Datasets for the assessment of changes in the incidence, extents, and spatial patterns of inundations in the Cambodian Mekong Delta, based on a water level – flood link calculated from in-situ water levels, and Sentinel-derived inundation maps

Christina Orieschnig, Jean-Philippe Venot, Sylvain Massuel, Khy Eam Eang, Kong Chhuon, Sambo Lun, Sokly Siev, Gilles Belaud

Data in Brief 43 (2022) 108469

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Datasets for the assessment of changes in the incidence, extents, and spatial patterns of inundations in the Cambodian Mekong Delta, based on a water level – flood link calculated from *in-situ* water levels, and Sentinel-derived inundation maps

Christina Orieschnig[a,*], Jean-Philippe Venot[b], Sylvain Massuel[b], Khy Eam Eang[c], Kong Chhuon[c], Sambo Lun[c], Sokly Siev[d], Gilles Belaud[b]

[a] *UMR Hydrosciences, AgroParisTech, University of Montpellier, IRD, 300 Av. du Professeur Emile Jeanbrau, Montpellier 34090, France*
[b] *G-Eau, IRD, AgroParisTech, Institut Agro, CIRAD, INRAE, University of Montpellier, 361 rue J.F. Breton, BP 5095, Cedex 5, Montpellier 34196, France*
[c] *Faculty of Hydrology and Water Resources Engineering, Institute of Technology of Cambodia, Russian Blvd., P.O. Box 86, Phnom Penh, Cambodia*
[d] *General Department of Science, Technology and Innovation, Ministry of Industry, Science, Technology and Innovation, 45 Preah Norodom Boulevard, Sangkat Phsar Thmey III, Khan Daun Penh, Phnom Penh 120203 Cambodia*

ARTICLE INFO

Article history:
Received 19 May 2022
Revised 1 July 2022
Accepted 8 July 2022
Available online 13 July 2022

Keywords:
Mekong Delta
Sentinel-1 and 2
Water levels
Monsoon inundations
Flood frequency

ABSTRACT

This brief contains the data needed to calculate and assess the robustness of a water level – flood link (WAFL) in the Cambodian Mekong Delta, which was used to analyze changes in the long-term behavior of Monsoon inundations in the region. The data comprises the WAFL raster (.tif) files for two zones in the delta. Zone A is located on the right bank of the Bassac River, a distributary of the Mekong. Zone B is bracketed between the Mekong River and the Bassac River. The WAFL was calculated by linking water levels mea-

DOI of original article: 10.1016/j.jhydrol.2022.127902
* Corresponding author.
E-mail address: christina.orieschnig@ird.fr (C. Orieschnig).

https://doi.org/10.1016/j.dib.2022.108469

sured by the Mekong River Commission (MRC) at the hydrological station in Koh Khel, with inundation maps derived from Sentinel-1 and -2 images taken between 2017 and 2021. The final WAFL raster files provides a basis for estimating inundation extents using *in-situ* water levels.
Furthermore, this brief includes data used for the assessment of WAFL, including *in-situ* water level data and the extents of natural vegetation in the case study area in 1990, 2000, 2010 and 2020. The former was collected using a differential pressure logger. The latter was calculated from historical Landsat image composites.
Finally, raster files representing the incidence and duration of inundations in the case study area before and after the year 2008 are provided. These were calculated based on the WAFL and the MRC water levels. For each area, before- and after-images are available, as well as a raster representing the change between the two.
To simplify visualization and geographical location, shapefiles (.shp) of the study area and the location of the *in-situ* logger are also provided.

