

Deltares

Reproduceerbare projecten advies voor verdere uitwerking in NHI

Joeri van Engelen

Huite Bootsma

Gerrit Hendriksen

Januari 2022

Projecten meestal onreproduceerbaar

1. Niet alle stappen gescript
2. Geen versiebeheer scripts en data
3. Geen automatische workflow van basisdata naar resultaten

Onreproduceerbare projecten maakt samenwerken met collega's en anderen bijna onmogelijk



Overzicht van open-source gereedschappen



Python als scripting taal

Scripts alleen maken nog geen reproduceerbaar project!



git

Git voor versiebeheer scripts



DVC voor versiebeheer data



snakemake

A framework for reproducible data analysis

Snakemake als slimme workflow manager

Versiebeheer scripts



Folder zonder Git

```
Analyse.py  
Analyse_v1.py  
Analyse_v2.py  
Analyse_v2a.py  
Analyse_v2b.py  
Analyse_v2_final.py  
Analyse_v2_final_definitief.py  
Analyse_v3.py
```

Opruimen folder vernietigt ontwikkelgeschiedenis!

Folder met Git

```
.git/  
Analyse.py
```

Geschiedenis opgeslagen in .git/

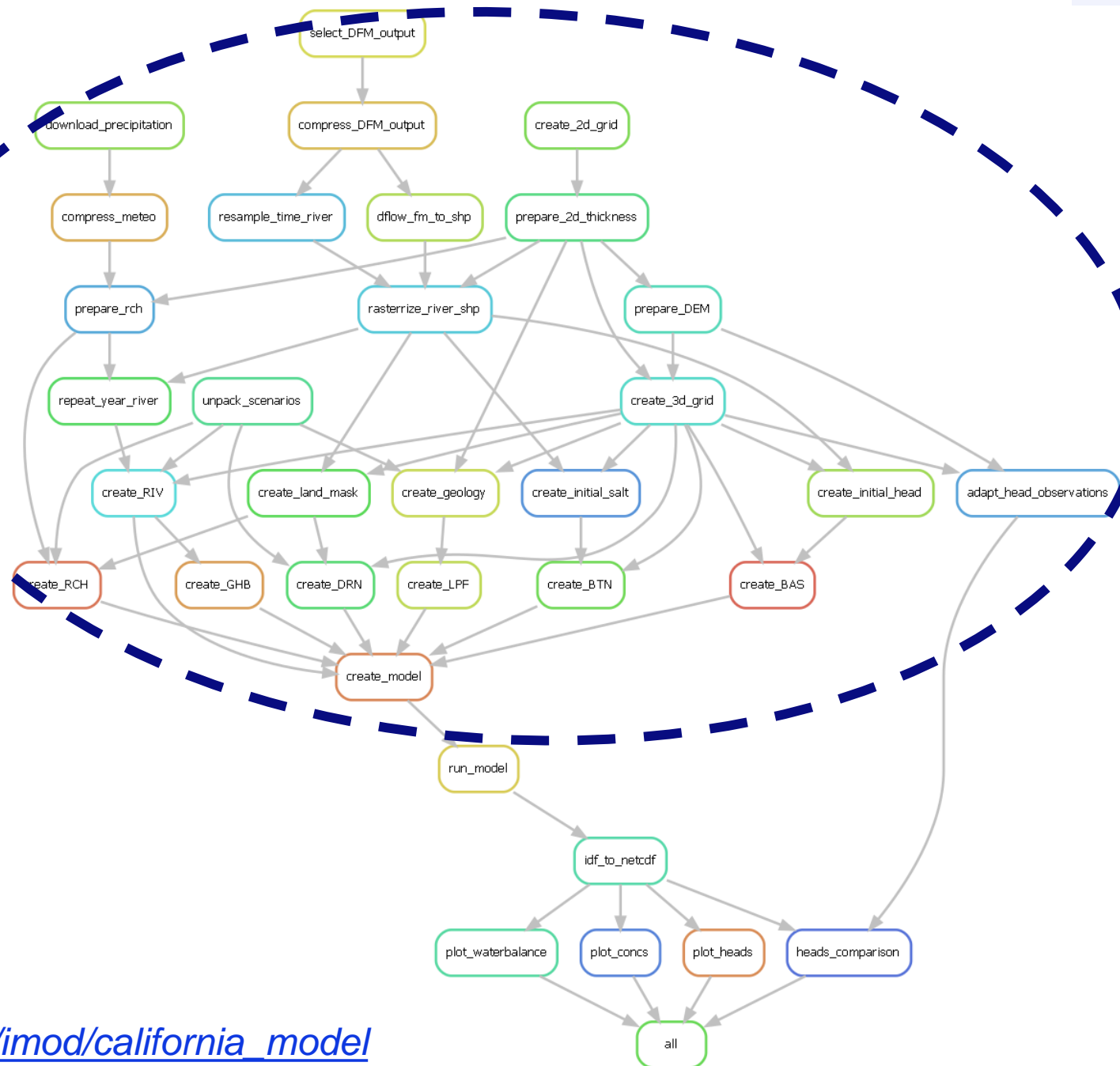
→ *Git geeft overzicht historie en helpt schakelen tussen versies*

Complexe workflows

Benodigde voorbereidingen om tot modelinvoer te komen!

