ORIGINAL ARTICLE

Potential of Large-Scale Inland Water Body Mapping from Sentinel-1/2 Data on the Example of Bavaria's Lakes and Rivers

Michael Schmitt[1]

Abstract

The mapping of water bodies is an important application area of satellite-based remote sensing. In this contribution, a simple framework based on supervised learning and automatic training data annotation is shown, which allows to map inland water bodies from Sentinel satellite data on large scale, i.e. on state level. Using the German state of Bavaria as an example and different combinations of Sentinel-1 SAR and Sentinel-2 multi-spectral imagery as inputs, potentials and limits for the automatic detection of water surfaces for rivers, lakes, and reservoirs are investigated. Both quantitative and qualitative results confirm that fully automatic large-scale inland water body mapping is generally possible from Sentinel data; whereas, the best result is achieved when all available surface-related bands of both Sentinel-1 and Sentinel-2 are fused on a pixel level. The main limitation arises from missed smaller water bodies, which are not observed in bands with a resolution of about 20 m. Given the simplicity of the proposed approach and the open availability of the Sentinel data, the study confirms the potential for a fully automatic large-scale mapping of inland water with cloud-based remote sensing techniques.

Keywords Remote sensing · Data fusion · Water mapping · Machine learning · Copernicus data

Zusammenfassung

Potenzial der großräumigen Binnengewässerkartierung mittels Sentinel-1/2-Daten am Beispiel der bayerischen Seen und Flüsse. Die Kartierung von Gewässern ist eine wichtige Anwendung der satellitengestützten Fernerkundung. In diesem Beitrag wird ein einfacher Ansatz gezeigt, der auf überwachtem Lernen und automatischer Annotierung von Trainingsdaten basiert und es ermöglicht, Binnengewässer aus Sentinel-Satellitendaten großräumig, d.h. auf Landesebene, zu kartieren. Am Beispiel des deutschen Bundeslandes Bayern und verschiedener Kombinationen von Sentinel-1 SAR und Sentinel-2 Multispektralbildern als Input werden Potenziale und Grenzen für die automatische Erkennung von Wasseroberflächen für Flüsse, Seen und Stauseen untersucht. Sowohl die quantitativen als auch die qualitativen Ergebnisse bestätigen, dass eine vollautomatische großflächige Kartierung von Binnengewässern aus Sentinel-Daten grundsätzlich möglich ist, wobei das beste Ergebnis erzielt wird, wenn alle verfügbaren oberflächenbezogenen Bänder von Sentinel-1 und Sentinel-2 auf Pixelebene fusioniert werden. Die Haupteinschränkung des Ansatzes ergibt sich aus den nicht detektierten kleineren Gewässern, die von den Bändern mit einer Auflösung von etwa 20 m nicht beobachtet werden. Angesichts der Einfachheit des vorgeschlagenen Ansatzes und der offenen Verfügbarkeit der Sentinel-Daten bestätigt die Studie das Potenzial für eine vollautomatische großräumige Kartierung von Binnengewässern mit Hilfe von Cloud-basierter Fernerkundungsdatenauswertung.

1 Introduction

Remote sensing has long been an important means for the mapping—and potential monitoring—of the surface water resources of the Earth (Palmer et al. 2015). Generally, for this purpose, data from all possible scales, i.e. from very-high-resolution UAV-borne data (Flener et al. 2013; Rivas Casado et al. 2015; Rusnak et al. 2018) to high- and medium-resolution satellite imagery (Isikdogan et al. 2017;

✉ Michael Schmitt
m.schmitt@tum.de

1 Signal Processing in Earth Observation, Technical University of Munich, Munich, Germany

Yang et al. 2017; Rishikeshan and Ramesh 2018), have been used. However, the greatest potential of spaceborne remote sensing lies in the possibility to observe surface water on large scales—from country-wide (Mueller et al. 2016) to even global coverage (Pekel et al. 2016).

The work presented in this contribution seeks to complement the existing large-scale approaches of Mueller et al. (2016) and Pekel et al. (2016). While they were based on information extraction from long-term optical archives and classification systems utilizing expert knowledge, in this contribution the focus lies on the following points:

– While much of the available literature focuses on the utilization of only optical imagery or only SAR data for water mapping, in this work, the joint use of the multi-spectral optical data provided by Sentinel-2, as well as the synthetic aperture radar (SAR) observations provided by Sentinel-1 is emphasised. An important reason for this choice is the fact that SAR imagery does not suffer from obstruction by clouds or other atmospheric effects. It is, thus, theoretically allowing for a more continuous monitoring of important water bodies or wetlands (see, e.g. Zeng et al. 2017). Furthermore, the great potential of SAR-optical data fusion for water mapping was only recently confirmed by Bioresita et al. (2019) and Slinski et al. (2019). While the former focused on decision-level fusion of water probability maps calculated from mono-sensor Sentinel-1 and Sentinel-2 time series, the latter used pixel-level fusion of SAR backscatter and multi-temporal modified normalised difference water index (MNDWI) composites derived from Landsat observations. Study results for test scenes located in Ireland and Ethiopia, respectively, confirmed the robustness of those fusion-based approaches.

– By utilising a simple classifier based on supervised learning and automated training data generation from globally available volunteered geographic information (VGI), freely available remote sensing data provided by the Sentinel missions of the European Copernicus program, and the cloud computation platform Google Earth Engine, the proposed approach can easily be applied to arbitrary regions of the world. It is assumed that this will enable also laypersons to map inland water on a large scale without the need for expert knowledge or manual intervention.

For the experimental investigations, the state of Bavaria is used as the study scene. The region is a typical representative of the temperate biome, located in the heart of Europe. Since Bavaria contains many kinds of possible inland water bodies, from small mountain creeks to large drainage rivers, from districts of small ponds to large freshwater lakes, the results of this work are expected to provide sufficient relevance to generalize beyond the study scene.

Thus, while using a central European region as an example, it has to be stressed that the main purpose of this work is to explore the general potential of a fully automatic, cloud-based framework that allows fast and cheap large-scale mapping of surface water for complete countries, continents, or even the whole globe.

In this context, it has to be stressed that the term "surface water" in this work refers to the very water extent that is visible in seasonal Sentinel-1 and Sentinel-2 composites (cf. Sect. 3.2.1).

The remainder of this article is organised as follows: Sect. 2 describes the Sentinel-1 and Sentinel-2 missions with a focus on system-inherent water mapping capabilities. In Sect. 3.2 a simple, yet fully automatic, water detection framework based on supervised learning is proposed, whose results for the study area of Bavaria are summarised in Sect. 4.

Finally, Sect. 5 discusses the quality of the presented water detection approach and possible sources of errors, while Sect. 6 presents summary and conclusion of the findings.

Table 1 Channels and their meaning of Sentinel-1 and Sentinel-2

System	Channels	Description	Resolution
Sentinel-1	VV	Vertical transmit/vertical receive, ideal for observation of wave-induced water surface roughness	20 m × 22 m
	VH	Vertical transmit/horizontal receive, much weaker backscatter return from water, higher return from volumetric targets	20 m × 22 m
Sentinel-2	B1, B9, B10	Related to atmospheric parameters, not relevant for Earth surface	60 m
	B2, B3, B4	Visible bands (blue, green, red), high reflectivity of water in blue and green bands	10 m
	B5, B6, B7	Vegetation red edge, relevant for vegetation sensing, water is highly absorbed	20 m
	B8	Broadband near infrared, main infrared band, water is highly absorbed	10 m
	B8a	Narrowband near infrared, designed to avoid contamination from water vapour, surface water is highly absorbed	20 m
	B11, B12	Short-wave infrared, important for determining materials, detecting water and fire	20 m

2 Sentinel-1 and Sentinel-2

The Copernicus program is the Earth Observation program of the European Union (Desnos et al. 2014). While it also provides data from in situ sources, its core is comprised of the Sentinel s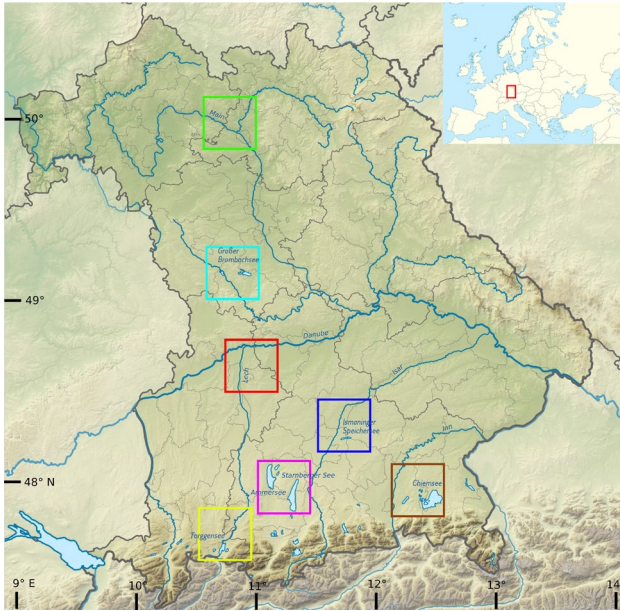atellite family. These Sentinel missions were specifically designed to meet versatile end user requirements. Among the Sentinel missions, Sentinel-1 and Sentinel-2 are classical Earth observation platforms and provide SAR and multi-spectral image data, respectively, in the metre-resolution domain. Thus, they are certainly the most relevant for inland water body monitoring; whereas, Sentinel-3 is specifically aimed at ocean and land monitoring with its low spatial resolution multi-spectral images and altimeter capabilities.

In the frame of this study, which focuses on inland freshwater (i.e. river and lake) mapping, thus only Sentinel-1 and Sentinel-2 are considered, whose channels are summarised in Table 1, and whose attributes are described in the following.