Specifications Table

Subject	Hydrology and water quality
Specific subject area	floodplain hydrology, inundation mapping based on remote sensing using Sentinel-1 and −2, linking *in-situ* information with remotely sensed data
Type of data	Table Image Shapefile
How the data were acquired	The *in-situ* data was acquired using a differential pressure logger (Onset Hobo MX2001) installed in a canal in the floodplain of the Cambodian Mekong Delta. The WAFL data was calculated on the basis of hydrological data provided by the Mekong River Commission (MRC) and inundation maps derived from Sentinel-1 (synthetic aperture radar) and −2 (optical) imagery through an unsupervised k-means classification on the open-access cloud computing platform Google Earth Engine. Furthermore, the tables listing the daily inundation extents (in %) in the case study area were also derived from the analysis. Missing values were interpolated and a rolling mean was used to eliminate fluctuations based on classification uncertainty. The delineation of natural vegetation (forest and shrubland) was conducted on the basis of historical Landsat-5 and −8 images, composited for the dry season (March-May) in 1990, 2000 ,2010, and 2020. The Normalised Difference Vegetation Index (NDVI) was calculated for each image, and pixels for which it lay over the threshold of 0.55 were identified. The calculation of inundation incidence and duration was conducted using the WAFL raster files and hydrological data from the MRC. Incidence was calculated for the overall year, as well as for dates significant to local agricultural practices: July 15, July 31, August 15, August 31 and September 25. The code for the calculations is available here: Flood detection: https://code.earthengine.google.com/646cd0fefcd8a93e6001833eb36991c6 Detection of natural vegetation: https://code.earthengine.google.com/8dc84bddf603f3d75ba2e28a8f7d68dc WAFL construction and analysis: https://github.com/christina-orieschnig/WAFL_Github_Files

(*continued on next page*)

Data format	Raw Analyzed
Description of data collection	*In-situ* data was collected by a differential pressure logger (Onset Hobo MX2001), which was installed in a canal in the floodplains of the Cambodian Mekong Delta in June 2020. Read-outs are carried out periodically, manually, in the field. Analyses of remote sensing imagery were carried out on the open-access cloud computing platform Google Earth Engine. Statistical analyses were carried out in Python.
Data source location	**Primary data source for the logger data:** • Institution: Institut de Recherche pour le Développement (IRD, Cambodian Team) and Institute of Technology of Cambodia (ITC) • City/Town/Region: Phnom Penh • Country: Cambodia • Latitude and longitude (11.062326,105.033056) **Raw data sources for the WAFL and natural vegetation calculation:**Landsat-5 and 8 images by the North American Space Agency (NASA), available via Google Earth Engine's Data Catalog (https://developers.google.com/earth-engine/datasets/catalog/landsat) Sentinel-1 and -2 images by the European Space Agency (ESA), also available via Google Earth Engine's Data Catalog (https://developers.google.com/earth-engine/datasets/catalog/sentinel-2/) Water level data at Koh Khel hydrological station was provided through the Mekong River Commission (MRC) data portal (https://portal.mrcmekong.org/time-series/chart?ts=64758ae8f24a46759d962ab9a2dea27a&sd=1990-09-05&ed=2022-02-28)
Data accessibility	Orieschnig, Christina; Venot, Jean-Philippe; Massuel, Sylvain; Eang, Khy Eam; Chhuon, Kong; Lun, Sambo; Siev, Sokly; Belaud, Gilles (2022), "Datasets for the assessment of changes in the incidence, extents, and spatial patterns of inundations in the Cambodian Mekong Delta, based on a water level - flood link calculated from *in-situ* water levels, and Sentinel-derived inundation maps", Mendeley Data, V1, doi: 10.17632/rjn5r4spkp.2 URL:https://data.mendeley.com/datasets/rjn5r4spkp/2
Related research article	Orieschnig, C., Venot, J.P., Massuel, S., Eang, K.E., Chhuon, K., Lun, S., Siev, S. and Belaud, G., 2022. A Multi-Method Approach to Flood Mapping: Reconstructing Inundation Changes in the Cambodian Upper Mekong Delta. *Journal of Hydrology, 610*, p.127902. https://doi.org/10.1016/j.jhydrol.2022.127902

Value of the Data

- Researchers can use the data provided here to estimate and analyze inundation extents in the Cambodian Mekong River Delta based on locally measured, publicly available water level data. They can harness the WAFL to retro- and forecast inundation extents by using a simple thresholding approach and the raster files.
- Furthermore, anyone interested in the hydrological dynamics of the Cambodian Mekong River Delta can make use of the data. No further analysis is necessary and the WAFL files can be used as-is, with input from publicly available water level data. This means that they are accessible for non-academic stakeholders as well.
- In addition, anyone wishing to implement the WAFL methodology in similar areas can benefit from the provided data as a point of reference. The methodology itself is transferable in nature and scientists working in other (tropical) deltas could use it to estimate inundation areas in their zones of interest as well. In this case, the data provided here can serve as a basis for their own analyses as they provide insights into both intermediate steps and the final result.
- Further studies can use the data as a point of reference in analysing the relationship between local changes in inundation patterns and factors such as upstream hydropower development and climate change.

