XIX International Conference on Water Resources
CMWR 2012
University of Illinois at Urbana-Champaign
June 17-22, 2012

BAYESIAN FUSION FOR HIGH-RESOLUTION RAINFALL SENSING USING PERVASIVE SENSORS

David J. Hill[*] and Farbod Farzan[*]

[*] Rutgers, the State University of New Jersey
Department of Civil and Environmental Engineering
Piscataway, NJ 08854 USA
e-mail: ecodavid@rci.rutgers.edu

Key words: Remote Sensing, Down Scaling, Bayesian Models

Summary. This paper develops a spatial Bayesian model to integrate radar-rainfall and qualitative pervasive sensor measurements to produce high-resolution rainfall field estimates.

1 INTRODUCTION

Recently, "smart" urban infrastructure has been suggested as a strategy to mitigate social and economic risks posed by extreme weather events and to improve the sustainability of urban centers.[1,2] Smart infrastructure is enabled by real-time sensing, modeling, optimization, and actuation, which permit the infrastructure to reason about the world around it and to autonomously reconfigure in order to meet changing needs. For example, researchers are exploring the use of smart sewer systems that combine real-time measurements of urban rainfall patterns with runoff models and remote actuated control structures to optimize the conveyance capacity of the sewer network during wet weather.[3] However, a significant barrier to the design and deployment of smart infrastructure is the inability of current sensing technology to characterize the environment at sufficient space and time resolutions to support the forecasts necessary to enable autonomous near-real-time decision making via predictive control algorithms. In the case of urban flooding, extrapolation from a recent study[4] indicates that the minimum resolution for rainfall observations is on the order of 0.5 km^2 and 1 minute. Unfortunately, this observational resolution is not feasible with traditional sensing technologies. Currently, information from multiple remote and embedded sensors, including satellites, weather radar stations, and rain gauges, is combined to achieve precipitation observations over broad areas. However, the resolution achievable by these methods is still too coarse to support many physics-based models that relate the rainfall process to actionable forecasts (e.g., flood depths).[5]

Because of their pervasiveness, the small, low-cost sensors that are embedded in everyday consumer products have begun to be explored for environmental monitoring. For example, sensors embedded in wireless communication devices[6,7] and vehicle-based automatic windshield wipers[8] have been explored for rainfall measurements. The latter of these studies showed that, through careful calibration, vehicle-based automatic windshield wiper sensors could provide quantitative point rainfall measurements at the vehicle location. As computational and

communication technology becomes smaller and cheaper, we expect that these devices will become increasingly pervasive and will possess increasing communication capabilities. For example, a survey of new vehicles shows that most automakers are beginning to equip their vehicles with rain-sensing windshield wipers. These wipers employ a sensor to determine the presence of rain and, based on the vehicle speed, they actuate and adjust the speed of the wipers. At the same time, most new vehicles are sold with integrated GPS chips and communication capabilities, which are used for navigation and emergency response, respectively.

This study develops a method to integrate radar-rainfall and qualitative pervasive sensor measurements to produce high-resolution rainfall field estimates. This method employs a Markov random field (MRF) to represent the variability of rainfall at scales smaller than the radar resolution. The MRF model uses noisy measurements from sparse binary sensors to accurately reconstruct a map of rainy versus not-rainy regions. This map is then used to redistribute the spatially averaged rainfall rate measured by the radar to concentrate it only in the rainy regions, leaving the dry regions with a zero rainfall rate. The next section describes this model in detail. The model's ability to reconstruct a synthetic rainfall field is then explored, particularly with respect to its requirements for binary sensor density and accuracy. Finally, conclusions and future work are discussed.

2 METHODS

The essence of the fusion method developed in this study is that qualitative, categorical measurements of the rainfall made by pervasive sensors can be used to identify the pattern of rainfall within each grid block of a radar image and that this pattern can be used to concentrate the rainfall into the wet portions of the radar block, such that the block average remains the same, but the rainfall rate is zero in the dry portions of the block. To accomplish this, we will sub-divide each grid block in a radar image into smaller pixels at the spatial scale of interest. To facilitate this discussion, the term "grid block" will hereafter be used to refer to the coarse-resolution grid-square of the radar image, while "pixel" will refer to the fine-resolution grid-square at which the rainfall field is reconstructed. For the remainder of the paper, we will assume that the categorical measurement at the pixel-scale will take the form of a binary variable in which a value of 0 represents dry conditions, and a value of 1 represents wet conditions. The method, however, is extensible to variables representing three or more categories.

