River gauging at global scale using optical and passive microwave remote sensing

Albert I. J. M. Van Dijk[1], G. Robert Brakenridge[2], Albert J. Kettner[2], Hylke E. Beck[3], Tom De Groeve[3], and Jaap Schellekens[4]

[1]Fenner School of Environment and Society, Australian National University, Canberra, Australian Capital Territory, Australia, [2]Dartmouth Flood Observatory, CSDMS, Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado, USA, [3]Joint Research Centre of the European Commission, Ispra, Italy, [4]Deltares, Delft, Netherlands

Abstract Recent discharge observations are lacking for most rivers globally. Discharge can be estimated from remotely sensed floodplain and channel inundation area, but there is currently no method that can be automatically extended to many rivers. We examined whether automated monitoring is feasible by statistically relating inundation estimates from moderate to coarse ($>0.05°$) resolution remote sensing to monthly station discharge records. Inundation extents were derived from optical MODIS data and passive microwave sensors, and compared to monthly discharge records from over 8000 gauging stations and satellite altimetry observations for 442 reaches of large rivers. An automated statistical method selected grid cells to construct "satellite gauging reaches" (SGRs). MODIS SGRs were generally more accurate than passive microwave SGRs, but there were complementary strengths. The rivers widely varied in size, regime, and morphology. As expected performance was low ($R < 0.7$) for many (86%), often small or regulated, rivers, but 1263 successful SGRs remained. High monthly discharge variability enhanced performance: a standard deviation of 100–1000 m^3 s^{-1} yielded ca. 50% chance of $R > 0.6$. The best results ($R > 0.9$) were obtained for large unregulated lowland rivers, particularly in tropical and boreal regions. Relatively poor results were obtained in arid regions, where flow pulses are few and recede rapidly, and in temperate regions, where many rivers are modified and contained. Provided discharge variations produce clear changes in inundated area and gauge records are available for part of the satellite record, SGRs can retrieve monthly river discharge values back to around 1998 and up to present.

1. Introduction

1.1. Background

Global change monitoring, water management, and flood risk assessment all require historical and current information regarding river discharge. Gauging stations are expensive to construct and operate; require well-defined river cross sections; require considerable and sustained efforts to produce reliable data (e.g., establishing and maintaining rating curves with field traverses); often are damaged during large floods or overtopped such that maximum peak discharges are not recorded; and, last but not least, records are very often not readily shared. For reasons such as these, the global network of gauging stations with recent available discharge data is limited, and the number of stations with adequately quality-controlled and near-real time accessible data is even smaller, and for many nations, in decline [*Hannah et al.*, 2011; *Pavelsky et al.*, 2014; *Shiklomanov et al.*, 2002]. The development of cheaper yet adequately precise methods to monitor river discharge is an important challenge in hydrology. Good quality and comprehensive observations have the potential to effect a major improvement in our understanding of the hydrologic cycle.

The most direct method of deriving such information at a *global* scale would seemingly involve satellite earth observation, and various approaches have been investigated (see *Alsdorf et al.* [2007] for a review). River discharge is monitored on the ground using stage (river level) as an indicator of discharge. River height measurements are possible through satellite altimetry and are conceptually closest to in situ flow gauging, but are currently only possible for some of the world's largest rivers [e.g., *Birkinshaw et al.*, 2014, 2010]. As an alternative, several studies have explored the potential of remote sensing methods to measure river width or areal inundation extent, and from this, estimate river discharge. An early literature review by *Smith* [1997] covered inundation extent mapping using radar [*Smith et al.*, 1996] and optical imagers [*Kruus*

et al., 1981]. About the same time, *Vörösmarty et al.* [1996] demonstrated that the inundation signal in passive microwave observations over the Amazon River could be correlated to monthly discharge estimates.

Research on predicting discharge from inundation extent since can generally be categorized as either semi-empirical (i.e., based on hydraulic geometry) or empirical (i.e., statistically relating inundation extent to discharge). In the first, satellite imagery is used to estimate river width for an individual river reach and used to constrain the semiempirical hydraulic geometry equations [*Leopold and Maddock*, 1953] relating width to discharge [*Bjerklie et al.*, 2005; *Gleason et al.*, 2014; *Pavelsky*, 2014; *Smith et al.*, 1996; *Smith and Pavelsky*, 2008]. In the second, observed correlations between inundation signal and station discharge are used to establish a "rating curve" and predict streamflow. It follows the same strategy employed at in situ gauging stations, where flow area and velocity measurements are matched to concurrent river stage records to predict discharge from stage [*Brakenridge et al.*, 2012, 2005; *Papa et al.*, 2008; *Vörösmarty et al.*, 1996]. At present, both approaches require either ground-based discharge measurements, or discharge estimates from a sufficiently accurate hydrological model. At the global scale, an advantage of the semiempirical approach is that full and detailed river inundation mapping is not required: it is immediately feasible to apply this method automatically to a very large number of reaches, as is shown in this paper.

Previous studies and reviews have shown that optical, radar, and passive microwave techniques all have their own strengths, as well as constraints that affect unsupervised processing over large areas. For example, optical observations suffer from cloud and vegetation obstruction and shadow effects, and radar observations can be affected by vegetation scattering (the so-called "double bounce" effect) [*Smith*, 1997]. Passive microwave observations from existing sensors have a relatively large observation footprint (e.g., 8–10 km) which does not permit actual width measurements, but, if calibrated well, can provide accurate in-pixel water extent area values [*Revilla-Romero et al.*, 2015, 2014; *Vörösmarty et al.*, 1996]. The current investigation was motivated by the potential benefits of any technique that could become a reliable source of global river discharge information. In particular, we consider that the integration of different remote sensing techniques may offer the strongest potential to overcome the known constraints.

Mitigation strategies have been developed for some of the observational problems described above. For example, to enhance the quality of optical methods, composites of imagery acquired during several overpasses can be constructed to reduce cloud and shadow effects. This can degrade the usefulness for near-real time flood monitoring applications, but can still support applications for less rapidly changing discharge regimes [*Guerschman et al.*, 2011]. Nonetheless, optical water extent mapping remains a processing intensive exercise because of the high resolution at which it needs to occur and, in some cases, allowance that needs to be made for the specific reflectance properties of dry compared to inundated land. In passive microwave applications, isolation of the hydrographic signal from the other factors affecting microwave radiance has been addressed by the use of nearby "reference" footprints in a ratio approach [*Brakenridge et al.*, 2007]. Subsequently, a novel processing method was developed by *De Groeve* [2010] for the automated selection of appropriate reference grid cells. Accordingly, *Kugler and De Groeve* [2007] reprocessed microwave observations obtained by a series of passive microwave satellites to produce a daily time series of gridded estimates of in-pixel water extent at 0.09° resolution and extending back to 1998. *Revilla-Romero et al.* [2014] tested the effectiveness of this technique in mapping large-scale flooding. Although the results are encouraging, the method does not work everywhere and there is considerable temporal noise in the discharge signal for some measurement reaches where water surface area changes with discharge are relatively small.

So far, there has been little exchange of processing methods between optical and microwave mapping approaches. The use of dry "reference" footprints developed for passive microwave data could potentially work equally to reduce processing time and calibration errors for optical data [cf. *Tarpanelli et al.*, 2011, 2013]. Conversely, the temporal compositing used for optical data might also reduce noise in microwave remote sensing [e.g., *Liu et al.*, 2012] and hence improve the quality of discharge estimates derived.

In both cases, the derivation of discharge estimates relies on the assumption that water extent increases as river discharge increases. Although this is true generally, that relationship is not unique, and depends on the relationship between discharge, stage, and inundation extent. Just as for river level measurements at gauging stations, hysteresis and change in the relationship can occur; that is, the same river surface area (or stage) may correspond to somewhat different discharge during the rising stage than during the falling

stage of a flood pulse, and the rating curve relationship may be affected also over time by fluvial morphological changes. A challenge specific to unsupervised remote sensing methods is that the radiative properties of the water and dry land vary in time and space, and it is therefore not always straightforward to accurately infer actual water extent.

1.2. Objective

Here we assess whether optical and microwave-derived water extent estimates derived at the global scale can provide a source of consistent, continuous and up-to-date monthly discharge estimates for many of the world's rivers. A correlation analysis is used to assess whether remote sensing grid cells (i.e., sensor footprint observations resampled to a regular grid) can be found in which water extent time series correlate strongly with measured monthly discharge. If such grid cells exist, they would allow the construction of the remote sensing equivalent of in situ gauging stations, which we will refer to as satellite gauging reaches (SGRs). These SGRs can then be calibrated individually to available historic discharge observations, and subsequently used to extend these past observations forward. The same approach could be used to estimate missing parts of existing gauging station discharge records where needed. To our knowledge, this study is the first to undertake such an analysis for a very large number of discharge gauging stations world-wide, covering a wide variety of river regimes and morphologies.

