Monitoring of Inland Water Levels by Satellite Altimetry and Deep Learning

Fernando Niño, Clement Coggiola, Denis Blumstein, Lea Lasson, Stephane Calmant

Monitoring of Inland Water Levels by Satellite Altimetry and Deep Learning

Fernando Niño, Clément Coggiola, Denis Blumstein, Léa Lasson and Stéphane Calmant

***Abstract*— Deep convolutional neural networks have proven their efficiency for image processing and are routinely used for image classification. In the present study, we use them to convert radar measurements into water distance and ultimately into water levels of inland waterbodies. The measurements used are the successive echoes of the spaceborne radar altimeter signal on a waterbody, the radargram. We show that by using forward modelling with an accurate altimetry simulator, we can generate a sufficient amount of radargrams and train a deep neural network accurately enough to obtain water level series from radargrams in a hydrology context. The method is validated at selected waterbodies by comparing these water level time series with those obtained by classical, largely manual, methods and also with in situ measurements provided by river gauges. Comparisons at crossovers of the orbits with close temporal colocations are also made. The validation shows that this automatic method performs generally as well as a carefully tuned manual method for removing outliers from the ranges provided by the state of the art classical retrackers used by the spatial hydrology community. This new tool is a big step towards a generic, global, and automated method to retrieve inland water levels from altimetry measurements. This goal is especially important in the context of continuously declining number of in situ measurements, and of utumost importance for adequate water resources management at the global scale.**

***Index Terms*—satellite altimetry, radar, hydrology, neural network, deep learning, artificial intelligence, Jason-3.**

I. Introduction

INLAND waters are an essential part of the water, carbon and energy global cycles. River and lakes water levels are Essential Climate Variables (ECV) as defined by the Global Climate Observing System (GCOS). From a societal point of view, surface waters are the most important source of water for domestic, agricultural and industrial usages. Furthermore, water resources over the continents have been and will continue to be affected by climate change and anthropogenic impact [1], [2]. The past three decades were exceptional in flooding in Europe, when compared to the last 500 years [3]. In many countries, not only floods but also droughts are having a strong impact on human societies. Measuring water levels on a global scale is of utmost importance for adequate water management.

Traditionally, water level observations are performed by *in situ* gauge stations whose continuous operation is costly and not always possible in countries of political and economical unrest. The number of publicly available water level temporal series has dwindled in the last decades, making databases of water levels sparse and very heterogeneous in space and time [4], [5]. By virtue of their global coverage and continuous operation, spaceborne radar altimeters can mitigate the lack of *in situ* data [6]–[8]. Indeed, these altimeters are designed to measure the distance (or *range*) to the surface beneath the satellite. Because the satellite position is known very precisely, when it flies above the ocean, a river or a lake, one can obtain the water height by subtracting the altimeter range from the satellite altitude (and other terms to account for propagation delays and Earth deformation, which will not be discussed here). To make possible the monitoring of geophysical quantities, altimetry satellites are placed on repeat-track orbits; this means that the radar altimeter will pass at regular intervals over a waterbody and can provide time series of water levels with a sampling period equal to the cycle duration of the orbit (ca. 10 days for the Jason-3 satellite).

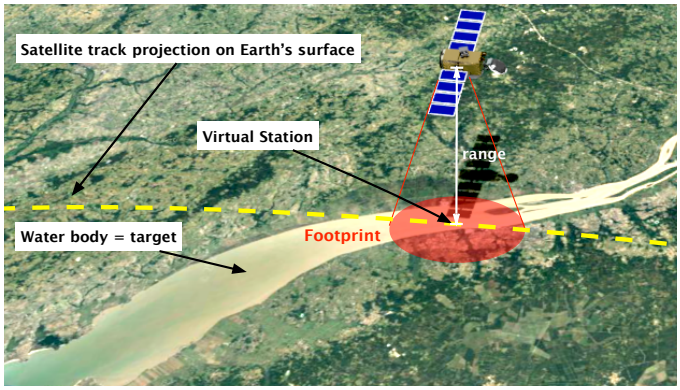

Fig. 1. Figure showing the terminology used: the hydrology target is the water body, the virtual station being the intersection of the waterbody and the satellite track projection. The distance between the satellite altimeter and the target (waterbody) is the range. The radar altimeter echo does not bounce off a single point, but integrates the echos of a region, called the footprint.

In the following text, we will call *hydrology targets* the intersections between satellite groundtracks and waterbody centerlines (Figure 1). By analogy with the *in situ* stations, hydrology targets will be called *virtual stations* when it is possible to extract good quality temporal series of water levels from the altimetry measurements.

As formulated above, obtaining temporal series of water levels from space seems straightforward; however, for over 20 years different research teams have been working on this problem, and the number of available virtual stations is only a small fraction (less than 10%) of the available number of

Manuscript received ...; revised ... This research was supported by CNES in the framework of the VOLODIA project.

F. Niño, C. Coggiola, D. Blumstein, S. Calmant and L. Lasson are with LEGOS, Université de Toulouse, CNES, CNRS, IRD, UPS. 18 avenue E. Belin - 31400 Toulouse, France

TABLE I
HYDROLOGY TARGETS FOR ALTIMETRY MISSIONS USING THE DEM TRACKING MODE

Mission	Hydrology Targets
Jason 3	33445
Sentinel 3A	74051
Sentinel 3B	73630
Total	**181126**

TABLE II
VIRTUAL STATIONS AVAILABLE IN ALTIMETRIC HYDROLOGY DATABASES

Database Name	Virtual Stations	Target %
HydroWeb	12966	7.2
DAHITI	2894	1.6
Hydrosat	1714	1
Total	**17816**	

hydrology targets. There are at least two explanations for this low ratio [9]:

1) radar altimeters, which were designed primarily to measure the oceans, can have difficulties to keep the radar echo on the acquisition window (or to have correct *tracking*) when faced with the rapid change of radiometry and topography over lands [10],
2) the validation of altimetry measurements over inland waters requires a great deal of expertise to remove the frequent outliers.

The first problem has been tackled in the last years by the operational use of the new onboard DEM tracking mode implemented on recent altimeters (Poseidon-3 on Jason-3, SRAL on Sentinel-3) [11], [12]; for past measurements, wrong onboard retracking means the radar echoes are unfortunately lost and for these cases, the retrieval of altimeter range is impossible.

The methods proposed in this article aim at solving the second problem by providing automated methods to convert altimetry measurements into range over inland waters. This automation is even more important when confronting the task of processing the increasing number of altimetry measurements in virtual stations made available by the onboard DEM tracking mode (compare the availability of virtual stations of table II with the number of hydrological targets attainable for missions using the DEM tracking mode in table I). In this paper we will focus on the Jason-3 mission because of the availability of this DEM tracking mode, but the methodology can be applied to other missions.

Altimeters do not measure directly a range value at each point, but the echo of a radar signal (or *waveform*) from which the range value must be retrieved. Radar echoes are processed onboard and transmitted to the ground at a relatively high frequency (20 Hz for Jason-3) which is equivalent to a spatial sampling every few hundreds of meters along-track. If we juxtapose consecutive altimetry waveforms along the time axis (or equivalently, by latitude), this waveform sequence can be viewed as a small grayscale or pseudocolor image of reflected power. These images are called *radargrams*. In figure 2, we show various representations of radar altimetry echoes for data

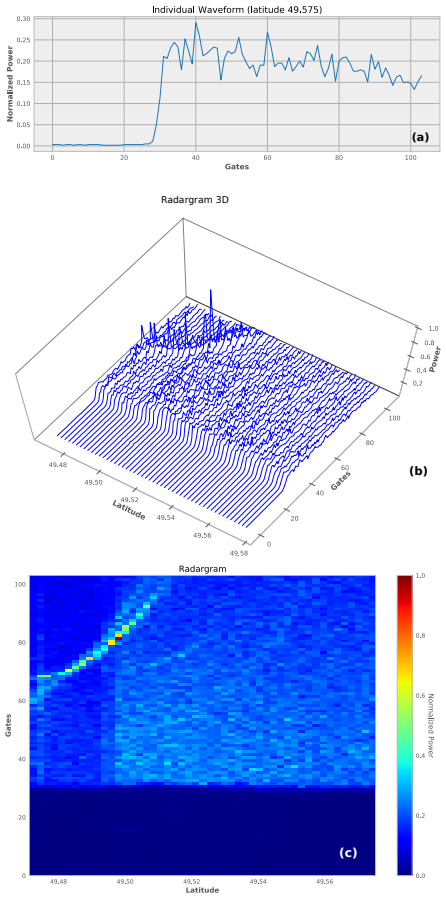

Fig. 2. Radar altimetry echoes using data of pass number 146 of Jason-3 on the North Atlantic on 2016/04/02 (see text for more information). (a) a view of one waveform, showing the evolution of measured signal power through time, as sampled in the so-called gates. (b) a 3D view of a sequence of echoes, with axes latitude, gates and power. (c) the radargram, which is a view of figure b from above, where power values (heights) are color coded.

corresponding to pass number 146 of Jason-3 on the North Atlantic on 2016/04/02, just offshore Normandy in the north of France. Far from the shore, one radar echo (figure 2a) behaves as expected from the Brown model, as discussed in Section II. This is a time vs returned-power plot, where the time axis has been converted to the equivalent gate samples. Figure 2b shows the 3D view of a radargram and figure 2c a pseudo-color plot of the radargram. For this particular case, the sand on the nearby beach appear as a bright parabola segment on the radargram, just as a river would stand out on a low reflectance landscape.

With some expertise, it is easy to recognize in the radargrams the signature of different objects (e.g. the parabola in figure 2). This expertise being essentially visual, we have used supervised learning to train a deep convolutional neural network [13], [14] to learn the function that identifies the correct value of radar range (and hence, the water level) that is to be associated to a waterbody, by analyzing its radargram (more details in the next section).

The rest of the paper is structured as follows: in Section II a review of radar altimetry waveform properties is presented. We also highlight the differences between ocean and inland

water processing for radar range retrieval, and how a radargram is used to compute the water height of a waterbody. Section III presents a summary of the alternative strategies put forward for retrieving the altimeter range of waterbodies, Section IV describes the methodology used to construct the neural network. Section V shows the experimental results and their validation with respect to in-situ data. We close this article with a discussion on neural networks for hydrology processing, possible follow-ups and conclusions.

