Interactive comment on "Hydrologic calibration of paired watersheds using a MOSUM approach" by H. Ssegane et al.

J. Seibert (Editor)

This manuscript addresses the important issue of quantifying land-use change impacts. The reviewers appreciate the data set and found the approach interesting.

However, they also expressed concerns, which are rather fundamental:

1) Does the short pre-treatment period really allow 'calibration' of the paired catchments for the subsequent impact quantification? (based on my experience I agree with the reviewer that the relationships of paired catchments might vary seasonally and that, thus, at least one year of pre-treatment observations seems to be fundamental)

>> *The water table elevation and flow calibration period identified by the MOSUM approach from the longer January 01, 2009 to March 31, 2012 period was only from March 1, 2010 – March 31, 2012, just over two years. This period of minimal disturbance was between pine planting in January 2010 and the final site preparation for switchgrass sowing (in May 2012) on the intercropped site. This period spanned the very wet periods of September 2010 and August 2011 to very dry periods of spring-summer of 2010 and 2011 and average rainfall in 2010. Recently, Bren and Lane (2014) demonstrated that good calibration (Nash-Sutcliffe efficiency > 0.8) by simple linear models could be achieved after 100 days of data. So this length of the calibration period defined by the MOSUM method should be adequate. Below are the total numbers of days for each pair of watersheds for the stable periods as estimated by the MOSUM approach.*

1. *D0 vs. D2: 01 January 2011 to 31 March 2012: 456 days (over one annual cycle: MOSUM did not detect any structural break or significant instability in the regression coefficients). Monitoring on D0 started in late 2010*
2. *D1 vs. D2: 01 March 2010 to 31 March 2012: 762 days (over 2 annual cycles)*
3. *D3 vs. D2: 01 December 2009 to 31 July 2011: 608 days (≈ 1 ½ annual cycles)*

Field monitoring on watersheds D1, D2, and D3 at the study site started in 1987 and the data was reported in several studies (e.g., McCarthy et al., 1991; Amatya et al., 1996; 1998; 2000; 2003; 2006; 2007; Amatya and Skaggs, 2011) while watershed D0 was established in 2009 at the onset of this study. The above listed studies provide the chronology of management activities at the three watersheds over the past 25 years. Therefore, there are two "true-calibration" periods (1988-1990 and 2007-2008) for watersheds D1 and D2 where both watersheds were under mature pine. The 1988 - 1990 calibration period has previously been used to quantify effects of controlled drainage and silvicultural operations using a paired watershed approach (refer to above references). Therefore, for purposes of quantifying treatment effects under switchgrass intercropping, the MOSUM based calibration relationships in addition to these historical calibration relationships will be compared to quantify treatment effects and associated uncertainty.

The uniqueness of the MOSUM based calibration relationships is the use of most recent watershed response data close to the treatment period by eliminating periods where the calibration relationships between e.g., D1 and D2 may be influenced by external factors. The value of "true calibration" period spanning wet and dry seasons is the fact that the assumption of "a consistent" relationship between *control and treatment watersheds is probably often met. The MOSUM approach provides a statistical technique to test this assumption incase external factors may shift this consistent relationship. Clausen and Spooner (1993) state that the paired watershed design assumes a consistent, quantifiable, and predictable relationship between watershed response variables while Loftis et al. (2001) illustrate that moderate correlation coefficients ($r \geq 0.6$) are adequate to detect treatment effects for paired watershed studies. For this study all developed coefficients of determination (R^2) are greater than 0.8 (or $r > 0.89$). The choice of data under a stable regression period meets the requirement of the relationship to be consistent by eliminating data that shifts this relationship and thus increase model uncertainty, while the high R^2 meets the requirements for quantifiable and predictable relationships. The robustness of the MOSUM approach to detect temporal shifts in model coefficients of time series (significant structural instability of a regression relationship or a model) is documented elsewhere (de Jong et al 2013; Verbesselt and Herold, 2012' Chu et al., 1995). These references were provided in the original manuscript.*

2) How is the serial correlation of the water table data considered?

