This is an Open Access document downloaded from ORCA, Cardiff University’s institutional repository: https://orca.cardiff.ac.uk/id/eprint/141826/

This is the author’s version of a work that was submitted to / accepted for publication.

Citation for final published version:


Publishers page: http://dx.doi.org/10.1016/j.scitotenv.2021.147762

Please note: Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
Estimation of river flow using CubeSats remote sensing

Adriano M. Junqueira a,b*, Feng Mao c, Tatiana S. G. Mendes d, Silvio J. C. Simões a,c, José A. P. Balestieri a, David M. Hannah b

a São Paulo State University (UNESP), School of Engineering, Guaratinguetá - Brazil
b University of Birmingham, School of Geography, Earth and Environmental Sciences, Birmingham – United Kingdom
c Cardiff University, School of Earth and Environmental Sciences, Cardiff - United Kingdom
d São Paulo State University (UNESP), Institute of Science and Technology, São José dos Campos - Brazil

*Corresponding author:
Email: adriano.junqueira@unesp.br (A. M. Junqueira)

Av. Dr. Ariberto Pereira da Cunha, 333 – Guaratinguetá – SP – Brazil – Zipcode: 12516-410

Abstract:

River flow characterizes the integrated response from watersheds, so it is essential to quantify to understand the changing water cycle and underpin the sustainable management of freshwaters. However, river gauging stations are in decline with ground-based observation networks shrinking. This study proposes a novel approach of estimating river flows using the Planet CubeSats constellation with the possibility to monitor on a daily basis at the sub-catchment scale through remote sensing. The methodology relates the river discharge to the water area that is extracted from the satellite image analysis. As a testbed, a series of Surface Reflectance PlanetScope images and observed streamflow data in Araguaia River (Brazil) were selected to develop and validate the methodology. The study involved the following steps: (1) survey of measurements of water level and river discharge using in-situ data from gauge-based Conventional Station (CS) and measurements of altimetry using remote data from JASON-2 Virtual Station (JVS); (2) survey of Planet CubeSat images for dates in step 1 and without cloud cover; (3) image preparation including clipping based on different buffer areas and calculation of the Normalized Difference
Vegetation Index (NDVI) per image; (4) water bodies areas calculation inside buffers in the Planet CubeSat images; and (5) correlation analysis of CubeSat water bodies areas with JVS and CS data. Significant correlations between the water bodies areas with JVS (R² = 88.83%) and CS (R² = 96.49%) were found, indicating that CubeSat images can be used as a CubeSat Virtual Station (CVS) to estimate the river flow. This newly proposed methodology using CubeSats allows for more accurate results than the JVS-based method used by the Brazilian National Water Agency (ANA) at present. Moreover, CVS requires small areas of remote sensing data to estimate with high accuracy the river flow and the height variation of the water in different timeframes. This method can be used to monitor sub-basin scale discharge and to improve water management, particularly in developing countries where the presence of conventional stations is often very limited.

**Highlights:**

- Use of remote sensing information from Planet CubeSats constellation to build and assess a methodology to river flow estimation;
- This method has significant opportunity for river flow estimation at ungauged sites at the daily and sub-basin scales;
- The improvement of river flow measurements is essential to understand the changing water cycle and underpin the sustainable management of freshwaters.

**Keywords:** River flow; CubeSat; Remote Sensing; Cerrado (Savannah); Change Detection; Spatiotemporal Resolution.
1 Introduction

Freshwater is a basic requirement for life but the knowledge of river flow rates is scarce (Gleason and Smith, 2014). A better understanding of the large-scale water cycle process is essential to underpin socio-economic development and sustainably manage water-dependent ecosystems (Döll et al., 2014; Hannah et al., 2011; Kingston et al., 2020). River flow characterizes the integrated hydrological response and water yield from watersheds. Nowadays, there is a clear decline of river gauging stations and a shrinking of ground-based observation networks (Dixon et al., 2020; Hannah et al., 2011).

Conventional gauging stations are well developed and have contributed to quantify the movement of water in river channels. However, conventional stations are not enough to determine more complex riverine environments that involve the movement of water over wetlands and floodplains in multiple channel types (Lettenmaier, 2007), requiring new multidisciplinary approaches to improve the observation networks.

The advances in remote sensing hydrology, particularly over the past 10 years, have demonstrated that hydraulic variables can be measured reliably from orbiting satellite platforms (Huang et al., 2018). As the deluge of big data continues to impact practically every commercial and scientific domain, geosciences have also witnessed a major revolution from being a data-poor field to a data-rich field (Karpatne et al., 2019; Reichstein et al., 2019). The use of remote sensing by satellite for streamflow analyses can be categorized into techniques based on satellite altimetry, Synthetic Aperture Radar (SAR), and optical images (Ahmad and Kim, 2019).

