ETH zürich

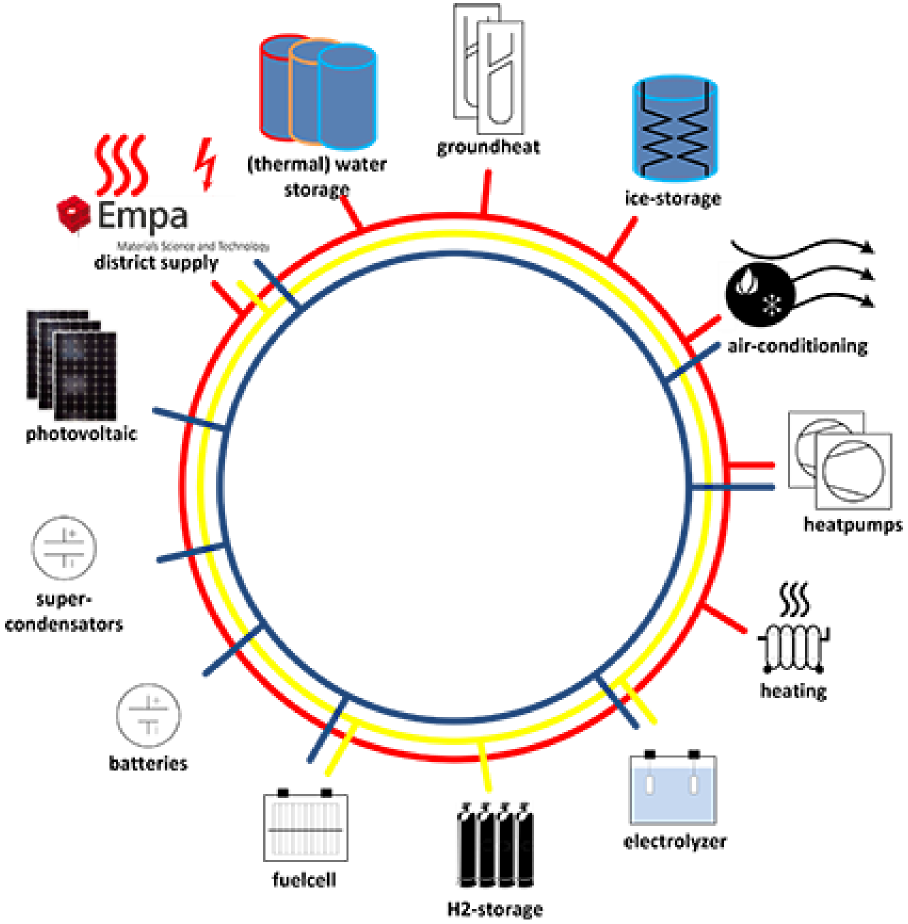
Data-Driven Adaptive Controller Parameterisation: A Bayesian Optimization Approach

Master Thesis Presentation

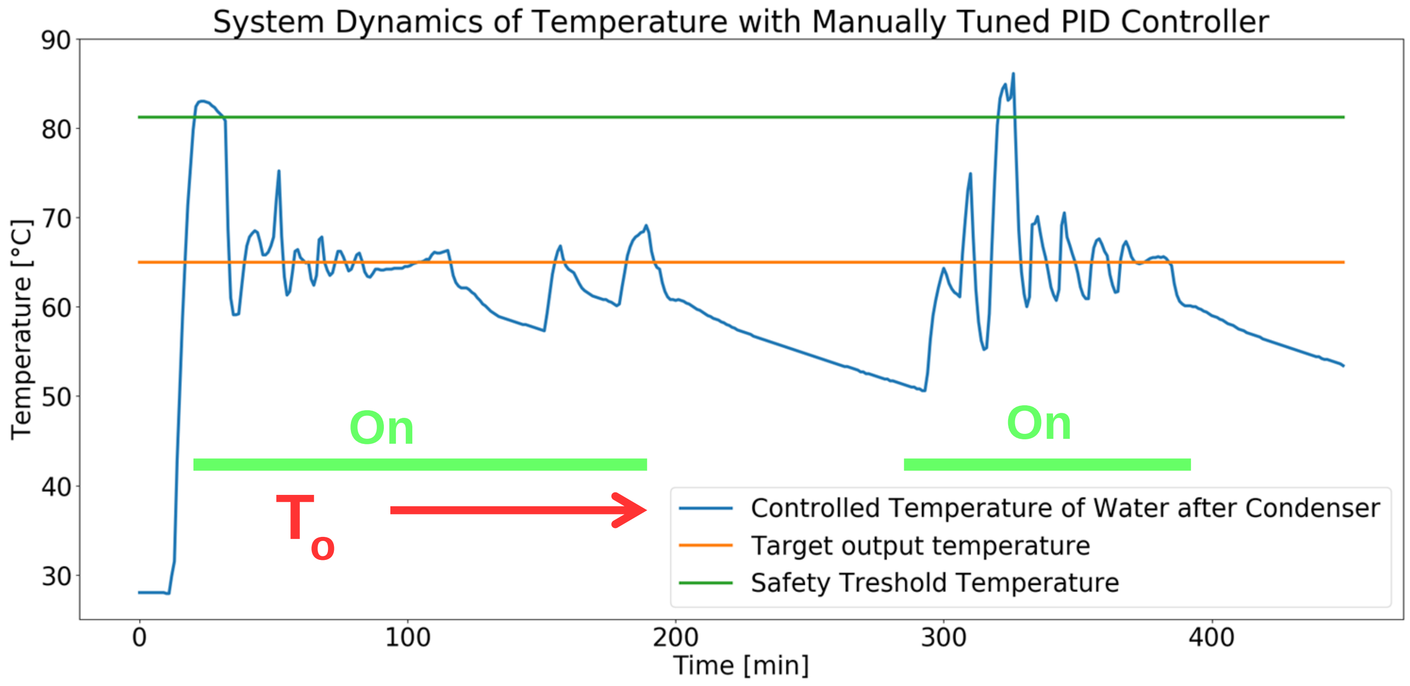
Nicolas Schmid

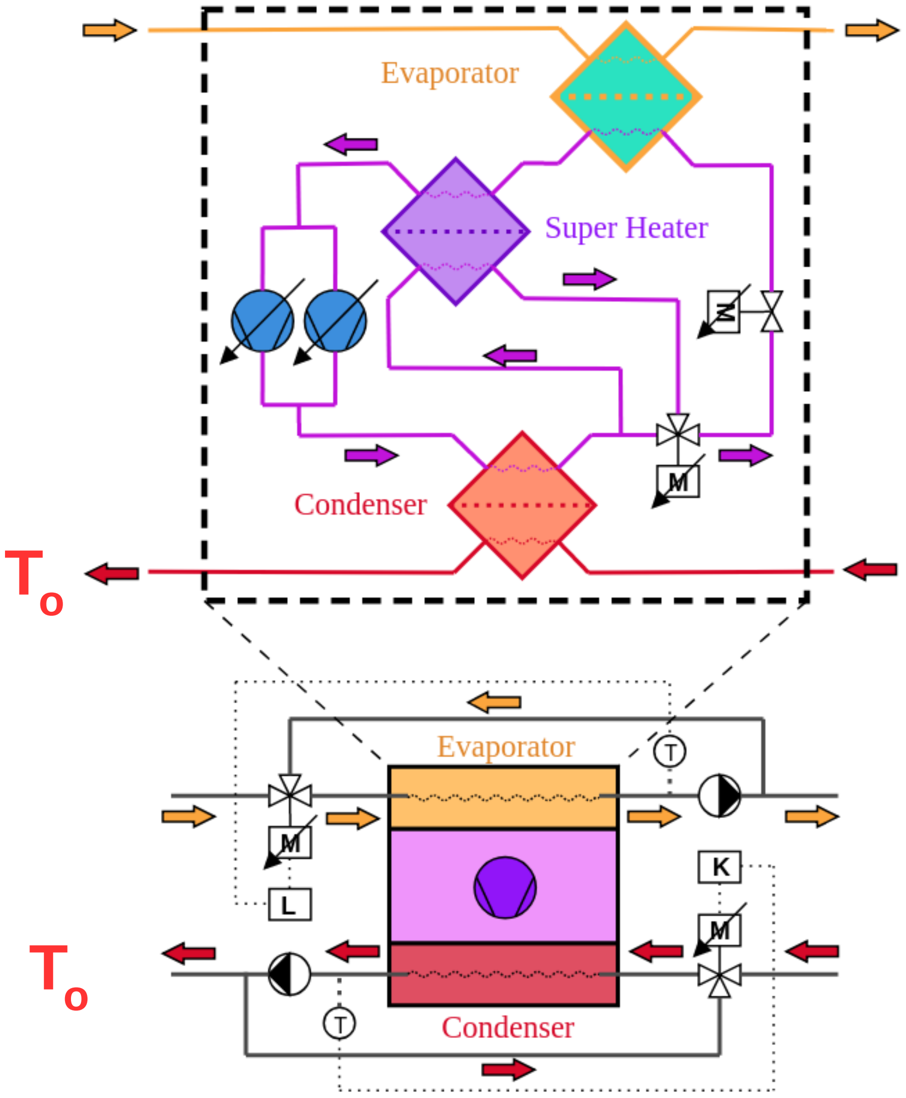
Supervision: Mohammad Khosravi, Philipp Heer (Empa), Annika Eichler, Roy Smith