We pursue this objective using a global database of monthly stream discharge from several thousand in situ gauging stations. In addition, we use satellite-based river water level observations for some of the world's largest rivers, for which contemporary gauging station data are unavailable. Our working assumption is that the likelihood of a strong relationship between remotely sensed water extent and discharge is greater at monthly rather than shorter time scales, because compositing can help to avoid data gaps and outliers in the satellite imagery, and because the influence of short-lived flood peaks and associated hysteresis is reduced.

2. Data

2.1. Passive Microwave Water Extent

The satellite flood extent data were derived following the method first developed by *Brakenridge et al.* [2007]. The technique uses microwave remote sensing observations in the Ka-band with H-polarization (37 or 36.5 GHz), which has been shown to be sensitive to surface water. Due to the different thermal inertia and emissivity of land materials of various type and water, the observed microwave radiation intensity (radiance, or brightness temperature) is much lower for water than for land. This allows detection of changes in open water area within a river measurement site (a selected grid cell or image pixel), as most river channels do not occupy the full ∼10 km diameter footprint. However, the raw brightness temperature observations are also a function of physical surface temperature, vegetation, and other environmental variables. Therefore, the analysis technique relies on the assumption that these factors affect river measurement site grid cells centered over rivers and floodplains in a similar way as nearby grid cells over the surrounding dry terrain. Hence, the time series signal of a partially inundated grid cell footprint along a river is compared to the driest (highest) value in data from a surrounding array of grid cells via a simple ratio. Because of the sensitivity to surface water, the ratio of that measurement grid cell microwave radiance to the surrounding, dry ground, background radiance should monitor hydrographic changes as river discharge and water surface extent rise and fall. In effect, this spatial ratio isolates the strong effect any surface water change will have on the measurement site radiance from other factors. Thus, discharge signal *s* is inferred from the ratio M/C: the brightness temperature for a target or "measurement" grid cell (M), divided by that of a background or "calibration" grid cell (C). The methodology uses nighttime (descending) observations as temperatures are more homogenous in space and time during the night.

Kugler and De Groeve [2007] developed an automated analysis system that underpins a Global Flood Detection System (GFDS) implemented at the European Commission Joint Research Centre [*De Groeve et al.*, 2015]. One of the challenges was establishing a consistent method of selecting C for each grid cell M. First, the incoming satellite observations are resampled to a global, daily regular grid with 0.09° (∼10 km) latitude and longitude resolution. Next, for each M cell, a 9 × 9 kernel is applied to calculate the highest (driest) nearby 5% percentile brightness temperature. This value is adopted as C for calculating the M/C ratio. GFDS output is currently derived from the Japanese Space Agency's ASMR-2 sensor and from NASA's GPM

instrument, whereas prior to 2012, data from the TRMM TMI and AMSR-E instruments were used (Table 1). It is noted that AMSR-E and AMSR-2 do not achieve full daily swath coverage toward the equator, causing temporal gaps in the daily time series of up to 3 days. The data used in the analysis here avoid such data gaps by averaging all available s values from the preceding 4 days for each daily image.

2.2. MODIS Water Extent

Optical remote sensing suffers from obstruction by cloud cover and various water classification errors, including "false positive" water mapping due to cloud shadow. At monthly time steps, there is a greater chance of obtaining cloud-free observations, which makes its use in river monitoring more consistent at this coarser time resolution, particularly for rivers with relatively slowly varying discharge rates. To test this, we used MODIS 8 day composites in a manner similar to the method described in the previous section. We used the MCD43C4.005 product available from NASA, which contains 8 day NBAR composites of imagery derived from both the AQUA and TERRA MODIS instruments for 2000 onward. These data are resampled to a regular global 0.05° grid, which makes it feasible for rapid processing at the global scale as part of this study. The water classification method is described in section 3.1.

2.3. In Situ Discharge Data

Beck et al. [2015] collated daily and monthly discharge observations from 14,852 gauging stations globally from a variety of sources, including the Global Runoff Data Centre and the USGS GAGES II (Geospatial Attributes of Gages for Evaluating Streamflow) database. The data were thoroughly checked for artifacts. For this study, we selected those stations with more than 40 months of daily or monthly discharge data during our satellite data period (2000–2014). This provided a total of 8848 stations available for analysis. The upstream catchment area represented at the stations varied widely, from <1 km^2 to very large river basins such as for stations on the Amazon River. Discharge values varied correspondingly, from a maximum monthly average flow of only 6.7×10^{-4} m^3 s^{-1} to as much as 9.2×10^4 m^3 s^{-1}. We did not limit the analysis to exclude any of these stations. Our prior expectation was to see better results for broad and low gradient river systems with large and variable discharge, and poorer results for narrow rivers in high relief terrain with small and perhaps less variable discharge.

2.4. River Altimetry

As part of a previous study we processed satellite water level altimetry time series for 1429 river locations world-wide [*van Dijk et al.*, 2014]. These data were derived from the ENVISAT and JASON-2 instruments and made available via the HYDROWEB web site hosted by Laboratoire d'Etudes en Geodésie et Océanographie Spatiales (LEGOS) and the River&Lake web site hosted by the European Space Agency, respectively. The data were quality-controlled, harmonized, merged, and gridded to provide monthly water level time series for 2003–2012, presenting the main river channel(s) in four-hundred and forty-two 1° × 1° grid cells world-wide. Each site had on average 61 (maximum 115) data points during 2003–2012. Because river altimetry requires a well-defined and sufficiently broad river channel, there is a bias in these data toward such rivers.

3. Methods

3.1. Microwave Water Extent

From the 4 day s data, monthly average s values were calculated. We used these to estimate the fraction water extent w, that is, the fraction of the footprint covered by surface water. We use the same linear mixing

Table 1. Characteristics of the Satellite Observations From Which Water Extent Was Derived

Sensor	TMI	AMSR-E	AMSR2
Platform	TRMM	Aqua	GCOM-W
Band (GHz)	37.0 ± 1.0	36.5 ± 0.5	36.5 ± 0.5
Incidence angle	53°	55°	55°
Footprint (km)	16° × 9/18° × 10	14° × 8	12° × 7
Pass	n/a	Descending	Descending
Observation rate	3 h	Near-daily	Near-daily
Data processed	7 Dec 1997–16 Jun 2015	May 2002–4 Sep 2011	2 Jun 2012–ongoing

Footprint increased after orbit boost in August 2001; sensor in semiequatorial orbit observing between 50°N and 50°S; full coverage is not achieved every 3 h.

theory as *De Groeve* [2010] and assume that water and land have the same surface temperature, such that observed brightness temperature is a function of their emissivity ε

$$\varepsilon_M = w\varepsilon_{water} + (1-w)\varepsilon_{dry},\tag{1}$$

where ε_M, ε_{water}, and ε_{dry} are the emissivities of the measurement footprint, of surface water, and of dry land, respectively. Furthermore, $s = \varepsilon_M/\varepsilon_{dry}$ and this can be substituted into equation (1) and rearranged to yield w [*De Groeve*, 2010]

$$w = (s-1)/\left(\frac{\varepsilon_{water}}{\varepsilon_{dry}} - 1\right),\tag{2}$$

Equation (2) was used to estimate w from the monthly average s values, assuming $\varepsilon_{water} = 0.50$ and $\varepsilon_{dry} = 0.85$ [*De Groeve*, 2010]. The relationship between w and s remains inversely linear regardless of the emissivities assumed, and therefore those assumptions do not in fact influence the correlation with observed discharge or water level.

