

# Artificial intelligence to advance Earth observation: a perspective

Devis Tuia<sup>1,\*</sup>, Konrad Schindler<sup>2</sup>, Begüm Demir<sup>3,14</sup>, Gustau Camps-Valls<sup>4</sup>, Xiao Xiang Zhu<sup>5</sup>, Mrinalini Kochupillai<sup>5</sup>, Sašo Džeroski<sup>6</sup>, Jan N. van Rijn<sup>7</sup>, Holger H. Hoos<sup>7,13</sup>, Fabio Del Frate<sup>8</sup>, Mihai Datcu<sup>9,10</sup>, Jorge-Arnulfo Quiané-Ruiz<sup>14,15</sup>, Volker Markl<sup>3,12</sup>, Bertrand Le Saux<sup>11</sup>, Rochelle Schneider<sup>11</sup>

<sup>1</sup> **Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland**

<sup>2</sup> **ETH Zurich, Switzerland**

<sup>3</sup> **Technische Universität Berlin, Germany**

<sup>4</sup> **Universitat de València, Spain**

<sup>5</sup> **Technische Universität München, Germany**

<sup>6</sup> **Jozef Stefan Institute, Ljubljana, Slovenia**

<sup>7</sup> **Leiden University, the Netherlands**

<sup>8</sup> **University of Rome “Tor Vergata”, Italy**

<sup>9</sup> **German Aerospace Center (DLR), Germany**

<sup>10</sup> **University Politehnica of Bucharest (UPB), Romania**

<sup>11</sup> **European Space Agency (ESA) Φ-lab, Italy**

<sup>12</sup> **German Research Center for Artificial Intelligence, Germany**

<sup>13</sup> **RWTH Aachen University, Germany**

<sup>14</sup> **BIFOLD - Berlin Institute for the Foundations of Learning and Data, Germany**

<sup>15</sup> **IT University of Copenhagen, Denmark**

**\* Corresponding author: [devis.tuia@epfl.ch](mailto:devis.tuia@epfl.ch)**

---

# Contents

<b>Introduction</b>	<b>4</b>
<b>I Modelling — Machine learning, computer vision and processing</b>	<b>6</b>
<b>1 Computer vision for Earth observation</b>	<b>6</b>
1.1 Specific challenges of Earth observation . . . . .	6
1.2 General computer vision challenges relevant for Earth observation . . . . .	8
1.3 Beyond computer vision . . . . .	9
<b>2 Machine learning for Earth observation</b>	<b>10</b>
2.1 Learning paradigms in Earth observation . . . . .	10
2.2 Learning from noisy labels for Earth observation . . . . .	11
2.3 Automated machine learning for Earth observation . . . . .	11
2.4 Domain adaptation for large-scale Earth observation data analysis . . . . .	12
2.5 Continual learning from growing archives . . . . .	13
<b>3 Advanced processing and computing</b>	<b>14</b>
3.1 Earth observation ecosystems – new system architectures . . . . .	14
3.2 Federated learning . . . . .	15
3.3 AI interoperability . . . . .	15
3.4 Scalable data processing . . . . .	16
3.5 On-board data processing . . . . .	16
<b>II Understanding — Physics-machine learning interplay, causality and ontologies</b>	<b>18</b>
<b>4 Knowledge-based AI and Earth observation</b>	<b>18</b>
4.1 Decision modelling and decision support with Earth observation . . . . .	18
4.2 Knowledge-based machine learning and Earth observation . . . . .	19
4.3 Ontologies for open science in machine learning and Earth observation . . . . .	20
<b>5 Explainable AI and causal inference</b>	<b>20</b>
5.1 Explainable AI . . . . .	21
5.2 Causal inference . . . . .	22
<b>6 Physics-aware machine learning</b>	<b>23</b>
6.1 Inverting physical models with machine learning . . . . .	24
6.2 Hybrid machine learning: interactions of transfer models and machine learning . . . . .	24

---

6.3	Discovering equations from data . . . . .	25
<b>III</b>	<b>Communicating — Machine-user interaction, trustworthiness &amp; ethics</b>	<b>26</b>
<b>7</b>	<b>User-centric Earth observation</b>	<b>26</b>
7.1	Decreasing the user load with active learning . . . . .	26
7.2	Language-based interaction . . . . .	27
7.3	Improved exploration and storytelling . . . . .	27
7.4	Enhanced visualisation . . . . .	28
<b>8</b>	<b>Earth observation and society: the growing relevance of ethics</b>	<b>29</b>
8.1	Trustworthy AI and AI ethics . . . . .	29
8.2	Data ethics . . . . .	30
	<b>Conclusions</b>	<b>30</b>

---

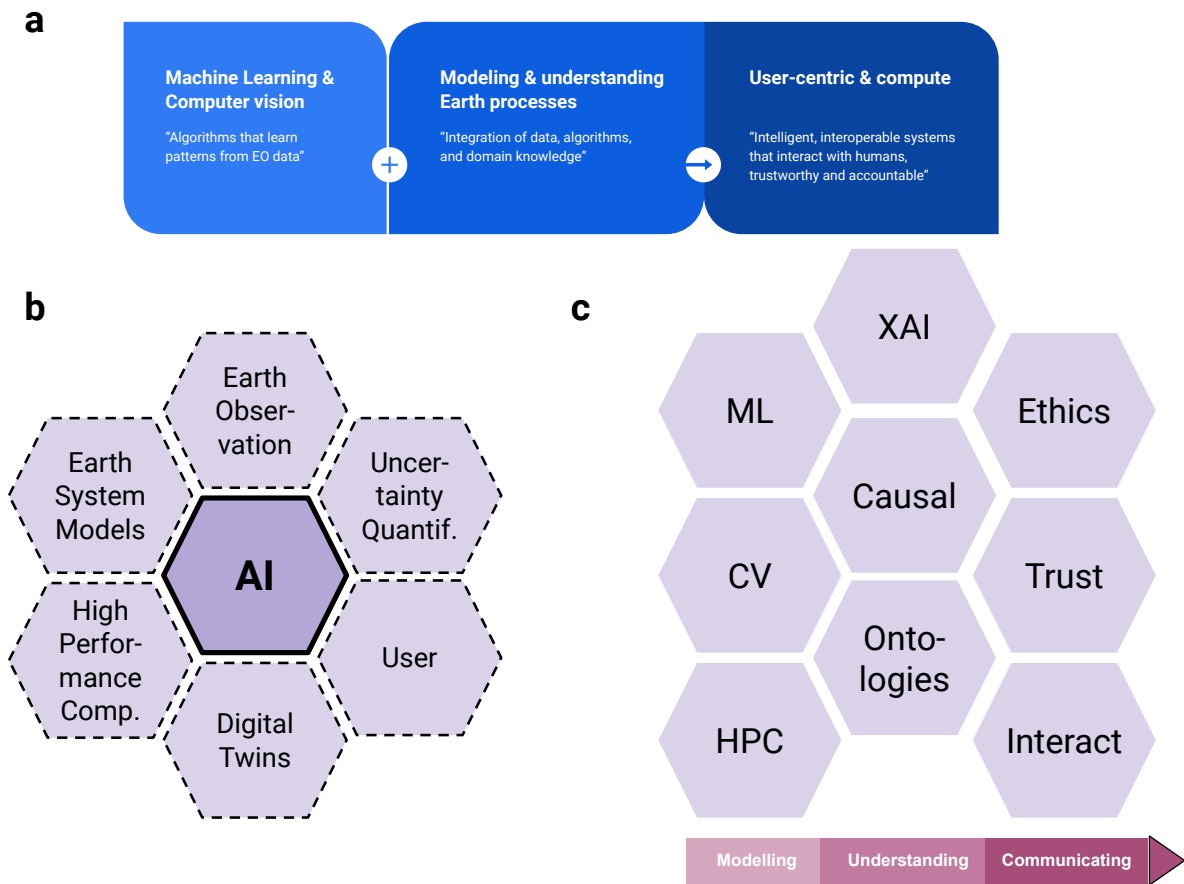
## Introduction

Earth observation (EO) is a prime instrument for monitoring land and ocean processes, studying the dynamics at work, and taking the pulse of our planet. A large variety of sensor data (active / passive / of many resolutions) are nowadays accessible to researchers, agencies, and the general public. However, a final barrier remains: the need for technology that can convert the enormous quantities of raw EO data being generated daily into the valuable information necessary for making decisions and for taking concrete action. The ability to extract meaningful information from raw EO data is an essential prerequisite to achieving significant impact: be it in terms of monitoring or documenting (e.g. progress towards the United Nation’s sustainable development goals), predicting and issuing timely warnings (e.g. related to future natural disasters [132] and need for emergency evacuations), or projecting the effects of human actions and natural processes on nature and society [68, 131, 212].

From the beginning of EO research, computer science and signal processing techniques have been of key importance for extracting meaningful information from raw data. However, we have observed rapid technological advances in these fields over the last few years, particularly owing to the increasing use of artificial intelligence (AI) methods – and in particular, techniques from machine learning (ML) in the context Earth observation. Notably, the use of deep learning (DL) and other ML techniques has had broad and major impact on EO and remote sensing [138, 222], where they are used across entire processing chains, from data compression and transmission to image recognition and predictions of environmental variables (land cover, land use, biomass, etc.). Similarly, ML techniques are increasingly widely used in many areas of environmental science [32].

However, while classical processing tasks, such as land cover classification, seem to have reached a certain degree of maturity, partially due to the better integration of EO and ML methods, substantial work remains to be done to render these approaches truly useful for users and society. In a recent article, an agenda for future technical achievements required in this context has been proposed [192], which triggered this effort from AI4EO scholars and the European Space Agency  $\Phi$ -Lab. Accordingly, with this overview article, we aim to explore and discuss in more detail the development and use of AI methods in the field of EO, with an emphasis on ML techniques, and to thus inspire the EO community to realise the transformative advances made possible by these techniques.

More specifically, this article gives a bird’s eye view of the essential scientific tools and approaches informing and supporting the transition from raw EO data to usable EO-based information. The promises, as well as the current challenges of these developments, are highlighted under dedicated sections. Specifically, we cover the impact of (i) Computer vision; (ii) Machine learning; (iii) Advanced processing and computing; (iv) Knowledge-based AI; (v) Explainable AI and causal inference; (vi) Physics-aware models; (vii) User-centric approaches; and (viii) the much-needed discussion of ethical and societal issues related to the massive use of ML technologies in EO. Figure 1 summarises the content and organisation of this perspective paper.



**Figure 1.** Conceptual overview of this perspective paper: (a) different levels of algorithms emerge from the areas of machine learning (ML) and interact with computer vision (CV), computer science, and statistics to learn patterns and associations from observational data. The models must integrate domain knowledge and biogeophysical constraints to advance in the modelling and understanding the Earth's processes. The ultimate goal is to provide intelligent, interoperable, actionable, trustworthy, robust systems whose decisions should be accountable. (b) The field of AI– and specifically, the area of ML within it – interacts (and is embedded into) several systems to realise such ambitious goal; from high-performance computing platforms in digital twins to Earth system's model simulations and products, a wide range of Earth observation data, the characterisation and quantification of uncertainty, as well as the (active) role of the users. (c) The processing chain in AI goes from simple modelling (e.g. classification, detection, parameter retrieval) with ML, CV and high performance computing techniques that answer 'what questions', to the more ambitious goals of explainable AI, causal relations and ontologies that answer 'what if' questions, and finally to communicate decisions, which involves ethical issues, trust and interaction with the user.

---

## Part I

# Modelling — Machine learning, computer vision and processing

## 1 Computer vision for Earth observation

A significant portion of EO data has the form of images, i.e. measurements that are distributed densely and regularly over the Earth’s surface. This includes optical imagery (panchromatic, multi-spectral, hyper-spectral), thermal imagery, but also other surface scanning techniques like synthetic aperture radar (SAR), as well as image-like raster layers derived from raw observations (e.g. elevation models).

Computer vision (CV), the discipline concerned with computational methods to extract information from image data, is a natural tool for EO. Indeed, the two fields have many fundamental tasks in common: visual recognition and detection of objects, segmentation and semantic labelling of image regions, change detection and time series analysis based on image sequences, spatially explicit regression (retrieval) of continuous surface variables including counting and density estimation, geometric 3D reconstruction, etc. Some of the earliest examples of computer vision algorithms were developed for EO [20, 64, 65].

CV has been an early adopter of machine learning and AI methods, which has benefited Earth observation, e.g. by developing deep neural network architectures that are particularly well-suited for image analysis, and are now in wide use for EO. Also, the quest for high-level, human-like perception inspires new research directions in EO, like monocular 3D perception or visual question answering.

On the other hand, particular challenges specific to EO do not typically arise in generic CV studies and call for more targeted research efforts.

### 1.1 Specific challenges of Earth observation

In remote sensing, one routinely needs to process data from multiple sources to solve a problem jointly, sometimes referred to as “data fusion” [163]. Obvious examples include, e.g. (i) pan-sharpening to combine high spatial resolution images (obtained by integrating over a wide spectral range) with multiband, high spectral resolution images (obtained at the cost of coarser spatial sampling), (ii) the fusion of optical and radar data to mitigate their weaknesses like occlusion by clouds, respectively radar shadows, or (iii) joint processing of imagery and elevation data to implicitly or explicitly address the impact of the surface topography. From the AI perspective, this is an interesting case of representation learning, where a joint latent representation (or encoding) shall be derived for data from different sources. Different strategies are conceptually related to the classical notions of early or late fusion and intermediate variants [159]. Still, it remains a largely open question how to optimally combine measurements from

---

different, possibly varying sensors and modalities into a generic representation [162, 186, 188]. While at the same time, we are faced with an increasingly broad, diverse, and ever-changing range of EO data, including not only satellite and aerial imagery, but also opportunistic data sources such as ground-based images [23, 104, 205] or social media posts [92]. Arguably, an ideal remote sensing system promptly delivers downstream users the required geospatial information using any sensor data pertinent to the task. This view suggests that user-centric abstractions, tailored to specific information extraction tasks but which can be derived in an invariant manner from a range of different inputs, could significantly extend the impact and quality of EO services (see Section 7).

