Remote Sensing in Ecology and Conservation
Open Access

ORIGINAL RESEARCH

Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality

Jonathan R. B. Fisher[1], Eileen A. Acosta[1], P. James Dennedy-Frank[2], Timm Kroeger[1] & Timothy M. Boucher[1]

[1]The Nature Conservancy, 4245 N Fairfax Dr, STE 100, Arlington, Virginia 22203, USA
[2]Department of Earth System Science, Stanford University, 473 Via Ortega, Room 140, Stanford, California 94305, USA

Keywords
Land cover, land use, remote sensing, spatial resolution, value of information, water fund, water quality

Correspondence
Jonathan R. B. Fisher, The Nature Conservancy, 4245 N Fairfax Dr, STE 100, Arlington, Virginia 22203, USA.
Tel: 703 841 5300;
E-mail: jon_fisher@tnc.org

Editor: Nathalie Pettorelli
Associate Editor: Graeme Buchanan

Received: 29 March 2017; Revised: 15 June 2017; Accepted: 19 July 2017

doi: 10.1002/rse2.61

Remote Sensing in Ecology and Conservation 2018;**4** (2):137–149

Abstract

Remote sensing offers an increasingly wide array of imagery with a broad variety of spectral and spatial resolution, but there are relatively few comparisons of how different sources of data impact the accuracy, cost, and utility of analyses. We evaluated the impact of satellite image spatial resolution (1 m from Digital Globe; 30 m from Landsat) on land use classification via ArcGIS Feature Analyst, and on total suspended solids (TSS) load estimates from the Soil and Water Assessment Tool (SWAT) for the Camboriú watershed in Southeastern Brazil. We independently calibrated SWAT models, using both land use map resolutions and short-term daily streamflow (discharge) and TSS load data from local gauge stations. We then compared the predicted TSS loads with monitoring data outside the model training period. We also estimated the cost difference for land use classification and SWAT model construction and calibration at these two resolutions. Finally, we assessed the value of information (VOI) of the higher-resolution imagery in estimating the cost-effectiveness of watershed conservation in reducing TSS at the municipal water supply intake. Land use classification accuracy was 82.3% for 1 m data and 75.1% for 30 m data. We found that models using 1 m data better predicted both annual and peak TSS loads in the full study area, though the 30 m model did better in a subwatershed. However, the 1 m data incurred considerably higher costs relative to the 30 m data ($7000 for imagery, plus additional analyst time). Importantly, the choice of spatial resolution affected the estimated return on investment (ROI) in watershed conservation for the municipal water company that finances much of this conservation, although it is unlikely that this would have affected the company's decision to invest in the program. We conclude by identifying key criteria to assist in choosing an appropriate spatial resolution for different contexts.

Introduction

New technology in remote sensing (satellites with more advanced sensors, as well as drones) is providing imagery at higher spatial and temporal resolutions than previously available (along with additional spectral bands), driving interest in using these new data for potentially more accurate analyses (Boyle et al. 2014). Many studies have used a variety of remote data sources, yet relatively few have examined how the choice of data source (e.g. relatively low-spatial resolution satellite imagery, relatively high-spatial resolution satellite imagery (<30 m), aerial photography, or drones; Turner 2014) impacts the accuracy, cost and utility of analyses. Where accuracy has been assessed, data with higher spatial resolution usually led to more accurate estimates (Geza and McCray 2008; Boyle et al. 2014), although often studies assumed that higher-resolution data are superior and that differences in coarser-resolution data represent errors (Cotter et al. 2004; Chaubey et al. 2005; Lin et al. 2010). As "high-resolution" is a subjective term and what is considered "high" changes over time, we focus here on comparing "higher" (1 m) "lower" (30 m) resolutions instead; we also focus on spatial rather than temporal resolution.

Remote sensing is used for a wide array of purposes in conservation science (Turner et al. 2003), and previous studies have analyzed the impact of changing the spatial resolution of data on analyses including land cover classification (Boyle et al. 2014), coral bleaching (Andréfouët et al. 2002), soil erosion (Molnár and Julien 1998), and both streamflow and water quality measures (Cotter et al. 2004; Chaubey et al. 2005; Geza and McCray 2008; Lin et al. 2010; Moriasi and Starks 2010). Coarser digital elevation model (DEM) resolution generally leads to lower predictions of erosion, streamflow, nitrogen, and phosphorous, presumably because coarse DEMs underestimate slope length and steepness (Molnár and Julien 1998). However, very few of these studies evaluated very high-resolution data (≤5 m) or compared the costs of different data sources; in addition, they typically resampled a single data set rather than comparing actual data sets at their native resolutions (with several exceptions: Boyle et al. 2014; Di Luzio et al. 2005; Moriasi and Starks 2010; Lin et al. 2010). Depending on the landscape pattern, resampling has the effect of increasing some classes and decreasing others, changing edge patterns and patch distribution as the scene is aggregated (Moody and Woodcock 1995). It also ignores that lower-resolution imagery often has many more spectral bands which are used in the classification process.

The cost (both for imagery acquisition and labor) of using higher-resolution data requires evaluating the value of information (VOI) gained from presumably increased accuracy in a given decision-making context. Value of information is a decision analytic concept (Howard 1966) not yet widely applied in environmental or infrastructure analyses (Keisler et al. 2014), which identifies the maximum amount a decision maker would be willing to pay for additional information prior to making a decision.

The case examined here involves managers of a municipal drinking water treatment company (EMASA) who seek information about how much watershed conservation reduces concentrations of total suspended solids (TSS) at the plant raw water intake. The VOI of the higher-resolution (1 m) data is the expected avoided cost of investing in a sediment control measure (watershed conservation or a competing alternative) that is not the most cost-effective one. Errors in the TSS concentration estimates could lead to either unmet peak-season water demand or unnecessarily large expenditures on new conventional treatment infrastructure. However, a high-cost analysis increases the cost of the natural infrastructure option (watershed protection), reducing its competitiveness with conventional engineering solutions.

We had two objectives for this study: assessing the influence of spatial resolution on the accuracy of land use classification and estimated TSS (sediment) loads at the plant intake (the TSS concentration can be calculated from loads and streamflow); and evaluating the technical and economic trade-offs associated with using spatial data of different resolutions. We report both streamflow and TSS loads to provide a more complete model.