3.2. MODIS Water Extent

In using the global MODIS composites, we use a method analogous to the microwave method of *De Groeve* [2010], but instead use the shortwave infrared (SWIR) spectral band 7 (2105–2155 nm). This band was chosen because of the strong contrast between water and dry surfaces in this wavelength, in which even sediment-rich water exhibits low reflectance. For each 0.05° cell M the fraction water extent w was estimated as (cf. equation (2))

$$w = \frac{\rho_M - \rho_{dry}}{\rho_{water} - \rho_{dry}},\tag{3}$$

where ρ_M is the band 7 reflectance of target cell M, ρ_{dry} the fifth percentile highest reflectance in a 7 × 7 cell window (~35 × 35 km) around M, and ρ_{water} the reflectance of surface water. The latter is not truly constant, but visual exploration of large freshwater bodies suggested that 0.008 is a sufficiently accurate reflectance for most water bodies, considering the typically much higher value of ρ_{dry}. Analogous to the application of this method to brightness temperatures, the accuracy of the derived value of w depends mainly on how well ρ_{dry} represents dry surfaces within the target cell, whereas the actual ρ_{water} value chosen again does not affect the temporal correlation between w and river flow or level. It is noted that phenomena other than channel width per se can affect the temporal evolution of w values, e.g., the response of floodplain vegetation to flooding or the filling of reservoirs. It is possible that such phenomena correlate sufficiently strongly with observed discharge to cause the affected grid cells to be selected as highly correlated to in situ discharge in the construction of SGRs.

3.3. Satellite Gauging Reach Selection and Performance Evaluation

The use of water extent measurement for discharge estimation requires that an increase in river discharge corresponds to an increase in inundated area within the SGR. This probably typically requires not just wider channel flows but also filling of connected wetlands, in-channel bar and low floodplain surface overtopping, tributary stream backwater, or overbank flows within the SGR. By contrast, in situ river gauging locations are typically chosen in relatively stable and well-contained river channel sections, where discharge increase leads primarily to an increase in river stage. Therefore, most suitable SGRs will not coincide with the location of gauging stations, and may be a considerable distance up or downstream. The river level altimetry data also do not apply to one specific river transect. These data were derived from orbital measurements at different locations along the river course, and in this case too, more suitable and less suitable locations may exist for useful retrieval of river discharge changes.

To accommodate these issues, we applied a search window. For the gauging stations, all cells in a ca. 1° × 1° (ca. 100 × 100 km at the equator) window centered on the station were considered. This contained 400 MODIS or 100 GFDS grid cells, each with a time series of water extent. Subsequently, these time series were compared with the observed discharge data, and a series of common data pairs was constructed. The series was split equally into two; the first part for SGR selection and the second part for validation:

Selection: SGRs were selected by calculating the correlation between observed discharge and each of the (400 or 100) water extent time series, and subsequently selecting the N_M cells with the highest coefficient of correlation.

Validation: The average water extent among the grid cells was calculated and used as a predictor of the remaining half of the observed discharge time series.

Because of the large sample size, the probability of spuriously high correlation coefficients due to statistical coincidence needs to be considered. Splitting the time series in a "training" and "validation" helps avoid such results. The inclusion of more than only the single "best" correlating time series (i.e., $N_M > 1$) further mitigates the selection of suboptimal SGRs, and should be evident from a smaller reduction from selection to validation performance than if only the strongest correlating grid cell was chosen (i.e., $N_M = 1$). To test this, the same experiment was repeated using N_M of 1, 2, 4, 8, 16, and 32. These initial experiments were done for a systematic sample of 500 out of the 8844 gauges. The influence of the choice of correlation coefficient was also investigated, by repeating the experiment, respectively, using the parametric Pearson's R or the nonparametric Spearman's rank R' in both analysis steps. The suitability of either metric depends on the relationship between remotely sensed inundation extent and actual river discharge, which does not have to be linear [*Gleason et al.*, 2014; *Leopold and Maddock*, 1953; *Pavelsky*, 2014]. Which metric is to be preferred also depends on the purpose of discharge estimation; e.g., for predicting high flows R would seem more appropriate than R'; and vice versa for low flows. These analyses were performed on the entire set of station records.

Finally, a similar analysis was undertaken using the satellite river water level altimetry time series. The overall approach was the same, except (i) the search window was expanded outside the original $1° \times 1°$ altimetry merging window, producing windows of $2° \times 2°$ (ca. 200×200 km at the Equator), or 40×40 grid cells for the MODIS data and 20×20 for the GFDS data; and (ii) only one analysis configuration was used on the basis of the discharge comparison experiments (i.e., $N_M = 8$ for MODIS and $N_M = 4$ for GFDS).

4. Results

4.1. Configuration Experiments

Experiments using a subset of 500 gauges helped to find the optimal number of predictor cells (N_M). For the GFDS data, N_M values between 1 and 16 all produced very similar results in validation mode, with $N_M = 1$ producing the optimal result the largest number of times, $N_M = 2$ producing the highest median R, and $N_M = 8$ the highest mean R. All subsequent experiments with the GFDS data were done using an intermediate value of $N_M = 4$. For the MODIS data, the optimal value of N_M was somewhat better defined: $N_M = 8$ produced optimal validation result for the largest number of sites, and also produced the highest mean and median R values. Using $N_M = 4$ or $N_M = 16$ produced very similar results, however. All subsequent experiments using the MODIS data were done using $N_M = 8$.

4.2. Predictive Value of GFDS Water Extent

Splitting the time series left an average 41 months (range 20–90) of data available for validation at 8847 gauging sites. The average correlation between predicted and observed discharge was similar for parametric and nonparametric correlation coefficients alike, with Pearson's $R = 0.21$ (standard deviation, SD ± 0.28) and Spearman's rank $R' = 0.20 \pm 0.28$, respectively. Values of R exceeded R' in 62% of cases. It is reiterated that low mean R and R' values were expected given the database includes many small catchments. Greater values of R or R' (or both) did occur, however, with maximum values >0.6 for 1172 stations (i.e., 13% of all stations), >0.7 for 537 stations (6%), >0.8 for 170 stations (2%), and >0.9 for 26 stations (0.3%) (Figure 1).

Figure 1 suggest that the most successful sites (R or $R' > 0.8$) are concentrated in the southeast of the USA and particularly the Mississippi valley, as well as parts of tropical South America, Africa, and part of Australia. Poor results are obtained in the remainder of the North America, Europe, and Asia, whereas mixed results were found for South America, Africa, and southeast Australia and New Zealand. It is noted that data for regions with few gauging stations are generally biased toward larger rivers.

Repeating the experiment using the river water level altimetry data also produced geographically varying results (Figure 2). R' values were >0.8 for the largest rivers (e.g., Amazon, Mississippi, Niger, and Ganges),
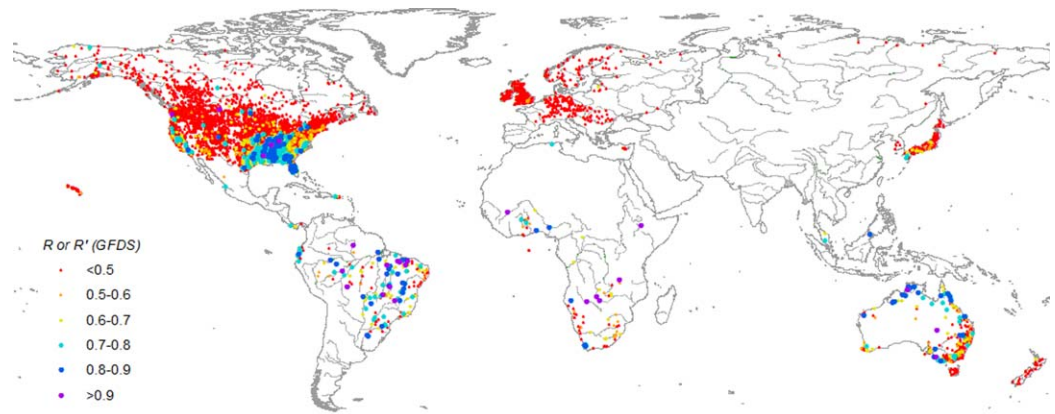
Figure 1. GFDS-based river remote sensing evaluated against monthly average station discharge observations, expressed as the greater of parametric (R) and nonparametric (R') correlation coefficient in validation.

but results were mixed for most other rivers. It is reiterated that river altimetry requires a relatively broad and well-defined river channel and therefore the sample is already biased toward such locations.

4.3. Predictive Value of MODIS Water Extent

For the MODIS-based water extent estimates, splitting the time series left an average 72 months (range 20–85) of validation data available for 8844 gauges. The average correlation between predicted and observed discharge in validation was again similar for nonparametric and parametric correlation coefficients, with $R = 0.43 \pm 0.19$ and $R' = 0.44 \pm 0.19$, respectively. The value of R exceeded R' in 54% of cases. Encouragingly, R or R' (or both) was >0.6 for 2125 stations (i.e., 24% of all stations), >0.7 for 863 stations (10%), >0.8 for 218 stations (2.5%), and >0.9 for 23 stations (0.26%) (Figure 3). For the remaining 6719 stations (76%) R and R' were both <0.6.

