2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Stochastic Optimal Control for Nonlinear Systems based on Sampling & Deep Learning

Andreas Bott, Kirill Kuroptev, and Florian Steinke

***Abstract*— Stochastic optimal control requires an adequate representation of state uncertainty. For stochastic nonlinear systems, the probability distribution over the states given measurements can often not be represented in closed form. In this paper, we thus propose to address this control task based on Monte-Carlo sampling, integrating the state estimation step with stochastic gradient descent-based control optimisation. A deep neural network approximation of the nonlinear system is the key to speeding up both parts. We motivate and demonstrate the approach for district heating systems, where the security of supply shall be guaranteed with high probability in the face of uncertain consumer demands. Our conceptually simple approach enables representing multimodal distributions and achieving computation times feasible for the online operation of district heating systems.**

***Index Terms* – Nonlinear systems, Stochastic optimal control, Neural networks**

I. Introduction

District heating systems (DHS) play a key role for the decarbonisation of the heating sector. Traditional DHS are dominated by single thermal power plants, but modern DHS contain multiple heat generators. This raises many new challenges including the online control of the generators to ensure the security of supply for all customers. Specifically, sufficiently high consumer supply temperatures have to be guaranteed with high probability while keeping the grid temperatures overall as low as possible to minimise losses and maximise the efficiency of heat pump-based generators [1]. This control task is non-trivial since the DHS temperatures depend strongly nonlinearly on the current heat demands and these are typically not known to the operator. The posterior probability distribution over the current supply temperatures given typical measurements at the heat generators, can therefore often not be represented well by closed form distributions such as the Gaussian distribution. It may even be multimodal in important, realistic application cases.

Previous work on the control of DHS has, for example, employed a nonlinear dynamic model of the hydraulic part of DHS for a distributed control scheme to drive generators and storages to predefined reference values while guaranteeing closed-loop system stability [2]. Determining the optimal generator values in DHS is often based on model predictive control (MPC) and linearisation [3], [4] or piecewise linearisation [5]. Demand uncertainty is dealt with via robust optimisation [3], stochastic optimisation [4], or the certainty equivalence principle [5]. Note that [5] assumes that a reference trajectory for the generator set points is already available from day-ahead planning and focuses on adapting these schedules online to optimally react to the current grid situation; a setup, we also consider in this paper. For the related control problem of heating, ventilation, and air conditioning in buildings with uncertain thermal loads, discrete Markov decision processes can be used to compute optimal control laws [6]; however, discretisation becomes impractical for DHS with many consumers and thereby exploding state space dimension.

More generally, our work falls into the large group of approaches for the optimal control of nonlinear systems affected by uncertainties [7] and learning-based MPC [8]. It shares some similarity to scenario-based MPC [9] but more tightly integrates state estimation and stochastic optimisation. During state estimation for the nonlinear system, it does not target a Gaussian distribution like, e.g., the unscented Kalman filter [10], but uses a sampling-based approach. Compared to sequential Monte-Carlo approaches [11], it integrates the sampling with stochastic optimisation and approximates a known, complex physical system model by a deep neural network (DNN) for speed purposes; a currently, common idea in various disciplines, e.g. [12], [13].

Motivated by the online control of DHS, we develop a novel approach for the stochastic optimal control (SOC) of nonlinear systems. The method is specially tailored to situations where the current state distribution cannot be described accurately by any parametric distribution. The key idea is to integrate state estimation and optimal control based on Monte-Carlo sampling and deep learning: probability distributions are represented via samples and the system model is approximated with an DNN. The latter allows the fast evaluation of the nonlinear system model and its derivative during stochastic gradient descent-based optimisation. The key contributions of our approach are that it

- circumvents the necessity for system linearisation or parametric probability distributions,
- is able to represent all kinds of probability distributions over states, including multimodal ones,
- has low computation times for evaluating the approximated nonlinear system model as well as for computing gradients.

Unlike classic approaches, we deliberately neglect the separation principle between state estimation and control optimisation: as we perform stochastic gradient descent, the esti-

A. Bott, K. Kuroptev and F. Steinke are with the Department of Electrical Engineering, Technical University of Darmstadt, Darmstadt Germany andreas.bott@eins.tu-darmstadt.de

This research was funded by the German Federal Ministry of Transport and Digital Infrastructure in project CO2030 (grant number 19OS22005D). It was performed in the context of the LOEWE center emergenCity.

mated posterior distribution in each iteration of the algorithm does not have to be perfect; only the derived gradients of the optimal control problem have to point in the right direction. This allows us to significantly reduce the required sample size and, thus, the computation times. All these benefits are demonstrated for a small, but relevant and challenging DHS example.

The remainder of this paper is structured as follows: Section II formally introduces the problem setting and Section III describes our solution approach. A case study for DHS is presented in Section IV, before concluding in Section V.