The analysis in this paper is part of an evaluation of the economic viability and impact of planned conservation and restoration activities on TSS concentrations in a watershed protection payment for ecosystem services (PES) program in the Rio Camboriú watershed in Brazil (Kroeger et al. in preparation). The program, funded primarily by EMASA, aims to reduce TSS concentrations in municipal intake water and their associated costs. The study area is the source watershed for Balneario Camboriú, a coastal city in southeastern Brazil that is attracting increasing numbers of tourists (Ferreira et al. 2009). The combined population of Balneario Camboriú and Camboriú city quadruples from about 200,000 to over 800,000 during the peak tourist season from December to early March, stressing the water supply in a watershed lacking large-scale water storage infrastructure (Kroeger et al. 2017). Higher TSS loads and associated water losses (from sludge discharge and filter flushing) increase costs for EMASA, through chemicals used in sediment removal, electricity use, and landfilling, with treatment water losses imposing additional costs from forgone revenue from reduced water sales during peak demand times (Kroeger et al. 2017). Water users bear additional costs from averting measures such as the recently imposed requirement that new multi-unit residential structures install cisterns to bridge potential short-term municipal supply shortfalls. Observational evidence and studies from similar watersheds suggest key sources of sediment stream loading include: dirt roads lacking best management practices (Minella et al. 2008; Guimarães et al. 2011), cattle entering streams causing bank and channel erosion, and a lack of riparian buffers on pasturelands (Palhares et al. 2012; Teixeira Guerra et al. 2014).

Based on initial feasibility analyses, EMASA decided to invest in reducing treatment costs and water losses by protecting the remaining natural forests from conversion to other land uses and restoring degraded areas with high TSS loading. Initial analyses indicated that these much smaller investments would be cost-competitive with developing new water sources from inter-basin transfers or adding large-scale storage capacity.

To investigate the effectiveness of conservation in reducing TSS, we tested the accuracy of modeling streamflow and TSS loads with two data sources at different spatial resolutions: 1 m (Digital Globe and a local DEM) and 30 m resolution (Landsat and SRTM DEM). For each data set (1 m and 30 m), we classified land cover (in

2004 and 2012/2013), reclassified land cover into land use, and tested that land use classification against ground-truth data. We developed a Soil and Water Assessment Tool (SWAT: Arnold et al. 1998; de Almeida Bressiani et al. 2015) model to estimate streamflow and water quality (both total and peak TSS loads) for each dataset, and compared the streamflow and water quality estimates at each resolution to each other and to out-of-sample data (streamflow and TSS load data that was not used to train the model). Finally, we compared the cost and time to process each data source.

We anticipated that the higher spatial resolution data would allow us to capture finer details of the landscape, and detect individually small land use changes that might be missed with coarser satellite data such as Landsat (30 m), which could significantly affect sediment export and targeting of TSS-reducing conservation and restoration interventions (Kroeger et al. in preparation). Observational evidence and expert consultations suggested that most land use change occurred in small (sub-30 m size) patches (e.g. at the forest edge) and thus might not be detected with 30 m data.

During our initial analysis (limited to 1 m data), we realized that the 1 m data required significant disk space and processing time (e.g. 2 weeks for a land use change model run with 1 m data versus 9 h with 30 m data). This led us to evaluate the impact of varying the resolution on the performance of the SWAT model (compared to out-of-sample data) as well on the costs of the overall analysis.

Materials and Methods

Field work and imagery

Field Work

Field work was conducted from October 28 to 30, 2014 to provide ground reference points to properly classify the imagery. Real-time satellite visualization software (OziExplorer) was used with the 2012 higher-resolution imagery to identify different apparent land use types and landmarks (such as intersections and bridges), which were then selected and visited. Photographs and a GPS waypoint were taken (using a Garmin GPSMAP64 with approximately 2 m accuracy) at each of 539 ground reference points visited (on average, 61 per land use class, from 27 to 105), and the land cover and land use of the point was described (for the 430 points used for land cover/land use rather than other features like road intersections). While the GPS accuracy was insufficient to ensure that each point was within the correct 1 m pixel, along with the photos it sufficed to describe the area that the point was within. These points were later used to calibrate and test the accuracy of the land use classification at both 1 m and 30 m resolution.

Higher-resolution

The first imagery source analyzed was Digital Globe's Quickbird and Worldview 2 satellites with a spatial resolution of 0.6 m for Quickbird (images taken on 10/10/2003 and 7/14/2004) and 0.5 m for Worldview 2 (06/25/2012). Although the rainfall in the study area varies somewhat throughout the year, the land cover is consistent, so we were not concerned about using field data in October for imagery taken in June or July. For example, the areas used to grow rice are not allowed to return to pasture or other land covers, ponds are permanent (ephemeral or rain dependent), and of course forest persists as well.

While the two satellites have different sensors, we used the same four spectral bands (red, green, blue, and near-infrared) from both (see the supporting information section for more detail). Due to cloud cover in 2003 and 2004, images from two separate dates were needed to cover the area. To merge the two data sources, we georeferenced the 2003 and 2004 imagery to the 2012 data using a spline methodology (with >200 tie points for each image). Note that while the imagery would have allowed for a land use classification at 0.6 m and 0.5 m, predicting future land use change (Kroeger et al. in preparation) required resampling the data to 1 m to match the best available DEM (a 1 m aerophotogrammetric product from Secretaria do Desenvolvimento Econômico Sustentável [SDS] 2010). As such, all derived products from the higher resolution imagery have a spatial resolution of 1 m.

Lower resolution

The second imagery source was Landsat 7 (03/29/2003) and Landsat 8 (04/17/2013) satellites with 30 m resolution (path 220, row 79), from which we used the full set of spectral bands (USGS 2015). We chose 2003 and 2013 because among available cloud-free images they were the closest in time to 2004 and 2012. The 30 m imagery was georeferenced to the 2012 1 m imagery to ensure proper alignment.

Land cover and land use classification

Land cover (physical land type) and land use (purpose land is put to) were classified for each image. For example, a recently cut plantation may have a relatively bare land cover at the moment a satellite image was taken (and be classified accordingly), but tree regrowth will change the cover (while the land use remains as a plantation). As such, identifying the *land use* for a given parcel as plantation better represents the land cover (and sediment export) over time than an initial classified land cover of either forest (unlogged) or bare ground. However, the first step in determining land use is to identify land cover.

Land cover classification

Local experts in Camboriú identified seven land cover classes that were both prevalent and relevant in thinking about land cover changes likely to occur that affect water quality: Water, Bare, Pasture, Rice, Impervious, Plantation, and Forest. Land cover image processing was done using Feature Analyst 5.1.21 for ArcGIS Desktop 10.2.

We ran the land cover classification process separately with the higher-resolution data and the lower-resolution data. For each land cover class, feature class polygons were created and a supervised classification was run using different land cover classification parameters (LCCP) that most suited that particular land cover class. Variations in LCCP included the imagery input bands and type (whether treated as reflectance or texture for instance), the input representation matrix pattern, any masking of the input imagery using other spatial data or pixel values to be excluded, setting output as vector or raster, and any post processing such as aggregation of small regions. After an initial run, each class was evaluated for accuracy through visual comparison with imagery and details from ground reference points. Each feature class polygon was then adjusted, added or removed, and other parameters changed to increase accuracy. The individual land cover class products were merged serially to produce a complete land cover map. Additional steps were taken to increase land cover classification accuracy as described in the supporting information section.

