



**HAL**  
open science

# Characterization and Space-Time Downscaling of the Inundation Extent over the Inner Niger Delta Using GIEMS and MODIS Data

F. Aires, F. Papa, Catherine Prigent, J. F. Cretaux, M. Berge-Nguyen

► **To cite this version:**

F. Aires, F. Papa, Catherine Prigent, J. F. Cretaux, M. Berge-Nguyen. Characterization and Space-Time Downscaling of the Inundation Extent over the Inner Niger Delta Using GIEMS and MODIS Data. *Journal of Hydrometeorology*, 2014, 15 (1), pp.171-192. 10.1175/jhm-d-13-032.1 . hal-00990910

**HAL Id: hal-00990910**

**<https://hal.science/hal-00990910>**

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.

## Characterization and Space–Time Downscaling of the Inundation Extent over the Inner Niger Delta Using GIEMS and MODIS Data

FILIPPE AIRES

*Estellus, France, and Laboratoire d'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, and Indo-French Cell for Water Sciences, IRD-IISc Joint International Laboratory, Indian Institute of Science, Bangalore, India*

CATHERINE PRIGENT

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

JEAN-FRANÇOIS CRÉTAUX AND MURIEL BERGE-NGUYEN

*Laboratoire d'Etude en Géophysique et Océanographie Spatiales, CNES/CNRS/IRD/UPS, Toulouse, France*

(Manuscript received 4 February 2013, in final form 30 August 2013)

### ABSTRACT

The objective in this work is to develop downscaling methodologies to obtain a long time record of inundation extent at high spatial resolution based on the existing low spatial resolution results of the Global Inundation Extent from Multi-Satellites (GIEMS) dataset. In semiarid regions, high-spatial-resolution a priori information can be provided by visible and infrared observations from the Moderate Resolution Imaging Spectroradiometer (MODIS). The study concentrates on the Inner Niger Delta where MODIS-derived inundation extent has been estimated at a 500-m resolution. The space–time variability is first analyzed using a principal component analysis (PCA). This is particularly effective to understand the inundation variability, interpolate in time, or fill in missing values. Two innovative methods are developed (linear regression and matrix inversion) both based on the PCA representation. These GIEMS downscaling techniques have been calibrated using the 500-m MODIS data. The downscaled fields show the expected space–time behaviors from MODIS. A 20-yr dataset of the inundation extent at 500 m is derived from this analysis for the Inner Niger Delta. The methods are very general and may be applied to many basins and to other variables than inundation, provided enough a priori high-spatial-resolution information is available. The derived high-spatial-resolution dataset will be used in the framework of the Surface Water Ocean Topography (SWOT) mission to develop and test the instrument simulator as well as to select the calibration validation sites (with high space–time inundation variability). In addition, once SWOT observations are available, the downscaled methodology will be calibrated on them in order to downscale the GIEMS datasets and to extend the SWOT benefits back in time to 1993.

### 1. Introduction

The Global Inundation Extent from Multi-Satellites (GIEMS) provides multiyear monthly variations of the

global surface water extent at about  $25\text{ km} \times 25\text{ km}$  resolution from 1993 to 2007. It is derived from multiple satellite observations (Prigent et al. 2012; see section 2b). Its spatial resolution is compatible with climate model outputs and with global land surface model grids but is clearly not adequate for local applications that require the characterization of small individual water bodies. There is today a strong demand for high-resolution inundation extent datasets for a large variety of applications

---

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

such as water management, regional hydrological modeling, or for the analysis of mosquito-related diseases. If the inundation extent is combined with altimetry measurements to obtain water volumes (Frappart et al. 2008, 2011, 2012), a better-resolved inundation extent also improves the accuracy of the estimates.

The Surface Water Ocean Topography (SWOT) mission ([swot.jpl.nasa.gov](http://swot.jpl.nasa.gov)), planned for launch in 2019, is designed for the measurements of surface water area and storage change with a high spatial resolution (Alsdorf et al. 2007; Rodriguez 2012). It will provide a mask of water bodies larger than  $250\text{ m} \times 250\text{ m}$  and rivers with width greater than 500 m, with an average revisit time around 11 days for low latitudes (shorter for higher latitudes). SWOT's main instrument is a Ka-band interferometer (KaRIN) with 50–70-m postings to derive high-resolution surface water maps as well as water height for rivers, lakes, inundated areas, and wetlands. Before the launch of this mission and in the framework of its preparation, long-term datasets of high-spatial-resolution surface water extent are in demand.

Would it be possible to develop downscaling methodology to derive high-resolution surface water extent from the existing GIEMS low-resolution dataset? Since GIEMS has a global coverage, the ideal situation would be to develop a downscaling technique general enough to work in all environments. However, each hydrological basin has its own characteristics, such as its topography or space–time variability. The downscaling algorithm needs to take into account these specificities and the availability of the a priori high-spatial-resolution information for a particular basin.

Two main approaches have been used so far for the downscaling of inundation extent. The first approach used digital elevation model (DEM) information to distribute the inundation extent at a finer scale following topographic information. Galantowicz (2002) and Li et al. (2013) use such an approach to downscale inundation over the Mississippi basin, and Brakenridge et al. (2012) and Wang et al. (2002) focus on the coastal areas. The second general approach utilizes statistical or image-processing methods. In Fluet-Chouinard and Lehner (2011), a downscaling algorithm derives first an inundation probability map by using a decision tree classifier trained on regional remote sensing wetland maps. Then, a seeded region growing segmentation process is used to redistribute the inundated area at the finer resolution. Recently, in Aires et al. (2013), a downscaling method is developed for the Amazon basin. It is based on an image processing technique that uses high-spatial-resolution information from a synthetic aperture radar (SAR) for low and high states of the inundation. The a priori high-resolution information is only available for these two

extreme states of the inundation, as the SAR observations do not provide a higher temporal sampling. Visible and/or infrared satellite measurements cannot help in these regions with dense vegetation and persistent cloud coverage. In other regions, vegetation density and cloud cover are less of a problem and other types of satellite observations with better temporal sampling can be exploited as a priori high-resolution information for the downscaling. For example, the Moderate Resolution Imaging Spectroradiometer (MODIS) provides surface reflectances sensitive to the inundation with a spatial resolution of  $\sim 250\text{-m}$  resolution in semiarid regions under cloud-free conditions.

In this study, we will investigate the downscaling of the GIEMS inundation extent [spatial resolution of about  $25\text{ km} \times 25\text{ km}$  from visible (VIS), infrared (IR), and microwave (MW) observations] to a spatial resolution of about 500 m, with the assistance of MODIS observations ( $\sim 250\text{-m}$  resolution) in a semiarid region. The objective is to develop a methodology that can benefit from the availability of enough temporal sampling at high spatial resolution. In Crétaux et al. (2011), MODIS observations are used to retrieve inundation extent over the Inner Niger Delta, located in central Mali in the Sahelian zone. The study provides an exceptional time and space sampling of the inundation extent in this region over 12 yr from 2000 to 2012, with an 8-day sampling at about 500-m resolution. This makes it possible to develop the method more easily and to test its sensitivity to the amount of available a priori high-spatial-resolution information. It could be then applied over other regions where less high-resolution information is accessible. Nevertheless, note that over the Inner Niger Delta, the existence of GIEMS since 1993 will make it possible to extend the high-resolution time series backward as far as 1993.

In section 2, the databases (GIEMS and MODIS water surface extents) are presented over the Inner Niger Delta, and these two estimates are compared. Section 3 uses a principal component analysis (PCA) to characterize the space–time variability of the inundation extent over the Inner Niger Delta at high resolution. It also considers the time interpolation and the filling of missing values in the inundation fields. Section 4 proposes two downscaling techniques, both based on the previous PCA analyses. The first one is derived from a linear regression, and the second one from a linear algebra solution. Conclusions and perspectives are drawn in section 5.

## 2. Databases over the Inner Niger Delta

The Inner Niger Delta is located in central Mali in a semiarid region south of the Sahara Desert. It consists of channels, swamps, and lakes. Water extent varies

seasonally from about 4000 to 20 000 km<sup>2</sup> with large interannual changes. The hydrology of this region has been extensively studied (e.g., Pedinotti et al. 2012).

#### a. GIEMS

The first component of our analysis is a global, monthly, multiyear dataset of surface water dynamics quantifying the variations of the surface water extent at about 25 km × 25 km resolution from multiple satellite observations. The complete methodology is described in detail in Prigent et al. (2001, 2007) and Papa et al. (2010) and is summarized here. The algorithm uses complementary satellite observations covering a large wavelength range. Combining different observations helps separate the effects of the different 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, 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); and
- active microwave observations (backscattering) at 5.25 GHz from the European Remote Sensing (ERS) Satellite scatterometer.

Observations are averaged over each month and mapped to an equal-area grid of 0.25° × 0.25° resolution at the equator (about 25-km interval; each pixel equals 773 km<sup>2</sup>). An unsupervised classification of the three sources of satellite data is performed, and the pixels with satellite signatures likely related to inundation are kept. 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 for the effects of vegetation cover (Prigent et al. 2001). As the microwave measurements are also sensitive to snow, snow and ice masks are used to avoid confusion with snow-covered pixels (Armstrong and Brodzik 2005). The ERS scatterometer encountered serious technical problems after 2000 and the processing had to be adapted. Among various options investigated, using a mean monthly climatology of ERS and NDVI–AVHRR observations in the methodology gives consistent results (Papa et al. 2010).

Fifteen years of global monthly water surfaces extent from 1993 to 2007 are already available (180 months).

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 multi-year coverage (15 yr).

#### b. MODIS inundation estimates

Estimation of surface water extent using visible or infrared measurements provides relatively high spatial resolution but shows limitations to detect surface water beneath clouds or dense vegetation. Yet, the potential of MODIS to monitor flood temporal changes has been demonstrated over specific regions such as the Mekong River (Sakamoto et al. 2007) or arid regions such as the Inner Niger Delta (Crétau et al. 2011), the Aral Sea (Crétau et al. 2009), or the Andean Altiplano (del Rio et al. 2012).

A method for wetland mapping and flood event monitoring was developed using surface reflectance measurements from the MODIS *Terra* instrument in order to detect open water and the spread of aquatic vegetation on a weekly basis with an about 250-m spatial resolution (Berge-Nguyen et al. 2008). The methodology uses a combination of surface reflectance measured by different spectral bands, from Band 1 (B1, 620–670 nm) to Band 7 (B7, 2107–2155 nm). The surface reflectance product (MOD09GHK, now encapsulated in MOD09GA), defined as the reflectance that would be measured at the land surface if there was no atmosphere, is used. Shallow depths and high concentrations of suspended sediment such as those observed along the Inner Niger Delta increase considerably the amount of solar energy reflected by a water body (Bukata 1992). Li et al. (2003) showed that the strong water absorption at wavelengths greater than 1 μm in MODIS B5, B6, and B7 does not allow illumination of the sediments in the water or at the shallow bottom of a water column. A simple combination of threshold technique was performed on MODIS B5 and NDVI [ $NDVI = (B2 - B1) / (B2 + B1)$ ] to delineate the shallow, sediment-laden, open water of the delta flood plain and to also distinguish aquatic vegetation from dry-land vegetation. It is assumed that very low surface reflectance in B5 characterizes open water, without any consideration on NDVI index. When surface reflectance in B5 increases to medium value over a certain threshold value, depending on the NDVI index, there is partial cover of dry land by water, aquatic vegetation, or vegetation on dry land. Small NDVI and high surface reflectance in B5 is a clear signature for dry land (except for over sun-glitter-contaminated water surfaces). This methodology has been validated with in situ measurements in Australian floodplains (J.-F. Crétau et al. 2008, unpublished manuscript). Over the Inner Niger Delta, the method was carefully evaluated using independent datasets such as radar altimetry water level variations

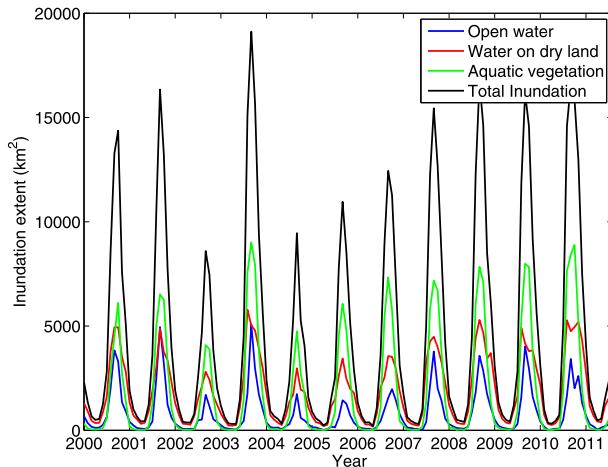


FIG. 1. Time series of the four classes (open water, water on dry land, aquatic vegetation, and vegetation on dry land) in the original MODIS dataset from 2000 to 2012.

over the Niger and inundated floodplains and in situ measurements (Berge-Nguyen et al. 2008). Since these MODIS data are used for the calibration of the downscaling methods using the GIEMS data (25 km  $\times$  25 km), it has been decided here to degrade the MODIS resolution to 500 m (i.e., four MODIS pixels). The area of these 500-m MODIS pixels is 0.2283 km<sup>2</sup>.

