water

MDPI

Article

Exploring the Cyber-Physical Threat Landscape of Water Systems: A Socio-Technical Modelling Approach

Georgios Moraitis *, Georgia-Konstantina Sakki, George Karavokiros, Dionysios Nikolopoulos, Ioannis Tsoukalas, Panagiotis Kossieris and Christos Makropoulos

Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Heroon Polytechneiou 5, 15780 Athens, Greece; sakkigeorgina@gmail.com (G.-K.S.); gkaravo@itia.ntua.gr (G.K.); nikolopoulosdio@central.ntua.gr (D.N.); itsoukal@mail.ntua.gr (I.T.); pkossier@mail.ntua.gr (P.K.); cmakro@mail.ntua.gr (C.M.)

* Correspondence: georgemoraitis@central.ntua.gr

Abstract: The identification and assessment of the cyber-physical-threat landscape that surrounds water systems in the digital era is governed by complex socio-technical dynamics and uncertainties that exceed the boundaries of traditional risk assessment. This work provides a remedy for those challenges by incorporating socio-technical modelling to account for the adaptive balance between goal-driven behaviours and available skills of adversaries, exploitable vulnerabilities of assets and utility's security posture, as well as an uncertainty-aware multi-scenario analysis to assess the risk level of any utility against cyber-physical threats. The proposed risk assessment framework, underpinned by a dedicated modelling chain, deploys a modular sequence of processes for (a) the estimation of vulnerability-induced probabilities and attack characteristics of the threat landscape under a spectrum of adversaries, (b) its formulation to a representative set of stochastically generated threat scenarios, (c) the combined cyber-physical stress-testing of the system against the generated scenarios and (d) the inference of the system's risk level at system and asset level. The proposed framework is demonstrated by exploring different configurations of a synthetic utility case study that investigate the effects and efficiency that different cyber-security practices and design traits can have over the modification of the risk level of the utility at various dimensions.

Keywords: risk assessment; cyber-physical attacks; agent-based model; sociotechnical system; probability of attack; cybersecurity; uncertainty; urban water systems; resilience; decision support

Citation: Moraitis, G.; Sakki, G.-K.; Karavokiros, G.; Nikolopoulos, D.; Tsoukalas, I.; Kossieris, P.; Makropoulos, C. Exploring the Cyber-Physical Threat Landscape of Water Systems: A Socio-Technical Modelling Approach. *Water* **2023**, *15*, 1687. https://doi.org/10.3390/w15091687

Academic Editor: Chin H Wu

Received: 25 March 2023
Revised: 13 April 2023
Accepted: 16 April 2023
Published: 26 April 2023

1. Introduction

The current and future landscape of the water industry is reshaped by the transformational power of digitalisation and the proliferation of IoT technologies. The emergent modus operandi of urban water systems builds on an integrated cyber-physical architecture and forthcoming information schemes [1]. These incorporate novel informatics and computer technologies, such as Big Data, IoT and Cloud computing, as well as innovations from the field of information and communication technologies (ICT), such as optical fibres and 5G cellular connectivity [2], along with the hydraulic infrastructures. The operations of such cyber-physical systems (CPS) rely on a continuous information, computation and action loop between the associated cyber and physical layer devices that synthesise them [3]. In an urban water CPS, this loop employs sensors (e.g., pressure or water quality sensors) for on-site data collection, a wireless and/or wired data transmission network to pass information, and a set of computational decision systems to define actions and remotely control the operation of actuators in the field (e.g., valve settings) to regulate the system. A conceptual representation for the monitoring, transmission and actuation loop in a water CPS can be seen in Figure 1.

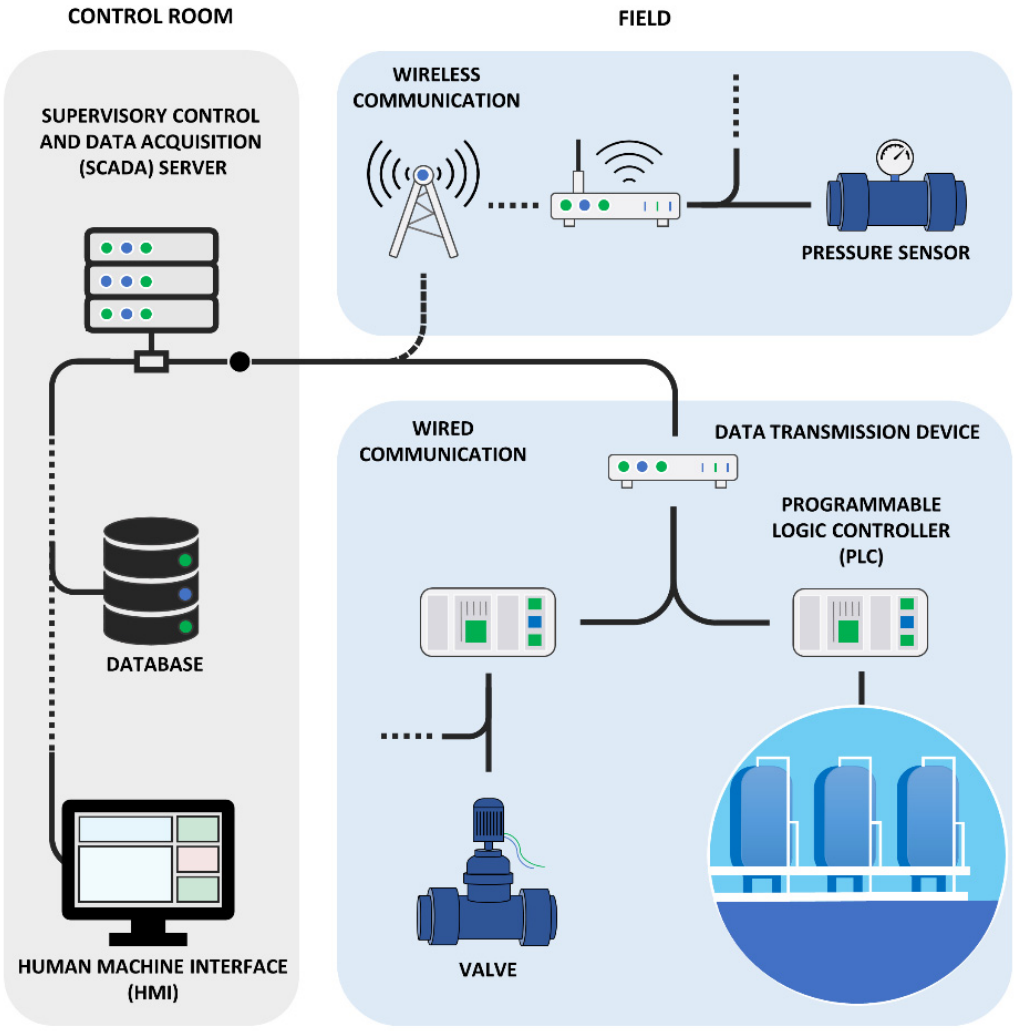

Figure 1. Conceptual representation of a water CPS (monitoring/observation, transmission, computation and actuation components).

Besides technical opportunities and merits associated with the transition of conventional systems into cyber-physical ones [4,5], the systems are also challenged by previously unknown and complex threats, stemming from the cyber domain [6]. In particular, the transition into CPS inevitably expands the previously available attack surface of the systems, as they inherit the vulnerabilities of the cyber layer and allow the implementation of the tactics, techniques and procedures (TTPs) of cyber-attackers [7]. In this emerging threat scene, the physical barriers are subordinated, and the potential attackers can leverage cyber access gates to infringe upon the system, gain control and eventually harm the system remotely. This has been the case for several recent attempts against the water sector world-wide, successful or otherwise. Examples of such incidents include the remote manipulation of a dosing system in Oldsmar's water treatment plant, that led to harmful concentration levels of sodium hydroxide [8], the 60-day PLC manipulations in the anonymised "Kemuri Water Company" that tampered with various asset settings and caused service disruptions [9], as well as a series of allegedly state-affiliated cyber-physical attacks against Israel's water infrastructure [10]. Such incidents have proven the capacity of the new attack vectors to cause deviations from the legal/regulatory levels of quality, quantity, continuity and pressure levels in supply systems [11], threaten the wellbeing of communities and possibly lead to severe reputational and/or financial damage for water utilities. Thus, being a force to be reckoned with, cyber-physical threats introduce a new challenge for the resilience of the sector and push the boundaries of traditional risk assessment.

This new era for critical infrastructure resilience is acknowledged by the EU and reflected in the provisions of both Directive (EU) 2022/2557 [12] on the resilience of critical entities and NIS2 Directive (EU) 2022/2555 [13]↓ on cybersecurity, published on December 2022. Under the new EU legislative umbrella in force, member states and critical infrastructure stakeholders are asked to break the silos between cyber and physical risk management and follow a new path towards combined cyber-physical resilience of their critical systems and services. This transition, however, requires not only a shift of mentality by the engaged parties, but most importantly, a rethinking of the risk assessment practices and techniques that will help to overcome the challenges and limitations posed by the complexity and obscure nature of the emerging threat landscape.

Typically, risk is expressed as a function of (a) the probability of an event's occurrence and (b) its potential consequences to the system [14–17]. To derive and assign probabilities, conventional practices rely either on statistical analysis of past recorded events or on pure expert judgment procedures. The latter are known to be susceptible to biases and misleading heuristics [18,19], while the event databases used for statistical analysis are often incomplete, non-representative or debased [20,21]. However, the success or avoidance of a cyber-physical threat is the result of a balance between, inter alia, the available skills, resources and motives of the attackers, and the exploitable vulnerabilities and applied cyber-security of the utility. Thus, even in the utopic case of "perfect event records", the implied statistical assumption of a stationary relationship between events and their generating mechanisms is dismissed by the continuous evolution of cyber-physical attacks and the expanding adversarial ecosystem [22]. This is also true when we consider the continuous structural and operational changes in utilities, as the level of exploitable vulnerabilities in a system is mainly a function of its structure and adopted strategies [23–25]. As a result, conventional approaches fall short on deriving vulnerability-induced attack probabilities and rendering the dynamic socio-technical system that constitutes the threat landscape of a cyber-physical water system. Although recent reviews over the state of hydroinformatics and cyber-security in water systems [1,26] suggest a pool of innovative models to realistically simulate cyber-physical water systems under attack (see e.g., [27–29]), the challenge of characterising the emerging threat landscape remains open, and propagates downstream in the risk assessment process when experts seek to parameterise threat events for analysis.

Critical infrastructure legislation, standards and contemporary cyber-physical risk assessment frameworks propose an event-based analysis of threats, following the scenario approach [14,17,23,30]. To formulate the scenarios for analysis, potential threats are assigned event-like characteristics, e.g., targeted asset, duration, etc., which in the case of cyber-physical threats, are strongly related to the attackers profile (e.g., skills, motives, resources) [22]. In the absence of sufficient knowledge that describes the interactions between social and technological aspects of threats, their assessment either omits them [31] or relies on expert elicitation [32,33]. Expert-centred techniques, however, are not bound by subjectivity, but rather produce arbitrary parameterisation of the risk based on the available knowledge that comes to mind easily, erroneously used as an objective measure [34,35]. False perceptions over security [36] coupled with cognitive and motivational biases in the process can seriously distort the risk analysis inputs [37] and affect the quality of the analysis. Consequently, they may lead to a fragmented narration of the cyber-physical threat landscape that surrounds urban water infrastructures, and thus to its incomplete assessment, as they are unable to characterise the underlying reasoning and motivation. In addition, it is recognised that the scenarios' capabilities are further narrowed in practice, due to deductive and deterministic approaches [38]. Under the norms of deterministic analysis, risk assessors attempt to formulate threats into a single, "most representative" scenario that usually adheres to a "fail-safe" mentality [39]. It is worth noting, however, that the prominence of deterministic analysis can be further attributed to technical limitations, as more inductive and uncertainty-aware scenario analyses are computationally expensive and time-consuming. Nevertheless, deterministic scenario assessments are unable to account for uncertainties, both epistemic and aleatory, which are known to have great

influence over the risk assessment results [38,40–42]. All of the abovementioned factors narrow the capacity, if not the validity, of evidence-based resilience planning of the sector and indicate the need for enhancements on both theoretical and technical grounds in the domain of risk assessment under the emerging cyber-physical threats.