Transforming land cover into land use

We identified which transitions were expected to be rare (e.g. pasture changing into natural forest), and which were common (e.g. pasture being replaced by plantation) through consultation with local experts, and devised a set of rules (Table 1) to alter the land cover for both time periods to reflect land use (see the supporting information section for additional detail). An area threshold for each land cover patch was applied to some rules to improve accuracy as described in the supporting information section.

Land use classification accuracy assessment

Ground reference points were used to assess land use classification accuracy at both resolutions, using the procedure of Landis and Koch (1977). While these points informed the land cover classification, they were not formally used in the classification process (e.g. to seed supervised classification, or manually reviewed to ensure classification was correct at each point) so they serve as a valid comparison dataset.

Each of the 430 points with land use information was classified among our seven classes (discarding points used for image rectification or other purposes) and converted to raster. We snapped this ground reference raster to each land use layer and compared them to produce a confusion matrix (see Results below).

We used four metrics of accuracy. "Producer's accuracy" (sometimes simply called "accuracy") indicates the fraction of ground reference points in each class classified correctly (e.g. if 18 of the 20 places with impervious surface on the ground were correctly classified as impervious in the 1 m analysis, then accuracy would be 90.0%). "User's accuracy" (also sometimes called reliability) asks what fraction of classified pixels in each class are correct (e.g. if 27 pixels classified as impervious had a ground reference point, and 18 were verified as impervious, then

Table 1. Rules to transform land cover to land use. For example, if we detected pasture in 2004, and forest in 2012, we reclassified both time periods to be plantation.

Original		Modified			
2003/2004 Land Cover	2012/2013 Land Cover	2003/2004 Land Use	2012/2013 Land Use	Area Threshold (1 m)	Area threshold (30 m)
Pasture	Forest	Plantation	Plantation	>1,220 m^2	None
Forest	Pasture	Plantation	Plantation	>1,347 m^2	None
Forest	Plantation	Plantation	Plantation	None	None
Plantation	Forest	Plantation	Plantation	None	None
Bare	Forest	Plantation	Plantation	None	None
Forest	Rice	Plantation	Rice	>473 m^2	None

Where an area threshold is listed, this reclassification was only done on patches larger than the specified threshold.

reliability = 66.7%). The "overall accuracy" divides the number of point / pixel combinations that matched by the total number of point / pixel combinations to provide a single number for accuracy across all classes. Finally, the overall kappa statistic for the classification indicates how much better the classification is than would be expected by chance.

Modeling water quality with SWAT

We used the Soil and Water Assessment Tool (SWAT: Arnold et al. 1998; Gassman et al. 2007; de Almeida Bressiani et al. 2015) to assess the effects of data resolution on: (1) SWAT's estimates of the TSS loads (total and peak) at the water treatment plant and one upstream gauge; and (2) SWAT's predictive power in estimating future loads. SWAT models the complex dynamics of a watershed through simple representation of processes such as surface runoff, infiltration and shallow groundwater flow; evaporation and transpiration as well as vertical soil flow; and plant growth, which responds to local conditions. It also includes both surface erosion and in-channel sediment transport (including channel erosion and deposition), which are important for modeling sediment in the Camboriú River due to observed sediment deposition at the EMASA intake. SWAT's simple process representations, including its lack of full spatial connectivity, do not perfectly represent actual hydrologic processes; instead, connectivity and simple process adjustments are contained in parameters used to fit the model to observed data. Thus, care is needed to avoid over-fitting the model to calibration data. We address this using a split-sample calibration approach in which 2015 streamflow and TSS load estimates are held out of the calibration sample for out-of-sample model performance tests.

We built two SWAT models: one (higher-resolution) with the 1 m land use layer (derived from Quickbird and Worldview 2 imagery) and DEM data from Secretaria do Desenvolvimento Econômico Sustentável, and one (lower-resolution) with the 30 m land use layer derived from Landsat imagery and a DEM from the Shuttle Radar Topography Mission (SRTM) (USGS 2014). For both models, we used a soil map provided by Santa Catarina state and data from 5 climate stations. A detailed rural road polygon was converted to raster and included in the 1 m land use model; that same polygon was resampled at 30 m to include in the 30 m model. Both models had 13 sub-basins; the higher-resolution model had with 971 HRUs and the lower-resolution model 779 HRUs. All HRUs were kept (no threshold applied) to ensure all potential sediment sources could contribute to the model. Each model was developed independently and parameterized separately, using a defined set of 11 parameters (Table 2). We chose these parameters based on their applicability and likelihood to simulate streamflow processes in this watershed.

We first calibrated the model to daily streamflow and then to daily TSS load at two sampling locations within the watershed: "Canoas", the outlet of a northwestern subwatershed of the study area with a drainage area of 48 km^2, using streamflow data from 1/1/2014 to 12/31/14 and TSS load data from 4/9/14-12/31/14; and "EMASA", the water treatment plant intake located near the full watershed outlet, with an area of 137 km^2, using both streamflow and TSS load data from 5/27/14-12/31/14. Both TSS load time series have some gaps, but include measurements on more than 50% of the dates. Streamflow and TSS load data were collected by EPAGRI (the Santa Catarina State agricultural and fisheries extension agency) in 2014 and 2015 for this project, using pressure transducers and optical turbidity sondes at each site. Rating curves were built for head to streamflow using a standard power curve and 57 and 31 in-situ streamflow measurements for Canoas and EMASA, respectively, following standard processes (Kennedy 1984). Linear rating

Table 2. Parameters adjusted within the SWAT (Soil and Water Assessment Tool) models at both 30 m and 1 m resolution.

Parameter	Meaning	Component	Adjust
ALPHA_BNK	Exponential baseflow recession factor for bank storage	Water	Value
ALPHA_BF	Exponential baseflow recession factor for shallow aquifer flow	Water	Value
ESCO	Soil evaporation compensation factor	Water	Value
GW_DELAY	Exponential aquifer recharge delay	Water	Value
GWQMN	Threshold aquifer depth for baseflow	Water	Value
LAT_TTIME	Exponential later soil flow travel time	Water	Value
SOL_AWC ()	Available water capacity of soil layer; done per layer	Water	Relative
ADJ_PKR	Peak rate adjustment for sediment routing in sub-basin	TSS	Value
SPCON	Linear parameter to control sediment reentrainment	TSS	Value
SPEXP	Exponential parameter to control sediment reentrainment	TSS	Value
USLE_K	USLE equation soil erodibility	TSS	Relative

curves were built for turbidity to TSS concentration (then converted into load) using laboratory measurements of TSS concentration based on 35 grab samples at each site, following standard processes (Rasmussen et al. 2009). No log transformation was performed because the errors were equally symmetric, linear, and homoskedastic in the linear case as log-transformed case, so the simplest approach was used, as has been done in other cases (e.g. Tena et al. 2011).

