SPN 10

2022 GRAZ, AUSTRIA

IWA
the international
water association

IAHR
AIRH
International Association
for Hydro-Environment
Engineering and Research
Hosted by
Spain Water and IWHR, China

TU
Graz.

CONFERENCE ON SEWER PROCESSES AND NETWORKS

Markus Pichler, David Camhy, Georg Arbesser-Rastburg,
Jannik Schilling
Workshop: Data and Models

Дмитрий Пешехонов –AdobeStock

Agenda

David Camhy and Georg Arbesser-Rastburg

- Data management
- Model and algorithm management

Markus Pichler

- Python interaction with SWMM (swmm-api)

Jannik Schilling

- Generate_SWMM_inp: An Open-Source QGIS Plugin to Import and Export Model Input Files for SWMM

SPN10
2022 GRAZ, AUSTRIA

Data Management

Conference on Sewer Processes and Networks
24–26 August 2022

SPN10
2022 GRAZ, AUSTRIA

Collected Data Types

- **Time series data:** Measurement data and simulation results
    - Several hundred stations —> billions of data points (precipitation, water quality, pressure, flow, temperature, infrastructure monitoring, …)
    - Different temporal resolution (100 ms and upwards)
    - Long time series (up to 70 years)
    - Complex values (Arrays/Spectra)
    - Simulation results (several hundred values / timestamp)

- **Geographic Information System** data
- **Arrays** (climate forecast data)
- **Media** (pictures and videos)
- **(Arbitrary analysis results)**

SPN10
2022 GRAZ, AUSTRIA

The Right Tool for the Job

- **Time series data: InfluxDB**
  - Easy visualization and alerting possibilities using Grafana
  - Very fast aggregation and analysis, efficient storage
- **Array data: TileDB and Parquet**
  - TileDB: Dense and sparse (!) array data support, parallel r/w support, versioning, fast subsetting
  - Parquet: Efficient storage format for local and distributed analysis
- **Vector GIS data: PostGIS (based on PostgreSQL)**
  - Geo-relational database
- **Dissemination/Publication: netCDF4**
  - Scientific data formats with standardized metadata (Climate and Forecast Metadata Conventions)

Measurement Metadata
based on sensorthings (OGC Standard)

- Station and sensor metadata management
  - Geolocation, units, operational data, project metadata, ...
- Custom developed web application (UI) based on Fraunhofer's FROST Server
- Standardized description of things, sensors, datastreams, observed properties,...

Data and Analysis Visualization – Grafana

- Powerful web-based time series visualization solution
- Easy to setup and use
- Supports alerting and annotations