The most successful SGRs (R or $R' > 0.8$) are concentrated in the boreal regions of North America and Russia, the Rocky Mountains and Pacific coast of North America, the Mississippi valley, Florida, and Ireland (Figure 3). Poor results are obtained in the remainder of North America, Europe, and Japan, whereas mixed results are found for South America, Africa, and Australia.

As for the GFDS data, agreement with the river water level altimetry data varied regionally and even along rivers (Figure 4). R' values >0.8 were found for the main channels in most large river basins, but with some clear exceptions: results were poor for the Danube, Wolga, Enisey, and Yellow Rivers, and mixed for some of the other rivers.
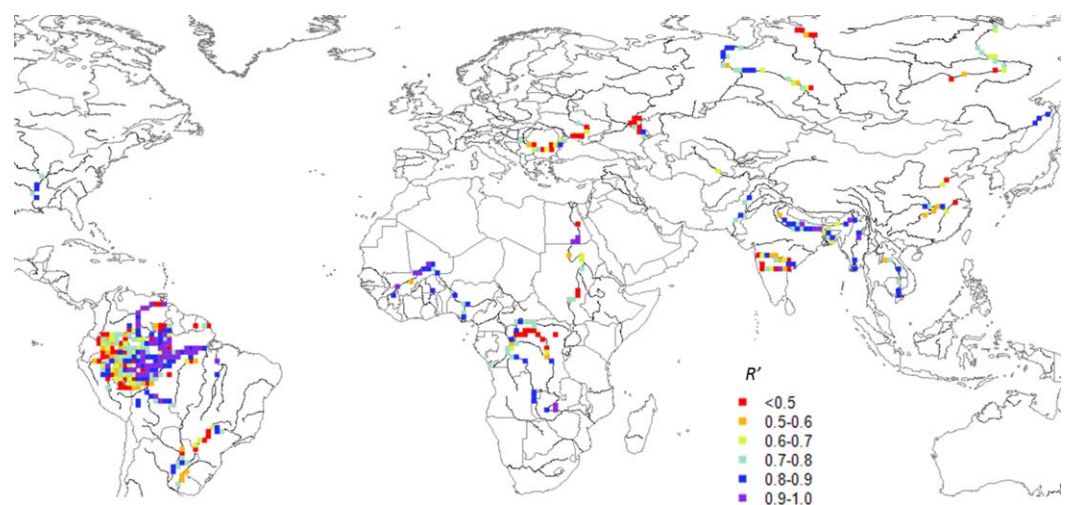

Figure 2. GFDS-based river remote sensing evaluated against monthly average river water level altimetry observations, expressed as the nonparametric (R') correlation coefficient in validation.
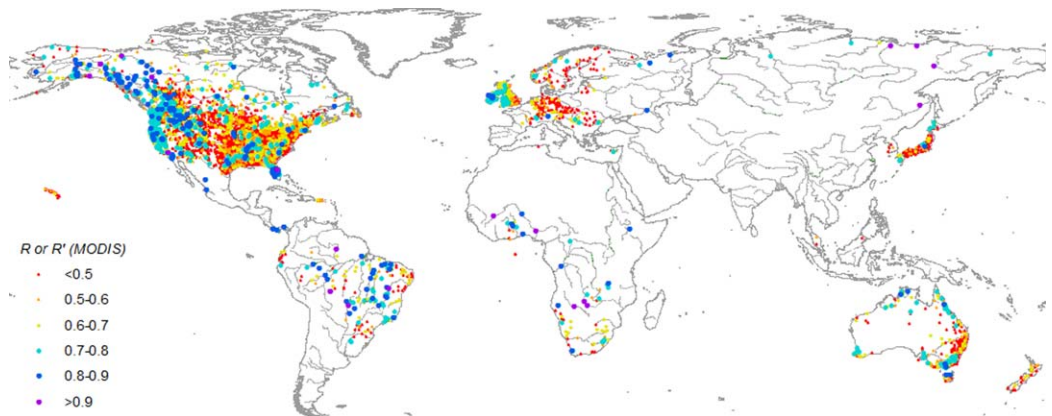
Figure 3. MODIS-based river remote sensing evaluated against monthly average station discharge observations, expressed as the greater of parametric (R) and non-parametric (R') correlation coefficient in validation.

5. Discussion

5.1. Requirements for Space-Based River Discharge Estimation

The ability to measure discharge from remote sensing varied widely among SGR locations. A priori, we expected to see better results for broad and low gradient river systems with large and variable discharge, and poorer results for narrow rivers in high relief terrain with small and less variable discharge [cf. *Revilla-Romero et al.*, 2014]. This was tested by calculating the standard deviation in monthly discharge, which itself was found highly correlated ($R > 0.98$) with other measures of river size such as maximum, median, and mean monthly discharge, whether log-transformed or not. This metric was compared to the highest coefficient of correlation found in validation mode (R_{max}) for either remote sensing approach tested, i.e., the greatest of R or R' derived for GFDS (Figure 1) or MODIS (Figure 3) data. It may be expected that R is most indicative of high-flow estimation, whereas R' assigns equal weighting to the full range of flows (cf. section 3.3). The three metrics had similar values for the majority of stations and overall statistics and spatial patterns were also very similar (median values were $R_{max} = 0.51$, $R = 0.48$, and $R' = 0.49$, Table 2).

In general terms, there was indeed a relationship between the standard deviation in discharge and R_{max} (Table 2 and Figure 5). Interestingly, there were still a considerable number of small rivers for which the validation performance appears sufficiently high to construct potentially useful SGRs. The influence of river size suggests that it may be possible to establish further SGRs if observations were analyzed at a higher
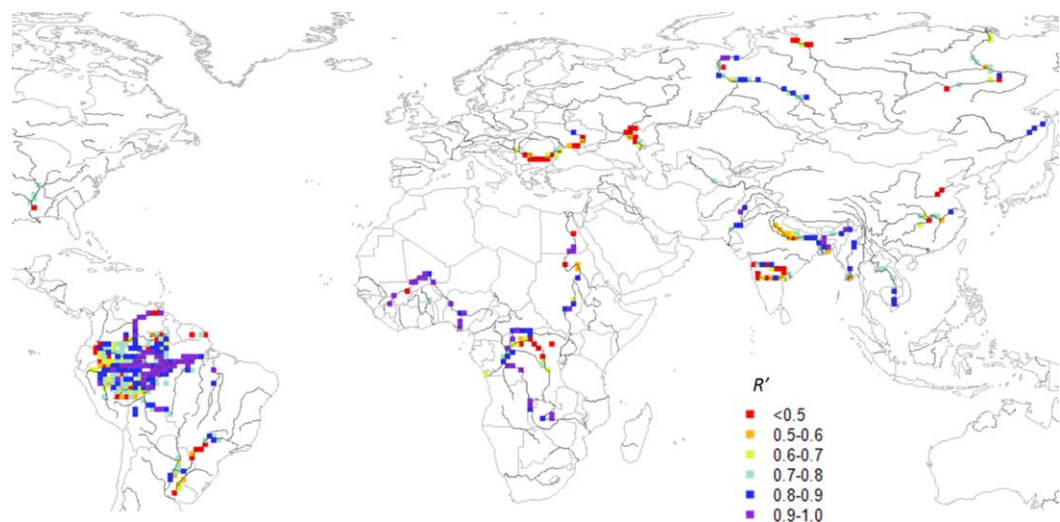

Figure 4. MODIS-based river remote sensing evaluated against monthly average river water level altimetry observations, expressed as the nonparametric (R') correlation coefficient in validation.

Table 2. Increase of the Feasibility of Satellite-Based Gauging With Increasing Absolute Variability in Monthly Discharge[a]

			Percentage With Maximum R/R' Greater Than			
σ_Q (m^3 s^{-1})	N	$<R>/<R'>$	0.6	0.7	0.8	0.9
0.001	8	0.34/0.33	0/0	0/0	0/0	0/0
0.01	272	0.42/0.41	15/17	4/5	1/0	0/0
0.1	1,708	0.43/0.44	18/21	6/7	1/1	0/0
1	3,437	0.46/0.47	21/24	7/8	1/1	0/0
10	2,653	0.51/0.52	32/34	13/15	3/4	0/0
100	685	0.56/0.57	42/43	24/26	12/11	2/1
1,000	75	0.70/0.71	72/72	55/59	44/43	9/15
10,000	5	0.78/0.83	80/100	80/100	60/60	0/20

[a]Listed are the standard deviation in monthly discharge (σ_Q, number represents the lower bound of the interval), the total number of stations in each interval (N), the mean achieved parametric and nonparametric correlation coefficient ($<R>$ and $<R'>$, resp.) in validation (for each station representing the greater of values from the optical or passive microwave approaches), and the percentage of stations for which R and R', respectively, exceeded values of 0.6–0.9.

resolution. However, mitigating against this may be that flood pulses dissipate faster in smaller rivers, which would increase challenges related to the temporal misalignment of satellite and station observations. Flooding along some of the rivers (e.g., lower Ganges, southern African, inland Australian, and Artic rivers) produces inundation widths well in excess of 0.09°, exceeding the monitoring capability of any single grid cell. In such cases, composite SGRs are needed to maintain sensitivity across the discharge range.

