



**HAL**  
open science

# A Long-Term, High-Resolution Wetland Dataset over the Amazon Basin, Downscaled from a Multiwavelength Retrieval Using SAR Data

F. Aires, F. Papa, Christophe Prigent

► **To cite this version:**

F. Aires, F. Papa, Christophe Prigent. A Long-Term, High-Resolution Wetland Dataset over the Amazon Basin, Downscaled from a Multiwavelength Retrieval Using SAR Data. *Journal of Hydrometeorology*, 2013, 14 (2), pp.594-607. 10.1175/jhm-d-12-093.1 . hal-00990909

**HAL Id: hal-00990909**

**<https://hal.science/hal-00990909v1>**

Submitted on 15 May 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## A Long-Term, High-Resolution Wetland Dataset over the Amazon Basin, Downscaled from a Multiwavelength Retrieval Using SAR Data

FILIPPE AIRES

*Estellus, and Laboratoire de l'Etude du Rayonnement et de la Matière en Astrophysique, CNRS, Observatoire de Paris, Paris, France*

FABRICE PAPA

*Laboratoire d'Etude en Géophysique et Océanographie Spatiales, IRD, Toulouse, France*

CATHERINE PRIGENT

*Laboratoire de l'Etude du Rayonnement et de la Matière en Astrophysique, CNRS, Observatoire de Paris, France*

(Manuscript received 29 June 2012, in final form 31 October 2012)

### ABSTRACT

A climatology of wetlands has been derived at a low spatial resolution ( $0.25^\circ \times 0.25^\circ$  equal-area grid) over a 15-yr period by combining visible and near-infrared satellite observations and passive and active microwaves. The objective of this study is to develop a downscaling technique able to retrieve wetland estimations at a higher spatial resolution (about 500 m). The proposed method uses an image-processing technique applied to synthetic aperture radar (SAR) information about the low and high wetland season. This method is tested over the densely vegetated basin of the Amazon. The downscaling results are satisfactory since they respect the spatial hydrological features of the SAR data and the temporal evolution of the low-resolution wetland estimates. A new long-term and high-resolution wetland dataset has been generated for 1993–2007 for the Amazon basin. This dataset represents a new and unprecedented source of information for climate and land surface modeling of the Amazon and for the definition of future hydrology-oriented satellite missions such as Surface Water and Ocean Topography (SWOT).

### 1. Introduction

Terrestrial surface water is a key parameter of the global water and biogeochemical cycles and plays an important role in the climate system and its variability. It is also crucial for terrestrial life and the human environment as a resource for water consumption, agriculture, and industry. Furthermore, natural disasters of hydrological origin regularly affect human activities and infrastructures with large economic losses during water-related extreme events such as floods or droughts. However, the knowledge of the global distribution and dynamics of this resource remains limited. First, a global survey is extremely challenging because of the large

range of spatial scales of individual water bodies (hundreds of square meters to over  $100\,000\text{ km}^2$ ) with significant temporal variability at daily to seasonal time scales. Second, surface freshwater measurements are still limited mostly to sparse in situ networks of gauges, the number of which has dramatically decreased during the last two decades, especially in remote areas. Finally, despite the advent of satellite remote sensing techniques for hydrologic investigations over the last 20 years, frequent accurate high-resolution characterization of the temporal and spatial variation of surface water is beyond the capabilities of current satellite methods (Alsdorf et al. 2007). Two future missions will address this data gap: Soil Moisture Active Passive (SMAP; [smap.jpl.nasa.gov](http://smap.jpl.nasa.gov)), planned for launch in 2014 by the National Aeronautics and Space Administration (NASA) (Das et al. 2011), and Surface Water and Ocean Topography (SWOT; [swot.jpl.nasa.gov](http://swot.jpl.nasa.gov)), planned for launch in 2019 by the Centre National d'Études Spatiales (CNES)

---

*Corresponding author address:* F. Aires, Estellus/LERMA, Observatoire de Paris, 61 avenue de l'Observatoire, 74014 Paris, France.  
E-mail: [filipe.aires@estellus.fr](mailto:filipe.aires@estellus.fr)

DOI: 10.1175/JHM-D-12-093.1

© 2013 American Meteorological Society

and NASA (Alsdorf et al. 2007; Rodriguez 2012). SMAP is designed for soil moisture measurement, yet its 1–3-km resolution synthetic aperture radar (SAR), 40-km-resolution L-band (1.4 GHz) radiometry, and 3-day revisit period make it an excellent candidate for providing inundation maps more frequently than other SARs and at a higher resolution than other radiometers. SWOT is being designed specifically for surface water area and storage change measurement, and it will have as a standard product a mask of water bodies greater than  $250 \times 250 \text{ m}^2$  and rivers of width greater than 100 m, with average revisit time around 11 days for low latitudes (shorter for higher latitudes). SWOT's principle instrument is a Ka-band radar interferometer (KaRIN) with 50–70-m postings that will provide high-resolution maps as well as water height for rivers, lakes, inundated areas, and wetlands.

A global inundation extent from multisatellite (GIEMS) dataset of monthly inundation and surface water dynamics at about  $25 \times 25 \text{ km}^2$  resolution has been produced by a multisensor analysis covering 1993–2007 (Prigent et al. 2001, 2007; Papa et al. 2010; Prigent et al. 2012). This is the only dataset we are aware of that can provide surface water extent (flooded fraction) with comparable resolution and revisit period over such a long time period. The dataset was produced from analysis of multisatellite passive and active microwaves and visible and near-infrared observations, as described in section 2a. At about 25 km, the GIEMS dataset resolution is comparable to climate model and some global land surface model grids but is clearly not adequate for local applications that require resolution of individual water bodies, wetland extent, and floodplains. Furthermore, subgrid-scale land surface or climate model parameterizations may include other land surface cover types (e.g., vegetation, snow, and permafrost) and topography that might be combined with surface water to form composite types with unique model behaviors. For example, a model may include the fractional coverage of woody and herbaceous vegetation types drawn from a higher-resolution map. When surface water area changes, the type and area of inundated vegetation and the length of time that inundation occurs can affect processes such as biogenic gas flux, land–air exchange of heat and moisture, radiation balance, and vegetation health (Aragão et al. 2007; Lewis et al. 2011; Marengo et al. 2008; Zeng et al. 2008).

General principles of upscaling (i.e., aggregation) or downscaling (i.e., disaggregation) have been developed to link one scale to another, but the definition of such general terms is sometimes confusing. In this study, “downscaling” refers simply to a technique that increases the spatial resolution of the original data of a

particular field, with the same variable in both scales. Most of the downscaling techniques have been developed for meteorological fields (Storch et al. 1993; Maraun et al. 2007; Pryor et al. 2005; Coulibaly et al. 2005; Schmith 2008). This type of approach cannot easily be applied to the problem of the wetland downscaling because high-resolution maps from SAR observations are binary (i.e., wetland presence or not). Physical downscaling techniques, often called “regionalization,” use a dynamical atmospheric model to interpolate meteorological fields: this type of method cannot be applied on surface parameters like wetland maps. Therefore, there is a strong need for innovative downscaling methods based on statistics. Geostatistics could be used on this type of binary data, but a simple image-processing technique is preferred here.

In this paper, an innovative downscaling approach based on image processing applicable to the GIEMS dataset is proposed. The method uses a statistical technique with coincident data from the low-resolution flooded fraction dataset and higher-resolution (500 m) SAR of low- and high-water conditions.

In section 2, the low- and high-resolution inundation datasets used in this study are presented. The downscaling methodology is described in section 3. Results are stated in section 4. Finally, conclusions and perspectives are discussed in section 5.

## 2. Databases

### a. Wetlands from a multiwavelength algorithm

The first component of our analysis is a global, multi-year dataset of inundation/surface water dynamics quantifying the monthly variations of the surface water extent at about  $25 \times 25 \text{ km}^2$  resolution, derived from multiple satellite observations. The complete methodology is described in detail in Prigent et al. (2001, 2007), Papa et al. (2010), and Prigent et al. (2012). It is summarized here. The algorithm uses a complementary suite of satellite observations covering a large wavelength range. Combining different observations helps disentangle the effects of the various surface characteristics contributing to the measured signals (i.e., vegetation, surface roughness, and soil texture). The following observations are available at a global scale:

- Visible and near-infrared reflectances and the derived normalized difference vegetation index (NDVI) from the Advanced Very High Resolution Radiometer (AVHRR).
- Passive microwave emissivities from 19 to 85 GHz. They are estimated from the Special Sensor Microwave Imager (SSM/I) observations by removing the

contributions of the atmosphere (water vapor, clouds, and rain) and the modulation by the surface temperature (Prigent et al. 2006). The technique uses ancillary data from the International Satellite Cloud Climatology Project (ISCCP; Rossow and Schiffer 1999) and the National Center for Environment Prediction (NCEP) reanalysis (Kalnay et al. 1996).