II. Problem Setting

The goal of this work is to propose an approach that allows solving the SOC problem for nonlinear systems of the form

$$\mathbf{x} = \mathbf{h}(\mathbf{u}, \mathbf{d}), \tag{1}$$

where $\mathbf{x}$ is the system state, $\mathbf{u}$ are the control inputs, and $\mathbf{d}$ are unknown disturbances. The formulation presented in this paper assumes that the application can be described by a steady state model: $\mathbf{u}$, $\mathbf{d}$, and $\mathbf{x}$, thus only cover one time step. This applies to two types of systems: first, for systems whose dynamics are either much faster, than the control intervals, such that a steady-state is reached in each time step. Second, for systems whose deviations from a steady-state are small compared to the uncertainties in the system. If dynamics were to be incorporated, the state and control variables could be extended to cover several time steps. We also assume that the plant $\mathbf{h}(\cdot)$ is deterministic and bijective, i.e., the same input always yields the same output and every output is unique to one input. We show in Section IV that both assumptions holds for the DHS use case.

The system's state $\mathbf{x}$ is observed via measurements $\mathbf{y}$ that depend probabilistically on the system state $\mathbf{x}$. The measurement is assumed to be partial in the sense that it does not uniquely determine the system state but only yields some information about it. The state estimation task is thus addressed in a probabilistic Bayesian way: given prior knowledge $p(\mathbf{d})$ about the uncertain factors and a known measurement likelihood $p(\mathbf{y}|\mathbf{x})$, the posterior $p(\mathbf{d}|\mathbf{y}, \mathbf{u}_{-})$ for the unknowns given the measurements and the control signal prior to optimisation $\mathbf{u}_{-}$ is expressed via Bayes' theorem as

$$p(\mathbf{d}|\mathbf{y}, \mathbf{u}_{-}) \propto p(\mathbf{y}|\mathbf{h}(\mathbf{d}, \mathbf{u}_{-}))p(\mathbf{d}). \tag{2}$$

This holds since $\mathbf{h}(\cdot)$ is bijective and the probability of state $\mathbf{x} = \mathbf{h}(\mathbf{d}, \mathbf{u}_{-})$ and the corresponding disturbance $\mathbf{d}$ can thus be identified.

The optimised control input $\mathbf{u}$ shall minimise the SOC problem

$$\begin{aligned} \min_{\mathbf{u}} \quad & \mathbb{E}_{\mathbf{d}|\mathbf{y},\mathbf{u}_{-}}[c(\mathbf{u}, \mathbf{w}, \mathbf{h}(\mathbf{u}, \mathbf{d}))], \\ s.t. \quad & \underline{\mathbf{u}} \leq \mathbf{u} \leq \bar{\mathbf{u}}, \\ & P_{\mathbf{d}|\mathbf{y},\mathbf{u}_{-}}[\underline{\mathbf{x}} \leq \mathbf{h}(\mathbf{u}, \mathbf{d}) \leq \bar{\mathbf{x}}] \geq 1 - \delta, \end{aligned} \tag{3}$$

where $\delta \geq 0$ is a, typically small, risk-controlling parameter, $\mathbb{E}[\cdot]$ the expectation and $P[\cdot]$ the probability that the condition holds. $c(\mathbf{u}, \mathbf{w}, \mathbf{x})$ denotes the cost for a certain control

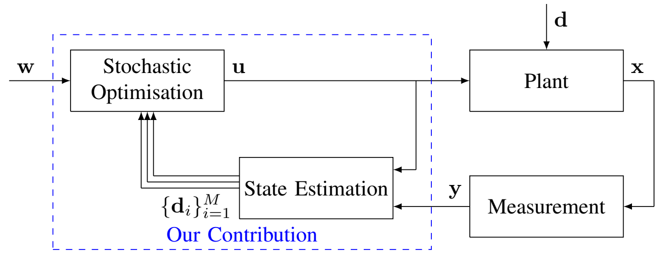


Fig. 1: System setup. The utilisation of a DNN in a combined state estimation and stochastic optimisation enables a sample-based uncertainty representation at acceptable compute times.

input and state, and also depends on a given reference value $\mathbf{w}$ for $\mathbf{u}$. $\underline{\mathbf{u}}$, $\bar{\mathbf{u}}$, $\underline{\mathbf{x}}$ and $\bar{\mathbf{x}}$ denote the minimal and maximal values for the control input and the system state, respectively.

III. Proposed Approach

A common approach for solving the estimation step (2) and the optimisation step (3) is to assume that all involved (conditional) probability distributions are Gaussian, the system model $\mathbf{h}(\cdot)$ is linear, and the cost function $c(\cdot)$ is convex quadratic. In this case, the posterior $p(\mathbf{d}|\mathbf{y}, \mathbf{u}_{-})$ is also Gaussian and the optimisation problem (3) becomes a convex quadratic problem, or a convex second order cone problem if constraints with $\delta > 0$ are considered [14].