The dataset used in this study is composed of an 8-day classification of the Niger Delta region (3° to 7°W, 13° to 16°N) into four classes: 1) open water, 2) water on dry land, 3) aquatic vegetation, and 4) vegetation on dry land. This dataset is available from February 2000 to September 2011. Figure 1 represents the time evolution of the number of pixels in each one of these classes over the considered spatial domain.

In this study, the goal is to downscale the GIEMS inundation estimates to the spatial resolution of the MODIS data (i.e., about 500 m). A MODIS pixel is set to an inundation value equal to one if the MODIS class is equal to the first or second class. Other pixels are set to an inundation value equal to zero. Figure 2 represents a climatology of the probability of each pixel been inundated.

### c. Space–time collocation of the two datasets

To develop the downscaling methodology, it is necessary to project on the same grid the two inundation datasets from GIEMS and MODIS. First, missing pixels are filled using a simple spatial interpolation scheme. Then, the MODIS 8-day files are converted to monthly means. The monthly-mean value considers only the 8-day files that intersect with the month. Each week has a weight proportional to the number of days included in the month. For each pixel, the monthly MODIS class is chosen to be the most frequent class.

Let  $[\mathbf{HR}(i); i = 1, \dots, P]$  be the inundation extent at the MODIS high resolution (HR) for a particular month (the time index is suppressed here for clarity of the presentation), and  $P = 649\,838$  is the number of HR pixels in a MODIS image of the delta. The low-resolution (LR) inundation is noted  $[\mathbf{LR}(j); j = 1, \dots, Q]$ , where  $Q$  is the number of LR boxes in GIEMS. The binary upscaling matrix  $\mathbf{S}$  has dimension  $Q \times P$  and is defined such that 1) each line  $j$  represents a low-resolution box ( $1 \leq j \leq Q$ ), and 2) for each line  $j$ , the columns of the high-resolution pixels included in the low-resolution box  $j$  are set to a value of 1. Using this upscaling matrix, it is possible to link the two spatial resolutions:

$$\mathbf{LR} = \text{area} \times \mathbf{S} \times \mathbf{HR}, \quad (1)$$

where  $\text{area} = 0.2283 \text{ km}^2$  is the surface of a HR pixel.

Using Eq. (1), the 500-m pixels of MODIS estimates are then upscaled to the  $0.25^\circ \times 0.25^\circ$  GIEMS boxes.<sup>1</sup> Figure 3 represents the GIEMS and MODIS climatology for June–January and at the  $0.25^\circ \times 0.25^\circ$  equal-area grid of the LR. GIEMS seem to have higher values than MODIS, but the seasonality appears similar. Low-inundation boxes are not retrieved by GIEMS, but this limitation of the dataset has already been documented (Prigent et al. 2007).

The two global inundation extent time series are compared in Fig. 4. The seasonality is relatively close: correlation is 0.63 for the raw times series, but it increases to 0.86 when a 1-month lag is introduced, with the GIEMS inundation season starting 1 month earlier. This can be explained by possible cloud contamination in the MODIS data that prevents the observations of the beginning of the wet season with MODIS or by the high sensitivity of the passive microwave observations used in GIEMS to very saturated soil. The amplitude of the GIEMS is also larger compared to MODIS estimate. It should be pointed out here that because of this time lag, the downscaling model will be difficult to calibrate. Another important discrepancy is the interannual variability that is quite different for the two estimates. The comparison, evaluation, and validation of both GIEMS and MODIS estimates need to be further studied. The tools developed in this study can facilitate this task.

In Fig. 5, the correlation between the two inundation datasets is represented for each  $0.25^\circ \times 0.25^\circ$  GIEMS box. The correlation in the boxes with no variability in the

<sup>1</sup> In this paper, in order to simplify the presentation, the low-resolution GIEMS points will be called “boxes” and the high-resolution points from MODIS will be named “pixels.”

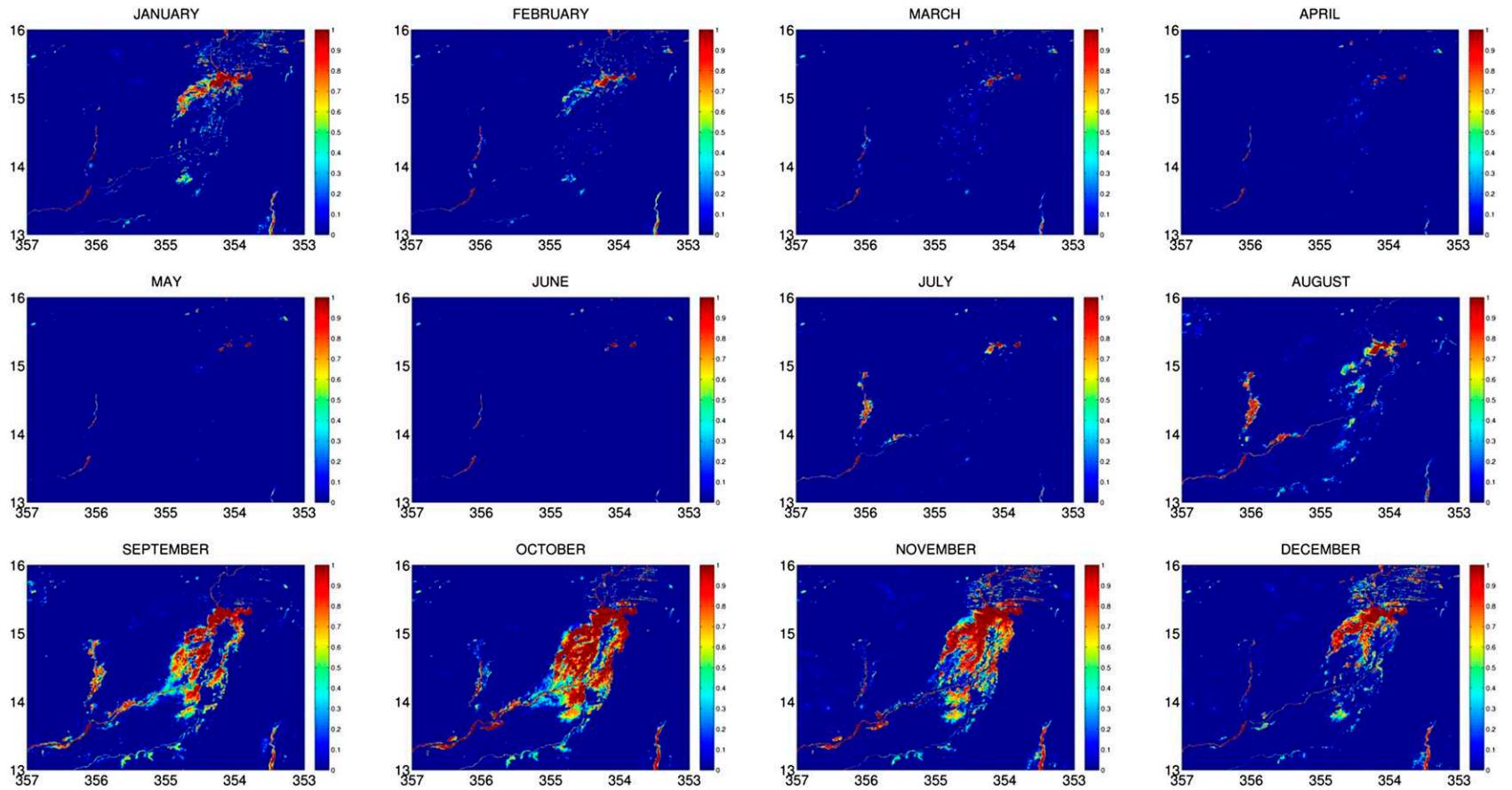


FIG. 2. Monthly–seasonal variability of MODIS inundation at a spatial resolution of about 500 m for the area 13°–16°N and 3°–7°W. The variable is the probability for a pixel to be inundated, estimated over the period 2000–07.

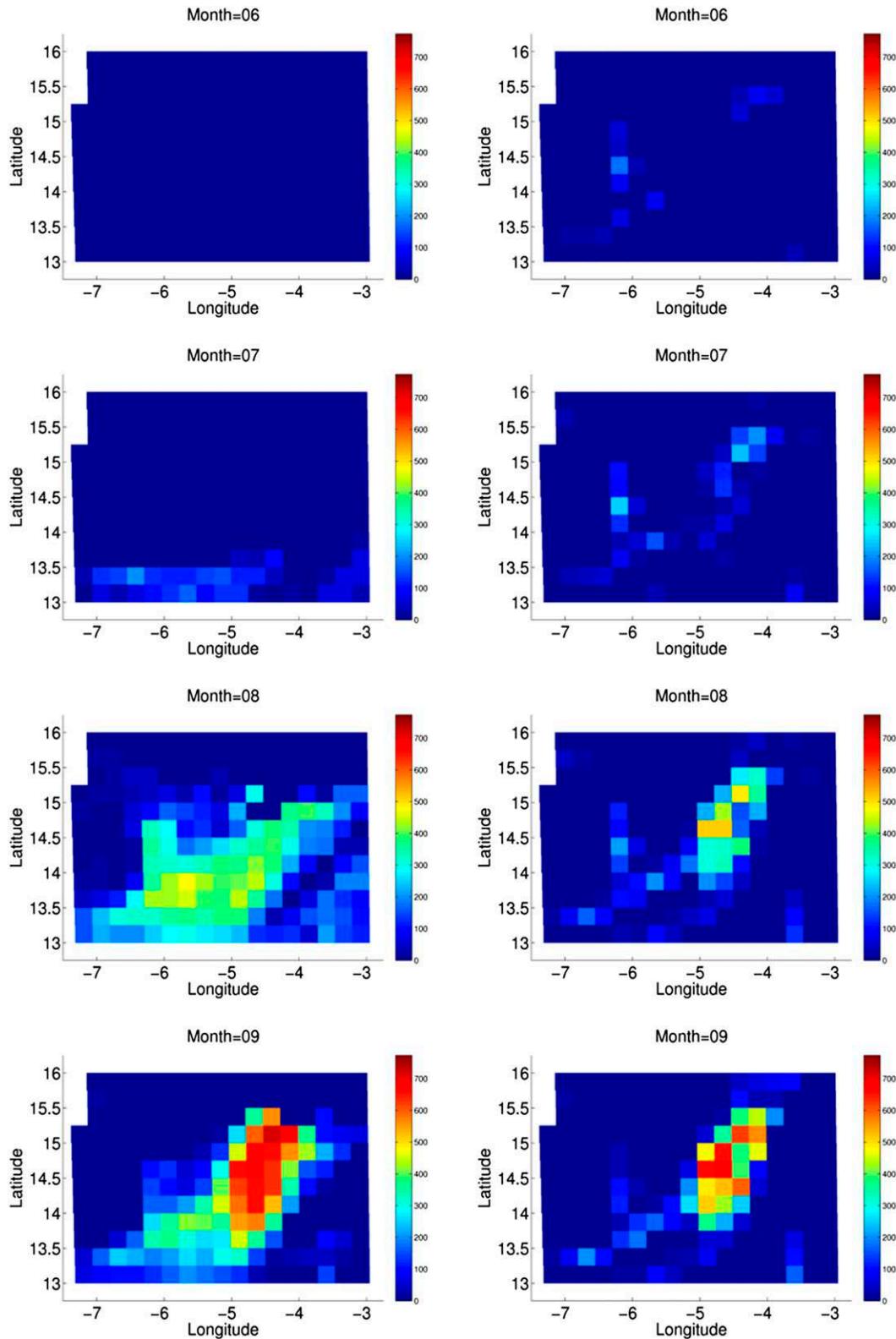


FIG. 3. June–December variation of the inundation extent (km<sup>2</sup>) over the Inner Niger Delta (13°–16°N and 3°–7°W) for (left) GIEMS and (right) MODIS estimates. The spatial resolution of the MODIS inundation estimates has been degraded to the GIEMS resolution (0.25° × 0.25° equal-area grid).