II. Waveforms in Satellite Altimetry

As briefly presented in the introduction, the measurements in radar altimetry are radar echoes. Satellite altimetry was designed for measuring the sea surface. The Brown model [15] is a theoretical model of the response of the radar signal to an ocean-like surface (a surface whose heights follow Gaussian statistics). A simplified schema of the Brown model, is shown in figure 3. By fitting the observed waveform to this theoretical model, one can estimate several parameters (altimeter range, backscatter coefficient, satellite mispointing) for oceanography applications. Of these, we will only be concerned with the altimeter range.

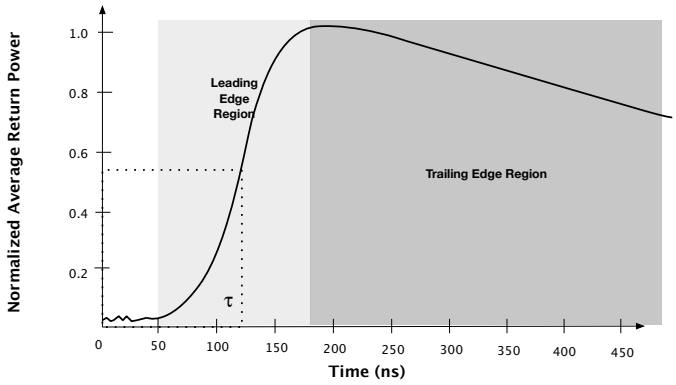

Fig. 3. A simplified schema of the Brown model for ocean waveforms. The midpoint of the leading edge of the waveform (epoch, or τ) corresponds to the range of the altimeter for this echo.

For waveforms of a form such as that of figure 3, the altimeter range will be obtained by the expression:

$$\text{range} = r_{\text{tracker}} + (\tau - \tau_{\text{ref}})\delta r \tag{1}$$

where r_{tracker} is the distance between the reference ellipsoid and a fixed reference point on the sampling window, τ_{ref} (gate 32 for the Jason-3 altimeter). The epoch of the waveform, τ is the time associated to water height (the midpoint of the leading edge in Figure 3). The conversion of gates to distance is made by factor δr which is the radar distance resolution (ca. 47 cm for Jason-3).

As shown in Figure 1, the radar echo is not the result of the interaction of the radar signal with a point on the surface, but with a *footprint* region. This footprint is just the region located at such a distance that the radar signal reflected by the ground reaches the receiving antenna in the expected time window to contribute to the waveform. In the case of the ocean, this footprint is approximately a circle. Ocean water height varies

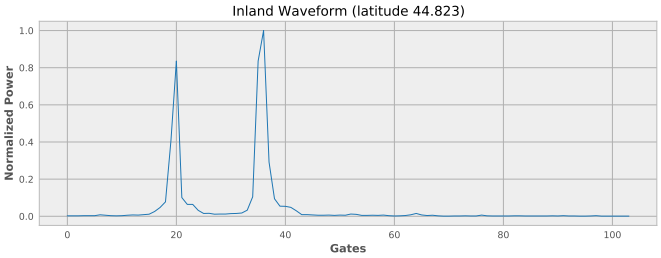

Fig. 4. An example of inland waveform not correctly described by the Brown model (cycle 18, pass 35 of Jason-3 overpass of the Dordogne river in South Western France,latitude 44.823, longitude 0.134). Two peaks show the ambiguity of deciding which peak corresponds to what waterbody below.

with approximately random Gaussian statistics, where the Brown model applies very well. For the case of inland water this is clearly not the case; land embedding the waterbody is generally not described by random heights and thus the Brown model does not apply. Moreover, a radar waveform of the ocean type fits easily to the Brown model and presents no ambiguity to define a range: there is a clear maximum that separates a leading edge and a trailing edge. By contrast, as shown in Figure 4, multiple peaks can appear on an inland waveform, which may correspond to different waterbodies (ponds or swamps for instance), and not necessarily the main hydrological target. However, if we look at the corresponding radargram of Figure 5 we see that high reflectance bodies (e.g. water on a low reflectance background such as cropland) appear as parabolas on the radargrams. A simple geometrical calculation shows that a bright object on the ground is visible on several waveforms and the loci of the maximum of these waveforms in the radargram follow a parabola (or the branch of a hyperbola, strictly speaking).

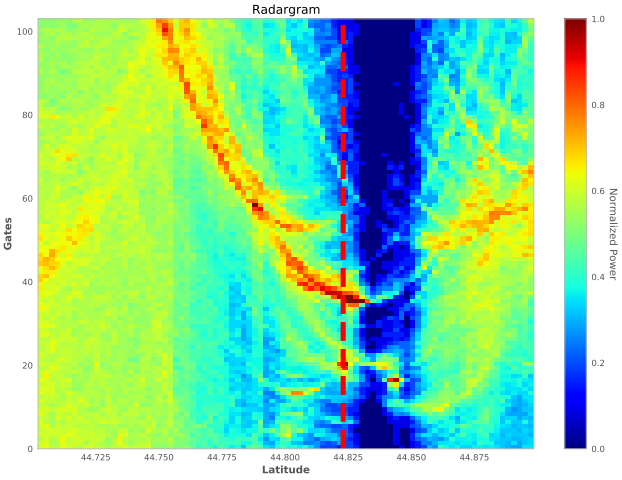

Fig. 5. An inland waters radargram (to make features stand out, we plot here the logarithm of the echo signal). The position of the waveform of figure 5 is shown with a dotted red line. The parabolas appear as a consequence of the high reflectance of waterbodies, which appear on different waveforms as seen at different angles and hence, distance).

The main insight is thus that bright waterbodies affect several successive waveforms in the neighborhood of the one corresponding to nadir, and that the estimate of their height

and location can benefit of a radargram analysis, instead of relying only on individual waveforms.

Efforts have been made to automate this process, and our hydrology group has been using semi-automated software tools for over a decade: VALS [16], MAPS [17], and more recently ALTIS [18] in which the operator must manually select the measurements to take into account when building a time series, eventually taking into account the slant ranges.

When the number of available hydrological targets is in the hundreds of thousands (since this analyses must be made for every measurement over a particular hydrological target), this approach loses its appeal, and further automation is required. Radargram parabola analysis can be automated, but our tests with this approach showed that only ca. 30 % of the cases can be treated algorithmically, many fuzzy cases remaining intractable.

Rather than complexifying the algorithmic approach to take into account more specific cases, we propose a machine learning approach in which a neural network is able to return the range of the main waterbody given its radargram for the general case of hydrology targets.

III. Related work

For hydrology applications, altimetry has the appeal of global coverage, insensitivity to cloud cover, and good accuracy. Many approaches have been taken to exploit these advantages for obtaining hydrological time-series from altimetry waveforms in continental surfaces.

As mentioned in Section II, altimetry was designed for ocean surfaces, and the waveforms obtained in hydrology contexts are not compatible with ocean waveform processing algorithms. The processing algorithms that take waveforms as input and obtain the corresponding range (among other parameters), are called retrackers. For ocean surfaces, retrackers are based on fitting the Brown model, and operational processing chains use a maximum likelihood estimator with 3 or 4 parameters. The shapes of continental surfaces waveforms being very different and having great diversity, many retrackers have been developed to cope with this situation.

Koblinsky et al. [6] were the first to study great river basins, and proposed two different models for Geosat waveforms following the work of Zwally [19]: a 5-parameter single-ramp similar to the Brown model, and a 9-parameter double-ramp function. These models were to be chosen according to a simple classification: if the waveform was diffuse, the single-ramp model should be used; if it was specular, the double-ramp model should be preferred. The results they obtained were not adequate for routine hydrological measurements because of their accuracy and the problems with the Geosat altimeter, particularly the uncertainty in its orbit. Birkett [7] constructed one of the first accurate lake level datasets, showing the usefulness of Topex/Poseidon altimetry in a hydrological context. Rather than developing a new retracker, she constructed a time series of height data relative to a reference pass profile, whose only constraint was that its slope shouldn't exceed 0.25 m/km. The dataset consisted of 24 big lakes, with most smaller lakes being missed because of the effect of topography variations which hindered radar echo acquisition (problem now solved with on-board DEM tracking mode) and low accuracy due to the quality of the atmospheric corrections. Birkett [8] later extended her analysis to 8 large rivers (width greater than 2 km) and 9 wetlands. Multi-mission altimetry data ([9], [20]) was also successfully used for monitoring big rivers and supporting the case for global hydrology.

Frappart et al. [21] analyzed the 4 different retrackers used in the Envisat mission's official products: an ocean retracker (based on the Brown model, and using Hayne's expressions [22]), the Ice-1 (or Offset Center of Gravity, OCOG) retracker [23], the Ice-2 retracker [24], and Laxon's Sea-ice retracker [25]. Remarkably, none of these retrackers was created for hydrological applications. Two of them were refinements of the Brown model (ocean and Ice-2), the other two being intended for ice applications. Ice-1 calculates the center of gravity, amplitude and width of a rectangular box using the maximum of waveform samples, and finds the point on the waveform where the amplitude exceeds a threshold. As such, the Ice-1 retracker doesn't take into account the shape of the waveform and thus is very robust. The Sea-ice retracker is just finding, by interpolation, the point on the waveform where the echo is greater than a threshold corresponding to half the waveform amplitude. The comparison of these 4 retrackers showed that despite its simplicity, the most accurate water stages were obtained when using the Ice-1 retracker.

Following the pioneer work of Berry [9], the Pistach project [26], [27] proposed the use of a waveform classifier prior to retracking altimetry data in coastal and hydrology contexts. According to the class the waveform was assigned to, among the 16 they defined (1:Brown echoes, 2: Peaky echoes, 3: very noisy echoes, 4: linear echoes, etc.), a particular retracker was invoked. The results were promising but were never compelling enough to be of widespread use, the main problem with their product being the lack of along-track consistency when changing from one retracker to another.

A different approach for constructing time-series is that of the DAHITI database [20], that uses automatic outlier rejection and Kalman filtering to use different tracks and missions into consistent water-level series estimations. The dataset is promising but the number of virtual stations is clearly lacking with respect to the available hydrology targets (cf. table II).

The approach we present is much in the spirit of that of Enjolras and Rodriguez [28], but the technical details are completely different, and with outcomes greatly enhancing usability for hydrologists. In particular, we deal with global adjustment of virtual stations while they inverted the waveforms of a selected site. We are not aware of any uses of machine learning to tackle the problem of creating hydrological time series through radargrams. Nevertheless, we do take advantage of the advances in feature extraction and image processing with deep learning.