To accurately detect river flow regimes at a field scale from space, a new high spatial and temporal resolution remote sensing source is necessary to improve water-level time series (Bogning et al., 2018). While satellites such as Sentinel-2 and Landsat may have an adequate spatial resolution for different applications (Sadeh et al., 2019), their temporal resolution (5 and 16 days revisit time, respectively) is not ideal for detection of flow regimes changes, as there may be weeks between the acquisition of two clear-sky images (Houborg and McCabe, 2018a).
Applications of satellite remote sensing in hydrological surface water modelling, mapping, and parameter estimation were reported in some reviews based on Earth Observing Systems (Huang et al., 2018; Joshi et al., 2016; Musa et al., 2015; Wagner et al., 2018) mainly using SAR, optical, altimetry and DEM data. Focusing flow river estimation, different techniques were applied to estimate the river width by averaging multiple cross-sections over an area (Gleason and Smith, 2014), use the Synthetic Aperture Radar (SAR) to calculate water area (Ahmad and Kim, 2019), and fusion topography dataset (Anh and Aires, 2019; Moramarco et al., 2019). The spatiotemporal restrictions for using remote sensing satellite data have been partly overcome via multisensor data fusion (Houborg and McCabe, 2018b) but all solutions proposed can not be replicated for daily measurements at a high spatial resolution.

Thus, remote sensing has the potential of conducting rapid, cost-effective, and continuous surveys of river management practices over large scales. Notably, the constellations of micro or nano-satellites, known as CubeSats, are revolutionizing the high spatiotemporal resolution possibilities in remote sensing and can potentially be used to capture many observations over time (Sadeh et al., 2019) and monitor dynamics surface water changes (Cooley et al., 2019, 2017). The repetitive observation mechanism of multiples CubeSats enables studying the river dynamics and observe different physical properties (Marinho et al., 2020). Among the different Cubsats on orbit, the CubeSat constellation provided by Planet Labs Inc. has the advantage of providing a near-daily revisit time globally at 3-meter orthorectified spatial resolution. It is also worth noting that Remote Sensing (RS) has several methods and techniques to identify land and water areas considering bands variation of different multi-spectral images (Acharya et al., 2018). Index methods are mostly used to estimate surface water that separates the water from the background based on a threshold value. Among these indexes, Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) are frequently adopted as they include visible and Near Infrared (NIR) bands provided by satellite optical images (Elsahabi et al., 2016).

The aims of this study were (1) to develop an innovative methodology for semi-automated river flow estimation using visible and NIR from Planet CubeSats bands to detect changes in the surface of flood area and (2) to compare Planet CubeSats derived river flow with Conventional
flow meters Station (CS) and JASON (Joint Altimetry Satellite Oceanography Network) Virtual gauging Station (JVS). The Araguaia River, in the Cerrado biome of Brazil, was selected as the research area.

2. Description of the study area

The Tocantins-Araguaia hydrographic region, with a total area of 920,087 km² and 13,779 m³/s of average discharge (Brasil/ANA, 2015), is the most important fluvial system draining the tropical savannah ecoregion of Brazil (Cerrado biome). Some areas of this region have been affected by severe water scarcity events since 2012 (Naturatins, 2017). In general terms, the hydrologic regime depends on the dominant climate (tropical wet-dry) with floods from January to May (rain period) and low water between June to September. This region is the confluence of two major rivers: the Tocantins River (~1,960 km extension) and the Araguaia River (~2,600 km extension). The Araguaia sub-basin (around 386,765 km²) represents around 42% of the hydrographic region and along its path, is placed the largest river island in the world, Bananal Island (Figure 1).

The Araguaia River is one of the priority areas for conservation of the aquatic biodiversity of the Cerrado biome and has been the target of political and environmental debates due to the intense and indiscriminate expansion of agricultural activities, with a greater degradation of the natural environment during the last four decades (Latrubesse and Stevaux, 2006). The entire sub-basin has just 166 conventional river gauging stations for an area of 386,765 km² and more than 6,000 river stretches mapped.

This region needs to be better monitored due to the increase in intensive agriculture and the use of water (Althoff et al., 2020). Especially in the Araguaia River, the irrigated area increased
more than 116% between 2006 and 2012, and the cultivated area increased by about 20% in the
same period (ANA, 2015). In the last decade, the Araguaia sub-basin has been suffering from
water scarcity in several tributaries of the Araguaia river, obliging managers to take actions for
rationing the water use in agriculture to prioritize human and animal supply (Lauris, 2019;
Naturatins, 2017).

As a testbed to develop and validate the proposed methodology, two river gauging stations
that are calibrated by ANA and used as an official instrument for public policies in this region
were selected. One of the rivers gauging stations was the in-situ Conventional Station (CS) ID
26350000 (http://gestorpcd.ana.gov.br) managed by ANA, located in São Felix do Araguaia –
MT (11°37’8.6” S; 50°39’75.0” W – WGS84 Datum). This station employs an instrument for
monitoring the river flow for 24/7 hours with records registered every 15 minutes. The other river
gauging station was the JASON-2 Virtual Station (JVS) ID 1055S05036WO
(http://hidrosat.ana.gov.br), which is monitored by ANA in cooperation with Institut de
Recherche pour le Développement (IRD) with barycenter data at 10°54’42.0” S; 50°36’48.6” W
– WGS84 Datum. This station acquires altimetry information by satellite to monitor the water
level and derived river flow, with no equipment locally installed. The JVS dataset is related to the
average altitude of all available altimetry data, along the JASON-2 track, over the river area in
each satellite cycle. JVS observations occur every 9.9 days from the JASON-2 satellite in the
same ground track within a margin of ±1 km (CNES; NASA, 2011).