Conference on Sewer Processes and Networks
24–26 August 2022

SPN10
2022 GRAZ, AUSTRIA
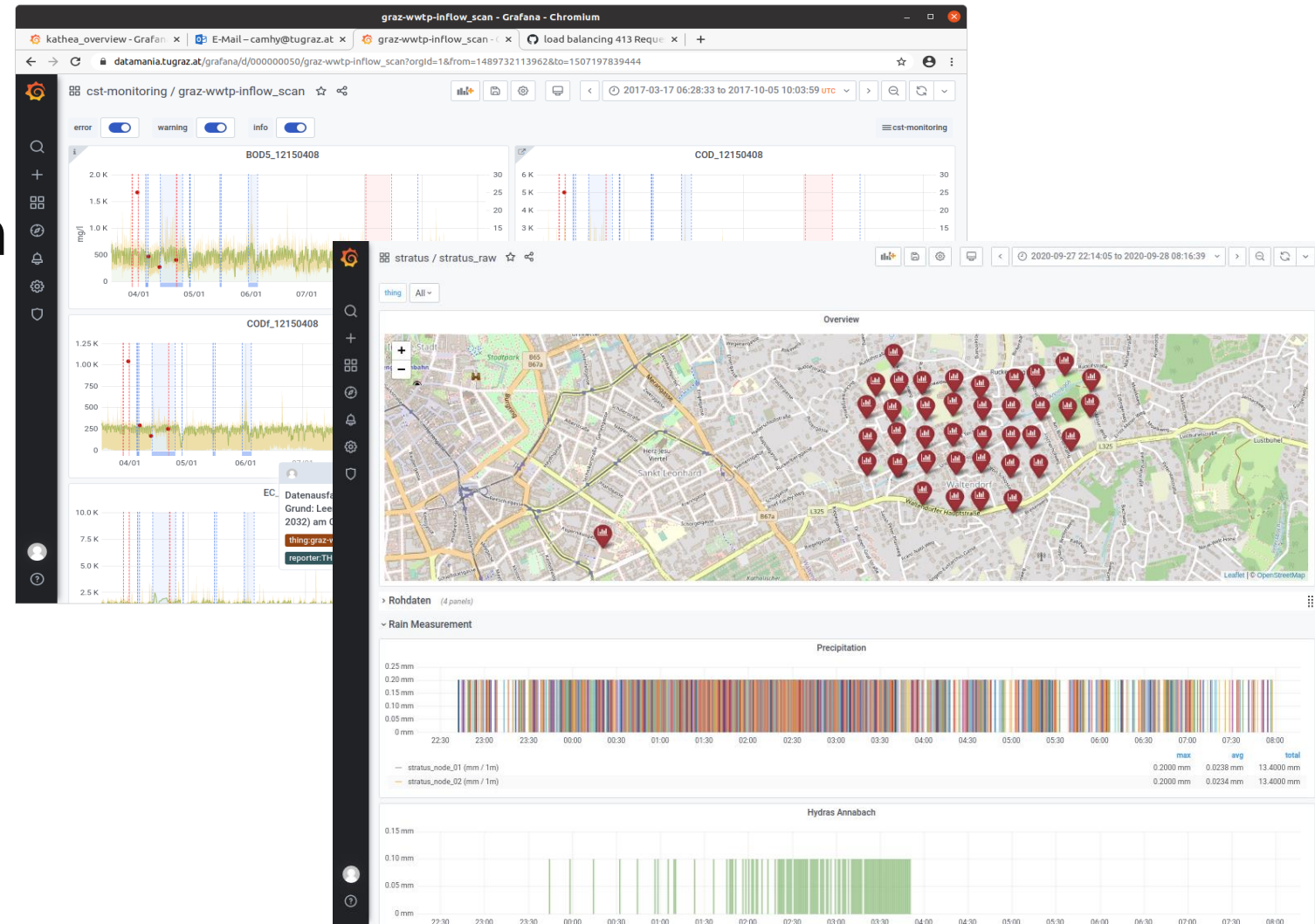
Data and Analysis Visualization – Dash