• Active microwave observations (backscattering) at 5.25 GHz from the European Remote Sensing Satellite (ERS) scatterometer.

Observations are averaged over each month and are mapped to an equal-area grid of  $0.25^\circ \times 0.25^\circ$  resolution at the equator (about 25-km interval; each pixel equals  $773 \text{ km}^2$ ) (Prigent et al. 2001). An unsupervised classification of the three sources of satellite data is performed, and the pixels with satellite signatures likely related to inundation are retained. For each inundated pixel, the monthly fractional coverage by open water is obtained using the passive microwave signal and a linear mixture model with end-members calibrated with scatterometer observations to account for the effects of vegetation cover (Prigent et al. 2001). As the microwave measurements are also sensitive to snow cover, snow and ice masks are used to edit the results and avoid any confusion with snow-covered pixels (Armstrong and Brodzik 2005). Because the ERS scatterometer encountered serious technical problems after 2000, the processing scheme had to be adapted to extend the dataset. Among various investigated options, using a mean monthly climatology of ERS and AVHRR NDVI observations in the methodology gives consistent results (Papa et al. 2010).

Fifteen years of global monthly water surface data from 1993–2007 are already available (180 months) (Prigent et al. 2012). This dataset is unique not only in its content (surface extent of open water) but also in terms of its domain (global), high temporal sampling (monthly), and multiyear coverage (almost two decades).

This dataset has been intensively evaluated at the global scale (Prigent et al. 2007; Papa et al. 2010) and for specific regions, including the Amazon basin (Papa et al. 2008b; Frappart et al. 2008, 2011). It is intensively used for climatic and hydrological analyses, such as the evaluation of methane surface emission models (Ringeval et al. 2010) or the validation of river flooding schemes in land surface models (Decharme et al. 2008, 2011).

### b. SAR data

Despite their lack of temporal coverage, many studies (e.g., Lewis 1998) have successfully used synthetic aperture radar sensors to map inundation and wetland

vegetation with high spatial resolution (Rosenqvist et al. 2000). One comprehensive study of flooding in the Amazon basin for low-water (September–October 1995) and high-water (May–June 1996) conditions is that of Hess et al. (2003), where wetland extent was mapped for the central Amazon region using L-band SAR imagery acquired by the Japanese Earth Resources Satellite 1. For the wetland area extending  $18^\circ$  in longitude by  $8^\circ$  in latitude of the central Amazon region, dual-season radar mosaics were used to map inundation extent and vegetation at 92.6-m resolution. As described in detail in Hess et al. (2003), polygon-based segmentation and clustering were used to delineate wetland extent with an accuracy of 95% along with a pixel-based classifier to map wetland vegetation and flooding state based on backscattering coefficients of two-season class combinations, producing the first high-resolution wetlands map for the region. These SAR inundation results are highly dependent on the thresholds that are used by nature in any classification technique [see Hess et al. (2003) for more details]. In this study, we will use the Hess et al. (2003) dataset as a reference for the downscaling.

The original SAR spatial resolution is very high, and there is little justification to downscale 25-km boxes to such small pixels. As a consequence, and for practical reasons, it has been decided to slightly degrade this spatial resolution by grouping  $5 \times 5$  original pixels into bigger 463-m pixels. The new SAR data are kept binary: if more than half of the  $5 \times 5$  original pixels are inundated, the new pixel is considered to be inundated; otherwise, it is considered to be water-free. The maximum area for the new SAR pixels is  $A = 0.463 \times 0.463 \text{ km}^2$ . Figure 1 represents the minimal (top panel) and maximal (middle panel) water extent provided by the SAR. This image is originally binary: it indicates if a pixel is inundated (white) or not (black). However, the figure is represented in this paper with a limited spatial resolution, and intermediate colors (gray pixels) can be observed. In the bottom panel of Fig. 1, the minimal inundated pixels are represented in black, and the red pixels indicate the pixels that are included in the maximal but not included in the minimal estimates. These red pixels will be used by the following downscaling algorithm to “oscillate” between the minimal and maximal SAR inundation. Again, the limited resolution of Fig. 1 introduces some blurring of the original figure.

## 3. Methodology

Most statistical downscaling techniques require coincident low- and high-resolution datasets used to calibrate

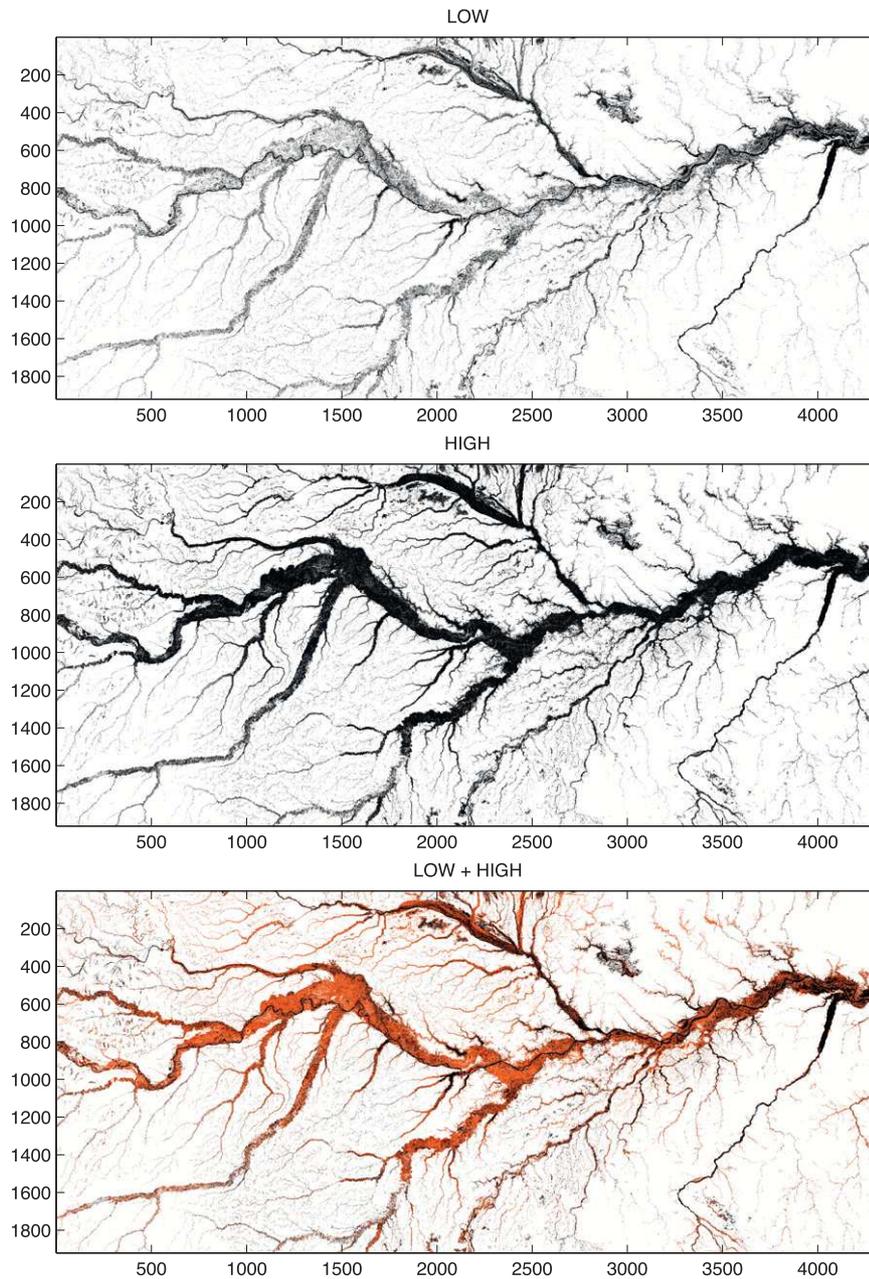


FIG. 1. (top) Low inundation ( $HR_{\min}$ ) and (middle) high inundation ( $HR_{\max}$ ) from the SAR data (at  $\sim 500$ -m resolution). (bottom) Both low and high inundation: the black pixels represent  $HR_{\min}$  and the red pixels represent  $HR_{\max} - HR_{\min}$  (i.e., the pixels that can be inundated by the downscaling process).

a model that will perform the transformation from one space to the other. In this study, the high-resolution SAR estimates will be used (section 2b). Only a few snapshots are available per location, usually the high- and low-wetland extents. Thus, the statistical technique proposed in this paper needs to be adapted to this particular situation with very limited time information in the high-resolution scale.