Apart from flow characteristics, the global distribution of correlation coefficients also reflects broad geomorphological and hydroclimatological patterns. To emphasize the geographic variation in relative potential for SGRs for other reasons than flow variability, R_{max} for each gauge was normalized by subtracting the expected value ($<R_{max}>$) for that flow interval (Table 2).

In addition to river size, it was anticipated a priori that river and floodplain modification, topography, and hydroclimate would be the primary determinants of SGR performance, as these variables impact the temporal distribution of flows and the relationship between discharge and inundated area [*Revilla-Romero et al.*, 2014]. Figure 6 tests this idea and indicates that relatively good SGR performance is found for little or unmodified low gradient rivers in humid regions (e.g., western coast of North America, southeastern US, central Brazil, (sub-) humid Africa, and most of Australia), whereas poor performance occurred for modified rivers (e.g., continental Europe and Japan) and rivers in regions with significant upland relief (e.g., the Rocky Mountains and eastern Australian coast). The poor performance for much of the North American Great Plains was not anticipated, although it is noted that medium to high river flows are regulated for many of the rivers. Also notable was the good performance for most of the British Isles (with the exception of

Figure 5. Box plots showing the distribution of measured SGR performance for river stations by interval of increasing monthly flow variability. Listed at the top is the number of SGRs in each interval (cf. Table 2).

England) and for rivers in Australia's Great Dividing Range in Queensland and Victoria. To examine possible reasons for these results, the river morphology and the selected SGRs for representative locations were inspected.
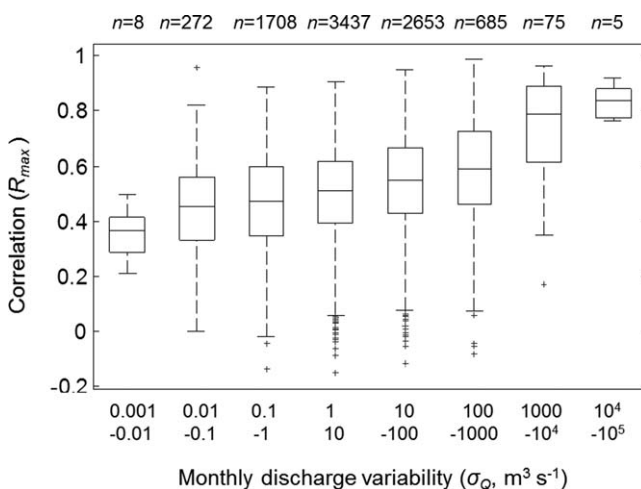
Although the evaluation of alternative methods to derive quantitative SGR-based discharge estimates was beyond the scope of this paper, the results of one of several possible approaches are shown for illustrative purposes. These estimates were obtained using the cumulative distribution function matching method, which provides a bias-free estimate, the quality of which depends on the accuracy of the temporal pattern in water extent with
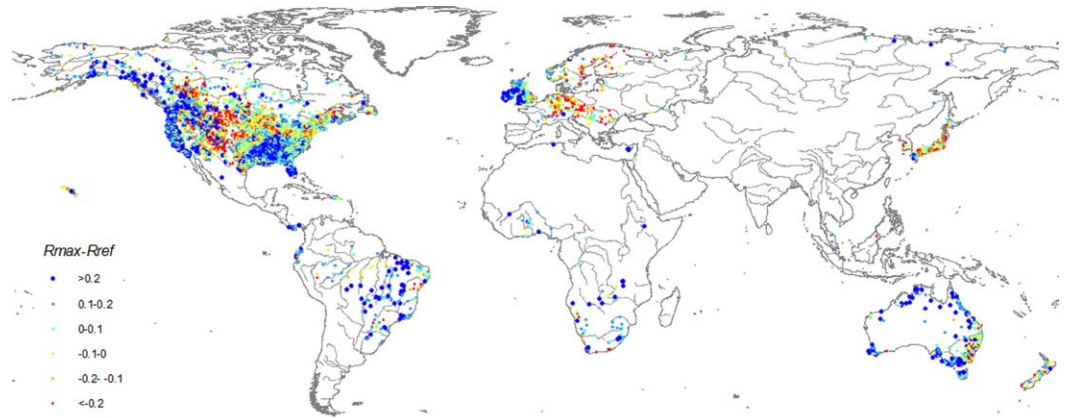
Figure 6. The difference between measured satellite river gauge performance (R_{max}) and the expected value based on the observed flow variability ($<R_{max}>$, cf. Table 2).

respect to observed gauging station observations [*Brown and Seo*, 2013; *Hashino et al.*, 2007; *Madadgar et al.*, 2014; *Verkade et al.*, 2013; *Wood et al.*, 2002]. In summary, the method uses paired SGR water extent and gauged monthly discharge for a training period where both are available. Subsequently, the paired data are both sorted in ascending order to create a Look-Up Table (LUT). Next, for each newly observed water extent value, the nearest two water extent values in the LUT are found, and discharge is estimated by linearly interpolating between the two discharge values with corresponding ranks in the LUT. Discharge for any water extent value outside the range of the LUT was extrapolated from the nearest available interval.
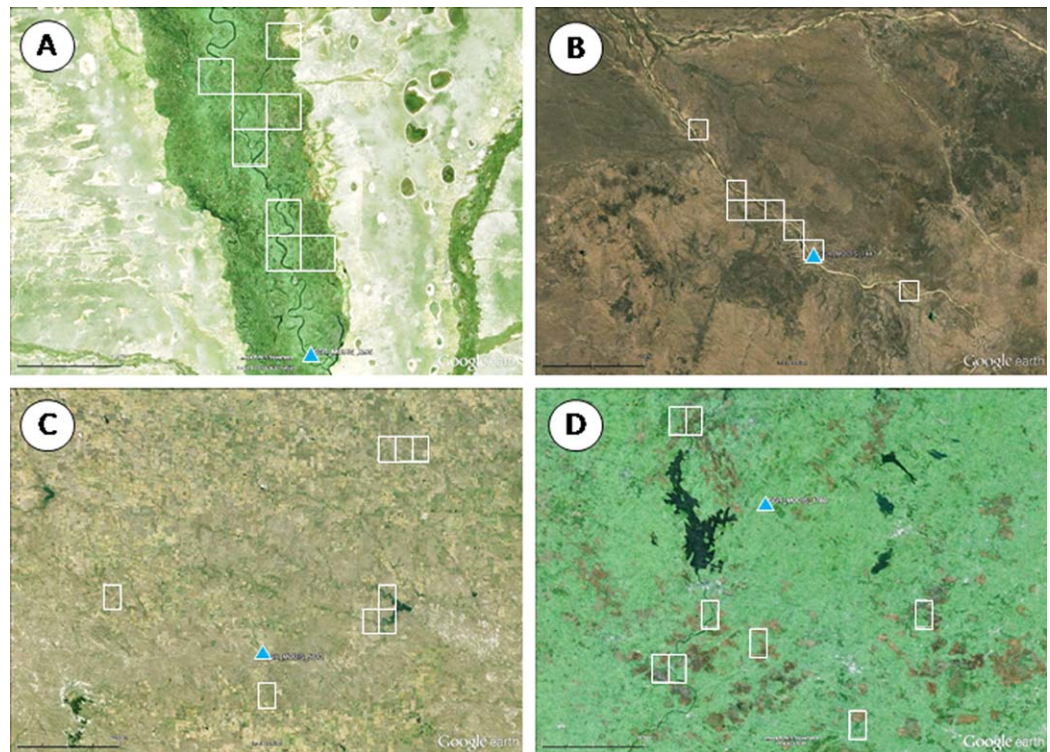