- Python toolkit
- Generates interactive visualization and analysis web-apps
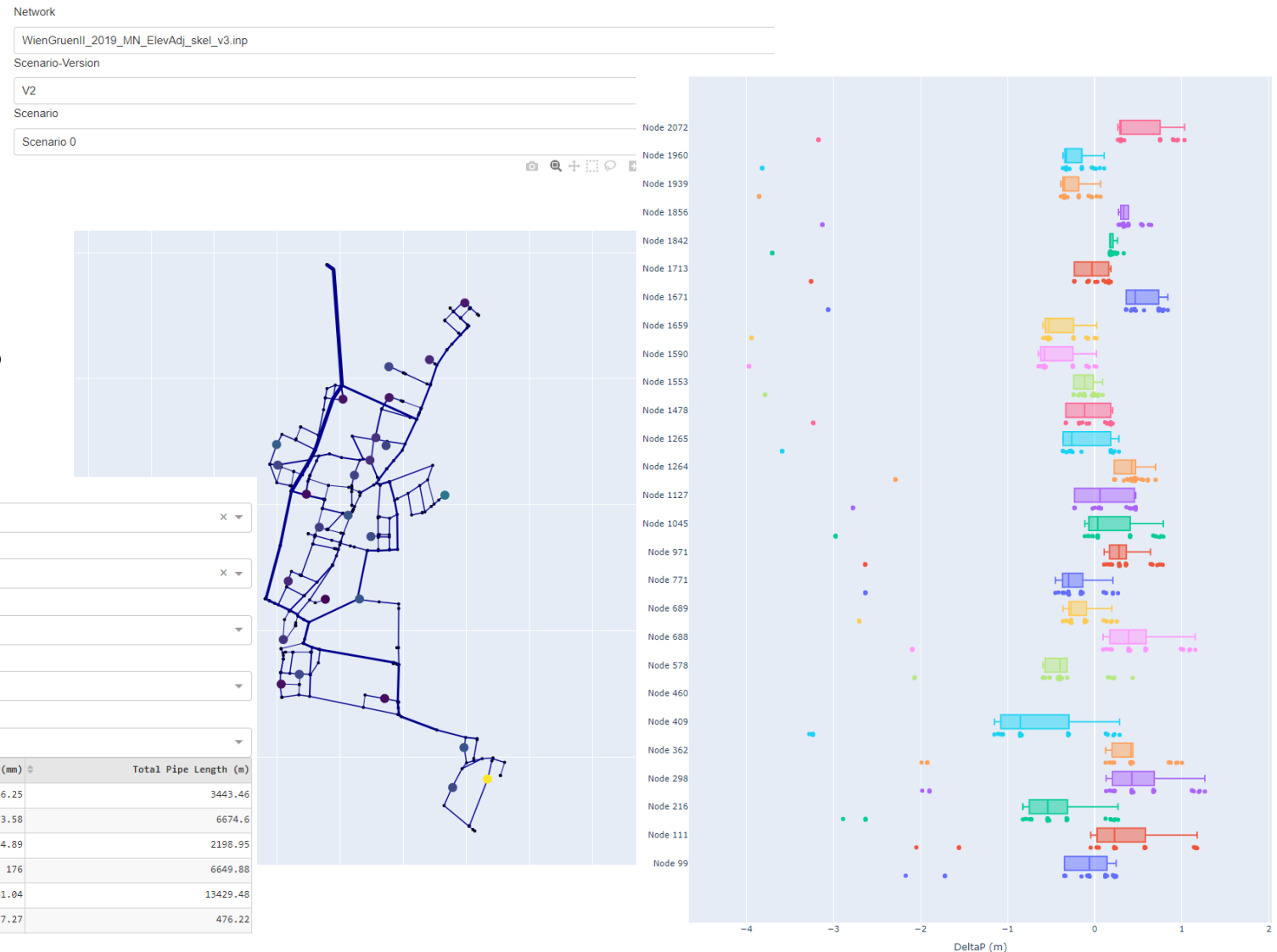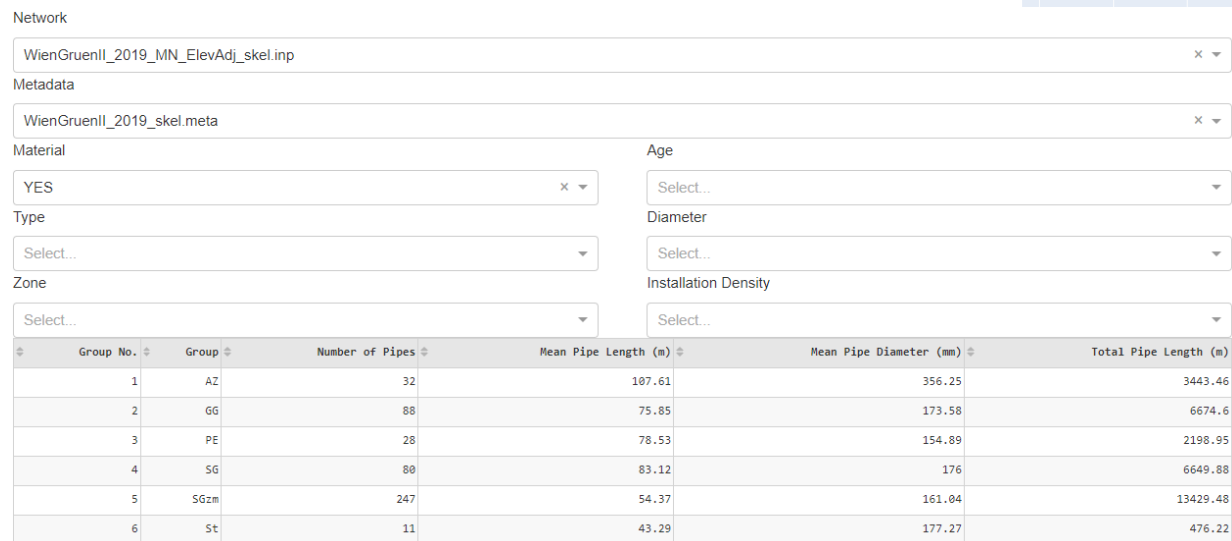
Reporting