The methodology first estimates the number of high-resolution pixels that are needed in each low-resolution box (this is the goal of sections 3b and 3c). This information is based on the GIEMS low-resolution dataset. Then, an image-processing approach is used to choose the high-resolution pixels that need to be switched to inundation inside the low-resolution boxes (sections 3d and 3e).

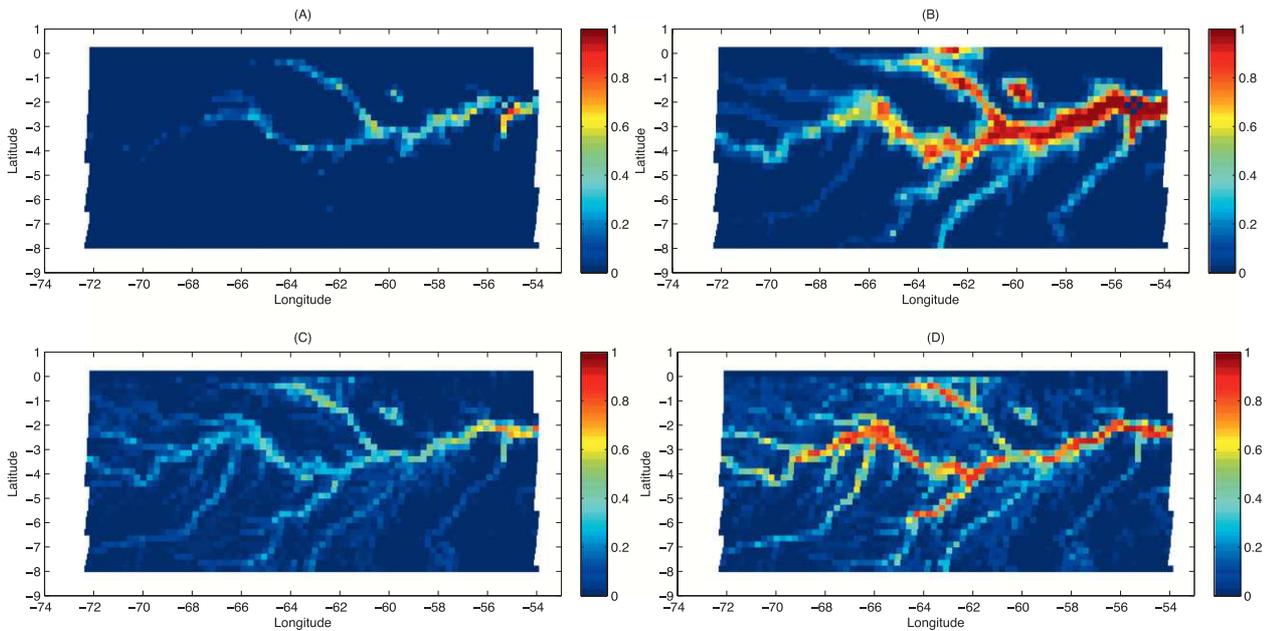


FIG. 2. Inundation estimates at the low resolution of GIEMS in fraction of inundation in each pixel (from 0 to 1). The high resolution of the SAR estimates is upscaled using a simple compositing of every SAR pixel into the GIEMS boxes. (a) Minimum of GIEMS, (b) maximum of GIEMS, (c) minimum of the high-resolution SAR, and (d) maximum of the SAR. The minimum and maximum for GIEMS have been estimated independently for each box so GIEMS maps are not for a particular month.

a. Notation

GIEMS low-resolution data (LR)

Let  $LR(i, j, t)$ <sup>1</sup> be the inundated area for box  $(i, j)$  and for month  $1 \leq t \leq 180$  provided by the GIEMS low-resolution dataset of section 2a.  $LR(i, j, t)$  includes real values oscillating between 0 and 773 km<sup>2</sup> (i.e., the maximum area for a GIEMS low-resolution box). Figures 2a and 2b represent the minimal and maximal values computed on each low-resolution pixel of the GIEMS dataset for the Amazon basin.

SAR high-resolution data (HR)

$HR_{min}(k, l)$  and  $HR_{max}(k, l)$  represent the minimum and maximum matrices of the inundation information provided by the SAR (section 2b). In this paper,  $1 < k < 1920$  and  $1 < l < 4320$ . These two matrices are binary: a value of zero is for the absence of inundation and a value of one is for an inundated pixel. Figures 2c and 2d represent the minimal and maximal matrices  $HR_{min}$  and  $HR_{max}$ .

Downscaled data (D)

The goal of this study is to obtain a downscaled dataset  $D(k, l, t)$  of the same spatial dimension as

$HR_{min}$  and  $HR_{max}$ , with spatial features from the SAR, but with an additional time dimension  $1 \leq t \leq 180$  ( $T = 180$  months) information from the GIEMS dataset.

Linkage (L)

Let  $L(i, j) = [(k, l), \text{such that } (k, l) \in (i, j)]$  be the list of pixels of the HR grid that are included in the low-resolution box  $(i, j)$ . In general, each  $0.25^\circ \times 0.25^\circ$  low-resolution box from GIEMS includes  $60 \times 60 = 3600$  high-resolution pixels from the SAR. With this list of pixels, it is possible to compare the inundated area  $LR(i, j, t)$  of the low-resolution dataset with its HR equivalent:

$$A \sum_{(k,l) \in L(i,j)} HR_{min}(k,l) \quad \text{or} \quad A \sum_{(k,l) \in L(i,j)} HR_{max}(k,l),$$

where  $A = 0.463 \times 0.463 \text{ km}^2$  is the area of the low-resolution SAR pixels. These LR/HR comparisons will be used in the following to define the number of inundated high-resolution pixels that need to be included in each low-resolution box.

Figure 2 compares the minimal and maximal surface water extents from GIEMS (Figs. 2a,b) and SAR (Figs. 2c,d) datasets. The minimal and maximal values are computed independently for each low-resolution box.

<sup>1</sup> In this paper, for clarity, the term “box” will be used for low resolution ( $0.25^\circ \times 0.25^\circ$ ) and “pixel” will be used for high resolution ( $500 \times 500 \text{ m}^2$ ). Indices  $(i, j)$  will be used for the low-resolution boxes and  $(k, l)$  for the high-resolution pixels.

For comparison, the SAR estimates are projected onto the  $0.25^\circ \times 0.25^\circ$  equal-area grid (773 km<sup>2</sup>). The spatial structures of the inundation are very similar. However, with its much better spatial resolution, the SAR can more accurately estimate small areas that are flooded during generally dry conditions or small dry areas during generally flooded conditions, whereas the lower-resolution observations may miss such small fractional coverage (Prigent et al. 2007).

It is important to note that, by downscaling the GIEMS data to the SAR resolution, we not only add finer spatial structures to the GIEMS data but also introduce a bias in the total water extent area since the new downscaled inundated areas will oscillate between  $HR_{\min}$  and  $HR_{\max}$ .