Another challenge is the synthesis between the purely statistical, data-driven modern AI paradigm and the rich physical domain knowledge about the phenomena we aim to observe in the Earth system, leading to hybrid modelling approaches [30]. Informally speaking, physical models excel at efficiently and coherently describing well-understood relations so that one can quantify them in a (largely) unbiased manner with explicit, low-dimensional equations. Such models exist for many phenomena of interest in Earth observation, such as atmospheric physics, hydrology, vegetation dynamics, etc. Whereas the strength of (statistical) AI is to construct good predictive models for phenomena where the mechanistic relations are either poorly understood or intractable by fitting generic, high-dimensional (appropriately regularised) functions to large volumes of data. One particular class of tasks where AI models presently shine are perception tasks like computational vision (but also the processing of sound signals and written text). In EO, physical models exist for the processes we aim to observe, but extracting them from the sensor data is a perception task. Given the complementary nature of the two paradigms, it can be expected that a hybrid approach should be beneficial. Not only would the results be consistent and explainable in terms of the underlying, known physical/chemical/biological processes. One would also need a lot less training data and, at the same time, be more immune to overfitting if the model exploits relations and constraints that are known a priori and focuses on learning patterns that are not known yet (see Section 6).

A specific challenge in EO is the type of reference data (i.e. the ground truth, in the form of class labels or prediction targets) available for training. Modern machine learning approaches rely on large volumes of labelled training data that one might not be able to collect for EO, for instance if fieldwork is required to acquire them. In the case of spatially explicit computer vision tasks, it is often also assumed that the prediction target comes in the form of dense 2D maps and hence includes detailed information about spatial context and layout. Moreover, in various standard computer vision benchmarks, the datasets are often reasonably balanced. On the contrary, EO reference data is often less ready to be used in AI-based systems. First, reference data used for the ground truth on a given location can be associated with various acquisition procedures and instruments, with multiple angles of view and resolutions, resulting in misalignment among data. Also, the reference data may have semantics unrelated to what can be seen in the image data, e.g. parcels or cartography information. Second, ground truth may come as sparse, point-wise observations without surrounding context, making it hard to learn spatial patterns [80, 101]. Third, often the data distributions are extremely unbalanced with

---

long tails [201], and the problem is exacerbated by the fact that the rare, under-represented extreme cases may be significant (e.g. extremely high vegetation, rare high-value crops, local artificial changes in the uninhabited wilderness, etc.). Finally, the output spaces in remote sensing may contain additional structure, like crops or vegetation types in a taxonomic hierarchy [193], or in general, correlations between different output variables.

## 1.2 General computer vision challenges relevant for Earth observation

Besides the open problems listed so far, several unsolved computer vision questions are essential for the endeavour of visual EO. Several of them are open problems of AI and machine learning as a whole that frequently emerge in the context of computer vision (also see Section 2).

Arguably one of the main frontiers of present-day AI is the step from supervised to unsupervised learning and self-supervised learning (SSL). In SSL, we use part of the data properties as source of supervision: for example, one can rotate the image and predict the rotation applied [70] or predict whether a series of images has been taken in the same season [118]. This topic is further discussed in Section 2. The success of learning-based computer vision, and the current generation of computer vision systems, is primarily due to supervised learning from large volumes of *labelled* training data, for which the desired outputs are known. Biological vision [61], as well as information-theoretic considerations [75], suggest that visual perception can be learned with a lot fewer annotations by discovering inherent structure in the (unlabeled) data and abstracting it into a representation. Some progress towards (partly) unsupervised learning has been made over the past years, especially in text processing and language models [25]. Still, unlabeled data, which is available in virtually unlimited quantities, at this point plays at most an auxiliary role [13, 118] in computer vision and EO, and the supervision does the heavy lifting with known training labels.

Alternatively, rather than ambitiously start with unsupervised learning and then pragmatically include some supervision as required, one can start from the well-established supervised schemes and search for ways to reduce the supervision. Simple forms like transfer learning, i.e. adapting a well-trained system to a related task by fine-tuning with a smaller training set, are routinely used [40, 81]. On the contrary, more challenging scenarios are at most partially solved and are currently very active research topics. Examples include unsupervised domain adaptation [116, 180], where one has a well-trained system but only unlabeled data for the target task; few-shot learning [24, 83, 111, 181], where the range of target values (e.g. object classes, terrain heights, etc.) shall be expanded with only a few examples for the new, previously unseen cases; zero-shot learning [208], where the learned representation should have enough metric structure to handle previously unseen output values outside the training range; and meta-learning to obtain flexible representations that can be transferred to new tasks with little effort and training data [24, 63, 83, 155]. A rich literature already exists about all these attempts to increase the generalisation and label efficiency, c.f. Section 2.

An open question specific to computer vision, and its application for EO, is the role of AI and learning in



---

the context of 3D scene reconstruction and how to integrate learning into conventional 3D vision based on well-established stereo triangulation, shading, and direct range measurements (e.g. with LiDAR or InSAR). Exciting attempts have been made to integrate machine learning components into 3D surface reconstruction [41,50,100,144,174], and to incorporate 3D reconstruction and semantic interpretation [21]. But we are still far from a complete understanding of the opportunities and limitations of AI for 3D vision. The field partially overlaps with the above-mentioned hybrid, physics-informed ML, since parts of the 3D vision system, like the shading/rendering equations, are well-understood physical models. One may speculate that combining the well-known multi-view geometry with learning-based models may give rise to geometry-informed machine learning models.

### 1.3 Beyond computer vision

Finally, we argue that CV methods should not be viewed in isolation. In the context of EO, CV is a tool that needs to be embedded in the larger context of integrated Earth system monitoring and decision support.

From that perspective, one important question is how to best integrate AI systems into the EO infrastructure. Arguably, computer vision capabilities and modules should not be added to an EO program ex-post as independent components; instead, they should already be considered in the planning phase and inform the system design. For instance, it may make sense to plan space missions synergistically, such that they support AI methods to combine their respective outputs into a higher-value product. At present, such combinations are employed opportunistically where they arise, e.g. commercial very high-resolution images feature only a few coarse spectral channels and are thus often fused with spectrally richer, but spatially coarser imagery like Landsat or Sentinel-2 [66,158]. Another example considers combining accurate but sparse spaceborne LiDAR data with optical imagery (in some cases from different, independent national space programs) to generate dense maps of vegetation structure [101,140]. We postulate that systematically planned combinations that factor in the potential of joint processing could significantly increase the impact of many satellite missions. A further example where computer vision and machine learning must be included as a mandatory component of the planning are onboard AI processors that deploy AI in space, for instance, to coordinate constellations or to compress and preprocess data.

Once one includes AI in the system design, the question naturally arises why it should be limited to low-level processing and perception. We speculate that in the long run computer vision will only be a component of much larger AI systems that go all the way to spatial and high-level reasoning and direct decision support. One might even argue that turning EO data into maps and stopping there is an arbitrary and unnatural cut, as there is no clear boundary between perception and reasoning. Some machine learning researchers claim [43] that, at the technical level, one may have to integrate statistical AI with relational, symbolic AI (see Section 4) as reasoning gains more importance.

---

## 2 Machine learning for Earth observation

Modern applications of Earth observation face new challenges. For example, increasing amounts of data become available through continuous data streams with extremely high frequency. Moreover, more complex types of data are being introduced. Additionally, the data and labels can be noisy. All these issues have further raised the expectations and challenges placed on ML for EO.

### 2.1 Learning paradigms in Earth observation

Most of the tasks for EO (e.g. land-cover maps generation and updating) are usually achieved by supervised classification techniques, which require a set of labelled samples for training the classification algorithm. Land-use land-cover (LULC) classes are related to physical phenomena characterized by specific properties (spatial variability of classes, non-stationary spectral signatures, etc.) that are very difficult to be modelled without genuine labelled data [27]. Gathering a sufficient number of high-quality labelled samples is time-consuming, costly and may not be feasible in large-scale EO applications. The quality of samples is related to different factors: (i) the degree to which the training data is sampled independently (this is often violated because training data is often collected in neighbouring locations on the ground); (ii) the intrinsic variability of the spectral signatures of land-cover classes which is related to the acquisition and the ground conditions [27]. This limits the design of effective classification models with supervised learning schemes.

To overcome the difficulties of assembling large-scale labelled training samples, DL models pre-trained on publicly available datasets (e.g. ImageNet) have been favoured within the CV community. However, recent studies have demonstrated that this approach is not adequate for remote sensing (RS) due to the differences between the characteristics of the images in CV and RS (e.g. Sentinel-2 multispectral images contain 13 spectral bands associated with varying and lower spatial resolutions) and their semantic content (and thus the considered semantic classes) [133, 177]. Alternatively, one can enrich the information given as input to the supervised classifier by iteratively expanding the original training set via human/machine interaction. This approach, known as active learning, is detailed in Section 7.

To avoid the extensive cost of collecting large-scale, labelled training sets, self-supervised methods have attracted substantial attention in RS in the context of learning general image features from large-scale unlabeled data without using any human-annotated labels (see Section 1). Interesting research directions for EO include: 1) integrating active learning and self-supervised learning methods for large-scale EO data analysis and 2) developing augmentation-free self-supervised learning methods. Augmentation-free self-supervised learning has recently attracted significant attention in CV and ML communities [103] and can be highly relevant for different EO problems. Using first self-supervised learning and then fine-tuning approaches leverage unlabeled and labelled data and thus constitute a semi-supervised learning scheme. More generally, semi-supervised learning alternatively uses data with labels to learn a representation function to the target objective (e.g. classification) and data only to get the underlying data structure. Various strategies co-exist, including self-training [34, 76] on

---

unsupervised tasks (e.g. image reconstruction), consistency regularization and pseudo-labeling [73, 170], and generative modeling [35, 152]. These last research directions seem the most promising for addressing standard EO tasks at a large scale.

## 2.2 Learning from noisy labels for Earth observation

Another conceptually different approach to gathering representative training images covering wide areas is to exploit publicly available thematic products (e.g. the Corine Land Cover [CLC] map provided by the Copernicus Land Monitoring Service) as labeling sources. Since these products already exist, a large number of labels are already available at no cost. However, these products are subject to (i) errors in the map (depending on the data/algorithm used to generate them, e.g. the accuracy of the CLC map is reported to be 85% [88]); (ii) changes in land-use/land-cover after the derivation of the maps; and (iii) geolocation errors (e.g. due to the residual misalignment between a digital map and a satellite image) [134].

Missing and noisy labels can simultaneously appear across different spatial areas of a given image. Their direct use results in uncertainty in the ML models and, thus, uncertainty in the LULC maps/scene predictions. Therefore, there is a need for methods robust to label noise that: 1) identify and correct the label-noise associated with different thematic products and 2) are architecture-independent and can allow training of any DL model under input-dependent label noise.

To address these issues, several studies have been presented recently to demonstrate learning from noisy labels in RS [6, 46, 60, 79, 194]. Most existing methods in RS are sensitive to long-tailed class label distributions (a common issue observed in the analysis of large-scale EO data). They can provide inaccurate results, particularly for the minority classes in the training set, represented by insufficient reliable labelled samples. Accordingly, label noise-sensitive methods robust to imbalanced training sets are needed.

## 2.3 Automated machine learning for Earth observation

The quality of ML-derived products depends (i) on the choice of the model and training algorithm to be used and (ii) on the settings of various hyperparameters. A poorly chosen or ill-configured model can lead to mediocre performance, whereas a well-chosen, carefully tuned model can yield drastically better results and sometimes outperform human experts. As a consequence, there has been substantial work and progress by the automated machine learning (AutoML) community on model selection and, to an even larger degree, hyperparameter optimization [86], with methods including random search [19], Bayesian optimization [18, 59, 84, 169, 184], evolutionary optimization [114], meta-learning [24, 77, 83] and bandit-based methods [108].

While these methods are general enough to be applied out-of-the-box to ML problems encountered in EO, we believe that the Earth observation community would benefit from methods tailored to the specific kinds of data arising in EO problems. Several of those specifics are large amounts of observational

---

data, relatively few high-quality labels, and data showing substantial variation between different spatial regions. Therefore, to best support EO research, AutoML should go beyond the standard paradigm of selecting models and optimising their hyperparameters and address more complex topics, particularly in AutoML for spatio-temporal data. This includes, for example, investigations on which type of search spaces would lead to good results within the constraints of the application [85, 137, 195], or of the impact of acquiring more labelled data by modelling and analysing the learning curves of certain algorithms [125, 126].

Furthermore, we see neural architecture search (NAS) [57, 124] as a very promising direction of AutoML in the area of DL. NAS aims to automatically find the best DL model for a given task and dataset, e.g. for image classification, which can be done by using reinforcement learning [15, 224]. Current research directions focus on speeding up the search process, e.g. by optimisation in the search space using stochastic gradient descent [112] (including for remote sensing classification [136]) or direct NAS without retraining [78, 122].

To realise these and other advances in AutoML for EO, new benchmarks consisting of EO data should be created. While initiatives like BigEarthNet [177] can be seen as a first step towards this goal, additional and prominent benchmarks would motivate researchers from the AutoML community to invest in these types of data [133].

## 2.4 Domain adaptation for large-scale Earth observation data analysis

Domain adaptation is another approach to improve the statistic in the learning of a classifier. Domain adaptation, which can be seen as a particular case of transfer learning [141, 190], aims at classifying an image for which no prior information is available (target domain) by using existing information acquired from another image on the same area at a different time or from a spatially disjoint site (source domain).

