 1993). Basic statistics, geostatistics and landscape indices can all characterise certain features of the spatial pattern. These numerical summaries can then be compared to measure the similarity between spatial patterns.

Spatial patterns are usually grid-based representations of an area, comprising a number of cells. Each cell has a location and a value. Some spatial patterns, such as simulations from a contour-based model, are made up of regions rather than cells, but they can be readily converted into a grid-based format. For local similarity methods, it is usual to examine the similarity for every cell in the spatial pattern. The most common method used is root mean squared error (RMSE), which provides a summary of the squared residuals (differences between the similarly located pixel values in the two spatial patterns). Local similarity methods are far more sensitive to differences between the spatial patterns than global methods because they evaluate <u>every</u> location and use the spatial data in its complete form. At each cell, a measure is calculated (e.g. the squared residual) to represent the similarity and this is stored in an intermediate spatial pattern (e.g. of residuals). The intermediate spatial pattern is useful for closer inspection of the differences and is finally summarised to produce the resultant local similarity measure.

More specialised methods involve additional processing both before and during the calculation of similarity measures. For example, preprocessing can involve smoothing or aggregation to change the scale of the observed or simulated spatial patterns to account for such things as measurement error, mismatches in scale and so on. These adjusted spatial patterns are then used for computing the similarity measure. A problem with these cell-by-cell comparisons is that minor errors in spatial location can severely degrade correlations. Alternatively, each location can be compared against neighbouring locations in the other spatial pattern, with the most similar measure from the whole neighbourhood used to represent the similarity for that location. Both these examples illustrate how a minor modification can alter a standard method, thus making it more specialised.

LOCAL COMPARISON METHODS

The aim of local comparison methods is to emulate the major features of detailed visual comparison, so that these can be quantified and automated. Three methods used to produce local similarity measures are i) fuzzy comparison, ii) importance maps and iii) image segmentation (Baatz & Schäpe 2000, Tompa et al. 2000, Hagen 2003).

Fuzzy comparison is a method used for tolerating shifts and differences during the calculation of the similarity measure. This allows the user to specify weights for locational matching (i.e. what amount of displacement is acceptable) and value matching (i.e. what amount of error is considered acceptable). The method processes each cell, computing a similarity value between the respective location and its neighbouring locations in the second spatial pattern (Hagen 2003, Wealands et al. submitted). These values reflect the level of similarity when considering the specified weights, with 0 denoting values that are not similar and 1 denoting identical values and locations. For each cell, the final similarity is computed as the average of a two-way comparison. More details are given using a detailed example.

Figure 1 i) shows two different sets of residual and location weights that have been used to calculate the fuzzy similarity between observed and simulated soil moisture data. The more tolerant residual weights (b) produce a higher overall similarity value than the more limiting weights (a). When multiple sets of observed and simulated spatial patterns are compared, this method can help reveal similar spatial patterns that are not detected by standard local similarity methods due to shifts or minor differences.

Weighting spatial patterns before computing similarity measures is a way of focusing on the important areas. Visual comparison does this automatically as a result of both visual cues (e.g. bright spots) and background knowledge (e.g. focusing only on areas the user knows are gullies). While there is literature on what draws visual attention in an image, the findings are often related to the type of image (e.g. in an image of a human face, viewers are drawn to the eyes). However, it has been recognised that features occurring infrequently in images (e.g. extreme values) are of high perceptual importance, regardless of the context (Tompa et al. 2000). This can be used to produce perceptually weighted spatial patterns, in which the infrequent values are given higher weights than those that are common. Due to the weightings, calculation of the standard RMSE measure will lead to a larger residual where the infrequent values do not match (see Wealands et al. 2004 for examples). Weighting can also be applied to limit the areas in which the similarity measure is computed. If the user is only interested in the similarity of certain areas (e.g. north facing slopes), then a weighting that either enhances or separates these areas will focus the meaning of the similarity measure accordingly.

Figure 1ii) shows the differences between standard RMSE calculations when using different slope weights to focus the comparison. By using the weights to limit the influence of slopes greater than 10 degrees (a), the similarity measure is focused more on similarity in flatter areas. If the weights exclude the steeper areas entirely (b), then a measure that is only related to flatter gully areas is produced.

Segmentation is the process of breaking up an image into regions using a set of rules. The simplest approach to segmentation is thresholding, where a value is chosen to separate an image into two or more regions. During visual comparison, spatial patterns are viewed as regions rather than pixels (Hay et al. 2003), with the regions detected at varying scales. Emulating this computationally is a difficult task.