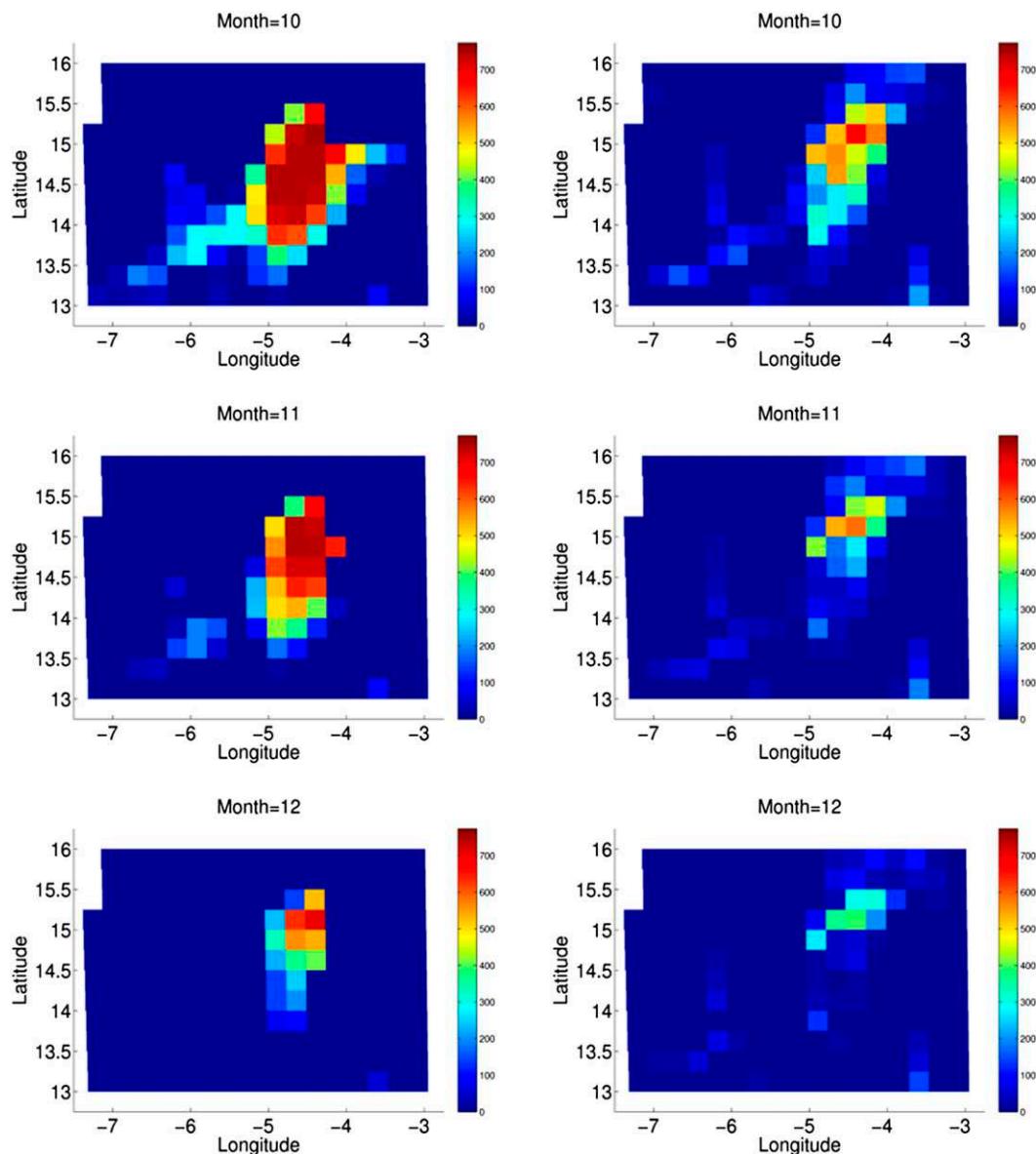


FIG. 3. (Continued)

GIEMS dataset cannot be estimated and is not represented. The main information for downscaling the GIEMS estimates to the MODIS resolution will be extracted from the pixels with maximum correlation. The correlations are high in pixels where the seasonality of inundation is high. This can be explained by the fact that GIEMS does not capture very limited inundation extent, contrarily to MODIS, as it is derived from observations with low spatial resolution. The spatial patterns are similar on both times series, and the correlations between the low and high resolution can be exploited to link the two spatial scales, making downscaling possible.

### 3. Principal component analysis of the MODIS inundation

The goal of this section is to perform a PCA of the inundation variability in the MODIS dataset of section 2b. This is intended to facilitate the downscaling process of the next section by reducing the dimensionality of the inundation space–time variability. However, it will be shown that this is an interesting general methodology for the analysis of the surface hydrology. We will also briefly comment on other potential applications of the PCA.

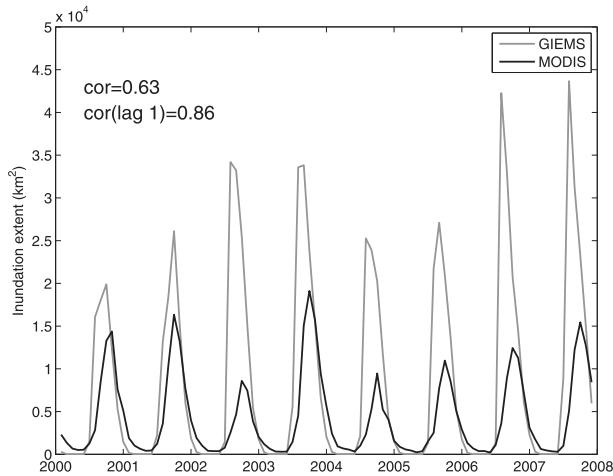


FIG. 4. Time series of the total inundation extent over the Inner Niger Delta for downscaled GIEMS (gray line) and MODIS (black line) estimates.

### a. PCA methodology

PCA is a very general statistical methodology that intends to describe the variability of multivariate observations using a new base defined by a limited set of patterns. PCA has been widely used in geophysics to isolate important spatial patterns, that is, the so-called empirical orthogonal functions (EOFs) (Aires et al. 2000; Jolliffe 2002). PCA can also be used on spectral (Aires et al. 2004) or temporal data (Aires et al. 2002).

Let  $\mathbf{HR}$  be the high-resolution inundation extent from MODIS in binary values (0 for no inundation and 1 for inundation). A spatial map of inundation will be noted as  $\mathbf{HR}(t)$ , for month  $1 \leq t \leq T$  ( $T = 140$  months). The inundation time series for pixel  $i$  will be noted as  $\mathbf{HR}_i$ , for pixels  $i = 1, \dots, P$  ( $P = 649\,838$  pixels in this application).

A PCA is used here to decompose the  $\mathbf{HR}$  space–time variability into a limited number of simple components (Jolliffe 2002). This PCA decomposition can either be done on the time or on the space dimension. However, once the PCA is done in one of these dimensions, it is possible to switch to the other (Legler 1984). Since the spatial dimension  $P$  is much larger than the time dimension  $T$ , it is highly recommended to perform the PCA in the time space: the covariance matrix that needs to be decomposed in the PCA will have dimension  $T \times T$  instead of  $P \times P$ .

Before using the PCA on  $\mathbf{HR}$  data, a preprocessing must be done. The monthly mean  $m(t) = \text{mean}[\mathbf{HR}(t)]$  for  $1 \leq t \leq T$  must be estimated globally over the spatial dimension of the whole Inner Niger Delta. The preprocessing is limited here to this centering:  $\overline{\mathbf{HR}}(t) = [\mathbf{HR}(t) - m(t)]$ . In this study, no normalization is used in

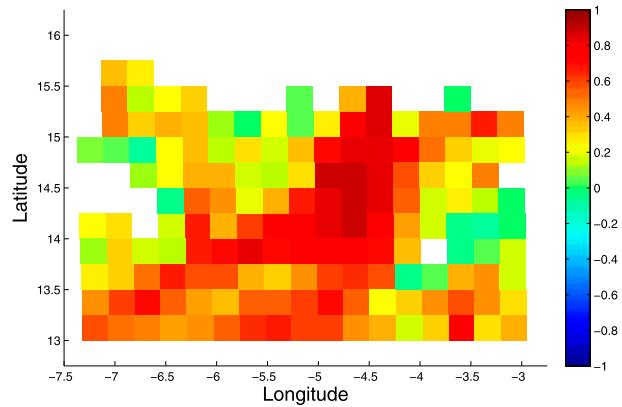


FIG. 5. Map of the correlation between the GIEMS and the MODIS inundation estimates at the GIEMS spatial resolution.

the preprocessing step in order to not artificially emphasize months with no or very limited inundation variability. The goal is here to study the inundation variability so it is reasonable to focus on months with significant inundation extent. Furthermore, the GIEMS estimates give no inundation extent information for these low-inundation months, so it would be illusory to work on these months.

The PCA decomposition in time allows for the estimation of two matrices:  $\mathbf{C}$ , the  $P \times T$  matrix of the components (or scores), and  $\mathbf{F}$ , the  $T \times T$  matrix with the  $T$ -dimensional temporal base function ( $f_k$ ;  $k = 1, \dots, T$ ). Matrix  $\mathbf{F}$  describes the axis of a new space of dimension;  $T$  and  $\mathbf{C}$  are the coordinates of the data in this new space.

The temporal decomposition can be written

$$\begin{aligned} \overline{\mathbf{HR}}_i(t) &= \mathbf{C}(i, :) \cdot \mathbf{F}, \quad \forall 1 \leq t \leq T \\ &= \mathbf{C}(i, 1) \cdot \mathbf{F}(1, t) + \dots + \mathbf{C}(i, T) \cdot \mathbf{F}(T, t) \\ &\simeq \mathbf{C}(i, 1) \cdot \mathbf{F}(1, t) + \dots + \mathbf{C}(i, K) \cdot \mathbf{F}(K, t) \end{aligned} \quad (2)$$

for  $1 \leq K \leq T$ ,

where the colon ( $:$ ) in the matrix location indicates that all the lines (or columns) are being considered. In the third line of Eq. (2), instead of using the full  $T$  PCA components, only the first  $K$  components are used ( $K \leq T$ ). This allows us to compress the data, and each time series can be described using these first  $K$  temporal base functions. Of course, this can introduce relatively small compression errors. The space decomposition can be written

$$\begin{aligned} \overline{\mathbf{HR}}(t) &= \mathbf{F}'(t, :) \cdot \mathbf{C}', \quad \forall 1 \leq t \leq T \\ &= \mathbf{F}'(t, 1) \cdot \mathbf{C}'(1, :) + \dots + \mathbf{F}'(t, T) \cdot \mathbf{C}'(T, :) \\ &\simeq \mathbf{F}'(t, 1) \cdot \mathbf{C}'(1, :) + \dots + \mathbf{F}'(t, K) \cdot \mathbf{C}'(K, :) \end{aligned} \quad (3)$$

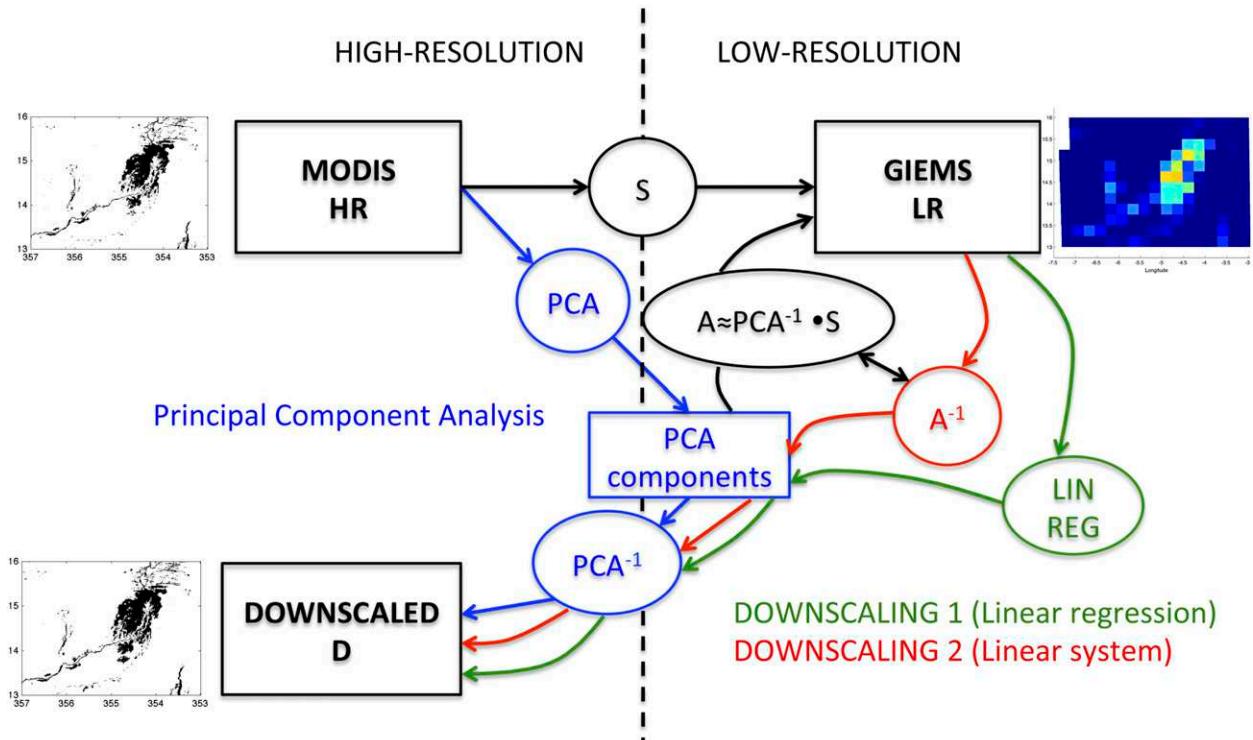


FIG. 6. Scheme representing the link between the (left) high- and (right) low-resolution areas. The PCA operators (section 3c) are represented in blue. The first downscaling method using a linear regression (section 3a) is represented in green and the linear algebra downscaling (section 3b) is represented in red.

where the prime is the transpose sign. In this decomposition, the terms  $\mathbf{F}'(t, k)$  for  $1 \leq k \leq K \leq T$  play the role of the components, and the spatial patterns  $\mathbf{C}'(t, k)$  play the role of the  $P$ -dimensional spatial base functions. A spatial inundation map at time  $t$  can be decomposed in the new basis of these spatial patterns.

