Investigating Spatial Pattern Comparison Methods for Distributed Hydrological Model Assessment

S.R. Wealands [a, b], **R.B. Grayson** [a] **and J.P. Walker** [b]

[a] *Cooperative Research Centre for Catchment Hydrology, Australia*
[b] *Department of Civil and Environmental Engineering, The University of Melbourne, Australia*

Abstract: Distributed hydrological models combine observations and knowledge about a hydrological system to make spatial predictions of hydrological attributes. These models require methods to assess their performance at spatial prediction. The current practice for assessment is simplistic. For qualitative assessment, simulated spatial patterns are compared visually against an observed pattern to assess their spatial similarity. To obtain a quantitative measure of similarity, each individual location is numerically compared to produce either a mean squared error (MSE) or correlation statistic. Both of these comparisons have their limitations. The visual comparison is subjective and the numerical comparison generally ignores the spatial structure of the patterns. There is demonstrable need for repeatable methods that can capture and quantify the important aspects of visual comparison. This paper demonstrates such a method from the image processing literature. It is a modification of the MSE statistic, called the information mean squared error (IMSE). This method weights each location in the spatial pattern by the 'informativeness' of 'an event' at that location. The weighted spatial patterns are then compared using a standard MSE statistic. IMSE aims to emulate human vision by more heavily weighting informative pixels. This paper applies IMSE to spatial patterns of soil moisture content. It is found to work well when using local variance as the 'event', as this helps enhance the general spatial trends that humans readily recognise. However, when the two spatial patterns are vastly different, IMSE proves to be less reliable due to the inconsistent weightings calculated for each spatial pattern.

Keywords: Spatial pattern; Model assessment; Comparison; Hydrology; Distributed models; Self information

1. INTRODUCTION

There is increasing recognition in hydrological modelling of a need for improved methods for comparing spatial patterns [Grayson et al., 2002; Jetten et al., 2003]. This has arisen from the increased availability of observed spatial patterns for testing the spatial predictions from distributed models. Grayson and Blöschl [2000] provide many examples of modelling projects where spatial patterns have been observed, with a purpose to assess the spatial component of the model predictions. However, within these projects there has been little use of new spatial pattern comparison methods. Most studies rely on standard statistical measures (like mean squared error) or subjective visual comparisons to tell the story of how well the model is predicting the spatial patterns. This work pursues new approaches for the spatial pattern comparison task.

Spatial patterns in hydrology are usually grid-based representations of an area (or catchment), with a value provided at every grid cell (or pixel). Each grid cell is square and has dimensions specified as the cell size (or resolution). Spatial patterns are effectively the same as grey level images, although the number of discrete pixel values is usually larger than in a standard (8-bit) image band. Observed spatial patterns can be obtained via grid-based field measurements, by interpolation of sparse field measurements or from remote sensing. These measurements then require processing to ensure they are consistent with the predicted spatial patterns from a hydrological model (i.e. with equivalent support, spacing, extent). The spatial patterns must be carefully prepared so that they are comparable. This work does not focus on this aspect of spatial patterns, preferring to concentrate on the comparison once the spatial patterns have been prepared correctly.

2. COMPARING SPATIAL PATTERNS

When comparing spatial patterns, the ability to obtain a measure of similarity is essential. Methods that provide a quantitative measure can

be used to compare an observation with multiple simulations. The resultant measure can then be used to determine which of the simulations are more similar. To understand which features of the simulations are more similar, the user needs to understand how the comparison method computed the similarity measure. By interpreting the performance of these measures with hydrological spatial patterns, certain methods may emerge that are more suitable.

A review of the current suite of methods used for spatial pattern comparison in hydrology identifies the features that are currently compared. The spatial pattern comparisons presented in Grayson and Blöschl [2000] and Grayson et al. [2002] provide a comprehensive cross-section of the commonly used methods. The most widely used method is visual comparison, which allows the user to draw on their background knowledge about the study area and model structure to interpret the similarity. This method will always be used when presented with two figures depicting spatial patterns. It is also used to compare time series data or transects that have been extracted from the observed and predicted spatial patterns. This type of comparison is too subjective for repeatable and rigorous comparison. It is also very time-consuming, unable to interpret large spatial patterns completely, and it cannot produce a quantitative measure.