The detailed characteristics of both selected stations are shown in Figure 2. The distance
between the CS and JVS is relatively close with 78 km (Euclidean distance) allowing their
comparison with the proposed methodology.

*Figure 2*
3. Data and Methods

For this study, the available Planet CubeSat data was explored to establish an innovative method for estimating river flow based on daily remote sensing images.

3.1. CubeSat data

This study used Planet images to develop a flow estimation model. These images are captured by Planet using an approach based on ‘fast design’ to launch satellites, mission control and operations systems, and the development of a web-based platform for imagery processing and delivery (www.planet.com). Planet is a commercial satellite operator that enhanced observation capacity offered by constellations of small and standardized satellites and employs an “always-on” image-capturing method (Planet Team, 2020).

This CubeSat constellation consists of 130 small-satellites at an altitude of approximately 475 km, following each other on two near-polar orbits of roughly 8° (descendent orbit) and 98° (ascendent orbit) inclination respectively, imaging the Earth at local morning time (Planet Team, 2020). The distance along the orbit between the CubeSats is constructed such that the longitudinal progression between them over the rotating Earth leads to the scan of the surface. Thus, the full constellation provides daily sun-synchronous coverage of the entire Earth (except the polar hole) with the resolution of 3.7 meters at nadir (Ground Sample Distance), 12-bit radiometric resolution, and 4 spectral bands (Blue [455 - 515 nm], Green [500 - 590 nm], Red [590 - 670 nm] and Near-InfraRed [780 - 860 nm]) (Planet Team, 2020).

The CubeSat across-track off-nadir viewing angle used for imaging usually to be lower than 5° (Planet Team, 2020) reducing the complexity to evaluate bands’ variation of different multi-spectral images arising from environmental noise such as shadow, however, these CubeSat do not present any spectral bands as Middle Infrared (MIR) and Shortwave Infrared (SWIR), used in different indexes applied to water such as TCW, MNDWI, NDWI, AWEI (Elsahabi et al., 2016).
The images can be downloaded from the Planet portal (https://www.planet.com/explorer) and include three types of raster images: (a) digital number (DN) commonly represented by an uncalibrated image into physically meaningful units; (b) top of atmosphere reflectance (TOA) that is the reflectance measured by a space-based sensor flying higher than the earth's atmosphere (calibrated to a radiance image), and (c) surface reflectance (SR) that is the radiance image atmospherically corrected and ready to be used to extract quantitative information about features on the Earth surface. As SR reflects the difference among land covers more accurately than other remotely sensed measurements (Huang et al., 2018). In this study, it was used the SR image products, orthorectified with 3 meters spatial resolution, positional accuracy with less than 10-meters Root Mean Square Error (RMSE) suitable for analytic and visual applications.

3.2. A method for estimating river flows using CubeSats

This study was divided into five steps: (1) survey of measurements of water level and river discharge using in-situ data from gauge-based Conventional Station (CS) and measurements of altimetry using remote data from JASON-2 Virtual Station (JVS); (2) preparation of river section and buffers and survey of CubeSat images for selected dates (correlated with step 1 and without cloud cover); (3) image preparation, including clipping based on different buffer areas followed by NDVI calculation per image, and data processing over CS and JVS stations using Extract-Transform-Load (ETL); (4) water bodies areas calculation inside buffer for river flow estimation in the Planet CubeSat images; and (5) correlation analysis of CubeSat water bodies areas with JVS and CS data (Figure 3).

Figure 3

Firstly, the CS and JVS measurements were collected in the period of 01/01/2018 to 30/07/2018, a period that historically includes the greatest variation in the flow of this river over the years. The dataset for CS included the water level and river flow every 15 minutes. The JVS
were collected altimetry data in repetitive periods of 10 days according to the measures provided by the altimetry missions satellite. The use of JVS data was independent of the time of collection and resulted in a median of 11:03 a.m. local time.

In the second step, a cross-section of the river was used as a reference for creating side buffers with 50, 250, 500, and 1000 meters. This cross-section was determined nearby to the JASON’s tracking to provide better conditions of comparison between the virtual stations and these buffers sections were created to reduce the satellite image processing area in the water surface calculation. Using the location of these buffers, Planet CubeSat images were searched for all dates with JVS data. The CubeSat images search also included the completed cover of buffer areas, the absence of clouds, and the possibility of using the surface reflectance product (SR) as input (Table 1).