Figure 7. The eight MODIS grid cells (white outlines) selected in constructing SGRs (blue triangle indicates location of the gauging station). Shown are (a, b) examples of plausible grid cell selection resulting in strong correlation; (c) implausible selection resulting in poor correlation; and (d) implausible selection resulting in strong correlation. Details are as follows: (a) Zambezi River at Senanga (Zambia; 16.08°S, 23.26°E; $R = 0.96$), (b) Gilbert River at Rockfields (Queensland, Australia; 18.20°S, 142.88°E; $R = 0.85$), (c) Little Nasty Creek (South Dakota, USA; 45.64°N, 102.64°W; $R = 0.00$), and (d) Inny River at Ballymahon (Ireland; 53.56°N, 7.76°W; $R = 0.88$).
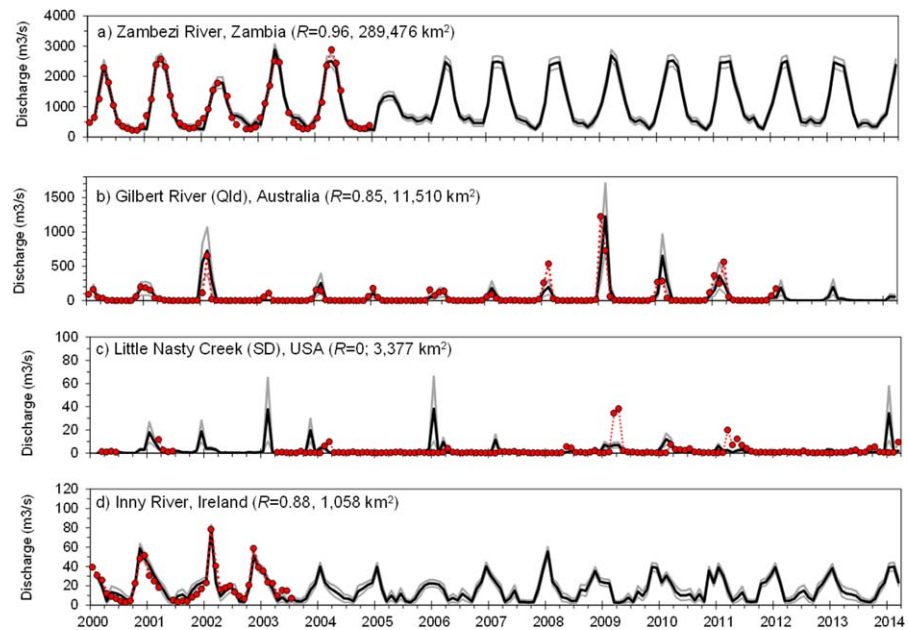
Figure 8. Comparison of monthly discharge as observed at gauging stations (red dots) and as predicted by the four MODIS SGRs in Figure 7 (black line represents the best estimate, shaded lines the interquartile range of estimates based on validation performance).

Four example SGRs are shown in Figures 7a–7d, with corresponding hydrographs in Figure 8. The same SGR information and hydrographs are available for all other SGRs through http:/www.wenfo.org/wald/.

Overall, good performance was usually associated with a plausible selection of SGR grid cells located along the river reach on which the gauging station was located (Figures 7a and 7b). Similarly, stations for which SGRs could not be constructed typically showed a less plausible distribution of selected grid cells (Figure 7c). Closer inspection of this station and other stations in the Great Plain showed a hilly topography in which minor rivers are incised and confined within a relative narrow floodplain with few wetlands. This could explain why SGR grid cell selection was troublesome and often occurred in nearby larger rivers or dam reservoirs, assuming the dynamics of these may correlate with discharge at the target station. Finally, some well-performing SGRs showed a selection of grid cells that appears implausible (Figure 7d). This appeared to occur most often for smaller rivers in poorly drained landscapes. The explanation appears to
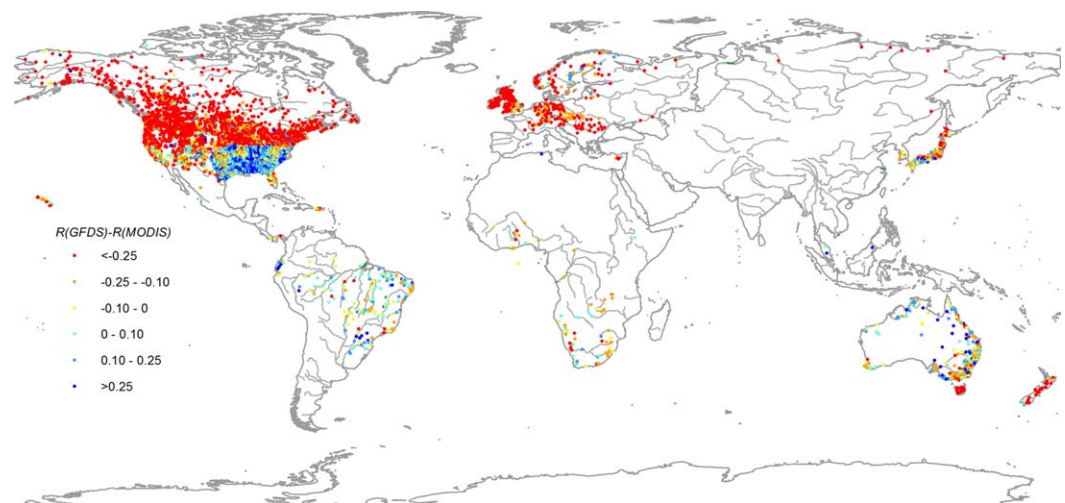

Figure 9. The difference in the maximum correlation coefficient (i.e., the greater of R and R') in validation mode obtained using water extent derived from passive microwave (GFDS) and optical (MODIS) observations, respectively.
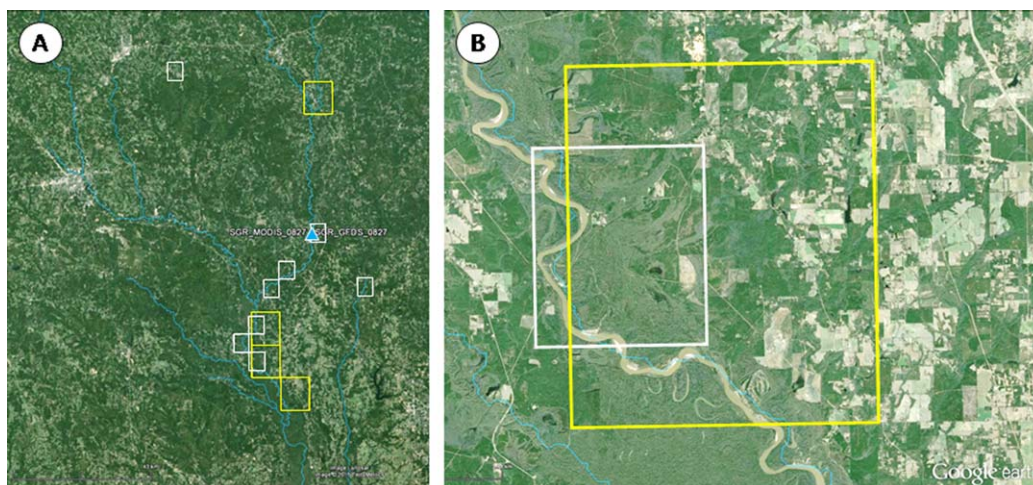
Figure 10. MODIS and GFDS SGR grid cells selected for a station in the southern USA, showing (a) all SGR cells for MODIS (white outline, 0.05° × 0.05°) and GFDS (yellow outline, 0.09° × 0.09°), and (b) closer view of two colocated cells covering the river floodplain. Example shown is Chickasawhay River at Leakesville, Mississippi (USGS code 02478500; 88.5479°N, 102.64°W), where GFDS achieved $R = 0.89$ and MODIS achieved $R = 0.46$, respectively.

be that water extent in nearby catchments was strongly correlated with discharge at the gauging station. In such cases, the ongoing performance of the SGR clearly depends on such relationships continuing, which is not guaranteed. Useful further work would be to examine in detail the surface water dynamics in such cases and the reasons for strong correlation to nearby discharge gauging.

5.2. Relative Merits of Microwave and Optical Remote Sensing
We assessed the effectiveness of SGRs based on passive microwave and on optical observations, both aggregated to monthly mean water extent estimates. Overall, optical observations appeared more suitable for constructing SGRs than the passive microwave-derived data. To some extent this may be attributable to the higher resolution of the global resampled data (0.05° versus 0.09°) and the smaller footprint of the original observations (500 m versus 8–10 km diameter). Another factor may be the strong influence of snow cover on passive microwave brightness temperature in high latitudes. Nonetheless, there were some regions where GFDS-based SGRs consistently performed better than MODIS-based ones, including the southern USA, parts of South America, and Northern Australia (Figure 9).

