Technical Specification Double-side Page

1. **TECHNICAL SCOPE:** Summarize the solution developed during the EXPERIMENT phase: how have you finally addressed the challenge/Theme Challenges and tackled with its requirements and data. Include a diagram.

The project consists in generating a water level prediction of the Po and Dora Riparia rivers in the City of Turin in order to prevent the drainage pipes, managed by the SMAT (DataProvider), from clogging and pressurising the sewage network. The problem has been addressed with a physically based approach merged with machine learning (ML) technologies to ensure the highest level of accuracy. The model design can be divided into 3 blocks:

- Data Ingestion: according to the data provider requirements, we implemented a shared directory accessible via SFTP to upload and download the data. We focused on the data we need to make the MVP but we also developed software to cover different data scenarios.

- Modeling: we built a flexible system to integrate different triggering scenarios for the ML model. During the experimentation the data provider asked to integrate the official forecast from the local regional meteo agency (Arpa Piemonte) as the trigger of the ML model that can forecast the river level up to 12h.

- Results distribution: outputs are uploaded to a SFTP directory that is accessed by the data provider who eventually feeds a specific web interface inside the official data controller.

2. **ALGORITHMS, TOOLS AND CONCLUSIONS:** Detail the algorithms and tools finally selected to accomplish the challenge/Theme Challenges. Summarize the main results that you have obtained during the EXPERIMENT phase: data, insights, conclusions and the main contributions to solve the challenge/Theme Challenges.

- Data ingestion: the river level measurements, provided by Arpa Piemonte to the data provider, are fetched by us and stored in a PostGreSQL database. The data are made available every 30 min (update time of the data) and managed by Celery (synchronous task queue).

- Modeling: if the forecasts from ARPA Piemonte suggest the exceedance of an "attention threshold", the solution triggers a ML procedure (Ridge Regression Model) that delivers hourly forecasts of river level up to +12 h lead time on the target points defined by the data provider. The results are updated on the shared SFTP directory and made accessible to the Data Provider.

- Results distribution: The results are fetched by the Data provider and are then sent to the specifically developed page in WinCC (the current SCADA system) in order to be available to the control room for the decision making.

3. **SCALABILITY AND FLEXIBILITY OF THE SOLUTION:** Explain how the solution copes with the challenge/Theme Challenges requirements and how can it be adapted to other similar problems. What work is still pending to create a real/stable product if any? What TRL level is it in?

The flexibility of the solution is provided by the fact that it can integrate external forecasts (see Arpa Piemonte in this case) transparently by excluding the specific hydrological part in the modelling chain, without disrupting the whole system. In case the forecasts were not available, the system could be reprogrammed to also launch the hydrological model.

Scalability is ensured on several factors:

- the use of a microservices approach, based on Docker technology (e.g. databases, ingestor and whole procedure are dockerized), which communicates using Rest API protocols for data transmission (via intranet, internet or cloud). In order to optimise the storage of ever-increasing data, the use of different databases for different types of data ensures the best data management from the IT and data usage point of view.
- in case of missing in situ meteo data, it can use global datasets (GFS, ERA5) that can

complement or replace meteo data acquired from local providers;

- The solution is also applicable in other business sectors where different update frequencies and forecast horizons are needed. For example: a) civil protection: they need predictions for flood alert; b) river transportation companies: they need to predict the river navigation with a 10 days advance; c) hydropower companies: they expect a forecast with 72 h lead time for the optimization of energy trading.

4. **DATA GOVERNANCE AND LEGAL COMPLIANCE:** Describe the security level of the solution, i.e. how authentication, authorization policies, encryption or other approaches are used to keep data secure. Explain how the solution is compliant with the current data legislations concerning security and privacy (e.g. GDPR).

The results are accessible only using ssh protocols with the use of public/private rsa keys, as well as the source code on the versioning platform.

