5th International Workshop on Climate Informatics
September 24-25, 2015
Hosted by the National Center for Atmospheric Research in Boulder, CO

ADDITIVE GAUSSIAN PROCESSES FOR BLENDING GAUGE AND SATELLITE RAINFALL DATA

Homer Strong[1], Andrew W. Robertson[2], Padhraic Smyth[1]

***Abstract*—Predicting ground rainfall from satellite estimates is useful as input for many applications, especially for areas with sparse rain gauges. We propose a predictive model based on an Additive Gaussian process (AGP) which can be viewed as the sum of a GP for the influence of the satellite estimate and a GP for the spatial distribution of rainfall between gauges. The hyperparameters for the covariance function estimates maximize the leave-one-out predictive densities. Initial results indicate that the proposed AGP model provides more accurate predictions compared with traditional kriging and inverse weighting methods.**

I. DATA AND MOTIVATION

Spatial estimates of ground rainfall are widely used for scientific applications such as modelling agricultural output and evaluating flood risk. Rain gauges can be sparse, especially in developing countries and in rural areas. Satellite estimates of rainfall provide uniform spatial coverage but are indirect estimates of precipitation, and tend to suffer from systematic bias.

The proposed model takes as input satellite measurements for the entire region and gauge observations at specific locations, and can predict the expected monthly gauge rainfall at arbitrary points. The predicted gauge rainfall is a smooth spatial field. This allows for predictions of ground rainfall at new points, for which there have been no gauge observations. Satellite pixels are measured on a grid, and rain gauges are both observed and predicted at arbitrary locations.The satellite estimate of rainfall for a particular location is taken to be the value of the pixel which contains the location.

The datasets used are the Tropical Rainfall Measuring Mission (TRMM) 3B43v7 satellite estimates, and gauge observations from a region covering Pakistan and northwest India in which gauges are sparse. Only data from principal monsoon months, July and August, are used. Figure 1 shows an example predictive mean surface based on some gauge observations for a particular month. The TRMM input is not shown.

Corresponding author: H Strong, University of California, Irvine, Irvine CA hstrong@uci.edu [1] University of California, Irvine [2] International Research Institute for Climate and Society, Columbia University, New York

II. METHODOLOGY

The mean function of gauge observations is modelled as the sum of 2 independent Gaussian processes (GPs). For a location s, let $Y(s)$ be the gauge observation of rainfall and X_s be the satellite measurement of rainfall in the pixel containing s. Then the model for ground observation given satellite estimates is $Y(s) = f(X_s) + g(s)$. The first GP, f, models the mean of the gauge rainfall as a nonlinear function of satellite measurement. The hyperparameters for the covariance functions are jointly learned.

We assume that the GPs f and g are independent, in which case their sum is again a GP with covariance $k_y(s, s') = k_f(s, s') + k_g(s, s')$. An interpretation of this assumption is that the satellite estimate is independent of location. Thus this model is analogous to regression kriging or universal kriging [1], but instead with a non-parametric mean function. The hyperparameters for k_y are the union of the hyperparameters of k_f and k_g. We select 2 covariance functions, which is natural since the component GPs operate on qualitatively different inputs and have different dimensionality. This model can be viewed as non-parametric regression on 3-dimensional inputs.

III. COMPARISON WITH PREVIOUS MODELS

This work proposes a model which has 2 major differences from past models which have been applied to this problem.

- Use of a non-parametric mean function for satellite estimates. Methods such as regression kriging, for example [5], treat the mean function of the gauge observations as a linear function of satellite measurements.
- We use hyperparameters which are minimized with respect to the leave-one-out (LOO) log validation density loss L_{LOO} [2], [3] using a gradient-based optimization algorithm. The more common

