

Online learning-based model predictive control with Gaussian process models and stability guarantees

Michael Maiworm¹  | Daniel Limon²  | Rolf Findeisen¹

¹Laboratory for Systems Theory and Automatic Control,
Otto-von-Guericke-University,
Magdeburg, Germany

²Department of Systems Engineering and Automation, Universidad de Sevilla,
Seville, Spain

Correspondence

Rolf Findeisen, Institute for Automation Engineering,
Otto-von-Guericke-University,
Magdeburg, Germany.
Email: rolf.findeisen@ovgu.de

Michael Maiworm, Laboratory for Systems Theory and Automatic Control,
Otto-von-Guericke-University,
Magdeburg, Germany.
Email: michael.maiworm@ovgu.de

Funding information

Erasmus+; MINECO and FEDER,
Grant/Award Number:
DPI2016-76493-C3-1-R

Abstract

Model predictive control allows to provide high performance and safety guarantees in the form of constraint satisfaction. These properties, however, can be satisfied only if the underlying model, used for prediction, of the controlled process is sufficiently accurate. One way to address this challenge is by data-driven and machine learning approaches, such as Gaussian processes, that allow to refine the model online during operation. We present a combination of an output feedback model predictive control scheme and a Gaussian process-based prediction model that is capable of efficient online learning. To this end, the concept of evolving Gaussian processes is combined with recursive posterior prediction updates. The presented approach guarantees recursive constraint satisfaction and input-to-state stability with respect to the model–plant mismatch. Simulation studies underline that the Gaussian process prediction model can be successfully and efficiently learned online. The resulting computational load is significantly reduced via the combination of the recursive update procedure and by limiting the number of training data points while maintaining good performance.

KEYWORDS

Gaussian processes, input-to-state stability, machine learning, online learning, predictive control, recursive updates

1 | INTRODUCTION

Model predictive control (MPC)¹ is naturally capable of dealing with multi-input multi-output systems and constraints on the input, state, and output already in the design process. This has led to manifold scientific interest, as well as practical applications.^{2,3} In terms of performance, MPC can be superior to other control approaches because the prediction of the process under consideration allows to compute control actions based on future outcomes and facilitates to take preview information about references and disturbances into account. Hence, the prediction model plays a crucial role in MPC. Unfortunately, there is always a certain process-model error or model uncertainty present in practice and the system might change over time, which limits the prediction quality of the model. One way to deal with this situation is to resort to

[Correction added on 5 February 2021, after first online publication: Michael Maiworm was designated as co-corresponding author.]

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *International Journal of Robust and Nonlinear Control* published by John Wiley & Sons Ltd.

robust MPC schemes, such as, for instance, min-max MPC,⁴ tube-based MPC,⁵ multiscenario approaches,^{6,7} or stochastic approaches⁸ that take the uncertainty explicitly into account.

Prediction models are often based on first principles approaches, which can be very time consuming or even impossible in practice. Furthermore, if the underlying process or environmental conditions change, a once good model can degrade and thus needs to be adapted. An alternative to first principle approaches is to derive prediction models directly from measured data. The resulting models, so-called black or grey box models,⁹ can in principle be learned or refined during operation by including newly available data. Thereby, they can account for changing process dynamics or a changing process environment. Combining data-driven with first principles models is another possibility.¹⁰⁻¹²

Although data-driven modeling is not a new field of research, it gained significant attention over the last years due to increasing computational power, the possibility to widely collect data, and the rise of machine learning algorithms, such as neural networks, deep learning, support vector machines, or Gaussian processes (GPs).^{13,14} Especially the use of GPs within MPC has attracted significant interest in recent years.^{10,11,15-18} However, combining GPs with MPC leads to multiple challenges, such as the cubical increase of the computational load with the number of training data points. This also increases the overall necessary computations to solve the resulting optimal control problem. Furthermore, the utilization of GPs in an optimal control problem can render the resulting optimization very nonlinear, even for a small number of data points, which increases the probability of obtaining suboptimal or infeasible solutions. Despite these challenges, GPs are employed together with MPC as they provide several advantages. For instance, they do not only allow to compute a prediction of the system evolution but also a prediction variance (an effective measure of the uncertainty of the learned model), they are less susceptible to overfitting, and they have, under certain circumstances, universal approximation capabilities for a large class of functions,¹⁹ thereby allowing to model the underlying dynamics of a wide variety of systems.

In order to reduce the computational load of GPs one can distinguish two main approaches. The first approach basically fixes the maximum number of training data points, while the second approach employs so-called sparsity.^{20,21} The first approach often entails the drawback that the GP might not be able to model the system with sufficient accuracy throughout the full operation space. To compensate for this, one can resort to online learning (or adaptation) of the GP during operation, which also allows to account for time-varying systems or changing environmental conditions. On the downside, some of the computation time that is saved by reducing the number of training data points is in turn spent by the learning process, which includes updates of the training data set and covariance matrix, recalculation of the covariance matrix inverse, and hyperparameter optimization in each time step. While these often computationally extensive calculations can be performed offline, only very few publications exist that combine MPC with online learning of GPs. The required computations often take too long to control most processes. Thus, GPs are mostly trained/learned offline.^{17,22,23} Exceptions are, for instance, the works by Ortman et al.,²⁴ where the system had a large time constant in the order of hours or Klenske et al.,¹⁶ which provided a hyperparameter optimization tailored to the specific application.

Another important aspect when combining GPs and model predictive control is safety, constraint satisfaction, and stability, for which different approaches have been proposed. One can, for example, avoid to enforce stability by design and include instead the GP posterior variance in the cost function of the optimal control problem. This avoids steering the plant into regions where the model validity is questionable.²⁵⁻²⁷ Also, one can perform a posteriori stability verification. For instance, Berkenkamp et al.²⁸ proposed to learn the region of attraction of a given closed-loop system, whereas Vinogradskaya et al.²⁹ calculated invariant sets for the validation of stability in a closed-loop with GP models. Another possibility is to use invariant *safe sets* and employ a two-layer control framework, where a safe controller is combined with a control policy that optimizes performance.³⁰⁻³³ For instance, in the works by Aswani et al.³⁴ and Bethge et al.³⁵ two different prediction models were used in parallel, where the first is a nominal model, used to guarantee robust stability using tubes, and the other can be a general learning-based model (e.g., a GP) used to optimize performance. In the work of Soloperto et al.³⁶ tube-based MPC was considered together with GPs, which were also used to derive robust stability. To this end, uncertainty sets that are based on the GP variance were used to construct tightened state and input constraint sets. Since the uncertainty sets hold probabilistically, the same goes for the stability result. The two-layer framework was extended to three layers in Bastani.³⁷ The aforementioned approaches are based on the assumption of full state information and the use of invariant terminal regions.

In Maiworm et al.¹⁸ we considered an output nominal MPC scheme (which neither does require full state information nor terminal region in the optimal control problem) with an offline trained GP prediction model and combined it with input-to-state stability (ISS), a framework that covers inherent robust stability of nominal MPC and stability of robust MPC schemes in the presence of constraints.³⁸ If a system under a predictive controller is shown to be ISS, then this property is preserved even in the case of suboptimal solutions of the involved optimal control problem. We outlined conditions under which the GP-MPC scheme is inherently robustly stable (i.e., bounded disturbances lead to bounded effects on the output) and guarantees recursive constraint satisfaction. To this end, the uncertainty or disturbance has to

be bounded deterministically. At the expense of a potentially smaller domain of attraction, the advantage of guaranteeing inherent robust stability lies in its simplicity. The already involved ingredients in MPC merely have to satisfy certain properties (e.g., uniform continuity). The aforementioned methods in the literature on the other hand are conceptually more complex and/or more computationally expensive than the nominal MPC case because different control layers with backup controllers are required,^{30-33,37} different prediction models are employed that have to be evaluated in parallel,³⁵ or tubes have to be computed.³⁶ Furthermore, since the employed MPC formulation provides guarantees without a terminal region, then if also no state constraints have to be fulfilled, the resulting optimal control problem is easier to solve.

In this work, we extend our previous results to the case of a limited training dataset of the GP and aim toward online learning for a wide class of applications. To reduce the computational load we do not consider online hyperparameter optimization. Instead, we focus on a recursive approach to adapt the training dataset and compute the inverse covariance matrix tailored to MPC. The main contributions of this work are:

- Online learning of the GP model, by means of adaptation of the training dataset, at reduced computational cost. This facilitates the possibility of deployment for faster processes. For this purpose, we employ a recursive formulation to update the GP prediction model online.
- Guaranteed ISS with constraint satisfaction for the presented online learning approach. The result is not confined to GPs but holds for general prediction models that are learned online and satisfy the presented conditions.
- The extension of the method such that it yields good performance with only limited prior process knowledge (e.g., lack of training data in important regions of the operation space). To this end, we incorporate the concept of evolving GPs to facilitate online learning by means of adaptation of the training dataset.^{14,39} We derive criteria that use the GP prediction error and the variance to determine which points to add to the training dataset.
- The use of analytic linearized GP models for the determination of the MPC terminal components.

The paper is structured as follows: The considered problem setup is formulated in Section 2. The concept of GPs, together with the recursive formulation for online learning, is outlined in Section 3 and used for the formulation of the optimal control problem in Section 4. The same section also contains the stability results. Section 5 presents simulation results with focus on online learning of the GP before Section 6 concludes the paper.

Notation: Vectors, matrices, and sequences (of vectors or scalars) are set using bold variables. For matrices we use upper case (\mathbf{Y}), for vectors slanted lower case (\mathbf{y}), and for sequences upright lower case (\mathbf{y}). Sets are denoted by calligraphic upper case variables (\mathcal{Y}). The distance of a point $\mathbf{z} \in \mathbb{R}^p$ to a set $\mathcal{Y} \subset \mathbb{R}^p$ is defined as $d(\mathbf{z}, \mathcal{Y}) = \inf_{\mathbf{y} \in \mathcal{Y}} \|\mathbf{y} - \mathbf{z}\|_\infty$, where $\|\cdot\|_\infty$ is the infinity norm (i.e., $d(\mathbf{z}, \mathcal{Y}) = 0$ if $\mathbf{z} \in \mathcal{Y}$). If not stated otherwise, $\|\cdot\|$ denotes the Euclidean vector norm. A function $\alpha : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is a \mathcal{K} -function if it is continuous, $\alpha(0) = 0$, and if it is strictly increasing. A function $\alpha : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is a \mathcal{K}_∞ -function if it is a \mathcal{K} -function and unbounded. A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is a \mathcal{KL} -function if $\beta(s, t)$ is \mathcal{K}_∞ in s for any value of t and $\lim_{t \rightarrow \infty} \beta(s, t) = 0, \forall s \geq 0$.

2 | PROBLEM FORMULATION

We consider nonlinear discrete-time systems represented by a nonlinear autoregressive model with exogenous input (NARX)*

$$y_{k+1} = f(\mathbf{x}_k, u_k) + e \quad (1a)$$

$$\text{s.t. } u_k \in \mathcal{U} \quad (1b)$$

$$y_k \in \mathcal{Y}. \quad (1c)$$

Here k denotes the discrete time index, $u_k \in \mathbb{R}$ the input, $y_k \in \mathbb{R}$ the output, and $\mathbf{x}_k \in \mathbb{R}^{n_x}$ is the NARX “state vector”

$$\mathbf{x}_k = [y_k \ \dots \ y_{k-m_y} \ u_{k-1} \ \dots \ u_{k-m_u}]^T \quad (2)$$

*Under certain observability assumptions,⁴⁰ a NARX model is sufficient to describe the dynamics of a wide class of systems.

that consists of the current and past outputs and inputs, and where m_y, m_u determine the NARX model order $n_x = m_y + m_u + 1$. The output is corrupted by Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$ with zero mean, noise variance σ_n^2 , and bounded support $|\epsilon| \leq \bar{\epsilon} < \infty$.[†] Inputs and outputs are restricted to lie in the constraint compact sets $\mathcal{U} \subseteq \mathbb{R}$ and $\mathcal{Y} \subseteq \mathbb{R}$, where \mathcal{U} are hard constraints and \mathcal{Y} can be hard or soft constraints that we denote by \mathcal{Y}_h and \mathcal{Y}_s respectively. The NARX state and the output are connected via $y_k = \mathbf{c}^T \mathbf{x}_k$ with $\mathbf{c}^T = [1 \ 0 \ \dots \ 0]$.

The considered control objective is set-point stabilization and optimal set-point change, that is, we want to steer the system from an initial point (\mathbf{x}_0, u_0) to a target reference point $(\mathbf{x}_{\text{ref}}, u_{\text{ref}})$, while satisfying the constraints and stabilizing the system at the target. To this end, we employ model predictive control, which requires a model

$$\hat{y}_{k+1} = \hat{f}(\mathbf{x}_k, u_k), \quad (3)$$

of the process (1a) that is capable of predicting future output values with sufficient accuracy. The hat notation ($\hat{\cdot}$) denotes an estimated quantity. We outline an approach to learn the system model approximation $\hat{f}(\mathbf{x}_k, u_k)$ from measured input-output data using a GP, which is capable of online learning during operation based on newly available data. This results in a GP-based NARX prediction model.

