



DATA PROCESSING, DATA QUALITY AND COPYRIGHT

A QUICK SERIES OF REMARKS ON THE PITFALLS OF TRUSTING THE PROCESS AND YOUR DATA PROVIDERS

Luc Girod
GEO(3|4)460 – Spring 2025

THE DATA PROCESSING PIPELINE

Input data



Processing



Results

THE DATA PROCESSING PIPELINE – IN THEORY

Great, complete
input data

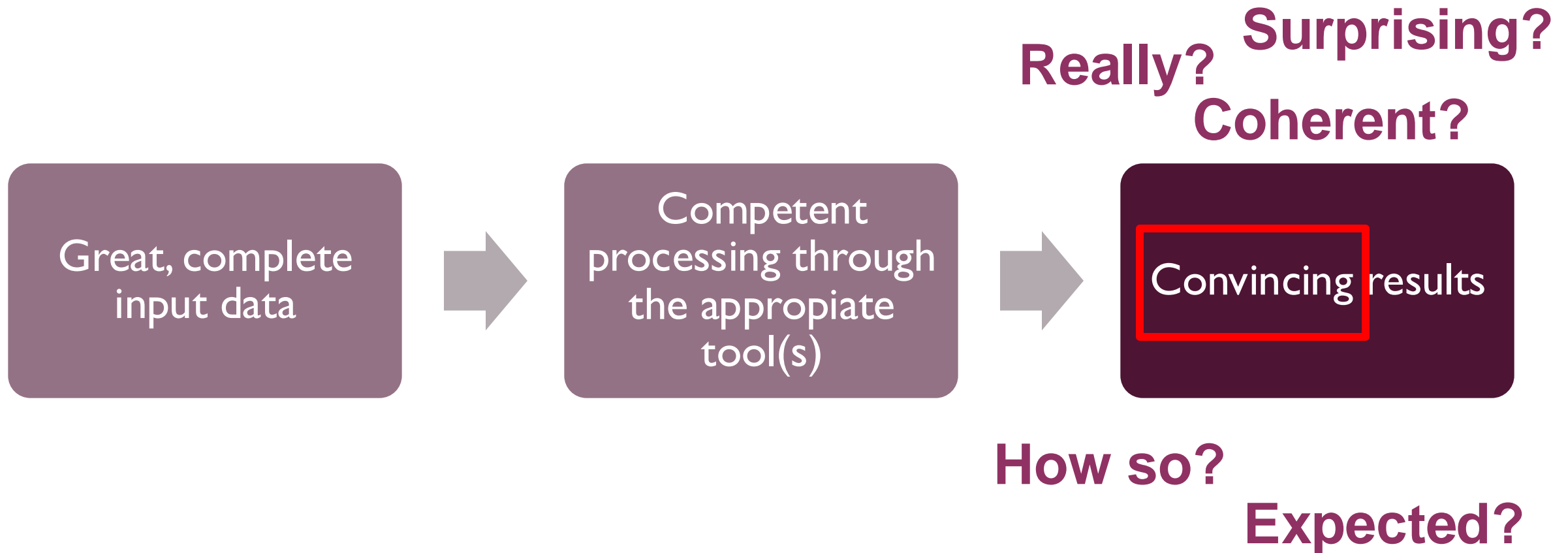


Competent
processing through
the appropriate
tool(s)



Convincing results

THE DATA PROCESSING PIPELINE – IN THEORY



VALIDATING RESULTS – DIFFERENCE FROM EXPECTATIONS



In accordance to a priori – can be logically explained



Not raising alarm – no strong a priori



Different from expectations – defies a priori



VALIDATING RESULTS – USING GROUND TRUTH

Input data



Processing



Results

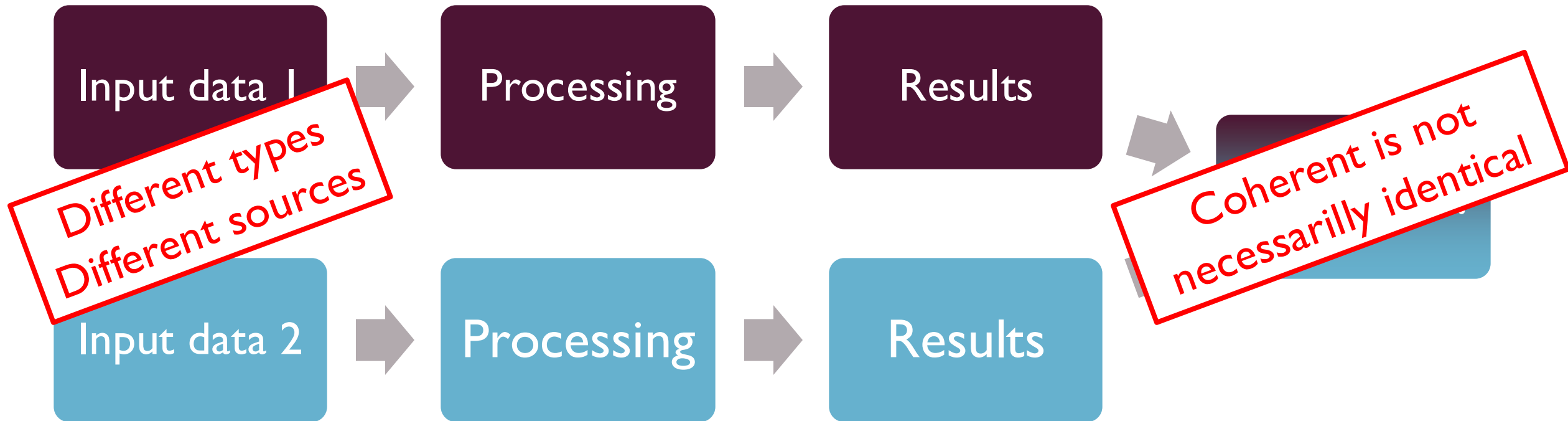


Validation data

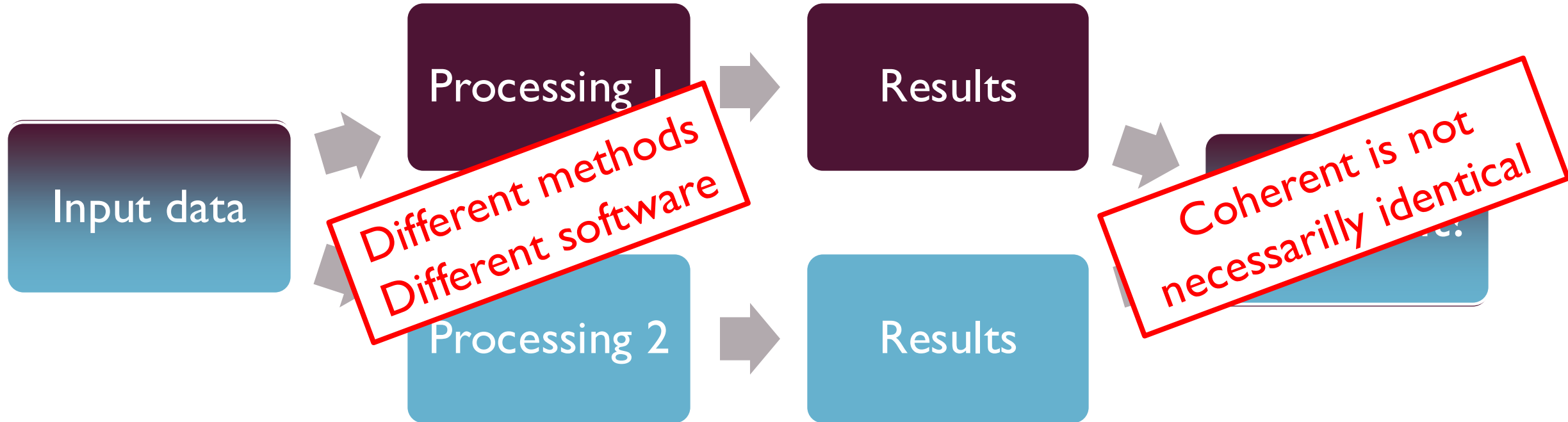
- On some limited sample points
- From another point in time (expected consistency through time)
- From another location (expected statistical similarity)

*Coherent is not
necessarilly identical*

VALIDATING RESULTS – THROUGH ANOTHER DATA SOURCE



VALIDATING RESULTS – THROUGH ANOTHER PROCESSING METHOD



WHAT IF THE RESULTS AREN'T CONVINCING?