IV. Methodology

The problem of determining inland waters level from altimetry data can be posed in different ways. Here we suppose that waterbodies (lake, river, etc.) of interest are known and that the

radargrams of the regions in the vicinity of these waterbodies are readily available.

We want to build a neural network that, given as input a radargram over a particular waterbody, returns the epoch corresponding to the waterbody's surface height (as in equation 1 and figure 3). By feeding to this neural network a set of radargrams of that waterbody, one for each pass of the altimeter over it, we can then construct a water level time series. For doing so, we will create a deep neural network by supervised learning. This technique requires a great number of training examples, defined by the couples (input=radargram, output=epoch). The number of examples required to train a deep network (more than 710000 examples in our current training set) is much greater than the number of the *in situ* measurements available. Instead, we choose to use radargrams produced by an altimetric waveform simulator with which we control the "ground truth", since its input is the description of a scenery as defined by two maps: surface heights and backscatter coefficients. Given these maps, the simulator is able to compute the response of the scene to a radar altimeter, and produce the simulated radargrams. These radargrams can, in turn, be used by a retracker to compute the surface heights below (which are nothing else than the input surface heights). The next section describes the model used for these simulations.

A. Simulations

1) The radar altimeter: A radar altimeter can be schematically described as an instrument emitting an instantaneous radar signal and recording the echos returned from the target surface below during a very short time (roughly the time needed for the speed of light to travel a distance of 50m, in the case of the Jason-3 altimeter considered in this paper). This recording is the waveform described in Section I, and is discretized in gates (104 for Jason-3), the width of each gate corresponding to a distance δr of approximately 47 cm. Because the radar altimeter travels onboard a satellite, it moves at a ground speed of ca. 7 km/s. With Jason-3's sampling of 20 Hz, waveforms are spaced every 350m.

A complete characterization of the radar instrument is beyond the scope of this paper, but we require its position at every measurement point (given by precise orbit calculations for real altimetry missions), its bandwidth, carrier frequency, antenna diameter, PTR (point target response), etc.

2) The scene: To simulate the formation of an altimetry waveform we consider a *scene* in which all the scattering objects are placed. This scene includes a description of the topography, rivers, lakes, urban areas, grass, sand, snow, etc. For our simulation, this scene will be discretized with a grid of 20 m resolution.

The portion of the scene that interacts with the radar signal can be seen as a set of elementary scattering reflectors. This scene itself can be described by the positions of these reflectors and their backscattering characteristics.

If we consider a single point reflector P receiving a radar signal from a radar altimeter at position R, lets denote the euclidean distance between them as $d(P, R) = r$. If we neglect the radar signal delays due to atmospheric effects, the echo created by P appears in the waveform as a peak at position g given by

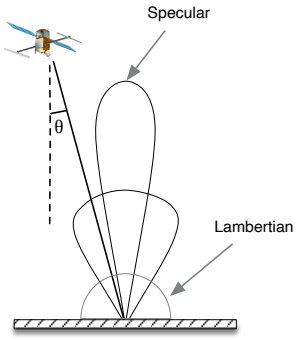

Fig. 6. Diagram showing 3 scattering behaviors for a scene element: diffuse (Lambertian), specular, and a mix of both.

$$g = (r - r_0)/\delta r \tag{2}$$

where r_0 is the distance correspondng to the first waveform gate and δr the gate width.

This energy peak contributes to different gates for successive waveforms. When P is at the nadir of the antenna, g is minimum and the peak is maximum. As P moves away from the nadir, g increases (late arrival) creating the parabolas in the gate / latitude plane shown in figures 2 and 5. The contribution of the peak decreases laterally according to distance and the antenna pattern.

A very important point here is that peak position depends on the distance between the altimeter and the reflector. A consequence of this observability issue is that if we suppose that the reflector is directly underneath the satellite, we can deduce the height of the reflector; however, if it is at a distance from the ground track, the deduced height will be wrong unless there is *a priori* information on the horizontal position of the object (for example a water mask or optical image of the zone).

3) Backscatter and antenna pattern: The backscatter coefficient σ_0 is a measure of the reflecting power of the surface, and is expressed in decibels (dB). The higher the coefficient, the stronger the echo. On the other hand, this coefficient depends on the angle of the surface with respect to the incident radar signal θ. Because our applications concern hydrology, the reflectors we work with are water and the normal direction to them are in general vertical.

We consider two scattering behaviors (Figure 6):

- diffuse or Lambertian reflector: no angular dependence, σ_0 is independent of θ.
- specular reflector (mirror): very marked dependence on the incidence angle θ. The function $\sigma_0(\theta)$ has a strong peak near the vertical, for value $\theta = 0$. The specularity function is modelled as $e^{-g_0\theta^2}$. This gaussian function is parametrized in terms of its full width at half maximum, $fwhm$.

In reality, a mix of both behaviors is frequently observed. The fraction of diffuse reflectance in the mix is called x_{diff}, and that of specularity $(1 - x_{diff})$.

The radar echo is captured by an antenna whose performance is crucial for measurement acquisition. The antenna pattern is defined by the gain $G(\theta)$ with which the radar power is damped at reception. $G(\theta)$ is modelled by a gaussian function with a $fwhm$ value of $1.3°$, typical of most altimeters.

4) The simulator: Putting all this together, the power bounced back by the reflector P from the signal of the altimeter at position R, and received at gate g (as defined above) is:

$$dp(g(P,R)) = A \cdot G(\theta(P)) \cdot \sigma_0(P,\theta(P)) \cdot dS \qquad (3)$$

where dS is a surface element associated to P and A is a scaling coefficient which takes into account many system parameters, such as the emitted power, the electronic gain of the instrument and free space losses.

The simulator then models the construction of a complete waveform by integrating the contributions of all the grid surface (Equation 4).

$$p(g(P,R)) = \int_S A \cdot G(\theta(P)) \cdot \sigma_0(P,\theta(P)) \cdot dS \qquad (4)$$

It is to be noted that this simulation makes use of the incoherent sum of the individual reflector echoes, as it was done in [28]. A more complete approach is also possible [29] in which the waveform is constructed by a coherent sum of the contributions of the reflectors (i.e. taking into account not only the amplitude of the received signal, but also its phase).

The variability of the simulated radargrams depends both on geophysical variations (the landscape changes through time, as well as weather conditions), and on differences in the location of altimetry measurements. Even if the satellite altimeter follows a repeat orbit, there are variations on the actual trajectory, which can wander up to 1 km (cross-track distance) away from the theoretical ground path (orbital maneuvers conducted by the space agencies correct these drifts). For hydrological observations, the radargrams will not always have the targets at exactly the same position. The altimeter acquisition window (when the recording of the echo of the emitted signal is made) is defined on-board and is used to define the distance of the altimeter to a reference gate in the radargram (the r_0 of Equation 2, whose value is 32 for Jason-3). This distance is the tracker range.

It is worth to note that other technical details must be accounted for in order to obtain realistic waveforms: such as pulse compression for the emitted signal (since the signal is not infinitely short), variable bandwidth for some altimeters (e.g. the Envisat mission's RA-2 altimeter changes resolution depending on the observed surface). The simulator takes them also into account but a technical presentation of these features are not relevant for the analysis that follows.

The simulation of realistic waveforms requires not only good knowledge of the technical details of the altimeter instrument but also good ancillary data of the scene and of the instrument (tracker range and its automatic gain control, or AGC). Surface heights are derived from SRTM-30 [30], a global digital elevation model; backscatter characteristics for hydrology can be defined by water masks (e.g. [31], [32]), using the fact that the reflectivity of inland water is significantly higher than that of the land surrounding it, for the case of the Ku band used here. We hereto describe how the water reflectance is constructed when preparing the simulations of the machine learning database.

The radargrams in the database are generated in several steps, and satisfying several conditions:

- locate the virtual stations (intersections of Jason-3 ground tracks with known waterbodies [32]). A map of the 2517 selected virtual stations is shown on Figure 7. Each virtual station will be used at least as many times as there are flyovers using open loop mode (later than september 2017, on cycle 58).
- a scene is created using a realistic digital elevation model around each virtual station using the SRTM-30 dataset [30]. The scene is typically 30 km wide and 56 km long, its length parallel to the satellite ground track.
- a water mask is added to the scene using only the 20 biggest objects from the SWBD [32] water mask database that are present in the scene. The hydrology target is always at the center of the scene.
- for each object in the water mask (simplifying the statistical description of parameters we observed in real data):
  - the one at the center is always a river or lake
  - it will be labelled as a lake if the length to width ratio is < 40 or if its size is greater than 300000 pixels (of 30m side length). Otherwise, they are labelled as rivers.
  - if it is a lake:
    - reflectance is mostly Lambertian, so a random value is chosen for the fraction of diffuse reflectance: $x_{diff} \in U(0.7, 1)$ where $U(a, b)$ is the uniform distribution in $[a, b]$.
    - the fraction of specularity is $(1 - x_{diff})$ and $fwhm \in 10^{U(log10(0.01),log10(0.1))}$.
    - the backscatter coefficient is $\sigma_0 \in U(15, 30)$ dB.
  - if it is a river:
    - $fwhm \in 10^{U(\log(0.01),\log(0.6))}$
    - reflectance is mostly specular, so $x_{diff} \in 10^{U(\log 10(0.00005),\log(0.02*fwhm))}$
    - the backscatter coefficient is computed from the the original measurement.
- we generate several backscatter configurations for the same scene and cycle in the spirit of data augmentation [33], [34], although we simply regenerate the random coefficients as described above. We refer to the number of configurations as the augmentation factor.
- the position of the altimeter will be given by its real trajectory, and we can then also use the real AGC and track range.
- to add realism to the radargrams, a convolution of each simulated waveform with the instrument PTR (a $sinc$ function) is made, creating saturation (clipped) peaks as observed in real data. However, to obtain a realistic

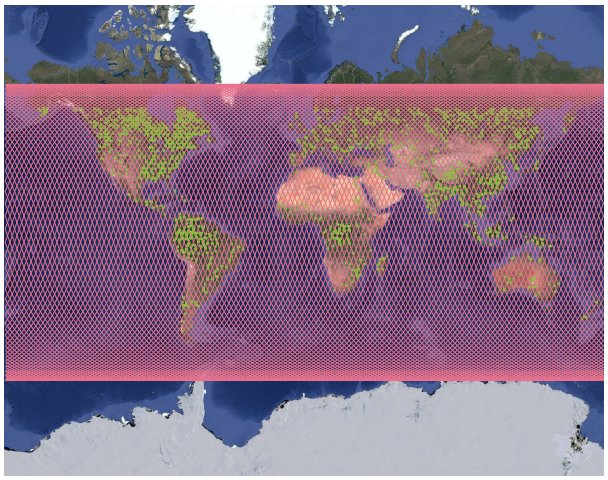

Fig. 7. Map showing as green dots the location of the virtual stations used. The grid shows the ground tracks of the Jason-3 mission.

proportion of saturated data, an iterative algorithm was implemented to obtain the correct saturation distribution.