Remark 1. We consider a NARX model with one output that is modeled by a GP. The presented approach can be extended to more outputs, where for each output an individual GP is used, compared to Ostafew et al.^{11,22} or Klenske et al.¹⁶ The theoretical results obtained in Section 4 are also valid for the multi-output case.

3 | GAUSSIAN PROCESSES

We first review the basics of GP regression and then present a recursive formulation that is based on the concept of evolving GPs. This facilitates the generation of a NARX prediction model capable of adapting to changing conditions. To reduce the online computational cost, we do not consider online hyperparameter optimization. Instead, we focus on updating the training dataset efficiently and how to perform the required computations online. To this end, we combine this concept with a recursive update of the involved Cholesky decomposition.

3.1 | Basics

A GP is a *collection of random variables, any finite number of which have a joint Gaussian distribution*.¹³ It generalizes the Gaussian probability distribution to distributions over functions and can therefore be used to model/approximate functions that can be used to capture dynamic systems.¹⁴ They can be utilized for models purely derived from data or combined in a hybrid way with other, for instance, deterministic models.^{10-12,28,30,31,36,41,42}

For regression, GPs are employed to derive or approximate maps of the form $z = f(\mathbf{v}) + \epsilon$ with input \mathbf{v} , output z , and where $f(\cdot)$ is the underlying but unknown *latent* function. The output is assumed to be corrupted by Gaussian noise[‡] $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$ with zero mean and noise variance σ_n^2 . The objective is to infer the function $f(\cdot)$ using measured input-output data (\mathbf{v}, z) with a GP $g(\mathbf{w})$ with input $\mathbf{w} \in \mathbb{R}^{n_w}$, called *regressor*. In the present case (1a), we have $z = y_{k+1}$ and $f(\mathbf{v}) = f(\mathbf{x}_k, u_k)$. The regressor of the GP will be $\mathbf{w}_k = (\mathbf{x}_k, u_k) \in \mathbb{R}^{n_w}$ with regressor order $n_w = n_x + 1$. For the sake of brevity we omit the dependence on the discrete time step k in the remainder of this section whenever possible.

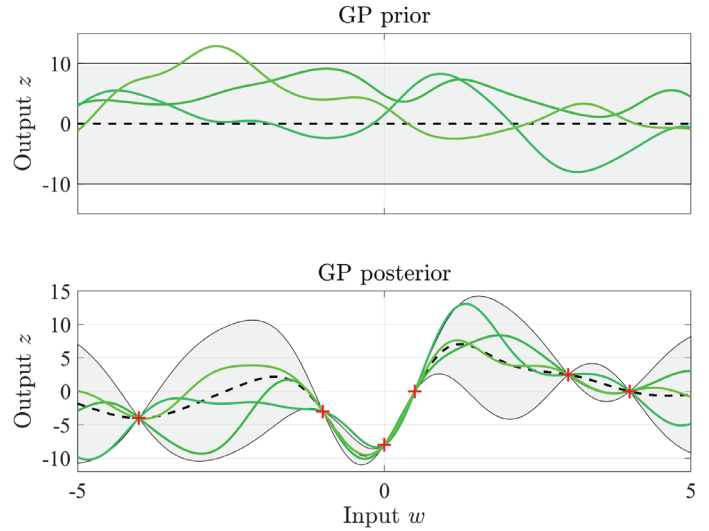
The first required element is a GP prior distribution $g(\mathbf{w}) \sim \mathcal{GP}(m(\mathbf{w}), k(\mathbf{w}, \mathbf{w}'))$ that is specified via the mean function $m(\mathbf{w}) = \mathbb{E}[g(\mathbf{w})]$ and the covariance function[§] $k(\mathbf{w}, \mathbf{w}') = \text{cov}[g(\mathbf{w}), g(\mathbf{w}')] = \mathbb{E}[(g(\mathbf{w}) - m(\mathbf{w}))(g(\mathbf{w}') - m(\mathbf{w}'))]$ with $\mathbf{w}, \mathbf{w}' \in \mathbb{R}^{n_w}$ and $\mathbb{E}[\cdot]$ denoting the expected value. The mean and covariance function together with a set of so-called hyperparameters θ , detailed later, fully specify the GP.

[†]In real processes the measurement noise is always bounded, for instance, due to the limitations of the involved data acquisition systems.

[‡]The concept of GPs assumes Gaussian noise in the measurements, that is, noise with unbounded support. The considered real system (1), however, is corrupted by Gaussian noise with bounded support. The resulting approximation error can be absorbed in the prediction error (8) defined further below. On the other hand, Gaussian noise with unbounded support can be regained by *GP warping*.⁴³ The smaller the bounded support, the larger the difference between the distributions and the larger the correcting effect of warping.

[§]The covariance function is also denoted as kernel.

FIGURE 1 Gaussian process (GP) inference: The top figure depicts a GP prior distribution with the dashed black line representing the mean function $m(\mathbf{w})$ and the green lines representing random function realizations drawn from the prior distribution. The grey shaded area is the 95% (twice the SD) confidence interval computed via $k(\mathbf{w}, \mathbf{w}')$. When data points D are added (bottom figure, red crosses), the GP posterior with $m_+(\mathbf{w}|D)$ and $\sigma_+^2(\mathbf{w}|D)$ is inferred from this data



The GP prior is trained/learned using a set of n measured input-output data points, where the input dataset is $\mathbf{w} = [\mathbf{w}_1 \dots \mathbf{w}_n]^T \in \mathbb{R}^{n \times n_w}$ and the output data set $\mathbf{z} = [z_1 \dots z_n]^T \in \mathbb{R}^{n \times 1}$. The combined data $D = \{\mathbf{w}, \mathbf{z}\}$ is denoted as training dataset and is used to infer the posterior distribution

$$g(\mathbf{w}|D) \sim \mathcal{GP}(m_+(\mathbf{w}|D), \sigma_+^2(\mathbf{w}|D)).$$

This is also a GP with posterior mean $m_+(\mathbf{w}|D)$ and posterior variance $\sigma_+^2(\mathbf{w}|D)$ given by

$$m_+(\mathbf{w}|D) = m(\mathbf{w}) + k(\mathbf{w}, \mathbf{w})\mathbf{K}^{-1}(\mathbf{z} - m(\mathbf{w})) \tag{4a}$$

$$\sigma_+^2(\mathbf{w}|D) = k(\mathbf{w}, \mathbf{w}) - k(\mathbf{w}, \mathbf{w})\mathbf{K}^{-1}k(\mathbf{w}, \mathbf{w}), \tag{4b}$$

with $m(\mathbf{w}) = [m(\mathbf{w}_1) \dots m(\mathbf{w}_n)]^T \in \mathbb{R}^{n \times 1}$, $k(\mathbf{w}, \mathbf{w}) = [k(\mathbf{w}, \mathbf{w}_1) \dots k(\mathbf{w}, \mathbf{w}_n)] \in \mathbb{R}^{1 \times n}$, $k(\mathbf{w}, \mathbf{w}) = k(\mathbf{w}, \mathbf{w})^T$, and $\mathbf{K} = k(\mathbf{w}, \mathbf{w}) = [k(\mathbf{w}_i, \mathbf{w}_j)] \in \mathbb{R}^{n \times n}$.

Note that realizations of the posterior can yield infinitely many function outcomes but as it is conditioned on the training data points, it rejects all possible functions that do not go through or nearby (if $\sigma_n^2 \neq 0$) these points (Figure 1).

The posterior mean function (4a) is the desired estimator of the unknown output latent function $f(\mathbf{x}_k, u_k)$ in (1a), which we highlight by defining

$$\hat{y}_{k+1} = \hat{z} := \hat{f}(\mathbf{x}_k, u_k) = m_+(\mathbf{w}_k|D_k). \tag{5}$$

The key elements for a GP to yield a sensible model are the prior mean and covariance function. Both depend generally on a set of hyperparameters θ , that is, $m(\mathbf{w}|\theta)$ and $k(\mathbf{w}, \mathbf{w}'|\theta)$. Very often just a constant zero prior mean $m(\mathbf{w}|\theta) = c = 0$ is used.^{15,44,45} However, other choices include, for instance, the use of a deterministic base model $\mathbf{x}_{k+1} = f(\mathbf{x}_k, u_k)$ as the prior mean function.^{10,46} Regarding the covariance function, it is often assumed or known that the system dynamics can be modeled by a member of the space of smooth functions C^∞ . A covariance function that provides this property is the *squared exponential covariance function with automatic relevance determination*

$$k(\mathbf{w}_i, \mathbf{w}_j|\theta) = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{w}_i - \mathbf{w}_j)^T \Lambda (\mathbf{w}_i - \mathbf{w}_j)\right) + \sigma_n^2 \delta_{ij}, \tag{6}$$

where $\mathbf{w}_i, \mathbf{w}_j \in \mathbb{R}^{n_w}$, $\theta = \{\sigma_f^2, \Lambda\}$, and $\Lambda = \text{diag}(l_1^{-2}, \dots, l_{n_w}^{-2})$. The measurement noise σ_n^2 is added via the Kronecker delta δ_{ij} in (6). The minimal required number of regressors n_w can be determined through optimization of the length scale parameters l in Λ .^{47,48} Other choices include, for instance, the combination of (6) with a linear kernel.^{47,49}

A common approach to determine the hyperparameters θ , given a training dataset $D = \{\mathbf{w}, \mathbf{z}\}$, is to maximize the log marginal likelihood¹³

$$\log(p(\mathbf{z}|\mathbf{w}, \theta)) = -\frac{1}{2}\mathbf{z}^T\mathbf{K}^{-1}\mathbf{z} - \frac{1}{2}\log|\mathbf{K}| - \frac{n}{2}\log(2\pi). \quad (7)$$

An advantage of GPs is that (4b) naturally provides a quantification of the model uncertainty in the form of its variance. On the other hand, the involved computations in (4a) and (4b) scale with $\mathcal{O}(n^3)$ due to \mathbf{K}^{-1} , where n is the number of training data points. This severely limits the application of GP models for fast processes, where small sampling times are required; especially in the case of relatively large training datasets with several hundred or thousands of data points. If online or close to online hyperparameter optimization is needed, this drawback becomes even more pronounced.

3.2 | Evolving GPs

In order to efficiently refine the GP model online we seek to update the training dataset D_k , possibly at each time step k , during operation. To this end, we resort to the concept of so-called *evolving* GPs,^{14,39} which can be used, for instance, if the training data is only available for certain regions of the operating space and one wants to expand operation beyond these regions online. The concept basically leads to GPs whose training data set D_k is updated online using some type of information criterion. Different criteria can be used to select new data points to be added and already existing points to be removed if necessary.

The general idea is to include an incoming data point to the training dataset only if it contributes enough new valuable information, which can be defined in different ways and depends on the respective application. Possible options are the use of the information gain, entropy difference, or the expected likelihood.^{50,51} We employ the GP as a prediction model in MPC and are therefore particularly interested in how accurate the current model is able to predict the output value at the next time step and how confident this prediction is. To this end, given a new data point (\mathbf{w}_k, y_{k+1}) , we first define the prediction error via

$$e^p := y_{k+1} - \hat{y}_{k+1} = f(\mathbf{x}_k, u_k) + \epsilon - m_+(\mathbf{w}_k|D_k), \quad (8)$$

and define the following rule that determines a new training dataset candidate D'_{k+1} .

Rule 1 (New training dataset candidate). At the current time step k with regressor \mathbf{w}_k and training dataset D_k compute $\hat{y}_{k+1} = m_+(\mathbf{w}_k|D_k)$ and $\sigma_+^2 = \sigma_+^2(\mathbf{w}_k|D_k)$. Once the next output y_{k+1} is available, the new data point (\mathbf{w}_k, y_{k+1}) is considered as a candidate for inclusion into the training dataset D_k

If $|e^p| > \bar{e}$ OR $\sigma_+^2 > \bar{\sigma}^2$ **then**

$$D'_{k+1} = D_k \cup (\mathbf{w}_k, y_{k+1})$$

end if

where \bar{e} and $\bar{\sigma}^2$ are prespecified thresholds and D'_{k+1} is the new training dataset candidate for $k+1$.

Thus, if the prediction error e^p is larger than the threshold \bar{e} , the data point is considered to be included in the training data set D_k because the current posterior model is not able to predict the output with the specified accuracy. If it is smaller but the resulting posterior variance $\sigma_+^2(\mathbf{w}|D_k)$ is larger than the threshold $\bar{\sigma}^2$, the data point is also a candidate because the current posterior model is not sufficiently confident in its prediction. This allows to include data points that are relevant to attain a certain prediction quality and effectively allows to limit the necessary number of data points in D_k . This becomes especially important for long operation times and many encountered data points with new information during operation.

