

SEMANTIC SEGMENTATION OF WATER BODIES IN MULTI-SPECTRAL SATELLITE IMAGES FOR SITUATIONAL AWARENESS IN EMERGENCY RESPONSE

M. Wieland*, S. Martinis, Y. Li

German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Oberpfaffenhofen, D-82234 Wessling, Germany
- marc.wieland@dlr.de

KEY WORDS: Convolutional neural network, Semantic segmentation, Flood disaster, Emergency response, Multi-spectral

ABSTRACT:

Satellite-based crisis information is frequently requested in the context of flood disasters to gain rapidly situational awareness and to prioritize response actions under often limited resources during emergency response. To assure that information products have the highest possible spatial, temporal and thematic resolutions, it is critical to be able to simultaneously analyse data from a large variety of satellite sensors. In this contribution, we present a solution to rapidly extract water bodies from Landsat TM, ETM+, OLI and Sentinel-2 for up-to-date situational awareness during emergency response. A convolutional neural network is used to segment water extent in these images, while clouds, cloud shadows and snow / ice are specifically handled by the network to remove potential bias from any downstream analysis. Atmospheric correction, post-processing and ancillary data are not required. To distinguish flood from permanent water we present a reference water mask that is derived by means of time-series analysis of archive imagery. Compared to widely-used mono-temporal reference water masks, it can be adapted to any area and time of interest. This study builds up on previous work of the authors and presents new results from recent flood disasters in Germany, Peru, China, India and Mozambique, as well as a flood monitoring application centred on the Indian state of Kerala. The processing chain produces very high overall accuracy and Kappa coefficient (>0.87) and shows consistent performance throughout a monitoring period of 12 months that covers 143 Landsat OLI and Sentinel-2 images.

1. INTRODUCTION

Near-real time information about flood water extent is an important input to decision making during flood disasters. Emergency responders increasingly request satellite images to support flood mapping and monitoring over large-areas (Martinis et al., 2017). In this context, Synthetic Aperture Radar (SAR) data are widely used due to their capability of penetrating through clouds and acquiring images during day and night (Schumann and Moller, 2015). However, to guarantee the highest possible spatial and temporal resolutions of the mapping products, any satellite-based flood monitoring system should be able to process data from various platforms and sensors (Lin et al., 2016). Therefore, multi-spectral satellite images are an important supplement, despite their sensitivity to atmospheric effects and the presence of clouds.

Existing methods for water segmentation from multi-spectral satellite images can broadly be categorized into rule-based and machine learning methods. Rule-based methods present good results for single sensors, but largely lack generalization ability and transferability between sensors, geographies and scene properties (Feng et al., 2016; Zhou et al., 2017). Classical machine learning methods can overcome some of these limitations by learning characteristics of water pixels from a set of labelled samples across a hand-crafted feature space at pixel- or object-level (Ko et al., 2015; Mueller et al., 2016). However, their generalization ability at global scale and across images from different sensors has proven difficult. Recently, studies that use Convolutional Neural Networks (CNNs), which learn

features directly from raw images by combining convolutional and pooling layers, reported superior accuracy and generalization ability compared to rule-based and classical machine learning approaches with hand-crafted features (Chen et al., 2018; Isikdogan et al., 2017).

To differentiate between temporarily flooded areas and water bodies of normal water extent (permanent water) in a single image proved to be unreliable even when performed manually by expert analysts. Hence, it is common to compare a co-flood image with a pre-flood image or an independent reference water mask (Franci et al., 2015; Twele et al., 2016). These methods are generally fast to compute but depend on the quality and date of the pre-flood dataset, which may strongly bias the flood water delineation. This becomes particularly problematic in locations with highly dynamic surface waters. A more reliable estimation would be to consider a time range, which identifies permanent water as being continuously present throughout an observation period (Pekel et al., 2016).

In this study, we present a solution to rapidly segment water bodies in Landsat TM, ETM+, OLI and Sentinel-2 images for up-to-date situational awareness during emergency response. To distinguish flood water from permanent water we introduce an adaptive reference water mask that is derived through time-series analysis of archive imagery. This study builds up on previous work of the authors (Wieland et al., 2019; Wieland and Martinis, in review) and presents new results from recent flood disasters as well as a flood monitoring application centred on the Indian state of Kerala.