For nonlinear system models $\mathbf{h}(\cdot)$, however, the posterior distribution does not in general belong to any parametric class of distributions, even if the prior distribution and measurement likelihood are assumed to be Gaussian; the experimental section shows an example thereof. The expectations and chance constraints of the SOC problem can then not be expressed and optimised in terms of the distribution parameters. Therefore, the key idea of this paper is to represent the posterior distribution via a flexible Monte-Carlo sampling procedure. The SOC problem can then be written in terms of these samples using the sample average approximation (SAA) [14] and the optimisation can be performed via (stochastic) gradient descent. Figure 1 schematically illustrates our control approach: the state estimator generates a set of posterior samples $\{\mathbf{d}_i\}_{i=1}^{N}$ based on the control input before optimisation $\mathbf{u}_{-}$ and the measurement $\mathbf{y}$.

State estimation (2) and stochastic optimisation (3) based on samples requires many evaluations of $\mathbf{h}(\cdot)$ and its derivative. In the case of heating grids, $\mathbf{h}(\cdot)$ is defined implicitly as the solution of a set of nonlinear, hydraulic and thermal equations; solving this set of equations is done via iterative algorithm and takes considerable time [15]. In order to significantly speed up the process, we propose to train a DNN $\mathbf{h}_{\boldsymbol{\theta}}(\cdot)$ to approximate the system model $\mathbf{h}(\cdot)$, i.e.,

$$\mathbf{h}_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{d}) \approx \mathbf{h}(\mathbf{u}, \mathbf{d}) = \mathbf{x}, \tag{4}$$

for all relevant $\mathbf{u}$, $\mathbf{d}$. Determining optimal parameters θ, i.e., training the DNN can be done in advance and, therefore, does not influence the online computation times. Evaluating the DNN and its derivative, however, can be performed many orders of magnitude faster than evaluating the original system model, at least for the DHS use case. Practically, the DNN approximation is the key step to enable the sampling-based approach in the first place.

The SOC controller then solves (3) in a SAA-based fashion using these samples. In the following, we detail each step. An overview of the whole process is given in Figure 2

A. Learning the DNN model approximation

To learn $\mathbf{h}_{\boldsymbol{\theta}}(\cdot)$ we generate a training set of N^{train} training samples by sampling $\mathbf{d}_i$ from the prior $p(\mathbf{d})$. The corresponding control inputs are sampled to cover the range between the control limits $\underline{\mathbf{u}}, \bar{\mathbf{u}}$ and slightly beyond. We then evaluate $\mathbf{h}(\cdot)$ for each sample. For DHS, $\mathbf{h}(\cdot)$ is defined implicitly by a set of nonlinear equations. By exploiting the form of these equations, the generation of training samples can be made very efficient [16]. Afterwards, we use standard least squares DNN training.

B. State estimation via SIR

The goal of the state estimation is to gather N samples $\{\mathbf{d}_i\}_{i=1}^N$ that represent the posterior probability distribution (2). This can be achieved by different Monte Carlo (MC) algorithms; we use sampling importance resampling (SIR) [17]. To this end, we first draw $M \gg N$ samples $\{\mathbf{d}'_i\}_{i=1}^M$ from the prior distribution $p(\mathbf{d})$ and weight each sample by its approximated likelyhood

$$\omega'_i = p(\mathbf{y}|\mathbf{h}_{\boldsymbol{\theta}}(\mathbf{u}_{-}, \mathbf{d}'_i)) \approx p(\mathbf{y}|\mathbf{h}(\mathbf{u}_{-}, \mathbf{d}'_i)). \tag{5}$$

Using the DNN approximation and batch processing, these weights can be computed very fast, as demonstrated below. This set of weighted samples represents the posterior distribution well, however, since M is large and many weights are close to zero, we reduce the number of samples for the following processing steps by resampling. Hence, we draw N samples with replacement form this weighted set, where the probability for drawing each sample is given by its normalised weight $\omega_i = \omega'_i / \sum_j \omega'_j$. Note that during SOC optimisation only the gradient derived from the sample set is required to point in the right direction, but an exact representation of the posterior distribution is not needed.

C. Stochastic Optimisation

The SOC problem (3) can be expressed using the generated samples via SAA principles, i.e., replacing population-based measures with their equivalents for the finite sample. Again, $\mathbf{h}(\cdot)$ is replaced by $\mathbf{h}_{\boldsymbol{\theta}}(\cdot)$ in (3) to speed up the calculations. However, the resulting optimisation problem remains nonlinear and non-convex. Depending on the tightness of the bounds and the variance of the inferred disturbance samples, there might be no control input $\mathbf{u}$ that satisfies the boundary constraints for all realisations of $\mathbf{d}$ or there may be multiple locally optimal solutions.