We calibrated with the goal of finding parameterizations that provided roughly similar (and maximal) values of Nash–Sutcliffe Equilibrium (NSE: Nash and Sutcliffe 1970), and similar (and minimal) percent bias (PBIAS, a measure of the average model error). We also considered, with a lesser emphasis, maximizing the models' R^2. We focused on matching the NSE since it most strongly represents the highest TSS load peaks, which we expect to carry the largest volume and concentration of TSS and thus cause the largest expenses to the water treatment plant. We included PBIAS in the calibration because the total TSS load has significant effects on the amount of sediment that will be filtered by the plant, which correlates well with chemicals application (Kroeger et al. in preparation).

We assessed model accuracy based on the same calibration metrics, but calculated on out-of-sample monitoring data from 1/1/2015-11/6/2015 compared against model outputs for streamflow and TSS load. These statistics provide one indication of how well the parameterized models predict future behavior. NSE and PBIAS values during this out-of-sample period that are nearly the same as those during the in-sample calibration period indicate that a model has predictive power similar to model performance during calibration, while a reduction in the statistical fit with out-of-sample data indicates that the model may have been overfitted and cannot predict watershed response with the accuracy suggested by the calibration statistics. If the higher-resolution and lower-resolution models predict out-of-sample responses that vary in different ways from the observations, and thus have different out-of-sample statistics, it indicates that input data resolution affects the predictive power of these models. This approach tests the model's predictive power, rather than being a sensitivity analysis which simply investigates how model output changes. Additional detail on the development of the models is available in the supporting information section.

Results

Land use

There were significant differences in the total area of each land use class between the resolutions (although the ranking of classes by area was the same, Table 3). Resolution also impacted accuracy of land use classification, with the overall accuracy of 82.3% at 1 m higher than the overall accuracy of 75.1% at 30 m, and 1 m generally having higher producer's accuracy and user's accuracy for most classes (Table 3). Table S1 shows the impact of the land use classification rules on the area of each class (it was especially significant for plantation, the area of which was more than doubled, see the Supporting Information section for more details). Table S2 further highlights the different results for the different resolutions (e.g. ranging from rice which was only 1% different between the two, up to bare ground which was 86% different). Maps showing classified land use are available in the supporting information section (Fig. S1 for 1 m data, Fig. S2 for 30 m data).

The rules used to adjust land cover to land use had a major impact on the classification (Table S1). The amount of plantation almost tripled at both resolutions, offset by a decrease in natural forest and pasture. The

Table 3. Comparison of the area and accuracy of each land use class at 1 m and 30 m for 2012/2013.

	2012 Area by Class (1 m), m^2	Producer's Accuracy (1 m)	User's Accuracy (1 m)	2013 Area by Class (30 m), m^2	Producer's Accuracy (30 m)	User's Accuracy (30 m)
Water	456,430	100.0%	100.0%	382,500	50.0%	93.3%
Bare	2,324,451	75.0%	63.6%	4,311,900	63.6%	59.2%
Pasture	21,876,744	77.5%	88.6%	27,962,100	90.5%	66.4%
Rice	9,044,171	92.3%	88.9%	8,968,500	88.9%	96.0%
Impervious	610,447	90.0%	66.7%	597,600	33.3%	81.8%
Plantation	10,659,607	80.6%	72.0%	18,496,800	69.3%	75.4%
Forest	91,709,658	81.6%	94.7%	75,814,200	84.0%	88.7%
All	136,681,508			136,533,600		
Overall Accuracy and Kappa:			2012 (1 m)			2013 (30 m)
Overall accuracy			82.3%			75.1%
Kappa			0.786			0.696

validity of this approach was tested on the 1 m land cover for areas that had the rule applied to them by comparing the before and after rule-applied land cover of 100 randomly distributed (spatially) 500 m × 500 m sample areas. The before and after rule-change land cover data was visually checked against the 2004 and 2012 images to see if it matched the newly assigned land use class. For all the rules combined, the accuracy improved from 24% to 85% (note: this is only for areas we could apply the rules to, and is not an overall accuracy assessment).

The total area of each land use class is one important factor in determining water quality, but the spatial distribution of land use is also important. Accounting for spatial location as well as total area by land use classes, 22.8% of the study area is in a different land use class (Fig. 1) at the two resolutions.

Water quality

While differences in land use are interesting, we are more focused on whether any differences in estimated water quality (TSS load) might impact the watershed protection program. Despite substantial differences in land use, differences in estimated water quality between the two resolutions were more moderate. The spatial resolution of the land use (and DEM) was found to affect total predicted TSS load, peak values of predicted TSS load, the accuracy of TSS load predictions (Fig. 2, Table 4), and the spatial allocation of modeled sediment contributions (Fig. S3).

In the larger EMASA watershed, the 1 m model significantly better predicts total TSS load (represented by PBIAS) than the 30 m model, although both are "satisfactory" according to the evaluation criteria for monthly models proposed by Moriasi et al. (2007). For peak TSS loads (to which NSE is more sensitive), the 1 m TSS model is significantly better than the 30 m model in EMASA. While it is slightly below the "satisfactory" cutoff value of 0.5 for monthly models, the corresponding value for a daily model is substantially lower (Moriasi et al. 2007), so 0.48 NSE is satisfactory for our purposes while an NSE of 0.16 (30 m model) may still be useful but is significantly worse. Both models are "satisfactory" at

