Pix2Streams: Dynamic Hydrology Maps from Satellite-LiDAR Fusion

Dolores Garcia[*][†]
Frontier Development Lab

Gonzalo Mateo-Garcia[*][†]
University of Valencia

Hannes Bernhardt
Arizona State University

Ron Hagensieker
osir.io

Ignacio G. Lopez-Francos
NASA Ames Research Center

Jonathan Stock
USGS

Guy Schumann
RSS-Hydro
University of Bristol

Kevin Dobbs
Frontier Development Lab

Freddie Kalaitzis[*][†]
University of Oxford
Frontier Development Lab

Abstract

Where are the Earth's streams flowing right now? Inland surface waters expand with floods and contract with droughts, so there is no one map of our streams. Current satellite approaches are limited to monthly observations that map only the widest streams. These are fed by smaller tributaries that make up much of the dendritic surface network but whose flow is unobserved. A complete map of our daily waters can give us an early warning for where droughts are born: the receding tips of the flowing network. Mapping them over years can give us a map of impermanence of our waters, showing where to expect water, and where not to. To that end, we feed the latest high-res sensor data to multiple deep learning models in order to map these flowing networks every day, stacking the times series maps over many years. Specifically, **i)** we enhance water segmentation to 50 cm/pixel resolution, a $60\times$ improvement over previous state-of-the-art results. Our U-Net trained on 30-40cm WorldView3 images can detect streams as narrow as 1-3m ($30-60\times$ over SOTA). Our multi-sensor, multi-res variant, **WasserNetz**, fuses a multi-day window of 3m PlanetScope imagery with 1m LiDAR data, to detect streams 5-7m wide. Both U-Nets produce a water probability map at the pixel-level. **ii)** We integrate this water map over a DEM-derived synthetic valley network map to produce a snapshot of flow at the stream level. **iii)** We apply this pipeline, which we call **Pix2Streams**, to a 2-year daily PlanetScope time-series of three watersheds in the US to produce the first high-fidelity dynamic map of stream flow frequency. The end result is a new map that, if applied at the national scale, could fundamentally improve how we manage our water resources around the world.

[*]Equal contribution (listed in alphabetical order).
[†]`dolores.garcia@imdea.org; gonzalo.mateo-garcia@uv.es; freddie@fdl.ai`

Figure 1: Narrow (5-2m) river in the South Dakota AoI. **Left to right:** WorldView 3 (WV3) (50cm/px), PlanetScope (3m/px), Sentinel-2 (30m/px), Digital Elevation Model (DEM) (1m/px), water probability map (50cm/px) of the same area. WV3, All satellite images are acquired on the same day, and are shown with a false RGB composite of Near-Infrared (NIR), Red, and Green bands.

1   Introduction

The United Nations 2020 Sustainable Development Goals report [1] states that *"water scarcity affects more than 40% of the global population and is projected to increase. Over 1.7 billion people are currently living in river basins where water use exceeds recharge"*. Small and medium sized streams make up a significant portion of the flowing network. The receding tips of this network could serve as an early warning for where droughts are born [2, 3, 4]. In the United States, the National Hydrography Dataset (NHD) [5] contains large rivers, some smaller streams, as well as lakes that have been mapped during aerial and on-ground mapping campaigns since the early 20th century. A subset of these streams are monitored by gages that continually record the water levels at discrete locations. However, small and intermittent streams (widths < 90 m) are inadequately mapped and not monitored, even though they contribute 48% of the total surface water extent of the stream network globally [6].

Open data from satellite missions, like Sentinel and Landsat, have been used for water mapping applications [7, 8, 9, 10]. Sentinel and Landsat offer temporally sparse images, in 5 and 16 day intervals, respectively, and at resolutions (10-30m/px) that are too low for detecting small and intermittent streams. At the opposite extreme, very-high resolution (VHR) imagery, like from Maxar's WorldView satellites, can be processed into very high resolution (<5m) maps over smaller areas. Such satellites are generally task-based, that is, they acquire images on customer demand, so multi-temporal coverage is not guaranteed. As a trade-off, the PlanetScope constellation has high-res (3m), daily, wall-to-wall coverage, offering a VHR-grade imagery that is temporally dynamic and spatially extensive.

