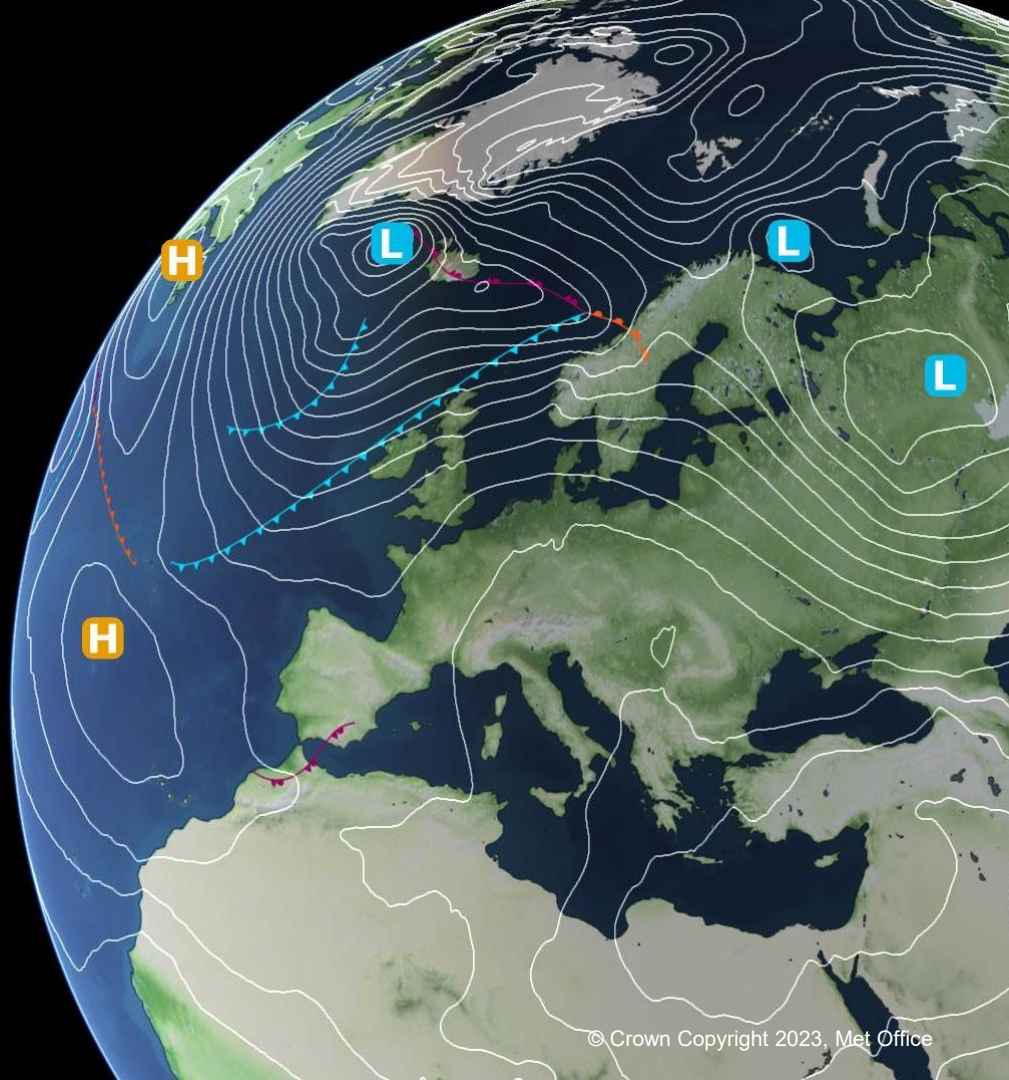
Met Office
Hadley Centre

A new blended rainfall database

Extending climatological UK rainfall series using Rainfall Rescue data



Stephen Packman, *Philip Whybra*
UK Meteorological Office

Background

- The **UK Meteorological Office** contains a wealth of historic rainfall records, extending as far back as the **17th century**, preserved in scanned in weather journals and logs.