>> *In the original regression analysis, we used a bootstrap geometric mean regression analysis (Efron and Tibshirani, 1994) using a MATLAB code (Trujillo-Ortiz et al., 2014). The bootstrap method implemented is a non-parametric case approach developed by Efron (1979) which assumes the residual errors to be independent and identically distributed (i.i.d). The method independently resamples a single case [Xi, Yi] from the original data [X, Y] with replacement, n times (n = size of original data) to form a single bootstrap. We applied 1000 bootstraps to generate the regression coefficients in Figure 1. One thousand bootstraps are considered adequate to determine confidence intervals (Efron and Tibshirani, 1994). However, the higher the number of bootstraps, the higher is the accuracy. Regression coefficients and the corresponding 95% confidence intervals were estimated as the averages of 1000 bootstrap resampled results and the corresponding percentiles.*

However, because of serial correlations inherent in daily time series (an issue the reviewers have correctly highlighted), we have re-analyzed the regression relationships between the control and treatment watersheds using a resampling technique that accounts for serial correlation in daily time series. This technique is called "block bootstrap" (Politis, 2003; Kundzewicz and Robson, 2004; Khaliq et al., 2009). In this approach, the original data is resampled in predetermined blocks for a large number of times to estimate regression coefficients. This method also incorporates effects of serial correlations higher than the first order dependencies. Sørensen et al. (2009) used a block length of 14 days. His choice of greater lengths gave similar results. For the re-analysis in our study here, block bootstrapping was performed using a time series function "tsboot" with a fixed block length of 50 days and 10,000 bootstrap resamples in R (R Development Core Team, 2015). Use of 20, 30, and 100 days block lengths gave similar results. Geometric mean regression was used to estimate the regression coefficients for each bootstrap resample. To ensure replicability of the results the same arbitrary number of 4711 was used to seed the random number generator. Refer to Table 1 for the regression coefficients (intercept, slope, and their respective 95% confidence intervals) and the corresponding performance metrics (coefficient of determination: R^2, Nash-Sutcliffe Efficiency: NSE, and root mean squared error: RMSE). Also, refer to figure 2 below for the recomputed distributions of the regression coefficients for the water table elevation (WTE) between D1 vs D2 and D3 vs. D2 based on 10,000 bootstrap resamples (compare to figure 8 in the manuscript).