Most common quantitative measures used are global measures, which characterise the spatial pattern first using statistics or indices (e.g. mean error to identify bias, spatial correlation length to compare spatial statistical structure). These summaries are then compared numerically. For local measures, pixel-by-pixel comparisons such as mean squared error (MSE) (to assess the local agreement between values) are the dominant measures. Here, the residuals are computed between two spatial patterns and then squared and averaged. The residuals are usually analysed to detect relationships with topographic variables. For spatial patterns to be judged as being similar with all of these local measures, there must be close agreement between the pixel values at coincident locations, so these techniques are very sensitive to minor shifts. They are also influenced by disagreement between coincident pixels, even though there may be close agreement with neighbouring pixels. This is especially evident when the 'support' of the two spatial patterns does not match (i.e. the observed pixel value has a support of much less than the cell size, whereas the predicted value represents the average of the entire pixel). In these situations, small-scale variance can mask the overall pattern and thus lead to poor results for local similarity measures.

These current methods are useful for the analysis of similarity between spatial patterns, but are limited in their ability to measure certain aspects of similarity. Other methods for characterising and comparing more detailed features of spatial patterns are necessary. By understanding the strengths of new similarity measures and experimenting with hydrological data sets, alternative methods for the comparison of spatial patterns can be further developed.

3. OTHER COMPARISON METHODS

There is a large amount of research in other disciplines, such as image processing and computer vision, that can suggest methods for working with spatial patterns in hydrology (e.g. segmentation, image filtering) [Scheibe, 1993]. However, not all techniques in other fields are applicable to hydrological patterns. For example, with face recognition, it is usual for the observed image (a face) to be processed down to a set of features (such as eye locations) that are stable in all conditions (e.g. different lighting). This set is then compared against a large database of features to find a match. The spatial patterns present in hydrology rarely have any known features and therefore need solutions that are more generic. A review of this literature is given in Wealands et al. [submitted; 2003]. In this paper, the focus is on an approach for image comparison that was initially developed for assessing the effect of image filtering on the original image.

3.1. Information Mean Squared Error

When an image is filtered (and often distorted), a measure of its similarity to the original is desired. In Tompa et al. [2000], a measure called the information mean squared error (IMSE) is developed. This measure aims to reflect the level of similarity that a human observer would perceive. When humans compare images, it is common for differences in the main features to be weighted more heavily than differences in the background values. Similarly, for spatial pattern comparisons, the less common values (e.g. highs or lows) attract more attention during visual comparison. The basic premise of the IMSE is quite simple – weight 'events' that occur less frequently in the spatial pattern more highly in the comparison. The basic events that occur within a spatial pattern are the actual pixel values. However, other events like local variance (i.e. the variance of pixel values within a neighbourhood) can also be used for the calculation of weights.

The size of the neighbourhood is related to the size of the features (e.g. small features have high variance in small neighbourhoods).

The level of weighting applied in this method represents the 'informativenss' of the particular event. This is measured using Shannon's self-information measure, as applied in Topper and Jernigan [1989]. This is defined to be

$$I(x) = -\log_Y P(x)$$
$$P(x) = \frac{n_x}{N}, \qquad (1)$$

where I(x) is the self-information for event x, $_Y$ is the base for the logarithm (base e is used here), P(x) is the probability of event x occurring, n_x is the number of pixels with event equal to x, and N is the total number of pixels in the spatial pattern.

Due to the logarithm, these weights are maximum when P(x) is close to zero, and minimum when P(x) approaches one. These effects are desirable, so that an event that is almost everywhere in the spatial pattern would contain little information (i.e. it is the background), while the most infrequent events would have the maximum. The weights produced can vary from almost zero to very large numbers, depending on the number of pixels in the spatial pattern (and the base of the logarithm).

Once the self-information is computed for each pixel, the original spatial pattern is multiplied by the weights. The calculation of weights is done for all spatial patterns being compared. To compute the IMSE similarity measure, a standard MSE calculation is done between two weighted spatial patterns.