- Weekly report of all precipitation measurements in Graz
- Multiple visualizations and statistical analysis
- Sent to all partners automatically once per week

Markus Pichler

Discussion

What solutions do you use to store which kind of data?

How do you store your measurement metadata?

Do you use web based visualization and analysis solutions?

How do you exchange data with project partners?

Which formats do you use for dissimination and publication?

Workshop: Data and Models
23 August 2022

SPN10
2022 GRAZ, AUSTRIA

Model and Algorithm Management

SPN10
2022 GRAZ, AUSTRIA

Model Storage

- Models
  - EPANET models (inp-files)
  - SWMM models (inp-files)

- Current solution
  - Models stored on network drives
  - Models stored in Git repositories
  - Models only locally available on workstations

- No central model storage, versioning only in Git repositories

Conference on Sewer Processes and Networks
24–26 August 2022

SPN10
2022 GRAZ, AUSTRIA

Algorithm Management

- Algorithms managed in Git repositories
  - General algorithm and script repository
  - Project-related repositories
  - User-related repositories
  - Algorithm-related repositories

Conference on Sewer Processes and Networks
24–26 August 2022

SPN 10
2022 GRAZ, AUSTRIA

Algorithm Execution

- **HTCondor**
  - Supports many (in our case up to several 100000) long running simulations
  - Allows execution on user workstations (Linux and Windows)
- **Jenkins**
  - Scheduled and interactive jobs
  - Data collection and analysis, educational jobs
- **Argo Workflows (work in progress)**
  - Complex pipelines for distributed cluster workflows
- **External Clusters/Supercomputers**
  - Is used when Institute infrastructure is not sufficient (RAM, CPUs, …)

Issues we face

- Results are stored (different databases, files on a server, …) but there is no link to the algorithms and models used

- Running old algorithms again takes time to set up (unclear dependencies, no dependency versions)

- There is no central model storage → models are scattered and hard to find

- No or not standardized model metadata

Conference on Sewer Processes and Networks
24–26 August 2022

SPN10
2022 GRAZ, AUSTRIA

Open Model Data Management - OpenMDM

- Tool with web-based user interface for
    - Managing hydraulic water distribution and sewer models, model metadata and model versions as well as results
    - Starting containerized computations in a computing cluster

- **Goals**
    - Creating a framework that allows for containerized workflows that
        - Extract metadata from models
        - Store (versioned) models and the corresponding metadata in a central storage
        - Start modelling tasks in computing clusters
        - Support different execution backends like HTCondor or Argo Workflows
        - Store simulation artifacts in a central storage

