Link to abstract

A method for hydrometric data rescue: Challenges and solutions for working with archival data

Kate de Smeth[1], Joanne Comer[2], and Conor Murphy[3]

[1] School of Geography, University College Dublin
[2] Hydrometrics Unit, Office of Public Works (OPW)
[3] Irish Climate Analysis and Research Units (ICARUS), Maynooth University

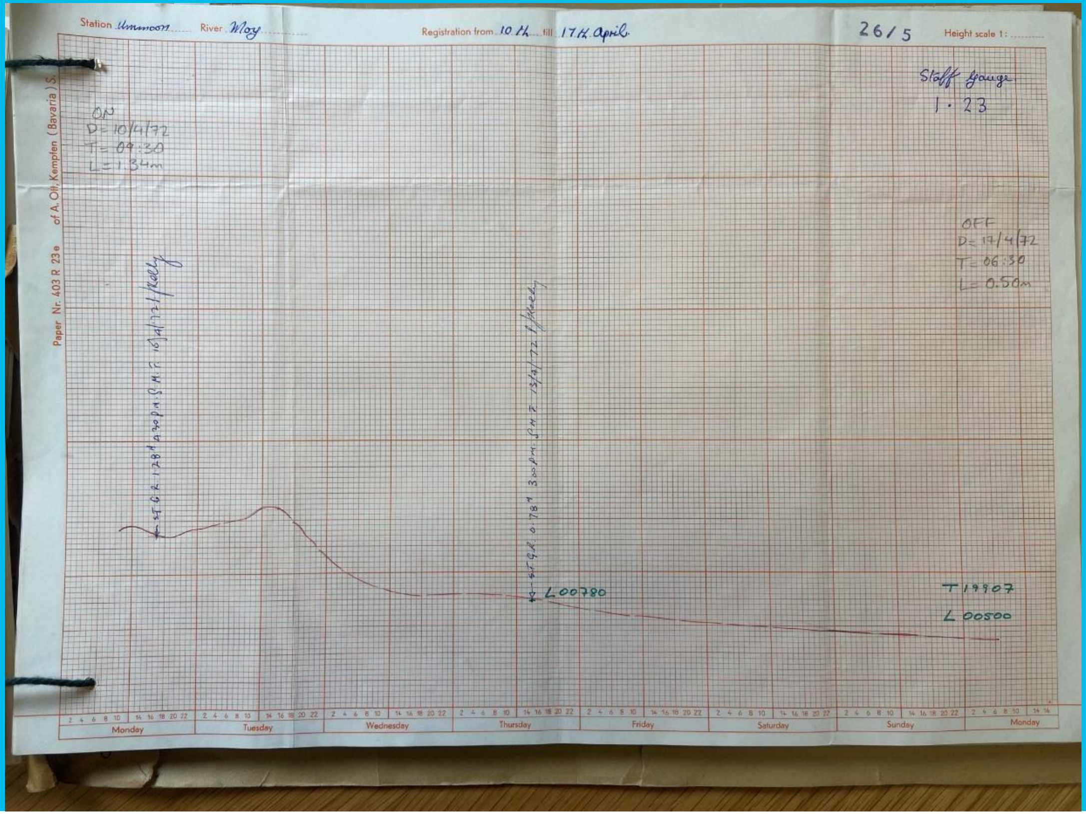


Image: Example of an autographic chart (K. de Smeth, OPW)

This presentation participates in OSPP

Outstanding Student & PhD candidate Presentation contest

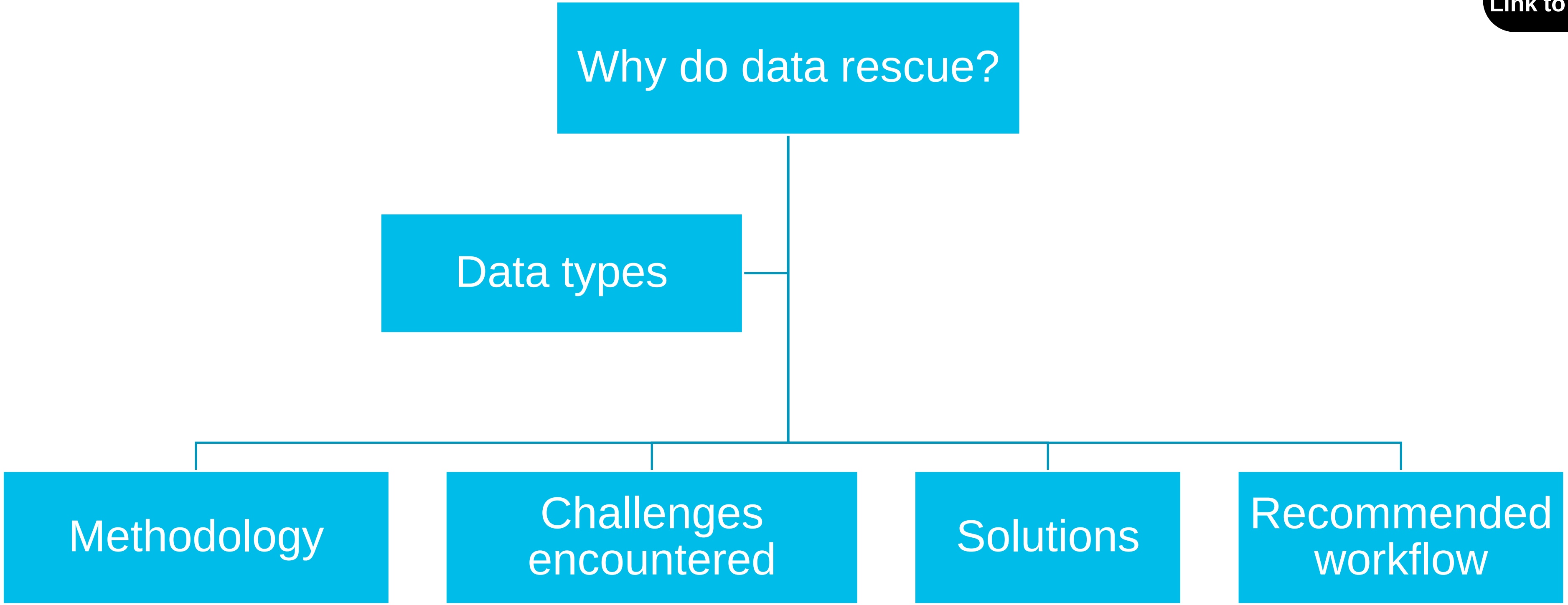
UCD School of Geography

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn

OPW
Oifig na nOibreacha Poiblí
Office of Public Works

Let's discuss

Link to abstract



Ireland example

Contacts + References

Why do data rescue? (1/1)

Motivations for data rescue

- Investigate non-stationary river flow dynamics.
- Validate flow reconstructions.
- Train hydrological models e.g. water resource management, flood dynamics.

