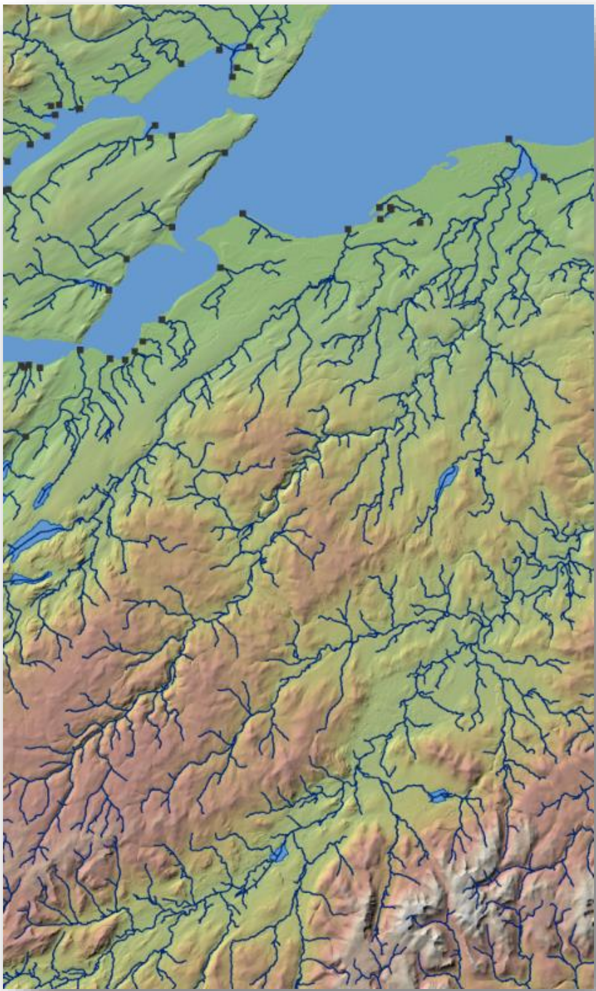
OpenRiverNetwork (ORN)

CONNECTING THE CATCHMENT: A topologically correct and attributed open-source river network for Great Britain



Report of Joint Project: GeoData Institute, University of Southampton, River Restoration Centre and Ordnance Survey (Research and Innovation)

November 2023

the River Restoration Centre
Working to restore and enhance our rivers

University of Southampton

OpenRiverNetwork (ORN)

A topologically correct and attributed open-source river network for Great Britain

Suggested citation: Hornby, D.D., Januchowski-Hartley, S., Naura, M., Hill, C.T., and Cavazzi, S. (2023) OpenRiverNetwork (ORN): a topologically correct and attributed open-source river network for Great Britain

Online resource: http://openrivers.net/

GeoData Institute, River Restoration Centre (RRC), Ordnance Survey
Version 1.0

November 2023

Funding:

This project has been supported by the Ordnance Survey (Research Department). Additional support was provided by Nature Scot to support the development of the ORN in Scotland and the Rivers Trust.

Project Officer: Stefano Cavazzi, Principal Innovation & Research Scientist, Ordnance Survey, Research Department.

Disclaimer

This report has been compiled based on the processing, analysis, correction and modification of the Ordnance Survey's Open River Network (OS Open Rivers(TM) by Stephanie Januchowski-Hartley and Duncan Hornby using data quality control tools (RivEX). The river network was attributed by Stephanie Januchowski-Hartley, GeoData, University of Southampton (GDI) and the River Restoration Centre's (RRC). The network was produced using the latest version of the OS Open River dataset which is a simplification of the comprehensive OS Water Layer(TM). The product is provided as a research output; the University of Southampton, RRC and the OS do not accept any liability for any errors in digitising polyline or in index attribution.

RRC is a national centre for information and advice and holds a dataset of river restoration and best practice management works. To inform the further development of this network, please let us know of any issues you may find or if you any suggestions how we could make it an even more valuable resource (rrc@therrc.co.uk).

GeoData Institute is a research and enterprise group within the School of Geography and Environmental Sciences at the University of Southampton. It specialises in the use of geospatial information for a sustainable environment and society.

Contents

OpenRiverNetwork (ORN)

A topologically correct and attributed open-source river network for Great Britain

Background

Ordnance Survey (OS) produce OS Open Rivers [1] as a vector polyline dataset for Great Britain that is a derivative generalisation of the more detailed OS MasterMap Water Network Layer [2]. The OS Open Rivers dataset is released under an Open Government License (OGL)[3].

Although widely used by the practitioner and the academic communities OS Open Rivers is not fully topologically connected; limiting its suitability for several uses and research applications such as linear referencing and reach analysis, hydro-ecological analysis, water quality monitoring, restoration and remediation prioritisation, connectivity planning and integration in decision support tools. Furthermore, it has certain levels of line generalisation of the river system.

The Environment Agency have developed a separate network based upon OS Master Map (OS MM) known as the Detailed River Network (DRN). The layer delineates rivers and artificial water bodies at a fine scale (1:1,250 to 1:10,000). The layer is, however, subject to copyright and cannot be shared with the wider public. The EA is also producing an 'Analysis Ready Water Network' that will be used for data derivation and analyses but again that cannot be made publicly available because of IPR restrictions. It is currently planned that derived outputs from the Analysis Ready Water Network will be transferred to the OS Open Rivers network for public use.

The only fully topologically linked network available in the UK is based on the CEH River Network[4], itself derived from OS data (1:50k raster mapping). This was generated many years ago (first in the 1990s) and has licence costs and restrictions for derived data products that restrict its availability and make it largely unsuitable for many applications and uses.

Although the OS Open Rivers data is partially topologically connected there are several exceptions which need to be resolved and a formal data structure enforced to allow for full network geospatial analysis.

The objective of this project is to explicitly maintain the open data position, such that no proprietary or third-party data rights are integrated into the product or methodology. This precludes the use of data such as OS MM / OS Water Layer and Open Street Map (OSM) data to help improve the network and each of these has restrictions for generating open derived datasets. The derived, topologically corrected and attributed data will be published under the OGL.

This project produced:

I. a topologically correct and attributed open access river GIS product for Great Britain under an Open Government Licence

---

[1] https://www.ordnancesurvey.co.uk/products/os-open-rivers
[2] https://www.ordnancesurvey.co.uk/products/os-mastermap-networks-water-layer
[3] Open Government Licence https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/
[4] https://www.ceh.ac.uk/data/15000-watercourse-network

II. standardised metadata
III. a report of the methods, applications and potential for further enhancements.

Introduction

A consortium of organisations agreed to part fund the enhancement of the OS Open Rivers product to a topologically correct river network for England, Wales and Scotland. The funders were Ordnance Survey, NatureScot, the University of Southampton, and the Rivers Trust.

