UK Parliament
POST

# POSTNOTE

Number 628 June 2020

# Remote sensing and machine learning


There is increasing interest in using machine learning to automatically analyse remote sensing data and increase our understanding of complex environmental systems. While there are benefits from this approach, there are also some barriers to its use. This POSTnote examines the value of these approaches, and the technical and ethical challenges for wider implementation.

## Overview

- Environmental remote sensing involves the use of satellites and other airborne instruments to collect data about the environment.
- Machine learning is a branch of artificial intelligence (AI) that can be used to analyse data with greater automation.
- Applying machine learning to environmental management can improve weather predictions, and climate and water resource monitoring.
- Wider implementation of remote sensing is limited by the availability of robust and representative datasets for training the machine learning algorithms.
- Quality assurance and control measures associated with the use of machine learning algorithms are not regulated, raising ethical and legal concerns.

## Background

Environmental systems, such as weather and climate, are complex and dynamic.[1] Increasingly, data on these systems are collected using remote sensing technologies, from Earth Observation (EO) satellites to unmanned aerial vehicles (UAVs), including drones (POSTnote 566).[2–5] This is generating substantial quantities of data, and environmental remote sensing is now considered an area of Big Data (POSTnote 468). For example, the Sentinel satellites, (an element of Copernicus, the European Space Agency (ESA) and European Commission EO programme, POSTbrief 37) are expected to produce 10 terabytes of EO data per day once fully operational (equivalent to around 2.5 million high definition photos).[6–8]

In order to address the challenge of processing large amounts of environmental data, experts are using artificial intelligence (AI) tools, such as machine learning for more efficient data analysis (Box 1).[9,10] Machine learning is a purely data-driven approach, which means it can be used to extract valuable information about a natural phenomenon from the data alone, without building in any model of how the data were generated from an underlying system.[11] This has benefits such as being able to manage more complex environmental data, but has challenges such as data accessibility.[10,12]

## Algorithms and environmental monitoring

While there is no universally agreed definition of AI, it is generally considered to include technologies that would otherwise require human intelligence.[13] All AI programs are underpinned by algorithms, which are a set of instructions used to perform computational tasks.[14] Machine learning algorithms allow a system to learn and improve from data and experience without being specifically programmed (Box 1).[14,15] These algorithms have been used since the 1990s to automate data collection, make sense of remote sensing data and extract previously undiscoverable patterns in the data.[16–18]

The availability of greater computational power and access to larger amounts of data have improved the capability of these systems.[19] Applications of machine learning in environmental remote sensing include weather forecasting, flood and drought prediction, and precision agriculture (Box 2). More recently, there has been increasing interest in a type of machine learning known as 'deep learning' (Box 1).[20–22] Because deep learning algorithms can easily be increased in size (they are 'scalable'), they can be expanded to manage larger data sets with increasing performance.[23,24] Although its use is not yet widespread in environmental management, experts are exploring applications of deep learning in research.[20]

The Parliamentary Office of Science and Technology, Westminster, London SW1A 0AA
**02072192840 post@parliament.uk parliament.uk/post @POST_UK**

**Box 1: Machine and deep learning**

**Machine learning**

Machine learning is a branch of AI where algorithms can learn and improve from experience and data without being specifically programmed, reducing the level of human intervention.[14] The quality of the output from the model depends on the quality of the data used to build it.[25] Machine learning models can be supervised, unsupervised or a combination. Supervised learning models are built on labelled training data.[26] For example, using images of known trees in order to identify trees in a new image. Unsupervised learning models are built on unlabelled training data.[26] For example, through exposure to millions of unlabelled images of trees in order to extract the common features that make up a tree and identify it in a previously unseen image.

**Deep learning**

Deep learning algorithms are a more modern type of machine learning algorithm, which are designed in some ways to mimic the neural networks of the human brain.[24] The algorithms are built of networks with large numbers, often hundreds of layers of individual components (hence the name "deep"), to carry out tasks such as image classification where advanced feature extraction can take place.[27] The decision to use a particular algorithm is based on algorithm suitability, the type and availability of training data, how the data will be used, as well as ethical considerations.[28–30]

The Government identified AI and the data revolution as one of four Grand Challenges (areas where the UK could be a global leader) for UK industry.[31] The importance of EO for the growth of the UK space sector has also been highlighted by the Government's space strategy, created in partnership with the UK space industry.[32,33] Experts have suggested that remote sensing and AI technologies could be used to monitor progress towards environmental targets.[34–37] The recently formed Centre for Data Ethics and Innovation provides recommendations to the Government on best practice and regulatory guidelines for AI and data-driven technologies.[38] Similar work is carried out by the Information Commissioner's Office, the Office for Artificial Intelligence and the Alan Turing Institute.[39–41]

## Benefits of machine learning

Machine learning in combination with remote sensing can help to improve experts' understanding of land, ocean and atmosphere systems (Box 2).[25] This can lead to benefits, including improved predictions about the behaviour of such environmental systems, improved automation of data analysis, better management of resources, and the discovery of new insights from complex data sets.