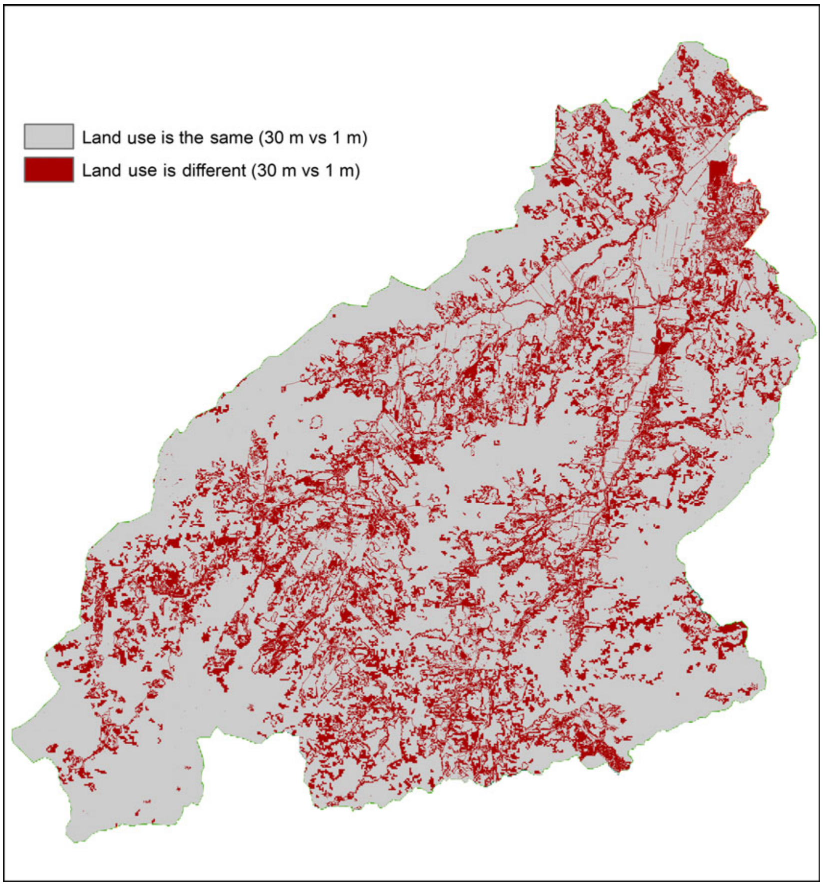

Figure 1. Map showing 1 m pixels of agreement (gray) and disagreement (red) between 30 m and 1 m land use for 2012/2013. 22.8% of the 1 m pixels within the study area had a different land use class in the 30 m data.

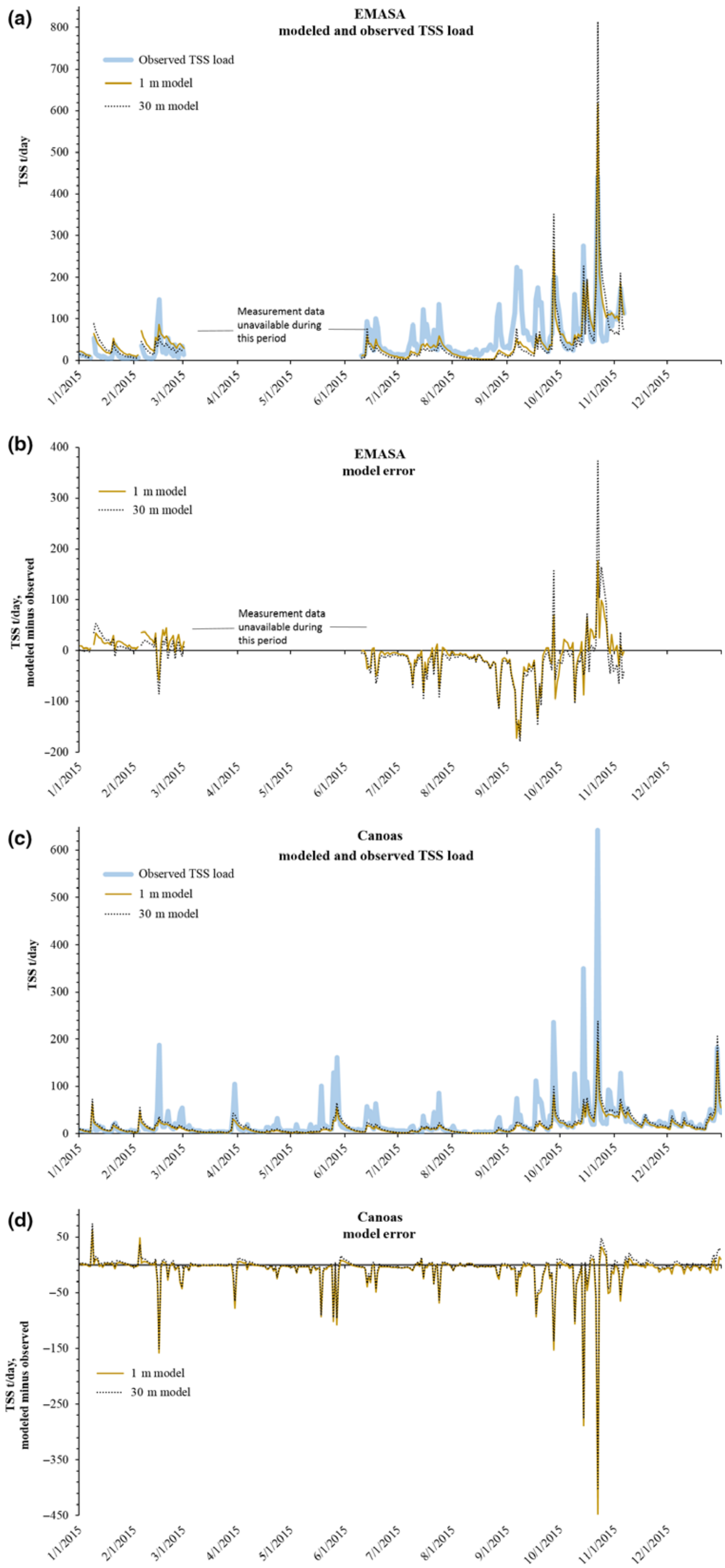

Figure 2. Comparison between observed and modeled TSS load at the two sampling stations. (A) EMASA observed and modeled TSS, (B) EMASA model error (modeled minus observed TSS), (C) Canoas observed and modeled TSS, (D) Canoas model error (modeled minus observed TSS). TSS, total suspended solids.

Table 4. TSS load and streamflow model performance statistics for out-of-sample testing (1/1/2015-11/06/2015) on two sampling stations in the river. TSS, total suspended solids.