To compare with in-situ data, we used cycles 58 to 104 (september 2017 to december 2018). After compiling the database, we have 2517 virtual stations over the 47 cycles with an augmentation factor of 6, obtaining 709 794 different simulations. These were divided into a training dataset of ca. 570 000 simulations and a validation dataset of ca. 140 000 simulations.

B. Deep Convolutional Neural Networks

As stated at the beginning of this section, our objective is to define a neural network that takes as input a radargram, and gives as output the epoch corresponding to the water surface. This problem is hence one of finding the function $f : R^2 \rightarrow R_{\geq 0}$ that predicts one positive real value:

$$f(\text{radargram}) = r = \begin{cases} 0, \text{if no water is detected} \\ \text{water surface epoch, otherwise} \end{cases}$$

In machine learning terms, this problem is one of regression analysis. Because we can view the problem as one of finding a point in the radargram, our model is based on the ResNet-50 neural network architecture, which performs well in image recognition [35]. Nevertheless, the model has to be adapted to our particular needs; ResNets are designed for classifiying RGB color images, whereas in this study the input radargrams are like greyscale (or pseudocolor) images, and thus the input is juste one matrix instead of three.

All of the changes we made are summarized in Figure 8.

We trained a convolutional neural network using PyTorch [36]. The 50-layer ResNet is an interesting compromise between the computation time and the time needed to finetune the hyperparameters in order to obtain satisfying results (hyperparameters are the parameters controlling the learning process as well as the topology and size of the neural network).

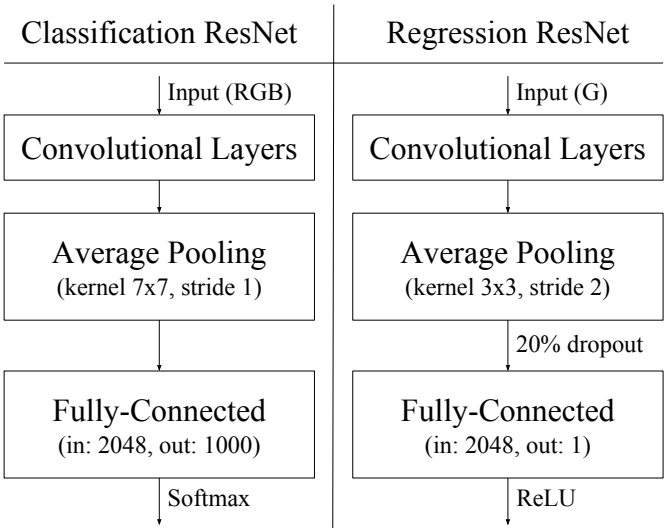

Fig. 8. Modifications made to the ResNet-50. Left: the original ResNet architecture [35]. Right: the modified ResNet architecture to fit our regression problem. Inputs are modified from 3 RGB layers to 1. Softmax and ReLU are the activation functions (smooth step-like function and ramp-like function) which translate the final state of the neural network into its final output.

We followed recommendations published in the literature to finetune the hyperparameters [37], [38]. The neural network weights were initialized like in [39] because of the use of Rectified Linear Units (ReLU) in the network [40]. The optimizer was Adam [41]. The learning rate is one of the most crucial hyperparameters to finetune. It determines how much the neural network weights are adjusted at each step. In order not to depend on a single value, a learning rate schedule which adjusts the learning rate during training was adopted. Empirically, we favored cyclical schedules [42], especially Loshchilov and Hutter's cosine annealing with restarts [43], [44], as seen in Figure 9. We added a maximum value decay at every restart, because without it the loss function value rose when the learning rate restarted.

Also, our networks benefited from dropout [45] as long as the training database was too small (less than 500k examples). 25% of convolutional units were dropped, as well as 20% of the fully-connected layer units. When the database became big enough, adding dropout slightly degraded the performance of the neural network whereas there was not much overfitting to prevent anymore. Hence, we chose to keep a slight overfitting to benefit from a better validation accuracy. Moreover, ResNet convolutional layers are paired with batch normalization layers. According to [46], batch normalization "acts as a regularizer, in some cases eliminating the need for dropout". We only kept 20% of dropout before the output fully-connected layer. This did indeed enhance the performance.

Optimizing for the available hardware (GPU and memory), we trained our neural networks with batch sizes varying from 64 to 512 samples. We did not notice significant differences in the prediction performance, although the use of bigger batches obviously reduced the training time. Table III summaries the measured timings depending on the hardware configurations

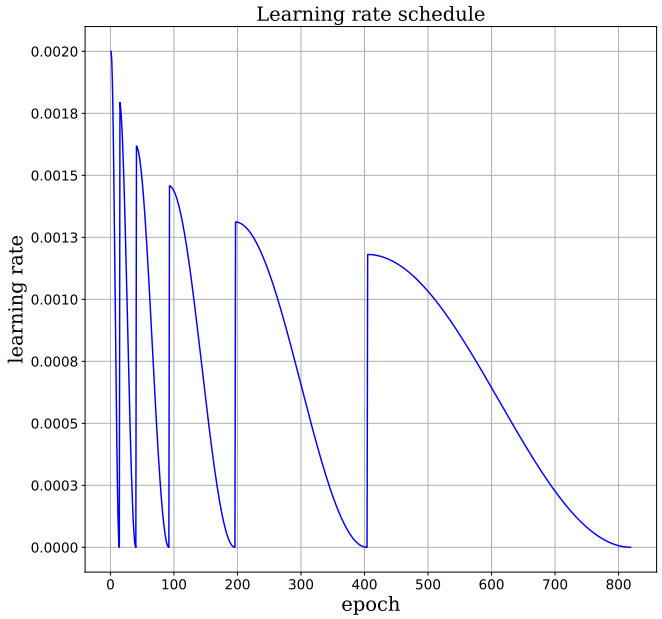

Fig. 9. Evolution of the learning rate using the cosine annealing with restarts and decay schedule with the following parameters: initial learning rate of 0.002, initial period of 13 epochs (NB: these are machine learning epochs, or computation cycles, completely unrelated to waveform epochs, τ), period multiplier of 2, restart decay of 0.9.

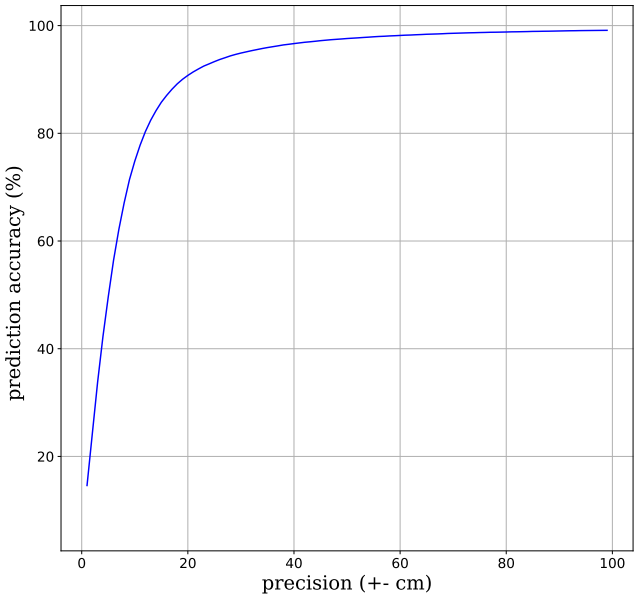

Fig. 10. Network performance on the validation set. This network reached 90.75% of correct predictions within a range of $\pm$20 cm on the validation set. If we wanted a precision better than 30 cm, network's prediction accuracy would be of 95%.

TABLE III
PERFORMANCES SUMMARY

CPU (cores/threads)	GPU (memory)	maximum batch size	time/epoch (minutes)	time/sample (seconds)
Xeon E5-1650 3.6 GHz (6/12)	GTX 1080 (8 Go)	64	19.6	0.01
2x Xeon Silver 2.1 GHz 2x (12/12)	2x V100 (2x 32 Go)	512	5.5	0.02

used during the study. Note that the processor used along with the Nvidia GTX 1080 was faster than the one used with the V100. We could still save time when working with the V100 if we had a faster CPU. Most of our work was done on a personal computer.

V. EXPERIMENTAL RESULTS

As usual when working with deep learning, the simulations previously described where split in two groups: training and validation simulations. During the training phase, the labels where obtained from the simulation's scene. During the training phase, we strived for an accuracy of 90% within a range of $\pm$20 cm. That is, the result was deemed accurate enough if it was within the range $label \pm 20cm$. Real measurements were tested against the resulting neural network to validate its real world performance.

A. Validation on Simulations

We created many different neural networks to tackle our radargram analysis problem. All of these networks used the ResNet architecture, but differed in the neural network weight initialization strategy, how the neural network weights were updated (the optimizer) and the learning rate schedule. Figure 10 shows the accuracy of the best model we found. This neural network, used on the remaining of the paper, achieved 90.75% of success. It was trained during 819 epochs on a 709494-sample dataset (with 80% of the set used for training and 20% for validation). The learning rate schedule was the cosine annealing with restarts and decay shown on Figure 9. 20% of the fully-connected output layer units were dropped out.

B. Validation on In-Situ Data

In order to test the performance of our model on real radargrams, we evaluated the results obtained against the values measured by *in situ* water level gauges. The regions chosen for validation depended mostly on the availability of such gauges, and either published by hydrologists or public agencies.

For each virtual station in the study domain, we processed the radargrams corresponding to successive Jason-3 flyovers with our best neural network in order to build a water height time series. The comparison of this time series with the *in situ* water level gauge time series permitted us to assess the performance of the neural network. Most of the time, *in situ* time series are sampled every fifteen minutes to one hour. It can reach up to twelve hours, depending on the region. We downsampled the *in situ* series to match the moments when the satellite flies over the hydrology target.