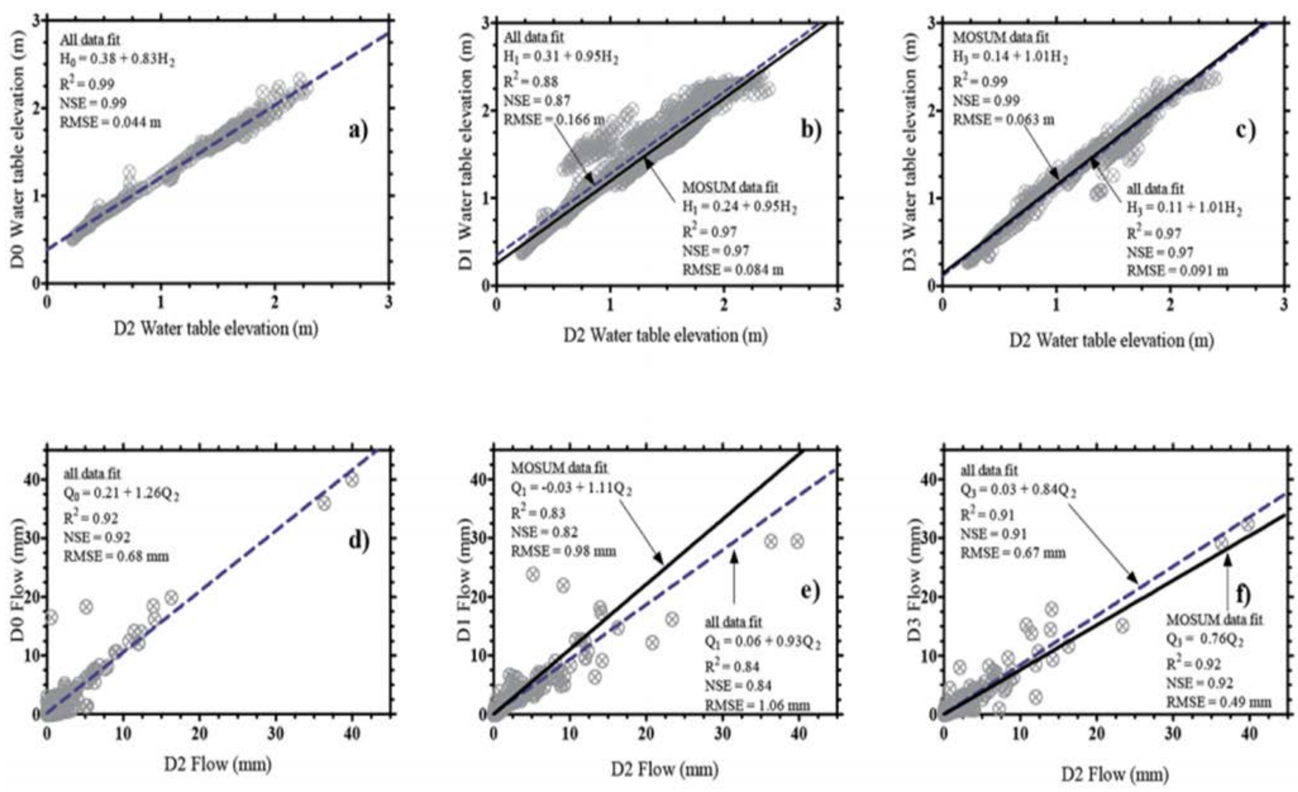

Figure 1: This is figure 7 in the manuscript based on case bootstrapped geometric regression analysis and 1,000 bootstrap resamples

Table 1: Results based on block bootstrapped geometric mean regression analysis and 10,000 bootstrap resamples

Paired	*Data*	*Intercept[a] A (Al, Au)*	*Slope[a] B (Bl, Bu)*	*R^2*	*NSE*	*RMSE*
WTE0 vs WTE2	*all data[b]*	*0.38 (0.33, 0.45)*	*0.83 (0.78, 0.86)*	*0.99*	*0.99*	*0.04*
WTE1 vs WTE2	*all data*	*0.32 (0.20, 0.59)*	*0.94 (0.80, 1.00)*	*0.88*	*0.87*	*0.16*
	MOSUM data	*0.24 (0.18, 0.29)*	*0.95 (0.91, 0.99)*	*0.97*	*0.97*	*0.08*
WTE3 vs WTE2	*all data*	*0.11 (0.03, 0.19)*	*1.01 (0.94,1.07)*	*0.97*	*0.97*	*0.09*
	MOSUM data	*0.14 (0.07, 0.21)*	*1.01 (0.97,1.05)*	*0.99*	*0.99*	*0.06*
Q0 vs Q2	*all data[b]*	*0.31 (0.20, 0.44)*	*1.24 (1.13, 1.38)*	*0.92*	*0.92*	*0.78*
Q1 vs Q2	*all data*	*0.07 (-0.07, 0.20)*	*0.94 (0.78, 1.30)*	*0.84*	*0.83*	*1.07*
	MOSUM data	*-0.02 (-0.13, 0.04)*	*1.10 (0.83, 1.58)*	*0.83*	*0.82*	*0.99*
Q3 vs Q2	*all data*	*0.03 (-0.02, 0.09)*	*0.85 (0.76, 1.05)*	*0.91*	*0.91*	*0.68*
	MOSUM data	*-0.02 (-0.07, 0.05)*	*0.77 (0.72, 0.82)*	*0.91*	*0.91*	*0.69*

[a]Letters l and u refer to the lower and upper 95% bootstrapped confidence intervals, respectively
[b]Water table elevation (WTE, m) and flow (Q, mm) relationships between watersheds D0 and D2 are based on all data because no structural instability was detected using the MOSUM approach

Results of the original analysis (Figure 1) and the re-analysis (Table 1) show similar coefficients with exact values in 9 out of the 20 coefficients and a negligible variation (second decimal place) in the remaining coefficients. Examination of the 95% CI (Table 1) shows that the Figure 1 and Table 1 regression coefficients are not significantly different. Both analyses use geometric mean regression to generate the coefficients; the difference is in the resampling approach. The close similarity of the results shows that for this dataset, the independent case bootstrapping with replacement may have scrambled the time dependencies and thus the results are similar to the 50-days block bootstrapping.