Estimation of the categorical rainfall pattern at the sub-grid scale will be performed using an MRF model. MRFs are statistical models of a set of spatially correlated random variables that describe the relationship of neighboring variables using a model that relies only on a finite set of neighbors (i.e., a model that has an *N*-order Markov property).[9,10,11] The Markov property permits the spatial field to be subdivided using a relationship that defines whether or not two pixels are neighbors. A random field has the Markov property with respect to a particular neighborhood structure if the state of the pixel is conditionally independent given its neighborhood. Therefore, the probability that a pixel is wet or dry, given values at all other

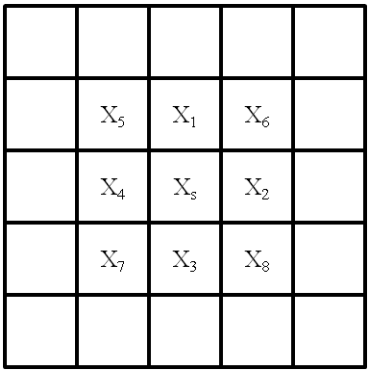

Figure 1: Neighborhood of pixel X_s.

pixels, depends only on the configuration in its neighborhood. Following accepted practice in image reconstruction, we will model wet/dry patterns using an autologistic model.[9] In this model, the neighborhood of a pixel is defined to be the set of all pixels immediately horizontally, vertically, and diagonally adjacent, as shown in Fig. 1.

Using this notation for neighboring pixels, the conditional probability for the qualitative rainfall measurement in pixel X_s can be computed as follows:

$$p(X_s|X_{i\neq s}) = \frac{exp\{-\beta_1 \sum_{i=1}^{4} I(X_s \neq X_i) - \beta_2 \sum_{i=5}^{8} I(X_s \neq X_i)\}}{\sum_{X_s=0}^{a-1} exp\{-\beta_1 \sum_{i=1}^{4} I(X_s \neq X_i) - \beta_2 \sum_{i=5}^{8} I(X_s \neq X_i)\}} \quad (1)$$

where β_1 and β_2 are parameters that are defined to be strictly greater than or equal to zero, and $I(X_s \neq X_i)$ is an indicator function that takes the value 1 when the condition is true and zero otherwise. The parameters β_1 and β_2 describe the dependence between a pixel and its type 1 (vertical and horizontal) and type 2 (diagonal) neighbors, respectively. Values close to zero reflect spatial fields where neighboring pixels are unlikely to be similar, whereas when β_1 and β_2 are strictly greater than 0, X_s is most likely to take the value of the majority neighboring class. By assigning different values to the parameters β_1 and β_2, the model can represent anisotropic conditions. If β_1 is greater than β_2, then similarity in the vertical/horizontal directions (i.e., along the principal axes of the image) is weighted more heavily, whereas if β_1 is less than β_2, then similarity in the diagonal directions is weighted more heavily.

If not all of the neighbors of pixel X_s are observed, then Eq. 1 can be modified to use likelihood weighting to calculate the contributions of the unobserved neighbors.

$$p(X_s|X_{i\neq s}) = \frac{exp\{-\beta_1 \sum_{i=1}^{4} p(X_s \neq X_i|X_j) - \beta_2 \sum_{i=5}^{8} p(X_s \neq X_i|X_j)\}}{\sum_{X_s=0}^{a-1} exp\{-\beta_1 \sum_{i=1}^{4} p(X_s \neq X_i|X_j) - \beta_2 \sum_{i=5}^{8} p(X_s \neq X_i|X_j)\}} \quad (2)$$

where X_j are the available neighbors of X_i. Note that if the sensor measurements are deterministic, then probability that an observed pixel takes the measured value is unity; thus, Eq. 2 reduces to Eq. 1 in the case that the pixels are fully observed by deterministic sensors. Thus, this method accounts for a partial contribution from unobserved, but estimated, pixels that is equivalent to the probability that those pixels are different than the value of X_s.

