ABSTRACT

of the thesis of Ainagul M. Seitbekova on the topic:

"Research and development of methods for optimizing the fuel consumption of district heating systems based on the detection of anomalies in big data" submitted for the degree of Doctor of Philosophy (PhD) in the educational programme 8D07103 – Automation and Control

Research relevance. District heating systems are a key element of Kazakhstan's energy and social infrastructure. A long heating season, substantial changes in outdoor temperature, and daily and seasonal variation in consumer demand require accurate coordination of heat production, transmission, and consumption. According to the statistics analysed in the dissertation, the total length of heat networks reaches 13.4 thousand km; 323 km were repaired or reconstructed in 2025, 377 km are scheduled for renewal in 2026, and 1.6 thousand km are to be modernised by 2029. The number of heat-supply sources in 2025 was 2,688. These figures demonstrate the scale of the sector and the importance of continuous technical-condition monitoring.

Ageing networks and equipment, elevated heat losses, insufficient measuring channels, incompatible information systems, and fragmented storage of process data limit efficient fuel use. Incomplete and unsynchronised information on temperature, pressure, heat-carrier flow, heat output, outdoor temperature, and equipment status reduces the speed and reliability of dispatch decisions. Sensor bias, a frozen signal, or communication loss may resemble a process fault, whereas a legitimate load change may be incorrectly classified as an anomaly. Monitoring individual fixed thresholds is therefore insufficient.

Kazakhstan's heat-sector digitalisation objectives call for systems that collect, process, monitor, and analyse data; observe temperature and pressure in real time; integrate heat-metering information; and forecast resource consumption using big-data methods. Meeting these objectives requires physical models, programmable logic control, SCADA, archives, machine-learning algorithms, and optimisation tools to operate within one computational and control environment.

State of the problem and research gap. Existing research has substantially developed thermo-hydraulic calculation, automated control, digital twins, anomaly detection, and fuel-consumption optimisation. Specialised software calculates network topology, steady and dynamic operating conditions, pressure and temperature distribution, and heat losses. Statistical thresholds, isolation methods, one-class classifiers, autoencoders, and temporal models are used in machine learning. Nevertheless, these tasks are commonly addressed separately: the physical model is isolated from the control system, while a data-driven detector does not fully represent physical laws and operating context.

The scarcity of labelled industrial data on real failures, different sensor sampling rates, seasonal variation, and uneven measurement quality make reliable algorithm evaluation difficult. Moreover, anomaly detection alone does not save fuel: a diagnostic indicator must be converted into a safe decision concerning heat-source output, supply temperature, differential pressure, pump operation, or load allocation. This gap defined the dissertation: a comprehensive method is needed that combines physical laws with intelligent data analysis and verifies a decision in Software-in-the-Loop before it reaches physical equipment.

Aim, object, and subject of the research. The aim is to develop methods and automated-control tools for optimising fuel consumption in a district heating system by detecting deviations in large volumes of operating and process data. The work goes beyond registering an anomaly and evaluates its calculated effect on heat production, network pressure and temperature, heat delivered to consumers, and fuel demand.

The research object is a district heating system comprising a heat source, supply and return pipelines, pumping equipment, heating substations, consumers, measuring instruments, and an automated control system. The subject includes process-data collection, synchronisation, and preprocessing; patterns of normal and abnormal operation; thermo-hydraulic models; anomaly-detection algorithms; and fuel-reduction methods subject to heat-quality and safety constraints.

Research tasks. The work analysed the current state and digitalisation level of Kazakhstan's heat-power sector; identified factors affecting fuel demand, including outdoor temperature, load, supply and return temperatures, pressure, flow, network heat losses, and equipment efficiency;

developed a digital-twin architecture that integrates PLC, SCADA, archive, and model data; constructed a Fluidit Heat model of a real network section; implemented an independent physical model in MATLAB/Simulink/Simscape; tested normal operation, increased load, sensor failure, communication loss, and process deviations in Software-in-the-Loop; verified component logic and timing in UPPAAL; compared six anomaly detectors and formed a majority-voting decision; and developed a heat-energy calculation algorithm and laboratory heat-meter prototype.