The work addressed the missing topological connections within the network to provide a seamless connected river network for Great Britain, and to enable the extraction of a series of attributes that can be used for river characterisation, data analysis and modelling. A suite of attributes were encoded directly within the vector network, and additional attributes were derived and attached to a separate layer of points generated every 100 metres.

This was principally achieved using the GIS tool RivEX (Hornby, 2023). Table 1 shows the additional data attributes that have been extracted so far and attached to individual polylines on the river network (left-hand column), and for a layer of points (right-hand column) regularly spaced along the network rivers (every 100 m). It is planned to expand the list of attributes encoded into the network as open.

Table 1 Summary of the reach and point based attributes generated with the OpenRiversNetwork dataset.

Reach-based attributes	Point-based attributes
▪ Shreve stream order ▪ Strahler stream order ▪ Segment ▪ Hack order ▪ Stream name ▪ Catchment ID ▪ Distance from network mouth ▪ Source ID ▪ Total upstream length ▪ Polyline length ▪ Country ▪ River Basin	▪ Altitude ▪ Slope ▪ Distance to source ▪ Height of source ▪ Sinuosity ▪ Underlying 625k solid and drift geologies

The combined datasets will lay the foundation for future development and derivation of:

- **indices** of interest to partners and potential users (e.g. stream power, channel substrate size – see Naura et al., 2016).
- **models** (e.g. habitat suitability models for riverine species – see water vole habitat mapping by NRW and Natural England).
- **outputs** (e.g. maps of agricultural sediment risk of impact on freshwater biota – see Naura et al., 2016; identification of river reaches based on stream power as for MImAS). These would require additional processing, such as widths (min/med/max) and the use of models to predict specific attributes and geomorphological indices (e.g. Qmed, stream power, specific stream power, shear stress, river restoration strategy based on stream power and including natural recovery, assisted natural recovery, designed restoration).

Methods

Connecting the network errors

The full OS Open Rivers dataset was downloaded from the Ordnance Survey's website on 9 February 2023; it is this version of the dataset that was corrected and became the ORN. All analyses were conducted using RivEX 10.38 (Hornby, 2023) and ArcGIS 10.8.2 (ESRI, 2023).

RivEX is an ArcGIS tool capable of identifying many types of topological errors in river networks as well as encoding into the network relevant catchment parameters.

To manage the editing tasks, the network was split into the 12 WFD river basins. An initial run on the Severn River basin was undertaken to identify any additional steps and to clarify categories of modification that would be needed.

RivEX quality control tools create log files that enabled targeted editing of topological errors.

Google Earth and OS Basemaps were used to visually identify polyline locations to assess the errors that were flagged and to situate each polyline location within the broader river network. As each error was accessed and corrected a log of the correction was kept in a separate Excel file. The correction is logged against a unique ID that was added to the original OS Open Rivers download so changes could be tracked and related back to the original downloaded version. Table 2 lists the 13 types of corrections that were made to the river network and are linked to each polyline in the network via the unique ID field.

Table 2 Types of corrections made to the river network.

Correction Code	Action	Notes
None	No edit was done to the polyline	The polyline was accepted as is
Canal-Deleted	A polyline is removed that was listed or appeared (based on visual inspection) to be a canal, drain, or sewer polyline	Removal of such channels re-established a dendritic pattern to the *river* network
Connected	Disconnected polyline extended to reconnect it back into main network	examples: where a dam or culvert occurred
Discon-Deleted	Polyline deleted	Polylines that were lakes or streams but not part of the larger river basin and if retained would create internal mouths or multiple mouths for the same river (mouth). In some instances, gaps between polylines were retained in the river network, for example where a canal splits a river and the waters of the river no longer connect from above and below the canal or other artificial structure or watercourse. In those cases, there are internal river mouths within a river basin network
Flipped	Direction of polyline was reversed so it flows in a downstream direction	Algorithms that traverse a river network expect the network to be a *directed graph* where polylines flow in a source to sea direction
Intrsct-Deleted	Intersecting polyline deleted	Polylines that intersected others were deleted so to support network connectivity

Small-Deleted	Small polylines connected to canals, partially digitised, or a spike in the network were deleted	Decision to delete a small polyline was based upon visual assessments, and not all small polylines have been systematically identified, assessed, and removed from the dataset.
Split	A polyline that needed to be split so another polyline could be added to the network	Splitting a line introduces a node into the network where a new line or disconnect can be connected into the main network
Braid-Deleted	A polyline that was digitized as a side or secondary channel that created a loop in the network was deleted	Removing braids/loops results in a single-threaded network. This simplifies many types of network analyses such as barriers to fish migration
Spikes-Deleted	1 or more vertices removed from a polyline	Digitizing errors and the generalisation methods used by OS introduce un-natural acute angles along the length of the polyline. Removing a vertex returns the polyline to a more natural planform
Geom-Modified	Polyline vertices removed, added or repositioned	A polyline that was modified to better represent the waterscape (e.g., a vertex removed to disconnect a stream that doesn't connect to a canal)
Insert	Polyline created to connect a disconnection to the main network	This is a simple straight line and inherits the attribution of its adjacent lines
Named	The name1 field was updated with the waterbody name. Can be river/lake/Loch.	Polyline is missing a name, yet it can be clearly determined from the OpenStreetMap base map

After the first pass editing, the 12 river basins were recombined to generate the GB-wide dataset.

RivEX has recently been ported to the Toolbox environment of ArcPro; ESRI are phasing out ArcMap (in which the current RivEX is written) to be replaced by their modern equivalent, ArcPro. This has allowed the addition of new checking tools in the latest release. The GB-wide corrected network was re-run through the new quality controls of the ArcPro version of RivEX and a few additional topological errors, not identified by the ArcMap version of RivEX, were picked up and corrected.

Figure 1 charts the type and number of edits made to the OS Open Rivers to create the topologically correct ORN. Only 4.1% of the network required editing to ensure topological correctness.