2.1 Sentinel-1

The Sentinel-1 mission (Torres et al. 2012) currently consists of two polar-orbiting satellites, which are equipped

Fig. 1 Map of the largest rivers and bodies of stagnant water in Bavaria (This work, "Largest water bodies of Bavaria", is a derivative of "Bavaria relief location map" by Alexrk2, used under CC-SA. "Largest water bodies of Bavaria" is licensed under CC-SA by Michael Schmitt.). The coloured boxes indicate the different test areas shown in Figs. 6, 7, 8, 9, 10, 11, 12
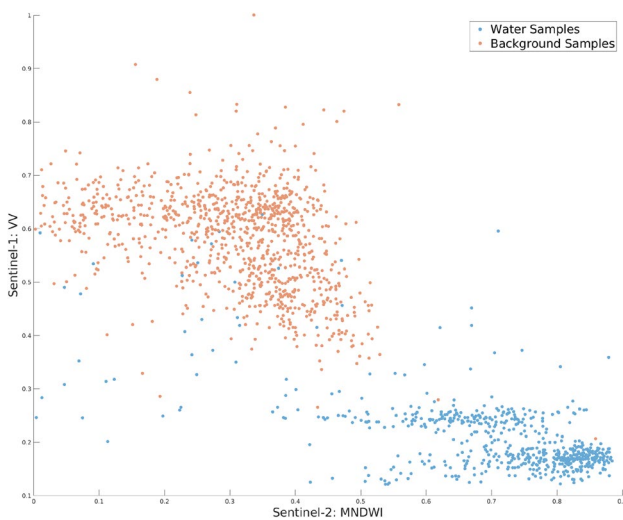

Fig. 2 Two-dimensional feature space created from Sentinel-1 VV backscatter (normalised to [0; 1]) and Sentinel-2 MNDWI magnitudes for the discrimination of water and background
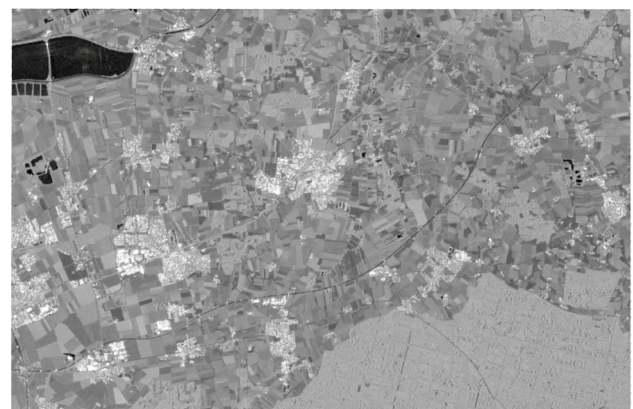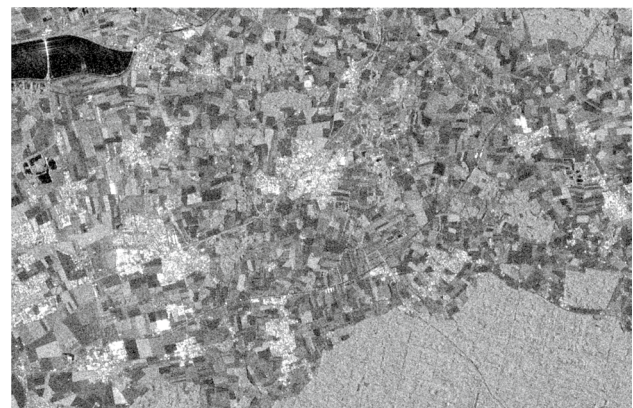

Fig. 3 The denoising effect of temporal averaging: **a** single Sentinel-1 scene, **b** temporal mean map achieved by averaging 59 multi-temporal acquisitions acquired within a time period of 3 months during summer 2018
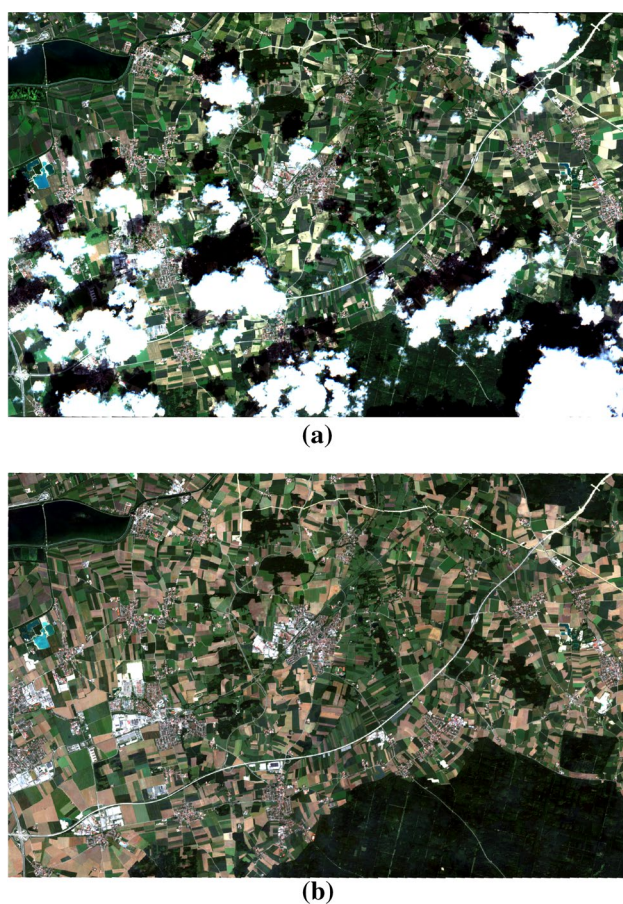
Fig. 4 Result of a simple multi-temporal cloud removal approach: **a** single Sentinel-2 scene constructed by combining data from two neighbouring granules, **b** cloud-masked median image constructed from 20 multi-temporal acquisitions per granule, acquired within a time period of 3 months during summer 2018

Fig. 5 NDVI (**a**) and MNDWI (**b**) magnitude image examples. While there is a high inverse correlation of NDVI magnitude and water, also impervious surfaces and some (barren) agricultural areas exhibit low NDVI magnitudes. In contrast, high MNDWI magnitudes clearly correlate with water

with C-band SAR sensors enabling them to acquire imagery regardless of metereological conditions.

Sentinel-1 works in a pre-programmed operation mode to avoid conflicts and to produce a consistent long-term data archive built for applications based on long time series. Depending on which SAR imaging mode is used, either high resolutions down to 5 m (at a swath width of 80 km) or wide coverages of up to 400 km (at a resolution of 20 m × 40 m) can be achieved.

Furthermore, Sentinel-1 provides dual polarisation capabilities and very short revisit times of about one week at the equator. Since highly precise spacecraft positions and attitudes are combined with the high accuracy of the range-based SAR imaging principle, Sentinel-1 images come with high out-of-the-box geolocation accuracy (Schubert et al. 2015). In the frame of the presented study, the most widely available standard product was used, i.e. ground-range-detected (GRD) imagery acquired in the so-called
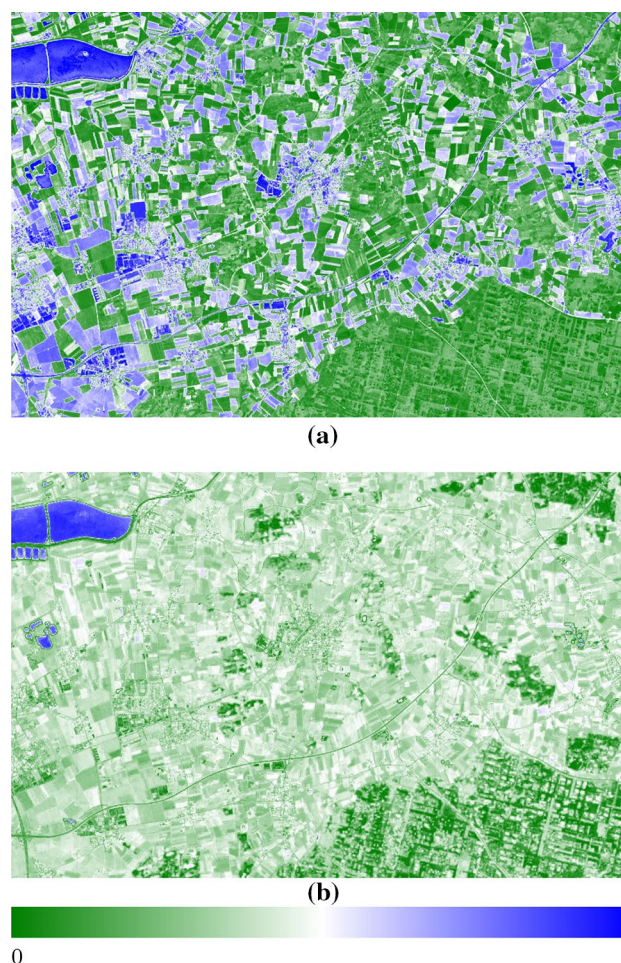
interferometric wide swath mode (IW). The GRD images display σ^0 backscatter coefficient values for both vertical (VV) and cross (VH) polarization in dB scale and are projected to ground range using the Earth ellipsoid model WGS84. For precise ortho-rectification, restituted orbit information is combined with the 30 m-SRTM-DEM or the ASTER DEM for high latitude regions where SRTM is not available. While the resolution of IW-GRD images is 20 m × 22 m in range and azimuth, respectively, the products are provided with a square pixel spacing of 10 m × 10 m to the end user. It has to be noted that SAR imagery is inherently affected by the so-called speckle effect, which appears as a form of noise to human analysts and most image processing algorithms, so that speckle filtering by dedicated algorithms or multi-temporal image fusion must be used during pre-processing to enhance the visual image quality.

Table 2 Input feature combinations used in the experiments

System	Number	Channels	Description
Sentinel-1	1	VV	Single polarization
	2	VH	Single polarization
	3	VV, VH	Dual polarization
Sentinel-2	4	B2, B3, B4, B8	10 m bands
	5	B2, B3, B4, B5, B6, B7, B8, B8a, B11, B12	All surface-related bands
	6	NDVI	Vegetation index
	7	MNDWI	Water index
	8	MNDWI, NDVI	Combination of water and vegetation indices
Fusion	9	VV, VH, B2, B3, B4, B5, B6, B7, B8, B8a, B11, B12	All surface-related bands of both systems
	10	B2, B3, B4, B8, VH	Combination of best single-sensor setups
	11	VH, MNDWI	Strongest features by expectation

2.2 Sentinel-2

The Sentinel-2 mission (Drusch et al. 2012) at the moment comprises two identical polar-orbiting satellites in the same orbit, phased at 180° to each other. The mission is meant to provide continuity for multi-spectral imagery of the SPOT and LANDSAT kind, which have provided information about the land surfaces of our Earth for many years. With its wide swath width of up to 290 km and its high revisit time of 5 days over the inhabited continents (based on two satellites) under cloud-free conditions, the Sentinel-2 mission is specifically well suited to the monitoring of vegetated land surfaces. Sentinel-2 data are provided in the form of so-called granules, which cover a ground area of 100 km × 100 km at pixel spacings of 10 m, 20 m and 60 m, depending on the respective spectral bands. At the most widely available

Level-1C standard, the end-user products provide orthorectified Top-Of-Atmosphere (TOA) reflectances at sub-pixel multispectral registration. Since cloud and land/water masks included in the products are rather coarse, these are often not very reliable, so that pre-processing is still necessary to deal with potential issues of cloud coverage.