In real networks of pervasive sensors, however, it is expected that the sensors will malfunction, resulting in erroneous and misleading data. Thus, we employ Bayes' Theorem to pose the problem as one of estimating the true random field (X) from an incomplete set of noisy measurements (Y).

$$p(X_s|X_{i\neq s}, Y) = \frac{p(Y_s|X_s)p(X_s|X_{i\neq s})}{p(Y_S, X_s)} \quad (3)$$

where $p(X_s|X_{i\neq s})$ can be solved by Eq. 2, $p(Y_s|X_s)$ is a probability density table that defines the observation model of the sensors, and $p(Y_s, X_s)$ is the normalization constant. Eq. 3 can be viewed as the 2-D equivalent of a hidden Markov model, which blends the information from the set of incomplete noisy measurements and the 2-D spatial MRF model. Because of the interdependencies of neighborhoods, this modification leads to a system of non-linear equations that must be solved simultaneously. However, Eq. 3 can be applied to each pixel in an image to define a single cycle in an iterative algorithm for estimating the true field *X*. This process is similar to the iterated conditional modes[12] method for reconstructing images corrupted by noise, except that by using Eq. 2, our method is applicable to incompletely observed, noisy images.

Once the spatial probability distribution of the binary random variable is computed for all pixels in the image, then the rainfall rate in each pixel can be computed as

$$R_s = \bar{R} \frac{N p(X_s = 1|X_{j\neq s})}{\sum_{i=1}^{N} p(X_i = 1|X_{j\neq i})} \quad (4)$$

where X is the binary random variable indicating not-raining/raining conditions in pixel *s*, $\bar{R}$ is the grid-block-averaged precipitation rate measured by the radar, $p(X_i = 1|X_{j\neq i})$ is the probability that it is raining in the i^{th} pixel, and N is the number of pixels in the radar block.

3 CASE STUDY

To evaluate the performance of downscaling radar-rainfall-based MRF sub-pixel models, experiments were conducted to explore the ability of the incomplete MRF (iMRF), defined in

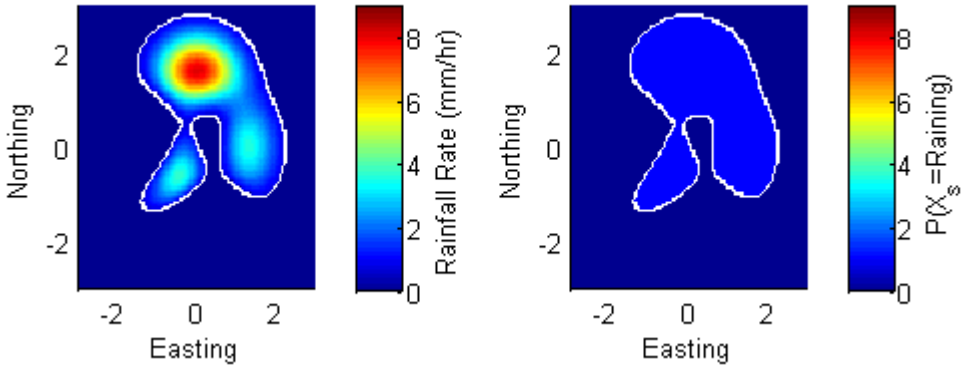

Figure 2: Synthetic rainfall field.

Eq. 2, and the incomplete, noisy MRF (inMRF), defined in Eq. 3, for estimating a binary field and for using that binary rainfall field to downscale a simulated radar block. These experiments are based on a synthetic rainfall field that was created by applying a threshold to the peaks function in MATLAB to remove negative values. The true rainfall field and its binary equivalent are shown in Fig. 2. From this figure, it can be seen that the field is composed of a continuous region of rain that covers approximately one-quarter of the total area. This rainy region has three areas of rainfall concentration, one in the north with the peak rainfall, and two side-by-side in the center of the field.

The first experiment explores the ability of the iMRF to estimate a binary field from a sparse set of observations. In this experiment, the true binary image was randomly sampled to create an incomplete set of observations, and the iMRF was used to estimate the true field from these observations. The iMRF parameters were set to be $\beta_1=\beta_2=0.5$. Accuracy of the estimated field is quantified in terms of false positives, where the model predicts rain when no rain actually exists, and false negatives, where the model predicts no rain when rain actually does exist. For each level of observational coverage, 600 realizations of the incomplete observation were used to compute the expected false positive and negative rates and their variance. Fig. 3 illustrates the relationship between estimation accuracy and level of observational coverage. These curves are characterized by a steep decline in the error when the observational coverage is low, which transitions to a shallow decline. The transition occurs around 10% and 20% coverage for the false positive and false negative rates, respectively. Furthermore, at 20% coverage, the estimation error is less than 2%. These results clearly indicate that the iMRF is capable of estimating a binary field from deterministic observations even when the level of observation is quite low.