In the third step, each image downloaded from the Planet portal was clipped on the buffer areas and processed to calculate the Normalized Difference Vegetation Index (NDVI) with a scale ranging from -1 to 1. The NDVI was calculated using the Red and Near-InfraRed regions of the electromagnetic spectrum with Equation 1 since these satellites do not have yet more specific spectral bands for the development of more sophisticated methods. For each image analyzed, the river flood areas were classified with NDVI values lower than 0.15.

\[ NDVI = \frac{(NIR\ Band - Red\ Band)}{(NIR\ Band + Red\ Band)} \] (1)

In the fourth step, the water bodies' areas were calculated for all clipped buffers in the dates analyzed, resulting in a temporal table of flooding areas. Then, in the fifth step, we performed the regression curves (exponential, linear, logarithmic, polynomial, power-law, and moving averages) between the water bodies areas for 4 different buffers (50, 250, 500, and 1000 m) and the 3 sets of reference data (JVS - altimetry, CS - water level, CS - flow). After that, 12 different Pearson coefficients (R²) were determined according to the best-fitted regression curves. The relationship between R² and the four buffer areas were analyzed for each set of reference data,
determining the equation used to predict the river flow from the CubeSat image flood area, which represented the CubeSat Virtual Station (CVS).

4 Results

The data collected from CS showed a river flow ranging from 868 to 4,739 m$^3$/s and a water level varying from 328 to 710 cm for the entire monitoring period in the Araguaia basin. A significant relationship between the river flow and the water level measurements ($R^2=1$) was found (Figure 4a). Concerning the JVS data, altimetry values between 177.9 to 181.1 m were observed, with a good correlation to CS water level ($R^2=0.85$) (Figure 4b).

Figure 4

In the survey of Planet CubeSat images, 8 image dates were selected with CS and JVS correspondent data, completed cover the buffer areas, absence of clouds, and available surface reflectance product (SR) (Table 1).

Table 1

In these 8 images, water bodies areas ranging from 39,924 to 54,846 m$^2$ for buffer 50 m, from 173,313 to 276,984 m$^2$ for buffer 250 m, from 320,805 to 557,586 m$^2$ for buffer 500 m, and from 563,895 to 1,097,469 m$^2$ for buffer 1000 m were determined (Table 2). Considering the maximum values of water bodies in each buffer, it was verified that the flood area corresponds to 77.3%, 63.9%, 52.5%, and 37.7%, respectively, for the 50 m, 250 m, 500 m, and 1000 m buffer areas.

Table 2
Based on the water bodies areas determined in each buffer area, the regression models were performed as demonstrated in Figure 5. Due to the obtained data values and its measure of greatness, the river flow equation was adjusted to the power-law regression model, while the water level and altimetry equations were adjusted to linear regression models.

**Figure 5**

Considering the results obtained in the regression models, it was found that the buffer of 500 m was the lowest buffer area capable of providing a high Pearson’s coefficient that remained stable even with the increase of buffer area (Figure 6). Therefore, the buffer of 500 m was selected to be the CVS and to estimate the river flow.

**Figure 6**

To estimate the river flow ($Qe$), water level ($Le$) and altimetry ($Ae$) were determined the Equations 2 ($R^2 = 96.49\%$), 3 ($R^2 = 94.38\%$) and 4 ($R^2 = 88.83\%$) respectively; in which $Af500r$ is the flooding area (m²) in the 500 m buffer.

\[
Qe = 10^{-1} \cdot Af500r^{2.8737} \quad (2)
\]

\[
Le = 0.0015 \cdot Af500r - 161.56 \quad (3)
\]

\[
Ae = 10^{-5} \cdot Af500r + 175.14 \quad (4)
\]

The root mean square error (RMSE) is used as the accuracy estimator (Equation 5), and the standard deviation (SD) is used as the precision estimator (Equation 6). These estimators were determined for river discharge ($Qe$), water level ($Le$), and altimetry ($Ae$) are presented in Table 3.
In summary, this CVS methodology established for the Araguaia river can be explored to estimate the river flow, water level, and altimetry for other rivers around the world.

5 Discussion

Hydrographic data obtained from satellites and other remote sources provide the possibility of broad global coverage for river discharge estimates (Bahadur and Samuels, 2013; Lakshmi, 2004). The advances in computing power and data storage capacity associated with the innovations in the satellite remote sensing area are enabling global monitoring of different variables related to the water cycle (Lettenmaier et al., 2015; Wagner et al., 2018). Nowadays, the increase in the number of Planet CubeSats brings images with more cost-effective and higher spatiotemporal resolutions than other commercial satellites (Houborg and McCabe, 2018b).

Currently, Planet CubeSat is the only commercial constellation available for capturing daily optical images with high resolution of the entire surface of the Earth. In Brazil, there is a national program (https://www.gov.br/mj/pt-br/acesso-a-informacao/acoes-e-programas/programa-brasil-mais/) that provides Planet images with high-resolution (3 m orthorectified per pixel) freely available to governmental institutions throughout the Brazilian territory.

