TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Lizaveta Miasayedava IVGM184361

AUTOMATED ASSESSMENT OF ENVIRONMENTAL FLOWS USING ESTONIAN HYDROLOGICAL OPEN GOVERNMENT DATA

Master's thesis

Supervisor: Jeffrey Andrew Tuhtan Dr.

Co-supervisor: Keegan David Braun McBride MSc.

Tallinn 2020

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Lizaveta Miasayedava IVGM184361

KESKKONNAVOOLUHULGA AUTOMAATNE HINDAMINE KASUTADES EESTI AVALIKU SEKTORI HÜDROLOOGILISI AVAANDMEID

Magistritöö

Juhendaja: Jeffrey Andrew Tuhtan Dr.

Kaasjuhendaja: Keegan David Braun McBride MSc.

Tallinn 2020

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Lizaveta Miasayedava

07.05.2020

Abstract

Rapid environmental changes emphasize the need for a responsive approach to scientific observation and interpretation. With the frequently changing dynamics of rivers, it is essential to enable the real-time monitoring and automation of the analysis of river ecological status, which can be numerically estimated with commonly adopted environmental flows calculated in multiple conventional ways.

The EU Water Framework Directive encourages better water management and protection of riverine ecosystems that can be achieved with the use of environmental intelligent information communication technologies. Hydrological open government data can substantially facilitate the analysis of the collected data in order to extract operational insights as well as the creation of innovative services allowing its automation.

The objective of the thesis project is the design and implementation of the environmentally intelligent web-service allowing automated estimation, analysis and forecasting of environmental flows of Estonian rivers using Estonian hydrological open government data to ensure the monitoring and reporting of the compliance with EU Water Framework Directive. The methodology used for the project is design science research methodology reinforced with the environmental intelligence framework.

Addressing the gaps of widely used hydrological standard-setting formula methods of environmental flows assessment, the Polish regionally applicable environmental low flow formula method is integrated to the developed web-service as a scalable alternative to existing techniques incorporating biological response along with the hydrological variability estimation.

This thesis is written in English and is 115 pages long, including 6 chapters, 60 figures and 1 table.

List of abbreviations and terms

ICT	Information Communication Technology
EU	European Union
SDG	Sustainable Development Goal
SWM	Sustainable Water Management
IoT	Internet of Things
NGO	Non-Governmental Organisation
OGD	Open Government Data
EIS	Environmental Information System
EIM	Environmental Information Management
INSPIRE	EU Directive establishing Infrastructure for Spatial Information in the European Community
WFD	EU Water Framework Directive
EI	Environmental Intelligence
Eflows	Environmental (or ecological) flows
EWS	Estonian Weather Service
RegTech	Regulatory Technology
EnvRegTech	Environmental Regulatory Technology
AAF	Average Annual Flow
RAELFF	Regionally Applicable Environmental Low Flow Formula
EMC	Environmental Management Class
FDC	Flow Duration Curve
FET	Fish Ecological Type
UCUT	Uniform Continuous Under Threshold
HST	Habitat Stressor Thresholds
DSRM	Design Science Research Methodology
EFCES	Environmental Flows Compliance Evaluation Service
SPA	Single Page Application
MTV	Model-Template-View

CRUD	Create-Read-Update-Delete
ERD	Entity Relationship Diagram
ORM	Object Relational Model
UI	User Interface
UX	User Experience
JS	JavaScript

Table of contents

List of figures

List of tables

1 Introduction

Digital transformation of governance is essential for unlocking various opportunities having a sustainable, social and economic impact. The use of Information Communication Technologies (ICTs) for delivering public services in a faster, less costly and more efficient way forms the basis of e-government enabling access to the services through ICTs and e-governance ensuring digital interaction between parties of different sectors. [1].

E-government also facilitates the transformation of existing practices and processes for increasing the efficiency of administering operations of governments and businesses [1]. ICTs performing data analytics and predictive modelling have innovative potential allowing not only timely reporting and communication but also the implementation of transparent, accountable and participatory governance resulting in better grounded, evidence-based decisions for the whole society [2].

Currently, climate change leads to various environmental challenges raising concern of sustainability and resilience and putting more stress on decision-makers to address the protection and preservation of the environment within natural limits taking into consideration complex relationships between environmental, social, and economic systems [3].

Natural weather-related disasters occur increasingly frequently across the globe, and around 90% of them are water-related caused by storms, floods, droughts, and extreme temperatures. Every year they take the lives of thousands of people, affect millions of people, destroy their livelihoods, and lead to large economic losses. With population growth, more people are going to be affected, which requires the strengthening of disaster-resilient governance - disaster risk prevention and reduction measures [4]-[5].

Water management faces various challenges: inadequate infrastructure, water scarcity, extreme events, etc. In order to address these challenges, massive investments into improvement and maintenance of the infrastructure as well as the deployment of smart solutions for water protection, conservation and management are required [6].

Supporting Sustainable Development Goals (SDG) agenda [3], monitoring of environmental changes facilitates better control and management of high-risk areas, helps to detect disaster-prone zones and enable early warning, address and mitigate the degradation of the affected ecosystems [4]. Sustainable Water Management (SWM) involves using various physical technologies to address water-related extreme events such as dams and dykes, as well as digital technologies. Supporting IoT Guidelines for Sustainability [7], IoT sensors embedded into water management infrastructures for weather forecasting, drainage control, alerting, etc. allow minimizing flooding, stormwater runoff or property damage [4], [8]-[9].

The water-related data (precipitation, river flows, lake and groundwater levels, water quality, etc.) can be collected by different parties from different sources (e.g. satellites, ground sensors, meters and others). Opening the data to public access can facilitate real-time monitoring, analysis and control of water systems by stakeholders of different sectors (researchers, citizens, NGOs, government agencies, etc.) allowing a better understanding of the observed system and development of smart solutions addressing the issues and serving to various purposes: drinking water, fisheries, hydropower, etc. [6]. Water data are often used for decision- and policy-making and to create services improving life conditions and supporting the environment [10] with the prerequisite of the proper data quality.

Sustainable future of aquatic communities inhabiting rivers depends on the careful management and regulations of water resources. Various hydrological changes, such as excessive water withdrawal, pollution, etc. affect their habitat state and availability. The natural dynamics of rivers is defined by the regime of natural flows - a historical pattern of their hydrologic variation and monitoring of the flows is important to maintain healthy river ecosystems [11]-[12].

The EU Water Framework Directive (WFD) [13] adopted in 2000 establishes a framework for water policies committing EU member states to achieve the good qualitative and quantitative status of all water bodies (rivers, lakes, coastal waters, groundwater). However, defined by WFD environmental objectives and deadlines were met neither in Estonia nor Europe [96] and needed extension (currently set for 2021 and 2027) because of various uncertainties in definition of monitoring measurements, assessment procedures and forecasts, shortage of nature conservation and water protection areas, the necessity of adaptation strategies for river basin management plans under climate change, etc. [14]-[15]. The implementation of the WFD framework is based on the water management plans [96] aiming at the implementation of the water and river basin management principles and measures [97]-[99].

The (combined) status of surface water bodies (very good, good, moderate, poor and very poor) is defined based on the ecological status and chemical status, whichever is worse. In order to achieve the objectives of WFD, all the water bodies should have achieved at least good (combined) status [100].

Environmental flows (eflows) are commonly used to assess river ecological status. According to the Brisbane Declaration and Global Action Agenda on Environmental Flows [59, p.1], "Environmental flows describe the quantity, timing, and quality of freshwater flows and levels necessary to sustain aquatic ecosystems which, in turn, support human cultures, economies, sustainable livelihoods, and well-being." The eflows approach is based on the natural flow regime [11] and was further developed by "research and practice focused on aquatic ecosystem protection, restoration and management" [59, p.2].

The WFD common implementation strategy involves the definition and assessment of eflows and their adoption in river basin management plans [16]. The eflows were defined as the "amount of water required for the aquatic ecosystem to continue to thrive and provide the services we rely upon" [16, p. 2].

The complexity of flow regimes and conflicting needs of different water users complicates the process of choosing an appropriate methodology for determining flow requirements. Addressing the ecological needs of water of aquatic organisms, the methodology is supposed to model necessary characteristics of a river system closely and be a trade-off between simplicity and expense of the use of the method [17].

There are more than 200 methodologies to assess eflows [18]. Different countries use different methodologies: hydrological, hydraulic-habitat modelling, holistic methodologies, etc., and their combinations [16]-[20]. The methods vary in complexity and resource demands and expenses, and the choice of the methods is usually made conventionally depending on the availability of resources. Simple affordable methods lead to gaps and limitations addressed in various studies. In many cases, these simple widely used methods (e.g. Tennant method) have poor results while their application to different locations [16], [20].

In Estonia, the monitoring of the ecological status of water bodies is established and in general, considered to be satisfactory [21]. Fishing is also regulated and monitored to ensure good conditions of fish populations [22]. However, by the initial WFD deadline (2015), only 57.6% of Estonian surface water bodies were at least in a good status [96], and this percentage is gradually decreasing: in 2018 it was 53.9% [101] which was also explained by over-optimistic estimations of the experts having an insufficient amount of data at the beginning of assessment and changes in the assessment procedures [96].

To assess the ecological status of rivers with eflows, Estonia uses hydrological standard-setting formula methods [23]-[24]. The most commonly used eflows assessment method in Estonia dates back to 1972 [23]-[24]. According to the regulation of the Ministry of the Environment specifying requirements for the expansion of a water body, environmental monitoring related to the expansion, protection of aquatic life, dam, elimination of the expansion and lowering of the water level, and methodology for determining the minimum ecological flow [102], the minimum eflow is determined for the ice-free period from May to October by calculating the average monthly minimum flow with a 95% probability of being exceeded. Tennant hydrological methods (30% and 20%) and other percentages of exceedance probability (75%, 90%, 95%) of a monthly average of the minimum flow rates found separately for winter (October-April) and the summer season (May-September) were also recommended to be used depending on the type of a river [23]-[24].

The hydrological standard-setting methods used in Estonia as well as in many other places using similar approaches have the following limitations:

- The methods are not designed to consider compound events: they cannot estimate eflows based on other than flow rate flow metrics such as water temperature.
- The biological response is not considered since the large scale application of biological rules for biological responses assessment is complicated, being too detailed and resource-consuming.
- The methods are scalable and adaptive: they are not able to dynamically link the eflow assessment with real-time monitoring in different locations.

These gaps can be addressed with the Regionally Applicable Environmental Low Flow Formula (RAELFF) which is easy to define (standard-setting formula), applicable over different regions and sufficiently detailed for capturing the biological response to environmental change [17]. This method is also suitable for real-time monitoring and can be used for the automation of eflow estimation.

In order to design and implement the automated assessment of eflows, the best practices of Regulatory Technologies (RegTech) [25] and Environmental Intelligence (EI) framework [26]-

[27] can be used. This enables automated real-time environmental data collection, processing, analysis, and in the same manner can involve automated real-time compliance evaluation. After embedding the functionalities into software, the resulting artefact can be seen as an Environmental RegTech (or EnvRegTech) artefact providing the automated assessment of eflows to ensure the compliance of rivers with WFD.

The main objective of the thesis project is to design and build a web-based service enabling the automated assessment of eflows (Environmental Flows Compliance Estimation Service, EFCES) applied to Estonian rivers using the RAELFF approach and hydrological Open Government Data (OGD) managed by the Estonian Environmental Agency and Ministry of the Environment [28]. The software should be designed to serve for results interpretation, scenario modelling and forecasting as well as decision support to various stakeholders: hydrologists – experts in river flow rates, freshwater biologists using the flow information and thresholds to assess risks for fish populations, and water managers making reports on the ecological status of rivers.

The methodology for the thesis aligns with its main research objective, and it is the Design Science Research Methodology (DSRM). The research is carried out following DSRM steps (refer to Chapter 3.1 for details).

The theoretical framework (Chapter 2) is devoted to the review of SWM, OGD, EI framework and eflows in general and in Estonia. Chapter 3 includes research questions and objectives, research methodology and the plan of its implementation. Chapter 4 describes the service design and implementation. It also includes the analysis of the Estonian hydrological data as well as forecasting modelling. Chapter 5 provides an evaluation of the developed service and discussion of the achieved results. The final Chapter 6 presents a summary of the thesis project research and further suggestions.

2 Environmental Intelligence Framework for Sustainable Water Management using Open Government Data

2.1 Environmental Intelligence

The Environmental Intelligence (EI) concept implies the use of technologies for the development, integration, and expansion of data processing methods (e.g. statistical predictive modelling) and actionable environmental information into a multilateral global communication network for observation, forecasting, maintenance, and control [29].

The EI paradigm offers various opportunities for the environmental community facilitating the decision-making of users of different sectors (Figure 1) [27], [30]:

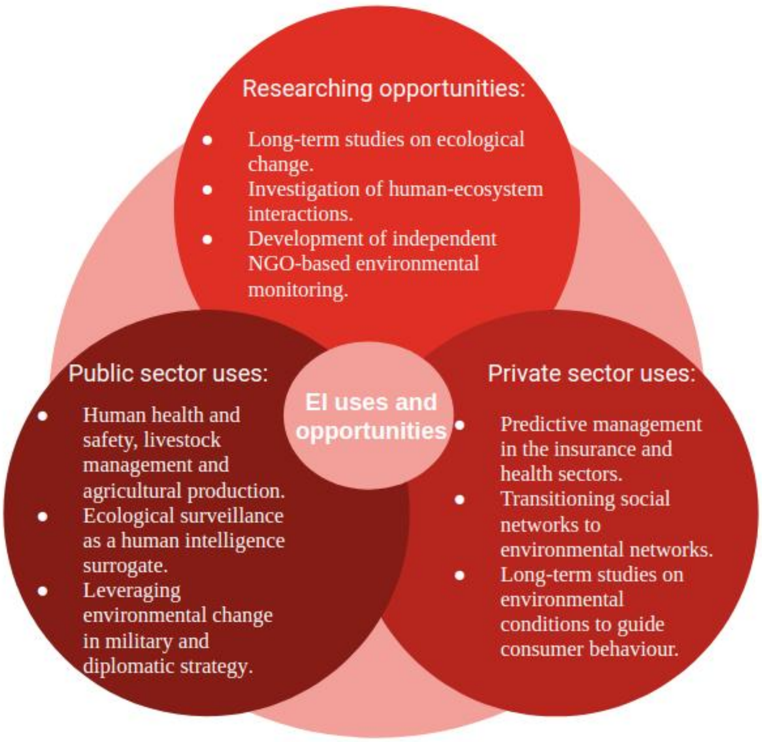


Figure 1. EI uses and opportunities. Author: [30]

The iterative process of gathering environmental knowledge can be described with the EI Cycle process (Figure 2), [26]-[27].

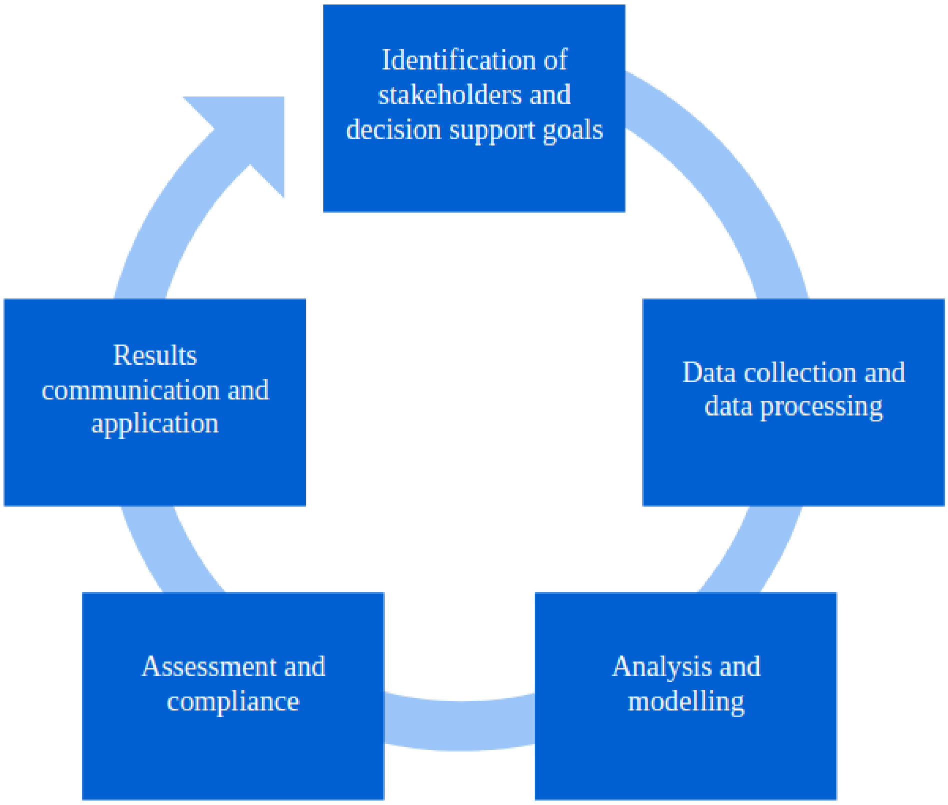


Figure 2. EI cycle. Author: [26]-[27]

The EI cycle starts with the identification of relevant stakeholders and decision support objectives depending on the environmental focus area. Data collection and processing involves spatial and temporal environmental data from the focus area. Analysis and modelling allow extracting functional evidence from the collected and processed data and capturing feedback from the explored system. The new environmental knowledge is evaluated during the assessment. Depending on the needs of the stakeholders, compliance estimations can also be carried out with the provision of options for action to decision-makers. After that, results are communicated and applied. At the same time, EI framework is iterative and can imply revisiting previous steps responding to the needs of stakeholders [26].

The efficiencies of the EI solution may be used for business risk management, critical infrastructure protection, disaster preparedness, response and recovery, community planning for resiliency, military operations, etc. [27].

The EI framework has a great operational value for environmental compliance management stakeholders (e.g. managing compliance with various environmental regulations). With the automated environmental compliance resolution, the software can automatically capture, store, analyse and visualise real-time data as well as provide forecasting, simulation and impact modelling abilities, satisfying the evolving operational needs of stakeholders.

2.2 Environmental Open Data Governance

2.2.1 Open Government Data Overview

Public sector bodies produce and collect various types of data and information that can be strategically valuable to public sector agencies, private businesses, academia, citizens and civic organisations. After providing open access to government data, the Open Government Data (OGD) can be further combined with data from other sources: smart devices, social media, etc. With the distribution, enrichment and reuse of the data, governments promote the creation of new businesses and innovative, value-added services involving the use of ICTs, data analytics and predictive modelling and improving public values: security, safety, transparency and accountability – through performance monitoring, reporting, planning, and policy-making. The creation of more agile and targeted to users' needs OGD-driven services can stimulate a competitive marketplace for the public sector and thereby ensure more inclusive service delivery and participatory economy [2], [31]-[32], [103]-[104].

The policies of OGD encourage the right of access to the governmental information supporting open government movement and handling information asymmetry. They also raise awareness of the government policies, activities and programmes as well as practices ensuring efficient use of the data [2], [103].

The government datasets can be found on OGD indices and portals if they are registered there or on the government services. The open datasets can be used for various purposes and applications (e.g. comparison, trend analysis, etc.) by all the interested stakeholders without any restrictions from copyright or patents on use or distribution [32]. Open Government Data (OGD) can include registers, legal, transport, social, geographic and meteorological information, etc. presented in human-readable and machine-readable formats [31].

With more diverse data available, the OGD-driven services can provide more crucial insights into a situation improving its overview and allowing zooming in into the situation. The extracted new knowledge is often interpreted through dashboards in the form of tables, graphs, heatmaps, etc. that can be integrated into e-governance services to support decision-making and policy-making processes or to communicate and interact with the public in a transparent and accountable way [2]. OGD fosters evidence-based decision-making expected to make the decisions better grounded and make the stakeholders aware of the consequences of their choices [2], [103].

However, the integration of the dashboards into decision-making processes might be challenging and risky [2]. The design of such services is challenging and has many prerequisites: deep understanding and analysis of the data, policy-making, organisation, legislation and public values, etc. The quality of data is essential since discrepancies in data can lead to inappropriate analysis. The services require proper design, interpretation and validation since the confusion and misconception about outcomes can result in wrong decision-making affecting the transparency, accountability and trust in the government. Moreover, the use of the services should be complemented by the organisational changes. If the communicated insights are not addressed anyhow, the services might turn out to be abandoned [2].

OGD requires leadership, active involvement and interest of all the parties. Not being considered as a tool to increase government efficiency and generate savings, the OGD practices may lag even with a mature e-government system [105].

In the spheres, where the strategic value of the OGD is acknowledged at the pan-European level, the EU legal obligations enforce the implementation of good OGD practices. One of the examples is the management practices of environmental OGD [106].

2.2.2 Environmental Open Data Management

The environmental information includes data on the state of the environment, implementation of measures and the effects of their environmental policies [109]. Environmental Information Management (EIM) involves reporting and monitoring for regulations ensured with Environmental Information Systems (EISs) [107], [109].

The environmental data collection is required to be organised in such a way that it is easy to find and share, thus environmental OGD often form the basis of EISs. However, environmental OGD and EISs of different domains are often scattered, spread out in different places, lack metadata and have different specifications and formatting [109].

Monitoring and reporting of environment legislation aims at streamlining monitoring obligations [108] and promotes the creation of data-driven services and tools as well as good practices of handling publicly available data through data harvesting and mining on the national and global levels serving environmental policy needs: reporting and e-reporting, efficient data harvesting and information gathering, active dissemination, transparency, participatory democracy, accountability, the accomplishment of the Digital Single Market Strategy, etc. [109], [111]. The practices promote targeted environmental data collection through active dissemination, unification practices ensuring compatibility and interoperability of the environmental data that are expected to lower administrative reporting burden [109].

In order to ensure accessibility and interoperability of the spatial geographical data, the INSPIRE Directive introduced in 2007 [112] established an infrastructure for spatial information in Europe targeting the purposes EU environmental policies and various environmental applications and purposes supporting sustainable development. According to INSPIRE, there should be ensured common legislation and implementation rules for metadata, interoperability and sharing of spatial data and services, network services, monitoring and reporting to make the spatial data infrastructures of the EU member states compatible and usable within the country and in the transboundary context [112].

2.2.3 Estonian Environmental Open Government Data

The environmental OGD management of Estonia is centred around 2 environmental portals: one of the Environmental Agency (keskkonnaagentuur.ee) and one of the Ministry of Environment (envir.ee) supplying information on a number of domains on the national environmental policies, legislation and plans, reports on the state of the environment, information on permits and licenses, monitoring data as well as high level indicators. These portals also show some monitoring data as tables/graphs with indicators, while spatial data can be found in the register portal (register.keskkonnainfo.ee). Some environmental domains have different EISs: for example, for the Air directive, a separate portal of the Estonian Environmental Research Centre is available (airviro.klab.ee). For INSPIRE, Estonia has a separate INSPIRE portal, inspire.maaamet.ee not integrated into the main environmental portal [110].

Government has various sources of hydrological data and making it public facilitates better analysis of the data. The analysis of hydrological data is essential for agriculture, e.g. harvesting depends on the amount of water available. The information obtained through analysis can be used for planning, farm advising, drought predictions and other decision-making purposes [33].

Open hydrological data represent data describing the statics and dynamics of ground and surface waters (water sources locations, flood zones, historical records of flooding, real-time water levels, water quality, etc.). The hydrological datasets can have different completeness and accuracy. There are maps of rivers, catchment data, hydrological models to predict water flows depending on the weather, topography, etc. [33].

Estonian monitoring water quality data of water bodies are available on the Environmental Agency portal along with the links to the text of the WFD and water management plans on the national environmental portal [110]. The meteorological and hydrological OGD can also be found on the Estonian national meteorological service - Estonian Weather Service (EWS) [34]. The content EWS is managed by the Estonian Environment Agency and Ministry of the Environment includes various meteorological observation data and forecasts [28]. The observations include satellite images, radar data, temperature, wind, sea levels, precipitation, pressure, etc. It also makes the historical hydrological data publicly available such as water levels, water temperature, flow rates, etc.

The provision of the access to national policies, legislation and plans and monitoring data, free sharing and exporting capabilities constitute an essential part of good OGD practices. The coverage of the communicating objective of OGD can be enabled with the improvement of the usability and structure of the environmental portals, organisation of active dissemination around one of the main environmental portals [110].

However, in order to unlock the sustainable, social and economic opportunities of the creation of national and cross-border innovative OGD-driven services, the proper data quality, unified format and metadata in line with the common standards such as INSPIRE should be ensured.

2.3 Environmental Regulatory Technologies and Environmental Regulations

Regulatory Technologies (RegTech) imply the automated management of regulatory processes such as monitoring, reporting, and compliance using big data analytics, artificial intelligence, cloud computing, etc. [25]. RegTech is common in the financial sphere to monitor transactions in real-time and identify issues or irregularities in the digital payments, resolve various data breaches, cyber hacks, money laundering, and other fraudulent activities [25].

Real-time tracking of the compliance status and quick response to changes can also feature the estimation of compliance with environmental regulations resulting in Environmental RegTech (EnvRegTech) solutions involving big data analysis for minimising risk and supporting decision-making.

In 2015, the United General Assembly set the Sustainable Development Goals (SDG) agenda intended to be achieved by 2030. The goals are interconnected and address global challenges such as poverty, inequality, climate change, environmental degradation, peace and justice, in order to achieve a better and more sustainable future for all [4].

Water management is one of the issues addressed by the SDG. The deterioration of waters has a major environmental and economic impact. The agricultural sector is exceptionally vulnerable to extreme weather events and other environmental changes. According to statistics published by the International Water Management Institute (IWMI), agriculture suffers from 25% of all the economic damage caused by climate-related disasters, and 84% of the damage resulting from drought [35].

There are numerous strategies undertaken to address SWM challenges. There are established government-protected aquatic areas and created regulations in order to protect vulnerable habitats, conserve biodiversity, reduce overfishing, marine pollution, and ocean acidification. Various technologies such as IoT sensors embedded into water management infrastructures can monitor local weather forecasts and control drainage to minimise flooding, stormwater runoff or property damage [4], [8]-[9] enabling farmers, resource managers, and policymakers to accelerate the adoption of improved practices [35]-[36].

Environmental regulations and their implementation can be the objects of interest of various stakeholders, especially polluting industries. There are multiple examples of the environmental consequences of unregulated industries all over the world [39]. At the same time, there are many regulations including the penalties for noncompliance with them such as the EU Environmental Liability Directive (Directive, 2004/35/CE) [40], the US Comprehensive Environmental Response, Compensation and Liability Act (CERCLA, 1980) [37], etc. Unfortunately, the enforcement of the regulations is uneven and complicated by different reasons (e.g. illegal activities) [39].

Early European water legislation dates back to 1970-1980s with the standards for rivers and lakes providing drinking, fish, shellfish, bathing waters, and groundwater and quality targets for them. In the 1990s, wastewater treatment and water pollution were addressed. The early water policies needed to be re-thought to create a more global approach - a single piece of framework legislation to resolve the fragmentation of the water policy previously tackling mostly individual issues. In response to this, the European Commission presented a Proposal for a WFD with the key aims such as expanding the scope of water protection, achieving a good status of waters by the deadline, ensuring water management based on river basins, streamlining legislation, and others [38]. The aims became the key operational elements of the WFD.

On 23 October 2000, the "Directive 2000/60/EC of the European Parliament and of the Council established a framework for the Community action in the field of water policy" or, in short, the EU WFD was adopted. This EU directive commits the European Union (EU) member states to achieve the good qualitative and quantitative status of all water bodies (rivers, lakes, coastal waters, groundwater). The ecological protection applies to all waters and involves the protection of the aquatic ecology, unique and valuable habitats, drinking and bathing water. The ecological and chemical statuses of surface waters are assessed according to the biological, hydromorphological, physical-chemical quality, chemical quality [13].

Defined by WFD environmental objectives and deadlines were not met in Europe and needed an extension because of various uncertainties in definition of monitoring measurements, assessment procedures and forecasts and shortage of nature conservation and water protection areas, etc. Implementation of the river basin management plans is also complicated by climate change since its effects, such as long droughts and increased flooding, requires adaptation strategies [14]-[15].

Nevertheless, the implementation of the WFD strategies ensuring sustainable water protection and outlining the water resource management models is important to agriculture, energy generation, transport policy (shipping), etc. It allows harmonising SWM with ecological objectives. Furthermore, in order to achieve ecological objectives in water protection, there is a need in the participation of all the stakeholders: political actors, water and agricultural users and authorities, municipalities, parties responsible for maintenance, and volunteers [14].

2.4 Sustainable Water Management in Estonia

2.4.1 Water Management and Fishing in Estonia

The water management in Estonia is attributable to the climate and territorial location of the country. Estonia is a small country (45277 square kilometres) located in Northern Europe [41].

Its topography is defined by its location in the north-western part of the East-European Plain: the country is a flat territory where uplands and plateaus vary with lowlands and valleys and coastal cliffs [42].

Since the country is located on the eastern coast of the Baltic Sea, it is in the transition zone from maritime to continental climate: coastal and inland areas have differences in climate. North-Atlantic stream of the Atlantic Ocean influences the climate, and the weather is characterised by strong winds, high precipitation and rapid changes of temperature [43].

Due to the features of the relief and climate, there are multiple inland water bodies. Annual precipitation exceeds evapotranspiration, and the excess water runs off via rivers. The territory is divided into 4 basins: the drainage basins of Lake Peipsi, Gulf of Finland, Gulf of Riga and the islands of West Estonia [44].

The rivers in Estonia are mostly short with a relatively small runoff. In summers rivers can dry out, in spring the river areas can be flooded. Narva River has the largest runoff, and Pärnu River is the longest river. In Tallinn, the water of Pirita River is discharged via a canal into Lake Ülemiste, which supplies Tallinn with municipal water [44].

The climate change in Estonia is observed through an increase in the average winter air temperature and precipitation leading to a decrease in the maximum discharge of spring floods and their earlier beginning (from the end of March-beginning of April to January-February). The shift in the spring runoff to an earlier time might result in the longer duration of the summer low-water period and decrease of the total runoff in the vegetation period (April-September). Warmer climate leads to milder winters with a decrease in the duration of snow and ice cover. In autumn, there is expected an increase in autumn precipitation. Increase in the annual runoff, decrease in the contribution to the runoff in spring and its increase in winter may have different impacts on the water management [45].

Hydrological surveys in Estonia were started already in 1867 over the Suur-Emajõgi river in Tartu, and by 1922 there was established systematic automated monitoring of nearly 50 rivers. Currently, Estonian hydrological observation network consists of more than 50 gauging stations [46].

The gauging stations measure various hydrological indicators such as water temperature, water level and discharge - the volume of water passing through the cross-section of the watercourse per time unit. The data obtained at gauging stations data helps to produce hydrological forecasts and review the state of water resources, and define potential uses [46].

The status of Estonian rivers mostly depends on the efficiency of wastewater treatment, the intensity of agriculture and protection measures used in agriculture. It is monitored by bodies of surface water; there are currently 750 bodies of surface water (644 for water courses, 90 for lakes and 16 in coastal waters [96]) whose status has to be determined. One body of surface water has a similar natural type, living environment, and human impact [21].

In the last decade, due to economic recession, changes in organising industry, and less pressure on the water environment by the water for human consumption, the status of Estonian surface waters has improved. From the perspective of the chemical status, the overall status is bad in four watercourses, and the measures to improve their status exist in a written format in river basin management plans on river basin districts. Overall, the water quality of Estonian rivers and lakes is considered to be satisfactory [21].

Fishing in Estonia is regulated. Estonian fisheries ensure that fish populations are diverse and in good condition, and monitor negative impacts of fishing on the ecosystem meaning that fish stocks populations have a natural age distribution and can reproduce naturally under current conditions. There are specific restrictions to protect more vulnerable populations, protected areas such as spawning grounds in order not to disturb spawning fish, preserved fish migration routes, catch limitations, fishing legislation, systematic recovery of fish stocks, etc. Since many species of fish (salmon, eel, sea trout) are threatened because of over-exploitation, lack of suitable reproduction areas and habitats and their natural low reproductive capacity, there are stocking areas for farming fish restoring and reinforcing their populations where fish are being reared. There were also created common fisheries online databases, and the spawning grounds are systematically mapped and assessed [22].

There are several offices managing fishing activities in Estonia that can be potentially interested in the service to be developed: Ministry of Rural Affairs, Veterinary and Food Board, Environmental Board, Environmental Inspectorate, Ministry of the Environment. These offices monitor fishing activities, issue permits for fishing, etc. In particular, the Ministry of the Environment prepares and implements policies on protection and use of fishery resources including reproduction of fish stocks and protection and restoration of spawning grounds and habitats. The ministry also provides permits for scientific research and special purpose fishing.

The Ministry of Rural Affairs manages the aquaculture sector and is responsible for policymaking regarding commercial fishing [22].

2.4.2 Water Framework Directive Implementation in Estonia

Defined by WFD environmental objectives and deadlines were met neither in Europe nor in Estonia [96]. The implementation of WFD framework is based on the water management plans [96] aiming at the implementation of the water and river basin management principles and measures [97]-[99].

According to the Water Act [100], the status of surface water bodies (very good, good, moderate, poor and very poor) is assessed on the basis of the ecological status (very good, good, moderate, poor and very poor) and chemical status of the body of surface water, whichever is worse. The ecological status of a body of surface water indicates the quality of the structure and functioning of aquatic ecosystems and the chemical status indicates the content of priority substances, priority hazardous substances and other pollutants in surface water and aquatic life.