*b. Basin-scale area normalization*

Let us first consider the surface low-resolution (SLR) inundated area of the GIEMS dataset for month  $t$ :

$$SLR(t) = \sum_i \sum_j LR(i, j, t)$$

The goal is to downscale LR into  $D$  so that, for each box  $(i, j)$ , the range of variability of  $D[L(i, j)]$  is equal to the range of variability of  $HR[L(i, j)]$ , with a temporal behavior driven by the GIEMS dataset. For this purpose, we first define the “total” low-resolution ratio of surfaces in the low-resolution grid, for each month  $t$ :

$$\text{Ratio } T(t) = \frac{[SLR(t) - SLR(m)]_{1 \leq m \leq 180}}{[\max SLR(m) - \min SLR(m)]_{1 \leq m \leq 180}} \quad (1)$$

where  $T$  stands for “total” and  $m$  covers the 180 months. Then, the target number of inundated pixels in  $D[L(i, j)]$  for each low-resolution box  $(i, j)$  and month  $t$  is given:

$$\begin{aligned} \text{Nb}(i, j, t) = & \sum_{(k,l) \in L(i,j)} HR_{\min}(k, l) + \text{Ratio } T(t) \\ & \times \sum_{(k,l) \in L(i,j)} HR_{\max}(k, l) \\ & - \sum_{(k,l) \in L(i,j)} HR_{\min}(k, l) \end{aligned} \quad (2)$$

If there is no variability in the LR dataset, then  $\text{Ratio } T(t) = 0$  and  $\text{Nb}(i, j, t)$  is simply equal to  $\sum_{(k,l) \in L(i,j)} HR_{\min}(k, l)$ . This corresponds to a situation where the low resolution from the GIEMS dataset is not providing any information, so the choice is to set the downscaled dataset  $D$  to the minimal value provided by the SAR.

Furthermore, if there is no difference between  $HR_{\min}$  and  $HR_{\max}$  for a particular box  $(i, j)$ , then, again,  $\text{Nb}(i, j, t)$  is simply equal to  $\sum_{(k,l) \in L(i,j)} HR_{\min}(k, l)$ . This corresponds to the case where the SAR is not providing any temporal information.

It should be clear that, in this methodology, the basic hypothesis is that SAR provides the minimal (low-inundation season in 1994) and maximal (high-inundation season in 1996) states of the inundation at the basin level (here) or at the pixel level (next section). Figure 8 will actually show that GIEMS minimal and maximal values are for 1997 and 2002, respectively. This time resolution limitation has a direct impact on the estimates that will give about 70 000 instead of 90 000 km<sup>2</sup> for the low-inundation state and 185 000 instead of 195 000 km<sup>2</sup> for the high-inundation state. These errors directly impact the downscaled estimates following a similar ratio. It should be noted, however, that if more temporal information on the high resolution was available (more than two SAR images, as would be the case using MODIS information on high-resolution wetlands in some less cloudy basins), then the estimation of these low- and high-inundation states would be directly improved.

*c. Low-resolution box-area normalization*

In this second approach, the surface area normalization is performed for each low-resolution box instead of being done, as in the previous section, at the Amazon basin scale. If the GIEMS dataset has noisy inundated-area boxes because of uncertainties (from the retrieval method or from the input datasets), then the geographical patterns of the downscaled dataset  $D$  will inherit these uncertainties. However, if it is believed that spatial information is valuable in the GIEMS dataset, the box-area normalization will allow for keeping this spatial information in the downscaled dataset  $D$ .

The goal is to downscale LR into  $D$  so that, for each box  $(i, j)$ , the range of variability of  $\sum_{(k,l) \in L(i,j)} D(k, l)$  is equal to the range of the high-resolution  $\sum_{(k,l) \in L(i,j)} HR_{\min}(k, l)$  and  $\sum_{(k,l) \in L(i,j)} HR_{\max}(k, l)$ . For this purpose, the ratio of surfaces that is necessary for each month  $t$  at box  $(i, j)$  is defined as

$$\text{Ratio}(i, j, t) = \frac{[LR(i, j, t) - \min LR(i, j, m)]_{1 \leq m \leq 180}}{[\max LR(i, j, m) - \min LR(i, j, m)]_{1 \leq m \leq 180}} \quad (3)$$

We then define a new target number of inundated pixels with Eq. (2) but using normalization coefficient  $\text{Ratio}(i, j, t)$  instead of  $\text{Ratio } T(t)$ .

In Fig. 3, the box-area normalization is illustrated for a given low-resolution box from the GIEMS dataset.

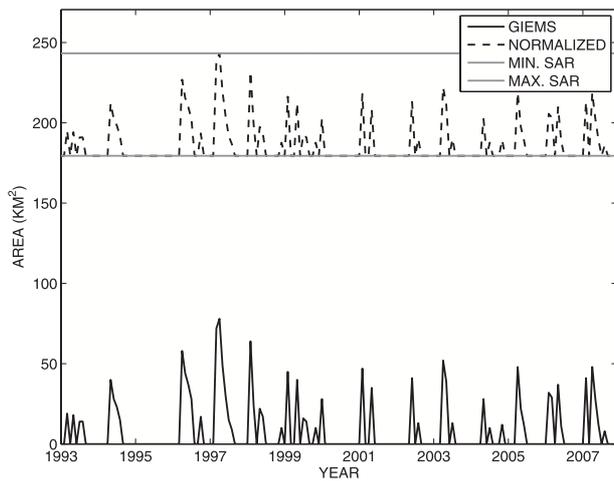


FIG. 3. Normalization at the pixel level: the original GIEMS time series is the dark continuous line. The normalized time series range from the minimum and maximum constant SAR values.

The area of  $HR_{min}$  and  $HR_{max}$  SAR data is indicated with horizontal lines at about 170 and 245  $km^2$ . The GIEMS low-resolution estimate is represented by the black continuous line. It can be seen that it is often equal to zero, illustrating that the GIEMS data have difficulties in retrieving low-inundation values. The downscaling with the box-area normalization introduces a bias and a normalization of the GIEMS estimates so that its minimal and maximal values equal  $HR_{min}$  and  $HR_{max}$ , respectively.

*d. Neighborhood system*

The downscaling procedure proposed in this study is based on an image-processing technique. First, several pixel configurations are listed. The choice of increasing or not the inundation for each one of these pixel configurations is made based on a priori completion probabilities. The list of these pixel configurations and their associated completion probabilities define a neighborhood system.<sup>2</sup>

For the downscaling algorithm to be well adapted to the application, it is important to use a neighborhood system adapted to the Amazon basin. Therefore, the first step is to analyze the inundation neighborhood system in this region by 1) identifying a list of pertinent pixel configurations and 2) estimating their associated a priori completion probabilities.

Sixteen pixel configurations have been selected for this particular application (Fig. 4). In these pixel

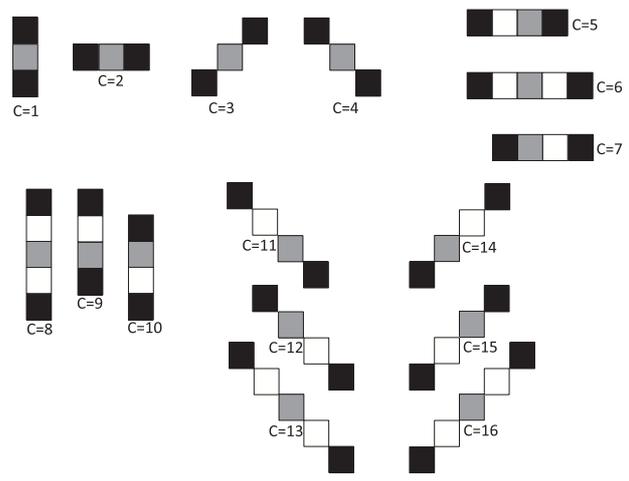


FIG. 4. The 16 pixel configurations used in the neighboring system.

configurations, the black pixels are inundated, the white pixels are water-free, and the question is whether or not to switch the gray pixels from no inundation to inundation. These pixel configurations have been chosen to foster the completion of lines associated to horizontal, vertical, or diagonal hydrological structures.

There are two types of information that this approach wants to exploit. First is the proximity of inundated pixels: a pixel that is surrounded by inundated pixels is more likely to become inundated. Second, some spatial features can increase the probability of being inundated. Since water flows from one pixel to another, lines are important for this application. Other applications would require different neighborhood systems (e.g., geometric squares would be emphasized for urban images).

Once these pixel configurations are identified, the estimation of  $P(p = 1 | C = c)$ , the a priori completion probabilities of central pixel  $p$  being inundated given configuration  $c$ , is done using statistics performed on both  $HR_{min}$  and  $HR_{max}$ . For this purpose, every pixel in  $HR_{min}$  and  $HR_{max}$  is examined to check if it is in one or more of the 16 pixel configurations. The total number of situations is estimated for each one of the 16 pixel configurations. For each one of the 16 configurations, the number of situations where the central gray pixel  $p$  is inundated is also kept. The ratio of latter and former numbers provides an estimation of  $P(p = 1 | C = c) = \text{Sit with } C = c \text{ knowing that } p = 1 / \text{Sit with } C = c$  (where Sit is the number of situations). Table 1 represents the results that were obtained in the Amazon basin. It can be seen that the orientation of the water structures is rather symmetrical since no direction seems to be dominant. It can also be noted that configurations  $c = 13$  or  $16$ , with three noninundated pixels, have lower completion probabilities than configurations  $c = 1, 2, 3$ , or  $4$ . More

<sup>2</sup> This general approach is similar to a Gibbs sampler (Geman and Geman 1984), except that this is a stochastic algorithm and the method proposed here is deterministic.