This paper provides a remedy for those challenges with a standardised and transferable modelling framework for resilience assessment, operationalised though the PROCRUSTES platform [43], which, inter alia incorporates socio-technical modelling to account for the goal-driven behaviours that drive threat actors and an uncertainty-aware scenario analysis to render the cyber-physical threat landscape of any utility. To introduce the proposed modelling chain in a nutshell, we apply a modular sequence of processes for (a) the estimation of vulnerability-induced probabilities and characteristics for the entire attack surface of the system, (b) its formulation to a representative set of threat scenarios that embed stochastic variability, (c) the cyber-physical stress-testing of the generated scenarios and (d) the inference of the system's risk level by combining the probabilities and consequences at system and asset level. The individual steps of the approach and their relevant technical component in the modelling chain are explained in the sections below. To showcase the framework and platform's capabilities to render and analyse the threat landscape of real-world utilities, we provide a proof-of-concept application in a semi-hypothetical water utility, which explores the adoption of different cyber-security practices and their effect on the utility's risk exposure. To the best of the authors' knowledge, the proposed framework is the first to provide a consolidated analytical (expert-independent) process to identify and assess the exposure of any utility against the ensemble of potential threats and adversaries that constitute the cyber-physical threat landscape, while accounting for the socio-technical drivers and inherent uncertainties. Thus, this work provides a holistic re-thinking of cyber-physical resilience assessment practices and can be leveraged to increase understanding over the current and future threat landscape as well as to design evidence-based strategies and resilience planning for water CPS.

2. Materials and Methods

A highly desired trait that compliments traditional risk assessments is that of resilience [16]; this can be defined as "the degree of continued performance of the system under disturbance" [44], whether natural or man-made, accidental or intentional. To ensure resilience, water utilities are asked to determine the nature and extent of their risks by identifying and analysing relevant threats and vulnerabilities that could lead to an incident and evaluate the potential impacts in the provision of their essential service [12]. Over the years, utilities have become familiar with traditional physical threats and natural hazards, developed a deeper understanding of their mechanisms, and established methodologies and tools to efficiently assess them, although gaps still exist [45]. However, water utilities are unaccustomed to the emerging cyber-physical threat landscape, with limited understanding over the obscure nature and complex mechanisms associated with it, while traditional approaches and tools lack the capacity to efficiently address them. Thus, for the assessment of cyber-physical threats, a seamless modelling chain is needed that can achieve the following:

- Typify and account for goal-driven behaviours of threat actors based on key attributes (i.e., skills, resources, motives, etc.) to allow a better understanding over the nature of the cyber-physical threats and their potential characteristics;
- Explore the effects of design aspects and operational security measures (applied or planned) that allow the identification and characterisation of vulnerability-induced opportunities against key components of the system, such as sensors, actuators, PLC or the SCADA;
- Render the threat landscape through alternative threat scenarios and parameters to account for the different attack paths, targeted assets and the profile of potential adversaries, and derive a bird's eye-view of the system's risk level and weak points;

- Simulate water systems as integrated cyber-physical systems to capture the dynamic interplay between cyber and physical elements under attack and quantify the potential consequences to the water services;
- Encompass the stochasticity of model inputs (e.g., water demands) and/or parameters to encapsulate the uncertainty throughout the assessment process.

This paper builds on the cyber-physical risk assessment framework for water systems presented by [23,46], and enhances it, inter alia, with the addition of sociotechnical analysis over vulnerability-induced attacks and their probabilities, the incorporation of uncertainty-aware simulation disciplines, the deployment of parallel computing to overcome the computational constraints in multi-scenario analysis and the examination of the entire threat landscape at a system-wide analysis level. Being that malevolent actions are subject to complex and obscure behavioural mechanisms, this study utilises the well-established capabilities of agent-based modelling (ABM) approaches to simulate cyber-physical threats against any system. Specifically, we build on the capabilities of the CPRISK-ABM approach [47] to reproduce a digital sociotechnical environment and its subsequent norms and rules, that hosts two individual and conflicting parts, i.e., the utility and the attackers. The ABM simulates the dynamic balance between its elements to render attack characteristic against key assets and the associated probabilities of attack. Subsequently, and by adhering to the guidelines of scenario planning, the CPRISK-ABM is linked downstream to a scenario generator. The scenario generator maps the ABM derived threat distribution and associated characteristics into an ensemble of scenarios that represent the entire cyber-physical threat landscape against the system and its multiple attack paths. To account for aleatory and epistemic uncertainties in the process, the scenario generator further builds on the capabilities of the *anySim* R package [48] and random sampling methods to enable a Monte-Carlo-type scenario analysis. This leads to an uncertainty-aware examination of multiple scenarios under a spectrum of threat parameters and synthetic, yet realistic, states of the system driven by stochastically generated demands. To streamline the quantitative risk assessment, the modelling chain encompasses a dedicated version of Risknought [27] downstream, as a CPS modelling platform which allows stress-testing of the system and analysis of its performance against the threat scenarios. To overcome the limitations of high computational loads associated with the multi-scenario approach, this study innovates by building on a multi-core analysis architecture to allow the full exploitation of the available processing power though parallel scenario analysis. Ultimately, the aforementioned stepwise approach, integrated in the PROCRUSTES platform [43], quantifies the total risk level of the system against the emergent cyber-physical threat landscape and allows its resilience assessment against multiple metrics of performance, to identify weak spots and prioritise risk reduction strategies. Figure 2 illustrates a schematic of the proposed framework and the connection among the elements of the modelling chain.

2.1. The CPRISK-ABM: Socio-Technical Analysis of Cyber-Physical Attacks

Cyber-physical attacks against urban water systems constitute a complex dynamic ecosystem that includes technical, technological and social subcomponents that continuously interact with each other. Technical and technological components include infrastructures, assets and operational norms, while the social component includes ethics, motivates, social behaviours and security culture. To capture both ingredients and their cascading effects across the cyber-physical attacks, those systems are examined as sociotechnical systems, with the social factor being associated to both the attackers and the stakeholders of the utility. This is achieved through the first tool in the modelling chain, the CPRISK-ABM, which aims to simulate the cyber-physical threat landscape of the utility and derive probabilities of attacks and their relevant characteristics. We utilised the capabilities of agent-based approaches to integrate complex adaptive system theory and distributed artificial intelligence to model the system as a collection of autonomous decision-making entities that act according to a set of rules [49]. To outline the purpose, variables, design concepts

and details of the CPRISK-ABM, this paper provides a modelling description following the Overview, Design concepts and Details (ODD) protocol [50–52].

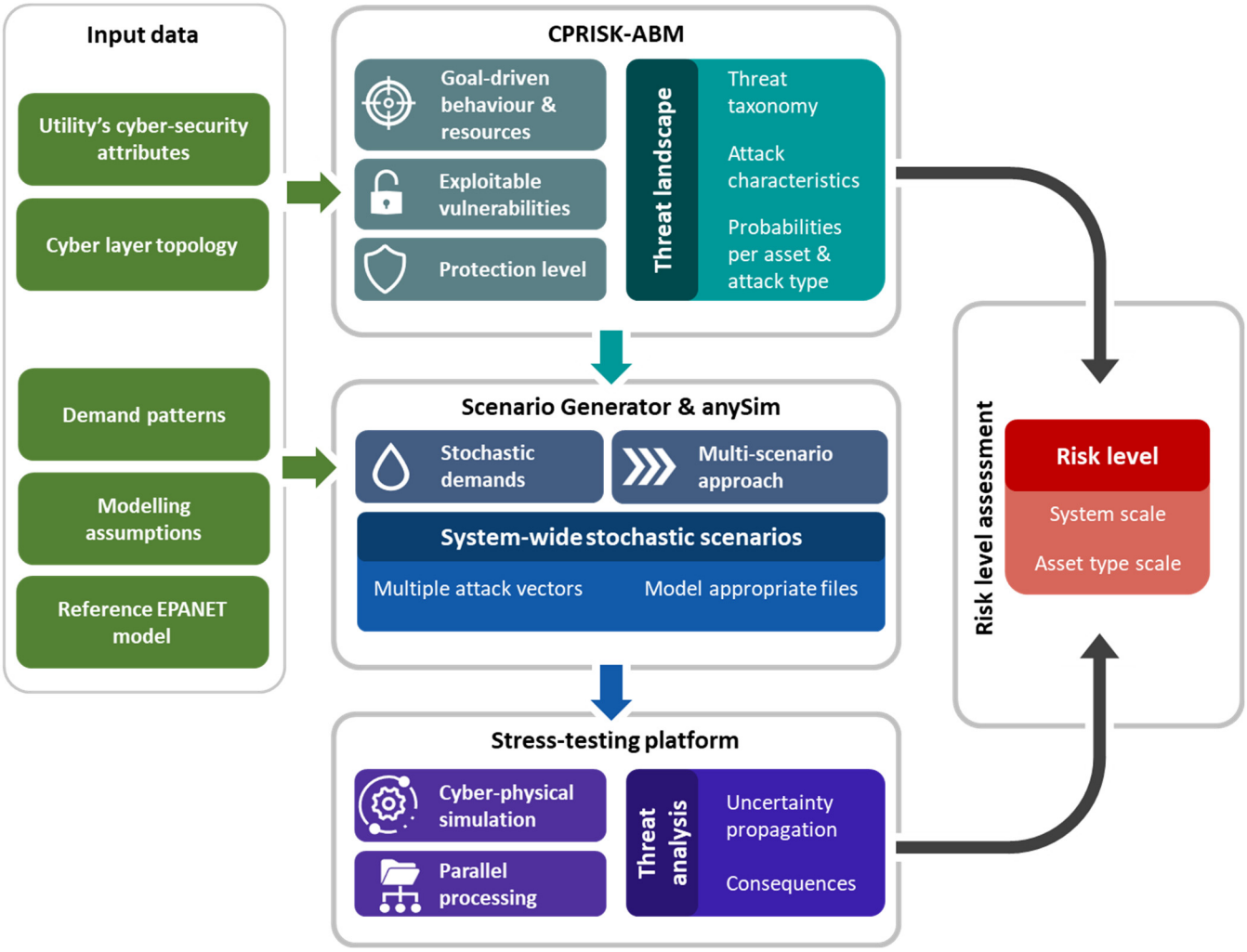

Figure 2. Visual overview of the core attributes and interactions of the modelling chain under the enhanced cyber-physical risk assessment scheme.

2.1.1. Purpose and Patterns

The CPRISK-ABM was developed using the Mesa framework, an Apache2 licensed agent-based modelling framework in Python [53], to simulate the dynamic and bidirectional interaction between the preferences, skills and motives of adversaries, and the opportunities and exploitable vulnerability conditions of the system. This allows the derivation of probabilities for vulnerability-induced attacks under a spectrum of attackers and system conditions that define the threat landscape, as well as the relevant attack characteristics at the asset and system level.

The model builds on the cyber security concept of a red team/blue team, which resembles a prey–predator relationship between the elements. Under this conceptualisation, the red team is comprised of adversaries that seek to compromise the system's integrity using different TTPs according to their profile, and the blue team, i.e., the utility, defends its assets and services by applying protection measures and using cyber-security practices that reduce the exploitable vulnerabilities in the system.

Although each adversary uses its own specific tools and attack patterns ad hoc, NIST [54], ENISA [22,55] and Verizon [56–58] have identified a series of trends and patterns that correlate various characteristics, including skills, motives and attack paths to groups of attacks and attackers, based on observations from real-world incidents. Moreover, the

patterns of exploitable vulnerabilities within SCADA-centred OT systems revolves mainly around asset-related weaknesses, insufficient authentication, remote access, wireless data communication paths, lack of cybersecurity mentality/training of employees and delayed patch/updates [9,59,60]. The CPRISK-ABM incorporates those real-world observations into its design, parameterisation and evaluation in a pattern-oriented manner [61,62], typically associated to lower sensitivity in parameters' uncertainty [62,63]. It should be noted that, the aggregation of specific behavioural patterns across different occasions and forms of actions has the capacity to cancel the influence of factors that are unique to each event and produce a more valid characterisation of the underlying general behavioural disposition, according to the theory of planned behaviour [64]. This is especially true in the case of behaviours with partial volitional control, which rely jointly on the motivation and the ability of individuals to achieve the goal in question—such as an attack or the protection of the system.

The CPRISK-ABM is designed as a generic model that expands the fundamental behavioural principles and patterns into a rich representation of real-world occasions for each individual system, adapting to the relevant design, security posture and population served. This allows the derivation of utility-specific probabilities of attack and characterisation of the characteristics across the entire cyber-physical threat landscape, that can be utilised downstream in a quantitative risk assessment. Overall, the tool is designed to streamline the cyber-physical risk assessment framework with the classic approach to risk in infrastructure planning and natural hazards, by providing the required data to derive the probabilities of relevant events—and model the complex socio-technical mechanisms for both successful events and near misses, under the existing or to-be-implemented protection level of each utility.