2.3 Suitability of Sentinel Data for Water Mapping

The SAR data provided by the Sentinel-1 mission are expected to be a good source of information for water mapping based on the fact that to C-band radar signals water is a smooth surface. This leads to a mirror-like reflection and consequently very low backscatter values (i.e. "dark" pixels). On the other hand, smooth artificial surfaces (e.g. roads or other paved areas) can also lead to specular reflections

Table 3 Test metrics for the different feature combinations resulting from tenfold cross-validation

System	Number	PA (water)	PA (non-water)	UA (water)	UA (non-water)	OA	κ
Sentinel-1	1	*95.08 ± 1.30*	**99.65 ± 0.32**	**99.63 ± 0.33**	95.31 ± 1.15	97.37 ± 0.68	0.95 ± 0.01
	2	**97.40 ± 0.75**	*98.73 ± 0.65*	*98.72 ± 0.64*	**97.42 ± 0.80**	98.07 ± 0.58	0.96 ± 0.01
	3	96.94 ± 0.90	99.39 ± 0.38	99.38 ± 0.39	97.05 ± 0.77	**98.18 ± 0.40**	**0.96 ± 0.01**
Sentinel-2	4	**97.26 ± 1.06**	99.27 ± 0.56	99.26 ± 0.51	**97.26 ± 1.18**	**98.25 ± 0.64**	**0.96 ± 0.01**
	5	97.02 ± 0.93	99.24 ± 0.55	99.22 ± 0.57	97.11 ± 0.82	98.13 ± 0.55	0.96 ± 0.01
	6	96.65 ± 0.85	*83.39 ± 1.15*	*85.33 ± 1.40*	96.16 ± 0.92	*90.03 ± 0.97*	*0.80 ± 0.02*
	7	*89.90 ± 1.93*	**99.64 ± 0.48**	**99.61 ± 0.55**	*90.86 ± 1.28*	94.79 ± 0.85	0.90 ± 0.02
	8	95.05 ± 1.70	99.33 ± 0.43	99.30 ± 0.45	95.23 ± 1.69	97.17 ± 0.81	0.94 ± 0.02
Fusion	9	97.29 ± 0.91	**99.64 ± 0.33**	**99.63 ± 0.34**	97.35 ± 0.91	***98.46 ± 0.43***	***0.97 ± 0.01***
	10	*96.99 ± 1.05*	99.64 ± 0.29	99.63 ± 0.28	*97.00 ± 1.23*	98.30 ± 0.60	0.97 ± 0.01
	11	***97.75 ± 0.55***	*98.72 ± 0.52*	*98.71 ± 0.53*	***97.73 ± 0.73***	98.23 ± 0.40	0.96 ± 0.01

Bold numbers indicate the best and italic numbers indicate the worst performance per group (Sentinel-1 only, Sentinel-2 only, Fusion), respectively; while, bold-italic cells indicate the overall best performance for the corresponding metric

PA Producer's accuracy, *UA* User's accuracy, *OA* overall accuracy
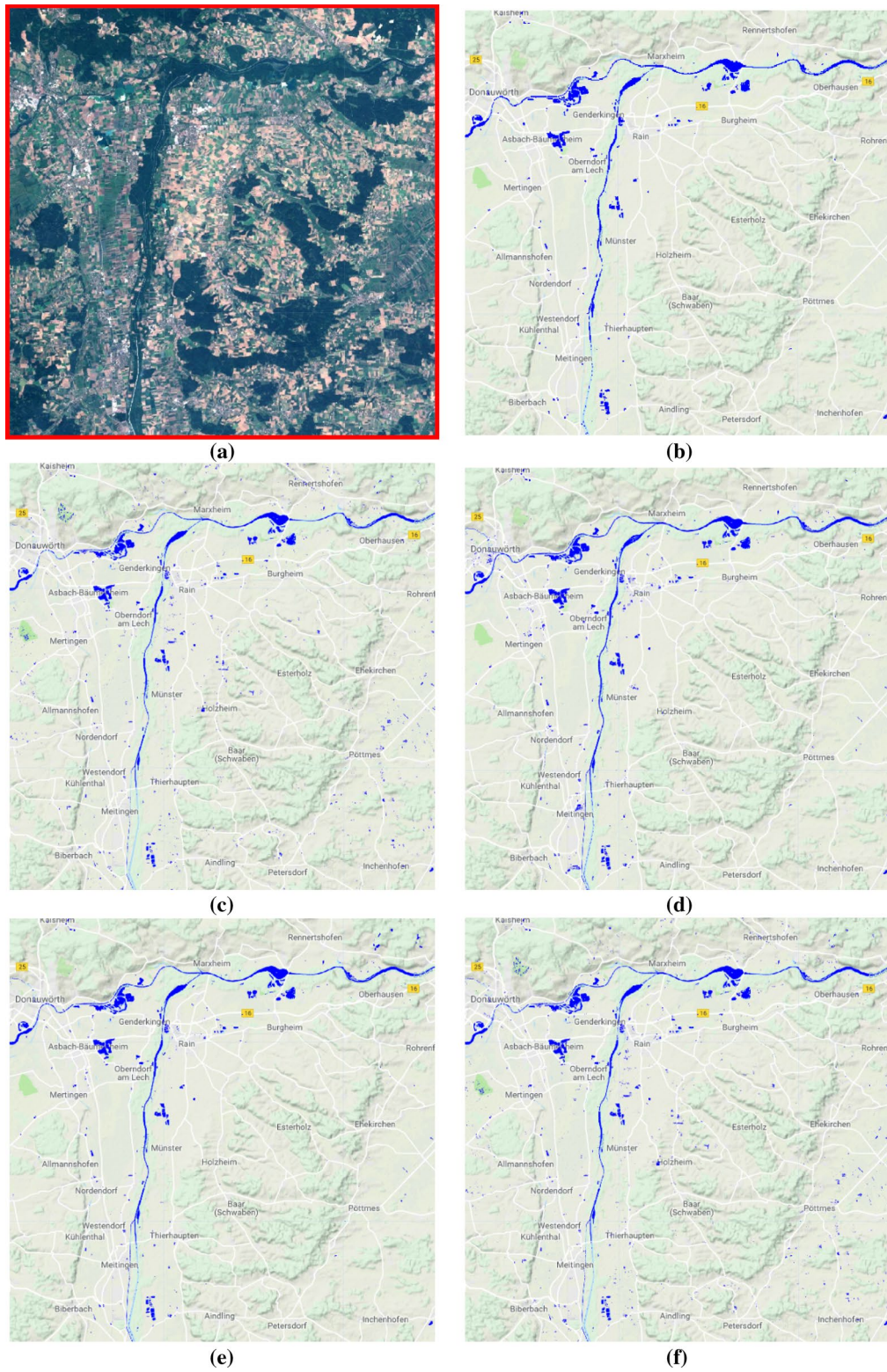
Fig. 6 Example area showing the Danube river (flowing in west–east direction), with the Lech river entering from the south. **a** Optical reference image, **b** GSWL reference layer, **c** Sentinel-1 VH polarisation result, **d** Sentinel-2 10-m bands result, **e** Sentinel-1/Sentinel-2 full-band fusion, **f** Sentinel-1/Sentinel-2 VH/MNDWI fusion

causing confusion with water bodies and leading to a high false-positive rate, if only SAR data were used as input to the water detection procedure.

Besides these radiometric considerations, in the interpretation of SAR imagery geometric effects caused by the side-looking imaging principle need to be taken into account: if the top of an object (e.g. a mountain) is closer to the sensor than its bottom, the object appears to "collapse" towards the sensor. This is called the layover effect. In correspondence, any target lying behind this elevated object will be within the so-called radar shadow. Since rivers are often carving themselves into canyons, these two effects can impact the detection of rivers and lead to both false positives (in the case of radar shadow) and false negatives (in the case of a river hidden in layover).

The multi-spectral data provided by Sentinel-2 support the identification based on the interaction of electromagnetic radiation from the optical part of the spectrum with water: While for pure water only less than 2–3% of the incoming radiation of the visible domain are reflected, almost complete absorption in the near-infrared (NIR) and short-wave infrared (SWIR) bands is observed (Hobson and Williams 1971). On the other hand, in particular, shallow water areas are affected by differing penetration depths from the blue to the NIR wavelength regions (Bhargava and Mariam 1991). Therefore, water level fluctuations will have an impact on the robustness of water detection from multi-spectral data. Besides, also high turbidity values or floating vegetation can lead to mis-classifications with both datasets.

Taking these considerations into account, it is advisable to make use of data fusion strategies to mitigate the weaknesses of the two sensor principles while benefitting from the complementarity of the information contained in the two data sources (Schmitt and Zhu 2016).

3 Materials and Methods

3.1 The State of Bavaria as Study Area

For the presented investigations on the potential of Sentinel remote sensing data for large-scale inland freshwater mapping, the state of Bavaria is used as study area. Bavaria is a landlocked federal state of Germany, occupying an area of about 70,550 km^2. From the surface water perspective, Bavaria is characterised by its location at the Alpine foothills with numerous lakes and small- to medium-sized rivers. In addition, two major rivers flow through the state: the Danube, which is Europe's second longest river, drains about 48.200 km^2 with a 380-km-long Bavarian section; while, the Main drains about 23.350 km^2 on a 408-km-long stretch.

The largest natural lakes belonging exclusively to Bavaria are Chiemsee with a surface of about 80 km^2, Starnberger See with about 58 km^2, and Ammersee with about 47 km^2; whereas, the largest reservoirs are Forggensee with about 15 km^2, Großer Brombachsee with about 9 km^2, and Ismaninger Speichersee with about 6 km^2. The mentioned water bodies are indicated in Fig. 1.

3.2 Water Detection by Support Vector Machines

In spite of the massive success modern deep learning approaches showed also in the field of remote sensing in recent years (Zhu et al. 2017), it is argued that complicated, data-hungry convolutional neural networks are not necessary for water detection, as water can very clearly be identified even in simple hand-crafted feature spaces. An example for this argument is shown in Fig. 2. As numerous researchers have confirmed before, even with only SAR backscatter or the modified normalised difference water index (MNDWI) (Xu 2006), water and non-water pixels can already be distinguished to some extent, e.g. by thresholding techniques (Liebe et al. 2005; Conrad et al. 2016). When only slightly more complexity is introduced, i.e. when the feature space is extended to two or three dimensions, water becomes clearly separable from the non-water background class.