From pixels to streams   [3] We show that, by combining VHR with high cadence, we can monitor previously unseen streams at near real-time (daily). Our **Pix2streams** pipeline consists of two functional primitives:

1. **Pixel-level mapping:** A U-Net (SOTA semantic segmentation model [11]) creates pixel-level water probability maps. Our variant, WasserNetz, fuses multiple sensors at different resolutions and outputs a water probability map at the resolution of the ground-truth annotation layer (0.5m/px).

2. **Stream-level mapping:** We aggregate these probabilities over a synthetic valley network, derived from 1m LiDAR-based digital elevation maps (DEM), to obtain a stream-level hydrology map.

We apply this pipeline to a 2-year daily time-series of 3m PlanetScope images, in three US watersheds, to obtain a stream-level **dynamic hydrology map**, that is, a collection of N time-series (for N streams) of the stream's surface % covered by water, daily. Through this map, we can now monitor stream flow frequency, and therefore quantify water availability on a daily, monthly, or yearly basis.

Contributions

A We develop WasserNetz, a **water segmentation model for PlanetScope imagery** trained on a custom dataset of manually labeled polygons, derived from same-day very high-resolution WV3 images (0.5m/px). We show that water detection accuracy is improved by fusing the PlanetScope imagery with USGS high-resolution Light Detection and Ranging (LiDAR) derived products.

---

[3]Video on the steps from the pixel-level map, to the **stream-level map**: bit.ly/pix2streams

B We analyze and compare the performance of WasserNetz **with models trained on higher resolution (30-40cm/px) WV3 images** on a manually annotated test dataset. Also, we study the sensitivity of the models **w.r.t. the stream width**.

C We produce a **fundamentally new hydrology map at the stream level**, by fusing the pixel-level output of WasserNetz with the synthetic valley network, derived from the LiDAR-based DEM.

2 Related Work

Most of the work on mapping and monitoring of rivers uses medium-resolution (30m/px) imagery from Landsat or Sentinel-2 [7, 9, 10], and only recently has VHR (sub-meter) imagery been used, albeit with a broad focus on all water bodies [12, 13]. Landsat imagery was used to derive continental and global maps of large river systems, and to estimate river width and surface area at the global scale [6, 14]. Through Google's Earth Engine, Landsat was also used to conduct a multi-annual global inventory of inland waters [8], enabling monthly estimations of water permanence. Limited by Landsat's $30\,\mathrm{m/px}$ resolution, both studies can reliably detect only streams wider than $60 - 90$m (2-3 pixels). Until today, this spatial limitation has been endemic in the detection of narrower streams [15].

Some aspects of our work are not directly comparable to these studies: on the data side, due to the **unique spatial and temporal resolution** of our imagery and LiDAR derivatives. On the application side, our pipeline for **daily VHR river mapping** is the first of its kind. On the algorithmic side, the U-Net [11], with its concurrent encoder-decoder bottlenecks of different resolution inputs, learns multi-scale features to ultimately inform a fine-grained segmentation. This simple and intuitive idea is accepted as **a staple of semantic segmentation** in remote sensing [16, 17, 18, 13, 19].

Instead of comparing architectures, our ablations in Section 5 focus on the resolving capacity of similar U-Nets w.r.t. two fronts: i) training on 30-40cm WV3 vs. 3m PlanetScope + 1m LiDAR derivatives (WasserNetz); ii) stream width.

3 Data

We study three Areas of Interest (AoIs) across the western United States. With limited resources for expert labeling, these AoIs are chosen to provide a representative sample of the morphology, geology and vegetation of terrains. Fig 1 shows an example of the three types of data used per AoI:

WV3: a single mosaic acquired over multiple dates, with 30-50cm ground sample distance (GSD), 8 spectral bands (coastal blue to near-infrared (NIR)), atmospherically corrected and pansharpened.

PlanetScope: 2 years (April 2018 - April 2020) of daily snapshots, with 3m GSD, four bands (rgb + NIR). We used an experimental version, corrected for the presence of clouds, radiometrically harmonized to Landsat, Sentinel-2 and MODIS information [20].