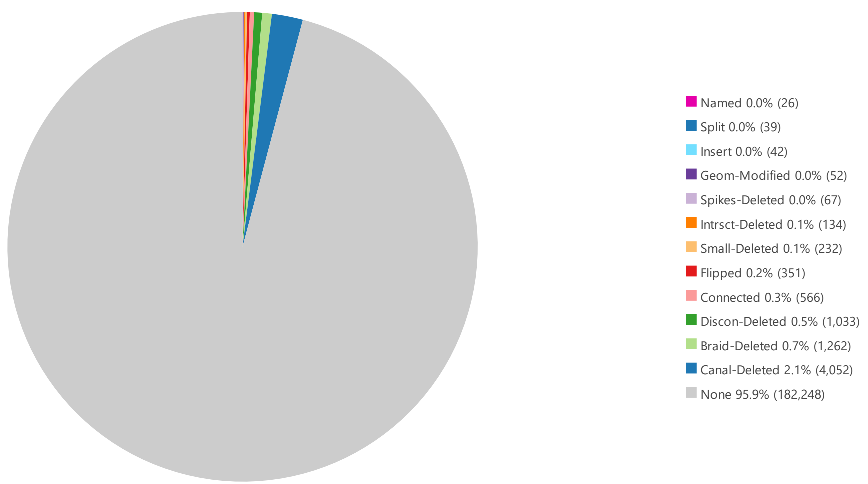


Figure 1 Type and number of edits made to the OS Open Rivers. 95.9% of the network was not altered.

Attribution

With the river network topology resolved it possible to run the dataset through RivEX to add a range of catchment variables. These are encoded into each polyline alongside the attributes that were kept from the OS Open Rivers dataset. Country and Basin codes were added manually through a series of selections and field calculations.

The GB-wide dataset is stored in an ESRI File Geodatabase for performance and data integrity; thus field naming conforms to Geodatabase conventions. Table 3 lists the fields in this release of the ORN.

Table 3 Fields in the ORN vector dataset.

Field name	Field Type	Description
OBJECTID	Object ID	A unique row ID maintained by the Geodatabase.
Shape	Geometry	The polyline geometry field, has no Z or M values.
Shape_Length	Double	The length of the polyline in metres, maintained by Geodatabase.
Name1	Text	The waterbody name, this field is inherited from the OS Open Rivers. It is not always populated.
Form	Text	Classification of the type of watercourse that is formed by the polyline, this field is inherited from the OS Open Rivers. Classifications are *canal*, *inlandRiver*, *lake* or *tidalRiver.*
Flow	Text	Direction of water flow in the polyline relative to digitisation of the geometry, this field is inherited from the OS Open Rivers. Classifications are *in direction* or *unknown.*
Name2	Text	An alternative name of the watercourse the polyline is part of. This field is inherited from the OS Open Rivers. It is not always populated.
UniqueID	Long	A sequential unique ID added to the dataset as a first step at the beginning of this project to track and relate changes back to the source OS Open Rivers that the ORN is derived from. This field was created as the ObjectID field changes during editing/exporting/merging of the data.
RivID	Long	A sequential unique ID created by RivEX used to build and extract the topological information of the network.
Fnode	Long	The FROM node ID assigned by RivEX to the start node of the polyline. Used to build and extract the topological information of the network.
Tnode	Long	The TO node ID assigned by RivEX to the end node of the polyline. Used to build and extract the topological information of the network.
CatchID	Long	The catchment ID the polyline belongs to. This is built by RivEX. A catchment are all the connected lines that drain to a single mouth. Mouth is typically the tidal limit of the network but can be internal drainage networks that do not connect to the wider river basin network.
Dist2Mth	Double	The distance to the network mouth in metres. This is built by RivEX. Distance includes the length of the polyline itself.
US_Accum	Double	The sum of the upstream polyline lengths that flow into the polyline. This is built by RivEX. Distance includes the length of the polyline itself.
Shreve	Long	The Shreve order. This is built by RivEX. Shreve stream ordering orders the network by starting at the source and travels down the network incrementing the order of the polyline and can be thought of as a count of the number of upstream sources for the polyline.
Strahler	Short	The Strahler order. This is built by RivEX. The Strahler number or Horton–Strahler number of a mathematical tree is a numerical measure of its branching complexity. These numbers were first developed in hydrology, as a way of measuring the complexity of rivers and streams, by Robert E. Horton and Arthur Newell Strahler. The Strahler Order was assigned by RivEX using the algorithm created by Gleyzer et al. (2004).

Segment	Long	Strahler Segments are unique numbers for sections of the river corresponding to a single Strahler stream order. This field is created by RivEX during the Strahler Ordering process.
SourceID	Long	The source ID is the ID number of the node that is the source of a river. This is built by RivEX. The source is defined as the furthest point upstream from a polyline.
Hack	Long	The Hack stream order is an alternative method for assigning a hierarchy to the river network. Starting at the mouth of the river the route to source is labelled 1. All tributaries and their routes to source are labelled 2, all tributaries of these and their routes to source are labelled 3 and so on. This is built by RivEX.
RBD_ID	Short	The WFD River Basin District code that the water course is within. This was manually added after the editing phase. This value ranges from 1 to 12. A Geodatabase domain has been assigned to this field to link the river basin name to the code.
Country	Text	ISO country codes[ⱡ] used to identify which country the water course is within. Country code was assigned using a set of rules rather than a geometric intersection of boundary data with network. This was done to avoid cutting up the network, so topology is maintained. If a polyline fell entirely within a country, it was given the country code, cross border lines were assigned to the country they predominantly fell within and polylines that tracked along borders were assigned both nation codes.

ⱡ = https://www.gov.uk/government/publications/open-standards-for-government/country-codes

Development of point-based attributes

With the river network topologically corrected it is possible to create a dataset of sampling points aligned to the river network using RivEX (Figure 2). The regularly spaced points were used to sample the network and secondary datasets. This point dataset provides powerful analysis of the network through querying the point(s) of interest. You can gain instant insight into location, geology, slope, distances, and elevation by simply clicking on the point. There is no need for complex and time-consuming processing requiring additional advance GIS tools. Such attribution can aid in the characterization of river networks: their ecology, geomorphology, and hydrology.

A sampling distance of 100m created a dataset of 1,472,056 points across the entire network. A variety of attributes encoded into the ORN were transferred to the sampling points through spatial joins, elevation and geology were transferred from their respective source.