3.2. Selecting the Event for Weighting

The event used for calculating self-information measures does not have to be the actual pixel value. Rather, the event chosen should be the characteristic of the spatial pattern that is responsible for separating the features of interest from the background. For example, in a spatial pattern with a few areas of very high pixel values on a background of very low values, then pixel value is a good event. For a more homogeneous spatial pattern (with less obvious features), local variance is better, as this is the visual cue for something of interest in the spatial pattern (as there is variance within the neighbourhood).

Another consideration in selecting the event is the number of distinct categories in which the event occurs. In Tompa et al. [2000], the images were always single band, 8-bit images (i.e. having 256 individual values). With spatial patterns, there are often thousands of different values, which should be categorised (or quantised) prior to having the weights calculated. If this is not done, a spatial pattern containing 400 pixels could have 400 different values, resulting in every pixel being weighted equally. As this method was developed to represent perceptual similarity, the number of different categories in the spatial pattern should correspond to the number of categories the human observer can discern in the spatial pattern. A firm value cannot be placed on this, as it varies with the observer. As such, this should be determined empirically. In Figure 1, a human observer can probably discern about 10-20 individual categories (when displayed like this), while in fact there are 146 different values in the non-categorised spatial pattern.

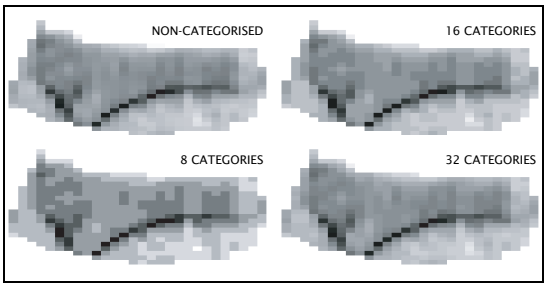

Figure 1. Observed spatial pattern of soil moisture displayed in different categories. The increasing level of grey denotes higher soil moisture content.

3.3. Using the IMSE Measure

The method described produces a measure that indicates the level of similarity between the spatial patterns, with high weights assigned to the 'more informative' pixels. In the resulting measure, a smaller IMSE value denotes more similar spatial patterns. As the self-information weights are specific to the spatial pattern being compared, they cannot be used for inter-comparison. For example, this measure cannot be used to compare a pair of spatial patterns from spring, then a pair from winter, with a view to stating <u>which pair</u> of spatial patterns are more similar. Instead, this method is suitable for comparing multiple spatial patterns of the same event.

One example that is common in modelling projects is for a single observed spatial pattern to be compared with many different model simulations (with different parameter sets). By obtaining measures of similarity between the observed spatial pattern and each simulation, the modeller can help decide which parameter sets lead to best agreement.

4. COMPARISON DEMONSTRATION

This section investigates the use of IMSE for hydrological spatial patterns. It will be presented along with measures of bias, correlation and root mean squared error to help interpret the results. The aim of this demonstration is to characterise which simulated spatial patterns are most similar to the observed spatial pattern for two different dates. The methods all provide measures that can be used to judge different aspects of similarity. This demonstration will undertake analysis of the similarity measures first and then look at the spatial patterns afterwards to discuss the performance.

4.1. Observations and Simulations

The observed spatial patterns analysed here are from the Tarrawarra project [Western et al., 2000]. In this study, soil moisture was measured in the field at regularly spaced grid intervals. This data has then been smoothed using geostatistical methods, to make the support of the field measurements compatible with the model simulations and to add in variability of the measurement technique (with a nugget effect on the variogram used for smoothing), the details of which are in Western and Grayson [2000]. The two observed spatial patterns represent vastly different soil moisture conditions related to the season in which they were measured.

Simulations have been produced using the Thales modelling framework, with the details of these particular simulations given in Western and Grayson [2000]. The 10 different simulations represent different parameterisations for the model. The simulations numbered 1-3 ignore spatially variable evapotranspiration (ET), while the others allow spatially variable ET (which is often related to slope and aspect). All simulations have been resampled from an element-based network onto a regular grid to correspond with the observed spatial patterns.

