Hydrol. Earth Syst. Sci. Discuss., 9, 7801–7834, 2012
www.hydrol-earth-syst-sci-discuss.net/9/7801/2012/
doi:10.5194/hessd-9-7801-2012

Hydrology and Earth System Sciences Discussions

Automated global water mapping based on wide-swath orbital synthetic aperture radar

R. S. Westerhoff[1], **M. P. H. Kleuskens**[1,*], **H. C. Winsemius**[1], **H. J. Huizinga**[2], and **G. R. Brakenridge**[3]

[1]Deltares, Utrecht, The Netherlands
[2]HKV Consultants, Lelystad, The Netherlands
[3]University of Colorado, Boulder, Colorado, USA
[*]now at: Alten PTS, Eindhoven, The Netherlands

Received: 30 May 2012 – Accepted: 5 June 2012 – Published: 22 June 2012

Correspondence to: R. S. Westerhoff (rogier.westerhoff@deltares.nl)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This paper presents an automated technique, embedded in an online service, which ingests orbital synthetic aperture radar (SAR) imagery and outputs surface water maps in near real time and on a global scale. The service anticipates future open data dissemination of water extent information using the European Space Agency's Sentinel-1 data. The classification methods used are innovative but practical and different per 1 × 1 degree tile. For each tile, a probability distribution function of a pixel, being covered with water or being dry is established based on a long SAR training dataset. These probability distributions are conditional on the backscatter and the incidence angle. In classification mode the probability of water coverage is calculated, conditional on the current backscatter – incidence angle combination. The overlap between the probability distributions of a pixel being wet or dry is used as a proxy for the quality of our classification. The service has multiple uses, e.g. for water body dynamics in times of drought or for urgent inundation extent determination during floods. The service generates data systematically: it is not an on-demand service activated only for emergency response, but instead is always up-to-date and available. We demonstrate its use in flood situations using Envisat ASAR information during the 2011 Thailand floods. A first merge with a NASA near real time water product based on MODIS optical satellite imagery shows excellent agreement between these independent satellite-based water products.

1 Introduction

The consequences of inland and coastal flooding can be devastating and flooding needs to be detected and mapped as quickly as possible, so that appropriate measures can be taken by governments or disaster management agencies, pre-warnings may be issued, and downstream forecasts may be initiated (Carsell et al., 2004; Werner et al., 2005). In-situ networks of hydrological gauges are increasingly being complemented

HESSD
9, 7801–7834, 2012

by satellite imagery, which plays an important role in the European Global Monitoring of Environment and Security (GMES; Brachet, 2004) Emergency Response Core Service. The service is meant to provide "Rapid Mapping": fast retrieval of information from satellite imagery in order to map consequences related to hazards and civil protection.

Fast retrieval and systematic retrieval are different terms. Thus, a number of commercial and non-commercial agencies can respond to flood disasters within a short amount of time (fast retrieval). However, these agencies react only on demand and when an emergency response has already started. Also, due to the required manual expertise and labour requirements, such response cannot be accomplished on a daily basis and, commonly, not within a processing time comparable to the capture time of satellite images (Hostache et al., 2012). "Systematic" water mapping can instead be developed; wherein water extent information is routinely provided through the consistent and automated generation of maps and associated GIS (geographic information system) data. In surface water mapping, these maps can then be used within a GMES Service for different purposes, such as flood status, environmental monitoring of lake and reservoir extents or initializing hydrodynamic models. In this article we focus on the development and use of such an automated system specifically for use in flood response.

2 Relative advantages of SAR and optical imaging

The quality of C-band (e.g. Envisat) synthetic aperture radar (SAR) images is independent of the time of the day and cloud cover. Water can often be visually distinguished due to the low backscattering exhibited by relatively flat water surfaces (with very low return to the side-looking sensor due to specular reflection, also known as backscatter). In contrast, and although not capable of observing through clouds, the MODIS optical sensor on NASA's Terra and Aqua satellites has some important advantages: the two MODIS bands provide global, twice daily coverage at 250 m spatial resolution in two wavelengths, and optical multispectral classification methods may

HESSD
9, 7801–7834, 2012

better distinguish land and water in some areas, including in deserts, where general SAR backscatter may be very low and highly variable (a.o. Ridley et al., 1996; Raghavswamy et al., 2008). The utility of MODIS for flood-related work has been repeatedly demonstrated by maps disseminated from the Dartmouth Flood Observatory (http://floodobservatory.colorado.edu/). For example, these water area products are usefully compared to numerical 2-D model output in the case of catastrophic storm surges (Brakenridge et al., 2012). Improvements in wide-swath SAR data processing can be undertaken to the same end, as the addition of all-weather, day-night imaging capability provides a major benefit.

The anticipated data output from ESA's Sentinel-1 satellites (Attema et al., 2009) will further open opportunities. However, at present, semi-automatic classical water extraction techniques, such as thresholding or change detection applied on SAR images, may fail due to windy conditions, or partially submerged vegetation, resulting in higher backscattering values (Yesou, 2007; Prathumchai and Samarakoon, 2005). According to Silander et al. (2006) misclassifications may also be caused by the dependency of backscattering on detection angle. O'Grady et al. (2011) concludes that misclassification due to low backscatter values from non-flooded areas can be reduced via image differencing approaches. Matgen et al. (2011) present a method relying on the calibration of a statistical distribution of "open water" backscatter values inferred from SAR images of floods. Given the many circumstances that can affect classification results, it is difficult to derive a consistent classification technique that, ideally, also includes an error or accuracy assessment, and for all incidence angles. Up to the present, for example, some manual interpretation is still normally required to translate SAR data into water maps. However, Hostache et al. (2012) research an automated way of selecting the best reference image for change detection.

HESSD
9, 7801–7834, 2012

3 Need for automated data processing and map generation

Automation is required for any systematic mapping approach unless an army of human operators are kept employed. Consider the case of a single ESA Sentinel-1 satellite: (1) the expected amount of data for only Level 0 data across all acquisition regions will reach 320 TB per annum, amounting to 2.3 PB (petabytes) in the course of 7.5 yr (Snoeij et al., 2009 and Attema et al., 2008), and (2) when processed further, Hornacek et al. (2012) expects the matching Level 1 data volumes for baseline soil moisture products to be 4 to 5 times larger than those for Level 0. In order to cope with these amounts of data, the need for automation, and to fully utilize the very high information content of these new sensor data streams, new techniques are needed.

4 Methodology

We now present a prototype automated technique, embedded in an online service, which classifies SAR imagery to probability of water for each image pixel, in near real time and at global scale. The service used Envisat ASAR data while that sensor was operating (it failed 8 April 2012, after 10 yr), which was made available in Level-1 format by the European Space Agency (ESA) in near real time from a 15-day rolling archive. The data were processed in near real time (NRT), so within 3 h (but usually faster) after the data had been put on the ESA NRT Rolling Archive. Output results were subsequently placed on an open data server in open data formats (i.e. NetCDF and Google Earth KML files).

Previous employment of relatively high resolution (small swath) SAR for flood classification includes Kasischke et al. (1997) and McCandless and Jackson (2004). Two attributes of SAR that are of importance for the present algorithm are: backscatter and incidence angle.