THE DATA PROCESSING PIPELINE – GOOD PROCESSING



Check your work

- Were there warnings or error messages?
- Read tool(s) documentation
 - Is the tool appropriate for the task?
 - Is the input data appropriate for the tool?
 - Are there options I forgot to (de)activate?

Great, complete
input data

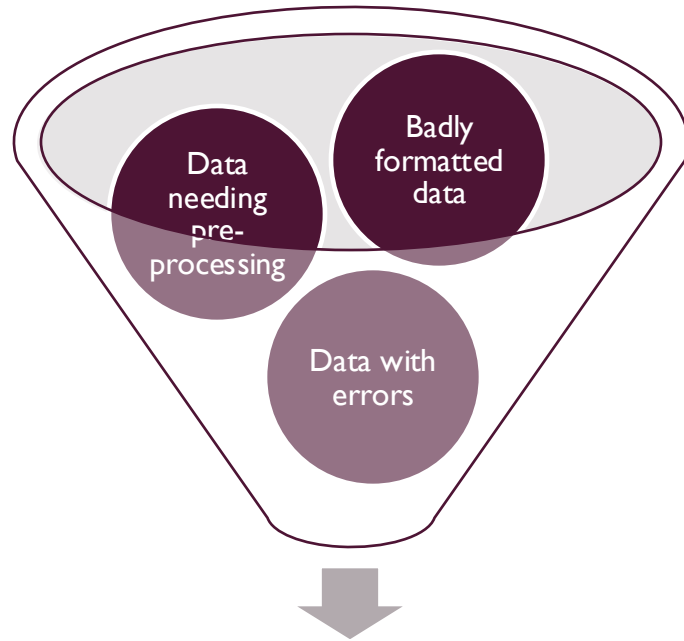


Competent(?)
processing through
the appropriate (?)
tool(s)



Unconvincing
results

THE DATA PROCESSING PIPELINE – GOOD INPUT?



Check input data

- Reliable source? (Official source vs 'somewhere online')
- Appropriate for the tools (formatting details can have big impact, check the tool documentation for info)
- Verify data integrity after topological operation!

Not that great
input data

Competent
processing through
the appropriate
tool(s)

Unconvincing
results

SCENARIOS WITH BAD DATA

The tool
refuses it and
gives a usefull
error message

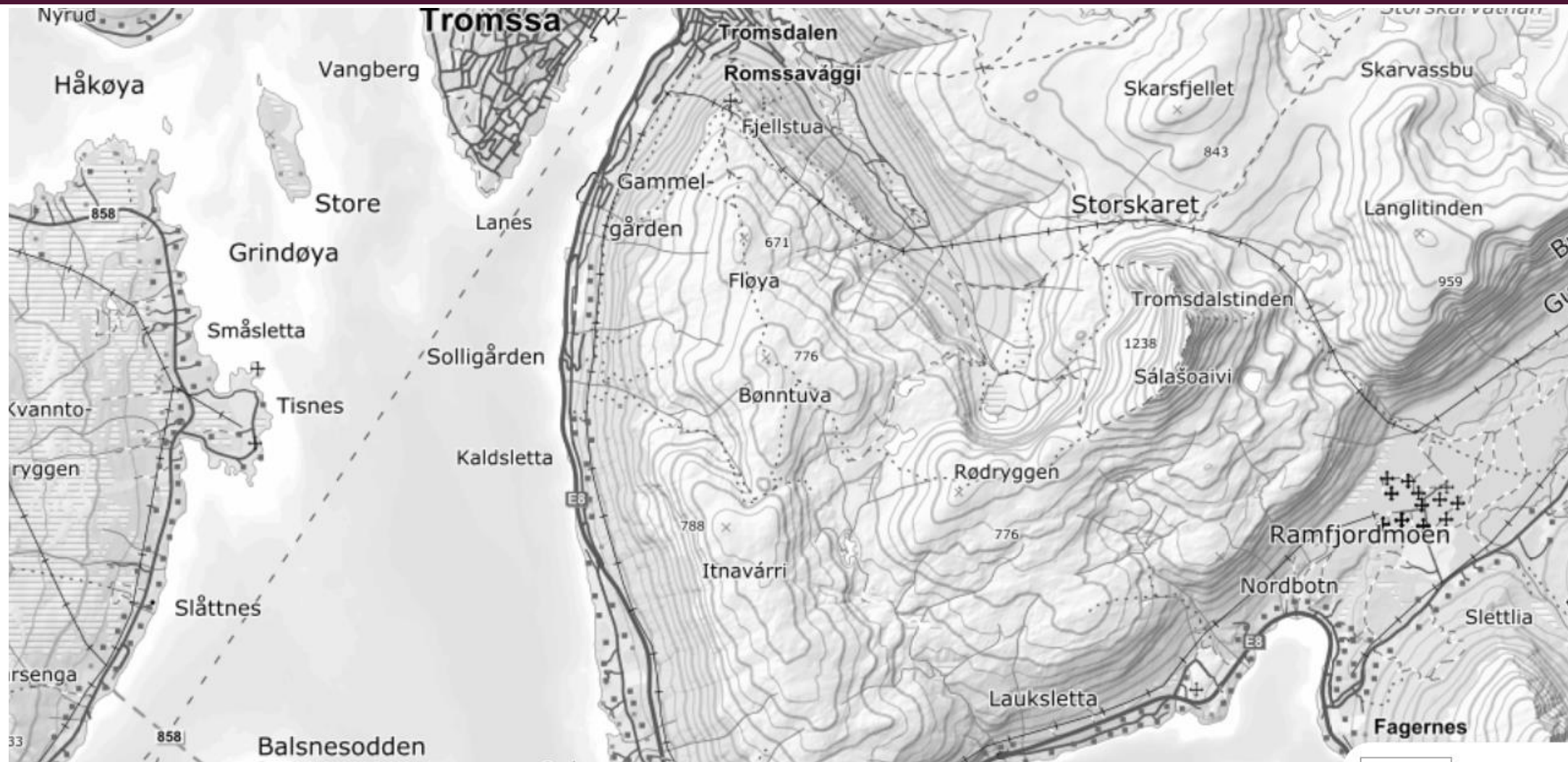
The tool
refuses it with
a generic
error

The tool
crashes

The tool
accepts the
bad data

Severity of the issue

AN EXAMPLE – TOPO DATA IN TROMSØ TO DEM



TIN TO RASTER SEEMS LOOKS REASONABLE, BUT LET'S VALIDATE

Output TIN
N50_TIN

Coordinate System
ETRS_1989_UTM_Zone_33N

Input Feature Class

Input Features: N50_Høyde_sent
Height Field: høyde
Type: Mass_Points
Tag Field: <None>

Input Features: N50_LakeAndSea
Height Field: <None>
Type: Soft_Erase
Tag Field: <None>

+ Add another

TIN To Raster

Parameters Environments

Input TIN
N50_TIN

Output Raster
N50_TINRaster


Output Data Type
Floating Point

Method
Linear

Sampling Distance
Cell Size

Sampling Value
20

Z Factor
1



TOPO TO RASTER IS...WEIRD

← Topo to Raster +

Parameters Environments ?