4.2. Similarity Measures

With each comparison (between the observed and simulated patterns) there are five similarity measures computed (Table 1). These are bias, R^2 correlation (which ignores bias), root mean squared error (RMSE), IMSE using pixel value (pv) and IMSE using local variance (lv) (within a 3 pixel square window). Using a small window ensures that the variance is only computed for the pixel and its 8 neighbours. Two different IMSE measures are given to highlight the impact of the event chosen. Bias is used to assess if one spatial pattern has an overall higher or lower value. This could be removed from subsequent comparison if desired. RMSE provides an overall summary of the difference between the spatial patterns at each location, with a penalty for large discrepancies (the squaring of residuals). IMSE also measures the difference between the spatial patterns at each location, but with an emphasis on areas having high information content. If a pixel is in a frequently occurring event category, its value will be reduced. If the pixel is in a rarely occurring event category, its value will be increased. During the subsequent comparison, if the high information areas between the spatial patterns are vastly different, then the measure will be high, while differences between the 'low information' pixels have far less effect.

4.3. Similarity Measures for April

The different measures are interpreted to identify which simulations are judged more similar to the observed spatial pattern. All of the simulations for April had minimal bias. R^2 correlation is best between the observation and simulation 09. The RMSE values are all around 2.5% V/V, apart from simulations 01-03 and 07. RMSE finds simulation 09 to be the best match. IMSE with pixel value as the event finds simulations 04 and 10 to be the best, with 08 also close. With local

Table 1. Comparison of the observed spatial pattern to 10 different model simulations for two occasions in 1996 (Apr, Oct). Bias and RMSE values are in % V/V. The most similar measures are in bold, other similar ones are in italic.

	13-Apr-96				
Sim.	**Bias**	**Corr.**	**RMSE**	**IMSE (pv)**	**IMSE (lv)**
01	-0.48	0.29	2.84	3087	1666
02	-0.51	0.19	2.84	3745	2190
03	-0.54	0.12	2.86	4664	2807
04	0.19	0.29	2.53	**2185**	1278
05	0.15	0.31	**2.51**	2659	**1152**
06	0.11	0.31	2.52	3148	1269
07	0.67	0.00	3.60	5442	1730
08	**0.02**	0.25	2.56	*2271*	*1212*
09	0.16	**0.40**	**2.39**	4313	**1058**
10	**0.04**	0.27	2.55	**2146**	1266
	25-Oct-96				
Sim.	**Bias**	**Corr.**	**RMSE**	**IMSE (pv)**	**IMSE (lv)**
01	5.96	0.25	6.98	2172	*3778*
02	5.51	0.38	6.60	1843	4222
03	4.62	0.46	5.99	1814	4369
04	6.24	0.32	7.19	1983	**3731**
05	5.85	0.44	6.83	1845	4044
06	5.04	**0.52**	6.23	1872	4236
07	**4.13**	**0.53**	6.13	**1678**	4380
08	4.30	0.40	**5.49**	*1773*	3946
09	5.61	0.42	6.84	**1668**	4286
10	4.73	0.45	**5.81**	1858	**3736**

variance as the event, simulations 05 and 09 are best. Further inspection of 04, 05 and 08-10 would be recommended, as these were judged similar by multiple measures.

4.4. Similarity Measures for October

The simulations for October are all biased, over predicting by about 4-5% V/V. However, reasonable R^2 was present with 06 and 07. The RMSE measure, which incorporates bias and other errors, found simulation 08 as the best. IMSE with pixel value suggests 07-09 to be most similar, while IMSE with local variance finds 04 or 10. On these findings, further inspection of 07-10 would be suggested.