LiDAR derivatives: With the TauDEM suite [21] we produce hydrology products from the bare-earth 1m resolution 3D Elevation Program (3DEP) DEM, like the synthetic valley network, reach catchments, Strahler stream order, and a stream buffer (the distance from a given pixel to the nearest reach). See A.1 for details on the derivatives used for training.

Manually labeled polygons. To train the model in a supervised manner, a set of georeferenced delineated polygons corresponding to 'wet' and 'dry' classes is annotated. These annotations focus on narrow streams and contain labels with shadows for both wet and dry, which are difficult to classify. Polygons are drawn using the WV3 images as a reference to exploit its higher resolution. The dataset contains a total of 5664 labels with 46.39% water polygons, distributed on the three AoIs. We split the dataset into train (5121) and test sets (543), so that they do not overlap spatially.

4 Methodology

Semantic segmentation model We consider the architecture U-Net [11] as a starting point. We modify it to have two downstream branches (PlanetScope images and LiDAR products) each taking a different size input. The branches are merged when the size of the feature maps is the same. To provide the output at the resolution of the ground-truth (50cm/px) we add an upsampling convolution step. We refer to this network as **WasserNetz**. See Appendix A.4 for an illustration.

Loss function   We use the binary cross entropy loss, back-propagating only the pixels inside the rasterized polygon, where ground-truth is available. This is required since only pixels within those polygons are annotated. Also, the loss contribution of each example is weighted by the number of labeled pixels in its ground truth, i.e polygon-level loss. This seeks to mitigate the bias due to polygon size.

Dynamic Hydrology Map   The derived synthetic valley network is a vector layer that delineates all the reaches where water can flow. For each reach we consider its catchment area, that is, the hydrological area unit in which all water ends up in the corresponding reach. Then we aggregate the model prediction (pixel-level water probability map) over each catchment area, to give the probability of having water. For an example, see Figure 2. This reach-level probability is computed by weighting the pixel-level probability map by the distance to the reach, and it represents the likelihood of finding water at any part of that reach. The reaches and their associated water probabilities give a static snapshot of flowing water in the stream network. Applying this process to a time-series of images we obtain the near-real time stream network[4].
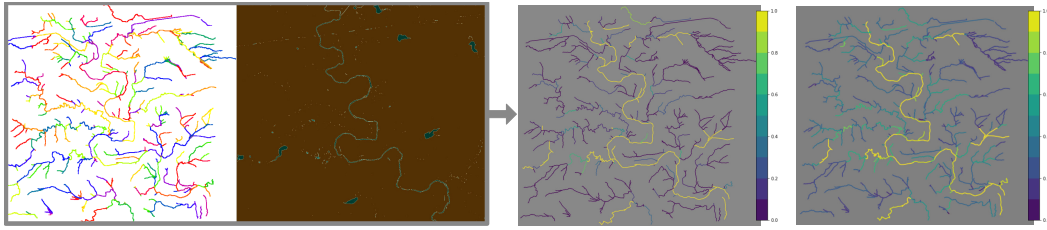

Figure 2: **Left to right:** 1) Synthetic valley network where each reach in the network is shown with a different color. 2) Pixel-level water probability map (colormap: [0-"dry"/ 1-"water"] → [brown, blue]) 3) derived stream-level hydrology map for a single day. 4) Annual stream-level hydrology map produced for 2019.

5   Results

We measure the performance of our water segmentation model with the accuracy at the pixel-level and polygon-level. The latter seek to balance wet and dry classes since dry polygons are usually larger. PR-curves,included in the appendix A.3, show similar behaviors. Table (1) shows the results for different model configurations. The multi-sensor segmentation (LiDAR+PlanetScope) outperforms by 3% the model trained only on Planet data. This shows that LiDAR-derived products complement 3m imagery yielding a significant improvement. The U-Net trained on WV3 (0.5m) outperforms WasserNetz by 7%. This is expected, since below a critical width some streams can be resolved at 0.5m while not at 3m resolution. The Normalized Difference Water Index (NDWI) approach [22], shown as a baseline, shows the worst performance, with a 30-40% loss in accuracy over all models.