It has to be mentioned, however, that simple thresholding or linear discrimination procedures are not sufficient (see, e.g. Klein et al. 2014). As can be seen from Fig. 2, quite some overlapping background and water pixels exist. This is mainly due to the confusions described in Sect. 2.3. In the SAR case, background samples (e.g. paved areas) might appear dark and be confused with water, or water samples might appear bright and, thus, be confused with background due to roughness caused by waves or top-water vegetation. In the MNDWI case, bare earth or built-up areas with very high surface temperatures can lead to strong reflections in the short-wave infrared (SWIR) band and, thus, lead to high MNDWI magnitudes; whereas, smaller water bodies might be mixed with soil reflectance in the 20 m× 20 m pixels of the Sentinel-2 SWIR band.

In this work, therefore, a simple support vector machine (SVM) with radial basis function (RBF) kernel is employed. Focus is put mainly on the following questions:

– Which input information that can be derived from Sentinel-1 and/or Sentinel-2 data constitutes the best source for freshwater mapping?
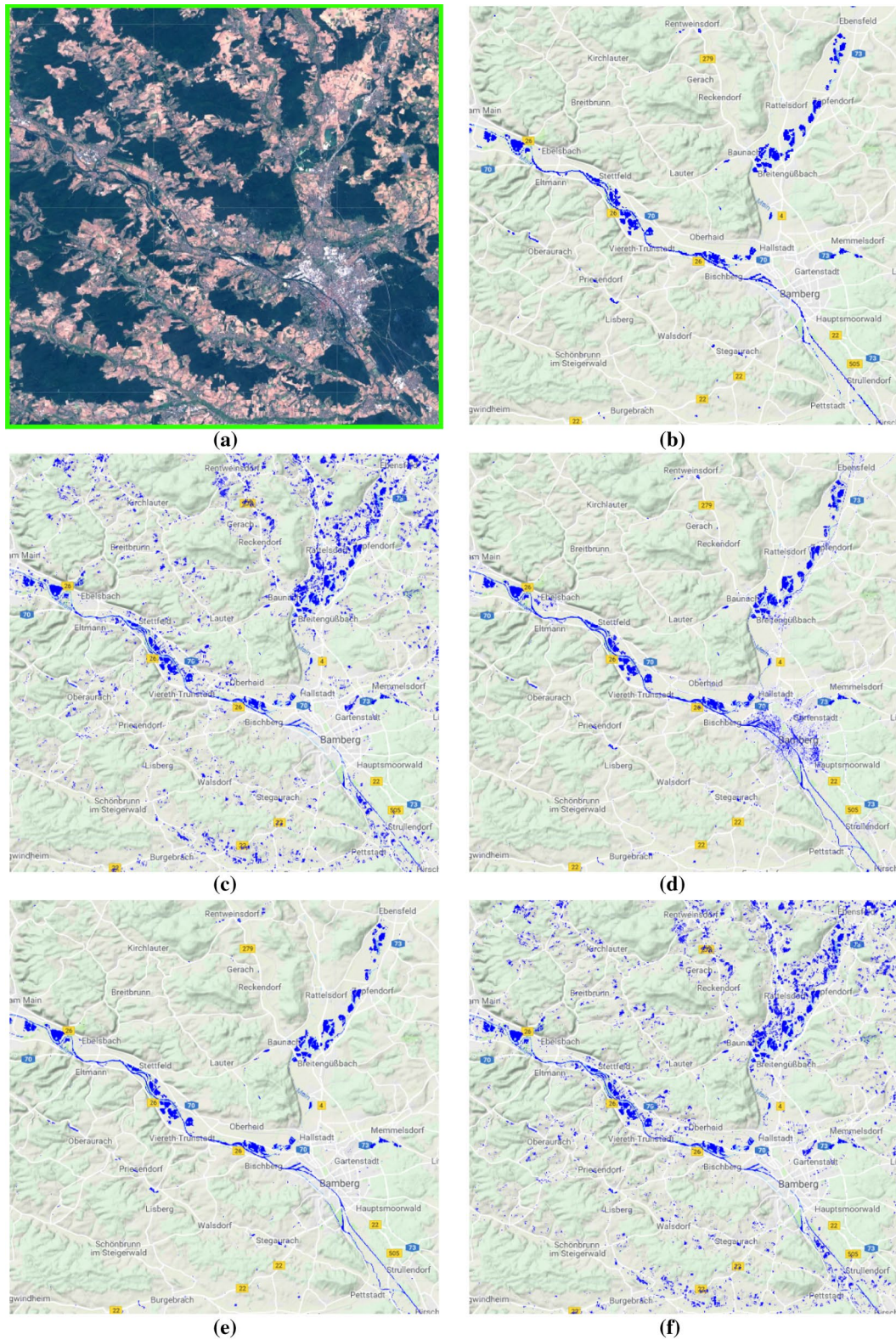– Does the fusion of Sentinel-1 SAR and Sentinel-2 optical data support the water mapping task?

Fig. 7 Example area showing the Main river (flowing in east–west direction), with the Regnitz river entering from the south close to the city of Bamberg. **a** Optical reference image, **b** GSWL reference layer, **c** Sentinel-1 VH polarisation result, **d** Sentinel-2 10-m bands result, **e** Sentinel-1/Sentinel-2 full-band fusion, **f** Sentinel-1/Sentinel-2 VH/MNDWI fusion

– What are the data-inherent limitations of Sentinel-based freshwater mapping, e.g. with respect to granularity or spectral information content?

The application of SVMs for land cover classification has a long tradition in the remote sensing community (Mountrakis et al. 2011). In this contribution, a fully automatic, cloud-based workflow for SVM-based freshwater mapping of extensive areas is described.

The workflow is implemented in Google Earth Engine (Gorelick et al. 2017) to utilize its curated data catalogue and cloud-processing capabilities. The procedure is detailed in the following sections.

3.2.1 Data Pre-Processing

Both Sentinel-1 SAR imagery and Sentinel-2 multi-spectral data come with their own system-inherent limitations, which make a reliable deduction of mapping information from them more difficult. In the Sentinel-1 case, one is mainly dealing with the already mentioned speckle effect, which appears as a form of noise to most image analysis approaches. While manifold methods for speckle filtering exist, most of them degrade the quality of the underlying image contents to some extent. Therefore, another option is used, which is often applied if precisely aligned multi-temporal SAR images are available—the creation of a temporal mean map. For this purpose, all Sentinel-1 IW datasets acquired in a specified time frame (i.e. the summer months June, July and August in 2018) are loaded from the data catalogue. For every pixel, then the mean value of all corresponding acquisitions is calculated. Afterwards, the σ_0 backscatter values are clipped to the interval $[-25; 0]$ as a form of data calibration. The effect of this pre-processing is illustrated in Fig. 3. The example shows how especially finer structures, such as roads, or water bodies, e.g. located in the top left corner of the scene, appear much more clearly in the temporal mean map than in the single-date SAR image.

When it comes to Sentinel-2, the most severe problem is cloud cover. To mitigate information loss caused by cloud-affected pixels, different pre-processing approaches can be utilised (see, e.g. Schmitt et al. 2019). In the frame of the presented work, a simple standard method is used, which combines multi-temporal median filtering and the cloud mask provided with the Sentinel-2 products in form of a quality assurance band. In analogy to the Sentinel-1 case, all Sentinel-2 scenes with less than 30% of clouds (on granule level) acquired in the summer months of 2018 (i.e. June, July, and August) are collected from the data catalogue.

Then, every pixel is masked, which is affected either by dense clouds or cirrus clouds (according to the QA60 quality band). Finally, the median for every pixel from all non-masked acquisitions is calculated, since the median is supposed to exclude very dark values (often caused by shadows) and very bright pixels (often caused by remaining clouds). The effect of this simple cloud masking is shown in Fig. 4. It becomes obvious that in single-date imagery, many scene parts are obstructed by cloud cover or cloud shadow; while, combining multi-temporal data allows to create a synthetic cloud-free input to subsequent water detection tasks.

While this proceeding is suitable for the creation of water maps at loosely sampled, discrete time points, it prevents an actual quasi-continuous monitoring of inland water at the temporal resolution that would be provided by the Sentinel-2 acquisition opportunities. In the future, the presented approach has to be extended by a more sophisticated cloud-mitigation procedure, which still makes use of SAR and optical data fusion instead of only relying on SAR data, but samples the optical imagery more densely in time. A quite inspiring technique tailored to the mapping of croplands from Sentinel-2 time series based on a deep neural network, which also learned to ignore clouds, was recently presented by Rußwurm and Körner (2018).

3.2.2 Automatic Training Data Generation

Since the aim of this work is detecting water surfaces with a model based on supervised learning, a suitable training dataset is needed, which contains a balanced sample of target examples, i.e. water pixels, and clutter examples, i.e. background pixels. To train a robust and well-generalizing model, it is important that the versatility of water observations (ranging from deep lakes to shallow rivers) is well represented in the training data. This holds even more so for the background-related observations, as here one has to deal with many different forms of land cover. To avoid labour-intensive, and often biased, manual annotation of training samples, a fully automatic approach is proposed, which exploits voluntary geographic information (VGI) in the form of the OpenStreetMap (OSM) water layer.