2.1.2. Design Concepts

Following the ODD terminology, the design concepts of the CPRISK-ABM can be summarised as follows:

- *Emergence*: The dynamic interaction at micro-level between the individual behaviour and the motives of adversaries, the security posture and capacity of the utility and the design-embedded vulnerability of the assets that synthesise the key cyber layer of the system leads to the emergence of real-world patterns and the estimation of the threat landscape (macro-level). The simulation output includes both failed and successful attacks of various types against different assets, which are subsequently mapped into probabilities.
- *Adaptation*: From the adversaries' side, the adaptation mechanism relies on their ability to gain intelligence and insight regarding the system from successful penetrations and attacks, thus increasing their capacity to perform new attacks against more protected targets. However, this ability is reversed when they fail to detect or compromise the utility assets, as their attack paths and TTPs are revealed to the utility personnel. The attackers may also temporarily increase their capacity and attempt attacks beyond their actual know-how by accessing the Darknet to gain intelligence over vulnerabilities and obtain advanced tools (e.g., anonymisation, encryption, malwares, etc.). On the other hand, the system also adapts from the deployment of protection actions from the utility's side, which increases the required cost of attacks against protected assets. The blue team is also able to deploy distributed honeypot network technology, which deceives attackers to interact with fake, isolated and monitored non-critical nodes of the network, thus gaining intelligence over adversaries. This translates to a reduction in the attacker's available resources as the selected tools, attack paths and TTPs—if not their identity—have been compromised and are more easily detectable from the utility.
- *Objectives*: The adaptive traits of both teams relate to the objectives of each. The attackers seek to locate vulnerable nodes of the system and compromise desirable targets according to their skills and motives, and/or increase their resources to perform attacks against more protected/critical targets. The utility's objective is to provide

additional protection over assets and deflect compromises and/or deployment of attacks against them, according to its available resources and security posture.

- *Sensing*: Attackers are aware of their contact with a potential target on the grid and initiate their attack procedure to gain access over the system node. This behavioural rule relies on their self-efficacy trait, i.e., the individuals' belief that they can successfully execute the required actions to achieve the goal of gaining access and exploiting the asset. The second sensing process in the CPRISK-ABM is that attackers can form a perception over the assets they have gained access to and decide according to their motives and intentions to a) deploy an attack or b) gather intelligence and stay stealth. This is related to the outcome expectancy of individuals, i.e., the perception that an attack against the specific asset will or will not result to an outcome that satisfies the goal-driven motives of the adversary. Lastly, attackers sense if the target at hand is already explored by another attacker. In this case, the model follows a strict "first come, first served" rule, to mimic the logical choice of real-world adversaries of avoiding competition with like-minded individuals during an attack campaign and the risk of exposure.
- *Interaction*: The interactions of agents occur on the grid cells that represent the feasible model interaction space. The attackers interact with the asset targets that fall within the same grid cell and initiate the attack procedure. In case the target is already locked by another attacker, the interaction with the second attacker results in the activation of the first-come, first-served rule. Regarding the interaction of the utility agents with the assets, this also occurs at the cell level and initiates the protection procedure. In the case of distributed honeypot deployment, the attackers and utility directly interact over the honeypot nodes on the grid, which allows the utility to deceive attackers to engage in an attack procedure against network-imitating nodes and collect counterintelligence.
- *Stochasticity*: The model employs the principles of the random walk stochastic process, which in turn, allows agents to move along the feasible interaction space. This allows the simulation of the search process of attackers to identify available IPs, open ports, transmitted data, etc., and spot potential vulnerable assets. In addition, the intrinsic agent parameters are assigned pseudo-randomly during the initialisation step using random sampling techniques and uniform distribution within value ranges associated with the different types of agents. The initial position of all moving agents is assigned randomly within the feasible model space.

2.1.3. Entities and State Variables

As mentioned, the CPRISK-ABM builds on the interaction of independent agents in a red team/blue team logic, that also defines the agents' collectives in a high-level taxonomy. The red team is comprised of autonomous moving agents aliased *Attackers*, which seek to detect, compromise and then attack cyber nodes of the system. The blue team is comprised of both the utility agents that cover the entire feasible space, aliased *Utility*, and seek to protect the system and the technical layer of the system through independent moving agents, aliased *Targets*, comprised of the individual cyber nodes of the system.

The *Attackers'* collective is further anatomised to groups of individuals that share similar state variables and behavioural traits. Within the CPRISK-ABM, this lower-level taxonomy accounts for the characteristics of resources, skills and intelligence as well as the underlying motivation of the forthcoming attack, based on the real-world patterns mentioned earlier. Attackers are assigned different motivations, namely no motivation, financial and reputational, and associated behavioural rules, that reflect their goal-driven decision-making mechanisms. The four distinct *Attackers'* groups and their main descriptions are as follows:

- *Amateurs*—Their resources and expertise are limited, and they are mainly driven by the thrill of the hacking process, with little motivation over the end result.
- *Experts*—They are experienced and respectably skilled attackers, predominantly driven by financial motivations.

- *Highly skilled*—Their sharper skills and increased resources allow them to attack key components of the system, in view of their increased financial or reputational motivations.
- *Nation state affiliated*—They are highly skilled individuals with increased access to intelligence and resources, able to deploy sophisticated TTPs on demand, and only undertake cyberattacks in support of a country's strategic objectives.

The Attackers' collective is characterised by the available "resources" that aggregate the state parameters of technical skills, available tools and resources, and intelligence for each individual. The individual parameter values are assigned pseudo-randomly within a range that reflects each group's distinct intrinsic characteristics and the divergence from other groups, e.g., nation state affiliated *Attackers* are assigned higher values than experts (see Figure 3).

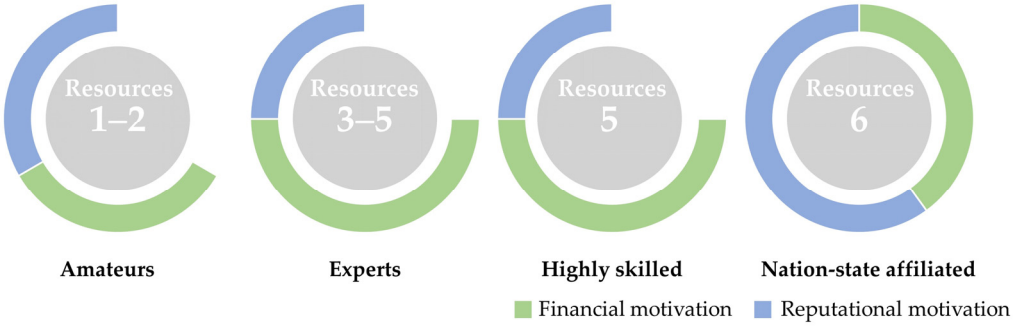

Figure 3. Motivation and resources level associated with the taxonomy of Attackers.

The *Utility* agents are considered as one team with uniform characteristics, governed by the security posture of the utility, and a common behavioural trait—to offer protection to the system and the individual assets. The protection of the technical system depends on design traits, system maintenance and the security posture of the utility, which can either mitigate or heighten its weaknesses. The CPRISK-ABM defines the level of protection that the *Utility* agents offer though a semi-quantitative questionnaire that scores the performance of the system against major relevant factors found in the patterns, i.e., implemented level of authentication and encryption processes, deployment of remote access outside the utility's intranet, use of wireless connections, cybersecurity mentality/training of employees, patch/updates frequency, etc. The questionnaire can be found in Table A4 of Appendix A.

As cyber-physical systems rely on a sensing, computation and remote action loop, the key assets and potential targets are the devices that serve that loop. Thus, the *Target* collective is further analysed to represent the critical cyber nodes of each step of the loop and their in-built and system-based vulnerabilities as found from the relevant patterns. The four groups that comprise the *Targets* and their main description are as follows:

- *Sensors*—are edge devices with limited memory and computational capacity that result in lower or negligible protection protocols in respect to other IT systems, and thus, higher security weaknesses.
- *Actuators*—are edge devices that share similar traits and security weaknesses with the sensors, also due to their restricted computational and memory capacity.
- *PLCs*—are the edge clients of SCADA with an intermediate level of communication and control weaknesses with respect to the edge devices, yet enjoy a more protected accessibility—both in the field and digitally, through firewalls.
- *SCADA*—are typically the most technically protected and least accessible assets, yet still suffer from in-built vulnerabilities that make them susceptible to persistent threats.

The *Targets* collective is characterised by a relevant "cost" attribute that aggregates the vulnerability level of the asset at hand and the required skills and intelligence that *Attackers* need to successfully carry out an attack against them. In addition, Targets are assigned a protection status variable that reflects the deployment of additional protection measure

by *Utility* agents at asset-specific level. The CPRISK-ABM also models the effects that a *Target*'s attractiveness has on the process, by further categorising them as interesting or not, linking their attributes to the motives and outcome expectancy of individual *Attackers.*

The low-level categorisation and parameterisation of all the agents' collectives, as well as the cyber-security assessment questionnaire are provided in Appendix A (Tables A1–A5).

2.1.4. Process Overview and Scheduling

The CPRISK-ABM focuses on the simultaneous interaction between red and blue team agents, under various interchangeable processes. At each model step, the moving agents (*Targets* and *Attackers*) take a random step within the feasible model space, overseen by the *Utility* agents, which mimics the scanning of the network by adversaries to locate potential exploitable points. Based on the protection level and available resources, the *Utility* agents can monitor and protect a given portion of the system. The *Utility* agents continuously monitor a portion of the system for malicious activity and take actions to prevent it, e.g., patch vulnerabilities, restrict access, etc., making it more difficult to attack the assets they oversee.

When an *Attacker* locates a *Target* in the landing grid, it locks-in and they engage in a two-step attack process. The first step includes the performance of reconnaissance actions to gather information over the *Target*, its characteristics and exploitable vulnerabilities to gain unauthorised access. If the targeted asset is monitored, the *Utility* agents are alerted by suspicious activities and undertake the appropriate actions to remediate any threat from amateurs. Higher tier *Attackers* can deploy more skilful techniques and/or tools to avoid intrusion protection systems (IPS) and remain stealth while gathering information and gaining unauthorised access. Any failed intrusion discloses the hacker and leads to a loss of available "resources" for the individual, as the part of the attacker's techniques is revealed to the blue team. On the other hand, if the *Target* is not actively overseen by a *Utility* agent, any type of *Attacker* can perform its TTPs to breach the node and proceed with the information-gathering process. The successful breach rewards the adversaries with additional knowledge over the system at hand, which translates to higher "resources" that can be utilised for an attack.

After gaining intelligence and access over the targeted asset, the *Attackers* decide if they will proceed with the deployment of an attack against it based on their ingrained behavioural traits and motives. Amateurs and expert attackers will opportunistically seek to deploy an attack against the target, while highly skilled and nation-state affiliated attackers will also co-examine the attractiveness and recognisability of the *Target*. If the targets on sight do not meet their criteria, the two higher tier *Attackers* decide that the intelligence-gathering process is the most profitable action and remain stealth. Thus, those agents seek to gain the maximum possible knowledge over the system before deploying an attack against their desirable target. When Attackers identify a suitable *Target*, they deploy an attack against it. This process is successful if the attacker's technical skills, tools and intelligence (i.e., available "resources"), are sufficient to exploit the vulnerabilities and overcome the protection of the *Target* (i.e., "cost" of attack). Otherwise, their attack fails to reach the end goal while remaining stealth and the attackers are revealed to the blue team, which mitigates the attack and diminishes their available resources.

The *Attacker*'s also lose resources if they wander around the system, scanning for exploitable targets, but do not engage in an attack against *Targets*, as the *Utility* agents oversee the system and can trace fragments of their long-lasting activities. If an *Attacker* reaches a low critical point, the agent is removed and re-instantiated under the same type with the respected attributes and parameter values, to maintain the ratio in every step. The decision-making and the individual agent actions are realised at combined time intervals. As the time variable is not pivotal in the scope of the process, the model's step aggregates the multiple time windows required from various attackers to scan the network, locate potential entry points and decide on the deployment of the attack procedure. By exploring the continuous interaction of agents in sequential steps, the model derives the balance

between them on a macro-scale. The overall interaction of agents within the CPRISK-ABM modelling environment can result in four possible outcomes, i.e., no attack, a simple attack, a motivated attack, a sophisticated attack, or a failed attack against four types of cyber assets, i.e., sensors, actuators, PLCs or the SCADA. They synthesise the probability distribution for attacks and successful attacks at both the asset and system level, under the specific design, characteristics and security posture of a utility.

2.1.5. Initialisation

The CPRISK-ABM initiates by re-creating (a) the assets and their relevant types (*Targets*) according to the system design at hand, (b) the *Attackers* and their types, whose population depends on the size of the utility and the population it serves, and c) the *Utility* agents with the given portion of the system they have under surveillance according to the protection level derived by the questionnaire. The population and the ratio of the types for both the *Attackers* and the *Targets* is maintained stable throughout the simulation steps. At the initialisation step, the descriptive characteristics, motivational traits and individual's cost- or resources-related parameters are assigned by sampling the distribution range that represents each type of agent. The distribution and range of values for each agent is derived though the pattern-oriented set-up that allows the emergence of the observed, real-world traits through the model. In the first step, all agents are placed in a random position within the feasible interaction space.

2.2. The Scenario Generator: Mapping the ABM-Derived Treat Landscape into Scenarios

As mentioned, the proposed framework adheres to industry practices and conceptualises the risk level (R) as a linear relationship between the probability of occurrence (P_r) and the potential consequences (C_r) of a threat. The CPRISK-ABM, described in Section 2.1, provides the required data to derive the vulnerability-induced probabilities of attack per asset and attacker type, and describe the cyber-physical threat landscape of a utility, under a given design and security posture. The second component of risk, that of potential consequences, is the product of risk analysis in which "what-if" scenarios are typically formulated to translate threats into potential (future) events. To streamline the process, this study introduced a scenario generator that maps the probability distribution of the potential attacks, and their relevant characteristics derived by the CPRISK-ABM into a set of representative scenarios, suitable for quantitative analysis.