1. Data Description

Generally, the terminology of the files is self-explanatory, with the A suffix referring to the case study sub-zone A and the B suffix referring to the case study sub-zone B. Shapefiles showing their geographical delineation are provided in the sub-folder Study_Area_Delineation. Extensive descriptions of their terrain and land cover characteristics are given in [1]. These shape files were delineated manually and encompass a total area of approximately 500 km2.

The two raster (.tif) files in the sub-folder WAFL_Rasters provide the calculated **Wa**ter Level - **F**lood **L**ink (WAFL). The value of each pixel (between 0 and 10) gives the water level in m asl measured at the Koh Khel hydrological station maintained by the Mekong River Commission at which the pixel in question becomes flooded. This is the result of linking water levels measured at the hydrological station at Koh Khel with inundation extents extracted via k-means clustering from Sentinel-1 and −2 imagery using the cloud computing platform Google Earth Engine and a correlation model. Details on the methodology are provided in [1].

The folder Nautural_Vegetation contains four sets of raster files that give the extent of natural vegetation (forest and shrubland) in the case study area in 1990, 2000, 2010, and 2020. They were calculated using dry-season composites of historical Landsat imagery on Google Earth Engine. The Normalised Difference Vegetation Index (NDVI) was calculated and a threshold of 0.55 was set to delineate forests and shrubland. A value of 1 means that the pixel in question contains natural vegetation. A value of 0 means that it does not.

The sub-folder Inundation_Extents provides tables that give the daily inundation extents (in % of the sub-zone) calculated from Sentinel-1 and −2 between 2017 and 2020. This was accomplished by calculating a daily mean water level at the station at Koh Khel and then thresholding the WAFL raster so that all pixels with values below that water level were set to inundated.

The folder *In-Situ*_Data provides a .csv file with data collected using a differential pressure logger (Onset Hobo MX2001) installed in the case study area – the daily averaged water levels in m asl.

Finally, the folder Flood_Frequency_Duration_Results_Raster includes the full set of raster files generated using the WAFL and locally measured water levels. They represent the average inundation incidence (at different dates) and duration pre- and post-2008. Each pixel value gives the fractional increase or decrease in the average incidence of inundations before and after that year, a well-known hydrological break point in the region.

2. Experimental Design, Materials and Methods

2.1. In-Situ *Water Level Measurements*

The water level logger was installed at the bank of Prek Chann, a channel in the floodplain of the Bassac River. The pressure probe for atmospheric pressure and the data logger itself were installed in a PVC shaft (diameter 100 mm) and encased in cement for stability. The pressure probe to measure the water level was installed at the deepest point of the channel's bed and secured using a steel bar and clamp. It is connected to the logger via a data cable protected by a PVC pipe (diameter 35 mm). The logger measures water levels through differential pressure every 15 min. Read-outs are conducted manually.

2.2. Detection of Inundation Extents

The detection of inundations was carried out using synthetic aperture radar (SAR) Sentinel-1 and optical Sentinel-2 imagery on the cloud computing platform Google Earth Engine (GEE, [2]).

Sentinel-1 images were subjected to orthorectification, border noise removal, radiometric calibration, thermal noise removal, and a refined Lee speckle filter [3]. Sentinel-2 images underwent

atmospheric correction and cloud masking. In addition, three spectral indices were calculated for these optical images: the Normalised Difference Vegetation Index (NDVI, [4]), the Normalised Difference Water Index (NDWI, [5]), and the Normalised Difference Moisture Index (NDMI, [6]).