Backscatter is the portion of the outgoing satellite radar signal – usually looking sideways in different incidence angles (as shown in Fig. 1) – that the target redirects back

HESSD
9, 7801–7834, 2012

towards the radar receiver antenna. If the target is horizontal, the backscatter is a measure of the electromagnetic roughness of the first very thin layer of the subsurface (a.o. Verhoest et al., 2008). As already widely known from ground penetrating radar and other microwave techniques, the electromagnetic roughness, creating a "subsurface microtopography" is depending on the physical contrasts between the conductivity and permittivity within this layer, causing a reflection coefficient: a measure of the reflective strength of a radar target. Usually for the solid Earth, this contrast is caused by differences in soil moisture; differences in soil type within this thin layer play a minor role (Beres and Haeni, 1980). The normalised measure of the radar return from a distributed target is the backscatter coefficient (σ^0), defined per unit surface area.

If the beamed radar is to the side of the sensor, instead of vertically downward, then an incidence angle applies, and a lesser amount of total energy returns to be recorded by the sensor Such radar backscatter is dependent on both the incidence angle α and on land cover, land topography, and soil moisture. Incidence angles in operating SAR sensors commonly range between between 15° (closest to the satellite) and 45° (furthest from the satellite), as shown in Fig. 1.

Our algorithm calculates the probability of a pixel within a satellite imaging swath being water, by matching its backscatter signal to a probability distributions of the pixel being dry, or being wet. These probability distributions are conditioned on geographic location, incidence angle and polarisation of the signal and were established using a training dataset of three years of Envisat ASAR data (Global Mode (GM), Wide Swath Mode (WSM), Image Mode (IMM) and Alternating Polarisation Mode (APM)). The probability distribution is distributed over each 1 × 1 latitude – longitude tiled dataset of the land covered globe. In general, for most incidence angles, the backscatter-incidence angle ($\sigma - \alpha$) pair of land is different than that from water. In practice, an empirical distribution function is estimated per geographical area by building 2-D histograms of $\sigma - \alpha$ pairs for (a) pixels within a 1 × 1 degree tile, which are permanently wet; and (b) pixels which are permanently dry.

HESSD
9, 7801–7834, 2012

An example of a trained histogram is shown in Fig. 2, where for the Netherlands thousands of $\sigma - \alpha$ WSM pairs have been gathered, smoothed and plotted for land and water in a land 2-D histogram (bottom panel) and a water 2-D histogram (top panel). The figure shows that backscatter characteristics for most incidence angles on land differ from the ones over water. These $\sigma - \alpha$ pair 2-D histograms can be used for classification.

5 Training method details

As noted, a training period is first used to derive a spatially distributed probabilistic model to distinguishing land and water. Then the application of this model in near real-time is accomplished.

First it should be noted that the parameter α as used here does not take into account local topographic features. Next, for building the histogram training set, an ancillary dataset is used, called the "water mask". The water mask is derived from the NASA SRTM Water Body map (SWBD), documented by USGS (2005) to be used for classification of water and land boundaries at the time of the Shuttle mission in February 2000. The SWBD divides the Earth in three types of classes:

- land (defined as –1);
- sea (defined as 1), which is not used in the training set;
- freshwater, consisting of large rivers and lakes (defined as 2).

During the training each combination of latitude, longitude, local incidence angle, polarization and backscatter in the SAR file is added to two possible training histograms, being either a land or a freshwater histogram. These histograms in turn consist of the following information:

- backscatter σ, in 28 discrete evenly distributed values of the value in dB in between 1.55 and 4.25;

HESSD
9, 7801–7834, 2012

- local incidence angle α, in 29 discrete evenly distributed values in between 15.5 and 43.5 degrees;
- polarization of the SAR signal, being either 0 = HH, 1 = HV, 2 = VH, 3 = VV;
- latitude information, in discrete evenly distributed values between –54.5 and 68.5 degrees;
- longitude information, in discrete evenly distributed values between –179.5 and 179.5 degrees.

For each 1 × 1 degree tile (in latitude – longitude) a histogram is made for land for discrete values of the backscatter and the local incidence angle, for each polarization. For the freshwater dataset, one global histogram file is made. The resulting multidimensional trained histograms consist of one or more hydrological years of SAR data. The trained land and water histograms are used as the reference set for classifying newly downloaded SAR data to land or water pixels. The training sets are built as separate entities for each SAR mode (e.g. ASAR-GM, ASAR-WSM, ASAR-IMM, ASAR-APM). It should be mentioned that by doing this, some 'noise' is created, since flood events that occur while building the training set are not filtered out.

6 Classification and quality assessment details

Because of the difference found in the $\sigma - \alpha$ pair 2-D histograms (for dry land and water), a distinction can be made between dry land and water. This is shown in a visual example in Fig. 3.

The probability that a pixel in a SAR dataset is wet or dry is established using Bayes' law, considering two empirical distribution functions for wet and dry pixels as posterior distributions. The procedure to establish a single probability of a pixel being wet is given below. All equations are written as if continuous probability distributions are used and

HESSD
9, 7801–7834, 2012

are applicable on a limited area within the earth's surface for which the set of empirical distributions for land/water apply.

Bayes' law in its general form can be written as

$$P\left[M|D\right] = \frac{P\left[D|M\right]P\left[M\right]}{P\left[D\right]} \tag{1}$$

where $P[M|D]$ represents the probability of a model M given demonstrative data D. $P[D|M]$ is the probability of data D occurring when model M applies and $P[M]$ is the prior distribution. In our case, we may write this as

$$P\left[s = w|b\right] = \frac{P\left[b|s = w\right]P\left[s = w\right]}{P\left[b\right]} \tag{2}$$

where $s = w$ means that a pixel s is classified as water (w) and b represents a certain $\sigma - \alpha$ pair. $P[b|s = w]$ is the probability that a certain $\sigma - \alpha$ combination is experienced when a pixel is classified as water. This probability distribution is approximated empirically based on discrete slices of the trained 2-D histograms per discrete incidence angle value as described in Sects. 4 and 5. $P[s = w]$ represents prior knowledge that the pixel within the SAR scene is water. Since we have no prior knowledge about this, and a pixel can only have two states (land or water), this probability is set on 0.5. Finally, $P[b]$ is the normalization constant. The same equation can be established for the probability that a pixel should be classified as dry land, being:

$$P\left[s = d|b\right] = \frac{P\left[b|s = d\right]P\left[s = d\right]}{P\left[b\right]} \tag{3}$$

where $s = d$ means that a pixel s is classified as dry/land.

Equations (2) and (3) both share the same denominator. Furthermore, both priors have the same value being 0.5. Therefore, the following applies:

$$P\left[s = w|b\right] = cP\left[b|s = w\right] \tag{4}$$

HESSD
9, 7801–7834, 2012

and

$$P\left[s = d|b\right] = cP\left[b|s = d\right]. \quad (5)$$

Furthermore, it is known that the sum of probabilities of a pixel being dry land or water is equal to unity, given that dry land or water are the only two states possible. Therefore we can write

$$c = \frac{1}{P\left[b|s = w\right] + P\left[b|s = d\right]}. \quad (6)$$

Substituting Eq. (6) in Eq. (4) gives:

$$P\left[s = w|b\right] = \frac{P\left[b|s = w\right]}{P\left[b|s = w\right] + P\left[b|s = d\right]}. \quad (7)$$

Equation (7) is used to determine the probability that a pixel is water. Finally, knowing the empirical probability distributions for dry land and water, we define a quality indicator q at a certain latitude, longitude and polarisation, as defined in Eq. (8):

$$q = \left[\left(\int P\left[b|s = w\right] \cup \int P\left[b|s = d\right]\right) - 1\right] \quad (8)$$

in which the shared area of the normal distributions of the land and water probabilities $P(b|s = d)$ and $P(b|s = w)$ are a measure for the quality of the probability calculation. In other words, when the two probability distributions overlap completely in a trained dataset of a 1 × 1 degree tile – which could for example happen at some low and intermediate local incidence angles and in very dry areas like deserts – q will be close to 0 %. If the two distributions are separated completely q will be 100 %. This indicator is dependent on backscatter, incidence angle and geographical location (latitude-longitude). It can be used as a post-processing tool to filter out data that is already pre-defined as inferior for calculating water probability by defining threshold values (e.g. to only show probabilities where the quality indicator higher than 70 %).