Using a multiresolution segmentation technique from image processing (Baatz & Schäpe 2000), the spatial patterns of soil moisture have been segmented into homogeneous regions in Figure 1 iii). Using the mean values for each region, an RMSE measure has been calculated between the segmented spatial patterns. This value is less than the RMSE calculated between the original spatial patterns (shown in Figure 1 ii)) due to the removal of noise via the averaging within regions. This method seeks to emulate the region detection process that is done visually, by simplifying the spatial pattern prior to comparison. It may be particularly useful for detecting similarity between noisy data sets, in which the noise precludes the use of standard methods like RMSE.

FUZZY COMPARISON

Fuzzy comparison has been applied to the task of comparing spatial patterns to produce a measure that tolerates errors in value and location. The user can specify the tolerances as fuzzy membership functions, which relate the error amount to a level of similarity (see Power et al. 2001). By then combining the various aspects of similarity (e.g. similarity of location, similarity of value), an intermediate fuzzy similarity map is produced. This can be used to visually interpret the spatial arrangement of similarity, but is usually reduced down to the average similarity to summarise the entire spatial pattern.

When comparing spatial patterns from hydrological models, we are interested in tolerating some error in the values and also in the location of the values. If the process understanding or model parameterisation are wrong, this can lead to shifted values (e.g. due to drainage being too rapid). The situation detailed in this paper applies to continuous value spatial patterns. In the past these methods have been applied to categorical data (Hagen 2003, Güntner et al. 2004).