The ehub at the Empa



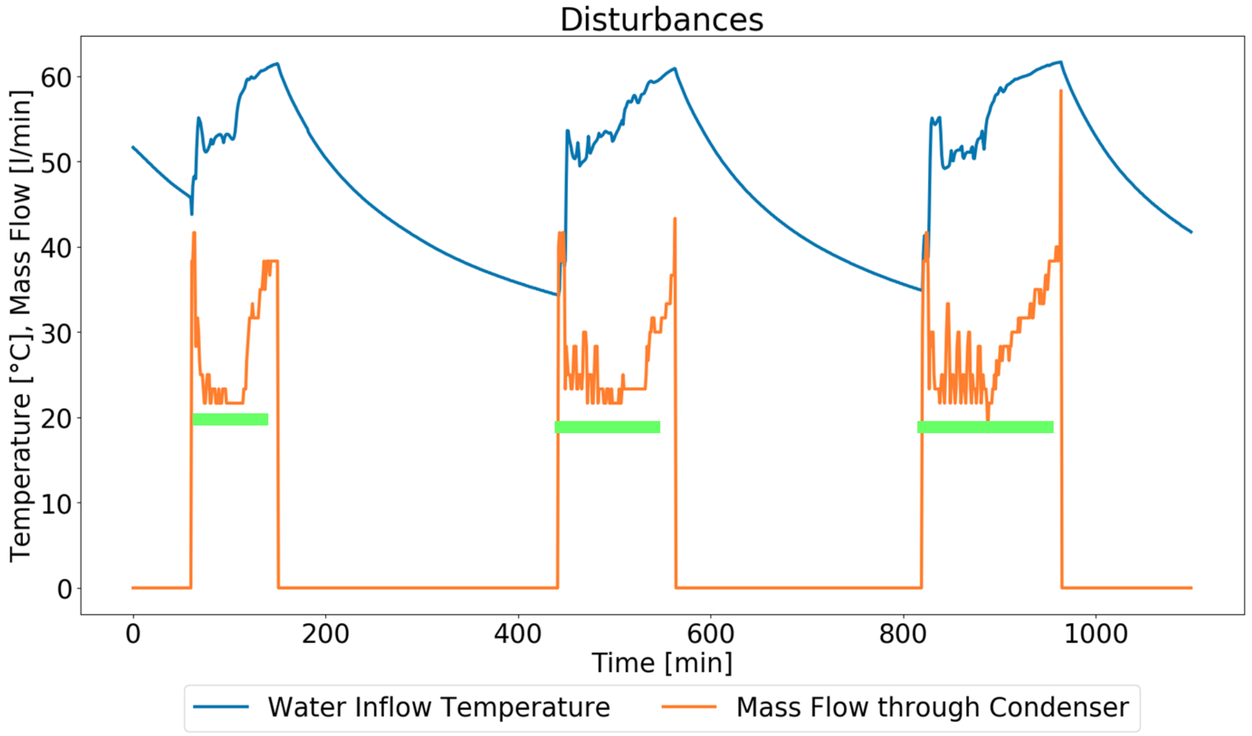
The High-Temperature Heat Pump



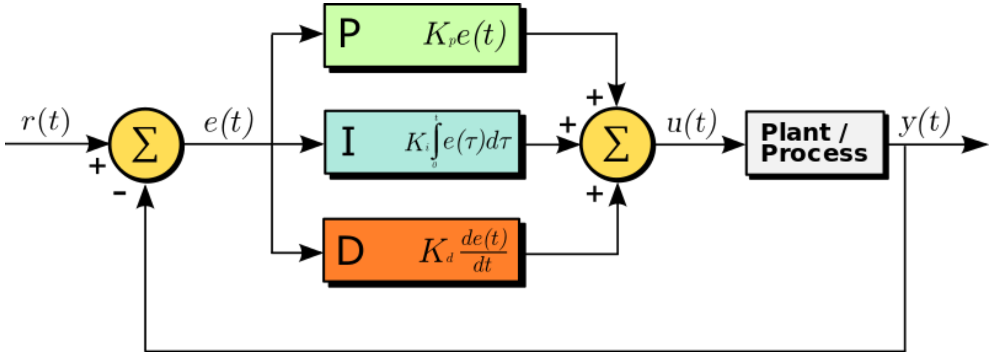


Reasons for Unsatisfying Performance

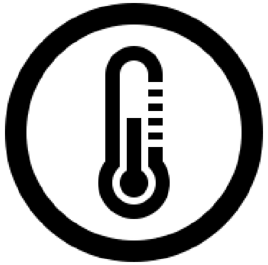
- Suboptimal parameterization of the PID controller
- Temperature challenging to control:
  - Strongly changing disturbances
  - Time delays
  - Non-linearities



Motivation

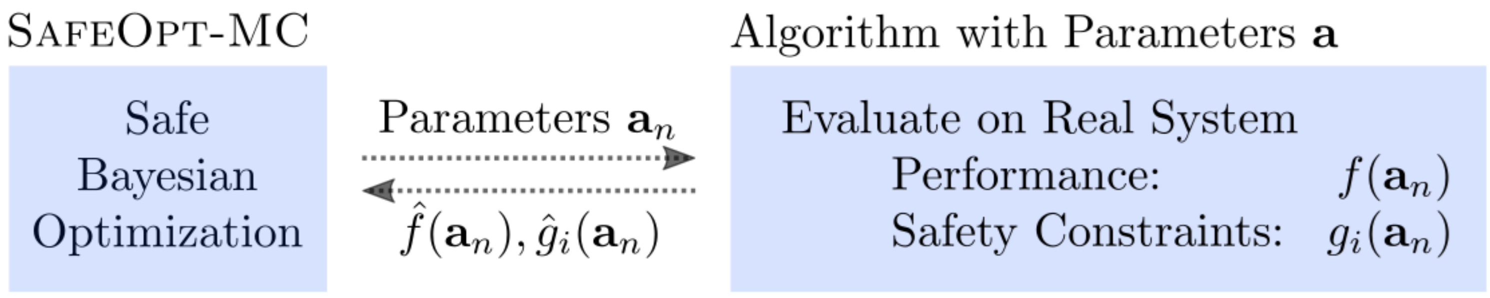


- 95% of Industrial Controllers are PID like
- Physical modelling and system identification time-consuming and costly
- Optimization without physical model or system identification unfeasible
- Leads to dangerous and non-optimal tuning
- But sensor data of dynamic systems, especially with „Industry 4.0“, frequently available



Solution: Safe Bayesian Optimization [Berkenkamp,2016]

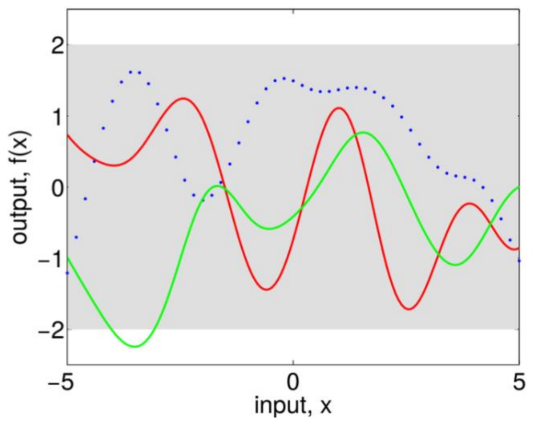
- Needs no physical modelling or system identification
- Guarantees safety with high probability
- Bounds for optimality



$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right), \text{whereas}$$
$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})],$$
$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x} - m(\mathbf{x}))(f(\mathbf{x}' - m(\mathbf{x}'))]$$

[Berkenkamp,2016] “Bayesian Optimization with Safety Constraints”, CoRR

Gaussian Processes with Noisy Observations



- Prior:

$$m(\mathbf{x}_p) = 0$$
$$\mathrm{cov}(y_p, y_q) = k(\mathbf{x}_p, \mathbf{x}_q) + \sigma_n^2 \delta_{pq}.$$

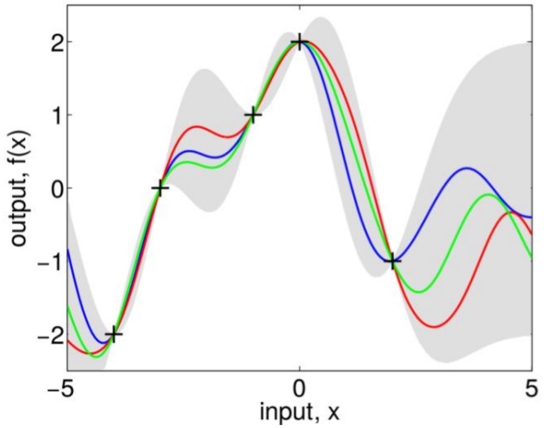
- Likelihood Function:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) + \sigma_n^2 \mathbb{I} & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right)$$

- Bayes Rule:

$$\mathcal{D} = (\mathbf{x}_i, y_i), i = 1:N \text{ and } y = f(\mathbf{x}) + \epsilon$$

$$\textit{Bayes Rule:} \quad p(f|\mathcal{D}) = \frac{(p(\mathcal{D}|f)p(f)}{p(\mathcal{D})}$$

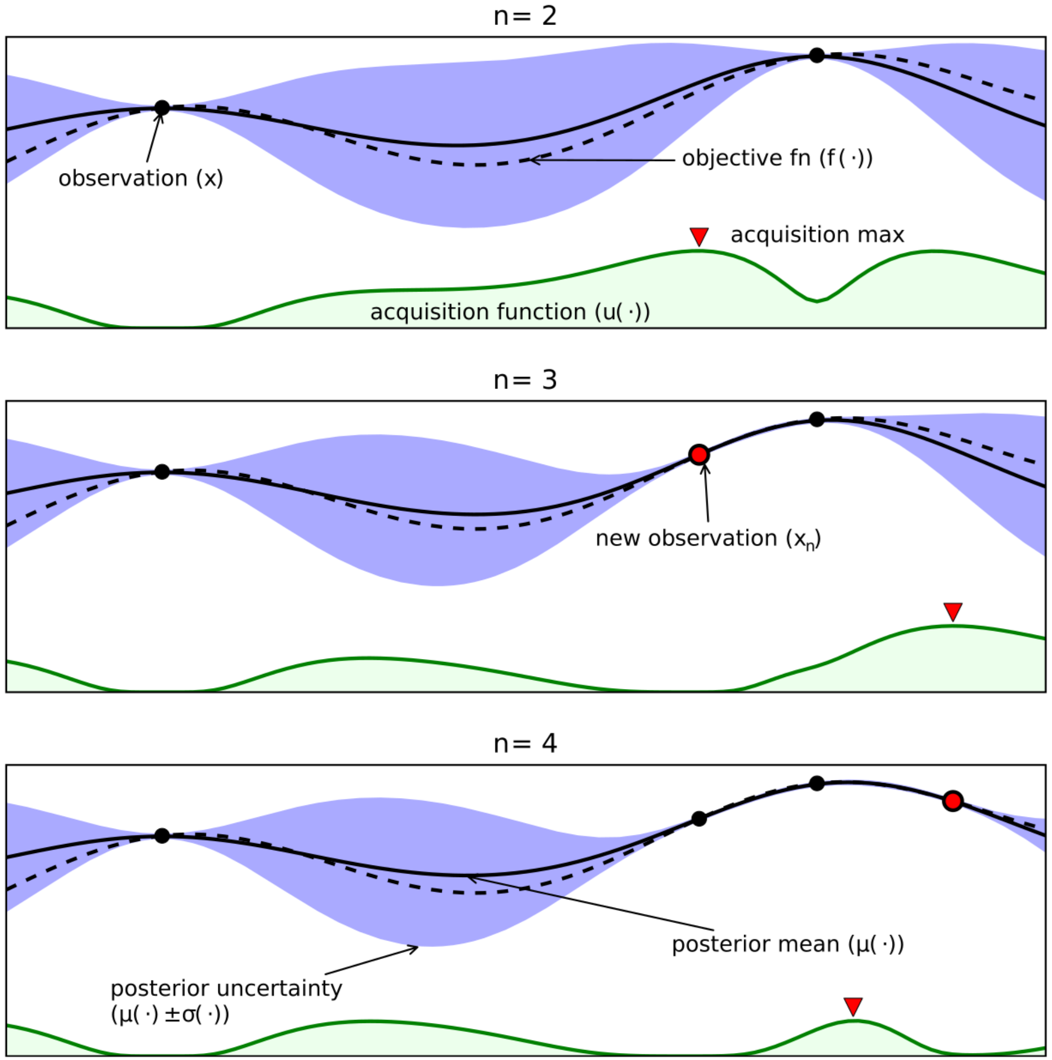


- Posteri distribution:

$$\mathbf{f}_* | X, \mathbf{y}, X_* \sim \mathcal{N}\left(\bar{\mathbf{f}}_*, \mathrm{cov}(\mathbf{f}_*)\right), \text{ with}$$
$$\bar{\mathbf{f}}_* = \mathbb{E}[\mathbf{f}_* | X, \mathbf{y}, X_*] = K(X_*, X)[K(X,X) + \sigma_n^2 \mathbb{I}]^{-1}\mathbf{y},$$
$$\mathrm{cov}(\mathbf{f}_*) = K(X_*, X_*) - K(X_*, X)[K(X,X) + \sigma_n^2 \mathbb{I}]^{-1} K(X, X_*)$$

ETH zürich

Basics of Bayesian Optimization



- Model objective function f(x) as a Gaussian process:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right), \text{whereas}$$