→ **Need long, quality assured records**

- Declining hydrological monitoring means a greater reliance on remaining stations where lengthy records are available.

→ **Limited opportunity to obtain such records**

- Poor archiving practices = data deterioration and/or loss

→ **Risk of not rescuing data now is that we might not be able to in the future**

- Data collection was often funded by public money (and still is…).

→ **Obligations to make data publicly available?**

Context for data rescue in hydrology

- Lags efforts in meteorology for:
  - Procedure and workflow development
  - International efforts to facilitate data rescue
  - Volume of rescue work completed

→ **Extensive archival records remain at risk and under-utilised**

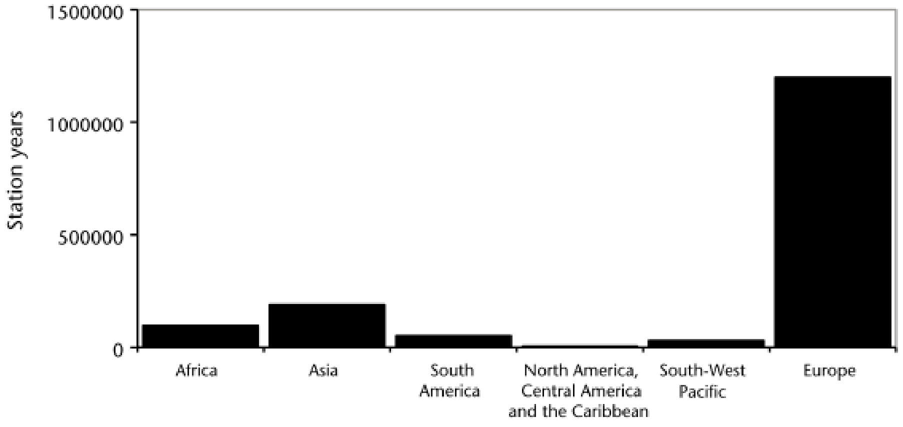


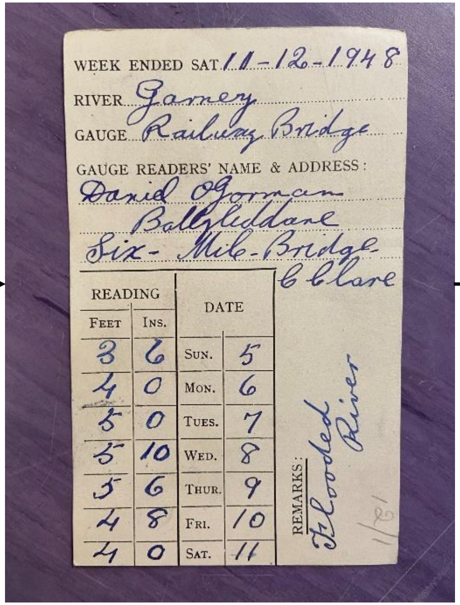
Figure 1. Volumes of data in need of rescue by region

Image: Fry, 2014. WMO-1146. Appendix B.

Data types (1/2)

Staff gauge water level readings

- Daily resolution i.e. “snapshot” data
- Imperial units, later metric
- Occasional checks on staff gauge zero
- Limited to no checks on practitioner responsible for making daily readings



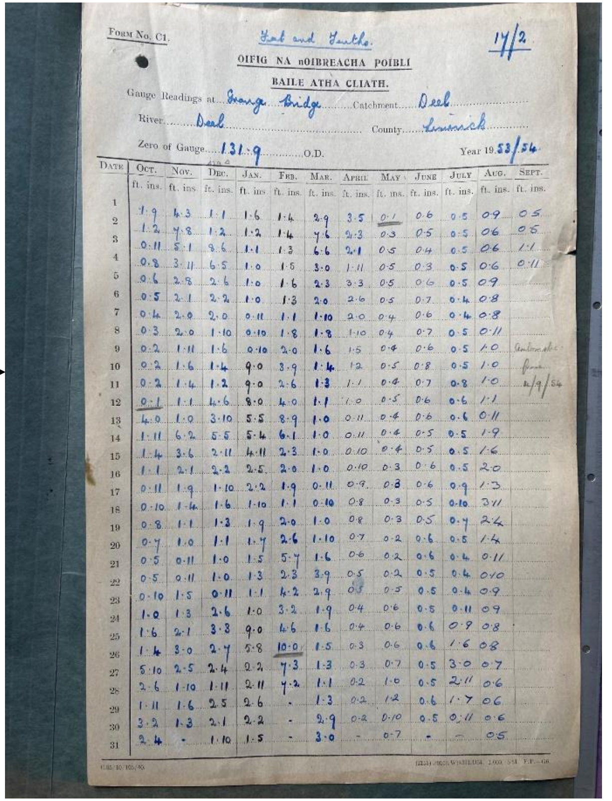


Image: Example of staff gauge and historical staff gauge readings (K. de Smeth, OPW)

Data types (2/2)

Autographic records of water level readings

- Often referred to as “chart” data
- Continuous resolution
- 7 days to one chart
- Imperial units or decimal feet, later metric
- Semi-regular calibration by trained hydrometrics engineer

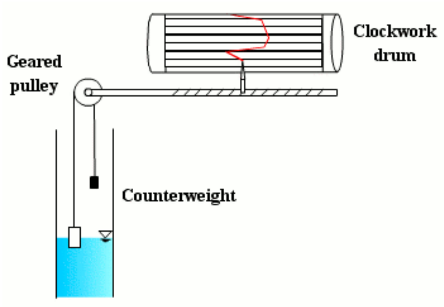


Image: How an autographic recorder works (https://echo2.epfl.ch/VICAIRE/mod_1a/chapt_9/text.htm)

