Satellite Imagery and AI: A New Era in Ocean Conservation, from Research to Deployment and Impact

Patrick Beukema
AI2
patrickb@allenai.org

Favyen Bastani
AI2
favyenb@allenai.org

Piper Wolters
AI2
piperw@allenai.org

Henry Herzog
AI2
henryh@allenai.org

Joe Ferdinando
AI2
joef@allenai.org

Abstract

Illegal, unreported, and unregulated (IUU) fishing poses a global threat to ocean habitats. Publicly available satellite data offered by NASA and the European Space Agency (ESA) provide an opportunity to actively monitor this activity. Effectively leveraging satellite data for maritime conservation requires highly reliable machine learning models operating globally with minimal latency. This paper introduces three specialized computer vision models designed for synthetic aperture radar (Sentinel-1), optical imagery (Sentinel-2), and nighttime lights (Suomi-NPP/NOAA-20). It also presents best practices for developing and delivering real-time computer vision services for conservation. These models have been deployed in Skylight, a real time maritime monitoring platform, which is provided at no cost to users worldwide.

1 Introduction

Unprecedented environmental catastrophes compounded by ruthlessly efficient fishing are pushing our oceans to the brink. Entire species have gone missing seemingly overnight. Last year, 10 billion snow crabs vanished from the Bering Sea prompting the fishery to close for the first time in its history [1]. Worldwide, it is estimated that 34% of fisheries are unsustainably harvested [2], a concerning trend that continues to escalate.

Remote satellite data coupled with artificial intelligence provide a means to monitor and deter unregulated and illegal fishing, one of the biggest threats to marine ecosystems. Although no single satellite can provide adequate coverage of the entire planet, employing many satellites with a variety of passive and active sensors enhances the likelihood of identifying destructive behavior as it occurs. Large, publicly available image data from a diverse constellation of satellites enables real time monitoring of the entirety of the world's oceans.

This paper provides an overview of three novel computer vision models designed for near real time vessel detection. Each of these models has been deployed in Skylight [3], a maritime intelligence platform, supporting international conservation efforts through real-time monitoring. Skylight is provided for free to users worldwide, spanning 308 organizations and over 60 countries.

Computational Sustainability Workshop at NeurIPS 2023.

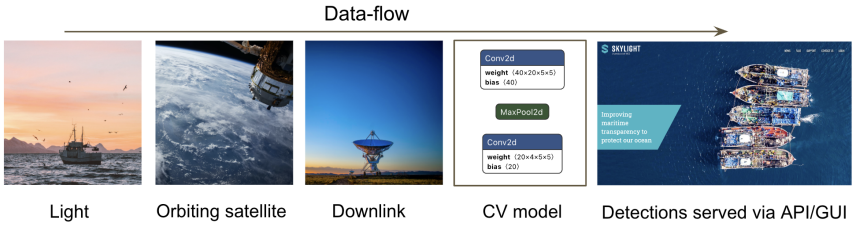

Figure 1: Data-flow depiction of a real-time streaming computer vision service for vessel detection in satellite imagery. An orbiting satellite images a vessel. The image is downlinked to a ground station. That data is copied to Skylight owned servers and processed by a computer vision model [4]. The resulting vessel detection is reported to our users through a GUI and available via an API.

Table 1: Overview of each satellite detection service

	Provider	PX (m)	Signal	Revisit rate	Latency	Vessel Count (10/2023)
VIIRS	NASA	750	watts	2x/night	2.5 hrs	145,063
Sentinel-1	ESA	10	radar	14 days	5 hrs	182,234
Sentinel-2	ESA	10	optical	5 days	5 hrs	430,467

2 Building computer vision models for maritime intelligence

Achieving high performance is critical for our users, who cannot afford to expend limited resources, such as fuel, on pursuing non-existent vessels.. However, while performance is paramount, there are other important considerations that anchor research and development, including minimizing latency and ensuring adequate interpretability. It is essential to report a vessel's presence as quickly as possible. Although the overall latency–from the vessel to the satellite, then to the model, API, and finally the user–is dominated by the downlink latency (see table 1 "Latency"), computational efficiency is important. This efficiency facilitates high throughput iteration and regular upgrading. In addition, the model outputs should be interpretable. If the model commits egregious errors or its reasoning is opaque, our users cannot (and should not) trust its outputs. For this reason, the platform outputs a simple crop centered on each vessel detection (Fig. 2.2 D-F) to allow users to visually inspect every detection. We aim for transparency and share documentation about the model creation and ML strategy to help establish confidence in the machine intelligence.