It is worth noting that the pre-trained source models can be fine-tuned for the given target domain. However, fine-tuning still needs considerable amounts of labelled training data, which may not be available for typical EO applications. Through domain adaptation, the need and effort to recollect labelled samples are significantly reduced by adapting the already available information on the source domain to classify the target domain. To this end, the domain adaptation methods define strategies that adapt and use the information available on the source domain to classify the target domain, assuming that the two domains may have different (but related) distributions [89, 190]. In RS, the current research studies in the context of domain adaptation have recently shown the importance of developing methods that are capable of (i) efficiently and accurately learning domain-invariant data representations; (ii) generalising well to the different spatially disjoint large-scale areas; and (iii) mitigating uncertainties of the classification model associated to imbalanced and incomplete source domain training sets [141, 190]. These terms are crucial, as classifiers trained on one image seldom perform well on other images acquired at a different time or for other geographical areas under different conditions. To address these issues,

discrepancy-based methods [216, 217, 221], adversarial methods [110, 210, 218, 220], and self-supervised methods [39, 149, 165] have been recently demonstrated to be effective for different EO problems. Similarly, meta-learning approaches have also been developed [24, 77, 83]. Meta-learning takes a slightly different approach, where data from the source domain is utilized to determine a learning procedure that can learn more efficiently from limited amounts of available data, for example by determining good initialisation weights [63] or even by learning an optimisation algorithm [11, 82, 145]. A meta-learning approach to land-cover classification (cast as a few-shot learning problem) was proposed in [155], using model-agnostic meta-learning (MAML) [63]. Most existing domain adaptation methods mainly aim at adapting a model to a single target domain without considering a more practical scenario where the target consists of multiple data distributions. To investigate a more realistic setting for domain adaptation, an interesting approach is to study the problem of multiple-target domain adaptation and develop domain adaptation methods compatible with both single-target and multi-target domain adaptation settings [149]. We also note that, considering that the labelled source domains may contain multi-modal data (e.g. multispectral, SAR, etc.), the joint use of different EO data modalities can boost the performance of domain adaptation.

## 2.5 Continual learning from growing archives

Once an ML model is trained with the available training samples, it is fixed to be employed to realise the considered EO tasks. However, the repeat-pass nature of satellite orbits with high speed (i.e. high temporal resolution) results in fast and regular growth of the EO data archives. For example, through the Sentinel satellites, a volume of roughly 12TB of satellite images is acquired daily. Accordingly, continuously learning new information (e.g. new LULC classes, sudden/abrupt changes) from a stream of recently acquired EO data is essential. This requires updating the model parameters by considering the recently acquired training images. On the one hand, re-training the entire ML model is impractical for large-scale EO applications (due to constraints on computational and storage resources). On the other hand, updating the model parameters based on only recently available training samples may degrade the capability of the model for characterizing the previous data (this issue is known as catastrophic forgetting [209]). Open world (or open set) recognition [16, 67] represents the exemplar use-case: knowledge about the world is only given partially, and new classes can be presented at inference time. Thus, the methods (and related tools) that achieve learning over time in a streaming fashion (with changing data distributions) are needed. Few deep continual learning studies have recently been presented in RS [10, 105, 171, 185]. The primary goal of these methods is to overcome the forgetting of already learned LULC classes and to leverage the earlier knowledge for obtaining better performance or faster convergence/training speed on the new LULC classes (considered as new concepts). However, these methods require a set of accurate labels to describe the new classes, which may not be feasible in large-scale EO applications. In addition, these methods assume that the concept boundaries/identities are known. However, such a setup is unsuitable in operational EO applications since the concept boundaries/identities are not known a priori. For example, the LULC classes that appeared and

---

disappeared in different geographical areas may not be known a priori at a large scale. To address this problem, concept-agnostic learning (i.e. online learning), which has recently attracted attention in CV communities [215], can be highly relevant. Motivated by advances in reinforcement learning, an interesting approach is to investigate memory-based online continual learning methods. Another interesting direction is to research unsupervised domain adaptation methods in continual learning settings, aiming to update a model continually to learn distributional shifts across arriving concepts with unlabeled data while retaining the earlier knowledge.

### 3 Advanced processing and computing

The efficient processing and intelligent analysis of very large, complex and heterogeneous EO data are crucial to revolutionizing and substantially improving our capabilities for EO. In this perspective, the management of big EO data and ML is the scientific and technical pillar powering the current wave of innovation in AI for EO. This section describes the state-of-the-art and the challenges in the advanced processing and computing of EO data. It outlines the most promising research directions to address these challenges.

#### 3.1 Earth observation ecosystems – new system architectures

The growing operational capability of global EO provides scientists and practitioners with a wealth of information that can be used for various EO applications. However, to take advantage of this data, users need significant expertise to validate their assumptions and gain insights. Recently, different EO exploitation platforms — such as the Google Earth Engine, Open Data Cube, and Sentinel Hub – have emerged to ease large-scale analyses by offering a certain level of processing and data abstraction [175]. Yet, despite these benefits, all exploitation platforms rely on a heterogeneous set of technologies with varying sets of interfaces and data formats, making the federated use of these platforms difficult [202]. The openEO [164] project goes in the right direction, but it requires users to do all the binding of the different EO platforms. We believe that EO technology must be interoperable and accessible to everyone in a seamless manner. The AgoraEO project is exploring this direction toward a decentralized, open, and unified EO ecosystem, where one can share, find, compose, and efficiently execute cross-platform EO assets [202]. With such EO ecosystems, the project aims at fostering innovation, by putting EO technology at the disposal of all, and improving EO data literacy for the EO application and service developers, space agencies, the space industry, the science community and the general public. However, building EO ecosystems is still a big challenge. First, assets are highly heterogeneous, ranging from datasets (e.g. Sentinel data), processing tools (e.g. SNAP and GDAL), and ML algorithms to high-performance computing clusters and human expertise (e.g. an RS expert providing consultancy services). This high heterogeneity renders asset management (storing, querying, and composition) an open problem [187]. Second, most EO analytics are exploratory, which requires fast query results. Third, data assets are naturally scattered across multiple sites and platforms, making conventional (AI-based)

---

analytics hard. It is necessary to devise a framework to support federated analytics.

### 3.2 Federated learning

An increasing amount of EO applications operate in highly distributed settings, especially with the emergence of EO ecosystems. Nowadays, it is not rare at all to find EO data highly scattered across multiple sites. However, standard ML algorithms require centralising the data before training, which is problematic in the EO settings today because EO and other geo-referenced data are often large in volume. Additionally, centralising datasets might be impossible due to data constraints, such as privacy regulations or legal issues. Indeed, many public authorities and private industries remain reluctant to share their data with third parties for confidentiality, privacy, and legal reasons, but often also due to its economic value. Federated learning is a way to address this problem of associated governance, privacy concerns, and technical challenges [2, 99]. Federated learning aims to learn a local model on each distributed site and aggregate the resulting local models to a shared global model [121]. Federated learning reduces data transfer costs to a minimum, and models can generalise across EO datasets, leading to more accurate and less biased models. Applying federated learning to geo-distributed EO data is challenging because federated learning relies on the general assumption that the underlying data are independent and identically distributed (the i.i.d. assumption). This is rarely the case in real-world EO data scenarios. Furthermore, in RS, the data naturally tends to be remarkably heterogeneous – not only because of the variety of sensors, modalities, and characteristics of RS missions but even within the same product. Most current federated learning algorithms cannot effectively deal with these conditions, leading to a sub-optimal performance [219]. Moreover, independent third parties cannot build standard ML models while respecting data privacy constraints. This feature is crucial in EO ecosystems where participants do not know each other beforehand.

### 3.3 AI interoperability

The need for distributed or federated analytics and learning is increasing rapidly. An increasing amount of applications require building models in a distributed or federated fashion (i.e. over multiple data sources). Interoperability, in terms of data and model communication, across different areas to achieve a common goal, e.g. to build a global model, is critical in these applications. Without interoperability, achieving common goals efficiently and effectively is almost impossible. Primarily, we expect interoperability as a de facto standard in EO ecosystems (environments), where multiple participants can join to contribute with their assets. This calls for standards representing input data (e.g. satellite imagery) and intermediate data (e.g. local models) to optimise multiple data sources and information models. Unfortunately, most data representations and systems are not interoperable at all. Examples in this direction are the Hub project by ActiveLoop, which aims at providing an AI database, i.e. a common storage representation, for input data [3], the Open Neural Network eXchange (ONNX) format, an open-source standard for machine and deep learning [4], or OpenML [196], an eco-system for sharing machine learning datasets, algorithms and results. However, real interoperability is still



---

far from reality in EO. AI interoperability must be solved if we expect EO ecosystems to succeed and increase shortly.

### 3.4 Scalable data processing

RS is developing so fast that data is not only getting increasingly diverse but also the amount of available data is increasing at a very fast pace [17]. This poses many new challenges for scalable and efficient data processing to support emerging EO applications [9, 42]. We have to revisit the entire processing stack. First, we must redesign how data is ingested into data processing systems for analysis. Instead, we must be able to process the data in-situ with fast *virtual* data ingestion, i.e. without physically moving or copying the data into a processing system. Additionally, one could take inspiration from works in the data management community [7, 117] whose idea is to load or prepare only those portions of the data required by incoming queries. This allows the system to scatter the data loading cost across the processing of multiple incoming queries. Second, scanning the entire available datasets every time a query comes is impractical. There is a need to develop new indexing techniques that allow for fast querying of EO data. Recent works propose hashing-based indexing and approximate nearest neighbours search in extensive EO data archives [178]. Hashing methods encode high-dimensional image descriptors into a low-dimensional Hamming space. The image descriptors are represented by binary hash codes, indexed into a hash table that enables real-time search. Such binary hash codes can also significantly reduce the memory required for storing the RS images in the auxiliary archives. The success of deep neural networks in image feature learning has inspired research on developing DL-based hashing methods (i.e. deep hashing methods). Recently, several deep hashing-based indexing systems that simultaneously learn image representations and hash functions based on the suitable loss functions were introduced [151, 178]. Given the heterogeneity of data and processing in EO applications, one must be able to run different data analyses on a diverse set of processing hardware devices (e.g. CPUs, GPUs, FPGAs), depending on the input data and query. In particular, a significant challenge is efficiently processing EO data on the space segment (see Section 3.5).

### 3.5 On-board data processing

One of the most impelling goals in EO is to transfer satellite data processing from Earth (ground segment) to space (space segment) through the development of so-called onboard payload data processing. Indeed, quite often, the raw data generated by modern instruments is more than what can be transmitted to the ground. This makes using various signal processing and compression techniques necessary to reduce the amount of data transmitted, especially for small satellites. However, onboard processing requires several characteristics, such as flexibility, robustness and low computation burden. Moreover, the design of optimum schemes generally involves an appropriate combination of software and hardware solutions. For this reason, AI models represent a valid technology [223]. Indeed, in this case, the highest processing demand is required during the training phase, which is operated on the ground segment, while less effort is necessary to yield the onboard products. Hardware prototyping can be done in fast



---

loops, utilizing for example the neural compute stick. ESA has already launched a first demonstration mission named Phi-sat-1, characterized by onboard AI processing for cloud detection. More challenging AI-based applications are being developed for implementation on the payload of the Phi-sat-2 mission.

---

## Part II

# Understanding — Physics-machine learning interplay, causality and ontologies

## 4 Knowledge-based AI and Earth observation

AI is indeed much broader than just deep learning (and machine learning in general) and includes knowledge representation, reasoning (incl. constraint satisfaction) and planning (incl. scheduling). With the need for explanations and the rise of the topic of explainable AI, knowledge-based AI is again attracting a significant amount of research interest. This section discusses several issues of knowledge-based AI relevant to EO, including decision modelling and decision support, knowledge-based machine learning, and knowledge representation in ontologies for environmental sciences.

### 4.1 Decision modelling and decision support with Earth observation

A prototypical example of the knowledge-based approach to AI is expert systems, designed to mimic the behaviour of human experts in narrow areas of expertise [87]. Introduced in the mid-1960s, expert systems were among the first truly successful forms of AI software. They consist of a knowledge base (capturing human knowledge) and an inference engine (emulating human reasoning) and overall support the decision-making process.

The spirit of expert systems has survived in its purest form at the intersection of AI and decision support, and is exemplified by the DEX (Decision EXpert) approach [22]. DEX is a hierarchical multi-criteria decision modelling method, which represents decision preferences on alternatives described by qualitative variables in the form of decision rules, very similar to the IF-THEN rules of expert systems, and supports a variety of knowledge-explanation methods. DEX helps decision-makers in tackling difficult decisions by ranking the available options and supporting explanations (e.g. what-if questions), thus allowing humans to explore a given situation in depth and to make informed, well-justified decisions.

DEX has been successfully and extensively used in various areas, with increasing applications concerning agriculture and its environmental impacts. One example of such an application is SoilNavigator, a field-scale decision support system for assessing and managing soil functions [47]. It jointly considers multiple soil functions, including primary productivity, nutrient cycling, water purification and regulation, climate regulation and carbon sequestration, and finally, soil biodiversity and habitat provision. In addition, it provides management advice for improving the performance of prioritised soil functions.

Due to the lack of measured data on soil functions, the SoilNavigator decision support system assesses all of them in a knowledge-based manner. However, when combined with machine learning, EO can

---

provide more objective and precise information on at least some of the soil functions: for example, recent work by Wolanin et al. [203] has applied machine learning to EO data to estimate primary productivity.

The potential of EO data in decision-making is increasingly recognised, even if applications are still rare. For example, the OECD [1] has prepared a document entitled “EO for decision making”, which clearly states the importance of EO data for decision making, especially when combining it with other geospatial (e.g. ground data on air pollution), administrative and socio-economic data, in the context of assessing environmental risks and their impacts on humans and ecosystems. Combining knowledge-based approaches to decision support with EO data or downstream products derived by machine learning will unlock a whole new range of applications of AI that can help in achieving the sustainable development goals [199].

## 4.2 Knowledge-based machine learning and Earth observation

Most ML approaches today are data-driven and require large quantities of data to produce accurate and reliable models, which are typically not transparent and do not provide explanations. However, ML approaches that focus on cause and make use of formally represented domain knowledge, in addition to data, have a long tradition in the field, starting with explanation-based learning [123]. Inductive logic programming [102, 128] uses logic programs to represent data (examples), domain knowledge (background knowledge) and learned models (hypotheses). This approach has given rise to a broader learning paradigm known as relational learning or relational data mining [55]. Combining logical representations with probability theory has led to the development of approaches that learn probabilistic models expressed in logic within the paradigm of statistical relational learning [69] and, more broadly, statistical relational AI [143].

The knowledge used in ML, in addition to data, can take different forms. Taxonomies are a form of domain knowledge that represent relations among concepts in the domain of study: in predictive modelling and classification, such taxonomies can represent hierarchical relations among the classes predicted. Typically, examples in this context have multiple labels from the taxonomy and the task at hand is known as hierarchical multi-label classification [167, 198].

Land cover classification schemes [211], such as the Corine Land Cover map, are typically hierarchical. However, most ML approaches to land cover classification ignore this and learn models for each land cover class separately. Only recently has a multi-label classification of land cover [172] and classification with label hierarchies [193] begun to attract attention.

