MATTHEW PAICE

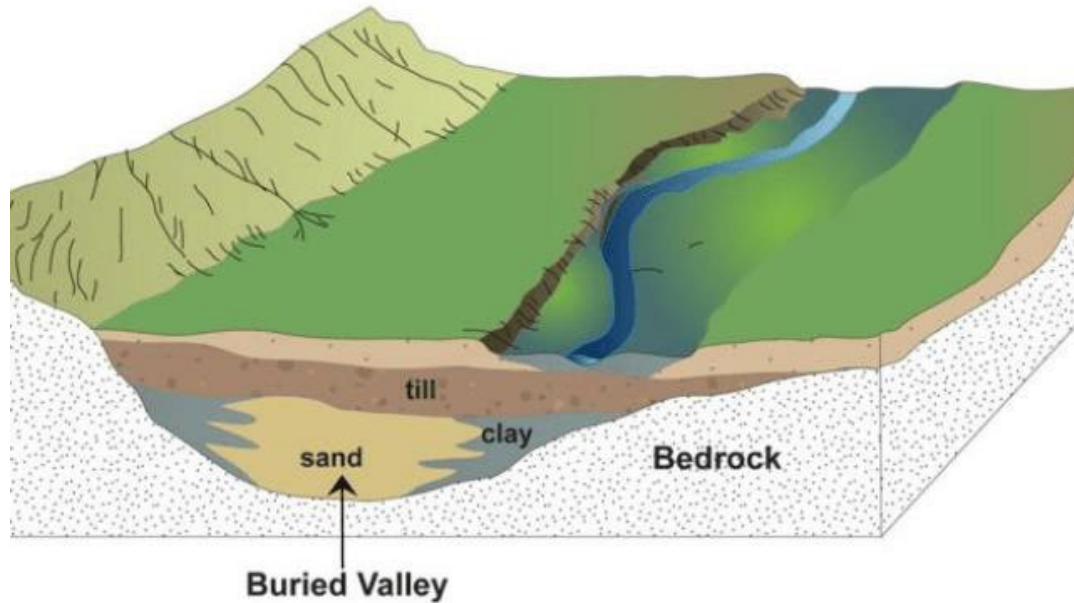# Updating the SDTM - challenges and the future

British Geological Survey

Introduction

- An introduction to superficial deposits and the SDTM

- Previous workflow and challenges

- The new data processing pipeline

- The potential future of the SDTM
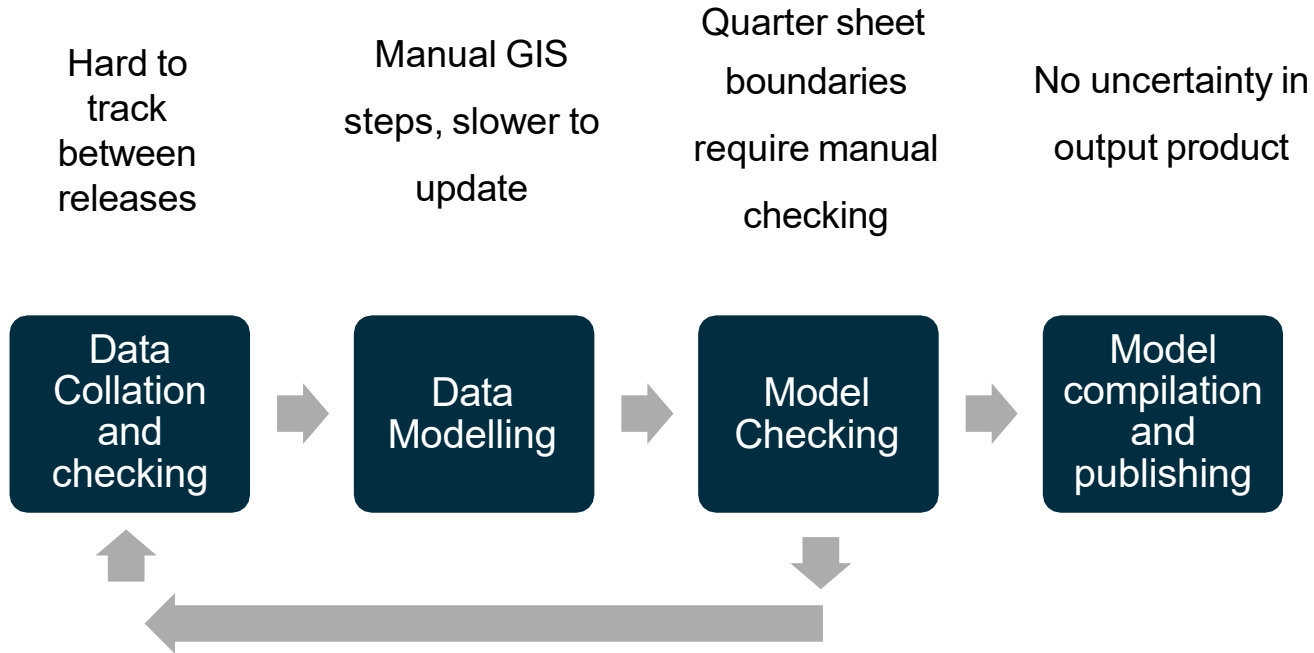
- Remaining challenges