The data shared by Arpa Piemonte are released according to a licence Creative Commons BY 2.5 Italy;

5. **QUALITY ASSURANCE AND RISK MANAGEMENT:** Describe the quality process followed for the final product. Technologically, which problems have you encountered and how you have solved them, and any processes followed that guarantee that the solution fulfills the challenge/Theme Challenges and data provider requirements.

The KPIs chosen with the DataProvider concern the goodness of prediction itself (e.g. R2),

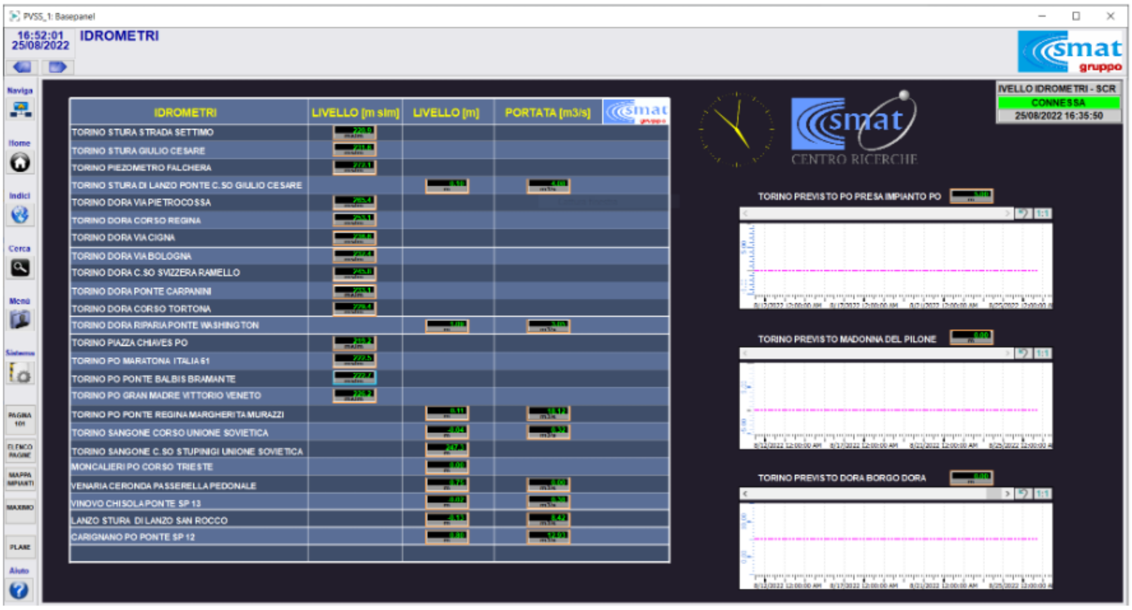
The critical issues we faced were:

- Unavailability of real time in situ meteo stations: initially Arpa Piemonte didn't want to share the meteo data in real time, needed to feed the hydrological model. → We solved this issue by fetching the flood bulletin directly to trigger the ML model.
- Real time river measurements: the data are not robustly updated every 30 min as in the specifications. → The data provider will negotiate with Arpa Piemonte to fix the problem.
- Operational phase: in case the real time river measurements are provided with an excessive delay, the forecast procedure may not work → we want to reconstruct the in situ measurements based on the data we have.

Annex 1. Means for accessing the MVP

Please, indicate in 1 page maximum the means for accessing the MVP for a potential customer (login information, website address, link to a demo video or whatever means are needed to check that the MVP exists and works).

The solution is already implemented in the SCADA system (WinCC) of our data provider. The system is triggered only if the bulletin from ARPA Piemonte exceeds a given “attention threshold”.
In the image below (updated 2022-08-25) one can see an example of the output inside the WinCC interface from the data provider. This system has been officially presented inside the Festival dell’acqua, the most important exhibition in Italy on the water sector, in a dedicated presentation given on 23 September 2022.



Operative Diagram for SMAT