Input feature data ▾

Feature layer N50_Høyde_sei ▾

Field høyde ▾

Type Contour ▾

Feature layer N50_LakeAndS ▾

Field

Type Lake ▾

Feature layer N50_Arealdekk ▾

Field

Type Stream ▾

+ Add another

⚠ Output surface raster

TopoToR_N50_20m_BadRiver ▾

Output cell size

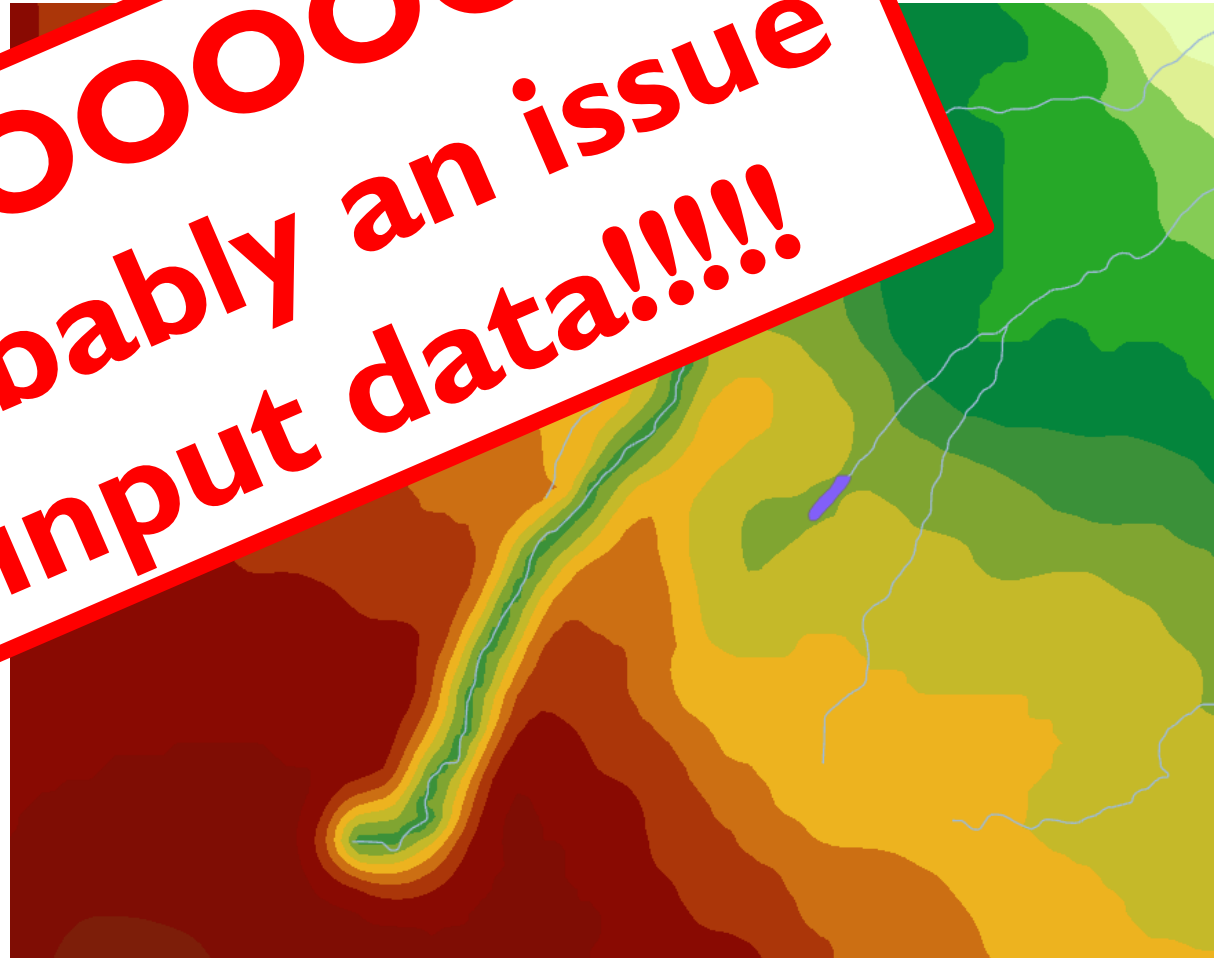
20 ▾



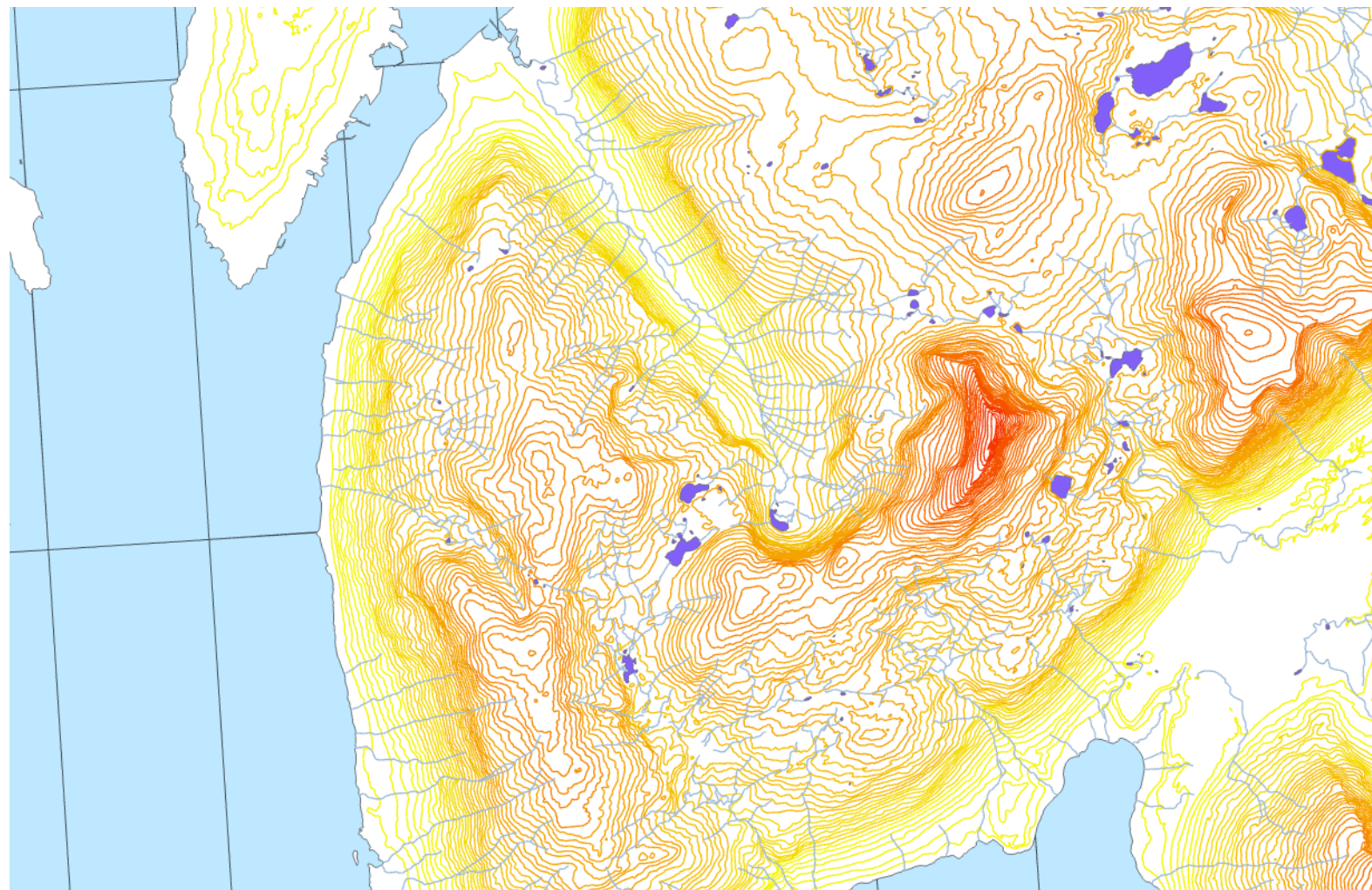
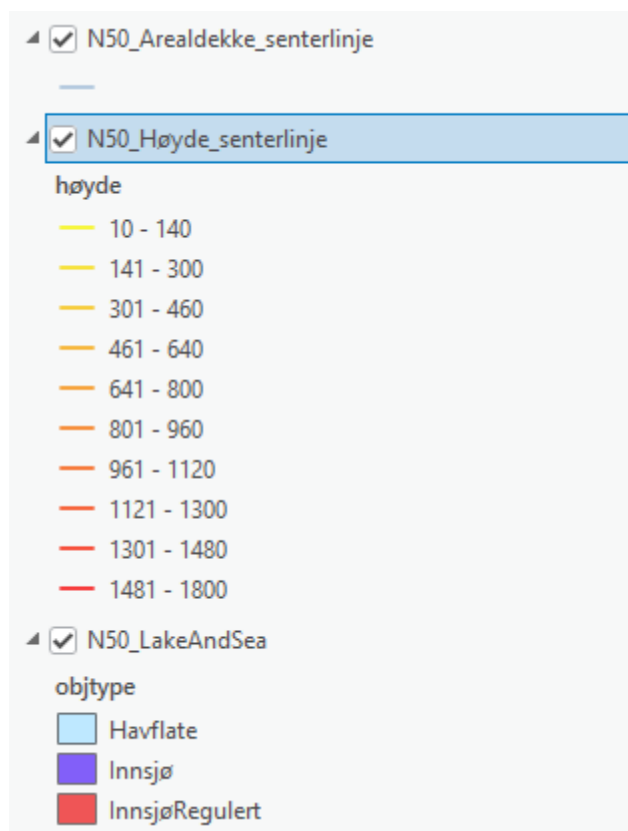
WHAT'S THAT????