In the following sections we provide a brief description of the unique characteristics and modeling strategy for each satellite. Table 1 provides a high level overview of each vessel detection service. The code and model architectures alongside complete processing pipelines and additional details have been open sourced on GitHub [5, 6].

2.1 Vessel detection in VIIRS imagery

The Visible Infrared Imaging Radiometer Suite (VIIRS) sensor on board the Suomi-NPP and NOAA-20 satellites collect visible and infrared images during both the day and night [7]. While not originally intended as a real time vessel monitoring data source, the low latency ($\sim$ 2 hrs), global coverage, satellite redundancy, and unique signal characteristics make the VIIRS sensor a useful tool in the fight against illegal fishing. However, the low spatial resolution (750m) precludes distinguishing vessels from non vessels prima facie (see example detection in Fig. 2.2D). Therefore, care must be taken to achieve high precision.

The modeling strategy adopted a three stage approach. The first stage consisted of a classical computer vision model, trained without supervision, to extract all possible sources of light. This was achieved with a simple 2D kernel. In the second stage, all known non-vessel light sources (lightning, gas flares, moonlit clouds, the northern and southern lights, and ionospheric particles from within the South Atlantic Anomaly) are removed through a series of postprocessing steps [8]. These non-vessel light sources often exhibit stereotyped distributional patterns (unlike vessels) that is amenable to rules based logic. Additionally, we implemented statistical tests to identify unusually geographically distributed vessels coincident with scan lines and suppressed false positives at the frame's extremeties due to the "noise smile" [9] to control the false positive rate. The final stage involved filtering all positive detections through a regularly updated 2d CNN. This CNN was trained on human annotated

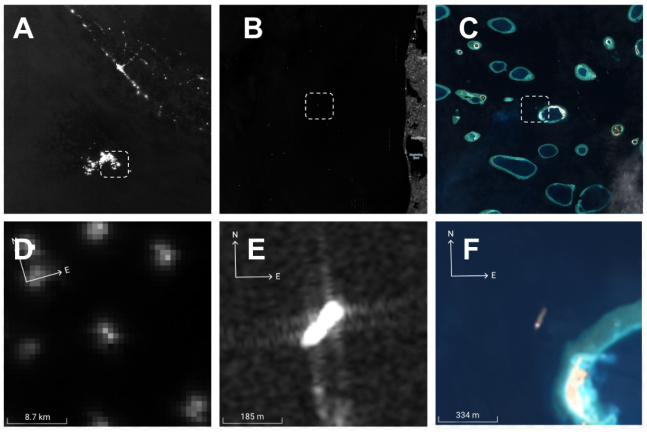

Figure 2: Example satellite imagery (top row) and sample detections (bottom row) from a VIIRS image (A, D) near the Ecuadorian coast, an S1 image (B, E) from the North Sea and an S2 image (C, F) from the Maldives. Scale bars are approximate. Confidence scores ≥ 0.95.

image labels (correct/not) with four channels (nanowatts, land water masks [10], moonlight, and clouds [11, 12]). This model was specifically designed to run in resource constrained environments, requiring only modest hardware (4 GB RAM, no GPU).

End-to-end deep learning based approaches were also evaluated. However, given the simplicity and limited spatial extent of the objects (1-2 pixels, as shown in 2.2D), and their sparse distribution, end-to-end deep learning based models required significantly more computational power to achieve performance comparable to our hybrid design. Our hybrid approach is designed to be highly efficient, which allows for regular and economical retraining using new labeled data. Such efficiency is particularly beneficial for machine learning-specific continuous integration and continuous delivery (CI/CD) pipelines including extensive model-specific integration testing. Thorough testing (especially within the CICD framework) is beneficial for preventing regressions during phases of fast-paced development, and is typically prohibitively expensive with conventional large-scale DNNs.