Remark 2. Since Rule 1 would also consider outliers for inclusion, we propose to combine it with an additional update rule presented in Theorem 1 (Section 4.3). The application of both update rules is contained in Algorithm 1.

Algorithm 1. Recursive GP model predictive control

MPC Parameters: Prediction horizon N , stage cost $\ell(\cdot)$ with respective parameters, hard input constraint set \mathcal{U} , output constraint set \mathcal{Y} .

rGP Parameters: Prior mean $m(\cdot)$, covariance function $k(\cdot, \cdot)$, initial hyperparameters θ , thresholds \bar{e} and $\bar{\sigma}^2$, maximum number of training points M .

Initialization

Training data set \mathcal{D} .

Optimize hyperparameters θ (7) with initial data set \mathcal{D} .

Initialize GP posterior mean function $m_+(\mathbf{w})$ with covariance matrix \mathbf{K}^{-1} , Cholesky factor \mathbf{R} , and α (Section 3.3).

Compute GP posterior mean gradient $\nabla m_+(\mathbf{w})$ (Section B).

Compute linear GP model at \mathbf{x}_{ref} (Section 5.5).

Compute terminal cost function $V_f(\cdot)$ (Section 5.5).

Recursion

for each time step k **do**

Solve optimal control problem (12) for initial condition \mathbf{x}_k and obtain optimal input sequence $\hat{\mathbf{u}}_{k|k}^*$.

Apply first element $u_k = \kappa_{\text{MPC}}(\mathbf{x}_k | \mathcal{D}_k) = \hat{u}_{k|k}^*$.

Obtain new output y_{k+1} .

Construct new GP data point (\mathbf{w}_k, y_{k+1}) with $\mathbf{w}_k = (\mathbf{x}_k, u_k)$.

Update GP:

Compute $\hat{y}_{k+1} = m_+(\mathbf{w}_k | \mathcal{D}_k)$ and $\sigma_+^2 = \sigma_+^2(\mathbf{w}_k | \mathcal{D}_k)$.

if $|y_{k+1} - \hat{y}_{k+1}| > \bar{e}$ OR $\sigma_+^2 > \bar{\sigma}^2$ **then**

$\mathcal{D}'_{k+1} = \mathcal{D}_k \cup (\mathbf{w}_k, y_{k+1})$.

Using \mathcal{D}'_{k+1} , compute \mathbf{K}' and \mathbf{R}' via (A1).

if number of training points $> M$ **then**

Remove oldest data point and downdate \mathbf{K}' and \mathbf{R}' via (A2).

end if

Compute α' via (9).

if $V_N^*(\mathbf{x}_k | \mathcal{D}'_{k+1}) \leq V_N^*(\mathbf{x}_k | \mathcal{D}_k)$ **then**

$\mathcal{D}_{k+1} = \mathcal{D}'_{k+1}$

Make \mathbf{K}' , \mathbf{R}' , and α' effective.

else

$\mathcal{D}_{k+1} = \mathcal{D}_k$

Reverse \mathbf{K}' , \mathbf{R}' , and α' .

end if

else

$\mathcal{D}_{k+1} = \mathcal{D}_k$

end if

end for

As the available computational power is always limited and depending on the concrete system, this can require the limitation of the maximum number of points in \mathcal{D}_k by a constant $M \in \mathbb{N}$.[‡] If this limit is reached, data points have to be removed to maintain the size of \mathcal{D}_k . Again, different criteria can be employed to determine which data point shall be deleted. For instance, the point in the training dataset with the lowest benefit for the model quality (e.g., the data point that is most accurately predicted under the current posterior) can be deleted. This, however, can be computationally expensive because the prediction has to be evaluated for every of the M training data points at each time instant k . For online implementation, we employ a more simple approach that deletes the oldest point contained in \mathcal{D}_k .

[‡]This approach is also sometimes denoted as *truncated GP*.⁵²

Remark 3. The concept of evolving GPs, in particular the outlined data handling approach, leads to a training dataset D_k that captures the system dynamics in an (evolving) subregion of the whole operating region. Thus, information about already visited regions can be lost when moving towards other regions and have to be regained when visited again. This could be counteracted, for instance, by exploiting multiple GPs for different regions or by GP blending.³⁵

Remark 4. In principle, the smaller the thresholds $\bar{\epsilon}$ and $\bar{\sigma}^2$, the better the prediction. However, then also the overhead for the computational evaluation for adding and removing data points becomes larger. In addition, the smaller the thresholds, the smaller the region in which the training dataset captures the system behavior, given the case that only a finite number of training data points is allowed. Hence, the selection of the thresholds $\bar{\epsilon}$ and $\bar{\sigma}^2$ is an application specific trade-off and might be chosen heuristically by the user. Some general guidelines are, (i) a lower bound for $\bar{\sigma}^2$ is the measurement noise variance, and (ii) $\bar{\epsilon}$ could be chosen proportional to $\frac{1}{M} \sum_{i=1}^M |\mathbf{z} - m_+(\mathbf{w}|D_k)|$, that is, to the mean value of all the absolute values of the prediction errors, based on the current training data set D_k . In the same way $\bar{\sigma}^2$ could be chosen.

3.3 | Avoiding numerical illconditioning for MPC by Cholesky decomposition

The squared exponential covariance function (6) and other smooth covariance functions lead to a poor conditioned covariance matrix \mathbf{K} .^{53,54} This results in numerical problems when computing the inverse \mathbf{K}^{-1} with computational cost $\mathcal{O}(n^3)$, as required for (4a), (4b), or (7). These problems become even worse if (4a) and (4b) are nested within an optimization procedure like model predictive control. One way to alleviate this problem is by adding an additional noise or *jitter* term⁵³ to the diagonal of the covariance matrix. An effective approach however is to avoid the numerical instabilities that arise in the explicit computation of the matrix inverse by performing the required computations using the Cholesky decomposition, which is numerically more stable.

Given a system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ with a symmetric positive matrix \mathbf{A} , we denote the solution by $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} := \mathbf{A} \setminus \mathbf{b}$. The Cholesky decomposition of \mathbf{A} is $\mathbf{A} = \mathbf{R}^T\mathbf{R}$, where $\mathbf{R} = \text{chol}(\mathbf{A})$ is an upper triangular matrix that is called the *Cholesky factor*. It can be used to obtain the solution via $\mathbf{x} = \mathbf{R} \setminus (\mathbf{R}^T \setminus \mathbf{b})$. In order to use the Cholesky factor to solve (4a) and (4b), we define

$$\begin{aligned}\alpha &:= \mathbf{K}^{-1}(\mathbf{z} - m(\mathbf{w})) \\ \beta &:= \mathbf{K}^{-1}k(\mathbf{w}, \mathbf{w}),\end{aligned}$$

which can then be computed with the Cholesky decomposition $\mathbf{K} = \mathbf{R}^T\mathbf{R}$ via

$$\begin{aligned}\alpha &= \mathbf{R} \setminus (\mathbf{R}^T \setminus (\mathbf{z} - m(\mathbf{w}))) \\ \beta &= \mathbf{R} \setminus (\mathbf{R}^T \setminus k(\mathbf{w}, \mathbf{w})).\end{aligned}\tag{9}$$

The computational cost of computing \mathbf{R} is $\mathcal{O}\left(\frac{n^3}{6}\right)$ and the cost of computing α and β is $\mathcal{O}(n^2)$.¹³

If the training data set D_k does not change, the Cholesky decomposition $\mathbf{K} = \mathbf{R}^T\mathbf{R}$ and the computation of α have to be performed only once at the beginning, whereas β has to be recomputed for every new test point \mathbf{w} . If D_k changes, that is, with each inclusion or removal of a data point, the covariance matrix \mathbf{K} has to be updated for an appropriate evaluation of the GP posterior. If a data point is included, a row and column have to be added to \mathbf{K} . If a data point is removed, the respective row and column associated with this point have to be removed. These changes require in principle a full recalculation of the Cholesky factor \mathbf{R} , which is the most expensive computation. To reduce this computational load we employ the approach of Osborne⁵⁴ to recalculate the Cholesky factor recursively, taking advantage of the available factor of the previous step. The precise procedure is outlined in the Appendix in Section A1.

Remark 5. The recursive update of the Cholesky factor can only be applied if the hyperparameters θ do not change because otherwise, every single element of \mathbf{K} changes and a recursive approach is not applicable anymore.

Remark 6. Note that in many works⁵⁵⁻⁵⁷ not the Cholesky decomposition but the covariance matrix inverse \mathbf{K}^{-1} is recursively computed, which is based on the partitioned block inverse using the Woodbury matrix identity. Presumably for the numerical issues outlined above, this approach has never been used in combination with MPC. It has, however, in the signal processing literature, where it is strongly connected to the concept of kernel recursive least-squares.^{56,57}

Due to the recursive nature, both in the data inclusion approach and the Cholesky decomposition, we denote the resulting GP as recursive GP (rGP). The most important steps of the resulting rGP-MPC formulation are presented in Algorithm 1.

4 | GP-BASED OUTPUT FEEDBACK MODEL PREDICTIVE CONTROL

In this section, we present the output feedback model predictive control formulation, based on the rGP NARX model for prediction. We highlight the necessary components and show under which conditions stability can be guaranteed even if the GP model changes online.

4.1 | Prediction model

In Section 4.3 we establish input-to-state stability for the considered system, which is defined using the evolution of the state and not the output. For this reason, we first reformulate the GP output prediction in terms of the NARX state $\hat{\mathbf{x}}_k$. We start by setting $k := k + 1$ in $\hat{\mathbf{x}}_k$ and arrive at

$$\hat{\mathbf{x}}_{k+1} = [\hat{y}_{k+1}, y_k, \dots, y_{k+1-m_y}, u_k, \dots, u_{k+1-m_u}].$$

Since the predicted output \hat{y}_{k+1} is computed by (5) we obtain the NARX prediction model

$$\hat{\mathbf{x}}_{k+1} = \hat{F}(\hat{\mathbf{x}}_k, u_k | \mathcal{D}_k) := [m_+(\mathbf{w}_k | \mathcal{D}_k), y_k, \dots, y_{k+1-m_y}, u_k, \dots, u_{k+1-m_u}], \quad (10)$$

which we also denote as the *nominal* model.

Correspondingly, for the NARX model of the real process (1a) we have

$$\begin{aligned} \mathbf{x}_{k+1} &= [y_{k+1}, y_k, \dots, y_{k+1-m_y}, u_k, \dots, u_{k+1-m_u}] \\ &= [f(\mathbf{x}_k, u_k) + \epsilon, y_k, \dots, y_{k+1-m_y}, u_k, \dots, u_{k+1-m_u}], \end{aligned}$$

and due to (8) this can be reformulated as

$$\begin{aligned} \mathbf{x}_{k+1} &= [m_+(\mathbf{w}_k | \mathcal{D}_k) + e^p, y_k, \dots, y_{k+1-m_y}, u_k, \dots, u_{k+1-m_u}] \\ &= \hat{F}(\mathbf{x}_k, u_k | \mathcal{D}_k) + \mathbf{d}e^p =: F(\mathbf{x}_k, u_k, e^p), \end{aligned} \quad (11)$$

with $\mathbf{d} = [1 \ 0 \ \dots \ 0]^T$, that is, the real NARX model can be represented as the superposition of the nominal/prediction model and the prediction error.