HESSD
9, 7801–7834, 2012

7 Creating a topography mask

With proper correction for topography, SAR classification methods can be improved in mountainous areas (a.o. van Zyl, 1993). Resulting errors in this topography correction will depend on the spatial resolution and quality of this topography data. Instead of correcting for topography, however, and because our concern is surface water, we have chosen to improve efficiency of the automated NRT calculation methods by using a preprocessing filter or mask, prior to classification to water probability. By using threshold values of the Height Above Nearest Drainage (HAND) index (Rennó et al., 2008), areas that are unlikely for long term flooding are filtered out. The HAND index is calculated by expressing the relative height of a location to its drainage outlet in an associated channel. It has now been calculated globally based on the HYDRO1k dataset, developed at the US Geological Survey (2008) and based on GTOPO30, a global Digital Elevation Model (DEM) at 30 arc second (approximately 1 km at the equator) resolution (Gesch et al., 1999). An example of the HAND index for Thailand is shown in Fig. 4.

8 Creating a Google Earth KML viewer

To enable an end-user to view relevant results from a planetary view (many floods affect very large land areas), as well as a detailed regional or local views, a smart visualisation is required. We investigated the use of Google Earth for this purpose. Google Earth's KML (Keyhole Markup Language) format allows for the use of scalable visualisation with different detail at different zoom levels. The result of this technique is that a user may view a low detail image of the whole globe when viewing Google Earth's globe from far above. As a user zooms in, more detailed sub-figures are downloaded and displayed; these show a small portion of the complete dataset. In this way, a user may view a large region, but only download the parts that are relevant. Tools from the open-source OpenEarth SVN repository (de Boer et al., 2012) were used to generate such tiled KML structures. Examples of global KML outputs are presented below.

HESSD
9, 7801–7834, 2012

Figure 5 shows a daily compilation, but multi-day compilations are also available. Figure 6 shows an example of a 10-daily compilation of all data.

9 Preliminary results

The downloaded SAR files are temporarily stored in (NetCDF) grid files. The trained histograms for each 1 × 1 degree (in latitude – longitude) tile have been generated for land for discrete values of the backscatter and the local incidence angle, for each polarization. Figure 7 shows a global compilation for ASAR Global Mode (GM) data and three examples of discrete histogram training sets, being rainforest, desert, and freshwater.

Different areas in the world show different backscatter characteristics. For example, desert (as well as savannah and mountainous regions) have a low backscatter, rainforest generally has a rather constant backscatter value, whereas ice (not shown in the figure) exhibits a very high backscatter. This regional dependency is the reason that the probability distributions of land are stored per latitude–longitude tile.

Water probabilities and quality indicators are calculated and stored together in (daily) folders containing all calculated water probabilities, and designed to be made publicly available. The water probabilities are made available in gridded (NetCDF) files of 10×10 degrees tiles. An example of an output in a resolution of 0.009 × 0.009 lat-lon degrees is shown in Fig. 8, where areas are shown where the water probability is set to 0 and the quality indicator to 100, independent if there is satellite data available. This is the result of the incorporation of the HAND filter, automatically setting water probabilities as a pre-processing step before classification.

10 Validation of the method in a case study, Thailand, 2011

We provide here the results of the algorithms as tested in an area in and north of Bangkok, Thailand. The region along the Chao Phraya River suffered from severe

HESSD
9, 7801–7834, 2012

flooding in the fall of 2011, caused by heavy rains. The area is rather flat, but surrounded by hills. During this flood, ESA decided to switch on the image mode (IMM) of the Envisat-ASAR sensor, in order to capture the flood in more detail. The best overpass of the satellite during the flood propagation over de Chao Phraya was on 13 October 2011. The resulting map generated by our algorithms, showing water data thresholded on probabilities higher than 70 % and quality higher than 70 %, is provided as a Google Earth flood map on the left of Fig. 9. During processing it was decided to not incorporate the HAND filter, in order to analyse what the best threshold is for this filter is in this region. Analysing the flood map in more detail, leads to a number of interesting observations:

- Elevated features, such as roads, embankments and railways can be distinguished in our image. These objects constrain the flood water and are marked by a sharp boundary between pixels detected as dry land, and pixels detected as flooded.
- Bangkok itself did not suffer extensive flooding. This qualitatively corroborates our result.
- Just East of Bangkok, there are many flooded rice paddies. These paddies and their borders are clearly visible in our image as square like features.

To show the predictive value of the HAND in Thailand, the thresholded detected flooding by the Global Flood Observatory on the 13 October 2011 using ASAR IMM is superimposed on the HAND image in Fig. 9 (right panel). The Thailand region shows a strong correlation between flooded areas and low HAND index values between 0 and 1.

HESSD
9, 7801–7834, 2012

11 Comparison/combination of ASAR and MODIS maps

The Dartmouth flood observatory (DFO) heavily utilizes the two MODIS sensors aboard the NASA Terra and Aqua satellites. Currently, a team at NASA is also assisting this effort by performing the classification procedure in an automated way (their NRT Flood product). In this automated process, the NRT processor collects and combines 4 images over each 10 degree latitude by 10 degree longitude subset, and over a forward running period of 2 days (thus, two images/day worldwide; four images/two days). The resulting GIS file shows surface water as boundary polygons: each such "daily" file actually includes two days of imagery, using a MODIS band 1/band 2 threshold approach to detect water and requiring at least two detections per pixel in order to exclude cloud shadows (which have similar spectral characteristics to water, but which change location over time).

Because the DFO approach does suffers from cloud cover and the present SAR-based approach provides less frequent temporal coverage, it is desirable to merge the two independent approaches for flood mapping. A first attempt to merge the two products was accomplished using IMM mode data from Envisat-ASAR, collected on 13 October 2011, and the DFO MODIS based data from 13 October 2011. Because DFO provides a binary map (flooded, or non flooded) and our Global Flood Observatory (GFO) produces a probability map of flooding, along with a quality indicator for this probability, it was decided to establish a binary map from the GFO product as well. This was done by thresholding the ASAR data as described in Sect. 10. The results are shown in Fig. 10. The image shows the high correlation between the two complementary and independent mapping methods. We conclude also that our algorithm is useful to detect floods in cloud-covered areas. In particular the most right image shows that there are highly plausible flood areas located by the SAR processor, which were cloud-obscured in the DFO product. These additional flood patterns clearly follow elevated features in the landscape such as roads and railways.