Other forms of knowledge that ML approaches can use include constraints, as considered in constraint-based data mining [54]. Rules can be hard or soft, and can (in predictive modelling) impose preferences on the predictive power, form (language constraints), or size of the learned models. Constraints on clustering can specify that specific data points should belong to the same cluster (or different ones) and indicate the number and size of groups.

---

Last but not least, explanation and the use of existing domain knowledge play a prominent role in computational scientific discovery [56], which includes the topic of automated modelling of dynamical systems in science and engineering. Approaches such as process-based modelling, exemplified by the tool ProBMoT [168], integrate data-driven and knowledge-driven approaches into modelling, where ontological domain knowledge in the form of templates can be used as model parts. Thus, understandable models can be obtained, which can not only simulate the behaviour of the modelled system, but also explain its structure in terms of the processes involved. Up-to-date land use data from EO would greatly help in the design of effective models.

### 4.3 Ontologies for open science in machine learning and Earth observation

Ontologies formally represent knowledge about a domain of interest, focusing on entities in the field and the relations between them. They mainly comprise classes (or concepts), properties (or attributes), instances (or class members) and relationships. Ontologies are crucial in providing controlled vocabularies for semantic annotation of datasets, which allows datasets to be FAIR, i.e. findable, accessible, interoperable and reusable [120]. With the rapidly increasing amounts of EO data, it must be FAIR according to the aforementioned criteria, to maximise its added value. More broadly, we should strive for open Earth science, where all products of the scientific process (e.g. data and models) are stored and available for reuse according to FAIR principles.

Currently, EO data are primarily managed through spatial data infrastructures, which facilitate the exchange, sharing, use and integration of geospatial data based on interoperability standards. Spatial data infrastructures follow the FAIR principles in some aspects (e.g. have registered or indexed meta-data in a searchable resource and are also accessible, even when the data are no longer available). However, they are not yet entirely FAIR-compliant [71], mainly due to the lack of ontologies for EO data that would provide vocabularies for meta-data that follow FAIR principles. Ontologies for describing EO data, such as RESEO (the REmote SEnsing Ontology), are only starting to appear [8]. In the near future, we expect significant further development of ontologies for EO as well as ontologies for ML and other areas of AI. These should be connected to commonly accepted upper-level ontologies, facilitating their interoperability. Linking the ontologies for EO and AI will significantly facilitate open science at the intersection of EO and AI.

## 5 Explainable AI and causal inference

Experimentation and observational data analysis provide the basis for the accurate modelling of physical phenomena. Nevertheless, model fitting is typically not enough, and one often wants to also understand and characterise the system's behaviour. Scientific consistency, reliability, and explainability of the results obtained from AI-based models are of paramount relevance when working with complex systems like the Earth or its climate systems [146, 153, 192]. A prerequisite is to design models that are robust and whose inner functioning can be visualised, queried, and interpreted. In this context, we

---

should aim for transparency, interpretability, and explainability of machine learning models for several reasons: 1) to ensure consistency, generalisability, and robustness of the model performance, 2) to better understand the system, and 3) to achieve wider adoption and confidence by domain scientists. Yet, model interpretability comes in many flavours and is complicated by the fact that the model under scrutiny already assumes a particular *causal* relation (inputs cause outputs), which is not necessarily correct or complete. Retrieving causal relations from observations has been the goal of humanity for centuries, and it is one of the key goals when using AI approaches, and notably machine learning techniques, for scientific enquiries.

## 5.1 Explainable AI

Many successful neural network models are so large and complex that it is challenging to understand or study their inner workings at a level of abstraction that gives insight into the predictions obtained from them. Many eXplainable AI (XAI) methods exist for various learning paradigms [150, 160]. These techniques range from *post-hoc interpretability*, via either advanced feature ranking techniques [91, 204] or analysis of the activation maps within the learned network (e.g. GradCam or the layer-wise relevance propagation method described below); to *interpretability by design*, i.e. networks that are learned to be explicitly explainable [119]. These techniques can derive spatially explicit and temporally resolved explanations of what the AI models have learned and what are the most expressive space and temporal scales [91].

“Post-hoc” methods include (but are not limited to) backpropagation-based or feature attribution techniques, such as layer-wise relevance propagation, DeepLift, and integrated gradients [150, 160]. Input feature ablation, occlusion, and perturbation-based approaches, such as model reliance and Shapley value sampling, also belong to this category. They all try to address the problem of meaningfully relating the output or the intermediate activations of a complex neural architecture to its input features. Attention mechanisms and transformers have recently shown great potential in Earth sciences [154] and provide ante-hoc explanations by assigning coefficients to the different spatio-temporal locations, which correspond to their importance of the latter when the model provides an output. These coefficients constitute saliency maps, which can be used for interpretation purposes. Unlike conventional recurrent neural networks, they allow introspection into the long-range interactions in space-time series.

“Explainable by design” methods are methods whose interpretability is built into the model from the start. Examples are models based on semantic bottlenecks, where intermediate processing pipelines (e.g. a classifier of concepts relevant to the downstream task) are learned parallel to the final task. In recent EO literature, such models have been proposed to understand reasons behind perceived subjective appreciation of landscapes (e.g. if a specific type of land use is connected to a more positive appreciation of a landscape [107]), aligning forests type classifiers by replicating the concepts behind the decision-making process employed by foresters [130] or the unsupervised exploration of what makes a landscape wild [173].

---

## 5.2 Causal inference

Despite the outstanding predictive capabilities of the current machine and deep learning methods, there is still little actual *learning*. Machine learning algorithms excel in fitting complex functions but do not produce clear understanding of the underlying causal relations [52, 135, 153]. Analysis of the Earth’s complex systems faces the problem of extracting information from multivariate and non-gridded datasets. Missing data, non-linearities, and non-stationarities are present in the wild. Variables and physical processes are often coupled in space and time, and (tele)connections can be long-range, discontinuous, and varying in strength and intensity. Addressing this problem can allow us to identify the right set of predictors, develop robust models and to avoid obtaining correct answers for the wrong reasons.

*Causal inference* aims to discover and explain the system’s causal structure based on data, models and assumptions [135, 139, 153]. Causal inference and discovery tackle fundamental problems in Earth sciences: 1) hypothesis testing, 2) discovery of latent factors and relations beyond spurious correlations, and 3) understanding the system from observational data and assumptions. True understanding is about knowledge of the entire causal chain; without it, we cannot predict the consequences of our actions (*interventions*) or analyse where things went wrong (*counterfactuals*).

**Interventional analysis.** So far, causal inference methods applied to observations have not produced scientific discoveries in climate science. We attribute this to the inherent complexity of the problem and the lack of controlled experiments (i.e. the lack of interventional data). To remedy this, one can perform interventions on the simulations. Incorporating interventions, if done carefully, will make the causal discovery process much more efficient and robust.

Working on interventions is typically a non-unique challenge. For example, intervening on global mean temperature could mean setting it to a constant value globally, but that would be a poor intervention not respecting the spatio-temporal trajectory of the system. Not all interventions make physical (and causal) sense, because some variables are impossible to intervene on in isolation. For instance, it is physically impossible to intervene in temperature without changing other climate variables. An intervention only makes sense if it leads to a true distributional shift in the system’s dynamics. We also note that interventions can occur at different levels: on exogenous variables, on observed trajectories, by removing interactions, or even in a latent space. This sort of intervention translates into a fantastic playground for doing science with data and simulations and answering *what if* questions.

**Counterfactuals analysis** In interventional analysis, one asks what will happen *on average* when acting on a particular variable. In counterfactual analysis, one tries to *imagine* what would have happened had a different course of action been taken in a specific situation. Attributing causes of extreme events to climate change is intrinsically a counterfactual thought experiment: would this event have happened in a world without humans? Counterfactual modules (e.g. relying on MACE, DACE, and NSGA-II algorithms) can help interpret the output of a model in human-like language. They could – and should – find application in the Earth and climate sciences. Using counterfactuals would have

---

an additional positive impact; the interaction between the model, the machine, and the user is made explicit, allowing the user to gain knowledge, while improving both the model of the environment and the machine.

**Causal discovery.** Interventions may be infeasible and unethical in the Earth sciences and only possible in the model world. In this case, *observational causal discovery* comes into play to extract cause-effect relationships from multivariate datasets, going beyond the commonly adopted correlation approach. Several approaches for causal discovery are available nowadays: probabilistic graphical models [98], structure learning methods [153], dynamical systems [52], Granger-causality [53, 72], instrumental variable paradigms and additive noise models [153]. There is a significant body of literature that, under some assumptions, allows us to learn causal relations even when time is not necessarily involved [139].

To tackle climate change, one needs to understand the complex phenomena occurring on the planet. Discovering teleconnection patterns is an essential part of this endeavour. Events such as El Niño-Southern Oscillation (ENSO) impact essential climate variables at large distances and influence the underlying Earth system dynamics. However, their automatic identification from the observational data is still unresolved. Nonlinearities, nonstationarities, and the (ab)use of correlation analyses hamper the discovery of true causal patterns. Time series feature extraction in combination with traditional Granger causality [72] or extensions to account for nonlinearities and nonstationarities [53] have given good results. However, extracting modes of variability and causal discovery are typically two independent steps which may give rise to false detections. Recently, deep learning has allowed the extraction of causal feature representations from spatio-temporal Earth data [197].

## 6 Physics-aware machine learning

Machine learning has been widely and openly criticised because of the difficulty to interpret models and the lack of physical consistency. Models are so big and overparameterised that accuracy compromises fulfilling physical laws, such as mass or energy conservation. Losing contact with physics should always be avoided if we want to design robust and reliable (machine learning) models. This disconnection has implications for the consistency of products and parameter estimation in data assimilation schemes and forecasting tools and in understanding the underlying processes from Earth’s observational data.

Historically, physical modelling and machine learning have been treated as two fields under very different scientific paradigms (theory-driven versus data-driven). However, integrating domain knowledge and physical consistency has been suggested as a principled way to provide solid theoretical constraints on top of the observational ones [146]. The synergy between the two approaches has gained attention by either redesigning model’s architecture, augmenting the training dataset with simulations, or including physical constraints in the cost function to be minimised [93, 146, 183]. Integrating physics in machine learning models may achieve improved performance and generalisation and, more importantly, may lead to improved consistency and credibility of the machine learning models. The hybridisation has

---

an interesting regularisation effect, given that the physics limits the parameter space to search and thus discards implausible models. Therefore, physics-aware machine learning models combat overfitting better, typically become simpler (sparser), and require less training data to achieve similar performance. Physics-aware ML thus leads to enhanced computational efficiency and constitutes a stepping stone towards achieving more interpretable and robust machine learning models [161, 200].

### 6.1 Inverting physical models with machine learning

Biogeophysical parameter estimation and retrieval with machine learning learn a mapping from an observed satellite spectrum to an underlying biophysical parameter. Unfortunately, relying only on observational data (both spectra and in-situ measurement) to perform regression is expensive, time-consuming, and very often infeasible [31, 33]. This problem can be better posed by exploiting the wealth of simulated data available through physical models such as radiative transfer models (RTMs) and electromagnetic models (EMs). They solve what is called the *forward problem*, that is, they generate physically meaningful observational data (e.g. spectra) from state parameters (e.g. leaf-level or canopy parameters). Once the simulations of a particular observed bio-geophysical scenario (e.g. a forest) are given, they provide the electromagnetic quantity that would be measured by the EO sensor (e.g. the back-scattering coefficients or reflectances). Using these models, we can generate as many input-output pairs as we need for training and testing our ML scheme to do the inversion, i.e. to estimate parameters from spectral data [48]. The use of an RTM/EM model can be beneficial from different points of view: (1) they can help data generation balancing by compensating for the lack of ground measurements; (2) they can help in understanding the bounds within which the desired parameters can be retrieved less ambiguously; and (3) they can help in the selection of the optimal inputs of the ML model.

### 6.2 Hybrid machine learning: interactions of transfer models and machine learning

ML algorithms and physics can be fully blended in several ways extending traditional approaches based on data assimilation. Still, simple constrained optimisation methods often work well in practice [94, 206]. Indeed, by inferring the constraints to the input/output pairs based on physical models, machine learning models can obtain improved performance. Another option is to learn machine learning emulators combined with purely data-driven algorithms for model inversion [29]. Finally, one can also consider a fully coupled neural network where layers that describe complicated and uncertain processes feed physics layers that encode known relations of intermediate feature representation with the target variables. The promise of such a hybrid modelling approach [30] is to combine the excellent predictive abilities of modern machine learning, fuelled by the “unreasonable effectiveness of data” [74] with the causal consistency, explainability, and data efficiency of classical physical modelling.

Physical models require setting parameters that are rarely derived from first-domain principles. ML models can learn such parameterisations. For example, instead of assigning vegetation parameters empirically to plant functional types in Earth system models, one can learn these parameterisations from



---

proxy covariates using ML, and thus achieve flexibility, adaptability, dynamics, and context-dependent properties [127].

Emulating models in geosciences, climate sciences, and remote sensing is gaining popularity [31, 33, 146]. Emulators are ML models that mimic the forward physical models using a small, yet representative data set of simulations. Once trained, emulators can provide fast-forward simulations, allowing improved model inversion and parametrisations. However, replacing physical models with machine learning models require running expensive offline evaluations first, and alternatives exist that construct the model and choose the proper simulations iteratively [29, 182].

### 6.3 Discovering equations from data

As in many fields of science and engineering, Earth system models describe processes with a set of differential equations encoding our prior belief about the dynamics and variable interactions. Learning ordinary and partial differential equations (ODE/PDEs) from stochastic observations is perhaps one of the most challenging problems nowadays in statistical learning. Several approaches exist for solving this problem, ranging from equation-free modelling [213] and empirical dynamic modelling [52, 176] to and automated inference of dynamics [45]. Imposing sparsity on equation discovery is a sensible assumption that allows us to identify the governing equations of the biosphere more sharply [5, 26]. Other criteria, such as compositionality, unification of theories, and interestingness, can help in learning laws and equations from data, while trying to alleviate the *equifinality* (non-identifiability) issue, that is, finding many equally accurate models for the wrong reasons. For a comprehensive overview of equation discovery techniques with case studies in Earth and climate sciences, we refer the interested reader to the work of Camps-Valls et al. [28].