Finding the global optimum of large, non-convex optimisation problems is mostly very difficult and takes considerable computation time. Solving (3), therefore, remains impractical

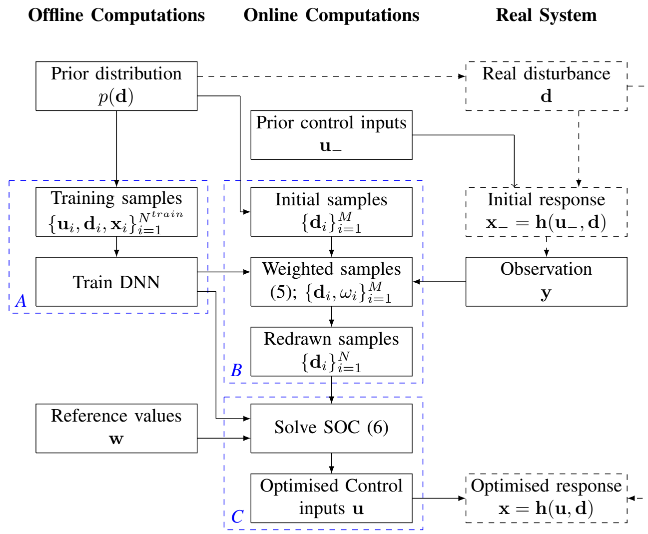


Fig. 2: Our approach can be separated into computations which are done offline (left column) and those which are done online (middle column). Only the online computations depend on the observations gathered from the real system (right column). Dashed elements indicate, that they can not directly be accessed by the controller. The blue boxes mark the parts of the algorithm detailed in the respective subsections III-A - III-C.

for an online controller, even after integrating the DNN approximation. However, considering that the primary goal of the controller is to keep the system's state within predefined bounds and that a reference trajectory exists, we assume that two trade-offs are made: first, having a low computation time and a high probability for a valid system state is valued higher than finding the globally optimal solution.

Second, having a small risk of violating the boundary conditions, especially regarding the system's state, is preferred over not making any decision at all. We thus conclude that, for many practical applications, it is sufficient to find a local optimum of the relaxed optimisation problem

$$\begin{aligned}\min_{\mathbf{u}} \frac{1}{N} \sum_{i=1}^{N} \quad & \Big[c(\mathbf{u}, \mathbf{w}, \mathbf{h}_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{d}_i)) \\ & + \lambda \|\max(\bar{\mathbf{x}} - \mathbf{h}_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{d}_i), 0)\|_2^2 \\ & + \lambda \|\min(\mathbf{h}_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{d}_i) - \underline{\mathbf{x}}, 0)\|_2^2 \Big] \\ + \lambda \|\max(\mathbf{u} - \bar{\mathbf{u}}, 0)\|_2^2 & + \lambda \|\min(\mathbf{u} - \underline{\mathbf{u}}, 0)\|_2^2,\end{aligned} \tag{6}$$

where λ is a hyperparameter balancing cost-optimality versus constraint satisfaction and $\|\cdot\|_2^2$ denotes the squared Euclidean norm. This problem can be solved using off-the-shelf stochastic gradient descent algorithms.

IV. Numerical Case Study for DHS

Our test case represents a typical, important new challenge in modern DHS with several decentral heat sources. The grid is supplied by two heat generators from different ends, see Figure 3a. The supply temperatures in the grid depend strongly non-linearly on the exact demand and supply conditions and are in danger of being too low. Even though the number of nodes is small, the example captures a lot of the key challenges of this novel control task as we discuss in detail after introducing the setup. We then describe our implementation and three baseline models and demonstrate that the proposed approach achieves good control performance while being sufficiently fast for this application.

A. Setup

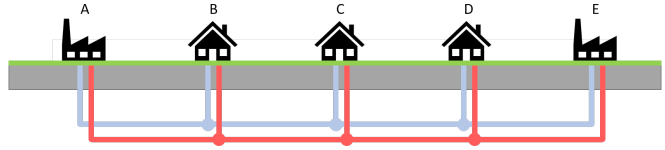
Taking the role of the DHS operator, the task is to control the heating plants A and E such that the consumers B, C, and D are supplied with a sufficiently high temperature without using a too high temperature overall. The heat demands are unknown to the operator, who can only measure at the heating plants.

We use a set of steady-state physical grid equations to implicitly define the system model $\mathbf{h}(\cdot)$ of the DHS [15]. The system's state $\mathbf{x}$ consists of all temperatures, pressures, and mass flows in the grid. It is the only solution of the steady state equations given the unknown heat demand $\mathbf{d}$ and the manipulated variables $\mathbf{u}$. The manipulated values $\mathbf{u}$ are the supply temperatures of plants A and E as well as the supplied heat of plant E. For these values and the heat output of plant A a day-ahead schedule $\mathbf{w}$ is assumed to be available. Plant A serves as a slack generator, required to ensure the energy balance; its heat output is a flexible consequence of the remaining system conditions. The mass flow at plant A and the return temperatures at both plants are the measurements $\mathbf{y}$ available to the operator. Fixed system parameters are the supply and return pressures at plant A and the return temperatures of the consumers.