In Fig. 6, the blue arrows and operators indicate the PCA analysis. The PCA provides the spatial EOF and the inverse PCA operator  $\text{PCA}^{-1}$  allows it to transform the PCA components into a high-resolution estimate  $D$ . The other parts of this scheme will be commented on below.

*b. PCA results*

Figure 7 represents the temporal base functions ( $f_k$ ;  $k = 1, \dots, 4$ ) in the left column. The monthly-mean inundation evolution  $m(t)$  used to center the data before the PCA (section 3a) is represented in the top of the right column. In the following graphs of the right column [below  $m(t)$ ], the first base function  $f_1$  is added with components 2, 3, and 4 in order to measure their role in the PCA representation:  $f_1 \pm 0.2 \cdot f_i$ , for  $i = 2, 3$ , and 4. The first component is a low-frequency global time evolution that includes the seasonal and interannual variability at the global scale. Even if the monthly-mean average  $m(t)$  has been suppressed from the data before

their use in the PCA, there are still seasonal and interannual anomalies among the different locations, and this component is used to describe this type of variability. The second base function has an annual cycle and is used to modulate the starting date of the inundation season. The inundation season starts earlier when this second component is added (in black) and is delayed when it is suppressed (dashed line). The third component has a similar behavior for the end of the inundation season. Adding or suppressing this component can delay or advance the end of the season. Both components 2 and 3 are used to represent the skewness in the inundation season. The fourth component has an annual and biannual frequency. It can be seen that this component has a peak in the start and the end of the inundation season, so this component can be used to increase or decrease the inundation duration. It is used to modulate the inundation cycle duration and kurtosis. The higher-order components even with a localized anomaly in time (not shown in this figure) have a higher time frequency and are more difficult to interpret in the time dimension. These temporal components need to be interpreted using their associated spatial patterns.

Similarly to Fig. 7 for the temporal base functions  $f_k$ , Fig. 8 represents the corresponding spatial components

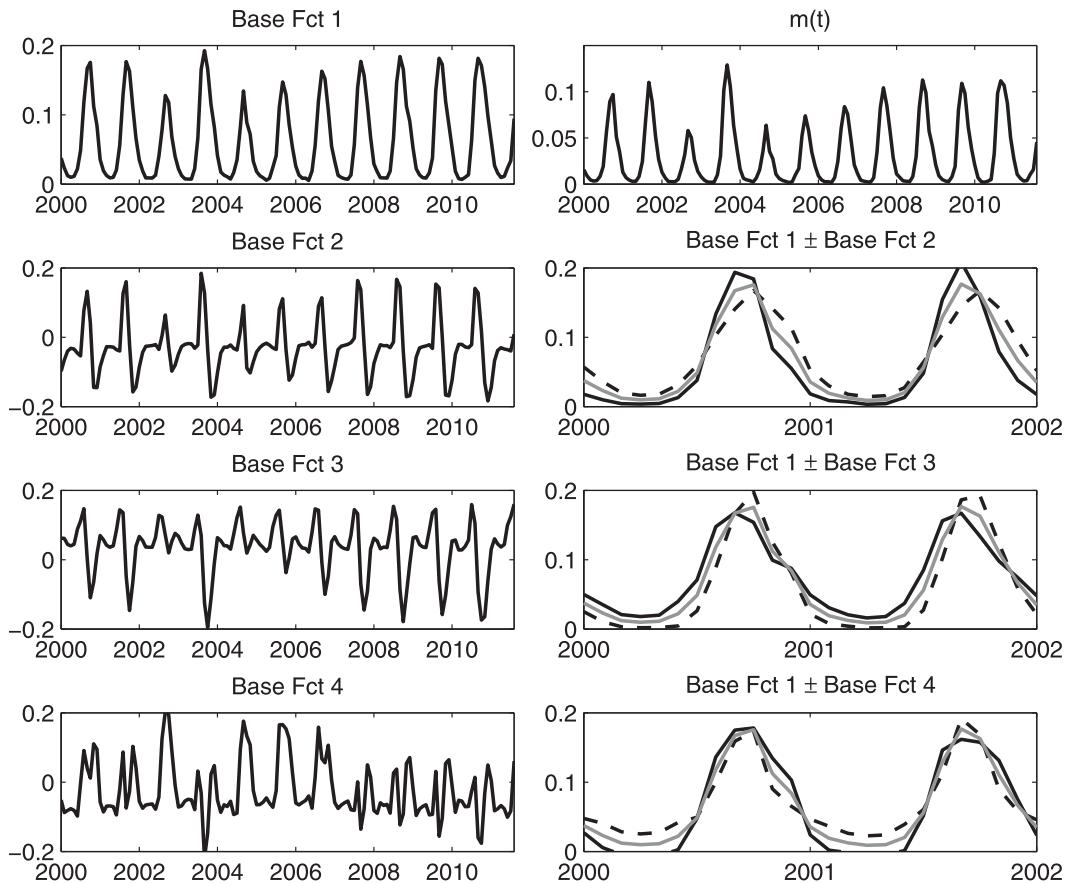


FIG. 7. (top to bottom) (left) The first four temporal PCA base functions of the MODIS inundation. (top to bottom) (right) The global monthly-mean inundation  $m(t)$  defined in section 3 for selected years and used to center the MODIS data; and the sum and difference ( $\pm$ ) of base function 1 and base functions 2, 3, and 4, respectively, being used to better understand their impact on the inundation time series. Gray is for the base function 1, and black solid and dashed lines are for base function  $1 \pm$  the other base functions, respectively.

$\mathbf{C}(:, k)$  in the right (left and middle columns will be commented on in section 3d). As explained earlier, the first components are global in nature, and the higher-order components become more localized in space. An interesting remark comes from the fact that these spatial patterns are relatively smooth. This means that a simple spatial interpolation could be used to obtain high-resolution fields at an even greater spatial resolution, for example, 500 m (compatible with the SWOT spatial resolution) instead of the 500 m used in this study (section 2b).

### c. Number of PCA components

It is not possible to use the  $T = 140$  PCA components in the downscaling scheme. The low-resolution estimates from GIEMS provide  $Q = 201$  pieces of information, so in principle, it should be possible to constrain the  $T$  components using them. However, there are not 201 pieces of independent information in GIEMS estimates:

GIEMS has no information on the low seasonal variability boxes, so some of the  $Q$  boxes have no variability at all (see section 2c). Furthermore, low-resolution boxes are correlated to each other. Moreover, some HR components are very localized. If two of them hold inside a same low-resolution box, it is impossible to constrain both of them using only one low-resolution estimate from GIEMS. Using these local principal components would make this problem ill posed, and numerical instabilities will affect the downscaling.

Therefore, the question of the number of necessary components to represent satisfactorily the inundation variability appears. In Fig. 9, a sample of 4 yr (2000–03) of the time series of the MODIS inundation extent in the Niger Delta (similar to Fig. 4), together with the compression/reconstruction of the time series using the PCA with an increasing number of components ( $k = 1, 2, 3, 5, 10, \text{ and } 20$ ). Using only one component (in blue) provides already a good seasonal cycle phase, but this is

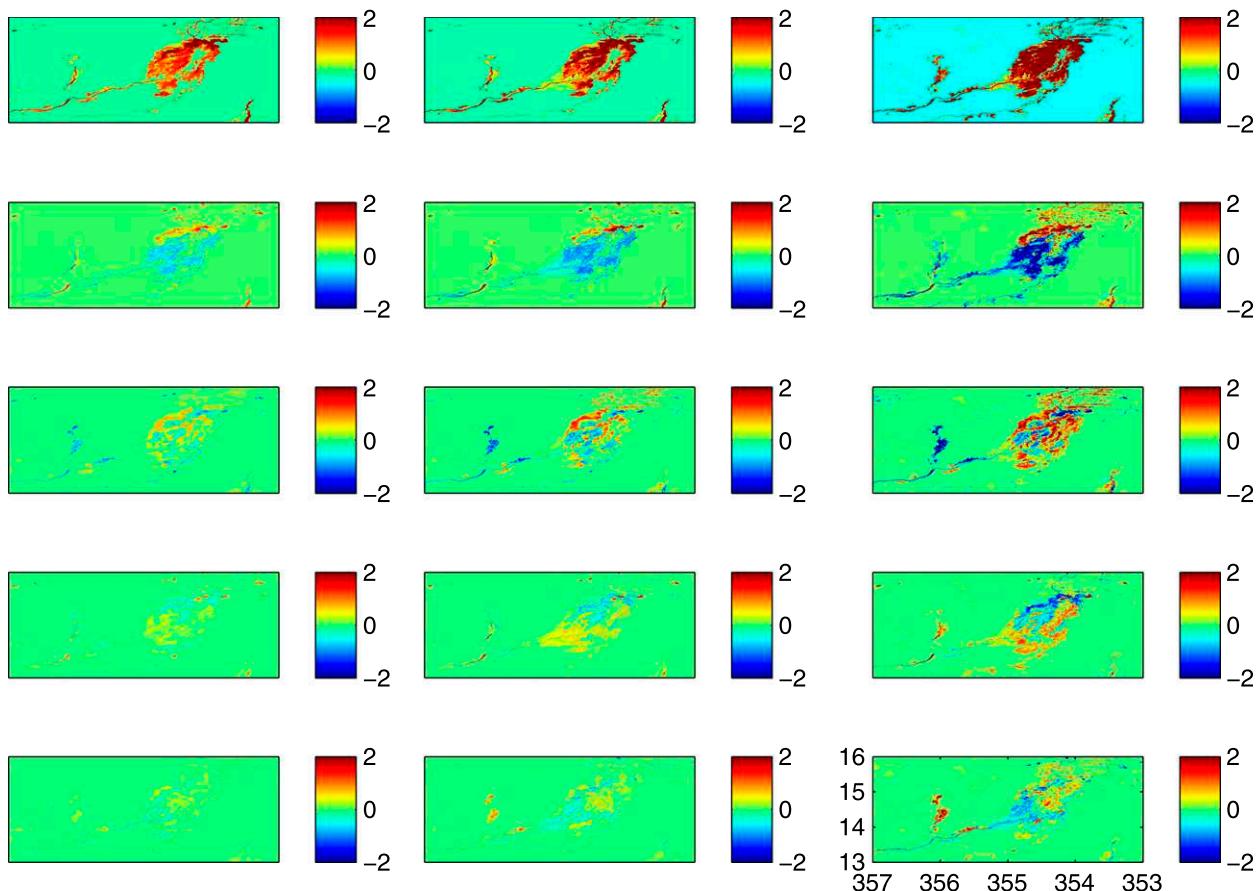


FIG. 8. (top to bottom) The first five spatial EOFs of the MODIS inundation estimates, when (left) 12, (middle) 36, and (right) 140 months are used in the PCA.

clearly not satisfactory. When the number of components increases, the times series increases its seasonal cycle amplitude and becomes closer to the MODIS estimate, with both high and low states better represented. Twenty components appear to be satisfactory in terms of total inundation extent. At this stage, many more components would be required to obtain a rather limited improvement.

Similar to the previous figure for the PCA compression errors in time, it is possible to represent the PCA compression–reconstruction of a spatial map to illustrate its spatial error patterns. Figure 10 represents first a sample original MODIS inundation map (ORIG). Below, the left column represents the PCA compression/reconstruction when an increasing number of components ( $k = 2, 8, 14,$  and  $20$ ) are used. The PCA representation becomes more precise when the number of components increases. The right column represents the difference between the reconstructed and the original inundation. Gray pixels are for pixels well classified, and black and white pixels are for classification errors. The

percentage of well-classified pixels is indicated on the right. Again, 20 components appear to be satisfactory with 99.1% of well-classified pixels.

By increasing the number of PCA components, it can be seen in Fig. 11 that both the ability to identify the inundation measured by the sensitivity (the percentage of positive inundations that are correctly identified), and the ability to identify the noninundation cases (the proportion of noninundation cases that are well classified, or specificity) are increasing regularly when increasing the number of components. The sensitivity increases from 60% with one component, to 90% with 20 components, and reaches 100% with a little more than 30 components.

Similarly, by increasing the number of components, the PCA representation needs about 30 components to reach the 100% level. However, with only one component, the specificity is higher than 98%. This shows that the PCA representation privileges the no inundation state, which is because the goal of the PCA is to represent as well as possible the variability of the dataset that,

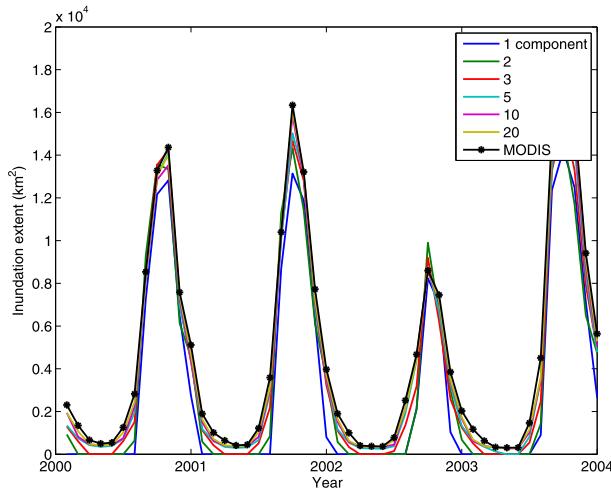


FIG. 9. Impact of the number of PCA components in the reconstruction of the total inundation extent times series.

in this case, includes much more inundated than water-free pixels. This could be corrected if necessary (e.g., by decreasing the thresholds to define the inundation state or by equating the number of inundated and water-free pixels in the database used to perform the PCA), but it will be seen in the following that this regular PCA representation is satisfactory for the downscaling purpose.