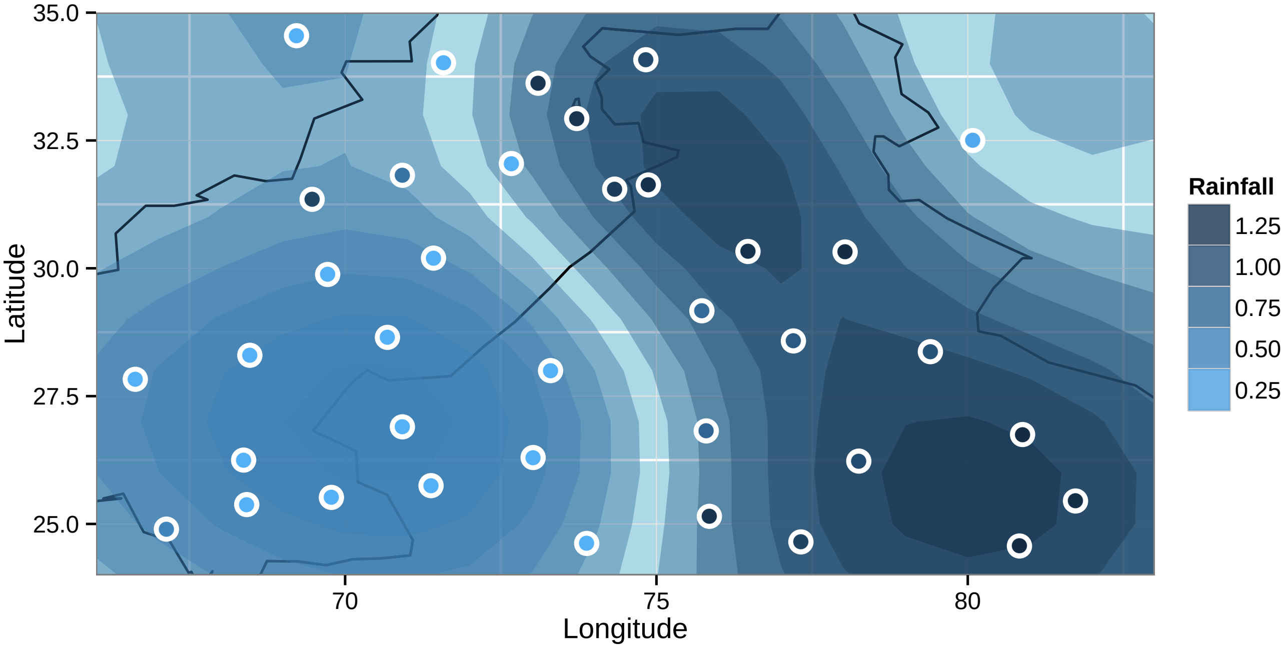

Fig. 1. Example map of observations and predicted surface for July 2002. Darker blue areas have more rainfall. The predictive mean surface is conditioned on the rainfall observations from gauges located at the dots.

marginal maximum likelihood (ML) estimates can be sensitive to covariance misspecification [4] and might overfit the training data.

	MAE	% bias	% RMSE change	RMSE
TRMM	0.45	8.31	0.00	0.64
Mean	0.36	5.26	17.25	0.52
IDW	0.37	5.14	16.07	0.53
RK	0.33	2.42	28.33	0.45
AGP	**0.30**	**1.53**	**32.89**	**0.43**

TABLE I
MAE = MEAN ABSOLUTE ERROR, RMSE=ROOT MEAN SQUARE ERROR. % RMSE CHANGE IS THE PERCENT IMPROVEMENT IN RMSE COMPARED WITH TRMM. THESE ARE MEDIAN VALUES FROM 300 TRIALS OF EVALUATION MONTHS. FROM EACH MONTH, 30% OF STATIONS WERE HELD OUT AND USED TO EVALUATE PREDICTIONS. THE TOTAL NUMBER OF STATIONS EACH MONTH VARIES FROM ROUGHLY 20 TO 40. BOLD VALUES IN THE TABLE INDICATE THE BEST VALUE IN EACH COLUMN.

IV. Evaluation and future directions

The predictive performance of this model is compared primarily with those of residual kriging (RK) which treats the satellite estimate as a linear covariate, i.e. $Y(s) = X_s\beta + g(s)$. Other baseline models include the raw TRMM estimates, an inverse distance weighting (IDW) model, and constant spatial mean of observed station values in the region. As seen in Table 1, the AGP model outperforms the others in each of the performance metrics.

There is significant interest from an applications perspective to make predictions at much finer time-scales, e.g., daily. However, the highly non-Gaussian nature of precipitation at a daily time-scale requires adaptations to the type of GP model proposed in this paper. For this study, a simple log transformation of rainfall was satisfactory for yielding a bell-shaped response distribution, but such a transformation is not sufficient for the distributional characteristics of daily rainfall observations.

References

[1] N. Cressie and C. K. Wikle, *Statistics for Spatio-Temporal Data*. Wiley, 2011.
[2] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.
[3] S. Sundararajan and S. S. Keerthi, "Predictive approaches for choosing hyperparameters in gaussian processes," *Neural Computation*, vol. 13, pp. 1103–1118, 1999.
[4] F. Bachoc, "Cross validation and maximum likelihood estimations of hyper-parameters of gaussian processes with model misspecification," *Computational Statistics Data Analysis*, vol. 66, pp. 55 – 69, 2013.
[5] A. Verdin, B. Rajagopalan, W. Kleiber, and C. Funk, "A bayesian kriging approach for blending satellite and ground precipitation observations," *Water Resources Research*, vol. 51, no. 2, pp. 908–921, 2015.