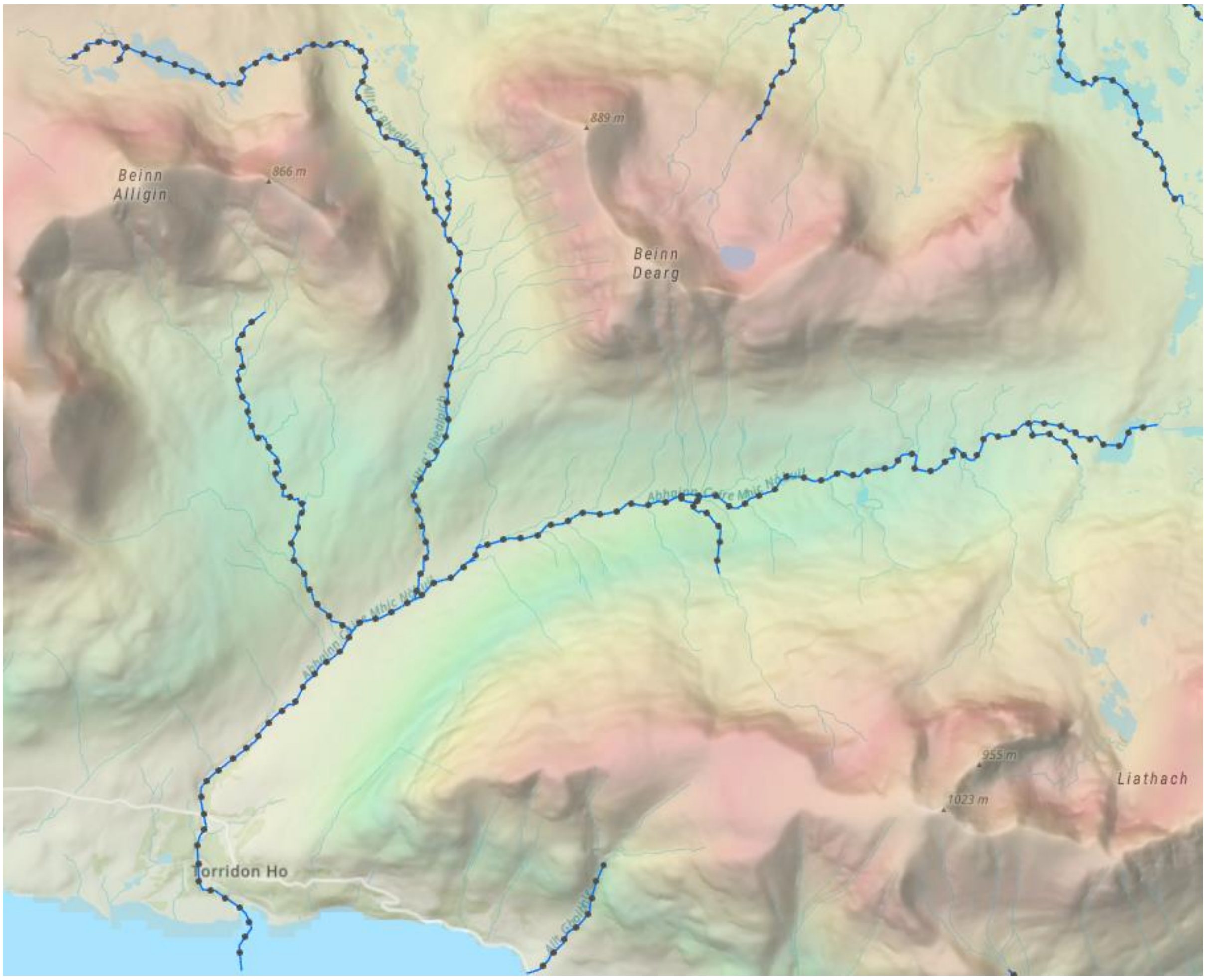


Figure 2 An example of the 100m spaced sampling points along the Open River Network. Each point is attributed with 23 catchment/site derived values. Contains OS data © Crown Copyright [and database right] (2023).

Most attributes were a simple spatial join operation, the exception being slope which was an iterative process to reduce the number of negative slope values created by spatial alignment issues. RivEX can generate reach lines centred on the sampling point. The initial length of the reach was scaled by Strahler order (Table 4) and clipped to the source ID encoded into the ORN. With the reach created, end points were extracted and used to sample the OS Terrain 50 elevation dataset and slope was computed over the reach length. Initial runs identified negative slopes, these negative reaches were extended by 500m and re-run. The remaining negative slopes had their reaches extended a further 2 more times. It was decided to stop at the fourth iteration as such negative errors were the product of gross DEM errors or spatial alignment issues which would never be resolved without significant editing in the base network and terrain datasets. Remaining negative slopes were recoded to NULL and account for 0.8% of all sampling points.

Table 4 Initial reach length of a sampling point was scaled by the Strahler Order of the river channel.

Strahler Order	Initial reach length (m)
1	1000
2	1000
3	1000
4	1000
5	1500
6	2000
7	2500

The sampling points were attributed with 23 values, Table 5 lists the fields in this release of the data.

Table 5 Fields in the sampling point dataset.

Field name	Field Type	Description
OBJECTID	Object ID	A unique row ID maintained by the Geodatabase.
Shape	Geometry	The point geometry field, has no Z or M values.
SampleID	Long	A unique sequential integer assigned to the point as it is created by RivEX. This field is used extensively by RivEX during processing so data can be joined back to this dataset.
PolylineID	Long	The ID of the polyline that the sampling point was created along. This is the RivID field in the ORN.
X_Coord	Double	The X coordinate of the point. Coordinates are in British National Grid.
Y_Coord	Double	The Y coordinate of the point. Coordinates are in British National Grid.
LineLength	Double	The length of the polyline that the sampling point was created along. Units are in metres.
Per_along	Double	The distance along the polyline the sampling point is created at expressed as a percentage of the line length.
Dist_along	Double	The distance along the polyline the sampling point is created at. Units are in metres.
Dist2Mouth	Double	The distance from the network mouth encoded into ORN was transferred to the sampling point using the RivEX transfer network metrics tool. This tool adjusts the distance based upon the position of the point along the river line segment to reflect the true distance from the network mouth. Units are in Kilometres.
CatchID	Long	The catchment ID encoded into ORN was transferred to the sampling point using the RivEX transfer network metrics tool. No importance should be assigned to the actual ID number, they were sequentially created by RivEX when the network was attributed with catchment ID.
RBD_ID	Short	The WFD river basin district code that the water course is within. This value ranges from 1 to 12. A Geodatabase domain has been assigned to this field to link the river basin name to the code.
Country	Text	ISO country codes[†] used to identify which country the water course is within.
Strahler	Short	The Strahler order of the channel. The Strahler Order encoded into ORN was transferred to the sampling point using the RivEX transfer network metrics tool.
SourceID	Long	The source ID is the ID number of the node that is the source of a river. The source is defined as the furthest point upstream from a polyline. The Source ID encoded into ORN was transferred to the sampling point using the RivEX transfer network metrics tool.
OS_T50_Ele	Float	The elevation in metres extracted from the open dataset OS Terrain 50[††]. This was transferred to the point using the ArcPro tool Extract Multi Values to Point. As the river network extends out into estuaries/bays and parts of it in low lying areas, it is possible for the elevation to have a negative value as encoded into the OS Terrain 50 dataset.
Site2Src	Double	The distance the sampling point is from its source. The source is defined as the furthest point upstream from a polyline. This is a value that is computed using the RivEX tool Find Sources and the resulting table joined to this dataset.