- Summary

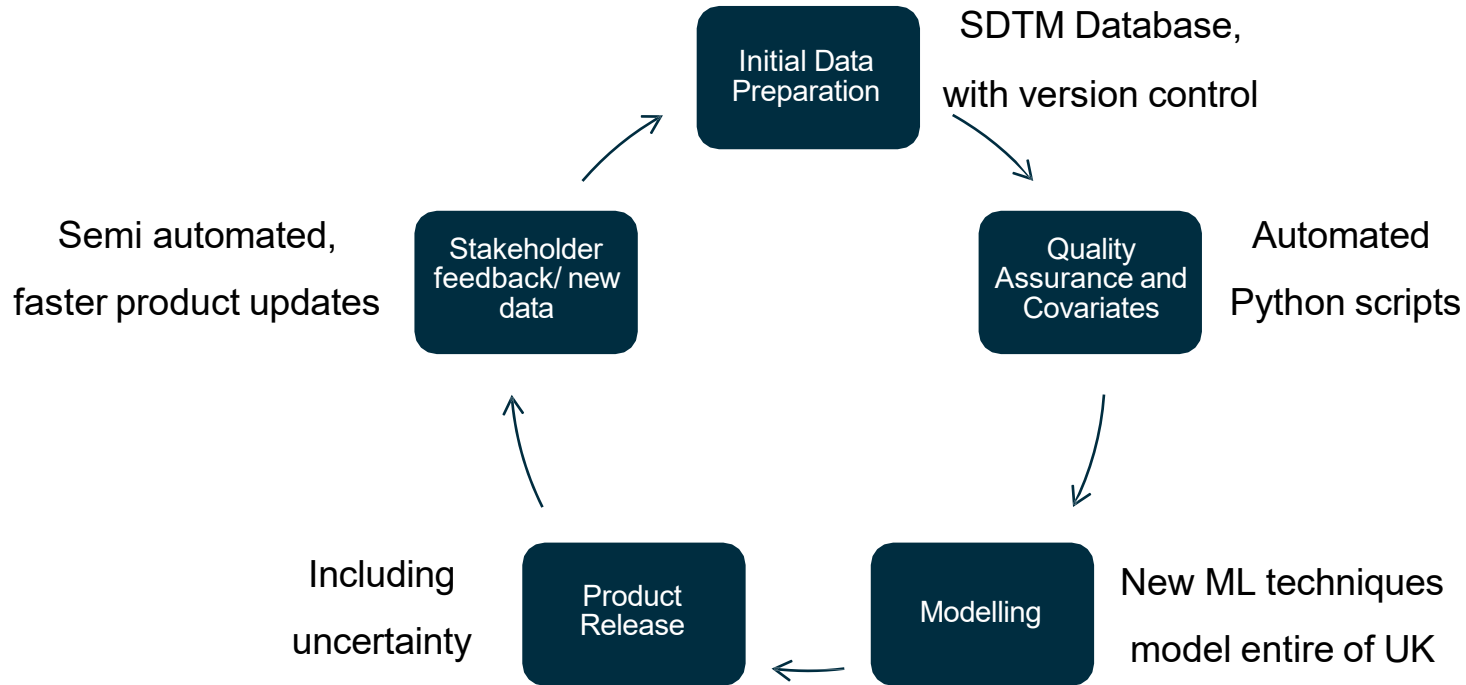# An introduction to superficial deposits and the SDTM

Dr Tim Kearsey
https://www.bgs.ac.uk/news/britains-ancient-valleys-now-uncovered/