- Difference between TIN and Raster
Topo to Raster reproject
meter
- Cany
SOME
- The tool
(checked.

STOOOOOOOOOOP
There's probably an issue
with the input data!!!!



DATA IS FROM A TRUSTWORTHY SOURCE

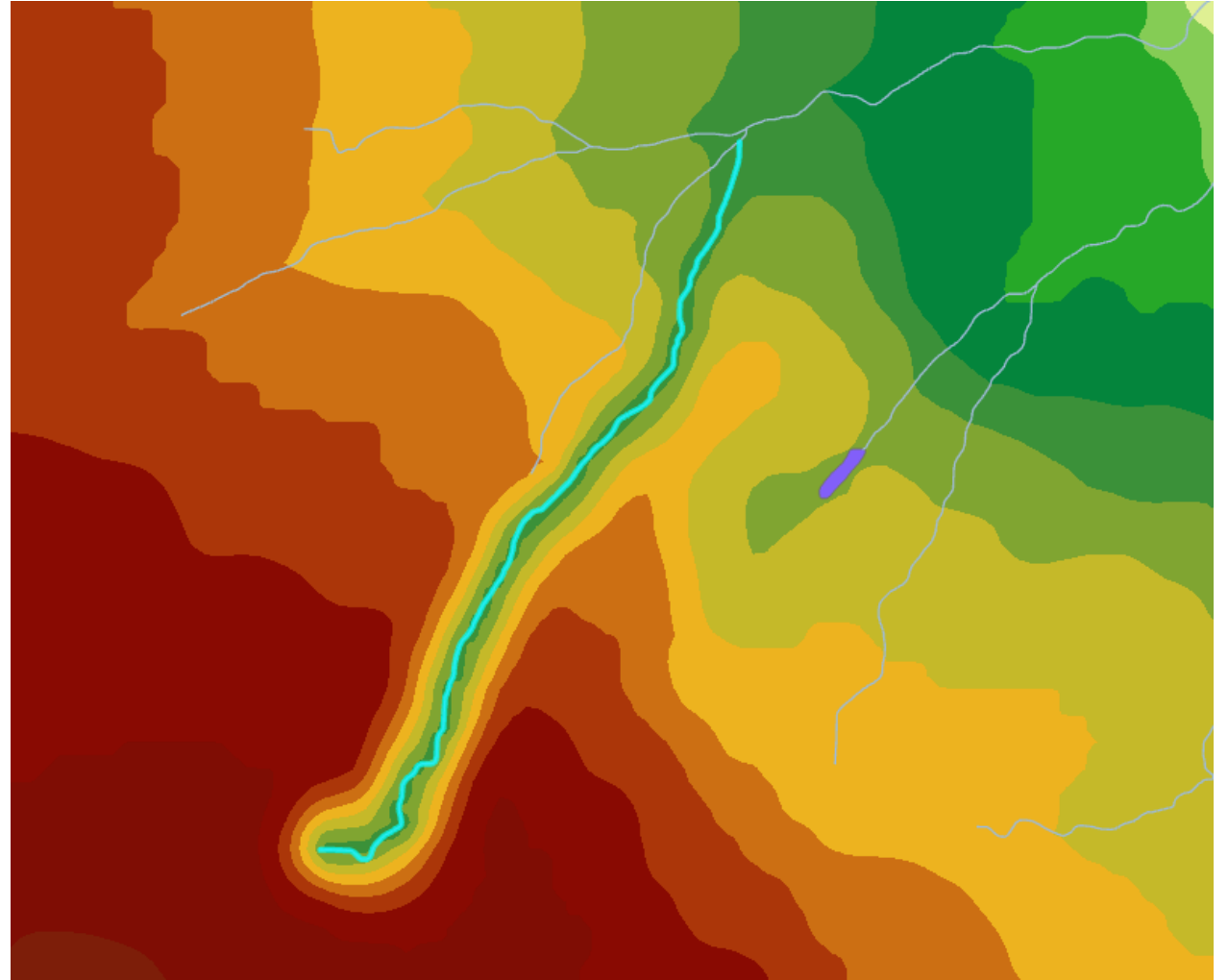


FIGURING OUT THE ISSUE

- The river seem to have an undesirable effect...sometimes
- They are used in the Topo to Raster tool through the 'Stream' option
- Let's lookup the doc : <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/topo-to-raster.htm>
 - **Stream**—A line feature class of stream locations. All arcs must be oriented to point downstream. The feature class should only contain single arc streams. ~~There is no Field option for this input type.~~

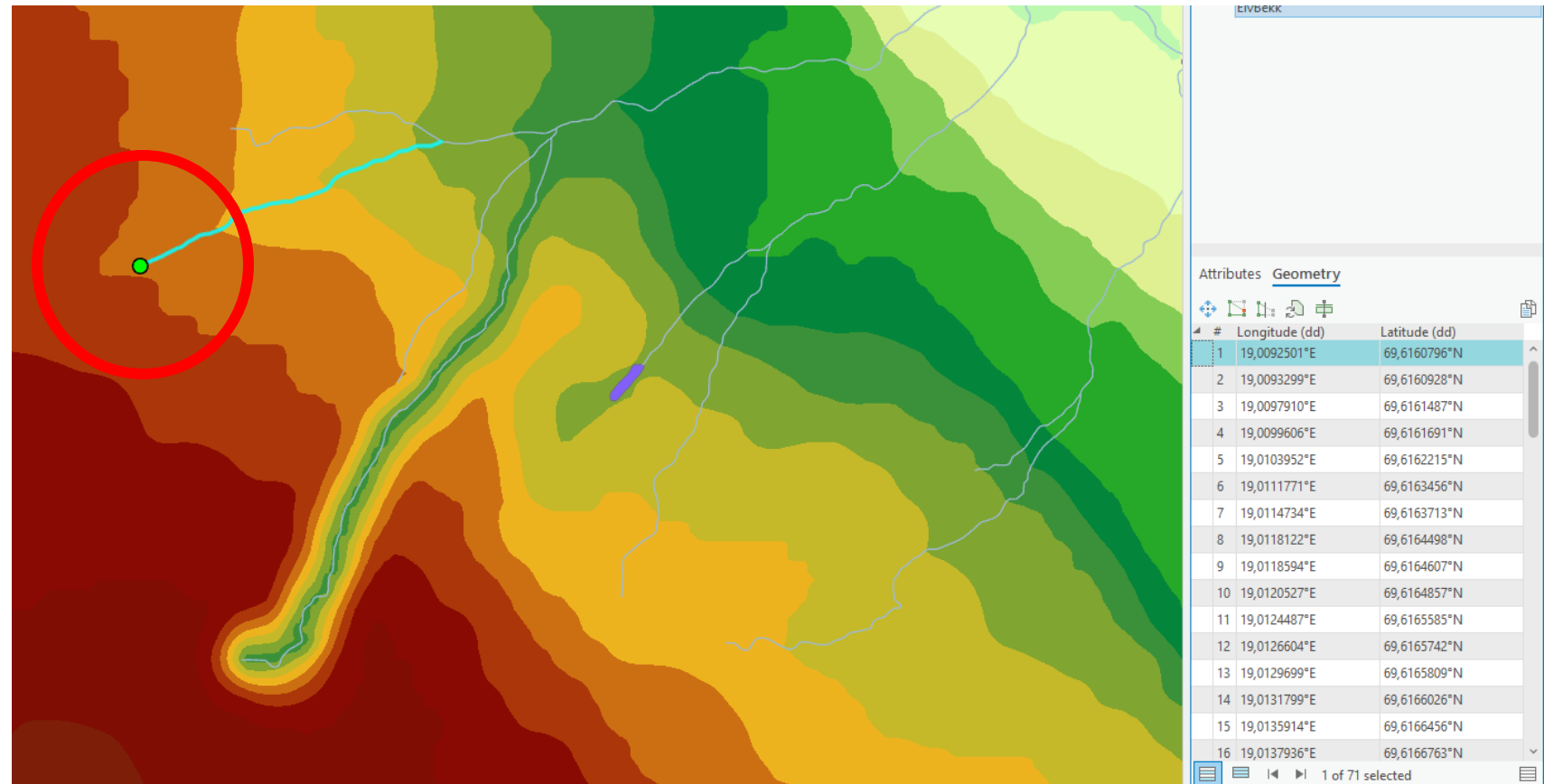
SINGLE ARC STREAMS?