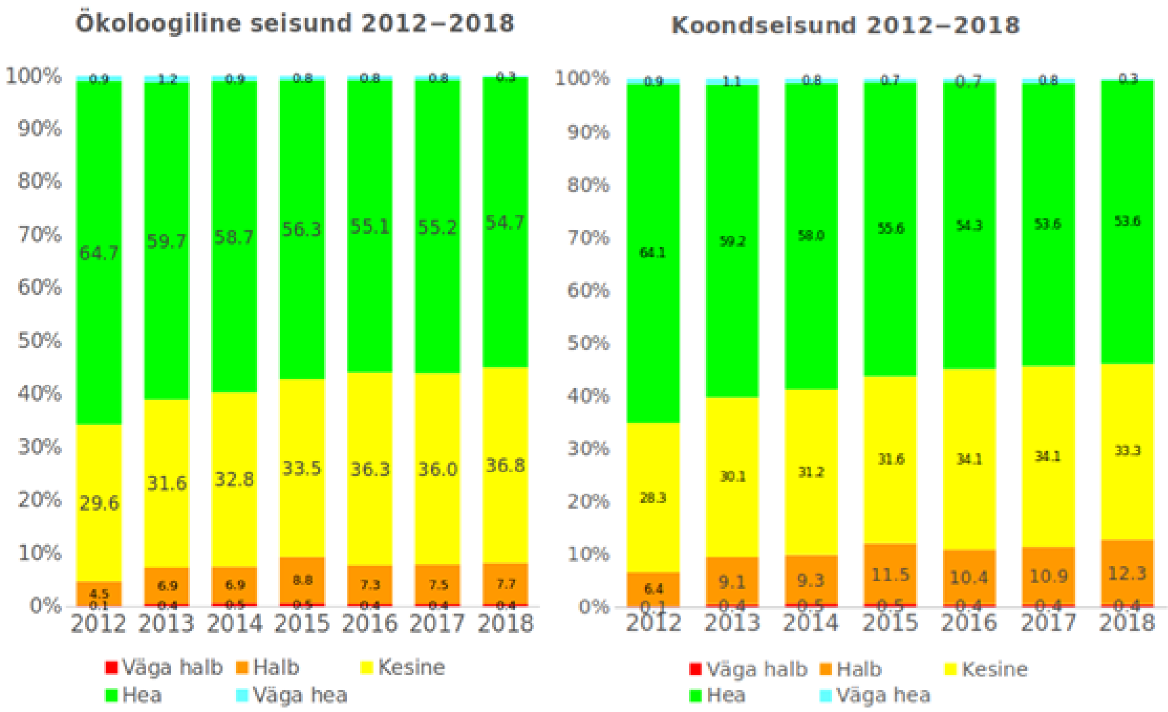


Figure 3. Statuses of Estonian surface water bodies in 2012-2018: ecological (left) and combined (right). Source: [101]

In order to achieve the objectives of WFD, all (100%) the water bodies should have achieved at least good (combined) status. By the initial WFD deadline (2015), only 57.6% of Estonian surface water bodies were at least in a good status [96], and this percentage is gradually decreasing: in 2018, it was 53.9% [101] which was also explained by initial over-optimistic assumptions of the experts not having sufficient amount of data in the beginning of assessment and changes in the assessment procedures [96]. The evolution of the ecological and combined (ecological and chemical) statuses is shown in Figure 3.

The evolution of the number of WFD-compliant surface waters from 2010 to 2017 is presented in Figure 4.

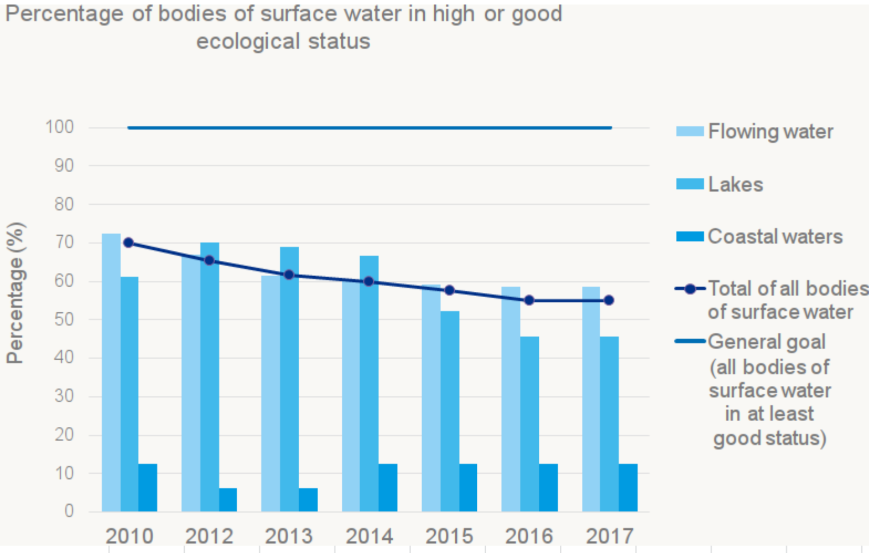


Figure 4. Percentages of WFD-compliant Estonian surface water bodies in 2010-2017. Source: [96]

The map with the combined statuses of Estonian rivers in 2018 is shown in Figure 5.

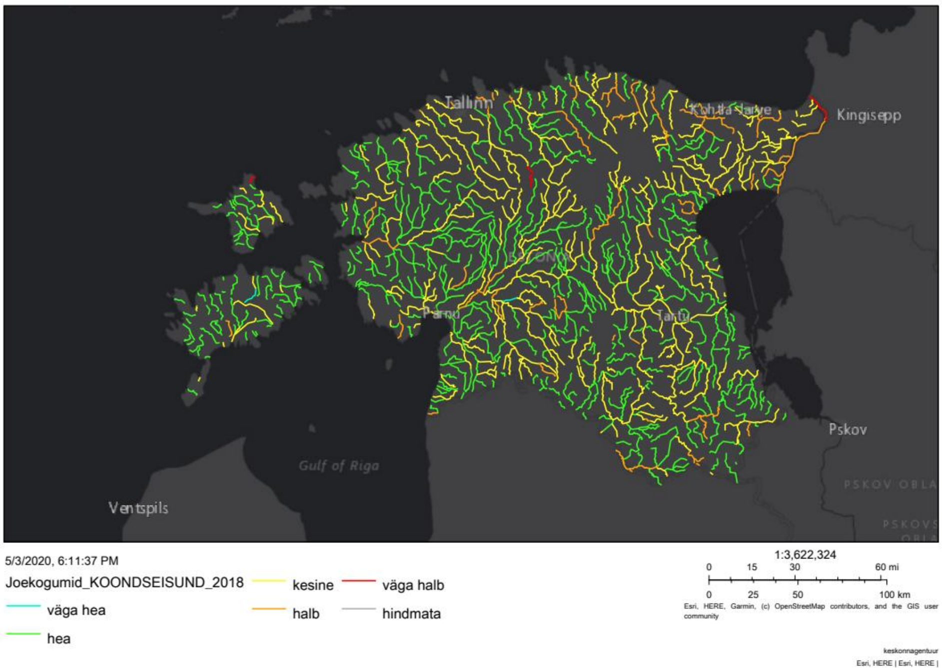


Figure 5. The map with the combined statuses of Estonian rivers in 2018. Source: [103]

On the map, the rivers having neither green nor blue indicators can be considered not WFD-compliant.

2.5 Environmental Flows

2.5.1 Environmental Flows Overview

Rivers are highly dynamic systems, and their natural dynamics differ from river to river. It is defined by the regime of natural flows - a historical pattern of hydrologic variation for a river, and monitoring of the flows is important to maintain healthy river ecosystems [11]. The flow can be described with its hydrograph: magnitude, frequency, duration, timing, predictability, and rate-of-change of streamflow conditions. Other hydrologic metrics are based on these 5 components [47].

The natural hydrologic variation can be caused by climate (seasonal timing, amount of precipitation, and whether it occurs as rainfall or snow), draining and runoff varying across different landscapes depending on a geographic location. The variation can also be caused by human activities and environmental changes affecting biophysical processes responsible for ecological self-sustainability over time [11]. The structure and function of river ecosystems are strongly influenced by the natural flow regimes since they make different aquatic species develop adaptations regulating key ecological processes in a river ecosystem [11], [48]-[49].

The change of the natural flow regime of a river can modify the suitability of physical habitat for species as well as the strength of biological interactions [50]-[51]. Rivers that have suffered human interventions have significantly modified natural flow regimes and thus have experienced ecological degradation due to significant alteration of ecological processes. Information gathered during monitoring of natural flows can help in the river restoration processes.

Various studies provide examples of how streams and rivers can be classified based on an examination of the similarities and differences among flow regimes. Thus, rivers that have similar flow regimes are likely to have similar ecological and evolutionary constraints, and thus can be managed in similar ways [52]-[53].

Water flows supporting various ecological processes can be simulated using eflows [54]. There are different types of eflows (Figure 6, [55]), and two major types of eflows are low flows and high flows [17].

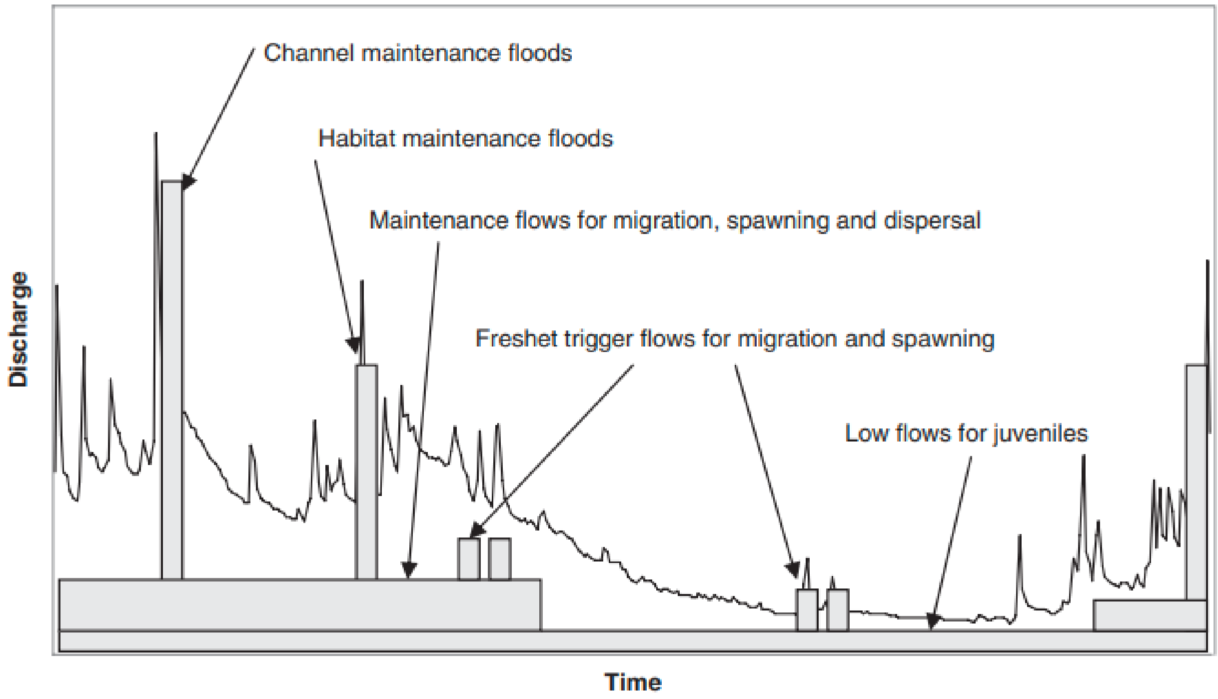


Figure 6. Eflow types. Source: [55]

Low flows create pools of slow-moving water, protecting young fish for feeding and growing. [56]. *High flows* (or maintenance flows) provide access to fish to new habitats and refresh water quality conditions. “Extreme phenomena, such as floods and droughts, play the role of environmental features regulators and change the structure of the population by selective elimination of species” [17, p.2].

There are various case studies of implementation of eflows [57]. However, eflow requirements have still not been adequately assessed for most aquatic ecosystems, and have been implemented in even fewer. Implementation of eflows is complicated by insufficient resources, lack of commitment and support of governments and stakeholders, conflicts of interest, lack of knowledge, training and institutional capacity to manage water resources [58].

2.5.2 Environmental Flows in Environmental Regulations

According to the Brisbane Declaration and Global Action Agenda on Environmental Flows [59, p.1], “Environmental flows describe the quantity, timing, and quality of freshwater flows and levels necessary to sustain aquatic ecosystems which, in turn, support human cultures, economies, sustainable livelihoods, and well-being.” The eflows approach is based on the natural flow regime [11] and was further developed by “research and practice focused on aquatic ecosystem protection, restoration and management” [59, p.2].

The eflows definition goes in line with the UN Sustainable Development Agenda 2030 [60] and its SDG and targets that promote the wise use of water, other natural resources and global life support systems. The SDG 6 implies ensuring access to water and sanitation “which targets to improve water quality by reducing pollution (6.3) and protect and restore water-related ecosystems including rivers, wetlands, aquifers, and lakes (6.6, 15.1). Environmental water requirements are explicitly referenced and defined in SDG indicators 6.4.2 (Level of water stress) and 6.6.1 (Change in the extent of water-related ecosystems over time)” [59, p.5].

Facilitating improvements in the production of freshwater and estuarine foods such as fisheries (14.2), eflows indirectly contribute to other SDG: “SDG 1 (no poverty), SDG 2 (zero hunger), SDG 3 (good health and well-being), SDG 8 (decent work and economic growth), SDG 12 (sustainable management and efficient use of natural resources), and SDG 16 (peaceful and inclusive societies for sustainable development, and access to justice for all)” [59, p.5].

The WFD common implementation strategy involves the definition and assessment of eflows and their adoption in river basin management plans. The eflows were defined as the “amount of water required for the aquatic ecosystem to continue to thrive and provide the services we rely upon” [16, p.2].

Accordingly, aquatic ecosystems demand specific eflow requirements being estimated based on water quantity and dynamics - the sustainability of aquatic ecosystems is defined by the hydrological regime. Therefore eflows can also be described as “a hydrological regime consistent with the achievement of the environmental objectives of the WFD” [16, p.3].

The environmental objectives of the WFD refer to [16, p.3]:

- non-deterioration of the existing status;
- achievement of good ecological status in natural surface water body;
- compliance with standards and objectives for protected areas, including the ones designated for the protection of habitats and species where the maintenance or improvement of the status of water is an important factor for their protection.

To sum up, SWM should include compliance with WFD that requires quantitative water management based on the measurement of eflows from low flows to flood regime.

2.5.3 Environmental Flows Assessment Methodologies

The complexity of flow regimes and conflicting needs of different water users complicates the process of choosing an appropriate methodology for determining flow requirements. Addressing the ecological needs of water of aquatic organisms, the methodology is supposed to model necessary characteristics of a river system closely and be a trade-off between simplicity and expense of the use of the method [17].

The eflow rate values are usually defined conventionally for each river: these are numeric values to be maintained to guarantee conditions suitable for fish for migration, spawning, etc. processes depending on bioperiods [23].

There are more than 200 methodologies to assess eflows. Different countries use different methodologies. There are hydrological, hydraulic, habitat simulation, holistic methodologies, and their combinations (Figure 7). The methods differ in complexity and resource demands and expenses [16], [18]-[19], [61]-[63].

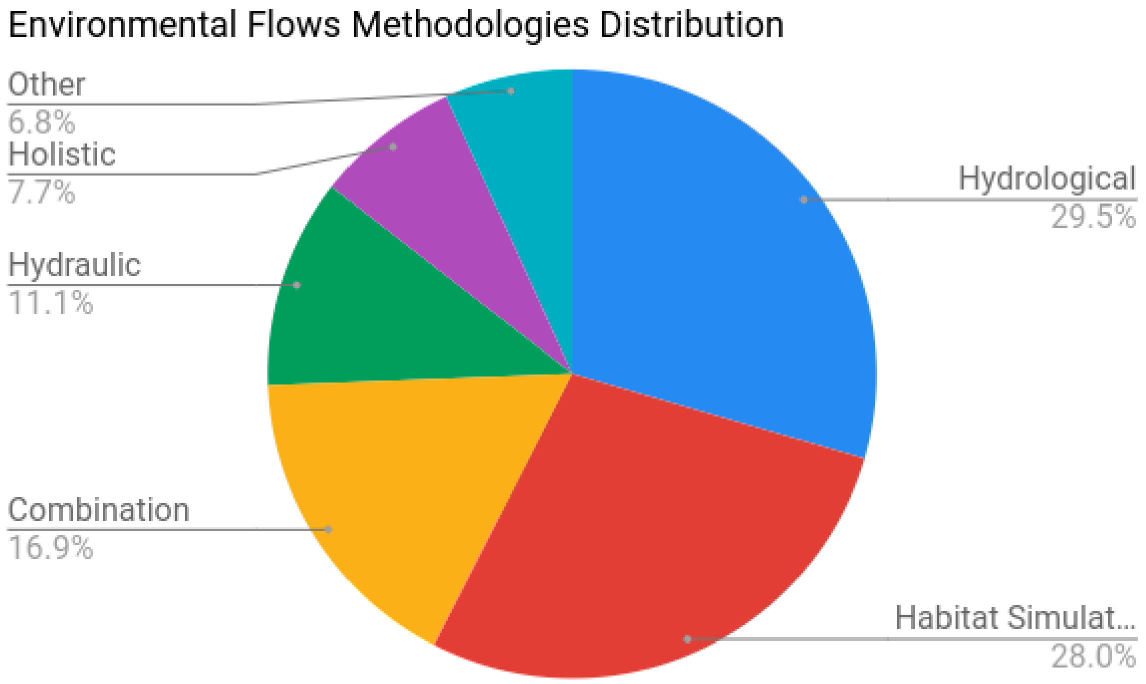


Figure 7. Eflow assessment methodologies. Author: [18]

One of the widely accepted and inexpensive techniques is to use a percent of Average Annual Flow (AAF) to define a minimum flow which is relatively simple to apply since the rules can be made with very limited data collection, but they do not take into account for interannual system variation. [20].

Eflow calculators estimate Environmental Management Classes (EMC) of a river system - specific conditions defining the state of an ecosystem (from modified to critically modified) [61]. Global Environmental Flow Calculator [64] implements the eflow estimation technique using "monthly time step series reflecting natural unregulated flow conditions and its corresponding Flow Duration Curve (FDC) - a cumulative distribution function of flows" [64] (relation between flow rates and the probability of their occurrence). Eflows aim to maintain an ecosystem, or upgrade its EMC. Each EMC is represented by its unique FDC that can be converted into an eflow time series. "The higher the EMC, the more water is needed for ecosystem maintenance and more flow variability needs to be preserved" [62, p.17].

Hydrological methodologies are among the most widely used methodologies due to their ease of use and low cost. They do not operate at a species- or community-specific level and assume that "the full range of natural variability in the hydrological regime is necessary to conserve river ecosystems" [16, p.52]. They are based on the analysis of historic (existing or simulated) streamflow data which requires hydrological and some ecological studies. The methodologies "do not directly include any ecological and morphological characteristics and processes of rivers" [16, p.53] and are recommended to be used at the planning level or in low risk, low controversy situations [18]. They can also be used for habitat modelling, holistic or combined methodologies.

The calculations of the hydrological methodologies involving standard-setting formulas are suitable for real-time automated assessment and cheaper to implement compared to other methodologies, but their simplicity often leads to poor results during implementation. Currently, these methods are considered to be outdated, and various studies seek to address their gaps [16]-[20].

Hydraulic–habitat methods are based on "determining when and for how long habitats are available to aquatic and riparian communities" [16, p.54]. These methodologies involve both physical (hydraulic) modelling of the river channel and modelling of the biological associations with the physical environment (different habitat parameters, such as water depth, flow velocity, substrate composition, channel geometry, cover availability, water temperature) and address the sustainability of communities and ecosystems within the whole river corridor. The hydraulic-habitat simulation methods "estimate only the amount of habitat as a function of hydromorphological conditions, not accounting for more complex ecological and biological factors (e.g., food availability, interspecific interactions and presence of alien species)" [16, p.55]. Complex physical habitat simulation systems such as PHABSIM are suitable to be used in negotiations, in situations of debate about the use of the water [20].

Holistic methodologies involve both human expertise and ecosystem flow requirements: eflow standards are "developed in a workshop setting where river-specific data is considered by a multidisciplinary team of experts (typical areas including hydrology, geomorphology, water quality and various disciplines of ecology)" [16, p. 57]. Deep evaluation, massive data collection and extensive expert consultation can result in the time-consuming and expensive application of the holistic framework [16].

2.5.4 Environmental Flows Assessment in Estonia

The most commonly used eflows assessment method in Estonia dates back to 1972 [23]-[24], and it is still enforced in the legislation. According to the regulation of the Ministry of the Environment specifying requirements for the expansion of a water body, environmental monitoring related to the expansion, protection of aquatic life, dam, elimination of the expansion and lowering of the water level, and methodology for determining the minimum ecological flow [102], the minimum eflow is determined for the ice-free (low water) period from May to October by calculating the average monthly minimum flow with a 95% probability of being exceeded. The average monthly flow is determined for each reference year from hydrographs, it uses 30-day minimum flows, calculates their 30-day average, and then the average of all the 30-day averages is applied [23]-[24].

This method results in the very low flow rates (water scarcity) which rarely occur in a river having a very negative effect. However, with the environmental changes, the natural state and flow regime of rivers changes over years, spring high water flows decreased, and winter minima increased then it may be considered that such a methodology for calculating the ecological minimum flow rate may not be appropriate to be suitable in the present circumstances and the result obtained does not guarantee a healthy ecosystem of the river sustainability [23]-[24].

Some of the hydrological standard-setting formula methods were also recommended to be used for eflows assessment in Estonia [23]-[24]:

1) Tennant Hydrological Method, 30% of the long-term average flow rate.
2) Tennant Hydrological Method, 20% of the long-term average flow rate.

3) 75% exceedance probability of a monthly average of the minimum flow rates found separately for winter (October-April) and the summer season (May-September).

4) 95% (or 90%) exceedance probability of a monthly average of the minimum flow rates found separately for winter (October-April) and the summer season (May-September).

Tennant method. The Tennant method is an internationally applied method developed in the 1970s. The method uses a percentage of AAF to determine fish habitat quality in two different segments of the year, October to March and April to September. The percentages define minimum for short-term fish survival, sustaining fair survival conditions, outstanding habitat, etc. [20] (Figure 8, [20]).

Narrative Description of flows*	Recommended base flow regimens Oct.-Mar.	Apr.-Sept.
Flushing or maximum	200% of the average flow	
Optimum range	60%-100% of the average flow	
Outstanding	40%	60%
Excellent	30%	50%
Good	20%	40%
Fair or degrading	10%	30%
Poor or minimum	10%	10%
Severe degradation	10% of average flow to zero flow	

*Most appropriate description of the general condition of the stream flow for all parameters listed in the title of this paper.

Figure 8. Tennant method eflow percentages. Source: [20]

Tennant method implies the following calculation of eflow rate [23]:

$$Q_{env} = \frac{\sum_{i=1}^{n} Q_i \cdot X\%}{n \cdot 100\%}$$

where

Q_{env} - eflow rate [meters cubed per second],

Q_i - discharge [meters cubed per second],

n - total number of discharge measurements Q_i,

X - percentage (for example, for the Tennant 30% method *X=30*).

The percentages defined by Tennant allow water managers to set flow rate environmental regulations using a certain percent of the AAF without further onsite data collection. However, the applications of the method (without further modifications) have shown its inability to protect habitat adequately [20].

It is expected that 30% of the long-term average flow will provide the most better habitats for aquatic benthos, aquatic plants and fish; and secure water supply in the river. Percentage of the selected environmental flow rate (the percentages essentially mean the hydrograph cutting at a given flow rate) also allows you to set the number and duration of days, how many days the flow rate was less than the predetermined value (Figure 9 [24]) [23]-[24]. Periods marked with + show that during that time the given flow exceeds the given flow rate. Periods marked with "-" how water shortages [24].

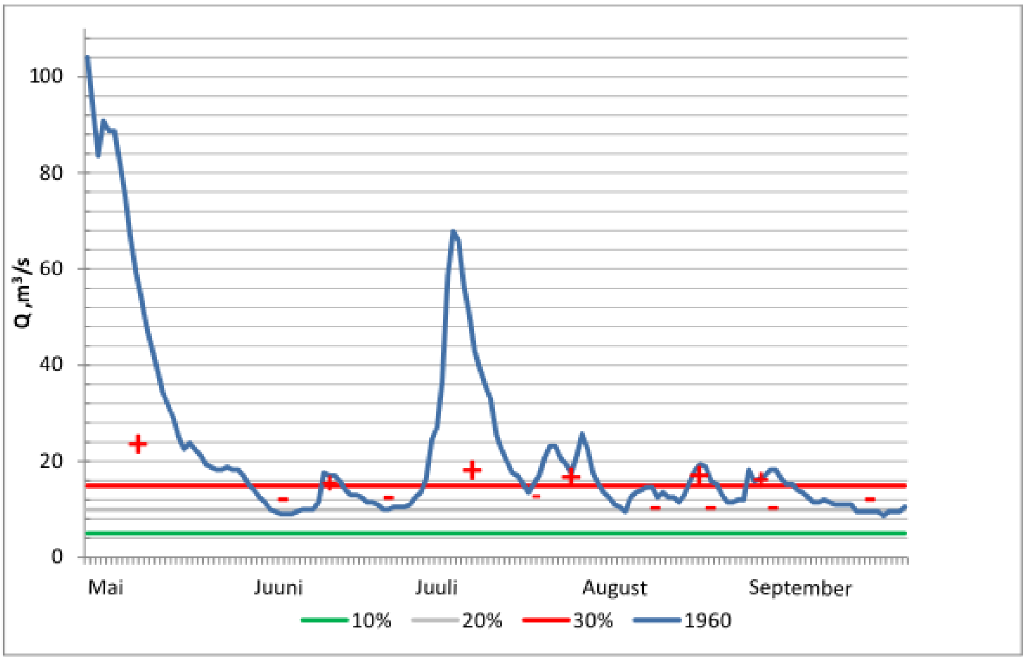


Figure 9. Application of Tennant method to Oore gauging station. Source: [24]

p% exceedance probability of a monthly average of the minimum flow rates. In order to find the p% exceedance probability of a monthly average of the minimum flow rates, one needs to select summer and winter average monthly flows are selected, and then the monthly flow rates shall be ranked in descending order. The winter and summer flows should be assessed with separate probability curves, and p% probability flow rate should be taken as an eflow rate [23].

$$p = \frac{m}{n+1} \cdot 100\,\%$$

where

p is the probability of an observation value occurring;

m is the sequence number of an observation sorted in the descending order (starting from 1);

n is the total number of observations.

2.5.5 Regionally Applicable Environmental Low Flow Formula Approach

One of the limitations of hydrological standard-setting formulas is that they do not consider the biological response of fish to environmental change.

Different species of fish have a wide range of needs, and it is hard to determine eflows at the regional scale based on only hydrological patterns. Large scale application of biological rules of biological responses assessment is also complicated, being too detailed and resource-consuming. "From a management perspective, eflow guidelines need to be easy to define and inexpensive to apply" over whole regions. From an environmental resource protection perspective, they should be sufficiently detailed for capturing the biological response to environmental change [17, p.2].

The approach suggested by the Polish researchers allows the use of a simple environmental low flow formula that can be applied regionally to estimate eflows [17] (further RAELFF - Regionally Applicable Environmental Low Flow Formula).

The formula to calculate the eflow threshold values on any cross-section of the catchment area in any water body from the same FET is [17]:

$$Q = p \cdot \bar{q} \cdot A$$

where

p - tabulated value index obtained from the pilot studies specific for the bioperiod and fish ecological river type [unitless],

$\bar{q}$ - site-specific mean low flow for the bioperiod at the cross-section [meters cubed per second per kilometres squared],

A - catchment area at the cross-section [kilometres squared].

Site-specific low flows $\bar{q}$ should be scaled by the catchment area to make the formula applicable to other locations along the water body (considering their different sizes).

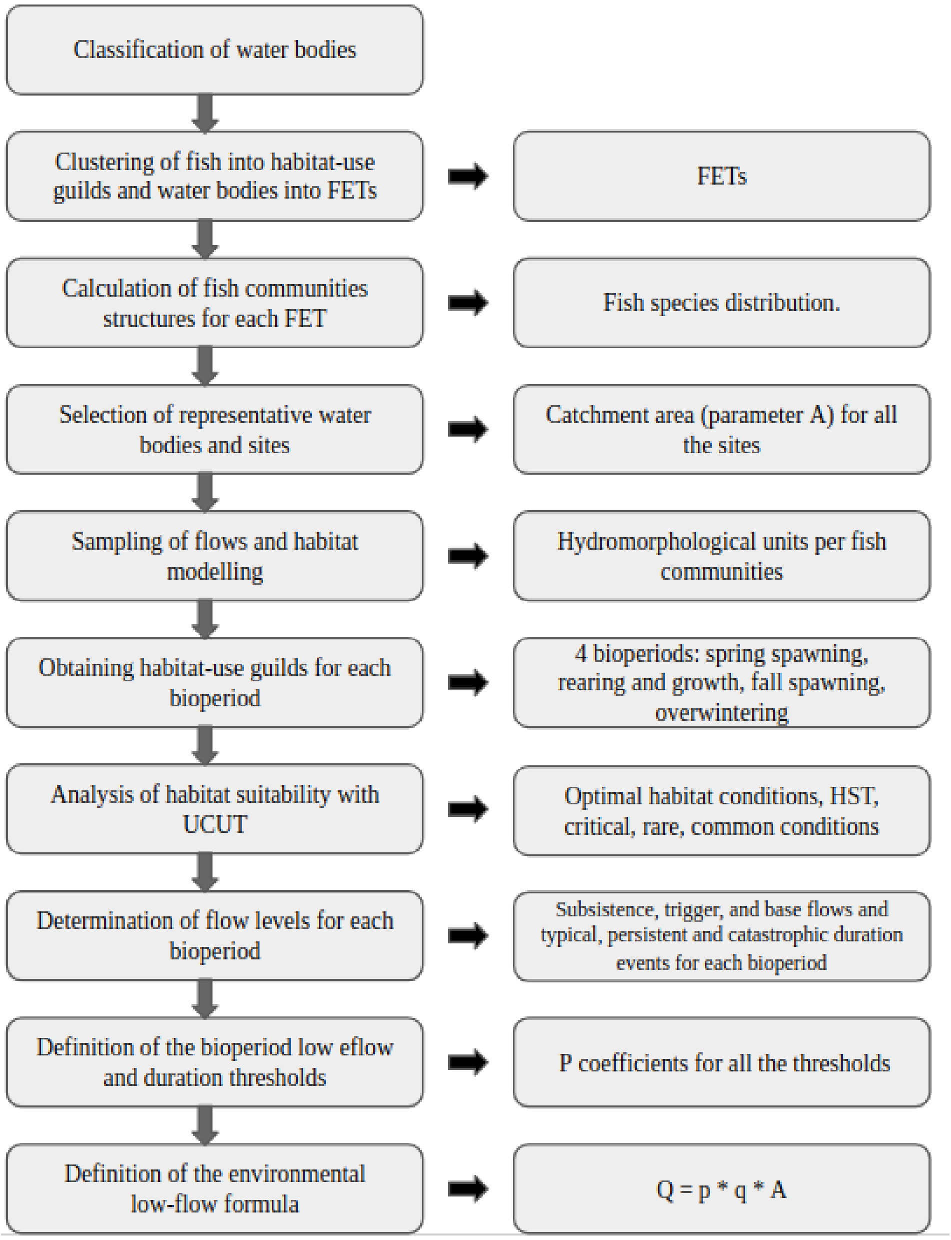


Figure 10. Determination of RAELFF. Author: [17]

The procedures used to determine p coefficients for Estonian rivers are suggested by the author of the approach [17] and outlined in Figure 10. The approach requires additional field research activities to enable distributed data collection to apply the method to any location, determine river types and their species population, etc. and can be carried out with the cooperation of the Polish researchers.

The approach leading to the determination of the RAELFF can be described with the following steps [17]:

1. Classification of water bodies according to the target fish communities using fisheries monitoring data.
2. Clustering of rearing life stages of fish species into habitat-use guilds and water bodies into FETs through the non-hierarchical cluster analysis.
3. Calculation of target fish communities for each FET.
4. Selection of the representative water bodies following multiscale hierarchical framework.
5. The sampling of flows according to the mesohabitat simulation methodology into hydromorphological units, building of the habitat model for fish guilds and community structures.
6. Division of the time scale into bioperiods representing different life stages. Establishment of habitat-use guilds based on the fish community structure for each bioperiod.
7. Determination of the habitat suitability criteria in order to define optimal habitat conditions. Habitat time series analysis with the help of the UCUT methodology to identify HST (e.g. pulse and ramp disturbances that affect fauna densities and species composition).
8. Determination of the flow levels creating rare, critical, and common conditions (subsistence, trigger, and base flows) and types of duration events: typical, persistent (unusual events that are likely to occur every few years but not more than twice in a year), and catastrophic (occur on a decadal-scale) for each bioperiod.
9. Definition of the bioperiod specific low eflow thresholds and duration thresholds to persistent and catastrophic conditions. Definition of the FET specific coefficients p for all the thresholds (Figure 11 [17]).
10. Definition of the environmental low-flow formula.

Compared to the other hydrological methods, the RAELFF approach takes into account not only hydrological variability of site-specific discharges but also river type, site-specific fish species and fish habitats at each location. It allows classifying rivers by their types and applying type-specific coefficients for each bioperiod.

RAELFF can be used as a base for environmental regulations. The compliance is determined by comparing discharge values with eflows rates calculated with the environmental low flow formula. As it has been suggested in the method research paper [17], water management rules can be defined by the duration of crossing the subsistence, trigger/critical and base lines.

The eflows in the approach are distinguished as base, subsistence, and critical eflows. These thresholds correspond to different ecological thresholds: base eflow is the eflow considering natural hydrological variability at a given site, critical eflows may occur every few years and are often flagged for action because they can cause damage to fish populations, and subsistence eflows are eflows below which fish populations will not survive [17], [66]-[67].

Base, critical and subsistence eflow thresholds are not trivial to define. For demonstration, the eflow thresholds can be found based on the ratios of exceedance probabilities at the crossing x-axis (exceedance probability) point of FDC and discharge values filtered by low pulses (long-term means) corresponding to them. For example, the subsistence discharge value corresponds to 25%, critical - 50%, base - 75% of the exceedance probability found at the crossing point.

In the result of the field research activities, it should be possible to classify rivers into different Fish Ecological Types (FETs), determine bioperiods and *p* coefficients for different eflow thresholds (Figure 11, [17]).