In situ gauges have to be located as close as possible to the studied virtual stations. However, it is difficult to find publicly available gauge data at the precise position of the

virtual station. To avoid problems with different reference heights we compare anomaly time series instead of absolute water height. The height bias induced by the presence of the river slope is also partially removed when comparing water elevation anomalies. Anomaly time series were computed by removing the temporal means of the time series. Comparisons were made using the correlation coefficient and the Root Mean Square Error (RMSE) following [47]. Note that the RMSE includes the uncertainties due to the satellite ground track jitter and noise from both the measure and the method.

1) Rio Negro (Brazil) (Figure 11): Rio Negro is among the ten largest rivers in the world. An *in situ* gauge is available in Manaus harbor, under Jason-3 track 63 on the Rio Negro near its confluence with the Amazon, where the river width is about 2500 m. *In situ* data was provided by the Brazilian Water Agency ANA (http://www2.ana.gov.br)

The gauge is 10 km upstream from the virtual station. The river slope is known to hugely vary during the seasonal cycle and its width is much greater near the harbor than near the confluence. Moreover, the sampling rate of the *in situ* series is of 12 hours only and there can be up to 6h difference when comparing measurements. All this can explain the gap between the two series around June 2017. Nevertheless, the correlation between these time series is 0.998 and the RMSE is 28.8 centimeters.

2) La Garonne (France) (Figure 12a): Garonne is a mid-latitude small size river, with rapid level variations. At the crossing with the Jason ground tracks, it is about 150 m wide. Observations collected at in situ gauges are released publicly online by the SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations- http://www.hydro.eaufrance.fr). Water elevations are measured with a non-uniform time step which depends on the water elevation variations. The median time step on the studied gauge is 30 minutes.

We compare the neural network output for the target identified at the crossing with this river with an in situ gauge operating 6 km upstream (Marmande). The river is too narrow to be in the SWBD water mask. Consequently, the neural network has never seen such rivers of this size during its training and it is an interesting challenge for our model. The correlation between the time series is 0.985 and the RMSE is 30.3 centimeters. We notice that the radar altimetry and the neural network were able to follow the rapid rise in water elevation that occurred from February to June 2018.

3) Pend Oreille River (USA) (Fig. 12b): We used the United States Geological Survey Water Data Bank (http://waterdata.usgs.gov) to find gauges close to Jason-3 virtual stations. One of them is on the Pend Oreille River, on a section whose width is 190 meters, in a mountainous region of the state of Washington. The virtual station is 3 km downstream from the gauge. The *in situ* water elevation is measured every fifteen minutes. The Pend Oreille region is hilly, the river width is not constant and there is a lake a few kilometers away from the station. This contributes to disrupt the radar measure, and creates more complex radargrams.

The correlation between the time series is 0.983 and the RMSE is 36.6 centimeters.

TABLE IV
SUMMARY OF VIRTUAL STATION COMPARED TO IN SITU RESULTS

Location	River width	Distance from gauge (km)	Gauge sampling rate	Correlation	RMSE (cm)
Rio Negro (BR)	4 km	10 (us)	12h	0.998	28.8
Pend Oreille (US)	190 m	3 (ds)	15'	0.983	36.6
Garonne (FR)	150 m	6 (us)	30'	0.985	30.3
La Marne (FR) track 44	50 m	9 (ds)	30'	0.955	31.9
La Marne (FR) track 35	50 m	13.5 (ds)	30'	0.884	43.8

C. Validation on Satellite Ground-Track Crossovers

For a few ground-track crossovers, we can find two virtual stations on the same waterbody, on two different tracks. It means the satellite passes over the same river with two different points of view. Then we have two radargrams which can be very different, although measured on the same object.

Among these stations, we were particularly interested in those which are overflown within a few hours only. Thus, we can compare the time series computed by the neural network from the two virtual stations, presuming the water height has to be almost the same. In this paper, we focused on two cases.

1) Orange River (South Africa) (Fig. 13): The region is a dry mountainous plateau and the river width varies between 50 to 200 meters. There are 37 km between the two virtual stations, and are overflown at a 10 hour interval. The correlation of the two series is 0.767 and the RMSE is 27.4 centimeters.

2) La Marne (France) (Fig. 14): The case is particularly interesting since there is an *in situ* gauge near the crossover. Virtual station on track 44 (respectively 35) is 9 km (respectively 13.5 km) downstream from the gauge.

The river has many meanders and a mean width of only 50 meters. Moreover, there are many small ponds close to the river which is in an agricultural area. This can explain the differences between the three time series on Fig. 14. Because the three stations are not so close from each other, our model probably measures the dynamics between the river and its environment between the stations. Then, this contributes to degrade both the correlation coefficient and the RMSE.

Virtual station on track 44 (resp. 35) has a correlation of 0.955 (resp. 0.884) with the gauge and a RMSE of 31.9 centimeters (resp. 43.8 centimeters). The two neural network processed time series have a correlation of 0.889 and a RMSE of 49 centimeters.

A summary of the results is presented on Tables IV (where "us" means upstream, and "ds" means downstream) and V.

D. Visualizing the behavior of Neural Network Components

The behaviour of deep neural networks is still subject to much scientific research work [48]. For a particular network, it is interesting to know how the network reacts to a specific

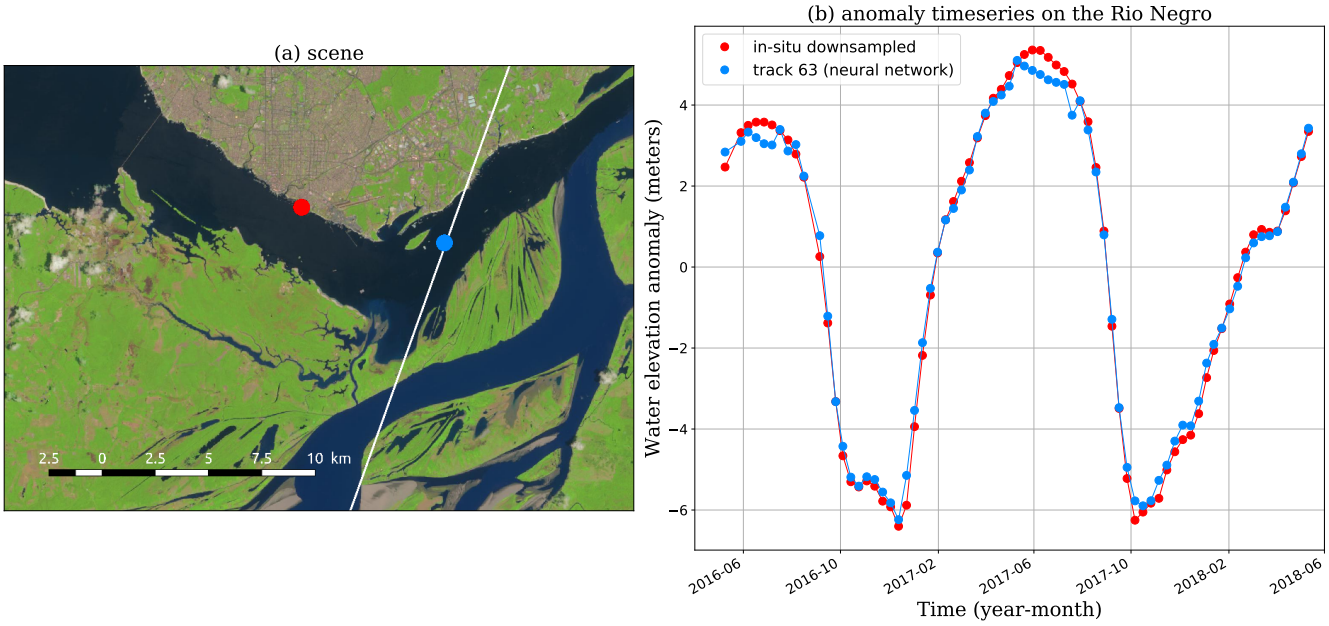

Fig. 11. (a) The *in situ* gauge (red dot) and the virtual station (blue dot). The white line corresponds to Jason-3 ground track. (b) Jason-3 water elevation anomaly time series on the Rio Negro near Manaus (blue line and dots) compared to *in situ* anomaly time series (red line and dots). *In situ* series is downsampled to match the 10-day satellite sampling. Correlation: 0.998, RMSE: 28.8 cm.

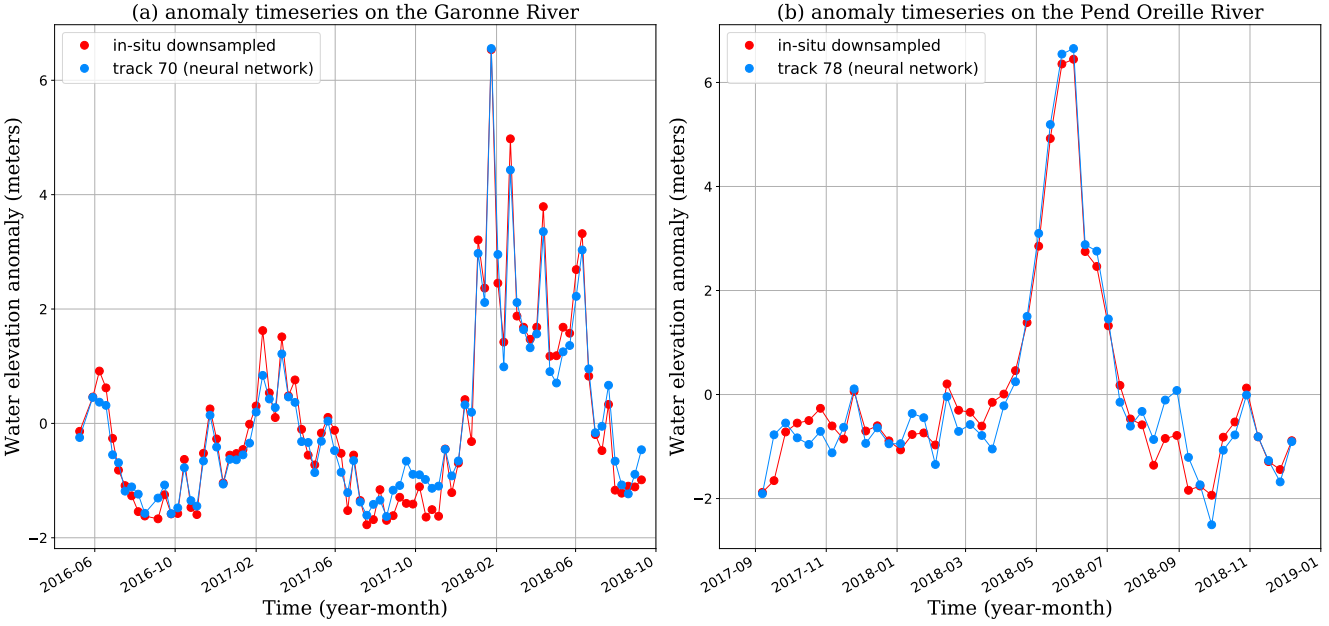

Fig. 12. Jason-3 water elevation anomaly time serie (blue lines and dots) compared to *in situ* anomaly time series (red lines and dots). *in situ* series are dowsampled to match the 10-day satellite sampling. (a) The Garonne River near Marmande, France. Correlation: 0.985, RMSE: 30.3 centimeters. (b) The Pend Oreille River in northeastern Washington state, United States. Correlation: 0.983, RMSE: 36.6 centimeters.