HESSD
9, 7801–7834, 2012

12 Improvements and present application of the topographic (HAND) index

The process to reach a topologically sound and accurate drainage network introduces occasional canyon-like artefacts into any DEM, as a result of aberrant height differences adjacent to the drainage network. These artefacts are transferred to the HAND grid during computation. An example of these artefacts is shown in detail in Fig. 11 in the area indicated by the red ellipse. We processed the currently used HAND indices from the HYDRO1k dataset, based on GTOPO30 data. In the currently used HAND we use the empirically based data threshold of 15 (i.e. all data higher than 15 will not be processed and set to 0 % water probability and 100 % quality indicator). This is mainly done to be "on the safe side": to prevent flood-prone areas with artefacts to be unfairly filtered out.

Rennó et al. (2008) state that when using original SRTM data for the HAND grid computation, these artefacts associated with the corrected DEM can be avoided. To filter more efficiently an improved version is recommended, using the HAND-data based on the HydroSHEDS 30 arc sec DEM (Lehner et al., 2006), in which the data are upgraded to streams that are burned less deep in the DEM.

Also in regard to topographic effects, the simplified explanation shown in Fig. 12 (left panel) shows that terrain slopes, when assuming a small swath width and thus neglecting the ellipsoid of the geoid, cause most of the difference between incidence angle and local incidence angle. Note that a wrongly assumed incidence angle causes a different backscatter returned to the satellite: the $\sigma - \alpha$ pair shifts in our 2-D histograms. Figure 12 (right panel) shows a slice of the 2-D histogram at a certain incidence angle for a certain place on the globe. Correcting for this slope will shift the position of $\sigma - \alpha$ pairs and cause a noticeable shift in the 2-D histograms. For the present algorithm, however, this improvement is not deemed efficient, as we already filter out non-flood prone areas. Removing all HAND values higher than 15 in fact means that the pixels we do use are never higher than 15 m above the nearest drainage point. The largest shift of incidence angle is thus expected directly near the drainage point.

HESSD
9, 7801–7834, 2012

Looking again at Fig. 12, and assuming a GFO pixel size of 1 × 1 km, the shift in incidence angle is

$$\alpha - \theta = \beta = \arctan \frac{z}{x} \tag{9}$$

where β is the terrain slope, α is the incidence angle, θ the local incidence angle, z the elevation and x the length for which the slope is calculated. The shift is less than 1° when z = 15 m and x = 1 km.

Lastly, in regard to topographic corrections: when working with 1 × 1 km pixel scales, we can consider the errors in global topography models with roughly the same resolution, such as GTOPO30 or the newer 30-arc second global mosaic of the Shuttle Radar Topography Mission (SRTM; USGS, 2004). According to Rodríguez at al. (2006) SRTM can give average absolute height errors per continent up to almost 10 m and locally even higher. Harding et al. (1999) indicate that GTOPO30 can reach even higher errors of 30 m. When we consider this error in z in Eq. (9), it is clear that a global topography model can also cause shifts in incidence angles of the same order and higher than the maximum shift we expect in our 2-D histograms after applying the HAND-index based filter. When furthermore taking into account computer processing efficiency, it is practical to avoid the correction to local incidence angle, at least until higher quality global digital elevation models are available.

13 Use of these algorithms for Sentinel-1 data

When looking at the NetCDF output of the water probabilities and quality indicator and using only Level-1 Envisat-ASAR GM1 data, the automated algorithm presented in this article generates a total daily output of approximately 500 MB without the temporary local storage needed for the processing sequences. This is slightly less and in the same order as the total daily data volume of the Level 1 GM1 data stored on the ESA Rolling Archive (in between 600–800 MB per day of GM1 data). The main reasons for this evenly large data size are:

HESSD
9, 7801–7834, 2012

– There is a “downscaling” from 500 × 500 m pixels to 1000 × 1000 m resolution, also commonly accepted as the “ground resolution”;

– wherever possible, numbers are stored in low-size formats (e.g. bytes instead of integers).

When we extrapolate the findings of data volumes to the expected Sentinel-1 data sizes, and we assume that a Level-1 base product is 4 times larger than the Level 0 data size (Hornacek et al., 2012), we estimate a total data size of 4 × 320 TB/365 d ∼ 3.5 TB of Level-1 data volume per day for all acquisition regions when assuming one Sentinel-1 satellite. If the further assumptions is made that half of these data consists of the main Interferometric Wide Swath (IW) mode, generating pixels of 5 × 20 m (Attema et al., 2009), and the pixel size of the NetCDF water probability and quality indicator output will be 20 × 20 m, the estimated data volume of a Sentinel-based NRT Global Flood Observatory (GFO) based on the presented algorithms will be approximately 0.44 TB per day, almost 1000 times more than the current output. Any changes in chosen spatial resolutions and foreseen data sizes will directly influence the size of the GFO output.

14 Conclusions

In preparation for the Sentinel-1 SAR satellite, and in order to address the urgent need for fast flood water detection and mapping, systematic and automated processing algorithms are needed. A binary product (e.g. water/not water or flooded/not flooded) is not optimum, as SAR-based classification products include noise and therefore an uncertainty indication is desirable. In this article an automated method to calculate probability of water, including a quality indicator, from Level-1 Envisat or Sentinel 1 SAR data is presented. The method is a new contribution, primarily because in our approach data over a certain time span are stored in 2-D histogram training sets in the incidence angle – backscatter domain and because the algorithms automatically calculate a water

HESSD
9, 7801–7834, 2012

probability and a quality index for each image pixel. Also, the application of the HAND index as a pre-processing filter improves the final result.

Water probabilities and quality indicators are calculated and stored in (daily) folders that can be made publicly available as soon as they are generated. The average total daily size when working with Envisat ASAR Global Mode data is ~0.5 GB. Projected to future Sentinel-1 data from one satellite, we calculate that the total data size would be ~0.44 TB d^{-1}, which is about 1000 times more.

A first merge with MODIS imagery in a case study in Thailand shows strong resemblance between the ASAR and MODIS derived results. At locations where MODIS suffers from clouds, ASAR shows hydrologically correct results, as observed through the clouds and as verified by other knowledge from the ground. We recommend more research in merging SAR and MODIS derived water imagery, to combine the strengths of both methods and improve the desired operational global surface water product.

Acknowledgements. We thank the Global Flood Observatory project team and especially Nicky Villars for flexible project management and ideas during the project within the Flood Control 2015 program. Furthermore, the merging of MODIS and Envisat ASAR data was done with the help of Dan Slayback (NASA Goddard Space Flight Centre). In this research, use was made of the following satellite sensors/products. Envisat ASAR NRT data (courtesy of ESA); SRTM (courtesy of NASA), MODIS (courtesy of NASA).

References

Attema, E., Davidson, M., Floury, N., Levrini, G., Rosich, B., Rommen, B., and Snoeij, P.: Sentinel-1 ESA's New European Radar Observatory, 7th European Conference on Synthetic Aperture Radar (EUSAR), 1–4, 2008.

Attema, E., Davidson, M., Snoeij, P., Rommen, B., and Floury, N.: Sentinel-1 mission overview, Int. Geosci. Remote S., 1, 36–39, 2009.

Beres, M. and Haeni, F. P.: Application of ground-penetrating radar methods in hydrogeologic studies, Ground Water, 29, 375–386, 1991.