FET	Threshold	Spring Spawning	Rearing and Growth	Fall Spawning/ Overwintering	Overwintering
		III–VI	VII–IX (X)	X (XI)–XII	I–II
1	Base	0.65	0.87	0.83	1.77
	Critical	0.52	0.71	0.68	0.56
	Subsistence	0.46	0.56	0.60	0.52
2	Base	2.57	4.41	0.90	2.20
	Critical	1.15	0.74	0.21	0.80
	Subsistence	0.82	0.61	0.16	0.60
3 *	Base	4.08	3.83	4.56	1.62
	Critical	1.28	1.17	0.73	0.62
	Subsistence	1.04	0.85	0.55	0.37
4 *	Base	2.76	2.98	2.63	4.43
	Critical	1.03	0.93	0.75	0.74
	Subsistence	0.90	0.69	0.56	0.55
4s	Base	1.54	1.44	1.39	1.08
	Critical	1.11	0.95	0.85	0.89
	Subsistence	1.05	0.91	0.82	0.86

Figure 11. FET specific coefficients p for the eflow thresholds. Source: [17]

The rivers classified by their geomorphic water body type were grouped into Fish Ecological Types (FET), distinguished by specific fish communities. The needs of the fish communities can be outlined with bioperiods: rearing and growth (July–September/October), fall spawning (October–November), overwintering (November/December–February), and spring spawning (March–June). Summing up the surface area of suitable and optimal habitats of these fish communities and weighting the sum by the proportion of guilds in the expected community allowed to calculate habitat rating curves for guilds – sufficient habitat for communities occurring in each bioperiod [17], [68].

The results can be interpreted with the graphs such as the one presented in Figure 12 [17], where horizontal lines show specific eflow types.

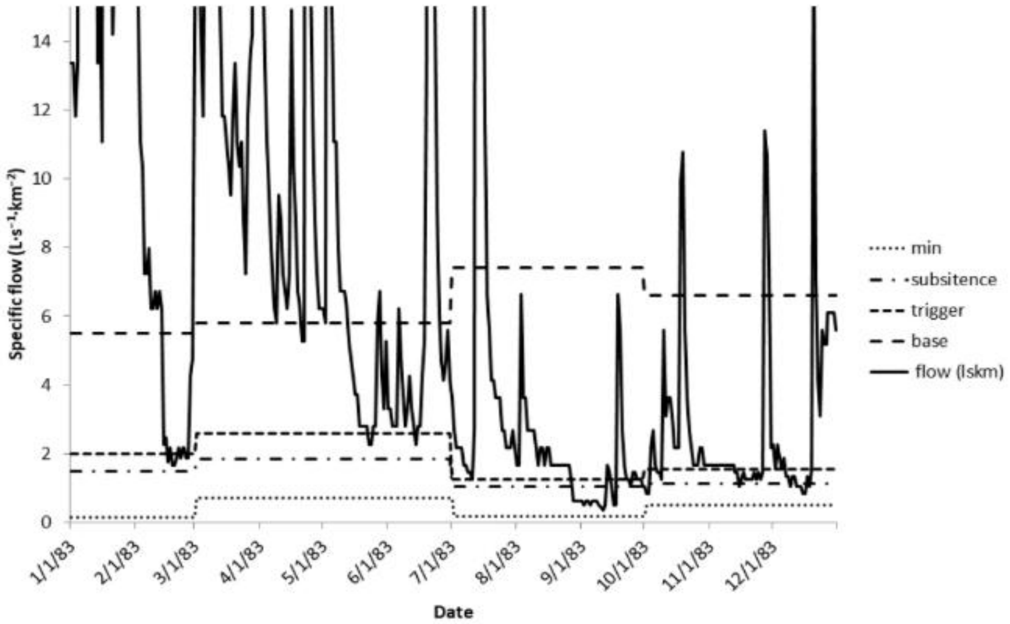


Figure 12. Skawa river flows with the eflow thresholds (1983). Source: [17]

The habitat time series analysis with the help of the Uniform Continuous Under Threshold (UCUT) methodology allowed to identify Habitat Stressor Thresholds (HST): the pulse stressors cause an immediate alteration of aquatic organisms densities and the ramp disturbance causes an alteration of species composition. The UCUT analysis with flow time series allows defining low eflow thresholds for each bioperiods and durations of persistent and catastrophic conditions [17], [63].

The determination of the durations and HST can be performed analysing the cumulative duration of events when habitat is lower than a certain eflow threshold building UCUT curves (Figure 13, [17]), [17], [67], [69].

After that, the possible management actions of the approach depending on the water availability can include passive continuing observation, restriction of water withdrawals, flow augmentation, morphologic modification. The objective of the mentioned actions is to shorten the duration of habitat deficits [17].

The assessment of the eflows should be accompanied by further restoration and mitigation measures addressing various water use. New uses shouldn’t prevent the achievement of a good ecological status. The permitting conditions of water bodies defined by WFD with the deteriorating ecological status due to hydrological alteration should be reviewed and adapted [16].

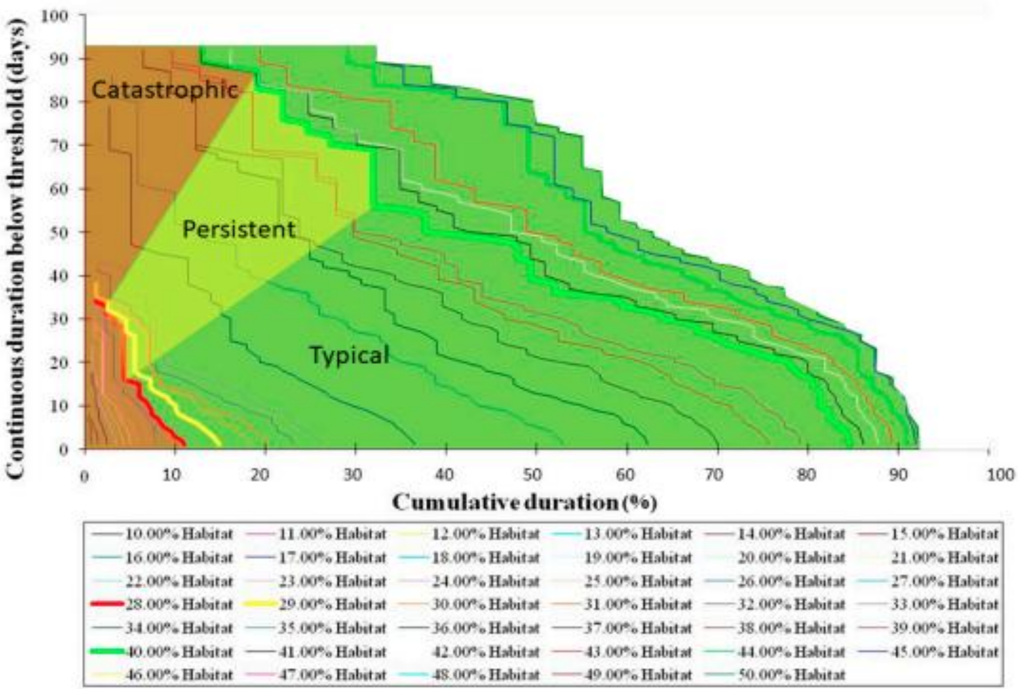


Figure 13. Example of UCUT curves demonstrating the cumulative duration of events when habitat is lower than a threshold for a continuous duration of days. Source: [17]

The ecological conditions can be improved through restoring a more natural flow regime: modifying the timing in which water is taken, additional provision of flow to the river, restoration of flow variability, installation of balancing reservoirs in the river channel, reduction of the flow rate, etc. [16]. "The combination of hydrological measures (ensuring the discharge of an appropriate flow regime by all abstractions and regulation) and morphological measures (improving the aquatic habitats in order to reduce their vulnerability to flow impairments) may be the most cost-effective approach" [16, p. 62] to achieve environmental objectives.

The timely detection of the river's ecological status degradation can be addressed with the automated eflow assessment and analysis. With the automation of the eflows assessment, it is also possible to enable scenario and impact modelling to see whether a water use activity alters the hydrological flow regime and ecological status of the water body taking into account cumulative effects over time.

2.6 Summary and Propositions

The protection and preservation of the environment guided by sustainability and climate resilience goals require regulatory measures, investments and deployment of smart technological solutions in order to improve existing practices. Under frequently occurring water-related disasters affecting millions of people, disaster-resilient governance is needed for better water protection, conservation and management reducing and preventing disaster risks. Evidence-based decision-making facilitates well-grounded decisions, whereas proper analysis and predictive modelling contribute to risk assessment and planning.

The design and creation of EnvRegTech artefacts guided by the EI framework using environmental OGD of proper quality can support decision-making and enable real-time situational awareness about the ecological status of rivers needed to ensure compliance with the regulatory framework of WFD. With the Estonian hydrological OGD and using the best existing practices of eflow estimation such as RAELFF, the assessment of river ecological status can be automated and enhanced with analytics and forecasting providing insights complementing measures to address flow alteration.

With access to the hydrological OGD, the EnvRegTech service to be designed can help Estonian hydrologists, fish biologists, water managers and other environmental authorities to improve existing practices of eflow estimation.

Addressing the gaps of hydrological standard-setting methods used and recommended to be used in Estonia, the embedded RAELFF approach can address the biological response of fish to flow alteration, water levels and temperatures can be used for compound event estimation and automation of the assessment along with the time series forecasting allowing to model and evaluate risks and impact of the hydrological changes. All these features facilitate the creation of more profound compliance rules for eflows and allow the authorities to respond quickly to environmental changes and ensure a good status of rivers according to WFD.

Supporting the needs of decision-makers, the design of the EnvRegTech service as software should involve the dynamic implementation of the graph building for compliance time series (Figure 9, 12) and UCUT plots (Figure 13) that can be useful for reporting. Dynamic change of estimation parameters and forecasting can be helpful for scenario and impact modelling. Making the service available on the web (web service) and its further deployment simplifies access to the insights for a larger audience and unleashes the new opportunities for service improvement and reinforcement.

Finally, the design of an EI artefact promotes further research and development activities in EnvRegTech currently being far behind other RegTech solutions. Enabling the Polish RAELFF approach for the Estonian eflows estimation contributes to the method validation and application potentially triggering the subsequent research of the other use cases and presumably leading to the eflows estimation unification.

The steps of the artefact formation are described in the next chapters.

3 Research Methodology

The main research question of the thesis is **how to design a service providing the automated assessment of the ecological status of Estonian rivers for the WFD compliance monitoring and reporting using OGD.**

Addressing the gaps of the existing methods of eflow estimation in Estonia and globally, the following additional research questions are considered:

1. How to make the designed automated system suitable to be used for decision-making?
2. How to improve currently used in Estonia methods of eflow estimation having the Estonian hydrological OGD accessible?

3.1 Design Science Research Methodology

The methodology for the thesis was chosen following its main research question, and it is the Design Science Research Methodology (DSRM). According to the definition [70], DSRM is a research paradigm in which a designer answers questions relevant to human problems via the creation of innovative artefacts - thereby contributing new knowledge to the body of scientific evidence with the designed artefacts that are both useful and fundamental in understanding the problem. DSRM in information systems research involves the creation of an artefact (application) for a specific design problem which requires a prior investigation of the problem [70].

In this thesis project, DSRM is adopted to design and develop a web service as an artefact that can potentially contribute to the scholarly work on the automation of the eflows estimation and lay a foundation for the further distribution of RAELFF as a unified approach that can be used on a large scale by different countries and further research on the modification of the approach for various use cases.

EI framework (Figure 2) can be seen as a particular implementation of DSRM involving the design of an EI artefact. The corresponding steps of both methodologies [27], [70] are summarised in Table 1.

Table 1. DSRM and EI steps alignment

Step	DSRM step	EI step
1	Problem identification and motivation	Identification of stakeholders and decision support goals
2	Definition of the objectives for a solution	
3	Design and development	Data collection and data processing
		Analysis and modelling
4	Demonstration	Assessment and compliance
5	Evaluation and communication	Results communication and application

The thesis paper covers the steps of the research in the following way:

1. **Problem identification and motivation.** The design problem is defined in Chapter 1 as well as the gaps addressed by the solution under development. The theoretical background of the problem is covered in Chapter 2.

2. **Definition of the objectives for a solution.** The objectives of the solution, its stakeholders and decision support goals are described in Chapter 1.

3. **Design and development.** The design of the artefact, data analysis, forecasting modelling, and usage of the artefact are described in Chapter 3.

4. **Demonstration.** The demonstration of the artefact involves its usability testing carried out by the stakeholders and a feedback questionnaire.

5. **Evaluation and communication.** The results of the artefact demonstration are reported in Chapter 5.

3.2 Data Sources and Data Collection

The data used for the artefact implementation are Estonian hydrological OGD. It represents the hydrological measurements made in 54 Estonian gauging stations [71] in the period from 01.01.2009 to 31.12.2018. The data has the following features:

- Minimum/Maximum/Average daily water level [centimetres]
- Minimum/Maximum/Average daily water temperature [°C]
- Minimum/Maximum/Average daily discharge [meters cubed per second]

For each of the rivers, its watershed area has also been obtained. General information of the gauging stations used for analysis and modelling is presented in Appendix 1.

The determination of p coefficients for RAELFF requires additional research of fish species populations. The publicly available online resources of fish species are fisheries databases [72]-[73]. In general, there are around 75 species of fish living in Estonian waters. Most of the fish are freshwater, semidiadromous (ide, vimba bream) or diadromous fish (salmon, sea trout, eel) [73].

Within the scope of the thesis project, the fish coefficients p were taken from the Polish project [17] through superficial comparison of the rivers characteristics and fish species living there assuming that the more precise values should be obtained through further research. Due to Estonia's relatively uniform hydroclimate and homogeneous river morphologies, all the Estonian rivers are likely to have FET of 3. Further configuration of FETs is left to the discretion of users.

The thesis project involves exploration and analysis of the described data and its further use for feature selection, modelling and forecasting. The results were obtained and interpreted using Python scripts with such packages as scipy [74], numpy [75], pandas [76], statsmodels [77], matplotlib [78], sklearn [79], and others (the scripts are presented in Appendices 2-4).

The forecasting involves the training of regression models and aims to predict hydrological values based on the other stations' data (e.g. to handle interruptions of data collection at gauging stations).

3.3 Risks and Limitations

For the current project, there can be distinguished *risks and limitations of the service to be developed* and *risks and limitations of the methods used for eflows calculations*.

The service to be developed is supposed to provide a tool to be used by the environmental specialists (such as the authorities of the Ministry of Environment) and other stakeholders to estimate whether the monitored river flow discharges exceed the calculated by a specific method (including RAELFF) eflow rate. It does not provide rules to be used to trigger water management actions (as well as specific measures to be undertaken) and assumes further research and possible modifications of the suggested method. However, it interprets the results of estimations and can be used to create the rules simulating various scenarios and analysing the obtained results, which is not in the scope of the research.

Since there was no research on the applicability of the Polish method (RAELFF) to the Estonian rivers use case, the fish ecological coefficients are taken from the Polish research and assume to be validated and modified if needed. The bioperiod months are also taken from the Polish study but should also be adjusted by the specialists following climate conditions of the analysed area.

The automation requires the input data to be in the same format and have a proper data quality. If the obtained Excel spreadsheets with the hydrological data are generated not in an automated manner, there is a greater probability of service failures.

The forecasting model developed within the current project may require additional modifications and validation on the other data since the provided datasets have a large amount of missing data that required to put additional restrictions on the regressors' fitting.

With all the benefits of RAELFF described in Chapter 2.5.5, the assessment approach has risks and limitations resulting from the formula determination research activities, its components, etc.

First of all, the formula is comparatively brand new which requires its further validation, verification with additional data and the research on its applicability to other geographical locations (Estonia in particular) despite the overall concept involving extrapolation having been considered feasible [17]. The specific mean low flow approach should also be chosen among a large number of various hydrological formulas as well as the method of definition of base, critical and subsistence eflows.

While determining RAELFF, there was a low number of rivers used for testing the approach. It was caused by the limited availability of hydrological data, abnormal changes in weather conditions and hydrological modifications due to specific activities (unregistered flow augmentation, flow alterations caused by channelisation and melioration upstream). That restricted the choice of water bodies and resulted in uneven sizes of clusters [17].

The formula to calculate river discharge differs from place to place, which may add variability and deviations of the results. That is why the accuracy of the metric estimates should be further tested, and if needed, it could be replaced by another flow metric. There is also the variability of p coefficients within the same FET which requires the uncertainty analysis with a larger number of sites [17].

Another limitation of the approach is that hydromorphologically modified rivers require additional modifications of the rules since the rivers used for testing had low hydromorphological modification. To address this limitation, site-specific studies are required to be carried out [17].

Taking into account these risks and limitations, the service to be designed, implemented and delivered will serve only as a prototype - demonstration of the chosen approach application which is going to mark the beginning of the further research to make it more applicable to a particular use case. It does not guarantee that the solution is supposed to replace existing methods of estimation without any modifications because of the limitations mentioned above.

Nevertheless, the solution can address many problems that face the research community and other practitioners bringing the benefits of the RAELFF approach. The formula is simple, but it still captures a complicated relationship between flows and biological response, allowing using it universally and applying for legal regulations at both regional and global scale. The verified method can be incorporated into the legal framework of eflows regulations for better management and protection of riverine environments [17] including the context of WFD.

4 Service Design and Implementation

4.1 Service Design

4.1.1 Service Architecture and Technology Stack

The potential general architecture of the system involving real-time data collection is presented in Figure 14.

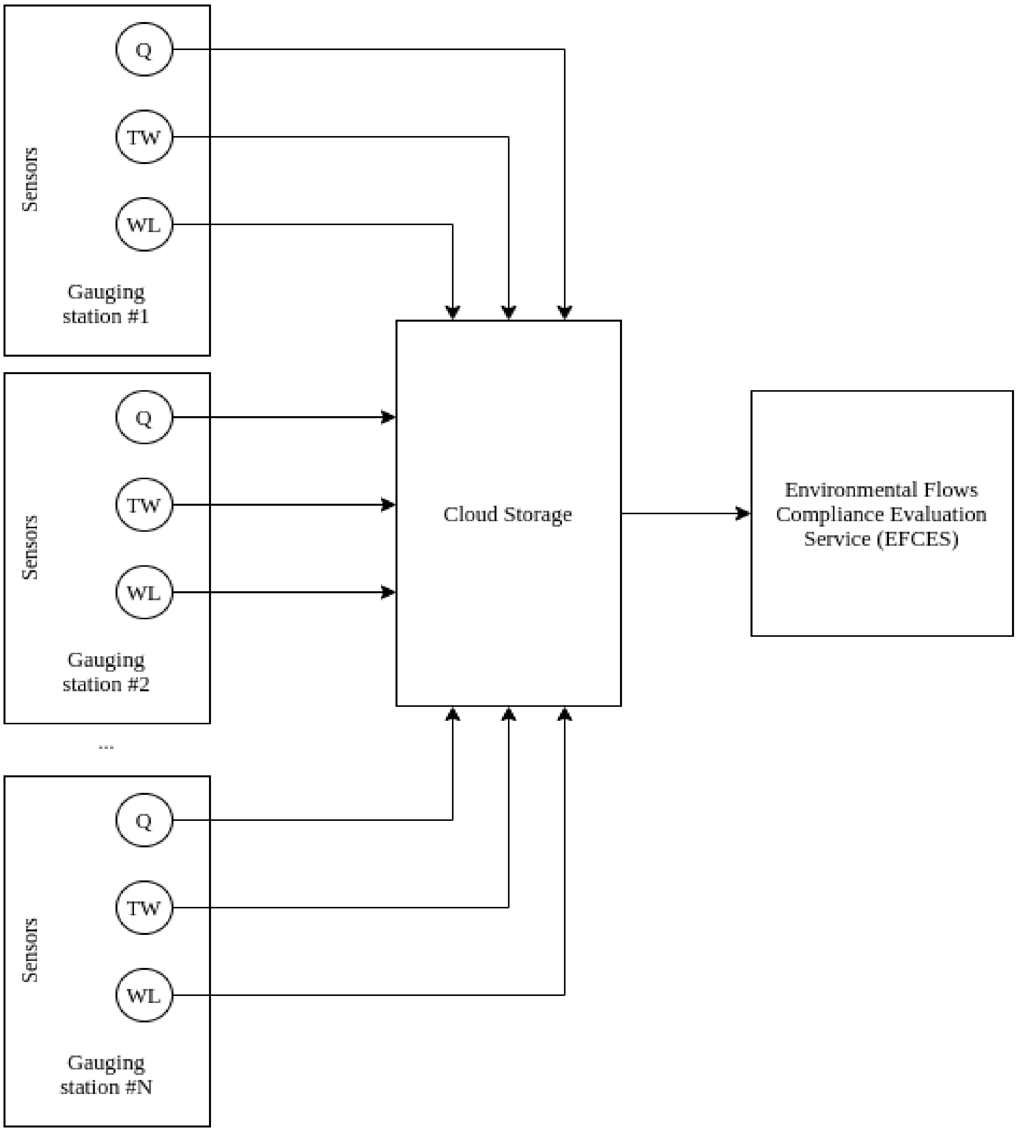


Figure 14. General architecture of the system

In such a system, at each of the gauging stations sensors collect hydrological measurements such as discharge (Q), water temperature (TW), and water level (WL), and pass it to cloud storage. The data can be collected by different parties. Environmental Flows Compliance Evaluation Service (EFCES) is supposed to access the storage and analyse the data to estimate the compliance of the eflows of the evaluated river. The results are supposed to be interpreted in the way suitable for decision-making.

The EFCES system is implemented as a web service; its architecture is presented in Figure 15.

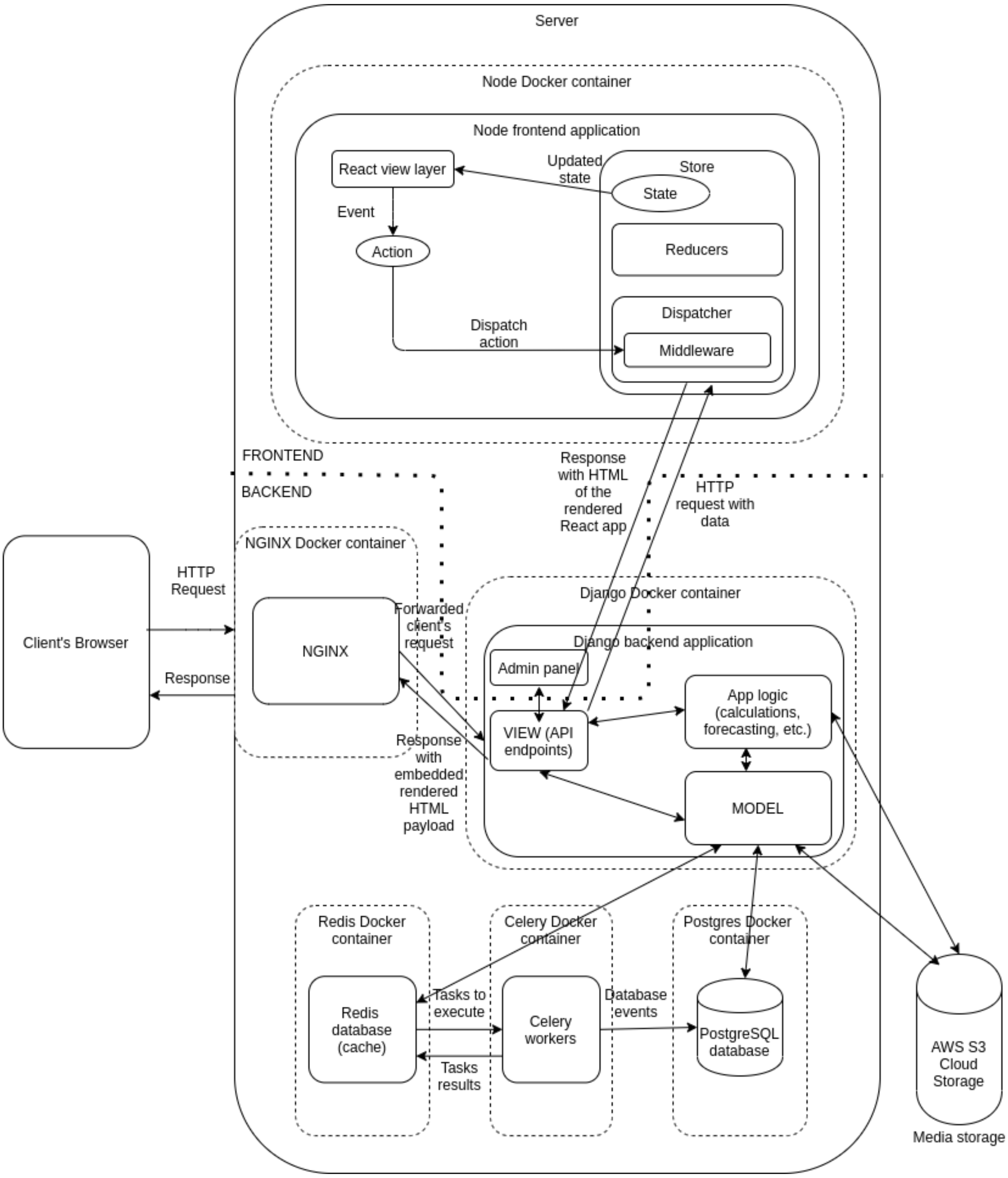


Figure 15. EFCES architecture

Since the direct access to the sensor network is not established, the project is focused on the use of the Estonian hydrological OGD for the automated eflow assessment, analysis and compliance evaluation. The hydrological OGD are provided in the form of Excel spreadsheets.

The station data are stored on the EFCES relational database, whereas spreadsheets with the hydrological data are stored in the media storage on cloud per each gauging station database entry.

The core of the web-service and configuration is generated with the use of the publicly accessible Single Page Application (SPA) project template created and maintained by the Estonian digital product company Thorgate supporting the promotion of open-source software [80]. The service is dockerized and consists of multiple Docker [81] containers of subservices responsible for various functions:

- Django (backend application) [82]
- Node (frontend application) [83]
- PostgreSQL (relational database management) [84]
- Redis (in-memory key-value database - caching) [85]
- Celery (asynchronous tasks execution) [86]

The backend application is implemented with the RESTful Django web-framework following the Model-Template-View (MTV) architectural pattern [87]. The main programming language of the backend application is Python. The frontend application is written with the use of HTML, CSS, JavaScript (React JS library rendering data to DOM). Redux library [88] is used for state management. Server-side rendering of dynamic web pages (running JS scripts) is carried out within Node.js runtime environment. For storage, the PostgreSQL relational database management system is used, Redis broker is used for caching, Celery - for efficient tasks processing.

4.1.2 Service Requirements

In order to make the EFCES suitable to be used for decision-making, the following functional service requirements have been set:

1. The storage and manipulation (CRUD operations) of the hydrological data and parameters used for calculations (hydrological time series, FET information, p

coefficients, bioperiod ranges, etc.) should be enabled and accessible to the administrators of EFCES.

2. The calculation of low flows using the methods currently used in Estonia as well as RAELFF should be implemented and accessible to all the end-users for selection.
3. Compliance estimation should be enabled for discharges (exceedance of eflows), and water temperatures and water levels (exceedance of a threshold) and accessible to all the end-users.
4. Visualisation requirements are time series plots with compliance indication, UCUT graphs estimating exceedance probability over time, tables with compliance/noncompliance proportions per bioperiods of the selected date range.
5. UI requirements are configuration of parameters: data range, catchment area and area factor, eflow calculation method, FET and types of fish coefficients eflow thresholds, secondary time series selection (water temperature or water level), measurement type (minimum, average, maximum values, or ranges of values), forecasting parameters (fitting input feature type).
6. Export of the estimated results (PNG and Excel) should be enabled for the interpreted estimation results.
7. Forecasting of the mean values of hydrological data (time series forecasting) should be added.
8. User management should be enabled.

The main non-functional service requirement is maintainability of the solution: implementation of a modular structure allowing most efficiently modifying and replacing specific modules on-demand. Since the results of the calculations are supposed to be used to make decisions, it is significant to cover the service with tests and ensure common code styling and formatting.

4.1.3 Database Design

The database representing Object Relational Model (ORM) of EFCES is described with the Entity Relationship Diagram (ERD) demonstrated in Figure 16.

The relational data are stored in the PostgreSQL database [84].

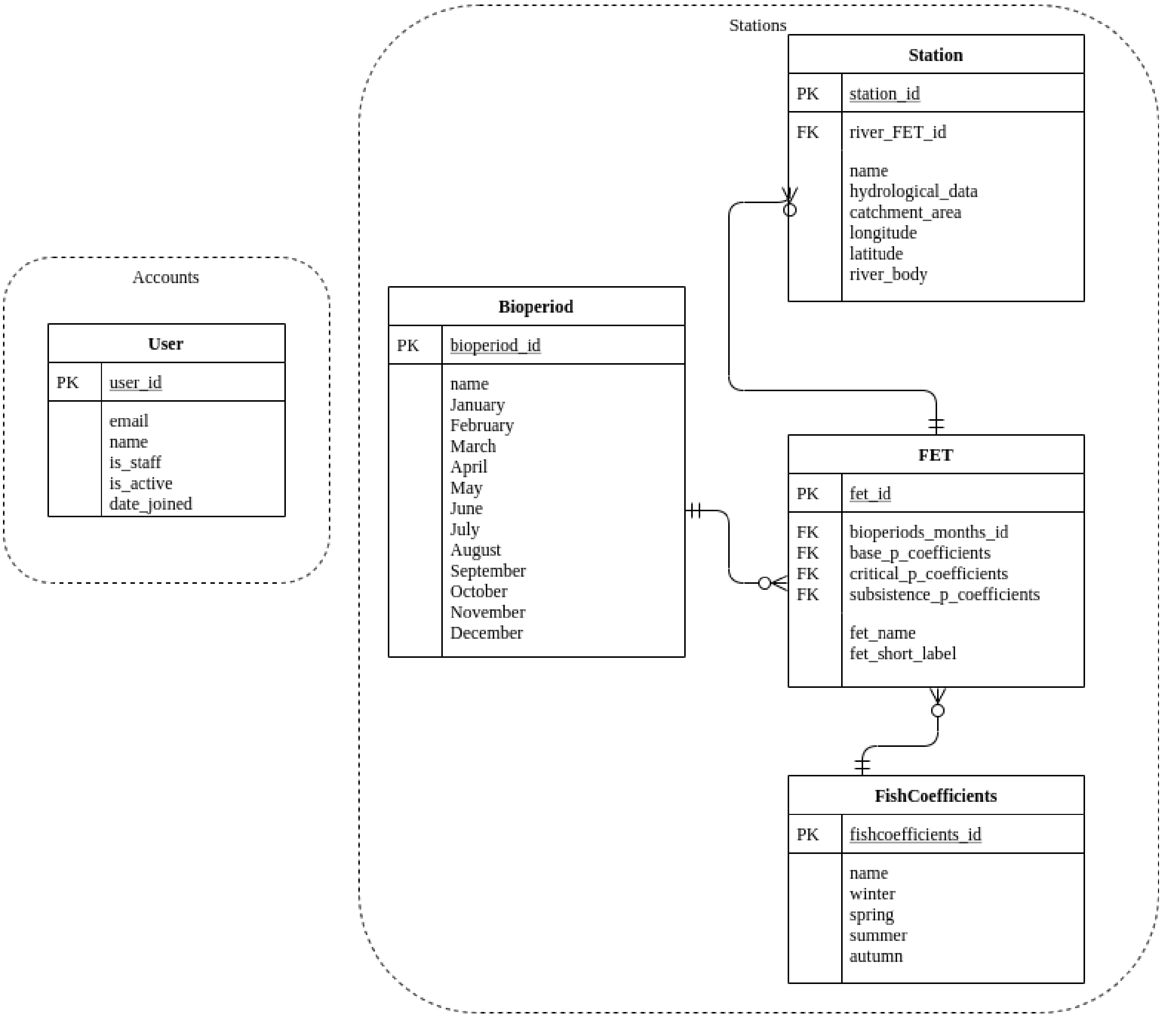


Figure 16. EFCES ERD

Non-relational data (spreadsheet files, metadata, forecasting information, etc.) are stored in the media storage on the AWS S3 cloud [89].

4.2 Data Analysis and Modelling

4.2.1 Data Exploration and Analysis

As it has been described in Chapter 3.2, the hydrological data to be analysed are Estonian hydrological OGD.

For each of the variables of 54 river bodies monitored by gauging stations, there has been carried out an automated analysis indicating missing values, distributions of values and their correlation.

In the paper, the results of the time series analysis are presented for the river bodies of Aesoo and Narva linn gauging stations. The script allowing to generate profiles for all the river bodies is added to Appendix 2. The heatmaps of the Pearson correlation matrices are presented for all the rivers.

Daily mean discharge values:

Mean discharge time series plot of Aesoo and Narva linn gauging stations is presented in Figure 17.