- Each stream has to be a simple arc, not a multi-line
- Edit→select→stream are selected individually, so they are 'single arcs'



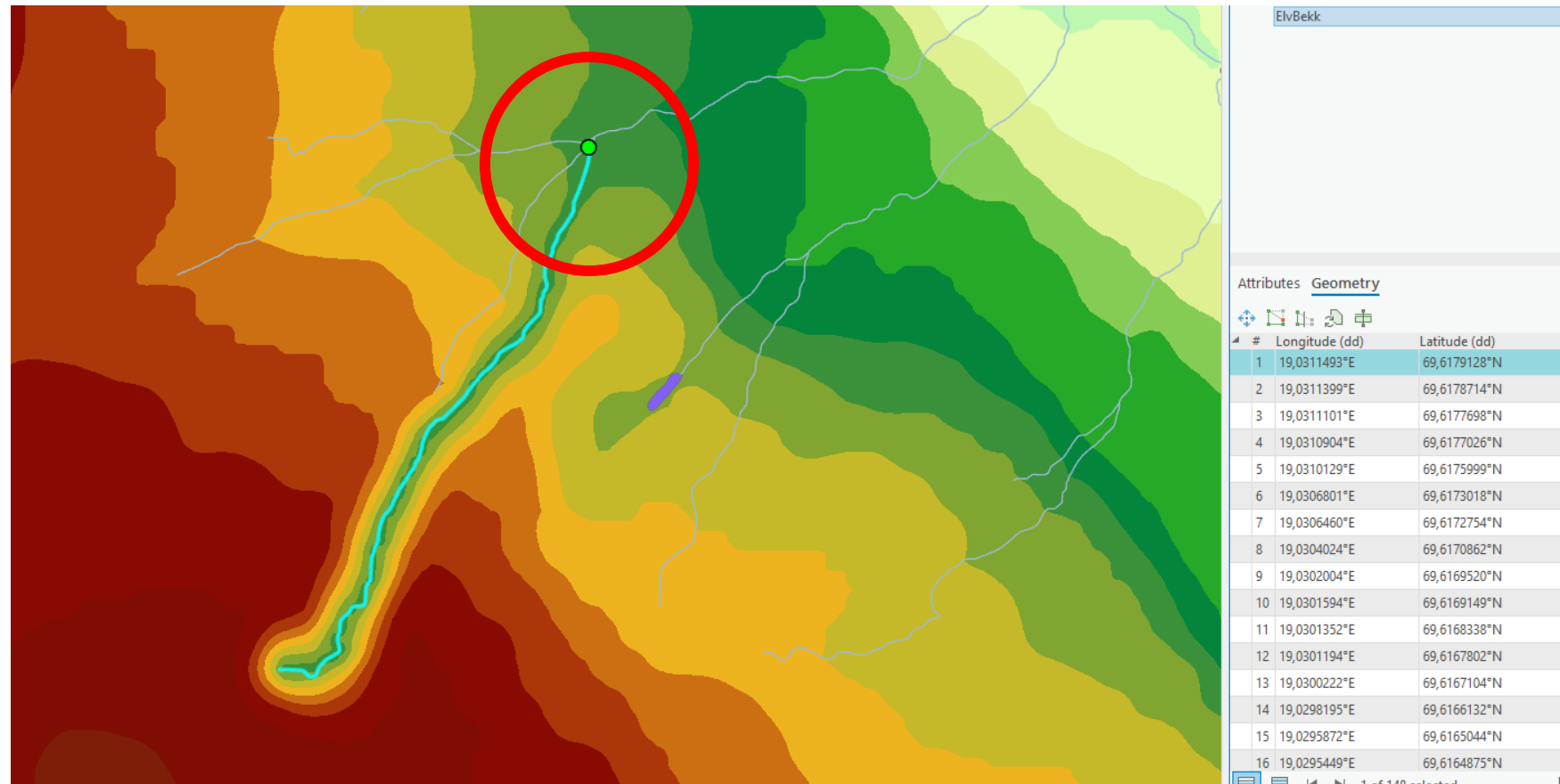
ORIENTED TO POINT DOWNSTREAM?

- Edit→
attribute→geometry
- Non-problematic
arcs have their first
point at the
source/high point of
the river section



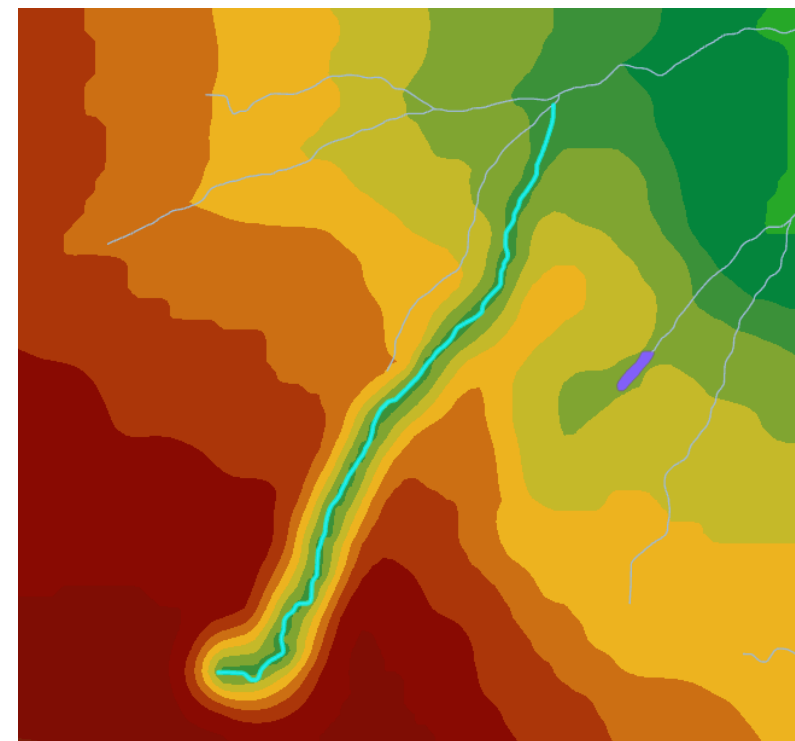
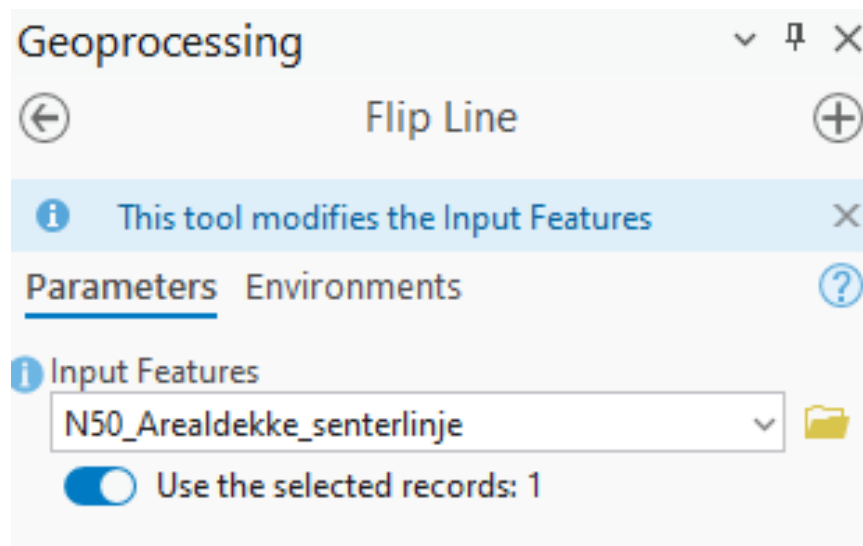
ORIENTED TO POINT DOWNSTREAM?