After loading the layer, it is filtered to remove any polygons related to the *wetlands* class, so that only actual surface water bodies (i.e. rivers, lakes etc.) are kept. Then, a negative buffering of 20 m to each polygon is applied to make sure that each polygon really only contains water surfaces and not ambiguous areas such as shorelines. Finally, the list of polygons is filtered to remove all water polygons with an area smaller than 10,000 m^2 for the sake of memory requirements
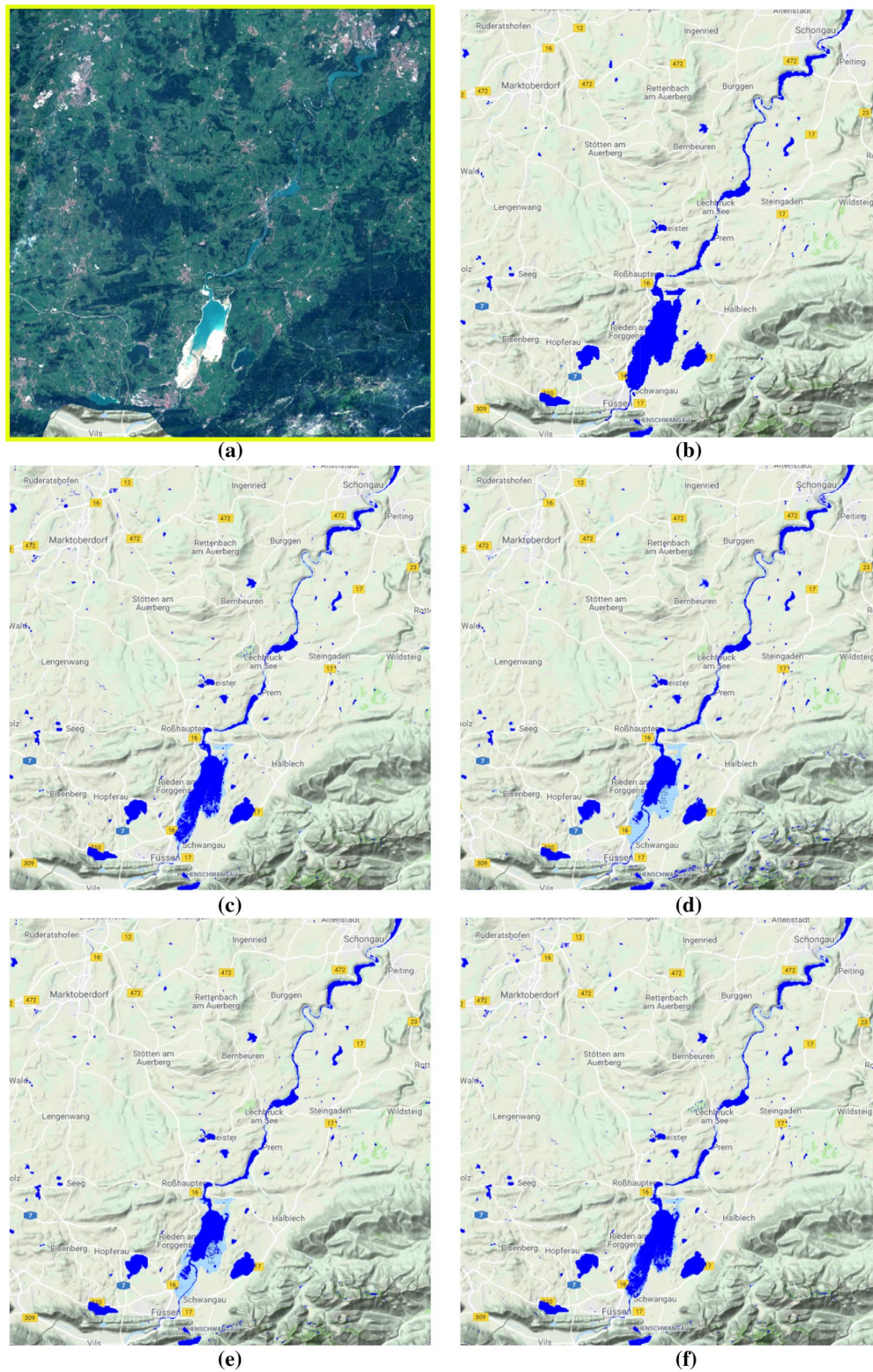
Fig. 8 Example area showing the reservoir Forggensee, some smaller natural lakes, and the Lech river exiting Forggensee to the north. **a** Optical reference image, **b** GSWL reference layer, **c** Sentinel-1 VH polarisation result, **d** Sentinel-2 10-m bands result, **e** Sentinel-1/Sentinel-2 full-band fusion, **f** Sentinel-1/Sentinel-2 VH/MNDWI fusion

and processing speed. In parallel, an inverse, i.e. positive, buffering of 50 m is applied to the previously reduced water polygons and those enlarged polygons are subtracted from the outer polygon describing the Bavarian state borders. Thus, one finally has one list of water polygons, and one list of non-water polygons. This allows to randomly sample 2500 training pixel locations each for both the water and the background class, as well as 500 pixels each for testing. From those locations, later the desired remote sensing-derived information are extracted to train and test the proposed SVM-based classification models.

3.2.3 Features for Water Mapping

From Sentinel-1, both the VV and VH polarisation channels are used. From Sentinel-2, all bands with 10-m and 20-m resolution are used, which carry highly relevant information from the visible, the near infrared, and the short-wave infrared parts of the electromagnetic spectrum. The 60-m bands are ignored, as those are primarily intended to provide measurements about atmosphere parameters such as aerosoles and clouds. In addition, two spectral indices are employed. The normalised difference vegetation index (NDVI) (Rouse 1974) was developed for mapping vegetation status and is probably the most frequently used spectral index in environmental remote sensing (Pettorelli 2013). It combines information from the red and near-infrared parts of the spectrum and is calculated by

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} = \frac{\text{Band 8} - \text{Band 4}}{\text{Band 8} + \text{Band 4}}. \tag{1}$$

The NDVI is ued based on the assumption that its response in case of water is inverse to that of vegetation.

In addition to that, the more relevant modified normalised difference water index (MNDWI) (Xu 2006) is employed, which is an extension of the normalised difference water index (NDWI) (Gao 1996) and is dedicated to the detection of water surfaces. It combines information from the green and short-wave infrared parts of the spectrum and is calculated by

$$\text{MNDWI} = \frac{\text{Green} - \text{SWIR}}{\text{Green} + \text{SWIR}} = \frac{\text{Band 3} - \text{Band 11}}{\text{Band 3} + \text{Band 11}}, \tag{2}$$

whereas the original NDWI is calculated by

$$\text{NDWI} = \frac{\text{NIR} - \text{SWIR}}{\text{NIR} + \text{SWIR}} = \frac{\text{Band 8} - \text{Band 11}}{\text{Band 8} + \text{Band 11}}. \tag{3}$$

Since the MNDWI was developed to improve the sensitivity of the NDWI against a confusion of water with impervious

surfaces, it is expected to be the most powerful feature for water detection from optical imagery. To illustrate the use of both NDVI and MNDWI, the corresponding feature images are displayed in Fig. 5. The illustration also shows the higher resolution of the NDVI, which is based on 10-m bands; whereas, the MNDWI uses one of the 20 m SWIR bands.

Table 2 summarizes the different feature combinations used in the experiments for this work. To analyse the full potential of the available Sentinel-1 and Sentinel-2 data, the potential of Sentinel-1 only, Sentinel-2 only, and a fusion of both data sources is investigated.

4 Results for Rivers and Lakes in Bavaria

Different SVM models for the 10 feature combinations summarised in Table 2 were trained based on the 5,000 training samples as described in Sect. 3.2.2. Using tenfold cross-validation, the 10 models are evaluated on 1,000 test samples each. The results are summarised in Table 3. It has to be noted that in spite of cross-validation, both training sets and test sets are randomly sampled from the same basic distribution. Thus, the results should be considered as optimistic estimates for the study area of Bavaria and cannot be used to extrapolate accuracies to be expected in unseen areas. Furthermore, due to the automatic training data generation procedure, small water bodies are not represented in the ground truth; this also leads to positively biased values.

Looking at these quantitative results, one can observe the following phenomena:

1. *Dual-pol SAR does not provide higher detection accuracy*

   Accuracy-wise, the best SAR configuration for detecting water is a single-channel analysis of only VH-polarised intensity images (setup 2). This setup provides the best producer's accuracy for water, while overall accuracy and kappa coefficient are not significantly different from the numerically better dual-pol result. VV polarization (setup 1) only performs better with respect to producer's accuracy of non-water and user's accuracy of water. This indicates that it misses some water areas, but does usually not lead to a misclassification of background. Dual polarisation (setup 3) does not provide significantly better metrics in any category, but is one of the results with the smallest standard deviations, i.e. the highest prediction stability.
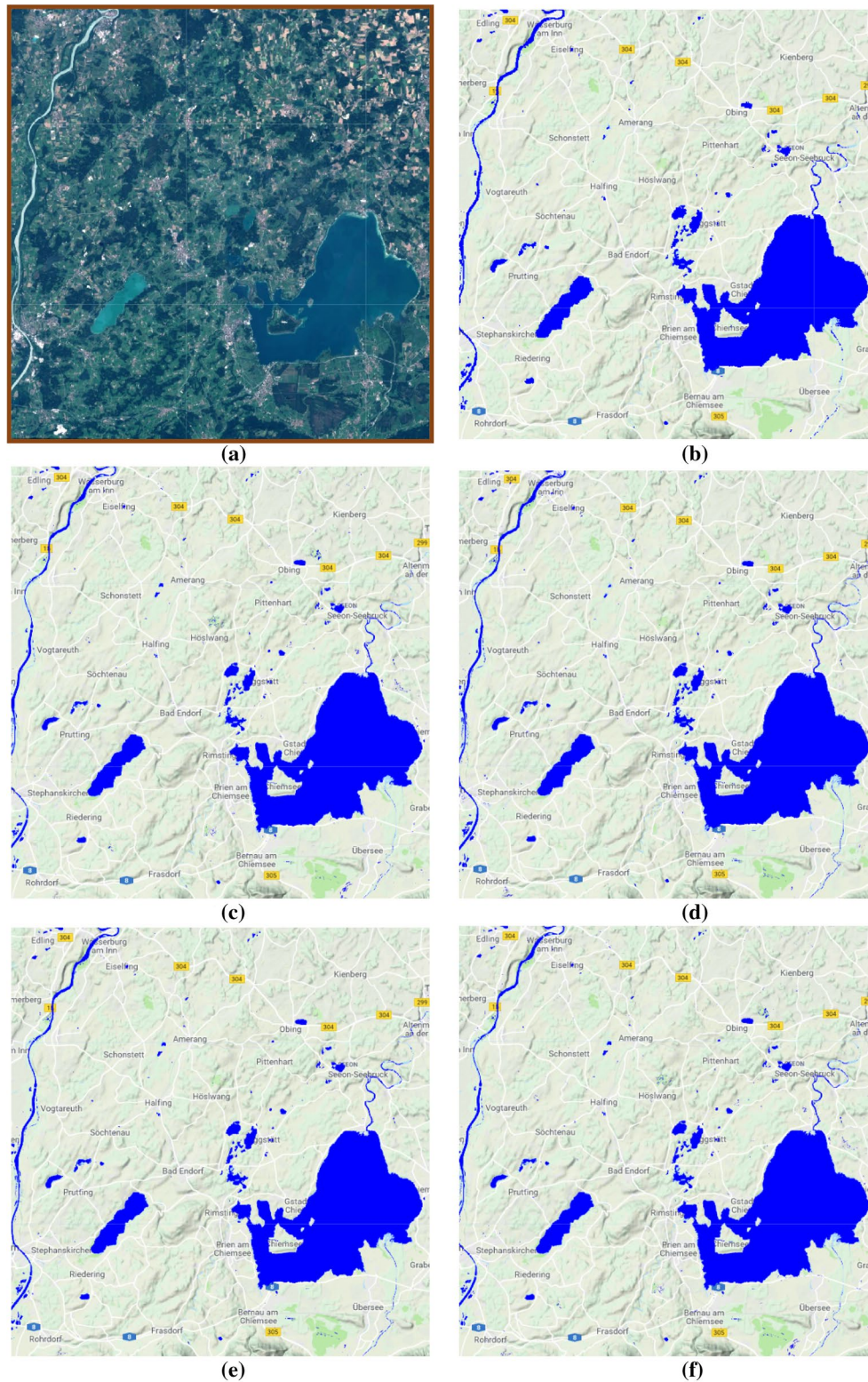
Fig. 9 Example area showing river Inn (flowing in south–north direction), the Simsee, and the Chiemsee, with the small Alz river exiting Chiemsee to the north. **a** Optical reference image, **b** GSWL reference layer, **c** Sentinel-1 VH polarisation result, **d** Sentinel-2 10-m bands result, **e** Sentinel-1/Sentinel-2 full-band fusion, **f** Sentinel-1/Sentinel-2 VH/MNDWI fusion