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})],$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x} - m(\mathbf{x}))(f(\mathbf{x}' - m(\mathbf{x}'))]$$

Safe Bayesian Optimization

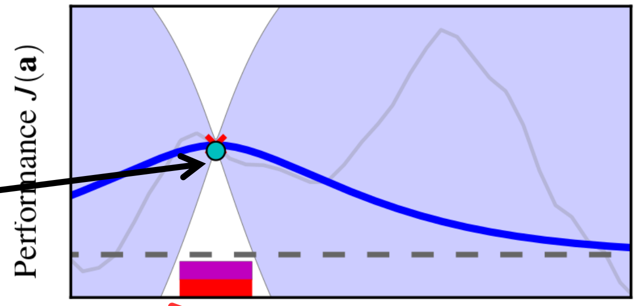
Initial safe parameters

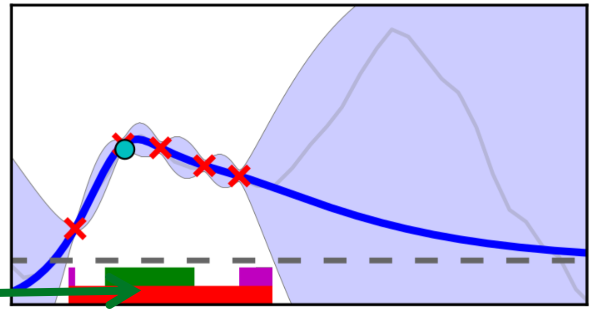
Safe parameter set

Exploitation:
Potential maximizer subset

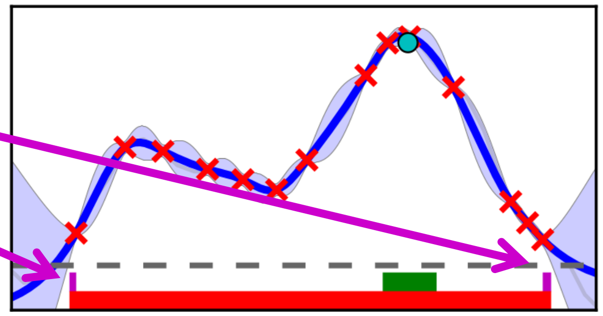
Exploration:
Potential safe set expander subset

Next query: Parameters with highest uncertainty
within the two subsets



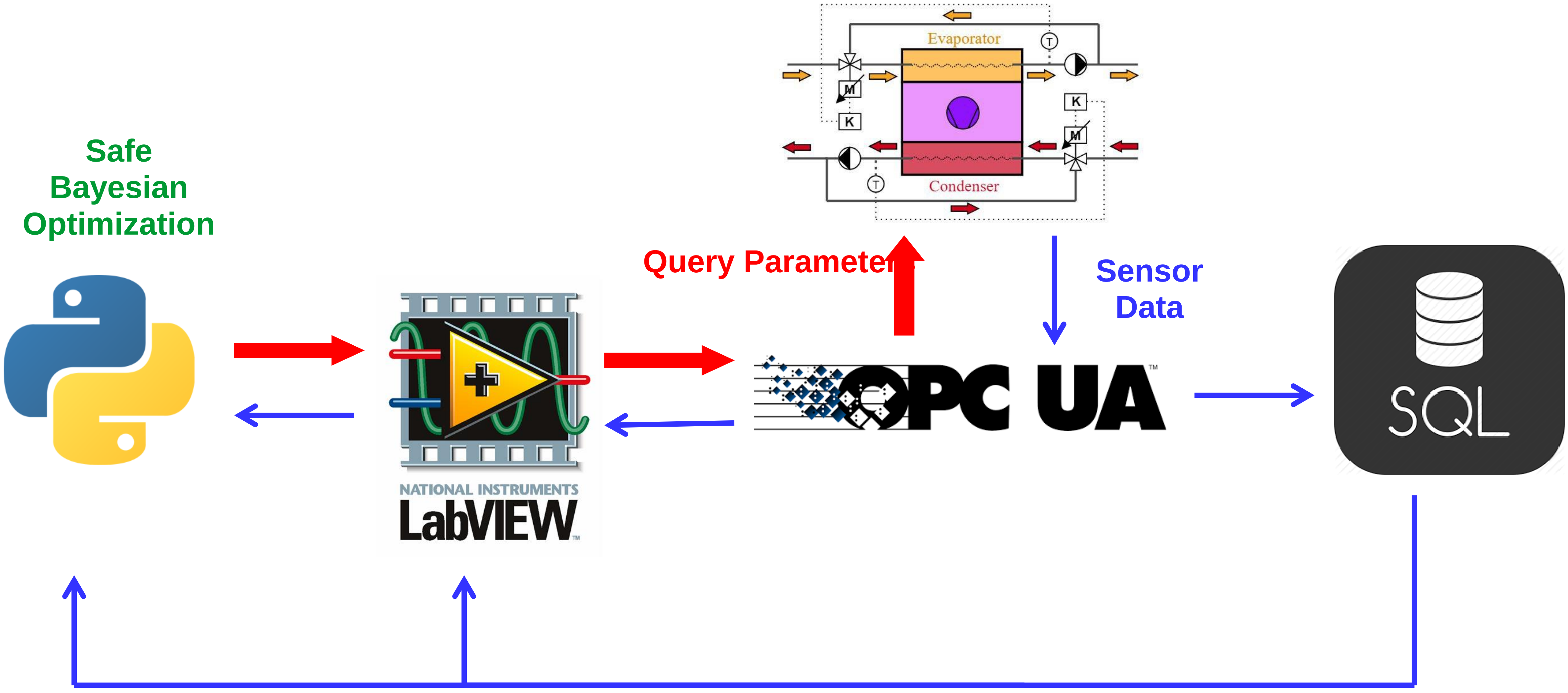


(b) After 5 evaluations: local maximum found.

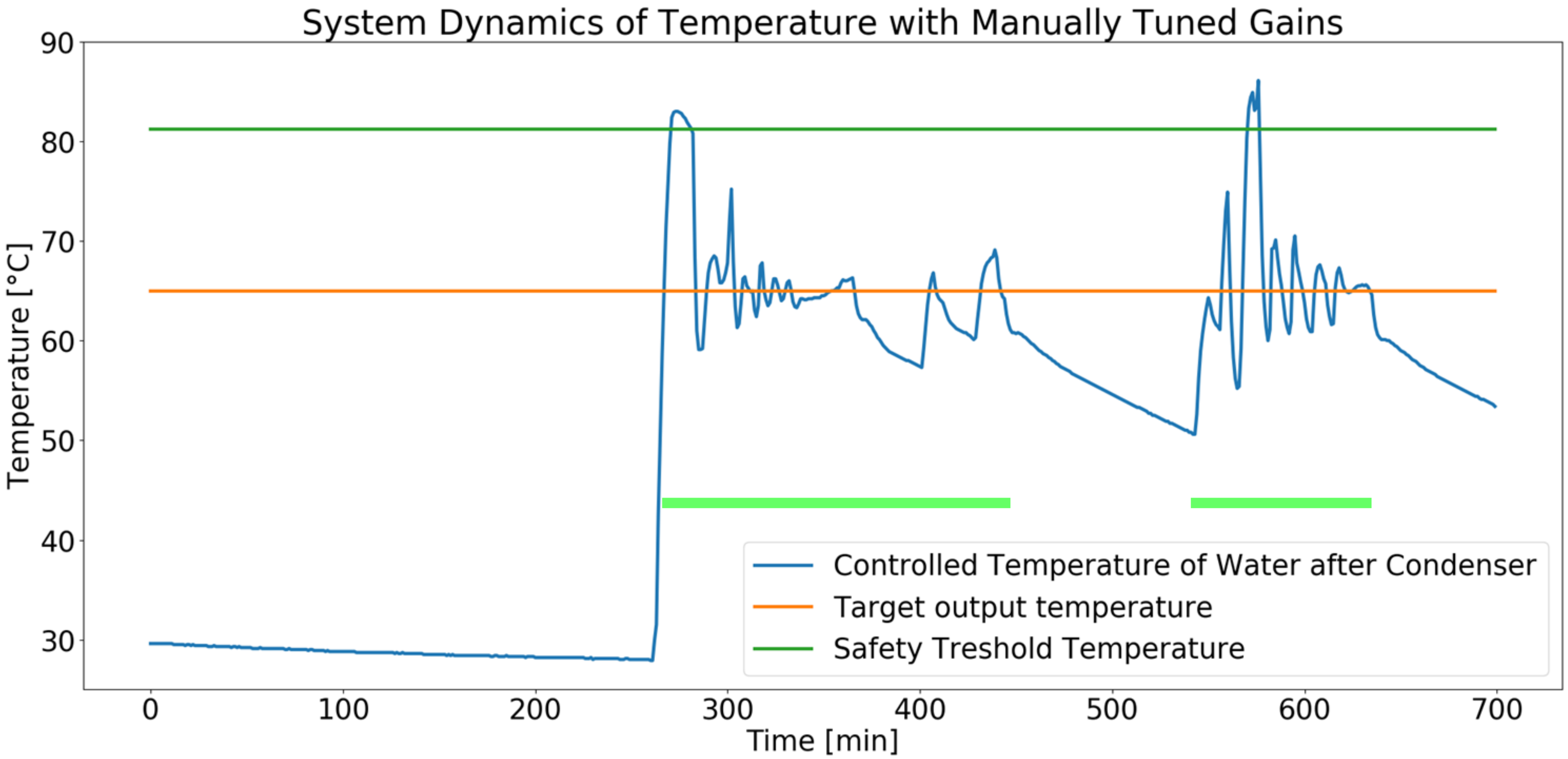


(c) After 13 evaluations: global maximum found.

Experimental Setup for Online Optimization



Experiment: Reducing the Overshoot



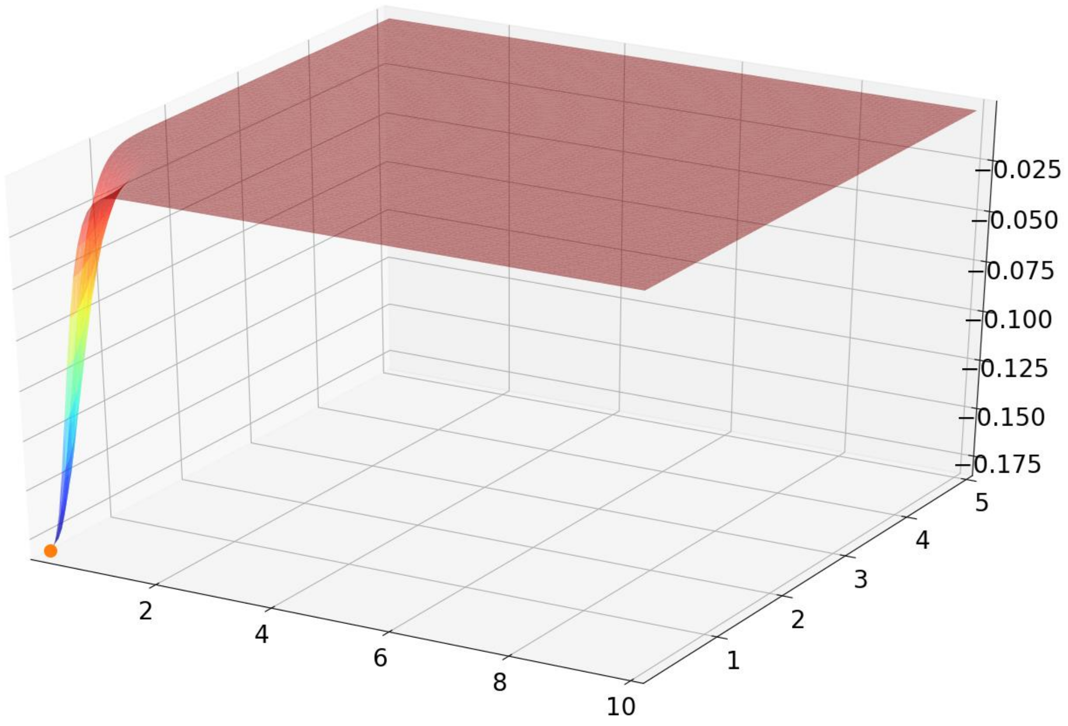
Experiment: Reducing the Overshoot cont.