### Improved environmental predictions

An ongoing objective in long-term climate modelling is reducing uncertainties and improving predictions.[42,43] Machine learning has been used to reduce uncertainties about the role of aerosols in the climate system by combining 50 different aerosol models into an atmospheric map.[11] Machine learning has also been used to combine climate model projections in a way that better takes into account individual models' strengths and weaknesses, which may lead to more realistic projections.[44,45]

**Box 2. Examples of environmental applications**

- **Weather forecasting.** Physical models are the established operational weather forecasting tools.[46] However, machine learning is used alongside established methods to address challenges and advance analysis.[47] For example, the Met Office uses machine learning to remove interference in rainfall measurements.[48] These measurements are fed into forecasting tools and used by the Environment Agency for flood forecasting.[49] The Met Office is collaborating with DeepMind (a UK-based AI company) to apply deep learning to forecasting.[50]
- **Flood and drought prediction and management**. Water resource management, rainfall forecasting and advances in climate modelling all feed into flooding and drought predictions and their management (POSTnote 623).[51] Satellite and drone capabilities are being combined with machine learning for detection, mitigation, alerts, emergency response and recovery efforts.[52–55]
- **Precision agriculture.** The uptake of new technology to boost farming efficiency, referred to as precision agriculture (or farming), is low in the UK (POSTnote 505). To improve the uptake of AI, the Government has created several £1 million grants as part of the £22 million made available for Transforming Food Production.[56,57] Machine learning is applied to soil, livestock, water and crop management.[58] For example, crop health can be determined from remotely sensed images, where the abundance of chlorophyll, the green pigment responsible for photosynthesis, acts as indicator for health.[59]
- **Managing forests.** Machine learning is applied to forest management (harvesting, cropping, logistics) and in the monitoring of disease in forest species.[60,61] It can also be used in deriving forest cover change and deforestation from remote sensing images.[28] For example, machine learning is used to track and police illegal logging in tropical rainforests in real-time, with information sent to local services via a phone application.[62]
- **Marine research**. Machine learning with satellite and drone data can be used for marine habitat mapping, monitoring coastal change due to landslides, and in coastal plastic clean-up projects.[63–66] For example, an algorithm has been developed by researchers to allow detection of aggregated patches of plastics floating in coastal waters from satellite images.[67]

### Improving resource management

Information for managing vital natural resources such as water and soil (for farming and other land uses) can be generated using machine learning. The components of the hydrological cycle (the transfer of water between the ocean, land and atmosphere systems), such as the amount of surface water available, rates of rainfall and evaporation, are measured on the ground and with remote sensing technologies.[68,69] Machine learning algorithms are used to merge complex hydrological datasets, such as the satellite data collected on the total terrestrial water amounts, into global hydrological models.[70] The fusing of data in this way can address knowledge gaps and increase the accuracy of water resource estimates.[71] Improved estimates have implications for flood and drought predictions (Box 2). Machine learning is used to convert global satellite soil moisture data, an indicator of soil health, into more accurate estimates at the regional and local levels.[72,73] This has implications for local agricultural decisions (Box 2).

### Automating and improving image classification

Remotely sensed images are one of the biggest data sources for environmental science.[74–76] The classification of these images relates to the identification of features or objects in the digital image.[77] Machine learning algorithms have advanced the automation of image classification, using more of the available data, saving on time and reducing errors introduced by human analysts[78,79] For example, the UK Land Cover Map 2015 was created using satellite imagery classified into 21 types of land cover, such as 'freshwater', 'urban', and 'arable and horticulture'.[80] With the use of more sophisticated machine learning methods, the 2015 Land Cover Map was produced faster and with greater classification accuracy than the previous 2007 version.[81] This land cover map is used across sectors from government to NGOs, research and commercial stakeholders. The Scottish Government, in partnership with industry, is developing its own land cover map using machine learning.[82]

Deep learning (Box 1) is an emerging type of machine learning that is particularly useful for image classification. It has shown positive results when applied to environmental data, with some areas outperforming other machine learning algorithms.[20,83–91] However, deep learning typically has a higher energy consumption than other types of machine learning as it needs a larger amount of computing power to run. This requires the creation of new data processing services (Box 3).

