

---

## The Rivers Trust

# Machine learning approaches for Environmental data analysis

### TINDSG-009

#machinelearning #datastudygroup

### About the Organisation

[The Rivers Trust](#) is the umbrella body for the Rivers Trust movement, comprising 60 local charities delivering environmental improvements to rivers and their catchment areas throughout the UK and Ireland. The Rivers Trust has nearly 50 employees and has expanded rapidly in recent years. Find out more about the impact of our work in our [annual review](#).

With our expertise, reach and our member trusts' local knowledge, we work with farmers, government, businesses and local communities to deliver lasting environmental improvements. We bring together the people, knowledge, data and intelligence to help rivers thrive again; for us and future generations.

We strive to be an inclusive and collaborative movement and believe that by working together, we can make our shared vision a reality: wild, healthy, natural rivers, valued by all.

### Role Description and Responsibilities

You will be part of an exciting transformational change project called the Catchment Systems Thinking Cooperative (CaSTCo) which is a three-year Ofwat-funded innovation partnership between The Rivers Trust, 12 water companies and 11 other academic, commercial and NGO partners.

The CaSTCo project is launching this summer and will revolutionise the way crucial data about England and Wales' water environment is gathered and shared, in particular on the health of the nation's rivers. Only 14% of rivers in England are in 'good ecological health' – one of the worst records in Europe, yet we lack the density of information needed to fully understand the nature and sources of intermittent problems such as pollution, or the effectiveness of Nature Based Solutions at delivering multiple beneficial outcomes such as flood protection and water quality improvements. We need to fill evidence gaps with enough certainty to drive action and, crucially, to bring together the partners and funding streams which could tackle these complex problems. This project will start to build a more robust evidence base for tackling environmental challenges by developing a national framework of standardised monitoring tools, training and data platforms. A major focus of the project is that it will directly engage and empower local communities in gathering local evidence in eight demonstrator catchments, feeding this information into decision support tools and decision-making processes.

The CaSTCo project will upskill and mobilise citizen scientists to collect environmental data alongside wider deployment of 'continuous' monitoring sensor networks. Your role will be to streamline the interpretation and extraction of insight from this sensor data so that it can be used and applied by all project stakeholders. Through the project partners you will have

## TIN + DSG Spring 2022

access to a team of water quality monitoring experts that can provide interpretation of the data, help define the scope and test the machine learning approaches which you develop.

You will have the opportunity to work with businesses, environmental NGOs, government, local communities and campaign groups to support greater use and understanding of data for detecting pollution sources and events, and for underpinning sustainable catchment management decision-making.

The Rivers Trust has secured access to a number of data sources for use on this project, including:

- Continuous water quality sensor data from the Environment Agency's National Water Quality Instrumentation Service – data variables include 8 key water quality parameters (temperature, turbidity, conductivity, pH, dissolved oxygen, chlorophyll, blue-green algae and ammonium)
- Water chemistry data from the Environment Agency's water quality archive – over 58 million rows of data from the last 20 years of spot sampling undertaken by the EA
- Continuous water quality data from a network of simple sensors (measuring temperature, water level and conductivity) deployed at high spatial resolution across a river catchment (7 sensor nodes in a catchment of 14 sq km area). This network has specifically been deployed to collect a 'training' data set to enable greater understanding of dynamic water quality characteristics at the catchment scale
- Continuous water quality data from water quality sondes and tryptophan-like fluorescence sensors (an indicator of bacterial contamination)
- Rainfall and river level data from the Environment Agency's flood monitoring network
- Citizen Science data comprising 8000 surveys across the Westcountry between 2016 and 2022. The data set is still growing and includes records of wildlife sightings, pollution sources, problem plants, water quality and flow conditions – with accompanying photographs

You will work within The Rivers Trust Technical Team – a team of around 10 technical experts in the fields of spatial data analysis and visualisation, environmental modelling and monitoring. We do not currently have machine learning expertise within the team, so you will help build our understanding of the potential for application of machine learning and artificial intelligence approaches to environmental data science. Members of the Technical Team will be working on the CaSTCo project – liaising with the catchment demonstrators, developing data collection and monitoring plans and protocols for sharing data from diverse sources. You will work closely with the team to integrate your work in to the CaSTCo programme.