+

=

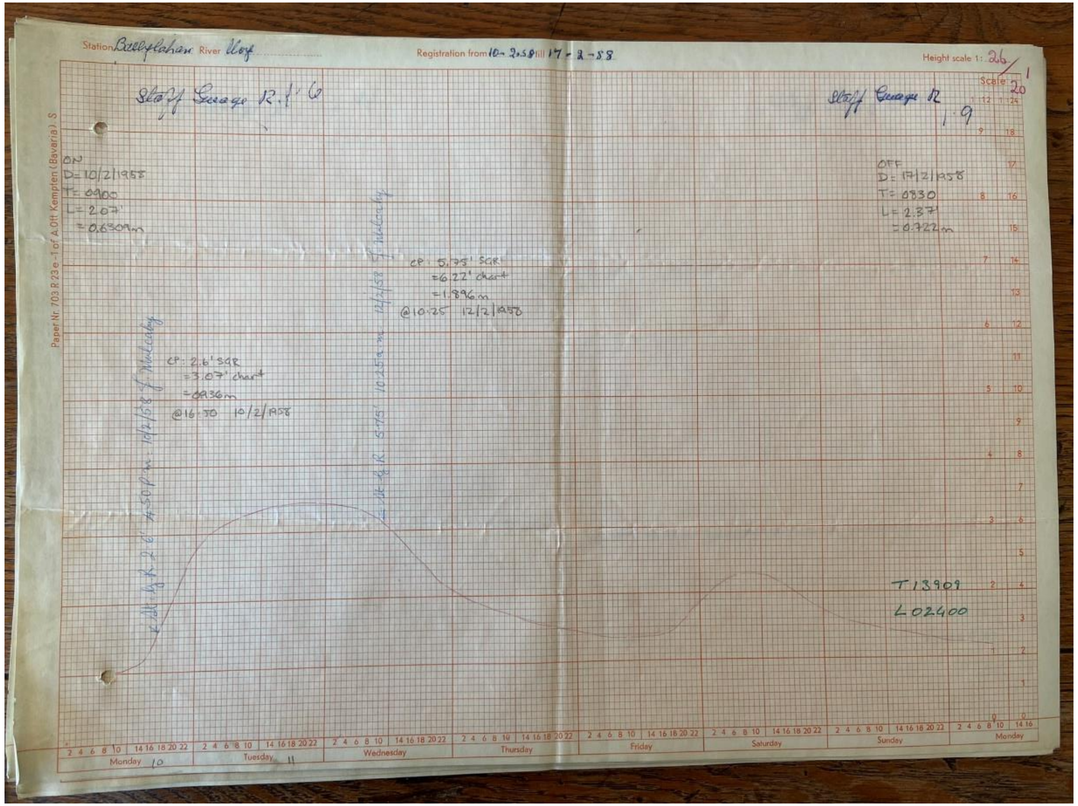


Image: Example of staff gauge and historical autographic chart (K. de Smeth, OPW)

Methodology (1/4)

Guidelines for Hydrological Data Rescue (WMO-1146)

The World Meteorological Organisation released guidelines in 2014 that encourage Members to engage in data rescue by providing generalised guidance and links to further resources.

Topics covered include:

- How to plan a data rescue project and prioritise between records.
- Physical archive management e.g. storage conditions.
- Types and formats of data likely to be encountered.
- Tools to assist with digitisation.
- Data organisation and management, including database storage.
- Possible issues in water level data to look for.
- The link between data rescue and application of their Guide to Hydrological Practices (WMO-168) i.e. that modern data processes methods and standards are also appropriate for data rescue work.
- Metadata.

Also presented the results of a Member State survey about possible extent of, and constraints regarding, historical data rescue of national records.

This got us thinking...

What kinds of data quality issues do we have here in Ireland?

How can we apply modern hydrometric practices to historical data?

Does historical data require additional metadata?

Is the workflow really as "simple" as digitise/collect, process and check?

Methodology (2/4)

We designed a three-phase workflow:

Phase 1: Data collection and review

- Identify and source key historical data sets in raw format including water level data and historical rating information.
- Review modern water level and flow data.
- Review historical hydrometric methods and associated metadata including:
  - Datums and units of measurement.
  - Quality assurance processes (if any) that were in place (such as calibration checks and quality coding).
- Determine whether missing information can be sourced or derived and with what level of confidence.

↓

Phase 2: Data processing

- Transcribe and digitise paper records.
- Develop any required historical rating equations.
- Apply historical rating equations to rescued water level data to convert to historical flow data.
- Append rescued water level and flow data to existing available data sets to create extended records.

! Method depends on format of historical data

↓

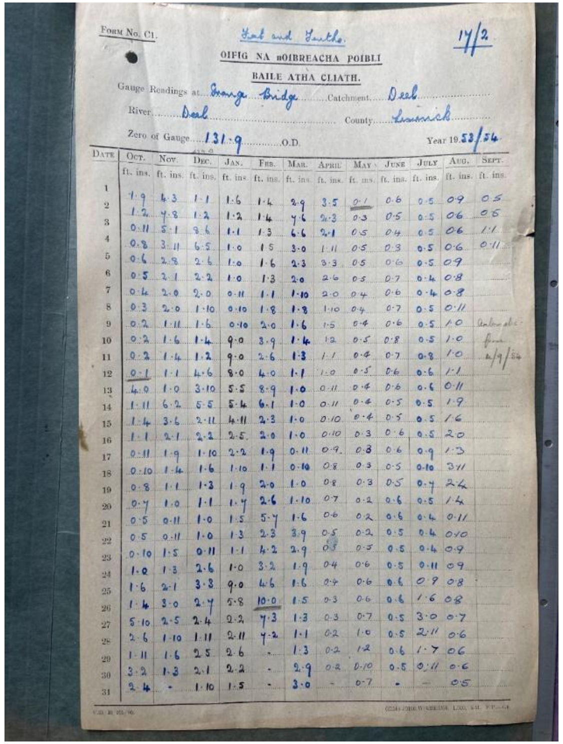
Phase 3: Quality assurance

- Develop and apply quality assurance procedures (e.g. quality coding).

Daily water level readings

i) Transcription to digital format

ii) Unit conversion to metric

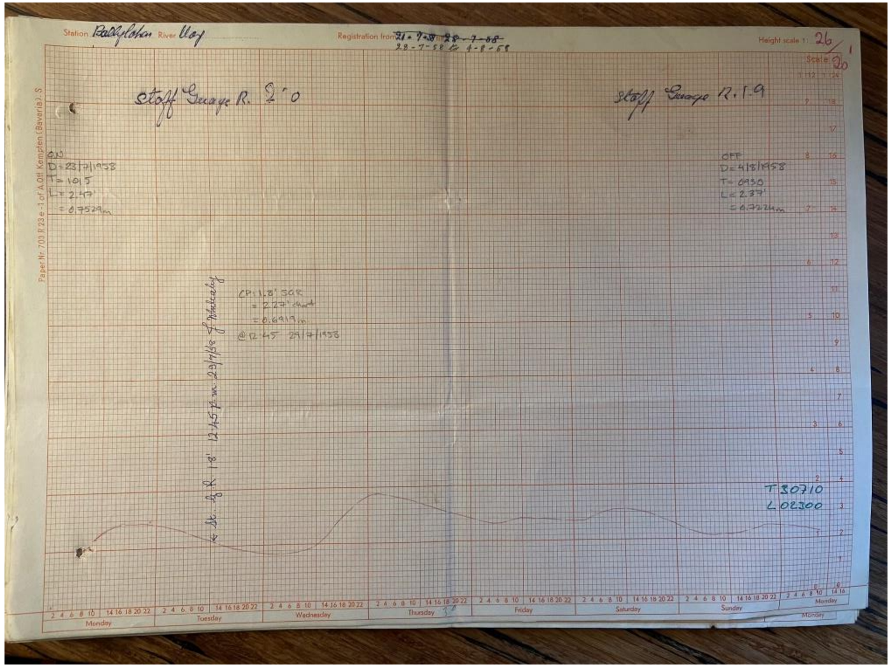


Methodology (3/4)