Een snakemake workflow:

- Definieert expliciete data afhankelijkheden
- Hieruit volgt de volgorde scripts
- Na veranderingen: herhaalt automatisch alleen afhankelijke scripts
- Visualiseert workflow



Model San Francisco Bay

Vb toepassing: NHI-zoet-zout

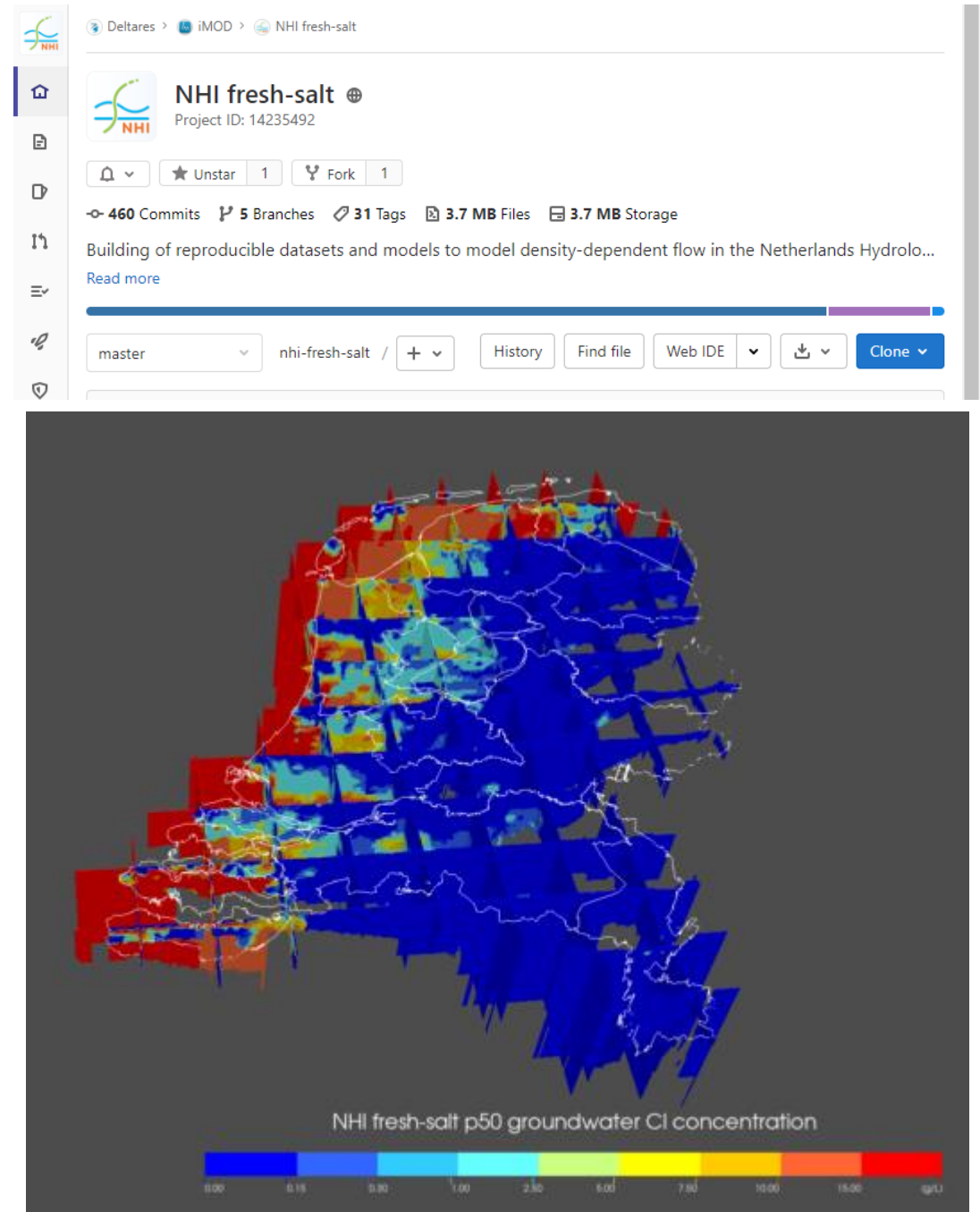
iMOD-WQ model

50 lagen, 1300 rijen, 1200 kolommen

- Volledig gescript
- In versiebeheer
- 1 workflow van brondata naar plaatjes uitvoer

<https://gitlab.com/deltares/imod/nhi-fresh-salt>

Deltares



Conclusies

1. Enkel scripts maken een project nog niet reproduceerbaar
2. Reproduceerbaar maakt overdraagbaar
3. Open-source gereedschappen beschikbaar om project reproduceerbaar te maken

Wij adviseren voor toekomstig versiebeheer NHI (data, scripts, modelcodes):

Ontwikkeling van het gereedschap

- Open-source
- Gedeeld op een *NHI Github* pagina

Data en NHI toepassingen (regionale en landelijke modellen):

- modelbouw reproduceerbaar
- scripts in Git, gekoppeld met een Snakemake workflow
- (basis)data moet ergens versiebeheerd zijn (regionaal of NHI dataportaal)

Meer informatie o.a. op <https://deltares.gitlab.io/imod/imod-practical-tips/>