### Discovering new insights from complex datasets

Machine learning makes processing large datasets more feasible. Large datasets can be processed with algorithms to extract data and derive patterns that might otherwise be too complex or time consuming for human analysts.[18] For example, new air-pressure patterns over the Tasman sea were discovered inadvertently, which have implications for the regional climate.[44] BioDAR, a UK-based insect diversity monitoring programme, makes use of the background data collected by weather stations to derive insect biodiversity information in the skies and relate that to the ground.[92] Other examples include the development of digital twins (a virtual interactive model of a physical system). For example, Belmap, a 3D model of Belgium, is being developed with machine learning enhanced remote sensing data.[93]

### Managing unwanted data

Clouds are a common feature of satellite images blocking the line of sight to the ground or creating shadows that prevent analysis.[95] Machine learning algorithms can be used to distinguish between clouds and useful Earth imagery with increased accuracy and automation in a technique referred to as 'cloud masking'.[96,97] Advances to this step can reduce the delay between when data are collected and when they are ready to be used (See Analysis Ready Data section below). Non-machine learning algorithms used for cloud masking are complicated, and have high computing costs and patchy accuracy.[98]

**Box 3. Massive GPU cluster for Earth observation**

The Plymouth Marine Laboratory will host a dedicated cloud computing (POSTnote 629) facility for processing Earth observation (EO) data on behalf of the National Centre for Earth Observation. This new computing cluster, called MAGEO (Massive Graphical Processing Unit Cluster for Earth Observation), will facilitate the application of deep learning to EO data. MAGEO was funded through a £1 million Natural Environment Research Council (NERC) grant, and from 2021 will be operated as a service for scientific researchers.[94]

### Detecting ongoing change

The ability to detect change in the natural environment is important for monitoring and allows for timely decision making.[99] For example, machine learning algorithms are used to identify changes in satellite images of land as an indicator of whether land use is changing over time, such as between recreational, conservation or economic use.[100,101] Change detection can also be used in preparing for and managing natural disasters, such as in detecting flash flooding events and in monitoring urban expansion.[102–104] At the local scale, drone technology can be used to monitor system changes such as changes to river habitats.[105]

## Technical and ethical challenges

While there are many benefits to using machine learning in environmental remote sensing, there are also some barriers to its use. Widespread implementation of machine learning is limited by data accessibility, building and validating the training data, and uncertainties in algorithm design.[106–108] There are also some associated ethical and legal considerations.[109–111]

### Data accessibility

*Analysis Ready Data (ARD)*

Copernicus EO data is free at the point of use and data can be accessed directly from ESA and a range of UK agencies.[112–117] Downloaded data needs to be processed into ARD, which is data that are ready to be analysed with minimal additional user effort.[118,119] This can later be developed into an ARD product, such as a map, or an ARD platform, such as an interactive and accessible system.[120–122] The preparation of ARD requires expertise and computational power and can consume 80% of an analyst's time.[123] Steps include removing the effects of the atmosphere, removing digital noise, unifying the data format from multiple instrument types and including contextual information.[10,124]

The international Committee on Earth Observation Satellites (CEOS), representing the major EO agencies, identifies ARD as an important part of making EO data more accessible, and in reducing the time and associated costs of data preparation.[125] ARD for some data types can be accessed from CEOS and ESA.[126–128] In the UK, the Joint Nature Conservation Committee led technical innovation to automate the processing of Sentinel-1 and Sentinel-2 data to global standards and played a key role in establishing the Defra Earth Observation Data Service.[122,129] The processed ARD is available under an Open Government License from the Centre for Environmental Data Analysis (CEDA).[130] Private EO companies, such as 4 Earth Intelligence, sell ARD products from a range of satellite data, recognising the demand for data in a ready-to-use format.[131]

*Storage and computational needs*

The demands on storage, transfer and processing of large datasets are high.[132–134] Academic users can usually access storage and high-performance computing via institutional access or research council funding.[135–137] National cloud-computing services such as CEDA, the Joint Analysis System Meeting Infrastructure Needs (JASMIN) and MAGEO (Box 3) are accessible to grant funded UK researchers.[138,139] Google Earth Engine, IBM, and Earth on Amazon Web Services provide cloud EO access and analysis, which can be free or with a charge for providing online training and customer support.[133,140–143] The possibility of sudden changes to these services and non-transferability of data create uncertainty.[133]

## Data management

The effectiveness of machine learning algorithms is dependent on having an accurate and well-developed training dataset, which can take months or years to build and can be a bottleneck to wider implementation.[25,29,144,145]

*Collecting and validating the training data*

The amount of training data needed is application specific and depends on the complexity of the problem and the algorithm.[146] Accessing an adequate amount of training data is a constant challenge, which can be eased by the existing and free datasets like Copernicus EO data.[52,112] The data to train and test the machine learning model can come from higher resolution remote sensing datasets, such as UAV, aerial imagery and local field data.[4,5,147,148–150] The data and algorithm also need to be validated (checked for accuracy) to the specifics of the application.[29,52,151] For example, a machine learning system built to detect images of rivers in temperate regions, where rivers are surrounded by green vegetation, would need to be re-trained to detect rivers in desert regions. Some applications, such as biodiversity monitoring, are more dependent on local training data, which can limit the global applicability of machine learning models.[29,152,153,]