The cells associated with a number of such SGRs were examined to interpret why GFDS observations may be more effective in these areas. It appeared that many of the SGRs in the southern USA included cells over forest-covered floodplains (Figures 10a and 10b). Similar results were found for most of the South American and some of the Australian SGRs where GFDS appeared a more powerful source of information. In these conditions, the forest canopy cover overhead challenges flood detection by optical methods, whereas detection by passive microwave remote sensing remains possible due to greater sensitivity to subcanopy conditions. This agrees with the analysis by *Revilla-Romero et al.* [2014], who conclude that forest-covered floodplains present little impediment to GFDS-based discharge estimation when compared to other factors. The reasons for the better performance of GFDS in arid inland Australia are less clear. The corresponding stations appear to be mostly on ephemeral rivers covering vast floodplains when in flood, and therefore we speculate that the spatial range of inundation is too large to be captured at the 0.05° MODIS resolution and is better captured at the 0.09° GFDS resolution (i.e., the situation noted previously in which larger rather than smaller SGRs are needed to the full dynamic range of discharge and inundation).

5.3. Prospects for Further Development
One important limitation of the methodology tested here is that ground station discharge observations during the satellite observation period were required to guide the selection of SGR grid cells and to calibrate the discharge signal to actual discharge values. This means that SGRs cannot be established for river reaches without ground station observations or other independent discharge data such as those from hydrologic modeling. Further investigation of the characteristics of SGRs by the method presented here

may allow the development of a set of rules for automated selection in ungauged reaches, however. For example, inspection of the successful gauges suggest that the variation in water extent itself may be used to select cells, provided cells outside the river floodplain in question are not considered (Figure 7). Better constrained selection of cells may be achieved using a combination of comparatively high-resolution mapping of floodplain elevation [e.g., *Nobre et al.*, 2015] or historic inundation [e.g., *Brakenridge et al.*, 2005].

The conversion of water extent to river discharge in truly ungauged reaches is challenging. For larger rivers, the Surface Water and Ocean Topography satellite mission envisaged for launch in 2020 could allow simultaneous measurements of river water surface elevation, slope and width for large rivers, thus potentially allowing discharge to be intermittently estimated without station data [*Durand et al.*, 2014; *Garambois and Monnier*, 2015; *Pavelsky et al.*, 2014]. Another approach being considered by the SWOT mission is to couple model-estimated discharge to the satellite data constrain the range of likely discharge rates. In such an approach, the error in model estimates can be reduced and quantified by multimodel ensemble techniques [*van Dijk et al.*, 2014] and machine learning approaches [*Beck et al.*, 2015, 2013]. Indeed, the satellite-based river inundation estimates themselves may be useful in calibrating hydrological models [*Revilla-Romero et al.*, 2015], which suggests that an iterative process of SGR construction and model calibration may be feasible.

A second important limitation is that the approaches tested generally only produce both good performance and a plausible selection of SGR cells for relatively large rivers. Better results for smaller rivers, and for larger rivers during extended low flow periods, likely require inundation observations at higher spatial resolution [*Pavelsky*, 2014]. Finer spatial resolution microwave data at the needed near-daily frequency are not planned for the near future. However, MODIS shortwave infrared observations are available at 500 m and therefore could be used for SGR-based water extent estimation at much higher resolution. Global observations at even higher resolution of 60 m are also becoming available from the European Space Agency's Sentinel-2, albeit with a lesser temporal frequency of ca. 5 days. A challenge to successfully constructing SGRs for smaller rivers is that they are likely to dissipate flood pulses faster, and therefore observations will need to be available not only at higher spatial, but also higher temporal resolution. Currently, the best opportunities for this may be offered by the Sentinel-1 dual radar instrument mission, which combines a spatial resolution of <20 m with revisit time of less than 3 days. SGRs would have to be specifically targeted, however. Alternatively, geostationary instruments such as those on Himarawi-8 (launched in 2014) and GOES-R (planned for launch in 2016) offer quasi-continuous daytime shortwave infrared observations at ca. 1 km resolution [*Schmit et al.*, 2005] and hold promise for detecting more rapidly evolving flood conditions and possibly even for quite small streams and rivers.

6. Conclusions

We examined whether fully automated moderate to coarse (>0.05°) resolution global remote sensing discharge signals are sufficiently strongly associated with ground-observed monthly river discharge to enable monitoring of the world's rivers. Satellite gauging reach (SGR) inundation extents were derived from the optical MODIS instrument and a series of passive microwave sensors. These data were compared to monthly discharge records from more than 8000 gauging stations and to satellite altimetry observations in 442 reaches of large rivers world-wide. A quantitative and consistent method was developed to statistically select the best grid cells to be used in defining SGRs, and because the site-specific characteristics strongly control how accurately discharge changes can be sensed, as is the case for in situ gauging stations. The main conclusions are:

1. The data set included rivers of widely varying size, regime, and floodplain morphology, and as expected the majority of SGRs (86%) showed relatively low performance ($R < 0.7$), which still left a total of 1263 SGRs. Success rate increased with the magnitude of monthly discharge variability, with a standard deviation of 100–1000 m^3 s^{-1} yielding a ~50% chance of $R > 0.6$ (Table 2).
2. Overall, SGRs appear a viable approach to measuring monthly discharge changes in many large as well as some smaller rivers world-wide, provided gauging records are available during part of the satellite record. This opens opportunities also to continue discharge observation for reaches with stations that are no longer in use or with a long delay in data availability.

3. At the monthly time scale considered, MODIS-based SGRs were generally more effective than passive microwave SGRs, but there are some complementary strengths. In particular, passive microwave appeared to produce better results for rivers with forest-covered floodplains. As well, denser temporal sampling is possible, because the currently used microwave sensors provide at least daily revisits, whereas cloud cover is a limitation for any optical sensor.

4. The SGRs assessed included many relatively small rivers, but even so potentially useful correlations (R > 0.7) were obtained. As expected, best results (R > 0.9) were found for large and unregulated lowland rivers, particularly in tropical and boreal climate zones. Conversely, generally poor results were obtained in arid and temperate regions. This may be attributed to few and rapidly receding flow pulses in arid regions, and a greater prevalence of river regulation and floodplain modification in temperate regions.

5. Although the magnitude of river discharge variations was less important than anticipated, it nonetheless appears likely that higher resolution remote sensing would further increase the number of useful SGRs. However, because the associated rivers are likely to have faster receding flood pulses, increased temporal resolution is also likely to be required in many such cases.

6. The method for automated construction of SGRs that was applied in this study requires discharge or water level observations. It may be possible to relax this requirement in future by using sufficiently accurate river discharge modeling, combined with detailed floodplain elevation data or maximum flood extent mapping to enhance the selection of SGR cells.

References

Alsdorf, D. E., E. Rodríguez, and D. P. Lettenmaier (2007), Measuring surface water from space, *Rev. Geophys.*, *45*, RG2002, doi:10.1029/2006RG000197.

Beck, H. E., A. I. J. M. van Dijk, D. G. Miralles, R. A. M. de Jeu, L. A. Bruijnzeel, T. R. McVicar, and J. Schellekens (2013), Global patterns in base flow index and recession based on streamflow observations from 3394 catchments, *Water Resour. Res.*, *49*, 7843–7863, doi:10.1002/2013WR013918.

Beck, H. E., A. de Roo, and A. I. J. M. van Dijk (2015), Global maps of streamflow characteristics based on observations from several thousand catchments, *J. Hydrometeorol.*, *16*(4), 1478–1501, doi:10.1175/JHM-D-14-0155.1.

Birkinshaw, S. J., G. M. O'Donnell, P. Moore, C. G. Kilsby, H. J. Fowler, and P. A. M. Berry (2010), Using satellite altimetry data to augment flow estimation techniques on the Mekong River, *Hydrol. Processes*, *24*(26), 3811–3825, doi:10.1002/hyp.7811.

Birkinshaw, S. J., P. Moore, C. G. Kilsby, G. M. O'Donnell, A. J. Hardy, and P. A. M. Berry (2014), Daily discharge estimation at ungauged river sites using remote sensing, *Hydrol. Processes*, *28*(3), 1043–1054, doi:10.1002/hyp.9647.

Bjerklie, D. M., D. Moller, L. C. Smith, and S. L. Dingman (2005), Estimating discharge in rivers using remotely sensed hydraulic information, *J. Hydrol.*, *309*(1), 191–209.

Brakenridge, G. R., S. V. Nghiem, E. Anderson, and S. Chien (2005), Space-based measurement of river runoff, *Eos Trans. AGU*, *86*(19), 185–188.