These indices, as well as the VH and VV bands of the SAR images served as inputs for the classification. Said classification itself was carried out using an unsupervised k-means clustering algorithm with two clusters. Composite images dating to September 2018 were chosen as training data for the classifier as the considerable flood extents at the time showed a sufficient contrast for the binary division.

The accuracy of the classification was tested using 2340 manually set points, determined on the basis of SPOT-6 images. A confusion matrix and kappa coefficient were calculated and the accuracy found to be sufficient at 96.5%, with a kappa value of 0.785.

Overall, 360 Sentinel-1 and 119 Sentinel-2 from the years 2017–2021 were processed. In several cases, daily composites had to be created since the case study area is located at the overlap of different Sentinel tiles.

Once the classification was concluded, two actions were performed. First, the inundation extents for both sub-zones of the study area were calculated for each available image. Second, all images were stacked and a count created for each image related to how often it was detected as flooded.

The detected inundation extents were exported and down-scaled to a daily resolution using linear interpolation. Furthermore, a rolling mean was applied to reduce the effect of any remaining classification uncertainty.

2.3. Calculation of WAFL

The water level - flood link (WAFL) was calculated on the basis of the stack of classified Sentinel-1 and 2 images, as well as water level data provided by the Mekong River Commission. Each image was linked to a water level, taking into account the temporal delay between a water level being measured and said level unfolding its full effect in the downstream floodplains. To achieve this, the Lamagat correlation model (cf. [7]) was employed, which merges statistical methods and kinematic elements to pinpoint optimal correlation factors between two time series in a set number of sub-sections of these series (also cf. [1]). In this way, the lowest water level at which each individual pixel in the case study area becomes flooded could be identified, as well as the typical temporal delay depending on the water level.

2.4. Detection of Natural Vegetation

To identify areas of natural vegetation, meaning shrubland and regularly flooded forests in the context of the Mekong River Delta, Landsat-5 and −8 images on GEE were employed. Both Landsat-5 and −8 images underwent orthorectification for several bands, and both were atmospherically corrected - using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) for Landsat 5 and the Land Surface Reflectance Code (LaSRC) for Landsat-8. Following this, a cloud filter and a cloud mask were applied and cloud-free scenes were collected for the dry season months between March and May in 1990,2000, 2010, and 2020, when vegetation would normally be at a low point in the case study area. Then, composites were calculated and gaps left behind by masked clouds were filled by stacking all available scenes in each year and calculating a mean value for each pixel. Subsequently, the NDVI was calculated and a threshold of 0.55 (cf. [8]) was applied to identify likely patches of natural vegetation (marked as 1 in the results).

2.5. Calculation of Flood Frequency and Duration

The calculation of flood frequency and duration, finally, was carried out on the basis of the WAFL raster files the creation of which is outlined above, and the river water levels measured by the MRC. It should be noted that "frequency" is here defined as being synonymous with "incidence", referring to the average number of years a particular pixel is reached by the annual monsoon flood. This was calculated for the overall year, as well as for particular dates, namely July 15, July 31, August 15, August 31, and September 25. These dates carry particular significance for the agricultural harvesting and planting cycles in the region (cf. [1]).

In general, two periods were compared: 1991–2008 and 2009–2020. The reason for this is that the locally measured hydrological series provided by the MRC begins in 1991. 2008 is known to be a hydrological break point in the region (cf. [1,9,10]). For each period, the average flood frequency for each date and the total flood duration were calculated for each pixel and a raster file generated on this basis. As a final step, the values of the two raster files were subtracted to generate a file that shows the change in inundation incidence and duration between the two periods. Inundation incidence is given as a fraction of years (dimensionless, number of years in which a pixel is flooded by a certain date divided by number of years in the time series). Inundation duration is given in days.