Although, there are no significant differences between the original analysis and the re-analysis, the revised manuscript will describe the "block bootstrap" resampling method and report results based on this method because it was primarily developed to address issues of serial correlation.

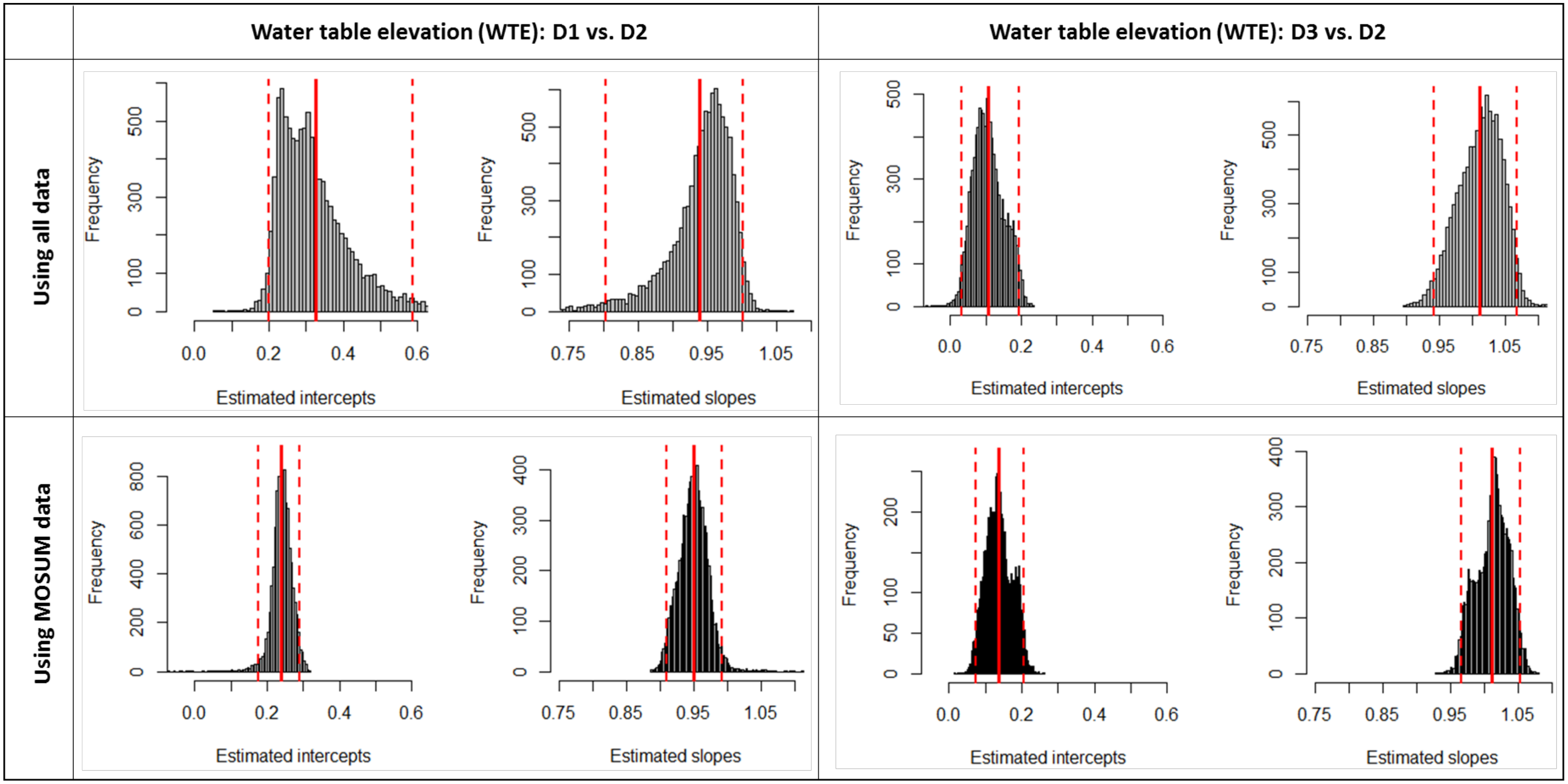

Figure 2: Differences in uncertainty of regression coefficients due to difference in the data (all data: top plots versus temporally stable data determined by moving sums of recursive residuals: MOSUM). The dotted vertical lines represent the 95% CI while the solid line is the estimated coefficient (mean of 10,000 bootstrap resamples). Uncertainty is represented by the width of the 95% CI. A small width is indicative of less uncertainty.

3) What can be said about the potential influence of the side-by-side located watershed and their water tables?

>> The natural soil surface gradient from the north boundary of D0 to the south boundary D3 is less than 0.3 m over the 1600 m distance for a slope of less than 0.025 %. The low gradient continues for over 5 km both north and south of the borders of the research site. Consequently, lateral subsurface flow under natural conditions is very low. Subsurface drainage from the watersheds, however, is driven by a system of parallel 1.2 to 1.5 m deep drainage ditches spaced 100 m apart. According to "drainage theory", the midpoint between parallel drains of equal elevation can be treated as a no-flow boundary. The boundaries between the watersheds are along the midpoints of the fields separating the parallel ditches of the different watersheds. The elevations of all of the drains across all of the watersheds are the same. It is possible that a small amount of water seeps from one watershed to the other due to the uneven boundary conditions at the surface of the fields. These conditions would occur when the water tables are lower or near the elevation of the bottom of the ditches. However, in these