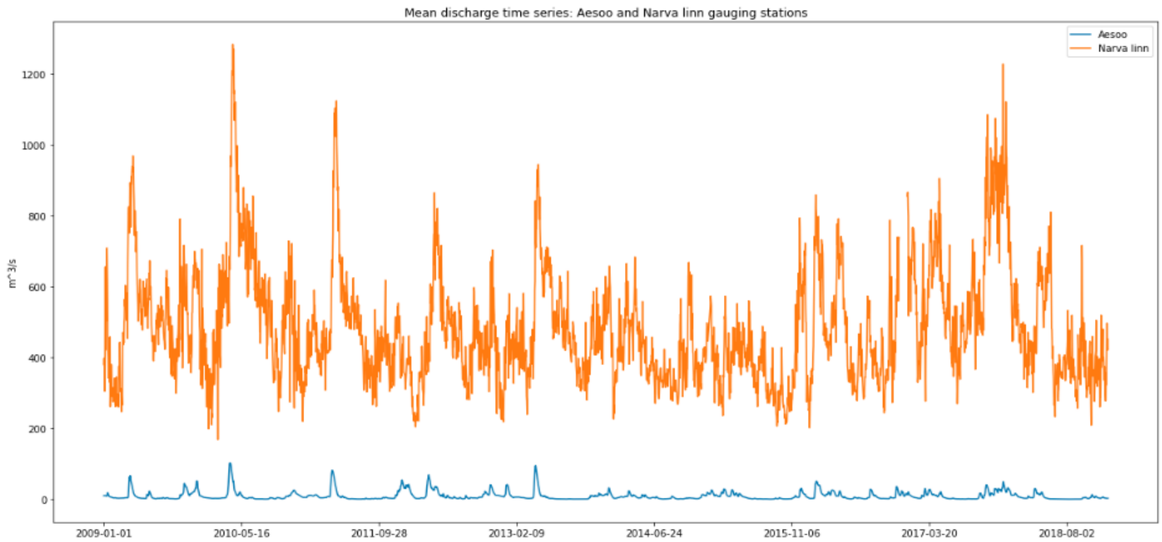
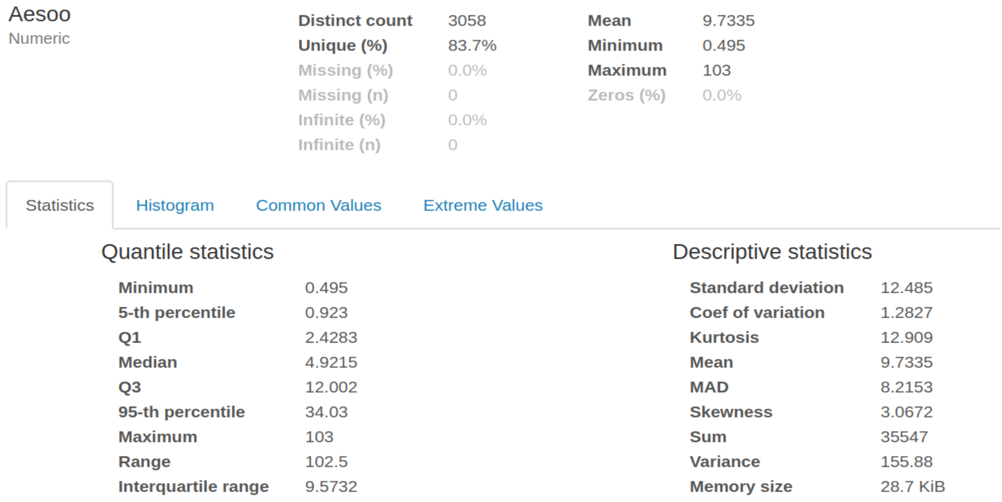


Figure 17. Mean discharge time series of the Aesoo and Narva linn gauging stations

The descriptive statistics for the mean discharge values of the Aesoo gauging station is presented in Figure 18, and for the Narva linn gauging station - in Figure 19.



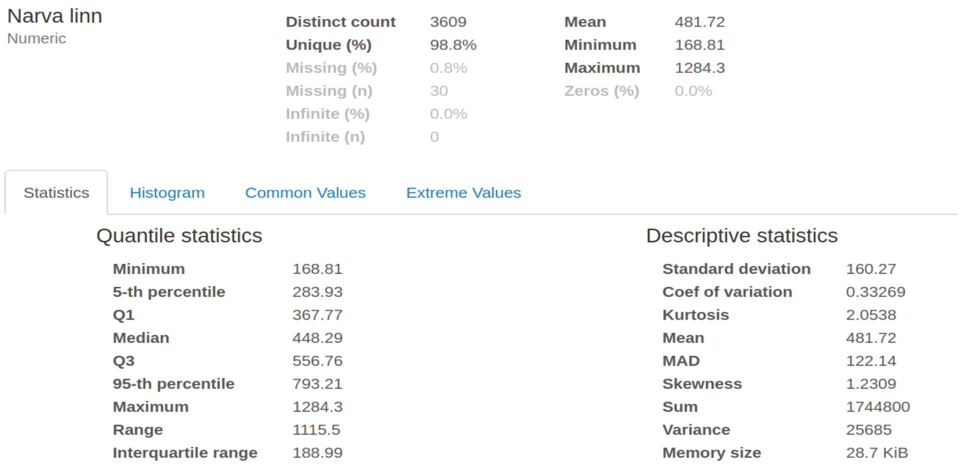
Aesoo
Numeric

Distinct count	3058	Mean	9.7335
Unique (%)	83.7%	Minimum	0.495
Missing (%)	0.0%	Maximum	103
Missing (n)	0	Zeros (%)	0.0%
Infinite (%)	0.0%		
Infinite (n)	0		

Statistics Histogram Common Values Extreme Values

Quantile statistics		Descriptive statistics	
Minimum	0.495	Standard deviation	12.485
5-th percentile	0.923	Coef of variation	1.2827
Q1	2.4283	Kurtosis	12.909
Median	4.9215	Mean	9.7335
Q3	12.002	MAD	8.2153
95-th percentile	34.03	Skewness	3.0672
Maximum	103	Sum	35547
Range	102.5	Variance	155.88
Interquartile range	9.5732	Memory size	28.7 KiB

Figure 18. Descriptive statistics of the mean discharge values of the Aesoo gauging station



Narva linn
Numeric

Distinct count	3609	Mean	481.72
Unique (%)	98.8%	Minimum	168.81
Missing (%)	0.8%	Maximum	1284.3
Missing (n)	30	Zeros (%)	0.0%
Infinite (%)	0.0%		
Infinite (n)	0		

Statistics Histogram Common Values Extreme Values

Quantile statistics

Minimum	168.81
5-th percentile	283.93
Q1	367.77
Median	448.29
Q3	556.76
95-th percentile	793.21
Maximum	1284.3
Range	1115.5
Interquartile range	188.99

Descriptive statistics

Standard deviation	160.27
Coef of variation	0.33269
Kurtosis	2.0538
Mean	481.72
MAD	122.14
Skewness	1.2309
Sum	1744800
Variance	25685
Memory size	28.7 KiB

Figure 19. Descriptive statistics of the mean discharge values of the Narva linn gauging station

For the Aesoo gauging station, there are 3058 mean discharge observations provided (which is around 8.3 years of daily observations). The mean value of all the provided Aesoo discharge values is 9.7335 cubic meters per second, the recorded minimum is 0.495 cubic meters per second, and the maximum is 103 cubic meters per second.

For the Narva linn gauging station, there are 3609 mean discharge observations provided (which is around 9.8 years of daily observations) with 0.8% of missing values (around a month). The mean value of all the provided Narva linn discharge values is 481.72 cubic meters per second, the recorded minimum is 168.81 cubic meters per second, and the maximum is 1284.3 cubic meters per second. On average, the river body of the Narva linn station is around 50 times faster than the river monitored by the Aesoo station.

The histogram of the Aesoo mean discharge values is presented in Figure 20, and the histogram of the Narva linn mean discharge values is presented in Figure 21.

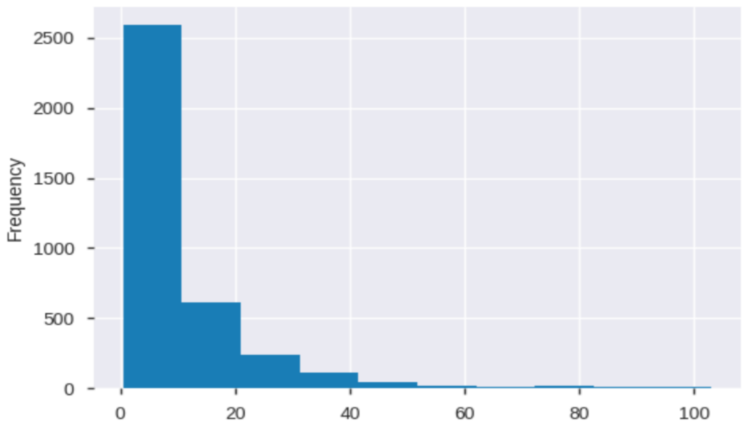


Figure 20. Histogram of the mean discharge values of the Aesoo gauging station

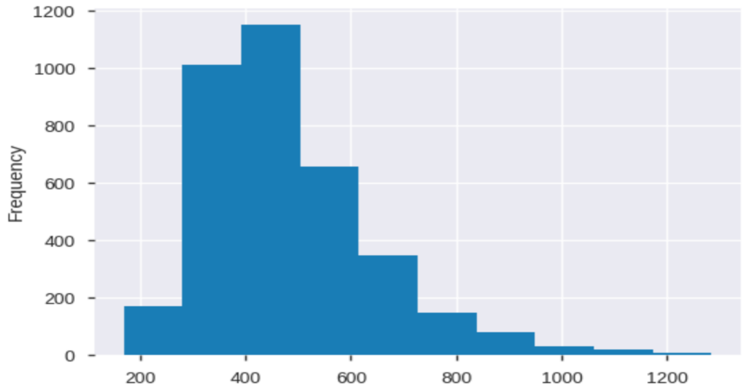


Figure 21. Histogram of the mean discharge values of the Narva linn gauging station

The estimation of Pearson correlation [90] between mean discharge values of different stations (Figure 22) shows high positive correlation among most of the rivers which is going to be further used for forecasting modelling based on the data from other stations.

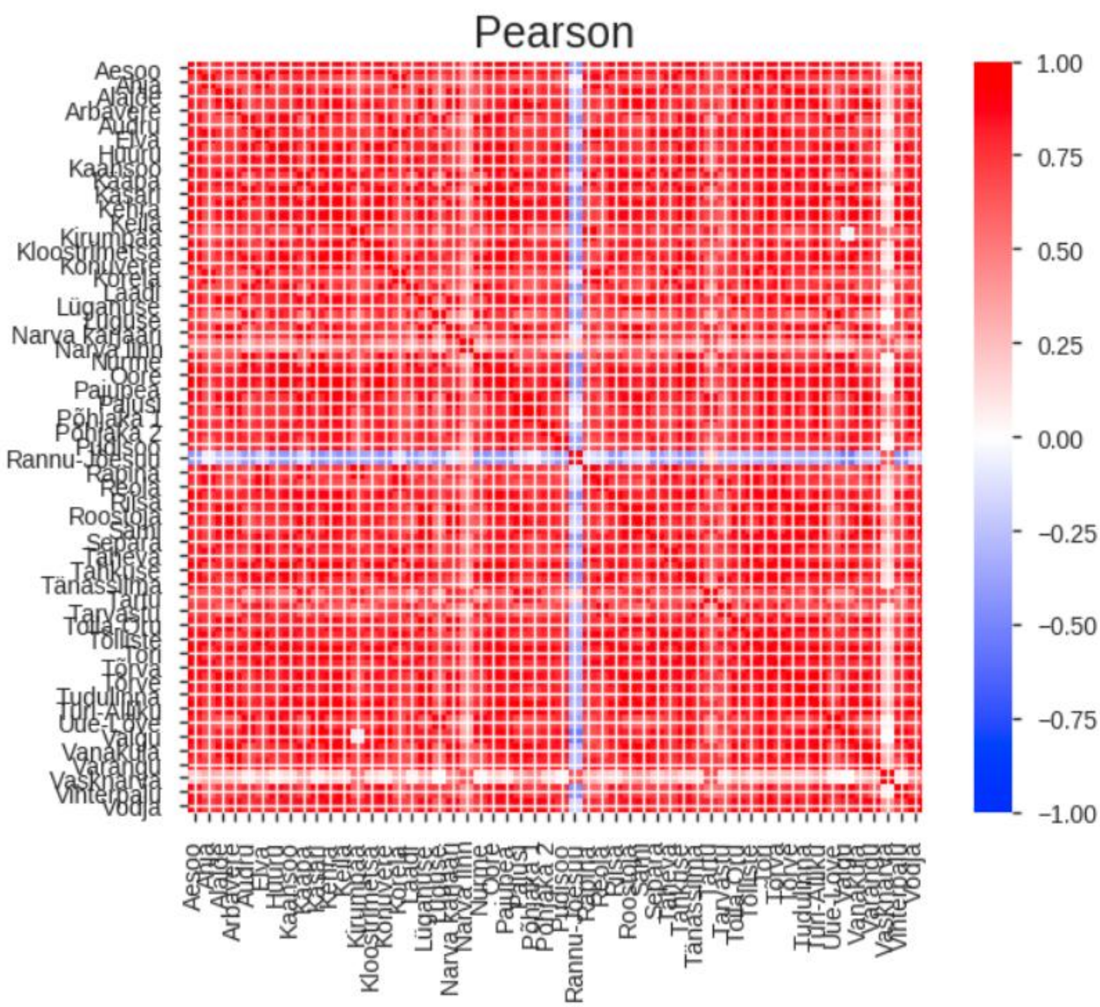


Figure 22. Mean discharge Pearson correlation matrix

Daily mean water temperature:

Mean water temperature time series plot of Aesoo and Narva linn gauging stations is presented in Figure 23.

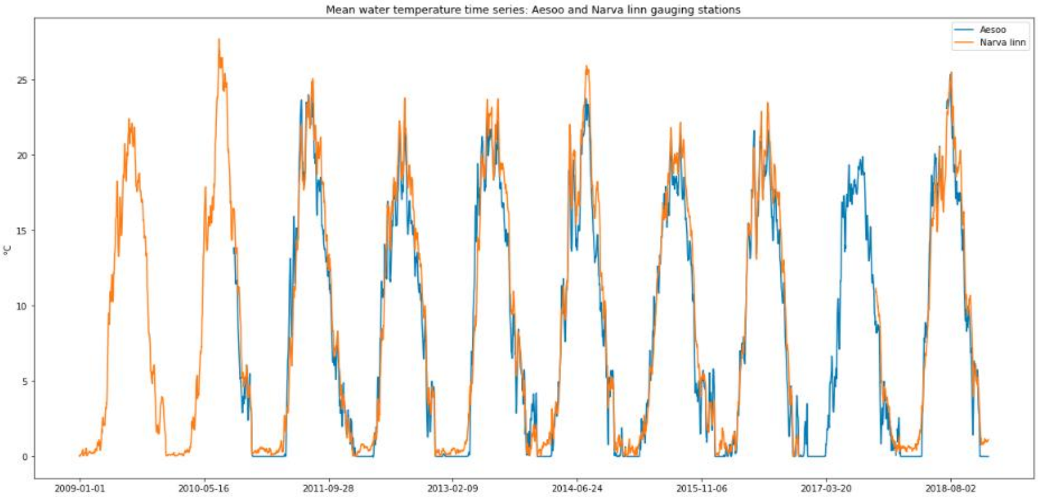


Figure 23. Mean water temperature time series of the Aesoo and Narva linn gauging stations

As it can be seen from the time series plot, the mean water temperature values of the Narva linn gauging station are similar to the ones of the Aesoo gauging station. That is why for this hydrologic feature the descriptive statistics and histogram are presented only for the Aesoo gauging station.

The descriptive statistics of the mean water temperature values of the Aesoo gauging station is presented in Figure 24.

Aesoo
Numeric

Distinct count	1389	**Mean**	8.2117
Unique (%)	38.0%	**Minimum**	-0.04
Missing (%)	17.3%	**Maximum**	25.31
Missing (n)	633	**Zeros (%)**	18.7%
Infinite (%)	0.0%		
Infinite (n)	0		

Statistics | Histogram | Common Values | Extreme Values

Quantile statistics

Minimum	-0.04
5-th percentile	0
Q1	0.1
Median	6.01
Q3	15.455
95-th percentile	20.862
Maximum	25.31
Range	25.35
Interquartile range	15.355

Descriptive statistics

Standard deviation	7.6378
Coef of variation	0.93011
Kurtosis	-1.3221
Mean	8.2117
MAD	6.8851
Skewness	0.39777
Sum	24791
Variance	58.336
Memory size	28.7 KiB

Figure 24. Descriptive statistics of the mean water temperature values of the Aesoo gauging station

There are 1389 observations provided (which is around 3.8 years of daily observations), 17.3% of which is missing which is around 8 months. Compared to the provided discharge values, there are more than 2.5 times fewer data available. The mean value of all the provided Aesoo mean water temperatures values is 8.2117 degrees Celsius, the recorded minimum is -0.04 degrees Celsius, and the maximum is 25.31 degrees Celsius.

The histogram of the Aesoo mean water temperature values is presented in Figure 25.

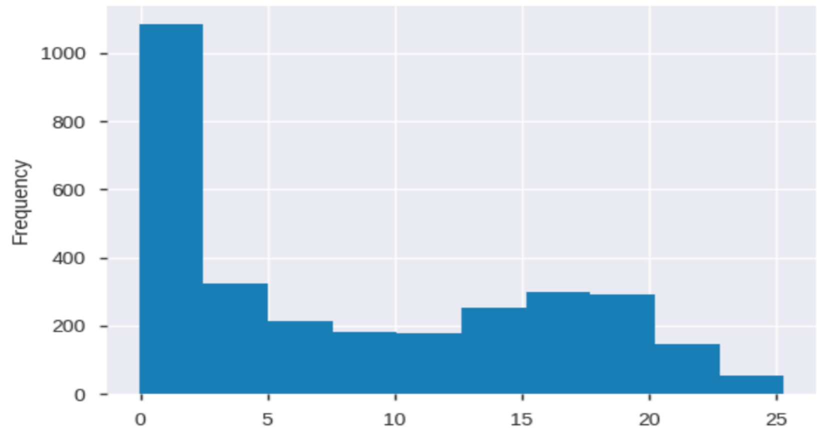


Figure 25. Histogram of the mean water temperature values of the Aesoo gauging station

The mean water temperature observations of different stations show high Pearson correlation, as it can be seen from Figure 26.

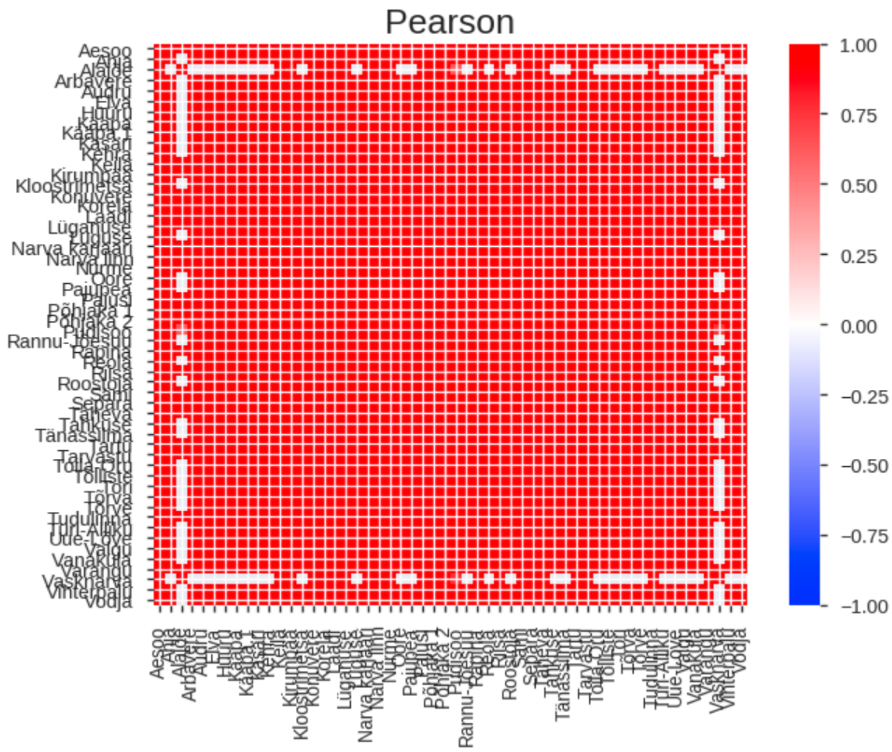


Figure 26. Mean water temperature Pearson correlation matrix

Daily mean water levels:

Mean water level time series plot of Aesoo and Narva linn gauging stations is presented in Figure 27.

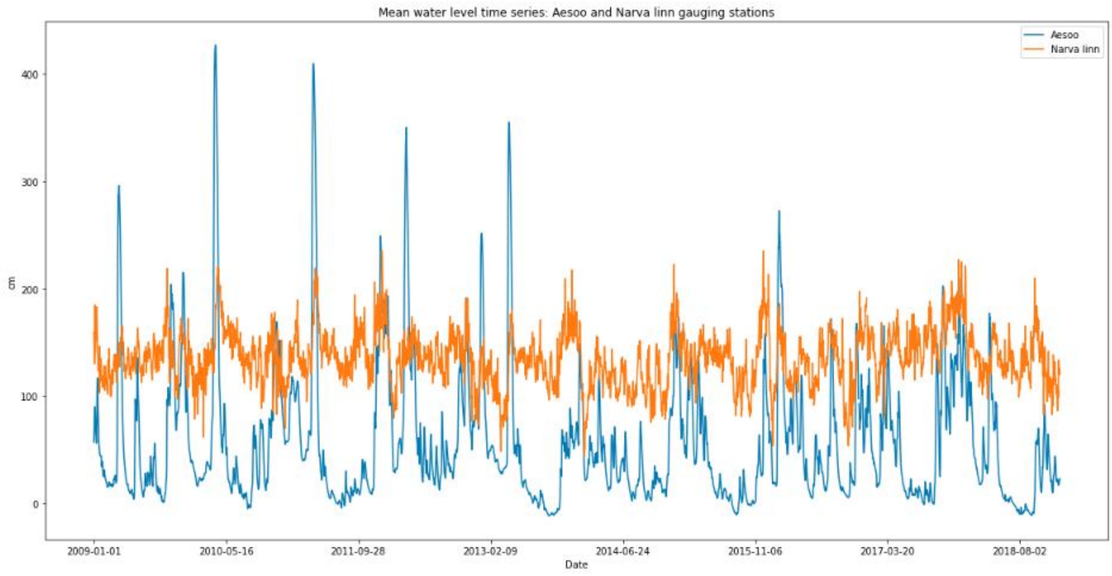


Figure 27. Mean water level time series of the Aesoo gauging station

The descriptive statistics of the mean water level values of the Aesoo gauging station is presented in Figure 28. The descriptive statistics of the mean water level values of the Narva linn gauging station is presented in Figure 29.

Aesoo
Numeric

Distinct count	2770	**Mean**	56.714
Unique (%)	75.8%	**Minimum**	-11.61
Missing (%)	0.0%	**Maximum**	427
Missing (n)	1	**Zeros (%)**	0.0%
Infinite (%)	0.0%		
Infinite (n)	0		

Statistics | Histogram | Common Values | Extreme Values

Quantile statistics

Minimum	-11.61
5-th percentile	-1.47
Q1	14.62
Median	37.59
Q3	78.015
95-th percentile	174.22
Maximum	427
Range	438.61
Interquartile range	63.395

Descriptive statistics

Standard deviation	62.756
Coef of variation	1.1065
Kurtosis	6.788
Mean	56.714
MAD	44.803
Skewness	2.2092
Sum	207060
Variance	3938.3
Memory size	28.7 KiB

Figure 28. Descriptive statistics of the mean water levels values of the Aesoo gauging station

Narva linn
Numeric

Distinct count	3035	**Mean**	135.26
Unique (%)	83.1%	**Minimum**	44.21
Missing (%)	0.0%	**Maximum**	235.22
Missing (n)	1	Zeros (%)	0.0%
Infinite (%)	0.0%		
Infinite (n)	0		

Statistics | Histogram | Common Values | Extreme Values

Quantile statistics

Minimum	44.21
5-th percentile	94.925
Q1	118.82
Median	134.82
Q3	151.3
95-th percentile	179.18
Maximum	235.22
Range	191.01
Interquartile range	32.47

Descriptive statistics

Standard deviation	25.618
Coef of variation	0.1894
Kurtosis	0.57256
Mean	135.26
MAD	19.84
Skewness	0.19737
Sum	493830
Variance	656.26
Memory size	28.7 KiB

Figure 29. Descriptive statistics of the mean water levels values of the Narva linn gauging station

The histogram of the Aesoo mean water level values is presented in Figure 30. The histogram of the Aesoo mean water level values is presented in Figure 31.

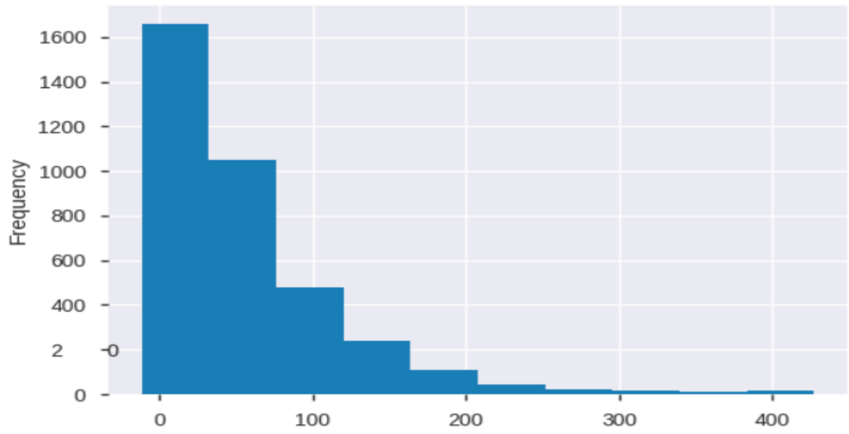


Figure 30. Histogram of the mean water levels values of the Aesoo gauging station

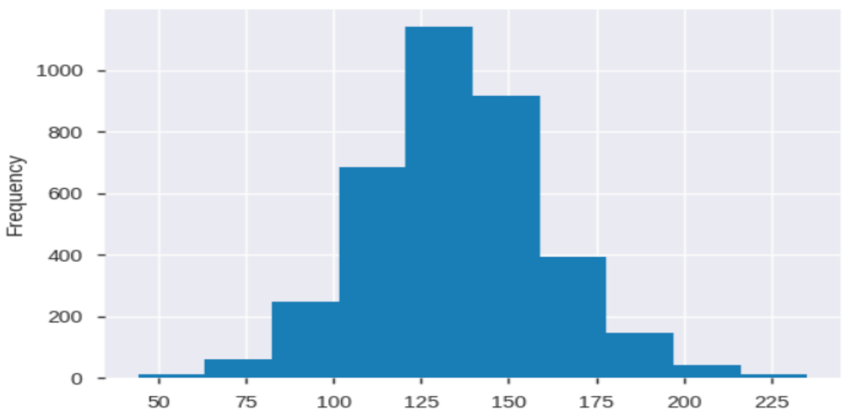


Figure 31. Histogram of the mean water levels values of the Narva linn gauging station

For the Aesoo gauging station, there are 2770 mean water level observations provided (which is around 7.5 years of daily observations). The mean value of all the provided Aesoo water level values is 56.714 cm, the recorded minimum is -11.61 cm, and the maximum is 427 cm.

For the Narva linn gauging station, there are 3035 mean water level observations provided (which is around 8.3 years of daily observations). The mean value of all the provided Narva linn water level values is 135.26 cm, the recorded minimum is 44.21 cm, and the maximum is 235.22 cm.

The estimation of Pearson correlation [90] between mean water level values of different stations is presented in Figure 32.

The matrix shows that there are quite many highly correlated stations but compared to mean discharge and water temperature values, water level correlation between stations seems to vary more (lighter zones on the heatmap not considering missing data).

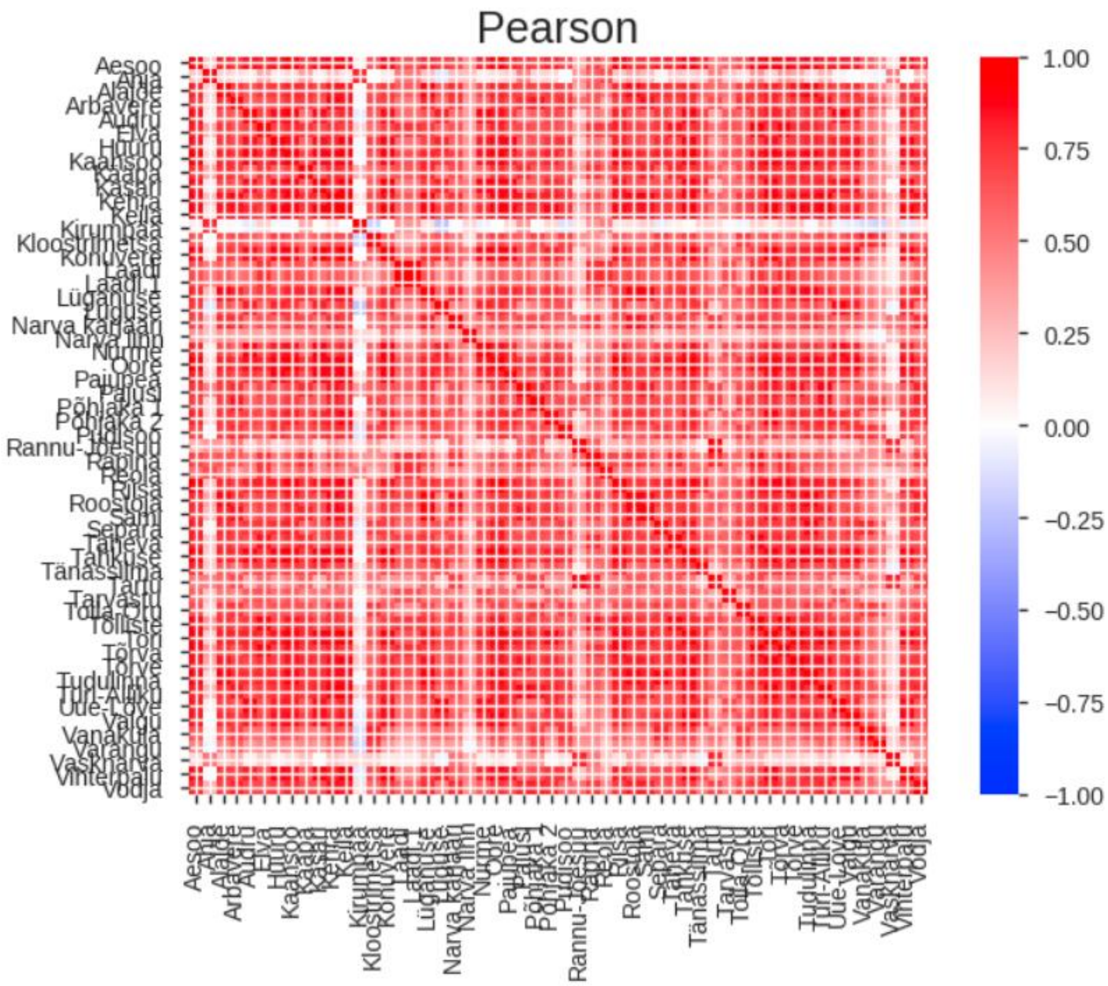


Figure 32. Mean water level Pearson correlation matrix

Summary. The analysed daily mean hydrological time series are going to be used to fit regression models and make hydrological forecasts using them. The goal is to build forecasters allowing prediction of a station's hydrological values based on the other stations' data of both the same and different hydrological measurement types (e.g. predict discharge based on other stations' discharge and predict discharge based on the other stations' water temperature).

The correlation analysis has shown that for the same measurement type (discharge, water temperature and water level), there are many highly correlated station variables meaning that their values can be used to predict the other station's values.

However, it is more difficult in case of different measurement types. Thus, the correlation analysis of daily mean water temperatures has shown that all the stations have correlation coefficients close to 1, meaning that the dependency between any stations can be defined with a linear equation. The water temperature time series data are largely determined by Estonia's relatively uniform hydroclimate and homogeneous river morphologies. As a result, the differences observed between water temperature time series data at different gauging stations are too small to be used for the prediction of the more dynamic changes found at each gauging station's depth and discharge relations.

In cases where a river's channel has a simple geometry and flows change slowly, discharge can be used to predict water levels and vice versa.

Rivers discharge and water levels dynamics have different variability. For example, the river monitored by the Aesoo station (Navesti jõgi) compared to the river monitored by the Narva linn station is much slower (smaller discharge), but has higher variability of the water levels that can be seen from the time series plots and histograms. To describe the variation of the time series with linear models that are going to be used for fitting, longer time series of different variations or more sophisticated models might be required for more precise fitting.

The rivers also have a natural spatial dependency: being affected by the localized precipitation, the quantity of water in the rivers monitored by neighbouring stations is supposed to change similarly. Since some of the stations have a large number of highly correlated variables (more than 10), there should be established an approach for feature selection that is described in Chapter 4.2.2.

4.2.2 Modelling and Feature Selection

The map with the locations of gauging stations [71] is presented in Figure 33.

Given the latitude and longitude of different stations, there has been calculated a Euclidean distance matrix representing the distance between stations in kilometres. After that, the obtained set of Euclidean distances $\{dist(i, j)\}$ between stations (i and j) were normalized to have values from 0 to 1 resulting in $\{normdist(i, j)\}$ with 0 corresponding to the same location $i = j$, and 1 - the longest distance between any of the stations. The longest distance $\max(\{dist(i, j)\})$ turned out to be between Narva linn and Uue-Lõve gauging station (around 324.54 km).

Figure 33. Estonian gauging stations map (56 stations are shown)

In order to take into account both spatial dependencies and correlation between stations $corr(i, j)$ in the same manner, a matrix of station weights $\{w(i, j)\}$ has been calculated, where

$$w(i, j) = \frac{|corr(i, j)| + (1 - normdist(i, j))}{2}$$

$$normdist(i, j) = \frac{dist(i, j)}{\max(\{dist(i, j)\})}$$

The set of $normdist(i, j)$ corresponds to the values obtained for the normalized Euclidean distance matrix (Figure 34). The $corr(i, j)$ are Pearson coefficients between measurements of different stations [90].