Access to the CubeSat images is an important political and economic decision. The use of satellite information is an economical way of measuring river discharge using in situ gauges stations that are costly to install, maintain, and operate (Zaji et al., 2018). According to U.S.
Geological Survey (USGS), the cost for a typical in situ gauge station involves several costs associated with its activities, which are estimated at 41% for labor staff (field and office), 25% for administrative activities, 10% for building and utilities, 10% for field equipment, 7% for data management and delivery, 5% for vehicles and 2% for travel (Norris, 2010). These percentages can vary according to location and conditions, especially in remote areas where in situ gauges stations require expensive field works (Norris, 2010). Besides reducing costs, the implementation of technology-based remote sensing for river discharge can avoid exposing surveyors to dangerous and reacher inaccessible rivers (Samboko et al., 2020).

Monitoring of rivers requires a reliable system, being the water level and the river discharge the two essential parameters in this analysis (Mao et al., 2020, 2019, 2018; Mostafavi, 2018). Besides that, the monitoring requires integrated modeling tools that cover adequate spatial and temporal scales involving mathematical applications (Mannschatz et al., 2015). In this context, an innovative methodology for river flow estimation was developed using Planet CubeSat images to detect changes in the flood area surface, which can be used as a CubeSat Virtual Station (CVS).

Although methods of river discharge from the direct measurement of width, depth, and velocity (based on velocity-area method) provides a higher level of accuracy than orbital remote sensing (Bjerklie et al., 2005), the proposed methodology used an approach that relies on identifying the water surface from morphologic features that are easier to recognize from space. The geomorphic features and structural dynamics related to river discharge as channel type, channel slope, channel roughness, depth, and velocity were assumed associated with the river hydraulic geometries and can be used to develop more robust calibration methods.

The CVS data obtained were compared with the measures of water level and river discharge of a Conventional Station (CS) and altimetry of the JASON Virtual Sation (JVS), located in the Araguaia River (Brazil).

Analyzing the CS data collected in this study, it was found a complete correlation ($R^2 = 1$) between the measurements of water level and river discharge, indicating that the in-situ reference station used was well-calibrated. It was also observed a high correlation of the CS water level measurements with the JVS altimetry data ($R^2 = 0.85$) confirming that the satellite remote sensing
can be a useful tool for river flow estimation. Bogning et al., (2018) also found a good correlation (R² > 0.82) when in-situ gauge records were compared to altimetry-based water levels from multiple satellites, composed of a network of altimetric virtual station (ENVISAT, SARAL, ERS-2, Sentinel-3A, JASON-2, and JASON-3 data). This analysis was performed in the Ogooué river basin, located at Gabon, with an annual river discharge of 4,750 m³/s and a hydrologic wet-dry regime similar to the characteristics of this study in the Araguaia River. Smith and Pavelsky, (2008) demonstrated that remotely sensed width variations were well correlated to ground measurements of river discharge (R² = 0.81) when taken days later and hundreds of kilometres downstream. Gleason and Smith (2014) showed that useful estimates of absolute river discharge may be obtained solely from river width using multiple satellite Landsat images, through a characteristic scaling law named At-Many-station Hydraulic Geometry (AMHG), with no ground-based or a priori information. The AMHG was calculated with the monitoring of large extensions (10 to 13 km) from remote sensing along the river.

JVS and Landsat can be employed in the case of lacking river rating curves and cross-sectional geometries, as well as when the water levels or flow rates measurements are missing in situ station historic data. However, the use of JVS and Landsat data in a river hydrodynamics context is limited by data coverage in both time and space, which may be insufficient to capture key spatiotemporal variations in water surface elevation daily (Houborg and McCabe, 2018a). Besides that, JVS can be used only to monitor the level of wider rivers (Huang et al., 2018) that intersect with JASON-2 satellite tracks.

In this study, the innovative method using Planet CubeSat images provides a possibility to monitor river narrower than those evaluated by JVS and Landsat due to its higher spatial resolution (Houborg and McCabe, 2018b, 2018a). Also, Planet constellation allows monitoring rivers around the world (Kääb et al., 2019) with more flexibility to establish Virtual Stations concerning JVS that are limited to the track satellite intersections. The fact of use high-resolution Planet CubeSat images, with low acquisition inclination, reduces the effect of shadow and increases the river border identification details. This agrees with Bjerkli et al. (2003), who reinforces that even if the river could always be distinguished from the surrounding landscape,
narrower channels would have greater uncertainty in the width estimates due to the relative width
concerning the resolution of the sources. The altimetry accuracy for the JVS stations varies with
the river width, with better precision the wider the channels (Bjerklie et al., 2018), while for the
proposed CVS stations the accuracy in determining the level varies with the flooded area close to
the station, with precision dependent on the slope at the edges of the channel, as it reflects in the
expansion of the flooded area.

For the period of study, 8 images were selected with dates correspondent to CS and JVS
reference measures. However, many other Planet CubeSat images were available in the period,
with 110 cloud-free SR images against 22 JVS measures, representing at least 5 times more
information than JVS data for the Araguaia river. Although many Planet images were available,
it was decided to use only the images that allowed the comparison with CS and JVS on the same
date.