- Performance: Overshoot
- Safety Constraint: Overshoot
- Safety Treshold: const.=81.25 °C
- Inital safe parameters:
  - Kp=0.02
  - Ki=0.01
- GP Prior ~ N(0,k_c)
- Composite Kernel k_c:

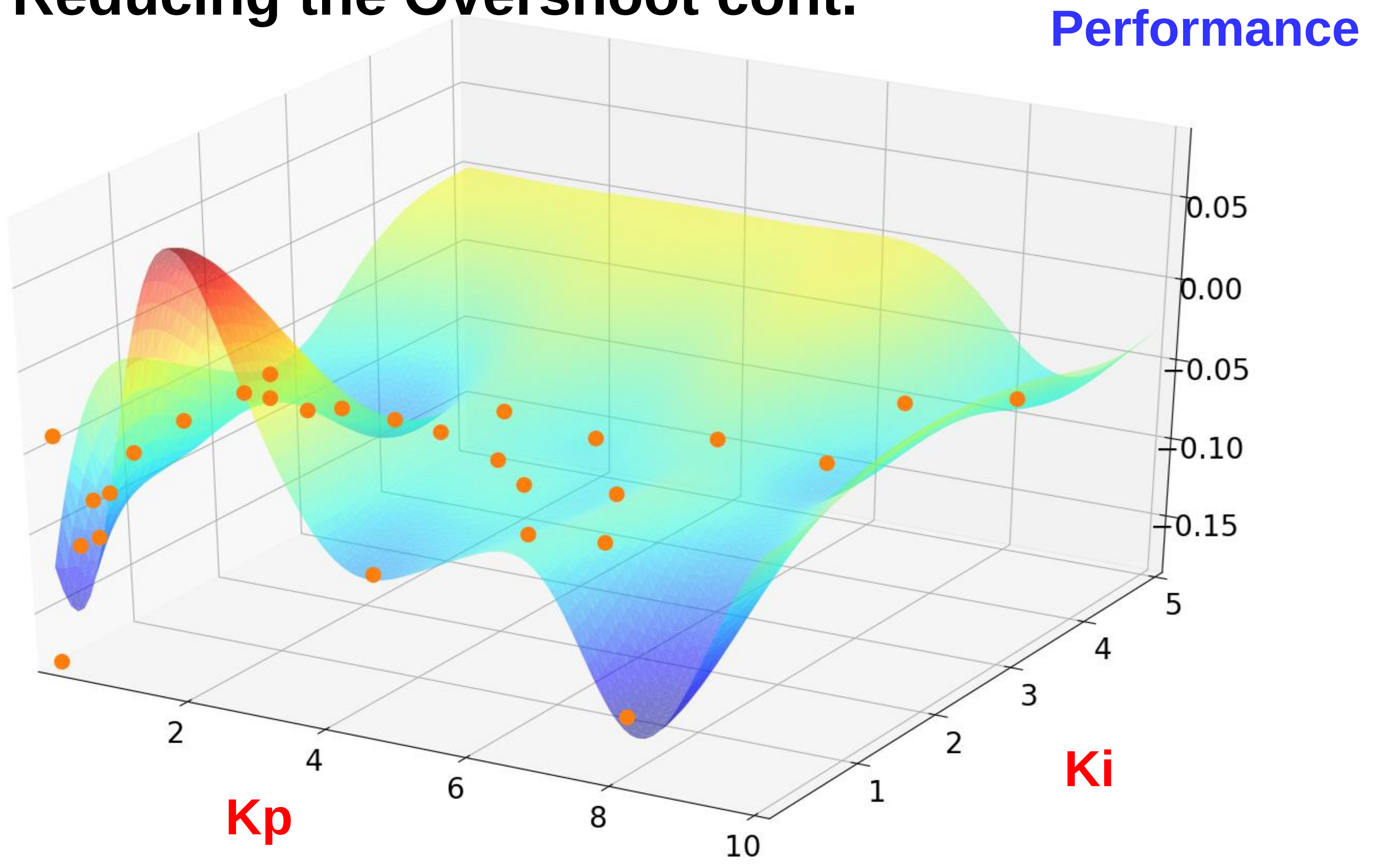
$$k_c(\mathbf{x}_p, \mathbf{x}_q) = k_1(\mathbf{x}_p, \mathbf{x}_q) + k_2(\mathbf{x}_p, \mathbf{x}_q) + \sigma_n^2 \delta_{pq} \, , whereas$$

$$k_1(\mathbf{x}_p, \mathbf{x}_q) = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{x}_p - \mathbf{x}_q)^\mathsf{T} M (\mathbf{x}_p - \mathbf{x}_q)\right),$$

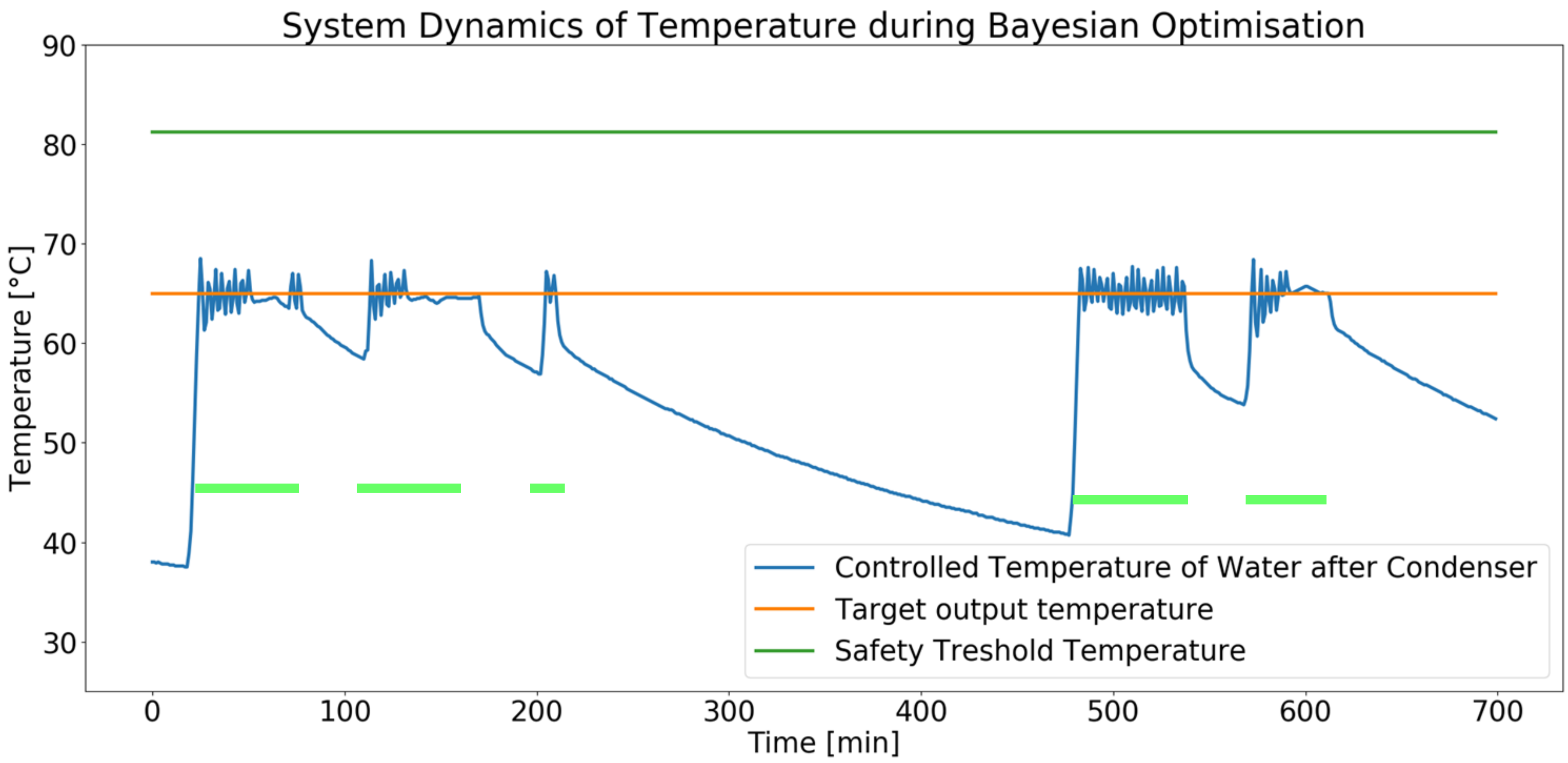
$$k_2(\mathbf{x}_p, \mathbf{x}_q) = const.$$



Experiment: Reducing the Overshoot cont.



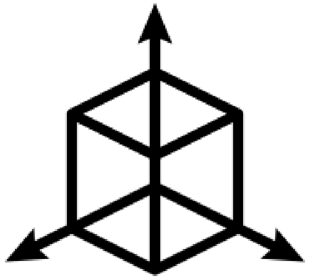
Result: Reducing the Overshoot



The Burden of Optimizing the Gains of the Heat Pump

- Just a few queries per day, i.e. 2-4 per day
- Limited in experimental settings due to safety issues

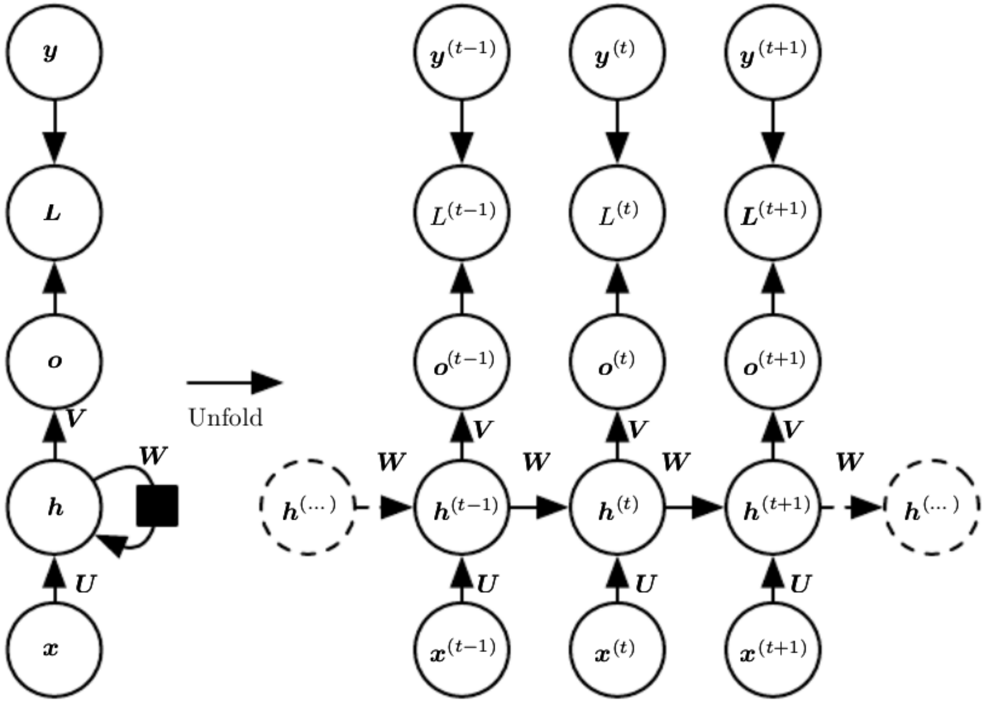
Solution: Data-Driven Simulations of the Dynamic System