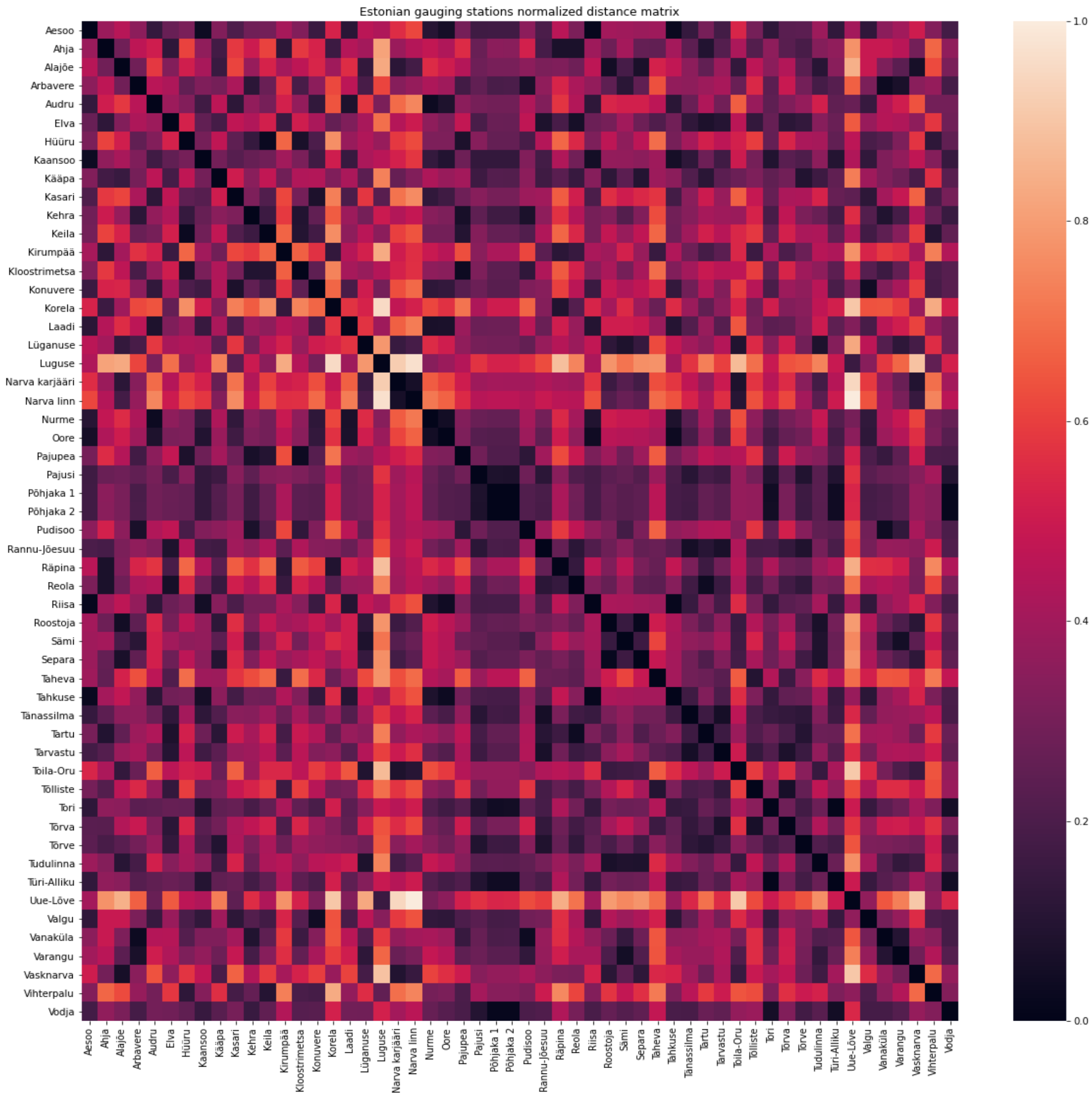


Figure 34. Normalized Euclidean distance matrix of Estonian gauging stations

Taking Pearson coefficients between mean discharge measurements, the resulting weights matrix is presented as a heatmap in Figure 35. Weights matrices for mean water temperatures and mean water levels are presented in Figures 36-37.

Having a lot of missing values in the historical data, the network of stations is modelled as a network of regressors that use available data from other stations to predict the values of its station. The regressors are built taking a training sample of one preceding year of available data to predict the selected date range.

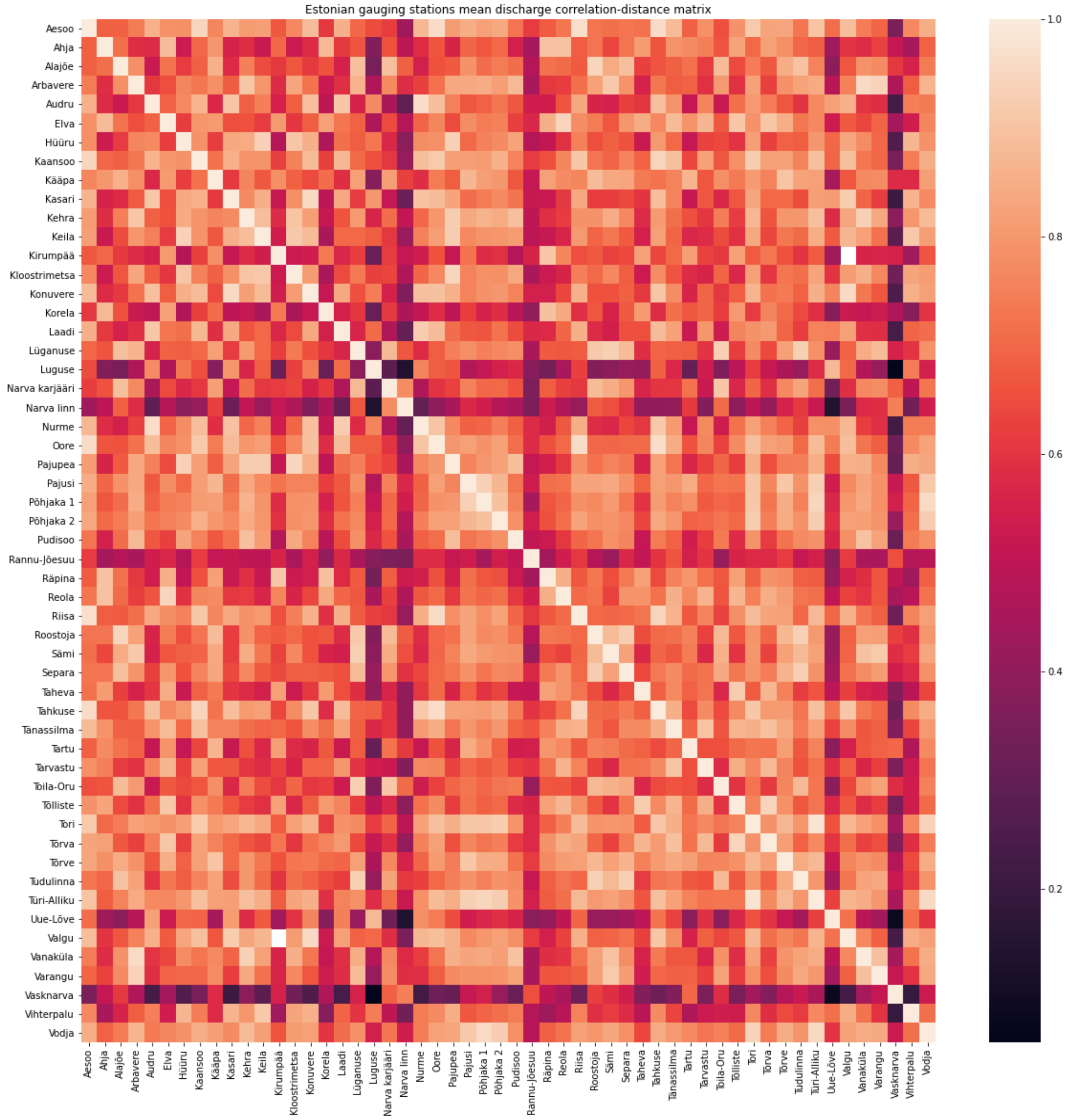


Figure 35. Mean discharge distance-weighted correlation matrix of Estonian gauging stations

Since the network of regressors is supposed to function real-time, and stations' measurements are highly correlated, it is suggested to use simple regression algorithms assuming their simplicity, interpretability, and fast performance to be re-fitted in case of any disruptions (missing data).

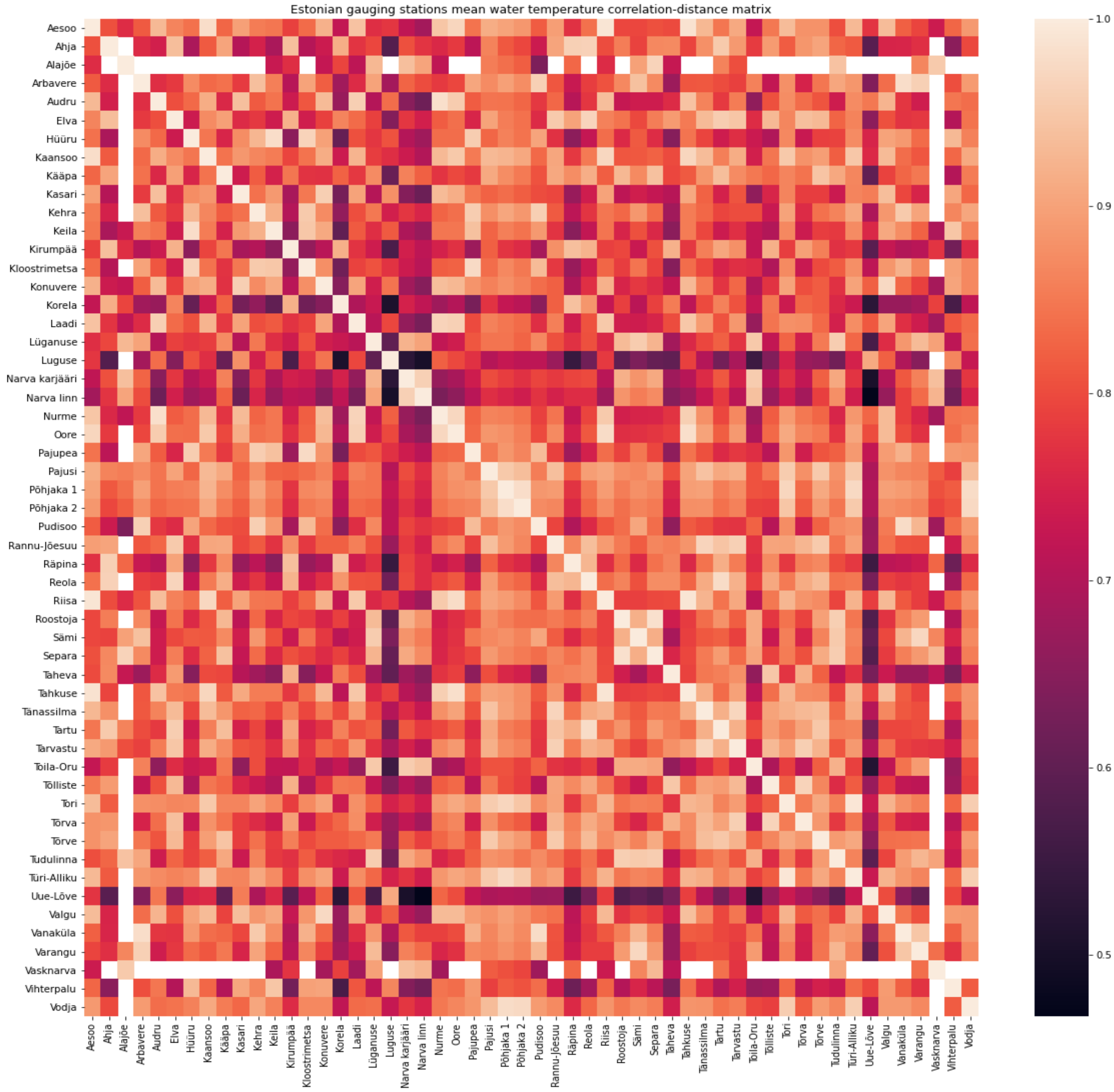


Figure 36. Mean water temperature distance-weighted correlation matrix of Estonian gauging stations

Due to the high correlation between stations measurements, it was possible to use simple regression algorithms for multi-station time series forecasting:

- Linear regression [91]
- Ridge regression [92]
- Lasso regression [93]

These methods are suitable for real-time forecasting and can be retrained quickly compared to more complex methods that do not boost the performance for the given dataset.

Regressors are fitted until one of the regressors fits with R2 (R squared - determination coefficient [94]) is not less than 0.95. If none of them fits with the mentioned determination metric, the best one of the trained regressors is used.

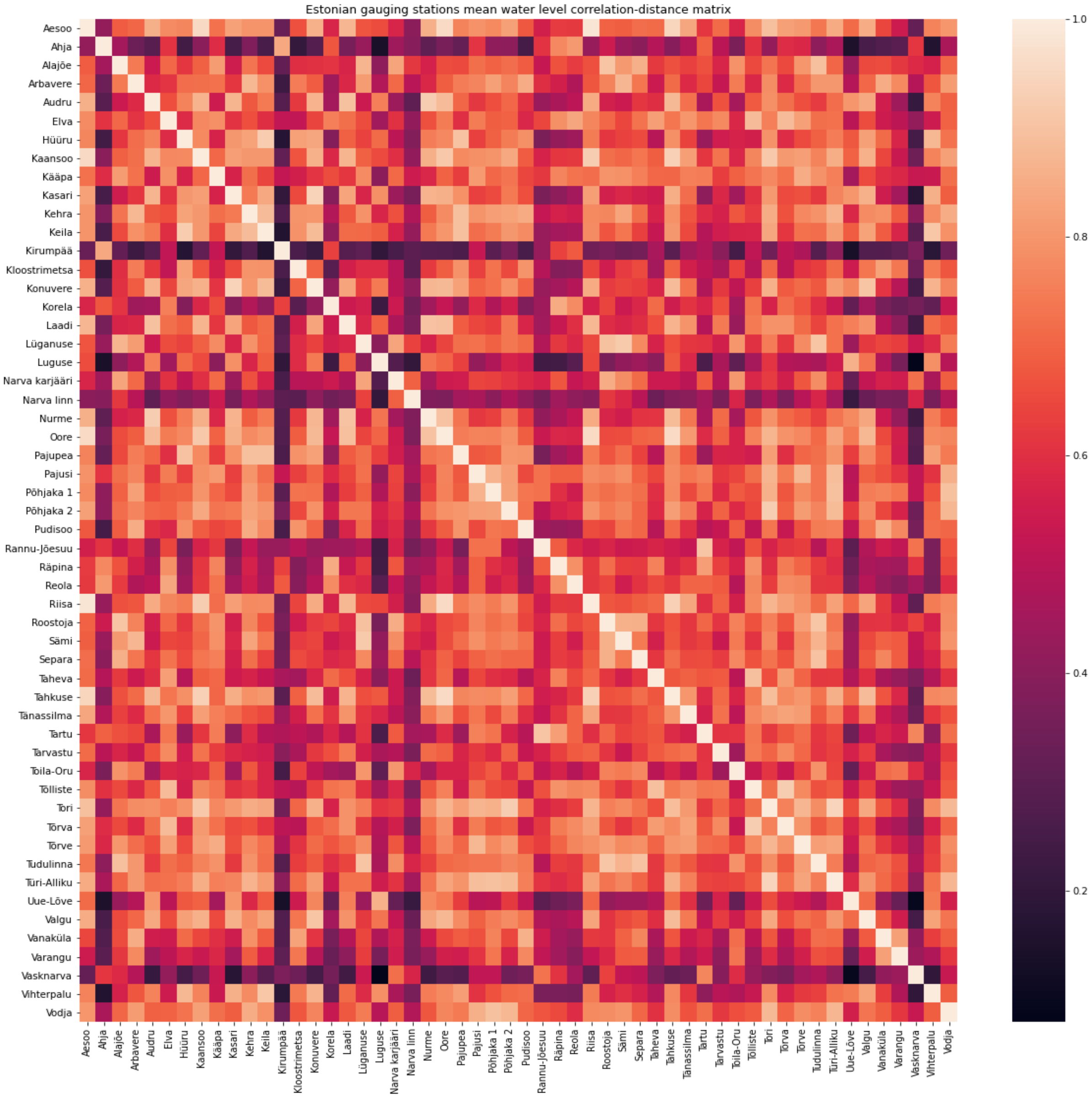


Figure 37. Mean water level distance-weighted correlation matrix of Estonian gauging stations

Feature selection based on distance-correlation weights is carried out in case of more than 10 available stations for linear regression and ridge regression. Assuming that the Lasso regression algorithm is capable of doing feature selection on its own assigning zero weights to not needed features, distance-correlation-based feature selection for this regressor is not carried out.

4.2.3 Forecasting

The network of gauging stations is modelled as the network of fast regressors that use other station's data for prediction and/or other types of data and re-fit to take into account spatial dependency, the correlation between measurements and recent one year long temporal dependencies. The overall network forecasting model is presented in Figure 38.

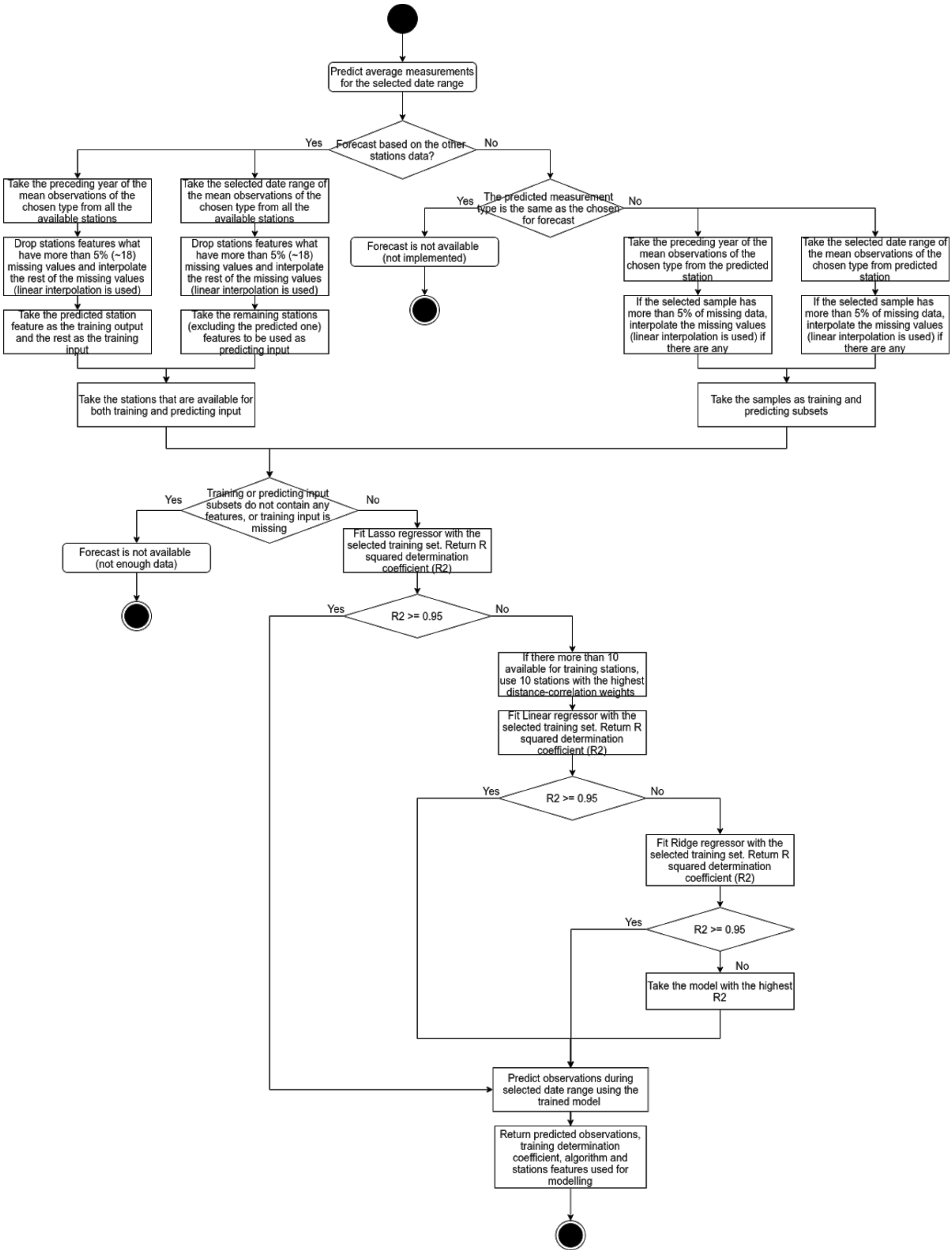


Figure 38. Estonian gauging station hydrology forecasting model

The performance of the model is demonstrated for different measurements made at the Aesoo and Narva linn gauging stations in 2015-2016. The actual and forecasted time series are presented in Figures 39-44.

Examples of forecasting based on the measurements from other stations is presented in Figures 45-46.

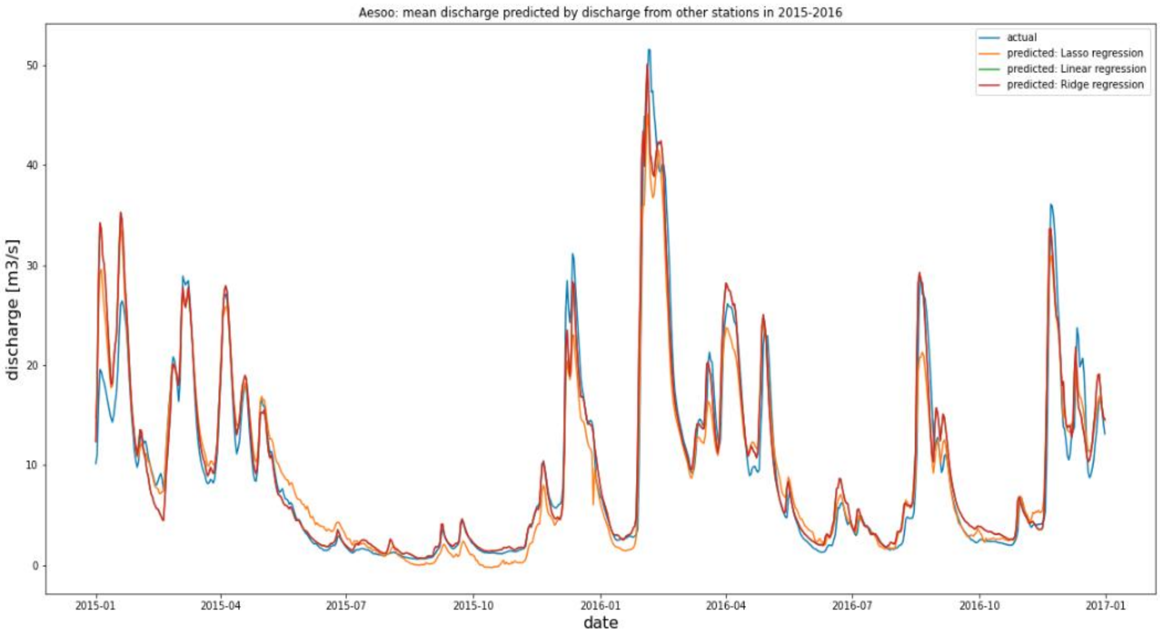


Figure 39. Mean discharge time series forecasting for Aesoo gauging station based on other stations data in 2015-2016

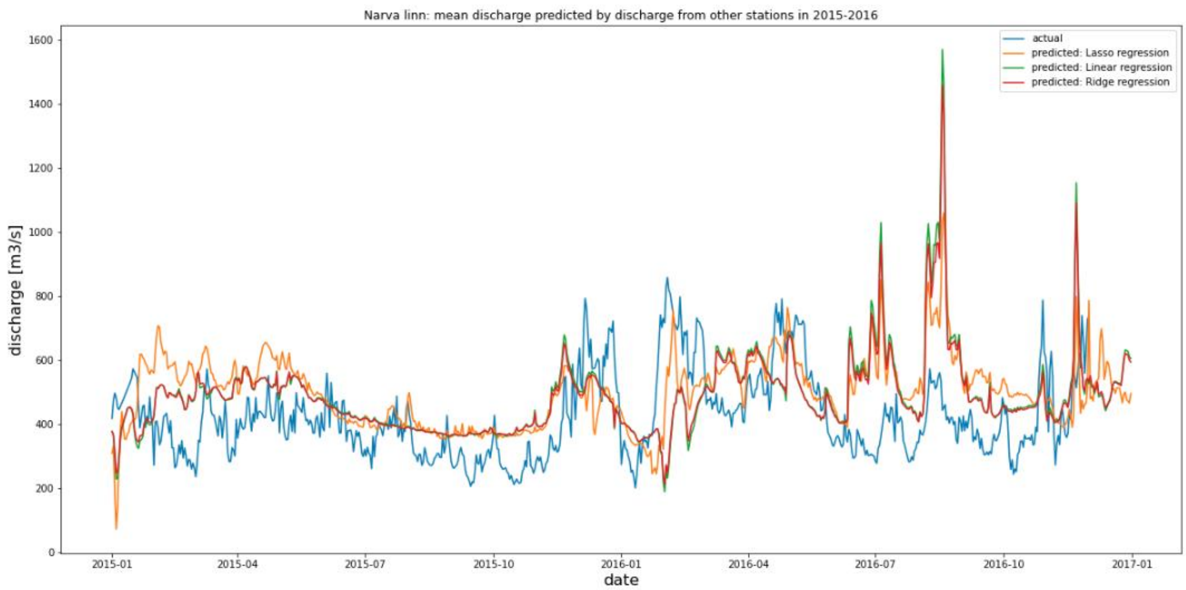


Figure 40. Mean discharge time series forecasting for Narva linn gauging station based on other stations data in 2015-2016

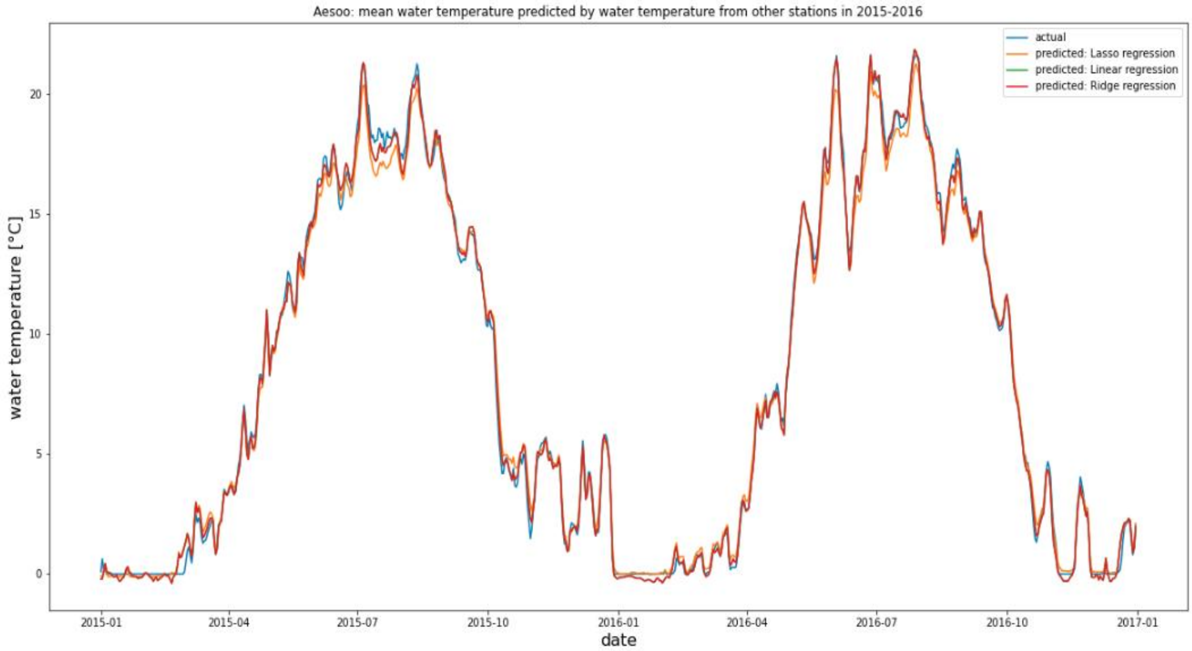


Figure 41. Mean water temperature time series forecasting for Aesoo gauging station based on other stations data in 2015-2016

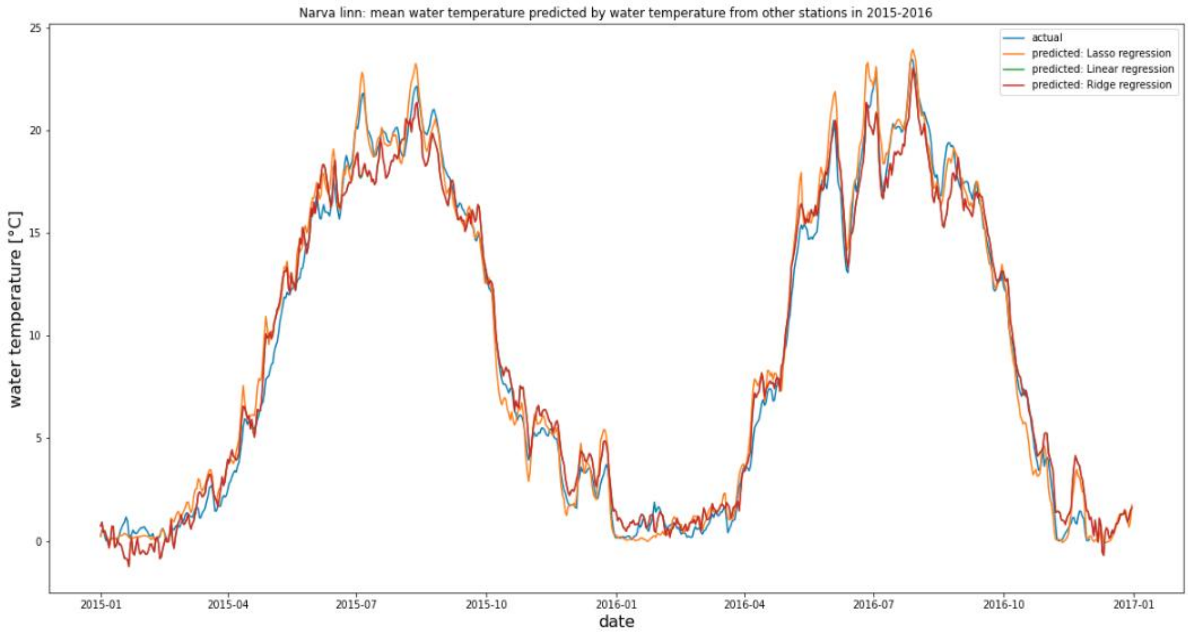


Figure 42. Mean water temperature time series forecasting for Narva linn gauging station based on other stations data in 2015-2016

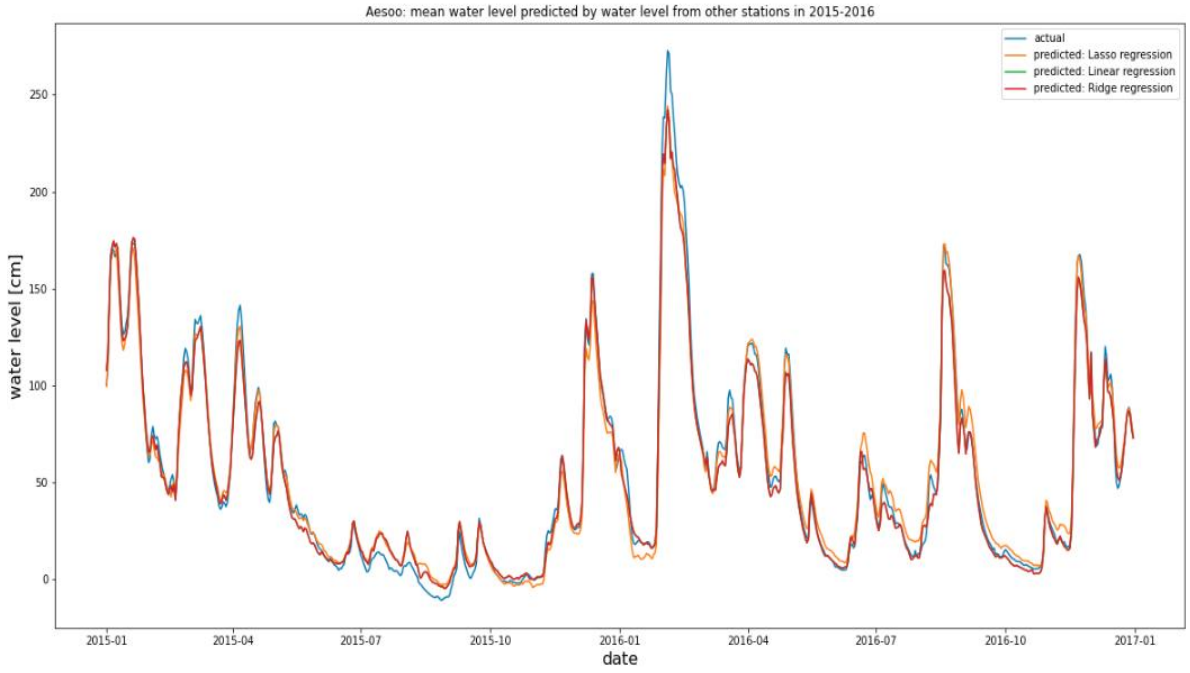


Figure 43. Mean water level time series forecasting for Aesoo gauging station based on other stations data in 2015-2016

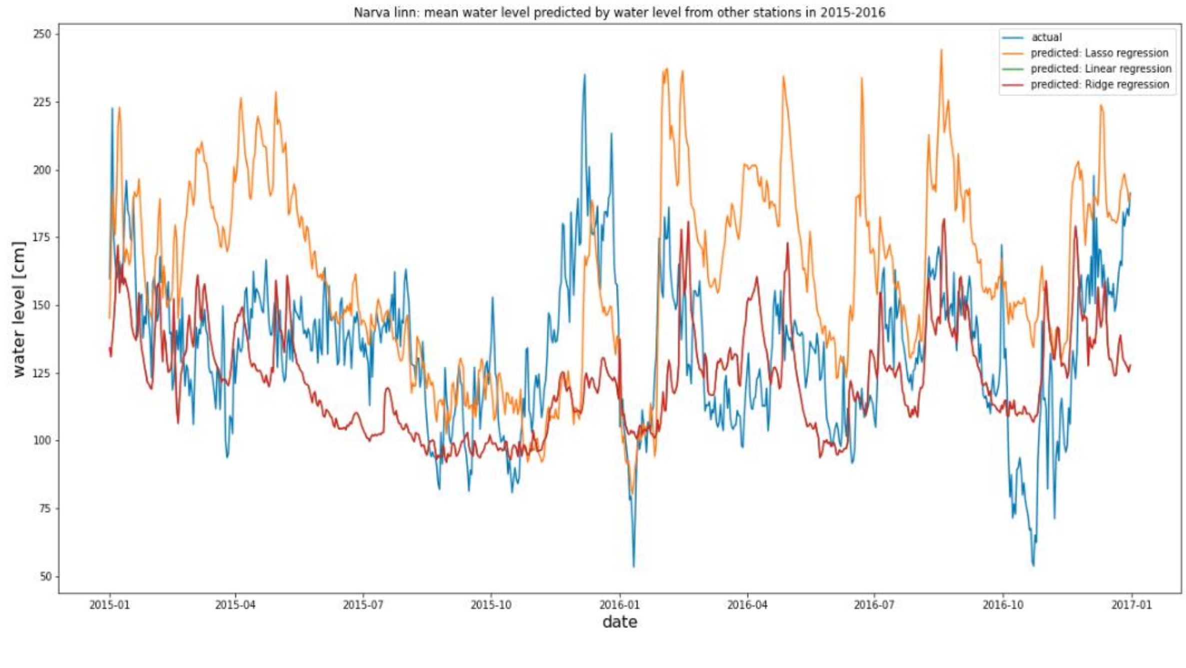


Figure 44. Mean water level time series forecasting for Narva linn gauging station based on other stations data in 2015-2016