4.5. Visually Assessing the IMSE Measures

Figure 2 provides a visual representation of how the IMSE weightings influence the standard MSE similarity measure. In the first column, there are two simulations shown for the soil moisture content on 13 April 1996. The other columns contain the IMSE weighted spatial patterns that are subsequently compared to produce the IMSE measures in Table 1. By visually inspecting the first column of spatial patterns, it appears that simulation 09 does a better job than 10 of reproducing the linear high-moisture feature. These high pixel values occur less frequently and thus receive a high weighting in the second column. In simulation 10, there is not such a distinct difference between the high and low pixel values, resulting in a more even weighting across the spatial pattern. For local variance, both simulations are enhanced. The variable areas are given higher weightings than the homogeneous background. This appears a more suitable event when comparing these spatial patterns, as it helps discern the feature (i.e. the pixels with more information) from the background. This is a logical characteristic of spatial patterns to use for weighting, as human vision is often drawn to these areas of larger variation. This is similar to the use of 'edges' by Topper and Jernigan [1989], which are a measurement of local gradient widely used in image processing.

The degree of weighting for local variance that is applied to the spatial patterns is more pronounced for simulation 09 than 10. In 09, there are many areas with low variance, but only a few with very high variance. As such, the few pixels are heavily enhanced, while the remainder are heavily reduced. For 10, there is certainly more weighting for the areas with high variance, but the distribution of variances is not as extreme. As a result, the feature is not as greatly enhanced, which visually appears more correct.

As with the standard MSE statistic, even when two spatial patterns look quite similar (e.g. the

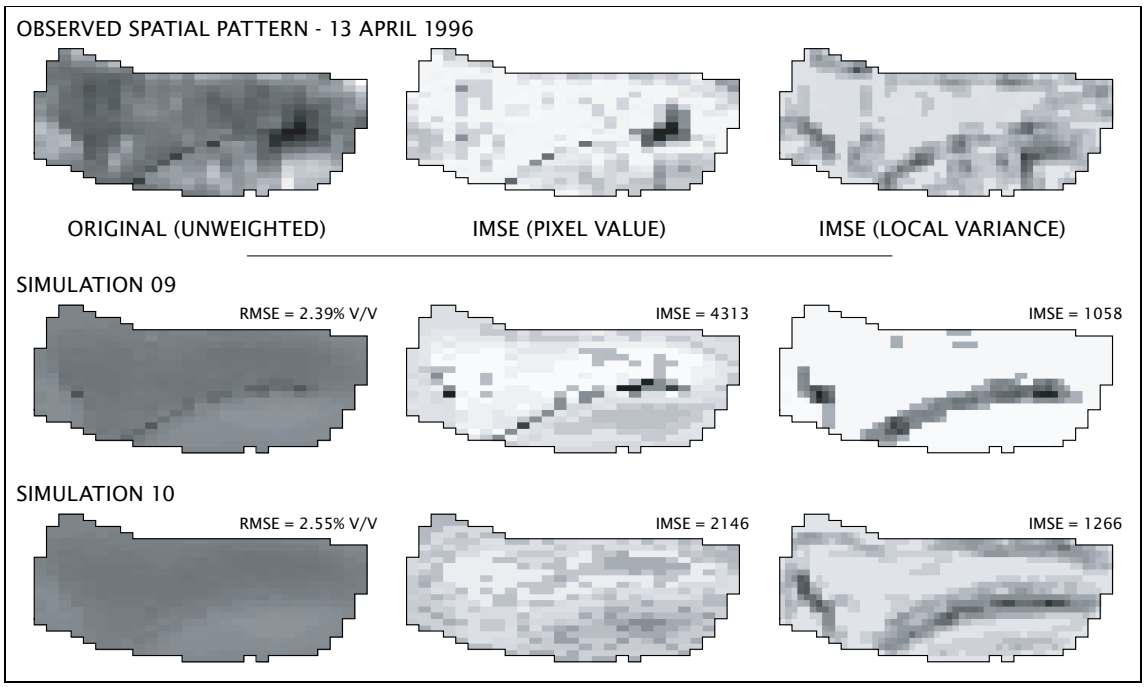

Figure 2. Spatial patterns of soil moisture from 13 April 1996. The observed spatial patterns are shown for two alternative simulations. The original and IMSE weighted spatial patterns are shown to help interpret the meaning of similarity measures. Comparable grey scales are used, with darker greys denoting higher values. Values have been placed into 20 equal interval categories.

IMSE pixel values for observed and simulation 09), if there is not local agreement between the pixel values then the measure states that they are dissimilar. This happens here, with simulation 09 having a similarity measure that is twice simulation 10, although it appears more similar. At present, the IMSE measure does not account for minor shifts between pixels, although this could be one avenue for improvement.