The data from CubeSat images showed a well-correlated estimation with river discharge
(R² = 96.49%) when small lengths of the river (500 m buffer) were analyzed. This correlation was
higher than the JVS data and river discharge correlation (R²=85.01%) that is one of reference
adopted by the ANA in Brazil. The results using CubeSat images were also well correlated with
CS water level (R² = 94.38%) and JVS altimetry (R² = 88.83%) when evaluated the flooding area
(m²). According to Papa et al. (2010), ideally, the goal for discharge data accuracy is within ±5%
related to the true value, but the community agrees that 15% to 20% accuracy is in general
acceptable for discharge measurements.

Virtual stations with CubeSat images showed greater accuracy and precision at lower river
discharge rates (Fig. 5) whereas JASON virtual stations have greater accuracy and precision for
higher river discharge rates (Fig. 4b). This is observed by the fact that in the CVS, a better
refinement of the flooded area is possible even with the presence of sandbanks while the JVS
presents greater noise in the identification of the altimetry related to these areas. At higher river
discharges, we observed that the JVS presented better details of the altimetry, especially when the
increase in the level of the river occurred inside the channel without accompanying the expansion
of the flooded area.
In this study, it was visually observed a refined design of the flooded areas due to the resolution of the CubeSat images. From this observation, an attempt was made to find a relationship between the width of the river and a buffer size that provided less demand of image areas to achieve a high Pearson coefficient. It was observed that a buffer ranging from 0.5 to 1 times the average width of the studied river section allowed to reach these results, remaining stable even the value of $R^2$. Then, the proposed methodology allows the use of less than 388 km² of images/year for this virtual station, representing an advantage in comparison with the AMGH methodology, which suggests the evaluation of multiple river widths by stretches greater than 10 km. Using 3-meter Planet images as reference, Pôssa et al. (2018) observed a slightly increased precision in the water surface delineation compared to Landsat and Sentinel images due to spatial resolution of satellite images.

Overall, these experiments allow us to employ a simple exponential equation model with daily CubeSat images for well predict the river flow at a monthly scale, based on the surface hydrological information measured from space as proof of concept and utility of the method. The use of new constellations, new hydrological science methods, and advancements in resolutions (spatial, temporal, and radiometric) of remote sensing make possible the application of monitoring increasingly smaller watersheds, as they have coverage of several pixels and more frequent data acquisitions (Lakshmi, 2004). Besides that, to flow estimation measurement, the CubeSat images can be used to study other processes related to data assimilation; flood monitoring and prediction; floodplain connectivity (Cooley et al., 2017); land surface characteristics (land use, temperature, snow cover) (Reichle, 2008); image fusion to land use mapping and monitoring (Houborg and McCabe, 2018b; Joshi et al., 2016); and water quality monitoring (Maciel et al., 2020).
Conclusion

An innovative method for semi-automated river flow estimation using Planet CubeSat data was developed to detect changes in the surface of the flood area, using the Araguaia River as a testbed. In summary, the flood areas detected by CubeSat data showed significant correlations with river discharge and water level measurements from gauge-based Conventional Station (CS) in small areas of the river using a reduced amount of satellite images. CubeSat data also presented a significant correlation with altimetry measurements from JASON-2 Virtual Station (JVS) officially adopted by the Brazilian agency, with the advantage to provide greater records numbers per year, more flexibility of position for Virtual Stations establishment, and the possibility to monitor narrower rivers. The river discharge extracted from Cubesat data showed a higher correlation with CS than JVS, indicating that CVS had more capacity of reproducing the ground truth compared to JVS.

In the future, this method can be completely automated to fill the gaps in the streamflow series, to compute different riverine contributions in the sub-basin, and to promote an understanding of river discharge spatial distribution on a near-real-time for entire continents. This method can also be used as a cost-effective alternative to monitoring the sub-basin discharges, improving water management, particularly, in developing countries where the presence of conventional stations is limited.

Acknowledgements

The authors are grateful to Brazilian National Water Agency (ANA – Agência Nacional de Águas) for providing the fluvial data in Brazil (conventional and virtual stations) and, to SCCON/PLANET data images accessed from the online portal (https://www.planet.com/explorer).

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.
Adriano M. Junqueira was hosted at the University of Birmingham, UK, to undertake this work.

Jose Antonio Perrella Balestieri is grateful to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), process 301853/2018-5.

References


Bogning, S., Frappart, F., Blarel, F., Niño, F., Mahé, G., Bricquet, J.P., Seyler, F., Onguéné, R., Etamé, J.,


Houborg, R., McCabe, M.F., 2018b. Daily retrieval of NDVI and LAI at 3 m resolution via the fusion of CubeSat, Landsat, and MODIS data. Remote Sens. 10. https://doi.org/10.3390/rs10060890

https://doi.org/10.1029/2018RG000598

https://doi.org/10.3390/rs8010070


https://doi.org/10.1109/TKDE.2018.2861006


https://doi.org/10.1029/2006RG000197.1


Fact Sheet 2010-3025).