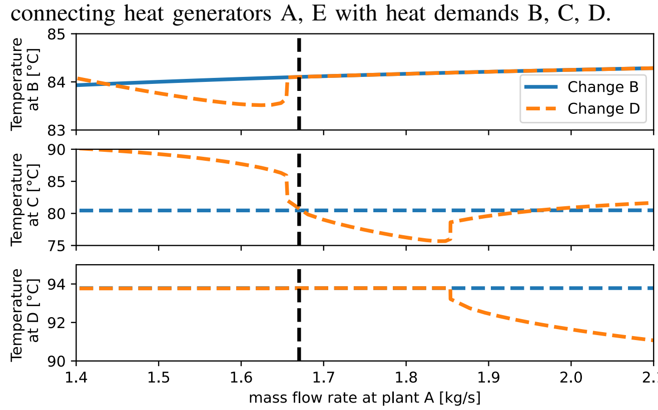
In the test case, the demands' thermal power $\mathbf{d}$ is modelled as a zero-truncated multivariate normal distribution, $p(\mathbf{d}) = N^{+}(\mu_{\mathbf{d}}, \Sigma_{\mathbf{d}})$ with mean $\mu_{\mathbf{d}}$ and covariance matrix $\Sigma_{\mathbf{d}}$. The mean consumption of the middle node C is much smaller than the mean demands at B and D. The demands at B and D are negatively correlated to each other, and independent of the demand at C.

The cost function $c(\cdot)$ penalises the deviation of $\mathbf{u}$ from the corresponding entries of the day-ahead schedule $\mathbf{w}$ as well as the deviation of the slack plant's power output to the corresponding day-ahead reference. Limits for the heat power output of the heat generators are defined relative to the day-ahead schedule and in absolute terms for their supply temperatures. The default, non-optimised input $\mathbf{u}_{-}$ is set equal to $\mathbf{w}$. A minimum supply temperature of $80\,°\mathrm{C}$ is defined for all consumer nodes. The exact model equations are presented in [15]. The equations and all parameter values are also found in our open-source implementation [1].

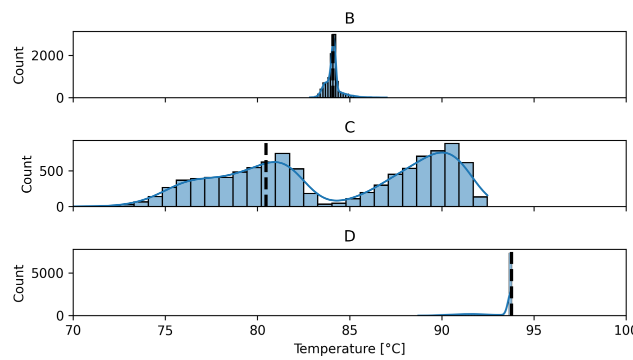
[1] The code of our implementation is available at https://github.com/EINS-TUDa/DNN_SOC4DHS



(a) The test DHS consists of supply (red) and return (blue) pipes connecting heat generators A, E with heat demands B, C, D.



(b) Supply temperatures at nodes B, C, D over the measured mass flow at plant A, using the day-ahead schedule $\mathbf{w}$ as the control input. The mass flow at plant A can be varied by changing the heat demand at B or D, respectively. The dashed black line denotes the flow measurement for the mean a priori demand.



(c) Histogram and kernel density estimate of the supply temperatures at nodes B, C, and D, derived from 10000 samples from the prior distribution over heat demands and the day-ahead schedule $\mathbf{w}$. The dashed black line denotes the temperatures for the mean a priori demand.

Fig. 3: The considered DHS test case.

B. DHS specific goals & challenges

A key goal of DHS control is ensuring a sufficiently high supply water temperature at the inlet of all heat demands. To understand the resulting operational challenges the physical processes in DHSs can be separated into two parts: fast hydraulic processes, which determine the flow directions in all pipes, and slower thermal processes, which describe the temperature losses. The hydraulic system reaches a steady state within few seconds, whereas the thermal system may still be in transition for much longer. While steady-state thermal consequences of changing flows can be large and non-linearly complex, the impact of transients in thermal state is typically limited, as argued in the Appendix of

[15]. Therefore, the state uncertainty can be represented reasonably well using a steady-state model and the proposed approach is suitable for the task of online control of DHSs.

As shown in Figure 3b the following operation challenge arises for the test grid in the face of unknown heat demands: the day ahead schedule is selected such that both heating plants supply demand C, with a major share coming from generator E. If the thermal demand of consumer B changes, the flow directions in all pipes remain the same, as the additional power is provided by the slack supply in A. As a result, the temperatures remain mostly constant. If, on the other hand, the thermal power at demand D changes, the temperatures change strongly non-linearly. If the demand decreases, the flow direction in the pipes between demands B and C changes, indicated by a sudden temperature change for these demands. Demand C is now fully supplied from the plant at E and the flows from both supplies mix at demand B. Similarly, if the demand for D increases, the demand C is increasingly supplied from power plant A. This leads to a slowly reducing temperature at demand C until the flow directions in the pipes between C and D changes.

Figure 3c examines the consequences of these non-linearities on the supply temperature distributions when the heat demands are uncertain. For node C a multi-modal state distribution can be observed. Linear systems and Gaussian uncertainty models cannot represent this behaviour well.