The code for the calculations is available here:

Flood detection: https://code.earthengine.google.com/646cd0fefcd8a93e6001833eb36991c6
Detection of natural vegetation:
https://code.earthengine.google.com/8dc84bddf603f3d75ba2e28a8f7d68dc
WAFL construction and analysis: https://github.com/christina-orieschnig/WAFL_Github_Files

Ethics Statements

All the data available in the repository may be used and reproduced freely as it is either the result of analyses or original data collection. The WAFL raster fies are based on analyses of Sentinel images by the European Space Agency and the Mekong River Commission, with raw data available via the Copernicus Data Hub (https://scihub.copernicus.eu) and the MRC Data Portal (https://portal.mrcmekong.org/home). The original Landsat images used for the calculation of vegetated areas are freely available via the USGS' Earth Explorer portal (https://earthexplorer.usgs.gov).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Datasets for the assessment of changes in the incidence, extents, and spatial patterns of inundations in the Cambodian Mekong Delta, based on a water level - flood link calculated from *in-situ* water l (Original data) (Mendeley Data).

CRediT Author Statement

Christina Orieschnig: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization; **Jean-Philippe Venot:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition; **Sylvain Massuel:** Conceptualization, Investigation, Resources, Writing – review

& editing, Project administration, Funding acquisition; **Khy Eam Eang:** Investigation, Resources; **Kong Chhuon:** Investigation, Resources; **Sambo Lun:** Investigation, Resources; **Sokly Siev:** Investigation, Resources; **Gilles Belaud:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Acknowledgment

The authors thank the French Development Agency (AFD) for its support through the COSTEA project, the Wat-Health project (FSPI-2021), and our esteemed colleagues at the Institute of Technology of Cambodia as well as the Royal University of Agriculture for their invaluable partnership and support during field work.

Referances

[1] C. Orieschnig, J.P. Venot, S. Massuel, K.E. Eang, K. Chhuon, S. Lun, S. Siev, G. Belaud, A multi-method approach to flood mapping: reconstructing inundation changes in the Cambodian upper Mekong delta, J. Hydrol. 610 (2022) 127902 (Amst), doi:10.1016/j.jhydrol.2022.127902.

[2] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, R. Moore, Google Earth Engine: planetary-scale geospatial analysis for everyone, Remote Sens Environ. 202 (2017) 18–27, doi:10.1016/J.RSE.2017.06.031.

[3] Lee, J.S. (1983). Digital image smoothing and the sigma filter. Comput. Vis. Graph. Image Process., 24(2), pp. 255–269. doi:10.1109/TSMC.1983.6313036.

[4] J.W. Rouse, R.H. Haas, J.A. Schell, D.W Deering, in: Proceedings of the Third Earth Resources Technology Satellite-1 Symposium: Section AB. Technical presentations, 1, Scientific and Technical Information Office, National Aeronautics and Space Administration, 1974, p. 309.

[5] S.K. McFeeters, The use of the normalized difference water index (NDWI) in the delineation of open water features, Int. J. Remote Sens. 17 (7) (1996) 1425–1432, doi:10.1080/01431169608948714.

[6] E.H. Wilson, S.A. Sader, Detection of forest harvest type using multiple dates of Landsat TM imagery, Remote Sens. Environ. 80 (3) (2002) 385–396, doi:10.1016/S0034-4257(01)00318-2.

[7] G. Belaud, L. Cassan, J. Bader, N. Bercher, T. Feret, Calibration of a propagation model in large river using satellite altimetry, in: Proceedings of the 6th International Symposium on Environmental Hydraulics, 2010, pp. 23–25, doi:10.1201/b10553-143.

[8] B. Nath, S. Acharjee, Forest cover change detection using normalized difference vegetation index (NDVI): a study of reingkhyongkine lake's adjoining areas, Indian Cartogr. 33 (2) (2013) 348–403.

[9] S. Heng, A. Horton, P. Hok, S. Chung, J. Koponen, M. Kummu, The Cambodian Mekong floodplain under the future development plans and climate change, Nat. Hazards Earth Syst. Sci. Discuss. (2021) 1–25, doi:10.5194/nhess-2021-65.

[10] X. Yun, Q. Tang, J. Wang, X. Liu, Y. Zhang, H. Lu, Y. Wang, L. Zhang, D Chen, Impacts of climate change and reservoir operation on streamflow and flood characteristics in the Lancang-Mekong River Basin, J. Hydrol. 590 (2020) 125472 (Amst), doi:10.1016/J.JHYDROL.2020.125472.