Also, we quantify the sensitivity of our pixel-level hydrology maps w.r.t. the width of the streams (Fig. 3). We perform this analysis on a subset (280 polygons) of our test set that include width annotations enclosing the opposing banks of a water body. A positive relationship between width and performance is observed for all models. In South Dakota, the WV3-trained model is the most robust but without a significant performance margin. Also, South Dakota's landscape conditions seem to be favorable for WasserNetz (LiDAR+PlanetScope), as they allow it to be competitive to the WV3-trained U-Net. Idaho (ID), with extensive shadows and canopy cover, poses an unfavorable environment for PlanetScope-trained models, whose performance degrades significantly on streams narrower than 5m. However, the performance gain of a multi-sensor model over a Planet only model is 18% for narrower streams (<5m). This suggests that our approach can lead to a large improvement in complex landscapes, especially for narrow streams (<5m).

Fig. 2 shows the annual (2019) stream-level water probabilities, produced by averaging daily stream-level probabilities, derived from PlanetScope images through our proposed pipeline.

---

[4]Video on steps from pixel-level map, to stream-level, to **dynamic hydrology map**: bit.ly/pix2streams
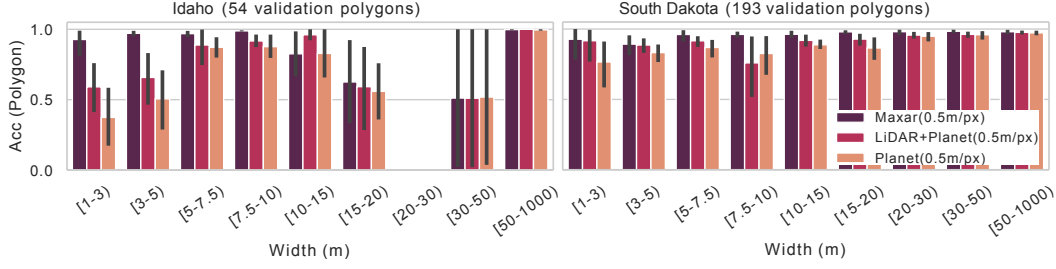
Figure 3: Models performance on wet polygons of the test dataset stratified by stream width.

Table 1: Polygon-level, pixel-level accuracy on test data. Mean (stdev) across 10 runs of different random seeds.

Data	Model	Input m/px	Output m/px	Polygon-level	Pixel-level
LiDAR+PlanetScope	WasserNetz	3 / 1	0.5	88.24 (0.36)	94.34 (0.43)
WV3	U-Net	0.5	0.5	95.26 (0.23)	96.87 (0.21)
PlanetScope	U-Net	3	0.5	85.42 (0.24)	91.60 (0.52)
WV3	NDWI	0.5	0.5	56.96	71.25

6 Conclusions

Our multi-sensor deep-learning model can segment water at unprecedented scales, from a fusion of multispectral satellite imagery and LiDAR-derived products, processed at their native resolutions. Fusing with LiDAR, compared to an optical-only model, improves the detection of narrow streams (< 5m) by 18% in difficult areas (with shadows and canopy cover), and by 8% in optically cooperative areas. We derive the first **dynamic hydrology map** from the predictions of the semantic segmentation model (WasserNetz) over a time-series of 2 years. We emphasize the trade-off between spatial and temporal resolution in existing VHR instruments: even though this work pushes the observation frontier down to the scale of 1-3m stream widths ($30\text{-}60\times$ over SOTA) with WorldView3 data, our proposed **pix2streams** pipeline, and the main idea of our paper, rests on the availability of daily VHR imagery, as exemplified by the PlanetScope constellation.

Future work We plan to extend our results to other regions that are optically cooperative and open-source a larger dataset of annotated polygons. To further ease the interpretation of the stream-level probabilities in our dynamic hydrology maps, more work must be done to disentangle the spatial proportion of water coverage in a stream, from the uncertainty in a model prediction. The latter can be further broken down into *aleatoric* uncertainty (noisy data and annotations) and *epistemic* uncertainty (eg. ambiguity in features of shadow vs. water). Finally, we believe that the connectedness in dendritic stream networks, as an inductive bias, can inspire new architectures and losses that improve the data-efficiency and robustness in stream mapping.

