ASSIMILATION OF SVM-BASED ESTIMATES OF LAND SURFACE TEMPERATURE FOR THE RETRIEVAL OF SURFACE ENERGY BALANCE COMPONENTS

Giorgio Boni[2], *Federica Martina*[2], *Gabriele Moser*[1], *and Sebastiano B. Serpico*[1]

[1] University of Genoa, Dept. of Biophysical and Electronic Eng. (DIBE),
Via Opera Pia 11a, I-16145, Genoa (Italy), e-mail: sebastiano.serpico@unige.it
[2] CIMA Research Foundation, Savona University Campus,
Via Armando Magliotto 2, I-17100 Savona (Italy), e-mail: giorgio.boni@cimafoundation.org

1. EXTENDED ABSTRACT

Thanks to the recent missions for spaceborne Earth observation, remote-sensing data acquired with very different spatial resolutions and revisit times currently offer the opportunity to monitor and analyze the behaviors of the Earth surface at the required (local, regional, or global) observation scales. In particular, geostationary satellites (e.g., Meteosat Second Generation, MSG) currently allow wide-scale images to be acquired with very short revisit times (e.g., around 15 minutes). In order to exploit this great amount of information into dynamic physical models used to predict the evolution of hydrological processes, suitable techniques must be used. In this framework data assimilation (DA) represents the most effective approach to couple in a statistically optimal way dynamic models and observations to retrieve and estimate land surface variables [1, 2, 3].

In this paper, the DA perspective is adopted in order to address the problem of the estimation of the mass and energy exchanges at the soil surface in the framework of the application to flood prevention. DA can play an important role in inferring fluxes from successive remote sensing measurements of state variables such as land surface temperature (LST). LST time series over land contains a large amount of information about energy and mass fluxes at the soil surface: these exchanges have significant variations on a diurnal basis [4]. The signature of these large variations is evident in LST observations. On the other hand ground-based data of latent and sensible heat fluxes are available only for limited time periods and over small field experiment areas. Tower instruments within measurement networks are costly to install and maintain.

Here, a novel integrated method is proposed that assimilates LST estimates generated by a recently proposed algorithm based on support vector regression [5, 6, 7] into a model for surface energy flux estimation [8]. The

model is based on a simplified version of the surface energy balance equation, namely the *Force-Restore equation* [9], coupled with a soil wetness dynamic equation based on the antecedent precipitation index (API) [10]. The system dynamic equations are incorporated in an adjoint-state variational scheme as given by [11]. In particular, the proposed DA scheme aims at taking benefit from the capability of the techniques introduced in [5, 6, 7] of providing both a fully automatic LST estimation from satellite infrared data and a pixelwise evaluation of the statistics of the related LST-estimation error.

Many physically-based algorithms have been devised to retrieve LST from infrared space radiometry, typically involving prior information about the atmosphere (e.g., temperature and water vapor profiles) and the surface (e.g., emissivity) [12]. A different approach has recently been proposed in [5] based on pattern recognition using support vector machines (SVMs) [13]. SVMs represent a general family of supervised learning (i.e., classification, regression, or probability density estimation) techniques. In the case of LST-estimation, SVMs compute a nonparametric approximation of the relationship between satellite data and corresponding *in-situ* measurements, which are used for training purposes and can be collected, for instance, by a network of micrometeorological stations in the monitored area. This strategy can be theoretically proven to exhibit very good generalization and robustness [13] and, when applied to LST estimation, it proved complementary with respect to the aforementioned physically-based methods. Indeed, it was experimentally remarked that SVMs can generate more accurate estimates than these methods, even though with increased computational burden. Moreover, SVMs do not require prior information on atmospheric and surface properties, but they rely on the availability of *in situ* measurements to be used for training purposes [7].

Given these training data, a fully automatic SVM-based LST-estimation method was developed in [5] by automatically optimizing the values of the related model parameters. The formulation of SVM-based regression intrinsically involves regularization and kernel parameters [13, 14], whose values are often chosen through time-expensive and human-error prone "trial-and-error" procedures. The approach in [5] formalizes the parameter-optimization problem as the numerical minimization of the span-bound functional, which is a (tight) analytical upper bound on the leave-one-out regression error [14] and has also been found to be often strongly correlated with test-set hold-out errors [14, 5]. As the span bound is a nondifferentiable function of the SVM parameters, Powell's method is used to address the related numerical minimization [15, 5].

Moreover, in order to effectively integrate LST estimates in the above mentioned DA variational scheme, a further critical piece of information is represented by the LST regression-error statistics. However, classical error-estimation procedures, such as cross-validation or leave-one-out sampling, provide only global information. They characterize each map of LST estimates on the basis of a single numerical error index (e.g., a root-mean square error) without a pointwise characterization of the error associated with each pixel. Unlike this classical approach, the problem of pixelwise modeling the statistics of the LST regression error was addressed in [6]. This modeling problem has only recently been explored in the SVM literature. It has been proven in [16] that the intrinsically nonbayesian SVM approach to regression can be equivalently reformulated as a Bayesian "maximum-*a-posteriori*" regression with respect to suitably defined conditional likelihood and prior distributions. According to this formulation, the regression error on each sample is expressed as the sum of two independent multidimensional stochastic processes related to the SVM functional approximator and to the intrinsic uncertainty in the input data, respectively [16]. In this framework, two techniques were developed in [6] to model the pixelwise regression-error statistics by combining the nonstationary kernel-based characterization proposed in [16] for the error contribution due to the SVM functional approximator, and case-specific maximum-likelihood and confidence-interval estimators for the stationary error contribution due to data uncertainty.