Methods and experimental environment. The study used system analysis, mass and energy balances, Darcy-Weisbach hydraulic resistance, the Colebrook–White relation, heat-transfer calculations, pump characteristics, simulation modelling, statistical processing, comparative analysis, machine learning, and timed automata. Diagnostic features were formed from residuals between measured and calculated parameters, mass- and energy-balance discrepancies, and departures of multivariate time series from their normal structure.

The Software-in-the-Loop environment integrated TIA Portal V20, a virtual SIMATIC S7-1500 controller configured as CPU 1513F-1, S7-PLCSIM Advanced V7.0, WinCC Runtime Advanced V20, OPC UA, Python, and Fluidit Heat 2.8. The Python service read PLC tags through OPC UA, checked data type, unit, completeness, timestamp, and engineering range, mapped inputs to Fluidit Heat parameters, executed the calculation, processed the results, and returned them to the controller. MATLAB/Simulink/Simscape provided an independent cross-check, while UPPAAL formally assessed control logic and timing constraints.

Reliability was supported by comparing two independent physical models, repeating SIL tests with an unchanged configuration, evaluating every detector on the same data split, formal verification, and laboratory testing of a measuring instrument. The applicability limits were also defined: the anomaly detectors were assessed on synthetic data, so industrial accuracy and actual fuel savings require a full heating-season archive and controlled field trials.

Dissertation content. The dissertation contains an introduction, four sections, a conclusion, 115 references, and four appendices, with 45 figures and 21 tables. Its logic proceeds from assessment of the sector to development of a digital twin, verification of mathematical and simulation models, and finally anomaly detection and substantiation of control decisions.

Content of Section One. Section One examines global and Kazakhstan heat-power trends, fourth- and fifth-generation district-heating concepts, network digitalisation, smart metering, and predictive control. It statistically analyses heat production, transmission, and consumption in Kazakhstan and systematises network wear, losses, repair volumes, the number of supply sources, and the state of measurement infrastructure.

The analysis shows that district-heating operation depends on outdoor temperature, building thermal inertia, demand schedules, pipeline condition, pump operation, and heat-source efficiency. Their combined influence changes the normal operating state over time. One fixed threshold therefore increases false alarms; temperature, pressure, flow, heat output, and equipment status must be evaluated jointly.

The literature review demonstrates the benefits of digital twins exchanging data with a physical object, adapting model parameters from measurements, identifying leakage through model residuals, applying machine learning, and using model-predictive control. However, a physical model, automation system, anomaly detector, and fuel-optimisation module are rarely integrated in a closed loop. This conclusion motivated the architecture developed in Section Two.

Digital-twin architecture in Section Two. Section Two develops a functional digital twin of a district heating system. The architecture includes the physical-process layer, PLC and SCADA, a Python integration and analytics service, an SQL archive, and the Fluidit Heat thermo-hydraulic core. A virtual PLC in PLCSIM Advanced performed the controller functions in the experiment. WinCC displayed parameters, commands, and execution status; OPC UA served as the primary real-time channel, and the archive supported long-term trends and future model training.

Two PLC data blocks supported bidirectional exchange: an output block contained operating parameters and an input block received model results. Python validated mandatory variables and mapped PLC tags to Fluidit Heat boundary conditions. After a successful calculation, pressure, temperature, flow, delivered heat, heat deficit, and network losses were returned to the controller. If data were incomplete, outside their engineering range, or inconsistent with process

constraints, the write operation was blocked and an error code was generated.

The information-exchange structure remains unchanged when a physical controller is replaced by a virtual PLC. Transition to an actual facility requires adaptation of controller addresses, the process-tag map, and network-model parameters. The architecture can therefore test dispatch commands, failure scenarios, and optimisation results without affecting operating equipment.

Computational and experimental results of Section Two. Six modes were investigated in Fluidit Heat: space heating, domestic hot water, combined heating and hot water, supply temperature reduced to 90 °C, differential pressure reduced to 3 bar, and demand increased to 120%. In the combined baseline mode, 2,369.84 kW was delivered to consumers, source output was 2,377.14 kW, flow was 12.288 L/s, minimum differential pressure was 4.0 bar, and network heat loss was 7.298 kW, with no heat deficit.