Scenarios can be described as a combination of potential system states and threat events that lead to undesired consequences [17], and are thus comprised of two distinct components. The first component describes the potential threats that influence specific asset characteristics and the associated parameters that transform them from generic into event-like incidents. The second component is the cyber-physical model of the water network, i.e., the combined physical and cyber layers of the system, and its internal characteristics, i.e., rules, interconnections, demand, etc., that define its potential state when a threat occurs.

2.2.1. From Socio-Technical Analysis to Threat Scenarios

The CPRISK-ABM renders all possible interactions between the utility and the adversaries' ecosystem and accounts for both failed and successful attacks against assets. For the purposes of quantitative analysis, however, the focus shifts towards successful attacks, as those have the potential to disturb the services of an urban water system. Each asset type, based mainly on attributes of vulnerability and attractiveness, has a different probability of being targeted and successfully attacked (P_{asset}) over other asset types. In the process, this is represented in the assemble of scenarios of user-defined size S as a subset, with size $S_{asset} = P_{asset} * S$. Each subset is then further divided according to the attack types carried out against the specific asset type, each having a probability $P_{asset,\ attack}$. Thus, the scenario generator synthesises the overall threat landscape through subsets of scenarios with size $P_{asset,\ attack} * S$, which are then parameterised accordingly. Within this context, in each scenario the attackers focus either on rendering a service inaccessible through Denial

of Service (DoS) attacks or secretly relay and alter the communication between devices, to manipulate the system behaviour through bogus data.

The three key parameters that transforms those generic threats to event-like occasions are (a) the network-specific asset targeted by the adversary, (b) the start time and (c) the duration of the attack. Considering that this step accounts for successful attack scenarios, any additional attack-specific parameter required for the scenario is assumed to be sufficient and within a feasible range to account for the success of the attack, e.g., falsified water level signals fall within the minimum and maximum stages of the tank. The network-specific asset targeted in each scenario is randomly sampled from a list of feasible targets that contains all the system assets of the specific type, e.g., all of the tank level sensors, all of the remotely controlled valves, etc. In the absence of relevant evidence, the framework samples the potential targets following a uniform distribution, apportioning the probability equally among all assets of the same type within the system. However, this step can be coupled with asset-specific analysis tools (quantitative or qualitative) to derive a different probability distribution function, either fixed or for a given period, e.g., after receiving an alert from state intelligence for high activity against specific industrial sensors found in the system. The second scenario parameter, that of the start time of the event, is actually a source of epistemic uncertainty, given that it is impossible to know or predict exactly when the attacker will decide to attack the system. As such, the start time of the event is randomly sampled within the entire simulation duration, i.e., an attack can occur at any given moment within the simulation. On the other hand, the duration of the attack is closely associated with the security posture of the utility and the skills, motives and techniques of individual adversaries, while, as indicated by relevant real-world incidents, cyber-physical attacks can last from minutes to even months before being detected and shut down. The scenario generator utilises the CPRISK-ABM results and associates the duration for each scenario to the respected attack type that emerged from the socio-technical analysis. Within the quantitative analysis process, the duration of the attack is interpreted as a portion of the available time remaining after the attack start time until the completion of the simulation. Scenarios that resemble threats from lower skilled and unmotivated attackers (simple attacks) have an attack duration of 10–25% of the remaining time, while higher tier and more motivated last longer, i.e., 25–50% of the remaining time. Lastly, scenarios that render more sophisticated attacks consider even longer durations, lasting 50–75% of the remaining simulation time, to account for the more refined TTPs and efficient camouflage of the attack by the relevant adversaries. In all cases, the events are assumed to end before the simulation completion, thus allowing modellers to assess the performance of the system's recovery or detect cascading effects that surpass the attack duration.

2.2.2. Stochastically Generated Scenarios

Within the context of deterministic scenario-based risk analysis, the threats, as external drivers, act upon a pre-defined model of the system—assuming that this model is representative of the usual state. Thus, the simulated deviation from the operating conditions and the estimated consequences are also assumed representative. However, urban water systems are dynamic by nature and, even under normal conditions, exhibit significant variability in their behaviour, driven principally by the stochasticity of demands [65]. Hence, uncertainty pertains to the simulation outcomes and the resulting risk information [42,66,67]. This work couples the capabilities of the *anySim* R package [48] as a stochastic time-series generator with the scenario generator, to create stochastically enhanced realisations of the threat scenarios. This allows the examination of threat scenarios under various synthetic, yet realistic, system states that may emerge, and the derivation of uncertainty-aware estimations over their potential consequences.

The *anySim* package offers state-of-the-art stochastic simulation methods that preserve both the marginal distribution and the dependence structure of the underlying stochastic process. This is achieved by coupling linear stochastic models with Nataf's joint distribution model, i.e., Gaussian copula [68,69]. When simulating stochastic processes, *anySim*

translates via the inverse cumulative distribution function (ICDF) an auxiliary Gaussian process (Gp) to recreate processes with the desired marginal distribution and correlation structure. Within the context of this work, the *anySim* models the marginal behaviour of each individual demand pattern though the Beta (bounded) or the Gamma (bounded only from the left) distribution and reproduces a) the auto-correlation structure of each pattern and b) the *lag-0* correlation between all DMA patterns. Thus, the tool allows the generation of multiple synthetic demand patterns, of any temporal scale and duration, with the same characteristics and correlation structures as the reference "historical" demand patterns across the DMAs of the system. The scenario generator randomly samples the poll of synthetic demand patterns and assigns each correlated set to the model of each scenario. In this Monte-Carlo-type experiment, the various threats against the system are examined under a spectrum of realistic model behaviours driven by stochastic demand patters, to account for the inherent uncertain system state under which a cyber-physical attack might take place.

2.3. Cyber-Physical Stress-Testing to Quantify System-Wide Consequences

The next step in the proposed modelling chain is the quantitative analysis of the threat scenarios to derive their potential consequences and define the risk level at both the asset and utility level. To streamline the analysis process, this work builds on the capabilities of Risknought [27,70] which allows the simulation of both quantity and quality related cyber-physicals attacks. A PROCRUSTES-dedicated version of Risknought is integrated with the scenario generator which allows for seamless interaction between the tools and the formulation of a stress-testing platform that can analyse the threat landscape scenarios.

The stress-testing platform explores water distribution networks as cyber-physical systems by simulating the assets' behaviour at the cyber and physical layers and their interactions in a unified process. The simulation renders the sensing, computation and remote action loop based on the system's control logic and automations, that subsequently affect the hydraulic behaviour of the system. Besides individual asset behaviour, the stress-testing platform is able to simulate the cascade both from edge devices upstream of the connected PLCs and SCADA, and from the control devices (i.e., PLC and SCADA) to the downstream connected edge devices. The platform is thus capable of quantifying the "physical" consequences of composite cyber-physical attacks against the various SCADA elements, including sensors, actuators and PLCs, i.e., the targets of the cyber-physical attacks. For the hydraulic simulations, the platform relies on the EPANET 2.2 solver [71], as an industry standard, which enables the pressure-driven demand analysis (PDA) to reproduce the effects that pressure deficiency can have on the service availability within an urban water distribution system. The stress-testing platform utilises a set of metrics, inspired by [72], to capture the system's performance against key integrity objectives and quantify the consequences of each threat scenario. Under the proposed scheme, threat scenarios that affect the hydraulic performance of the system are analysed in respect to (a) the total unmet demand, (b) the number of customers affected and (c) the spatial extend of the consequences, i.e., portion of the network affected.

The generated stress-testing process involves a variety of attack scenarios that target the network infrastructure, covering a wide range of attack characteristics that result from the ABM analysis. To ensure the reliability of the results from the stress-testing procedure, a large number of simulations are typically required. This poses a computational challenge, as the duration of each simulation can be significant, depending on various factors such as simulation parameters, network size and characteristics, hardware capabilities and the efficiency of the solver. For processes that require the execution of a series of independent simulations in a multi-scenario analysis approach, such as sensitivity analysis and stress-testing, the platform adopts a concurrent computing architecture that allows the parallel execution of scenario simulations, to reduce the overall time required to complete the process. In particular, the platform employs Celery, a distributed task queue library that allows for the asynchronous processing of tasks. When a stress-testing task is submitted

for execution, Celery adds it to a queue, from which a broker (RabbitMQ) efficiently distributes the tasks across multiple worker nodes, allowing for their parallel execution. As each task is completed, the result is returned to the application, where it can be further processed, combined with other results and displayed to the user. The platform is designed to make efficient use of the available hardware resources by dynamically adapting the number of concurrent tasks to the capabilities of the processing system and thus being scalable by design. Considering the hardware constraints, the developed architecture is capable of supporting multiple users executing tasks in parallel, originating from different platform tools.

2.4. Evaluating Risk Level at Utility and Asset Level

Risk-related processes and decision-making mechanisms in a company rely on long-established practices, terminology and notions, often framed by standards and guidelines (see e.g., ISO series 31000 [14], European Standard CEN-EN 15975-2 for risk management in drinking water supply [11], American Water Works Association RAMCAP® Standard [73], etc.). Under such frames, a widely adopted conceptualisation of the risk level (R) is that of a linear relationship between the probability of occurrence (P) and the potential consequences (C) of a threat, while other approaches may also include a third individual component to the equation, that of vulnerability (V) (see e.g., [48]).

$$R = P * C \tag{1}$$

Despite differences across industries and fields of application, the appropriateness in the description of risk level depends on the situation examined [74]. In the context of cyber-physical threats, the vulnerabilities of a system become risk sources, waiting to be discovered and exploited by attackers to activate their attack path, and thus directly affect the probability of a successful event, unlike, for example, the probability of occurrence of natural hazards. Thus, in the case of cyber-physical attacks, the effects of vulnerability must be co-examined along with attackers' attributes to derive the probability of successful attacks. Our framework endorses this conceptualisation and takes vulnerability into account intrinsically within the ABM process and ingrains its effects to the probability of successful events and the characterisation of vulnerability-induced attacks. Moreover, the analysis of the total threat landscape, especially under a spectrum of stochastically driven system states, leads to a collection of consequences at various dimensions. To provide a concise and representative overview of the risk level the different dimensions of consequences need to be aggregated into a homogeneous metric that will represent the overall consequence of the threats at hand. To produce the homogeneous metric of consequences for each scenario s, we use the dimensionless form of each metric, such as the percentage of unmet demand, that accounts for the expected services under the stochastically driven reference state of the system. Each of the n total failure metrics examined can be assigned a utility-specific weight factor to account for its risk criteria, as in Equation (2).

$$C_s = \sum_{i=1}^{n} W_i * C_{\%i,s} \tag{2}$$

To derive the risk level at asset level, e.g., the risk level stemming from the sensors, the scenarios can be grouped accordingly and the estimated weighted average value of consequences for the group can be coupled with the ABM-derived probabilities. The relevant probabilities of successful attacks are then derived as the sum of probabilities for all attack type combinations against the specific asset type. Subsequently, the total consequences at the system scale are estimated as the weighted mean of its components, as in Equation (3).

$$C_{system} = \sum_{j=1}^{4} w_j * \widetilde{C}_j \tag{3}$$

where j represents the number of subsets $S_{asset\ j}$ that contain scenarios with the same asset type targeted and $w_j = n(S_j) / \sum_{j=1}^{4} n(S_j)$.

Thus, the risk level can be derived as the combination of consequences and the probability of vulnerability-induced threats at the system or asset level, which is typically visualised through risk matrices. Due to the sensitive nature of risk data and to avoid subjective view over their numerical values, both axes of the risk matrix should be assigned semi-qualitative scales that correspond to the risk criteria of each utility. In the case of consequences, the axis can be classified into acceptable, tolerable, major or critical impacts, according to the risk posture and criteria of the utility, for any given metric. To avoid arbitrary values and adjust the risk matrix in a utility specific manner, the axis of likelihood can be classified within the ranges from P(PL0) and P(PL8), i.e., the ABM-derived probabilities of attack that correspond to the minimum and maximum protection level that can be achieved by the utility. An example of a 3 × 4 risk matrix used in the platform, with semi-qualitative classes and their generic ranges can be seen in Figure 4.