As discussed above, with the number of components of about 20, it is possible to represent quite precisely the high-resolution inundation extent of MODIS estimates. However, this does not mean that 20 components is the ideal number of components to be used in the downscaling process; it is only an upper limit.

#### d. Number of necessary monthly MODIS data

In the Inner Niger Delta, a long time series of 140 monthly-mean MODIS data are available from 2000 to 2012 and the robustness of the PCA results to the number of available months can be tested. This analysis can also help decide if locations with much smaller time series can be downscaled or not. For the SWOT mission, it can help decide how many HR samples are needed to set up a downscaling of the LR observations.

The percentage of variance explained by the first PCA components (not shown) has been represented when the PCA is performed using 12, 36, and 140 monthly MODIS data. Of course, when fewer months are used for the PCA, fewer components need to be used to represent the variability. Specifically, with fewer months in the dataset, there is less interannual variability, so there is no need for components describing this interannual variability. However, at about 15–20 components, the curves of the explained variances start to become more linear. This elbow is often an important

feature in these curves, and it separates large and global features from the noise or more local features.

The spatial EOFs are represented in Fig. 8, again, when using 12, 36, and 140 months in the dataset. The spatial patterns for the first global EOF are rather stable, but the intensity and contrast of patterns becomes higher when more months are used. However, what will be used in the following for the downscaling are actually the spatial patterns, rather than their intensity. These results are encouraging for regions on Earth where short time series are available; even 1 yr (12 samples) of data seems to be exploitable.

#### e. Time interpolation

The PCA representation of the high-resolution inundation will be used for space interpolation (e.g., downscaling), but it can also be used for time interpolation (Aires et al. 2004). Let  $\mathbf{HR}_{t_1}$  and  $\mathbf{HR}_{t_2}$  be two consecutive monthly estimates from MODIS (section 2b). Using the PCA decomposition of Eq. (3),

$$\mathbf{HR}_{t_k} \simeq \overline{\mathbf{HR}} + \mathbf{F}'(t_k, 1) \cdot \mathbf{C}'(1, :) + \dots + \mathbf{F}'(t_k, K) \cdot \mathbf{C}'(K, :); \quad (4)$$

it can be noted that  $\mathbf{HR}_{t_1}$  is entirely determined by the vector of components  $\mathbf{F}'(t_1, :)$  and that  $\mathbf{HR}_{t_2}$  is determined by  $\mathbf{F}'(t_2, :)$ . It is then possible to perform a linear interpolation for each component  $k$  for time  $t$  with  $t_1 \leq t \leq t_2$ :

$$\mathbf{F}'(t, k) = \mathbf{F}'(t_1, k) + \frac{t - t_1}{t_2 - t_1} \times [\mathbf{F}'(t_2, k) - \mathbf{F}'(t_1, k)]. \quad (5)$$

These components can then be used to reconstruct  $\mathbf{HR}_t$ . With the help of the PCA representation, the complex binary interpolation in a space of dimension  $P = 649\,838$  (i.e., number of pixels in the high-resolution dataset) has been transformed to a simple linear interpolation of  $K = 10$  PCA components.

Figure 12 represents the results of such a time interpolation. The two consecutive monthly inundation estimates are for  $t_1$  equal to October 2000 (top map) and  $t_2$  for November 2000 (bottom map). Five intermediate times are represented between them. The behavior of time interpolation scheme appears very satisfactory. The total inundated area expressed in square kilometers is represented in each map, and this surface decreases regularly too from  $t_1$  to  $t_2$ .

#### f. Spatial interpolation for missing values

Similar to the temporal interpolation scheme described in the previous section, it is also possible to use

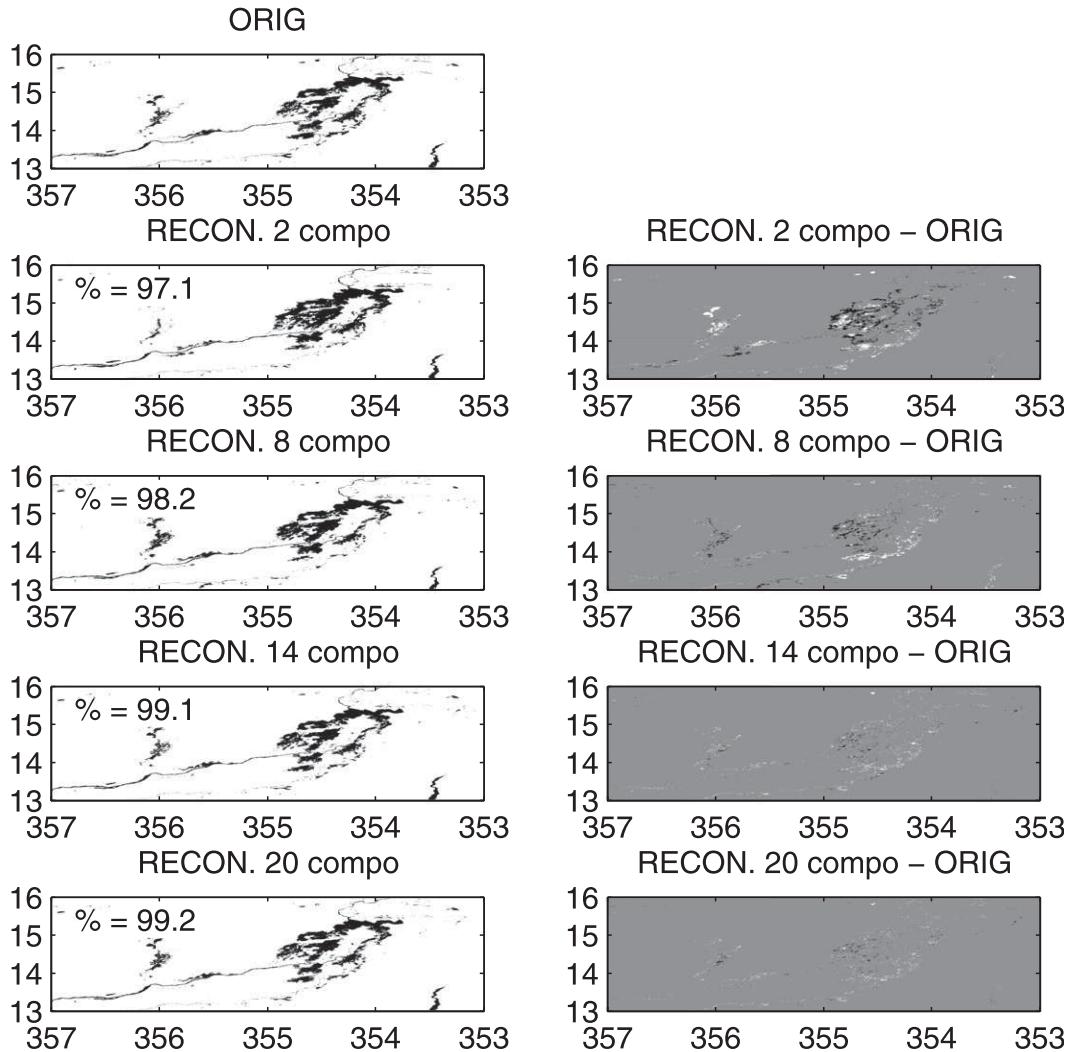


FIG. 10. (top left) A sample (August 2000) of high-resolution inundation from MODIS estimates for the area 13°–16°N and 3°–7°W in terms of the percentage of well-classified pixels. The graphs below represent the PCA compression–reconstruction inundation when an increasing number of PCA components are used (2, 8, 14, and 20). (right, top to bottom) The reconstruction using 2, 8, 14, and 20 components minus the original inundation: black pixels are inundated pixels not retrieved by the PCA representation, white pixels are not inundated but represented as such by the PCA, and gray pixels are well represented.

the PCA representation to interpolate spatially the inundation extent fields. This can be useful, for example, when the presence of clouds introduces missing values in an inundation field, with MODIS particularly sensitive to the presence of clouds. For the SWOT mission, missing values can be the result of high precipitation or the interferometric layover.

The approach is rather simple. Let us first consider a regular high-resolution inundation field  $\mathbf{HR}$  (Fig. 13a). It is supposed that the observed field  $\mathbf{HR}'$  includes missing values in an ensemble of pixels (see Fig. 13b, which includes a large square of missing values). These missing pixels are first set to a mean value  $\mathbf{HR}''$  in order to be compatible with

the PCA representation. Then, PCA components are computed by projecting  $\mathbf{HR}''$  into the spatial base functions. These PCA components are used to reconstruct the spatial field  $\widehat{\mathbf{HR}}$  (Fig. 13c). This simple interpolation scheme provides quite reasonable inundation estimates in the missing values square:  $\mathbf{HR}$  (Fig. 13a) and  $\widehat{\mathbf{HR}}$  (Fig. 13c) are quite similar. The total inundation area is also indicated for the three maps, and the inundation of  $\widehat{\mathbf{HR}}$  (14 463 km<sup>2</sup>) is much closer to that of  $\mathbf{HR}$  (14 364 km<sup>2</sup>) than that of the missing value field  $\mathbf{HR}'$  (13 256 km<sup>2</sup>).

This spatial interpolation procedure should be highly effective too when a high-resolution inundation map is perturbed by retrieval noise. This PCA projection and

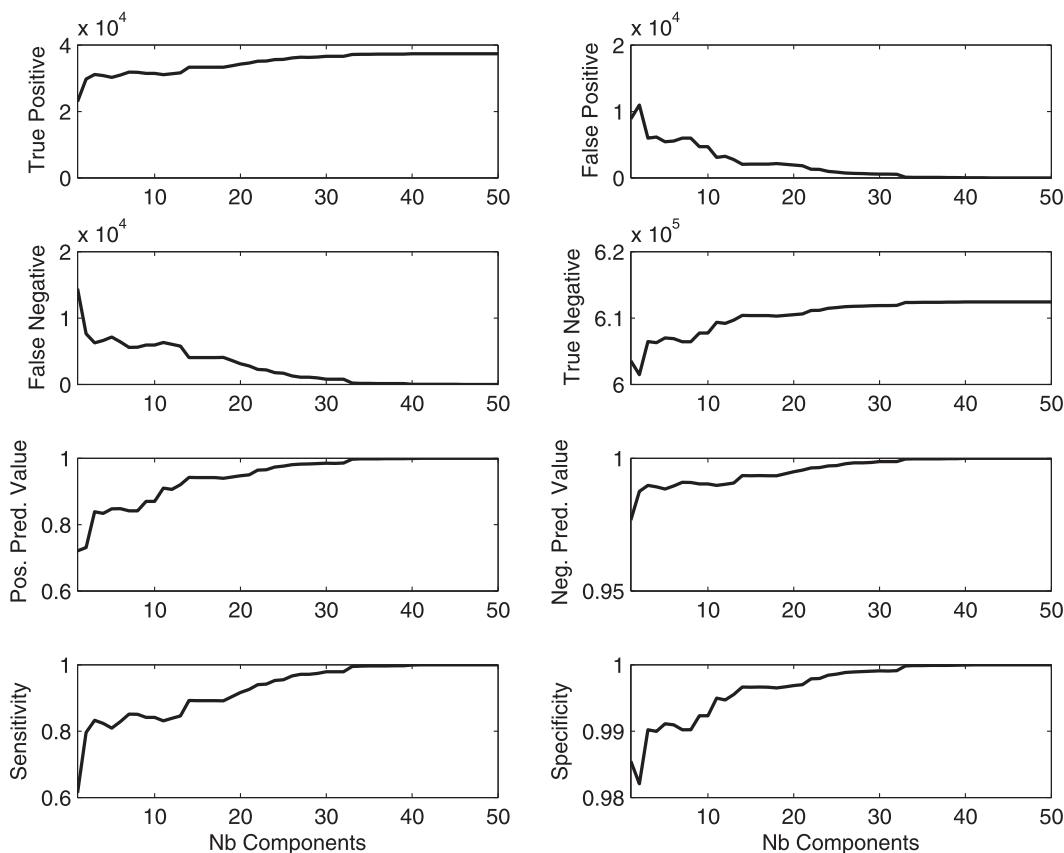


FIG. 11. Evolution of the quality criteria of the classification (see text for definitions) when increasing the number of components in the PCA representation of the high-resolution inundation.

reconstruction process should be an ideal “denoising” technique for inundation fields.

Technically, the fact that a large area of the inundation field is set to an averaged value that might be quite wrong does perturb the PCA representation. The resulting components can have a different magnitude, but the relative importance of the components appears to be correct. To solve this scaling issue, it is just necessary to transform **HR** into binary values.

#### 4. Downscaling

The two downscaling methodologies presented in this section are both based on the PCA representation of the MODIS HR inundation extent. The first approach uses a linear regression to link the high to the low resolution. The second approach is based on the resolution of a linear system using simple linear algebra.

##### a. Downscaling using a linear regression

###### 1) METHODOLOGY

The goal of this section is to obtain a statistical model that uses the GIEMS inundation data as inputs and

predicts the PCA components that represent the high-resolution inundation from MODIS.