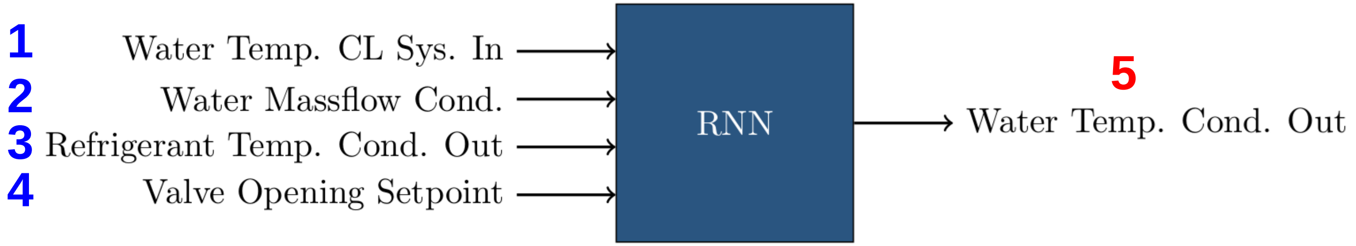
- Enables higher dimensional optimization problems:
- More parameters, contexts
- Fast rigourous testing and validation of different hyperparameters
- Assist and speed up optimization on the real system

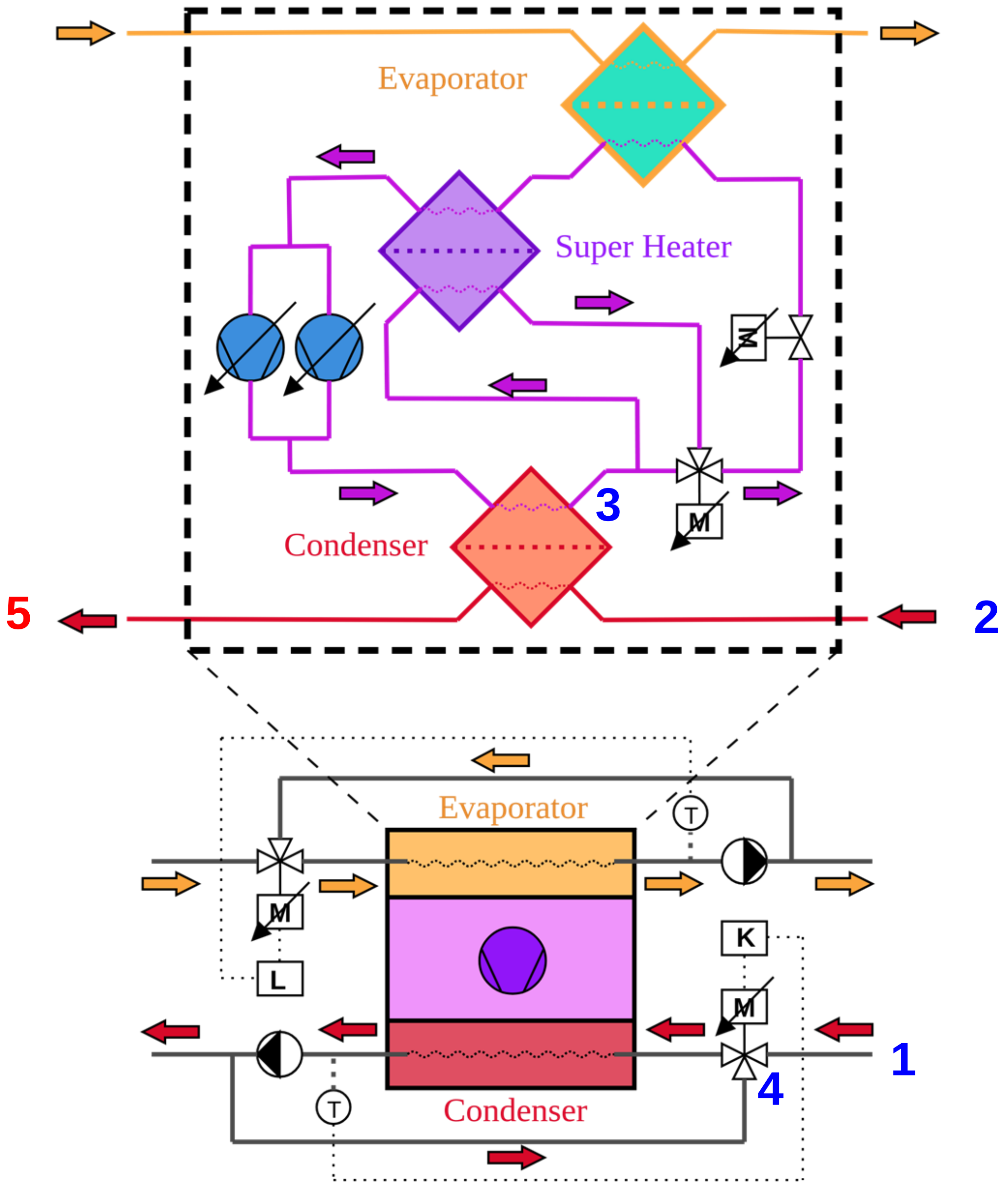
Recurrent Neural Networks

- Emulate dynamic systems
- Need fewer parameters than classical Feed Forward Neural Networks