HESSD

9, 7801–7834, 2012

Brachet, G.: From initial ideas to a European plan: GMES as an exemplar of European space strategy, Space Policy, 20, 7–15, doi:10.1016/j.spacepol.2003.11.002, 2004.
Brakenridge, G. R., Nghiem, S. V., Anderson, E., and Mic, R.: Orbital microwave measurement of river discharge and ice status, Water Resour. Res., 43, W04405, doi:10.1029/2006WR005238, 2007.
Brakenridge, G. R., Syvitski, J. P. M., Overeem, I., Higgins, S., Stewart-Moore, J. A., Kettner, A. J., and Westerhoff, R.: Global mapping of storm surges, 2002–present and the assessment of delta vulnerability, Nat. Hazards, in press, 2012.
Carsell, K. M., Pingel, N. D., and Ford, D. T.: Quantifying the Benefit of a Flood Warning System, Nat. Hazards Rev., 5, 131–140, doi:10.1061/(ASCE)1527-6988(2004)5:3(131), 2004.
de Boer, G. J., Baart, F., Bruens, A., Damsma, T., van Geer, P., Grasmeijer, B., den Heijer, K., and van Koningsveld, M.: OpenEarth: using Google Earth as outreach for NCK's data, University of Twente, Department of Water Engineering and Management, available at: http://proceedings.utwente.nl/177/ (last access: 27 March 2012), 2012.
De Grandi, G. D., Bouvet, A., Lucas, R. M., Shimada, M., Monaco, S., and Rosenqvist, A.: The K&C PALSAR Mosaic of the African Continent: Processing Issues and First Thematic Results, IEEE T. Geosci. Remote, 49, 3593–3610, 2011.
Dekker, R. J.: Speckle filtering in satellite SAR change detection imagery, Int. J. Remote Sens., 19, 1133–1146, 1998.
European Space Agency: ASAR User Guide, Chapter 5: Geometry Glossary, http://envisat.esa.int/handbooks/asar/CNTR5-5.htm (last access: June 2012), 2007.
Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D.: The shuttle radar topography mission, Rev. Geophys., 45, RG2004, 1-33, doi:10.1029/2005RG000183, 2007.
Gesch, D. B., Verdin, K. L., and Greenlee, S. K.: New land surface digital elevation model covers the Earth, Eos, Transactions, American Geophysical Union, 80, 69–70, 1999.
Harding, D. J., Gesch, D. B., Carabajal, C. C., and Luthcke, S. B.: Application of the shuttle laser altimeter in an accuracy assessment of GTOPO30, A global 1-kilometer digital elevation model, Proceedings of the International Society for Photogrammetry and Remote Sensing, Volume XXXII-3, 9–11 November 1999, La Jolla, USA, 1999.

HESSD
9, 7801–7834, 2012

Hong, Y., Adler, R., Huffman, G., and Pierce, H.: Applications of TRMM-Based multi-satellite precipitation estimation for global runoff prediction: prototyping a global flood modeling system, in: Satellite Rainfall Applications for Surface Hydrology, edited by: Gebremichael, M. and Hossain, F., Springer, 245–265, 2010.

Hornacek, M., Wagner, W., Sabel, D., Truong, H.-L., Snoeij, P., Hahmann, T., Diedrich, E., and Doubkova, M.: Potential for high resolution systematic global surface soil moisture retrieval via change detection using Sentinel-1, IEEE J. Sel. Top. Appl., submitted, 2012.

Hostache, R., Matgen, P., and Wagner, W.: Change detection approaches for flood extent mapping: How to select the most adequate reference image from online archives?, Int. J. Appl. Earth Obs., submitted, 2012.

Kasischke, E. S., Melack, J. M., and Dobson, M. C.: The use of imaging radars for applications – a review, Remote Sens. Environ., 59, 141–156, 1997.

Lehner, B., Verdin, K., and Jarvis, A.: HydroSHEDS Technical Documentation. World Wildlife Fund US, Washington, D.C., available at http://hydrosheds.cr.usgs.gov (last access: September 2010), 2006.

Matgen, P., Hostache, R., Schumann, G., Pfister, L., Hoffmann, L., and Savenije, H.: Towards an automated SAR-based flood monitoring system: Lessons learned from two case studies. Phys. Chem. Earth, 36, 241–252, 2011.

McCandless, S. W. and Jackson, C. R.: Principles of Synthetic Aperture Radar, SAR NOOA Marine Users Manual, Chapter 1, available at: www.sarusersmanual.com/ManualPDF/NOAASARManualCH01pg001-024.pdf (last access: January 2009), 2004.

O'Grady, D., Leblanc, M., and Gillieson, D.: Use of ENVISAT ASAR Global Monitoring Mode to complement optical data in the mapping of rapid broad-scale flooding in Pakistan, Hydrol. Earth Syst. Sci., 15, 3475–3494, doi:10.5194/hess-15-3475-2011, 2011.

Prathumchai, K. and Samarakoon, L.: Application of remote sensing and GIS techniques for flood vulnerability and mitigation planning in Munshiganj district of Bangladesh, Proceedings of the 25th Asian Conference on Remote Sensing, Hanoi, Vietnam, 2005.

Raghavswamy, V., Gautam, N. C., Padmavathi, M., and Badarinath, K. V. S.: Studies on microwave remote sensing data in conjunction with optical data for land use/land cover mapping and assessment, Geocarto International, 11, 2008.

Ridley, J., Strawbridge, F., Card, R., and Phillips, H.: Radar backscatter characteristics of a desert surface, Remote Sens. Environ., 57, 63–78, 1996.

HESSD
9, 7801–7834, 2012

Rodríguez, E., Morris, C. S., and Belz, J. E.: A global assessment of the SRTM performance, Photogramm. Eng. Rem. S., 72, 249–260, 2006.

Silander, J., Aaltonen, J., Sane, M., Malnes, E.: Flood maps and satellite, case study Kittilä, in: XXIV Nordic Hydrological Conference, Vingsted, Denmark, 69, edited by: Refsgaard, J. C. and Højberg, A. L., available at: http://wwwi9.ymparisto.fi/i9/en/110_Silander_final2.pdf, last access: 4 August 2006.

Small, E., Meier, E., and Nuesch, D.: Robust radiometric terrain correction for SAR image comparisons, IEEE International Proceedings to the Geoscience and Remote Sensing Symposium, 20–24 September 2004, 3, 1730–1733, 2004.

Snoeij, P., Attema, E., Davidson, M., Duesmann, B., Floury, N., Levrini, G., and Rommen, B.: The Sentinel-1 Radar Mission: Status and Performance, Radar Conference – Surveillance for a Safer World, Bordeaux, 1–6 October 2009.

Sohn, H. G. and Song, Y. S.: Water Area During Flood Event from SAR Image, chapter in: Computational Science and Its Applications – ICCSA 2005, Lecture Notes in Computer Science, 3481, 30–31, doi:10.1007/11424826_82, 2005.

US Geological Survey Center for Earth Resources Observation and Science (EROS): HYDRO1k Elevation Derivative Database, available at: http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/GTOPO30/hydro (last access: August 2011), 2008.

USGS: Shuttle Radar Topography Mission, 30 Arc Second Global Mosaic, Global Land Cover Facility, University of Maryland, College Park, Maryland, 2004.

USGS: Documentation for the Shuttle Radar Topography Mission Water Body Data Files, available at: http://dds.cr.usgs.gov/srtm/version21/SWBD/SWBDDocumentation/Readme_SRTM_Water_Body_Data.pdf, last access: September 2005.

van Zyl, J. J., Chapman, B. D., Dubois, P., and Shi, J.: The effect of topography on SAR calibration, IEEE T. Geosci. Remote, 31, 1036–1043, 1993.