- Much of this historic data remains disconnected from the modern station network, resulting in shorter and incomplete rainfall series.

- A new approach is needed to integrate these historic observations with the current network.

Data sources

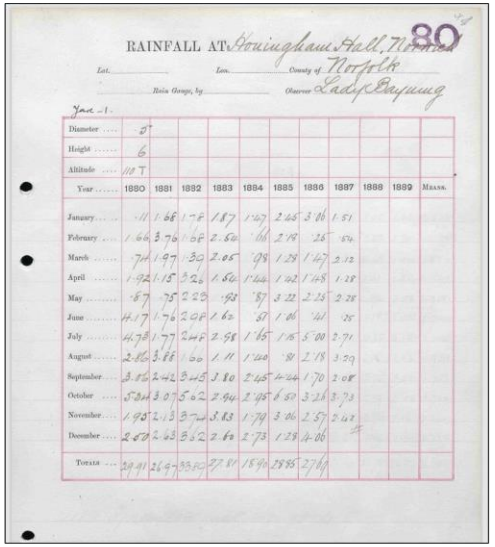
MIDAS Open

An open dataset product containing **hourly/daily observations** from land surface stations in the UK Met Office network back to 1853.

Rainfall Rescue

A citizen science project, led by Professor Ed Hawkins, to digitise historic monthly rainfall data back as far as the 17th century.

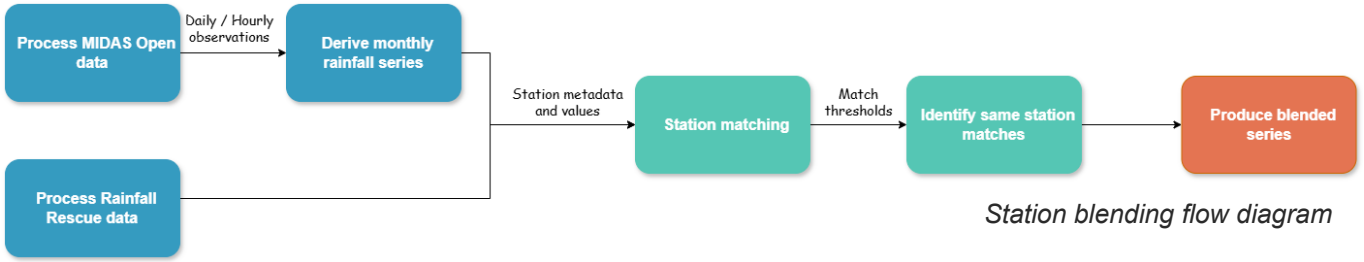
Over **16,000 volunteers** transcribed data from **65,000 archived records.**



Monthly and annual rainfall totals recorded by Lady Bayning at Honingham Hall, Norfolk between 1880-1887

Project Overview

- Develop new database of monthly rainfall, blending **MIDAS Open** and **Rainfall Rescue** data.
- Match rescued rain gauges to existing stations in the UK Meteorological Office land network.
- Produce traceable observation-based monthly rainfall series for use in climate datasets.



Station blending flow diagram

Deriving monthly rainfall series

- A **rule-based priority** system selects the most reliable daily values for each station, prioritising non-suspect observations from primary gauges.
- Monthly rainfall series maintain **full traceability** back to source observations (daily, hourly, 12-hourly).

Source observations

Day	Observation	Rain gauge	Value	Suspect value
2024-03-01	Daily	Primary	15.0 mm	Y
2024-03-01	1st 12-hourly reading	Secondary	2.5 mm	N
2024-03-01	2nd 12-hourly reading	Secondary	2.5 mm	N
2024-03-02	Daily	Primary	0.0 mm	N
2024-03-03	Daily	Primary	2.0 mm	N

Daily

Day	Rain gauge	Value	Priority (daily value)	Suspects
2024-03-01	Secondary	5.0 mm	2 – Two 12 hourly readings	0
2024-03-02	Primary	0.0 mm	1 – Daily reading	0
2024-03-03	Primary	2.0 mm	1 – Daily reading	0