Figures

Fig. 1. Overview of the study area. (A) Overview of South America with its 10 Hydrographic Regions (black), passes of JASON-2 over the area (red), and the Tocantins-Araguaia basin (light blue); (B) Zoom in the Tocantins (light-green) - Araguaia (yellow) basin boundaries, presence of rivers with federal regulation (blue) and all fluviometric in situ stations (black dots), complemented by tracks of Jason-2 (red) highlighting JVS (red circle; north) and CS (pink circle; south) ground references.
Fig. 2. The study area of the Araguaia sub-basin. (A) Zoom in the Araguaia River (center) highlighting the track of Jason-2 (red) over the surface drainage (blue), and the Conventional Station (CS) used as ground truth (pink dot). The region where the Jason-2 track crosses the Araguaia River is the Virtual Station (JVS) far 78 km from CS and, the region used to download the CubeSat SR images; (B) Zoom in the VS (red dot) over the Araguaia River (blue line) with the 4 different buffer areas used in the methodology (50 m – grey; 250 m – yellow; 500 m – orange; 1000 m – red). In the background it is presented the Surface Reflectance (SR) Planet image from July, 20th 2018 with 50% transparency and the representation of the Jason-2 track over the region (red dashed line).
Fig. 3. Methodological framework using CubeSat to estimate the river flow. The 1st step is used for gathering reference information to support the key-curves. In the 2nd step are prepared the specific buffer areas and selected the CubeSat images to investigate the correlation between inundation extent and river flow. The 3rd step is used for preprocessing selected CubeSat images with pixel classification (water/no-water) based on NDVI values. In the 4th step are calculated the flooding area for each image. The last step is to evaluate the regression analysis comparing the 3 reference data (CS - Flow river; CS – Water Level; JVS – Altimetry).
**Fig. 4.** Evaluation of the reference data. (a) Relationship between the river level (x-axis) and flow (y-axis) by the CS in the São Felix do Araguaia/MT station for the period of analyses. The blue curve represents the exponential equation \( y = 0.0026x^{2.1978} \) that are well adjusted with \( R^2 = 1 \).

(b) Relationship between the altitudes (x-axis) calculated by the JVS and the quotes determined on the water level (y-axis) by the CS in the Araguaia basin for the period of analyses. The blue curve represents the linear equation \( y = 128.65x - 22585 \) with \( R^2 = 0.8501 \). In both figures, the solid blue circles represent correlated values with 10 days interval.
Fig. 5. Regression curves against 3 ground references and 4 different buffer sizes
Fig. 6. Relationship between R² (x-axis) and 4 buffers distances (y-axis) for 3 sets of reference data (CS water level and river flow and JVS altimetry).
Table 1. Detail of measurements related to JVS, CS, and satellite images used in the period of analysis for river flow estimates.

<table>
<thead>
<tr>
<th>Label ID</th>
<th>Date</th>
<th>JASON Virtual Station (JVS) Altimetry (m)</th>
<th>Conventional Station (CS) Water Level (cm)</th>
<th>Conventional Station (CS) River Flow (m³/s)</th>
<th>Surface Reflectance (SR)</th>
<th>Planet CubeSat Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2018 01 03</td>
<td>179.049</td>
<td>579</td>
<td>3,022.10</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2018 01 13</td>
<td>178.917</td>
<td>561</td>
<td>2,819.00</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2018 01 23</td>
<td>180.478</td>
<td>623</td>
<td>3,551.62</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2018 02 02</td>
<td>180.404</td>
<td>597</td>
<td>3,233.03</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2018 02 12</td>
<td>180.165</td>
<td>608</td>
<td>3,365.81</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2018 02 22</td>
<td>180.615</td>
<td>679</td>
<td>4,294.38</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2018 03 04</td>
<td>181.030</td>
<td>708</td>
<td>4,709.77</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2018 03 13</td>
<td>181.103</td>
<td>710</td>
<td>4,739.20</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2018 03 23</td>
<td>180.913</td>
<td>685</td>
<td>4,378.59</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2018 04 02</td>
<td>180.931</td>
<td>691</td>
<td>4,463.70</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2018 04 12</td>
<td>180.651</td>
<td>667</td>
<td>4,128.66</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>2018 04 22</td>
<td>180.508</td>
<td>666</td>
<td>4,115.01</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>2018 05 02</td>
<td>180.419</td>
<td>627</td>
<td>3,602.11</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2018 05 12</td>
<td>179.721</td>
<td>536</td>
<td>2,549.77</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2018 05 22</td>
<td>179.557</td>
<td>464</td>
<td>1,856.67</td>
<td>TOA image</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>2018 06 01</td>
<td>179.220</td>
<td>454*</td>
<td>1,769.96*</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>2018 06 11</td>
<td>179.248</td>
<td>421</td>
<td>1,499.74</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>2018 06 21</td>
<td>178.612</td>
<td>372</td>
<td>1,143.57</td>
<td>TOA image</td>
<td></td>
</tr>
<tr>
<td>Label ID</td>
<td>Date</td>
<td>JASON Virtual Station (JVS) Altimetry (m)</td>
<td>Conventional Station (CS) Water Level (cm)</td>
<td>Conventional Station (CS) River Flow (m³/s)</td>
<td>Surface Reflectance (SR) Planet CubeSat Images</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>------------------------------------------</td>
<td>------------------------------------------</td>
<td>-------------------------------------------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2018 01 03</td>
<td>179.049</td>
<td>579</td>
<td>3,022.10</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2018 01 13</td>
<td>178.917</td>
<td>561</td>
<td>2,819.00</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2018 01 23</td>
<td>180.478</td>
<td>623</td>
<td>3,551.62</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2018 02 02</td>
<td>180.404</td>
<td>597</td>
<td>3,233.03</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2018 02 12</td>
<td>180.165</td>
<td>608</td>
<td>3,365.81</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2018 02 22</td>
<td>180.615</td>
<td>679</td>
<td>4,294.38</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2018 03 04</td>
<td>181.030</td>
<td>708</td>
<td>4,709.77</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2018 03 13</td>
<td>181.103</td>
<td>710</td>
<td>4,739.20</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2018 03 23</td>
<td>180.913</td>
<td>685</td>
<td>4,378.59</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2018 04 02</td>
<td>180.931</td>
<td>691</td>
<td>4,463.70</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2018 04 12</td>
<td>180.651</td>
<td>667</td>
<td>4,128.66</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>2018 04 22</td>
<td>180.508</td>
<td>666</td>
<td>4,115.01</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>2018 05 02</td>
<td>180.419</td>
<td>627</td>
<td>3,602.11</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2018 05 12</td>
<td>179.721</td>
<td>536</td>
<td>2,549.77</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2018 05 22</td>
<td>179.557</td>
<td>464</td>
<td>1,856.67</td>
<td>TOA image</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>2018 06 01</td>
<td>179.220</td>
<td>454*</td>
<td>1,769.96*</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>2018 06 11</td>
<td>179.248</td>
<td>421</td>
<td>1,499.74</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>2018 06 21</td>
<td>178.612</td>
<td>372</td>
<td>1,143.57</td>
<td>TOA image</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>2018 07 01</td>
<td>179.049</td>
<td>355</td>
<td>1,032.33</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2018 07 10</td>
<td>178.217</td>
<td>345</td>
<td>969.81</td>
<td>incomplete/cloud</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>2018 07 20</td>
<td>178.786</td>
<td>334</td>
<td>903.52</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>2018 07 30</td>
<td>177.912</td>
<td>328</td>
<td>868.46</td>
<td>available</td>
<td></td>
</tr>
</tbody>
</table>