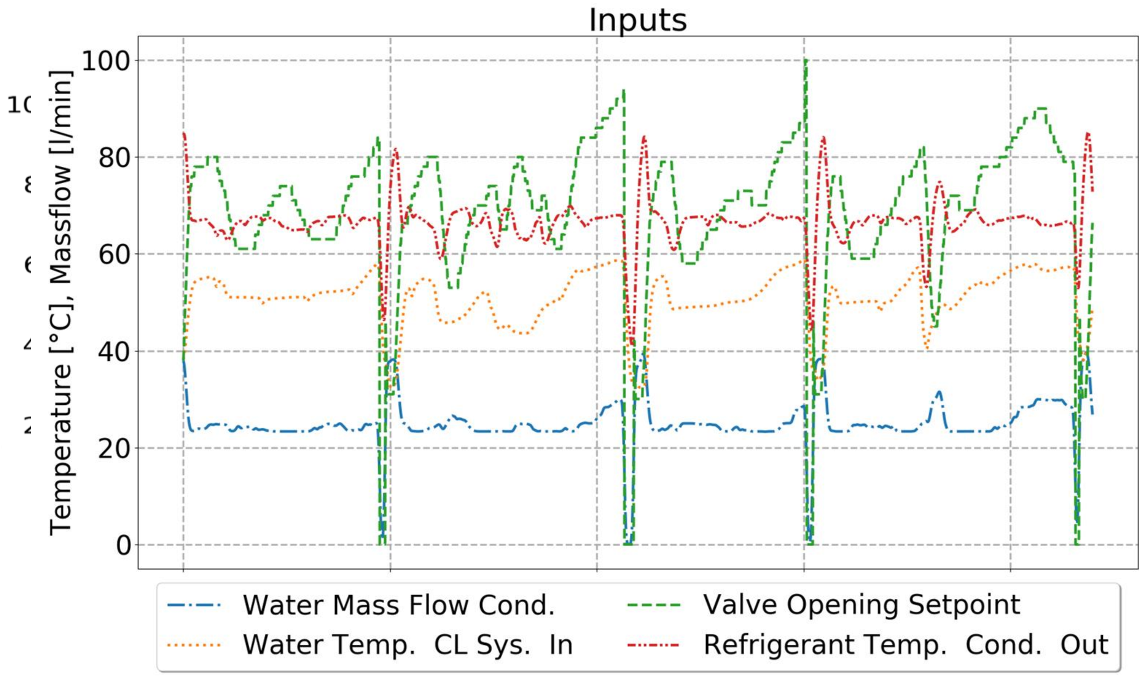


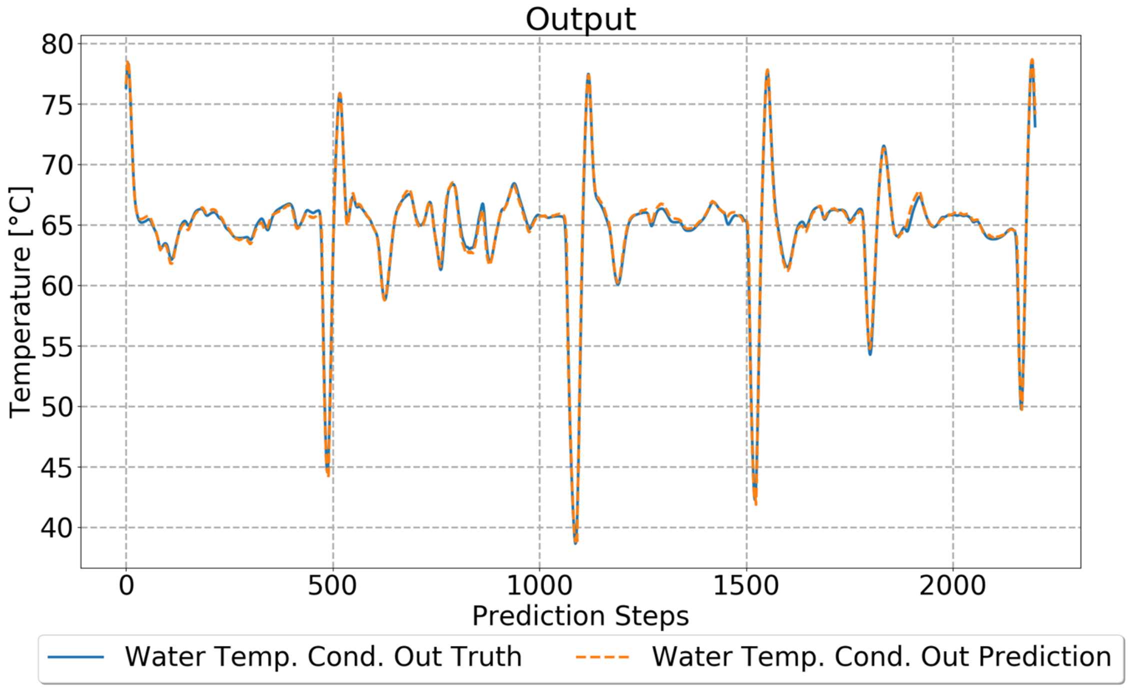
Modelling the Closed Loop System





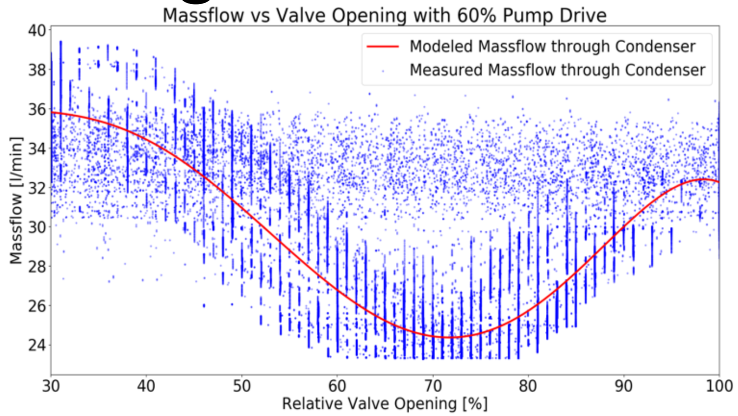
Modelling the Closed Loop System

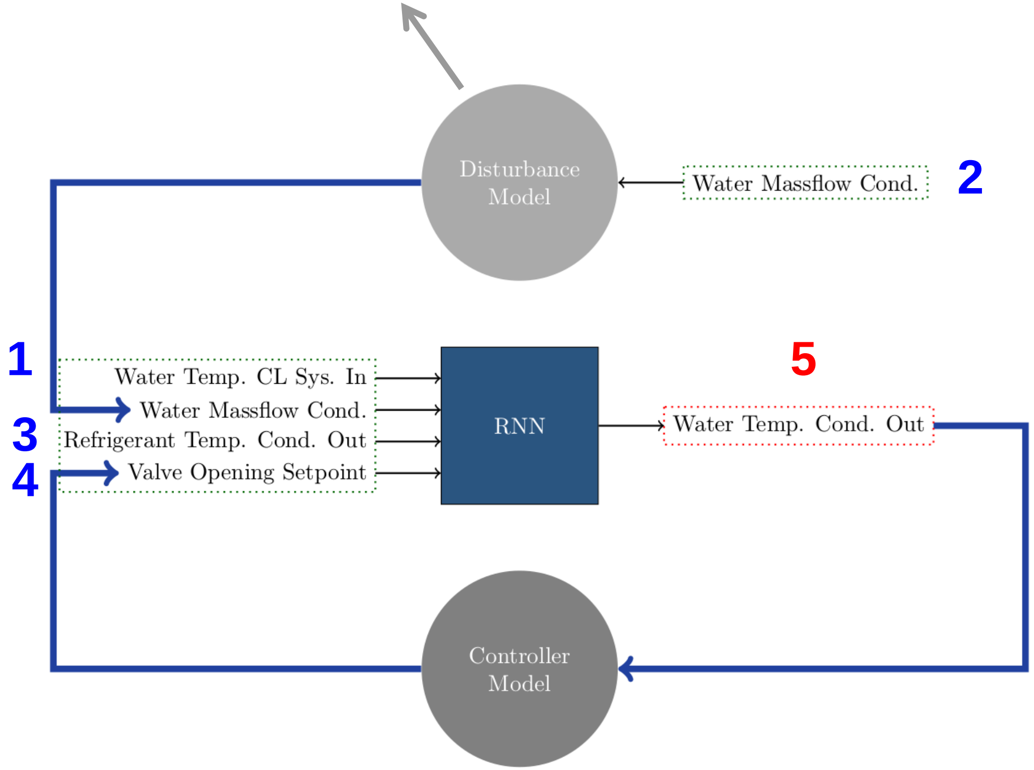


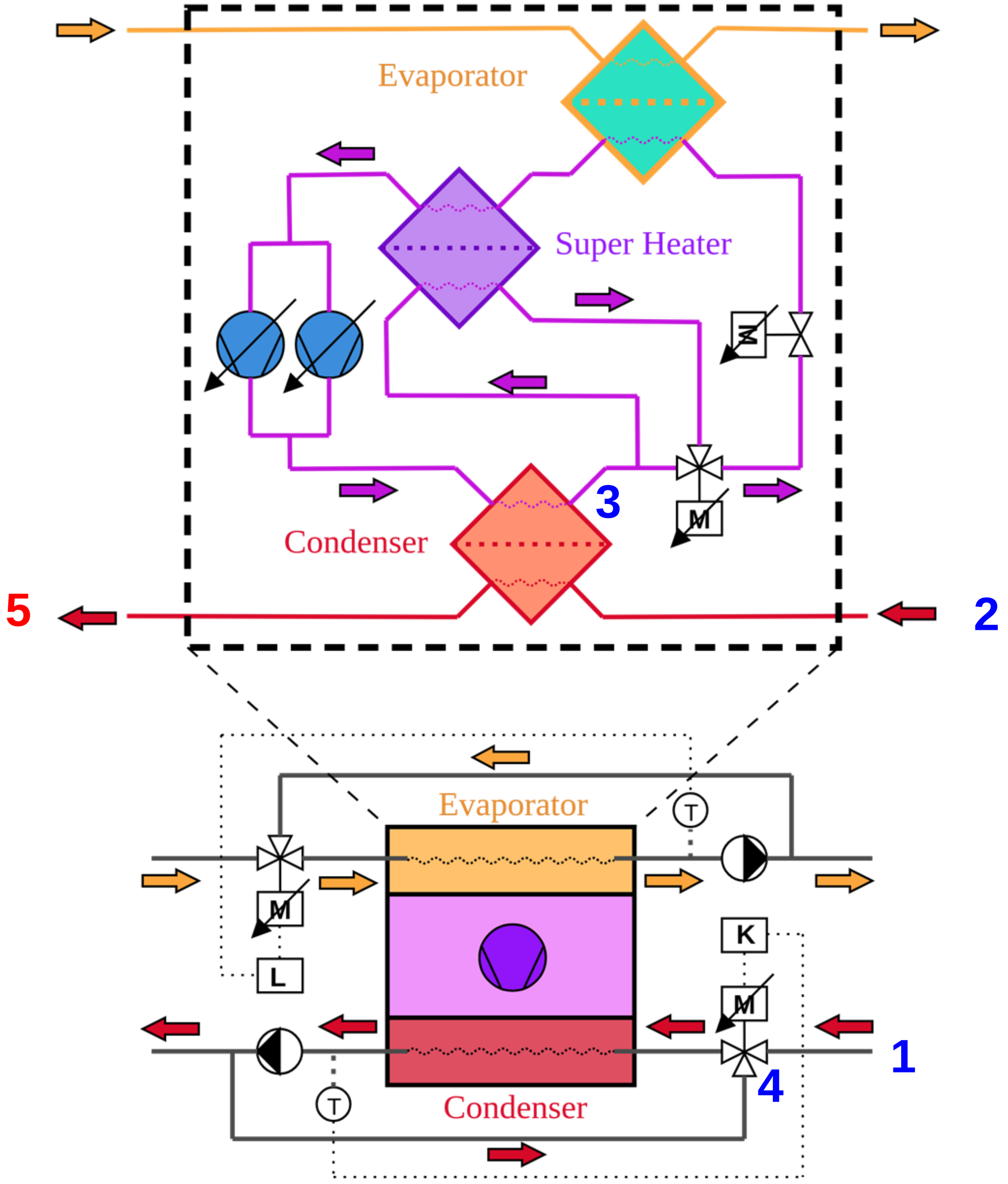


MSE=0.00054 °C^2

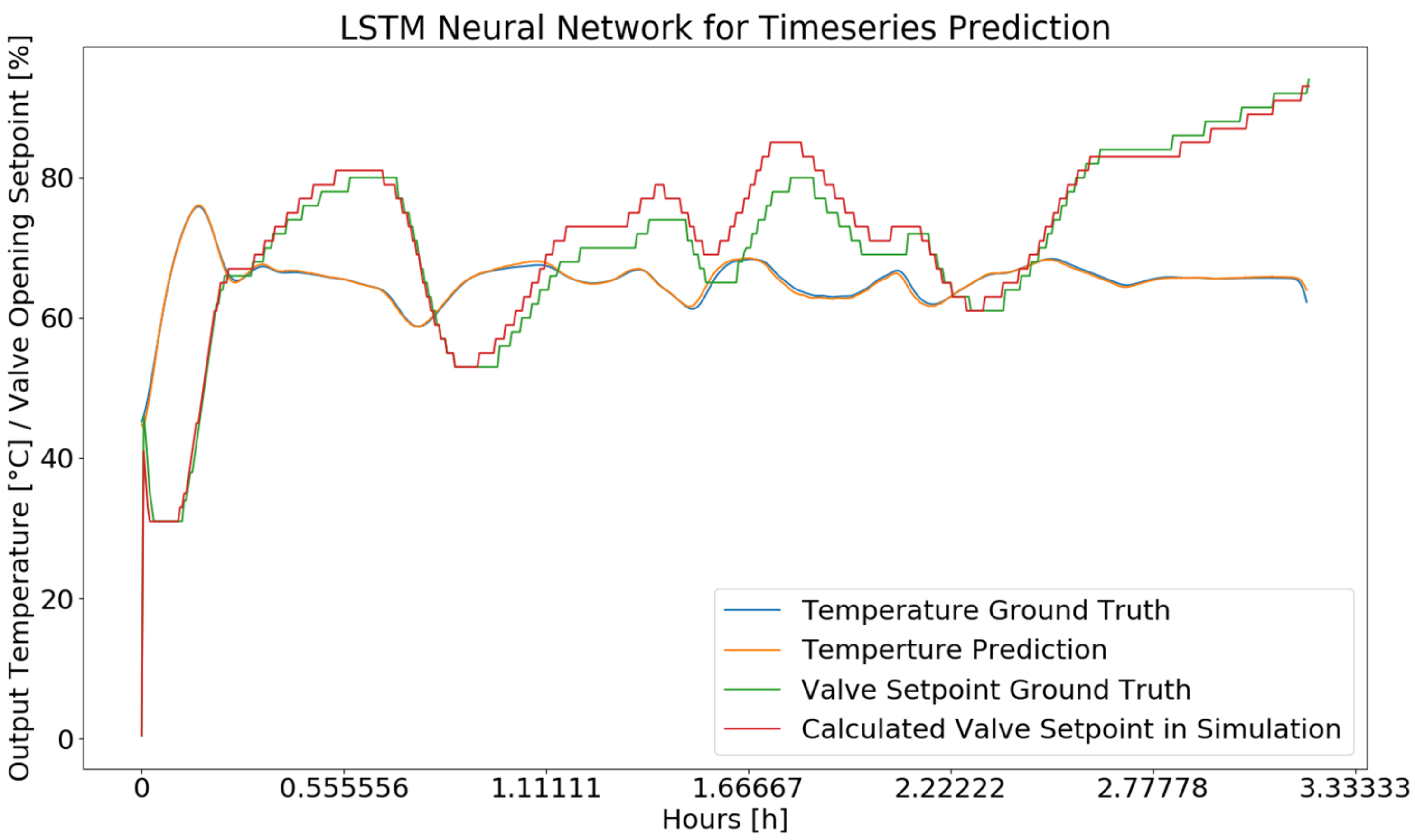
Simulating the Closed Loop System



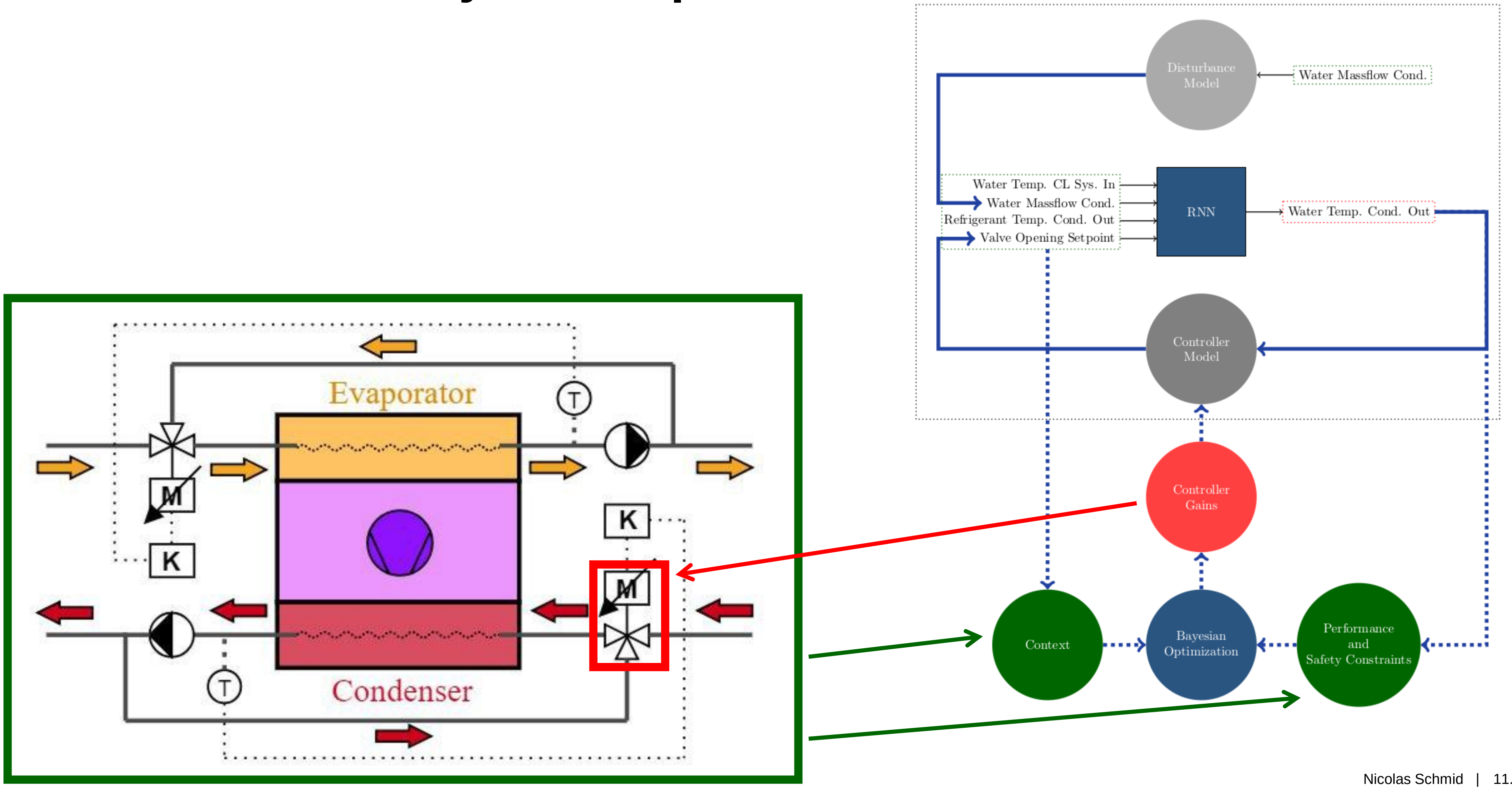