Monthly

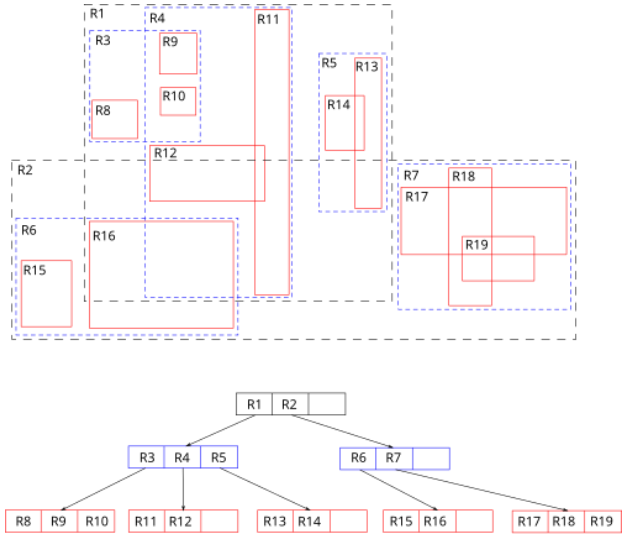
Month	Value	Day count	Missing days	Highest daily value	Day of highest daily value	Suspects
2024-03	7.0mm	3	28	5.0 mm	1st March	0

Station matching

- Rainfall Rescue rain gauges matched to MIDAS Open stations on:
  - **Location**
  - **Monthly rainfall patterns**
  - **Station name similarity**

Location matching

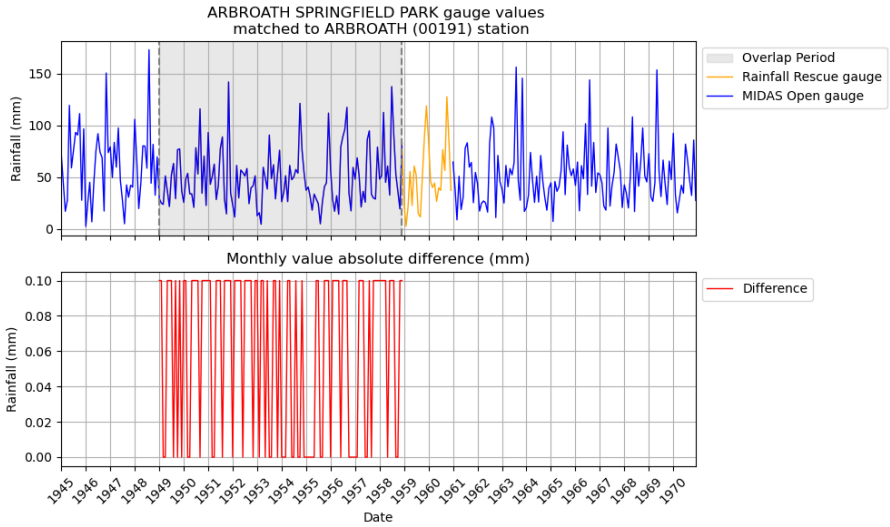
- Efficient location matching process developed utilising **R-tree** data structures.
- R-tree quickly identifies all stations within a specified search radius, avoiding pairwise comparisons.
- For example, with a 500m radius, **over 30 million** potential station pairs were processed in **under 10 seconds.**



Example of an R-tree data structure

Value matching

- Compare monthly rainfall time series, requiring at least **12 months** of overlapping data.
- **Adjustable value tolerance,** e.g. ± 1mm, defines acceptable differences.
- Similarity in rainfall patterns used to identify matches where location information is missing or unreliable.