* average between 02/06/2018 and 31/05/2018
Table 2. Flood area calculated inside buffers with NDVI value smaller than 0.15.

<table>
<thead>
<tr>
<th>Label</th>
<th>Selected dates of Surface Reflectance (SR) Planet CubeSat Images</th>
<th>Flood Area inside Buffer 50m (m²)</th>
<th>Flood Area inside Buffer 250m (m²)</th>
<th>Flood Area inside Buffer 500m (m²)</th>
<th>Flood Area inside Buffer 1000m (m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2018 01 23</td>
<td>53,181</td>
<td>273,024</td>
<td>547,929</td>
<td>1,046,745</td>
</tr>
<tr>
<td>7</td>
<td>2018 03 04</td>
<td>54,261</td>
<td>275,040</td>
<td>550,251</td>
<td>1,079,640</td>
</tr>
<tr>
<td>12</td>
<td>2018 04 22</td>
<td>54,846</td>
<td>276,984</td>
<td>557,586</td>
<td>1,097,469</td>
</tr>
<tr>
<td>14</td>
<td>2018 05 12</td>
<td>52,839</td>
<td>267,615</td>
<td>513,891</td>
<td>997,580</td>
</tr>
<tr>
<td>16</td>
<td>2018 06 01</td>
<td>44,964</td>
<td>222,336</td>
<td>431,730</td>
<td>819,477</td>
</tr>
<tr>
<td>17</td>
<td>2018 06 11</td>
<td>40,572</td>
<td>199,512</td>
<td>395,604</td>
<td>736,335</td>
</tr>
<tr>
<td>21</td>
<td>2018 07 20</td>
<td>39,924</td>
<td>177,471</td>
<td>332,415</td>
<td>585,909</td>
</tr>
<tr>
<td>22</td>
<td>2018 07 30</td>
<td>40,707</td>
<td>173,313</td>
<td>320,805</td>
<td>563,895</td>
</tr>
</tbody>
</table>
Table 3. Accuracy (RMSE) and precision (SD) estimators related to river flow ($Q_e$), water level ($L_e$) and altimetry ($A_e$)

<table>
<thead>
<tr>
<th>Regression Curves</th>
<th>Accuracy Estimator (RMSE)</th>
<th>Precision Estimator (SD)</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>River Flow ($Q_e$)</td>
<td>717.59</td>
<td>553.63</td>
<td>m³/s</td>
</tr>
<tr>
<td>Water Level ($L_e$)</td>
<td>59.45</td>
<td>37.16</td>
<td>cm</td>
</tr>
<tr>
<td>Altimetry ($A_e$)</td>
<td>0.33</td>
<td>0.34</td>
<td>m</td>
</tr>
</tbody>
</table>