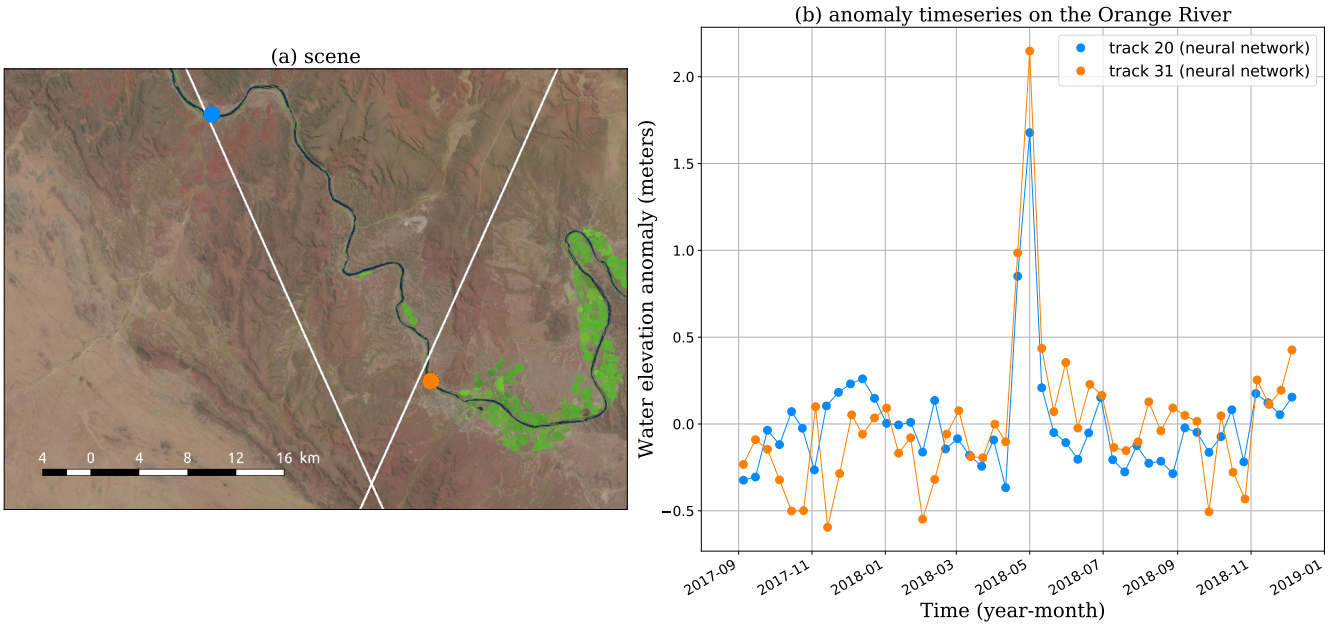

Fig. 13. (a) The Orange River in South Africa. The satellite flies over the river twice within 10 hours, on tracks 20 (blue dot) and 31 (orange dot). The two virtual stations are 37 km apart. The white line corresponds to Jason-3 ground track. (b) Jason-3 water elevation anomaly time series of the virtual stations on tracks 20 (blue line and dots) and 31 (orange line and dots). Correlation: 0.767, RMSE: 27.4 centimeters.

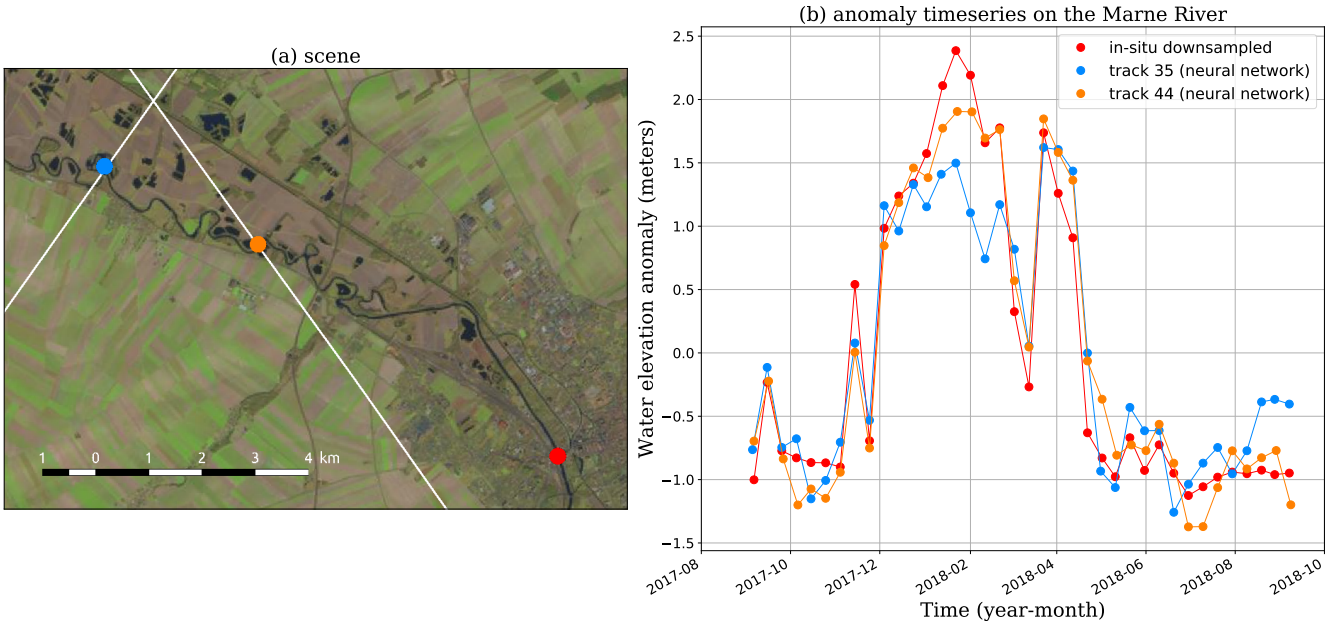

Fig. 14. (a) The Marne River at Châlons-en-Champagne. The satellite flies over the river twice within 8 hours, on tracks 35 (blue dot) and 44 (orange dot). The two virtual stations are 4 km apart and the *in situ* gauge is 9 km upstream the closest virtual station. The white line corresponds to Jason-3 ground track. (b) Jason-3 water elevation anomaly time series (blue and orange lines and dots) and *in situ* time serie at Châlons-en-Champagne (red line and dots). The two neural network processed time series have a correlation of 0.889 and a RMSE of 49 centimeters. The correlation between track 44 (resp. 35) and the *in situ* gauge is 0.955 (resp. 0.884) and the RMSE is 31.9 centimeters (resp. 43.8 centimeters).

TABLE V
SUMMARY OF CROSS-OVER RESULTS

Location	River width	Distance	Lag	Correlation	RMSE
Orange River (SA) tracks 20/31	50-200 m	37 km	10h	0.767	27.4 cm
La Marne (FR) tracks 33/45	50 m	4.5 km	7h18'	0.889	49 cm

input, and more specifically, to find out what are the features in the radargram that the neural network reacts to. We adapted Zeiler's [49] occlusion experiment to regression in order to visualize which radargram's features are important to our neural network and which ones are taken into account to output the water surface elevation. The method consists in sliding a 3×3 occluding window on the radargram and monitoring the neural network output. For each position of the window we compute the difference between the neural network output with occlusion and the default output, without occlusion. Once every position has been occluded, we have a heatmap of the essential radargram features.

Red pixels mean the neural network overestimates the water elevation, whereas blue pixels mean it underestimates the elevation. If an area is white, it means that the neural network does not give importance to this portion of the radargram. The stronger the color, the more important to the network is the associated radargram region.

From the heatmaps in Figure 15 we can draw the following conclusions:

- The trained convolutional neural network is sensitive to the parabolas when they exist, or to the areas with a high power level. This is something we expected since this is where we visually find the information about the water surface elevation.
- The neural network gives a strong importance to the main waterbody, near the center of the radargram. Other waterbodies are ignored or given a low importance as seen on cases 1 and 2. The parabola under the centered one is the reflection of a lake we want to ignore. As seen on the heatmap, this lake is given an extremely low importance by the neural network.
- The network takes the saturation into account because it gives importance to the side lobes (*sinc*, from the altimeter PTR), when they are visible. On case 1, there is a strong saturation. On the heatmap, the saturation peak is visible as well, meaning that the neural network gives importance to it and uses the side side lobes to extract information about the water surface elevation.
- Finally, on case 3 the satellite ground track is almost parallel to the river. Because of its meanders, there are many close encounters resulting in some waveforms having a higher signal power than the main central signal. The river slope is visible in the radargram: from waveforms 20 to 60, the river surface elevation increases. As expected, the neural network focuses on the middle waveforms. When waveforms from 30 to 40 are hidden, the network overestimates the height. Indeed, these waveforms correspond to lower water levels. The exact opposite happens for waveforms 40 to 50. For this radargram, the neural network needs approximately 20 waveforms to estimate the elevation.

To conclude, our neural network extracts the information from multiple waveforms, with a particular focus on the centered ones, and has learnt how to deal with saturated waveforms. This is crucial in order to process Jason-3 hydrology targets worldwide, because more than 80% of hydrological targets produce saturated waveforms. Version du 2021/02/15 utiliser mode Suggestions pour voir les différences.

VI. DISCUSSION AND CONCLUSIONS

This paper has shown an innovative method for retrieving hydrological time series from altimetry. We started by recalling the importance of constituting these time series, and the challenges to be tackled. The usage of the new DEM tracking mode onboard Jason-3 and Sentinel-3 satellites has radically increased the quantity of available data in hydrology contexts.