C. Implementation

The used DNN for $\mathbf{h}_{\boldsymbol{\theta}}(\cdot)$ has four fully connected trainable layers, consisting of 200, 400, 400, and 74 neurons with a rectifier activation function in the first three layers and a linear activation function in the last layer. It is trained with 150000 training and 10000 validation samples. The SOC formulation (6) is solved locally using the Adam optimiser [18] with an initial learning rate of 0.1 and mini-batches of 10 samples each. The SIR based state estimation generates $M = 1000000$ initial demand samples and redraws $N = 10$, 100, or 1000 samples, respectively, which are then used in the optimisation step.

D. Baselines

We compare our approach against three baselines. The first baseline, the "Slack Controller" (SC), makes no adaptations to the day-ahead schedule; the slack plant reacts automatically, according to the varying demand situations and grid losses.

The second baseline approach, the "Linear Controller" (LC), is an elaborate linearised approach. It linearises the system model around the a priori most probable demand and neglects the truncation of the demand's uncertainty. This leads to a Gaussian prior distribution for the system state. The posterior distribution given the measurements can then be computed analytically [15]. To avoid matrix singularity during this step, a small value is added to the diagonal of the state's covariance matrix. The LC baseline resembles a classical Bayesian filter, which is equivalent to a Kalman filter for linear systems. We then draw random samples from

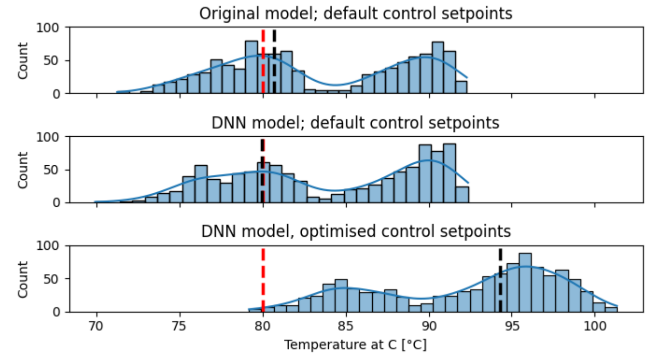


Fig. 4: Histograms of the posterior distribution of the supply temperature at node C based on 1000 samples, given exemplary demands and measurement values. (Top) using original system model $\mathbf{h}(\cdot)$ and non-optmised input $\mathbf{u}_{-}$, (middle) using approximate system model $\mathbf{h}_{\boldsymbol{\theta}}(\cdot)$ and non-optmised input $\mathbf{u}_{-}$, (bottom) using approximate system model $\mathbf{h}_{\boldsymbol{\theta}}(\cdot)$ and control input $\mathbf{u}$ optimised with the proposed approach. The red line denotes the minimum allowed supply temperature and the black line denotes the temperature given the (unknown) true underlying demand values.

the posterior state distribution, compute the corresponding demand values, and linearise the system model again around the most likely posterior demand estimate. The control inputs are optimised using these estimates and the second linearised model via stochastic gradient descent.

Last, we consider the hypothetical case where the operator has perfect knowledge of the demands (OPT). This approach serves as an optimality bound. To this end, we conduct a grid search for the optimal $\mathbf{u}$ where the plants' supply temperatures are changed in steps of $2.5\,°\mathrm{C}$ and the power output of E in $1.5\,\mathrm{kW}$ steps. To reduce the computational complexity, the grid search is pruned using various heuristics; nevertheless, the approach was only barely feasible for this small example. For larger grids with more degrees of freedom and for realistic settings with unknown demands, the grid search is impractical.

E. Numerical Results

First, we qualitatively access the performance of our proposed approach for an example demand setting. Second, we discuss the computation times and comparative performance of the results for randomised demand settings.

a) Qualitative Analysis: Figure 4 shows the posterior distributions for the supply temperature of demand C using the original system model $\mathbf{h}(\cdot)$ and its approximation $\mathbf{h}_{\boldsymbol{\theta}}(\cdot)$, given a measurement generated by simulating an arbitrarily chosen set of demand values. The two distributions visually match well. The finding that the supply temperature limit of $80\,°\mathrm{C}$ is violated with high probability for the non-optmised control inputs $\mathbf{u}_{-}$ is consistent in both evaluations. Using the optimised $\mathbf{u}$, the risk of violating the temperature requirements is significantly reduced.