---

## Part III

# Communicating — Machine-user interaction, trustworthiness & ethics

## 7 User-centric Earth observation

Making EO data accessible in catalogues is not enough. Improving users' experience when working with it must be high on the agenda too. Such experience is currently still not optimal, since EO data analysis and interpretation is still performed in a very laborious way, by repeated cycles of trial and error, without reaching the desired degree of flexibility and robustness. Furthermore, the overwhelming amount of data puts strong limitations on the extent (geographic coverage within acceptable cost bounds), depth (causal analysis, evaluation of the implications) and response time of human interpretation. As a consequence, a significant amount of EO data remains unfound and unused [62].

User-Centric EO is a concept to enable the interpretation EO data content, associate it with other sources of information, understand user inquiries in an interactive dialogue (thus refining the expression of needs), distil data content, and suggest the most appropriate algorithms, information items or the interpretation alternatives. Overall, it is a framework at the convergence of geo-information intelligence and human-machine interaction and communication.

### 7.1 Decreasing the user load with active learning

The first form of interaction is to be found in the support of human operators creating the information necessary for learning, i.e. the labels. Extracting labels directly from images, without guidance, is hugely time consuming and can be difficultly crowdsourced as it is done in other computer vision tasks. EO crowdsourcing may not be impossible, but it will not easily be implemented since agreement about class definitions is currently missing and class definitions are largely variable among communities, problems and even users [156]. The label collection process can be optimized via human-machine interaction with active learning [191]. The active learning process is conducted according to an iterative process. At each iteration, the learning algorithm automatically chooses the most informative unlabeled samples for manual labelling from a human expert (supervisor) and the supervised algorithm is retrained with the additional labelled samples. This way, the unnecessary and redundant labelling of samples that are not informative for the classifier is avoided, significantly reducing the labelling cost and time. Questions of sample informativeness (which are associated with uncertainty and diversity of samples) have been at the core of the debate. At the same time, equally important aspects such as accessibility costs [49] and user's skills in labelling have been so far relatively neglected [95, 189]. Also, the question of the applicability of active learning methods to the improvement of deep learning models is still

---

an open debate since single samples usefulness risks are diluted by the batch-learning nature of deep neural networks [106,148]. A comprehensive and systematic overview of deep active learning methods is provided in [147], showing promising avenues primarily on improving sample selection strategies, optimising training methods and improving task-independent models.

## 7.2 Language-based interaction

Besides label collection, another crucial form of interaction has the scope of making the content of EO images accessible to virtually anyone. To reach such a goal, one must work on the access interface to the information itself. The engineering and modelling phase is a barrier insurmountable for many users, who cannot exploit the potential of EO because of technical impediments. Creating access in a more inclusive way, e.g. through natural language, would open EO to a much broader set of users, including decision-makers, the press and virtually all areas of society. As for traditional search engines, this access can be granted by working on the interface between image processing (the content extraction process) and queries that would be asked in the simplest way possible, with natural language. Language opens EO access to society: methodologies ensuring this access via multimodal processing connecting natural language processing (NLP) models (understanding users' questions) and visual models (from computer vision, reviewed in the other sections) is critical. Visual question answering [12] is a set of methodologies aiming at creating such links and is one of the most exciting areas in User-centric EO. The first approaches are emerging, showing proofs of concept for Sentinel resolution and VHR resolution imaging [113] and change detection [214]. These first methods still have a large margin of progression, especially regarding the fusion strategies to mix language and images [36]; the way to generate, learn and exploit from more realistic language [109]; and finally the usage of large language models from NLP (BERT [51], GPT [25]) and the injection of EO context (or *prompting*) in these models [37]. All these research lines, together, will permit in the future to develop digital assistants able to understand real questions from users and retrieve the correct, action-ready information from EO. The emergence of EO chatbots, scanning EO data and retrieving information on demand or dialoguing with users (like ChatGPT does) is an exciting avenue for both research and the industry.

## 7.3 Improved exploration and storytelling

Besides answering user-specific questions, one could be interested in gathering general information about a scene for writing a journalistic story, exploring large stacks of data, or simply having a summary of the image content for storytelling. Such image transcription in the language is often referred to as image summarization and captioning. By summarizing the content of images in human-readable text, imaging captioning provides comprehensible information about EO images, making them understandable to everyone. In this context, recent research has created large datasets, developed practical algorithms, and extended potential applications for RS image captioning tasks [38,115,179]. Most existing remote sensing image captioning methods are based on encoder-decoder frameworks [142], which use a convolutional neural network or a recurrent neural network as a backbone. Such frameworks learn features from

---

the input image, and the decoder converts the encoded features into natural language descriptions. Attention mechanisms are also extensively explored in this task to adaptively attend to the relevant semantic region of each word [207]. Transformers models showed state-of-the-art results, as they can better model long-term dependencies and generate descriptions [166]. On the technical side, multi-task learning may further boost the performance of remote sensing imaging captioning by defining related auxiliary tasks such as multi-label classification and connecting to the visual question-answering task described in the last section. Looking ahead, promising directions are spatio-temporal reasoning for image captioning and its extension to retrieval/captioning systems focusing on changes and anomalies. The idea is to explore the long-range relations of semantics in space and time. With particular emphasis on this spatio-temporal context, higher-level information can be extracted in human-readable text, which leads to a better understanding of the scene.

## 7.4 Enhanced visualisation

The joint interpretation of big multimodal data, including EO data multi-mission images, meta-data, or text, is still seldom explored in research. This calls for new visualisation techniques able to represent data in a way that eases interpretation and explanations, but also the discovery of unexpected information, anomalies and causal relations (see Section 5). Succeeding in these new forms of exploration implies designing new visualisation techniques, enhancing scientific visualisation, information visualisation, visual analytics, human–computer interaction, augmented reality or virtual reality [157]. The EO data visualisation methods shall respond to users’ needs for interaction (Section 7.1), human-machine dialogues (Section 7.2) and exploration (Section 7.3) to support the decision process.

Early geographic EO visualisation attempts were ‘virtual flights’ on 3D terrains, rendered using 3D models from stereo or InSAR observations. Artefacts and noise removal, as well as multi-resolution integration, were the core issues. EO multi-, hyper-spectral or SAR image products could be visualised therein, with information extracted by dimensionality reduction or Fourier-based transformations for SAR complex-valued data and projected in the visual RGB representation [129].

Visual data mining can be enhanced by employing immersive visualization [14] of the EO information content with high user interaction. This can be achieved using immersive video environments like CAVE (Cave automatic virtual environment, a cube-sized room, where data are projected on 3 to 6 inner walls. The user is standing in its middle) or VR sets enriched with AI methods for multi-dimensional spaces adaptive projections. These techniques may become popular in EO due to the progress of virtual reality devices in the multimedia context.

The visualization methods and the analytics used interfaces shall be predictive, adaptive, learning and anticipating the user behaviour and collaborating with the user. This means understanding the user’s intentions and context and establishing a dialogue to transform non-visual sensor data and information into easily understandable human representations.

---

## 8 Earth observation and society: the growing relevance of ethics

In this section, we consider ethical issues linked to Artificial Intelligence, Earth observation and society under the light of AI and data ethics.

Ethical issues related to combined uses of AI for EO can emerge either in the context of AI ethics or data ethics. The principles linked to each of these overlap. However, it is helpful to understand them separately.

### 8.1 Trustworthy AI and AI ethics

According to the European Union’s high-level expert group on artificial intelligence, any AI system should be lawful, ethical and robust [58]. This is further specified into the following key requirements for achieving trustworthy AI, citing [58]: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability. Recently, Jobin et al. [90] reviewed 84 documents containing AI ethics guidelines issued by various countries and international agencies and categorized the most dominant/frequently re-appearing ones under 11 heads: (i) Transparency (ii) Justice and fairness; (iii) Non-maleficence (iv) Responsibility; (v) Privacy; (vi) Beneficence; (vii) Freedom and autonomy; (viii) Trust; (ix) Sustainability; (x) Dignity; and (xi) Solidarity. Notably, several of these categories overlap with the principles of trustworthy AI compiled in the EU ethics guidelines for trustworthy AI [58].

Even when neatly categorized above, ethics guidelines may appear vague or too high-level to EO scientists. At the most superficial level, ethics is about minimizing harm and maximizing benefits using appropriate (fair, honest) means. Accordingly, dominant Western approaches to ethics consider the duties of the actor and the consequences of the action/inaction (deontological and consequentialist approaches to ethics) to determine their acceptability.

Yet, these approaches may be impractical in emerging technologies such as AI4EO; scientists may not be able to predict the future consequences of their research in the present and may not be fully aware of their duties. Recent work recommends combining approaches that also consider the intentions/desires (of the individual researcher and the research project/program as a whole) driving the research and evaluating whether each action or research step corresponds with ethically motivated intentions or desires [96].

Consider the ethical principle of justice and fairness, which requires, for example, that appropriate measures be taken to ensure consistency and accuracy when applying sensitive data labels, such as slums. Even when accurate, such labels can lead to stigmatization or even to uprooting people from their living spaces (see [97]). Accordingly, ethics also calls for more constructive labeling, e.g. by re-thinking research designs that are more likely to go beyond labeling and help uplift or improve the socio-economic conditions of marginalized segments. Therefore, the principle of fairness can be applied

---

at the time of labeling data and even earlier in the research life-cycle, i.e. when selecting research focus areas and research questions. However, ethics is equally essential at other stages of research - from allocating research funding to converting research findings to practical applications.

Similarly, the issue of responsibility, in cases where human-in-the-loop features are included in AI designs, needs careful consideration. While active involvement of human actors (diverse stakeholders) in every aspect of AI4EO research may not be possible, human agency and oversight must not be limited within the confines of the system and its narrow research question. Such oversight is also needed to understand the context in which the designed system will be implemented or introduced and the impact that such an introduction may have on real people, their lives, livelihoods, cultural sensitivities, and their environment.

If the human agency and oversight requirement does not evolve considerably in the AI4EO research context while looking at the Earth and its physical elements from afar, complex human realities, relationships and expediencies can be overlooked or ignored, leading to significant ethically undesirable consequences. Ethics, therefore, calls for the humanization of projects and supplementing any simulations and recommendations emerging from there with multi-disciplinary social science and ground realities research. Here, ground realities are not limited to ground truth or in situ data as understood in conventional EO language but also include socio-cultural, economic and political realities of the region that can make the consequences of implementing recommendations emerging from AI4EO ethical or unethical.

## 8.2 Data ethics

Recently, the Data Ethics Commission (in German: ‘Datenethikkommission’), set up by Germany’s Federal Government on 18 July 2018, derived six additional general ethical and legal principles from 8 guiding motifs. These general ethical principles are: *human dignity, self-determination, security, democracy, justice and solidarity, and sustainability*. These principles signal the central relevance of human-centred and value-oriented technology design. However, the question of finding a balance between potentially divergent tenets requires case-by-case attention.

For instance, in the context of agriculture, the EU code of conduct for agricultural data sharing emphasizes that “*The farmer remains at the heart of the collection, processing and management of agricultural data*” [44]. The code recognizes that farmers who share data have the right to know and participate in the uses to which their data is put. The challenge in EO and AI4EO research is that such data is not collected or shared by the farmers. Indeed the data is remotely collected, often without the knowledge of farmers. There was little concern about ethical issues when such collections were limited to providing information. However, as the collection, analysis and distribution of such (EO) data start resulting in the formulation of policies that (adversely) impact individual farmers’ rights (e.g. their right to dignity or self-determination), the principles of the EU code of conduct will become increasingly relevant to ensure society’s participation in and eventual acceptance of the resulting policies.

---

## Conclusions

The field of AI for EO, and specifically, the use of ML methods in EO and remote sensing, is rapidly advancing, due to the increasing availability of large amounts of sensor data at various resolutions. In this perspective paper, we have provided an overview of the essential scientific tools and approaches supporting the transition from raw EO data to usable information. Since its inception, computer science and signal processing have been of key importance in this field. Recently, the integration of machine learning, deep learning and computer vision techniques has opened up numerous avenues for further, substantial advances in EO.

Here, we have discussed current challenges that still prevent the field to reach full maturity and highlighted promising directions for future research leveraging approaches from computer vision, explainable AI, physics-aware models, user-centric approaches, and advanced processing and computing. Additionally, we have emphasised the importance of carefully considering ethical issues related to the massive use of ML technologies in EO, which is further heightened in the context of large generative ML models for EO. This overview aims to inspire the EO community to achieve necessary transformative advances in the field and unlock future technical achievements, which are necessary to fulfill the monitor capabilities of using Earth observation to take the pulse of Earth.

## References

1. Earth Observation for Decision Making. <http://www.oecd.org/environment/indicators-modelling-outlooks/earth-observation-for-decision-making.htm>. Accessed: 2022-03-22.
2. Federated Learning: Collaborative Machine Learning without Centralized Training Data. Google AI Blog: <https://ai.googleblog.com/2017/04/federated-learning-collaborative.html>. Accessed: 2022-02-28.
3. Hub: A common database for AI. <https://www.activeloop.ai/>. Accessed: 2022-02-28.
4. Open Neural Network Exchange. <https://onnx.ai/>. Accessed: 2022-02-28.
5. J. E. Adsuara, A. Pérez-Suay, J. Muñoz-Marí, A. Mateo-Sanchis, M. Piles, and G. Camps-Valls. Nonlinear Distribution Regression for Remote Sensing Applications. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12):10025–10035, 2019.
6. A. K. Aksoy, M. Ravanbakhsh, T. Kreuziger, and B. Demir. A Consensual Collaborative Learning Method for Remote Sensing Image Classification Under Noisy Multi-Labels. In *2021 IEEE International Conference on Image Processing, ICIP 2021*, pages 3842–3846. IEEE, 2021.