		In-sample data (1/1/2014-12/31/2014)			Out-of-sample data (1/1/2015-11/06/15)		
		NSE	PBIAS	R^2	NSE	PBIAS	R^2
CANOAS	Streamflow (30 m)	0.53	−23.97	0.69	0.67	18.12	0.73
	Streamflow (1 m)	0.66	−12.47	0.71	0.64	22.15	0.72
	TSS (30 m)	**0.39**	5.85	0.40	0.51	23.97	0.62
	TSS (1 m)	**0.38**	27.44	0.41	**0.42**	39.27	0.61
EMASA	Streamflow (30 m)	0.63	−11.24	0.72	0.54	−4.00	0.66
	Streamflow (1 m)	0.71	−4.54	0.74	0.53	−6.17	0.81
	TSS (30 m)	0.50	18.85	0.60	**0.16**	24.38	0.52
	TSS (1 m)	0.63	8.42	0.63	**0.48**	15.01	0.57

Values in bold indicate unsatisfactory performance for monthly models according to Moriasi et al. (2007)'s guidance on standardized model evaluation (note that those criteria were developed for monthly models, and should be relaxed for daily models, per Moriasi et al. 2007). NSE, Nash–Sutcliffe Equilibrium; PBIAS, percent bias i.e. a measure of the average model error.

predicting total streamflow in EMASA according to Moriasi et al.'s (2007) criteria, with "very good" PBIAS numbers. In contrast, at Canoas (the subwatershed covering about 35% of the EMASA watershed) the 30 m model does a somewhat better job predicting total TSS load (with a "good" PBIAS), and the two models perform roughly similarly ("satisfactorily") predicting total streamflow. Since the total annual TSS load is of most importance for the EMASA water treatment plant, the lower TSS load PBIAS at EMASA for the 1 m model is particularly important, although the higher NSE value indicates that this model also better predicts peak TSS loads. However, the models underestimate some of the smaller peaks and may overestimate the largest peak, indicating that we are not capturing some nonlinear effects.

The difference in PBIAS between in-sample and out-of-sample runs is informative. We attempted to minimize PBIAS, a measure of relative error and thus of the integrated streamflow error, though were of course unable to do so perfectly. However, the 30 m and 1 m TSS models at EMASA have nearly the same change in PBIAS, suggesting that the lower PBIAS of the 1 m model might be at least as correctable as that of the 30 m model, further enhancing confidence in the 1 m model. In fact, the change in PBIAS in the 1 m model is always of lower amplitude than that of the 30 m model except for TSS at EMASA, where it is only larger by 1%; this suggests that the higher-resolution model may have better predictive power than the 30 m model for integrated streamflow. In addition, the changes in NSE similarly tend to have lower amplitude for the 1 m model, except for streamflow at EMASA, where they are slightly higher. While clearly not definitive, this is suggestive of the ability of finer scale data to better predict hydrologic outcomes. However, as a caution, we also note that the total PBIAS of the 1 m TSS estimate at Canoas is quite large during the out-of-sample testing, so this stability may not lead to better predictions.

Cost

Some costs of this assessment were independent of data resolution: the need to conduct field work (a week of time plus travel and equipment costs), and the need to have a reasonably powerful computer available for processing (we used a dual core 2.4 GHz PC with 24GB of RAM and SSD hard drive). Many of the data collection and processing steps are the same in both cases (for example, obtaining soil maps and entering appropriate parameters into the SWAT soil database).

Costs that differed included imagery acquisition ($6969 for the 1 m imagery vs. $0 for 30 m), and staff time. It is difficult to estimate the staff time requirements independently at 30 m and 1 m resolutions due to the experience gained from the earlier 1 m analysis prior to beginning the 30 m analysis. Nevertheless, the 1 m data was much more time-consuming to work with. The original land use classification took approximately 4 times as long at 1 m (560 h vs. 140 h), the land change modeling (not directly used in this analysis, but part of related work) took about 35 times as long for each run to complete at 1 m compared to 30 m, and the SWAT modelling took approximately 5 times as long for each run to complete at 1 m.

Discussion

We found that data resolution impacted both the land use classification and water quality modeling. Assessing the value of information (VOI) of the higher-resolution data (both in this case, and in general) requires a closer examination of the decision-making context.

It is likely that we have underestimated the overall accuracy of the land use classification. The selection of ground reference points favored areas where the land cover/land use was not apparent from the imagery alone, and this accuracy assessment would be different if a random or stratified random sampling strategy had been used. For example, 67% of the study area is forest (according to the 1 m classification), but only 17% of the ground reference points were in the forest class. An area-weighted assessment would likely have scored better at both resolutions as the large blocks of native forest are easy to classify.

For the 30 m analysis, the least accurate classes – "Bare," "Impervious," and "Water" – all tended to have small patches: either fragmented or narrow (for dirt roads and streams). Bare and impervious are subject to rapid change, and were difficult to correctly classify in the urban area as they were mixed together; in particular, bare areas consisting of hard packed gravel were sometimes classified as impervious.

Here, 90% of the discrepancy between the area-weighted 30 m and 1 m land use estimates is due to three classes: the 30 m data has (1) less forest, (2) more pasture, and (3) more plantation. These differences have significant effects on the hydrologic response of the watershed. Pasture has a curve number (a measure of surface runoff vs. infiltration as a function of precipitation; curve numbers typically go from ~30 to 100) about 5 larger than plantation, which in turn has a curve number about 5 larger than forest. The runoff vs. infiltration difference between these depends on the strength of precipitation, but for a reasonably strong storm of 50 mm (of which there are about 5 per year on average in Camboriú), pasture will have about 6.4 mm of surface runoff versus 3.7 mm for plantation and 1.7 mm for forest. In addition, pastures will have 3.5 times as much erosion as plantation, which will have about 25% more erosion than forest. Therefore, estimating these land uses, especially differentiating between plantation and pasture, is crucial to correctly target land management to reduce TSS load at the treatment plant.

Higher-resolution imagery improved land use classification accuracy, but for our purposes, land use was only an input to modeling water quality. The finding that the 30 m TSS load estimates had an unsatisfactory NSE for the larger EMASA watershed is a cause for concern, and supports the notion that higher spatial resolution enables more accurate hydrologic analysis results. The fact that the smaller Canoas watershed 30 m model had a higher NSE than the 1 m model demonstrates the importance of using out-of-sample validation data to test for model overfit (rather than simply assuming higher resolution data will always produce more accurate models). To inform future analyses, we need to understand whether the choice of data resolution could have altered the predicted impact and return on investment (ROI, or benefit-cost ratio) of the Camboriú watershed conservation program for EMASA and what resolution would be most informative in different contexts.

Using the 1 m data and discounting annual program costs and benefits using Brazil's estimated social discount rate of 3.85% (Fenichel et al. 2016), Kroeger et al. (in preparation) found that the ROI of the watershed conservation program for EMASA exceeded 1 for time horizons of 44 years or longer. These time horizons are typical for conventional water treatment infrastructure (U.S. EPA 2002), meaning the program generates net benefits for EMASA solely from its TSS load reduction impact, ignoring third-party co-benefits such as biodiversity protection and flood risk reduction. In contrast, using the 30 m data, predicted annual TSS loads (without the conservation program) at the EMASA intake are 11.7% lower than for the 1 m data (13,964 t vs. 15,823 t). Assuming the relative effectiveness of conservation interventions remains unchanged, estimated TSS reductions and associated benefits therefore are also 11.7% lower, with the ROI < 1 for any time horizon. Consequently, use of the 30 m model could have changed EMASA's decision to adopt the program. In this case, however, even before our modeling had been completed, the municipal government considered the expected co-benefits of the program sufficiently important to pursue regulatory changes that will allow EMASA to incorporate program operational costs into municipal water user fees. We do not know whether the 12% lower modeled TSS reduction benefits or the higher PBIAS for estimated annual TSS loads of the 30 m model would have changed that decision, but their interest in co-benefits makes this unlikely. Thus, we cannot exclude that the VOI of using the 1 m versus 30 m data might have been zero in this case.