* Corresponding author

2. DATA AND STUDY AREA

We compile a global multi-sensor reference dataset for training, validation and testing of the CNN for water segmentation (Figure 1). We apply a stratified random sampling on the basis of a global biomes map (Olson et al., 2001) to derive samples that are representative for a variety of climatic, atmospheric and land-cover conditions. For each of the 14 biomes we pick one sample location, acquire Landsat-TM, -ETM+, -OLI and Sentinel-2 images and manually delineate thematic masks. We add the freely available Spatial Procedures for Automated Removal of Cloud and Shadow (SPARCS) dataset, which consists of Landsat OLI images with respective thematic masks (Hughes and Hayes, 2014). We resample all imagery to 30 m spatial resolution, stack the image bands together (Red, Green, Blue, Near-Infrared (NIR), Shortwave-Infrared (SWIR)-1 and SWIR-2) and convert Digital Numbers (DN) to Top of Atmosphere (TOA) reflectance. We split all images and masks into non-overlapping tiles with 256 x 256 pixels size, shuffle and distribute them into training (60 %), validation (20 %) and testing (20 %) parts. The training dataset is augmented with random contrast, brightness, and rotation. Thematic masks consider classes “Water”, “Land”, “Snow / Ice”, “Cloud” and “Cloud shadow”. The final dataset covers 94 locations, 136 images from four different sensors, and is split into 1,075 tiles for training (5,375 tiles with augmentation) and 358 tiles for validation and testing respectively.

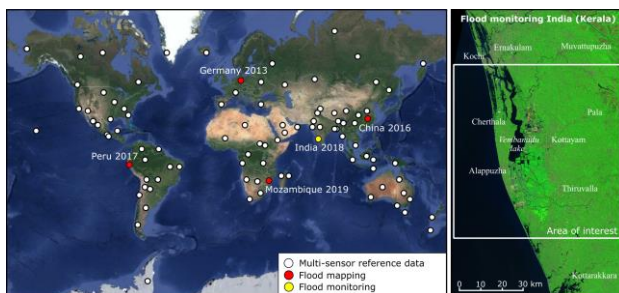


Figure 1. Spatial distribution of reference data, flood mapping tests and flood monitoring application.

To further evaluate the performance of water segmentation and flood mapping methods, we select five major flood disasters, for which we acquire satellite images and derive independent test data in the form of manually labelled point samples. Namely these are Germany (June, 2013), China (June, 2016), Peru (March, 2017), India (August, 2018) and Mozambique (March, 2019).

For a systematic flood monitoring application all available Landsat OLI (23 images) and Sentinel-2 data (140 images) are acquired, processed and analysed over an area of interest (80 x 80 km) in Kerala, India during 12 months in 2018 (Figure 1). Kerala is located in southwest India, has a wet and maritime tropical climate and is seasonally affected by monsoon rains. The state receives an average annual rainfall of nearly 3,000 mm, of which around 65% fall between June and August (southwest monsoon) and the rest between September and December (northeast monsoon). The majority of rivers originate in the Western Ghats and flow towards the sea. Their water discharge is entirely led by monsoon rains. Hence, the main contributing factor to disastrous floods is high intensity rainfall. Dominating land-use / land-cover classes include small scale agricultural fields with a variety of crops and growing patterns, patches of forest, wetlands as well as urban and rural

settlements. Water body types include sea, rivers, ponds, natural and artificial lakes and waterways. Particular for the study area are the backwaters, which are an interconnected system of brackish water lakes and river estuaries that stretches along the state’s coastline.