A brief presentation of the principles of radar altimetry allowed to make clear the difference between radargrams in ocean and in hydrology contexts; namely, the fitness of purpose of using the Brown model. In the literature, approaches to obtaining hydrology time series from radargrams use retracking algorithms based either on physical models of more complexity than the Brown type, or by heuristics ones which have proven useful, such as the OCOG/Ice-1 retracker. These classical approaches do not act on the radargram, but treat each waveform independently. In contrast, by restating the problem as an interpretation of the complete radargram, we can use deep learning techniques and extract information of the radargram as a whole.

A deep neural network is defined by an architecture and the weights of its neurons. The architecture is chosen according to state of the art knowledge (ResNet in our case), but many techniques are used to create the best weights when using a training dataset. In this paper we used a very complete altimetry simulator to constitute a database of 710 000 realistic simulated radargrams, all based on the trajectories and ancillary data of real measurements. The scene used for the simulations merges several sources to establish a "ground truth" with which we can label the resulting radargrams. The scene uses topography from the SRTM-30 dataset [30] and water mask from SWBD [32]. The reflectance of the background land cover being much less than that of water, we describe it by a constant of very low reflectance. The objects of the water mask are more realistically produced by modelling both lake and river objects. The first ones are mostly diffuse, the latter mostly specular, but both are a mixture of these two. Random distributions are used to create the reflectance characteristics of these objects, while maintaining some constraints so as to make them physically plausible.

From a machine learning perspective, the problem we seek to solve is obtaining the water height from the radargram of a known target (whose position is near the center of the

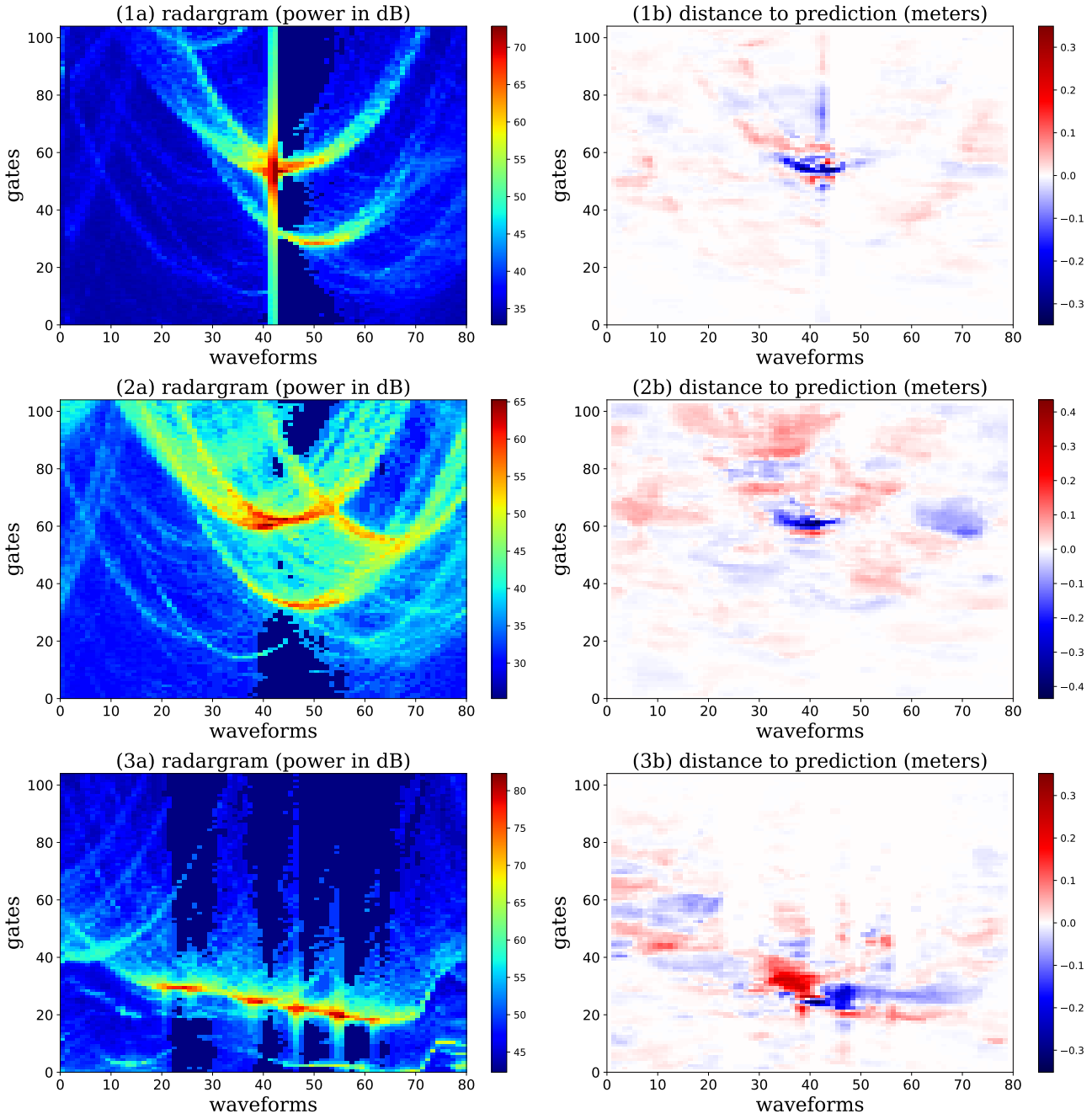

Fig. 15. Radargrams and their corresponding heatmaps. (1a) (1b) Pend Oreille River, USA. Cycle 66. (2a) (2b) Pend Oreille River, USA. Cycle 102. (3a) (3b) La Garonne, France. Cycle 90.

radargram). By iterating over all available radargrams over a given target, the neural network constructs a water level time series, point by point.

We used PyTorch [36] and obtained a fairly satisfactory neural network for Jason-3 altimetry data, with 90.75% success on the validation dataset. When applied to real measurements, we compared 5 virtual stations to in-situ gauge data, finding correlations of 88% for one station and of over 95% for the other four (Table IV). These correlations have very high values and correspond to very low time differences between the altimetry measurement and the gauge measurement (less than 30' except for the Rio Negro, which is the largest one and one that flows unhindered). In contrast, validation with track cross-overs over France and South Africa had time lags of 7 and 10 hours and were in rivers subject to water management maneuvers (Gariep Dam in South Africa, Joinville over La Marne). In these cases, crossover correlations are not as good as those presented for *in situ* data, but remain nonetheless, worthwhile.

We also tested the performance of the neural network using partial occlusion of the radargrams, as a way of creating a heatmap of the features in the radargram that were more

important for the regression output value. As expected, we found that the central features were given priority, and that waveform saturation was correctly taken into account, when present.

With respect to the original target of creating an automatic processing method for retrieving water heights from altimetry data, we believe the progress in this paper is substantial. A first takeaway message of this work is that there *is* information in the radargrams that can be extracted for hydrology. As with all machine learning techniques, creating a neural network is cumbersome and computationally expensive but, once created, its use is very fast and efficient. Furthermore, the neural network is able to use generalization, reacting appropriately to radargrams even when their kind has never been seen before. Using this neural network makes possible the analysis of the hundreds of thousands of virtual stations available with the advent of satellite altimeters using the DEM tracking mode.

However, this is not a perfect solution at global scale; there is still work to be done. Firstly, the final neural network is very sensitive to the training dataset, which must exhibit a representative sampling of the data to be processed. The crux of our method is using our altimetry simulator (with which we can obtain appropriate labels for our training data) as a proxy of real altimetry data. It is nevertheless very difficult to create a perfectly realistic training database.

Secondly, there is a wealth of historical data from missions not using the DEM tracking mode. These missions show radargram variations that are more difficult to tackle with the present network: the tracker range can shift its position substantially when approaching the waterbodies (and it can even create a radargram made of echoes completely unrelated to them). As a first step using this methodology and restricting ourselves to the DEM tracking mode altimeters, this problem is simply ignored, but this problem should clearly be overcome to be able to exploit historical altimetry data.

Thirdly, we noticed that the neural network can infer a great number of hydrological timeseries where only a couple of points are outliers. In our approach we use the information of successive waveforms to create a radargram and obtain information from the successive waveforms, all of them from the same satellite flyover. However, we can go even further and use the information from successive radargrams, at different dates. Using this information, we believe these outliers can be detected and the general accuracy improved if the temporal information was included in the network (using e.g. Temporal Convolutional Networks [50]).

Finally, the network returns results, but there is no justification for them. In line with recent approaches for explainable AI [51]–[53], we believe there are tools to be developed, the first one of which is the heatmap of section V

The neural network obtained is very usable and completely automatic, giving good results when compared to what is obtained with manual processing.

ACKNOWLEDGMENT

The authors would like to thank Chedy Raïssi for a fruitful discussion at the beginning of this project and the CTOH (LEGOS) for providing access to the database of altimetry products, and the Brazilian Water Agency ANA for the in situ data of the Rio Negro.

REFERENCES

[1] K. E. Trenberth and G. R. Asrar, "Challenges and Opportunities in Water Cycle Research: WCRP Contributions," *Surveys in Geophysics*, vol. 35, no. 3, pp. 515–532, May 2014.

[2] C. Prigent, F. Papa, F. Aires, C. Jimenez, W. B. Rossow, and E. Matthews, "Changes in land surface water dynamics since the 1990s and relation to population pressure: LAND SURFACE WATER DYNAMICS," *Geophysical Research Letters*, vol. 39, no. 8, pp. n/a–n/a, Apr. 2012.

[3] G. Blöschl, A. Kiss, A. Viglione, M. Barriendos, O. Böhm, R. Brázdil, D. Coeur, G. Demarée, M. C. Llasat, N. Macdonald, D. Retsö, L. Roald, P. Schmocker-Fackel, I. Amorim, M. Bělínová, G. Benito, C. Bertolin, D. Camuffo, D. Cornel, R. Doktor, L. Elleder, S. Enzi, J. C. Garcia, R. Glaser, J. Hall, K. Haslinger, M. Hofstätter, J. Komma, D. Limanówka, D. Lun, A. Panin, J. Parajka, H. Petrić, F. S. Rodrigo, C. Rohr, J. Schönbein, L. Schulte, L. P. Silva, W. H. J. Toonen, P. Valent, J. Waser, and O. Wetter, "Current European flood-rich period exceptional compared with past 500 years," *Nature*, vol. 583, no. 7817, pp. 560–566, Jul. 2020.