2.2 Vessel detection in S1 imagery

Sentinel-1 (S1) is a satellite constellation from the ESA providing 10 m/pixel synthetic aperture radar (SAR) imagery, which measures properties of energy reflections from the planet surface [13, 14]. It captures Earth's land surface and coastal waters roughly every two weeks. The visual features are more complex than VIIRS, requiring detailed visual discrimination to distinguish vessels from islands and fixed marine infrastructure, and we found that deep learning methods were required. We created a dataset of 55,499 vessels (point labels) annotated by maritime experts.

We developed a detection model consisting of a standard Faster-RCNN [15] head, with a customized backbone. The backbone consists of a small 13-layer fully-convolutional encoder, which outputs feature maps at four resolutions that are processed through a feature pyramid network (FPN) [16]. We adapt the backbone to input not only the current target S1 image (with two bands, VV and VH polarizations) in which vessels should be detected, but also one or more aligned historical images of the same region at different times. These historical images enable the model to learn that marine objects consistently in the same location, such as fixed platforms and islands, are unlikely to represent transient objects like vessels. We process the images independently through the backbone, and at each resolution we concatenate the features of the target image with the pooled features of the historical images and pass the result to the FPN. Note that an early version of this model architecture scored fourth place in the xView3 competition [17]. The training data used here does not overlap with the xView3 data, and in contrast to xView3, is 100% human annotated (rather than by machine).

2.3 Vessel detection in S2 imagery

Sentinel-2 (S2) is another ESA satellite constellation, providing optical imagery with four bands at 10 m/pixel, six bands at 20 m/pixel, and three bands at 60 m/pixel [18]. It captures Earth's land surface

and coastal waters roughly every 5 days. The same maritime experts annotated 43,102 vessels (point labels) in S2 imagery.

The various S2 bands provide rich information about the physical objects present in a scene. For example, RGB bands already enable distinguishing most vessels from other marine objects, but additional bands can be leveraged to further improve accuracy due to the different reflectance signatures of vessels and other objects. Thus, we developed a detection model that, like our S1 model, uses a Faster-RCNN detection head, but we couple it with a much larger Swin Transformer [19] backbone that has sufficient parameters to perform complex analysis of the S2 bands. We found that pre-training the backbone on SatlasPretrain [20], a large-scale remote sensing dataset, further improved performance. Unlike with S1, we did not observe a performance increase from inputting historical images. We speculate that this is because the visual signatures of vessels in S2 optical images are sufficiently distinct from stationary objects (e.g., platforms or wind turbines), so historical images are not needed.

2.4 Model evaluation, validation, and deployment

While there is no single metric that we use to determine when a model is ready to be released to our users, each model is continuously evaluated against a variety of criteria as it passes from research, to staging, to deployment. During the research phase, the primary method of evaluation consists of offline F1 scores against large and randomly held out validation datasets (where possible). For S1, we also compared the model against a previous version that had been submitted to the xView3 competition (the new model improved from 70.1% to 82.7% F1). The S2 model was not part of an external competition, but exhibited a similar F1 score of 0.81. There was no large annotated dataset for the VIIRS model, however, we did compare its performance against the industry standard model on previously human validated frames [9]. Due to the fact that offline performance of VIIRS was limited, we supplemented this evaluation with extensive unit and integration tests (CI/CD), covering a variety of known failure modes (aurora, moonlit clouds, imaging artifacts, etc).

Once all failure modes have been addressed, we transition models into an online staging environment that replicates production dataflows (i.e. streaming and real-time inference). Importantly, all of the inferenced data in the staging environment is new (out-of-sample) data. Models are evaluated in this staging environment for extended periods (months) by subject matter experts versed in maritime intelligence and regularly updated in response to that feedback.

Once confident that a new model is performant, we deploy models from staging into the production environment. In production, the primary method of evaluation consists of user feedback (internal and external) which gives us prompt notice of performance degradation or model drift.