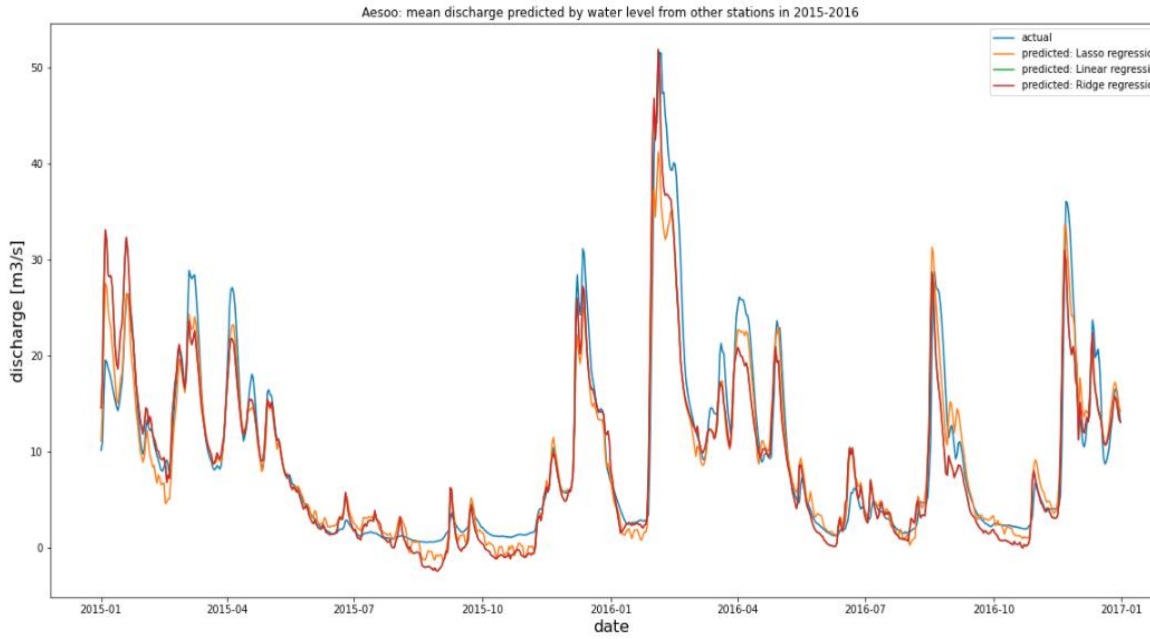


Figure 45. Mean discharge time series forecasting for Aesoo gauging station based on the water level at other stations data in 2015-2016

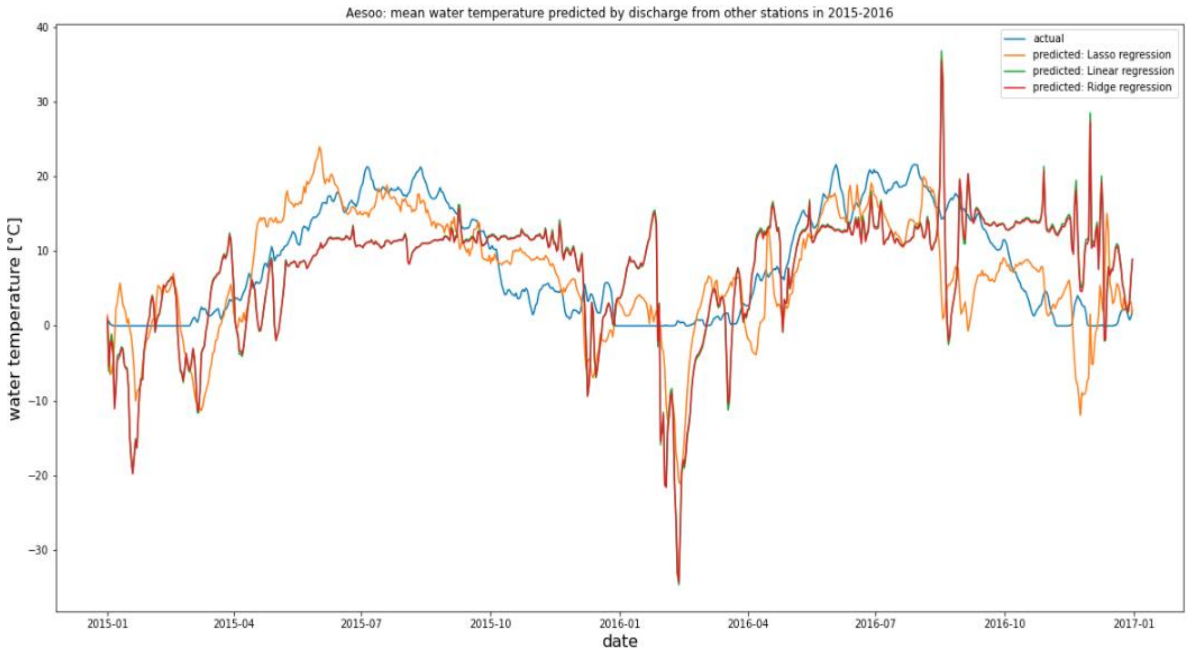


Figure 46. Mean water temperature time series forecasting for Aesoo gauging station based on discharge at other stations data in 2015-2016

The presented forecasting model is integrated into EFCES.

4.3 Service Implementation

4.3.1 User Interface

The frontend part of EFCES is implemented as a Single Page Application (SPA) that dynamically changes current pages instead of not reloading them entirely from the server. The SPA principles enact smoother transitions making the web application resemble native applications. Multiple JS frameworks adopt SPA techniques allowing keeping a single page even during communication with the server.

The JS framework used for the EFCES is React (ReactJS [95]). With Redux [88], it allows managing states providing vast opportunities for complex dynamic applications.

The main views are:

- Login view (Figure 47) providing user authentication and access to the stations' information in case of successful authorisation. User authentication involves entering valid credentials: email and password. Currently, the credentials can be obtained from the service administrator that is supposed to create an account for a new user from the administration panel. The successful authorisation implies access only to the frontend application but not the administration panel.

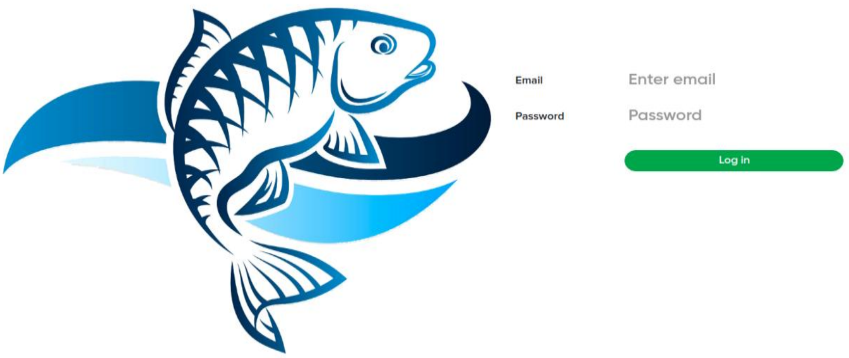


Figure 47. Login view UI of EFCES

- List of gauging stations view (Figure 48) providing the list of gauging stations along with their monitored river bodies, state catchment areas, default FETs (more precise coefficients of which are supposed to be found out while further research activities), locations (longitudes and latitudes).

The selection of one of the presented gauging stations redirects to the selected gauging station's details view. The view also includes "Logout" option and email of the authorised user. "Logout" link leads to the login view requiring authentication.

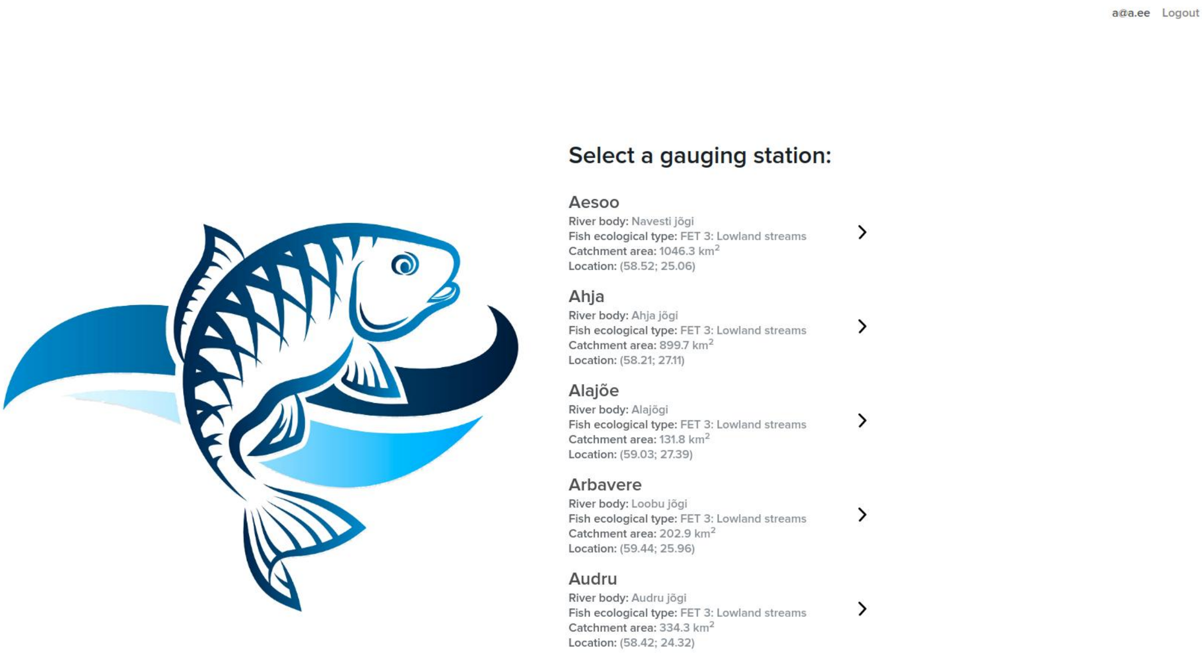


Figure 48. Gauging stations list UI of EFCES

- Gauging station details view (Figures 49-56) includes configuration panels and interprets the estimated values of eflows and their compliance based on the provided historical data as well as forecasted values. It also includes interpretations for the water temperature and water levels data and allows estimation of the exceedance of the specified by a user threshold.

The parameters of the assessment that can be configured by a user include:

- Data range (a year or custom selection),
- Long-term low flow calculation method (Tennant methods as well as p% exceedance probability),
- Long-term low flow calculation frequency (monthly, by bioperiod, season, all the data),
- Catchment area (catchment area at the gauging area by default),
- Catchment area factor (divides the eflow value by a factored catchment area assuming that it is constant),
- FET and corresponding eflow threshold types (coefficients values are provided in the table),
- Primary axis measurements: minimum, maximum, average, all (range) discharge,

- Secondary axis time series type (water temperature or water level),
- Secondary axis measurements: minimum, maximum, average, all (range) water temperatures or water levels,
- Second axis measurement threshold value,
- Forecasting configuration for both primary and secondary axes time series (time series type and single/multiple stations for input; filling missing historical data with the predicted values (if available).

The parameters are separated into tab panels:

- Low flow parameters (Figure 49)
- Environmental low flow formula (Figure 50)
- Compound event (Figure 51)
- Forecasting (Figure 52)

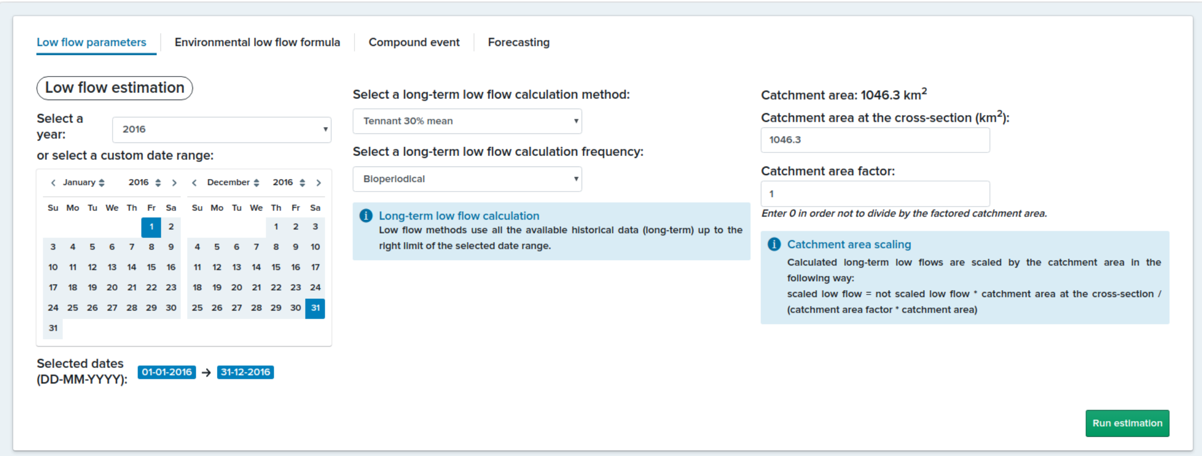


Figure 49. Low flow parameters configuration

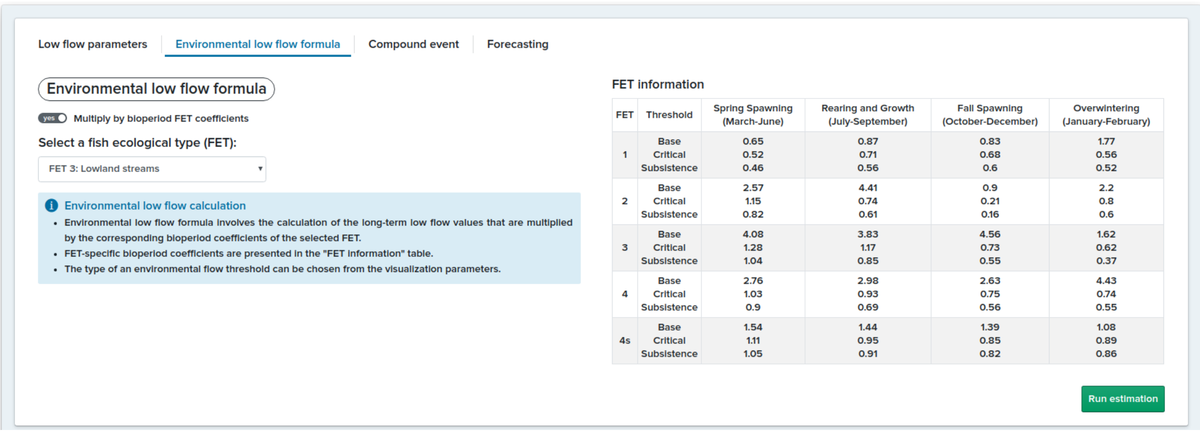


Figure 50. Environmental low flow formula configuration

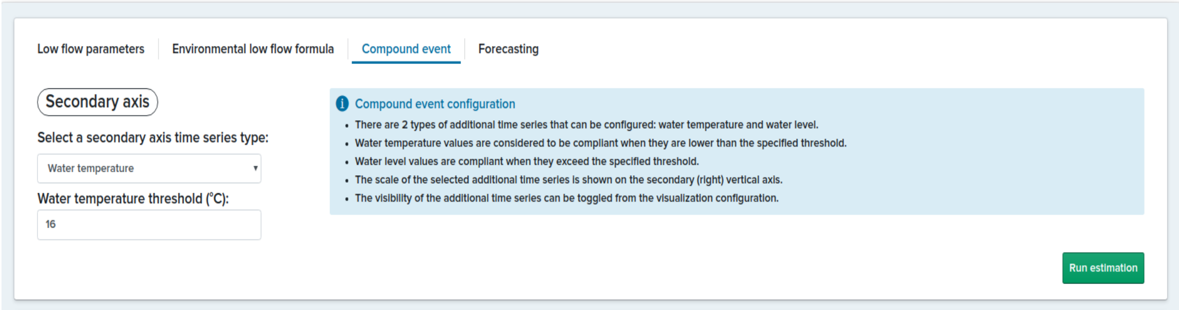


Figure 51. Compound event configuration

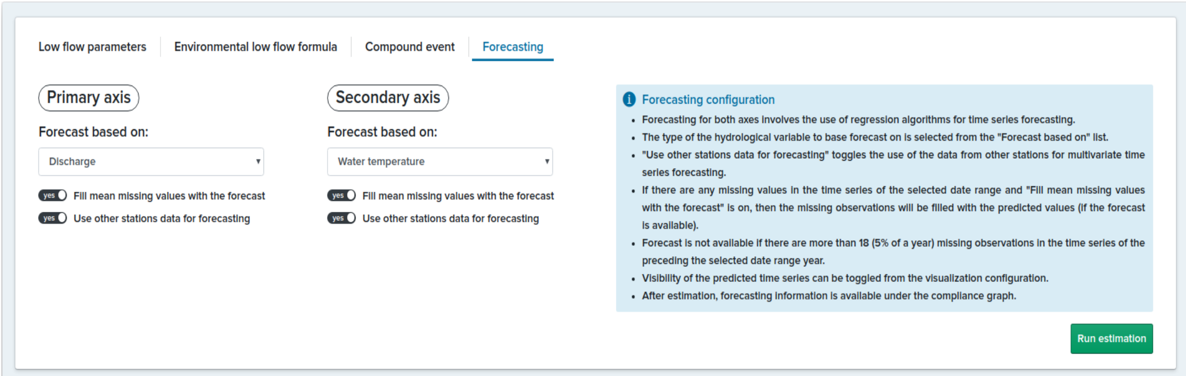


Figure 52. Forecasting configuration

Since the service involves time-consuming operations impeding near-instantaneous estimations, the calculations are triggered by the "Run estimation" button.

The tabs are supposed to handle different necessities of users. For example, users that are not aware of the environmental low flow formula (RAELFF) can only low flows methods to calculate eflow rates. "Compound event" feature allows estimation of water temperatures and levels for compound compliance estimation.

Additionally, there is a visualisation panel (Figure 53) allowing dynamical configuration of the parameters of graphs since it does not require server communication (only state update).

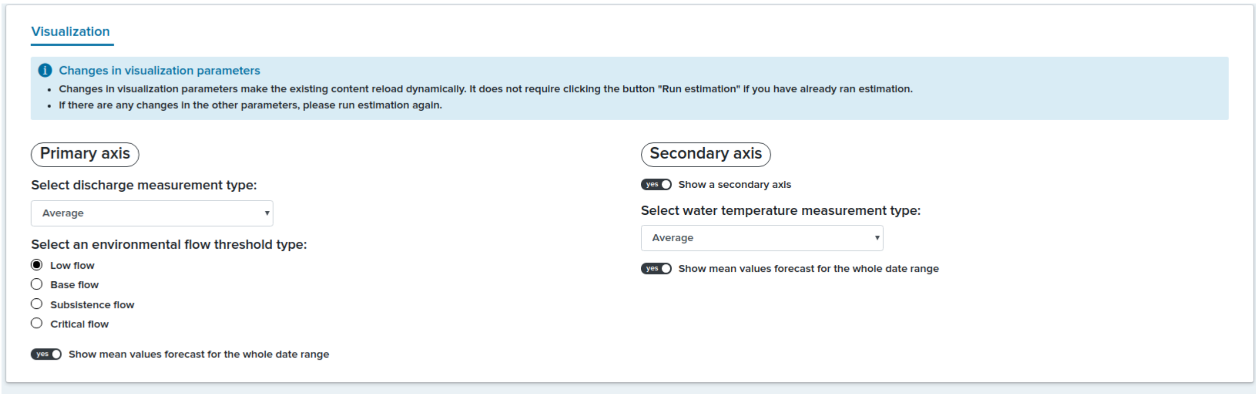


Figure 53. Visualisation configuration

The visualisation parameters include discharge measurement type (primary axis) and water temperature/water level (secondary axis, configured from the "Compound event" tab): minimum, average, maximum, all range; eflow threshold type: low flow (one of the Tennant or p% exceedance probability methods – refer to Chapter 2.5.4 for details), base flow, subsistence flow and critical flow. The determination of the subsistence, critical and base flows is carried out using the ratios approach (25%, 50%, 75% correspondingly) described in Chapter 2.5.5. The discharge values for their determination are taken using all the available data up to the right limit of the selected date range. The visibility of the additional time series (secondary axis) as well as the forecasted time series can also be configured from the "Visualisation" panel.

The estimation results are visualised with a compliance time series graph (Figure 54).

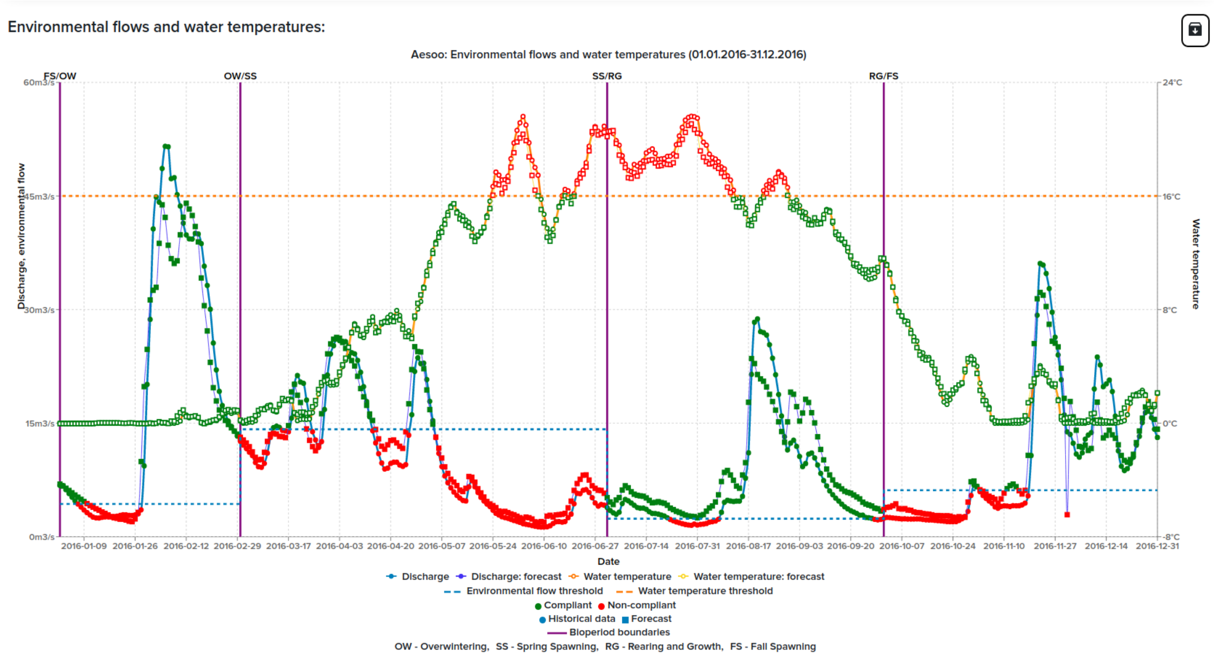


Figure 54. EFCES compliance graph

Compliance is estimated based on the threshold values (eflow thresholds for discharges and set by a user threshold for water temperatures and levels). The thresholds are visualised using dashed lines: blue for eflows, orange for water temperatures and pink for water levels. Red points on the graph demonstrate incompliance (water temperatures are above the corresponding threshold, and discharges and water levels are below); otherwise, points are green. The points of the time series of the primary axis are filled, whereas the points of the time series of the secondary axis are not filled. The historical time series lines are coloured with the same colours their thresholds have: blue for eflows, orange for water temperatures and pink for water levels. The historical time series lines are solid. The forecasted time series lines are thinner and have slight differences in colours compared to the corresponding historical time series lines. The historical observations have a circular shape, whereas forecasted points are squares. Purple vertical lines correspond to the bioperiod boundaries - the end of one and beginning of another bioperiod (OW - Overwintering, SS - Spring Spawning, RG - Rearing and Growth, FS - Fall Spawning). All the notation keys are presented in the legend at the bottom of the compliance graph. The visibility of the secondary axis time series and forecasts can be toggled from the visualisation panel (Figure 53).

Forecasting information is shown to users in the way presented in Figure 55.

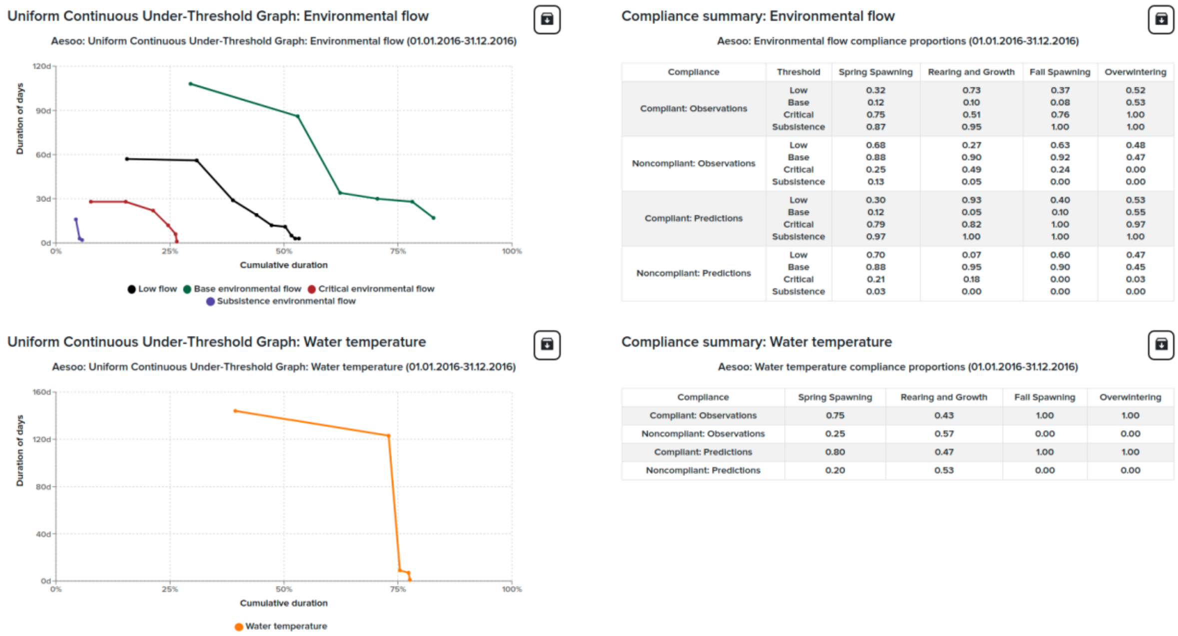
Figure 55. EFCES forecasting information

The forecasting information includes the algorithm chosen for the fitting of the forecasting model of particular time series – mean discharge (primary axis) and mean water temperature/water level (secondary axis). The choice of the regressor is made automatically according to the developed forecasting model described in Chapter 4.2.3. The choice of the type of the secondary axis time series is carried out on the “Compound event” tab. If a user toggles the “Show a secondary axis” off, the forecasting information of the secondary axis is not shown.

“Based on” field in the forecasting information corresponds to the type of time series used as input data for the model fitting. The type of the input time series is chosen from the “Forecasting” tab. If “Use other stations’ data” is toggled on, the forecasting is multivariate time series forecasting that uses available time series of the selected type from other gauging stations. The description of the data preparation and feature selection is provided in Chapter 4.2.2.

The service also builds UCUT graphs and calculates compliance proportions for both eflows and selected secondary axis hydrological measurements (Figure 56) visualising the cumulative duration (in %) of being under threshold per duration of days. UCUT graphs are built for average measurements of the low, base, critical, and subsistence eflows and water temperatures/levels for the selected date range for both historical and forecast time series. The compliance proportions tables show the proportions of compliance per bioperiod for the selected date range.

The service also allows downloading the estimation results by clicking the button on the top right angle of each of the graphs. The results can be saved in Excel (XLSX, data) or PNG format (image).



Compliance	Threshold	Spring Spawning	Rearing and Growth	Fall Spawning	Overwintering
Compliant: Observations	Low Base Critical Subsistence	0.32 0.12 0.75 0.87	0.73 0.10 0.51 0.95	0.37 0.08 0.76 1.00	0.52 0.53 1.00 1.00
Noncompliant: Observations	Low Base Critical Subsistence	0.68 0.88 0.25 0.13	0.27 0.90 0.49 0.05	0.63 0.92 0.24 0.00	0.48 0.47 0.00 0.00
Compliant: Predictions	Low Base Critical Subsistence	0.30 0.12 0.79 0.97	0.93 0.05 0.82 1.00	0.40 0.10 1.00 1.00	0.53 0.55 0.97 1.00
Noncompliant: Predictions	Low Base Critical Subsistence	0.70 0.88 0.21 0.03	0.07 0.95 0.18 0.00	0.60 0.90 0.00 0.00	0.47 0.45 0.03 0.00

Compliance	Spring Spawning	Rearing and Growth	Fall Spawning	Overwintering
Compliant: Observations	0.75	0.43	1.00	1.00
Noncompliant: Observations	0.25	0.57	0.00	0.00
Compliant: Predictions	0.80	0.47	1.00	1.00
Noncompliant: Predictions	0.20	0.53	0.00	0.00

Figure 56. EFCES UCUT graphs and compliance proportions tables

The further data such as spreadsheets with hydrological data, default values of the catchment area, FET types, etc. can be modified from the admin panel (Figure 57) accessible only to superusers of the system. The superuser users can also create accounts for other users to grant them access to the service.

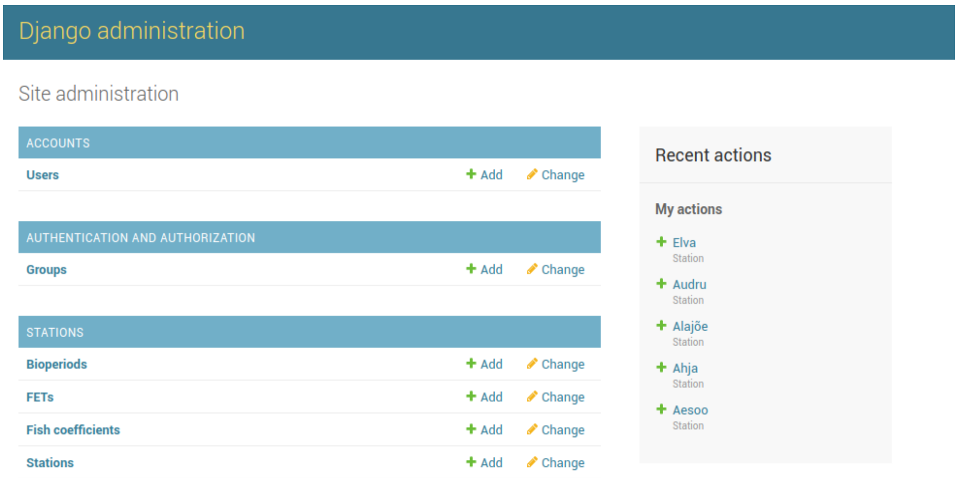


Figure 57. Administration panel of EFCES

The structure of the admin panel corresponds to the ORM of the service presented in Chapter 4.1.3. The description of service users is presented in Chapter 4.3.2.

4.3.2 Service Usage

The main workflow process diagram is presented in Figure 58.

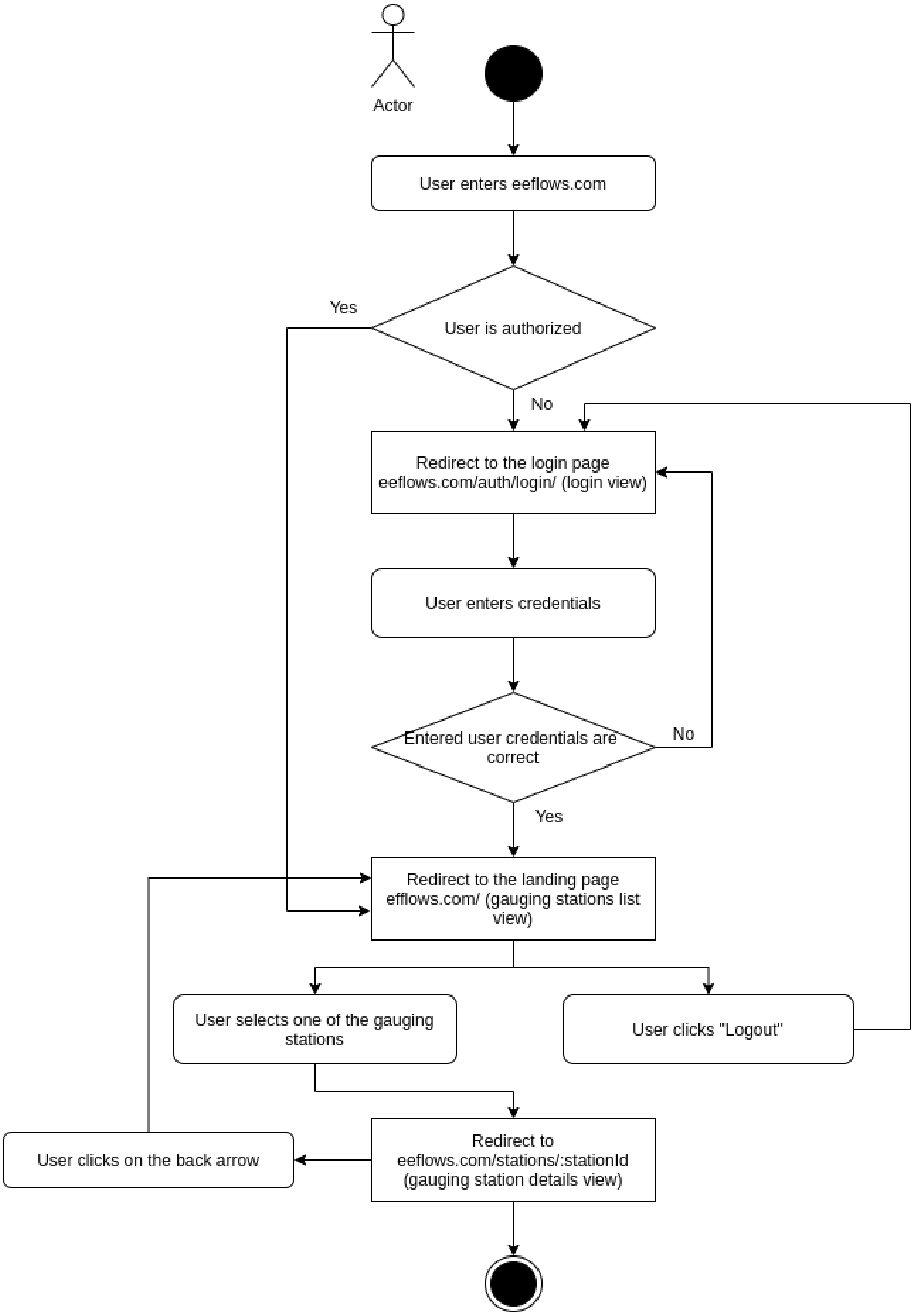


Figure 58. Main workflow process diagram of EFCES

The admin access process diagram is presented in Figure 59.