4.2 | MPC optimization problem

Using the prediction model (10), we consider at each time step k the optimization problem

$$\begin{aligned} \min_{\hat{\mathbf{u}}_{k|k}} & V_N(\mathbf{x}_k, \hat{\mathbf{u}}_{k|k}) \\ \text{s.t. } & \forall i \in \mathcal{I}_{0:N-1} : \\ & \hat{\mathbf{x}}_{k+i+1|k} = \hat{F}(\hat{\mathbf{x}}_{k+i|k}, \hat{u}_{k+i|k} | \mathcal{D}_k) \\ & \hat{\mathbf{x}}_{k|k} = \mathbf{x}_k \\ & \hat{u}_{k+i|k} \in \mathcal{U} \\ & \hat{\mathbf{x}}_{k+i|k} \in \mathcal{X}. \end{aligned} \quad (12)$$

The input sequence to be optimized is denoted by $\hat{\mathbf{u}}_{k|k} = \{\hat{u}_{k|k}, \dots, \hat{u}_{k+N-1|k}\}$, N is the prediction horizon, \mathbf{x}_k is the initial condition of the measured NARX state (2), and $\mathcal{X} \subseteq \mathbb{R}^{n_x}$ is the resulting constrained set of the NARX state that is a combination of multiple instances of \mathcal{Y}_h depending on the specific composition of \mathbf{x}_k .^{||} Since \mathcal{Y}_h is compact, the resulting \mathcal{X} is also compact. As cost function in (12) we consider

$$V_N(\mathbf{x}_k, \hat{\mathbf{u}}_{k|k}) = \sum_{i=0}^{N-1} \ell(\hat{\mathbf{x}}_{k+i|k}, \hat{u}_{k+i|k}) + \lambda V_f(\hat{\mathbf{x}}_{k+N|k} - \mathbf{x}_{\text{ref}}),$$

where $V_f(\cdot)$ is the terminal cost function that is weighted by a design parameter $\lambda \geq 1$. The employed positive stage cost is given by

$$\ell(\hat{\mathbf{x}}_k, \hat{u}_k) = \ell_s(\hat{\mathbf{x}}_k - \mathbf{x}_{\text{ref}}, \hat{u}_k - u_{\text{ref}}) + \ell_b(\hat{y}_k),$$

where $\ell_s(\cdot)$ penalizes input and state deviations from the reference and $\ell_b(\cdot)$ is a barrier function that can account for soft output constraints \mathcal{Y}_s . It is defined by

$$\ell_b(\hat{y}_k) \geq \alpha_b(d(\hat{y}_k, \mathcal{Y}_s)),$$

and must satisfy $\ell_b(\hat{y}_k) = 0, \forall \hat{y}_k \in \mathcal{Y}_s$, where $\alpha_b(\cdot)$ is a \mathcal{K} -function and $d(\cdot)$ the distance function as defined in Section 1.

The optimal solution of (12) is denoted by $\hat{\mathbf{u}}_{k|k}^*$, the resulting optimal state sequence by $\hat{\mathbf{x}}_{k|k}^*$. The first element of $\hat{\mathbf{u}}_{k|k}^*$, that is, $\hat{u}_{k|k}^*$, is applied to the process such that we obtain $u_k = \kappa_{\text{MPC}}(\mathbf{x}_k | \mathcal{D}_k) = \hat{u}_{k|k}^*$. Note that the implicitly defined control law $\kappa_{\text{MPC}}(\mathbf{x}_k | \mathcal{D}_k)$ is time-varying, as well as the resulting optimal cost function $V_N^*(\mathbf{x}_k | \mathcal{D}_k) = V_N(\mathbf{x}_k, \hat{\mathbf{u}}_{k|k}^* | \mathcal{D}_k)$, also denoted as *value function*, because they depend on the changing prediction model associated with \mathcal{D}_k . Note furthermore that (12) does not include any explicit terminal region constraint for stability. This makes its solution less computationally expensive, especially if only soft output/state constraints are considered.

4.3 | Stability

Establishing stability in MPC is often based on the use of a terminal cost function $V_f(\cdot)$ and a terminal region \mathcal{X}_f .⁵⁸ Here we employ an approach where the optimal control problem (12) does not require an explicit terminal region \mathcal{X}_f . Instead, we use $V_f(\cdot)$ weighted by a factor λ , as proposed by Limon et al.,⁵⁹ to establish ISS.

Definition 1 (ISS). Consider the closed-loop system $\mathbf{x}_{k+1} = F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k | \mathcal{D}_k), \mathbf{e}_k)$. The set-point \mathbf{x}_{ref} is ISS if there exist a \mathcal{KL} -function $\beta(\cdot, \cdot)$ and a \mathcal{K} -function $\gamma(\cdot)$ such that

$$\|\mathbf{x}_k - \mathbf{x}_{\text{ref}}\| \leq \beta(\|\mathbf{x}_0 - \mathbf{x}_{\text{ref}}\|, k) + \gamma\left(\max_{k \geq 0} \|\mathbf{e}_k\|\right), \quad (13)$$

holds for all initial states \mathbf{x}_0 , errors \mathbf{e}_k , and for all k .

ISS combines nominal stability as well as uniformly bounded influence of uncertainty in a single condition. It implies asymptotic stability of the undisturbed (nominal) system (with $\mathbf{e}_k \equiv 0$) and a bounded effect of the uncertainty on the state evolution. Furthermore, if the error signal \mathbf{e}_k fades, the uncertain system asymptotically converges to the reference point. We therefore consider stability first for the nominal case, i.e., when the prediction/nominal model (10) and the true system (11) are exactly the same. After that, we establish robust stability in the sense of ISS.

4.3.1 | Nominal stability

In the following, let the current deviation from the reference point and the deviation at the next time step be $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{x}_{\text{ref}}$ and $\tilde{\mathbf{x}}^+ = \mathbf{x}^+ - \mathbf{x}_{\text{ref}}$ respectively. This change of coordinates is required if $\mathbf{x}_{\text{ref}} \neq 0$.

^{||}If for instance $\mathbf{x}_k = [y_k, y_{k-1}, y_{k-2}]$, then $\mathcal{X} = \mathcal{Y}_h \times \mathcal{Y}_h \times \mathcal{Y}_h$.

Assumption 1. Assume that

1. the stage cost function $\ell(\mathbf{x}, \mathbf{u})$ is positive definite, that is, $\ell(\mathbf{x}_{\text{ref}}, \mathbf{u}_{\text{ref}}) = 0$ and there exists a \mathcal{K}_{∞} -function $\alpha(\cdot)$ such that $\ell(\mathbf{x}, \mathbf{u}) \geq \alpha(\|\tilde{\mathbf{x}}\|)$ for all $\mathbf{u} \in \mathcal{U}$, and
2. there exists a terminal control law $\kappa_{\text{f}}(\cdot)$ and a control Lyapunov function $V_{\text{f}}(\cdot)$ such that the conditions

$$\alpha_1(\|\tilde{\mathbf{x}}\|) \leq V_{\text{f}}(\tilde{\mathbf{x}}) \leq \alpha_2(\|\tilde{\mathbf{x}}\|),$$

and

$$V_{\text{f}}(\tilde{\mathbf{x}}^+) \leq V_{\text{f}}(\tilde{\mathbf{x}}) - \ell(\tilde{\mathbf{x}} + \mathbf{x}_{\text{ref}}, \kappa_{\text{f}}(\tilde{\mathbf{x}}) + \mathbf{u}_{\text{ref}}),$$

hold for all $\tilde{\mathbf{x}} \in \mathcal{X}_{\text{f}} = \{\tilde{\mathbf{x}} \in \mathbb{R}^{n_{\text{x}}} : V_{\text{f}}(\tilde{\mathbf{x}}) \leq \nu\} \subseteq \mathcal{X}$ with $\nu > 0$ and $\tilde{\mathbf{x}}^+ = \hat{F}(\tilde{\mathbf{x}} + \mathbf{x}_{\text{ref}}, \kappa_{\text{f}}(\tilde{\mathbf{x}}) + \mathbf{u}_{\text{ref}} | \mathcal{D}) - \mathbf{x}_{\text{ref}}$, and where $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ are \mathcal{K}_{∞} -functions. The constant ν is chosen such that $\mathcal{X}_{\text{f}} \subseteq \mathcal{X}$ and $\kappa_{\text{f}}(\tilde{\mathbf{x}}) + \mathbf{u}_{\text{ref}} \in \mathcal{U}$ for all $\tilde{\mathbf{x}} \in \mathcal{X}_{\text{f}}$.

Assumption 1 ensures that the system is locally asymptotically stable on the positive invariant set \mathcal{X}_{f} , while satisfying state and input constraints. It can be satisfied if we have at least a locally valid description of the process at the target point. This can, for instance, be a linearized version of the GP prediction model at the reference (e.g., $\mathcal{D} = \mathcal{D}_{\text{ref}}$), which in turn can then be used to derive a suitable terminal cost and controller. Possible options are then, for instance, the use of a linear-quadratic regulator and/or applying Lyapunov methods (see also Section 5.5). Although it is sufficient to determine the terminal components from the nominal model, one could also consider the design of a robust terminal controller and cost. For instance, using a GP model for the target region one could consider a specific probability bound given by the posterior variance and then based on this design a robust terminal controller.

Theorem 1 (Nominal stability). *Let $\kappa_{\text{MPC}}(\mathbf{x}_k | \mathcal{D}_k)$ be the predictive controller derived from the optimal control problem (12) and let Assumption 1 be satisfied. Furthermore, let \mathcal{D}_k be the training dataset at time k , \mathcal{D}_{k+1} the one that will be used at time $k+1$, and \mathcal{D}'_{k+1} the updated training dataset candidate resulting from Rule 1. If \mathcal{D}_k is updated using the additional rule*

If $V_N^*(\mathbf{x}_k | \mathcal{D}'_{k+1}) \leq V_N^*(\mathbf{x}_k | \mathcal{D}_k)$ **then**

$$\mathcal{D}_{k+1} \leftarrow \mathcal{D}'_{k+1}$$

else

$$\mathcal{D}_{k+1} \leftarrow \mathcal{D}_k$$

end if

then $\forall \lambda \geq 1$, there exists a feasible region $\mathcal{X}_N^0(\lambda) \subseteq \mathcal{X}$ such that $\forall \mathbf{x}_0 \in \mathcal{X}_N^0(\lambda)$ the target \mathbf{x}_{ref} of the nominal closed-loop system $\mathbf{x}_{k+1} = \hat{F}(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k | \mathcal{D}_k))$ is asymptotically stable. The size of the set $\mathcal{X}_N^0(\lambda)$ increases with λ .

Proof. Let $\hat{\mathbf{x}}_{k|k}^* = \{\hat{\mathbf{x}}_{k|k}^*, \hat{\mathbf{x}}_{k+1|k}^*, \dots, \hat{\mathbf{x}}_{k+N|k}^*\}$ be the predicted state sequence that results from applying the optimal input sequence $\hat{\mathbf{u}}_{k|k}^*$. Then we can write the optimal cost for initial condition $\mathbf{x}_k = \hat{\mathbf{x}}_{k|k}^*$ also as $V_N^*(\mathbf{x}_k | \mathcal{D}_k) = V_N(\mathbf{x}_k, \hat{\mathbf{u}}_{k|k}^* | \mathcal{D}_k) = V_N(\hat{\mathbf{x}}_{k|k}^*, \hat{\mathbf{u}}_{k|k}^* | \mathcal{D}_k)$. Let furthermore $\hat{\mathbf{x}}_{k+1|k}^* = \{\hat{\mathbf{x}}_{k+1|k}^*, \dots, \hat{\mathbf{x}}_{k+N+1|k}^*\}$ and $\hat{\mathbf{u}}_{k+1|k}^* = \{\hat{u}_{k+1|k}^*, \dots, \hat{u}_{k+N-1|k}^*, \kappa_{\text{f}}(\hat{\mathbf{x}}_{k+N|k}^* - \mathbf{x}_{\text{ref}}) + \mathbf{u}_{\text{ref}}\}$ be the respective sequences that start at $k+1$ computed at time k , where the last state is given by the terminal control law, that is, $\hat{\mathbf{x}}_{k+N+1|k}^* = \hat{F}(\hat{\mathbf{x}}_{k+N|k}^*, \kappa_{\text{f}}(\hat{\mathbf{x}}_{k+N|k}^* - \mathbf{x}_{\text{ref}}) + \mathbf{u}_{\text{ref}} | \mathcal{D}_k)$.

By Assumption 1 we have that the stage and terminal cost are positive definite. Hence, the cost function $V_N(\mathbf{x}_k, \hat{\mathbf{u}}_{k|k}^*)$ is positive definite. Furthermore we also obtain

$$V_N(\hat{\mathbf{x}}_{k+1|k}^*, \hat{\mathbf{u}}_{k+1|k}^* | \mathcal{D}_{k+1}) \leq V_N(\hat{\mathbf{x}}_{k|k}^*, \hat{\mathbf{u}}_{k|k}^* | \mathcal{D}_{k+1}) - \ell(\mathbf{x}_k, \mathbf{u}_k),$$

by Assumption 1, which is a well-known result in standard MPC (for the derivation see, for instance, Rawlings and Mayne¹ or Rakovic et al.⁶⁰). Given the update rule in Theorem 1 we have

$$V_N(\hat{\mathbf{x}}_{k|k}^*, \hat{\mathbf{u}}_{k|k}^* | \mathcal{D}_{k+1}) - \ell(\mathbf{x}_k, \mathbf{u}_k) \leq V_N(\hat{\mathbf{x}}_{k|k}^*, \hat{\mathbf{u}}_{k|k}^* | \mathcal{D}_k) - \ell(\mathbf{x}_k, \mathbf{u}_k).$$

Combining the previous two equations we obtain

$$V_N(\hat{\mathbf{x}}_{k+1|k}^*, \hat{\mathbf{u}}_{k+1|k}^* | \mathcal{D}_{k+1}) \leq V_N(\hat{\mathbf{x}}_{k|k}^*, \hat{\mathbf{u}}_{k|k}^* | \mathcal{D}_k) - \ell(\mathbf{x}_k, \mathbf{u}_k). \quad (14)$$

Thus, the value function is decreasing even if the prediction model changes. Hence the value function is a Lyapunov function.