*Algorithm selection and development*

The choice to use a particular algorithm is based on historical success and proof of concept, purpose, cost of development, computational costs, data availability and popularity.[154,155] It is also dependent on the availability of the algorithm across popular programming platforms.[156] In the development of a new algorithm, a prototype is built by a researcher and entered into an open access repository where it is tested and improved. Once the software is robust enough the algorithm can be commercialised.[157]

*Data labelling and building the algorithm*

Creating the necessary amounts of labelled data requires domain expertise and can be labour intensive.[25,108] An example of labelling could be tagging flooded features in hundreds of satellite images. Efforts are being made to reduce the demands for labelling on expertise.[158] One approach is to combine unsupervised learning (Box 1), using unlabelled data to train the model, with supervised learning for fine-tuning.[24] In general for machine learning applications, 75% of the training data are used to train the model and 25% are used to validate the model.[159,160]

Ideally this 25% should be new training data and not from the original batch, but this is not always realistic in terms of resources. How well the algorithm works depends on having enough accurate data to test decisions and improve the learning accuracy. Determining and quantifying how well-suited your machine learning algorithm design is to the application is also a challenge.[29,108,161] Natural changes to the environment that fall outside the range of the training data may undermine the reliability of the algorithm.[161]

Deep learning algorithms produce better results with unlabelled data but require significant amounts to prevent 'overfitting'. This is where the algorithm processes every aspect of the training data, producing results that may be misleading, instead of deriving only valuable information from it.[18] Crowdsourcing can be used for data labelling, but this requires additional systems to cross validate the work.[162]

## Uncertainty, error and bias

Inherent to all of the remote sensing and locally collected field data used to train and test the models is a level of uncertainty and error.[163,164] The human analyst can also introduce errors and bias in the preparation of ARD, data labelling and overall data management.[108,165] In classification problems, algorithm accuracy and performance is heavily reported as classification accuracy, which is an incomplete measure of model robustness.[166] Uncertainties in algorithms may be poorly understood by stakeholders. For example, an image classification of 'tree' may be based on a probability of 90%, but it can never be 100%, the risks of which may not be made explicit. Best practice guidelines vary and there is an overall lack of regulation by government or relevant public bodies, in spite of the wide range of application areas and the speed of algorithm development.[167]

## Ethical and legal issues

Machine learning models are sometimes referred to as 'black box', implying that it is not easy to determine how an algorithm made a decision.[18,13] This is especially common for deep learning algorithms because of the more complex processing architecture.[168] While algorithms have some level of inherently unexplainable features, they can be explained with varying levels of difficulty.[13,169,170] However, for a public sector organisation, where any aspect of a decision is opaque, it raises issues of accountability and having the authority to make a decision.[171] Commercial algorithms referred to as black box can relate to the propriety nature of the technology rather than the technical challenges of explaining how it works.[169]

The growing need for 'explainable AI' is met in part by the regulations set out in the EU General Data Protection Regulation (GDPR), which came into effect in 2018. However, the GDPR Article relating to 'a right to an explanation' only covers the use of personal data and does not cover all AI-enabled decision making.[170] Remote sensing data itself raises ethical challenges as it operates at a high resolution, creating issues around privacy and the use of such data by the state.[172] The lack of quality assurance and quality control measures associated with the use of machine learning algorithms poses legal issues around decision-making and liability. As machine learning becomes of greater value to the UK economy, challenges could be addressed by stakeholders such as NGOs, academics and commercial EO companies working with government to establish good practice codes, standards and regulatory frameworks.[13,37,173,174]

POST is an office of both Houses of Parliament, charged with providing independent and balanced analysis of policy issues that have a basis in science and technology. POST is grateful to Berivan Esen for researching this briefing, to NERC for funding her parliamentary fellowship, and to all contributors and reviewers. For further information on this subject, please contact the co-author, Dr Jonathan Wentworth. 