3. METHOD

3.1 Water segmentation

For semantic segmentation of water bodies we use an encoder-decoder CNN with U-Net architecture, which achieved state-of-the-art results on semantic segmentation benchmark datasets, while learning from very little data (Ronneberger et al., 2015). The encoder takes as input a multi-band image of size 256 x 256 pixels and feeds it through five convolutional blocks. Starting from 32, the number of feature channels is doubled per block. The basic convolutional block consists of two 3 x 3 convolutions with Rectified Linear Unit (ReLU) activation function, batch normalization and 2 x 2 max pooling. In the decoder part, the feature map is up-sampled by a 2 x 2 transpose convolution followed by a concatenation with the correspondingly cropped feature map from the decoder and two 3 x 3 convolutions with ReLU activation and batch normalization. At the final layer a 1 x 1 convolution with softmax activation function is used to map each feature vector x to the number of classes. The categorical output y is computed by maximizing the predicted probability vector $p(x) = \{p_1(x), p_2(x), \dots, p_l(x), \dots, p_n(x)\}$.

As loss function we use weighted categorical cross-entropy, where the weight vector $w \in \mathbb{R}$ is defined over the range of class labels $i \in \{1, 2, \dots, l\}$ and is computed on the training dataset for each class as the ratio of the median class frequency and the class frequency. The input image feature space is standardized to zero mean and unit variance with mean and standard deviation being computed on the training dataset and applied to the validation and testing datasets. We optimize the weights during training using the adaptive moment estimation algorithm with default hyper-parameters $\alpha = 10^{-4}$, $\beta_1 = 0.9$ and $\beta_2 = 0.999$ (Kingma and Lei, 2015). Additionally, we step-wise reduce α by a factor of 0.5 if no improvement is seen for five epochs. We track loss, Overall Accuracy (OA), Cohen’s Kappa and Dice coefficient for model evaluation. The network is trained in batches of 20 until convergence on a NVIDIA M4000 GPU using Keras with Tensorflow backend as deep learning framework (Figure 2).

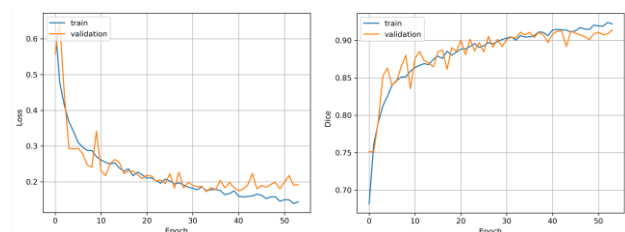


Figure 2. U-Net training history showing training and validation loss and dice coefficient.

U-Net predicts on small local windows, which may lead to reduced confidence towards the image borders. To overcome this limitation, we expand the image with mirror-padding, split it into overlapping tiles, run the predictions over batches of tiles, and finally blend the prediction tiles with a tapered cosine window function to generate a seamless segmentation output.

3.2 Flood mapping

To map temporarily flooded areas we derive a reference water mask that aims at approximating the normal water extent, which we define in the following as areas that are permanently flooded within a specified time-period. Therefore, we use time-series of pre-flood images, for which we run the water segmentation. The categorical output y of any input image in the time-series is reclassified to produce binary valid pixel and water masks. The sum of water masks over the sum of valid pixel masks computes the relative water frequency f , with which water is present throughout the observation time-period. This means that a pixels would be classified as being permanently flooded, if it belongs to class “Water” in all valid observations, which would result in a relative water frequency $f = 1.0$. According to the accuracy of the U-Net model for class “water” on the test dataset (Table 1), we relax this threshold and set it to $f \geq 0.9$. This means that a pixel is identified as being permanent water if it belongs to class “water” in 90 % or more of the valid observations during the observation time-period. Finally, the set difference of the water mask derived for the flood image and the reference water mask maps the temporarily flooded areas. More details can be found in Wieland and Martinis (in review).

4. RESULTS

Over the globally distributed 358 test image tiles for Landsat TM, ETM+, OLI and Sentinel-2, it can be seen that the U-Net model produces highly accurate results with Overall Accuracy (OA) of 0.93 and Kappa of 0.87 (Table 1). In particular, “Water”, “Land” and “Cloud” classes show very high precision and recall across all test tiles and sensors. The least accurate class is “Cloud shadow”, which is mainly affected by confusions with the “Land” class.

Class	Precision	Recall	F1-score
Water	0.95	0.87	0.91
Land	0.94	0.97	0.96
Snow / Ice	0.84	0.95	0.89
Cloud	0.94	0.91	0.92
Cloud shadow	0.79	0.72	0.75
Total	0.93	0.93	0.93
OA	0.93		
Kappa	0.87		

Table 1. Prediction results of the U-Net model for water segmentation on 358 test image tiles for Landsat TM, ETM+, OLI and Sentinel-2.