Regarding the feasible region, we first review a result of Limon et al.⁵⁹ and show afterwards an extension to the present case. In particular, theorem 3 of Limon et al.⁵⁹ shows for the nominal and time-invariant case of (12) (i.e., constant prediction model and no model-plant mismatch) with value function $V_N^*(\mathbf{x}_k)$ that $\forall \lambda \geq 1$ there exists a feasible region $\mathcal{X}_N(\lambda)$ such that $\forall \mathbf{x}_0 \in \mathcal{X}_N(\lambda)$ the nominal closed-loop system $\mathbf{x}_{k+1} = \hat{F}(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k))$ is recursively feasible and asymptotically stable. The feasible region is characterized by $\mathcal{X}_N(\lambda) = \{\mathbf{x}_k \in \mathbb{R}^{n_x} : V_N^*(\mathbf{x}_k) \leq N \cdot d + \lambda \cdot \nu\}$, where ν is defined in Assumption 1 and d is a positive constant such that $\ell(\mathbf{x}_k, u_k) > d$, $\forall \mathbf{x}_k \notin \mathcal{X}_f$ and $\forall u_k \in \mathcal{U}$. The size of the set $\mathcal{X}_N(\lambda)$ increases with λ .^{**}

In this work, the value function $V_N^*(\mathbf{x}_k | \mathcal{D}_k)$ changes at certain time instances k whenever the data set \mathcal{D}_k changes. For this reason we extend the definition of the feasible region to $\mathcal{X}_N^k(\lambda) = \{\mathbf{x}_k \in \mathbb{R}^{n_x} : V_N^*(\mathbf{x}_k | \mathcal{D}_k) \leq N \cdot d + \lambda \cdot \nu\}$, which then also changes with k . Due to (14), the optimal cost is decreasing for a particular state sequence $\mathbf{x} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k, \dots\}$ with $Nd + \lambda\nu \geq V_N^*(\mathbf{x}_0 | \mathcal{D}_0) \geq V_N^*(\mathbf{x}_1 | \mathcal{D}_1) \geq \dots \geq V_N^*(\mathbf{x}_k | \mathcal{D}_k)$ and therefore $\mathcal{X}_N^k(\lambda)$ is increasing along the state sequence. Thus, if the initial state $\mathbf{x}_0 \in \mathcal{X}_N^0(\lambda)$, then the subsequent states $\mathbf{x}_k \in \mathcal{X}_N^k(\lambda)$ and the optimal control problem is recursively feasible. Hence, the target \mathbf{x}_{ref} is asymptotically stable for the nominal closed-loop system $\mathbf{x}_{k+1} = \hat{F}(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k | \mathcal{D}_k))$. ■

At \mathbf{x}_k (with the current output measurement y_k) the optimal control problem is solved with data set \mathcal{D}_k and the resulting input $u_k = \kappa_{\text{MPC}}(\mathbf{x}_k | \mathcal{D}_k) = \hat{u}_{k|k}^*$ is applied to the system. If the next data point (\mathbf{w}_k, y_{k+1}) is a candidate for updating the GP, the previous optimal cost is recomputed using the updated GP. If the cost does not increase, the GP update becomes effective. Thus, the update rule in Theorem 1 is executed additionally after the data selection process of Rule 1. This is also reflected in Algorithm 1.

Remark 7 (Conflicting objectives). Theorem 1 establishes nominal stability despite a changing training dataset \mathcal{D}_k . In order to determine the new dataset candidate \mathcal{D}'_{k+1} we use Rule 1, whose objective is to refine the current prediction model. Note that also other rules, which utilize different selection criteria for model refinement (e.g., statistical methods, see Section 3.2), can be employed. Now, one could assume that the additional update rule in Theorem 1 is not necessary because with every new data point the prediction model should become more accurate. This is, however, not necessarily the case if, for instance, the output is corrupted by noise or if outliers are present. In both cases, the apparent process behavior differs from the true behavior and it cannot be guaranteed that the prediction model becomes more accurate with every added data point, nor that the value function continues decreasing monotonically. Thus, the objective of the update rule in Theorem 1 is to make sure that safety, in the sense of stability and constraint satisfaction, is guaranteed. This is also illustrated in the simulations section in Figure 10. However, in the same simulations we also see that data points, selected by Rule 1 and which carry valuable information, are discarded by the update rule of Theorem 1 because the decreasing value function condition, and with that stability, could not be guaranteed. In other words, the two objectives of model refinement (expressed by Rule 1) and safety (in the sense of stability, expressed by the update rule of Theorem 1) are conflicting objectives, especially in the case of corrupted measurements. In this work we prioritize safety, thereby sacrificing a bit of the potential of model refinement.

On the basis of the nominal stability result for the online rGP-MPC scheme, we now establish robust stability.

4.3.2 | Robust stability

Based on Theorem 1 we show that the real process controlled by the proposed predictive controller is ISS w.r.t. the prediction error e^p .

Theorem 2 (ISS). *Let $\kappa_{\text{MPC}}(\mathbf{x}_k | \mathcal{D}_k)$ be the predictive controller derived from optimal control problem (12) satisfying Assumption 1 and Theorem 1. If*

- *the nominal model $\hat{F}(\mathbf{x}_k, u_k | \mathcal{D}_k)$ is uniformly continuous in \mathbf{x}_k for all $\mathbf{x}_k \in \mathcal{X}_N^0(\lambda)$, all $u_k \in \mathcal{U}$, and all \mathcal{D}_k during the prediction horizon^{††}, and*

^{**}Note that theorem 3 of Limon et al.⁵⁹ is stated the other way round, that is, for each region $\mathcal{X}_N(\lambda)$ and for all $\mathbf{x}_k \in \mathcal{X}_N(\lambda)$, there exists a $\lambda \geq 1$ such that the nominal closed-loop system is asymptotically stable at \mathbf{x}_{ref} .

^{††}Note that this condition does not prohibit the change of the nominal model from the current time instant k to the next $k + 1$.

- the stage cost function $\ell(\mathbf{x}_k, u_k)$ and the terminal cost function $V_f(\mathbf{x}_k)$ are uniformly continuous in \mathbf{x}_k for all $\mathbf{x}_k \in \mathcal{X}_N^0(\lambda)$ and all $u_k \in \mathcal{U}$,

then the target \mathbf{x}_{ref} of the closed-loop system $\mathbf{x}_{k+1} = F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k | \mathcal{D}_k), e^p)$ is ISS w.r.t. the prediction error e^p in a robust feasible set $\Omega_r^0(\lambda) \subset \mathcal{X}_N^0(\lambda)$ for a sufficiently small μ with $|e^p| < \mu < \infty$. The smaller μ , the larger the set $\Omega_r^0(\lambda)$.

Proof. We first establish the set $\Omega_r^0(\lambda)$ and prove recursive feasibility. Afterwards we prove the ISS property.

Regarding the nature of $\Omega_r^0(\lambda)$ we first review a result of Limon et al.³⁸ and then extend it to our case. Proposition 1 (C2) in³⁸ shows for the time-invariant case of (12) (i.e., for a nonchanging prediction model) that the closed-loop $\mathbf{x}_{k+1} = F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k | \mathcal{D}_k), e^p)$ is robustly feasible for all \mathbf{x}_k in a robust feasible set $\Omega_r(\lambda)$. In particular, it is proven that if $|e^p| < \mu$ with a sufficiently small μ , there exists a r such that $\Omega_r(\lambda) := \{\mathbf{x}_k \in \mathbb{R}^{n_x} : V_N^*(\mathbf{x}_k) \leq r\} \subset \mathcal{X}_N(\lambda)$ is a compact and positive invariant set (where $\mathcal{X}_N(\lambda)$ is the feasible set of the OCP with $e^p \equiv 0$) and for all $\mathbf{x}_k \in \Omega_r(\lambda)$ the resulting predicted state sequence remains in $\Omega_r(\lambda)$. Therefore the state constraints \mathcal{X} do not become active. Hence, for all $\mathbf{x}_0 \in \Omega_r(\lambda)$ the MPC scheme is recursively feasible and the constraints are robustly satisfied. Furthermore, larger values of λ lead to a larger region $\Omega_r(\lambda)$.

In the definition of $\Omega_r(\lambda)$ in Reference 38 the value function $V_N^*(\mathbf{x}_k)$ is time-invariant, whereas in this work $V_N^*(\mathbf{x}_k | \mathcal{D}_k)$ depends on the changing dataset \mathcal{D}_k . For this reason we extend the definition of the robust feasible region to $\Omega_r^k(\lambda) = \{\mathbf{x}_k \in \mathbb{R}^{n_x} : V_N^*(\mathbf{x}_k | \mathcal{D}_k) \leq r\} \subset \mathcal{X}_N^k(\lambda)$, which then also changes with k . In order for $\Omega_r^k(\lambda) \subset \mathcal{X}_N^k(\lambda)$ to hold we require $0 < r < N \cdot d + \lambda \cdot v$ because $\mathcal{X}_N^k(\lambda) = \{\mathbf{x}_k \in \mathbb{R}^{n_x} : V_N^*(\mathbf{x}_k | \mathcal{D}_k) \leq N \cdot d + \lambda \cdot v\}$. Thereby, $N \cdot d + \lambda \cdot v$ establishes an upper bound for r . Like the feasible set $\mathcal{X}_N^k(\lambda)$ (see the proof to Theorem 1), also $\Omega_r^k(\lambda)$ increases with λ and in particular with k along a particular state sequence $\mathbf{x} = \{\mathbf{x}_0, \mathbf{x}_1, \dots\}$. Therefore, if the initial state $\mathbf{x}_0 \in \Omega_r^0(\lambda)$, then the subsequent states $\mathbf{x}_k \in \Omega_r^k(\lambda)$ and the state constraints do not become active. The existence of $\Omega_r^0(\lambda)$ is established by proposition 1 (C2) in Reference 38 (as outlined above) and therefore, if $\mathbf{x}_0 \in \Omega_r^0(\lambda)$ then (12) is recursively feasible and the constraints are robustly satisfied.

Now we show that the closed-loop system $\mathbf{x}_{k+1} = F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k | \mathcal{D}_k), e^p)$ is ISS w.r.t. the prediction error e^p . To this end we start by showing that the cost function $V_N(\mathbf{x}_k, \hat{\mathbf{u}}_{k|k})$ is uniformly continuous in \mathbf{x}_k . Since the nominal model $\hat{F}(\mathbf{x}_k, u_k | \mathcal{D}_k)$ is uniformly continuous in \mathbf{x}_k during the prediction horizon, there exists a \mathcal{K} -function $\sigma_x(\cdot)$ such that $\|\hat{F}(\mathbf{x}_k, u_k | \mathcal{D}_k) - \hat{F}(\mathbf{z}_k, u_k | \mathcal{D}_k)\| \leq \sigma_x(\|\mathbf{x}_k - \mathbf{z}_k\|)$ for all $\mathbf{x}_k, \mathbf{z}_k \in \mathcal{X}_N^0(\lambda)$, all $u_k \in \mathcal{U}$, and for a given data set \mathcal{D}_k . In accordance with lemma 2 in Reference 38, the predicted state evolution then satisfies $\|\hat{\mathbf{x}}_{k+i|k} - \hat{\mathbf{z}}_{k+i|k}\| \leq \sigma_x^i(\|\mathbf{x}_k - \mathbf{z}_k\|)$ for $i \in \mathcal{I}_{0:N-1}$. Furthermore, since the stage and terminal cost are uniformly continuous in \mathbf{x}_k , there exists a couple of \mathcal{K} -functions $\sigma_\ell(\cdot), \sigma_{V_f}(\cdot)$ such that $\|\ell(\mathbf{x}_k, u_k) - \ell(\mathbf{z}_k, u_k)\| \leq \sigma_\ell(\|\mathbf{x}_k - \mathbf{z}_k\|)$ and $\|V_f(\mathbf{x}_k) - V_f(\mathbf{z}_k)\| \leq \sigma_{V_f}(\|\mathbf{x}_k - \mathbf{z}_k\|)$ for all $\mathbf{x}_k, \mathbf{z}_k \in \mathcal{X}_N^0(\lambda)$ and all $u \in \mathcal{U}$. Combining these properties we obtain

$$\begin{aligned} \|V_N(\mathbf{x}_k, \hat{\mathbf{u}}_{k|k}) - V_N(\mathbf{z}_k, \hat{\mathbf{u}}_{k|k})\| &\leq \sum_{i=0}^{N-1} \|\ell(\hat{\mathbf{x}}_{k+i|k}, \hat{\mathbf{u}}_{k+i|k}) - \ell(\hat{\mathbf{z}}_{k+i|k}, \hat{\mathbf{u}}_{k+i|k})\| + \|V_f(\hat{\mathbf{x}}_{k+N|k}) - V_f(\hat{\mathbf{z}}_{k+N|k})\| \\ &\leq \sum_{i=0}^{N-1} \sigma_\ell \circ \sigma_x^i(\|\mathbf{x}_k - \mathbf{z}_k\|) + \sigma_{V_f} \circ \sigma_x^N(\|\mathbf{x}_k - \mathbf{z}_k\|) =: \sigma_V(\|\mathbf{x}_k - \mathbf{z}_k\|), \end{aligned}$$

where \circ denotes the concatenation of functions (e.g. $\sigma_1 \circ \sigma_2(x) = \sigma_1(\sigma_2(x))$) and $\sigma_V(\cdot)$ is a \mathcal{K} -function. Therefore the cost function is uniformly continuous in \mathbf{x}_k for all $\mathbf{x}_k \in \mathcal{X}_N^0(\lambda)$ and all $\hat{\mathbf{u}}_{k|k}$.