Figure 5 shows for the same measurements the resulting

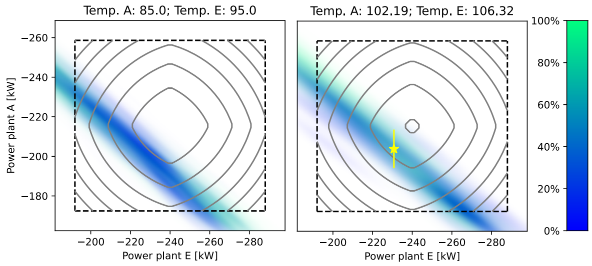


Fig. 5: The power provided by slack plant A depending on chosen power set point for plant E, for consumer demands following the posterior distribution over demands given the same measurement as in Figure 4; colour intensity encodes the probability of each combination. The green and blue colours indicate the posterior probability of the resulting grid state being feasible, given the measurement. Left: default supply temperatures of the heating plants, right: optimised temperatures. Optimisation of the cost function (isolines denoted grey) of the feasible control inputs (black dashed) yields the optimised heat power for plant E, shown yellow. The slack supply A remains uncertain and depends on the demands.

power output of slack power plant A given the set point for the power output of plant E, for two different supply temperature set points of the two heat generators, the default temperatures and the optimised ones. The combinations that, given the measured values, have a high probability of leading to feasible temperature form a complex non-convex set in this two dimensional slice of the full state space of the model. The set also depends on the supply temperatures of the heating plants. The optimised control inputs imply a high probability of the grid states being feasible while not deviating too much from the default reference values.

b) Quantitative Analysis: Truthfully representing the gradients of the original plant $\mathbf{h}(\cdot)$ is crucial for our plant approximation $\mathbf{h}_{\boldsymbol{\theta}}(\cdot)$ since they drive the employed stochastic gradient descent algorithm to optimise the control inputs. We examine the gradient of the supply temperature for demand C with respect to the three independent decision variables at the initial control set points as an example. We compute the gradients of the true plant $\mathbf{h}(\cdot)$ via the implicit function theorem and compare them with the gradients received by backpropagation from the DNN $\mathbf{h}_{\boldsymbol{\theta}}(\cdot)$. The length of the gradient vector is not relevant since the step-size of the gradient descent algorithm can be chosen accordingly. We thus compare the gradients by computing the scalar product between the normalised gradients, receiving a numerical value between -1, which indicates that the gradients point in opposite directions, and 1, if the gradient's directions align.

The quality of the gradients depends on the evaluated point, i.e., the demands' heat power values. We first draw 10000 random power values from the demands' prior distribution and compare the gradients to estimate an upper limit for the DNN's quality. We achieved an average value of 0.90 with an standard deviation of 0.20. Using only 100 samples, we obtain an average performance of 0.87 and a standard deviation of 0.25. This indicates, that 100 samples are sufficient to determine a gradient that is pointing in the right direction with high probability.

TABLE I: Numerical evaluation of proposed approach for 500 random demand situations. The evaluation step uses the true demands to determine costs and constraint violation.

	λ	# opt samples	mean cost	invalid states	invalid power A	total time [s]
SC	–	–	122	29.4%	5.6%	–
OPT	–	true $\mathbf{d}$	103	0%	0%	–
Ours	100	10	183	15.6%	0.4%	18.0
	100	100	182	11.8%	0%	13.5
	100	1000	189	10.4%	0%	33.9
	1000	10	249	9.0%	0.6%	19.9
	1000	100	278	3.6%	0%	21.3
	1000	1000	300	2.0%	0%	29.7
	10000	10	327	4.8%	0.8%	21.0
	10000	100	342	2.0%	0%	22.4
	10000	1000	376	0.4%	0%	31.4
LC	100	1000	678	16.2%	52.4%	15.2
	1000	1000	740	13.2%	53.4%	14.3
	10000	1000	742	14.0%	55.0%	14.5

Table I shows numerical indicators of our approach and the two baseline cases SC and LC. We generated 500 random realisation for the uncertain demand $\mathbf{d}$ and corresponding measurements $\mathbf{y}$ and used all considered methods to compute optimised control inputs. Using the true demands and the control inptus, we then computed the true grid state using $\mathbf{h}(\cdot)$ and verified whether the resulting grid state violated any constraints. The SC baseline approach reaches the lowest (re-)scheduling costs, as it only changes the power value of the slack heat supply to match the realised demands. However, this approach leads to high rate of violated grid states and requires the slack power plant to operate outside of its desired operation range occasionally. The LC baseline can adapt the temperature and power set points to reduce the risk of constraint violation. However, it is not capable of capturing the complex nonlinear system response and the chance of violating grid constrains remains high. Our approach is capable of capturing the nonlinearities and leads to a valid grid state in nearly all cases, at rescheduling costs below the linearised LC approach. The costs of our proposed approach are higher than lower bound given by the OPT approach. However, the OPT approach uses knowledge that is not realistically available.

As expected, increasing the value of λ increases the costs for rescheduling but reduces the risk of constraint violating for our proposed approach. Additionally, it can be observed that increasing the number of samples used for the optimisation has a similar effect at a smaller scale. If too few samples are used, no samples may violate the boundary conditions during optimisation, which leads to a less conservative rescheduling. For the LC baseline, no significant impact can be observed since the probability model is inadequate, independent of λ.

TABLE II: Computation times for 10000 Jacobian matrices.