References

[1] United Nations Statistics Division. *The Sustainable Development Goals Report 2020*. United Nations, 2020.

[2] Anne C Steinemann and Luiz FN Cavalcanti. Developing multiple indicators and triggers for drought plans. *Journal of Water Resources Planning and Management*, 132(3):164–174, 2006.

[3] Jamie Hannaford, Benjamin Lloyd-Hughes, Caroline Keef, Simon Parry, and Christel Prud-homme. Examining the large-scale spatial coherence of european drought using regional indicators of precipitation and streamflow deficit. *Hydrological Processes*, 25(7):1146–1162, 2011.

[4] Cecilia Svensson, Jamie Hannaford, and Ilaria Prosdocimi. Statistical distributions for monthly aggregations of precipitation and streamflow in drought indicator applications. *Water Resources Research*, 53(2):999–1018, 2017.

[5] U.S. Geological Survey. National hydrography dataset, 2016.

[6] George H Allen and Tamlin M Pavelsky. Global extent of rivers and streams. *Science*, 361(6402):585–588, 2018.

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

[8] Jean-François Pekel, Andrew Cottam, Noel Gorelick, and Alan S. Belward. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633):418–422, December 2016.

[9] Furkan Isikdogan, Alan Bovik, and Paola Passalacqua. RivaMap: An automated river analysis and mapping engine. *Remote Sensing of Environment*, 202:88–97, December 2017.

[10] Xiucheng Yang, Qiming Qin, Hervé Yésou, Thomas Ledauphin, Mathieu Koehl, Pierre Grussenmeyer, and Zhe Zhu. Monthly estimation of the surface water extent in France at a 10-m resolution using Sentinel-2 data. *Remote Sensing of Environment*, 244:111803, July 2020.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[12] Mark Randall, Rasmus Fensholt, Yongyong Zhang, and Marina Bergen Jensen. Geographic object based image analysis of WorldView-3 imagery for urban hydrologic modelling at the catchment scale. *Water*, 11(6):1133, 2019.

[13] Wenqing Feng, Haigang Sui, Weiming Huang, Chuan Xu, and Kaiqiang An. Water body extraction from very high-resolution remote sensing imagery using deep u-net and a superpixel-based conditional random field model. *IEEE Geoscience and Remote Sensing Letters*, 16(4):618–622, 2018.

[14] George H Allen and Tamlin M Pavelsky. Patterns of river width and surface area revealed by the satellite-derived north american river width data set. *Geophysical Research Letters*, 42(2):395–402, 2015.

[15] Renato Prata de Moraes Frasson, Tamlin M Pavelsky, Mark A Fonstad, Michael T Durand, George H Allen, Guy Schumann, Christine Lion, R Edward Beighley, and Xiao Yang. Global relationships between river width, slope, catchment area, meander wavelength, sinuosity, and discharge. *Geophysical Research Letters*, 46(6):3252–3262, 2019.

[16] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.

[17] Sisi Wei, Hong Zhang, Chao Wang, Yuanyuan Wang, and Lu Xu. Multi-temporal sar data large-scale crop mapping based on u-net model. *Remote Sensing*, 11(1):68, 2019.

[18] Wei Yao, Zhigang Zeng, Cheng Lian, and Huiming Tang. Pixel-wise regression using u-net and its application on pansharpening. *Neurocomputing*, 312:364–371, 2018.

[19] Sherrie Wang, William Chen, Sang Michael Xie, George Azzari, and David B Lobell. Weakly supervised deep learning for segmentation of remote sensing imagery. *Remote Sensing*, 12(2):207, 2020.

[20] Rasmus Houborg and Matthew F McCabe. A cubesat enabled spatio-temporal enhancement method (cestem) utilizing planet, landsat and modis data. *Remote Sensing of Environment*, 209:211–226, 2018.

[21] Ahmet Artu Yıldırım, Dan Watson, David Tarboton, and Robert M. Wallace. A virtual tile approach to raster-based calculations of large digital elevation models in a shared-memory system. *Computers & Geosciences*, 82:78–88, September 2015.

[22] S. K. McFeeters. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7):1425–1432, 1996.