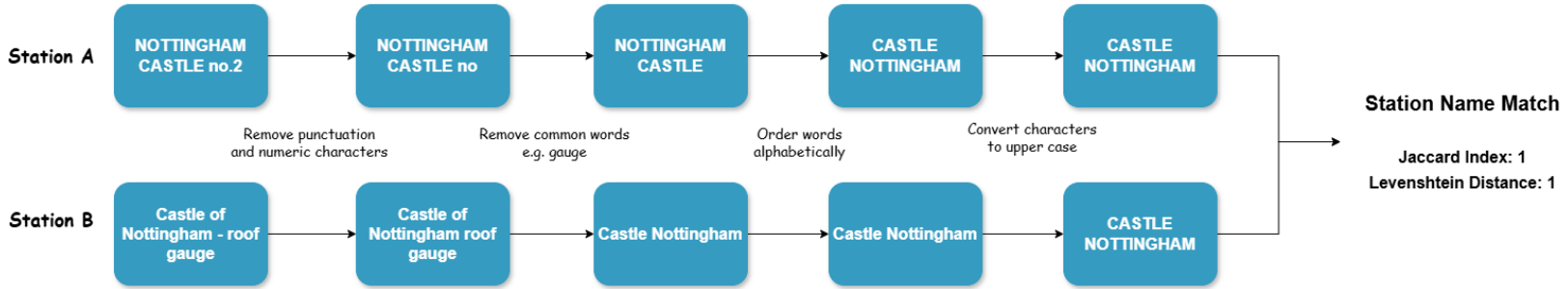


Rainfall trends for Arbroath and Arbroath Springfield Park stations

Station name matching

- Standardise the station names in MIDAS Open and Rainfall Rescue prior to calculating string similarity indices

- **Jaccard Index** – determines proportion of overlapping sets (words) in sample (station name).
- **String similarity score** – determines string similarity using **Levenshtein Distance** method.



Defining same station matches

- **Dynamic thresholds** allow for tighter or looser same station definitions.

- Match criteria are ranked and prioritised to keep the most likely station pair and remove duplicate matches.

Match criteria	Thresholds
Distance	< 500m
Value Match rate	>= 95%
Name similarity (max index)	> 0

Example

Select best match pair based on match criteria

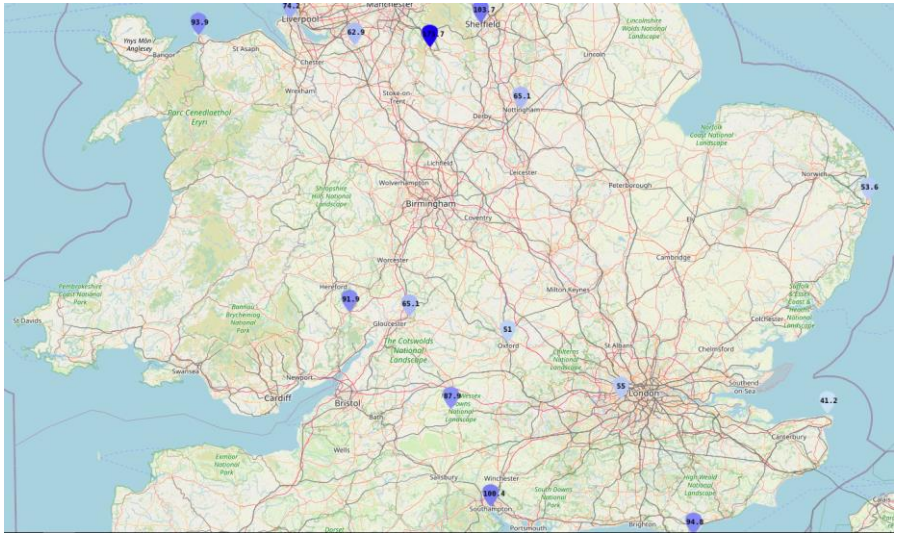
Rainfall Rescue ID1 → Station A

MIDAS Open Station	Matched Gauge	Distance	Value match	Name similarity (max index)	Match criteria
Station A	Rainfall Recue ID1	0m	100.0%	0.5	**1. Distance, value and name**
Station B	Rainfall Recue ID1	0m	50.0%	1	**2. Distance and name**
Station C	Rainfall Recue ID1	100m	99.0%	0	**3. Distance and value**
Station D	Rainfall Recue ID1	600m	100%	0.75	**4. Value and name**