Model:	Batch size	Total time [s]
true $\mathbf{h}(\cdot)$	1	19730
DNN model $\mathbf{h}_{\boldsymbol{\theta}}(\cdot)$	1	163
	10	32
	100	22
	1000	22

c) Speed: Table II reports the required times to compute the complete Jacobian matrix using the true state model $\mathbf{h}(\cdot)$ and the approximate DNN $\mathbf{h}_{\boldsymbol{\theta}}(\cdot)$, for varying batch sizes. The DNN model is about two orders of magnitude faster if each sample is computed individually and can be further sped up by one order of magnitude if batch processing is used. Note that the iterative process involved in evaluating $\mathbf{h}(\cdot)$ and the implicit computation of the gradients prohibit batch processing for the DHS use case. All reported computation times were derived on a laptop with an Intel i5-8265U CPU processor and 16 GB RAM.

The compute times for the complete optimisation process, as shown in Table I, imply that our approach is slower than the LC approach, but fast enough for the online control of DHS where time scales of minutes to quarters of hours are relevant. Additionally, our approach scales well with an increasing number of samples used during the optimisation, which might be important for larger DHS.

V. Conclusion

In this paper, we proposed a conceptual simple, integrated state estimation and control approach that is able to handle stochastic nonlinear system models. By approximating the plant by a DNN in order to speed up the evaluation of the model and of its derivatives we enable sampling-based MC state estimation and SAA-based algorithms for control optimisation. We demonstrated the applicability of this novel scheme for the online control of DHS. The model outperforms linear approximation methods due to its capability of representing complex posterior distributions, even multi-modal ones, without requiring a parametric representation. Even though the approach is based on sampling, it is fast enough for DHS online control due to the DNN model approximation.

Future work could extend the research by using a dynamic model for the DHS, which would increase the accuracy shortly after the new set points are applied to the heating plants and before a new steady state is reached. For significantly larger DHS more advanced MC algorithms, such as Markov Chain Monte Carlo algorithm may be preferable to generate samples more efficiently.

References

[1] A. Vandermeulen, B. van der Heijde, and L. Helsen, “Controlling district heating and cooling networks to unlock flexibility: A review,” *Energy*, vol. 151, pp. 103–115, 2018.

[2] J. E. Machado, M. Cucuzzella, N. Pronk, and J. M. A. Scherpen, “Adaptive Control for Flow and Volume Regulation in Multi-Producer District Heating Systems,” *IEEE Control Systems Letters*, vol. 6, pp. 794–799, 2022.

[3] G. Sandou, S. Font, S. Tebbani, A. Hiret, and C. Mondon, “Predictive Control of a Complex District Heating Network,” in *Proceedings of the 44th IEEE Conference on Decision and Control*, pp. 7372–7377, Dec. 2005.

[4] F. Verrilli, A. Parisio, and L. Glielmo, “Stochastic model predictive control for optimal energy management of district heating power plants,” in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 807–812, Dec. 2016.

[5] A. La Bella and A. Del Corno, “Optimal management and data-based predictive control of district heating systems: The Novate Milanese experimental case-study,” *Control Engineering Practice*, vol. 132, p. 105429, Mar. 2023.

[6] Y. Yang, G. Hu, and C. J. Spanos, “Stochastic optimal control of hvac system for energy-efficient buildings,” *IEEE Transactions on Control Systems Technology*, vol. 30, no. 1, pp. 376–383, 2022.

[7] A. Mesbah, “Stochastic model predictive control: An overview and perspectives for future research,” *IEEE Control Systems Magazine*, vol. 36, no. 6, pp. 30–44, 2016.

[8] L. Hewing, K. P. Wabersich, M. Menner, and M. N. Zeilinger, “Learning-based model predictive control: Toward safe learning in control,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, pp. 269–296, 2020.

[9] G. Schildbach, L. Fagiano, C. Frei, and M. Morari, “The scenario approach for stochastic model predictive control with bounds on closed-loop constraint violations,” *Automatica*, vol. 50, no. 12, pp. 3009–3018, 2014.

[10] E. A. Wan and R. Van Der Merwe, “The unscented kalman filter for nonlinear estimation,” in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, pp. 153–158, Ieee, 2000.

[11] O. Cappé, S. J. Godsill, and E. Moulines, “An overview of existing methods and recent advances in sequential monte carlo,” *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[12] X. Pan, T. Zhao, M. Chen, and S. Zhang, “Deepopf: A deep neural network approach for security-constrained dc optimal power flow,” *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 1725–1735, 2020.

[13] M. G. Alfarizi, M. Stanko, and T. Bikmukhametov, “Well control optimization in waterflooding using genetic algorithm coupled with Artificial Neural Networks,” *Upstream Oil and Gas Technology*, vol. 9, p. 100071, Sept. 2022.

[14] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.

[15] A. Bott, T. Janke, and F. Steinke, “Deep learning-enabled mcmc for probabilistic state estimation in district heating grids,” *Applied Energy*, vol. 336, p. 120837, 2023.

[16] A. Bott, M. Beykirch, and F. Steinke, “Efficient training of learning-based thermal power flow for 4th generation district heating grids,” 2024. (under review).

[17] D. B. Rubin, “The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The sir algorithm,” *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 543–546, 1987.

[18] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2017.