[4] C. Vorosmarty, A. Askew, W. Grabs, R. G. Barry, C. Birkett, P. Doll, B. Goodison, A. Hall, R. Jenne, L. Kitaev, J. Landwehr, M. Keeler, G. Leavesley, J. Schaake, K. Strzepek, S. S. Sundarvel, K. Takeuchi, and F. Webster, "Global water data: A newly endangered species," *Eos, Transactions American Geophysical Union*, vol. 82, no. 5, pp. 54–54, 2001.

[5] J. D. Milliman and K. L. Farnsworth, *River Discharge to the Coastal Ocean: A Global Synthesis*. Cambridge ; New York: Cambridge University Press, 2011.

[6] C. J. Koblinsky, R. T. Clarke, A. C. Brenner, and H. Frey, "Measurement of river level variations with satellite altimetry," *Water Resources Research*, vol. 29, no. 6, pp. 1839–1848, Jun. 1993.

[7] C. M. Birkett, "The contribution of TOPEX/POSEIDON to the global monitoring of climatically sensitive lakes," *Journal of Geophysical Research*, vol. 100, no. C12, p. 25179, 1995.

[8] ——, "Contribution of the TOPEX NASA Radar Altimeter to the global monitoring of large rivers and wetlands," *Water Resources Research*, vol. 34, no. 5, pp. 1223–1239, May 1998.

[9] P. A. M. Berry, "Global inland water monitoring from multi-mission altimetry," *Geophysical Research Letters*, vol. 32, no. 16, 2005.

[10] S. Biancamaria, T. Schaedele, D. Blumstein, F. Frappart, F. Boy, J.-D. Desjonquères, C. Pottier, F. Blarel, and F. Niño, "Validation of Jason-3 tracking modes over French rivers," *Remote Sensing of Environment*, vol. 209, pp. 77–89, May 2018.

[11] J. D. Desjonquères, G. Carayon, N. Steunou, and J. Lambin, "Poseidon-3 Radar Altimeter: New Modes and In-Flight Performances," *Marine Geodesy*, vol. 33, no. sup1, pp. 53–79, Aug. 2010.

[12] S. Le Gac, F. Boy, D. Blumstein, L. Lasson, and N. Picot, "Benefits of the Open-Loop Tracking Command (OLTC): Extending conventional nadir altimetry to inland waters monitoring," *Advances in Space Research*, p. S0273117719307811, Nov. 2019.

[13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov./1998.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'12. Curran Associates Inc., 2012, pp. 1097–1105.

[15] G. S. Brown, "The Average Impulse Response of a Rough Surface and Its Applications," *Society*, no. 1, 1977.

[16] J. S. Da Silva, S. Calmant, F. Seyler, O. C. Rotunno Filho, G. Cochonneau, and W. J. Mansur, "Water levels in the Amazon basin derived from the ERS 2 and ENVISAT radar altimetry missions," *Remote Sensing of Environment*, vol. 114, no. 10, pp. 2160–2181, 2010.

[17] C. Normandin, F. Frappart, A. T. Diepkilé, V. Marieu, E. Mougin, F. Blarel, B. Lubac, N. Braquet, and A. Ba, "Evolution of the Performances of Radar Altimetry Missions from ERS-2 to Sentinel-3A over the Inner Niger Delta," *Remote Sensing*, vol. 10, no. 6, p. 833, May 2018.

[18] CTOH, "ALTIS: Altimetric Time Series Software," 2020.

[19] H. Zwally, A. Brenner, J. Major, T. Martin, and R. Bindschadler, *Satellite Radar Altimetry over Ice. Volume 1: Processing and Corrections of Seasat Data over Greenland*, ser. NASA Reference Publication. NASA, Feb. 1990, vol. 1.

[20] C. Schwatke, D. Dettmering, W. Bosch, and F. Seitz, "DAHITI – an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry," *Hydrology and Earth System Sciences*, vol. 19, no. 10, pp. 4345–4364, Oct. 2015.

[21] F. Frappart, S. Calmant, M. Cauhope, F. Seyler, and A. Cazenave, "Preliminary results of ENVISAT RA-2-derived water levels validation over the Amazon basin," *Remote Sensing of Environment*, vol. 100, no. 2, pp. 252–264, Jan. 2006.

[22] G. Hayne, "Radar altimeter mean return waveforms from near-normal-incidence ocean surface scattering," *IEEE Transactions on Antennas and Propagation*, vol. 28, no. 5, pp. 687–692, Sep. 1980.

[23] D. Wingham, C. Rapley, and H. Griffiths, "New techniques in satellite altimeter tracking systems." in *Proceedings of IGARSS'86 Symposium*, Zurich, 8.

[24] B. Legresy, F. Papa, F. Remy, G. Vinay, M. Van Den Bosch, and O. Z. Zanife, "Envisat radar altimeter measurements over continental surfaces and ice caps using the ice-2 retracking algorithm," *Remote Sensing of Environment*, vol. 95, no. 2, pp. 150–163, 2005.

[25] S. LAXON, "Sea ice altimeter processing scheme at the EODC," *International Journal of Remote Sensing*, vol. 15, no. 4, pp. 915–924, 1994.

[26] F. Mercier, N. Picot, J. Lambin, A. Lombard, D. Gerald, C. Dufau, L. Carrere, E. Obligis, S. Labroue, A. Michaël, P. Sicard, G. Garcia, T. Moreau, A. Cazenave, J. Bouffard, F. Seyler, P. Kosuth, and N. Bercher, "Improved jason-2 altimetry products for coastal zones and continental waters (PISTACH project)," *AGU Fall Meeting Abstracts*, p. 0662, Nov. 2008.

[27] F. Mercier, V. Rosmorduc, L. Carrere, and P. Thibaut, *Coastal and Hydrology Altimetry Product (PISTACH) Handbook*, ser. SALP-MU-P-OP-16031-CN. CNES, Oct. 2010.

[28] V. Enjolras and E. Rodriguez, "Using Altimetry Waveform Data and Ancillary Information From SRTM, Landsat, and MODIS to Retrieve River Characteristics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 6, pp. 1869–1881, Jun. 2009.

[29] R. Abileah, A. Scozzari, and S. Vignudelli, "Envisat RA-2 Individual Echoes: A Unique Dataset for a Better Understanding of Inland Water Altimetry Potentialities," *Remote Sensing*, vol. 9, no. 6, p. 605, Jun. 2017.

[30] E. R. Observation and S. E. Center, "Shuttle Radar Topography Mission (SRTM) Non-Void Filled," 2017.

[31] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, no. 7633, pp. 418–422, Dec. 2016.

[32] N. JPL, "NASA Shuttle Radar Topography Mission Water Body Data Shapefiles & Raster Files," 2013.

[33] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima," *arXiv:1609.04836 [cs, math]*, Sep. 2016.

[34] M. Vono, N. Dobigeon, and P. Chainais, "Asymptotically exact data augmentation: Models, properties and algorithms," *arXiv:1902.05754 [cs, stat]*, Feb. 2019.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016.

[36] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035.

[37] Y. Bengio, "Practical Recommendations for Gradient-Based Training of Deep Architectures," in *Neural Networks: Tricks of the Trade: Second Edition*, G. Montavon, G. B. Orr, and K.-R. Müller, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 437–478.

[38] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *The IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015.

[40] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML'10. Omnipress, 2010, pp. 807–814.

[41] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *CoRR*, vol. abs/1412.6980, 2014.

[42] L. N. Smith, "Cyclical Learning Rates for Training Neural Networks," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Santa Rosa, CA, USA: IEEE, Mar. 2017, pp. 464–472.

[43] I. Loshchilov and F. Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts," *arXiv:1608.03983 [cs, math]*, Aug. 2016.

[44] ——, "Fixing Weight Decay Regularization in Adam," *CoRR*, vol. abs/1711.05101, 2017.

[45] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[46] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv:1502.03167 [cs]*, Feb. 2015.

[47] S. Biancamaria, F. Frappart, A.-S. Leleu, V. Marieu, D. Blumstein, J.-D. Desjonquères, F. Boy, A. Sottolichio, and A. Valle-Levinson, "Satellite radar altimetry water elevations performance over a 200 m wide river: Evaluation over the Garonne River," *Advances in Space Research*, vol. 59, no. 1, pp. 128–146, Jan. 2017.

[48] R. Shwartz-Ziv and N. Tishby, "Opening the Black Box of Deep Neural Networks via Information," *CoRR*, vol. abs/1703.00810, 2017.

[49] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," *arXiv:1311.2901 [cs]*, Nov. 2013.

[50] S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," *arXiv:1803.01271 [cs]*, Mar. 2018.

[51] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You?": Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*. San Francisco, California, USA: ACM Press, 2016, pp. 1135–1144.

[52] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52 138–52 160, 2018.

[53] C. Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *arXiv:1811.10154 [cs, stat]*, Sep. 2019.

Fernando Niño Fernando Niño received his degree of Computer Science in 1990 from the Universidad de los Andes in Bogotá, and his Ph.D. in Geophysics in 1997 from the Univeersité de Montepellier. After working as research fellow on numerical modelling at the University of Edimburg, he was System Architect at a consulting firm for 2 years, before joining the Institute for Research and Developement (IRD) in France, where he was been project leader in scientific application development for ca. 10 years, and since 2009, chief technical officer of the a french altimetry center, the CTOH. His interests are in developing new techniques for radar altimetry, with coastal and continental hydrology applications in particular, of which deep learning is one he has been studying for several years.

Denis Blumstein Denis Blumstein received an engineering degree from the Ecole Centrale de Paris, France in 1982. He joined the French National Space Agency (CNES) in 1983, and was involved for 30 years in many space projects, mainly in the field of remote sensing. In 2011 he moved to LEGOS where his research interests are the use of satellite radar altimetry for continental hydrology, coastal, and glaciology applications. He proposed the SMASH idea to CNES and is now the project scientist for the SMASH Phase A study.

Clement Coggiola Clément Coggiola received his M.Sc. degree in electronics and signal processing from the French graduate school ENSEA, Cergy, France, in 2018. One semester of his M.Sc. was +dedicated to electronics and space technologies at NTNU, Trondheim, Norway. He worked on satellite radar altimetry for over a year at LEGOS, studying how satellite hydrology could benefit from deep learning. He is currently working as on-board payload data processing engineer at the French National Space Agency (CNES). His work focuses on implementing complex algorithms such as deep neural networks or payload data compression on embedded hardware as well as studying hardware/software co-design.