		NULL values exist in this field and are the result of conflicting topology of catchments being linked across watersheds, these have occurred in low lying areas. Units are in metres.
OS_T50_SrcEle	Float	The elevation of the source of the river in metres extracted from the open dataset OS Terrain 50[¶]. The source is defined as the furthest point upstream from a polyline. This was transferred to the point using the ArcPro tool Extract Multi Values to Point. As the river network flows through low lying areas, it is possible for the source elevation to have a negative value as encoded into the OS Terrain 50 dataset. For example, these could be spurs flowing out of creeks in estuarine sections.
ReachLen	Long	The reach length to build for the sampling point. This was initially scaled by Strahler Order (Table 4). During processing and in an attempt to resolve negative slope values the reach length was incrementally increased by 500m and the final length used is recorded here. This initial reach length is half the desired length. For example, the initial desired reach length for a Strahler order 1 stream was 1Km. Thus, the recorded reach length is 500m for upstream and downstream directions. This desired length may not be achieved due to truncation, so the final reach length created to compute metrics such as sinuosity and slope is recorded in the field AchievedReachLen.
Truncated	Text	This flag field is created by RivEX during the process of creating a reach to compute slope/sinuosity. Reach length is scaled initially by Strahler order but may have been additionally extended to try and overcome negative values. This is the desired length and recorded in ReachLen. A reach can fail to achieve this desired length because the sampling point is near a source, mouth or a downstream tributary junction where a change in source ID along the network has occurred. In this case the reach is clipped to these limits and recorded as truncated. A worst-case scenario could very easily be a first order stream at the coastline which is less than 500m long. A sampling point in this case would be too near its source AND mouth and the reach is clipped at both ends. Truncation is not an error, but simply a scenario where RivEX was unable to build the reach to the desired length.
Sinuosity	Double	Sinuosity is computed from the reach that was built for the sampling point. A reach may not have achieved its desired length thus sinuosity would be calculated over a shorter distance. Sinuosity is computed as reach length / straight line distance of reach end points.
AchievedReachLen	Double	This is the length in metres of the reach. The desired reach length is twice the length of the value recorded in the field ReachLen. This can fail to be achieved due to the sampling points' proximity to source, mouth or downstream tributary junction with a change in source ID and the reach is

		clipped back. Thus, this field records the True reach length that slope and sinuosity were calculated over.
Slope	Double	RivEX can extract the end points during the creation step of the reach. These points are then assigned an elevation from the OS Terrain 50ᵻᵻ data using the Extract Multi Values to Points tool. The difference in elevation can then be computed and finally the slope calculated over the reach length. Elevation difference is Upstream elevation minus Downstream elevation. This value is then divided by reach length. Units are in m/m
LEX_RCS_BEDROCK	Text	Bedrock 625K BGS geologyᵻᵻᵻ. Generalised digital geological map data based on BGS's published poster maps of the UK (north and south). Bedrock-related themes were created by generalisation of 1:50 000 data to make the 2007 fifth edition bedrock geology map.
LEX_RCS_SUPERFICIAL	Text	Superficial 625K BGS geologyᵻᵻᵻ. Superficial geology-related themes were digitised from the 1977 first edition Quaternary map (north and south).

ᵻ = https://www.gov.uk/government/publications/open-standards-for-government/country-codes
ᵻᵻ = https://osdatahub.os.uk/downloads/open/Terrain50
ᵻᵻᵻ = https://www.bgs.ac.uk/datasets/bgs-geology-625k-digmapgb/

The reach length influences slope and sinuosity values. The desired reach length (as scaled by Strahler order and recorded in ReachLen field) is modified by the sampling points proximity to a mouth\source\tributary junction. To understand the truncation that can occur, Figure 3 shows a reach which has been truncated in a downstream direction at the tributary junction where a change in source ID had occurred.

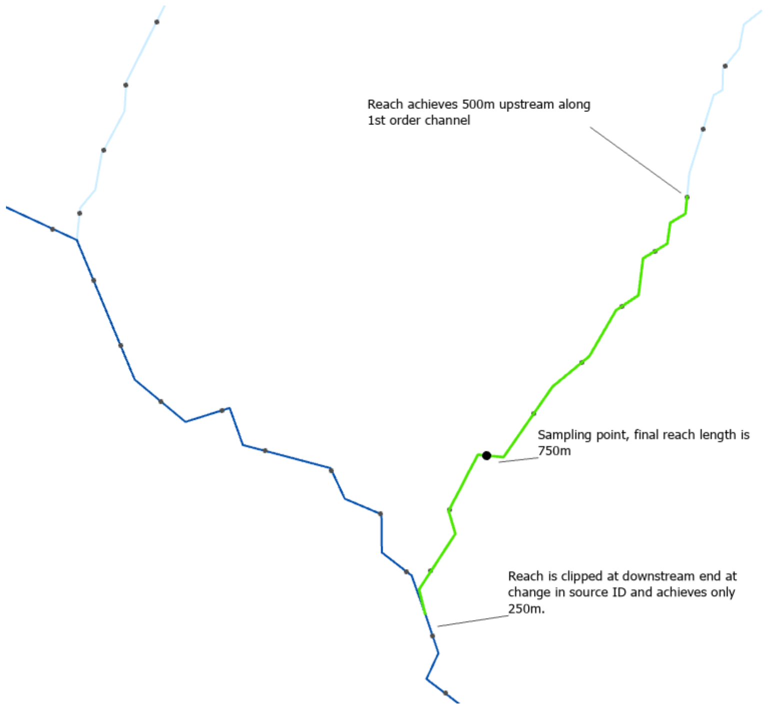


Figure 3 Impact of proximity of sampling point to tributary junction on reach length (green line). In this example the downstream end on this first order channel has not achieved the required 500m as it was clipped back at change of source ID. Thus, the reach length for this sampling point is 750m instead of the desired 1Km. As sampling points get nearer to their

source they will be truncated in an upstream direction. On short lengths of channel, it's quite possible for truncation to simultaneously occur in an upstream and downstream direction resulting in a short reach.