2. *Optical indices are not as helpful as expected*

While the MNDWI magnitude is at least the most powerful feature to reliably decide what pixels belong to the background (shown by high producer's accuracy for non-water and user's accuracy for water in setup 7), it does not perform very well in finding all water areas. One might think that this is caused by the fact that the SWIR bands necessary for MNDWI calculation have a resolution of only 20m, so that small water areas will simply be missed. However, due to the removal of small water bodies in the label pre-processing (cf. Sect. 3.2.2), the validation data also do not contain any small water bodies with an area less than $10,000 \, \mathrm{m}^2$.

In contrast to the MNDWI magnitude, the NDVI magnitude provides the overall least amount of useful information for water detection (see setup 6). At the same time, a combination of MNDWI and NDVI (setup 8) leads to a better but still mediocre water detection performance. Using only the 10-m bands of Sentinel-2 without any previous feature extraction performs best (see setup 4); whereas, using all surface-related bands (setup 5) provides no additional benefit besides slightly better prediction stability.

3. *Fusion performs better than single-sensor analysis*

With the highest overall accuracies and at least competitive κ coefficients, the fusion-based predictions on average perform better than the single-sensor results. While the fusion of just VH polarisation and MNDWI (setup 11) provides the overall highest producer's accuracy for water, the fusion of all available surface-related bands (setup 9) is better with respect to the user's accuracy of the water class as well as overall accuracy and κ coefficient—at least if rounding errors are ignored. Interestingly, leaving out the Sentinel-2 20m bands and the Sentinel-1 VV channel (setup 10) unexpectedly does not provide any advantage, although they could potentially downgrade the result due to their lower resolution.

4. *Detecting water is harder than detecting background*

The mean producer's accuracy for water detection is about 96%, whereas the mean producer's accuracy for background is almost 98%. This indicates that detecting the single, well-defined water class is harder than detecting background pixels, which inherently comprise a much more versatile sample distribution.

Based on these numerical results, selected visual results are shown for several example regions across Bavaria for

- the best Sentinel-1 configuration, i.e. the VH band (setup 2)
- the best Sentinel-2 configuration, i.e. the 10-m bands (setup 4)
- the two best fusion-based configurations, i.e. all bands (setup 9) and VH+MNDWI (setup 11)

in terms of mean accuracy achieved in the tenfold cross-validation experiments.

The selected maps contain sections of the Danube (Fig. 6), the Main (Fig. 7), the Lech (Figs. 6 and 8), the Inn (Fig. 9), and the Isar (Fig. 10) to illustrate the performance on large- and medium-sized rivers. With respect to a mapping of bodies of stagnant water, Starnberger See, Ammersee, Wörthsee, Pilsensee (all Fig. 11), Chiemsee, Simsee (both Fig. 9), Forggensee, Weißensee, Hopfensee, Bannwaldsee (all Fig. 8), Großer Brombachsee, Altmühlsee (both Fig. 12), and Ismaninger Speichersee (Fig. 10) are shown to illustrate the performance on large- and medium-sized lakes and reservoirs.

For comparison to an independent yet similar reference, all figures contain a surface water map extracted from the Global Surface Water Layer (GSWL) (Pekel et al. 2016), which was generated by a multi-temporal analysis of Landsat imagery. The displayed reference map is an aggregation of the binary monthly GSWL maps for the summer months of 2015, i.e. when a pixel was identified as water in at least two out of the three months June, July, and August 2015, it is considered to contain surface water.

In these visual results, the following phenomena can be observed:

- In the results that utilise Sentinel-1 SAR backscatter (subfigures c,e,f) very often false-positive detections occur. In Fig. 6, this happens on the eastern edge of the scene, which depicts a part of the bog land "Donaumoos". In Figs. 7 and 12 the false positives are spread out through most parts of the scene for the cases in which only VH polarisation and VH polarisation plus MNDWI are used. And in Fig. 10, the SAR input causes a mis-classification of the Munich airport as water.
- Interestingly, false-positive detections also occur in urban areas sometimes when only the 10-m bands of Sentinel-2 are used. This can be observed in Figs. 7d and 10d.
- The continuously most robust results from a visual point of view are obtained when all surface-related bands from both Sentinel-1 and Sentinel-2 are fused (subfigures e),
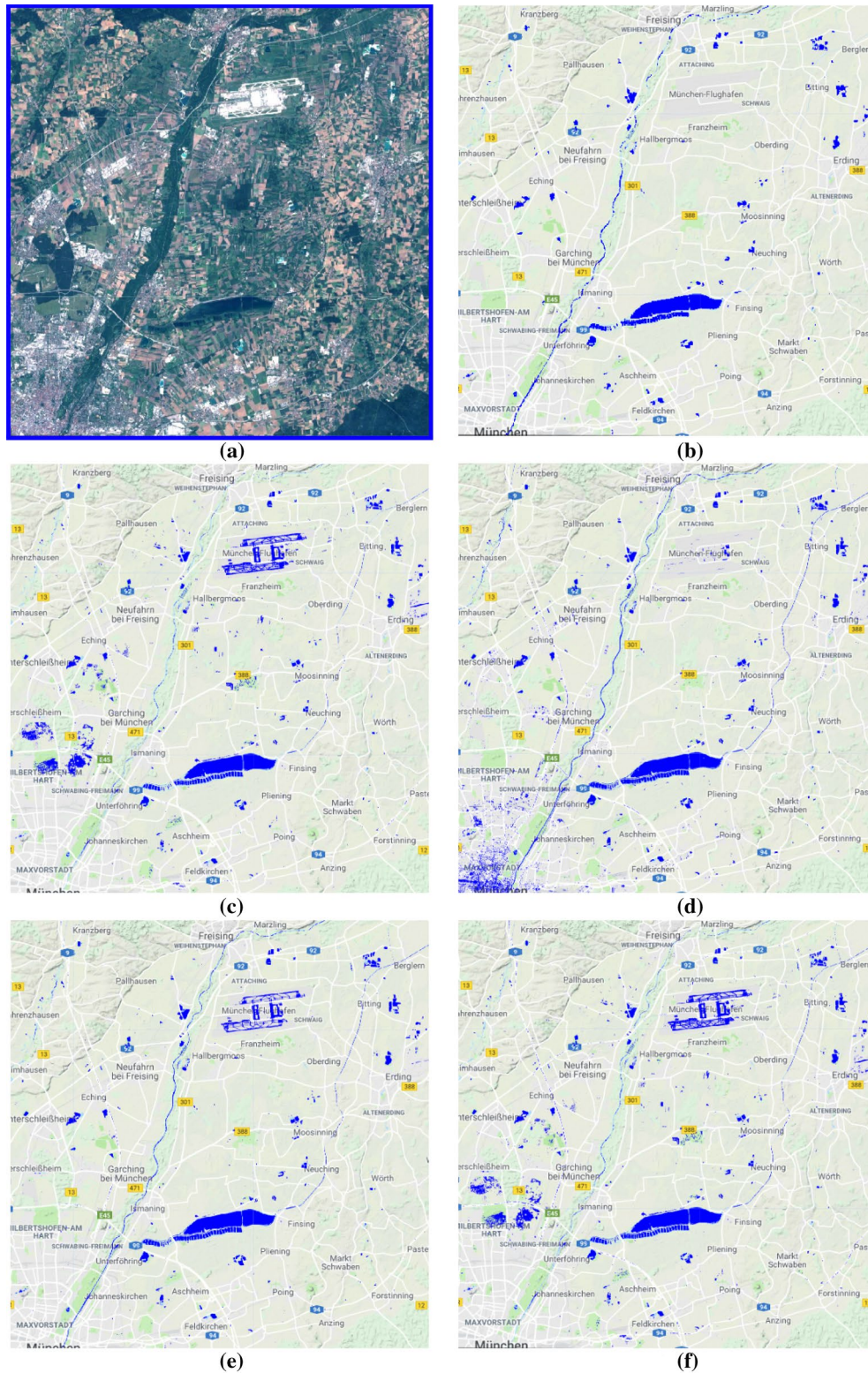
Fig. 10 Example area showing the northeastern surroundings of the city of Munich with the river Isar (flowing in south–north direction), the reservoir Ismaninger Speichersee, and the artificial channel Isarkanal exiting the Speichersee to the east. **a** Optical reference image, **b** GSWL reference layer, **c** Sentinel-1 VH polarisation result, **d** Sentinel-2 10-m bands result, **e** Sentinel-1/Sentinel-2 full-band fusion, **f** Sentinel-1/Sentinel-2 VH/MNDWI fusion

closely followed using only the 10-m bands of Sentinel-2 (subfigures d).

– Several smaller rivers are not present in the GSWL layer, but are found by the all-band fusion (e.g. visible in Fig. 6e, southern part of the scene, Fig. 7e on the southern edge of the city of Bamberg or Fig. 10e with the Isarkanal running parallel to the river Isar in the eastern half of the scene.

– The Forggensee (cf. Fig. 8) is not detected completely in terms of its cartographic outline. This is due to the fact that Forggensee is a man-made reservoir that is regularly emptied. This illustrates the fact that the proposed method only maps the observable water surface extent. It is interesting to note, though, that the models giving large weight to the VH-polarised SAR information (experiments 2 and 11, Fig. 8c, f, respectively) detect more water than the optical-only models.