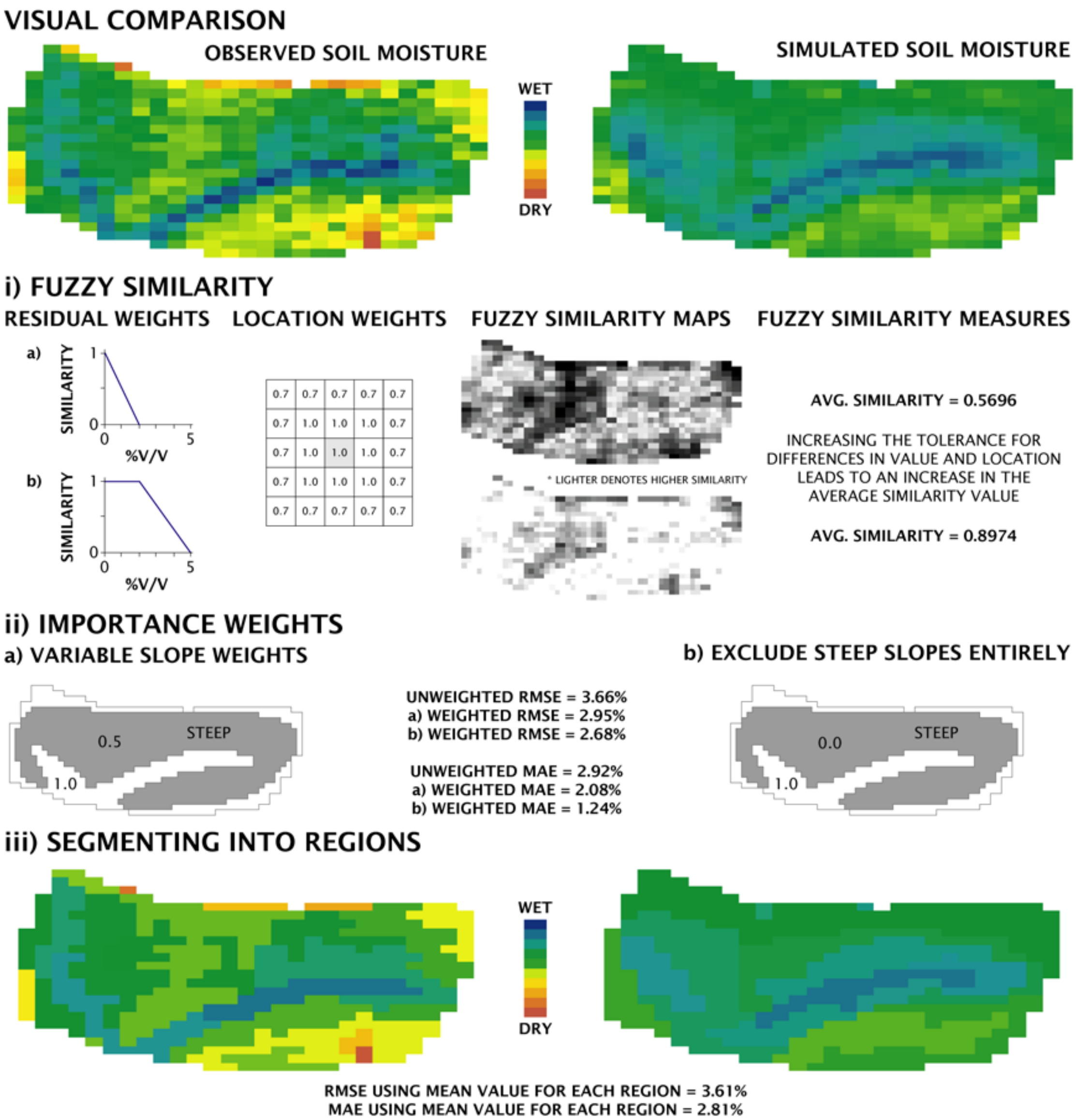

Figure 1. Example illustrating three specialised methods for assessing local similarity. The methods aim to emulate some aspects of visual comparison, including i) tolerance for differences in values and locations; ii) focus on certain parts of the spatial pattern more than others; and iii) comparison of regions rather than pixels.

Tolerance for value differences

For continuous data, a different approach for specifying the fuzzy membership functions is required compared to those used for categorical data. One possible method is to categorise the continuous data, but this removes the variability in the data and requires decisions to be made about category numbers and bounds. Instead, an approach was developed in which the user specifies the 'allowable error' and the 'error limit'. The allowable error is the amount of error between values that will be ignored (i.e. the value will be judged to be identical). The error limit is the amount of error at which values are judged to be completely different. In between these values, a linear decay is used to determine the membership value for similarity (see Figure 1 i)). For example, if the allowable error in rainfall was 1mm, and the limit was 3mm, then values that differed by 2mm would have similarity of 0.5 (within range 0 to 1). These values can usually be derived from analysis of measurement techniques or model expectations.

Tolerance for locational differences

As with errors in value, errors in location can be specified using a fuzzy membership function (except that the function is two-dimensional). The approach used here produces a window around the processing cell that contains membership values for each cell around it (out to a specified distance). These values reflect how similar the location of other cells is to the central cell. This is done with a decay function (e.g. linear, exponential), starting at a membership of 1 for the central cell and decaying as distance increases. Alternatively, the user can choose the membership values subjectively, as done in Figure 1 i). Here, the immediately neighbouring cells have membership 1 (i.e. they are considered as the same), while those two cells away have membership 0.7 (i.e. still similar, but not as much). Often in hydrology, the locational error tolerance would have an upper limit, as seen in Güntner et al. (2004) where the maximum shift tolerated was 5 cells (due to resampling the observed data from 10m to 50m resolution).

Computing fuzzy similarity values

At each cell where a value occurs in both the observed and simulated spatial pattern, the observed value is compared with the simulated value and its neighbours (those cells that have locational membership > 0). For each simulated value, the fuzzy similarity is computed by multiplying the locational membership by the value error membership. From all the fuzzy similarity values computed for the processing cell, the maximum similarity is kept. To do a complete comparison, we conduct this both ways and take the average maximum similarity (i.e. the observed value is compared against the window of simulated values, and the simulated value is compared against the window of observed values). This progresses until a similarity value is computed for every cell. The final similarity measure is the average similarity for all cells in the intermediate similarity map.

Demonstration of fuzzy comparison

A number of fuzzy comparisons have been undertaken using the method described here. The data used are spatial patterns of soil moisture that were observed and simulated as part of the Tarrawarra project (Western et al. 1998). A sample of the observed and simulated spatial patterns from May 2, 1996 is shown in Figure 1. There are 10 different simulations used for comparison, as well as two diagnostic spatial patterns — one has relabeled every cell with the mean value of the observed spatial pattern, the other has all values of the observed spatial pattern shifted by 1 cell to the east.

The fuzzy similarity values have been computed for six different sets of fuzzy tolerances. These tolerances are realistic for this data. There is approximately 2% V/V accepted as being measurement error, while 5% V/V is the maximum error desired between the observations and simulations. In terms of locational error, the correlation length of these spatial patterns is around 20m (Western et al. 1998), so any related value should fall within 2 cells (2 x 10m resolution). A RMSE measure is also provided for comparison of the fuzzy method against a standard method. The lower RMSE values and the higher fuzzy similarity values represent more similar spatial patterns.