Verhoest, N., Lievens, H., Wagner, W., Álvarez-Mozos, J., Susan Moran, M., and Mattia, F.: On the soil roughness parameterization problem in soil moisture retrieval of bare surfaces from synthetic aperture radar, Sensors, 8, 4213–4248, doi:10.3390/s8074213, 2008.

Werner, M., Reggiani, P., De Roo, A., Bates, P., and Sprokkereef, E.: Flood forecasting and warning at the River Basin and at the European Scale, Nat. Hazards, 36, 25–42, doi:10.1007/s11069-004-4537-8, 2005.

HESSD

9, 7801–7834, 2012

Yesou, H., Li, J., Malosti, R., Andreoli, R., Huang, S., Xin, J., and Cattaneo, F.: Near real time flood monitoring in p.r. China during the 2005 and 2006 flood and typhoons seasons based on envisat asar medium and high resolution images, Proc. "Envisat Symposium 2007", ESA SP-636, Montreux, Switzerland 23–27 April 2007.

Yilmaz, K. K., Adler, R. F., Tian, Y., Hong, Y., and Pierce, H. F.: Evaluation of a satellite-based global flood monitoring system, Int. J. Remote Sens., 31, 3763–3782, 2010.

HESSD

9, 7801–7834, 2012

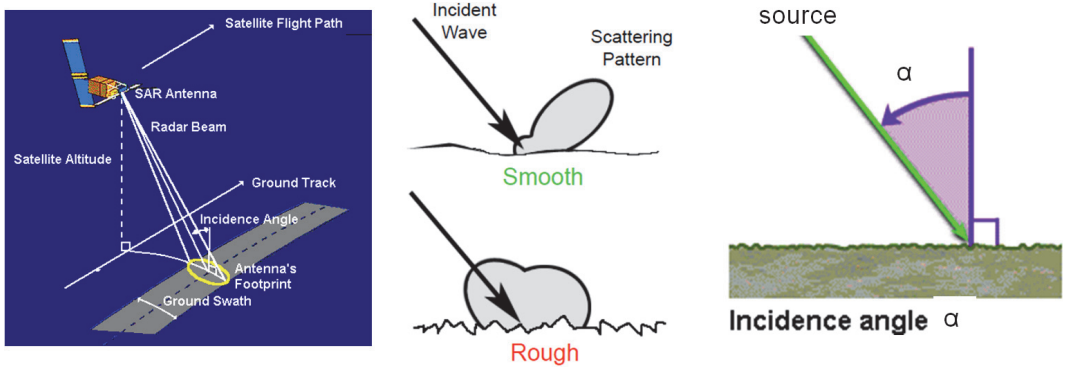

Fig. 1. The backscatter is the portion of the outgoing satellite radar signal, usually looking sideways (left panel) in different incidence angles α (right panel) is highly depending on the backscatter characteristics of the subsurface (middle panel). Adapted from ESA (2012).

HESSD
9, 7801–7834, 2012

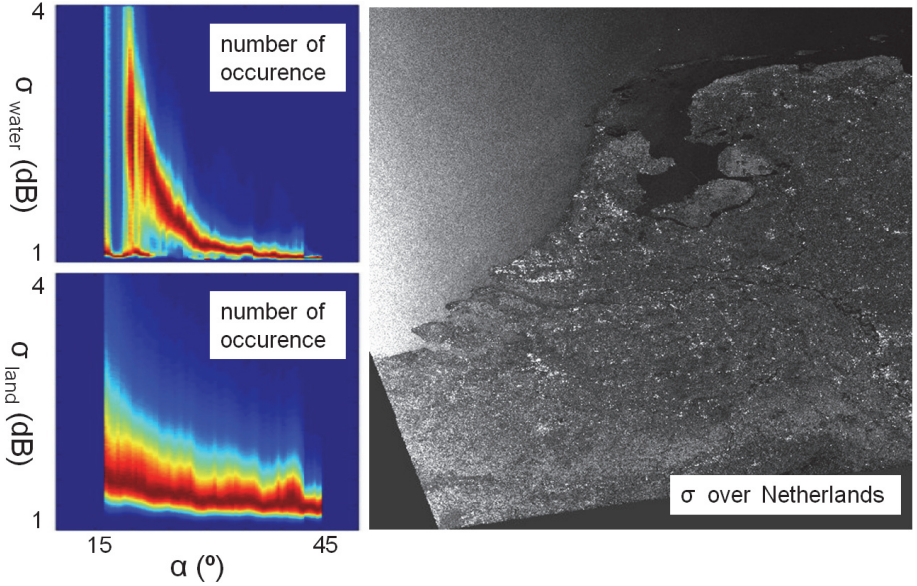

Fig. 2. Trained and smoothed 2-D histograms for land and water (left panels) as derived from two years of WSM backscatter data in the Netherlands (right panel).

HESSD
9, 7801–7834, 2012

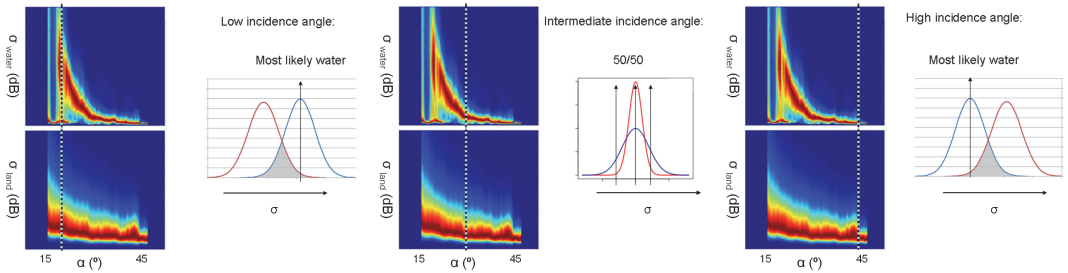

Fig. 3. Explanation of the ability to distinguish between water and land for different incidence angles.

HESSD
9, 7801–7834, 2012

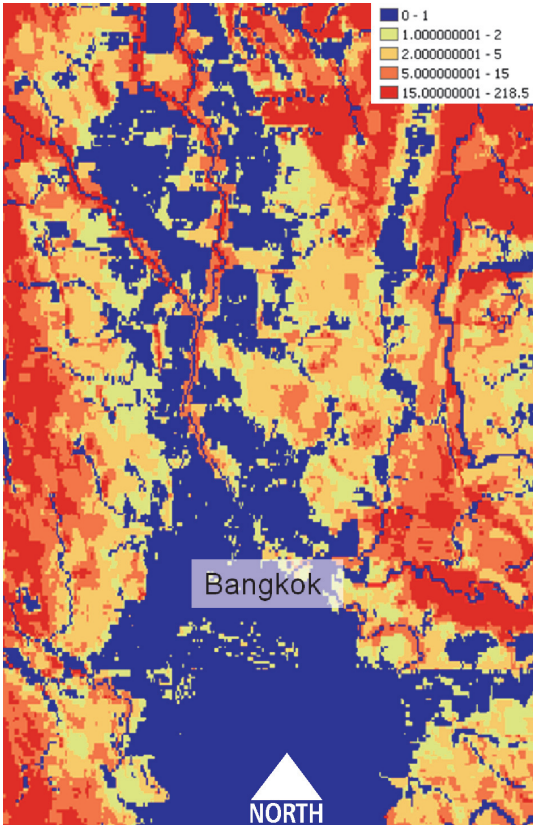

Fig. 4. The Height Above Nearest Drainage (HAND) Index for an example location in Thailand.