In particular, the DA approach that is followed here to exploit the LST estimates and the related regression-error statistics is an adjoint-state variational scheme [11]. This scheme is based on the definition of a performance cost function that incorporates, through Lagrange multipliers, the system (LST and API) dynamics. The cost function is a quadratic penalty function that weights the misfit errors between model and observed LST according to the statistics of the observation errors estimated with the procedure described above.

The proposed integrated DA-SVM method is experimentally validated over a time series of MSG images acquired over Italy between 7:00 am and 6:00 pm local time by MSG-SEVIRI between August and September 2005, and endowed with corresponding ground temperature measurements collected by the ARSIA network of micrometeorological stations in Tuscany (Italy). Validation is performed by forcing hydrological models with the energy fluxes estimated by assimilating both standard (physically-based) and SVM-based LST estimates. The errors in the estimation of the different terms of the hydrologic cycle (compared with ground-based observations) are then quantified for the two assimilations and the performance of the integrated DA-SVM algorithm is quantified.

2. REFERENCES

[1] D. Entekhabi, G.R. Asrar, A.K. Betts, K.J. Benven, R. L. Bras, C.J. Duffy, T. Dunne, R.D. Koster, D.P. Lettenmaier, D.B. McLaughlin, W.J. Shuttleworth, M.T. van Genuchten, M. Wei, and E.F. Wood, "An agenda for land surface hydrology research and a call for the second international hydrological decade," *Bulletin of the American Meteorological Society*, vol. 80-N10, pp. 2043–2058, 1999.

[2] G. Boni, F. Castelli, and D. Entekhabi, "Sampling strategies and assimilation of ground temperature for the estimation of surface energy balance components," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 165–172, 2001.

[3] F. Caparrini, F. Castelli, and D. Entekhabi, "Variational estimation of soil and vegetation turbolent transfer and heat flux parameters from sequences of multisensor imagery," *Water Resources Research*, vol. 40, pp. W12515, 2004.

[4] W. E. Nichols and R. H. Cuenca, "Evaluation of the evaporative fraction for parameterization of the surface energy balance," *Water Resources Research*, vol. 29 NO.11, pp. 3681–3690, 1993.

[5] G. Moser and S. B. Serpico, "Automatic parameter optimization for support vector regression for land and sea surface temperature estimation from remote-sensing data," *IEEE Trans. Geosci. Remote Sensing*, vol. 47, pp. 909–921, 2009.

[6] G. Moser and S. B. Serpico, "Modeling the error statistics in support vector regression of surface temperature from infrared data," *IEEE Geosci. Remote Sensing Letters*, vol. 6, no. 3, pp. 448–452, 2009.

[7] G. Moser and S. B. Serpico, *Land and sea surface temperature estimation by support vector regression*, in Kernel methods for remote sensing data analysis, ed.: G. Camps-Valls, L. Bruzzone, John Wiley & Sons., 2009.

[8] F. Sini, G. Boni, F Caparrini, and D. Entekhabi, "Estimation of large-scale evaporation fields based on assimilation of remotely sensed land temperature," *Water Resources Research*, vol. 44, 2008.

[9] Z. Hu and S. Islam, "Prediction of ground surface temperature and soil moisture content by the force restore method," *Water Resources Research*, vol. 31, pp. 2531–2539, 1995.

[10] R.K. Linsley and M.A. Kohler, "Predicting the runoff from storm rainfall," *U.S. Weather Bureau Research paper 34*, 1951.

[11] F. Caparrini, F. Castelli, and D. Entekhabi, "Mapping of land-atmosphere heat fluxes and surface parameters with remote sensing data," *Boundary Layer Meteorology*, vol. 107, pp. 605–633, 2003.

[12] P. Dash, F.-M. Gottsche, F.-S. Olesen, and H. Fischer, "Land surface temperature and emissivity estimation from passive sensor data: theory and pratice - current trends," *Int. J. Remote Sens.*, vol. 23, no. 13, pp. 2563–2594, 2002.

[13] V. N. Vapnik, *Statistical learning theory*, Wiley Interscience, 1998.

[14] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Mach. Learn.*, vol. 46, pp. 131–159, 2002.

[15] W. H. Press, S. A. Teukolsky, W. T. Wetterling, and B. P. Flannery, *Numerical recipes in C*, Cambridge University Press, Cambridge (UK), 2002.

[16] J. B. Gao, S. R. Gunn, and C. J. Harris, "A probabilistic framework for SVM regression and error bar estimation," *Machine Learning*, vol. 46, pp. 71–89, 2002.