RESULTS

Similarity of model simulations

The measures that are computed in Table 1 indicate the model simulations that have the most similar arrangement of values to the observed spatial pattern. Using only the RMSE (as is commonly done), model runs 2 and 3 are the most similar, while runs 4 and 7 are the least similar. These runs actually produce worse results than the mean observed pattern, suggesting that they poorly simulate the observations. This can be verified visually using Figure 2.

In all of the fuzzy measures computed, run 3 is the most similar. The RMSE indicates that the agreement between coincident pixel values sums to the lowest total error. As the tolerance for value differences increases (without locational tolerance), the similarity values also increase, but the relative ranking of the different simulations remains much the same. When there is no tolerance for locational error, there is simply a linear rescaling of the residuals.

Introducing some tolerance for locational errors (i.e. the 1 and 2 cell fuzzy measures) permits close matches with neighbours to be included in the comparison. As a result, the overall similarity for spatial patterns with shifted cells improves more than those where shifts do not occur. This is where the fuzzy similarity measure starts to provide different information to the standard RMSE. As the tolerance for locational error is increased, different simulations emerge as being more similar to the observed. The results for 1 and 2 cells, with 0% allowed and 2% limit, consistently identify the best three simulations as runs 3, 9 and 6. Each approach recommends the user investigate the similarity maps for runs 3, 6 and 9 more closely before deciding which is most similar. From previous visual comparison of this same data, Western & Grayson (2000) decided that run 5 was best, although this considered multiple different simulation dates. The reader can experience the subjective nature of visual comparison by contrasting the quantitative results of Table 1 with the spatial patterns shown in Figure 2.

Shifted spatial pattern

To observe how fuzzy similarity responds to major locational errors, the observed spatial pattern had every value shifted by 1 cell. In the results, it is evident that as tolerance for shifts is permitted, the similarity increases dramatically. In practice, this kind of shifting may occur if there are georeferencing problems with either the observed or simulated data. More subtle shifts would happen when processes in a model are wrong (e.g. draining is too slow on a hillslope). These would produce less dramatic improvements in the overall similarity, as they only occur in a part of the whole spatial pattern. As such, limiting the error analysis to smaller or focused regions (as shown in Figure 1 ii)) can be useful to highlight these effects.

DISCUSSION

In the demonstration example used in this paper, there are only 10 different simulations, making it feasible to visually inspect each one and decide based on visual similarity. However, if many different parameterisations are used to produce many different simulations, then using a method such as fuzzy similarity allows the user to reveal the most similar spatial pattern under different tolerances, without subjective visual inspection of hundreds of spatial patterns. It is better to reduce the number of visual comparisons needed, so that they can be done carefully as a final step.

The demonstration is of a method of fuzzy comparison, which tolerates errors in value and location. However, to fully use such a method to detect similarity, the concepts from the other methods need to be combined. For example, the spatial pattern can be broken into homogeneous units using image segmentation. These units can then be used to focus comparison on different parts of the spatial pattern. Similarly, importance maps (or weighting) can be used to control the influence of different parts of the spatial pattern in the overall similarity calculated. There are many ways to combine these relatively simple methods to test different hypotheses about similarity. Used together, they can provide specialised measures to assess particular aspects of similarity between spatial patterns. A comprehensive comparison of spatial patterns involves noting the various aspects of similarity that are important to the study of interest. From these, the user can then make their judgment about which simulations are the ‘most similar’.

Quantitative comparison of spatial patterns for model testing and calibration is in its infancy. In this paper we have illustrated three automated methods that emulate key elements of visual comparison. These focus on the importance of particular parts of a spatial pattern and the tolerance of small differences in location or value. As with the more sophisticated methods in time series comparisons (e.g. Boyle et al. 2000), user intervention is required to be explicit about what components of a pattern are critical. Nevertheless, the techniques enable rapid computation of summary similarity measures that enable a much richer quantitative comparison than has previously been possible.

Table 1. Fuzzy comparison of 12 simulated spatial patterns versus an observed spatial pattern of soil moisture for Tarrawarra on May 2, 1996. Results for six combinations of tolerances are shown. RMSE is computed to enable comparison of this method with a standard method. The best three results are shown in bold, while any results that are worse than the mean observed pattern are in *italic*.