The previous SDTM workflow and challenges

Hard to track between releases

Manual GIS steps, slower to update

Quarter sheet boundaries require manual checking

No uncertainty in output product

Data Collation and checking → Data Modelling → Model Checking → Model compilation and publishing

Process flow overview from Internal Report OR/10/035

The new data processing pipeline

Initial Data Preparation

SDTM Database, with version control

Quality Assurance and Covariates

Automated Python scripts

Modelling

New ML techniques model entire of UK

Product Release

Including uncertainty

Stakeholder feedback/ new data

Semi automated, faster product updates

Work in progress – not final

Semi automated – still including feedback from geologists at all stages

Quality Assurance

Flag	Description
1	**Accepted**
7	**Depth on montane domain >4m (accepted for now)**
0	Rejected at source
2	Rejected by geologist check
6	Pseudo older record check
10	Duplicate location, deeper borehole data used
11	More accurate point within precision range of BH
30	Max depth over 30m, no HV
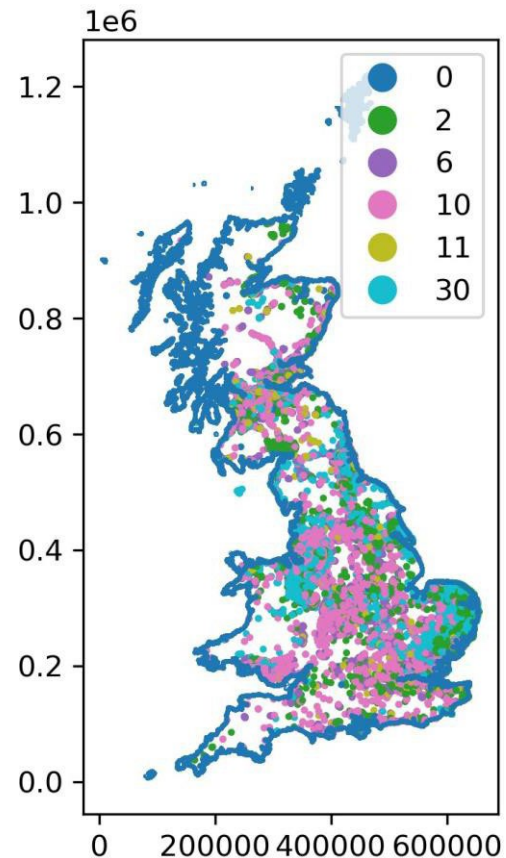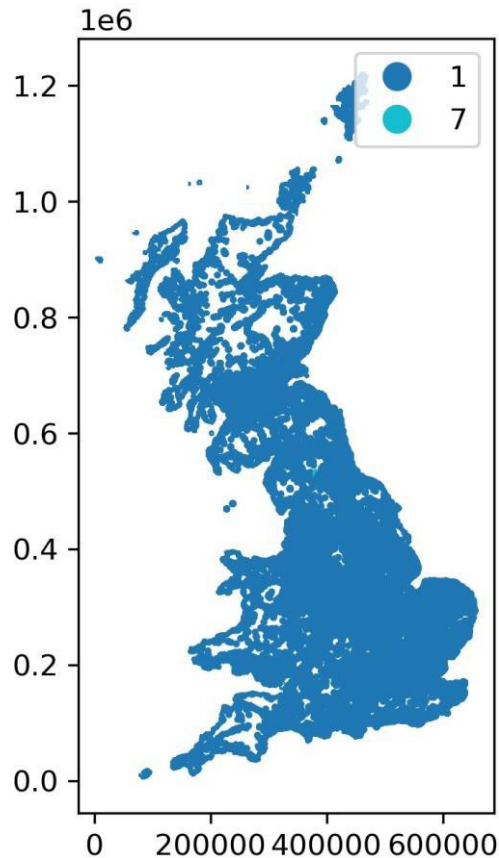
Current Prototype Model

SDTM prediction (23/3/23)