After deployment, models are regularly upgraded (monthly release cadence) in response to feedback and/or new information sources. It is worth noting that while the data sources are (largely) stable, ocean activity is highly dynamic. For example, marine infrastructure (wind turbines, oil platforms, etc) are constantly evolving. These dynamics must be tracked and addressed through regular maintenance to maintain high performance. For example, to improve precision of each of the above models, we recently added an additional postprocessing step that geofences false positive detections coincident with recently detected [21] marine infrastructure produced by Bastani et al. [20].

2.5 Satellite and GPS correlations

Every vessel detection is augmented by the addition of GPS information, when available, provided by the Automatic Identification System (AIS). Most vessels broadcast their locations, but some do not. Notably it is possible to obfuscate one's position by suppressing AIS but quite challenging to evade detection from a satellite. Vessels that neglect to broadcast their locations, but are still visible under radar or satellite imagery are especially relevant to analysts. In a typical satellite image, there are many detected vessels, and many possible candidate matches from AIS signals in the vicinity. Therefore, it is necessary to correlate the signals from these two information sources. Fig. 3 shows a depiction of the correlation process. Matching geolocations from an image and as determined by GPS can be formulated as a simple minimum weight matching problem in bipartite graphs. We apply a Jonker-Volgenant algorithm [22] as implemented by Pedregosa et al. [23] to assign matches.

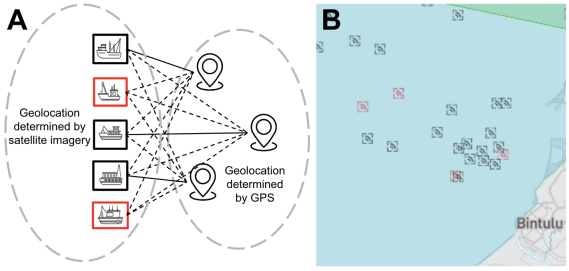

Figure 3: A. Depiction of the correlation process. We compute the haversine distance between vessels in imagery and as located by AIS, then minimize the distance over the pairs. B. Panel from the Skylight UI showing radar detections. Correlated = black, uncorrelated = red.

3 Best practices for shipping computer vision for maritime intelligence

Deploying computer vision models into a real time streaming context has offered many valuable lessons, particularly around closing the gap between offline batch and online streaming performance.

- Near real-time satellite data may exhibit unique features or artifacts that will be missed if training on historical imagery. If streaming inference is the goal, ensure that the model is trained against the same data that will ultimately be inferenced in production.
- The performance of static models, i.e. those with frozen weights, is at risk of regression due to model or data drift. Ensure that model iteration is as simple and as automated as possible in order to facilitate seamless retraining from feedback.
- Allow ample time to empirically assess the model performance in real world conditions and at appropriate temporal and spatial scales to identify and correct problems as they occur. For example, VIIRS is highly sensitive to the lunar cycle.
- Users cannot rely on machine intelligence unless it is consistently available and reliable. Expect to dedicate significant engineering resources beyond ML to ensure that the model is always online and maintained.
- Employ the best practices from software engineering during research and development. In particular, continuous integration and continuous deployment, unit and integration tests, code quality enforcement, and documentation are all essential.
- Satellite imagery exhibits massive variation at global scale and therefore it can be challenging to anticipate how performance will degrade on out-of-sample data. For example, after deployment we discovered hitherto unknown sources of false positives due to the Aurora Borealis/Australis for VIIRS, newly constructed wind turbines in the North Sea for the S1 model, and sargassum patches [24] in the Caribbean for the S2 model.

4 Equity considerations

Real time vessel detections from each these models are provided through Skylight, a free maritime intelligence platform. The purposes of these models specifically, and Skylight more broadly, is to help nations protect their marine resources and promote ocean health for future generations. While we believe that the benefits of this technology (both within our platform and as open source repositories) outweigh potential risks, there is a possibility these models could be used for ignoble ends. We do not have a straightforward response that obviates these concerns and we do not take the decision to open source this technology lightly. We chose to open source these models because we believe that both the machine learning research community and the conservation community using the vessel detections should have complete transparency and full access into the underlying model architectures and logic (both for inference and training). In addition, these models are made possible by the existing open geospatial community, especially NASA and the ESA which provides both historical data used for training and near real time data for inference and monitoring.