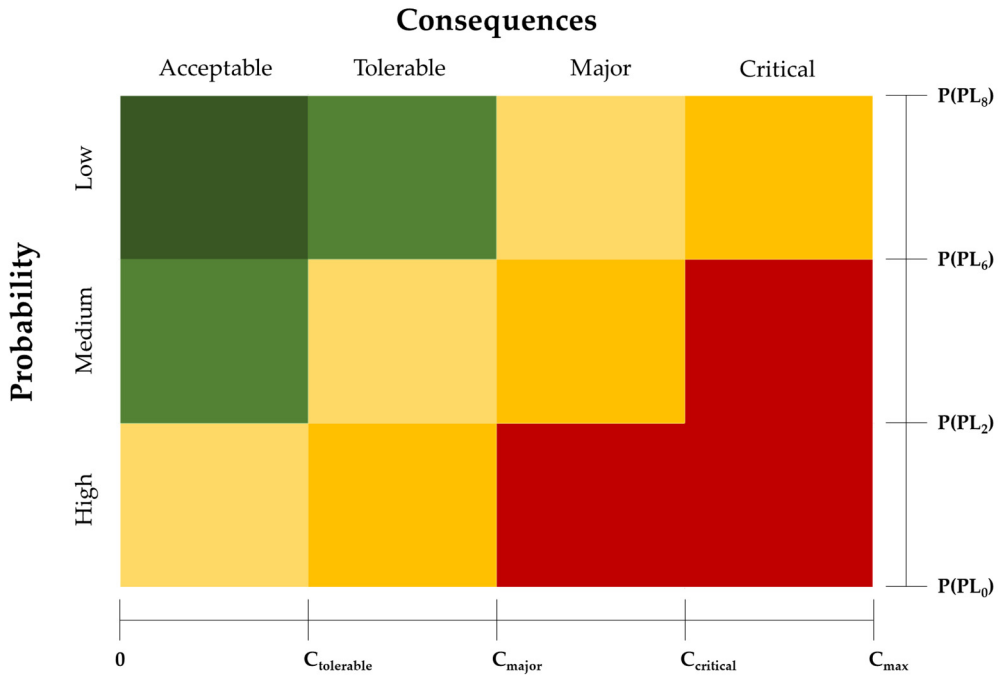

Figure 4. Generic 3 × 4 risk matrix with the respective ranges of classes according to the risk tolerance of the utility (consequences ranges) and the proposed classification of probability ranges—green, yellow and red colours represent low, medium and high risk levels respectively.

3. Demonstration of the Risk Assessment Framework

3.1. Case Study and Cyber-Security Configurations

To demonstrate the risk assessment framework and the underlying modelling chain, this paper presents a case study on a semi-hypothetical utility. The water distribution network of the utility is represented by the widely used C-Town, a benchmark WDN model of an anonymised real-world system of a medium-size city [75]. The system is comprised of five DMAs, with relevant demand patterns, that serve a total of 388 consumption nodes from a single source, i.e., the seasonal reservoir R1. Since water systems exhibit dynamic variability and seasonality at various scales, the reference model of C-Town configured for the risk analysis has a duration of 1 week, with a hydraulic timestep of 1 hr. This timescale allows the stochastically driven processes to randomise the system state, e.g., the status of actuators, the water level in tanks, etc., and also explore attack manifestations on random days of the week, and thus yield more representative results. The threat scenarios are simulated under the PDA approach for more realistic estimation of the consequences, and the parameters assigned for the Wagner equation are P_{nom} = 20 m and P_{min} = 0.0 m. Using as reference the five demand multiplier patterns that characterise the consumption of the system, 780 sets of synthetic patterns of 1-week duration and a 1 hr timestep were generated via the integrated *anySim* module. Each set is comprised of five auto- and cross-correlated

patterns, one for each DMA, that preserve the marginal properties of the original patterns. The median and the outer (upper and lower) bands of the 780 hourly demand multiplier patterns per DMA can be seen in Figure 5.

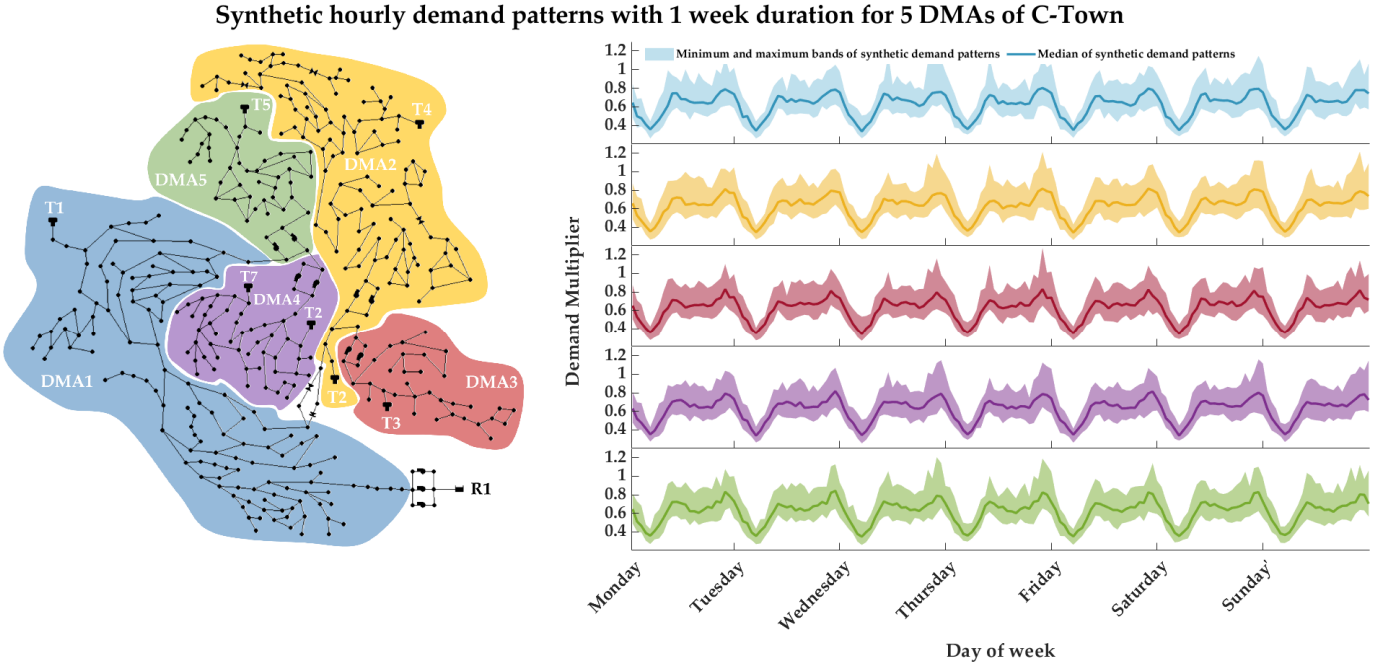

Figure 5. (**Left**) C-Town and DMAs visualisation; (**right**) Median and outer bands for the synthetic hourly demand patterns of 1-week duration generated for the five DMAs of C-Town. The corresponding DMA for each plot is indicated by colour, i.e., blue for DMA1, yellow for DMA2, red for DMA3, purple for DMA4 and green for DMA5.

In terms of its security posture, the hypothetical water utility is assumed to be an average utility, that implements cyber-security protocols and best practices at a moderate level. More specifically, the utility has a fully connected SCADA system that transmits signals wirelessly through privately owned infrastructure. The access to critical systems is provided to authorised personnel, mainly through the internal company network (intranet). However, under the COVID-19 restrictions, remote access and control were enabled, via relevant software, for personnel that worked from home. The utility updates the system and carries out backups at a moderately regular basis, and applies encryption and anomaly detection techniques, but only to a portion of its systems. Lastly, the utility has recently begun to hold annual training for personnel on the topic of cyber-security, necessitated by rising trends in the sector's threat landscape. In our case study analysis, this is denoted as *Configuration 1*, and under this set-up, the utility scored an average protection level (class 3) in the relevant questionnaire.

To investigate the effects and efficiency that different cyber-security practices can have on the risk level of the utility, we also explore a second, hybrid scenario. In this, the utility decides to (a) completely restrict the use of remote access and control software, as a result of the 2021 Florida hack that exploited relevant systems and (b) reduce the wireless transmission from their field devices and make wired connections the predominant path for data relaying. This scenario is characterised as hybrid because it explores both a technical investment that affects the design traits of the cyber layer to secure the communication of field devices and a cost-free change in practice with the restriction of remote access and control software. In the analysis this is denoted as *Configuration 2* and it leads to an increased protection level of the utility (class 5) in the relevant questionnaire.

To analyse the threat landscape and assess the potential consequences against the utility, we follow the modelling chain described previously and perform the stress-testing analysis with 2000 scenarios for each configuration. In respect to the risk assessment

process, the classes of the risk matrix are defined by the assumption of risk criteria for the hypothetical utility and correspond to the values $C_{tolerable}$ = 0.1, C_{major} = 0.2 and $C_{critical}$ = 0.3. The results of the analysis for each configuration at both system and asset level are presented below.

3.2. Results

The key results of the CPRISK-ABM simulations under both configurations of the utility, i.e., *Configuration 1* with protection level 3 and *Configuration 2* with protection level 5, are summarised and presented in Tables 1–3. The probabilities of attack and probabilities of successful attacks per asset type for the utility are presented in Table 1. Both are given as a percentage of the total actions undertaken from threat actors within the ABM simulations. It is worth noting that the remaining percentage of actions taken led to no interaction between attackers and the cyber assets, which can be attributed to either the attackers being uninterested in engaging or unable to successfully scan available ports or signals from the various assets, i.e., they did not identify a suitable potential target to engage. To observe the effects that the new cyber-security practices and system design can have over the threat landscape of the utility, the last section of Table 1 also presents a comparison between the probabilities of the two configurations, as a percentage change in respect to *Configuration 1*.

Table 1. CPRISK-ABM analysis results for Configuration 1 and 2 per asset type.

		Sensors	Actuators	PLC	SCADA
Config. 1	**Probability of attack (%)**	1.15	1.61	1.11	0.162
	Probability of success (%)	31.09	33.03	22.97	21.21
Config. 2	**Probability of attack (%)**	1.21	1.66	1.12	0.132
	Probability of success (%)	25.46	29.74	19.65	20.37
Comparison	**Change in probability of attack (%)**	+5.22	+3.11	+0.90	−19.75
	Change in probability of success (%)	−18.11	−9.94	−14.48	−3.97

Table 2. CPRISK-ABM analysis results for Configuration 1 and 2 organised per asset type, as percentage of the total actions against the system.

		Sensors	Actuators	PLC	SCADA
Config. 1	**Failed to gain access or remain stealth (%)**	0.79	1.08	0.85	0.13
	Simple attacks (%)	0.08	0.13	0.06	0.007
	Motivated attacks (%)	0.18	0.26	0.14	0.02
	Sophisticated attacks (%)	0.10	0.14	0.06	0.005
Config. 2	**Failed to gain access or remain stealth (%)** *(change from conf.1)*	0.90 *(+13.92%)*	1.16 *(+7.41%)*	0.90 *(+5.88%)*	0.11 *(−15.38%)*
	Simple attacks (%) *(change from conf.1)*	0.05 *(−37.50%)*	0.11 *(−15.38%)*	0.03 *(−50%)*	0.002 *(−71.43%)*
	Motivated attacks (%) *(change from conf.1)*	0.17 *(−5.56%)*	0.26 *(0%)*	0.14 *(0%)*	0.02 *(0%)*
	Sophisticated attacks (%) *(change from conf.1)*	0.08 *(−20%)*	0.13 *(−7.14%)*	0.05 *(−16.67%)*	0.007 *(+40%)*

Table 3. Distribution of successful attacks per asset and attack type for Configuration 1 and 2.

		Sensors	Actuators	PLC	SCADA
Config. 1	**Simple attacks (%)**	6.74	10.95	4.84	0.63
	Motivated attacks (%)	15.16	22.32	11.58	1.89
	Sophisticated attacks (%)	8.21	12.21	5.05	0.42
	Total %	30.11	45.47	21.47	2.95
Config. 2	**Simple attacks (%)**	5.19	10.14	3.07	0.24
	Motivated attacks (%)	16.27	24.29	13.21	1.65
	Sophisticated attacks (%)	8.02	12.50	4.72	0.70
	Total %	29.48	46.93	20.99	2.59

The results indicate that the additional measures undertaken by the utility have a significant effect on the entire system. Noticeably, the predominant influence appears for the probability of successful attacks over the edge devices of the system, i.e., sensors and actuators. This can be linked to the reduction in wireless data broadcasting and thus the reduction in easily accessible communication paths and related exploitable vulnerabilities. This also appears to affect the probabilities of successful attacks over PLCs, which operate as an intermediate node in the system, sending and receiving signals from and to the edge devices. The last effect comes both at the probability of attack and successful attacks against the utility's central SCADA. This effect can be associated with the restriction of remote access and control software that served personnel working remotely, and thus can reduce the ability of attackers scanning open ports of the SCADA outside the intranet or attack the system through compromised credentials and other related attacks.

The CPRISK-ABM results can be further analysed for both configurations based on the proposed threat taxonomy to derive the probability of attacks per asset and attack type, as seen in Table 2. It is worth noting that in Table 2, next to the probability values per asset and attack type for *Configuration 2*, we also present the change in respect to the previous, less protected configuration. Overall, the system is expected to have a reduced threat landscape at every asset type under *Configuration 2*.