26.807151

0.578894

0   100   200 km
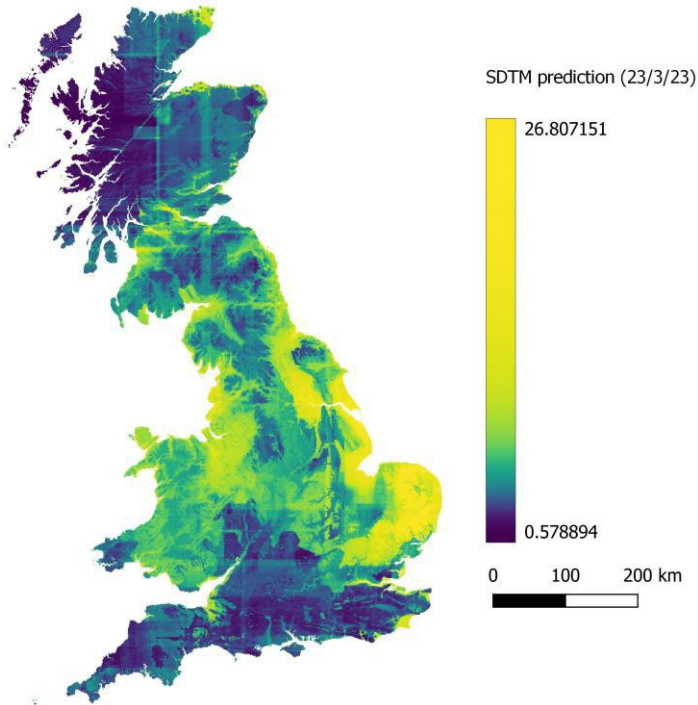
Trained on 50m UK point data of:

- Eastings and Northings

- Elevation (DTM)
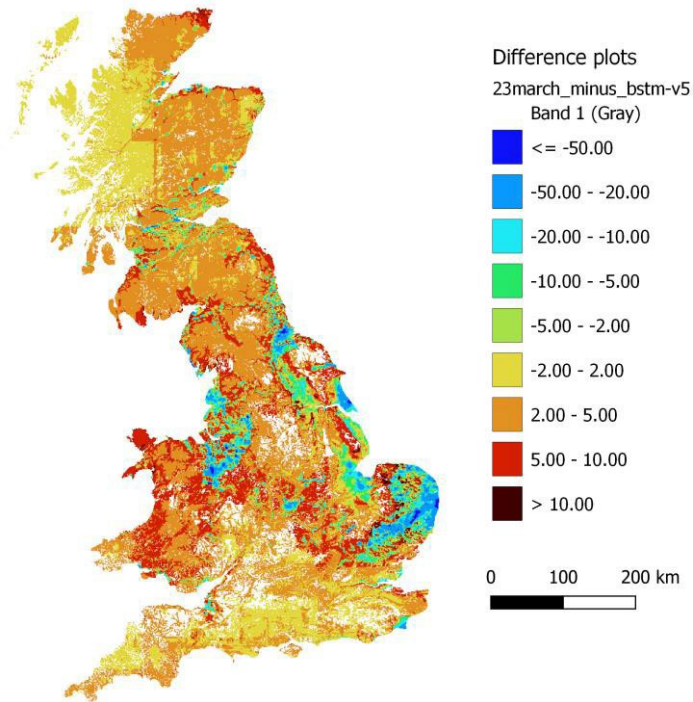
- One hot encoded QD

- Distance to next nearest borehole

Target variable:

- Rockhead borehole depths

Model:

- Quantile Gradient Boosted Tree

Remaining Challenges

Difference plots
23march_minus_bstm-v5
Band 1 (Gray)

■	<= -50.00
■	-50.00 - -20.00
■	-20.00 - -10.00
■	-10.00 - -5.00
■	-5.00 - -2.00
■	-2.00 - 2.00
■	2.00 - 5.00
■	5.00 - 10.00
■	> 10.00

0   100   200 km

- Incorporating geological rules

- Removing "subjective" rules – such as estimates for older records

- Removing "banding"

- Increasing superficial deposit depth in prototype model

- Improving uncertainty/ how to present uncertainty, especially in low borehole density regions

Summary

- Issues (mostly due to technological limitations):

  - Hard to update, mostly manual (QS boundaries), tracking data decisions

- Solutions

  - Semi automated data preparation, quality assurance and modelling, maintaining input from geologists at each stage.
  - Quantified rules around data prep and QA.

- Remaining issues:

  - Quantitatively identifying less reliable SOBI records
  - Model: lack of depth of deposit, banding
  - How to present modelled uncertainty

Thank yous

- Chris Williams – Initial project lead, initial datasets, project management
- Russel Lawley – Existing knowledge of previous SDTM and geological rules
- Rob Shaw – Initial datasets, additional geological datasets, project management
- Sophie Taylor – Data pipeline
- Tim Kearsey – Hidden valley dataset and geological interpretation
- Katie Whitbread – Geological interpretation
- Roman Roth – SDTM database
- All of the previous BGS geologists and mappers who have worked on the SDTM