HESSD
9, 7801–7834, 2012

Fig. 5. Water probabilities plotted in Google Earth, collected over 24 h from 5–6 October 2011 over an example location in Kazakhstan. In the more transparent pixels the ability of the GFO algorithm to distinguish dry land from water, expressed in the quality indicator q, is lower.

HESSD
9, 7801–7834, 2012

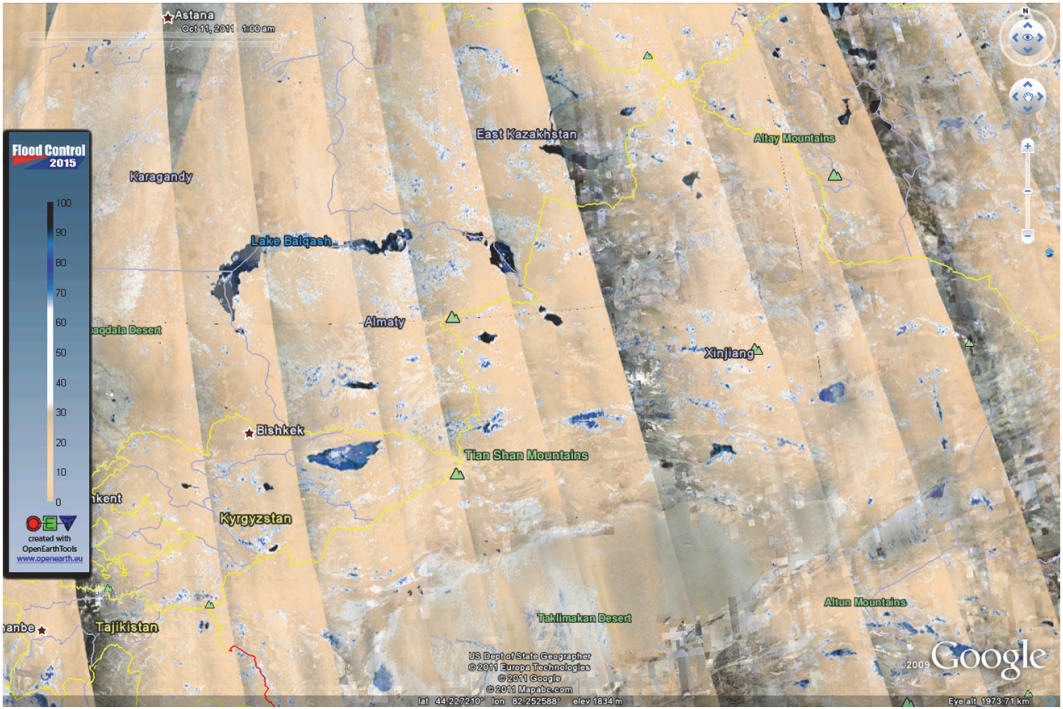

Fig. 6. Water probabilities, similar to Fig. 5, but now presenting the coverage from 1 until 10 October 2011.

HESSD
9, 7801–7834, 2012

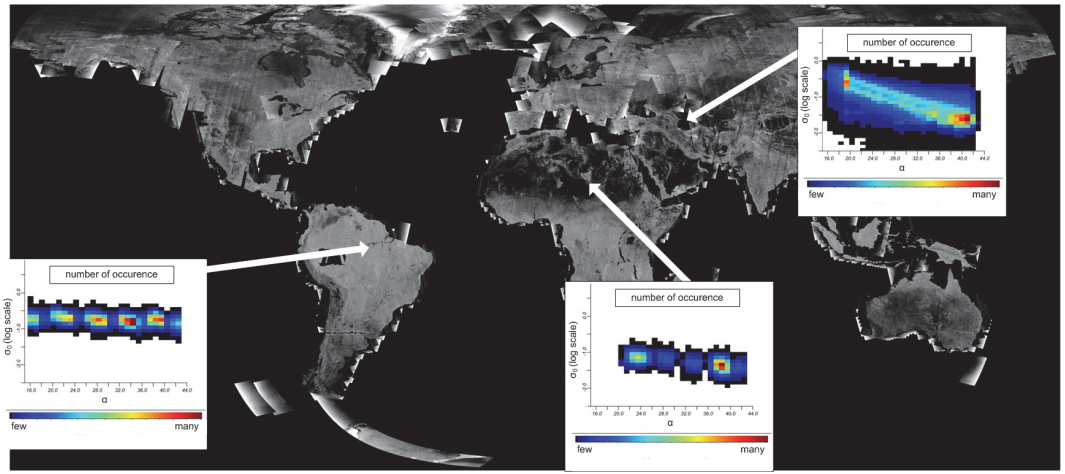

Fig. 7. Different histograms as shown on a global backscatter map. Different areas in the world show different backscatter characteristics. In this figure, the backscatter characteristics for rain forest (left panel), desert (middle panel) and lakes (right panel) are shown.

HESSD
9, 7801–7834, 2012

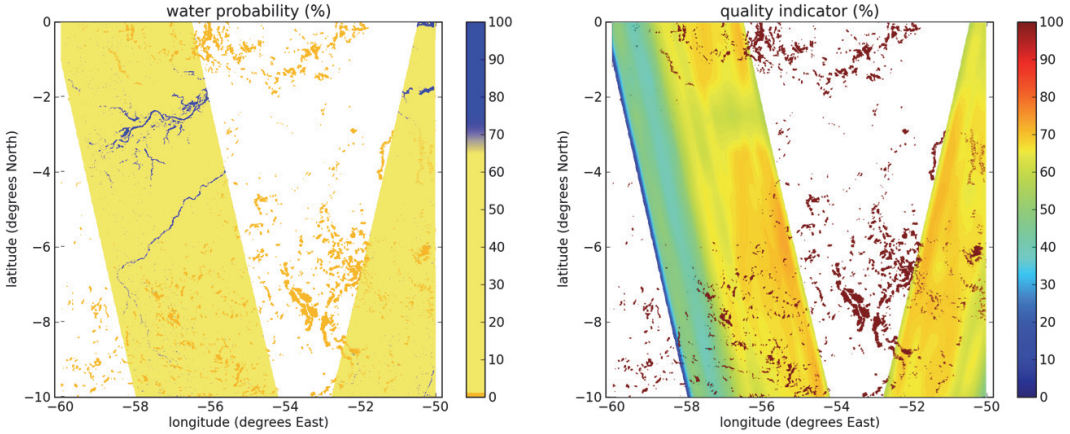

Fig. 8. Output of a single day 10 × 10 lat-lon tile of water probability (left panel) and the corresponding quality indicator q. Places where the HAND index are higher than 15 are set to $q = 100$, water probability = 0 in a pre-processing phase.