However, in a different decision context where EMASA had required program ROI to exceed 1, the VOI of using 1 m data would have been positive as it would have avoided the profit-reducing decision of pursuing conventional solutions instead of investing in the program. In that case, with a program ROI approaching 1 based on 30 m data, a more compelling case could be made to repeat the analysis at higher resolution to either confirm or refute those findings. It is also possible that the finding of a positive return on investment in this case will be helpful to convince other water treatment companies to consider conservation. To support others in choosing whether to use higher-resolution data (across a broader set of contexts), in Table 5, we have listed several factors that we believe (based on our experience) should be considered when choosing the appropriate spatial resolution for different contexts. While we found a varying

Table 5. Considerations for selecting the appropriate spatial data resolution for a given analysis.

Consideration	Use higher-resolution data	Use lower-resolution data
Project budget	Higher budget	Lower budget
Study area size (affecting imagery cost & processing time)	Smaller study area	Larger study area
Required accuracy/precision (and risk associated with error)	High accuracy & precision needed, low tolerance for error	Accuracy & precision less critical, more error acceptable
Need to explore/refine model (affecting elapsed time for multiple model runs)	Model inputs and process well known, few model runs likely required	Exploration of model development needed, many model runs and refinements expected
Thresholds in decision making (e.g. if the estimates change slightly, a different decision is needed)	Initial estimates are close to an important threshold (e.g. ROI near 1)	Initial estimates appear to be safely distant from key thresholds
Size of land cover/land use patches in the landscape	Smaller patches	Larger patches
Size distribution of individual land cover/land use change patches	Most land cover change occurs in patches smaller than lower-resolution pixels	Most land cover change occurs in patches larger than lower-resolution pixels
Size (and heterogeneity) of parcels identified for conservation interventions	Small and/or heterogeneous parcels	Large and/or homogeneous parcels
Scale of variations in elevation / slope	Elevation varies considerably across small areas/ steep slopes	Relatively gradual elevation changes/gentle slopes
Presence of small but important features or management practices that could impact your results (e.g. small dirt roads, water control bars on roads, strip-tillage, drainage ditches, thin riparian buffers or grass strips)	Important small features present	Small features absent or unimportant
Frequency of data updates needed (temporal resolution)	Infrequent updates acceptable	Need for frequent updates
Other uses for spatial data	The data can be used for multiple analyses	The data will solely be used for a single analysis

resolution to have a relatively small impact on modeled water quality (our primary area of concern), the difference in model error and ROI could be critical in some contexts.

Higher-resolution data may be preferable to lower-resolution data when a landscape has small features and/or fine-scale variation in land cover/land use, when a large portion of land cover/land use change patches are smaller than the pixel size at lower-resolutions, and when high accuracy is necessary to inform decisions. While our analysis was focused on satellite imagery, there is increasing interest in the use of drones as a source of higher-resolution data they offer the advantage of being able to control precisely when and where imagery is collected (even under cloud cover).

However, there are several trade-offs involved. In addition to higher cost, higher-resolution data is likely to require more processing time, raise challenges ensuring that data from different times are precisely spatially rectified or aligned (important for measuring land cover/land use change); have lower temporal resolution, and have more sensor angle variation and visible shadows. There are also legal and practical challenges to obtaining high-quality spatial data from drones (especially over large areas).

Fortunately, as new data sources become available, it should become easier to select the one that best balances the trade-offs for a given need. For example, the new Sentinel 2 satellites provide free imagery up to 10 m resolution (depending on which of the 13 multi-spectral bands are used) with global coverage every 5 days. This should offer some of the advantages of even higher-resolution data while avoiding some of the limitations. As lower-resolution data is available at no cost, downloading and examining the highest resolution data that is freely available is a good first step to determine whether higher-resolution data might be required.

Many previous studies have assumed that the higher-resolution data provides superior estimates of watershed behavior (e.g. Cotter et al. 2004; Chaubey et al. 2005; Lin et al. 2010), either using it as a basis for comparison or to set model parameters. Future work should empirically evaluate the value of this higher-resolution data for different decision contexts, carefully considering both the costs of higher-resolution data and the ability of each dataset to predict future behavior independently.

Acknowledgements

We are indebted to Esri for providing us with free licenses of ArcGIS products used in this analysis. Thanks also to Silvana Giberti, Claudio Klemz, Andre Targa Cavassani, Jeffrey Evans, Trisha Gopalakrishna and Jessica Garcia for their input on this work; and to Katie Dietrich, Trishna Dutta, Tyler Kartzinel, Joseph Kiesecker and Matt Miller who provided peer review on an earlier draft of this manuscript.

References

de Almeida Bressiani, D., P. W. Gassman, J. G. Fernandes, L. H. P. Garbossa, R. Srinivasan, N. B. Bonumá, et al. 2015. A review of soil and water assessment tool (SWAT) applications in Brazil: challenges and prospects. *Int. J. Agric. Biol. Eng.* **8**, 1–27.

Andréfouët, S., R. Berkelmans, L. Odriozola, T. Done, J. Oliver, and F. Müller-Karger. 2002. Choosing the appropriate spatial resolution for monitoring coral bleaching events using remote sensing. *Coral Reefs* **21**, 147–154.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams. 1998. Large area hydrologic modeling and assessment part I: model development. *J. Am. Water Resour. Assoc.* **34**, 73–89.

Boyle, S. A., C. M. Kennedy, J. Torres, K. Colman, P. E. Pérez-Estigarribia, and N. U. De La Sancha. 2014. High-resolution satellite imagery is an important yet underutilized resource in conservation biology. *PLoS ONE* **9**, 1–11.

Chaubey, I., A. S. Cotter, T. A. Costello, and T. S. Soerens. 2005. Effect of DEM data resolution on SWAT output uncertainty. *Hydrol. Process.* **19**, 621–628.

Cotter, A. S., I. Chaubey, T. A. Costello, T. S. Soerens, and M. A. Nelson. 2004. Water quality model output uncertainty as affected by spatial resolution of input data. *J. Am. Water Resour. Assoc.* **72701**, 977–986.