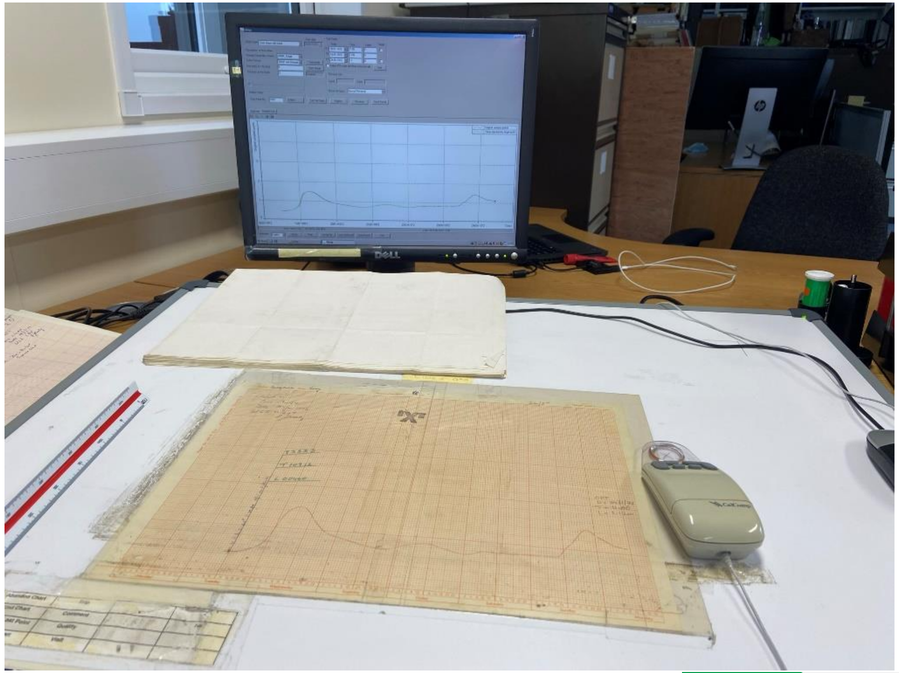
Image: Historical data courtesy of OPW. Digitiser image by K. de Smeth.

Continuous autographic records

1. Preparation of charts via editing process.
   a) Use annotations to check whether water level line is correctly positioned relative to zero line.
   b) Apply any unit conversions (charts digitised as metric series).



2. Digitisation using specialist equipment and software.
   a) Apply any corrections to water level line as determined during editing.



Why rescue? | Data types | Challenges | Solutions | Recommendations | Irish example

Methodology (4/4)

Hydrometric software (WISKI) employed with extended water level series

1. Datums added = Absolute water level series
2. Apply ratings = Flow series
3. Checks, including but not limited to:
    a) Outliers (check against raw data).
    b) Tie-in between historical and existing records
    c) Range of values
    d) Shapes of hydrographs
    e) Gaps
    f) Jumps
4. Aggregated data e.g. to 15 min series or daily mean

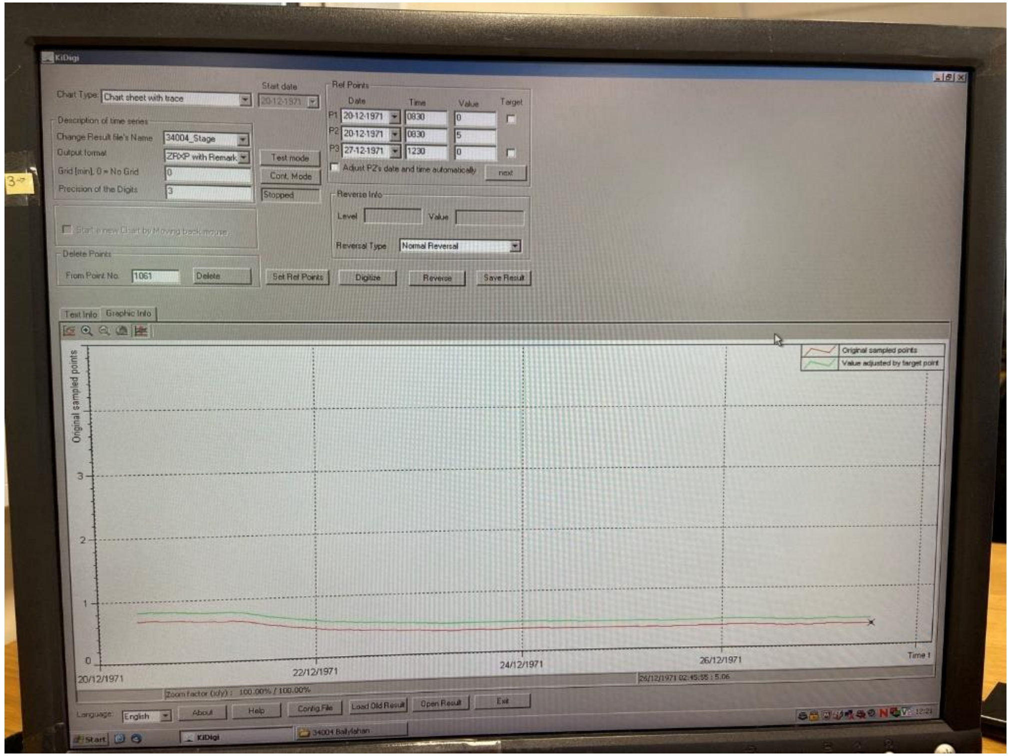


Image: Screen grab of WISKI (K. de Smeth).

Why rescue? | Data types | Challenges | Solutions | Recommendations | Irish example

Challenges (1/3)

We encountered several issues during this process, but by far the most significant were quality issues in the chart records, e.g.:

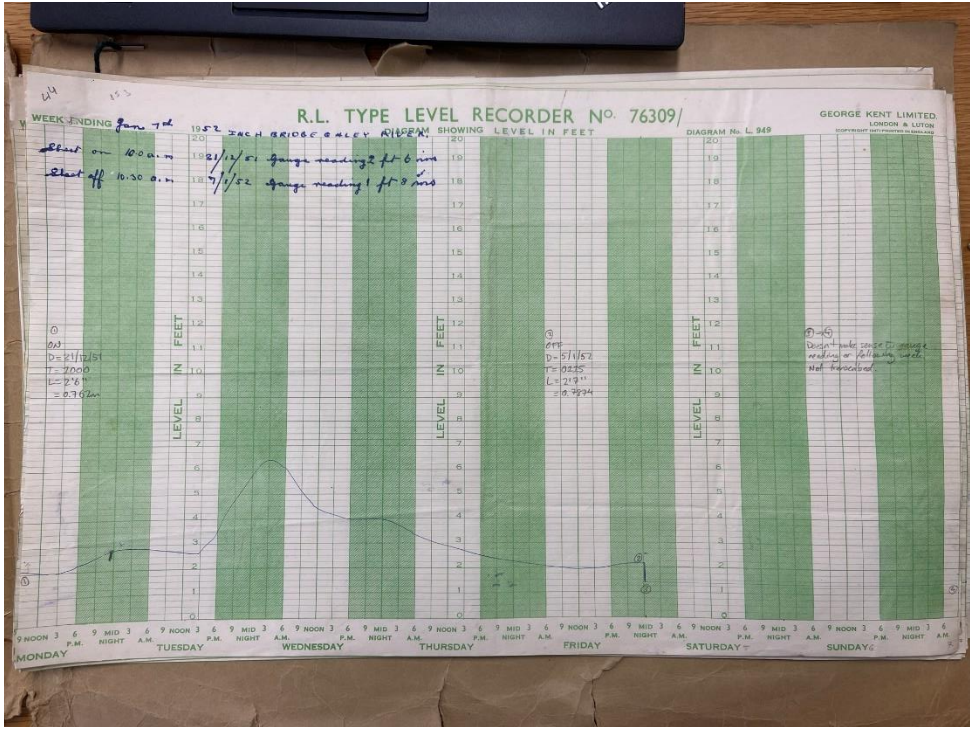
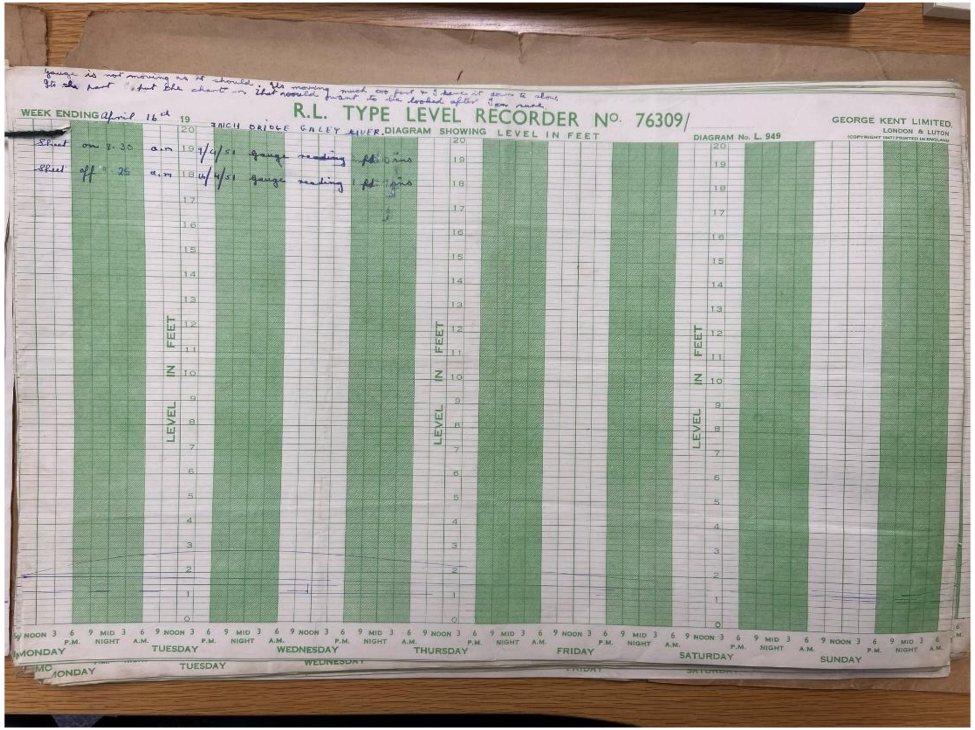


Chart has been put on at the wrong level = chart line needs adjusting upwards during digitising AND it malfunctioned towards the end of the week.



Clock has run fast meaning the drum has over-rotated and multiple lines appear for one week of data.

Why rescue? | Data types | Methodology | Solutions | Recommendations | Irish example

Challenges (2/3)

What kinds of data quality issues do we have here in Ireland?

Issues relating to chart data can be grouped:

Answer!

Hydrometric	**Equipment**	**Other**
Loss of data over time i.e. data archiving issue - Pencil line faded	Clock malfunction - Runs fast	Pump test
Inconsistent datums - Staff gauge zero ≠ Auto-recorder zero	Drum malfunction - Doesn't rotate	
Missing annotations - Local person are missing - No calibration checks for a long time	Pen malfunction - Not pressed down properly - Jumps rather than smooth	
Staff gauge unusable - e.g. silted up	Other malfunction - Gauge out of order, reason unknown - Cannot capture peaks i.e. overbanking	
Inconsistent station management - Leaving charts on for weeks at a time - Not placing chart on correctly		

Challenges (3/3)

As a result of the complexity in the editing process, we were struck by two major challenges:

1. How to deal with these quality issues during data rescue? I.e., during editing.

→ Working with historical data means we often lack additional data, metadata or contextual information to help us problem solve. **How do we work best with what we've got when issues arise?**

→ **How do we minimise subjectivity during data rescue?**

2. How to communicate the level of confidence in the data to the end user?

→ How confident a practitioner feels about the quality of rescued data will be nuanced depending on the type and extent of complexity they encountered during the rescue process. **Modern quality code definitions may be too generalised** to reflect this.

→ **What information, and level of detail, might the end user need?**

Solutions (1/3)

How to deal with these quality issues during data rescue?

= Standard Operating Procedures

For each quality issue, we defined a standard approach to dealing with it during editing (example below).

Table 2: Standard Operating Procedure to resolve data quality issues encountered in chart data (numbered and denoted H = Hydrometric, E = Equipment, O = Other).

Quality issue	Type	Description	Evidence on charts	Standard Operating Procedure during editing	Data rescued?
Loss of data	H	Water level line (pencil or pen) has degraded over time.	Faded water level line that is difficult to see. The indentation from the pen/pencil on the paper chart may be preserved. (H1)	Where the faded pencil line could confidently be restored, it was traced with a red dashed pen line.	Yes
			No or minimal markings are present, chart is essentially blank. (H2)	Weekly chart could not be digitised, data is lost.	No
Missing annotations	H	Chart annotations by local person are partial or missing entirely.	Chart annotations by local person of the start and end staff gauge levels, dates and/or times are partial or missing. (H3)	Missing information was determined from contextual sources e.g. dates, times and levels in the previous and following weeks, and annotations added to the charts.	Yes

Designing and implementing standard procedures aims to ensure consistent data rescue:

- **between individual practitioners on the same project; and**
- **between different projects across time.**

Solutions (2/3)

How to communicate the level of confidence in the data to the end user?