TABLE 1. A priori completion probabilities  $P(p = 1 | C = c)$  of pixel  $p$  being inundated given a pixel configuration  $c$  around it (for  $c$  defined in Fig. 4). These probabilities have been estimated on HR<sub>min</sub> and HR<sub>max</sub> SAR data.

Box configuration	Completion probability	Box configuration	Completion probability
1	0.842	9	0.832
2	0.845	10	0.832
3	0.832	11	0.821
4	0.832	12	0.824
5	0.835	13	0.683
6	0.822	14	0.822
7	0.835	15	0.822
8	0.818	16	0.683

generally, the completion probability decreases with the number of noninundated pixels inside the hydrological structure, which is a natural behavior.

The 16 pixel configurations and associated a priori completion probabilities constitute the neighborhood system. More sophisticated neighborhood systems could be defined and used, with, for example, more extended pixel configurations. However, it is interesting to show that the proposed downscaling algorithm is satisfactory and that robust results can be obtained with such a simple neighborhood system.

*e. Downscaling*

By using the basin-scale area or the low-resolution box-area normalizations, it is possible to obtain the target number  $Nb(i, j, t)$  of high-resolution pixels in  $D$  that needs to be inundated in each low-resolution box  $(i, j)$ . This number oscillates between

$$\sum_{(k,l) \in L(i,j)} HR_{min}(k, l) \leq Nb(i, j, t) \leq \sum_{(k,l) \in L(i,j)} HR_{max}(k, l)$$

The downscaling algorithm starts with  $D = HR_{min}$ . Then, for each low-resolution box  $(i, j)$ , a list  $L_2(i, j)$  of high-resolution pixels is created. This list is included in  $L(i, j)$ , but only high-resolution pixels present in HR<sub>max</sub> and not present in HR<sub>min</sub> are considered (red pixels in Fig. 1).

For each noninundated pixel  $p$  in  $L_2(i, j)$ , a criterion is computed:

$$Crit(p) = \sum_{c=1}^{16} P(p = 1 | C = c) \mathbb{1}_{\{C=c\}}, \quad (4)$$

where  $\mathbb{1}_{\{B\}}$  is the indicator function (i.e., equal to 1 when  $B$  is true and 0 otherwise). This criterion sums the completion a priori probabilities for all the pixel configurations  $C$  that apply. The higher  $Crit(p)$  is, the stronger the probability for pixel  $p$  to be equal to one (inundated). The iterative procedure first inundates the pixels  $p$  in  $L_2(i, j)$  with higher  $Crit(p)$ , then it inundates the pixels with decreasing  $Crit(p)$  until  $\sum_{(k,l) \in L(i,j)} D(k, l, t)$  equals  $Nb(i, j, t)$ .

Since all of the 16 pixel configurations are used together to decide if the gray pixel is switched or not, it is clear that a gray pixel completely surrounded by inundation will have large probabilities to become inundated. For example, in Fig. 5, the criterion of Eq. (4) is provided for a long hydrological horizontal line interrupted by a dry central pixel (structure A), a single dry pixel surrounded by inundated neighbors (structure B), and a thicker horizontal line (structure C). For structure A,  $Crit(p) = P_2 + P_5 + P_6 + P_7 = 3.337$ , where  $P_i = P(p = 1 | C = i)$  (probability values are given in Table 1). For structure B,  $Crit(p) = P_1 + P_2 + P_3 + P_4 = 3.351$ , and for structure C,  $Crit(p) = P_1 + P_2 + P_3 + P_4 + P_6 = 4.173$ . Therefore, dry pixels in configuration C will be inundated by the downscaling process first, followed the pixels in configuration B and then A.

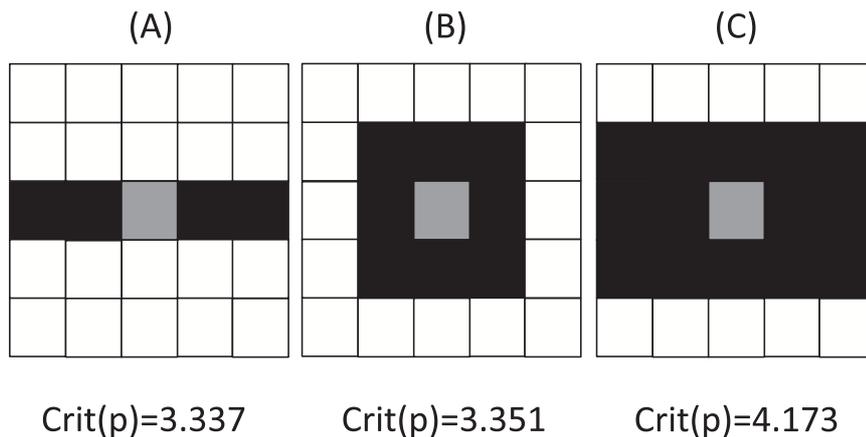


FIG. 5. (a)–(c) Integration of the pixel configurations (Fig. 4) into more complex pixel structures. The quality criterion of Eq. (4) is provided for the three samples.

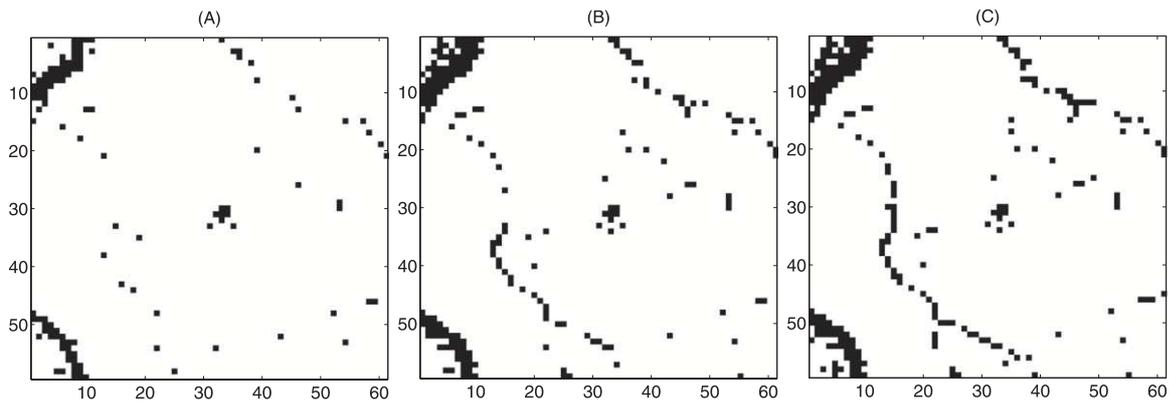


FIG. 6. One sample of a low-resolution box downscaling: (a)  $HR_{\min}$  from the SAR at resolution  $\sim 500$  m, (b) downscaling with a ratio of 0.7, and (c)  $HR_{\max}$  from the SAR.

## 4. Results

### a. Spatial maps

The downscaling process is illustrated in one low-resolution box in Fig. 6. The limited resolution of the figure introduces some blurring of the original binary figure. Figures 6a and 6c correspond to the  $HR_{\min}$  and  $HR_{\max}$  from the SAR data. Figure 6b is the downscaling when a ratio of 0.7 [Eq. (1) for the basin-scale normalization or Eq. (3) for the pixel normalization] is applied to the downscaling process. Figures 6a and 6c correspond to a ratio of 0 and 1. The behavior of the downscaling at the box level is identical for both normalization strategies. It can be seen that the hydrological lines in the box are being completed toward  $HR_{\max}$ . The larger structures at the top left and bottom left of the panels are also increasing.

As explained in the methodology sections, the downscaling process acts on the red pixels of the bottom panel of Fig. 1. This is a strong constraint; it supposes that the inundation occurs in the same locations and that the two available SAR snapshots are describing well the low- and high-inundation states of the Amazon basin. If a satisfactory neighboring system was available, or, even better, if some other a priori information such as the topography or the routing information was available (Galantowicz 2002), then this constraint could be attenuated or even suppressed. The  $HR_{\max}$  information provides the a priori information that is required for the downscaling. It should be noted that, if more SAR data were available, this a priori information would be more precise. For example, the minimal and maximal values for each box would be more reliable with more than two SAR snapshots.