Closed Loop Simulations

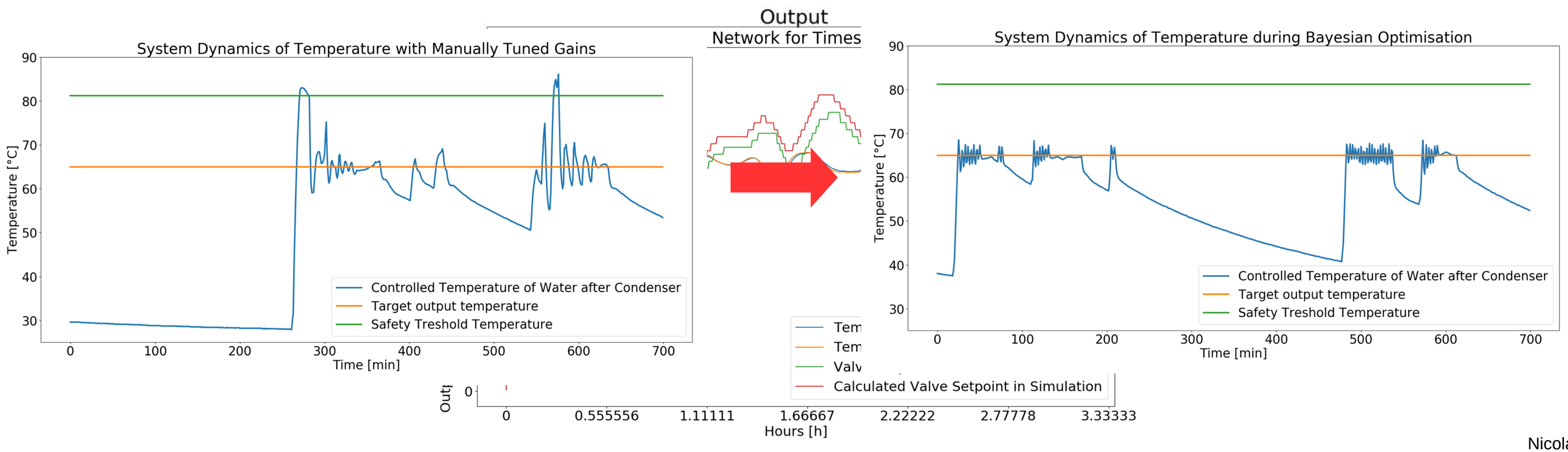


Contextual-Safe Bayesian Optimization



Conclusion

- ✓ Strongly optimized different performance metrics
- ✓ Approximated the closed loop dynamics realisticely
- ✓ Generated a fast and accurate data-driven simulation framework
- ✓ Integrated contextual-safe Bayesian optimization in the simulation framework

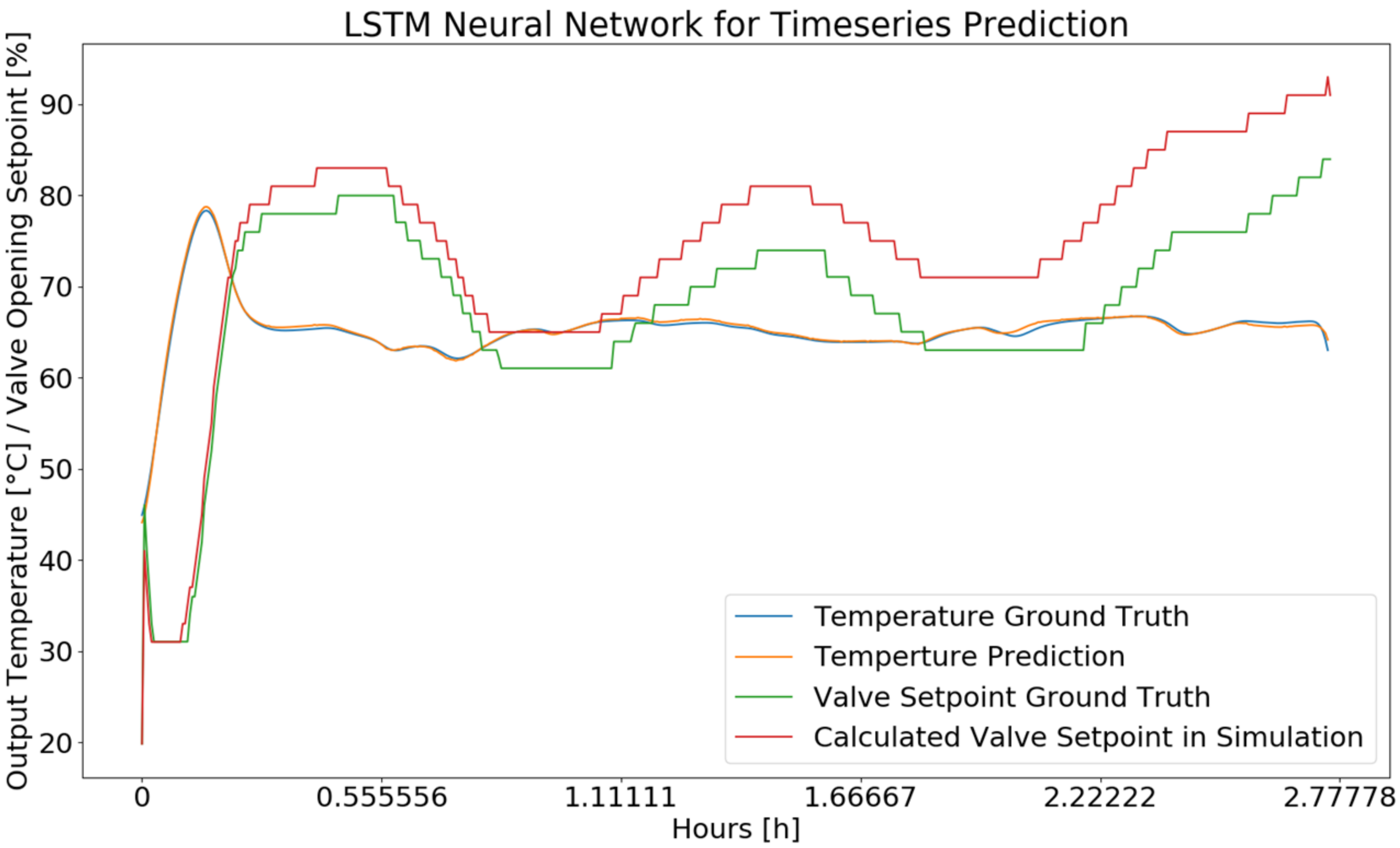


Thanks!