conditions, drainage rates would be low due to lower water surface gradients toward the ditches and the lower hydraulic conductivity values of the deeper soil layers. These seepage rates would be negligible compared to the overall subsurface rates to the ditches. Surface drainage does not occur on the watersheds due to microtopography created by the 25 cm raised beds for the pine trees that are parallel to the ditches.

Since seepage rates between watersheds are negligible, the water tables measured in the fields located two parallel drainage ditches from the watershed boundaries would not be affected by the adjacent watersheds. Fipps and Skaggs (1986) showed theoretically that the water table in a field located two parallel drains from a water source (such as an unlined irrigation canal) would not be affected by the seepage from the source. Water table gradients and drainage intensities in the Fipps and Skaggs (1986) study were much greater than those observed on our watersheds.

These additional clarifications will be incorporated in the revised manuscript.

While also the other reviewer comments are useful and important, these three questions on the experimental design are most fundamental to be addressed in the authors' response.

References

Amatya, D.M. and Skaggs, R.W. 2011. Long-term hydrology and water quality of a drained pine plantation in North Carolina, USA. *Trans of the ASABE*, 54(6):2087-2098

Amatya, D.M., Gregory, J.D, and Skaggs, R.W. 2000. Effects of controlled drainage on the storm event hydrology in a loblolly pine plantation. *J. of the Amer. Water Resou. Assoc*. 36(1):175-190.

Amatya, D.M., Gilliam, J.W., Skaggs, R.W., Lebo, M., and Campbell, R.G. 1998a. Effects of Controlled Drainage on Forest Water Quality. *J. of Environ. Qual*. 27:923-935(1998).

Amatya, D.M., R.W. Skaggs, and J.E. Hughes. 2003. Effects of an Orifice and a Weir on the Hydrology and Water Quality of a Drained Forested Watershed. *Southern Journal of Applied Forestry*, 27(2): 130-142.