From this secondary analysis, the most effective protection appears to be against low-skilled hackers that target the SCADA, as they may no longer gain quick access over the system though the remote access software vulnerabilities. It is important to note at this point that the probability of an attack against the utility's SCADA by an unmotivated amateur hacker is now lower, but not zero. Lastly, the ABM results are translated to the relevant distribution of successful attacks per asset and the attack type can be seen in Table 3, for both configurations.

Those distributions are utilised downstream by the scenario generator to render the threat landscape of the utility and synthesise the scenario ensembles for the stress-testing procedure. The last section of Table 3 summarises the distribution percentage assigned to each asset type. Each stress-testing procedure is performed for an ensemble of 2000 scenarios under the ABM-derived distribution, coupled with the stochastic demand pattern timeseries generated by the *anySim*. To estimate the overall consequences (C), we assume equal weight factors for the three key failure metrics, i.e., unmet demand (UD), customers affected (CA) and the spatial extend of the failure in the system (SE). Table 4 presents the statistical characteristics of the failure metrics for each configuration, at the system level, i.e., for all the threat landscape scenarios. The analysis indicates that under the second configuration, the system is also expected to face reduced consequences. The median and average of the overall consequences is estimated to be approximately 33% and 11% smaller, respectively, while the most critical attacks (95th percentile) also appears to have smaller consequences over the system, reduced by approximately 22%. This effect can be attributed to the change in attack distributions and the related attack

characteristics that are deployed against the various assets of the system. On the other hand, the lower tail of the attacks (5th percentile) appears to have the same impact under both configurations, which suggests a "typical" behaviour of the system over less complex or sophisticated attacks.

Table 4. Statistical characteristics of the failure metrics and the overall consequences metric for the ensemble of stochastically driven threat scenarios against the utility under the two different configurations.

		Median	Mean	Max	Min	Range	St. Dev.	95th Percentile	5th Percentile
Config. 1	***UD (%)***	0.01	0.05	0.57	0.00	0.57	0.07	0.20	0.00
	CA (%)	0.24	0.31	1.00	0.01	0.99	0.21	0.86	0.15
	SE (%)	0.22	0.27	0.86	0.01	0.85	0.18	0.74	0.13
	C	0.16	0.21	0.81	0.01	0.80	0.15	0.54	0.09
Config. 2	***UD (%)***	0.01	0.04	0.58	0.00	0.58	0.07	0.19	0.00
	CA (%)	0.16	0.28	1.00	0.01	0.99	0.19	0.61	0.15
	SE (%)	0.14	0.24	0.86	0.01	0.85	0.16	0.53	0.13
	C	0.11	0.19	0.81	0.01	0.80	0.13	0.42	0.09

The consequences metric can then be grouped per asset type, as in Table 5 below, to indicate the source of the consequences at system level. From the breakdown, it is indicated that the average consequences stemming from attacks against the edge devices practically remains at the same level. The different attack characteristics of the threat landscape under *Configuration 2* seem to affect the consequences related to PLC and SCADA targeting attacks.

Table 5. Overall estimated consequences metric per asset type targeted as derived from the stress-testing process under the two configurations.

		Sensors	Actuators	PLC	SCADA
Configuration 1	$\overline{C}$	0.26	0.13	0.26	0.28
Configuration 2	$\overline{C}$	0.25	0.13	0.23	0.17

By combining the probabilities of successful attacks with the overall consequence metric for each configuration in Table 4, we summarise the risk assessment results into the risk matrix at system level, as seen in Figure 6 and at asset level, as seen in Figure 7.

Based on the risk criteria of the utility, under *Configuration 1*, the system is considered to face relatively frequent successful attacks (medium probability) that may cause major consequences. Under the second configuration, the probability of successful attacks, although lower, is still considered medium, but the new level of consequences that may emerge from those attacks appeared to become tolerable. Thus, the system level analysis under *Configuration 2* indicates a new, more efficient response of the utility against its threat landscape and the cyber-physical threats that it is comprised of.

The analysis of the risk level of the utility at asset-type level grants a more detailed and concise view over the distribution of the threat landscape and the critical parts of the system that may need to be prioritised. Through this analysis, we identify that the risk level stemming from sensors and actuators remains in the same class under both configurations, although it is reduced after the implementation of better practices in *Configuration 2*. In the case of sensor-related attacks, the consequences are considered major, while for actuator-related threats, these are considered tolerable. The new behaviour of the system under *Configuration 2* alters the class of risk level stemming from PLC attacks, by the reduction in the probability to low levels, and the slight modification of the consequences. All of the

above signify risks that may need to be further modified under the risk criteria of the utility, as also indicated by the relevant colour code in the risk matrix. On the other hand, under *Configuration 2*, SCADA-related threats are expected to have a low probability and tolerable consequences over the system.

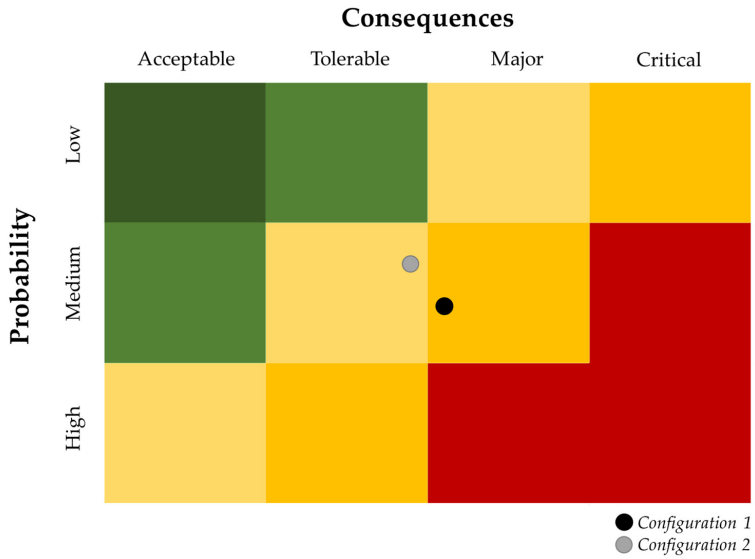

Figure 6. System's risk level under Configuration 1 and Configuration 2 presented in a 3 × 4 risk matrix with semi-qualitative classes adjusted to utility criteria.

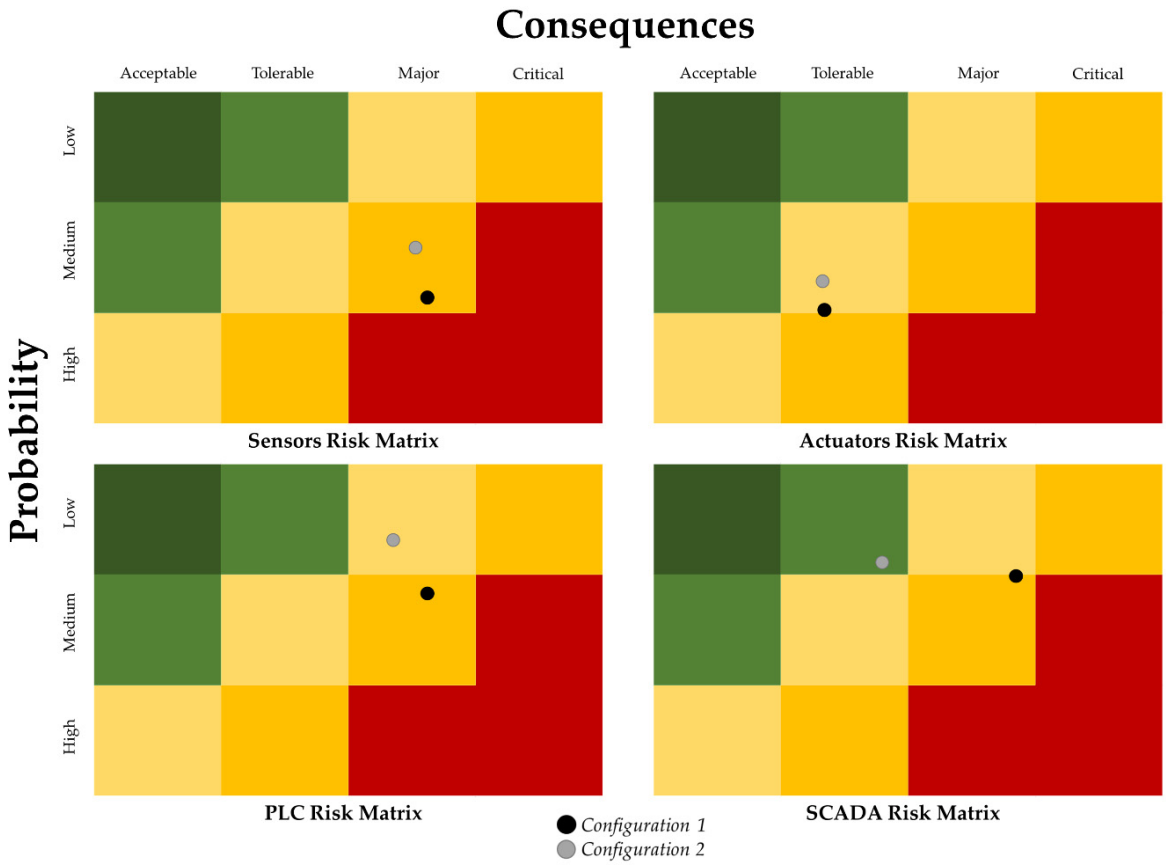

Figure 7. Risk level per asset type targeted under Configuration 1 and Configuration 2 presented in a 3 × 4 risk matrix with semi-qualitative classes adjusted to utility criteria.

The results presented herein demonstrate the current risk exposure of a system and the effects that new practices and design traits can have over the cyber-physical-threat landscape of a utility. The changes between the two configurations are predominantly focused on the modification of the probability and the characteristics of cyber-physical attacks. The change in the level of consequences should thus be considered as a cascade effect of the modified exploitable vulnerabilities of the system that leads adversaries to adjust their behaviour and tactics in order to achieve their goals. On the other hand, technical measures designed specifically to mitigate the potential consequences of attacks, such as the increase of a tank's storage, are also worth exploring, either individually or in combination with measures such as those presented in this paper.

4. Discussion and Conclusions

This paper has proposed and presented an ABM-enhanced and uncertainty-aware risk assessment framework for cyber-physical threats that employs a modelling chain able to render and analyse the threat landscape of any water utility. The introduction of the CPRISK-ABM in the risk assessment process allows for socio-technical analysis of cyber-physical threats under the complex, goal-driven behaviours of various threat actors and the effects of a utility's security posture, to derive the probabilities of vulnerability-induced attacks against assets of the system. Overall, the framework utilises the capabilities of the PROCRUSTES platform components to explore the effects that different practices and system design traits have over the threat landscape and allow the assessment of the risk at the system and asset level. The synthetic case study used to demonstrate the proposed framework is a hypothetical example based on a real-world system of a medium size. The results presented herein are indicative of the effects that the security posture of any utility can have over the threat landscape and the subsequent risk level that stems from it. As such, the PROCRUSTES test-bed platform and its modelling components can be utilised by real-world utilities to identify their current threat landscape and assess their risk level under their existing system design and cyber-security practices, while accounting for the uncertainties that govern the underlying processes. The analysis of risk at the asset level can also help the utility to gain a better understanding of the current state and the vulnerable components that require attention. The potential application of the framework also extends to the investigation of the periodic/seasonal changes in the risk level of utilities that service tourist destinations. The models can adjust to both seasonal demand conditions and the population changes that can affect the interest of attackers, to explore the variations in risk level that such utilities face during different periods, under the same cyber-security posture and system design.

Moreover, the proposed framework allows for the quantitative exploration of potential mitigation measures, both operational and technical, that the utility might wish to implement to modify the threat landscape and reduce the system's risk to lower and more acceptable levels. The effectiveness of such measures can be explored either individually or in combination, allowing the utility to explore the full potential of a strategic cyber-security plan and organise its stepwise implementation. Thus, the platform can serve both as an analytical tool for risk assessors and as a decision-support system that provides quantitative proof to stakeholders to support the design, selection and implementation of strategic plans against cyber-physical threats.

The threat landscape analysis can be performed for both shorter and longer timespans than the 1 week presented in the case study, and both smaller and larger scenario ensembles. Small ensembles, however, may be unable to properly sample the threat landscape or account for uncertainties in the final risk results. On the other hand, very large ensembles are bound to lead to higher computational times, which under the parallel processing architecture of the PROCRUSTES platform, can be significantly reduced. The computational load is also affected by the selected hydraulic analysis timestep. The hourly timestep is generally proposed for the analysis of distribution networks, as it can sufficiently describe the operational behaviour and the adjustments of the system to demand changes. In

particular, under the context of the proposed framework, which aims for a bird's eye view of the system's resilience against the entire threat landscape, any finer timescale analysis, e.g., 15 min, would most likely lead to additional computational loads without significant changes in the overall picture. Lastly, the proposed framework and the platform are expandable to the analysis of additional components within urban water systems, such as water treatment plants, with the provision of the relevant subsystem model and adjustment of the inputs. The examination of upstream components of the urban water system will also require the assessment of cascade effects such as quality-related attacks that may occur purely from the remote tempering of, for example, chemical treatment processes in the water treatment plant, and cascade to the distribution system. Nevertheless, utilities should also consider the risks stemming from complex cyber-physical attacks at the distribution level that combine a physical injection of chemical or biological factors and simultaneous blinding of the sensors, and establish a resilient design of their sensor network [76,77].