Acknowledgments and Disclosure of Funding

Allen Institute for Artificial Intelligence (AI2)

All authors are employees at AI2. Skylight is a product of AI2.

Defense Innovation Unit

The Defense Innovation Unit contributed funding which supported creation of the VIIRS and S2 vessel detection models and supported the improvements of the S1 model. The S1 model described in this paper was an extension of a previous model created by an AI2 team that was submitted to the DIU xView3 competition.

NASA & LANCE (VIIRS)

We acknowledge the use of data and imagery from NASA's Land, Atmosphere Near real-time Capability for EOS (LANCE) system (`https://earthdata.nasa.gov/lance`), part of NASA's Earth Observing System Data and Information System (EOSDIS). The data products that were used for VIIRS model are described here: `https://lance.modaps.eosdis.nasa.gov/viirs/`. The following products are used. For SuomiNPP, `VNP02DNB_NRT` (light), `VNP03DNB_NRT` (supporting data), `VNP02MOD_NRT` (gas flares). For NOAA20, `VJ102DNB_NRT` (light), `VJ103DNB_NRT` (supporting data), `VJ102MOD_NRT` (gas flares). In addition, we use the cloud masks created by the University of Wisconsin SSEC (`https://www.earthdata.nasa.gov/learn/find-data/near-real-time/viirs-a`) [12, 11]. Additional details and code to use these data can be found on the VIIRS GitHub repository [5].

The ESA (European Space Agency) and the Copernicus Data Space Ecosystem (S1 and S2)

Data for the S1 and S2 constellations are available at https://dataspace.copernicus.eu/. Additional details of the specific products used for each satellite model are provided below.

- S1: Level-1 GRD (Ground Range Detected) data were used. Models were trained with both VV and VH polarization modes.
- S2: Level-1C data (orthorectified top-of-atmosphere reflectance, with sub-pixel multispectral registration). In addition to the L1C data, we also apply cloud detection to suppress false positives due to clouds. To do so we use the s2 cloud detector (s2cloudless) available from PyPI [25]. More details on this algorithm are available in [26].

Note that we previously used Copernicus Open Access Hub which is deprecated as of October 2023, and replaced by the Copernicus Data Space Ecosystem.

References

[1] Alaska Department of Fish and Game. 2022/23 bering sea snow crab season closed, 2022. URL `https://www.adfg.alaska.gov/static/applications/dcfnewsrelease/1441272349.pdf`. Accessed on November 20, 2023.

[2] FAO. *The State of World Fisheries and Aquaculture 2020: Sustainability in Action*. Rome, 2020.

[3] Skylight, 2023. URL `https://www.skylight.global/`. Accessed on: September 1 2023.

[4] Lutz Roeder. Netron - visualizer for neural network models, 2023. URL `https://github.com/lutzroeder/netron`.

[5] Allen Institute for AI. Vessel detection with viirs. `https://github.com/allenai/vessel-detection-viirs`, 2023. Accessed on: September 1 2023.

[6] Allen Institute for AI. Vessel detection with sentinels. `https://github.com/allenai/vessel-detection-sentinels`, Year of access. Accessed on: Current date.

[7] Christopher D. Elvidge, Kimberly E. Baugh, Mikhail N. Zhizhin, Feng-Chi Hsu, and Tilottama Ghosh. Viirs night-time lights. *International Journal of Remote Sensing*, 38:5860 – 5879, 2017. URL https://api.semanticscholar.org/CorpusID:134847136.

[8] Andreas Nilsson, Neil Suttie, Joseph S. Stoner, and Raimund Muscheler. Recurrent ancient geomagnetic field anomalies shed light on future evolution of the south atlantic anomaly. *Proceedings of the National Academy of Sciences of the United States of America*, 119, 2022. URL https://api.semanticscholar.org/CorpusID:249434557.