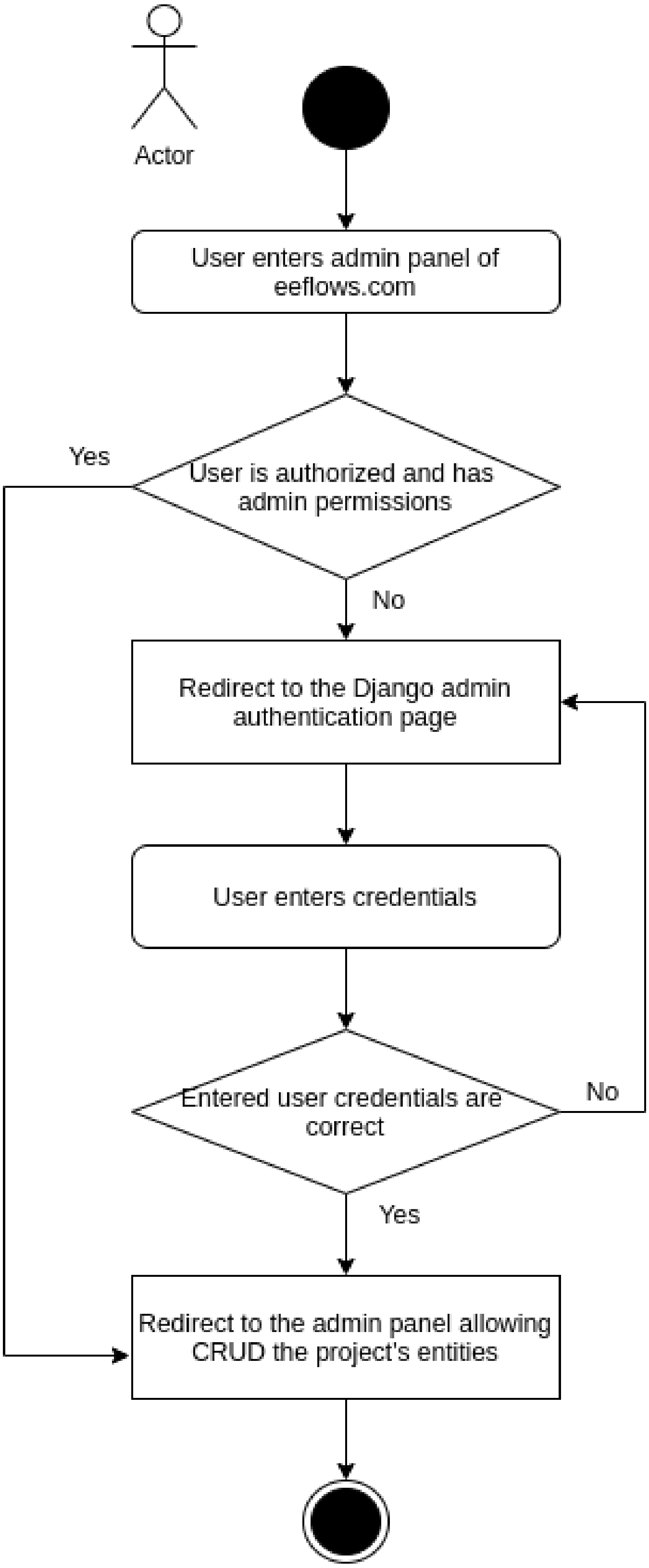


Figure 59. Administration access process diagram of EFCES

Users of the system can be classified as:

1. Guest users. Guest users do not have credentials to access the gauging stations' information. The only view accessible to them is the Login page.
2. Users (without admin access). Users without admin access or superuser rights have credentials to pass authentication. These users have access to all the pages of the publicly accessible frontend application and can read provided by the service. These users do not have access to the admin panel of the backend application allowing to modify shown by frontend information.
3. Superusers. Superusers can do everything that Users can do, but also have access to the admin panel (Figure 57), allowing them to modify data stored in the database.

4.3.3 Advisory Usage

The integration of the service to decision-making processes requires the target users to establish certain rules, e.g. the number of days a monitored river stays noncompliant, etc. The establishment of the rules is not in the scope of the project and left for the target end-users users.

The current version of the service allows for simulating various scenarios. With the historical data, it can help to model and evaluate the known historical cases, tune the parameters, and create the decision rules based on them. The forecasting capability allows modelling the inaccessible stations in case of any disruptions. Furthermore, additional users' training activities might be required.

With further research on the fish coefficient values, appropriate bioperiods boundaries, FETs, etc., the service model can be tuned to be a full-fledged advisory system interpreting the results and providing additional information to authorities for the decision-making.

4.4 Demonstration

The demonstration of the service involves the presentation of the service to the target users - Estonian Ministry of the Environment. As a result of the evaluation, the authorities were asked to leave feedback about the service and the suitability of its features for their professional needs.

The survey consisted of the questions presented in Appendix 5 - EFCES User Testing Feedback Questionnaire. The results of the evaluation are described in Chapter 5.

5 Evaluation and Discussion

In accordance with the “Results communication and application” step, the developed EnvRegTech artefact - EFCES has been evaluated by the potential stakeholders: Estonian Ministry of the Environment and other involved officials.

The feedback received from the Ministry of Environment and Environmental Agency of Estonia has shown that despite positively assessed UI, the system requires additional work on the UX. Since the service provides multiple features for different stakeholders and includes unfamiliar estimation methods and features (RAELFF in particular, since currently the environmental legislation does not involve the estimation of the eflow rates using this method), the resulting service turned out to be complex for perception. The testing process has also been complicated by the consequences of the incomplete test coverage of the system and data quality issues leading to unhandled errors.

The results of the UX/UI assessment of the system can be seen in Figure 60.

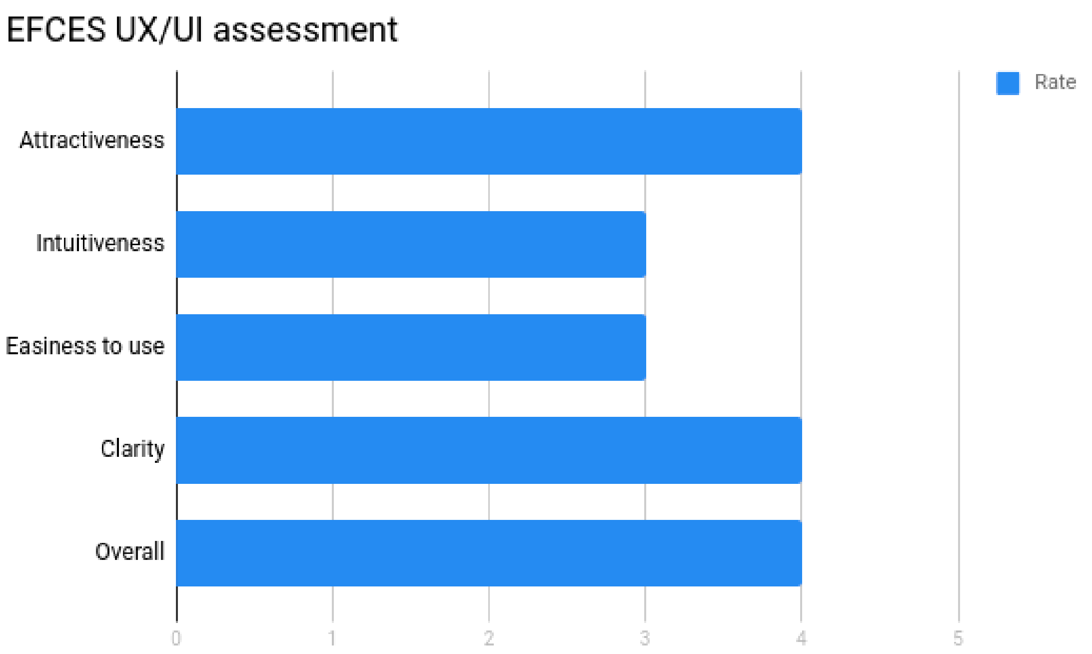


Figure 60. EFCES UX/UI assessment

In the feedback, the following potential uses of the service were mentioned: determination of the methodology to be used in environmental permits (scenario modelling with different configuration parameters) which also requires special knowledge and skills; monitoring of the compliance of holders of the environmental permits with the requirements stated in the permits.

Environmental permits of damming rivers require estimation of the flow rates in different spots allowing setting specific flow rates for each fish bioperiod that the permit holder will be obliged to obey. In order to define which flow rates are suitable to be set in the permit for the fish passing dams, additional data collection is required. Another suggestion addressing spatial attribute of the assessment was to add the map of the gauging stations and allow choosing the gauging stations from the map.

After results communication and application, the first iteration of the DSRM and EI cycles can be considered completed. Bridging the gaps of the eflow assessment with the creation of EnvRegTech artefact, the built EFCES service reveals the potential to be used for decision-making of governmental authorities as well as other stakeholders from the hydrological community. Repeating the accomplished research steps will help to refine the system to be more suitable and usable for the target users as well as make it visible for other researchers in the field to promote the methodology and opportunities for automation for the real-time eflow assessment.

The design of the service providing the automated assessment of the ecological status of Estonian rivers for the WFD compliance monitoring and reporting using OGD involved the web service design and development activities. The assessment of the river ecological status is carried out using various eflow assessment methods along with the RAELFF. The availability of the Estonian hydrological OGD allowed the creation of the web service for reporting and decision-making purposes supporting the objectives of WFD, environmental monitoring and reporting needs. The results of the assessment are visually interpreted with coloured indicators and summaries that can be exported for the WFD compliance monitoring and reporting purposes.

Comparing to the existing practices of eflow estimation in Estonia, the improvements can be seen in providing opportunities of the automated dynamic eflow assessment through an OGD-driven web-service, inclusion of compound event (water temperature/water level), forecasting, integration of RAELFF as a more profound, spatially scalable, fish-friendly alternative to average percentages of flow rates and other widely used but largely outdated hydrological methods. The RAELFF approach is expected to be a better grounded characterization of the river eflows ecological status, addressing the limitations of the currently used approaches. The approach is scalable and can be used for dynamic real-time compliance with environmental regulations estimation as well as scenario modelling and forecasting.

In addition to the automated eflow assessment constituting the compliance evaluation of rivers with WFD, the decision support is provided with the scenario modelling capabilities allowing configuring the parameters of estimation, build compliance graphs, UCUT graphs and compliance proportion tables that can be the objects of interest to environmental governmental authorities as well as other stakeholders. The specialists interested in flow rates (e.g. for environmental permits) can model scenarios with various hydrological time series and configuration parameters as well as forecast discharge, water temperature and water levels. Promoting the use of RAELFF, fish biologists can be interested in determining FETs and fish p coefficients of rivers to capture biological response along with the hydrological variation and test the determined coefficients to assess the river ecological status according to them. The reporting feature allows exporting of the results that can be potentially useful to water managers.

The developed within the scope of the thesis project and described in the paper EFCES can be considered only as a prototype. Since DSRM and EI are iterative, the further research activities involve repetition of the accomplished steps: extensive hydrological, fish and habitat data collection and further development activities towards improving UX and UI to sufficiently cover the professional necessities of target users as well as the improvement of their understanding of the suggested methodologies and the results interpreted by EFCES and potential transformation to real-time assessment with the alerting system. The forecasting model of the prototype is assumed to be refined to improve the precision of the forecasts as well as enhanced by the sensitivity analysis and uncertainty assessment. In order to deploy RAELFF for the regulatory eflow estimation, field research is required to determine FETs and necessary for the formula coefficients.

The implementation of the service has faced multiple difficulties, asides from the difficulties of the eflow assessment practices, resulting in the degradation of the performance and requiring optimization. For example, the methods using long-term (all available) data are not computationally efficient in real-time assessment because of the growing amount of memory allocated for more observations and more calculations to be made. At the same time, recent observations can have greater importance for the assessment, which should be further researched and addressed.

The quality of the spreadsheets with hydrological data was not sufficient for machine processing. Thus, there were spotted the traces of manual intervention and differences in formatting leading to reading errors as well as inaccuracies in the data (e.g. minimum values greater than maximum values, erroneous variables, etc.) resulting in improper distorted interpretation and affecting forecasts. With a large amount of missing data, forecasting modelling has been significantly restricted, potentially resulting in a decrease in the model precision.

Finally, since the service is the only system of its kind developed based on the research papers interpretation, this was harder to design an intuitive system understandable to all the users without additional training.

Sustainable water management has many challenges and requires technologically advanced approaches to address water deterioration issues and improve existing practices. Having a great economic potential, the adoption of digital technologies can be an object of interest for multiple stakeholders. Decision support and risk assessment enabled by intelligent digital solutions can provide an essential justification for major time- and resource-consuming decisions. Therefore, the contribution of the research results - design and implementation of the environmentally intelligent web-service - can be seen as an example of improved practices in SWM and innovative usage of OGD - technological handling of the uncertainties in the definition of monitoring measurements, assessment procedures and forecasts faced during WFD implementation. The project justifies the driving ability of OGD in the creation of innovative data-driven services providing decision support. It verifies the importance of the OGD quality to ensure appropriate machine processing as well as the data accessibility, compatibility and interoperability in order to fulfil scaling potential of such services to be used not only within the country but also abroad. The used framework of the automated eflow assessment as well as the developed service can be expanded to be used in the Baltic region and other countries.

6 Summary

Introduction of information communication technologies to improve the efficiency of public service delivery is essential for e-government and e-governance on the way of digital transformation, and the next step is to maximise the sustainable, social and economic benefits of the established infrastructure.

The government collects and produces a lot of data concealing operational knowledge and actionable information valuable to stakeholders of different sectors. Making the data open allows combining it with other sources of data, performing analysis and predictive modelling for getting the overview in different scales. The provision of open access to government data facilitates not only reporting and communicating making the government more transparent and accountable and ensuring participatory governance, but also the creation of new innovative services or enhancing of existing ones with decision support.

Decision support is vital for environmental information systems. The environmental systems are complex and highly dynamic. They involve multiple variables observed over time with different frequencies. The collected big data are beyond the manual processing and analysis and require automation. The results of the analysis can be used to make decisions supporting expert knowledge with the statistical evidence as well as reveal previously unknown issues. Predictive modelling can help to simulate various scenarios and forecast results to gain confidence while planning activities.

The environmental OGD are widely used to address climate change issues and environmental challenges requiring real-time monitoring and reporting as well as timely decision-making and governmental measures. Short terms, high risks and heavy responsibilities for the made decisions can be supported with the evidence and knowledge obtained from the data-driven services. And the scale of the evidence and knowledge to be used as the ground of the decisions is largely determined by the availability of the data and services allowing its processing and analysis.

Acknowledged strategical importance of the environmental data resulted into various measures addressing the environmental data collection, monitoring and reporting practices as well as its provision to public access engaging all the interested parties: citizens, academia, private sector, etc. to enrich the database and develop new value-added services and tools.

The design and implementation of such data-driven services have many prerequisites: a deep understanding of the domain, proper analysis, design of the tool, etc. Data quality is extremely important since machine processing is sensitive to formatting. The databases ensure the common formatting of the observations within one environmental information system, but with the multiple environmental information systems, processing and interpreting of the related data in different ways and places makes it harder to capture the overview and extract benefits of the automation and data availability as well as scale solutions. That is why the standardisation is important not only for data collection, storage and monitoring but also for reporting and service scalability.

The design of data-driven services requires addressing specific purposes and stakeholder's needs. Thus, the services can be used not only for decision-making but also for policy-making, which is complicated with the conflicting needs of affected parties. The governmental decisions require considering complex environmental, social and economic interests of the parties, which acknowledges the importance of the expert knowledge. Therefore, the data-driven services and tools are intended to assist expert decision-making.

The objective of the thesis project was to design and develop such a service facilitating monitoring, reporting and decision-making of the environmental governmental authorities on the way of the implementation of objectives set by the Water Framework Directive.

The implementation of the Water Framework Directive is complicated by the data collection, monitoring, determination of the assessment practices, reporting, etc. significantly slowed down without unified procedures of the achievement of the goals. The availability of a large number of eflow assessment practices rather introduces an additional source of uncertainty than improves the river ecological status assessment practices.

With the accessible hydrological data collected and stored in the standardised format, chosen assessment practices (even by the law of vital few), the developed eflow compliance evaluation service can be scaled to ensure the automated assessment of eflows at the pan-European level. The prerequisites for the achievement of this objective as well as implementation of the real-time eflow assessment can be defined while further research activities.

The developed service is not linked to the Estonian environmental portal which does not promote good practices of the environmental open data accessibility. With the sufficient extent of readiness of the service to be used for monitoring and reporting purposes, integration should be addressed.

The results provided by the data-driven service should be used for decision- and policy-making, and the rules for the advisory usage of the service should be established, taking into account various aspects and legislation purposes.

Biological rules are sophisticated to describe, but RAELFF approach addresses the gap providing a scalable, better grounded alternative to existing hydrological practices. However, the determination of fish ecological types of rivers and corresponding coefficients of the formula for the Estonian use case requires additional field research and experts collaboration.

After the determination of the fish ecological types, bioperiods and corresponding coefficients, the analysis of the river ecological status evolution can be carried out. The newly obtained statuses can be compared with the ones found with the previously existing methodology to figure out how different the initial expert estimations and data-driven estimations using different methods of assessment.

The environment is a complex system that operates as a whole, and the evaluation of isolated rivers without the context of other rivers especially under rapid environmental changes does not seem sufficient. The forecasting model developed within the scope of the thesis project using this assumption leaves the room for improvement. With the new data from other places available, the forecasting model should be scaled spatially.

Finally, further research activities can be devoted to the improvement of the UX/UI of the developed service addressing the needs of the Estonian experts. The design of the service was not trivial since it is supposed to be used by different target users in multiple ways, and the importance of the features might vary depending on the target user group which in result can seem too sophisticated. With the active participation and engagement of the stakeholders, additional interviewing and A/B testing activities, it will be possible to define the most optimal workflow.

In conclusion, the improvement and integration into decision- and policy-making processes of the eflow compliance estimation service most probably will not make the achievement of the

WFD Directive easier, but might help to set smart goals not postponing the achievement of the objectives for the near future but assessing time frames and directions of its implementation using statistical evidence in an automated data-driven way.

References

[1] Lonergr, “eGovernment & Digital Public Services,” *Shaping Europe's digital future - European Commission*, 27-Sep-2019. [Online]. Available: https://ec.europa.eu/digital-single-market/en/public-services-egovernment. [Accessed: 03-May-2020].

[2] R. Matheus, M. Janssen, and D. Maheshwari, “Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities,” *Government Information Quarterly*, 07-Feb-2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0740624X18300303. [Accessed: 03-May-2020].

[3] “Sustainable Development Goals Report,” United Nations, 2019. [Online]. Available: https://unstats.un.org/sdgs/report/2019/The-Sustainable-Development-Goals-Report-2019.pdf. [Accessed: 24-Nov-2019].

[4] CRED/UNISDR. The Human Cost of Weather Related Disasters, 1995–2015; CRED: Brussels, Belgium, 2015; pp. 5–16.

[5] UN-Water, “Disasters: UN-Water,” *UN*. [Online]. Available: https://www.unwater.org/water-facts/disasters/. [Accessed: 24-Apr-2020].

[6] L. Wilber, “How Open Data Can Lead to Smarter Water Management,” How Open Data Can Lead to Smarter Water Management. [Online]. Available: https://www.opendatasoft.com/blog/2017/04/26/smart-water-management-open-data. [Accessed: 24-Apr-2020].

[7] “Internet of Things: Guidelines for Sustainability,” World Economic Forum, 17-Jan-2018. [Online]. Available: https://www.weforum.org/whitepapers/internet-of-things-guidelines-for-sustainability. [Accessed: 24-Nov-2019].

[8] “The Internet of Things: Mapping the value beyond the hype. Executive summary,” McKinsey Global Institute, 2015.

[9] Z. Supalla, “The internet of things and the war against water,” TechTarget, 02-Oct-2018. [Online]. Available: https://internetofthingsagenda.techtarget.com/blog/IoT-Agenda/The-internet-of-things-and-the-war-against-water. [Accessed: 24-Nov-2019].

[10] “Open water data on the European Data Portal,” Open water data on the European Data Portal | Evropský Datový Portál. [Online]. Available:

https://www.europeandataportal.eu/cs/highlights/open-water-data-european-data-portal. [Accessed: 24-Apr-2020].

[11] N. L. Poff, J. D. Allan, M. B. Karr et al., “The natural flow regime: A paradigm for river conservation and restoration,” BioScience, vol. 47, no. 11, pp. 769-784, 1997.

[12] S. E. Bunn, A. H. Arthington, “Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity,” Environmental Management, vol. 30, pp. 492-507, 2002.

[13] European Parliament, Council of the European Union, “Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000: establishing a framework for Community action in the field of water policy.” 23-Oct-2000. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32000L0060. [Accessed: 24- Nov-2019].

[14] “The EU Water Framework Directive - integrated river basin management for Europe,” *European Commission*. [Online]. Available: https://ec.europa.eu/environment/water /water-framework/index_en.html. [Accessed: 24-Nov-2019].

[15] N. Voulvoulis, K. D. Arpon, and T. Giakoumis, “The EU Water Framework Directive: From great expectations to problems with implementation,” *Science of The Total Environment*, 13-Oct-2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S004896971632157X. [Accessed: 05-Dec-2019].

[16] European Commission, “Ecological flows in the implementation of the Water Framework Directive: guidance document no. 31.,” *Publications Office of the European Union*, 10-Feb-2016. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/b2369e0f-d154-11e5-a4b5-01aa75ed71a1/language-en. [Accessed: 24-Nov- 2019].

[17] P. Parasiewicz, P. Prus, K. Suska et al., “E = mc2 of Environmental Flows: A Conceptual Framework for Establishing a Fish-Biological Foundation for a Regionally Applicable Environmental Low-Flow Formula,” *Water*, vol. 10, Oct-2018. [Online]. Available: https://www.researchgate.net/publication/328463086_E_mc2_of_Environmental_Flows_A_Conceptual_Framework_for_Establishing_a_Fish-Biological_Foundation_for_a_Regionally_Applicable_Environmental_Low-Flow_Formula. [Accessed: 24-Nov-2019].

[18] R. E. Tharme, “A global perspective on environmental flow assessment: emerging trends in the development and application of environmental flow methodologies for rivers,” *River Research and Applications*, vol. 19, pp. 397-441, 2003.

[19] G. E. Petts, “Instream Flow Science For Sustainable River Management,” *Journal of the American Water Resources Association*, vol. 45, pp. 1071–1086, 2009.

[20] J. Mann, “Instream flow methodologies: an Evaluation of the Tennant Method for Higher Gradient Streams in the National Forest System Lands in the Western U.S.,” USA, 2006.

[21] “Surface water,” Ministry of the Environment, 05-Jun-2014. [Online]. Available: https://www.envir.ee/en/surface-water. [Accessed: 30-Nov-2019].

[22] “Fisheries,” *Ministry of the Environment*, 07-Nov-2019. [Online]. Available: https://www.envir.ee/en/fisheries. [Accessed: 30-Nov-2019].

[23] A. Reihan, E. Loigu, “Ökoloogilise miinimumvooluhulga arvutusmetoodika väljatöötamine,” Töövõtuleping 18-25/185, Tallinn, 2010.

[24] A. Reihan, E. Loigu, K. Reinsalu, “Sanitaarvooluhulkade arvutamine ja paisutustaseme mõõtmine,” Töövõtuleping 4/1.1/87, Tallinn, 2012.

[25] J. Frankenfield, “What You Should Know About RegTech,” *Investopedia*, 27-Apr-2019. [Online]. Available: https://www.investopedia.com/terms/r/regtech.asp. [Accessed: 24-Nov-2019].

[26] J. Rohde, “What is Environmental Intelligence?,” *IARPC*, 28-Feb-2017. [Online]. Available: https://www.iarpccollaborations.org/news/7704. [Accessed: 24-Nov-2019].

[27] J. T. Mathis, E. Osborne, S. Starkweather, “Collecting Environmental Intelligence in the New Arctic,” Arctic Program, 08-Jan-2018. [Online]. Available: https://arctic.noaa.gov/Report-Card/Report-Card-2017/ArtMID/7798/ArticleID/691/Collecting-Environmental-Intelligence-in-the-New-Arctic. [Accessed: 24-Nov-2019].

[28] “Hydrology,” Website of Estonian Environment Agency. [Online]. Available: https://www.keskkonnaagentuur.ee/en/hydrology. [Accessed: 29-Apr-2020].

[29] “Sustainability Insights. Environmental Intelligence: Unlocking opportunities at the nexus of Big Data and climate change,” Generation Investment Management LLP, Apr-2015. [Online]. Available: https://www.generationim.com/media/1458/generation-sustainability-insights-series-environmental-intelligence.pdf. [Accessed: 24-Nov-2019].

[30] “Environmental Intelligence,” *MDPI Encyclopedia*. [Online]. Available: https://encyclopedia.pub/397. [Accessed: 19-Apr-2020].

[31] “Open Government Data,” *OECD*. [Online]. Available: https://www.oecd.org/gov/digital-government/open-government-data.htm. [Accessed: 24-Apr-2020].

[32] “About,” *Estonian Open Government Data Portal*. [Online]. Available: https://opendata.riik.ee/en/about/. [Accessed: 24-Apr-2020].

[33] “Hydrological data,” *International Open Data Charter*. [Online]. Available: https://opendatacharter.net/agriculture-open-data-package/section-2-towards-open-data-

infrastructure-agriculture/natural-resources-earth-environment-data/hydrological-data/. [Accessed: 26-Apr-2020].

[34] “Ajaloolised vaatlusandmed,” *Riigi Ilmateenistus*. [Online]. Available: https://www.ilmateenistus.ee/siseveed/ajaloolised-vaatlusandmed/ . [Accessed: 24-Apr-2020].

[35] “Building Resilience,” International Water Management Institute (IWMI). [Online]. Available: http://www.iwmi.cgiar.org/what-we-do/building-resilience/. [Accessed: 24-Nov-2019].

[36] “Application of framework on 640 projects in the IoT database,” *World Economic Forum*. [Online]. Available: http://widgets.weforum.org/iot4d/. [Accessed: 24-Nov-2019].

[37] “Summary of the Comprehensive Environmental Response, Compensation, and Liability Act,” *EPA*, 15-Aug-2019. [Online]. Available: https://www.epa.gov/ laws-regulations/summary-comprehensive-environmental-response-compensation-and-liability-act. [Accessed: 24-Nov-2019].

[38] “Introduction to the EU Water Framework Directive,” *European Commission*. [Online]. Available: https://ec.europa.eu/environment/water/water-framework/info/intro_en.htm. [Accessed: 24-Nov-2019].

[39] M. K. Roach, B. Grundfelt, L. O. Höglund et al., “Chapter 18 - Environmental Legislation and Best Practice in the Emerging European Rare Earth Element Industry,” in “*Rare Earths Industry: Technological, Economic, and Environmental Implications*”, pp. 279-291, 2016.

[40] European Parliament, Council of the European Union, “Directive 2004/35/CE of the European Parliament and of the Council of 21 April 2004 on environmental liability with regard to the prevention and remedying of environmental damage,” 21-Apr-2004. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02004L0035-20190626&qid=1568193390794&from=EN. [Accessed: 24-Nov-2019].

[41] “Estonica,” *Estonica.org - Location*. [Online]. Available: http://www.estonica.org/en/Nature/Location_and_natural_conditions/Location/. [Accessed: 29-Apr-2020].

[42] “Estonica,” *Estonica.org - Topography*. [Online]. Available: http://www.estonica.org/en/Nature/Location_and_natural_conditions/Topography/. [Accessed: 29-Apr-2020].

[43] “Estonica,” *Estonica.org - Climate*. [Online]. Available: http://www.estonica.org/en/Nature/Location_and_natural_conditions/Climate/. [Accessed: 29-Apr-2020].

[44] “Estonica,” *Estonica.org - The hydrographic network*. [Online]. Available: http://www.estonica.org/en/Nature/Location_and_natural_conditions/The_hydrographic_network/ . [Accessed: 29-Apr-2020].

[45] “Fresh water resources in Estonia,” *Climatechangepost.com*. [Online]. Available: https://www.climatechangepost.com/estonia/fresh-water-resources/. [Accessed: 29-Apr-2020].

[46] A. Reihan, E. Loigu, “Vooluhulga mõõtmise juhend,” Töövõtuleping 4/1.1/87, Tallinn, 2012.

[47] B. D. Richter, J. V. Baumgartner, J. Powell et al., “A method for Assessing Hydrologic Alteration within ecosystems,” *Conservation Biology*, vol. 10, no. 4, pp. 1163-1174, 1996.

[48] M. A. Palmer, R. F. Ambrose, N. L. Poff, “Ecological theory and community restoration ecology,” *Restoration Ecology*, vol. 5, no. 4, pp. 291-300, 1997.

[49] D. A. Lytle, N. L. Poff, “Adaptation to natural flow regimes,” *Trends in Ecology and Evolution*, vol. 19, no. 2, pp. 94-100, 2004.

[50] J. T. Wootton, M. S. Parker, M. E. Power, “Effects of disturbance on river food webs,” *Science*, vol. 273, pp. 1558-1561, 1996.

[51] M. E. Power, A. Sun, G. Parker, “Hydraulic food-chain models,” *Bioscience*, vol. 45, no. 3, pp. 159-167, 1995.

[52] J. T. Puckridge, F. Sheldon, K. F. Walker et al., “Flow variability and the ecology of large rivers,” *Marine and Freshwater Research*, vol. 49, pp. 55-72, 1998.

[53] A. H. Arthington, E. B. Stuart, N. L. Poff et al., “The challenge of providing environmental flow rules to sustain river ecosystems,” *Ecological Applications*, vol. 16, no. 4, pp. 1311-1318, 2006.

[54] J. King, C. Brown, H. Sabet, “A Scenario-based Holistic Approach to Environmental Flow Assessments for Rivers,” *River Research and Applications*, 2003.

[55] M. C. Ackerman, A. J. D. Ferguson, “Environmental flows and the European Water Framework Directive,” *Freshwater Biology*, vol. 55, no. 1, pp. 32-48, Jan-2010. [Online]. Available: https://www.researchgate.net/publication/227778161_EFs_and_the_European_Water_Framework_Directive. [Accessed: 03-Apr-2020].

[56] “Understanding flows, some terminology explained,” VEWH. [Online]. Available: https://www.vewh.vic.gov.au/news-and-publications/stories/understanding-flows-some-terminology-explained. [Accessed: 03-Apr-2020].

[57] R. Hirji, R. Davis, "Environmental Flows in Water Resources Policies, Plans, and Projects," *The World Bank,* 2009. [Online]. Available: http://siteresources .worldbank.org/INTWAT/Resources/Env_Flows_Water_v1.pdf. [Accessed: 24-Nov-2019].

[58] T. Le Quesne, E. Kendy, D. Weston, "The Implementation Challenge: Taking stock of government policies to protect and restore environmental flows," *WWF*, 2010. [Online]. Available: http://d2ouvy59p0dg6k.cloudfront.net/downloads/the_implementation_challenge.pdf. [Accessed: 24-Nov-2019].

[59] A. H. Arthington, A. Bhaduri, S. E. Bunn et al., "The Brisbane Declaration and Global Action Agenda on Environmental Flows (2018)," Frontiers, 22-May-2018. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fenvs.2018.00045/full. [Accessed: 24-Nov-2019].

[60] "Transforming our world: the 2030 Agenda for Sustainable Development: Sustainable Development Knowledge Platform," *United Nations*, 25-Sep-2015. [Online]. Available: https://sustainabledevelopment.un.org/post2015/transformingour world. [Accessed: 24-Nov-2019].

[61] "Global environmental flow information system," *International Water Management Institute (IWMI)*, 16-May-2019. [Online]. Available: http://www.iwmi.cgiar.org/2018/06/global-environmental-flow-information-system/. [Accessed: 26-Nov-2019].

[62] V. Smakhtin, M. Anputhas, "An Assessment of Environmental Flow Requirements of Indian River Basins," 2006. [Online]. Available: http://www.iwmi.cgiar.org/Publications/IWMI_Research_Reports/PDF/PUB107/RR107.pdf. [Accessed: 26-Nov-2019].

[63] P. Parasiewicz, "Using MesoHABSIM to develop reference habitat template and ecological management scenarios," *River Research and Applications*, vol. 23, pp. 924-932, 2007.

[64] "Environmental Flow Calculators," *International Water Management Institute (IWMI)*. [Online]. Available: http://www.iwmi.cgiar.org/resources/models-and-software/ environmental-flow-calculators/. [Accessed: 26-Nov-2019].

[65] M. Moore, "Perceptions and Interpretations of Environmental Flows and Implications for Future Water Resource Management: A Survey Study." Master's Thesis, Department of Water and Environmental Studies, Linköping University, Linköping, Sweden, 2004.

[66] P. Parasiewicz, "Habitat time-series analysis to define flow-augmentation strategy for the Quinebaug River, Connecticut and Massachusetts, USA," River Research and Applications, 2008, vol.24, no.4, pp. 439–452.

[67] The science of instream flows: a review of the Texas Instream Flow Program. Washington, DC: National Academies Press, 2005.

[68] M. B. Bain, M. S. Meixler, “A target fish community to guide river restoration,” *River Research and Applications*, vol. 24, pp. 453-458, 2008.