= quality coding

Starting from the modern hydrometric quality coding approach used in Ireland, we redefined the same code levels for the historical data scenarios and linked the chart data error types to these codes.

Table 3: Quality codes assigned to historical water level data (Hx, Ex, refer to the quality issues outlined in Table 2 that may have been encountered).

Code	Symbol	Current OPW application		Application to historical data	
		Name	Description	Name	Description
NA	missing	Missing	Data is missing.	Missing	Data is missing i.e. no staff gauge reading or chart record exists for this day.
31	31	Inspected (Good)	Inspected water level data - Data may contain some error but has been approved for general use.	Good	Data is considered Good quality. The date, time, and start/end water levels are consistent with calibration checks and tie in to the preceding and following week. No corrections were required. Regular calibration checks have ensured the gauge was working properly or, where calibration checks may have been less frequent, there is no evidence of data quality issues and consistent and accurate staff gauge annotations have been made by the local person. (H1, H3, H4, H6)
32	C	Inspected (Good - modified)	As per Code 31, but digitised water level data has been corrected.	Good (modified)	As per Code 31, however some corrections were made during editing due to Hydrometric quality issues. The digitised water level data has been corrected. (H1, H3, H4, H6)

For the final extended data product(s), it is important that historical water level and flow data is quality coded in a way that is consistent and complementary to the modern approach.

Solutions (3/3)

How to communicate the level of confidence in the data to the end user?

= issue all data products with:

i. README file recording station-specific information e.g., metadata, notes on data quality, rating curves, missing data etc

ii. Quality assurance metadata.

Table 4: Example of quality assurance metadata table for Station #26021 Ballymahon.

HY	Chart type	Chart height	#local person notes	#technician notes	#calibration points	Staff gauge zero (SGZ)	Auto-zero (AGZ)	Quality notes	% missing data
26021 Ballymahon *Staff gauge is IMPERIAL and charts are in HYDROMETRIC FEET*									
1956	Hydrometric feet	10'	None	20/06/1957	18/01/1957 15/05/1957 12/06/1957 24/07/1957 05/09/1957 12/09/1957	147.47' OD	147.00' OD	There was only one calibration check between October and April, after which checks became more frequent. Generally correct water levels were determined by working forwards and backwards from calibration checks and comparison with the recorded water level line. Staff gauge annotations were generally not trusted because: i) they implied more variation than the recorded water level line showed and Ballymahon is known for long stable periods of flow (this decision to favour the recorded water level line was agreed with OPW Hydrometrics); and ii) several calibration checks showed a different annotated staff gauge level to the one annotated by the local person on the chart. We suspected that the staff gauge was not being consistently read accurately by the local person. Where there was missing data, or an absence of checks for several months, the start/end water levels were taken from the staff gauge annotations in the absence of other confirming information. If this resulted in a jump in water level at the next calibration check, then generally the error (i.e. the difference) was averaged out over several weeks so the water level tie-in was smooth.	3.56

Provide enough detail that an end user can decide whether the rescued data is suitable for their specific purpose.

The more information they have to make the decision, the better!

Recommendations (1/2)

Generalised workflow for future data rescue projects

How can we apply modern hydrometric practices to historical data?

= Some require modification e.g. quality coding

Is the workflow really as “simple” as digitise/collect, process and check?

= It is not a sequential process. Processing and quality assurance must be concurrent.

Does historical data require additional metadata?

= Yes, due to the complexities of working with autographic data, and to record subjectivity in decision making.

Phase 1: Data collection and review

- Identify and source key historical data sets in raw format e.g. water level data, historical rating information etc.
- Review modern water level and flow data.
- Review historical hydrometric methods and associated metadata including:
  - Datums and units of measurement.
  - Quality assurance processes (if any) that were in place (e.g. calibration checks and quality coding).
  - Temporal resolution (e.g. daily vs continuous).
- Determine whether missing information be sourced or derived and with what level of confidence.
- Identify Equipment, Hydrometric or Other issues and record with photographic examples.
- Understand the potential for subjective decision making during Phase 2:
  - Identify aspects of the data rescue process that will be objective (e.g. application of datums) vs those where the practitioner can influence the outcome (e.g. dealing with identified hydrometric and equipment issues).
  - Reflect on the implications for the quality and usability of the final data product.
  - Brainstorm how subjectivity could be constrained (e.g. standardised operating procedures, quality coding).

Phase 2: Data processing

- Transcribe and digitise paper records implementing the Standard Operating Procedures.
- Develop any required historical rating equations.
- Apply historical rating equations to rescued water level data to convert to historical flow data.
- Agree and implement Standard Operating Procedures to append rescued water level and flow data to existing available data sets:
  - Identify which data set takes precedence where overlap between the series exists.
  - Aggregate data to the desired temporal resolution of the final output.

For example, a continuous record of good quality should take precedence if the final data set will be aggregated to a representative value per time step.

Phase 3: Quality assurance

- Agree and implement Standard Operating Procedures to:
  - Record hydrometric metadata and information collected during Phase 1.
  - Handle the Equipment, Hydrometric or Other quality issues identified in Phase 1.
  - Assign quality codes that reflect relative levels of confidence around data reliability and accuracy building on existing modern hydrometric practices (if applicable).
  - Record relevant metadata about the data rescue process including dates associated with different data set inputs and key decision points.
  - Record relevant metadata about data quality including calibration checks and informative annotations.
  - Compile data quality summaries.
  - Undertake quality assurance checks.

Phase 4: Data provision

Make the final data set(s) available with all necessary information to enable future users to make informed decisions about whether it is fit for their specific purpose.

- Compile historical and/or extended (as appropriate) records of hydrological parameter (water level, flow):
  - Date and time stamp, parameter value and quality code.
- Compile all supporting information:
  - Outline of data rescue methodology for Phases 1 through 3, including standard operating procedures.
  - Phase 2 metadata, including dates associated with each data source input per historical/extended series.
  - Phase 3 quality assurance metadata and quality codes.
- Collate the above per station and make publicly available via an open access platform, along with contact information.

Recommendations (2/2)

General lessons

- Take time to understand the data and the task.
  - Types of data + methods.
  - Quality issues.
  - Where is there potential for subjectivity in the data rescue process.
- Process the data in a standard way.
  - SOPs (collection, QA)
- Communicate metadata to provide critical context for interpretation.
- Relationship between data rescuer(s) and hydrometrics team is crucial given the time and labour requirements for these projects.

Specific to Ireland

- SOPs, including quality coding approach, have been written for future rescue efforts to ensure consistent practice at national level.

Hydrometric data rescue in Ireland (1/4)

Opportunity in Ireland

- Hydrometric monitoring network established in 1939/1940.
- Yet not all data is publicly available.
- Currently data rescue undertaken by the responsible agency (Office of Public Work, OPW) on a case-by-case basis.