Not all points have logical values for a field; these are the result of spatial alignment issues, topology, or values in secondary datasets (Table 6).

Table 6 Sources of illogical values found within the point dataset.

Field	Illogical value(s)	Reason	Number of illogical values (% of entire dataset)
OS_T50_Ele	Negative elevations	A point can be assigned a negative elevation because it intersects a cell in the OS Terrain 50 dataset which has a negative value. These typically occur in estuarine regions where the ORN extends out into the estuary and the OS Terrain 50 dataset records a negative value. Negative values can occur inland, specifically low-lying areas where OS Terrain 50 records negative elevation.	50,708 (3.4%)
Site2Src	<Null>	The sampling point is on a section of network where the catchment has multiple mouths. Depending on which mouth you are draining to means you have a different source. RivEX is unable to resolve this topological scenario without the network being radically altered.	91 (0.006%)
OS_T50_SrcEle	Negative elevations	Negative values can occur inland, specifically low-lying areas or coastal regions (e.g. salt marshes) where OS Terrain 50 records negative elevation. This introduces the illogical scenario of the source of a river having a negative elevation. These channels tend to be small and often drainage ditches or short spurs off a main channel.	2,787 (0.18%)
Truncated	<Null>	An error caused by a precision issue in storing the XY coordinate resulting in the snapping of the point to a node and RivEX failing to generate a reach.	3 (0.0002%)
Sinuosity	<Null>	An error caused by a precision issue in storing the XY coordinate resulting in the snapping of the point to a node and RivEX failing to generate a reach.	3 (0.0002%)
AchievedReachLen	<Null>	An error caused by a precision issue in storing the XY coordinate resulting in the snapping of the point to a node and RivEX failing to generate a reach.	3 (0.0002%)
Slope	<Null>	Due to spatial alignment issues and gross errors in the DEM, negative values in slope are generated. These were recoded to Null.	12,331 (0.8%)

Metadata

The final stage of data preparation was to use ArcPro to add metadata to the dataset. General summary, keywords, field descriptions, processing lineage and contact details were added a full ISO 19115 and 19139 metadata record to describe an item that is also INSPIRE compliant.

Results

The original OS Open Rivers network that was accessed on 9 February 2023 contained 190,069 polylines covering a network length of 152,759 km.

After all topological errors were resolved the ORN contains 183,349 polylines covering a network length of 147,387 km. 2,751 km of network in the OS Open Rivers dataset have been attributed with a '*form*' attribute of canal, the bulk of this having been removed from the ORN as they break or "short-circuit" the dendritic nature of the river network.

With the network logically attributed, this allows for a variety of statistics to be generated. For example, within Great Britain, the ORN records the source of the River Churn (a sub-catchment within the Thames basin) as the furthest point away from the mouth of its catchment, a distance of 360 km.

Figure 4 identifies length of ORN within each country, this is further broken down into WFD river basins (Figure 5)

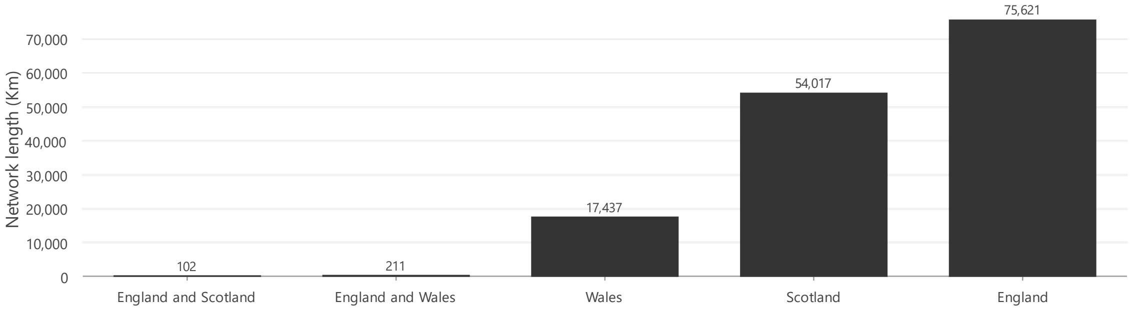


Figure 4 Length of ORN (km) within each country. 313 km of have been identified as rivers that track along the border of two nations.

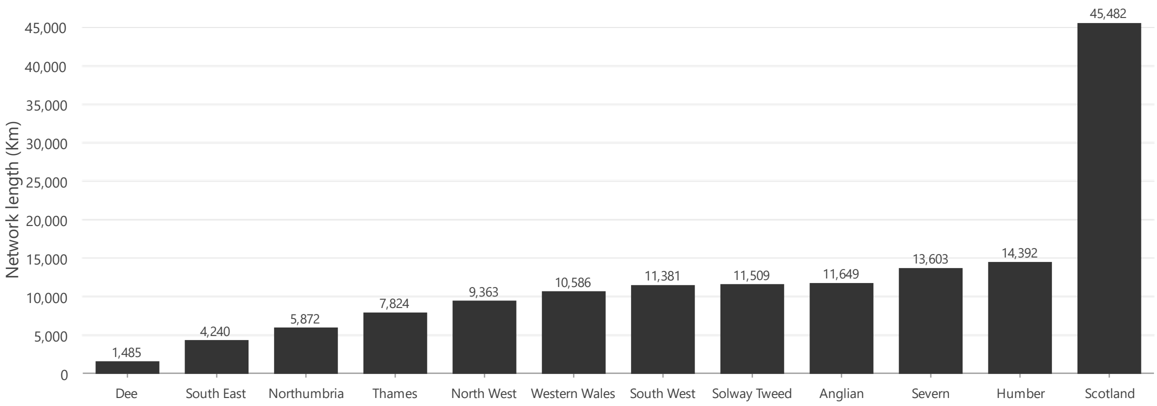


Figure 5 Length of ORN (km) within each WFD river basin.

Across the entirety of the ORN the largest Strahler order achieved was 7 on 4 rivers: River Severn, River Clyde, River Ouse and the River Great Ouse. Figure 6 plots the total length of river network by Strahler Order. First order streams account for 47% of the river network. Figure 7 plots the ORN colour coded by Strahler Order. It should be noted that these stream order data are very much dependent on the extent of the headwaters and tributaries captured in the dataset, such that it is partially a product of the line generalisations used.