5 Discussion

5.1 Quality of the Water Detection

Looking at both the quantitative as well as qualitative results summarised in Sect. 4, one can see that fully automatic water surface detection on a state-wide level is generally possible from Sentinel-1/2 data using a simple, SVM-based machine learning approach and OpenStreetMap data for training the classifier. While the accuracy metrics calculated from randomly distributed point samples indicate that the best configurations allow a water detection with up to 98% of producer's and overall accuracy, the visual examples show that the numeric results have to be taken with a grain of salt. Due to the limited resolutions of the Sentinel data (from 10-m GSD to about 20-m GSD), smaller water bodies of course cannot be detected. Since they are not part of the OSM-derived reference data nor the GSWL water layer used for comparison, this is not reflected by the accuracy metrics. Especially setup 11 (Fusion of Sentinel-1 VH and Sentinel-2 MNDWI) shows a larger amount of missed smaller water bodies in comparison to the other results. This is caused by the fact that this configuration provides the overall lowest resolution (about 20m GSD for both channels). However, given the powerful detections provided by the optical 10-m bands, a significant improvement could possibly be expected if higher-resolution SAR or optical data were available. The works of Du et al. (2016) and Yang et al. (2017), which aimed at the generation of a super-resolved SWIR band for

the calculation of a high-resolution MNDWI image, confirm the promising potential in this regard.

Apart from the sensor-inherent limitations with respect to GSD, the rate of missed water bodies is generally very low, i.e. a good level of completeness is achieved.

5.2 False-Positive Detections

In particular by looking at the visual results, it can be observed that some of the results suffer from a certain amount of false positives, e.g. by confusing an airport with water (e.g. in the test cases relying on Sentinel-1 data in Fig. 10). This is also reflected by the low producer's accuracy for the non-water class when using both setup 2 (Sentinel-1 VH channel only) and setup 11 (Fusion of Sentinel-1 VH and Sentinel-2-derived MNDWI), indicating that while the Sentinel-1 information hardly misses any water body that is of sufficient size, it also produces a considerable amount of false positive detections.

This is mainly caused by the fact that for C-band radar signals also smooth impervious surfaces (e.g. pavement) leads to a mirror-like reflection and, thus, very low backscatter amplitudes. Interestingly, also an occasional misclassification of urban areas as water is observed when the 10-m Sentinel-2 bands are used as inputs. This indicates that the visible plus broadband near-infrared bands (B2, B3, B4, B8) are not enough for a robust discriminative model trained with automatically generated training data as proposed in this contribution. Instead, the short-wave infrared information also provided by Sentinel-2 and the Sentinel-1 SAR backscatter turn out to be highly useful for robustifying the water detection—in spite of extreme cases such as the Munich airport (cf. Fig. 10).

This is confirmed by the impression that the overall best result is found to be setup 9, i.e. the straight-forward pixel-level fusion of all surface-related Sentinel-1 and Sentinel-2 bands, especially with respect to the robustness against false water detections. This is not only reflected by the accuracy measures provided in Table 3, but also by the visual results displayed in Figs. 6, 7, 8, 9, 10, 11, 12. Nevertheless, using only the 10-m bands of Sentinel-2 (i.e. setup 4) follows closely second with quite similar accuracy values and visual results, although the number of false positives is a bit higher. This indicates that fusion helps to enhance robustness to some extent; while, the detailed information provided by the 10-m bands is preserved.
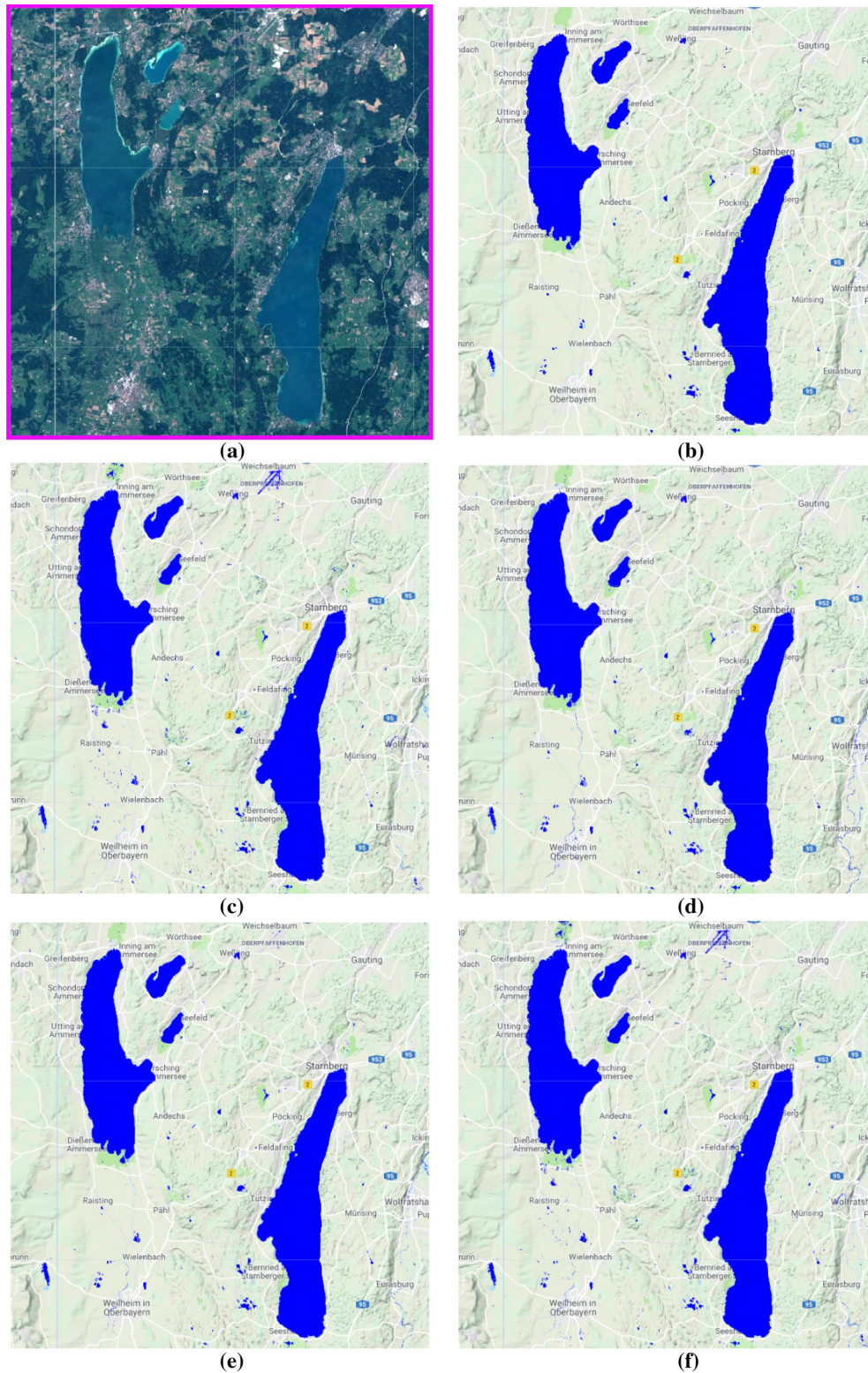
Fig. 11 Example area showing the "Fünfseenland" (with Starnberger See, Ammersee, Pilsensee, Wörthsee, and Weßlinger See). **a** Optical reference image, **b** GSWL reference layer, **c** Sentinel-1 VH polarisa-tion result, **d** Sentinel-2 10-m bands result, **e** Sentinel-1/Sentinel-2 full-band fusion, **f** Sentinel-1/Sentinel-2 VH/MNDWI fusion
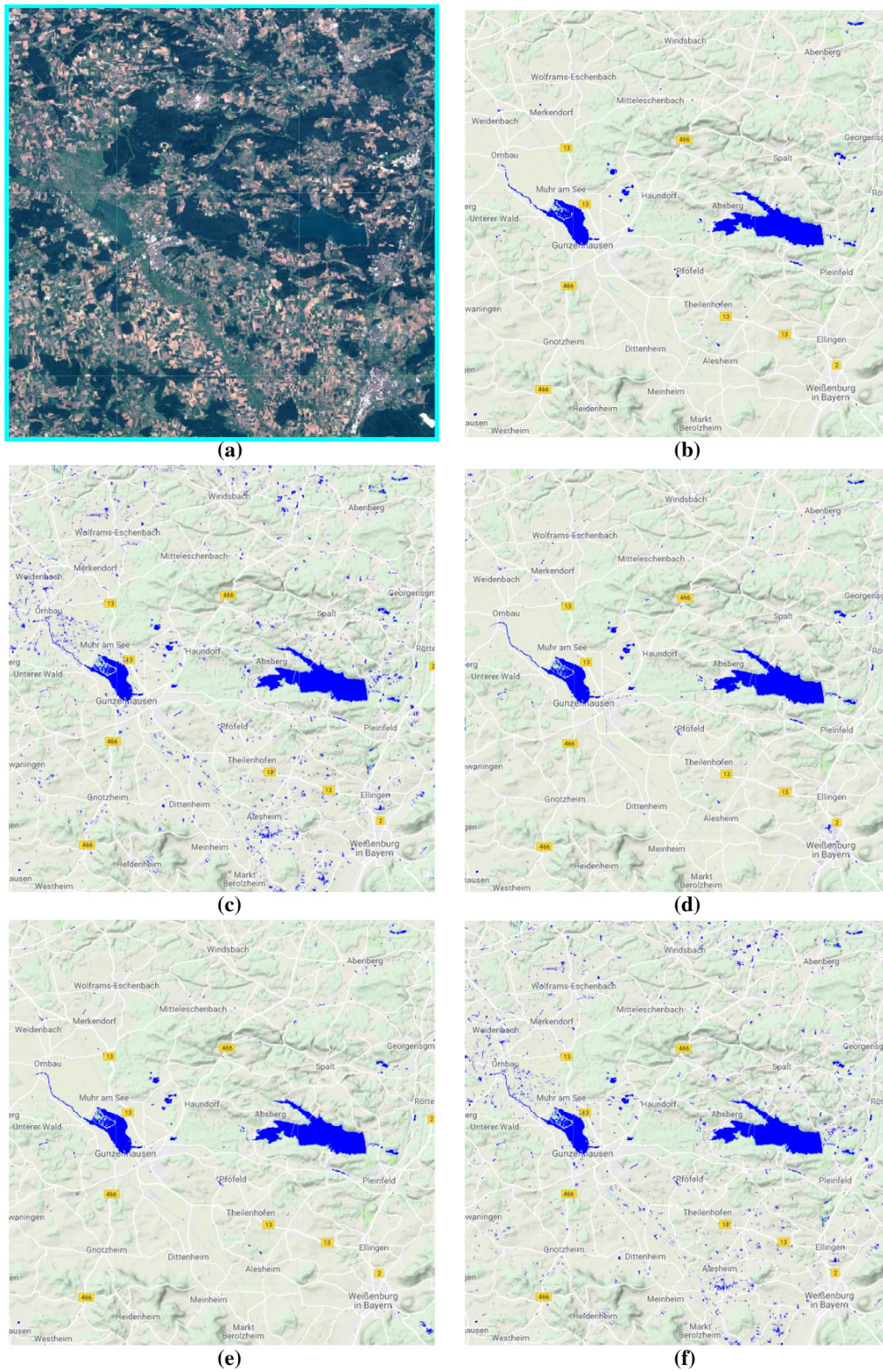
Fig. 12 Example area showing the reservoirs Altmühlsee (including the artificial channel Altmühlzuleiter entering on the west) and Brombachsee. **a** Optical reference image, **b** GSWL reference layer, **c** Sentinel-1 VH polarisation result, **d** Sentinel-2 10-m bands result, **e** Sentinel-1/Sentinel-2 full-band fusion, **f** Sentinel-1/Sentinel-2 VH/MNDWI fusion