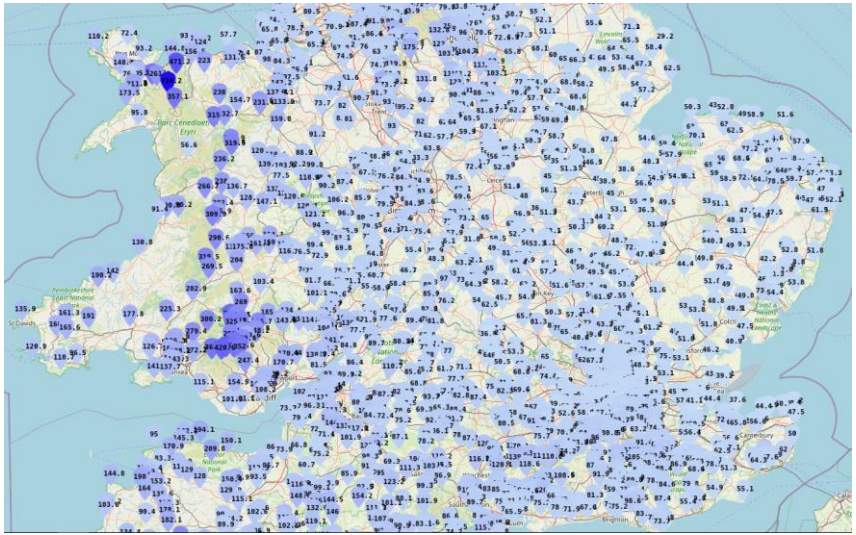
Key Results

- **Enhanced spatial coverage**
- **Extended rainfall records**
- **Infilling rainfall records**

Improved historic station coverage



January 1890 – MIDAS Open Only



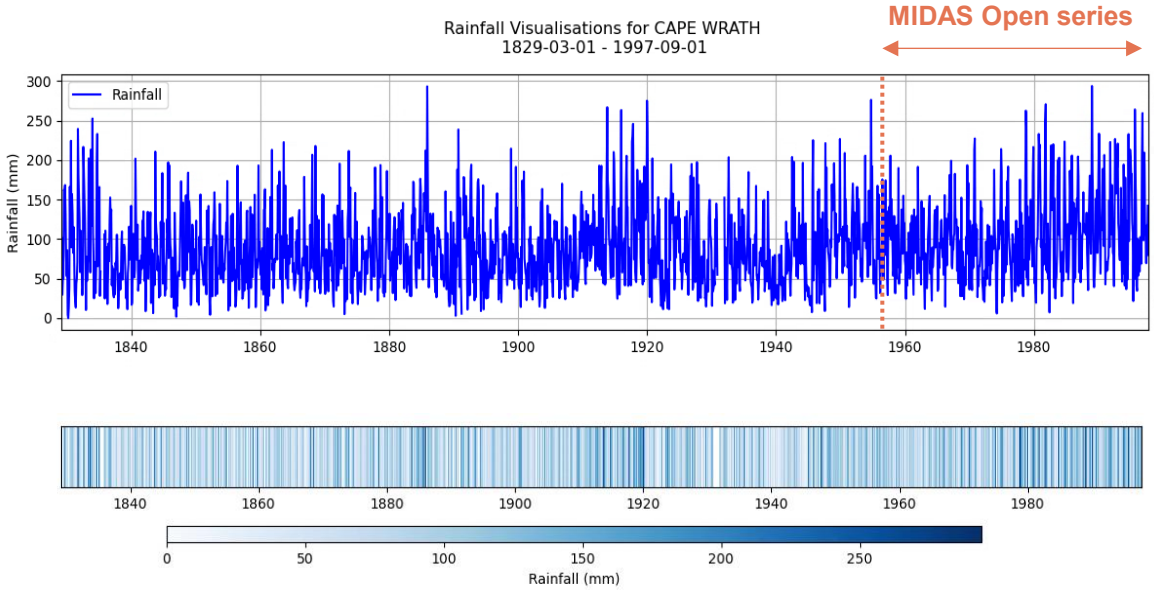
January 1890 – MIDAS Open and Rainfall Rescue (blended)

Extending station series

Example

Cape Wrath monthly rainfall series extended by nearly **128 years** back to March 1829.