Furthermore, we analyse five globally distributed flood disasters to evaluate the results of the model on flood water and to exemplify the flood mapping method. For each of the floods we acquire independent test data from manually labelled point samples. The samples are derived by means of a stratified random sampling using the categorical output of the model as strata with a fixed size of 100 samples per class. Figure 3 shows the input images and respectively predicted flood maps, whereas Table 2 depicts the results of the accuracy assessment for each flood disaster. Despite highly varying environmental and atmospheric conditions, land-use / land-cover patterns, sensors, locations and seasons, our model performs consistently well ($OA \geq 0.92$ and $Kappa \geq 0.90$) across the test images.

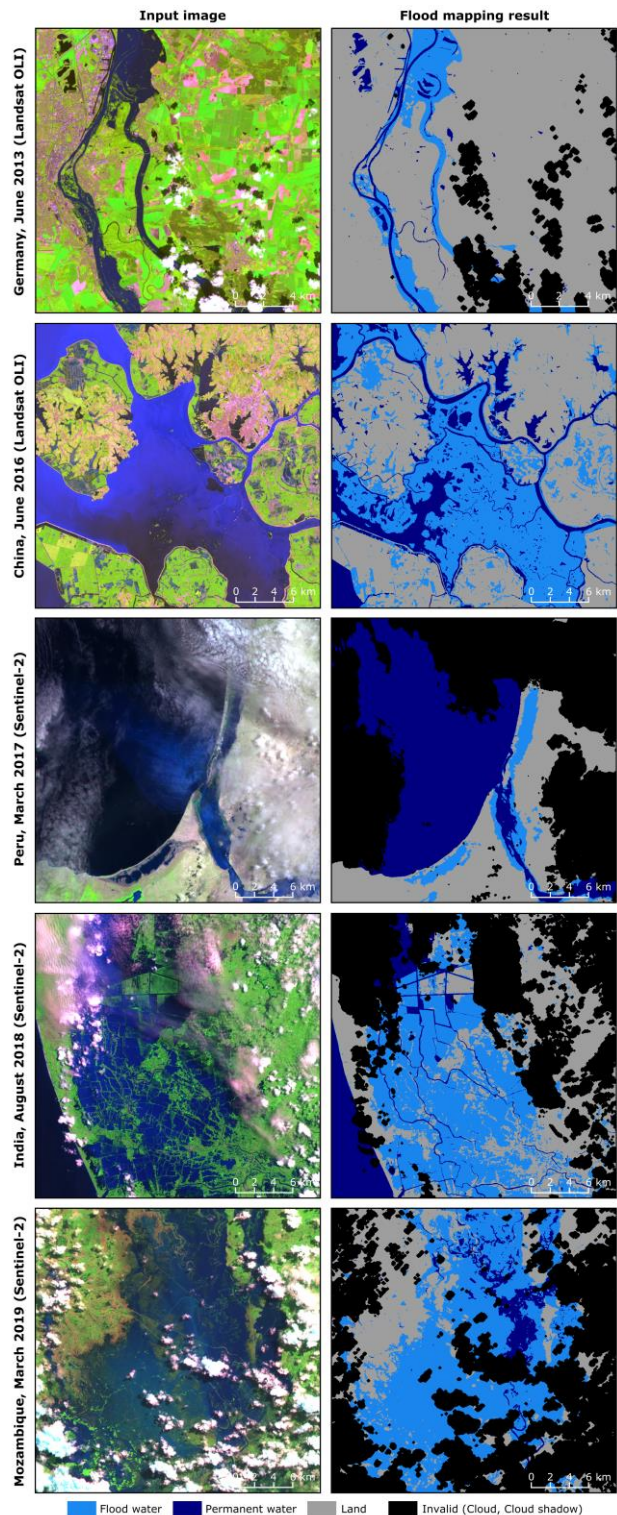


Figure 3. Test images of flood disasters with respective results of the flood mapping method.