As shown, for every $\mathbf{x}_k \in \Omega_r^0(\lambda)$ the state constraints do not become active. Thus, the optimal solution $\hat{\mathbf{u}}_{k|k}^*$ of (12) is feasible for every $\mathbf{x}_0 \in \Omega_r^0(\lambda)$ and we obtain

$$\|V_N^*(\mathbf{x}_k | \mathcal{D}_k) - V_N^*(\mathbf{z}_k | \mathcal{D}_k)\| = \|V_N(\mathbf{x}_k, \hat{\mathbf{u}}_{k|k}^*) - V_N(\mathbf{z}_k, \hat{\mathbf{u}}_{k|k}^*)\| \leq \sigma_V(\|\mathbf{x}_k - \mathbf{z}_k\|).$$

Therefore, the value function $V_N^*(\mathbf{x}_k | \mathcal{D}_k)$ is also uniformly continuous in \mathbf{x}_k for all $\mathbf{x}_k \in \Omega_r^0(\lambda)$ and a given dataset \mathcal{D}_k .

At last we show that the value function is a ISS-Lyapunov function. Since $V_N^*(\mathbf{x}_k | \mathcal{D}_k)$ is a Lyapunov function for the nominal system (Theorem 1) there exists \mathcal{K}_∞ -functions $\alpha_1(\cdot), \alpha_2(\cdot), \alpha_3(\cdot)$, such that $\alpha_1(\|\mathbf{x}_k\|) \leq V_N^*(\mathbf{x}_k | \mathcal{D}_k) \leq \alpha_2(\|\mathbf{x}_k\|)$ and $V_N^*(\mathbf{x}_{k+1} | \mathcal{D}_{k+1}) - V_N^*(\mathbf{x}_k | \mathcal{D}_k) \leq -\alpha_3(\|\mathbf{x}_k\|)$. Moreover, from (11) we have that $F(\mathbf{x}_k, u_k, e^p)$ is affine in e^p and is therefore uniformly continuous in e^p . Then, there exists a \mathcal{K} -function $\sigma_e(\cdot)$ such that $\|F(\mathbf{x}_k, u_k, e_1) - F(\mathbf{x}_k, u_k, e_2)\| \leq \sigma_e(|e_1 - e_2|)$ for

all $\mathbf{x}_k \in \mathcal{X}_N^0(\lambda)$, all $u_k \in \mathcal{U}$, and all $|e^p| \leq \mu$. From these facts, it can be inferred that

$$\begin{aligned}
V_N^*(\mathbf{x}_{k+1}|\mathcal{D}_{k+1}) - V_N^*(\mathbf{x}_k|\mathcal{D}_k) &= V_N^*(F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k), e^p)|\mathcal{D}_{k+1}) - V_N^*(\mathbf{x}_k|\mathcal{D}_k) \\
&= V_N^*(F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k), e^p)|\mathcal{D}_{k+1}) - V_N^*(F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k), 0)|\mathcal{D}_{k+1}) \\
&\quad + V_N^*(F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k), 0)|\mathcal{D}_{k+1}) - V_N^*(\mathbf{x}_k|\mathcal{D}_k) \\
&\leq \|V_N^*(F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k), e^p)|\mathcal{D}_{k+1}) - V_N^*(F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k), 0)|\mathcal{D}_{k+1})\| - \alpha_3(\|\mathbf{x}_k\|) \\
&\leq \sigma_V(\|F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k), e^p) - F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k), 0)\|) - \alpha_3(\|\mathbf{x}_k\|) \\
&\leq \sigma_V \circ \sigma_e(|e^p|) - \alpha_3(\|\mathbf{x}_k\|).
\end{aligned}$$

Hence, $V_N^*(\mathbf{x}_k|\mathcal{D}_k)$ is a ISS-Lyapunov function and the closed-loop system $\mathbf{x}_{k+1} = F(\mathbf{x}_k, \kappa_{\text{MPC}}(\mathbf{x}_k|\mathcal{D}_k), e^p)$ is ISS w.r.t. e^p for all $\mathbf{x}_0 \in \Omega_r^0(\lambda)$. ■

Remark 8 (Differences in soft and hard output constraints). In the case of soft constraints \mathcal{Y}_s , the proposed controller ensures robust stability and constraint satisfaction for all initial states that lie in the feasible region $\mathcal{X}_N^0(\lambda)$ of the optimal control problem. In the case of hard constraints \mathcal{Y}_h , the proposed controller ensures robust stability and constraint satisfaction for all initial states in a robust feasible set $\Omega_r^0(\lambda) \subset \mathcal{X}_N^0(\lambda)$ where the constraints are not active. Thus, from a practical point of view, if in the soft constraints case the initial state \mathbf{x}_k leads to a feasible solution, we then have $\mathbf{x}_k \in \mathcal{X}_N^0(\lambda)$ and the above guarantees hold. If in the hard constraints case the initial state \mathbf{x}_k leads to a feasible solution, then we also have $\mathbf{x}_k \in \mathcal{X}_N^0(\lambda)$. However, in that case, one cannot be sure if also $\mathbf{x}_k \in \Omega_r^0(\lambda)$ is satisfied. If $\mathbf{x}_k \notin \Omega_r^0(\lambda)$, then feasibility might be lost at one point. Thus, for safety critical applications the set $\Omega_r^0(\lambda)$ would required to be known in order to check $\mathbf{x}_k \in \Omega_r^0(\lambda)$, which is challenging because $\Omega_r^0(\lambda)$ (as well as $\mathcal{X}_N^0(\lambda)$) can in general not be computed but has to be estimated via simulations.¹ However, this issue could be circumvented if the hard constraints were tightened,³⁸ thereby enlarging $\Omega_r^0(\lambda)$.

Remark 9. Notice that the ISS property is based on the uniform continuity of the optimal cost function and this does not depend on the size of the error signal. Hence, even if e^p is larger than μ for a short period of time in which we assume that the feasibility of the optimal control problem is not lost, that is, \mathbf{x}_k remains in $\mathcal{X}_N^k(\lambda)$ and ends in $\Omega_r^k(\lambda)$, then the closed-loop ISS property and constraint satisfaction will still hold.

Remark 10 (Generalization). Theorems 1 and 2 are independent of the control input dimension and also hold for general errors \mathbf{e} independent of the concrete structure of the state \mathbf{x}_k , that is, whether \mathbf{x}_k is a vector comprised of NARX states or of physical states. Thus, the theorems also include the multi-input multi-output case. In addition, as long as the presented assumptions are satisfied, in particular the update rule in Theorem 1, the stability results also hold for the case of online hyperparameter optimization and even further, for general prediction models $\hat{F}(\mathbf{x}_k, u_k|\mathcal{D}_k)$ that are updated online, that is, the stability guarantees are not confined to the use of GP prediction models.

A necessary condition in Theorem 2 is that the nominal model $\hat{F}(\mathbf{x}_k, u_k|\mathcal{D}_k)$ is uniformly continuous in \mathbf{x}_k for all $\mathbf{x}_k \in \mathcal{X}_N^0(\lambda)$, all $u_k \in \mathcal{U}$, and all \mathcal{D}_k during the prediction horizon. In the case of GPs, this can be guaranteed by the following proposition.

Proposition 1 (GP Uniform Continuity¹⁸). *The nominal model (10) is uniformly continuous in \mathbf{x}_k if $\hat{f}(\mathbf{x}_k, u_k) = m_+(\mathbf{w}_k|\mathcal{D})$ is uniformly continuous in \mathbf{x}_k . Since the prior mean $m(\mathbf{w}_k)$ is added to $m_+(\mathbf{w}_k|\mathcal{D})$, the prior mean has to be uniformly continuous in $\mathbf{x}_k^{\ddagger\dagger}$. Then, one way to ensure that $m_+(\mathbf{w}_k|\mathcal{D})$ is uniformly continuous in \mathbf{x}_k , is to employ continuously differentiable kernels (e.g., the squared exponential covariance function, the Matérn class covariance function with appropriate hyperparameters, or the rational quadratic covariance function). In that case the process is mean square differentiable,^{13,61} that is, the posterior mean function is differentiable and therefore also uniformly continuous^{§§}.*

Remark 11. Although not required for Theorem 2, note that uniform continuity of the process $F(\mathbf{x}_k, u_k, e^p) = \hat{F}(\mathbf{x}_k, u_k|\mathcal{D}_k) + \mathbf{d}e^p$ in \mathbf{x}_k is ensured if $\hat{F}(\mathbf{x}_k, u_k|\mathcal{D}_k)$ is uniformly continuous in \mathbf{x}_k , which can be established via Proposition 1.

^{††}The prior mean is usually specified by the user and often set to zero. Thus uniform continuity of $m(\mathbf{w})$ is not an issue.

^{§§}Continuous differentiability is a stronger assumption than uniform continuity.

Resulting prediction errors

We finish this section with a discussion on the prediction error $e^p = y_k - \hat{y}_k$. According to Theorem 2, the smaller the error bound $|e^p| \leq \mu$, the larger the feasible set $\Omega_r^0(\lambda)$. Since the noise ϵ (affecting y_k) is in practice bounded by a finite $\bar{\epsilon}$, the error bound μ is finite if \hat{y}_k is finite (given of course that the original process y_k is finite), which translates to the necessity that the GP posterior mean (4a) is bounded.

From a theoretical point of view, such a bound exists under certain conditions. Note that the posterior mean (with zero prior mean $m(\mathbf{w}) = 0$) can also be expressed via $m_+(\mathbf{w}^*|\mathcal{D}) = \sum_{i=1}^n \alpha_i k(\mathbf{w}_i, \mathbf{w}^*)$, with $\mathbf{w}_i \in \mathbf{w}$, as a linear combination of n kernel functions¹³ that determines a *reproducing kernel Hilbert space* (RKHS). As shown in Steinwart and Christmann,¹⁹ a bound in the RKHS exists if *universal* kernels are employed. One such kernel is, for instance, the squared exponential covariance function[¶] (6) for which the existence of a bound had already been shown by Park and Sandberg.⁶² De Nicolao and Pillonetto⁶³ have presented a very similar result when modeling the impulse response via a spline kernel. The result has also been used in Pillonetto and Chiuso.⁶⁴ Furthermore, Engel⁶⁵ and Srinivas et al⁶⁶ provide ways to explicitly compute the bound, though only with high probability.

In practice, however, $m_+(\mathbf{w}^*|\mathcal{D})$ will generally be bounded assuming that the employed GP prior is well chosen and sufficiently informative training data \mathcal{D} is used. Thus, the actual bound depends on the designer's choices regarding the particular employed GP model and the involved tuning parameters. Among these, in particular the thresholds for the prediction error and posterior variance for the presented rGP approach.

5 | SIMULATIONS

In this section, we provide simulation results for the presented rGP-MPC scheme and consider a continuous stirred-tank reactor (CSTR) as simulation case study. We present the model equations, the training dataset generation, and the terminal components for the MPC based on the linearized GP posterior mean function. The closed-loop simulations involve investigations regarding the tuning parameters of the rGP-MPC, the influence of different initial training data sets, as well as comparisons with other MPC controllers.

5.1 | Continuous stirred-tank reactor

As exemplary case study, we consider the CSTR, where a substrate A is converted into product B .⁶⁷ The following set of differential equations describes the reactor dynamics:

$$\dot{C}_A(t) = \frac{q_0}{V} (C_{Af} - C_A(t)) - k_0 \exp\left(\frac{-E}{RT(t)}\right) C_A(t). \quad (15a)$$

$$\begin{aligned} \dot{T}(t) = & \frac{q_0}{V} (T_f - T(t)) - \frac{\Delta H_r k_0}{\rho C_p} \exp\left(\frac{-E}{RT(t)}\right) C_A(t) \\ & + \frac{UA}{V\rho C_p} (T_c(t) - T(t)). \end{aligned} \quad (15b)$$

$$\dot{T}_c(t) = \frac{T_r(t) - T_c(t)}{\tau}. \quad (15c)$$

The coolant temperature reference $T_r(\text{K})$ is the input and the concentration $C_A(\text{mol/l})$ the output, that is, $u = T_r$ and $y = C_A$. The tank and coolant temperatures are T and T_c , respectively. The model parameters are given in Table B1.