**Endnotes**


1. IPCC (2001). The Climate System: an Overview.
2. United Nations Office for Outer Space Affairs (2018). Annual Report.
3. Pixalytics (2019). How many satellites orbiting the Earth in 2019?
4. Datta, A. (2014). UAVs: The new trend in acquiring geospatial data through remote sensing. *Geospatial World*.
5. (2018). Special Issue: Remote Sensing from Unmanned Aerial Vehicles (UAVs). *Remote Sens.*, Vol 10,
6. ESA (2016). Sentinel-1B delivers.
7. Dropbox. How much is 1 TB of data storage?
8. Stanford University. Kilobytes Megabytes Gigabytes Terabytes.
9. ESA (2009). Working towards AI and Earth observation.
10. Sun, A. Y. *et al.* (2019). How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environ. Res. Lett.*, Vol 14, 073001.
11. Lary, D. J. *et al.* (2016). Machine learning in geosciences and remote sensing. *Geosci. Front.*, Vol 7, 3–10.
12. Soo Kim, B. *et al.* (2017). Data modelling versus simulation modelling in the big data era: case study of a greenhouse control system.
13. House of Lords (2017). AI in the UK: ready, willing and able?
14. Hutson, M. (2017). AI Glossary: Artificial intelligence, in so many words. *Science*, Vol 357, 19–19.
15. Duan, Y. *et al.* (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *Int. J. Inf. Manag.*, Vol 48, 63–71.
16. Lary, D. (2010). Artificial intelligence in geoscience and remote sensing. in *Geoscience and Remote Sensing New Achievements*. 105–128. InTech.
17. Hsieh, W. W. (2009). Machine learning methods in the environmental sciences: neural networks and kernels. *Cambridge University Press*.
18. Bergen, K. J. *et al.* (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, Vol 363, eaau0323.
19. Nguyen, G. *et al.* (2019). Machine Learning and deep learning frameworks and libraries for large-scale data mining: a survey. *Artif. Intell. Rev.*, Vol 52, 77–124.
20. Ma, L. *et al.* (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.*, Vol 152, 166–177.
21. Esteva, A. *et al.* (2019). A guide to deep learning in healthcare. *Nat. Med.*, Vol 25, 24–29. Nature Publishing Group.
22. Levine, S. *et al.* (2016). Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *arXiv:1603.02199*,
23. Brownlee, J. (2019). What is Deep Learning? *Machine Learning Mastery*.
24. Chen, X. *et al.* (2014). Big Data Deep Learning: Challenges and Perspectives. *IEEE Access*, Vol 2, 514–525.
25. Lary, D. J. *et al.* (2018). Machine learning applications for earth observation. *Earth Obs. Open Sci. Innov.*, Vol 165,
26. Ball, J. E. *et al.* (2017). Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. *J. Appl. Remote Sens.*, Vol 11, 54.
27. Mayer, R. *et al.* (2020). Scalable deep learning on distributed infrastructures: challenges, techniques, and tools. *Association for Computing Machinery*.
28. Holloway, J. *et al.* (2018). Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review. *Remote Sens.*, Vol 10, 1365.
29. Wagenaar, D. *et al.* (2019). Invited perspectives: How machine learning will change flood risk and impact assessment. *Nat Hazards Earth Syst Sci Discuss*, Vol 2019, 1–23.
30. Wearn, O. R. *et al.* (2019). Responsible AI for conservation. *Nat. Mach. Intell.*, Vol 1, 72–73.
31. BEIS (2019). The Grand Challenges.
32. UK space industry sets out vision for growth. *GOV.UK*.
33. Space Growth Partnership, UK Space (2017). Space Growth Partnership Statement of Aims.
34. DEFRA (2018). 25 Year Environment Plan.
35. Caldecott, B. (2019). Space tech and AI could be game changers for the environment. *Green Alliance Blog*.
36. Bonham, C. (2019). Green spaces in residential gardens. *Data Science Campus*.
37. London Economics (2018). Value of satellite-derived Earth Observation capabilities to the UK Government today and by 2020.
38. Centre for Data Ethics and Innovation. *GOV.UK*.
39. Information Commissioners Office. (2020). ICO consultation on the draft AI auditing framework guidance for organisations.
40. The Office for Artificial Intelligence: About us. *GOV.UK*.
41. The Alan Turing Institute: About us.
42. Snow, J. (2019). How artificial intelligence can tackle climate change. *National Geographic*.
43. Rolnick, D. *et al.* (2019). Tackling Climate Change with Machine Learning. *Arxiv*.
44. Jones, N. (2017). How machine learning could help to improve climate forecasts. *Nat. News*, Vol 548, 379.
45. Monteleoni, C. *et al.* (2013). Climate Informatics: Accelerating Discovering in Climate Science with Machine Learning. *Comput. Sci. Eng.*, Vol 15, 32–40.
46. MET Office History of numerical weather prediction. *Met Office*.
47. Holmstrom, M. *et al.* (2016). Machine learning applied to weather forecasting. *Meteorol Appl*,
48. Timothy Darlington, MET Office. Personal Communication. (04.03.2019).
49. MET Office Rainfall radar. *Met Office*.
50. Robinson, N. (2019). Machine learning at the Met Office.
51. NASA (2019). A Future for Freshwater at two extremes.
52. Mosavi, A. *et al.* (2018). Flood prediction using machine learning models: Literature review. *Water*, Vol 10, 1536.
53. Google (2018). Keeping people safe with AI-enabled flood forecasting.
54. Cranfield University. Using drones to manage flood risk.
55. Mircea, C. (2019). Using AI to predict droughts, floods and conflict displacements in Somalia. *Medium*.
56. Defra in the media. (2019). Government announces projects worth £22 million to transform food production.
57. Clark, P. (2019). £22m spending to raise food output and cut pollution. *Farmers Weekly*.
58. Liakos, K. G. *et al.* (2018). Machine learning in agriculture: A review. *Sensors*, Vol 18, 2674.
59. Chlingaryan, A. *et al.* (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput. Electron. Agric.*, Vol 151, 61–69.
60. McQueen, S. (2019). How Artificial Intelligence, Robots Enhance Forest Sustainability in Finland. *Esri*.
61. Rezatec. (2019). DEFRA use satellite technology to monitor disease in key species.
62. ESA (2017). Forest Sentinel – App-based Deforestation Alerts. *Copernicus Masters*.
63. Zhang, C. (2015). Applying data fusion techniques for benthic habitat mapping and monitoring in a coral reef ecosystem. *ISPRS J. Photogramm. Remote Sens.*, Vol 104, 213–223.