Brakenridge, G. R., S. V. Nghiem, E. Anderson, and R. Mic (2007), Orbital microwave measurement of river discharge and ice status, *Water Resour. Res.*, *43*, W04405, doi:10.1029/2006WR005238.

Brakenridge, G. R., S. Cohen, A. J. Kettner, T. De Groeve, S. V. Nghiem, J. P. Syvitski, and B. M. Fekete (2012), Calibration of satellite measurements of river discharge using a global hydrology model, *J. Hydrol.*, *475*, 123–136.

Brown, J. D., and D.-J. Seo (2013), Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions, *Hydrol. Processes*, *27*(1), 83–105, doi:10.1002/hyp.9263.

De Groeve, T. (2010), Flood monitoring and mapping using passive microwave remote sensing in Namibia, *Geomatics Nat. Hazards Risk*, *1*(1), 19–35.

De Groeve, T., G. R. Brakenridge, and S. Paris (2015), Global flood detection system: Data product specifications (version 2015), *Tech. Rep.* JRC97421, 16 pp., *Eur. Comm. Joint Res. Cent.*, Ispra, Italy.

Durand, M., J. Neal, E. Rodríguez, K. M. Andreadis, L. C. Smith, and Y. Yoon (2014), Estimating reach-averaged discharge for the River Severn from measurements of river water surface elevation and slope, *J. Hydrol.*, *511*, 92–104.

Garambois, P.-A., and J. Monnier (2015), Inference of effective river properties from remotely sensed observations of water surface, *Adv. Water Resour.*, *79*, 103–120.

Gleason, C. J., L. C. Smith, and J. Lee (2014), Retrieval of river discharge solely from satellite imagery and at-many-stations hydraulic geometry: Sensitivity to river form and optimization parameters, *Water Resour. Res.*, *50*(12), 9604–9619, doi:10.1002/2014WR016109.

Guerschman, J., W. Garth, G. Byrne, L. Lymburner, M. Norman, and A. I. J. M. Van Dijk (2011), MODIS-based standing water detection for flood and large reservoir mapping: Algorithm development and applications for the Australian continent, report, CSIRO, Canberra. [Available at http://www.clw.csiro.au/publications/waterforahealthycountry/wirada/TechReports/WIRADA-MODIS-standing-water.pdf.]

Hannah, D. M., S. Demuth, H. A. van Lanen, U. Looser, C. Prudhomme, G. Rees, K. Stahl, and L. M. Tallaksen (2011), Large-scale river flow archives: Importance, current status and future needs, *Hydrol. Processes*, *25*(7), 1191–1200.

Hashino, T., A. A. Bradley, and S. S. Schwartz (2007), Evaluation of bias-correction methods for ensemble streamflow volume forecasts, *Hydrol. Earth Syst. Sci.*, *11*(2), 939–950, doi:10.5194/hess-11-939-2007.

Kruus, J., M. Deutsch, P. Hansen, and H. Ferguson (1981), Flood application of satellite imagery, Proceedings of the 5th Annual William T. Pecora Memorial Symposium on Remote Sensing, Tech. Publ. Ser. TPS81-1, pp. 292–301, *Am. Water Resour. Assoc.*, Bethesda, Md.

Kugler, Z., and T. De Groeve (2007), The global flood detection system, *Off. for Off. Publ. of the Eur. Commun., Luxembourg*, Ispra, Italy.

Leopold, L. B., and T. Maddock Jr. (1953), The hydraulic geometry of stream channels and some physiographic implications, *Rep. 2330-7102*, USGS Professional Paper *252*, 57 pp.

Liu, Y., W. Dorigo, R. Parinussa, R. De Jeu, W. Wagner, M. McCabe, J. Evans, and A. Van Dijk (2012), Trend-preserving blending of passive and active microwave soil moisture retrievals, *Remote Sens. Environ.*, *123*, 280–297.

Madadgar, S., H. Moradkhani, and D. Garen (2014), Towards improved post-processing of hydrologic forecast ensembles, *Hydrol. Processes*, *28*(1), 104–122, doi:10.1002/hyp.9562.

Nobre, A. D., L. A. Cuartas, M. R. Momo, D. L. Severo, A. Pinheiro, and C. A. Nobre (2015), HAND contour: A new proxy predictor of inundation extent, *Hydrol. Processes*, *30*, 320–333.

Papa, F., C. Prigent, and W. B. Rossow (2008), Monitoring flood and discharge variations in the large Siberian rivers from a multi-satellite technique, *Surv. Geophys.*, *29*(4), 297–317, doi:10.1007/s10712-008-9036-0.

Pavelsky, T. M. (2014), Using width-based rating curves from spatially discontinuous satellite imagery to monitor river discharge, *Hydrol. Processes*, *28*(6), 3035–3040, doi:10.1002/hyp.10157.

Pavelsky, T. M., M. T. Durand, K. M. Andreadis, R. E. Beighley, R. C. Paiva, G. H. Allen, and Z. F. Miller (2014), Assessing the potential global extent of SWOT river discharge observations, *J. Hydrol.*, *519*, 1516–1525.

Revilla-Romero, B., J. Thielen, P. Salamon, T. De Groeve, and G. R. Brakenridge (2014), Evaluation of the satellite-based Global Flood Detection System for measuring river discharge: Influence of local factors, *Hydrol. Earth Syst. Sci.*, *18*(11), 4467–4484, doi:10.5194/hess-18-4467-2014.

Revilla-Romero, B., H. E. Beck, P. Burek, P. Salamon, A. de Roo, and J. Thielen (2015), Filling the gaps: Calibrating a rainfall-runoff model using satellite-derived surface water extent, *Remote Sens. Environ.*, *171*, 118–131.

Schmit, T. J., M. M. Gunshor, W. P. Menzel, and J. J. Gurka (2005), Introducing the next-generation Advanced Baseline Imager on GOES-R, *Bull. Am. Meteorol. Soc.*, *86*(8), 1079.

Shiklomanov, A., R. Lammers, and C. J. Vörösmarty (2002), Widespread decline in hydrological monitoring threatens pan-Arctic research, *Eos Trans. AGU*, *83*(2), 13–17.

Smith, L. C. (1997), Satellite remote sensing of river inundation area, stage, and discharge: A review, *Hydrol. Processes*, *11*(10), 1427–1439.

Smith, L. C., and T. M. Pavelsky (2008), Estimation of river discharge, propagation speed, and hydraulic geometry from space: Lena River, Siberia, *Water Resour. Res.*, *44*, W03427, doi:10.1029/2007WR006133.

Smith, L. C., B. L. Isacks, A. L. Bloom, and A. B. Murray (1996), Estimation of discharge from three braided rivers using synthetic aperture radar satellite imagery: Potential application to ungaged basins, *Water Resour. Res.*, *32*(7), 2021–2034.

Tarpanelli, A., L. Brocca, T. Lacava, M. Faruolo, F. Melone, T. Moramarco, N. Pergola, and V. Tramutoli (2011), River discharge estimation through MODIS data, *Proc. SPIE*, *8174*, 817408.

Tarpanelli, A., L. Brocca, T. Lacava, F. Melone, T. Moramarco, M. Faruolo, N. Pergola, and V. Tramutoli (2013), Toward the estimation of river discharge variations using MODIS data in ungauged basins, *Remote Sens. Environ.*, *136*, 47–55, doi:10.1016/j.rse.2013.04.010.

van Dijk, A. I. J. M., L. J. Renzullo, Y. Wada, and P. Tregoning (2014), A global water cycle reanalysis (2003-2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble, *Hydrol. Earth Syst. Sci.*, *18*(8), 2955–2973, doi:10.5194/hess-18-2955-2014.

Verkade, J. S., J. D. Brown, P. Reggiani, and A. H. Weerts (2013), Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales, *J. Hydrol.*, *501*, 73–91, doi:10.1016/j.jhydrol.2013.07.039.

Vörösmarty, C. J., C. J. Willmott, B. J. Choudhury, A. L. Schloss, T. K. Stearns, S. M. Robeson, and T. J. Dorman (1996), Analyzing the discharge regime of a large tropical river through remote sensing, ground-based climatic data, and modeling, *Water Resour. Res.*, *32*(10), 3137–3150, doi:10.1029/96WR01333.

Wood, A., E. Maurer, A. Kumar, and D. Lettenmaier (2002), Long-range experimental hydrologic forecasting for the eastern United States, *J. Geophys. Res*, *107*(D20), 4429, doi:10.1029/2001JD000659.