5.3 Influence of Data Pre-Processing

Besides the findings discussed so far, one important further point has to be mentioned. Both the Sentinel-1 and Sentinel-2 data as utilised in this study were products of multitemporal data fusion to mitigate speckle and cloud coverage effects (see Sect. 3.2.1). Therefore, the results must be seen as a kind of snapshot for a three-month observation period in the summer of 2018. If a denser monitoring of surface water extents is desired, other data processing strategies have to be used which employ, for example, speckle filtering for the SAR imagery, or sophisticated, cloud-robust techniques for time series analysis in the optical case. Alternatively, instead of a binary water/non-water map, one could seek to provide a map containing frequencies of water coverage as did (Pekel et al. 2016), whose goal was not to map surface water in a static manner, but rather monitor its dynamics as observed by optical Landsat image time series. As mentioned in Sect. 4, the static GWSL maps as displayed in Figs. 6, 7, 8, 9, 10, 11, 12 for reference, however, are the result of an aggregation of binary monthly water maps acquired over a three-month period in summer 2018 for the sake of comparability. As can be seen from the results presented in this contribution, using Sentinel data and an automatically trained supervised learning approach similarly robust maps with even slightly more details can be produced. This is particularly interesting, as (Pekel et al. 2016) relied on a more complicated expert system classifier, which additionally used different sources of auxiliary data (e.g. digital elevation models, glacier data, urban area data and lava masks) to robustify the water detection.

Finally, it has to be noted that both approaches rely on multi-temporal data fusion for surface water detection, so it has to be accepted that the resulting water map is neither a representation of a single time snapshot, nor following a technical definition of the term water body. That is, an area, which is mostly experiencing low-water conditions during the time frame used in the analysis (as, e.g. the southern end of the Forggensee reservoir in Fig. 8), will not necessarily be considered as part of the water body—although it might be in a hydrological sense.

5.4 Label Noise

The last point to discuss is the effect of using the OSM water layer polygons as labels, especially if pre-processed as described in Sect. 3.2.2. Most importantly, it has to be mentioned that, of course, this layer cannot be considered as *ground truth*. There are various reasons for that: On the one hand, water bodies often change their area or shape over time, which is not always reflected in OSM. On the other hand, the OSM quality is not homogeneously high in all regions of the world, which can lead to a significant amount

of label noise in the form of missed or wrongly mapped water bodies. Nevertheless, the advantage of using VGI such as OSM data for the automatic training data generation has to be highlighted. The main reason to use VGI instead of officially available geodata as provided by local governments is this work's goal to provide a simple water mapping tool that is potentially applicable to extended regions (e.g. whole continents). While for many regions across the world at least some amount of OSM information exists, governmental land cover information is often not available. As an alternative to OSM data, one could also consider the GWSL to be used to generate the training labels.

6 Summary and Conclusion

In this work, a simple procedure based on supervised learning via support vector machines and automatic training data annotation with information derived from OpenStreetMap for the large-scale detection of inland water bodies was proposed. The performance of the approach for the State of Bavaria was illustrated and example mapping results for the major rivers, lakes, and reservoirs in that state were provided. Both qualitative and quantitative results confirm that fully automatic large-scale water surface detection is generally possible. The best configuration of input data seems to be a pixel-based fusion of all surface-related Sentinel-1 and Sentinel-2 bands for this purpose, providing an acceptable trade-off between detection accuracy and a certain robustness against false positives. Nevertheless, for operational mapping, post-processing against remaining false positives, e.g. by including elevation data, is advisable. Besides, the proposed framework should be combined with dedicated signal processing algorithms for pre-processing, e.g. filters for SAR image despeckling, or single-image cloud-removal techniques. When this is done, an approximately weekly monitoring of inland water becomes imaginable, as the need to fuse data from a three-month period for speckle reduction and cloud-removal will be removed.

References

Bhargava DS, Mariam DW (1991) Light penetration depth, turbidity and reflectance related relationship and models. ISPRS J Photogr Remote Sens 46(4):217–230

Bioresita F, Puissant A, Stumpf A, Malet JP (2019) Fusion of Sentinel-1 and Sentinel-2 image time series for permanent and temporary surface water mapping. Int J Remote Sens 40(23):9026–9049

Conrad C, Kaiser BO, Lamers JPA (2016) Quantifying water volumes of small lakes in the inner aral sea basin, central asia, and their potential for reaching water and food security. Environ Earth Sci 75:952

Desnos Y, Borgeaud M, Doherty M, Rast M, Liebig V (2014) The European space agency's earth observation program. IEEE Geosci Remote Sens Mag 2(2):37–46

Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, Hoersch B, Isola C, Laberinti P, Martimort P et al (2012) Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sens Environ 120:25–36

Du Y, Zhang Y, Ling F, Wang Q, Li W, Li X (2016) Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. Remote Sens 8:354

Flener C, Vaaja M, Jaakkola A, Krooks A, Kaartinen H, Kukko A, Kasvi E, Hyyppä H, Hyyppä J, Alho P (2013) Seamless mapping of river channels at high resolution using mobile LiDAR and UAV-photography. Remote Sens 5(12):6382–6407

Gao BC (1996) NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens Environ 58(3):257–266

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017) Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens Environ 202:18–27

Hobson DE, Williams D (1971) Infrared spectral reflectance of sea water. Appl Opt 10(10):2372–2373

Isikdogan F, Bovik A, Passalacqua P (2017) RivaMap: An automated river analysis and mapping engine. Remote Sens Environ 202:88–97

Klein I, Dietz AJ, Gessner U, Galayeva A, Myrzakhmetov A, Künzer C (2014) Evaluation of seasonal water body extents in central asia over the past 27 years derived from medium-resolution remote sensing data. Int J Appl Earth Obs Geoinf 26:335–349

Liebe J, van de Giesen N, Andreini M (2005) Estimation of small reservoir storage capacities in a semi-arid environment: a case study in the upper east region of ghana. Phys Chem Earth, Parts A/B/C 30(6–7):448–454

Mountrakis G, Im J, Ogole C (2011) Support vector machines in remote sensing: a review. ISPRS J Photogr Remote Sens 66(3):247–259

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 Sens Environ 174:341–352

Palmer SC, Kutser T, Hunter PD (2015) Remote sensing of inland waters: challenges, progress and future directions. Remote Sens Environ 157:1–8

Pekel JF, Cottam A, Gorelick N, Belward AS (2016) High-resolution mapping of global surface water and its long-term changes. Nature 540(7633):418

Pettorelli N (2013) The normalized difference vegetation index. Oxford University Press, Oxford

Rishikeshan C, Ramesh H (2018) An automated mathematical morphology driven algorithm for water body extraction from remotely sensed images. ISPRS J Photogr Remote Sens 146:11–21

Rivas Casado M, Ballesteros Gonzalez R, Kriechbaumer T, Veal A (2015) Automated identification of river hydromorphological features using UAV high resolution aerial imagery. Sensors 15(11):27969–27989

Rouse JH (1974) Monitoring vegetation systems in the great plains with ERTS. In: Proc. Third Earth Resources Technology Satellite-1 Symposium, pp. 3010–3017. Greenbelt, USA

Rusnak M, Sladek J, Kidova A, Lehotsky M (2018) Template for high-resolution river landscape mapping using UAV technology. Measurement 115:139–151

Rußwurm M, Körner M (2018) Multi-temporal land cover classification with sequential recurrent encoders. ISPRS Int J Geo-Inform 7(4):129

Schmitt M, Hughes LH, Qiu C, Zhu XX (2019) Aggregating cloud-free Sentinel-2 images with Google Earth Engine. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. IV-2/W7, pp. 145–152

Schmitt M, Zhu XX (2016) Data fusion and remote sensing: an ever-growing relationship. IEEE Geosci Remote Sens Mag 4(4):6–23

Schubert A, Small D, Miranda N, Geudtner D, Meier E (2015) Sentinel-1A product geolocation accuracy: commissioning phase results. Remote Sens 7(7):9431–9449

Slinski KM, Hogue TS, McCray JE (2019) Active–passive surface water classification: a new method for high-resolution monitoring of surface water dynamics. Geophys Res Lett 46(9):4694–4704

Torres R, Snoeij P, Geudtner D, Bibby D, Davidson M, Attema E, Potin P, Rommen B, Floury N, Brown M et al (2012) GMES Sentinel-1 mission. Remote Sens Environ 120:9–24

Xu H (2006) Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int J Remote Sens 27(14):3025–3033

Yang X, Zhao S, Qin X, Zhao N, Liang L (2017) Mapping of urban surface water bodies from Sentinel-2 MSI imagery at 10 m resolution via NDWI-based image sharpening. Remote Sensing 9:596

Zeng L, Schmitt M, Li L, Zhu XX (2017) Analysing changes of the Poyang Lake water area using Sentinel-1 synthetic aperture radar imagery. Int J Remote Sens 38(23):7041–7069

Zhu XX, Tuia D, Mou L, Xia G, Zhang L, Xu F, Fraundorfer F (2017) Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci Remote Sens Mag 5(4):8–36