5. DISCUSSION

The use of IMSE with hydrological spatial patterns relies on choosing an event that can discern the features from the background. Figure 2 illustrates that local variance can be useful for discerning the features of interest within an otherwise homogeneous spatial pattern. This is also a logical surrogate for human vision, which uses variation as a means to identify features [Topper and Jernigan, 1989].

The nature of the spatial pattern and its complexity can also make a large difference to the use of the IMSE measure. In spatial patterns with a larger extent, the number of different events occurring can be far greater (due to having many more pixels). Here, the choice of the number of categories will influence how well the measure applies the weightings. Too few categories will result in rare events being lumped together with common events, whereas too many categories can lead to every event being treated as rare.

The application for which this method was initially developed looks at comparing a distorted image with the original. Both images therefore have a similar distribution of values (i.e. histogram), with some minor changes in the histogram of the distorted image. If there is a difference between the histograms for a common pixel value, this difference will be less influential. However, if the difference occurs in a rare pixel value, the weighting highlights the difference.

With applying this same idea to hydrological spatial patterns, the original image is synonymous with the observed spatial pattern, while the simulations should be like distortions of the original. In reality, the simulations are attempts at recreating the observations based on an understanding of the processes and forcings of the hydrological system. This can result in very different histograms for the observed and simulated spatial patterns. When the IMSE weights are calculated for each pixel value, there can be a large difference between the weightings applied to the observed and simulated spatial patterns. As such, IMSE appears more suitable for comparison when the spatial patterns have similar distributions of values. This measure could also be modified to reduce the impact of large differences. By calculating the mean absolute error rather than mean squared error, the impact of large residuals would be reduced.

The IMSE measure highlights the possibility for using weightings to make a standard MSE statistic compare something different. While Shannon's self-information has been used to define the weights here, other measures (e.g. terrain related measures) could alternatively be used to define informative locations.

6. CONCLUSION

This brief look at a method for comparing distorted images provides a number of ideas for the comparison of spatial patterns in hydrology. This method works predominantly in the measurement domain, but by using local variance as the event, some spatial characteristics can be incorporated. Further application of this method to spatial patterns from hydrological models will help in assessing its suitability for comparing spatial patterns.

7. REFERENCES

Grayson, R.B. and G. Blöschl eds., Spatial Patterns in Catchment Hydrology: Observations and Modelling, Cambridge University Press, 404 pp., Cambridge, 2000.

Grayson, R.B., G. Blöschl, A.W. Western and T.A. McMahon, Advances in the use of observed spatial patterns of catchment hydrological response, Advances in Water Resources, 25(8-12), 1313-1334, 2002.

Jetten, V., G. Govers and R. Hessel, Erosion models: quality of spatial predictions, Hydrological Processes, 17(5), 887-900, 2003.

Scheibe, T.D., Characterization of the spatial structuring of natural porous media and its impacts on subsurface flow and transport, PhD thesis, Department of Civil Engineering, Stanford University, 1993.

Tompa, D., J. Morton and E. Jernigan, Perceptually based image comparison, International Conference on Image Processing, Vancouver, BC, Canada, September 10-13, 2000.

Topper, T.N. and M.E. Jernigan, On the informativeness of edges, IEEE International Conference on Systems, Man and Cybernetics, Cambridge, MA, USA, 1989.

Wealands, S.R., R.B. Grayson and J.P. Walker, Hydrologic Model Assessment from Automated Spatial Pattern Comparison Techniques, International Congress on Modelling and Simulation, Townsville, Australia, 14-17 July, 2003.

Wealands, S.R., R.B. Grayson and J.P. Walker, Quantitative comparison of spatial patterns for hydrological model assessment - some promising approaches, Advances in Water Resources, submitted.

Western, A.W. and R.B. Grayson, Soil Moisture and Runoff Processes at Tarrawarra, In: Grayson, R. and G. Blöschl eds., Soil Moisture and Runoff Processes at Tarrawarra, Cambridge University Press, p. 209-246, Cambridge, 2000.