Amatya, D.M. and R.W. Skaggs. 2008. Hydrologic Effects of Thinning a Pine Plantation on North Carolina Coastal Plain. In E.W. Tollner and A. Saleh (Eds.), 21st Century Watershed Technology: Improving Water Quality and the Environment, Proceedings of the 29 March-3 April 2008 Conference, Concepcion, Chile, Publication date 29 March 2008, ASABE Publication Number 701P0208cd.

Amatya, D.M., Skaggs, R.W, and Gregory, J.D. 1996. Effects of controlled drainage on the hydrology of a drained pine plantation in the North Carolina coastal plains. *J. of Hydrology*, 181(1996), 211-232.

Amatya, D.M., Skaggs, R.W, Blanton, C.D, and Gilliam, J.W. 2006a. Hydrologic and water quality effects of harvesting and regeneration of a drained pine forest. In *Proc. of the ASABE-Weyerhaeuser sponsored Int'l Conference on Hydrology and Management of Forested Wetlands*, eds. Williams and Nettles, New Bern, NC, April 8-12, 2006.

Clausen, John C., and Jean Spooner. *Paired watershed study design*. No. PB-94-154820/XAB; EPA--841/F-93/009. Environmental Protection Agency, Washington, DC (United States). Office of Wetlands, Oceans and Watersheds, 1993.

Chu, C.-S. J., Hornik, K., and Kaun, C.-M.: MOSUM tests for parameter constancy, *Biometrika*, 82, 603–617, 1995.

de Jong, R., Verbesselt, J., Zeileis, A., and Schaepman, M. E.: Shifts in global vegetation activity trends, *Remote Sensing*, 5, 1117–1133, doi:10.3390/rs5031117, 2013.

DiCiccio, Thomas J., and Bradley Efron. "Bootstrap confidence intervals."*Statistical science* (1996): 189-212.

Efron, B. and Tibshirani, R. J.: *An Introduction to the Bootstrap*, CRC Press, Boca Raton, FL, 1994.

Fipps, G., and Skaggs, R.W. 1986. Effect of canal seepage on drainage to parallel drains. TRANSACTIONS of the ASABE 29(5):1278-1283.

Kundzewicz, Zbigniew W., and Alice J. Robson. "Change detection in hydrological records—a review of the methodology/revue méthodologique de la détection de changements dans les chroniques hydrologiques." *Hydrological Sciences Journal* 49, no. 1 (2004): 7-19.

Khaliq, M. N., Taha BMJ Ouarda, Philippe Gachon, Laxmi Sushama, and Andre St-Hilaire. "Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers." *Journal of Hydrology* 368, no. 1 (2009): 117-130.

Loftis, Jim C., Lee H. MacDonald, Sarah Streett, Hariharan K. Iyer, and Kristin Bunte. "Detecting cumulative watershed effects: the statistical power of pairing." *Journal of Hydrology* 251, no. 1 (2001): 49-64.

McCarthy, E.J., Skaggs, R.W, and Farnum, P. 1991. Experimental determination of the hydrologic components of a drained forest watershed. *Trans. Amer. Soc. Agr. Eng*., 34(5):2031-2039.

Politis, Dimitris N. "The impact of bootstrap methods on time series analysis."*Statistical Science* 18, no. 2 (2003): 219-230.

Sørensen, Rasmus, Eva Ring, Markus Meili, Lars Högbom, Jan Seibert, Thomas Grabs, Hjalmar Laudon, and Kevin Bishop. "Forest harvest increases runoff most during low flows in two boreal streams." *AMBIO: A Journal of the Human Environment* 38, no. 7 (2009): 357-363.

Trujillo-Ortiz, A., R. Hernandez-Walls and B. Martin-Atienza. (2014). *bootgmregresspi*: Bootstrap Interval Prediction of a Single Value for a Geometric Mean Regression (Bootstrap Reduced Major Axis Regression, RMA). A MATLAB file. [WWW document]. URL http://www.mathworks.com/matlabcentral/fileexchange/26251-bootgmregresspi

Verbesselt, J., Zeileis, A., and Herold, M.: Near real-time disturbance detection using satellite image time series, *Remote Sens. Environ.,* 123, 98–108, 2012.