Let **LR** be the low-resolution inundation estimates from GIEMS. There are  $Q = 201$  GIEMS boxes in the Niger domain under study, so  $\mathbf{LR} \in \mathbb{R}^{201}$ . However, correlations exist among these 201 box values, and the number of degrees of freedom is much more limited than the number of covariates. It is rather common to preprocess the data before using them as input of a statistical model, again with a PCA. The goal is to reduce the number of input variables in the downscaling process to regularize the process. Tests have been conducted (not shown), and 10 components provide satisfactory results. This is about the same number to be used in the output of the downscaling model coding the high resolution.<sup>2</sup> We note  $\mathbf{LR} \in \mathbb{R}^{10}$  in this PCA representation of the GIEMS data.

<sup>2</sup>If too many components are used as inputs of the statistical downscaling model, and if these high-order components do not bring any information for the downscaling, then they are just disregarded by the linear regression.

The inundation extent from MODIS is represented in the space of the first 20 PCA components (section 3). We note  $\mathbf{Y} \in \mathbb{R}^{10}$  in these components.

In this paper, a simple statistical model is first experimented to link  $\overline{\mathbf{LR}}$  and  $\mathbf{Y}$ , the linear regression:  $\mathbf{Y} = \mathbf{A} \times \overline{\mathbf{LR}} + \mathbf{B}$ . It is necessary to note here that only 95 points (February 2000 to December 2007; see section 2c) are available to calibrate this linear regression. This is a limiting factor. The stability of the results to the number of samples will be studied in a future study.

This downscaling methodology is illustrated in green in the scheme of Fig. 6. The linear regression tool links the GIEMS low-resolution data to the PCA components (in blue). The PCA components are then translated into the downscaled high-resolution fields using the reverse PCA.

## 2) RESULTS

In Fig. 14, five examples of MODIS inundation data (left) and corresponding downscaled GIEMS data (right) are presented. The overall comparison appears to be satisfactory. The spatial structures of the downscaled inundation are very similar to the structures of the original MODIS estimates: this is not a surprise, as the use of the PCA component  $Y$  ensures that the downscaled values are in the same space. Some discrepancies can be noticed when looking carefully. The main difference concerns low-inundation extent months where the GIEMS estimates are close to zero. In these cases, the downscaling is always similar, with a constant low-inundation map. The same behavior will be observed in the temporal results.

Figure 15 is similar to Fig. 4: the GIEMS (dashed) and MODIS (black) total inundation extent estimates are represented together with the inundation estimate from the downscaled data (gray). The seasonality of the downscaling is correct. The phase of the downscaled inundation is closer to MODIS than to GIEMS. Another behavior could be obtained with a different choice of the downscaling configuration. The minimal value of the inundation extent is constant with a non-zero value. It is not surprising to have difficulties in retrieving the low-inundation values knowing that the GIEMS estimate is of limited quality for the lower values. Technically, the problem is that, among the months where GIEMS is equal to zero, there are MODIS months that are not equal to zero. The linear regression is constrained to make a compromise and will associate a nonzero value for all of these months. It appears that this approach transforms the dynamical behavior of GIEMS too closely to the MODIS estimates. This is related to the low number of data samples in order to calibrate the linear regression model. In

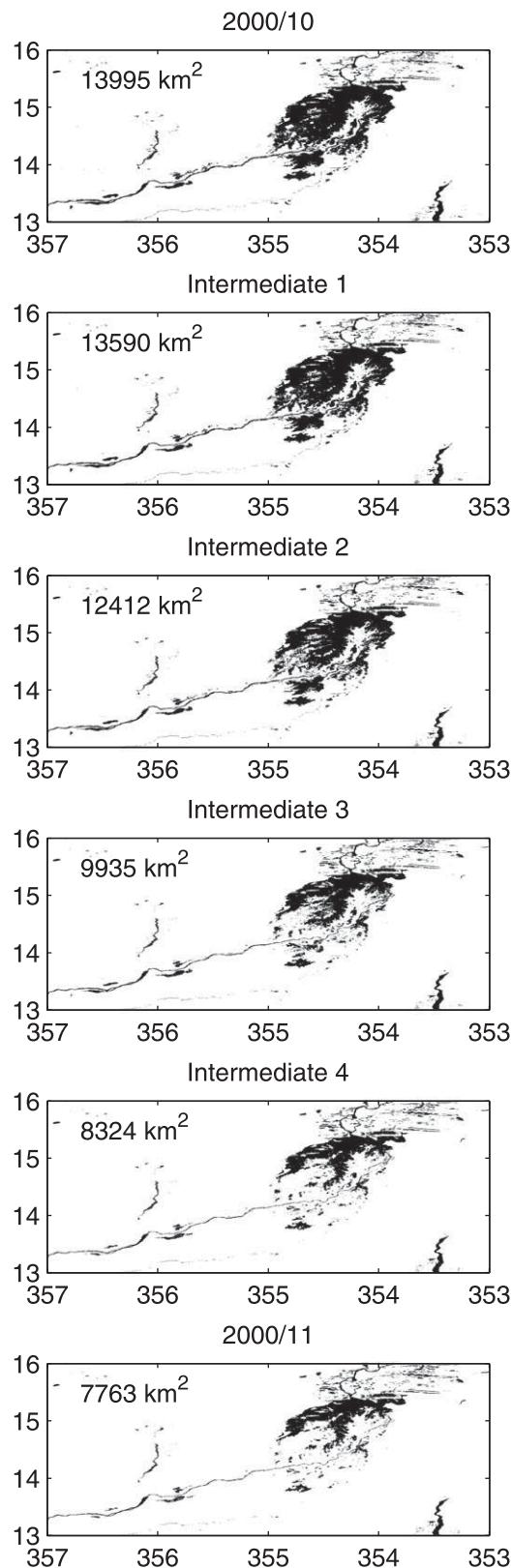


FIG. 12. (top to bottom) Temporal interpolation into five intermediate time steps in October–November 2000 of MODIS inundation estimates for the area 13°–16°N and 3°–7°W. The total inundation area is indicated in square kilometers on each map.

the following section, an another, less data-demanding downscaling technique is introduced.

### b. Downscaling using matrix inversion

#### 1) METHODOLOGY

In this section, we present a new downscaling methodology based on the resolution of a linear system. The technique uses again the PCA of the space–time variability of the high-resolution inundation.

Following Eq. (3), the PCA compression of a high-resolution map  $\mathbf{HR}(t)$  is given by

$$\mathbf{HR}(t) = \mathbf{F}'(t, 1:K) \cdot \mathbf{C}'(1:K, t) + m(t). \quad (6)$$

It is important, at this stage, to keep in the equations the monthly mean  $m(t)$  in the preprocessing step of the PCA (section 3a). This equation can be multiplied by the change-resolution matrix  $\text{area} \times \mathbf{S}$ , introduced in section 2c:

$$\begin{aligned} \text{area} \times \mathbf{S} \times \mathbf{HR}(t) &= \text{area} \times \mathbf{S} \times \mathbf{F}'(t, 1:K) \cdot \mathbf{C}'(1:K, t) \\ &+ \text{area} \times \mathbf{S} \times m(t). \end{aligned} \quad (7)$$

The first term corresponds to the upscaling of the high-resolution inundation toward the low resolution. This means that it can be equated to the low-resolution  $\mathbf{LR}(t)$  from GIEMS. Based on the low-resolution  $\mathbf{LR}$  from GIEMS, it is possible to define a linear system:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{B}, \quad (8)$$

where  $\mathbf{Y} = \mathbf{LR}(t)$  from the GIEMS estimate;  $\mathbf{A} = \text{area} \times \mathbf{S} \cdot \mathbf{F}'(t, 1:K)$  (this matrix is constant in time and can therefore be used for any time step, once estimated);  $\mathbf{X} = \mathbf{C}'(1:K, t)$  (the PCA components representing the high resolution); and  $\mathbf{B} = \text{area} \times \mathbf{S} \times m(t)$  (again, a constant matrix). This linear system relates the  $Q = 201$  observations of the low resolution in  $\mathbf{Y}$  to  $\mathbf{X}$ , and the  $K = 10$  unknowns represent the high-resolution PCA components  $\mathbf{C}'(1:K, t)$ . Matrix  $\mathbf{A}$  has dimension  $Q \times K$  (not square) and can be inverted using a pseudoinverse. In our application, the number of LR boxes is  $Q = 201$  and the number of components was set to  $K = 10$ . The inversion of matrix  $\mathbf{A}$  is an overdetermined problem. The PCA components can then be estimated using

$$\tilde{\mathbf{C}}'(1:K, t) = \mathbf{A}^{-1} \cdot \mathbf{S} \times \mathbf{HR}(t) - \mathbf{A}^{-1} \cdot \mathbf{S} \times m(t). \quad (9)$$

This means that once the  $K \times Q$  matrix  $\mathbf{A}^{-1}$  is obtained, any low-resolution inundation  $\mathbf{LR}(t)$  from GIEMS estimates can be converted into components  $\tilde{\mathbf{C}}'(1:K, t)$ , which allows, using Eq. (3) (third line), to obtain its HR downscaling:

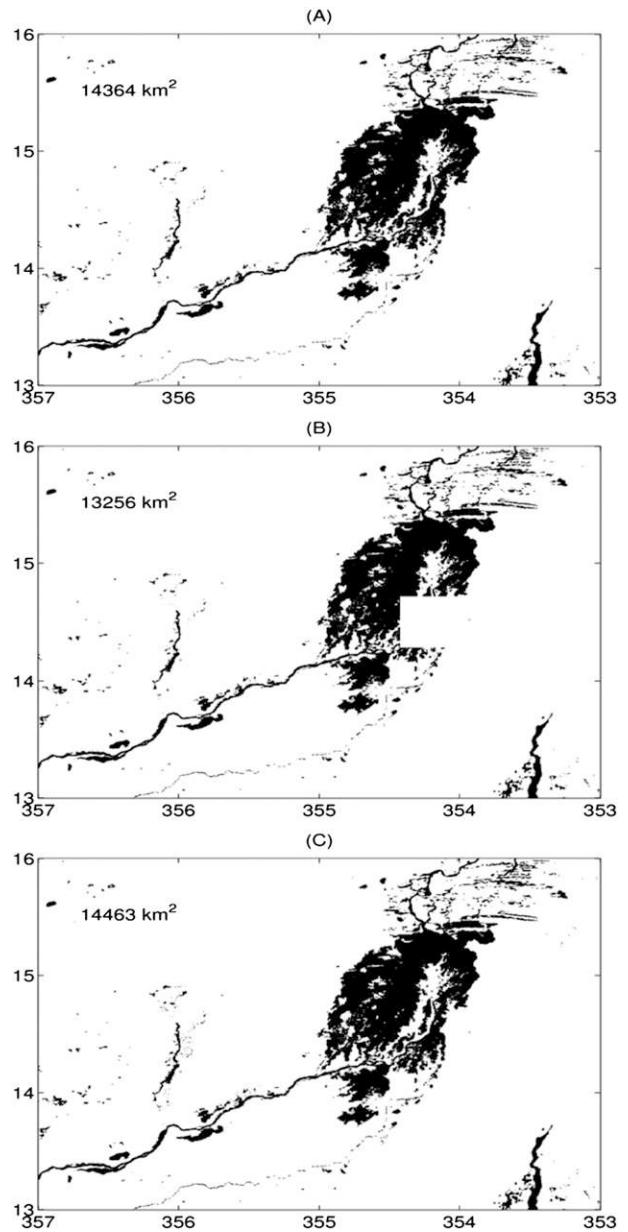


FIG. 13. Spatial interpolation of MODIS high-resolution inundation estimates for October 2000: (a) original inundation estimate, (b) inundation with missing values, and (c) interpolated inundation using PCA. The total inundation area is indicated in square kilometers on each map.

$$\mathbf{D}(t) = \mathbf{F}'(t, 1:K) \cdot \tilde{\mathbf{C}}'(1:K, t) + m(t). \quad (10)$$

This matrix inversion downscaling procedure is represented in red in the scheme of Fig. 6. The operator  $\mathbf{A}^{-1}$  is defined as the inverse of the PCA components to GIEMS low-resolution estimates ( $\text{PCA}^{-1} \times \mathbf{S}$ ). This is a purely linear algebraic solution to the downscaling problem. It is a deconvolution of the low resolution

toward the high-resolution patterns characterized by the PCA analysis of the MODIS dataset.

In this technique, the PCA patterns from the high resolution are applied to the low-resolution inundation data from GIEMS. It is important to facilitate this inter-resolution comparison. If the high-resolution patterns are equated to inundation values much larger or much lower than what was used to build them, the PCA representation becomes nonoptimal and the inversion problem is ill posed [matrix **A** in Eq. (9) becomes ill conditioned]. Therefore, before applying Eq. (9), the GIEMS estimates are first normalized toward MODIS values. A “basin normalization” is chosen (see Aires et al. 2013): the minimum and maximum GIEMS inundation areas at the basin scale are equated to the minimum and maximum MODIS areas. In Fig. 16, these normalized GIEMS estimates are shown in black.