- 
7. I. Alagiannis, R. Borovica, M. Branco, S. Idreos, and A. Ailamaki. NoDB: efficient query execution on raw data files. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2012*, pages 241–252. ACM, 2012.
  8. J. F. Aldana-Martín, J. García-Nieto, M. del Mar Roldán-García, and J. F. Aldana-Montes. Semantic modelling of Earth Observation remote sensing. *Expert Systems with Applications*, 187:115838, 2022.
  9. M. Amani, A. Ghorbanian, S. A. Ahmadi, M. Kakooei, A. Moghimi, S. M. MirMazloumi, S. H. A. Moghaddam, S. Mahdavi, M. Ghahremanloo, S. Parsian, Q. Wu, and B. Brisco. Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:5326–5350, 2020.
  10. N. Ammour, Y. Bazi, H. Alhichri, and N. Alajlan. Continual Learning Approach for Remote Sensing Scene Classification. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.
  11. M. Andrychowicz, M. Denil, S. G. Colmenarejo, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, and N. de Freitas. Learning to learn by gradient descent by gradient descent. In *Advances in Neural Information Processing Systems 29*, pages 3988–3996. Curran Associates Inc., 2016.
  12. S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick, and D. Parikh. VQA: visual question answering. In *2015 IEEE International Conference on Computer Vision, ICCV 2015*, pages 2425–2433, 2015.
  13. K. Ayush, B. Uzket, C. Meng, K. Tanmay, M. Burke, D. Lobell, and S. Ermon. Geography-aware self-supervised learning. In *IEEE/CVF International Conference on Computer Vision, ICCV 2021*, 2021.
  14. M. Babaee, S. Tsoukalas, G. Rigoll, and M. Datcu. Immersive visualization of visual data using nonnegative matrix factorization. *Neurocomputing*, 173:245–255, 2016.
  15. B. Baker, O. Gupta, N. Naik, and R. Raskar. Designing Neural Network Architectures using Reinforcement Learning. In *5th International Conference on Learning Representations, ICLR 2017*. OpenReview.net, 2017.
  16. A. Bendale and T. E. Boulton. Towards Open World Recognition. In *2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015*, pages 1893–1902. IEEE Computer Society, 2015.
  17. J. A. Benediktsson and Z. Wu. Distributed Computing for Remotely Sensed Data Processing. *Proceedings of the IEEE*, 109(8):1278–1281, 2021.



- 
18. J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl. Algorithms for Hyper-Parameter Optimization. In *Advances in Neural Information Processing Systems 24*, pages 2546–2554. Curran Associates, Inc., 2011.
  19. J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13:281–305, 2012.
  20. H. Bischof, W. Schneider, and A. J. Pinz. Multispectral classification of Landsat-images using neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 30(3):482–490, 1992.
  21. M. Blaha, C. Vogel, A. Richard, J. D. Wegner, T. Pock, and K. Schindler. Large-scale semantic 3d reconstruction: an adaptive multi-resolution model for multi-class volumetric labeling. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, pages 3176–3184, 2016.
  22. M. Bohanec. DEX (Decision EXpert): A Qualitative Hierarchical Multi-criteria Method. In *Multiple Criteria Decision Making*, pages 39–78. Springer Singapore, 2022.
  23. S. Branson, J. D. Wegner, D. Hall, N. Lang, K. Schindler, and P. Perona. From Google Maps to a fine-grained catalog of street trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, 135:13–30, 2018.
  24. P. Brazdil, J. N. van Rijn, C. Soares, and J. Vanschoren. *Metalearning: Applications to Automated Machine Learning and Data Mining*. Springer, 2nd edition, 2022.
  25. T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems 33*, 2020.
  26. S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.
  27. L. Bruzzone and B. Demir. A Review of Modern Approaches to Classification of Remote Sensing Data. In I. Manakos and M. Braun, editors, *Land Use and Land Cover Mapping in Europe: Practices & Trends*. Springer Netherlands, 2014.
  28. G. Camps-Valls, A. Gerhardus, U. Ninad, G. Varando, G. Martius, E. Balaguer-Ballester, R. Vinuesa, E. Diaz, L. Zanna, and J. Runge. Discovering causal relations and equations from data. *Physics Reports*, 2023.
  29. G. Camps-Valls, L. Martino, D. H. Svendsen, M. Campos-Taberner, J. Muñoz-Marí, V. Laparra, D. Luengo, and F. J. García-Haro. Physics-aware Gaussian processes in remote sensing. *Applied Soft Computing*, 68:69–82, 2018.

- 
30. G. Camps-Valls, M. Reichstein, X. Zhu, and D. Tuia. Advancing Deep Learning for Earth Sciences: From Hybrid Modeling to Interpretability. In *International Geoscience and Remote Sensing Symposium*, 2020.
  31. G. Camps-Valls, D. Sejdinovic, J. Runge, and M. Reichstein. A perspective on Gaussian processes for Earth observation. *National Science Review*, 6(4):616–618, 2019.
  32. G. Camps-Valls, D. Tuia, X. X. Zhu, and M. Reichstein. *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science and Geosciences*. Wiley and Sons, 2021.
  33. G. Camps-Valls, J. Verrelst, J. Muñoz-Marí, V. Laparra, F. Mateo-Jiménez, and J. Gomez-Dans. A survey on Gaussian processes for Earth-observation data analysis: A comprehensive investigation. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):58–78, 2016.
  34. J. Castillo-Navarro, B. Le Saux, A. Boulch, N. Audebert, and S. Lefèvre. Semi-supervised semantic segmentation in Earth Observation: the MiniFrance suite, dataset analysis and multi-task network study. *Machine Learning*, 111(9):3125–3160, 2022.
  35. J. Castillo-Navarro, B. Le Saux, A. Boulch, and S. Lefèvre. Energy-Based Models in Earth Observation: From Generation to Semisupervised Learning. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022.
  36. C. Chappuis, S. Lobry, B. Kellenberger, B. Le Saux, and D. Tuia. How to find a good image-text embedding for remote sensing visual question answering? In *Proceedings of MACLEAN: MACHine Learning for EArth ObservatioN Workshop co-located with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2021)*, volume 3088 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2021.
  37. C. Chappuis, V. Zermatten, S. Lobry, B. Le Saux, and D. Tuia. Prompt-RSVQA: Prompting visual context to a language model for Remote Sensing Visual Question Answering. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPR 2022 Workshops*, 2022.
  38. R. Chavhan, B. Banerjee, X. X. Zhu, and S. Chaudhuri. A novel actor dual-critic model for remote sensing image captioning. In *25th International Conference on Pattern Recognition (ICPR 2020)*, pages 4918–4925. IEEE, 2020.
  39. J. Chen, G. Zhai, G. Chen, B. Fang, P. Zhou, and N. Yu. Unsupervised Domain Adaption for High-Resolution Coastal Land Cover Mapping with Category-Space Constrained Adversarial Network. *Remote Sensing*, 13(8):1493, 2021.
  40. W.-Y. Chen, Y.-C. Liu, Z. Kira, Y.-C. F. Wang, and J.-B. Huang. A Closer Look at Few-shot Classification. In *7th International Conference on Learning Representations, ICLR 2019*, 2019.

- 
41. X. Cheng, Y. Zhong, M. Harandi, Y. Dai, X. Chang, H. Li, T. Drummond, and Z. Ge. Hierarchical Neural Architecture Search for Deep Stereo Matching. In *Advances in Neural Information Processing Systems 33*, 2020.
  42. M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu. Big Data for Remote Sensing: Challenges and Opportunities. *Proceedings of the IEEE*, 104(11):2207–2219, 2016.
  43. F. Chollet. On the Measure of Intelligence. *arXiv.org preprint [arXiv:1911.01547](https://arxiv.org/abs/1911.01547)*, 2019.
  44. C. Copa, F. CEMA, C. Ceettar, E. Ecpa, and E. Fefac. EU code of conduct on agricultural data sharing by contractual agreement. [https://www.copa-cogeca.eu/img/user/files/EU%20CODE/EU\\_Code\\_2018\\_web\\_version.pdf](https://www.copa-cogeca.eu/img/user/files/EU%20CODE/EU_Code_2018_web_version.pdf), 2018.
  45. B. C. Daniels and I. Nemenman. Automated adaptive inference of phenomenological dynamical models. *Nature communications*, 6:8133, 2015.
  46. R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau. Weakly supervised change detection using guided anisotropic diffusion. *Machine Learning*, 2021.
  47. M. Debeljak, A. Trajanov, V. Kuzmanovski, J. Schröder, T. Sandén, H. Spiegel, D. P. Wall, M. V. de Broek, M. Rutgers, F. Bampa, R. E. Creamer, and C. B. Henriksen. A Field-Scale Decision Support System for Assessment and Management of Soil Functions. *Frontiers in Environmental Science*, 7, 2019.
  48. F. Del Frate and D. Solimini. On neural network algorithms for retrieving forest biomass from SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 42(1):24–34, 2004.
  49. B. Demir, L. Minello, and L. Bruzzone. Definition of effective training sets for supervised classification of remote sensing images by a novel cost-sensitive active learning method. *IEEE Transactions on Geoscience and Remote Sensing*, 52(2):1272–1284, 2013.
  50. D. Derksen and D. Izzo. Shadow Neural Radiance Fields for Multi-view Satellite Photogrammetry. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPR 2021 Workshops*, 2021.
  51. J. Devlin, M. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019*, pages 4171–4186, 2019.
  52. E. Diaz, J. Adsuaara, A. Moreno-Martinez, M. Piles, and G. Camps-Valls. Inferring causal relations from observational long-term carbon and water fluxes records. *Scientific Reports*, 12:1610, 2022.
  53. G. C.-V. Diego Bueso, Maria Piles. Explicit Granger Causality in Kernel Hilbert Spaces. *Physical Review E*, 102:062201, 2020.

- 
54. S. Džeroski, B. Goethals, and P. Panov, editors. *Inductive Databases and Constraint-Based Data Mining*. Springer New York, 2010.
  55. S. Džeroski and N. Lavrač. *Relational data mining*. Springer, 2001.
  56. S. Džeroski and L. Todorovski, editors. *Computational Discovery of Scientific Knowledge*. Springer Berlin Heidelberg, 2007.
  57. T. Elsken, J. H. Metzen, and F. Hutter. Neural Architecture Search: A Survey. *Journal of Machine Learning Research*, 20(55):1–21, 2019.
  58. European Commission. Ethics Guidelines for Trustworthy AI: High-Level Expert Group on Artificial Intelligence, 2018.
  59. S. Falkner, A. Klein, and F. Hutter. BOHB: Robust and Efficient Hyperparameter Optimization at Scale. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 1436–1445. PMLR, 2018.
  60. K. Fatras, B. B. Damodaran, S. Lobry, R. Flamary, D. Tuia, and N. Courty. Wasserstein adversarial regularization for learning with label noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7296–7306, 2021.
  61. L. Fei-Fei, R. Fergus, and P. Perona. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):594–611, 2006.
  62. L. Filchev, L. Pashova, V. Kolev, and S. Frye. Challenges and Solutions for Utilizing Earth Observations in the “Big Data” era. In *BigSkyEarth Conference*, 2018.
  63. C. Finn, P. Abbeel, and S. Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 1126–1135. PMLR, 2017.
  64. M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
  65. W. Förstner. A feature based correspondence algorithm for image matching. In *ISPRS Commission III Symposium*, pages 150–166, 1986.
  66. M. Gašparović, D. Medak, I. Pilaš, L. Jurjević, and I. Balenović. Fusion of Sentinel-2 and PlanetScope imagery for vegetation detection and monitoring. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(1):155–160, 2018.
  67. C. Geng, S. Huang, and S. Chen. Recent Advances in Open Set Recognition: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3614–3631, 2021.

- 
68. R. Gens. Remote sensing of Coastlines: Detection, Extraction and Monitoring. *International Journal of Remote Sensing*, 31(7):1819–1836, 2010.
  69. L. Getoor and B. Taskar. *Introduction to Statistical Relational Learning*. The MIT Press, 2007.
  70. S. Gidaris, P. Singh, and N. Komodakis. Unsupervised representation learning by predicting image rotations. In *6th International Conference on Learning Representations, ICLR 2018*, 2018.
  71. G. Giuliani, H. Cazeaux, P.-Y. Burgi, C. Poussin, J.-P. Richard, and B. Chatenoux. SwissEnvEO: A FAIR National Environmental Data Repository for Earth Observation Open Science. *Data Science Journal*, 20, 2021.
  72. C. Granger. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, 37:424–438, 1969.
  73. P. Gómez and G. Meoni. MSMatch: Semisupervised Multispectral Scene Classification With Few Labels. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:11643–11654, 2021.
  74. A. Halevy, P. Norvig, and F. Pereira. The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2):8–12, 2009.
  75. G. E. Hinton, T. J. Sejnowski, et al. *Unsupervised learning: foundations of neural computation*. MIT press, 1999.
  76. D. Hong, N. Yokoya, G. Xia, J. Chanussot, and X. Zhu. X-ModalNet: A semi-supervised deep cross-modal network for classification of remote sensing data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167:12–23, 2020.
  77. T. M. Hospedales, A. Antoniou, P. Micaelli, and A. J. Storkey. Meta-Learning in Neural Networks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5149–5169, 2022.
  78. S. Hu, S. Xie, H. Zheng, C. Liu, J. Shi, X. Liu, and D. Lin. DSNAS: Direct Neural Architecture Search Without Parameter Retraining. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pages 12081–12089. Computer Vision Foundation / IEEE, 2020.
  79. Y. Hua, S. Lobry, L. Mou, D. Tuia, and X. X. Zhu. Learning multi-label aerial image classification under label noise: A regularization approach using word embeddings. In *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2020*, pages 525–528. IEEE, 2020.
  80. Y. Hua, D. Marcos, L. Mou, X. X. Zhu, and D. Tuia. Semantic segmentation of remote sensing images with sparse annotations. *IEEE Geoscience and Remote Sensing Letters*, 2021.