### Expected Outcomes

This internship offers a unique opportunity to help shape the future of water quality and environmental data analysis, interpretation and collection in England and Wales. Your work will be aligned with an innovation project which is being co-designed by the project partners, so you will have the opportunity to shape and design the processes and specific deliverables in a highly collaborative environment. You will be supported by leading national water quality and environmental data experts in interpreting and deriving insight from a range of archived and live data sets and your work could influence the direction of future environmental monitoring strategies.

Some of the outcomes which we anticipate from this internship, the Data Study Group and the CaSTCo project include:

- Improved understanding of the potential and limitations of machine learning and artificial intelligence approaches for water quality / environmental data analysis and

## TIN + DSG Spring 2022

interpretation.

- Improving the data literacy of communities and NGOs and enabling better understanding of the multiple factors affecting river health in space and time and the interdependencies between these factors.
- Better information for swimmers and recreational river users about dynamic changes in water quality and the factors that influence whether a river is fit to swim in.
- Application specific indicators derived from multiple data sources that help to identify pollution events and characterise catchment response to rainfall events i.e. automatic discrimination between point source pollution and rainfall-driven run off events
- Automatic flagging of suspect data by learning to recognise common features that are indicative of instrument failure or reduced performance (e.g. sensor biofouling, drift, etc.)
- Identifying long term trends against a very 'noisy' data background – continuous water quality data is subject to wide variations over a range of timescales – diurnal variations, weather and seasonally driven variations can obscure long term trends or changes due to land use, new developments or waste treatment asset performance
- Sensor based 'classification' of river health – typically the health of rivers has been summarised using spot testing results which, whilst subject to rigorous laboratory analysis is often collected at very poor temporal and spatial resolution. We would like to investigate the potential to provide an alternative approach to classifying some aspects of river health using high temporal resolution data from continuously recording water quality sensors

The impact of these outcomes will help to transform the use of sensors in the water environment, and we anticipate that this will increase demand and in turn influence sensor manufacturers to focus research and development into the areas of lower cost sensors and technologies that are deployable by local communities and citizen scientists.

The internship will be instrumental in the preparation of a Data Study Group challenge that will follow on and expand the intern's work with the organisation. The intern will have the opportunity to represent the organization during the Data Study Group.

### Supervision and Mentorship

You will be jointly supervised by The Rivers Trust's Technical Director and Technical Lead for the CaSTCo Project. You will also be mentored by experts from partner organisations, including member Rivers Trusts, the Environment Agency, water companies and academic and commercial partners. We will also seek opportunities to engage you in some of the innovative machine learning research being developed by our partners.

### Ideal Intern

We are looking for a skilled data or computer scientist with expertise in the design and application of machine learning approaches who is keen to work in a dynamic NGO environment. You will be passionate about using your skills to tackle some of the most pressing environmental challenges of our time and will relish the opportunity to engage with an interdisciplinary team to create innovative solutions to data science questions.

You will bring both technical skills and knowledge and an enthusiastic, people-oriented approach to this role. You will be willing and able to shape the approach to this challenge,

## TIN + DSG Spring 2022

liaising with project partners to build your understanding of the project and the gaps in knowledge which it is aiming to tackle. You will use your research skills and machine learning knowledge to explore the datasets, extract features and design and test a range of machine learning approaches, including supervised and unsupervised learning, deep learning and uncertainty quantification. You will liaise with data scientists, sensor manufacturers, environmental decision makers and project stakeholders to design and scope an appropriate Data Study Group challenge which can contribute to the overall aims of the CaSTCo project.

You will have great communication skills and a willingness to share your knowledge and build data literacy among a wide range of stakeholders that you will encounter on the project. The Rivers Trust has a strong open data and knowledge sharing ethos, and we will support you to produce open research outputs from your work through conferences, data impact case studies or peer reviewed publications.

Key Qualifications:

- Strong data science, machine learning and coding skillset is essential
- Experience with time series analysis
- Excellent problem solving, critical thinking, and communication skills
- Driven, self-motivated and highly organised
- Comfortable with dynamic, innovative and creative challenges
- Environmental science, especially water quality knowledge would be desirable

Education: Pursuing a PhD in computer science, data science or related technical field.

### Internship Logistics

Salary: £30,000 per annum pro rata

Length of Internship: Six months, fixed term. Full time (37.5 hrs per week with flexible working arrangements available). Part-time will be considered for exceptional candidates.

Start date: September 2022

Location: UK home based with some travel required for occasional face-to-face meetings in England and Wales.

For technical queries, please email: [michelle.walker@theriverstrust.org](mailto:michelle.walker@theriverstrust.org)