## 2) RESULTS

Figure 16 represents the time evolution of the scaled GIEMS and downscaled total inundation areas. The minimum and maximum values of the scaled GIEMS (0 and about 20 000 km<sup>2</sup>) have been set up to the minimum and maximum values of MODIS estimates. Interannual variability of the downscaled values is close to the GIEMS one. Some discrepancies still exist for the very high inundation months. The incoherencies between the low (from GIEMS) and high resolution (from MODIS) are higher for these months, so it affects the inversion of Eq. (8). However, note that the matrix inversion to solve this linear system could become more stable if a more sophisticated retrieval scheme was used. For a Bayesian inversion, taking into account in the retrieval a first guess of the solution and its uncertainties (Tarantola 1987) would be an interesting extension of this downscaling scheme.

Figure 17 represents two examples of downscaling for two different months: the GIEMS low-resolution estimates on the left, the high-resolution downscaling in the center, and the upscaling of the downscaling on the right. If the downscaling was perfect and if the GIEMS data were perfectly coherent with the PCA representation of the high-resolution inundation, then the left and right columns would be identical. This is not the case. However, this might not be a problem: the downscaling has introduced inundation values where GIEMS had none. For instance, around  $-6^{\circ}$  latitude and  $13.5^{\circ}$  longitude, the downscaling estimates inundation based on inundation patterns of the PCA. As commented earlier, GIEMS retrieval has difficulties with low-inundation values, but the downscaling has the potential to correct this problem.

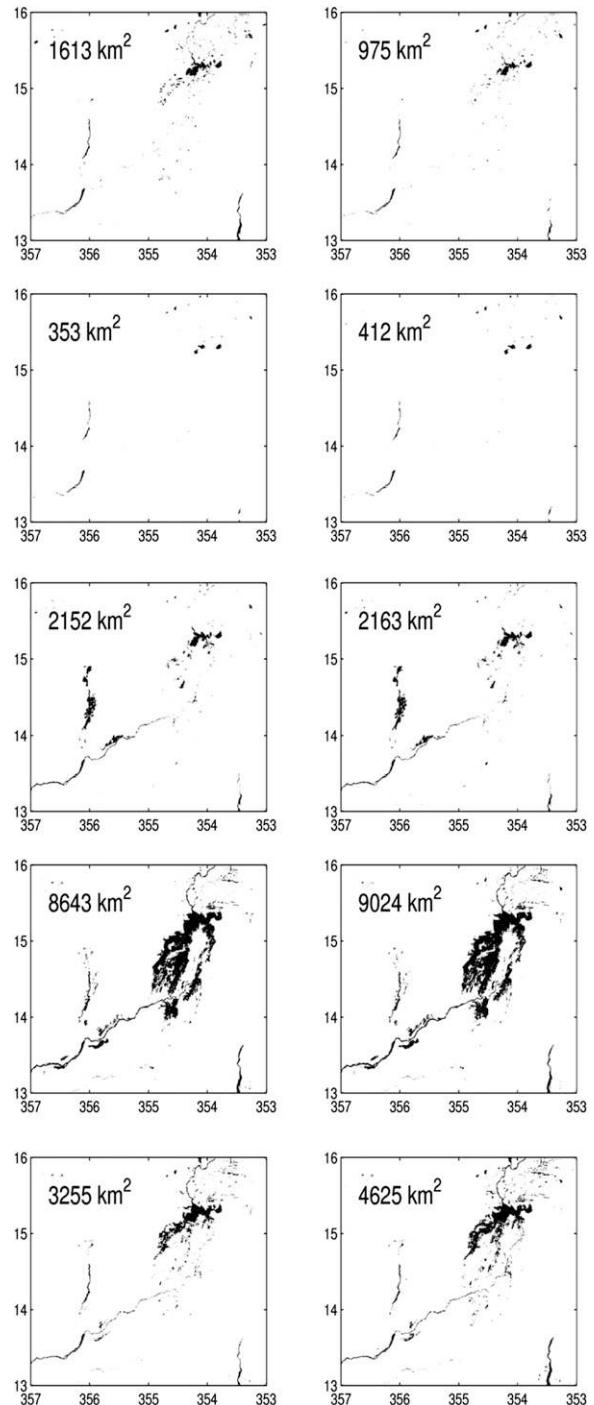


FIG. 14. (left) MODIS inundation and (right) associated down-scaled inundation from GIEMS estimates for (top to bottom) February, June, August, October, and December 2002. The total inundation area is indicated in each map.

To provide a quality measure for the downscaled data, we have upscaled the MODIS high-resolution inundation maps to obtain **LR**<sub>MODIS</sub>. These estimates become the reference state. We then estimate the

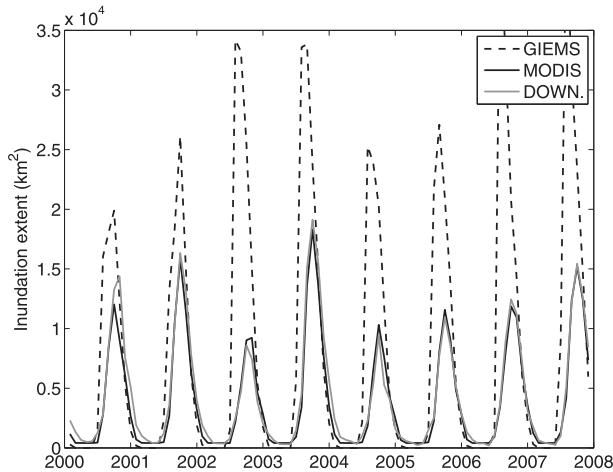


FIG. 15. GIEMS, MODIS, and first downscaled inundation extent time series over the Inner Niger Delta for 2000–08.

spatial correlation with the original GIEMS estimate  $\mathbf{LR}_{\text{GIEMS}}$  (left column in Fig. 17) and with the upscaling of the downscaled GIEMS estimate  $\mathbf{LR}_{\text{DownGIEMS}}$  (right column in Fig. 17). We obtained:  $\text{Cor}(\mathbf{LR}_{\text{MODIS}}, \mathbf{LR}_{\text{GIEMS}}) = 0.86$  and  $\text{Cor}(\mathbf{LR}_{\text{MODIS}}, \mathbf{LR}_{\text{DownGIEMS}}) = 0.92$ . This means that the downscaling process has actually “improved” GIEMS estimates toward the MODIS reference.

### c. Analysis of the downscaled dataset

In this section, the algebraic downscaling method developed in the previous section is used to produce a new, long-term, high-resolution (500 m) inundation dataset. Although the downscaling method has been calibrated using MODIS data, in the operational mode, it uses as inputs only the GIEMS observations. Figure 16 shows well that the downscaling has extended the MODIS period (February 2000 to September 2011) back in time to cover the full GIEMS period (1993–2007). Since the downscaling is only driven, in the operational mode, by GIEMS inputs, its dynamical behavior reproduces well the GIEMS time series and less so the MODIS ones. Its correlations are 0.96 with GIEMS and only 0.69 with MODIS.

Figure 18 represents, for the downscaled dataset, the probability for the pixels to be inundated for each month. This figure can be compared to Fig. 2, which was representing the same quantity from the MODIS dataset. The spatial structure and seasonality are very close. As a consequence, it can be said that the downscaling technique is able to reproduce the GIEMS dynamic and the MODIS spatial structures by obtaining the right compromise between the space and time variabilities.

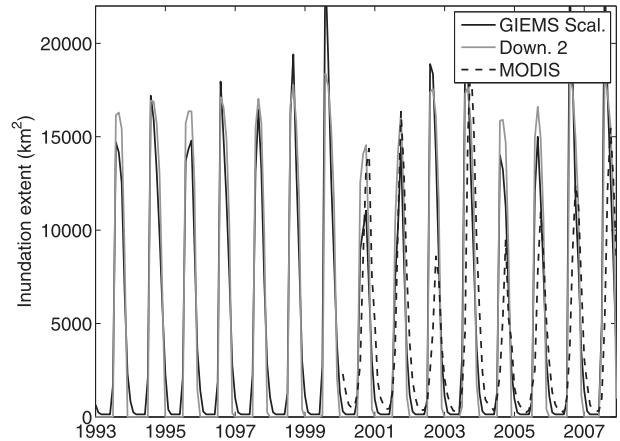


FIG. 16. GIEMS, MODIS, and second downscaled inundation extent time series over the Inner Niger Delta for 1993–2008.

## 5. Conclusions and perspectives

In this paper, an analysis of the space–time variability of the high-resolution inundation extent over the Inner Niger Delta was first conducted using a principal component analysis. This variability can be characterized using a limited number of temporal behaviors related to the season, such as duration or start and end dates of the inundation season. This PCA representation was shown to be an efficient tool to analyze and manipulate the inundation datasets. For example, time interpolation of the whole Inner Niger Delta was shown to be possible, allowing the increase of the temporal resolution of the satellite dataset at a very low computational cost. Spatial interpolation can also be performed using the PCA to compensate for missing values or to reduce retrieval noises in the satellite estimates.

Two innovative downscaling methodologies have been proposed to increase the spatial resolution of the GIEMS inundation estimates (about 25 km) toward the spatial resolution of about 500 m. Both methods are derived from the PCA representation of the space–time variability of the high-resolution inundation extent. The first technique is based on a linear regression and can be used if enough samples are available in both datasets and if the low- and high-resolution estimates are relatively coherent. The second downscaling technique is based on a simple linear algebra approach that solves a linear system. In this approach, the PCA patterns from MODIS need to be relatively coherent with the low-resolution estimates from GIEMS; otherwise, the PCA is nonoptimal and the inversion of the linear system can become ill posed. As a consequence, an inundation area normalization of the GIEMS estimates toward MODIS is required before the downscaling. This algebraic

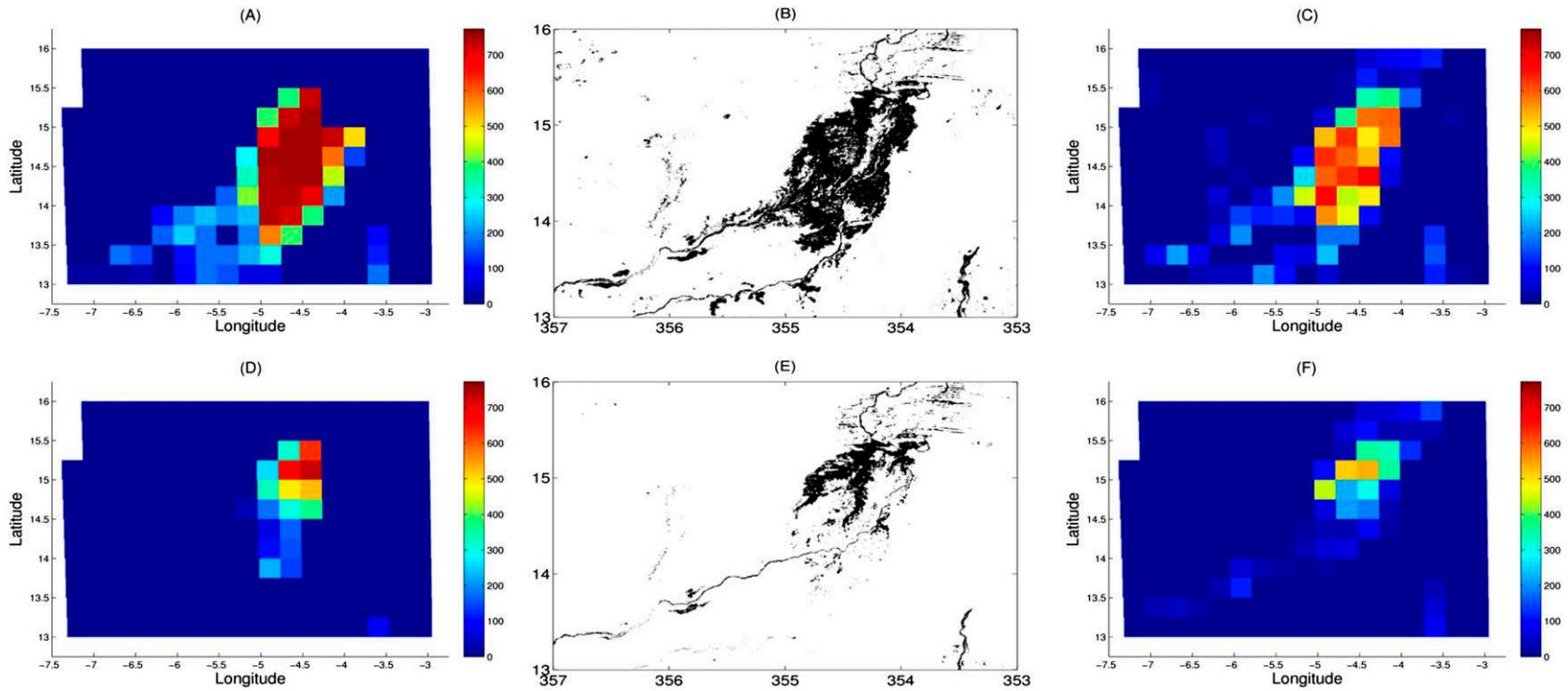


FIG. 17. Two samples of downscaling: (a),(d) original GIEMS low-resolution estimates, (b),(e) downscaling using the second method, and (c),(f) upscaling of the downscaled estimate for (a)–(c) October 2007 and (d)–(f) December 2007. The colored pixels represent the inundated area (km<sup>2</sup>) in the pixel.