[23] Manuel Guizar-Sicairos, Samuel T. Thurman, and James R. Fienup. Efficient subpixel image registration algorithms. *Optics Letters*, 33(2):156–158, January 2008.

[24] Abhinav Valada, Gabriel L Oliveira, Thomas Brox, and Wolfram Burgard. Deep multispectral semantic scene understanding of forested environments using multimodal fusion. In *International Symposium on Experimental Robotics*, pages 465–477. Springer, 2016.

[25] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, and Qingmin Liao. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 21(12):3106–3121, 2019.

[26] Michel Deudon, Alfredo Kalaitzis, Israel Goytom, Md Rifat Arefin, Zhichao Lin, Kris Sankaran, Vincent Michalski, Samira E Kahou, Julien Cornebise, and Yoshua Bengio. Highresnet: Recursive fusion for multi-frame super-resolution of satellite imagery. *arXiv preprint arXiv:2002.06460*, 2020.

A Appendix

A.1 LiDAR derived products

The derivative data layers derived from the 1 meter resolution 3DEP DEMs over all study areas are:

- Sink-filled DEM
- Flow direction
- Flow Accumulation
- Synthetic reach network, thresholded to 10K square meters minimum catchment size
- Horizontal distance buffer to reach network
- Reach catchments
- Strahler stream order
- Height Above Ground (HAG).

Inputs to the WasserNetz models are the DEM, the Synthetic reach network, horizontal distance buffer to reach network and Height Above Ground (HAG).

A.2 Co-registration

Co-registration issues are observed due to the different orthorectification methods used by the different data sources. In particular, we observed shifts of WV images of around 1-20 meters compared with the DEM and the PlanetScope images. An example of these shifts is shown in Figure. 4. This misalignment will affect models with PlanetScope and/or DEM inputs trained with the polygons, which were annotated using WV images as a reference. To fix this issue, we use off-the-shelf image registration algorithms [23], to shift the polygons for each example when training on Planet and/or LiDAR derived data. Although with this method we fix large misalignments between Planet/DEM images and labeled polygons, small shifts might still be present as the different resolution of the sources limit the shift corrections.
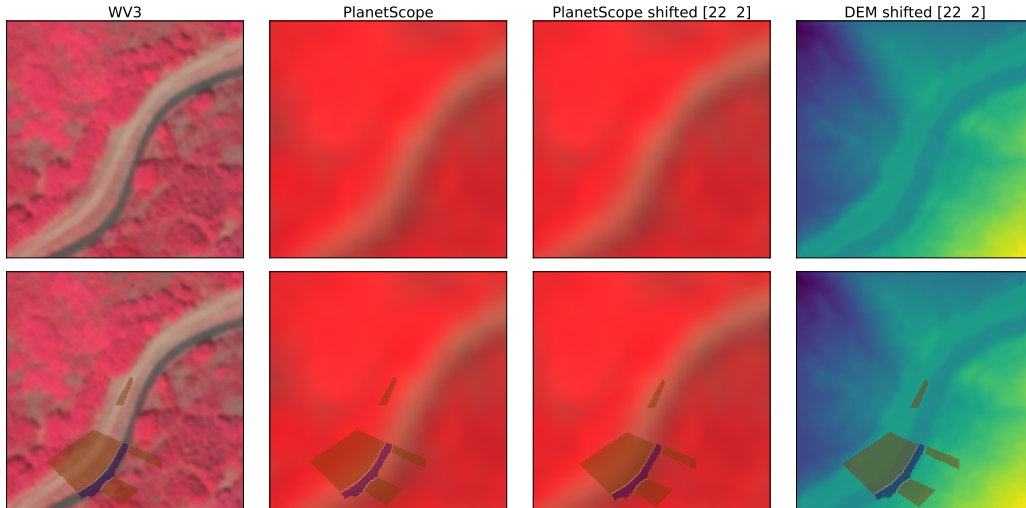

Figure 4: From left to right WV3 image, PlanetScope Image, PlanetScope image shifted to correct the polygon shift, DEM image shifted to correct the polygon shift. Second row shows the same image overlaid with training polygons.