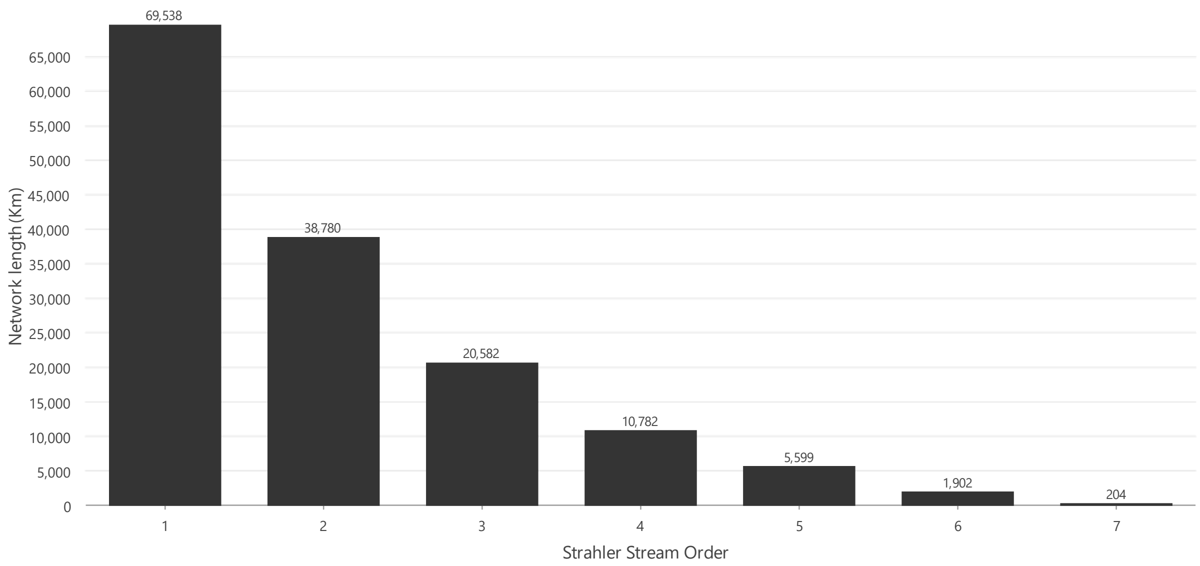


Figure 6 Length of ORN (Km) by Strahler Order.

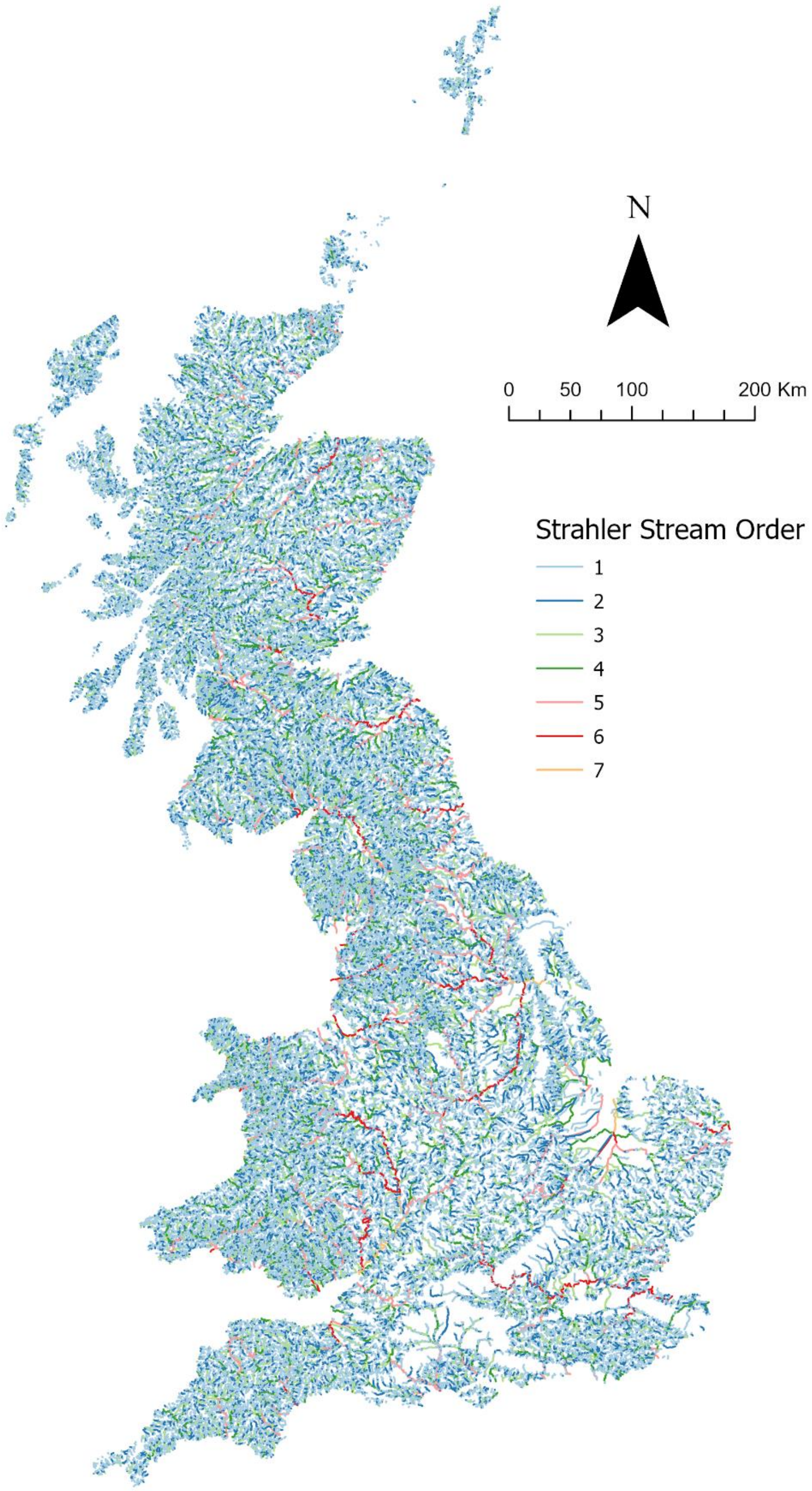


Figure 7 The ORN colour coded by Strahler Order.