64. Dickson, M. E. *et al.* (2016). Identifying the controls on coastal cliff landslides using machine-learning approaches. *Environ. Model. Softw.*, Vol 76, 117–127.
65. Greenemeier, L. (2017). The Plastic Tide. *Scientific American*.
66. Ellipsis.Earth.
67. Biermann, L. *et al.* (2020). Finding Plastic Patches in Coastal Waters using Optical Satellite Data. *Sci. Rep.*, Vol 10, 1–10. *Nature Publishing Group*.
68. Sheffield, J. *et al.* (2018). Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions. *Water Resour. Res.*, Vol 54, 9724–9758.
69. National Academies of Sciences, E. *et al.* (2018). Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space. *The National Academies Press*.
70. Sun, A. Y. *et al.* (2019). Combining Physically Based Modeling and Deep Learning for Fusing GRACE Satellite Data: Can We Learn From Mismatch? *Water Resour. Res.*, Vol 55, 1179–1195.
71. Chen, H. *et al.* (2019). Rainfall estimation from ground radar and TRMM Precipitation Radar using hybrid deep neural networks. *Geophys. Res. Lett.*, Vol 46, 10669–10678.
72. Im, J. *et al.* (2016). Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. *Environ. Earth Sci.*, Vol 75, 1120.
73. Guevara, M. *et al.* (2019). Downscaling satellite soil moisture using geomorphometry and machine learning. *PLOS ONE*, Vol 14, e0219639. *Public Library of Science*.
74. Anzilotti, E. (2017). Every day, this satellite company takes a snapshot of the entire planet. *FastCompany*.
75. Maclenan, A. (2019). Your eyes in the sky: Making satellites work for everyone. *Nesta*.
76. Cheng, G. *et al.* (2017). Remote sensing image scene classification: Benchmark and state of the art. *Proc. IEEE*, Vol 105, 1865–1883.
77. Shinozuka, M. *et al.* (2009). 4 - Synthetic aperture radar and remote sensing technologies for structural health monitoring of civil infrastructure systems. in *Structural Health Monitoring of Civil Infrastructure Systems*. (eds. Karbhari, V. M. et al.) 113–151. *Woodhead Publishing.*
78. Maxwell, A. E. *et al.* (2018). Implementation of machine-learning classification in remote sensing: an applied review. Vol 39, 2784–2817.
79. Weih, R. *et al.* (January 1). Object-based classification vs. Pixel-based classification: Comparitive importance of multi-resolution imagery. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, Vol 38,
80. Burns, P. (2015). Land Cover Map 2015. *UK Centre for Ecology & Hydrology*.
81. (2017). LCM2015 Dataset Documentation. *UK Centre for Ecology & Hydrology*.
82. (2020). Space Intelligence wins contract from Scottish Enterprise as part of their AI for Good funding. *Space Intelligence*.
83. Tsagkatakis, G. *et al.* (2019). Survey of Deep-Learning Approaches for Remote Sensing Observation Enhancement. *Sensors*, Vol 19, 3929.
84. Ronan, F. *et al.* (2019). Special Issue - Deep Learning Remote Sensing Data. *Sensors*,
85. Benediktsson, J. A. *et al.* (2019). Special Issue - Analysis of Big Data in Remote Sensing. *Remote Sens.*,
86. Chen, Y. *et al.* (2014). Deep Learning-Based Classification of Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, Vol 7, 2094–2107.
87. Liangpei, Z. *et al.* (2016). Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geosci. Remote Sens. Mag.*, Vol 4, 22–40.