→ **This limits change detection and attribution studies e.g., to understand the impacts of major catchment management policies such as arterial drainage.**

As part of PhD research to understand the influence of arterial drainage on the flow regime of affected rivers, we conducted historical data rescue for 8 stations

= extended water level and river flow records to 1939/1940.

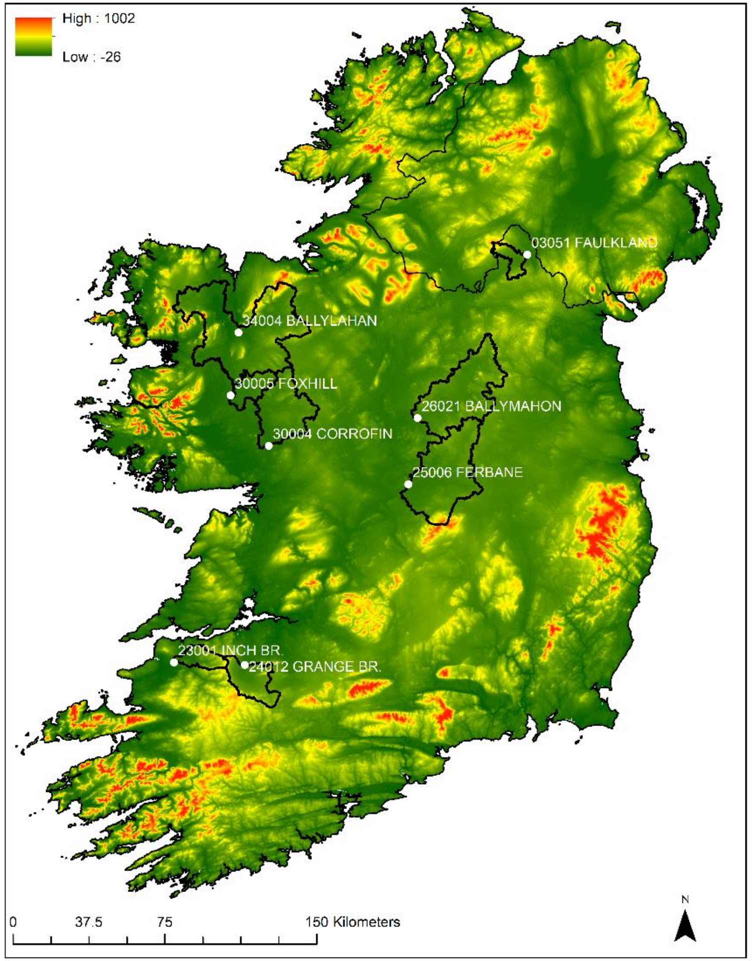


Image: Stations for which historical data rescue was completed (K. de Smeth)

Hydrometric data rescue in Ireland (2/4)

Outputs

- Extended water level and discharge series for Irish hydrometric year (HY) 1939-2020, with quality codes
- Readme text file (metadata + station-specific information about rescued data)
- For stations with rescued chart data, Excel file with summary of overall data quality.

Station #	Station name	Previous available record - start date	Extended record		Data rescued	
			Total period	Years added via data rescue	Staff gauge (daily)	Autographic recorder (continuous)
03051	FAULKLAND	25/03/1975	Dec 1939 - Sept 2021	35	12/12/1939 - 23/04/1977	None
23001	INCH BR.	01/01/1960(WL) 05/06/1972(Q)	Oct 1939 - Sept 2021	20	15/10/1939 - 26/02/1949	10/03/1949 - 31/12/1959
24012	GRANGE BR.	01/10/1954	Jan 1940 - Sept 2021	14	05/01/1940 - 04/09/1954	None
25006	FERBANE	24/10/1947 (WL) 01/01/1952 (Q)	Jan 1940 - Sept 2021	8	01/01/1940 - 26/07/1947	None
26021	BALLYMAHON	01/10/1972	Nov 1939 - Sept 2021	27	28/11/1939 - 02/05/1953 09/10/1965 - 26/11/1966	16/05/1953 - 27/06/1960 17/11/1966 - 30/09/1972
30004	CORROFIN	04/08/1951	Oct 1939 - Sept 2021	11	08/10/1939 - 12/06/1951	None
30005	FOXHILL	14/10/1955	Oct 1939 - Sept 2021	16	01/10/1939 - 10/12/1955	None
34004	BALLYLAHAN	17/05/1974	Oct 1939 - Sept 2021	21	01/10/1939 - 24/11/1951	27/10/1951 - 30/09/1960 01/10/1972 – 16/05/1974

Example extended flow series for 4/8 stations (red = extension to record via data rescue, black = existing available record).