To determine how well the iMRF and inMRF can estimate a binary field using a partial set of noisy observations, a series of experiments was undertaken, in which the true binary field was

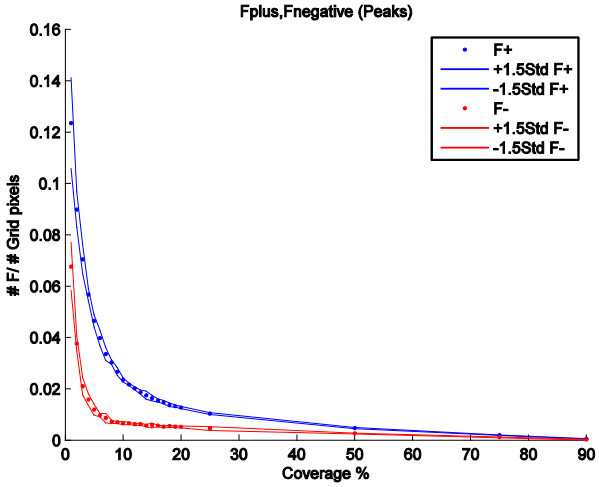

Figure 3: Expected false positive and false negative rate of iMRF for estimating the true binary wet/dry field as a function of the level of observational coverage.

sampled to produce a map with a specific observation density (as previously described). This map was then corrupted with noise from a simulated sensor malfunction. In this simulation, when a sensor malfunctioned, it produced a measurement with a value opposite that of the true field. The parameters (β) were again set to 0.5, and the observation model for the inMRF was specified to represent a 5% error rate in the sensors: $p(Y = 0|X = 1) = p(Y = 1|X = 0)$=0.05. We investigated observation rates of 10%, 20%, 30%, 40%, and 50% and sensor malfunction rates of 2%, 5%, 10%, 15%, and 20%. Again, 600 realizations of the observations were used to compute the expected false positive and false negative rates. Fig. 4 illustrates the trend in the error associated with increasing the observational coverage and the error rate. For low error rates, the accuracy of the iMRF improves with increasing observational coverage; however, for high error rates, the accuracy of the iMRF declines with increasing observational coverage. This result is due to the increasing number of misleading observations in the more highly observed cases, which threw off the iMRF because it is so strongly conditioned on the observations. This result implies that the sweet-spot for estimating a field from incomplete noisy images using the iMRF depends on balancing the observational coverage with the sensor accuracy, which would be challenging in real-world scenarios. On the other hand, the inMRF performs as expected: the error decreases with increasing coverage and increases with increasing sensor error rates. It is important to note that this behavior was achieved without tuning the error rate in the inMRF model to the actual error rate in the observations, indicating that the method is insensitive to this parameter.

If we assume that the field shown in Figure 2 represents a single radar block, then the block average rainfall rate is 0.68 mm/hr. Using this block average to represent the true field results in a maximum error and root-mean-squared-error (RMSE) of 7.4 and 1.4 mm/hr, respectively.

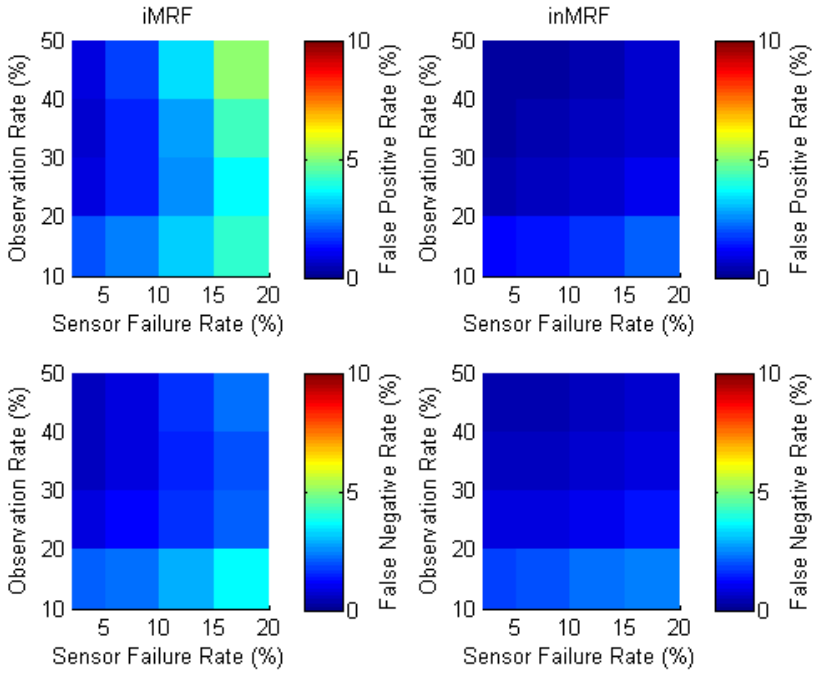

Figure 4: Expected false positive (top row) and false negative (bottom row) rates of iMRF (left column) and inMRF (right column) for estimating the true binary wet/dry field as a function of the level of observational coverage and rate of sensor malfunction.