[69] R. Ali, A. Kuriqi, S. Abubaker, “Hydrologic Alteration at the Upper and Middle Part of the Yangtze River, China: Towards Sustainable Water Resource Management Under Increasing Water Exploitation,” *Sustainability*, Sep-2019.

[70] K. Peffers , T. Tuunanen, M. A. Rothenberger et al., “A Design Science Research Methodology for Information Systems Research,” *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45-77, 2007.

[71] “Vaatlusvõrk,” Riigi Ilmateenistus. [Online]. Available: http://www.ilmateenistus.ee/ilmateenistus/vaatlusvork/#hydro. [Accessed: 30-Nov-2019].

[72] *Systematic list of Estonian Fishes*. [Online]. Available: http://bio.edu.ee/animals/Kalad/kalalist2.htm. [Accessed: 30-Nov-2019].

[73] *Estonian Fishes - general characterization*. [Online]. Available: http://bio.edu.ee/animals/Kalad/kalalist11.htm. [Accessed: 30-Nov-2019].

[74] “SciPy,” *SciPy.org - SciPy.org*. [Online]. Available: https://www.scipy.org/. [Accessed: 08-Mar-2020].

[75] “NumPy,” *NumPy*. [Online]. Available: https://numpy.org/. [Accessed: 30-Nov-2019].

[76] “Python Data Analysis Library,” *pandas*. [Online]. Available: https://pandas.pydata.org/. [Accessed: 30-Nov-2019].

[77] “About statsmodels,” *statsmodels*. [Online]. Available: https://www.statsmodels.org/stable/about.html. [Accessed: 08-Mar-2020].

[78] “Visualization with Python,” *Matplotlib*. [Online]. Available: https://matplotlib.org/. [Accessed: 08-Mar-2020].

[79] “Sklearn,” *scikit*. [Online]. Available: https://scikit-learn.org/stable/. [Accessed: 08-Mar-2020].

[80] “Thorgate Labs,” *Thorgate Labs*. [Online]. Available: http://thorgate.github.io/. [Accessed: 07-Mar-2020].

[81] “Empowering App Development for Developers,” *Docker*. [Online]. Available: https://www.docker.com/. [Accessed: 08-Mar-2020].

[82] “Django Software Foundation,” *Django*. [Online]. Available: https://www.djangoproject.com/foundation/. [Accessed: 08-Mar-2020].

[83] Node.js, “About,” *Node.js*. [Online]. Available: https://nodejs.org/en/about/. [Accessed: 08-Mar-2020].

[84] “About,” *PostgreSQL*. [Online]. Available: https://www.postgresql.org/about/. [Accessed: 08-Mar-2020].

[85] *Redis*. [Online]. Available: https://redis.io/. [Accessed: 08-Mar-2020].

[86] “Distributed Task Queue,” *Celery*. [Online]. Available: http://www.celeryproject.org/. [Accessed: 08-Mar-2020].

[87] R. Gour, “Working Structure of Django MTV Architecture,” Medium, 16-Apr-2019. [Online]. Available: https://towardsdatascience.com/working-structure-of-django-mtv-architecture-a741c8c64082. [Accessed: 08-Mar-2020].

[88] “Redux - A predictable state container for JavaScript apps.: Redux,” *A predictable state container for JavaScript apps*. [Online]. Available: https://redux.js.org/. [Accessed: 08-Mar-2020].

[89] “S3,” *Amazon*, 2002. [Online]. Available: https://aws.amazon.com/s3/features/nc=sn&loc=2. [Accessed: 08-Mar-2020].

[90] “SPSS Tutorials: Pearson Correlation,” *LibGuides*. [Online]. Available: https://libguides.library.kent.edu/SPSS/PearsonCorr. [Accessed: 09-Mar-2020].

[91] “sklearn.linear_model.LinearRegression,” *scikit*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html. [Accessed: 09-Mar-2020].

[92] “sklearn.linear_model.Ridge,” *scikit*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Ridge.html. [Accessed: 09-Mar-2020].

[93] “sklearn.linear_model.Lasso,” *scikit*. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html. [Accessed: 09-Mar-2020].

[94] “Coefficient of Determination (R Squared): Definition, Calculation,” *Statistics How To*. [Online]. Available: https://www.statisticshowto.com/probability-and-statistics/coefficient-of-determination-r-squared/. [Accessed: 09-Mar-2020].

[95] “React – A JavaScript library for building user interfaces,” – *A JavaScript library for building user interfaces*. [Online]. Available: https://reactjs.org/. [Accessed: 26-Apr-2020].

[96] “Water,” *Website of Estonian Environment Agency*. [Online]. Available: https://www.keskkonnaagentuur.ee/en/water-2. [Accessed: 03-May-2020].

[97] “Veepoliitika raamdirektiivi rakendamine,” *Keskkonnaministeerium*. [Online]. Available: https://www.envir.ee/et/eesmargid-tegevused/vesi/veemajanduskavad/veepoliitika-raamdirektiivi-rakendamine. [Accessed: 03-May-2020].

[98] “Veemajanduskavad,” *Keskkonnaministeerium*. [Online]. Available: https://www.envir.ee/et/veemajanduskavad. [Accessed: 03-May-2020].

[99] “River Basin Management plans,” *Ministry of the Environment*. [Online]. Available: https://www.envir.ee/en/river-basin-management-plans. [Accessed: 03-May-2020].

[100] Andmevara, “Water Act,” *Water Act – Riigi Teataja*. [Online]. Available: https://www.riigiteataja.ee/en/eli/527122019007/consolide. [Accessed: 03-May-2020].

[101] “Story Map Series,” *kaur.maps.arcgis.com*. [Online]. Available: https://kaur.maps.arcgis.com/apps/MapSeries/index.html?appid=fd27acd277084f2b97eee82891873c41. [Accessed: 03-May-2020].

[102] Webmedia, “Täpsustatud nõuded veekogu paisutamise, paisutamisega seotud keskkonnaseire, vee-elustiku kaitse, paisu, paisutuse likvideerimise ja veetaseme alandamise kohta ning ökoloogilise miinimumvooluhulga määramise metoodika,” *Riigi Teataja*. [Online]. Available: https://www.riigiteataja.ee/akt/124092014001. [Accessed: 03-May-2020].

[103] B. Ubaldi, “Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives,” *OECD instance*, 27-May-2013. [Online]. Available: https://www.oecd-ilibrary.org/governance/open-government-data_5k46bj4f03s7-en;jsessionid=exvDNru-yXpravDVAfi8vLoH.ip-10-240-5-116. [Accessed: 03-May-2020].

[104] M. Toots, K. Mcbride, T. Kalvet, and R. Krimmer, “Open Data as Enabler of Public Service Co-creation: Exploring the Drivers and Barriers,” *2017 Conference for E-Democracy and Open Government (CeDEM)*, 2017.

[105] K. Mcbride, M. Toots, T. Kalvet, and R. Krimmer, “Leader in e-Government, Laggard in Open Data: Exploring the Case of Estonia,” *Revue française dadministration publique*, vol. 167, no. 3, p. 613, 2018.

[106] “European Commission,” *INSPIRE*. [Online]. Available: https://inspire.ec.europa.eu/docs. [Accessed: 03-May-2020].

[107] *Environmental Information System*. [Online]. Available: https://www.eionet.europa.eu/gemet/en/concept/2853. [Accessed: 03-May-2020].

[108] REPORT FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS, Actions to Streamline Environmental Reporting COM/2017/0312 final. [Online]. Available: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52017DC0312. [Accessed: 03-May-2020].

[109] “Data,” *EIS*. [Online]. Available: https://www.eis-data.eu/eiss/. [Accessed: 03-May-2020].

[110] “Country Fiches: Estonia,” *EIS*. [Online]. Available: https://www.eis-data.eu/eiss/. [Accessed: 03-May-2020].

[111] “Good practice guidance Environmental Information Systems.” [Online]. Available: https://ec.europa.eu/environment/legal/reporting/pdf/SR3-Guidance-EIS_v10.pdf. [Accessed: 03-May-2020].

[112] “DIRECTIVE 2007/2/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE),” *Official Journal of the European Union*. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L:2007:108:FULL&from=EN. [Accessed: 03-May-2020].

Appendix 1 – General Estonian Gauging Stations Information

Station	Water body	Watershed (km^2)	Longitude	Latitude
Aesoo	Navesti jõgi	1046.3	25.0616666666666	58.5152777777777
Ahja	Ahja jõgi	899.7	27.1122222222222	58.2091666666667
Alajõe	Alajõgi	131.8	27.3924999996666	59.0308333266667
Arbavere	Loobu jõgi	202.9	25.9633333333333	59.4399999966667
Audru	Audru jõgi	334.3	24.3184138882116	58.4233166614834
Elva	Elva jõgi	232.2	26.4344444441111	58.2113888888889
Hüüru	Vääna jõgi	208.7	24.5347222218888	59.3799999933333
Kaansoo	Saarjõgi	184.1	25.2216666659999	58.5772222155556
Kääpa	Kääpa jõgi	267.7	26.8488888885555	58.7005555555556
Kasari	Kasari jõgi	2652.3	23.9969444441111	58.7263888822222
Kehra	Jägala jõgi	933.3	25.3402777774444	59.3441666633333
Keila	Keila jõgi	631.8	24.4347222218888	59.3086111111110
Kirumpää	Võhandu jõgi	582.4	26.9924999996666	57.8655555555556
Kloostrimetsa	Pirita jõgi	799.2	24.8796194440744	59.4663138876173
Konuvere	Vigala jõgi	597.9	24.3897222218888	58.8011111111110
Korela	Piusa jõgi	754.9	27.7261111104444	57.8852777744445
Laadi	Reiu jõgi	534.9	24.6463888885555	58.2674999933333
Lüganuse	Purtse jõgi	793.3	27.0388888885555	59.3836111077778
Luguse	Luguse oja	96.1	22.7133333333333	58.8105555555555
Narva karjääri	Mustajõgi	272.5	27.8573999998304	59.2671777712038
Narva linn	Narva jõgi	56951.7	28.1093333329094	59.3895277723252
Nurme	Sauga jõgi	533.5	24.4982638885555	58.4462749970905
Oore	Pärnu jõgi	5156.9	24.7674999993333	58.4630555555555
Pajupea	Leivajõgi	85.8	24.9688888882222	59.3808333266667
Pajusi	Põltsamaa jõgi	1018.4	25.9277777771111	58.7030555555556
Põhjaka 1	Esna jõgi	241.3	25.6672557	58.8919444
Põhjaka 2	Sargvere	4.1	25.674478	58.8888889

	peakraav			
Pudisoo	Pudisoo jõgi	125.7	25.5944444441111	59.5086111111111
Rannu-Jõesuu	Emajõgi	3408.8	26.1341666663333	58.3855555522222
Räpina	Võhandu jõgi	1110.4	27.4544444444444	58.0955555522222
Reola	Porijõgi	241.9	26.7419444441111	58.2733333266667
Riisa	Halliste jõgi	1885.9	24.9941666663333	58.4794444377778
Roostoja	Rannapungerja jõgi	335.5	27.1052777777777	59.0233333266667
Sämi	Kunda jõgi	408.2	26.5827777771111	59.3727777711111
Separa	Avijõgi	379.5	27.0366666663333	58.9661111111111
Taheva	Mustjõgi	1814.7	26.3491666663333	57.598333333
Tahkuse	Pärnu jõgi	2072.6	24.9155555555555	58.5183333266666
Tänassilma	Tänassilma jõgi	308.2	25.8219444437777	58.3947222188889
Tartu	Emajõgi	7857	26.7239224	58.38
Tarvastu	Tarvastu jõgi	88.2	25.8844444441111	58.2302777711111
Toila-Oru	Pühajõgi	215.4	27.5299999993333	59.4230555548888
Tõlliste	Väike-Emajõgi	1090.9	26.1324999993333	57.8508333333333
Tori	Prandi jõgi	283	25.4742002	58.7991667
Tõrva	Õhne jõgi	274.1	25.9211111104444	58.0036111111111
Tõrve	Pedja jõgi	740.9	26.3747222215555	58.6022222222222
Tudulinna	Tagajõgi	251.1	26.8524361112382	59.176313888095066
Türi-Alliku	Pärnu jõgi	591.4	25.4727777771111	58.8299999993334
Uue-Lõve	Lõve jõgi	135.6	22.8222222215555	58.3644444444444
Valgu	Velise jõgi	136.6	24.6449999996666	58.8177777711111
Vanaküla	Valgejõgi	390.2	25.7891666663333	59.4674999993333
Varangu	Selja jõgi	378.6	26.3516666666666	59.4722222215555
Vasknarva	Narva jõgi	53113.3	27.7402777774444	59.0008333333333
Vihterpalu	Vihterpalu jõgi	480.1	23.8663888888888	59.2519444444445
Vodja	Vodja jõgi	78.1	25.6422557	58.9388889

Appendix 2 – Hydrological Data Exploration and Analysis Script

```
import numpy as np
import pandas as pd # version 0.25.3
from pandas_profiling import ProfileReport
import matplotlib.pyplot as plt
%matplotlib inline

q_path = './rivers-q-mean.csv'
temp_path = './rivers-TW-mean.csv'
level_path = './rivers-WL-mean.csv'
# Load data
full_discharge_df = pd.read_csv(q_path, index_col=0)
full_temp_df = pd.read_csv(temp_path, index_col=0)
full_level_df = pd.read_csv(level_path, index_col=0)

def make_plots_and_stats_report(df, station_name1, station_name2, ts_type_name, unit):
  # Plot time series
  plt.figure(figsize=(20,10))
  plt.title(f"Mean  {ts_type_name}  time  series:  {station_name1}  and  {station_name2}  gauging
stations")
  plt.ylabel(unit)
  df[station_name1].plot(label=station_name1)
   df[station_name2].plot(label=station_name2)
   plt.legend()
  plt.savefig(f"{station_name1}_{station_name2}_{ts_type_name}_mean.png")

  # Generate HTML report with discharge statistics
  profile = ProfileReport(df)
  profile.to_file(f"{ts_type_name}_mean_report.html")

make_plots_and_stats_report(full_discharge_df, "Aesoo", "Narva linn", "discharge", "m^3/s")
make_plots_and_stats_report(full_temp_df, "Aesoo", "Narva linn", "water temperature", "°C")
make_plots_and_stats_report(full_level_df, "Aesoo", "Narva linn", "water level", "cm")
```

Appendix 3 – Feature Selection Script

```
from math import sin, cos, sqrt, atan2, radians
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import datetime
import numpy as np

meta_path = './Stations_Metadata.xlsx'
meta_df = pd.read_excel(meta_path, index_col=0)

# approximate radius of earth in km
R = 6373.0

# calculate distance matrix
stations = meta_df["Station"].values
distance_df = pd.DataFrame(columns=stations, index=stations)

for idx1, row1 in meta_df.iterrows():
  station1 = row1["Station"]
  lat1 = radians(row1["Latitude"])
  long1 = radians(row1["Longitude"])
  for idx2, row2 in meta_df.iterrows():
    if idx1 == idx2:
      distance_df.loc[station1, station1] = 0
    elif idx2 > idx1:
      station2 = row2["Station"]
      lat2 = radians(row2["Latitude"])
      long2 = radians(row2["Longitude"])
      dlong = long2 - long1
      dlat = lat2 - lat1
      a = sin(dlat / 2)**2 + cos(lat1) * cos(lat2) * sin(dlong / 2)**2
      c = 2 * atan2(sqrt(a), sqrt(1 - a))
      distance = R * c
      distance_df.loc[station1, station2] = distance
      distance_df.loc[station2, station1] = distance

# calculate normalized distance matrix
norm_distance_df = distance_df.copy()

row_max_values = list()
for idx, row in distance_df.iterrows():
  max_value = max(row.values)
  row_max_values.append(max_value)
longest_dist = max(set(row_max_values))

for idx, row in norm_distance_df.iterrows():
  for col in norm_distance_df.columns:
    norm_distance_df.loc[idx, col] = norm_distance_df.loc[idx, col] / longest_dist
norm_distance_df = norm_distance_df.astype(float)

plt.figure(figsize=(20,20))
plt.title('Estonian gauging stations normalized distance matrix')
sns.heatmap(norm_distance_df)
plt.savefig('norm-distance-heatmap.png')

# calculate correlation-distance matrices and make heatmaps
```

```
q_file_name = './rivers-q-mean.csv'
full_discharge_df = pd.read_csv(q_file_name, index_col=0)

temp_file_name = './rivers-TW-mean.csv'
full_temp_df = pd.read_csv(temp_file_name, index_col=0)

level_file_name = './rivers-WL-mean.csv'
full_level_df = pd.read_csv(level_file_name, index_col=0)

def make_corr_dist_heatmap(df, dist_df, ts_type):
  corr_matrix_df = df.corr()
  dist_corr_matrix_df = corr_matrix_df.copy()

  for idx, row in corr_matrix_df.iterrows():
    for col in corr_matrix_df.columns:
      if idx == col:
        dist_corr_matrix_df.loc[idx, col] = 1
      elif not np.isnan(dist_corr_matrix_df.loc[idx, col]):
        dist_corr_matrix_df.loc[idx, col] = (np.abs(corr_matrix_df.loc[idx, col]) + (1 -
dist_df.loc[idx, col])) / 2

  plt.figure(figsize=(20,20))
  plt.title(f'Estonian gauging stations mean {ts_type} correlation-distance matrix')
  sns.heatmap(dist_corr_matrix_df)
  plt.savefig(f'corr-dist-{ts_type}.png')

make_corr_dist_heatmap(full_discharge_df, norm_distance_df, "discharge")
make_corr_dist_heatmap(full_temp_df, norm_distance_df, "water temperature")
make_corr_dist_heatmap(full_level_df, norm_distance_df, "water level")
```

Appendix 4 – Forecasting Modelling Script

```
import datetime
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.linear_model import LinearRegression, Lasso, Ridge

def get_available_set_by_dates(whole_df, from_date, to_date, train,
                training_period_days=365):
  max_date = max(whole_df.index)
  min_date = min(whole_df.index)
  if train:
    from_train_dt = from_date - datetime.timedelta(days=training_period_days+1)
    to_train_dt = from_train_dt + datetime.timedelta(days=training_period_days)
  else:
    from_train_dt = from_date
    to_train_dt = to_date
  if min_date <= from_train_dt <= max_date and min_date <= to_train_dt <= max_date:
    filtered_dates = [dt for dt in whole_df.index.tolist() if from_train_dt <= dt <= to_train_dt]
    filtered_df = whole_df.loc[filtered_dates]
    return filtered_df
  return None

def get_multiple_columns_subset(filtered_df, max_missing_count):
  # calculate the number of NaN values in a column
  filtered_df_null = filtered_df.isnull().sum(axis=0)
  # find and drop columns with > max_missing_count NaNs
  cols_to_drop = [col for col in filtered_df_null.index if
          filtered_df_null[col] > max_missing_count]
  if len(cols_to_drop) > 0:
    df = filtered_df.drop(cols_to_drop, axis=1)
    # interpolate the ones that can be interpolated
    df = df.interpolate(axis=1)
    # drop columns if not possible to interpolate entirely
    df = df.dropna(axis=1)
  else:
    df = filtered_df.copy()
  if df.empty or len(df.columns) < 1:
    return None
  return df


def get_single_column_subset(filtered_df, max_missing_count):
  filtered_df_null = filtered_df.isnull().sum()
  if filtered_df_null > max_missing_count:
    return None
  if filtered_df_null > 0:
    df = filtered_df.interpolate()
    if df.isnull().sum() > 0:
      return None
  else:
    df = filtered_df.copy()
  return df
```

```python
def get_train_dfs(whole_input_df, whole_output_df, from_date, to_date,
         station_to_predict_name, is_multistations, is_same_variable,
         max_missing_count=18):
  if not is_multistations and is_same_variable:
    return None, None # not supported
  input_filtered_df = get_available_set_by_dates(whole_input_df, from_date,
                            to_date, True)
  output_filtered_df = get_available_set_by_dates(whole_output_df, from_date,
                            to_date, True)
  if input_filtered_df is None or output_filtered_df is None:
    return None, None
  if is_multistations:
    input_filtered_df = input_filtered_df.drop(station_to_predict_name, axis=1)
    input_df = get_multiple_columns_subset(input_filtered_df, max_missing_count)
  elif not same_variable:
    input_filtered_df = input_filtered_df[station_to_predict_name]
    input_df = get_single_column_subset(input_filtered_df, max_missing_count)
  output_filtered_df = output_filtered_df[station_to_predict_name]
  output_df = get_single_column_subset(output_filtered_df, max_missing_count)
  if input_df is None or output_df is None:
    return None, None
  return input_df, output_df

def get_test_df(whole_input_df, from_date, to_date, station_to_predict_name,
        is_multistations, is_same_variable, max_missing_count=18):
  if not is_multistations and is_same_variable:
    return None # not supported
  input_filtered_df = get_available_set_by_dates(whole_input_df, from_date,
                            to_date, False)
  if input_filtered_df is None:
    return None
  if is_multistations:
    input_filtered_df = input_filtered_df.drop(station_to_predict_name, axis=1)
    input_df = get_multiple_columns_subset(input_filtered_df, max_missing_count)
  elif not is_same_variable:
    input_filtered_df = input_filtered_df[station_to_predict_name]
    input_df = get_single_column_subset(input_filtered_df, max_missing_count)
  if input_df is None:
    return None
  return input_df

def get_selected_features(X_train, X_test, X, station_to_predict):
  if len(X_train.columns) <= 10:
    return X_train, X_test, X_train.columns
  dist_file_name = './DistanceMatrix.xlsx'
  dist_df = pd.read_excel(dist_file_name, sheet_name="Normalized distances", index_col=0)
  corr_matrix_df = X.corr()
  dist_corr_matrix_df = corr_matrix_df.copy()
  for idx, row in corr_matrix_df.iterrows():
    for col in corr_matrix_df.columns:
      if idx == col:
        dist_corr_matrix_df.loc[idx, col] = 1
      elif not np.isnan(dist_corr_matrix_df.loc[idx, col]):
        dist_corr_matrix_df.loc[idx, col] = (np.abs(corr_matrix_df.loc[idx, col]) + (
            1 - dist_df.loc[idx, col])) / 2
  weights = dist_corr_matrix_df[station_to_predict]
  sorted_weights_idx = [idx for idx in weights.index if idx in X_train.columns]
  sorted_weights = weights.loc[sorted_weights_idx].sort_values(ascending=False)
```

```
  best_stations_10 = list(sorted_weights[:10].index)
  selected_X_train = X_train[best_stations_10]
  selected_X_test = X_test[best_stations_10]
  dependent_stations = best_stations_10
  return selected_X_train, selected_X_test, dependent_stations

def get_models():
  models = dict()
  models['lasso'] = Lasso()
  models['lr'] = LinearRegression()
  models['ridge'] = Ridge()
  return models

def get_sets_for_model(X_train_whole, X_test_whole, y_train, X_train_selected, X_test_selected,
model_key, is_multistations):
  if is_multistations and model_key != 'lasso':
    X_train = X_train_selected
    X_test = X_test_selected
  else:
    X_train = X_train_whole
    X_test = X_test_whole
  if is_multistations:
    X_train_values = X_train.values
    X_test_values = X_test.values
  else:
    X_train_values = X_train.values.reshape(-1, 1)
    X_test_values = X_test.values.reshape(-1, 1)
  y_train_values = y_train.values.reshape(-1, 1)
  if is_multistations:
    X_train_cols = X_train.columns
  else:
    X_train_cols = None
  return X_train_values, y_train_values, X_test_values, X_train_cols

def get_dependent_stations_for_model(X_train_cols, dependent_stations_selected, model_key, model,
is_multistations):
  if is_multistations:
    if model_key == 'lasso':
      return [col for idx, col in enumerate(X_train_cols) if model.coef_[idx] != 0.0]
    else:
      return dependent_stations_selected
  return []

def get_R2_for_model(y_train, y_pred):
  df_train_results = pd.DataFrame()
  df_train_results['ground_truth'] = y_train.ravel()
  df_train_results['predictions'] = y_pred
  correlation_matrix = np.corrcoef(df_train_results['predictions'],
df_train_results['ground_truth'])
  correlation_xy = correlation_matrix[0, 1]
  R2 = correlation_xy ** 2
  return R2

def get_prediction(X_train_whole, y_train, X_test_whole, X, is_multistations,
station_to_predict):
  models = get_models()
  R2_dict = dict()
  trained_models_dict = dict()
  dependent_stations_dict = dict()
  # do feature selection for linear and ridge regression
  if is_multistations:
```

```
    X_train_selected, X_test_selected, dependent_stations_selected =
get_selected_features(X_train_whole, X_test_whole, X, station_to_predict)
  for key, local_model in models.items():
    print(key)
    if is_multistations:
      X_train_values, y_train_values, X_test_values, X_train_cols =
get_sets_for_model(X_train_whole, X_test_whole, y_train, X_train_selected, X_test_selected, key,
is_multistations)
    else:
      X_train_values, y_train_values, X_test_values, X_train_cols =
get_sets_for_model(X_train_whole, X_test_whole, y_train, None, None, key, is_multistations)
    local_model.fit(X_train_values, y_train_values)
    if is_multistations:
      dependent_stations_dict[key] = get_dependent_stations_for_model(X_train_cols,
dependent_stations_selected, key, local_model, is_multistations)
    else:
      dependent_stations_dict[key] = get_dependent_stations_for_model(X_train_cols, None, key,
local_model, is_multistations)

    print(dependent_stations_dict[key])
    y_pred = local_model.predict(X_train_values)
    R2_dict[key] = get_R2_for_model(y_train_values, y_pred)
    print(R2_dict[key])
    trained_models_dict[key] = local_model
    if R2_dict[key] >= 0.95:
      y_test = local_model.predict(X_test_values)
      return y_test.ravel(), R2_dict[key], key, dependent_stations_dict[key]
  best_model_keys = [key for key, value in R2_dict.items() if value == max(R2_dict.values())]
  best_key = best_model_keys[0]
  print(best_key)
  if is_multistations:
    X_train_values, y_train_values, X_test_values, X_train_cols =
get_sets_for_model(X_train_whole, X_test_whole, y_train, X_train_selected, X_test_selected,
best_key, is_multistations)
  else:
    X_train_values, y_train_values, X_test_values, X_train_cols =
get_sets_for_model(X_train_whole, X_test_whole, y_train, None, None, best_key, is_multistations)
  y_test = trained_models_dict[best_key].predict(X_test_values)
  return y_test.ravel(), R2_dict[best_key], best_key, dependent_stations_dict[best_key]

def get_algorithm_by_key(key):
  if key == 'lr':
    return 'Linear regression'
  if key == 'lasso':
    return 'Lasso regression'
  if key == 'ridge':
    return 'Ridge regression'
  return key

file_name = './rivers-q-mean.csv'
full_discharge_df = pd.read_csv(file_name, index_col=0, parse_dates=True)

file_name = './rivers-TW-mean.csv'
full_temp_df = pd.read_csv(file_name, index_col=0, parse_dates=True)

file_name = './rivers-WL-mean.csv'
full_level_df = pd.read_csv(file_name, index_col=0, parse_dates=True)

def run_forecasting_model_and_plot_result(predicted_station_name,
                     from_time_dt, to_time_dt,
                     whole_input_df,
                     whole_output_df,
                     multistations,
```

```
                    same_variable,
                    ts_type,
                    forecast_details,
                    units,
                    years):
  test_date_range = [dt for dt in whole_output_df.index if from_time_dt <= dt <= to_time_dt]
  ground_truth = whole_output_df.loc[test_date_range][predicted_station_name]
  X_train, y_train = get_train_dfs(whole_input_df, whole_output_df, from_time_dt, to_time_dt,
predicted_station_name, multistations, same_variable)
  X_test = get_test_df(whole_input_df, from_time_dt, to_time_dt, predicted_station_name,
multistations, same_variable)
  if not multistations and same_variable:
    print('Not implemented')
  elif X_train is None or y_train is None or X_test is None:
    print('Not available')
  else:
    if multistations:
      all_cols = list(set(list(X_train.columns) + list(X_test.columns)))
      cols_to_drop = [col for col in all_cols if not col in X_train or not col in X_test]

      if len(cols_to_drop) > 0:
        train_cols_to_drop = [col for col in X_train.columns if col in cols_to_drop]
        test_cols_to_drop = [col for col in X_test.columns if col in cols_to_drop]
        if len(train_cols_to_drop) > 0:
          X_train = X_train.drop(train_cols_to_drop, axis=1)
        if len(test_cols_to_drop) > 0:
          X_test = X_test.drop(test_cols_to_drop, axis=1)
    X_train_whole = X_train
    X_test_whole = X_test
    X = whole_input_df
    is_multistations = multistations
    station_to_predict = predicted_station_name
    models = get_models()
    R2_dict = dict()
    trained_models_dict = dict()
    dependent_stations_dict = dict()
    legend_labels = list()
    if is_multistations:
      X_train_selected, X_test_selected, dependent_stations_selected =
get_selected_features(X_train_whole, X_test_whole, X, station_to_predict)
    plt.figure(figsize=(20, 10))
    df_results = pd.DataFrame()
    df_results['ground_truth'] = ground_truth
    plt.plot(df_results['ground_truth'])
    legend_labels.append("actual")
    for key, local_model in models.items():
      if is_multistations:
        X_train_values, y_train_values, X_test_values, X_train_cols =
get_sets_for_model(X_train_whole, X_test_whole, y_train, X_train_selected, X_test_selected, key,
is_multistations)
      else:
        X_train_values, y_train_values, X_test_values, X_train_cols =
get_sets_for_model(X_train_whole, X_test_whole, y_train, None, None, key, is_multistations)
      local_model.fit(X_train_values, y_train_values)

      if is_multistations:
        dependent_stations_dict[key] = get_dependent_stations_for_model(X_train_cols,
dependent_stations_selected, key, local_model, is_multistations)
      else:
        dependent_stations_dict[key] = get_dependent_stations_for_model(X_train_cols, None, key,
local_model, is_multistations)
```

```
        y_pred = local_model.predict(X_train_values)
        R2_dict[key] = get_R2_for_model(y_train_values, y_pred)
        trained_models_dict[key] = local_model
        y_test = local_model.predict(X_test_values)
        y_test_df = pd.Series(y_test.ravel(), index=test_date_range)
        df_results[key] = y_test_df
        plt.plot(df_results[key])
        legend_labels.append(f"predicted: {get_algorithm_by_key(key)}")

    plt.title(f'{predicted_station_name}: mean {ts_type} {forecast_details} in {years}')
    plt.xlabel(xlabel='date',fontsize=16)
    plt.ylabel(ylabel=f'{ts_type} [{units}]',fontsize=16)
    plt.legend(legend_labels)
    plt.savefig(f'{predicted_station_name} - mean {ts_type} {forecast_details} in {years}.png')

# make and plot forecasts
from_time = '2015-01-01'
to_time = '2016-12-31'
from_time_dt = datetime.datetime.strptime(from_time, "%Y-%m-%d")
to_time_dt = datetime.datetime.strptime(to_time, "%Y-%m-%d")

run_forecasting_model_and_plot_result('Aesoo', from_time_dt, to_time_dt, full_discharge_df,
full_discharge_df, True, True, "discharge", "predicted by discharge from other stations", 'm3/s',
"2015-2016")
run_forecasting_model_and_plot_result('Narva linn', from_time_dt, to_time_dt, full_discharge_df,
full_discharge_df, True, True, "discharge", "predicted by discharge from other stations", 'm3/s',
"2015-2016")
run_forecasting_model_and_plot_result('Aesoo', from_time_dt, to_time_dt, full_temp_df,
full_temp_df, True, True, "water temperature", "predicted by water temperature from other
stations", '°C', "2015-2016")
run_forecasting_model_and_plot_result('Narva linn', from_time_dt, to_time_dt, full_temp_df,
full_temp_df, True, True, "water temperature", "predicted by water temperature from other
stations", '°C', "2015-2016")
run_forecasting_model_and_plot_result('Aesoo', from_time_dt, to_time_dt, full_level_df,
full_level_df, True, True, "water level", "predicted by water level from other stations", 'cm',
"2015-2016")
run_forecasting_model_and_plot_result('Narva linn', from_time_dt, to_time_dt, full_level_df,
full_level_df, True, True, "water level", "predicted by water level from other stations", 'cm',
"2015-2016")
run_forecasting_model_and_plot_result('Aesoo', from_time_dt, to_time_dt, full_level_df,
full_discharge_df, True, False, "discharge", "predicted by water level from other stations",
'm3/s', "2015-2016")
run_forecasting_model_and_plot_result('Aesoo', from_time_dt, to_time_dt, full_discharge_df,
full_temp_df, True, False, "water temperature", "predicted by discharge from other stations",
'°C', "2015-2016")
```

Appendix 5 – EFCES User Testing Feedback Questionnaire

Interviewee information

1. Name, email
2. Position, a short description of responsibilities

Reporting

1. Please rate from 1 (low) to 5 (high) the importance of the reporting feature to you as a user
2. What information do you find most important for reporting and exporting?
3. Would you add any other reporting and exporting features (e.g. other types of graphs, additional information for exporting)? Please specify.
4. What would you change in existing reporting and exporting features (e.g. another way of interpretation, other formats, etc.)? Please specify.

Environmental low flow

1. If you are aware of the environmental low flow approach, would you use the feature for environmental flows estimation reporting?
2. Would you like to change any functionalities related to the "Environmental low flow" section to make it more convenient/clear/usable to users? Please specify.

Compound event

1. Please rate from 1 (low) to 5 (high) the importance of the ability to add additional hydrological time series to you as a user.
2. Would you like to change any functionalities related to the additional time series to make it more convenient/clear/usable to users? Please specify.

Forecasting

1. Please rate from 1 (low) to 5 (high) the importance of the forecasting feature to you as a user
2. Making forecasts, what would you add/change (in forecasting summary, interpretation, etc.) to make the results more useful to you?

General feedback

1. What was your first impression when entering the web service?
2. Was there anything unexpected in the functionality of the web service? Please specify.
3. What is the most important feature for you in the web service (the feature that you would use most)?
4. What do you like most about web service?
5. What do you like least about web service?
6. If you could add/change anything about the web service, what would it be?
7. If you needed to start using the web service in your daily practices, please rate from 1 (poor) to 5 (excellent) the suitability of web service features to achieve your goals?
8. Rate the overall web service user experience from 1 (poor) to 5 (excellent)
9. Any further suggestions, comments