SPN10
2022 GRAZ, AUSTRIA

OpenMDM – Current Concept
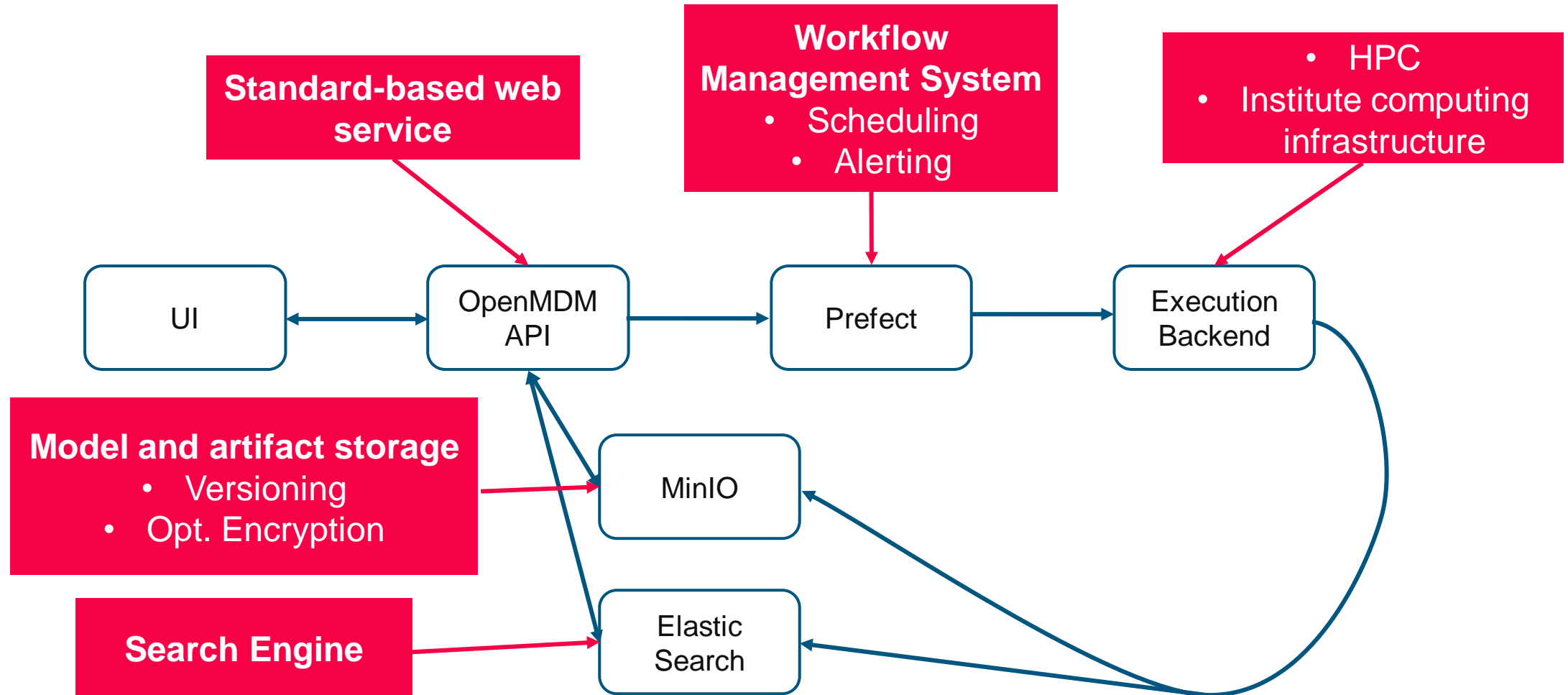
Discussion

How do you store models and simulation artifacts?

How do you manage your algorithms?

How do you ensure reproducible workflows?

What kind of model and simulation metadata would you need for your scientific work?

SPN 10
2022 GRAZ, AUSTRIA

swmm-api

Markus Pichler

SPN10
2022 GRAZ, AUSTRIA

Why?

Why SWMM?
- Open Source
- Widely used in science (so do we)

Why Python?
- also widely used in science
- reasonably fast
- relatively easy to learn

SPN10
2022 GRAZ, AUSTRIA

Why another one?

Bryant E. McDonnell (UDM 2022): "Everyone was just wrapping SWMM Input files and SWMM Output files with Python"

Aren't there already a lot of them?