Location error	0 cells	0 cells	1 cell	1 cell	2 cells	2 cells	**RMSE**
Allowed error	0% V/V	2% V/V	0% V/V	2% V/V	0% V/V	2% V/V	**(%V/V)**
Limit of error	2% V/V	5% V/V	2% V/V	5% V/V	2% V/V	5% V/V	
Run 1	0.178	0.662	0.460	0.861	0.589	0.886	3.7
Run 2	**0.210**	**0.699**	0.545	**0.915**	0.671	0.938	**3.4**
Run 3	**0.235**	**0.731**	**0.601**	**0.942**	**0.735**	**0.960**	**3.2**
Run 4	*0.162*	*0.563*	0.425	0.819	0.574	0.875	*4.1*
Run 5	0.170	*0.617*	0.489	0.879	0.641	0.919	3.8
Run 6	0.186	0.676	**0.555**	0.910	**0.700**	**0.941**	**3.6**
Run 7	*0.162*	*0.576*	0.511	0.860	0.679	0.924	*4.2*
Run 8	0.183	0.649	0.505	0.874	0.652	0.916	3.8
Run 9	0.188	0.660	**0.568**	0.912	**0.719**	**0.948**	3.7
Run 10	**0.198**	**0.678**	0.543	**0.912**	0.684	0.941	3.6
Mean observed pattern	0.170	0.621	0.407	0.799	0.500	0.809	4.1
Shifted observed pattern	0.322	0.766	0.973	0.990	0.980	0.992	3.1

NOTE: All fuzzy comparison values are unitless and have the range from 0 to 1.