[9] Christopher D. Elvidge, Mikhail N. Zhizhin, Kimberly E. Baugh, and Feng-Chi Hsu. Automatic boat identification system for viirs low light imaging data. *Remote. Sens.*, 7:3020–3036, 2015. URL https://api.semanticscholar.org/CorpusID:10198494.

[10] J. R. G. Townshend R. A. Sohlberg A. I. Elders S. Devadiga A. M. Sayer M. L. Carroll, C. M. DiMiceli and R. C. Levy. Development of an operational land water mask for modis collection 6, and influence on downstream data products. *International Journal of Digital Earth*, 10(2):207–218, 2017. doi: 10.1080/17538947.2016.1232756. URL https://doi.org/10.1080/17538947.2016.1232756.

[11] S. Ackerman et al. Viirs atmosphere l2 cloud mask product. NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA, 2017.

[12] S. Ackerman et al. Viirs atmosphere l2 cloud mask product. NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA, 2017.

[13] Ramon Torres, Paul Snoeij, Dirk Geudtner, David Bibby, Malcolm Davidson, Evert Attema, Pierre Potin, BjÖrn Rommen, Nicolas Floury, Mike Brown, et al. Gmes sentinel-1 mission. *Remote sensing of environment*, 120:9–24, 2012.

[14] Karen Fletcher. *SENTINEL 1: ESA's Radar Observatory Mission for GMES Operational Services*. European Space Agency, 2012.

[15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems*, 28, 2015.

[16] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature Pyramid Networks for Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2117–2125, 2017.

[17] Fernando Paolo, Tsu-ting Tim Lin, Ritwik Gupta, Bryce Goodman, Nirav Patel, Daniel Kuster, David Kroodsma, and Jared Dunnmon. xView3-SAR: Detecting Dark Fishing Activity Using Synthetic Aperture Radar Imagery. *Advances in Neural Information Processing Systems*, 35: 37604–37616, 2022.

[18] European Space Agency. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. ESA Publications, 2012. URL https://www.esa.int/About_Us/ESA_Publications/ESA_SP-1322_2_Sentinel_2. ESA SP-1322/2.

[19] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021.

[20] Favyen Bastani, Piper Wolters, Ritwik Gupta, Joe Ferdinando, and Ani Kembhavi. SatlasPretrain: A Large-Scale Dataset for Remote Sensing Image Understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

[21] Allen Institute for AI. Geospatial data products. https://github.com/allenai/satlas/blob/main/GeospatialDataProducts.md, 2023. Accessed: [insert access date here].

[22] Roy Jonker and Ton Volgenant. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38:325–340, 1987. URL https://api.semanticscholar.org/CorpusID:7806079.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[24] Kristie S. T. Alleyne, Donald Harry Johnson, Francis C. Neat, Hazel A. Oxenford, and Henri Valls. Seasonal variation in morphotype composition of pelagic sargassum influx events is linked to oceanic origin. *Scientific Reports*, 13, 2023. URL `https://api.semanticscholar.org/CorpusID:257367760`.

[25] Sentinel Hub. sentinel2-cloud-detector. `https://github.com/sentinel-hub/sentinel2-cloud-detector`, 2023. Accessed: 2023-11-18.

[26] Sergii Skakun, Jan Wevers, Carsten Brockmann, Georgia Doxani, Matej Aleksandrov, Matej Batič, David Frantz, Ferran Gascon, Luis Gómez-Chova, Olivier Hagolle, Dan López-Puigdollers, Jérôme Louis, Matic Lubej, Gonzalo Mateo-García, Julien Osman, Devis Peressutti, Bringfried Pflug, Jernej Puc, Rudolf Richter, Jean-Claude Roger, Pat Scaramuzza, Eric Vermote, Nejc Vesel, Anže Zupanc, and Lojze Žust. Cloud mask intercomparison exercise (cmix): An evaluation of cloud masking algorithms for landsat 8 and sentinel-2. *Remote Sensing of Environment*, 274:112990, 2022. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2022.112990. URL `https://www.sciencedirect.com/science/article/pii/S0034425722001043`.