5.2 | Training datasets

A raw data set \mathcal{D}_{raw} (depicted in Figure 2) is generated using the plant (15). The data points (z_i, \mathbf{w}_i) consist of values of $(y_{k+1}, y_k, \dots, y_{k-my}, u_k, \dots, u_{k-m_u})$, where $z = y_{k+1}$ is going be the GP output and $\mathbf{w} = (y_k, \dots, y_{k-m_y}, u_k, \dots, u_{k-m_u})$

[¶]The squared exponential covariance function is sometimes also denoted as *Gaussian radial basis function*. Especially in the field of neural networks or support vector machines.

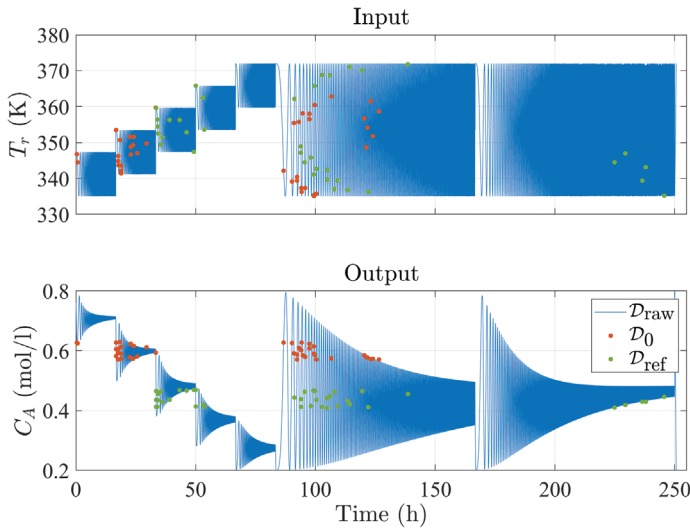


FIGURE 2 Training datasets: The raw data set D_{raw} was generated by chirp signals on the input. The sets D_0 and D_{ref} are local neighborhoods of the initial point y_0 and the reference point y_{ref} and their associated inputs

its corresponding regressor. Based on this data, we generate the three training data sets D_0 , D_{ref} , and D_{comb} . The set D_0 is a local subset around the initial point $y_0 = C_A = 0.6$ mol/l. The associated input is $u_0 = T_r = 353.5$ K. The set D_{ref} is a local subset around the target reference point $y_{\text{ref}} = C_A = 0.439$ mol/l with associated input $u_{\text{ref}} = T_r = 356$ K. The set $D_{\text{comb}} = D_0 \cup D_{\text{ref}}$ is the union of the two sets.

The sets D_0 and D_{ref} are generated by selecting first all points $z = y_{k+1}$ (and their corresponding \mathbf{w}) that are located within a local neighborhood of the respective set-points and second, by reducing the number of points via exclusion of those that add only little information. For a given data point (z_i, \mathbf{w}_i) , all following $(z_j, \mathbf{w}_j), j > i$, are removed, for which $\|\mathbf{w}_i - \mathbf{w}_j\| < \bar{w}$ with a chosen threshold \bar{w} . As a result, the sets are less dense but still contain enough informative data points. The thresholds for D_0 and D_{ref} are chosen such that both sets contain approximately 40 data points.

Remark 12. All input and output values are given in the original units of the system (15). However, it is beneficial for the modeling process with the GP to normalize the input–output data to the interval $[0, 1]$.

5.3 | GP prediction model

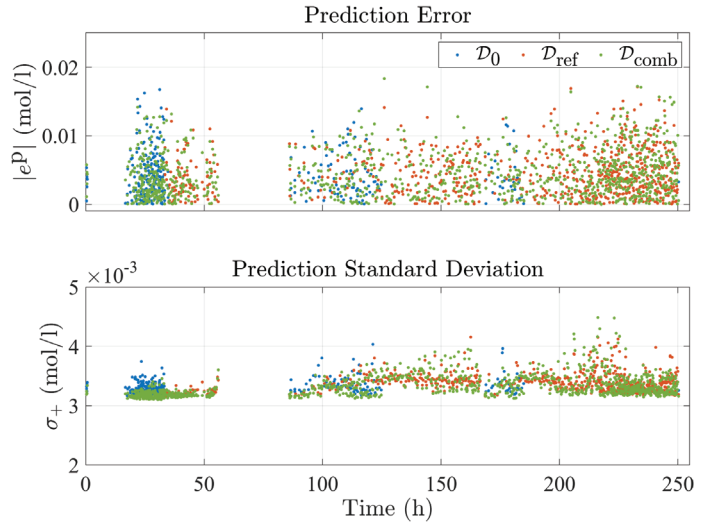
For the GP prior we employ a constant mean function with constant c . Since the underlying process equations are smooth and to obtain the universal approximation property (see Section 4.3.2) we employ the covariance function (6) with regressor $\mathbf{w} = [y_k, y_{k-1}, y_{k-2}, u_k]$. According to (2), the NARX state is then $\mathbf{x}_k = [y_k, y_{k-1}, y_{k-2}]^T$. The hyperparameters are $\theta = \{c, l_1, l_2, l_3, l_4, \sigma_f^2\}$ and are computed offline via maximization of (7) for each of the three datasets D_0 , D_{ref} , and D_{comb} . We obtain three sets of hyperparameters, respectively (Table B2) and with that three different GP prediction models that use the same prior but different training data sets and hyperparameters. The cross-validation results of these different GP models are shown in Figure 3, where we select test points throughout the regions of the respective training datasets. Test points are chosen such that they are not part of D_0 , D_{ref} , or D_{comb} . As can be seen, appropriate GP predictions are achieved with prediction error $e^p < \bar{e} = 0.02$ mol/l and posterior standard deviation $\sigma_+ < \bar{\sigma}^2 = 5 \cdot 10^{-3}$ mol/l for all three GPs.

5.4 | Optimal control problem

The continuous-time model (15) is discretized with Euler's method and a sampling time of $T_s = 0.5$ min. The input constraints are $\mathcal{U} = \{335K \leq T_r \leq 372K\}$, the output constraints $\mathcal{Y}_h = \{0.35\} \text{ mol/L} \leq C_A \leq 0.65\} \text{ mol/l}$. We add measurement noise $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$ to the output data with $\sigma_n^2 = 0.003^2$, which we furthermore bound^{***} by $\pm 4\sigma_n$. The employed

^{***}According to the considered system class we add Gaussian noise with bounded support. Since GPs are based on Gaussian noise with unbounded support, there is a small difference, which could be accounted for by GP warping.⁴³ However, due to the large bound of four SDs, the difference is so small that the following simulation results are equal to the ones with unbounded noise.

FIGURE 3 Cross-validation results: Top, the prediction error e^p (8) is depicted. Bottom, the posterior SD $\sigma_+(\mathbf{w}) = \sqrt{\sigma_+^2(\mathbf{w})}$



quadratic stage cost is given by

$$\ell_s(\mathbf{x}_k, u_k) = \|\mathbf{x}_k - \mathbf{x}_{\text{ref}}\|_Q^2 + \|u_k - u_{\text{ref}}\|_R^2,$$

with $Q = \text{diag}(100, 0, 0)$, and $R = 5$. The prediction horizon is set to $N = 5$. The resulting optimal control problem is solved in MATLAB using `fmincon`.

5.5 | Terminal controller and cost function

The terminal controller $\kappa_f(\cdot)$ and cost function $V_f(\cdot)$ can be determined arbitrarily, as long as the assumptions in Section 4.3 are satisfied. We choose the terminal controller as $\kappa_f(\mathbf{x}) = \mathbf{k}^T(\mathbf{x} - \mathbf{x}_{\text{ref}}) + u_{\text{ref}}$ and the terminal cost function as $V_f(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_{\text{ref}}\|_P^2$, where $\mathbf{k} \in \mathbb{R}^3$ and $\mathbf{P} \in \mathbb{R}^{3 \times 3}$ are computed using the linearization of the prediction model (10) with the GP model, based on the training dataset D_{ref} obtained near the reference \mathbf{x}_{ref} .

The linearization of the nominal NARX model $\mathbf{x}_{k+1} = \hat{F}(\mathbf{x}_k, u_k)$ with $\mathbf{x}_k = [y_k, y_{k-1}, y_{k-2}]$ takes the form

$$\begin{bmatrix} y_{k+1} \\ y_k \\ y_{k-1} \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}}_A \begin{bmatrix} y_k \\ y_{k-1} \\ y_{k-2} \end{bmatrix} + \underbrace{\begin{bmatrix} b_1 \\ 0 \\ 0 \end{bmatrix}}_b u_k.$$

As the next output is computed using the GP, that is, $y_{k+1} = m_+(\mathbf{w}_k)$, the parameters $a_{11}, a_{12}, a_{13}, b_1$ can be determined using the posterior mean gradient derived in the appendix, Section B1. In particular we have $[a_{11}, a_{12}, a_{13}, b_1] = \nabla m_+(\mathbf{w}_{\text{ref}})^T$ with $\mathbf{w}_{\text{ref}} = [y_{\text{ref}}, y_{\text{ref}}, y_{\text{ref}}, u_{\text{ref}}]$. The resulting linear model becomes

$$\mathbf{x}_{k+1} = \begin{bmatrix} 0.162 & 0.005 & -0.012 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} -0.034 \\ 0 \\ 0 \end{bmatrix} u_k. \tag{16}$$

We define the feedback vector as $\mathbf{k} = \mathbf{P}\mathbf{s}$ with $\mathbf{s} \in \mathbb{R}^3$, $\mathbf{P} = \mathbf{P}^T > 0$, and $\mathbf{G} = \mathbf{P}^{-1}$. We furthermore define the state constraint set $\mathcal{X} = \mathcal{Y}_h \times \mathcal{Y}_h \times \mathcal{Y}_h$ and reformulate \mathcal{X} and \mathcal{U} as polyhedral sets of the form $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^{n_x} : \mathbf{q}_i^T \mathbf{x} \leq r_i, i = 1, \dots, n_x\}$ and $\mathcal{U} = \{u \in \mathbb{R}^{n_u} : v_l^T u \leq t_l, l = 1, \dots, n_u\}$, where n_x and n_u are the respective numbers of inequalities. Then, we compute \mathbf{s} and \mathbf{P} offline by solving the semidefinite optimization problem⁷

$$\begin{aligned}
& \max_{\mathbf{G}, \mathbf{s}} \log(\det(\mathbf{G})) \\
& \text{s.t. } \mathbf{G} = \mathbf{G}^T > 0 \\
& \begin{bmatrix} \mathbf{G} & (\mathbf{A}\mathbf{G} + \mathbf{b}\mathbf{s}^T)^T \\ (\mathbf{A}\mathbf{G} + \mathbf{b}\mathbf{s}^T) & \mathbf{G} \end{bmatrix} \geq 0 \\
& \begin{bmatrix} \mathbf{G} & (\mathbf{G}\mathbf{q}_i) \\ (\mathbf{G}\mathbf{q}_i)^T & r_i^2 \end{bmatrix} \geq 0, \quad \forall i \in \{1, \dots, n_x\} \\
& \begin{bmatrix} \mathbf{G} & (\mathbf{s}\mathbf{v}_l) \\ (\mathbf{s}\mathbf{v}_l)^T & t_l^2 \end{bmatrix} \geq 0, \quad \forall l \in \{1, \dots, n_U\},
\end{aligned} \tag{17}$$

and obtain

$$\mathbf{k}^T = \begin{bmatrix} 1.745 & 0.082 & -0.001 \end{bmatrix} \text{ and } \mathbf{P} = \begin{bmatrix} 16.38 & -0.556 & -0.066 \\ -0.556 & 16.32 & -0.554 \\ -0.066 & -0.554 & 16.30 \end{bmatrix}.$$

The optimization problem (17)⁷ results from using the Schur complement in combination with the discrete-time Lyapunov equation and the support function concept of closed convex sets. The resulting \mathbf{s} and \mathbf{P} are such that the closed-loop linearized system is asymptotically stable in $\mathcal{X}_f = \{\mathbf{x} \in \mathbb{R}^{n_x} : V_f(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_{\text{ref}}\|_{\mathbf{P}}^2 \leq 1\} \subseteq \mathcal{X}$ and $\mathbf{k}\mathcal{X}_f \subseteq \mathcal{U}$.