- 
81. M. Huisman, A. Plaat, and J. N. van Rijn. A Preliminary Study on the Feature Representations of Transfer Learning and Gradient-Based Meta-Learning Techniques. In *5th Workshop on Meta-Learning at NeurIPS 2021*, 2022.
  82. M. Huisman, A. Plaat, and J. N. van Rijn. Stateless neural meta-learning using second-order gradients. *Machine Learning*, 111(9):3227–3244, 2022.
  83. M. Huisman, J. N. van Rijn, and A. Plaat. A survey of deep meta-learning. *Artificial Intelligence Review*, 54(6):4483–4541, 2021.
  84. F. Hutter, H. H. Hoos, and K. Leyton-Brown. Sequential Model-Based Optimization for General Algorithm Configuration. In *Learning and Intelligent Optimization - 5th International Conference, LION 5*, volume 6683 of *Lecture Notes in Computer Science*, pages 507–523. Springer, 2011.
  85. F. Hutter, H. H. Hoos, and K. Leyton-Brown. An efficient approach for assessing hyperparameter importance. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014*, pages 754–762, 2014.
  86. F. Hutter, L. Kotthoff, and J. Vanschoren. *Automated machine learning: methods, systems, challenges*. Springer, 2019.
  87. P. Jackson. *Introduction to Expert Systems*. Addison-Wesley, 3rd edition, 1999.
  88. G. Jaffrain, C. Sannier, A. Pennec, and H. Dufourmont. CORINE Land Cover 2012 - Final Validation Report. Technical report, European Environment Agency, 2017.
  89. S. Ji, D. Wang, and M. Luo. Generative Adversarial Network-Based Full-Space Domain Adaptation for Land Cover Classification From Multiple-Source Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(5):3816–3828, 2021.
  90. A. Jobin, M. Ienca, and E. Vayena. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1:389–399, 2019.
  91. J. E. Johnson, V. Laparra, A. Pérez-Suay, M. Mahecha, and G. Camps-Valls. Kernel methods and their derivatives: Concept and perspectives for the Earth system sciences. *PLOS One*, 2020.
  92. B. Jongman, J. Wagemaker, B. R. Romero, and E. C. De Perez. Early flood detection for rapid humanitarian response: harnessing near real-time satellite and Twitter signals. *ISPRS International Journal of Geo-Information*, 4(4):2246–2266, 2015.
  93. A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Transactions on Knowledge and Data Engineering*, 29(10):2318–2331, 2017.



- 
94. K. Kashinath, A. Albert, R. Wang, M. Mustafa, and R. Yu. Physics-informed Spatio-temporal Deep Learning Models. *Bulletin of the American Physical Society*, 64, 2019.
  95. B. Kellenberger, D. Marcos, S. Lobry, and D. Tuia. Half a percent of labels is enough: Efficient animal detection in UAV imagery using deep CNNs and active learning. In *Transactions on Geoscience and Remote Sensing*, volume 57, no. 12, pages 9524–9533. IEEE, 2019.
  96. M. Kochupillai. Outline of a Novel Approach for Identifying Ethical Issues in Early Stages of AI4EO Research. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 165–1168, 2021.
  97. M. Kochupillai, M. Kahl, M. Schmitt, H. Taubenböck, and X. X. Zhu. Earth Observation and Artificial Intelligence: Understanding emerging ethical issues and opportunities. *IEEE Geoscience and Remote Sensing Magazine*, 2022.
  98. D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
  99. J. Konečný, H. B. McMahan, F. X. Yu, P. Richtarik, A. T. Suresh, and D. Bacon. Federated Learning: Strategies for Improving Communication Efficiency. In *NIPS Workshop on Private Multi-Party Machine Learning*, 2016.
  100. S. Kunwar, H. Chen, M. Lin, H. Zhang, P. D’Angelo, D. Cerra, S. M. Azimi, M. Brown, G. Hager, N. Yokoya, R. Hänsch, and B. Le Saux. Large-Scale Semantic 3-D Reconstruction: Outcome of the 2019 IEEE GRSS Data Fusion Contest—Part A. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:922–935, 2021.
  101. N. Lang, K. Schindler, and J. D. Wegner. High carbon stock mapping at large scale with optical satellite imagery and spaceborne LIDAR. *arXiv preprint [arXiv:2107.07431](https://arxiv.org/abs/2107.07431)*, 2021.
  102. N. Lavrač and S. Džeroski. *Inductive logic programming. Techniques and applications*. Ellis Horwood series in artificial intelligence. Ellis Horwood, 1994.
  103. N. Lee, J. Lee, and C. Park. Augmentation-Free Self-Supervised Learning on Graphs. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022*, pages 7372–7380. AAAI Press, 2022.
  104. S. Lefèvre, D. Tuia, J. D. Wegner, T. Produit, and A. S. Nassar. Toward seamless multiview scene analysis from satellite to street level. *Proceedings of the IEEE*, 105(10):1884–1899, 2017.
  105. G. Lenczner, A. Chan-Hon-Tong, N. Luminari, and B. Le Saux. Weakly-supervised continual learning for class-incremental segmentation. *CoRR*, abs/2201.01029, 2022.
  106. G. Lenczner, A. Chan-Hon-Tong, N. Luminari, B. Le Saux, and G. Le Besnerais. Interactive Learning for Semantic Segmentation in Earth Observation. In *Proceedings of MACLEAN*:

107. A. Levering, D. Marcos, and D. Tuia. On the relation between landscape beauty and land cover: A case study in the UK at Sentinel-2 resolution with interpretable AI. *ISPRS journal of Photogrammetry and Remote Sensing*, 177:194–203, 2021.
108. L. Li, K. G. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar. Hyperband: Bandit-Based Configuration Evaluation for Hyperparameter Optimization. In *5th International Conference on Learning Representations, ICLR 2017*. OpenReview.net, 2017.
109. Y. Li, N. Duan, B. Zhou, X. Chu, W. Ouyang, X. Wang, and M. Zhou. Visual question generation as dual task of visual question answering. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2018*, pages 6116–6124, 2018.
110. Z. Li, M. Liu, Y. Chen, Y. Xu, W. Li, and Q. Du. Deep Cross-Domain Few-Shot Learning for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–18, 2022.
111. B. Liu, X. Yu, A. Yu, P. Zhang, G. Wan, and R. Wang. Deep few-shot learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(4):2290–2304, 2018.
112. H. Liu, K. Simonyan, and Y. Yang. DARTS: Differentiable Architecture Search. In *7th International Conference on Learning Representations, ICLR 2019*. OpenReview.net, 2019.
113. S. Lobry, D. Marcos, J. Murray, and D. Tuia. RSVQA: visual question answering for remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12):8555–8566, 2020.
114. I. Loshchilov and F. Hutter. CMA-ES for Hyperparameter Optimization of Deep Neural Networks. *ArXiv [cs.NE]*, 1604.07269v1, 2016.
115. X. Lu, B. Wang, X. Zheng, and X. Li. Exploring models and data for remote sensing image caption generation. *IEEE Transactions on Geoscience and Remote Sensing*, 56(4):2183–2195, 2017.
116. B. Lucas, C. Pelletier, D. Schmidt, G. I. Webb, and F. Petitjean. Unsupervised domain adaptation techniques for classification of satellite image time series. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1074–1077, 2020.
117. C. Lutz, S. Breß, S. Zeuch, T. Rabl, and V. Markl. Pump Up the Volume: Processing Large Data on GPUs with Fast Interconnects. In *Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020*, pages 1633–1649. ACM, 2020.



- 
118. O. Manas, A. Lacoste, X. Giró-i Nieto, D. Vazquez, and P. Rodriguez. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. In *IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2021*, pages 9414–9423, 2021.
  119. D. Marcos, S. Lobry, and D. Tuia. Semantically interpretable activation maps: What-where-how explanations within CNNs. In *IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2019*, pages 4207–4215, 2019.
  120. Mark D. Wilkinson. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3:160018, 2016.
  121. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017*, volume 54 of *Proceedings of Machine Learning Research*, pages 1273–1282. PMLR, 2017.
  122. J. Mellor, J. Turner, A. Storkey, and E. J. Crowley. Neural Architecture Search without Training. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021*, volume 139 of *Proceedings of Machine Learning Research*, pages 7588–7598. PMLR, 2021.
  123. T. M. Mitchell, R. M. Keller, and S. T. Kedar-Cabelli. Explanation-based generalization: A unifying view. *Machine Learning*, 1(1):47–80, 1986.
  124. R. Mohan, T. Elsken, A. Zela, J. H. Metzen, B. Staffler, T. Brox, A. Valada, and F. Hutter. Neural Architecture Search for Dense Prediction Tasks in Computer Vision. *International Journal of Computer Vision*, 2022.
  125. F. Mohr and J. N. van Rijn. Learning Curves for Decision Making in Supervised Machine Learning – A Survey. *CoRR*, abs/2201.12150, 2022.
  126. F. Mohr and J. N. van Rijn. Fast and informative model selection using learning curve cross-validation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
  127. Á. Moreno-Martínez, G. Camps-Valls, J. Kattge, N. Robinson, M. Reichstein, P. van Bodegom, K. Kramer, J. H. C. Cornelissen, P. Reich, M. Bahn, et al. A methodology to derive global maps of leaf traits using remote sensing and climate data. *Remote Sensing of Environment*, 218:69–88, 2018.
  128. S. Muggleton. *Inductive logic programming*. Academic Press, 1992.
  129. I. C. Neagoe, M. Coca, C. Vaduva, and M. Datcu. Cross-Bands Information Transfer to Offset Ambiguities and Atmospheric Phenomena for Multispectral Data Visualization. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:11297–11310, 2021.
  130. T.-A. Nguyen, B. Kellenberger, and D. Tuia. Mapping forest in the Swiss Alps treeline ecotone with explainable deep learning. *Remote Sensing of Environment*, 281:113217, 2022.

- 
131. D. Oberhauser. Blockchain for environmental governance: Can smart contracts reinforce payments for ecosystem services in Namibia. *Frontiers in Blockchain*, 2, 2019.
  132. P. C. Oddo and J. D. Bolton. The Value of Near Real-Time Earth Observations for Improved Flood Disaster Response. *Frontiers in Environmental Science*, 7:2–3, 2019.
  133. N. R. Palacios Salinas, M. Baratchi, J. N. van Rijn, and A. Vollrath. Automated Machine Learning for Satellite Data: Integrating Remote Sensing Pre-trained Models into AutoML Systems. In *Machine Learning and Knowledge Discovery in Databases*, volume 12979 of *Lecture Notes in Computer Science*, pages 447–462. Springer, 2021.
  134. C. Paris and L. Bruzzone. A Novel Approach to the Unsupervised Extraction of Reliable Training Samples From Thematic Products. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3):1930–1948, 2021.
  135. J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2nd edition, 2009.
  136. C. Peng, Y. Li, L. Jiao, and R. Shang. Efficient Convolutional Neural Architecture Search for Remote Sensing Image Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7):6092–6105, 2021.
  137. V. Perrone, H. Shen, M. W. Seeger, C. Archambeau, and R. Jenatton. Learning search spaces for Bayesian optimization: Another view of hyperparameter transfer learning. In *Advances in Neural Information Processing Systems 32*, pages 12751–12761, 2019.
  138. C. Persello, J. D. Wegner, R. Haensch, D. Tuia, P. Ghamisi, M. Koeva, and G. Camps-Valls. Deep Learning and Earth Observation to Support the Sustainable Development Goals: Current approaches, open challenges, and future opportunities. *IEEE Geoscience and Remote Sensing Magazine*, 10(2), 2022.
  139. J. Peters, D. Janzing, and B. Schölkopf. *Elements of Causal Inference - Foundations and Learning Algorithms*. Adaptive Computation and Machine Learning Series. MIT Press, 2017.
  140. P. Potapov, X. Li, A. Hernandez-Serna, A. Tyukavina, M. C. Hansen, A. Kommareddy, A. Pickens, S. Turubanova, H. Tang, C. E. Silva, et al. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, 253:112165, 2021.
  141. R. Qin and T. Liu. A Review of Landcover Classification with Very-High Resolution Remotely Sensed Optical Images, Analysis Unit, Model Scalability and Transferability. *Remote Sensing*, 14(3), 2022.
  142. B. Qu, X. Li, D. Tao, and X. Lu. Deep semantic understanding of high resolution remote sensing image. In *2016 International conference on computer, information and telecommunication systems (Cits)*, pages 1–5. IEEE, 2016.

- 
143. L. D. Raedt, K. Kersting, S. Natarajan, and D. Poole. Statistical Relational Artificial Intelligence: Logic, Probability, and Computation. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 10(2):1–189, 2016.
  144. R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2020.
  145. S. Ravi and H. Larochelle. Optimization as a Model for Few-Shot Learning. In *5th International Conference on Learning Representations, ICLR 2017*, 2017.
  146. M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat. Deep learning and process understanding for data-driven Earth System Science. *Nature*, 566:195–204, 2019.
  147. P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, and X. Wang. A Survey of Deep Active Learning. *ACM Computing Surveys*, 54(9), 2021.
  148. A. C. Rodríguez, S. D’Aronco, K. Schindler, and J. D. Wegner. Mapping oil palm density at country scale: An active learning approach. *Remote Sensing of Environment*, 261:112479, 2021.
  149. E. Rolf, J. Proctor, T. Carleton, I. Bolliger, V. Shankar, M. Ishihara, B. Recht, and S. Hsiang. A generalizable and accessible approach to machine learning with global satellite imagery. *Nature communications*, 12(1):4392, 2021.
  150. R. Roscher, B. Bohn, M. F. Duarte, and J. Garcke. Explainable machine learning for scientific insights and discoveries. *IEEE Access*, 8:42200–42216, 2020.
  151. S. Roy, E. Sangineto, B. Demir, and N. Sebe. Metric-Learning-Based Deep Hashing Network for Content-Based Retrieval of Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*, 18(2):226–230, 2021.
  152. S. Roy, E. Sangineto, N. Sebe, and B. Demir. Semantic-Fusion Gans for Semi-Supervised Satellite Image Classification. In *2018 IEEE International Conference on Image Processing, ICIP 2018*, pages 684–688. IEEE, 2018.
  153. J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Clymour, M. Kretschmer, M. Mahecha, J. Muñoz-Marí, E. van Nes, J. Peters, R. Quax, M. Reichstein, M. Scheffer, B. Schölkopf, P. Spirtes, G. Sugihara, J. Sun, K. Zhang, and J. Zscheischler. Inferring causation from time series with perspectives in Earth system sciences. *Nature Communications*, 10(2553), 2019.
  154. M. Rußwurm and M. Körner. Self-attention for raw optical satellite time series classification. *ISPRS journal of photogrammetry and remote sensing*, 169:421–435, 2020.