88. Zhu, X. X. *et al.* (2017). Deep learning in remote Sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.*, Vol 5, 8–36.
89. Francis, A. *et al.* (2019). CloudFCN: Accurate and Robust Cloud Detection for Satellite Imagery with Deep Learning. *Remote Sens.*, Vol 11, 2312.
90. Poggi, G. *et al.* (2018). Special Issue - Deep Learning for Remote Sensing. *Remote Sens.*,
91. Li, Y. *et al.* (2018). Deep learning for remote sensing image classification: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, Vol 8, e1264.
92. BioDar Project. *BioDAR*.
93. EO4Belmap - Enriching GIM Belmap with AI-enhanced Earth Observation data. ESA Business Applications.
94. Massive GPU Cluster for Earth Observation (MAGEO). *NEODAAS*.
95. Jedlovec, G. (2009). Automated detection of clouds in satellite imagery. *Adv. Geosci. Remote Sens.*, 303–316.
96. ESA Technical Guide - Cloud Masks. *Sentinel Online*.
97. Gómez-Chova, L. *et al.* (2017). Cloud detection machine learning algorithms for PROBA-V. in 2251–2254. IEEE.
98. Zupanc, A. (2019). Improving cloud detection with machine learning. *Medium*.
99. Lu, D. *et al.* (2004). Change detection techniques. *Int. J. Remote Sens.*, Vol 25, 2365–2401.
100. Cao, C. *et al.* (2019). Land-Use Change Detection with Convolutional Neural Network Methods. *Environments*, Vol 6, 25. *Multidisciplinary Digital Publishing Institute*.
101. Rogan, J. *et al.* (2008). Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *Remote Sens. Environ.*, Vol 112, 2272–2283.
102. Lamovec, P. *et al.* (2013). Detecting flooded areas with machine learning techniques: case study of the Selška Sora river flash flood in September 2007. *J. Appl. Remote Sens.*, Vol 7, 073564.
103. Daudt, R. C. *et al.* (2018). Urban change detection for multispectral earth observation using convolutional neural networks. in 2115–2118. IEEE.
104. Karpatne, A. *et al.* (2016). Monitoring land-cover changes: A machine-learning perspective. *IEEE Geosci. Remote Sens. Mag.*, Vol 4, 8–21.
105. Woodget, A. *et al.* (2017). Drones and digital photogrammetry: from classifications to continuums for monitoring river habitat and hydromorphology. *Wires Water*.
106. Philippe Mathieu, P. *et al.* (2018). Towards a European AI for Earth observation research & innovation agenda. Proceedings of a workshop at ESA Φ-lab. ESA.
107. Zuleta, I. (2019). Planet experts discuss future of analysis ready data at 2019 ARD workshop. *Planet*.
108. Begoli, E. *et al.* (2019). The need for uncertainty quantification in machine-assisted medical decision making. *Nat. Mach. Intell.*, Vol 1, 20–23.
109. Kaeser-Chen, C. *et al.* (2019).Towards Ethical Deployment of AI for Conservation Systems. *Cornell.edu*
110. Berman, G. *et al.* (2018). Ethical Considerations When Using Geospatial Technologies for Evidence Generation. *Unicef*
111. Atabekov, A. *et al.* (2018). Legal status of artificial intelligence across countries: Legislation on the move. *Eur. Res. Stud. J.*, Vol 21, 773–782.
112. Copernicus. Access to data.
113. Sentinel Data Access Overview. *Sentinel Online*.
114. (2018). Sentinel data weighs in at a 'horn'-ormous 2PB. *CEDA*.
115. (2018). New dataset: ESA Sentinel 3A SLSTR data. *CEDA*.
116. Sentinel Data Access Service (SeDAS). *Satellite Applications Catapult*.
117. Satellite Applications Catapult. Working with Industry.
118. CEOS Analysis Ready Data.