- Problematic arcs have their first point at the end/low point of the river section
- This forces the *Topo to Raster* tool to create a raster where the high point is lower than the low point

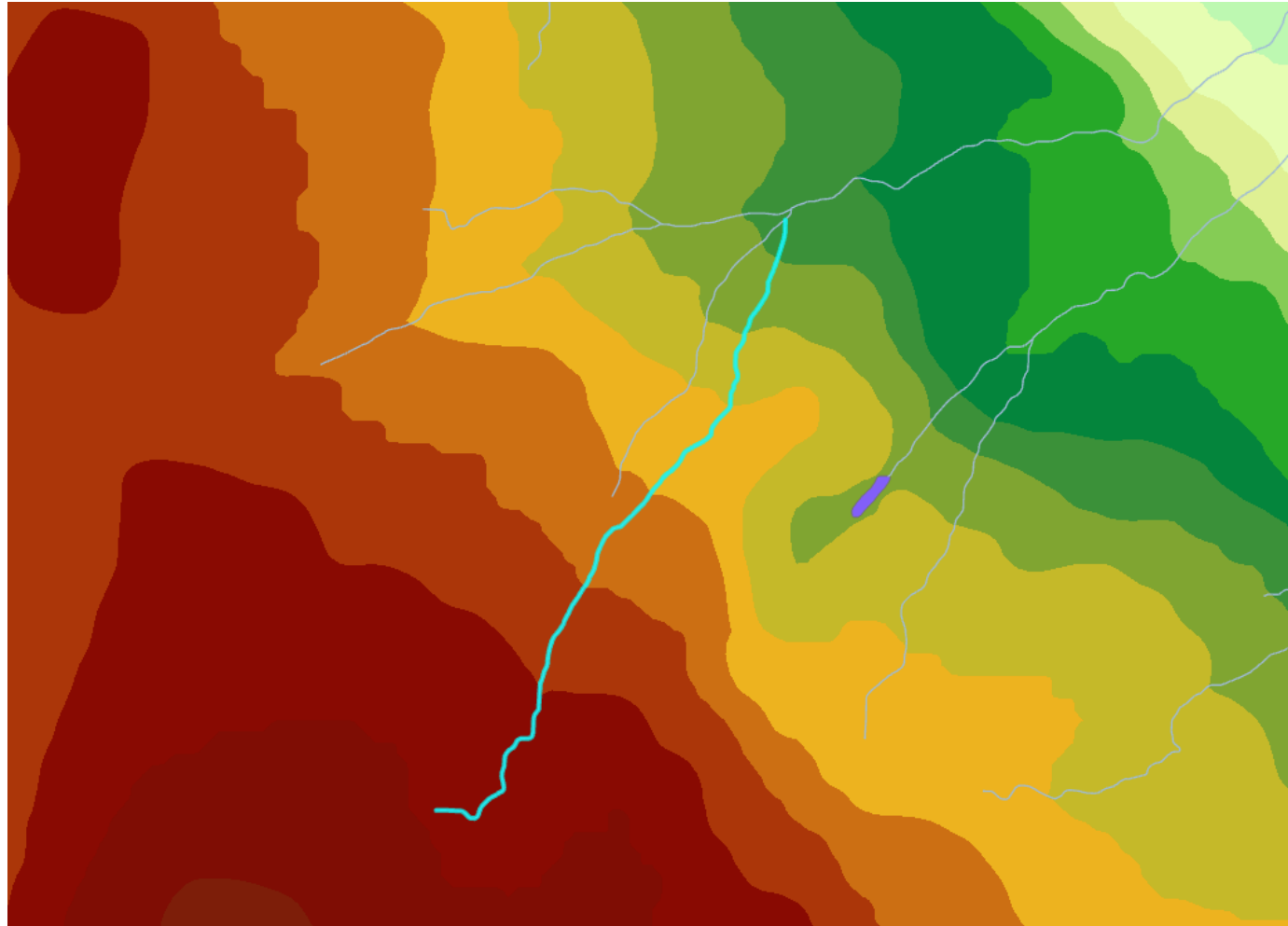


SOLVING THE ISSUE

- Select all bad lines
- Run Flip Line



RE-RUNNING THE TOOL ON CLEANED DATA → SUCCESS!!!

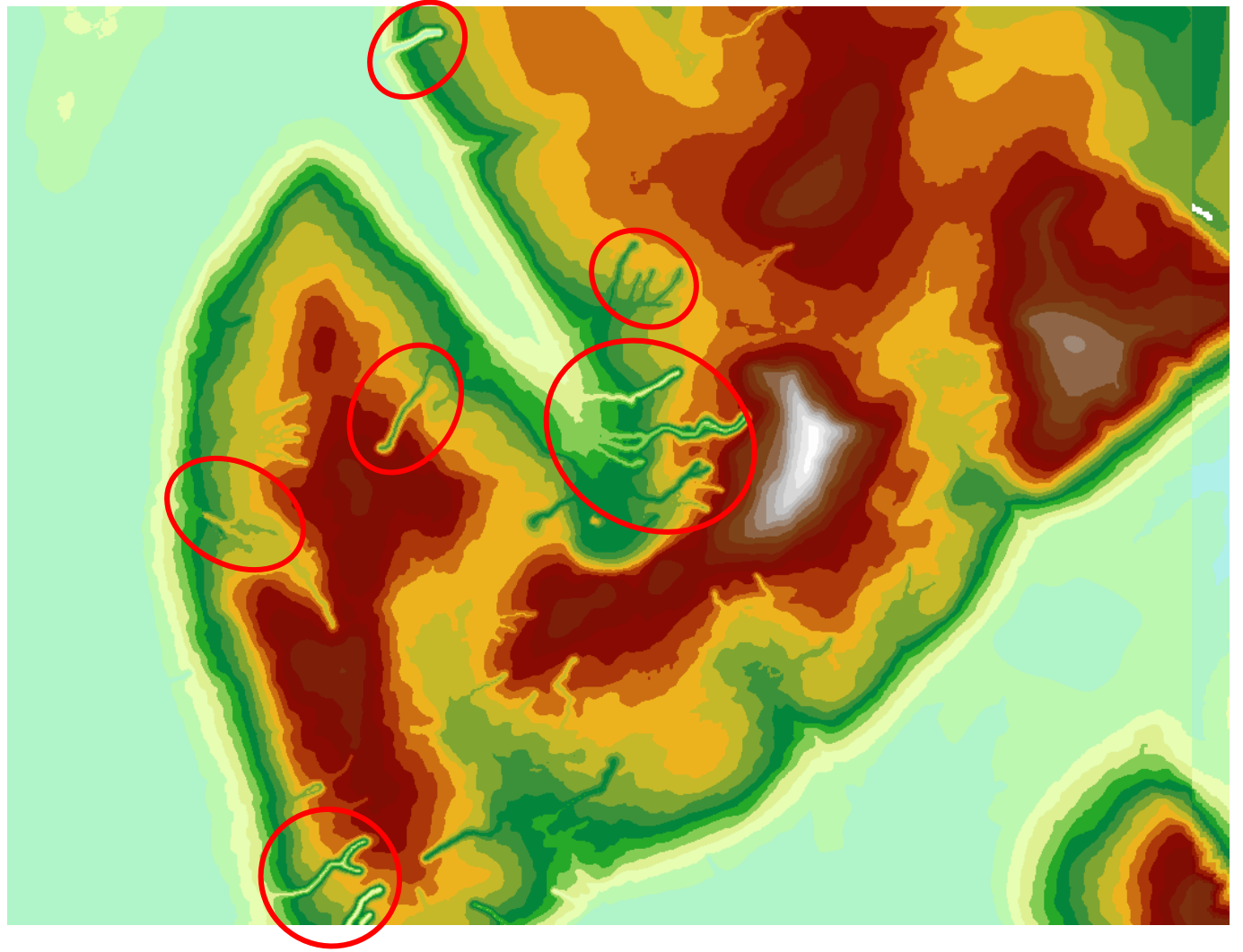


WELL...

Manually fixing all those rivers would be quite the task.