- 
155. M. Russwurm, S. Wang, M. Korner, and D. Lobell. Meta-Learning for Few-Shot Land Cover Classification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020 Workshops*, 2020.
  156. M. Rußwurm, S. Wang, and D. Tuia. Humans are Poor Few-Shot Classifiers for Sentinel-2 Land Cover. In *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2022*, pages 4859–4862. IEEE, 2022.
  157. D. Sacha, M. Sedlmair, L. Zhang, J. A. Lee, D. Weiskopf, S. North, and D. Keim. Human-centered machine learning through interactive visualization. *Neurocomputing*, 268:164–175, 2017.
  158. Y. Sadeh, X. Zhu, D. Dunkerley, J. P. Walker, Y. Zhang, O. Rozenstein, V. Manivasagam, and K. Chenu. Fusion of Sentinel-2 and PlanetScope time-series data into daily 3m surface reflectance and wheat LAI monitoring. *International Journal of Applied Earth Observation and Geoinformation*, 96:102260, 2021.
  159. V. Sainte Fare Garnot, L. Landrieu, and N. Chehata. Multi-modal temporal attention models for crop mapping from satellite time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 187:294–305, 2022.
  160. W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K.-R. Müller. Toward Interpretable Machine Learning: Transparent Deep Neural Networks and Beyond. [arXiv:2003.07631](https://arxiv.org/abs/2003.07631), 2020.
  161. W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K.-R. Müller. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, volume 11700 of *Lecture Notes in Computer Science*. Springer, 2019.
  162. L. Scheibenreif, M. Mommert, and D. Borth. Contrastive self-supervised data fusion for satellite imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:705–711, 2022.
  163. M. Schmitt and X. X. Zhu. Data fusion and remote sensing: An ever-growing relationship. *IEEE Geoscience and Remote Sensing Magazine*, 4(4):6–23, 2016.
  164. M. Schramm, E. Pebesma, M. Milenković, L. Foresta, J. Dries, A. Jacob, W. Wagner, M. Mohr, M. Neteler, M. Kadunc, et al. The openEO API—Harmonising the Use of Earth Observation Cloud Services Using Virtual Data Cube Functionalities. *Remote Sensing*, 13(6):1125, 2021.
  165. W. Shen, Q. Wang, H. Jiang, S. Li, and J. Yin. Unsupervised Domain Adaptation for Semantic Segmentation via Self-Supervision. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 2747–2750, 2021.
  166. X. Shen, B. Liu, Y. Zhou, and J. Zhao. Remote sensing image caption generation via transformer and reinforcement learning. *Multimedia Tools and Applications*, 79(35):26661–26682, 2020.

- 
167. C. N. Silla and A. A. Freitas. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, 22(1-2):31–72, 2010.
  168. N. Simidjievski, L. Todorovski, J. Kocijan, and S. Dzeroski. Equation Discovery for Nonlinear System Identification. *IEEE Access*, 8:29930–29943, 2020.
  169. J. Snoek, H. Larochelle, and R. P. Adams. Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems 25*, pages 2951–2959. Curran Associates, Inc., 2012.
  170. K. Sohn, D. Berthelot, C.-L. Li, Z. Zhang, N. Carlini, E. D. Cubuk, A. Kurakin, H. Zhang, and C. Raffel. FixMatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems 33*, 2020.
  171. S. Srivastava, M. Berman, M. Blaschko, and D. Tuia. Adaptive Compression-based Lifelong Learning. In *British Machine Vision Conference (BMVC)*, 2019.
  172. R. Stivaktakis, G. Tsagkatakis, and P. Tsakalides. Deep Learning for Multilabel Land Cover Scene Categorization Using Data Augmentation. *IEEE Geoscience and Remote Sensing Letters*, 16(7):1031–1035, 2019.
  173. T. Stomberg, I. Weber, M. Schmitt, and R. Roscher. Jungle-net: Using explainable machine learning to gain new insights into the appearance of wilderness in satellite imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:317–324, 2021.
  174. C. Stucker and K. Schindler. ResDepth: A Deep Residual Prior For 3D Reconstruction From High-resolution Satellite Images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021.
  175. M. Sudmanns, D. Tiede, S. Lang, H. Bergstedt, G. Trost, H. Augustin, A. Baraldi, and T. Blaschke. Big Earth data: disruptive changes in Earth observation data management and analysis? *International Journal of Digital Earth*, 13(7):832–850, 2020.
  176. G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch. Detecting causality in complex ecosystems. *Science*, 338(6106):496–500, 2012.
  177. G. Sumbul, M. Charfuelan, B. Demir, and V. Markl. Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding. In *2019 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2019*, pages 5901–5904. IEEE, 2019.
  178. G. Sumbul, J. Kang, and B. Demir. Deep Learning for Image Search and Retrieval in Large Remote Sensing Archives. In *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science and Geosciences*, chapter 11, pages 150–160. John Wiley & Sons, 2021.

- 
179. G. Sumbul, S. Nayak, and B. Demir. SD-RSIC: Summarization-driven deep remote sensing image captioning. *IEEE Transactions on Geoscience and Remote Sensing*, 59(8):6922–6934, 2020.
  180. B. Sun and K. Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *Computer Vision - ECCV 2016 Workshops*, volume 9915 of *Lecture Notes in Computer Science*, pages 443–450. Springer, 2016.
  181. F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales. Learning to compare: Relation network for few-shot learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2018*, 2018.
  182. D. Svendsen, L. Martino, and G. Camps-Valls. Active Emulation of Computer Codes with Gaussian Processes – Application to Remote Sensing. *Pattern Recognition*, 100(107103):1–12, 2020.
  183. D. H. Svendsen, L. Martino, M. Campos-Taberner, F. J. García-Haro, and G. Camps-Valls. Joint Gaussian processes for biophysical parameter retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, 56(3):1718–1727, 2017.
  184. K. Swersky, J. Snoek, and R. Adams. Multi-task Bayesian optimization. In *Advances in Neural Information Processing Systems 26*, pages 2004–2012, 2013.
  185. O. Tasar, Y. Tarabalka, and P. Alliez. Continual Learning for Dense Labeling of Satellite Images. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 4943–4946, 2019.
  186. M. Tom, Y. Jiang, E. Baltsavias, and K. Schindler. Learning a Joint Embedding of Multiple Satellite Sensors: A Case Study for Lake Ice Monitoring. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.
  187. J. Traub, Z. Kaoudi, J.-A. Quiané-Ruiz, and V. Markl. Agora: Bringing Together Datasets, Algorithms, Models and More in a Unified Ecosystem [Vision]. *SIGMOD Record*, 49(4):6–11, 2021.
  188. D. Tuia, D. Marcos, and G. Camps-Valls. Multi-temporal and multi-source remote sensing image classification by nonlinear relative normalization. *ISPRS Journal of Photogrammetry and Remote Sensing*, 120:1–12, 2016.
  189. D. Tuia and J. Muñoz-Marí. Learning user’s confidence for active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2):872–880, 2013.
  190. D. Tuia, C. Persello, and L. Bruzzone. Domain Adaptation for the Classification of Remote Sensing Data: An Overview of Recent Advances. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):41–57, 2016.

- 
191. D. Tuia, F. Ratle, F. Pacifici, M. Kanevski, and W. J. Emery. Active learning methods for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7):2218–2232, 2009.
  192. D. Tuia, R. Roscher, J. D. Wegner, N. Jacobs, X. X. Zhu, and G. Camps-Valls. Towards a Collective Agenda on AI for Earth Science Data Analysis. *IEEE Geoscience and Remote Sensing Magazine*, 9(2):88–104, 2021.
  193. M. O. Turkoglu, S. D’Aronco, G. Perich, F. Liebisch, C. Streit, K. Schindler, and J. D. Wegner. Crop mapping from image time series: deep learning with multi-scale label hierarchies. *Remote Sensing of Environment*, 2021.
  194. P. Ulmas and I. Liiv. Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification. *CoRR*, abs/2003.02899, 2020.
  195. J. N. van Rijn and F. Hutter. Hyperparameter Importance Across Datasets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018*, pages 2367–2376. ACM, 2018.
  196. J. Vanschoren, J. N. van Rijn, B. Bischl, and L. Torgo. OpenML: networked science in machine learning. *SIGKDD Explorations*, 15:49–60, 2014.
  197. G. Varando, M.-A. Fernández-Torres, and G. Camps-Valls. Learning Granger Causal Feature Representations. In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*, 2021.
  198. C. Vens, J. Struyf, L. Schietgat, S. Džeroski, and H. Blockeel. Decision trees for hierarchical multi-label classification. *Machine Learning*, 73(2):185–214, 2008.
  199. R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S. D. Langhans, M. Tegmark, and F. F. Nerini. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11(1), 2020.
  200. L. Von Rueden, S. Mayer, J. Garcke, C. Bauckhage, and J. Schuecker. Informed machine learning—towards a taxonomy of explicit integration of knowledge into machine learning. *Learning*, 18:19–20, 2019.
  201. F. Waldner, Y. Chen, R. Lawes, and Z. Hochman. Needle in a haystack: Mapping rare and infrequent crops using satellite imagery and data balancing methods. *Remote Sensing of Environment*, 233:111375, 2019.
  202. A. Wall, B. Deiseroth, E. Tzirita Zacharatou, J.-A. Quiané-Ruiz, B. Demir, and V. Markl. Agora-EO: A Unified Ecosystem For Earth Observation -A Vision For Boosting EO Data Literacy. *Big Data from Space Conference*, 2021.



- 
203. A. Wolanin, G. Camps-Valls, L. Gómez-Chova, G. Mateo-García, C. van der Tol, Y. Zhang, and L. Guanter. Estimating crop primary productivity with Sentinel-2 and Landsat 8 using machine learning methods trained with radiative transfer simulations. *Remote Sensing of Environment*, 225:441–457, 2019.
204. A. Wolanin, G. Mateo-Garcia, G. Camps-Valls, L. Gomez-Chova, M. Meroni, G. Duveiller, Y. Liangzhi, and L. Guanter. Estimating and Understanding Crop Yields with Explainable Deep Learning in the Indian Wheat Belt. *Environmental Research Letters*, 15(2):1–12, 2020.
205. S. Workman, M. Zhai, D. J. Crandall, and N. Jacobs. A unified model for near and remote sensing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2688–2697, 2017.
206. J. Wu, K. Kashinath, A. Albert, M. Prabhat, and H. Xiao. Physics-Informed Generative Learning to Emulate Unresolved Physics in Climate Models. In *AGU Fall Meeting*, 2018.
207. S. Wu, X. Zhang, X. Wang, C. Li, and L. Jiao. Scene attention mechanism for remote sensing image caption generation. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2020.
208. Y. Xian, B. Schiele, and Z. Akata. Zero-shot learning – the good, the bad and the ugly. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pages 3077–3086. IEEE Computer Society, 2017.
209. J. Xu and Z. Zhu. Reinforced Continual Learning. In *Advances in Neural Information Processing Systems 31*, pages 907–916. Curran Associates Inc., 2018.
210. Q. Xu, X. Yuan, and C. Ouyang. Class-Aware Domain Adaptation for Semantic Segmentation of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022.
211. H. Yang, S. Li, J. Chen, X. Zhang, and S. Xu. The Standardization and Harmonization of Land Cover Classification Systems towards Harmonized Datasets: A Review. *ISPRS International Journal of Geo-Information*, 6(5), 2017.
212. J. Yang, P. Gong, R. Fu, M. Zhang, J. Chen, S. Liang, B. Xu, J. Shi, and R. Dickinson. The Role of Satellite remote sensing in climate change studies. *Nature climate change*, 3(10):875–883, 2013.
213. H. Ye, R. J. Beamish, S. M. Glaser, S. C. Grant, C.-h. Hsieh, L. J. Richards, J. T. Schnute, and G. Sugihara. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proceedings of the National Academy of Sciences*, 112(13):E1569–e1576, 2015.
214. Z. Yuan, L. Mou, Z. Xiong, and X. Zhu. Change Detection Meets Visual Question Answering. *arXiv preprint [arXiv:2112.06343](https://arxiv.org/abs/2112.06343)*, 2021.



- 
215. C. Zeno, I. Golan, E. Hoffer, and D. Soudry. Task-Agnostic Continual Learning Using Online Variational Bayes with Fixed-Point Updates. *Neural Computation*, pages 1–39, 2021.
  216. B. Zhang, T. Chen, and B. Wang. Curriculum-Style Local-to-Global Adaptation for Cross-Domain Remote Sensing Image Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12, 2022.
  217. J. Zhang, J. Liu, B. Pan, and Z. Shi. Domain Adaptation Based on Correlation Subspace Dynamic Distribution Alignment for Remote Sensing Image Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(11):7920–7930, 2020.
  218. X. Zhang, X. Yao, X. Feng, G. Cheng, and J. Han. DFENet for Domain Adaptation-Based Remote Sensing Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022.
  219. Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra. Federated Learning with Non-IID Data. *arXiv:1806.00582 [cs.LG]*, 2018.
  220. A. Zheng, M. Wang, C. Li, J. Tang, and B. Luo. Entropy Guided Adversarial Domain Adaptation for Aerial Image Semantic Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.
  221. S. Zhu, B. Du, L. Zhang, and X. Li. Attention-Based Multiscale Residual Adaptation Network for Cross-Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.
  222. X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, 2017.
  223. M. Ziaja, P. Bosowski, M. Myller, G. Gajoch, M. Gumiela, J. Protich, K. Borda, D. Jayaraman, R. Dividino, and J. Nalepa. Benchmarking Deep Learning for On-Board Space Applications. *Remote Sensing*, 13(19), 2021.
  224. B. Zoph and Q. Le. Neural Architecture Search with Reinforcement Learning. In *5th International Conference on Learning Representations, ICLR 2017*. OpenReview.net, 2017.

---

## Contributions statement

- DT coordinated the writing group.
- Section contributions (in **bold**, lead writer): Introduction: **DT**; Computer vision: **KS**, DT; Machine learning: **BD**, JvR, HH; Advanced processing and computing: **BD**, JQR, VM, BLS, RS, FDF; Knowledge-based AI: **SD**, JvR, HH; Explainable AI and causality: **GCV**; Physics-aware ML: **FDF**, GCV, KS; User-centric: **DT**, XZ, MD; Ethics: **MK**.

ESA has partially funded the authors to work on this manuscript through the ESA visiting professors program.