Outlook

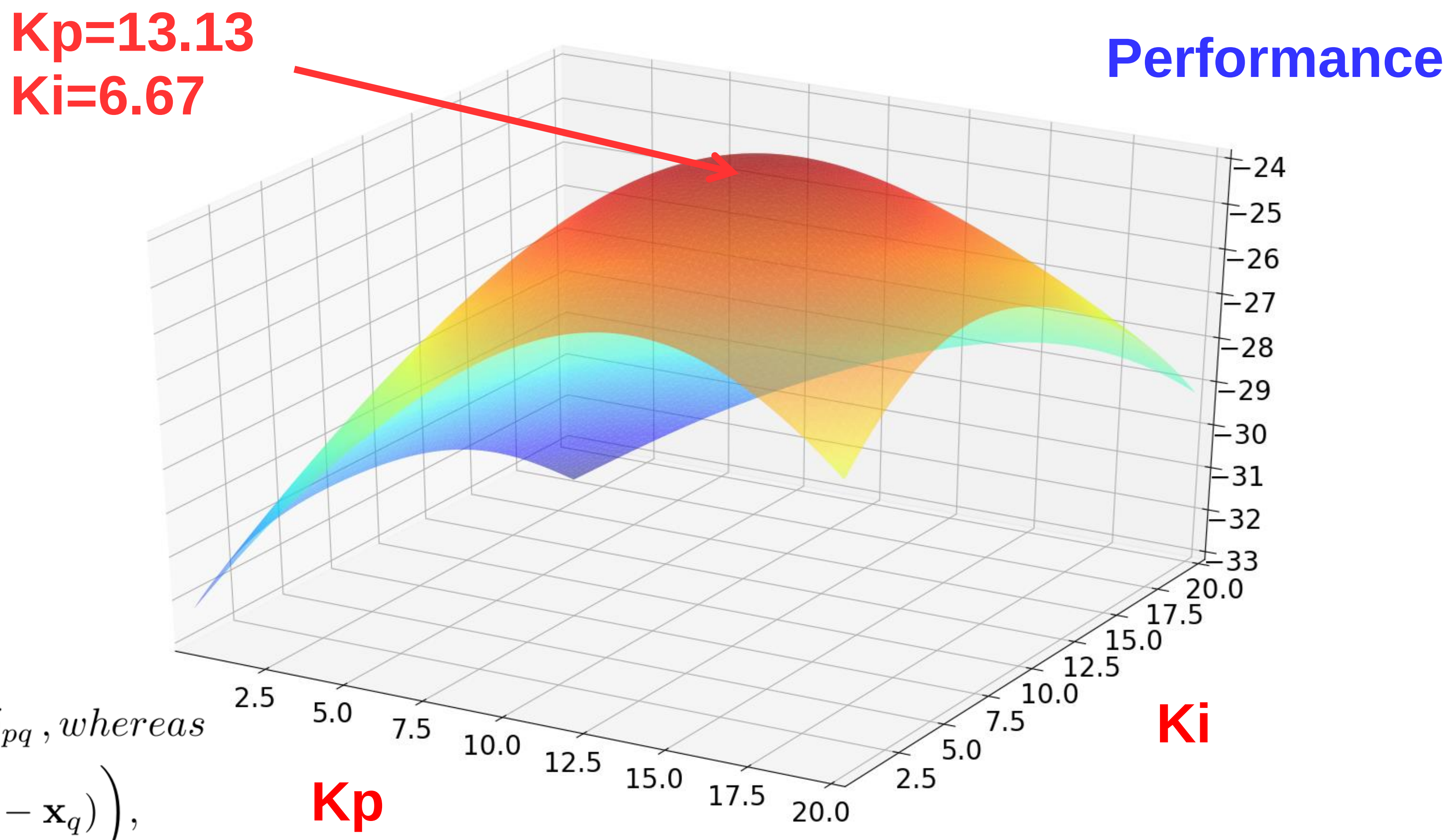
- Systematic experiments with the simulation framework including contexts
- Derive Hyperparameter settings from data-driven simulations
- Trade of simulations and experiments on the real system
- Use simulation framework for learning and optimizing other systems

Closed Loop Simulations cont.



Reducing the Mean Squared Tracking Error

erformance: MSTE
afety Constraint: Overshoot
afety Treshold: const.=81.25 °C
nital safe parameters:
p=0.02
i=0.01
GP Prior ~ N(0,k_c)
Composite Kernel k_c:

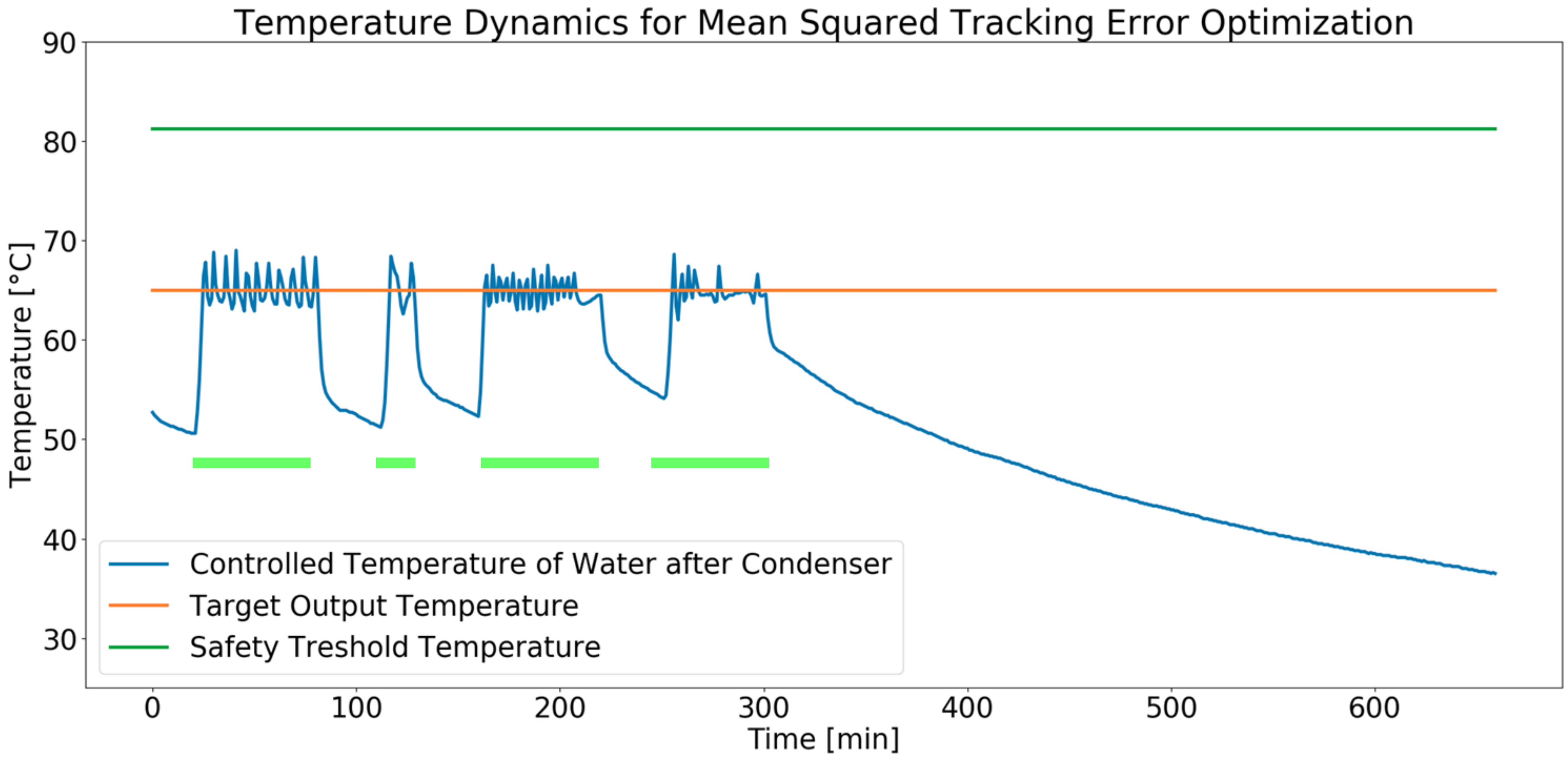


$$k_c(\mathbf{x}_p, \mathbf{x}_q) = k_1(\mathbf{x}_p, \mathbf{x}_q) + k_2(\mathbf{x}_p, \mathbf{x}_q) + \sigma_n^2 \delta_{pq} \,, whereas$$

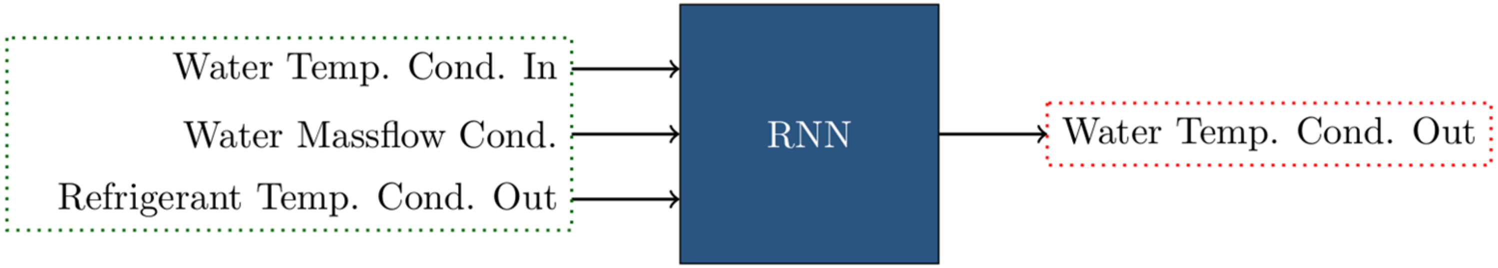
$$k_1(\mathbf{x}_p, \mathbf{x}_q) = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{x}_p - \mathbf{x}_q)^\mathsf{T} M (\mathbf{x}_p - \mathbf{x}_q)\right),$$

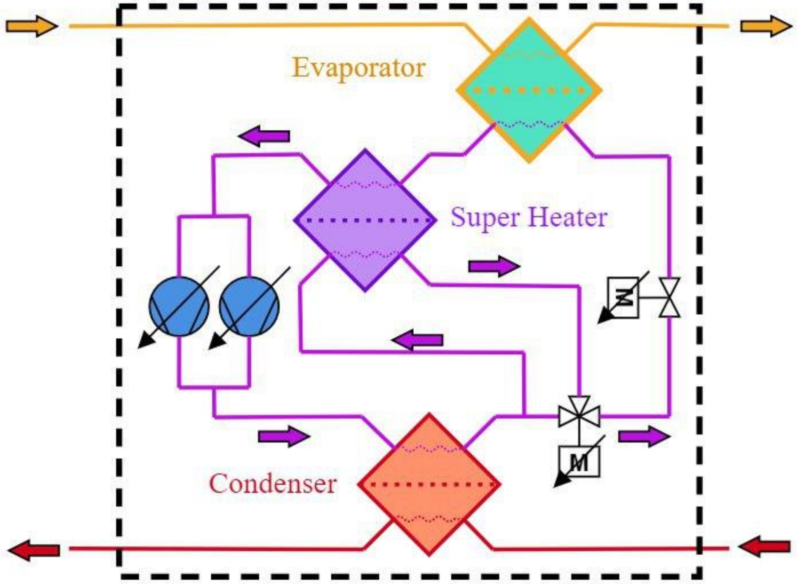
$$k_2(\mathbf{x}_p, \mathbf{x}_q) = const.$$

Reducing the Mean Squared Tracking Error



Modelling the Condenser





Long Short-Term Memory Recurrent Neural Network

The RNN Setup for the Closed Loop Model

1. **Learning Model:** RNN, similar to graph (3.7)
2. **Input Dimensions:** 4, see RNN in figure (**??**)
3. **Output Dimensions:** 1, see RNN in figure (**??**)
4. **Layers:** 128 = input sequence length
5. **Neurons per Hidden Layer:** 16
6. **Training Batch Size:** 256
7. **Total Number of Trainable Parameters:** 1457
8. **Output Layer:** Linear
9. **Prepossessing:** Gaussian filtering
10. **Resampling Interval:** $20s$, corresponds to $20s$ steps
11. **Prediction Length:** 1
12. **Error:** Mean Squared Error
13. **Optimizer:** Adam with Gradient Clipping and Decaying Learning Rates [23] [36]
14. **Regularization:** Recurrent Dropout [3]
15. **Gated Cell:** LSTM [21]
16. **Activation Function:** Leaky ReLU [48]
17. **Parameter Initialization:** Weights: He Initialization, Biases: zero [17]
18. **Normalization:** Layer Normalization [22]