The downscaling of the Amazon basin is represented in Fig. 7 for two months (January and July 2007). The blue pixels are inundated only for the pixel normalization, the

red pixels are inundated only for the basin normalization, and the black pixels are inundated for both of them. Again, the limited resolution of the figure introduces some blurring of the original binary figure. It is difficult to measure the differences between the two normalization techniques at this scale, but the pixel normalization is definitely using different ratios [Eq. (3)] for each box. Contrary to the basin-scale normalization, for the box-area normalization the inundation increase is not uniform over the whole Amazon basin, and some river tributaries are more inundated than others. The blue pixels from the pixel normalization are located more in the highly inundated boxes where it is known that GIEMS has more information (underestimation of low-fraction inundated boxes by GIEMS). In July, the blue pixels seem to be more localized in the right portion of the map. This means that the geographic repartition of the inundation in the GIEMS dataset is in good agreement with the pixel normalization.

### b. Time series

Figure 8 compares the total inundation area of the Amazon basin for the three different estimates: the low-resolution estimate from the GIEMS dataset (gray continuous line), the downscaled estimate when using the basin-scale normalization (dashed black line), and the downscaled estimate when using the pixel normalization (black continuous line). By definition, the two downscaled estimates oscillate between the SAR-derived total inundation of  $HR_{\min}$  and  $HR_{\max}$  indicated by the two horizontal lines in Fig. 8.

These three time series have a very similar time evolution with good agreement in the seasonal and interannual variations. The seasonality is very coherent for both downscaled estimates. The correlation between the GIEMS estimates is equal to 0.999 with the

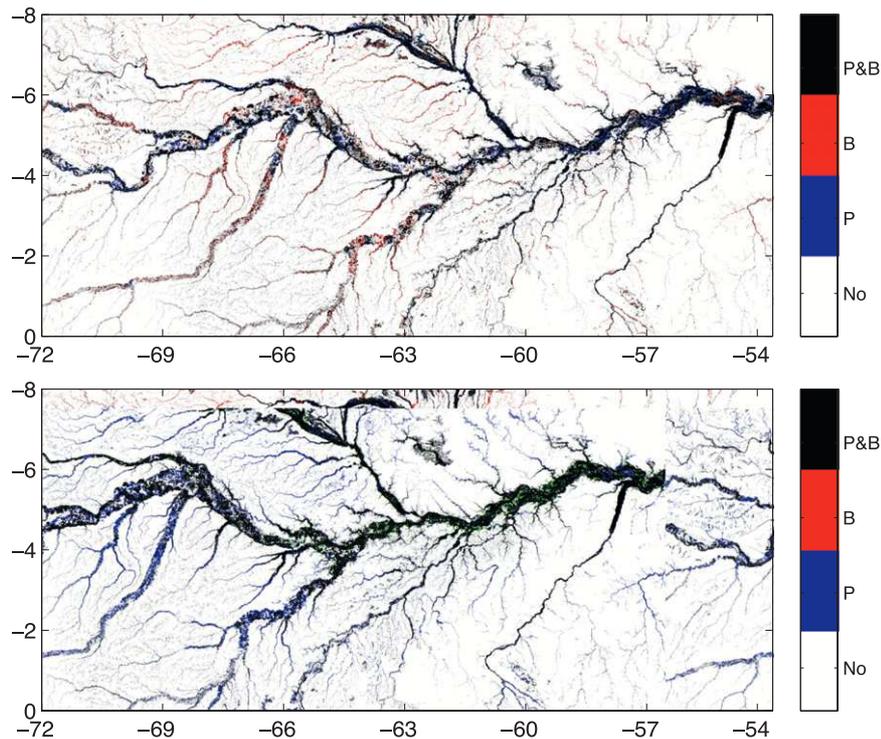


FIG. 7. Downscaling of the Amazon basin for the year 2007: (top) January and (bottom) July. Blue pixels are only inundated for the pixel normalization, red pixels are only inundated for the total normalization, and black pixels are inundated for both normalizations.

basin normalization downscaled estimate and 0.989 with the pixel normalization estimate (both highly significant for a sample of 180 months). Furthermore, the two downscaled estimates are correlated at the 0.988 level. This shows that the downscaling process does not modify the time evolution of the low-resolution estimate. This is very important, as it confirms that the downscaling technique acts with the appropriate time behavior provided by the low resolution.

However, one can notice differences between the basin and pixel normalization estimates. The basin normalization provides a downscaled inundation area equal to the estimates of  $HR_{\min}$  and  $HR_{\max}$  when it reaches its minimum (in 1997) and its maximum (in 2002). The downscaling procedure reproduces well the behavior that was asked for in its design (section 3b). For the pixel normalization, the SAR minimal and maximal values are never reached. This would happen only if the minimal or maximal values of the low-resolution GIEMS estimates were reached for the same month for each box, but this is not the case.

### c. Evaluation using independent observations

To evaluate the seasonal and interannual variations of the downscaled surface water extent, we compare here

the pixel normalization results with related hydrological variables, namely, in situ river discharges along reaches of the Amazon River. For the period 1993–2007, in situ discharge observations at four locations (see details in Table 2) and overlapping in time with the wetland extent estimates are available from the Environmental Research Observatory (ORE) Geodynamical, Hydrological, and Biogeochemical Control of Erosion/Alteration and Material Transport in the Amazon Basin (HYBAM) project (<http://www.ore-hybam.org>). The GIEMS dataset has already been similarly evaluated at several locations worldwide (Papa et al. 2006, 2008b), including the Amazon watershed (Papa et al. 2010). Note that these in situ data are entirely independent from our estimates. Figure 9 compares the downscaled surface water extent (black curve) and the in situ river discharge (red) for each location. The total wetland extent is estimated on a  $0.25^\circ \times 0.25^\circ$  region centered on the in situ station locations.

For all of the locations, both variables show good agreement in the strong seasonal cycle (Fig. 9, left panels). The right panels of Fig. 9 show good correspondence in interannual variations: large anomalous events are generally well reproduced in both datasets. The correlation coefficients between the time series are summarized in Table 2. Over 15 years, the maximum correlations

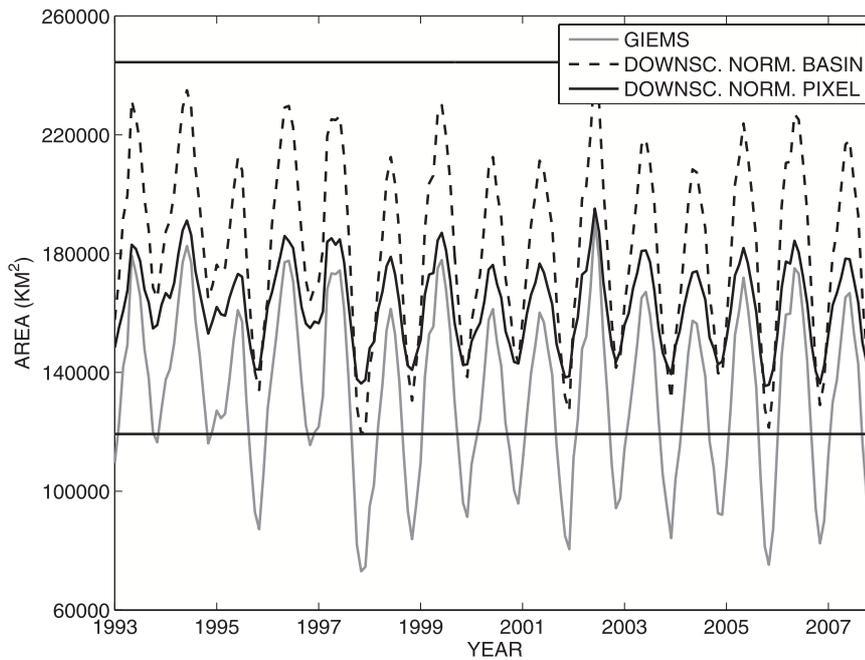


FIG. 8. Time series of the total inundation area over the Amazon basin for the GIEMS estimate (gray line), the downscaled with the basin normalization (black dashed line), and the downscaled with the pixel normalization (black continuous line). The total areas provided by  $HR_{min}$  and  $HR_{max}$  are represented by the continuous horizontal lines.

between the two monthly variables are relatively high (180 months is used to calculate the linear correlation coefficient, giving for each location a high significance with  $p$  value  $< 0.01$ ). For instance, for Labrea and Manacapuru, the discharge and the surface water extent agree well in the seasonal cycle with a maximum lagged correlation of 0.82 and 0.81 (with the extent preceding the discharge by one month for Manacapuru). However, the lagged correlations between the deseasonalized anomalies are lower, with values at 0.35 and 0.59, but are still significant ( $p < 0.01$ ). For Fazenda, the in situ discharge and the surface water extent agree well in the seasonal cycle ( $R = 0.68$ ) but show some limited agreement in the interannual variability with a maximum correlation of only 0.14. Indeed, even if wetland extent and river discharge are closely related, the inundation characteristics depend also on other factors, including the distribution and intensity of local or upstream precipitation, as well as on the land surface properties such as local topography.