- yes, but …
    - availability
    - speed
    - usability
    - extensible

Based on the SWMM command line interface

What information is in the .inp-file

How the .inp-file is structured

Main functionalities

- Reading SWMM input files (.inp)
- Modifying model components and settings
- Simulating models using SWMM command-line tool or PySWMM
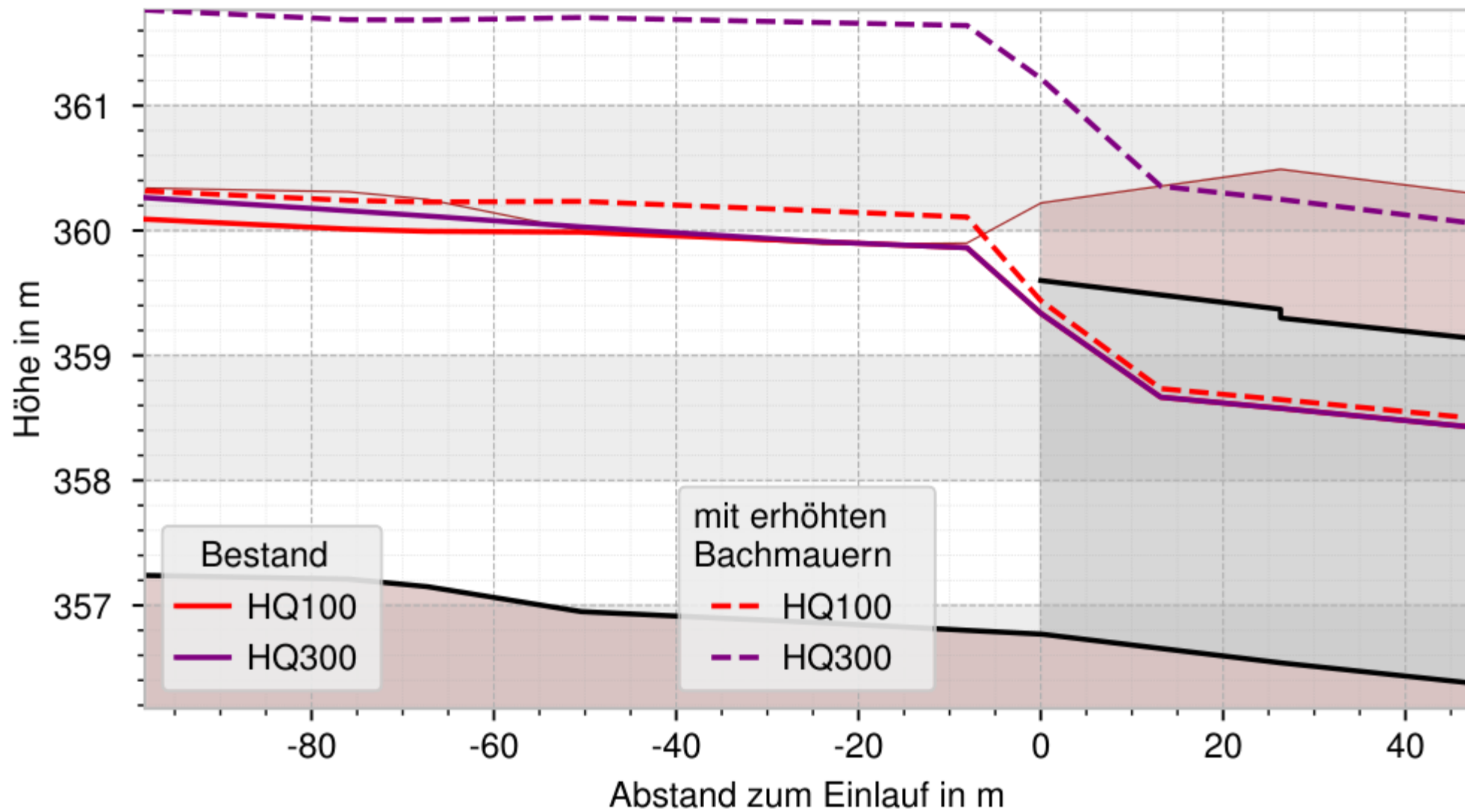- Reading results (.out and .rpt) as pandas data objects

Use cases

event based long term simulations

sensitivity analysis

creating a SWMM model with GIS data

working with a modified SWMM

creating beautiful plots for publications

SPN10
2022 GRAZ, AUSTRIA

Workshop: Data and Models
23 August 2022

SPN10
2022 GRAZ, AUSTRIA

Links

Installation

- `pip install swmm-api`

- https://pypi.org/project/swmm-api/

Documentation

- https://markuspichler.gitlab.io/swmm_api

Examples / Tutorials

- https://markuspichler.gitlab.io/swmm_api/examples/

SPN10
2022 GRAZ, AUSTRIA

Files in this workshop

https://gitlab.com/markuspichler/swmm_api/-/tree/master/examples/spn10

Workshop: Data and Models
23 August 2022

SPN10
2022 GRAZ, AUSTRIA