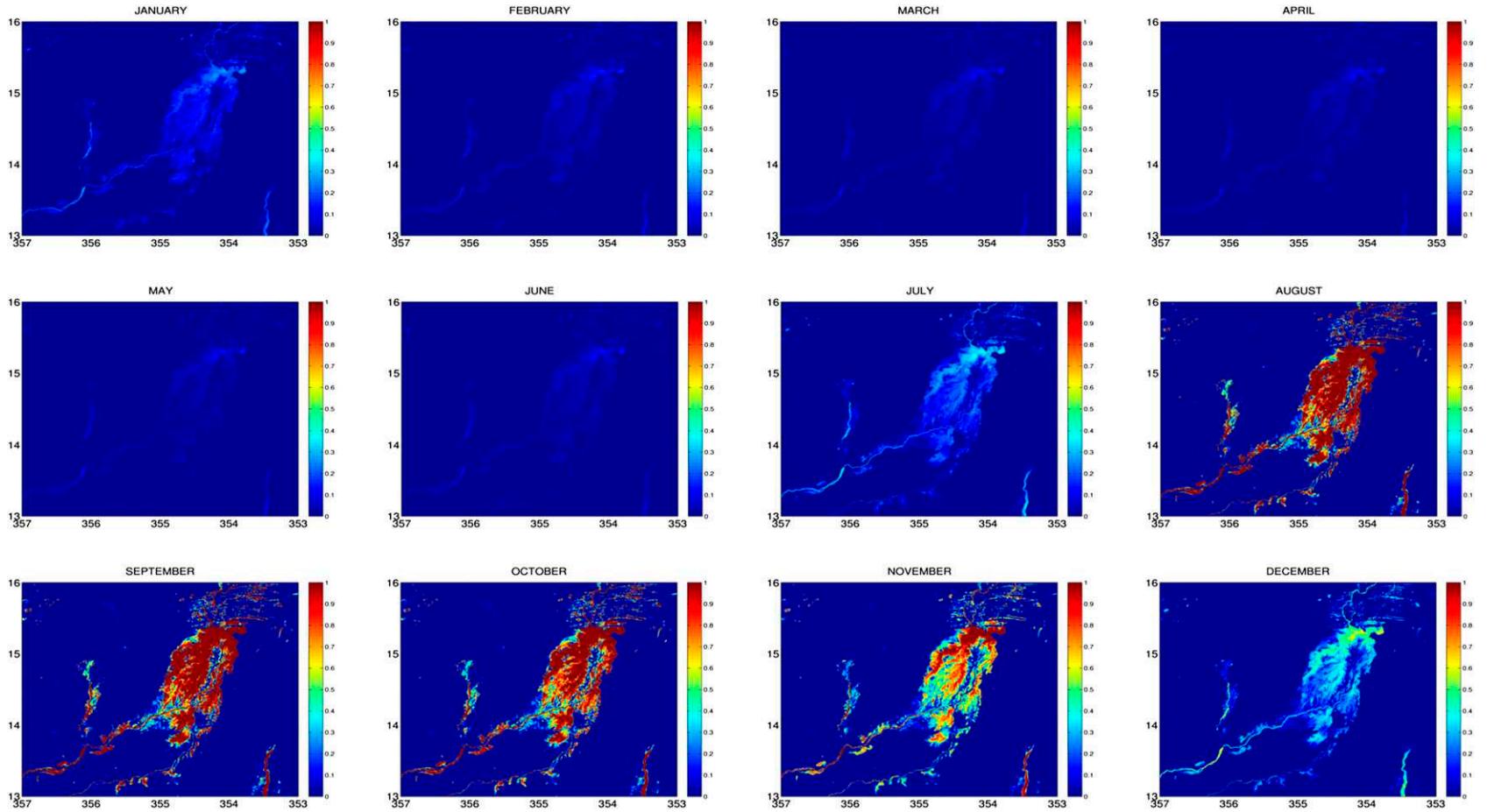


FIG. 18. As in Fig. 2, but for the downscaled inundation dataset at a spatial resolution of 500 m over the period 1993–2007.

downscaling is very general. The results of both downscaling techniques appear to be satisfactory: the spatial and temporal features of the new downscaled inundation data are coherent with MODIS estimates. The choice of one technique or the other should be based on the final use of the downscaled products and on the number of high-resolution samples. This work allowed us to build a new inundation dataset from 1993 to the present, at a 500-m spatial resolution, over the Inner Niger Delta. The PCA components could easily be interpolated spatially to obtain a better spatial resolution, closer to the SWOT resolution.

Additional work will analyze the sensitivity of the downscaling approach to the number of samples, its generalization capacity, and its ability to predict the extreme cases. Other models can be used to replace the linear regression in the first downscaling method. For example, a neural network could be tested: its non-linearity could be beneficial, but the limited number of samples is a real difficulty for the use of this technique in this particular application with the MODIS data. For the assimilation of the wetland information from satellite measurements in land surface models, bridges could be built between the satellite inundation datasets and the model outputs. For instance, the PCA could be performed on model outputs and the satellite observations would constrain these PCA components in an assimilation scheme. Another technical perspective would be to compare these two downscaling techniques with the approaches developed in Aires et al. (2013). Topographic information could also be introduced in the downscaling scheme.

This paper describes a path forward to integrate various remote sensing datasets to more fully describe (temporally and spatially) changes in surface hydrology. Surface inundation extent and their predictability vary strongly with the seasons in many areas. This “rhythm” is being and has been captured by 1) low-spatial-resolution sensors with global coverage at monthly time scales (i.e., GIEMS), 2) high spatial and temporal resolution but with cloud contamination (MODIS), and also 3) in more “snapshot” form by high-resolution sensors (SAR). There is good reason to develop bridging techniques that can exploit the higher spatial resolution back in time with the help of the temporal evolution of the lower resolution. The derived high-spatial-resolution dataset will be used in the framework of the SWOT mission to develop and test the instrument simulator as well as to select the calibration–validation sites. In addition, once SWOT observations will be available, it will be possible to use them to calibrate the downscaling techniques (instead of the MODIS high-resolution dataset). This will allow us to downscale the GIEMS inundation estimates, and these

new, SWOT-compatible inundations will be extended back in time until 1993.<sup>3</sup> The acquisition of data from the SWOT mission can be switched from a high- to a low-resolution mode. If low- and high-resolution SWOT measurements are available in a spatial domain, it is possible to calibrate a downscaling scheme that might be used to obtain high-resolution inundation from the low-resolution acquisitions. The downscaling scheme can also be used to exploit the synergy among various instruments. For example, MODIS observations might be downscaled toward the SWOT resolution, and the fusion of both retrievals should strengthen the robustness of the inundation estimates. This should also facilitate the evaluation of the inundation estimates from the various instruments (GIEMS, SWOT, MODIS, SAR, etc.). Finally, the spatial interpolation tools that we have developed in this paper should help mitigate SWOT retrieval problems due to high precipitations or interferometric layover problems.

*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.” We would also like to thank the three anonymous reviewers who have helped clarify the presentation of the paper.

#### REFERENCES

- Aires, F., A. Chédin, and J.-P. Nadal, 2000: Independent component analysis of multivariate time series: Application to the tropical SST variability. *J. Geophys. Res.*, **105**, 17 437–17 455, doi:10.1029/2000JD900152.
- , W. Rossow, N. Scott, and A. Chédin, 2002: Remote sensing from the infrared atmospheric sounding interferometer instrument: 1. Compression, denoising, first-guess retrieval inversion algorithms. *J. Geophys. Res.*, **107**, 4619, doi:10.1029/2001JD000955.
- , C. Prigent, and W. Rossow, 2004: Temporal interpolation of global surface skin temperature diurnal cycle over land under clear and cloudy conditions. *J. Geophys. Res.*, **109**, D04313, doi:10.1029/2003JD003527.
- , F. Papa, and C. Prigent, 2013: A long-term, high-resolution wetland dataset over the Amazon basin, downscaled from a multiwavelength retrieval using SAR data. *J. Hydrometeorol.*, **14**, 594–607, doi:10.1175/JHM-D-12-093.1.
- Alsford, D., E. Rodriguez, and D. Lettenmaier, 2007: Measuring surface water from space. *Rev. Geophys.*, **45**, RG2002, doi:10.1029/2006RG000197.

<sup>3</sup> Although GIEMS is currently available through the end of 2010, this database will be extended with the more recent satellite observations, which means that a time intersection with SWOT will be available.

- 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 [www.nsidc.org](http://www.nsidc.org).]
- Berge-Nguyen, M., J. Crétau, and S. Calmant, 2008: Combining of radar altimetry and MODIS for the monitoring of flood events: Application to the Inner Niger Delta. *Observing and Forecasting the Ocean: OSTST Meeting*, Nice, France, CNES/GODAE/IDS/Carte-Blanche, SG.12-132. [Available online at <http://meetings.copernicus.org/www.cosis.net/abstracts/EGU2008/08332/EGU2008-A-08332.pdf>.]
- Brakenridge, G., J. Syvitski, I. Overeem, J. Stewart-Moore, and A. Kettner, 2012: Global mapping of storm surges and the assessment of coastal vulnerability. *Nat. Hazards*, **66**, 1295–1312, doi:10.1007/s11069-012-0317-z.
- Bukata, R. P., 1992: *Satellite Monitoring of Inland and Coastal Water Quality: Retrospection, Introspection, Future Directions*. Taylor and Francis, 246 pp.
- Crétau, J.-F., R. Letolle, and S. Calmant, 2009: Investigation on Aral Sea regressions from mirabilite deposits and remote sensing. *Aquat. Geochem.*, **15**, 277–291, doi:10.1007/s10498-008-9051-2.
- , and Coauthors, 2011: Flood mapping inferred from remote sensing data. *Int. Water Technol. J.*, **1**, 48–62. [Available online at <http://iwjtj.info/wp-content/uploads/2011/12/v1-n1-p5.pdf>.]
- del Rio, R. A., J.-F. Crétau, M. Berge-Nguyen, and P. Maisongrande, 2012: Does Lake Titicaca still control the Lake Poopó system water levels? An investigation using satellite altimetry and MODIS data (2000–2009). *Remote Sens. Lett.*, **3**, 707–714, doi:10.1080/01431161.2012.667884.
- Fluet-Chouinard, E., and B. Lehner, 2011: Towards a high-resolution global inundation delineation dataset. *2011 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract H511-1313.
- 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 ground-water storage variations in large drainage basins with extensive floodplains. *Remote Sens. Environ.*, **115**, 1588–1594, doi:10.1016/j.rse.2011.02.003.
- , 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.
- Jolliffe, I., 2002: *Principal Component Analysis*. 2nd ed. Springer, 502 pp.
- Legler, D. M., 1984: Comments on “Empirical orthogonal function-analysis of wind vectors over the tropical Pacific region.” *Bull. Amer. Meteor. Soc.*, **65**, 162.
- Li, R.-R., Y. J. Kaufman, B.-C. Gao, and C. O. Davis, 2003: Remote sensing of suspended sediments and shallow coastal waters. *IEEE Trans. Geosci. Remote Sens.*, **41**, 559–566, doi:10.1109/TGRS.2003.810227.
- Li, S., D. Sun, M. Goldberg, and A. Stefanidis, 2013: Derivation of 30-m-resolution water maps from TERRA/MODIS and SRTM. *Remote Sens. Environ.*, **134**, 417–430, doi:10.1016/j.rse.2013.03.015.
- Papa, F., C. Prigent, C. Jimenez, F. Aires, and W. B. Rossow, 2010: Interannual variability of surface water extent at the global scale, 1993–2004. *J. Geophys. Res.*, **115**, D12111, doi:10.1029/2009JD012674.
- Pedinotti, V., A. Boone, B. Decharme, J. Crétau, N. Mognard, G. Panthou, F. Papa, and B. A. Tanimoun, 2012: Evaluation of the ISBA-TRIP continental hydrologic system over the Niger basin using in situ and satellite derived datasets. *Hydrol. Earth Syst. Sci.*, **16**, 1745–1773, doi:10.5194/hess-16-1745-2012.
- 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, doi:10.1029/2001GL013263.
- , F. Aires, and W. Rossow, 2006: Land surface microwave emissivities over the globe for a decade. *Bull. Amer. Meteor. Soc.*, **87**, 1572–1584, doi:10.1175/BAMS-87-11-1573.
- , F. Papa, F. Aires, W. 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.
- Rodriguez, E., 2012: Surface Water and Ocean Topography Mission (SWOT). Science Requirements Doc., Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 22 pp. [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).]
- Sakamoto, T., N. Nguyen, A. Kotera, H. Ohno, N. Ishitsuka, and M. Yokozawa, 2007: Detecting temporal changes in the extent of annual flooding within the Cambodia and the Vietnamese Mekong delta from MODIS time-series imagery. *Remote Sens. Environ.*, **109**, 295–313, doi:10.1016/j.rse.2007.01.011.
- Tarantola, A., 1987: *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier, 630 pp.
- Wang, Y., J. Colby, and K. Mulcahy, 2002: An efficient method for mapping flood extent in a coastal floodplain using Landsat TM and DEM data. *Int. J. Remote Sens.*, **23**, 3681–3696, doi:10.1080/01431160110114484.