### 5. Conclusions and perspectives

In this paper, a long-term (1993–2007) monthly dataset of a high-resolution (about 500 m) inundation dataset has been built for the Amazon basin. It results from the downscaling of a multi-instrument estimation of

wetland extent (Prigent et al. 2012) using low- and high-inundation information from SAR observations. Evaluation of the downscaled products was conducted first by comparison with the original datasets: the spatial features are very similar to the patterns in the SAR data, and the time evolution of the downscaled dataset is very close to the original low-resolution dataset. As a consequence, the downscaling procedure behaves exactly as expected. Furthermore, the downscaled inundation is highly correlated with the river discharge, that is, the information is totally independent from the satellite observations used for the downscaling.

TABLE 2. Location of stations and maximum time-lagged correlations during 1993–2007 between the downscaled wetland extent and corresponding in situ river discharge. Correlation 1 is the correlation between the raw time series (Fig. 8, left), and correlation 2 is for the interannual (deseasonalized normalized anomalies) time series (Fig. 8, right). Time lag in months to obtain the maximum correlation is in parentheses. For all locations, the total downscaled surface water extent is estimated over a region of  $0.25^\circ \times 0.25^\circ$  centered in the in situ discharge station locations.

Station	Latitude	Longitude	Correlation 1	Correlation 2
Fazenda	4.37°S	59.62°W	0.68 (−1)	0.14 (0)
Labrea	7.25°S	64.81°W	0.82 (0)	0.35 (0)
Manacapuru	3.31°S	60.63°W	0.81 (−1)	0.59 (−1)
Serinha	0.48°S	64.28°W	0.69 (0)	0.35 (0)

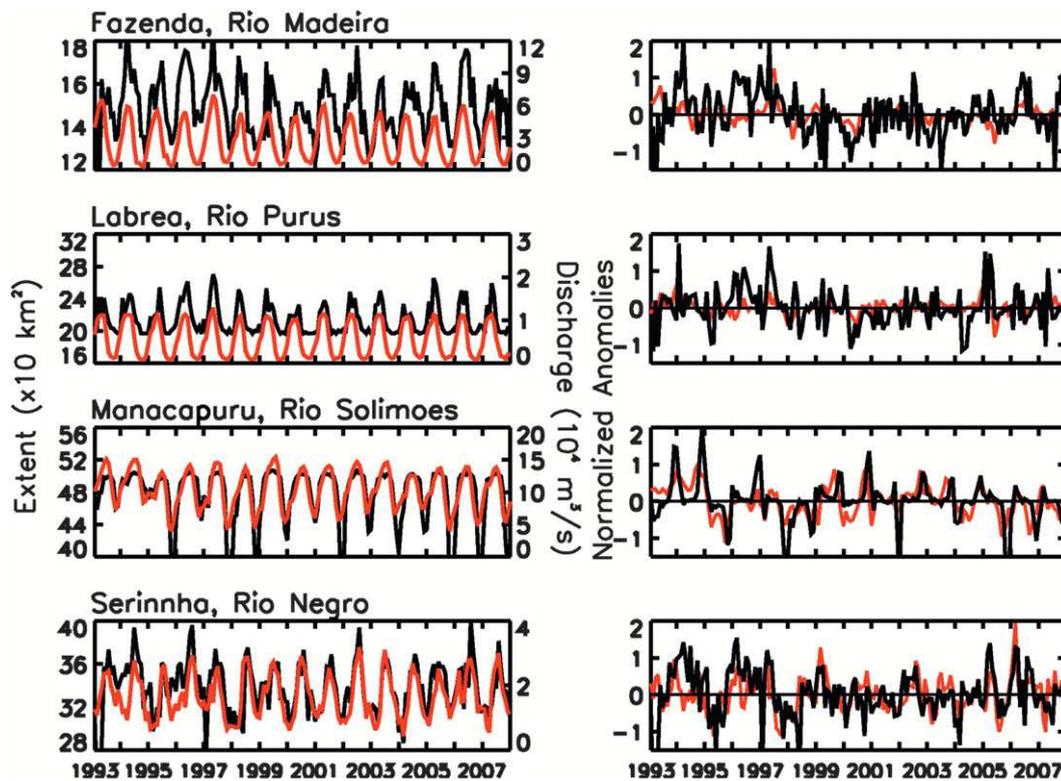


FIG. 9. Variability of the wetland extent between 1993 and 2007 and corresponding in situ river discharge variability. The total downscaled surface water extent is estimated over a region of  $0.25^\circ \times 0.25^\circ$  centered in the (top to bottom) in situ discharge station locations. (left) Monthly mean surface water extent 1993–2007 (black) and comparison with in situ river discharge (red). (right) Corresponding deseasonalized and normalized anomalies obtained by subtracting the 15-yr mean monthly value from individual months.

The perspectives of this work are numerous. First, as commented earlier, a more sophisticated neighboring system could be used. For this purpose, a thorough study would need to be done on  $HR_{\min}$  and  $HR_{\max}$  to determine the essential configurations characterizing the hydrological structures under consideration. Furthermore, if more information on the Amazon basin was available (more SAR data), then this could significantly improve the quality of the results. It would also be possible to introduce topography information in the downscaling. For example, in Galantowicz (2002), routing information is used. This type of information could be used in our neighboring system. Both types of information (image-processing constraints and routing information) would be used simultaneously, and the downscaling would benefit from the synergy of these two complementary approaches.

The downscaling methodology developed in this study can also be used to bridge the gap between climate model or global land surface model (LSM) scales (e.g.,  $0.25^\circ$ – $1^\circ$ ) and local hydrological models and data. For instance, Decharme et al. (2008) coupled an LSM—in

which a  $1^\circ$  flooded fraction is defined—with a river routing model that introduces a flood reservoir, and they demonstrated effects of standing water knowledge on land surface water and energy budgets. Downscaling methods similar to those developed here could be applied to low-resolution model outputs to obtain inundation patterns compatible with model states.

Several applications can be suggested for the downscaled dataset:

- 1) The downscaled dataset can be used in studies of surface water variation with spatial resolution (about 500 m) compatible with local hydrologic applications such as flood event and seasonal inundation mapping.
- 2) The downscaled inundation dataset can be readily aggregated to climate model or LSM scales for comparisons of model outputs to observations at equivalent scales.
- 3) Two efforts are now underway to combine the low-resolution multisensor dataset with altimeter data to assess terrestrial water storage change (Papa et al.

2008b; Frappart et al. 2012) and river discharge (Papa et al. 2008a). These methodologies could also be applied to the downscaled inundation products to study more localized conditions.

- 4) Land surface biogenic gas (e.g., methane) emission modeling depends on inundation duration, vegetation type, and other spatially distributed factors (Altor and Mitsch 2006; Bousquet et al. 2006). Downscaled inundation maps can be used in flux models to determine the time and place of inundation and distribute surface type conditions accordingly within a lower-resolution model grid.

Furthermore, this new downscaled high-spatial-resolution information on the wetland dynamics will be compatible with the SWOT spatial resolution. As a consequence, our high-resolution climatology will allow the analysis of the SWOT mission and the optimization of its instrumental characteristics. The new high-resolution wetland climatology will also help define the calibration/validation sites and will serve as a benchmark for the future SMAP and SWOT missions. The new downscaled dataset ties the future SMAP and SWOT missions to a more than two-decade record of similar inundation extent data, allowing the analysis of interannual variability and change.