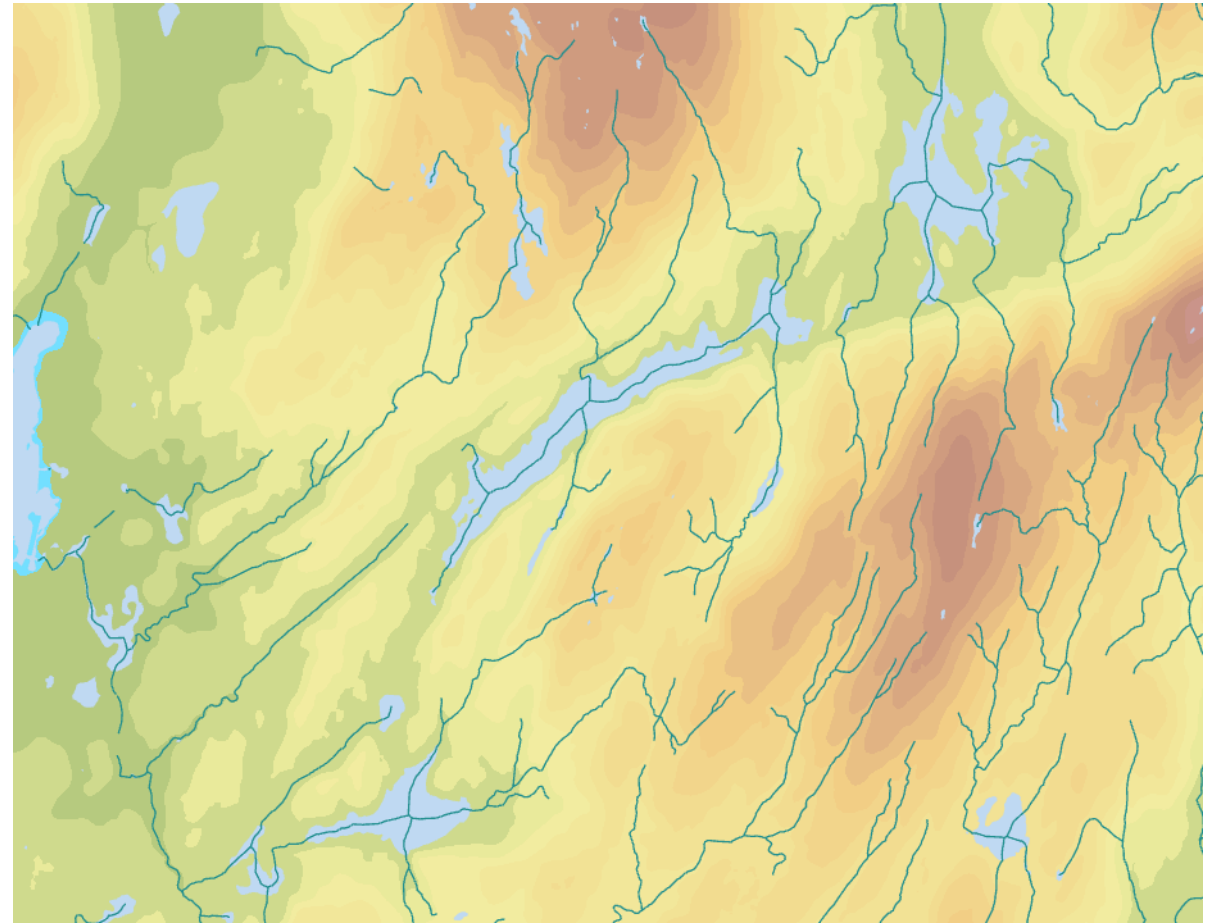
One could design an automated tool:

1. Compute a TIN based DEM
2. Compute the elevation of all start and end nodes in the stream lines
3. Compare the values
4. Flip the line if the start point is lower than the end.



ANOTHER LOCATION

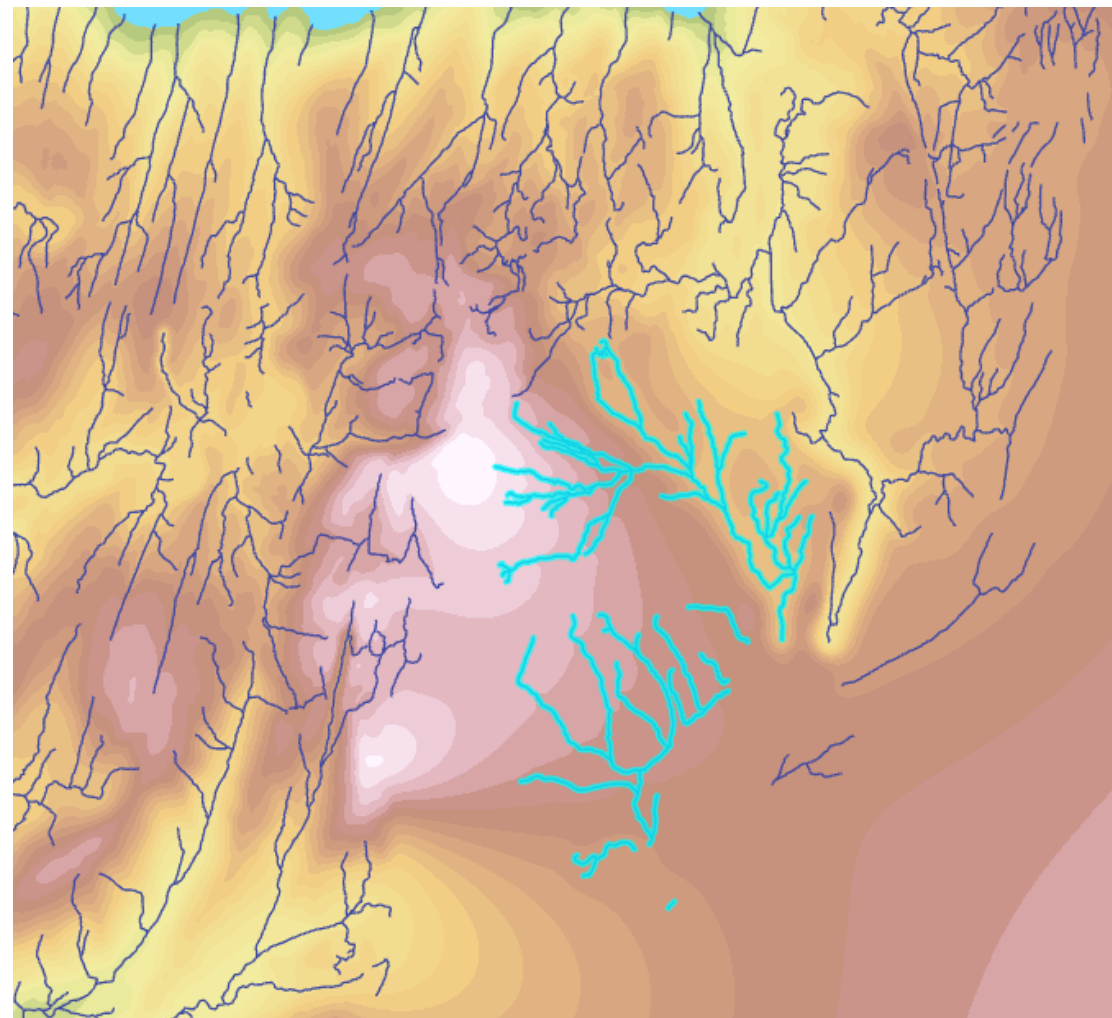
In Bergen, the issue is also widespread, and made worse (better?) by having a lot of lakes that the streams go through. Having those lakes AND rivers in the process leads to a crash as the forced gradients from the river cannot be reconciled with the forced flatness of lakes,