Di Luzio, M., J. G. Arnold, and R. Srinivasan. 2005. Effect of GIS data quality on small watershed stream flow and sediment simulations. *Hydrol. Process.* **19**, 629–650.

Fenichel, E. P., M. J. Kotchen, and E Addicott. 2016. The representative agent must die: Using demographics to inform social discount rates. Paper presented at the European Association of Environmental and Resource Economists (EAERE) 22nd Annual Conference, 22 - 25 June 2016, Zurich, Switzerland.

Ferreira, J., L. Silva, and M. Polette. 2009. The coastal artificialization process. Impacts and challenges for the sustainable management of the coastal cities of Santa Catarina (Brazil). *J. Coastal Res.* **56**, 1209–1213.

Gassman, P. P. W., M. M. R. Reyes, C. C. H. Green, and J. J. G. Arnold. 2007. The soil and water assessment tool: historical development, applications, and future research directions. *Trans. ASAE* **50**, 1211–1250.

Geza, M., and J. E. McCray. 2008. Effects of soil data resolution on SWAT model stream flow and water quality predictions. *J. Environ. Manage.* **88**, 393–406.

Guimarães, R. Z., C. Lingnau, N. E. Rizzi, R. G. Scheichi, and R. De Cassia Bianchi. 2011. Espacialização da perda de solo por erosão laminar na microbacia do Rio Campinas. Joinville Santa Catarina. *Raega-O Espaço Geográfico Em Análise* **23**, 534–554.

Howard, R. A. 1966. Information value theory. *IEEE Trans. Syst. Sci. Cybern* **2**, 22–26.

Keisler, J. M., Z. A. Collier, E. Chu, N. Sinatra, and I. Linkov. 2014. Value of information analysis: the state of application. *Environ. Syst. Decisions* **34**, 3–23.

Kennedy, E. F. 1984. Discharge Ratings at Gaging Stations. U.S. Geological Survey Techniques of Water-Resources Investigations. U.S. Geological Survey. Available at http://www.comm-tec.com/Library/technical_papers/USGS/twri_3-a9_ver1.pdf.

Kroeger, T., X. Klemz, D. Shemie, T. Boucher, J. R. B Fisher, E. Acosta, et al. 2017. Assessing the Return on Investment in Watershed Conservation: Best Practices Approach and Case Study for the Rio Camboriú PWS Program, Santa Catarina, Brazil. The Nature Conservancy, Arlington, VA. Available at https://global.nature.org/content/how-to-assess-the-return-on-investment-in-watershed-conservation

Kroeger, T., C. Klemz, T. Boucher, J. R. B. Fisher, E. Acosta, A. Targa Cavassani, et al. In preparation. Estimating the return on investment in watershed conservation: Best practices approach and case study for the Rio Camboriú PWS program, Santa Catarina, Brazil.

Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.

Lin, S., C. Jing, V. Chaplot, X. Yu, Z. Zhang, N. Moore, et al. 2010. Effect of DEM resolution on SWAT outputs of runoff, sediment and nutrients. *Hydrol. Earth System Sci. Discuss.* **7**, 4411–4435.

Minella, J. P. G., D. E. Walling, and G. H. Merten. 2008. Combining sediment source tracing techniques with traditional monitoring to assess the impact of improved land management on catchment sediment yields. *J. Hydrol.* **348**, 546–563.

Molnár, D. K., and P. Y. Julien. 1998. Computer and Geosciences. *Computer Geosci.* **24**, 183–192.

Moody, A., and C. E. Woodcock. 1995. The influence of scale and the spatial characteristics of landscapes on land-cover mapping using remote sensing. *Landscape Ecol.* **10**, 363–379.

Moriasi, D. N., and P. J. Starks. 2010. Effects of the resolution of soil dataset and precipitation dataset on SWAT2005 streamflow calibration parameters and simulation accuracy. *J. Soil Water Conserv.* **65**, 63–78.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Binger, R. D. Harmel, and T. L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. of the ASABE* **50**, 885–900.

Nash, J. E., and J. V. Sutcliffe. 1970. River flow forecasting through conceptual models part I-a discussion of principles. *J. Hydrol.* **10**, 282–290.

Palhares, J. C. P., A. L. Guidoni, R. L. R. Steinmetz, M. R. B. Mulinari, and G. C. Sigua. 2012. Impacto de los sistemas agropecuarios sobre la calidad del agua de la subcuenca del río Pinhal, Santa Catarina, Brasil. *Archivos de Zootecnia* **61**, 493–504.

Rasmussen, P. P., J. R. Gray, G. D. Glysson, and A. C. Ziegler. 2009. Guidelines and Procedures for Computing Time-Series Suspended-Sediment Concentrations and Loads from in-Stream Turbidity-Sensor and Streamflow Data. U.S. Geological Survey Techniques and Methods. U.S. Geological Survey. Available at https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf.

Teixeira Guerra, A. J., M. A. Fullen, M. do Carmo Oliveira Jorge, and S. Teixeira Alexandre. 2014. Soil erosion and conservation in Brazil. *Anuário Do Instituto de Geociências* **37**, 81–91.

Tena, A., R. J. Batalla, D. Vericat, and J. A. López-Tarazón. 2011. Suspended sediment dynamics in a large regulated river over a 10-year period (the lower Ebro, NE Iberian Peninsula). *Geomorphology* **125**, 73–84.

Turner, W. 2014. Sensing biodiversity. *Science* **346**, 301–302.

Turner, W., S. Spector, N. Gardiner, M. Fladeland, E. Sterling, and M. Steininger. 2003. Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* **18**, 306–314.

U.S. Environmental Protection Agency (EPA). 2002. The clean water and drinking-water infrastructure gap analysis. Office of Water (4606M). EPA-816-R-02-020. September.

U.S. Geological Survey (USGS). 2014. Shuttle Radar Topography Mission, 1 Arc Second scenes SRTM1S27W049V3 and SRTM1S28W049V3, Filled Finished 3.0, Global Land Cover Facility, University of Maryland, College Park, Maryland, February 2000.

U.S. Geological Survey (USGS). 2015. Landsat — Earth Observation Satellites, Available at: http://pubs.usgs.gov/fs/2015/3081/fs20153081.pdf. Accessed 22 March, 2017.

Supporting Information

Additional supporting information may be found online in the supporting information tab for this article.

Figure S1. Land use classification for 2012 at 1m spatial resolution.
Figure S2. Land use classification for 2013 at 30 m spatial resolution.
Figure S3. Comparison of spatial allocation of sediment yield in 1 m and 30 m models.
Table S1. Impact of the land use classification rules from Table 1 on overall area of each land use class.
Table S2. Comparison of 2012/2013 land use data at 1 m and 30 m.
Data S1. Land cover classification.