We argue that, under the uncertain and ever-expanding cyber-physical threat landscape, the risk information offered through the proposed modelling chain can lead to a better understanding of a system's current exposure. Furthermore, it can facilitate evidence-based decision making over a utility's design and security practices to achieve higher resilience and meet legislative provisions, such as those foreseen under the newly established EU Directives for cyber-security and the resilience of critical entities.

Author Contributions: Conceptualisation, G.M. and C.M.; methodology, G.M., G.K., G.-K.S., I.T., P.K. and D.N.; software, G.M., G.K., G.-K.S., P.K., I.T. and D.N.; validation, I.T., P.K., G.-K.S. and C.M.; formal analysis, G.M. and G.-K.S.; investigation, G.M.; resources, G.K. and C.M.; data curation, G.M., I.T., P.K. and D.N.; writing—original draft preparation, G.M. and G.-K.S.; writing—review and editing, G.M., G.-K.S., G.K., D.N., I.T., P.K. and C.M.; visualisation, G.M.; supervision, G.K. and C.M.; project administration, C.M. and G.M.; funding acquisition, C.M. and G.M. (PhD fellowship). All authors have read and agreed to the published version of the manuscript.

Funding: The research work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: HFRIFM17-2918) and the "Third Call for HFRI PhD Fellowships" (Fellowship Number: 6349).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. CPRISK-ABM State Variables

Table A1. Attacker's characteristics.

Variable	Value	Description	Set-Up
Type of attacker	Amateur, Experts, Highly skilled, Nation-state affiliated	Attackers are divided into four types that depict their different motives, behavioural traits and ability levels.	The ratio is calibrated from real-world patterns and the total number is automatically adjusted to the system's size and digitalisation level, i.e., more customers and more cyber-layer assets attract more adversaries.
Resources	[1,2,3,4,5,6]	Each Attacker is assigned a resource that aggregates its skills, experience and available tools, as well as access to intelligence, relevant to its type.	The range of resources relevant to each type is calibrated from real-world patterns. Amateurs ∈ [1,2] Experts ∈ [3,5] Highly skilled ∈ [5] Nation-state affiliated ∈ [6]

Table A1. *Cont.*

Variable	Value	Description	Set-Up
Motivation	None, Financial, Reputational	Different motives lead to different attack procedures and behavioural traits, related to the selection of a target, the decision to deploy an attack or remain stealth and gather more information over the system.	The range of motives is calibrated to allow the emergence of the real-world patterns. Amateurs: None 33%, Financial 33%, Reputational 33% Experts: None 25%, Financial 50%, Reputational 25% Highly skilled: None 25%, Financial 50%, Reputational 25% Nation-state affiliated: Financial 40%, Reputational 60%
Type of compromise	None, Failed, With protection, Without protection	The gain or loss of resources is affected by the type of compromise.	Defined by the Attacker, Target and Utility agents' interaction at each step.
Type of attack	None, Failed, Simple, Motivated, Sophisticated	Different behavioural traits per Attacker and asset characteristics lead to different attack characteristics.	Defined by the Attacker, Target and Utility agents' interaction at each step.

Table A2. Target's characteristics.

Variable	Value	Description	Set-Up
Type of target	Sensor, Actuator, PLC, SCADA	Targets are divided into four different types that depict the key nodes of the cyber layer, each characterised by different exploitable vulnerabilities.	The ratio is adjusted to the utility-specific network and remains the same throughout the simulation steps.
Cost	[0,1,2,3,4]	Each Target is assigned a cost attribute that aggregates the asset-specific protection, the in-built vulnerabilities, and the required resources from the attackers to gain access and perform a cyber-attack while stealth.	The initial cost per Target type is assigned on a range, calibrated from real-world patterns.
Recognisability	Common, Interesting	The Target is characterised based on its recognisability, which makes the Target more interesting to Attackers.	Recognisability of an asset is defined at the initialisation step.
Protection	Protected, Not protected	The Target is characterised based on its protection against cyber-attacks from the Utility agents, and affects the total cost required for Attackers to succeed.	Defined by the Target and Utility agents' interaction at each step.
State	Working, Compromised, Attacked	The Target at each time step saves its state in respect to its interaction with Attackers.	Defined by the Attacker, Target, and Utility agents' interaction at each step.
Attacker characteristics	Attacker ID, Attacker Type, Attacker Resources, Type of compromise, Type of attack	The Target at each time step saves the characteristics of the Attacker that they are in contact with.	Defined by the Attacker, and Target agents' interaction at each step.

Table A3. Utility agents' characteristics.

Variable	Value	Description	Set-Up
Protection	[0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%]	The Utility agents monitor and protect a given portion of the entire system and offer additional protection to the Targets within it.	The given portion of the system that Utility agents have under surveillance is related to the system's protection level, derived by a semi-quantitative questionnaire.

In order to derive the level of protection, prior to the ABM initialisation, the utility should complete the following questionnaire as a semi-quantitative cyber-security assessment. The elements and relevant scores assigned are associated with the patterns identified.

Table A4. Semi-quantitative protection level assessment.

Description of Cyber-Security Element	Score per Implementation Level			
	Full	**Moderate**	**Partial**	**Not Applied**
P1. Access to the central SCADA operation's room and other sensitive system's is restricted to authorised personnel and monitored.	4	3	2	1
P2. Anomaly detection system/software is deployed to detect data points that do not align with standard data patterns.	4	3	2	1
P3. Encryption techniques are applied to communication and data storage.	4	3	2	1
P4. Software systems' updates and patches are regularly checked and implemented, under a standardised protocol.	4	3	2	1
P5. Backups of the critical system components are regularly copied and preserved in a secondary storage location.	4	3	2	1
P6. The utility has a backup system able to by-pass the main system and operate key infrastructure assets.	4	3	2	1
P7. The utility personnel are trained and/or certified at regular intervals on the latest cyber-security protocols and best practices.	4	3	2	1
V1. The central SCADA system is connected to all the local PLC stations.	4	3	2	1
V2. Data relaying is achieved through wireless communication.	1	2	3	4
V3. The utility uses private (or leased with exclusive use) communication infrastructure for the communication between SCADA elements.	4	3	2	1
V4. System access is provided according to the staff member's role and relevant access rights.	4	3	2	1
V5. The SCADA elements and related systems are connected only to an isolated, internal company network (intranet).	4	3	2	1
V6. Remote access and remote-control computer software are allowed (e.g., Remote Desktop Connection, etc.).	1	2	3	4
V7. Local control settings can be reset or overridden manually or by the central system.	3	2	2	1

Based on the overall score, the protection level of the system used for the Utility agents, is translated as follows:

Table A5. Overall score and associated protection level offered by the Utility agents.

Protection Level	0	1	2	3	4	5	6	7	8
Score	14–27	28–34	35–37	38–40	41–44	45–48	49–52	53–55	56–59

It should be noted that the maximum protection level is achieved only by utilities that implement additional cyber-security technologies in addition to all of the best practices.

References

1. Makropoulos, C.; Savić, D.A. Urban hydroinformatics: Past, present and future. *Water* **2019**, *11*, 1959. [CrossRef]
2. Lu, Y. Industry 4.0: A survey on technologies, applications and open research issues. *J. Ind. Inf. Integr.* **2017**, *6*, 1–10. [CrossRef]
3. Rajkumar, R.; Lee, I.; Sha, L.; Stankovic, J. Cyber-Physical Systems: The Next Computing Revolution. *Cybern. Syst. Anal.* **2017**, *53*, 821–834. [CrossRef]
4. Lee, E.A. The past, present and future of cyber-physical systems: A focus on models. *Sensors* **2015**, *15*, 4837–4869. [CrossRef] [PubMed]
5. Savić, D. Digital water developments and lessons learned from automation in the car and aircraft industries. *Engineering* **2021**, *9*, 35–41. [CrossRef]
6. Loukas, G. Front-matter. In *Cyber-Physical Attacks*; Elsevier: Oxford, UK, 2015; pp. i–iii. ISBN 9780128012901.
7. Johnson, C.S.; Badger, M.L.; Waltermire, D.A.; Snyder, J.; Skorupka, C. *Guide to Cyber Threat Information Sharing*; NIST Special Publication: Gaithersburg, MD, USA, 2016.
8. Robles, F.; Perlroth, N. 'Dangerous Stuff': Hackers Tried to Poison Water Supply of Florida Town. Available online: https://www.nytimes.com/2021/02/08/us/oldsmar-florida-water-supply-hack.html (accessed on 5 February 2023).
9. Verizon. *Data Breach Digest. Scenarios from the Field*; Verizon: Jersey City, NJ, USA, 2016.
10. Cimpanu, C. Two More Cyber-Attacks Hit Israel's Water System. Available online: https://www.zdnet.com/article/two-more-cyber-attacks-hit-israels-water-system/ (accessed on 5 February 2023).
11. *CEN-EN 15975-2*; Security of Drinking Water Supply—Guidelines for Risk and Crisis Management Part 2: Risk Management. European Committee for Standardization (CEN): Brussels, Belgium, 2013; Volume 18.
12. *Directive (EU) 2022/2557 the European Parliament and of the Council of 14 December 2022 on the Resilience of Critical Entities and Repealing Council Directive 2008/114/EC*; Official Journal L 333; European Union: Brussels, Belgium, 2022; pp. 164–198.
13. NIS2. *Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on Measures for a High Common Level of Cybersecurity across the Union, Amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and Repealing Directive (EU) 2016/1148 (NIS 2 Directive)*; Official Journal L 333; European Union: Brussels, Belgium, 2022; pp. 80–152.
14. *ISO 31000*; Risk Management—Principles and Guidelines. ISO: Geneva, Switzerland, 2018.
15. Boyens, J.M.; Paulsen, C.; Moorthy, R.; Bartol, N. *Supply Chain Risk Management Practices for Federal Information Systems and Organizations*; NIST Special Publication: Gaithersburg, MD, USA, 2015.
16. Theocharidou, M.; Giannopoulos, G. *Risk Assessment Methodologies for Critical Infrastructure Protection. Part II: A New Approach*; Joint Research Centre, Institute for the Protection and Security of the Citizen, Publications Office: Luxembourg, 2015; ISBN 9789279492464. [CrossRef]
17. American Water Works Association. *Risk and Resilience Management of Water and Wastewater Systems*, 1st ed.; AWWA J100-10 (R13); American Water Works Association: Denver, CO, USA, 2010; ISBN 9781583217887.
18. Kahneman, D.; Frederick, S. Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In *Heuristics and Biases: The Psychology of Intuitive Judgment*; Griffin, D., Kahneman, D., Gilovich, T., Eds.; Cambridge University Press: Cambridge, UK, 2002; pp. 49–81. ISBN 9780521792608.
19. Sanfey, A.G.; Rilling, J.K.; Aronson, J.A.; Nystrom, L.E.; Cohen, D. The Neural Basis of Economic Decision-Making in the Ultimatum Game. *Science* **2003**, *300*, 1755–1758. [CrossRef]
20. Wangen, G. Quantifying and Analyzing Information Security Risk from Incident Data. In *Graphical Models for Security. GraMSec 2019*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2019; Volume 11720. [CrossRef]
21. Florêncio, D.; Herley, C. Sex, Lies and Cyber-Crime Surveys. In *Economics of Information Security and Privacy III*; Schneier, B., Ed.; Springer: New York, NY, USA, 2013; pp. 35–53. ISBN 978-1-4614-1981-5.
22. European Union Agency for Network and Information Security (ENISA). *ENISA Threat Landscape: Emerging Trends*; Lourenço, M.B., Marinos, L., Eds.; ENISA: Athens, Greece, 2020; ISBN 978-92-9204-354-4. [CrossRef]
23. Nikolopoulos, D.; Moraitis, G.; Makropoulos, C. 7. Strategic and Tactical Cyber-Physical Security for Critical Water Infrastructures. In *Cyber-Physical Threat Intelligence for Critical Infrastructures Security: Securing Critical Infrastructures in Air Transport, Water, Gas, Healthcare, Finance and Industry*; Soldatos, J., Praça, I., Jovanovic, A., Eds.; Now Publishers: Boston, MA, USA; Delft, The Netherlands, 2021; pp. 159–187. ISBN 978-1-68083-823-7.
24. Ten, C.W.; Liu, C.C.; Manimaran, G. Vulnerability assessment of cybersecurity for SCADA systems. *IEEE Trans. Power Syst.* **2008**, *23*, 1836–1846. [CrossRef]
25. European Union Agency for Network and Information Security (ENISA). *Analysis of ICS-SCADA Cyber Security Maturity Levels in Critical Sectors*; ENISA: Athens, Greece, 2015; ISBN 978-92-9204-135-9. [CrossRef]
26. Tuptuk, N.; Hazell, P.; Watson, J.; Hailes, S. A Systematic Review of the State of Cyber-Security in Water Systems. *Water* **2021**, *13*, 81. [CrossRef]
27. Nikolopoulos, D.; Moraitis, G.; Bouziotas, D.; Lykou, A.; Karavokiros, G.; Makropoulos, C. Cyber-Physical Stress-Testing Platform for Water Distribution Networks. *J. Environ. Eng.* **2020**, *146*, 04020061. [CrossRef]