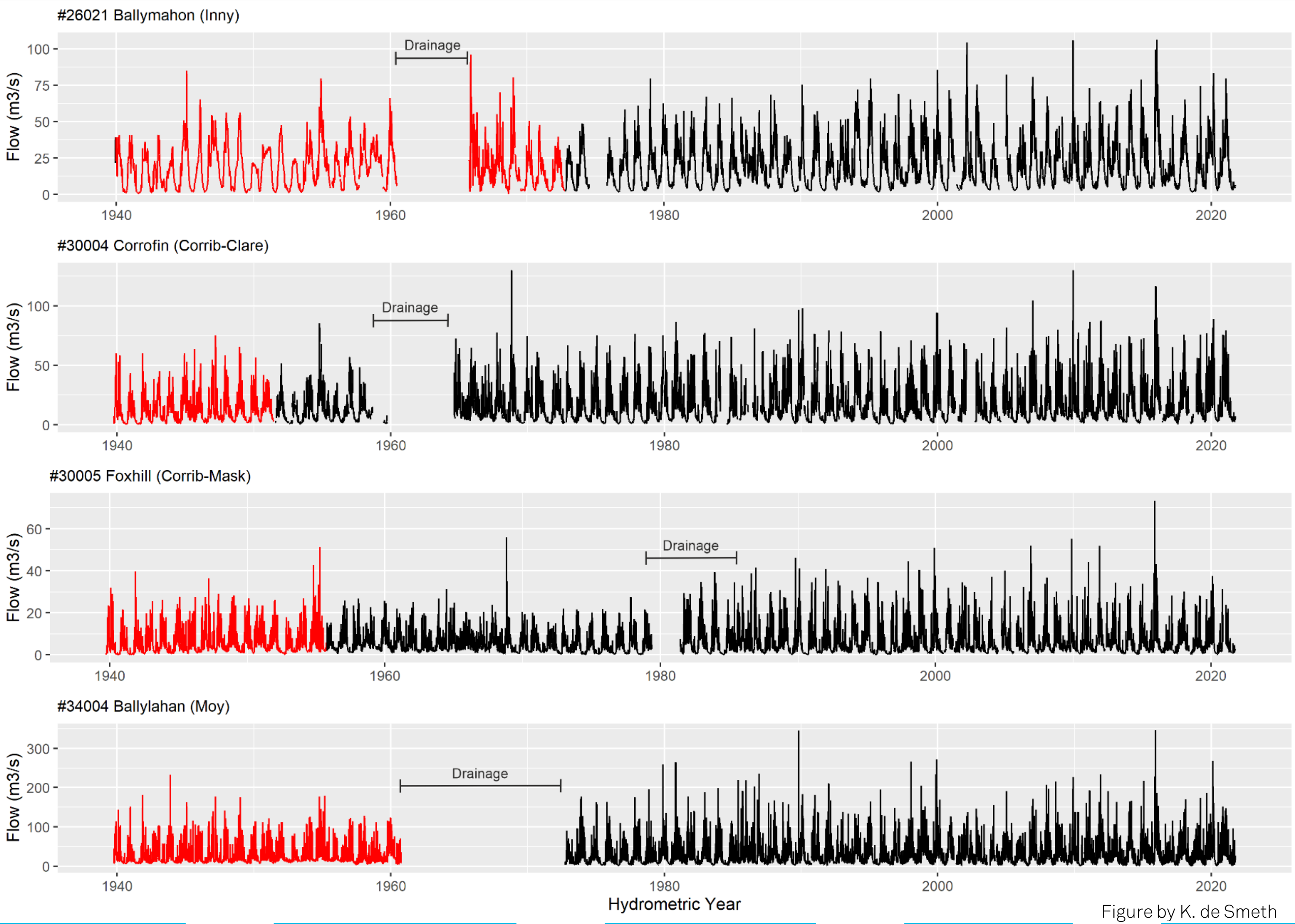


Figure by K. de Smeth

Example quality code visualisation for 4/8 extended records.

When the approach to quality coding rescued data complements that used in the modern hydrometric method, we can create a more complete picture of data quality over time.

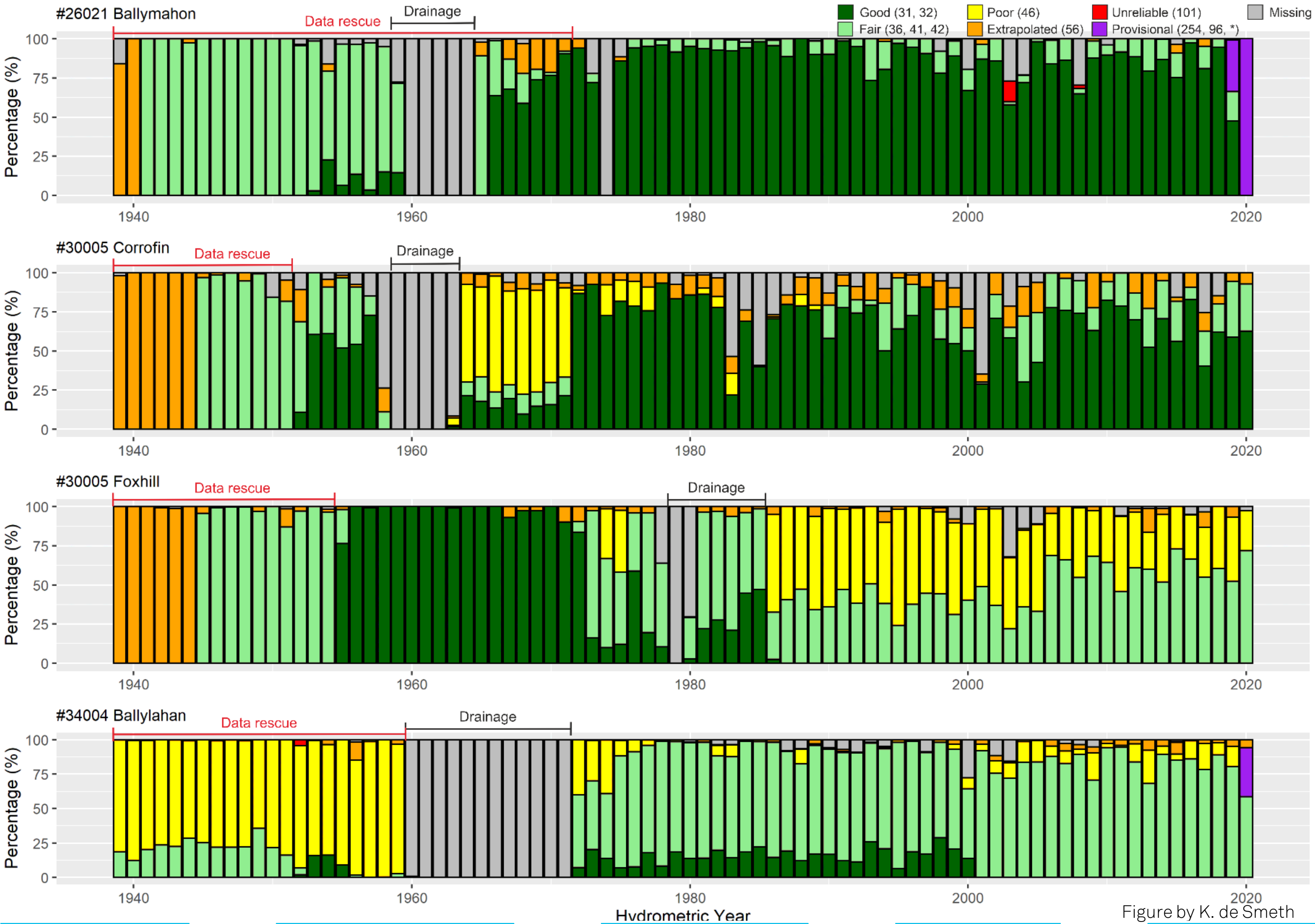


Figure by K. de Smeth

Contacts and References (1/1)

Data + methods:

- Paper accepted and soon to be published by Geoscience Data Journal (open access) titled:

Hydrometric data rescue and extension of river flow records: Method development and application to catchments modified by arterial drainage

- After publication, data will be available via PANGAEA® Data Publisher.

This presentation participates in OSPP

Outstanding Student & PhD candidate Presentation contest

Contact details:

Kate de Smeth kate.desmeth@ucdconnect.ie

Joanne Comer joanne.comer@opw.ie

Conor Murphy conor.murphy@mu.ie

Link to abstract

Funding acknowledgment:

This work was funded by an award from the Irish Research Council (IRC) and Office of Public Works (OPW) as part of the IRC Enterprise Partnership Scheme PhD Programme (EPSPG/2020/438).

References

Fry, M. (2014). Guidelines for hydrological data rescue. World Meteorological Organisation, WMO Technical Report, WMO-No 1146. Geneva, Switzerland.