At a supply temperature of 90 °C, delivered heat remained 2,369.83 kW, flow increased to 13.664 L/s, and network loss decreased to 6.820 kW. At 120% demand, delivered heat was 2,843.79 kW, source output was 2,850.80 kW, and flow was 14.737 L/s; network loss was 6.996 kW and the heat deficit remained zero. Flow velocity on the selected critical route increased by approximately 19.2%, while specific pressure loss increased by 41.5%. Because absolute network losses remained similar, their relative share decreased from 0.307% to 0.245%. These results confirm that thermal and hydraulic constraints must be assessed together.

The Software-in-the-Loop cycle was executed ten times with the same configuration. Mean duration was 23.509 s with a standard deviation of 0.907 s. Fluidit Heat simulation accounted for approximately 98.5% of total time, while OPC UA reading, data validation and mapping, post-processing, and verified PLC write-back together accounted for about 1.5%. The principal performance opportunity therefore lies in model loading and simulation rather than communication.

Formal-verification results. UPPAAL represented the system as a network of parallel timed automata: Scenario, Plant, PLC, HMI, Archive, Integration, Twin, and Observer. The model covered sensor failure, loss of OPC UA, archive unavailability, Fluidit calculation-service failure, and a process fault. Symbolic timing parameters specified the command deadline, response-confirmation time, and upper limit of model computation.

Verification confirmed transition to a safe state, blocking of an optimised command when data were invalid, issue of a safe command after calculation or communication failure, and absence of deadlocks under the stated assumptions. This result concerns the logic and timing discipline of TIA Portal-Python-Fluidit Heat interaction rather than the statistical accuracy of anomaly detection. Because the formal model checks design bounds rather than measured seconds, it complements the SIL timing experiment.

Mathematical model in Section Three. Section Three formulates a model that minimises fuel consumption while meeting the specified heat demand and constraints on temperature, pressure, and service quality. The network is represented as a directed graph; the model describes mass balance at nodes, hydraulic resistance and heat loss in pipelines, energy balance at mixing points, enthalpy difference at substations, and the relation between useful source heat and fuel input. Pump speed, valve resistance, network-water temperature, pressure, and consumer load are treated as controlled or constrained variables.

The optimisation criterion reduces total fuel input while accounting for source efficiency and fuel calorific value. Constraints include satisfying heat demand, maintaining an allowable supply-temperature range, minimum differential pressure, pipeline velocity, pump and valve operating regions, and zero heat deficit. Physical faults are represented by additional leakage or changes in resistance, thermal conductance, and equipment efficiency; measurement faults are modelled separately as bias and noise. This separation prevents a sensor fault from being automatically equated with a process failure.

Cross-verification of the Section Three models. An independent MATLAB/Simulink/Simscape model represented pipelines, heat-carrier flow, pressure losses, heat transfer, and consumer loads. It was compared with Fluidit Heat at 100% and 120% loads. The relative discrepancy in delivered heat did not exceed 0.003%, and the heat-source output discrepancy remained below 0.025%. MAPE for flow and heat-carrier velocity was 1.92-1.99%, while the discrepancy in consumer differential pressure was below 0.25%.

The MATLAB result for actual pressure loss was approximately 5.6% below Fluidit Heat. MATLAB network heat losses were 6.21% higher at 100% load and 10.81% higher at 120% load. The differences were attributed to heat-transfer formulation, soil temperature, and parameter alignment. The comparison confirmed agreement in the principal hydraulic and energy indicators while identifying the need to calibrate heat-transfer parameters before industrial use.

Anomaly-detection methodology in Section Four. Section Four augments the digital twin with six algorithms: Z-score, Isolation Forest, One-Class SVM, Local Outlier Factor, Elliptic Envelope, and an MLP autoencoder. The observational unit was one consumer's state at a 15-minute interval. The feature vector comprised space-heating load, domestic-hot-water load, supply and return pressure, supply and return temperature, and volumetric flow. A separate model was trained for each consumer, and features were standardised using training-period statistics only.

A synthetic 30-day series was created for five consumers: 14,400 observations in total, including 10,080 normal training observations and 4,320 test observations. The test set contained 669 controlled anomalies: supply temperature reduced to approximately 90 °C, differential pressure reduced to 3 bar, load increased to 120%, simulated leakage, and sensor failure. Increased demand was marked as a statistical deviation but was not automatically treated as a process violation.