28. Taormina, R.; Galelli, S.; Douglas, H.C.; Tippenhauer, N.O.; Salomons, E.; Ostfeld, A. A toolbox for assessing the impacts of cyber-physical attacks on water distribution systems. *Environ. Model. Softw.* **2019**, *112*, 46–51. [CrossRef]
29. Antonioli, D.; Tippenhauer, N.O. MiniCPS. In Proceedings of the First ACM Workshop on Cyber-Physical Systems-Security and/or PrivaCy—CPS-SPC'15; ACM Press: New York, NY, USA, 2015; pp. 91–100.
30. Department of Homeland Security. *NIPP 2013 Partnering for Critical Infrastructure Security and Resilience*; Homeland Security: Washington, DC, USA, 2013; pp. 1–57.
31. Renn, O. Three decades of risk research: Accomplishments and new challenges. *J. Risk Res.* **1998**, *1*, 49–71. [CrossRef]
32. Ryan, J.J.C.H.; Mazzuchi, T.A.; Ryan, D.J.; Lopez de la Cruz, J.; Cooke, R. Quantifying information security risks using expert judgment elicitation. *Comput. Oper. Res.* **2012**, *39*, 774–784. [CrossRef]
33. Wiedlea, A.C.K. Expert Elicitation for Risk Assessment. In *Wiley StatsRef: Statistics Reference Online*; Wiley: Hoboken, NJ, USA, 2014.
34. Tversky, A.; Kahneman, D. Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. In *Readings in Cognitive Science*; Collins, A., Smith, E.E., Eds.; Morgan Kaufmann: Burlington, MA, USA, 1988; pp. 440–451. ISBN 978-1-4832-1446-7.
35. Nowotny, H. *The Cunning of Uncertainty*; John Wiley & Sons: New York, NY, USA, 2015; ISBN 0745687652.
36. Loukas, G. Cyber-Physical Attacks on Industrial Control Systems. In *Cyber-Physical Attacks*; Elsevier: Amsterdam, The Netherlands, 2015; pp. 105–144.
37. Montibeller, G.; Winterfeldt, D. Cognitive and Motivational Biases in Decision and Risk Analysis: Biases in Decision and Risk Analysis. *Risk Anal.* **2015**, *35*, 1230–1251. [CrossRef] [PubMed]
38. European Environment Agency. *Looking Back on Looking Forward: A Review of Evaluative Scenario Literature*; European Environment Agency: Copenhagen, Denmark, 2009.
39. Ahern, J. From fail-safe to safe-to-fail: Sustainability and resilience in the new urban world. *Landsc. Urban Plan.* **2011**, *100*, 341–343. [CrossRef]
40. Kozak, M.W. Safety assessment for near-surface disposal of low and intermediate level wastes. In *Geological Repository Systems for Safe Disposal of Spent Nuclear Fuels and Radioactive Waste*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 475–498.
41. Klinke, A.; Renn, O. A New Approach to Risk Evaluation and Management: Risk-Based, Precaution-Based, and Discourse-Based Strategies 1. *Risk Anal.* **2002**, *22*, 1071–1094. [CrossRef] [PubMed]
42. Moraitis, G.; Tsoukalas, I.; Kossieris, P.; Nikolopoulos, D.; Karavokiros, G.; Kalogeras, D.; Makropoulos, C. Assessing Cyber-Physical Threats under Water Demand Uncertainty. *Environ. Sci. Proc.* **2022**, *21*, 18. [CrossRef]
43. Moraitis, G.; Nikolopoulos, D.; Koutiva, I.; Tsoukalas, I.; Karavokyros, G.; Makropoulos, C. The PROCRUSTES testbed: Tackling Cyber-Physical Risk for Water Systems. In Proceedings of the EGU General Assembly 2021, Online, 19–30 April 2021; p. EGU21-14903. [CrossRef]
44. Makropoulos, C.; Nikolopoulos, D.; Palmen, L.; Kools, S.; Segrave, A.; Vries, D.; Koop, S.; van Alphen, H.J.; Vonk, E.; van Thienen, P.; et al. A resilience assessment method for urban water systems. *Urban Water J.* **2018**, *15*, 316–328. [CrossRef]
45. Giannopoulos, G.; Filippini, R.; Schimmer, M. *Risk Assessment Methodologies for Critical Infrastructure Protection. Part I: A State of the Art*; Joint Research Centre, Institute for the Protection and Security of the Citizen, Publications Office: Luxembourg, 2012; ISBN 978-92-79-23839-0. [CrossRef]
46. Makropoulos, C.; Karavokiros, G.; Moraitis, G.; Nikolopoulos, D.; Bouziotas, D.; Lykoy, A. Introducing a Risk Assessment and Evaluation Toolkit (RAET) for cyber-physical preparedness and planning of critical water infrastructures. In Proceedings of the IWA Digital World Water Congress, Copenhagen, Denmark, 24 May–4 June 2021.
47. Koutiva, I.; Moraitis, G.; Makropoulos, C. An Agent-Based Modelling approach to assess risk in Cyber-Physical Systems (CPS). In Proceedings of the 17th International Conference on Environmental Science and Technology, Athens, Greece, 1–4 September 2021. [CrossRef]
48. Tsoukalas, I.; Kossieris, P.; Makropoulos, C. Simulation of Non-Gaussian Correlated Random Variables, Stochastic Processes and Random Fields: Introducing the anySim R-Package for Environmental Applications and Beyond. *Water* **2020**, *12*, 1645. [CrossRef]
49. Bonabeau, E. Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7280–7287. [CrossRef]
50. Grimm, V.; Berger, U.; Bastiansen, F.; Eliassen, S.; Ginot, V.; Giske, J.; Goss-Custard, J.; Grand, T.; Heinz, S.K.; Huse, G.; et al. A standard protocol for describing individual-based and agent-based models. *Ecol. Modell.* **2006**, *198*, 115–126. [CrossRef]
51. Grimm, V.; Berger, U.; DeAngelis, D.L.; Polhill, J.G.; Giske, J.; Railsback, S.F. The ODD protocol: A review and first update. *Ecol. Modell.* **2010**, *221*, 2760–2768. [CrossRef]
52. Grimm, V.; Railsback, S.F.; Vincenot, C.E.; Berger, U.; Gallagher, C.; DeAngelis, D.L.; Edmonds, B.; Ge, J.; Giske, J.; Groeneveld, J.; et al. The ODD Protocol for Describing Agent-Based and Other Simulation Models: A Second Update to Improve Clarity, Replication, and Structural Realism. *J. Artif. Soc. Soc. Simul.* **2020**, *23*, 7. [CrossRef]
53. Masad, D.; Kazil, J. Mesa: An Agent-Based Modeling Framework. In Proceedings of the 14th Python in Science Conference (SCIPY 2015), Austin, TX, USA, 6–12 July 2015; pp. 51–58.
54. NIST. *Guide for Conducting Risk Assessments*; NIST: Gaithersburg, MD, USA, 2012.
55. European Union Agency for Network and Information Security (ENISA). *ENISA Threat Landscape 2022*; ENISA: Athens, Greece, 2022.
56. Verizon. *2017 Data Breach Investigations Report (DBRIR) Tips on Getting the Most from This Report*, 10th ed.; Verizon: Jersey City, NJ, USA, 2017.

57. Verizon. *2019 Data Breach Investigations Report*; Verizon: Jersey City, NJ, USA, 2019.
58. Verizon. *2018 Data Breach Investigations Report*; Verizon: Jersey City, NJ, USA, 2018.
59. Markovic-Petrovic, J.D.; Stojanovic, M.D. Analysis of SCADA system vulnerabilities to DDoS attacks. In *2013 11th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services, TELSIKS 2013*; Institute of Electrical and Electronics Engineers: New York, NY, USA, 2013; Volume 2, pp. 591–594. [CrossRef]
60. European Union Agency for Network and Information Security (ENISA). *Communication Network Dependencies for ICS/SCADA Systems*; ENISA: Athens, Greece, 2017; ISBN 9789292041922.
61. Grimm, V.; Frank, K.; Jeltsch, F.; Brandl, R.; Uchmański, J.; Wissel, C. Pattern-oriented modelling in population ecology. *Sci. Total Environ.* **1996**, *183*, 151–166. [CrossRef]
62. Grimm, V.; Revilla, E.; Berger, U.; Jeltsch, F.; Mooij, W.M.; Railsback, S.F.; Thulke, H.-H.; Weiner, J.; Wiegand, T.; DeAngelis, D.L. Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science* **2005**, *310*, 987–991. [CrossRef] [PubMed]
63. Mooij, W.M.; DeAngelis, D.L. Uncertainty in Spatially Explicit Animal Dispersal Models. *Ecol. Appl.* **2003**, *13*, 794–805. [CrossRef]
64. Ajzen, I. The theory of planned behavior. *Organ. Behav. Hum. Decis. Process.* **1991**, *50*, 179–211. [CrossRef]
65. Kossieris, P.; Makropoulos, C. Exploring the Statistical and Distributional Properties of Residential Water Demand at Fine Time Scales. *Water* **2018**, *10*, 1481. [CrossRef]
66. Nikolopoulos, D.; Kossieris, P.; Tsoukalas, I.; Makropoulos, C. Stress-Testing Framework for Urban Water Systems: A Source to Tap Approach for Stochastic Resilience Assessment. *Water* **2022**, *14*, 154. [CrossRef]
67. Winkler, R.L. Uncertainty in probabilistic risk assessment. *Reliab. Eng. Syst. Saf.* **1996**, *54*, 127–132. [CrossRef]
68. Tsoukalas, I.; Efstratiadis, A.; Makropoulos, C. Building a puzzle to solve a riddle: A multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure. *J. Hydrol.* **2019**, *575*, 354–380. [CrossRef]
69. Tsoukalas, I.; Efstratiadis, A.; Makropoulos, C. Stochastic Periodic Autoregressive to Anything (SPARTA): Modeling and Simulation of Cyclostationary Processes with Arbitrary Marginal Distributions. *Water Resour. Res.* **2018**, *54*, 161–185. [CrossRef]
70. Nikolopoulos, D.; Makropoulos, C. Stress-testing water distribution networks for cyber-physical attacks on water quality. *Urban Water J.* **2022**, *19*, 256–270. [CrossRef]
71. Rossman, L.A.; Woo, H.; Tryby, M.; Shang, F.; Janke, R.; Haxton, T. *EPANET 2.2 User Manual—EPA/600/R-20/133*; United States Environmental Protection Agency (EPA): Washington, DC, USA, 2020.
72. Moraitis, G.; Nikolopoulos, D.; Bouziotas, D.; Lykou, A.; Karavokiros, G.; Makropoulos, C. Quantifying Failure for Critical Water Infrastructures under Cyber-Physical Threats. *J. Environ. Eng.* **2020**, *146*, 04020108. [CrossRef]
73. ASME-ITI. *All-Hazards Risk and Resilience: Prioritizing Critical Infrastructures Using the RAMCAP Plus Approach*; American Society of Mechanical Engineers (ASME): New York, NY, USA, 2009; ISBN 9780791802878.
74. Aven, T. Risk assessment and risk management: Review of recent advances on their foundation. *Eur. J. Oper. Res.* **2016**, *253*, 1–13. [CrossRef]
75. Ostfeld, A.; Salomons, E.; Ormsbee, L.; Uber, J.G.; Bros, C.M.; Kalungi, P.; Burd, R.; Zazula-Coetzee, B.; Belrain, T.; Kang, D.; et al. Battle of the Water Calibration Networks. *J. Water Resour. Plan. Manag.* **2012**, *138*, 523–532. [CrossRef]
76. Nikolopoulos, D.; Moraitis, G.; Karavokiros, G.; Bouziotas, D.; Makropoulos, C. Stress-Testing Alternative Water Quality Sensor Designs under Cyber-Physical Attack Scenarios. *Environ. Sci. Proc.* **2022**, *21*, 17. [CrossRef]
77. Nikolopoulos, D.; Makropoulos, C. A novel cyber-physical resilience-based strategy for water quality sensor placement in water distribution networks. *Urban Water J.* **2023**, *20*, 278–297. [CrossRef]