Downscaling the block average using a set of deterministic binary observations that cover only 10% of the radar block via the methods developed in this paper reduces the maximum error and RMSE to 5.4 and 0.68 mm/hr, respectively.

4 CONCLUSIONS

This study developed a method to integrate radar-rainfall and qualitative pervasive sensor measurements to produce high-resolution rainfall field estimates. This method uses a map of the rainy/non-rainy regions within the radar block to redistribute the spatially averaged rainfall rate measured by the radar to concentrate it only in the rainy regions. We compared two different models for estimating the binary rainfall field. The first model (iMRF) treats the binary field as an MRF that is only partially observed, while the second model (inMRF) extends the iMRF to include a noisy sensor observation model. Our comparison revealed that both models produced good results when the sensor error rate was low and required only about one-quarter of the field to be observed to produce an accurate estimate of the true field. However, as the sensor error rate increased, the accuracy of the iMRF began to degrade with increasing observational coverage. On the other hand, degradation of the inMRF with increasing sensor errors was well behaved, and the inMRF maintained good estimation accuracy even for many erroneous sensor measurements. Using the methods developed in this paper to downscale the radar block reduces

the maximum error and RMSE of the rainfall field estimation by 27% and 51%, respectively. This error reduction is due solely to the concentration of the rainfall into the wet regions of the block. Clearly, if only binary observations are used at the pixel scale, the MRF-based downscaling method cannot resolve variability in the rainfall field, such as the peak rainfall in the north of the simulated field discussed in this paper. However, future work will explore the utility of sensors with more levels of measurement (e.g., no-rain/light-rain/moderate-rain/heavy-rain) to address this limitation.

REFERENCES

[1] Boyle, C., Mudd, G., Mihelcic, J., Anastas, P., Collins, T., Culligan, P., Edwards, M., Gabe, J., Gallagher, P., Handy, S., Kao, J.-J., Krumdieck, S., Lyles, L., Mason, I., Mcdowall, R., Pearce, A., Riedy, C., Russell, J., Schnorr, J., Trotz, M., Venables. R., Zimmerman, J., Fuchs, V., Miller, S., Page, S., Reeder-Emery, K. (2010) Delivering sustainable infrastructure that supports the urban built environment. Environmental Science and Technology, 44, 4836-4840.

[2] Brown, C. (2010). The end of reliability. Journal of Water Resources Planning and Management, 136(2), 143–145.

[3] Schutze, M., Campisano, A., Colas, H., Schilling, W., and Vanrolleghem, P. (2004). Real time control of urban wastewater systems—where do we stand today? Journal of Hydrology, 229 (3-4), 335–348.

[4] Berne, A., Delrieu, G., Creutin, J.-D., and Obled, C. (2004). Temporal and spatial resolution of rainfall measurements required for urban hydrology. Journal of Hydrology, 299, 166–179.

[5] Seo, B.-C. and Krajewski, W.F. (2010). Scale dependence of radar uncertainty: Initial evaluation of NEXRAD's new super-resolution data for hydrologic applications. Journal of Hydrometerology, 11, 1191–1198.

[6] Zinevich, A., Messer, H., and Alpert, P. (2010). Prediction of rainfall intensity measurement errors using commercial microwave communication links. Atmospheric Measurement Techniques, 3, 1385–1402.

[7] Leijnse, H., Uijlenhoet, R., and Strickler, J.N.M. (2007). Rainfall measurements using radio links from cellular communication networks. Water Resources Research 43, W03201.

[8] Haberlandt, U. and Sester, M. (2010). Areal rainfall estimation using moving cars as rain gauges—a modeling study. Hydrology and Earth Systems Science, 14, 1139–1151.

[9] Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society, Series B, 36, 162–236.

[10] Isham, V. (1981). An introduction to spatial point processes and Markov random fields. International Statistics Review, 49, 21–43.

[11] Ripley, B.D. (1988). Statistical Inference for Spatial Processes. Cambridge University Press.

[12] Besag, J.E. (1986). On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society, Series B (Methodological), 48(3), 259-302.