The examples outlined above demonstrate the value of attributes built into the network and are only achievable because the network is topologically correct and thus “Analysis Ready”. A series of decisions have been made to generate the network that abstract away from a 'real world network'. For example, dropping loops simplifies all sorts of analysis, especially studies looking at barriers to fish migration. For that reason alone, it is a strong reason to exclude loops from the topological dataset. However, other applications may what these features in the network. The decision to maintain loops may allow them to be managed as a separate layer, which users can optionally merge back into the dataset. At present the dataset excludes loops. Alternatively the loops could be merged them back into the ORN and flag as participating in a loop, but merging them in would fundamentally change the topology of the network, which would require the re-computation of all RivEX fields, introduce topological oddities in the attribution. For example, you can travel up one side of a loop and arrive before the other side achieves the same distance, that's because one side is more sinuous than the other.

Next steps

There are a series of enhancements that can be considered, based on the experiences of developing the connected network and further discussions with users, although some of these options have quite wide re-processing implications. The potential developments include a mix of further processing, potential extensions to the network data and further attributions of both the vector network and the point-based dataset:

I. **Enhance the representation of the OS Open River sinuosity by use of natural algorithms-based data thinning.** The existing line generalisation of the OS Open Rivers dataset has been achieved by a simple vertex thinning processing. A more sophisticated approach might be adopted to present a more naturalistic river network. This would need to work on subsequent releases of the OS Open Rivers dataset. This would allow a more realistic representation of the lines and attribute calculation/derivation.

II. **River naming:** whilst the dataset contains watercourse names and alternate names where these were in the original OS Open Rivers dataset there are reaches that are not named. Addition of the naming to provide a comprehensive naming of the reaches and tributaries will add local relevance.

III. **Addition of headwater streams and additional branches** and re-run the processing of stream orders. The current dataset is very much the product of the resolution of the existing OS Open Rivers dataset and many of the tributaries and headwaters of the river systems are not captured in this process. This is a scale-related choice, but at least for individual catchments it would be useful to have a more finely resolved network that incorporated more of the network and potentially attributed these sections as 'headwaters'. Such an approach would need access to an open-source dataset from which to add these river reaches and would have implications for the subsequent derivation of the indices, such as distance to source and stream orders.

IV. **Loops in rivers have been excluded** from this version of the ORN, although there was some debate as to whether this is effective / appropriate. This is an issue of suitability for which purpose. As it is, it is a simplified, single channel version of the network. Dropping loops simplifies all sorts of analysis, especially studies looking at barriers to fish migration. For that reason alone, there may be a strong reason to exclude the loops from the dataset. However, someone using the network may want loops retained. It would be feasible to provide them as a separate layer, which users can optionally merge back into the dataset. Alternatively, loops could be merged back into the ORN (to create another dataset) and flag as participating in a loop, recognising that the topology of the network is changed. This would also require the re-computation of all RivEX fields, and these are likely to introduce topological oddities in the attribution.

V. **Additional attribution with morphological parameters.** A suite of additional morphological attributes of river morphological condition were conceived within the original ORN model. It has not been feasible to add these at this stage as they would require additional processing, such as widths (min/med/max) and the use of models to predict specific attributes and geomorphological indices (e.g. Qmed, stream power, specific stream power, shear stress, river restoration strategy based on stream power and including natural recovery, assisted natural recovery, designed restoration). The scope for additional processing and network and attribution enhancements will be reviewed and funding will be sought to add priority attributes:

   a) indices of interest to partners and potential users (e.g. stream power, channel substrate size – see Naura *et al.*, 2016);
   b) models (e.g. habitat suitability models for riverine species – see water vole habitat mapping by NRW and Natural England), and
   c) outputs (e.g. maps of agricultural sediment risk of impact on freshwater biota – see Naura *et al.*, 2016; identification of river reaches based on stream power as for MImAS).

VI. **Lakes naming within the vector layer.** The current dataset has generated a network through the lakes to enforce a connected network. Where the existing OS Open Rivers has attributes as 'lake' within the [Form] attribute, these are carried across to the new vector layer. However, some introduced vectors through lakes are not attributed as 'lake' within the network vector data. A route to attributing these links as an open dataset needs to be developed and in consultation with the OS. The cardinal rule of the development of the ORN is not to embed restrictive licence data; the current best approach to adding a lake attribute might be to use the Lake feature form OS MM, but this comes with licence restrictions and dependencies.

VII. **Considering the update strategy aligned to the Ordnance Survey release of OS Open Rivers**. The development of the ORN was undertaken on the latest release of the OS Open Rivers dataset, as at Feb 2023. It was always acknowledged that this would become more out of date as time progressed. This is similar to other networks generated from the authoritative OS mapped data. For example, the CEH River Network was generated c 30 years ago from the 1:50k maps, some of which were old at the time and only minor revisions have been made to the network mapping since The ORN has retained the list of corrections, additions and removals from the network to allow for easier future update based on a release of the OS Open Rivers dataset. Further discussions are needed with OS to explore how a new release of OS Open Rivers might incorporate the correction to generate a new ORN dataset, or perhaps to improve the network fidelity of the OS Open Rivers release.

These suggestions for further work may need to be integrated into a series of combined actions to provide the users with the most effective datasets for their needs. If any changes are made to the version of the network some of the processing would need to be re-run to ensure connected network. It is envisaged that at least some of these steps could be integrated into the OS Open River version releases by OS.

References

ESRI 2023. ArcGIS Desktop: Release 10.8.2. Redlands, CA: Environmental Systems Research Institute.

Gleyzer, A., Denisyuk, M., Rimmer, A. and Salingar, Y. (2004). A Fast Recursive GIS Algorithm for Computing Strahler Stream Order in Braided and Nonbraided Networks. *Journal of the American Water Resources Association*, 40(4), 937 - 946. doi: 10.1111/j.1752-1688.2004.tb01057.x

Hornby, D.D. (2023). RivEX (Version 10.38) [Software]. Available from http://www.rivex.co.uk.

Naura, M., Clark, M.J., Sear, Atkinson, P.M. Hornby, Kemp, P., England, G., Peirson, G., Bromley, C., Carter, M.G. (2016) Mapping habitat indices across river networks using spatial statistical modelling of River Habitat Survey data. *Ecological Indicators*. 66, 20-29. doi:10.1016/j.ecolind.2016.01.019

Naura, M., Hornby, D., Collins, A., Hill, C., Jones, I,J., Naden, P.S, Sear, D. (2016) Mapping the combined risk of agricultural fine sediment input and accumulation for riverine ecosystems across England and Wales.. *Ecological Indicators*, 70, 209-221. doi: 10.1016/j.ecolind.2016.03.055