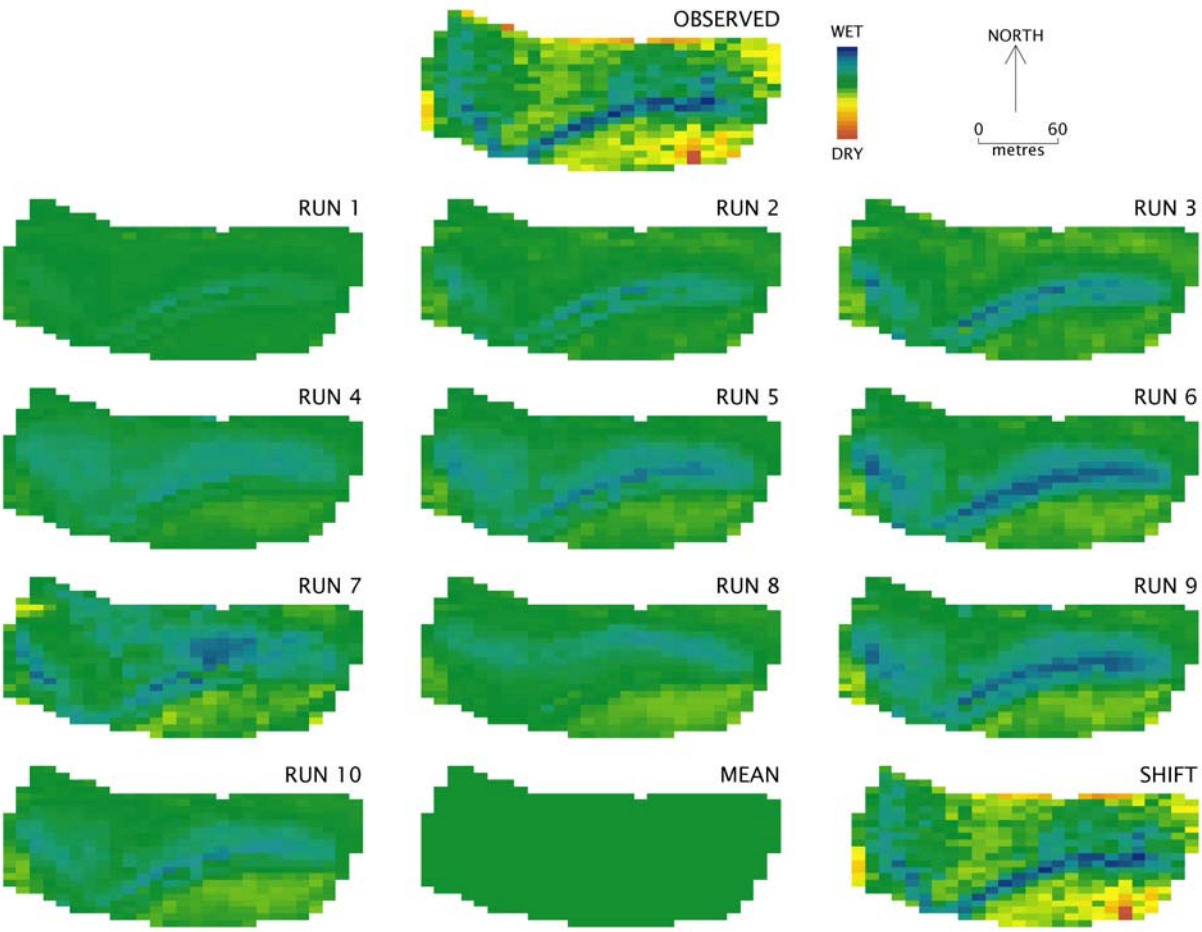

Figure 2. The soil moisture spatial patterns used for the comparison results shown in Table 1. All spatial patterns are shown on the same colour range, from 26% V/V (dry) up to 52% V/V (wet).

ACKNOWLEDGMENTS

We would like to acknowledge the provision of the Tarrawarra data by Andrew Western and technical support by the development team for The Invisible Modelling Environment (TIME) (http://www.toolkit.net.au). Financial support was provided by the Cooperative Research Centre for Catchment Hydrology, an Australian Postgraduate Award Scholarship and a Postgraduate Overseas Research Experience Scholarship from The University of Melbourne.

REFERENCES

Baatz, M & Schäpe, A 2000, 'Multiresolution segmentation: An optimization approach for high quality multiscale image segmentation', in J Strobl & T Blaschke (eds.), Angewandte geogr. Informationsverabeitung, Herbert Wichmann Verlag, Heidelberg, Germany, pp. 12–23.

Boyle, DP, Gupta, HV & Sorooshian, S 2000, 'Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods', Water Resources Research, vol. 36, no. 12, pp. 3663–3674.

Grayson, RB, Blöschl, G, Western, AW & McMahon, TA 2002, 'Advances in the use of observed spatial patterns of catchment hydrological response', Advances in Water Resources, vol. 25, no. 8–12, pp. 1313–1334.

Güntner, A, Seibert, J & Uhlenbrook, S 2004, 'Modeling spatial patterns of saturated areas: An evaluation of different terrain indices', Water Resources Research, vol. 40, no. 5.

Hagen, A 2003, 'Fuzzy set approach to assessing similarity of categorical maps', International Journal of Geographical Information Science, vol. 17, no. 3, pp. 235–249.

Hay, GJ, Blaschke, T, Marceau, DJ & Bouchard, A 2003, 'A comparison of three image-object methods for the multiscale analysis of landscape structure', ISPRS Journal of Photogrammetry and Remote Sensing, vol. 57, no. 5–6, pp. 327–345.

Legates, DR & McCabe, GJ 1999, 'Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation', Water Resources Research, vol. 35, no. 1, pp. 233–242.

Power, C, Simms, A & White, R 2001, 'Hierarchical fuzzy pattern matching for the regional comparison of land use maps', International Journal of Geographical Information Science, vol. 15, no. 1, pp. 77–100.

Scheibe, TD 1993, Characterization of the spatial structuring of natural porous media and its impacts on subsurface flow and transport, PhD thesis, Stanford University, 316 pp.

Tompa, D, Morton, J & Jernigan, E 2000, 'Perceptually based image comparison', in Proceedings of the IEEE International Conference on Image Processing, Vancouver, BC, Canada, 10–13 September, pp. 489–492.

Wealands, SR, Grayson, RB & Walker, JP 2004, 'Investigating spatial pattern comparison methods for distributed hydrological model assessment', in T Jakeman, C Pahl & S Schmidt (eds.), Proceedings of iEMSs 2004, Osnabrueck, Germany, 14–17 June.

Wealands, SR, Grayson, RB & Walker, JP submitted, 'Quantitative comparison of spatial patterns for hydrological model assessment — some promising approaches', Advances in Water Resources.

Western, AW, Blöschl, G & Grayson, RB 1998, 'Geostatistical characterisation of soil moisture patterns in the Tarrawarra catchment', Journal of Hydrology, vol. 205, pp. 20–37.

Western, AW & Grayson, RB 2000, 'Soil moisture and runoff processes at Tarrawarra', in R Grayson & G Blöschl (eds.), Spatial patterns in catchment hydrology: Observations and modelling, Cambridge University Press, Cambridge, pp. 209–246.