*Acknowledgments.* We thank the CNES (Centre National d'Études Spatiales) and, in particular, Selma Cherchali for funding this project named "Préparation à la mission SWOT: Désagrégation des surface inondées" and Aurélie Le Samedy, who helped us gather the datasets for the study.

#### REFERENCES

- Aldorf, D., E. Rodríguez, and D. Lettenmaier, 2007: Measuring surface water from space. *Rev. Geophys.*, **45**, RG2002, doi:10.1029/2006RG000197.
- Altor, A., and W. Mitsch, 2006: Methane flux from created riparian marshes: Relationship to intermittent versus continuous inundation and emergent macrophytes. *Ecol. Eng.*, **28**, 224–234.
- Aragão, L. E. O. C., Y. Malhi, R. M. Roman-Cuesta, S. Saatchi, L. O. Anderson, and Y. E. Shimabukuro, 2007: Spatial patterns and fire response of recent Amazonian droughts. *Geophys. Res. Lett.*, **34**, L07701, doi:10.1029/2006GL028946.
- Armstrong, R. L., and M. J. Brodzik, 2005: Northern Hemisphere EASE-grid weekly snow cover and sea ice extent version 3. National Snow and Ice Data Center, Boulder, CO, digital media. [Available online at <http://nsidc.org/data/nsidc-0046.html>.]
- Bousquet, P., and Coauthors, 2006: Contribution of anthropogenic and natural sources to atmospheric methane variability. *Nature*, **443**, 439–443, doi:10.1038/nature05132.
- Coulibaly, P., Y. B. Dibike, and F. Anctil, 2005: Downscaling precipitation and temperature with temporal neural networks. *J. Hydrometeorol.*, **6**, 483–496.
- Das, N., D. Entekhabi, and E. Njoku, 2011: An algorithm for merging SMAP radiometer and radar data for high-resolution soil moisture retrieval. *IEEE Trans. Geosci. Remote Sens.*, **49**, 1504–1512.
- Decharme, B., H. Douville, C. Prigent, F. Papa, and F. Aires, 2008: A new river flooding scheme for global climate applications: Off-line evaluation over South America. *J. Geophys. Res.*, **113**, D11110, doi:10.1029/2007JD009376.
- , R. Alkama, F. Papa, S. Faroux, H. Douville, and C. Prigent, 2011: Global off-line evaluation of the ISBA-TRIP flood model. *Climate Dyn.*, **38**, 1389–1412, doi:10.1007/s00382-011-1054-9.
- Frappart, F., F. Papa, F. Famiglietti, C. Prigent, W. Rossow, and F. Seyler, 2008: Interannual variations of river water storage from a multiple satellite approach: A case study for the Rio Negro River basin. *J. Geophys. Res.*, **113**, D21104, doi:10.1029/2007JD009438.
- , and Coauthors, 2011: Satellite-based estimates of groundwater storage variations in large drainage basins with extensive floodplains. *Remote Sens. Environ.*, **115**, 1588–1594.
- , F. Papa, J. S. da Silva, G. Ramillien, C. Prigent, F. Seyler, and S. Calmant, 2012: Surface freshwater storage and dynamics in the Amazon basin during the 2005 exceptional drought. *Environ. Res. Lett.*, **7**, 044010, doi:10.1088/1748-9326/7/4/044010.
- Galantowicz, J., 2002: High-resolution flood mapping from low-resolution passive microwave satellite observations. *Proc. 2002 IEEE Int. Geoscience and Remote Sensing Symp.*, Toronto, Ontario, Canada, IEEE, 1499–1502.
- Geman, S., and D. Geman, 1984: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, **6**, 721–741, doi:10.1109/TPAMI.1984.4767596.
- Hess, L., J. Melack, E. Novob, C. Barbosac, and M. Gastil, 2003: Dual season mapping of wetland inundation and vegetation for the central Amazon basin. *Remote Sens. Environ.*, **87**, 404–428.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Lewis, A., 1998: Geomorphic and hydrologic applications of active microwave remote sensing. *Principles and Applications of Imaging Radar*, F. M. Jenderson and A. J. Lexis, Eds., *Manual of Remote Sensing*, Vol. 2, John Wiley, 567–629.
- Lewis, S., P. Brando, O. P. M. van der Heijden, and D. Nepstad, 2011: The 2010 Amazon drought. *Science*, **331**, 554.
- Maraun, D., and Coauthors, 2007: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.*, **48**, RG3003, doi:10.1029/2009RG000314.
- Marengo, J. A., and Coauthors, 2008: The drought of Amazonia in 2005. *J. Climate*, **21**, 495–516.
- Papa, F., C. Prigent, W. B. Rossow, B. Legressy, and F. Remy, 2006: Inundated wetland dynamics from remote sensing: The use of Topex–Poseidon dual-frequency radar altimeter observations. *Int. J. Remote Sens.*, **27**, 4847–4866.
- , A. Güntner, F. Frappart, C. Prigent, and W. B. Rossow, 2008a: Variations of surface water extent and water storage in large river basins: A comparison of different global data sources. *Geophys. Res. Lett.*, **35**, L11401, doi:10.1029/2008GL033857.
- , C. Prigent, and W. Rossow, 2008b: Monitoring flood and discharge variations in the large Siberian rivers from a multi-satellite technique. *Surv. Geophys.*, **29**, 297–317, doi:10.1007/s10712-008-9036-0.
- , —, F. Aires, C. Jimenez, W. B. Rossow, and E. Matthews, 2010: Interannual variability of surface water extent at global

- scale, 1993–2004. *J. Geophys. Res.*, **115**, D12111, doi:10.1029/2009JD012674.
- Prigent, C., E. Matthews, F. Aires, and W. B. Rossow, 2001: Remote sensing of global wetland dynamics with multiple satellite data sets. *Geophys. Res. Lett.*, **28**, 4631–4634.
- , F. Aires, and W. B. Rossow, 2006: Land surface microwave emissivities over the globe for a decade. *Bull. Amer. Meteor. Soc.*, **87**, 1573–1584.
- , F. Papa, F. Aires, W. B. Rossow, and E. Matthews, 2007: Global inundation dynamics inferred from multiple satellite observations, 1993–2000. *J. Geophys. Res.*, **112**, D12107, doi:10.1029/2006JD007847.
- , —, C. Jimenez, W. Rossow, and E. Matthews, 2012: Changes in land surface water dynamics since the 1990s and relation to population pressure. *Geophys. Res. Lett.*, **39**, L08403, doi:10.1029/2012GL051276.
- Pryor, S. C., J. T. Schoof, and R. J. Barthelmie, 2005: Empirical downscaling of wind speed probability distributions. *J. Geophys. Res.*, **110**, D19109, doi:10.1029/2005JD005899.
- Ringeval, B., N. de Noblet-Ducoudre, P. Ciais, P. Bousquet, C. Prigent, F. Papa, and W. B. Rossow, 2010: An attempt to quantify the impact of changes in wetland extent on methane emissions at the seasonal and interannual time scales. *Global Biogeochem. Cycles*, **24**, GB2003, doi:10.1029/2008GB003354.
- Rodriguez, E., 2012: Surface Water and Ocean Topography Mission (SWOT). Tech. Rep., Jet Propulsion Laboratory, Pasadena, CA. [Available online at [swot.jpl.nasa.gov/files/SWOT\\_science\\_reqs\\_release2\\_v1.14.pdf](http://swot.jpl.nasa.gov/files/SWOT_science_reqs_release2_v1.14.pdf).]
- Rosenqvist, A., M. Shimada, B. Chapman, and A. Freeman, 2000: The global rain forest mapping project—A review. *Int. J. Remote Sens.*, **21**, 1375–1387.
- Rossow, W. B., and R. Schiffer, 1999: Advances in understanding clouds from ISCCP. *Bull. Amer. Meteor. Soc.*, **80**, 2261–2287.
- Schmith, T., 2008: Stationarity of regression relationships: Application to empirical downscaling. *J. Climate*, **21**, 4529–4537.
- Storch, H. V., E. Zorita, and U. Cubasch, 1993: Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime. *J. Climate*, **6**, 1161–1171.
- Zeng, N., J. Yoon, J. Marengo, A. Subramaniam, C. Nobre, A. Mariotti, and J. Neelin, 2008: Causes and impacts of the 2005 Amazon drought. *Environ. Res. Lett.*, **3**, 014002, doi:10.1088/1748-9326/3/1/014002.