119. Holmes, C. (2018). Analysis Ready Data Defined. *Medium*.
120. Doldirina, C. (2015). Open Data and Earth Observations The Case of Opening Up Access to and Use of Earth Observation Data Through the Global Earth Observation System of Systems. *JIPITEC*, Vol 6,
121. (2019). Analysis Ready Data (ARD). *JNCC*.
122. (2019). JNCC broadens access to satellite data – development of analysis-ready data. *JNCC*.
123. CEOS Data Users: Why CEOS Analysis Ready Data for land?
124. Dwyer, J. L. *et al.* (2018). Analysis ready data: Enabling analysis of the Landsat archive. *Remote Sens.*, Vol 10, 1363.
125. CEOS. (2019). Analysis Ready Data Strategy.
126. CEOS. CEOS Analysis Ready Data, Data Users.
127. EO Science for Society, ESA. (2019). European Data Cube Facility service: EO resource for researchers and value-adders.
128. Giuliani, G. *et al.* (2020). Data Cube on Demand (DCoD): Generating an earth observation Data Cube anywhere in the world. *Int. J. Appl. Earth Obs. Geoinformation*, Vol 87, 102035.
129. earthobs.defra.gov.uk.
130. The CEDA Archive.
131. 4 Earth Intelligence. Satellite Data Sales.
132. Ma, Y. *et al.* (2015). Remote sensing big data computing: Challenges and opportunities. *Future Gener. Comput. Syst.*, Vol 51, 47–60.
133. Sudmanns, M. *et al.* (2019). Big Earth data: disruptive changes in Earth observation data management and analysis? *Int. J. Digit. Earth*, 1–19.
134. Remedios, J. (2012). Processing EO data and its analysis with respect to models.
135. ARCHER national supercomputing service. EPCC at The University of Edinburgh. *EPCC*.
136. Durham University. (2020). Durham to host new national supercomputer.
137. University of Leicester. High performance computing: ALICE.
138. CEDA.
139. JASMIN.
140. Google Earth Engine.
141. IBM PAIRS Services.
142. Earth on AWS.
143. Google Earth Engine: FAQ.
144. Sivaraj, A. (2015). The founder's guide to machine learning: why you shouldn't build your own models. *Indico*.
145. Faria, T. (2019). We need to talk about the lifecycle of machine learning models. *Medium*.
146. Brownlee, J. (2017). How Much Training Data is Required for Machine Learning? *Machine Learning Mastery*.
147. Holloway, J. *et al.* (2018). Spatial and machine learning methods of satellite imagery analysis for Sustainable Development Goals. *16th Conference of IAOS*.
148. NERC Data centres.
149. Copernicus Open Access Data Hub.
150. Abdikan, S. *et al.* (2016). Land cover mapping using sentinel-1 SAR data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, Vol 41, 757.
151. Gonfalonieri, A. (2019). How to build a data set for your machine learning project. *Medium*.
152. Haas, A. de. (2017). Eyes in the sky and on the ground are helping forest conservation in Cameroon. *The Conversation*.
153. Bell, G. *et al.* (2015). Use of remote sensing to produce a habitat map of Norfolk. *Ecol. Inform.*, Vol 30, 293–299.
154. Schulz, K. *et al.* (2018). Machine learning methods for remote sensing applications: an overview. in Vol 10790, 1079002. *International Society for Optics and Photonics*.
155. Almaliki, Z. A. (2019). Do you know how to choose the right machine learning algorithm among 7 different types? *Medium*.
156. (2017). Commonly Used Machine Learning Algorithms | Data Science. *Analytics Vidhya*.
157. Crick, T. *et al.* (2014). 'Can I Implement Your Algorithm?': A Model for Reproducible Research Software. *ArXiv Prepr. ArXiv14075981*,
158. Xia, G.S. *et al.* (2015). Accurate annotation of remote sensing images via active spectral clustering with little expert knowledge. *Remote Sens.*, Vol 7, 15014–15045.
159. Pantazi, X. E. *et al.* (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Comput. Electron. Agric.*, Vol 121, 57–65.
160. Chi, M. *et al.* (2008). Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. *Adv. Space Res.*, Vol 41, 1793–1799.
161. Hyatt, J. S. *et al.* (2019). Requirements for Developing Robust Neural Networks. *ArXiv Prepr. ArXiv191002125*,
162. de Oca, A. M. M. *et al.* (2017). Earth observation image semantic bias: A collaborative user annotation approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, Vol 10, 2462–2477.
163. Zhang, Q. *et al.* (2019). An Uncertainty Descriptor for Quantitative Measurement of the Uncertainty of Remote Sensing Images. *Remote Sens.*, Vol 11, 1560.
164. Foody, G. M. *et al.* (2003). Uncertainty in remote sensing and GIS. *John Wiley & Sons.*
165. Wang, X. *et al.* (2016). Learning from uncertainty for big data: future analytical challenges and strategies. *IEEE Syst. Man Cybern. Mag.*, Vol 2, 26–31.
166. Brownlee, J. (2017). How to Report Classifier Performance with Confidence Intervals. *Machine Learning Mastery*.
167. House of Commons Science and Technology Committee (2017). Algorithms in decision-making.
168. Castelvecchi, D. (2016). Can we open the black box of AI? *Nat. News*, Vol 538, 20.
169. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.*, Vol 1, 206–215. Nature Publishing Group.
170. The Royal Society (2019). Explainable AI: the basics.
171. Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, Vol 376, 20170359.
172. Coffer, M. M. (2020). Balancing Privacy Rights and the Production of High-Quality Satellite Imagery. *Environ. Sci. Technol.*, Vol 54, 6453–6455. American Chemical Society.
173. BEIS (2019). Regulation for the Fourth Industrial Revolution.
174. Centre for Data Ethics (CDEI) 2 Year Strategy. *GOV.UK*.