Flood disaster	OA	Kappa
Germany, June 2013	0.92	0.90
China, June 2016	0.94	0.92
Peru, March 2017	0.98	0.97
India, August 2018	0.94	0.92
Mozambique, March 2019	0.95	0.94

Table 2. Accuracy assessment for different flood disasters.

Figure 4 shows the relative water frequency computed over all Landsat OLI and Sentinel-2 images in 2018 over the study area in Kerala, India. From the map a highly dynamic surface water environment becomes evident, with large patches of temporary flooded water bodies. Moreover, the maximum water extent indicates extended flooding throughout large parts of the study area. The 2018 Kerala flood, which had severe impacts on population, economy and housing sector in the state, is well represented in the water frequency map.

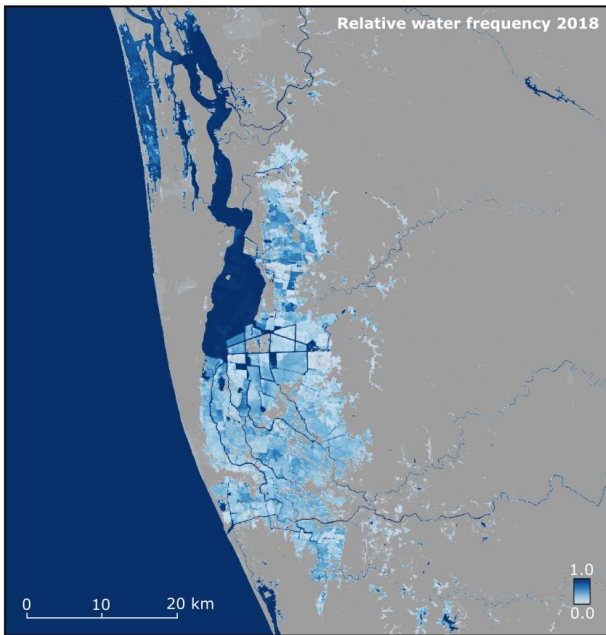


Figure 4. Relative water frequency over the study area in Kerala derived from 143 Landsat OLI and Sentinel-2 images in 2018.

In addition to the relative water frequency map, which depicts the spatial dimension of water changes in the study area, Figure 5 shows the temporal dimension of water changes in 2018 by means of the monthly mean ratio of flood to water pixels. The plot depicts well the annual water regime of the study area and also the major flooding in August 2018 is clearly visible in the timeline.

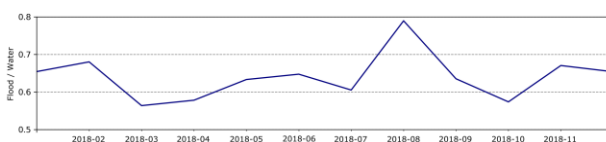


Figure 5. Monthly mean ratio of flood to water pixels over the study area in Kerala in 2018.

5. DISCUSSION AND CONCLUSIONS

In this study, we presented a solution to segment water bodies in Landsat TM, ETM+, OLI and Sentinel-2 images for rapid situational awareness in emergency situations. We introduced an adaptive reference water mask based on time-series analysis of archive imagery to distinguish flood water from permanent water and presented new results from recent flood disasters as well as a flood monitoring application in Kerala, India. Performance evaluations showed that our globally trained U-Net model produces highly accurate results and generalize well across different sensors, seasons, locations and disasters.

Our work presents an end-to-end solution that is specifically targeted towards an operational usage. No atmospheric correction and ancillary datasets are required, which reduces complexity and allows for rapid processing. We focus on multi-sensor generalization ability, simplicity and processing speed to produce timely, accurate and relevant information products for emergency responders in flood disaster situations.

Augmentation seemed to have an important effect on the network to learn invariance to changes in the target domain, which is strongly affected by atmospheric conditions, land-use / land-cover, seasonality, and other scene and image properties. Concerning the reference water mask, it should be noted that its definition is not universal and may vary depending on the application, targeted end-users, geographical region and time period. Therefore, an in-depth evaluation of use-cases and a more quantitative approach towards the definition of the term reference water and associated decisions about parameter settings should be targeted as part of future research.