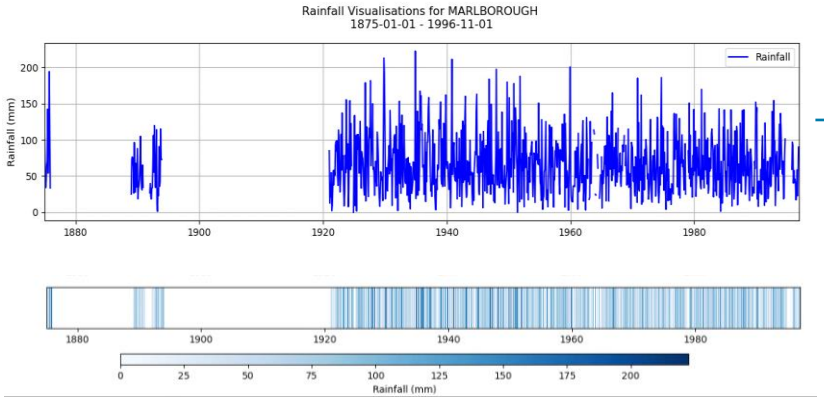
Only 1% of months missing in the **168-year** series.



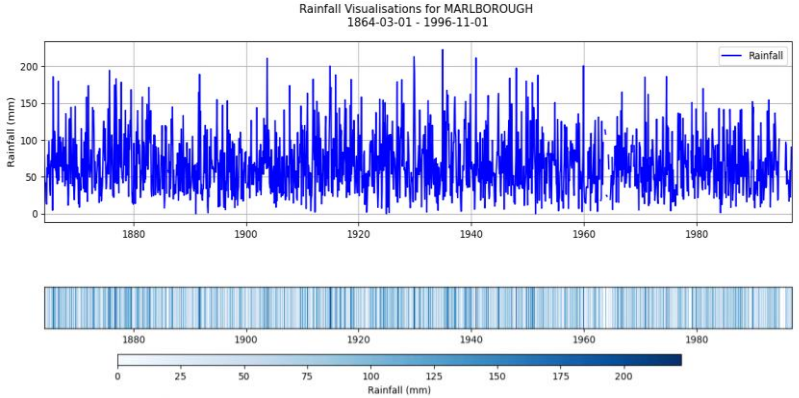
Infilling station series

Example

Before blending, more than a third of monthly values for Marlborough were missing in a rainfall series of over 130 years. After blending, **less than 2%** are missing.



MIDAS Open monthly rainfall series



Blended monthly rainfall series

Met Office

Summary

- Nearly **half of matched stations** had their monthly rainfall series extended by **25+ years.**
- For a **quarter of matched stations,** extending the series uncovered a new record rainfall event.
- Spatial coverage of the station network greatly expanded before 1950.
- Comparing monthly rainfall patterns proved useful for identifying matches where location metadata was missing or unreliable.
- Deriving station-level monthly rainfall series from raw observations is a complex task!

Future considerations

- **Define the purpose of the final blended product:** Should it provide a definitive ‘best case’ ruleset for the climate climate record’ or allow users to use bespoke rulesets?
- **Address station identification challenges:** Met Office land network stations have two sets of identifiers (station and gauge) which complicates the data aggregation and station matching.
- **Leverage additional metadata:** Useful metadata in Rainfall Rescue files could be utilized to flag suspect or inconsistent values.
- **Expand data coverage:** integrate additional historical sources, such as British Rainfall.

Met Office

Q&A