HESSD
9, 7801–7834, 2012

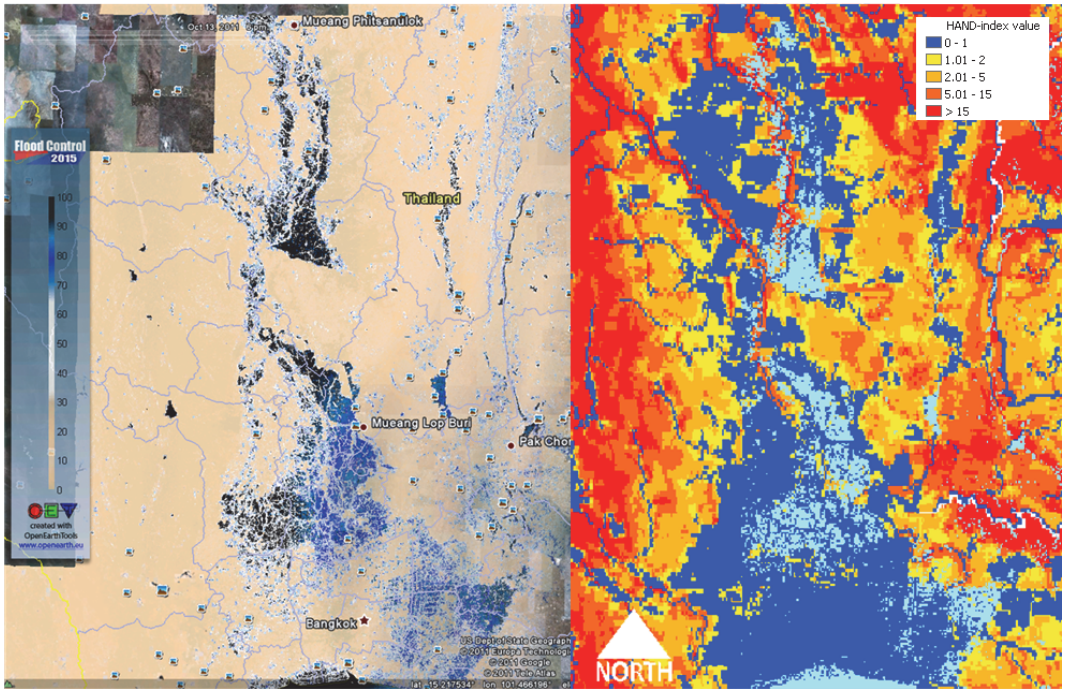

Fig. 9. Results of the GFO water probabilities over the Chao Phraya basin on 13 October 2011 as shown in Google Earth (left panel). HAND map with flooded areas (light blue) are shown on the right panel.

HESSD
9, 7801–7834, 2012

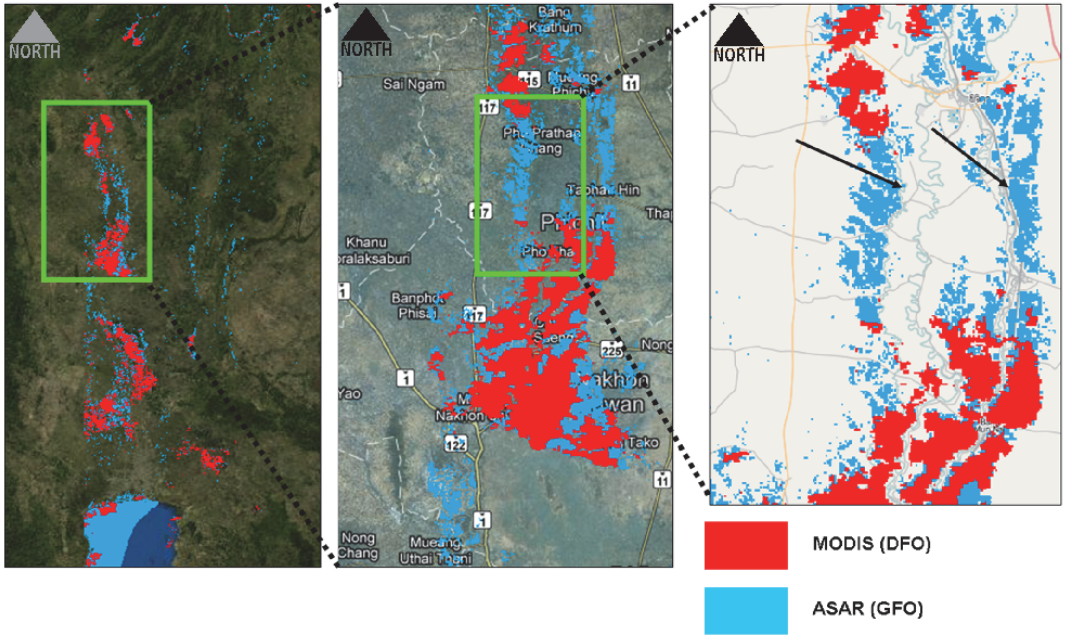

Fig. 10. The Chao Phraya basin at different zoom levels. Left panel: largest extent. Towards right panel: zoomed extents. Flood classification from Envisat ASAR (IMM) is shown in blue and is underlying the flood classification based on MODIS in red.

HESSD
9, 7801–7834, 2012

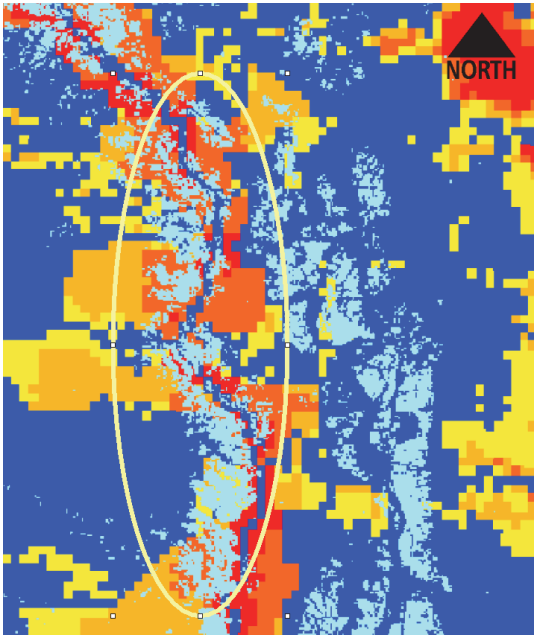

Fig. 11. Detail of artefact along the river in the HAND index.

HESSD
9, 7801–7834, 2012

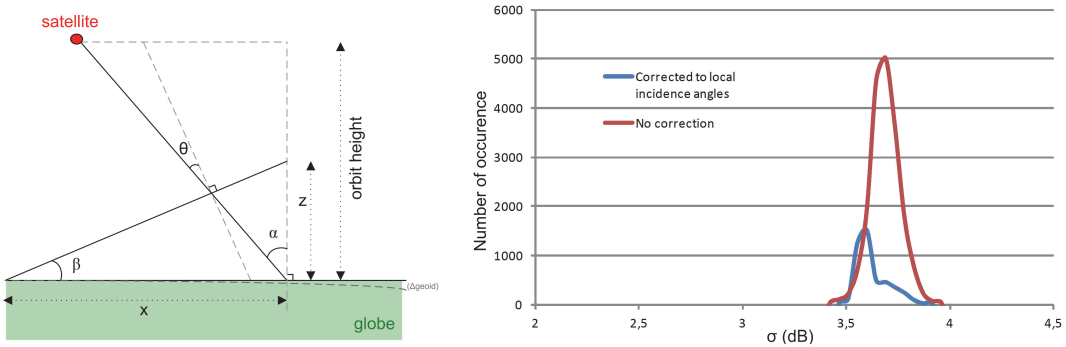

Fig. 12. Left panel: the difference between incidence angle α and local incidence angle θ is mainly caused by the terrain slope β. Right panel: a slice of the 2-D histogram at $\alpha = 32.5°$ (location lat – lon = 27.5 – 87.5). Corrections for the slope will shift the position of $\sigma - \alpha$ pairs.

HESSD
9, 7801–7834, 2012