Further ongoing and future works are dedicated to support additional high resolution sensors (e.g., PlanetScope), and to train a water segmentation model for very high resolution satellite (e.g., WorldView-3) and aerial imagery. Finally, the proposed processing chain will complement existing SAR-based flood monitoring services (Martinis et al., 2015; Twele et al., 2016).

ACKNOWLEDGEMENTS

This work was supported by DLR internal funding in the framework of cooperation with the Indian Space Research Organisation (ISRO) and by the German Federal Ministry of Education and Research (BMBF) as part of the RIESGOS project (Grant No. 03G0876).

REFERENCES

- Chen, Y., Fan, R., Yang, X., Wang, J., Latif, A., 2018. Extraction of Urban Water Bodies from High-Resolution Remote-Sensing Imagery Using Deep Learning. *Water*, 10, 585.
- Feng, M., Sexton, J.O., Channan, S., Townshend, J.R., 2016. A global, high-resolution (30-m) inland water body dataset for 2000: first results of a topographic-spectral classification algorithm. *International Journal of Digital Earth*, 9, 113–133.
- Franci, F., Mandanici, E., Bitelli, G., 2015. Remote sensing analysis for flood risk management in urban sprawl contexts. *Geomatics, Natural Hazards and Risk*, 6, 583–599.
- Hughes, M., Hayes, D., 2014. Automated Detection of Cloud and Cloud Shadow in Single-Date Landsat Imagery Using Neural Networks and Spatial Post-Processing. *Remote Sensing*, 6, 4907–4926.
- Isikdogan, F., Bovik, A.C., Passalacqua, P., 2017. Surface Water Mapping by Deep Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10, 4909–4918.
- Kingma, D.P., Lei, J., 2015. Adam: A Method for Stochastic Optimization. *ArXiv:1412.6980v9* 15.

Ko, B., Kim, H., Nam, J., 2015. Classification of Potential Water Bodies Using Landsat 8 OLI and a Combination of Two Boosted Random Forest Classifiers. *Sensors*, 15, 13763–13777.

Lin, L., Di, L., Yu, E.G., Kang, L., Shrestha, R., Rahman, Md.S., Tang, J., Deng, M., Sun, Z., Zhang, C., Hu, L., 2016. A review of remote sensing in flood assessment, in: *2016 Fifth International Conference on Agro-Geoinformatics*, Tianjin, China, 1–4.

Martinis, S., Twele, A., Plank, S., Zwenzner, H., Danzeglocke, J., Strunz, G., Lüttenberg, H., Dech, S., 2015. The International Charter ‘Space and Major Disasters’: DLR’s Contributions to Emergency Response Worldwide. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 85, 317–325.

Martinis, S., Kersten, J., Twele, A., 2017. A fully automated TerraSAR-X based flood service. *ISPRS Journal Photogrammetry and Remote Sensing*, 104, 203–212.

Mueller, N., Lewis, A., Roberts, D., Ring, S., Melrose, R., Sixsmith, J., Lymburner, L., McIntyre, A., Tan, P., Curnow, S., Ip, A., 2016. Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. *Remote Sensing of Environment*, 174, 341–352.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D’Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., Kassem, K.R., 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, 51, 933.

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540, 418–422.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, Cham, 234–241.

Schumann, G., Moller, D., 2015. Microwave remote sensing of flood inundation. *Physics and Chemistry of the Earth*, 83–84, 84–95.

Twele, A., Cao, W., Plank, S., Martinis, S., 2016. Sentinel-1-based flood mapping: a fully automated processing chain. *International Journal of Remote Sensing*, 37, 2990–3004.

Wieland, M., Li, Y., Martinis, S., 2019. Multi-sensor cloud and cloud shadow segmentation with a convolutional neural network. *Remote Sensing of Environment*, 230, 1–12.

Wieland, M., Martinis, S., in review. A modular processing chain for automated flood monitoring from multi-spectral satellite data. *ISPRS Journal of Photogrammetry and Remote Sensing*.

Zhou, Y., Dong, J., Xiao, X., Xiao, T., Yang, Z., Zhao, G., Zou, Z., Qin, Y., 2017. Open Surface Water Mapping Algorithms: A Comparison of Water-Related Spectral Indices and Sensors. *Water*, 9, 256.