A.3 Precision Recall Curves

A precision-recall curve shows the trade-off between precision and recall as we vary the threshold on the output of the network. Figure 5 shows precision recall on the raw pixel output and re-weighted by

the number of pixels of each polygon. The crosses show the 0.5 probability threshold that was used in the models for simplicity.
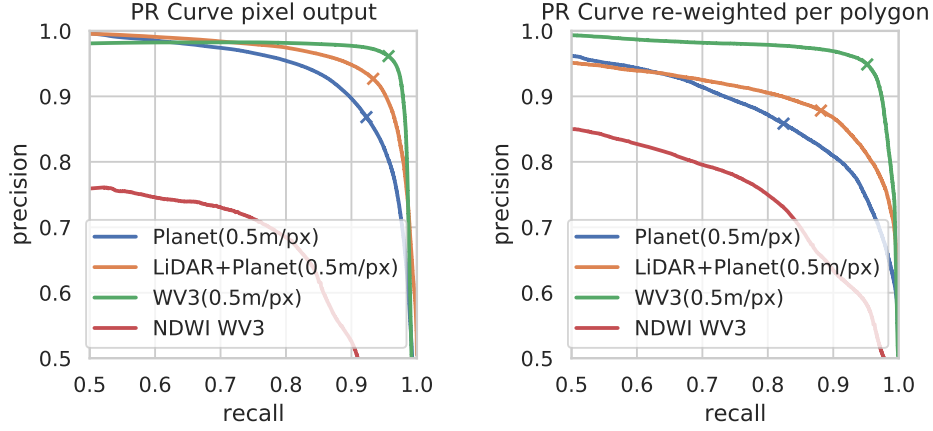

Figure 5: Precision-recall curves of the different models on the pixel-level output (left) and weighted per polygon, aka. polygon-level (right). Crosses shows the 0.5 threshold.

A.4   WasserNetz

Fig 6 shows the architecture of WasserNetz, our variant of the U-Net architecture adapted for multi-sensor multi-resolution input fusion, and an output at some desired target resolution.

With a U-Net backbone, the design of WasserNetz follows common practices of image processing, such as resampling, which defines our multi-sensor multi-res input fusion [24]. We also resample the output to evaluate it at the native resolution (30-40cm/px) of our ground-truth – a common practice in super-resolution [25, 26].

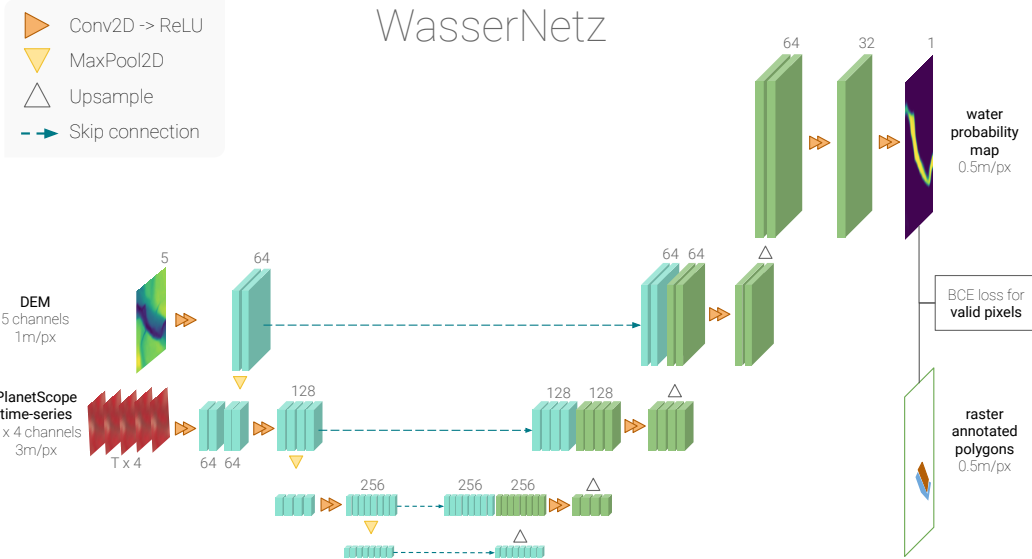In light of our limited labeling resources, our sparse but pixel-based annotations allow a trade-off between full and weak supervision [19], see Section 3.

Figure 6: WasserNetz architecture.