Z-score used a threshold of three standard deviations; Isolation Forest used 250 trees and contamination 0.03; One-Class SVM used an RBF kernel and nu 0.03; LOF used 35 neighbours; and Elliptic Envelope used support fraction 0.90. The MLP autoencoder had a 7-14-4-14-7 architecture, tanh activation, and early stopping. Except for Z-score, thresholds were set at the 97th percentile of training scores. The final label required agreement among at least four algorithms.

Anomaly-detection results. On 4,320 test observations, majority voting achieved a precision of 92.26%, recall of 98.06%, F1 score of 95.07%, and ROC AUC of 99.58%, with 55 false positives and 13 false negatives. The best individual model, Elliptic Envelope, achieved F1 of 91.66%, so the ensemble improved F1 by 3.41 percentage points. F1 was 91.06% for Z-score, 90.34% for the MLP autoencoder, 90.23% for LOF, 89.68% for One-Class SVM, and 73.11% for Isolation Forest. The different mathematical principles complemented one another and reduced excessive alarms from sensitive models.

The confusion matrix showed correct classification of 3,596 normal and 656 anomalous observations. In an industrial system, the safety cost of 13 missed events and the operator workload created by 55 false alarms must be assessed separately. The model decision is therefore not converted directly into an automatic command. At the initial stage it acts as an advisory signal, after which a Fluidit Heat scenario evaluates the possible fault's effects on pressure, temperature, flow, and heat deficit.

Heat-energy metering results. Section Four also develops a heat-energy calculation algorithm that accounts for summer and winter periods and for space-heating and domestic-hot-water modes. A Saiman-AUES electromagnetic heat-meter prototype was manufactured and tested for flow and supply and return temperatures. The reference volume was 180.960 m³ and the meter reading was 181.101 m³, giving a difference of 0.141 m³, which did not exceed the permissible limit. The measuring-instrument type was approved by certificate No. 2459 dated 28 December 2023.

Scientific contribution. First, the dissertation develops a heterogeneous software and hardware coordination method that combines OPC UA exchange, Python orchestration, Fluidit Heat calculation, and UPPAAL verification within one Software-in-the-Loop environment. It provides structured bidirectional exchange, scenario execution, result validation, and safe-command generation.

Second, it proposes a hybrid anomaly-detection approach that links physical-model results and autoencoder reconstruction error through model residuals. The method separates a measurement-channel fault from a process change and checks a statistical alert through a physical scenario.

Third, it substantiates a computational fuel-minimisation model that allocates heat load under technological constraints while accounting for source efficiency, fuel calorific value, and network hydraulic losses. Fourth, it develops a numerical cross-verification method for Fluidit Heat and MATLAB/Simulink/Simscape using heat output, flow, velocity, pressure, and heat-loss indicators.

Practical significance, reliability, and limitations. The practical outputs comprise a digital-

twin environment, a WinCC dispatch interface, a Python integration service, anomaly-detection software, a heat-energy metering algorithm, and a laboratory heat-meter prototype. Selected results were used by Saiman Corporation LLP in development and laboratory work. The environment allows an operator to compare modes, test a safe scenario, and assess deviations between measurements and the model.

The anomaly metrics apply to synthetic data and do not constitute proof of industrial diagnostic accuracy. Deployment requires at least one complete heating-season archive, sensor-quality metadata, a log of confirmed failures and operator actions, and temporally separated training, validation, and independent test periods. Actual fuel savings must be assessed by controlled comparison of baseline and modified control under the same outdoor conditions and heat demand.

Author contribution and dissemination. The author formulated the aim and tasks, analysed data, developed the digital-twin architecture and software integration, built and compared the Fluidit Heat and MATLAB/Simulink/Simscape models, conducted SIL tests, prepared the UPPAAL model, programmed and evaluated the anomaly detectors, and interpreted the results. The research was conducted within project AP25795155.

Six scientific papers were published on the dissertation topic, including one in a Scopus-indexed journal and five in journals recommended by the authorised committee; a paper was also published in international conference proceedings. Two copyright certificates were obtained, and utility-model patent No. 9014 was granted for the Hardware and Software Complex for Intelligent Heat Metering.

Conclusion. The results demonstrate the technical feasibility of integrating a physical district-heating model, automation components, and intelligent data analysis within one environment. The solution can flag an anomaly early, model its effects, and verify a control action before it is transmitted to the physical facility. The next stage is independent temporal validation on industrial archives and determination of actual fuel savings in controlled field trials.