TRY ANOTHER DATA SOURCE

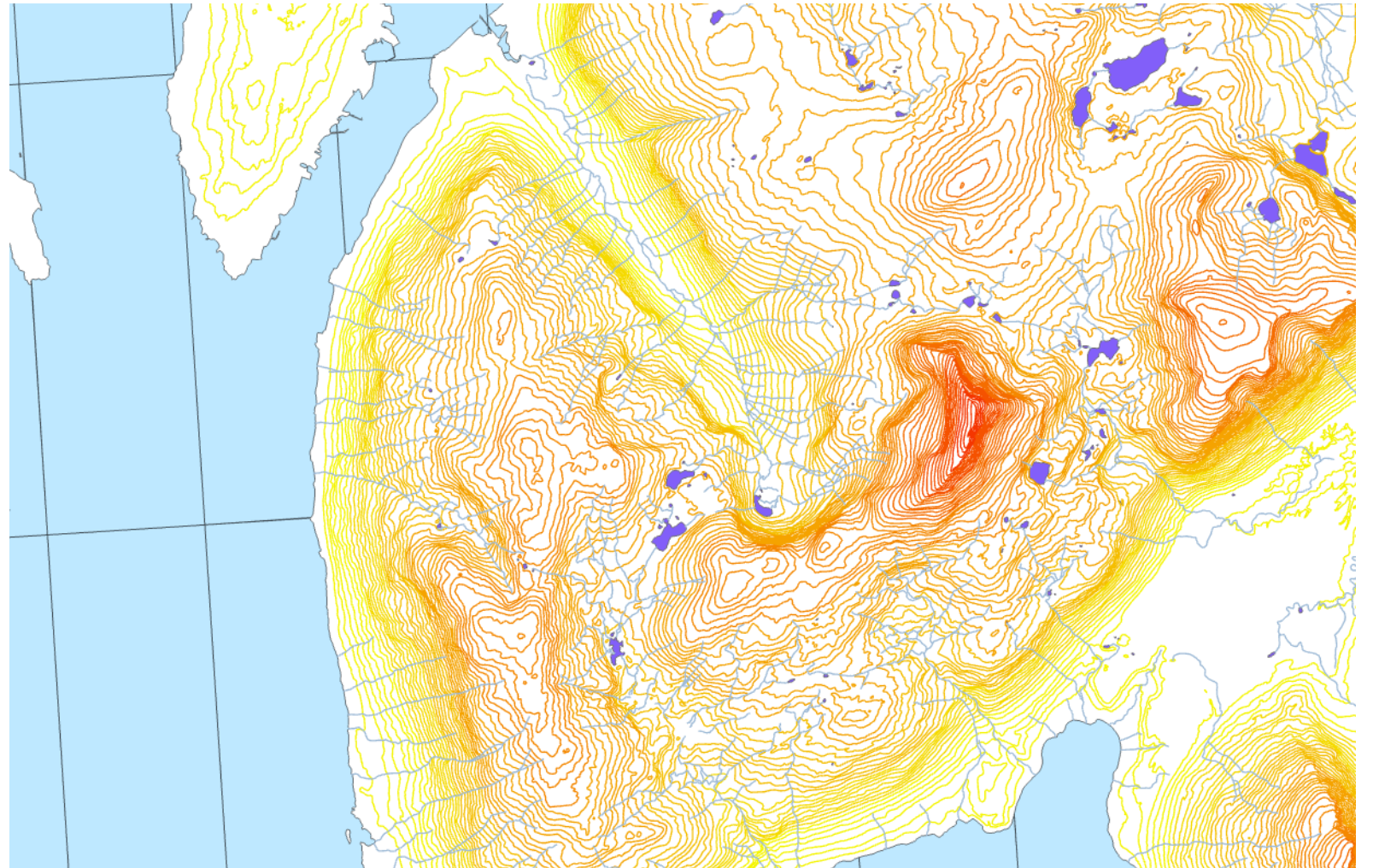
Trying the *Miljødirektoratet Vannforekomster* dataset didn't help, as it didn't satisfy the 'single arc' requirement of the Topo To Raster tool.

A single polyline is selected in this figure →



DATA IS FROM A TRUSTWORTHY SOURCE...BUT

Even trustworthy
data providers
might not provide
data that is fit for
YOUR purpose!



A FEW NOTES ON COPYRIGHT AND LICENSES

- (Geographical) data is expensive to capture, process, and distribute
- Data distributor are free to decide how their data is distributed:
 - Free? Paid? Only accessible to designated/cleared users?
 - Restrictions on use? Only in science/public service, all BUT some applications...
 - Restrictions on redistribution? Data is often only meant to be obtained from the source
 - Requirements for credit? Oftentimes a small ©*Data Provider* is necessary.

FAIR DATA

N50 Kartdata

FAIR-status: 90% 😊



FAIR

Findability, Accessibility, Interoperability, and Reusability

FAIR-status: 😊 90%



100%

Søkbarehet (Findable): Beskrivelser av datasett (metadata) og datasett skal være enkle å finne for både mennesker og datamaskiner. Maskinlesbare metadata er avgjørende for automatisk oppdagelse av datasett og tjenester

[Se detaljer](#)



80%

Tilgjengelighet (Accessible): Datasett skal være tilgjengelige gjennom standardiserte og åpne grensesnitt.

[Se detaljer](#)



100%

Interoperabilitet (Interoperabel): Dataene må ha samvirkningsevne og kunne integreres i ulike løsninger.

[Se detaljer](#)



80%

Gjenbrukbar (Reusable): For å optimalisere gjenbruk av data bør datasett beskrives godt slik at de kan tas i bruk i andre brukerløsninger.

[Se detaljer](#)

DISTRIBUTION – LICENSE

N50 Kartdata

FAIR-status: 90% 😊



☒ Kartkatalogen (9142)
☐ Artikler (320)



Meny ☰

Distribusjon

Representasjonsform: Vektor

Distribusjonstype:

Geonorge nedlastning

URL: <https://nedlasting.geonorge.no/api/capabilities/>

Geografisk distribusjonsinndeling: fylkesvis, kommunevis, landsfiler

Restriksjoner

Bruksbegrensninger: Ingen begrensninger på bruk er oppgitt

Tilgangsrestriksjoner: Åpne data

Bruksrestriksjoner: Ingen

Lisens: [Creative Commons BY 4.0 \(CC BY 4.0\)](#)

Sikkerhetsnivå: Ugradert

DISTRIBUTION – LICENSE

  CC BY 4.0

Attribution 4.0 International Deed

<https://creativecommons.org/licenses/by/4.0/>


You are free to:

Share — copy and redistribute the material in any medium or format for any purpose, even commercially.

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

 **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

FAIR DATA

N5 Kartdata

FAIR-status: 72% 😊



FAIR

Findability, Accessibility, Interoperability, and Reusability

FAIR-status: 😊 72%



95%

Søkbarhet (Findable): Beskrivelser av datasett (metadata) og datasett skal være enkle å finne for både mennesker og datamaskiner. Maskinlesbare metadata er avgjørende for automatisk oppdagelse av datasett og tjenester

[Se detaljer](#)



80%

Tilgjengelighet (Accessible): Datasett skal være tilgjengelige gjennom standardiserte og åpne grensesnitt.

[Se detaljer](#)



40%

Interoperabilitet (Interoperabel): Dataene må ha samvirkningsevne og kunne integreres i ulike løsninger.

[Se detaljer](#)



75%

Gjenbrukbar (Reusable): For å optimalisere gjenbruk av data bør datasett beskrives godt slik at de kan tas i bruk i andre brukerløsninger.

[Se detaljer](#)

DISTRIBUTION

N5 Kartdata

FAIR-status: 72% 😊

FAIR

Findability, Accessibility, Interoperability, and Reusability

N5 Kartdata

🔒 Datasett fra [Geovekst](#)

Type: Geonorge nedlastning

Formater **SOSI**

Restriksjoner

Bruksbegrensninger: Ingen begrensninger på bruk er oppgitt.

Tilgangsrestriksjoner: Norge digitalt begrenset

Brukerrestriksjoner: Lisens

Lisens: [No conditions apply to access and use](#)

Andre restriksjoner: Nedlasting av data begrenset til Norge Digitalt parter. For bestillinger og forespørsler kontakt en av Geovekst
- Forhandlerne

Sikkerhetsnivå: Ugradert

LICENSE

Restriksjoner

Bruksbegrensninger: Ingen begrensninger på bruk er oppgitt

Tilgangsrestriksjoner: Norge digitalt begrenset



Brukerrestriksjoner: Lisens

Lisens: [No conditions apply to access and use](#)

Andre restriksjoner: Nedlasting av data begrenset til Norge Digitalt

- Forhandlerne

Sikkerhetsnivå: Ugradert

English 

European Commission > INSPIRE > INSPIRE registry > INSPIRE metadata code list register > ConditionsApplyingToAccessAndUse >
no conditions to access and use

INSPIRE registry

no conditions to access and use

URI	http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse/noConditionsApply
This version	http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse/noConditionsApply:2
Version history	http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse/noConditionsApply:1
Label	no conditions to access and use
Definition	No conditions apply to access and use.
Collection	Conditions Applying To Access and Use
Governance level	Technical (EU)
Status	Valid
Insert date	2023-07-28 12:08 PM UTC
Edit date	2023-07-28 12:08 PM UTC
Available formats:	XML Registry XML ISO 19135 RDF/XML JSON ATOM ROR

CONCLUSION

- Data isn't always good
- Data isn't always free
- Data isn't always easy to access
- Data and tools don't always work well together, but a bit of convincing might help
- **If your results look weird or disagree with other part of your work, don't brush it off, figure out what's up!**
- **Thank you for coming to my TedTalk!**