Remark 13. It has been proven in the literature that the quadratic Lyapunov function holds for the nonlinear system in a certain neighborhood of the equilibrium point. The terminal region definition $\mathcal{X}_f = \{\mathbf{x} \in \mathbb{R}^{n_x} : V_f(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_{\text{ref}}\|_{\mathbf{P}}^2 \leq \nu\}$, parameterized with ν , could be used to characterize this neighborhood. Then we would need to take the nonlinear remainder term into account to calculate a particular value for ν , which would require the solution of a global optimization problem. Such a problem could be solved by using scenarios or a Monte Carlo approach. However, since the optimal control problem does not need the terminal region constraint, ν is not required.

5.6 | Simulation results

First, we simulate the set-point change from (u_0, y_0) to $(u_{\text{ref}}, y_{\text{ref}})$ and compare the closed-loop results of the rGP-MPC, a batch GP approach (bGP-MPC) that uses a fixed training dataset, and an output feedback MPC scheme (oMPC) that uses the model Equation (15) and acts as a performance bound. We evaluate the performance for the three cases, where D_0 , D_{ref} , and D_{comb} are used as initial training data sets. The bGP and rGP are initialized with the same initial training data and hyperparameters but the rGP updates its training dataset during operation. We set $\bar{e} = \bar{\sigma}^2 = 0$ such that every data point is considered as a candidate for inclusion^{†††} with no upper limit on the number of data points. Hence, no points are removed. Due to the stochastic nature of the noise component, we simulate each case $N_{\text{sim}} = 50$ times. The results are depicted in Figures 4–6. To quantify the performance we employ the measure

$$\bar{V} = \frac{1}{N_{\text{sim}}} \sum_{j=0}^{N_{\text{sim}}} \sum_{k=0}^{N_{\text{step}}} \ell(\mathbf{x}_k^j, \mathbf{u}_k^j), \tag{18}$$

which averages the stage costs of the resulting state and input sequences over all time steps $k \in \{0, 1, \dots, N_{\text{step}}\}$, as well as the individual simulations $j \in \{1, 2, \dots, N_{\text{sim}}\}$. The resulting \bar{V} values are presented in Table B3.

As expected, the oMPC scheme that uses the true model performs best and always the same (see Table B1) because it does not depend on any training data points. The rGP outperforms the bGP in the D_0 and D_{ref} cases due to the additional information gained during operation. The performance difference is especially large for D_{ref} , where the bGP, throughout

^{†††}Not every data point is added due to the update rule of Theorem 1.

FIGURE 4 Comparison of the three model predictive control schemes for the case of initial training data \mathcal{D}_0 . Thin lines represent individual simulations, thick lines represent mean values

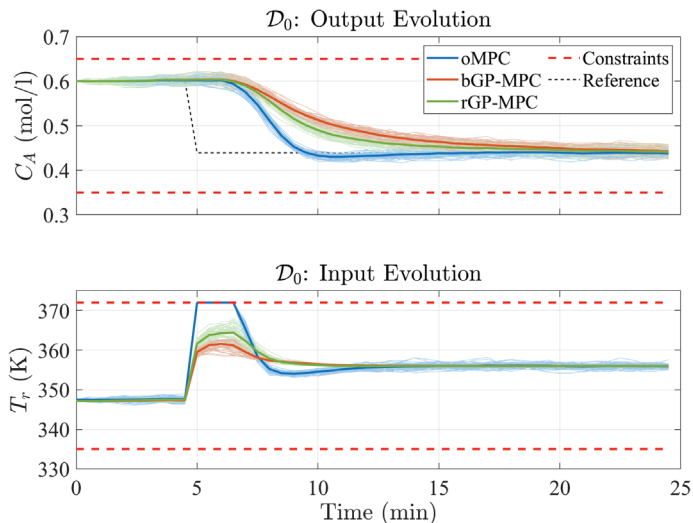
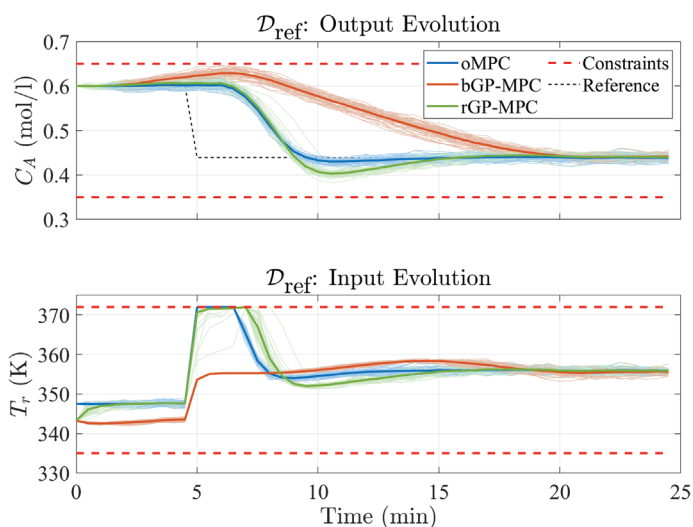


FIGURE 5 Comparison of the three model predictive control schemes for the case of initial training data \mathcal{D}_{ref} . Thin lines represent individual simulations, thick lines represent mean values



the whole operation, has only data points at the reference at its disposal but not at the initial condition. The rGP performs significantly better due to the added data points at the beginning of operation. In the $\mathcal{D}_{\text{comb}}$ case, the rGP and bGP performance is almost the same for the employed training data points.

Remark 14. The previous simulation results suggest that one should in general prefer the \mathcal{D}_{ref} case over the other cases, which is convenient for the used MPC scheme because knowledge at the reference is required anyway to determine the terminal cost and controller. Furthermore, this also suggests a practical rule for offline hyperparameter determination, namely that the hyperparameters should be optimized for a dataset that contains the target reference.

In the second set of simulations, we investigate the influence of different thresholds used in Rule 1, that is, different values for the maximum prediction error \bar{e} and the maximum prediction variance $\bar{\sigma}^2$. To this end, we start with Figure 7 that combines the rGP results of the previous figures for the three training data cases, together with the now plotted evolution of the prediction error e^p and the prediction variance σ_+^2 . In particular the prediction variance illustrates nicely the difference between the three cases. In the case of \mathcal{D}_0 , the variance is small at the beginning and increases around $t = 8$ min when the system leaves the neighborhood of the initial condition and moves towards the reference. The same holds, but the other way round, for the case with \mathcal{D}_{ref} , where the initial ($t < 3$ min) large error and variance is caused by their computation before the first data points are added to the training set. The prediction error bound μ is 0.033, 0.021, and 0.024 for the cases \mathcal{D}_0 , \mathcal{D}_{ref} , and $\mathcal{D}_{\text{comb}}$, respectively.

Figures 8 and 9 show results for different threshold values, where we focus for the sake of brevity on the simulation case with \mathcal{D}_{ref} . The results illustrate that instead of adding all data points, almost the same closed-loop performance can

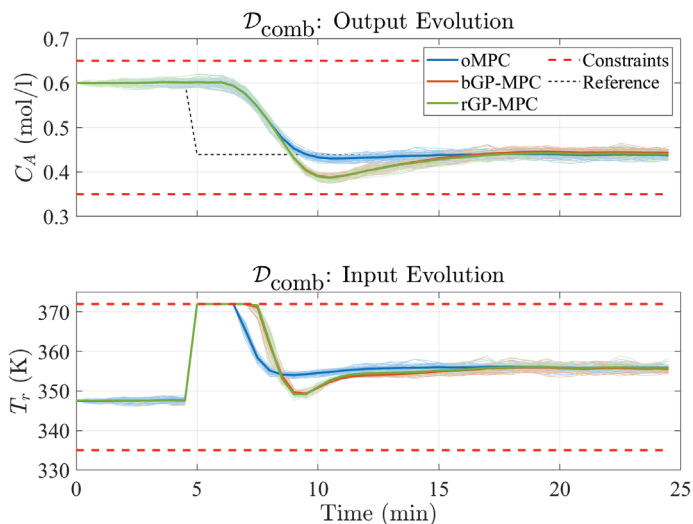


FIGURE 6 Comparison of the three model predictive control schemes for the case of initial training data $\mathcal{D}_{\text{comb}}$. Thin lines represent individual simulations, thick lines represent mean values

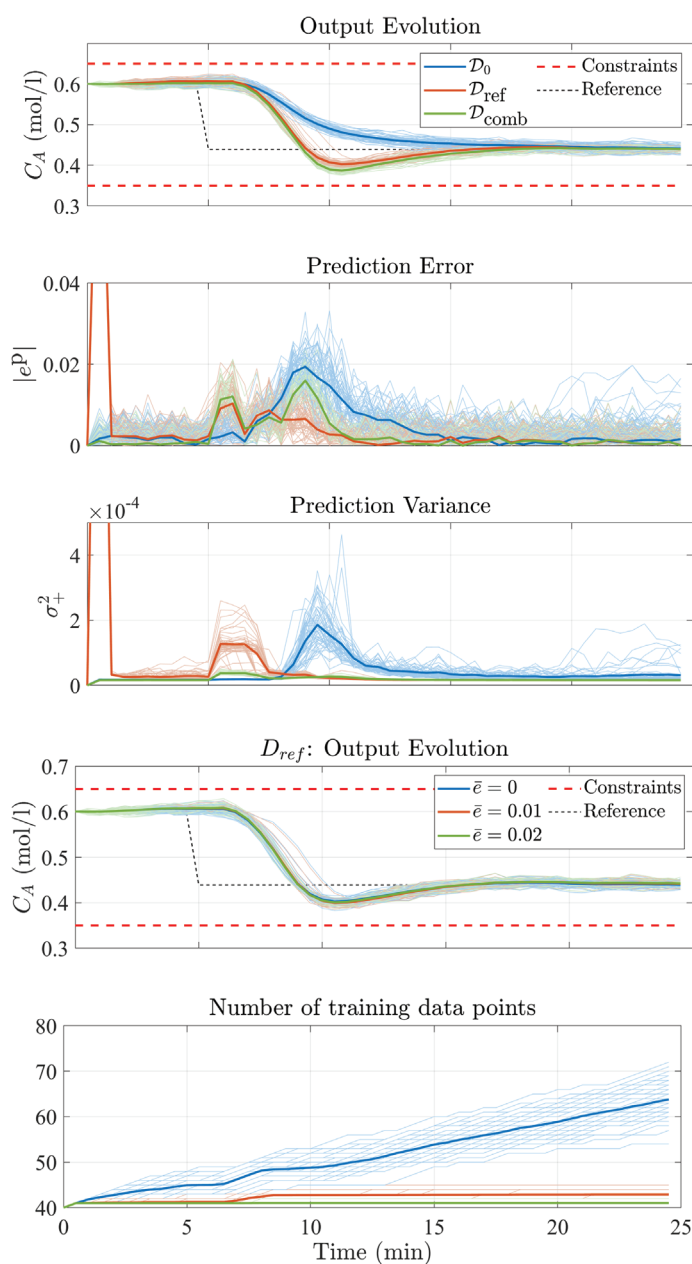


FIGURE 7 Simulation results with the rGP-MPC for the different training data cases together with the absolute value of prediction error $|e^p|$ and the prediction variance σ_+^2

FIGURE 8 Influence of $\bar{\epsilon}$ on the rGP-MPC with initial training data \mathcal{D}_{ref} . With $\bar{\epsilon} = 0$, every encountered data point is considered to be added to the training dataset. The variance threshold $\bar{\sigma}^2$ was set to a large value to not affect the result

FIGURE 9 Influence of $\bar{\sigma}^2$ on the rGP-MPC with initial training data D_{ref} . With $\bar{\sigma}^2 = 0$, every encountered data point is considered to be added to the training dataset. The prediction error threshold $\bar{\epsilon}$ was set to a large value to not affect the result

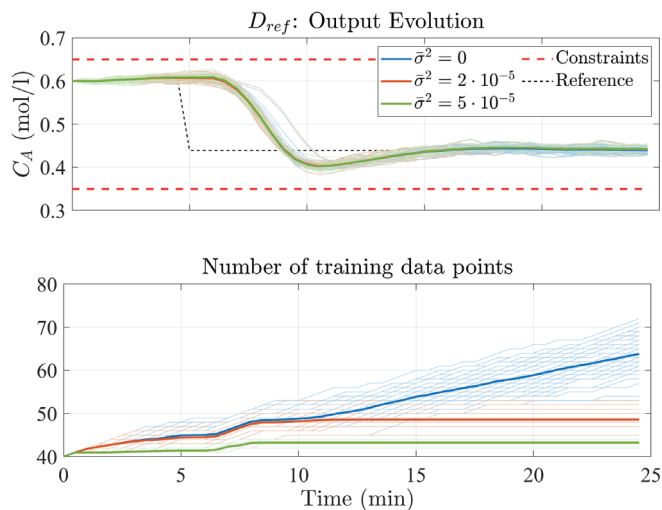
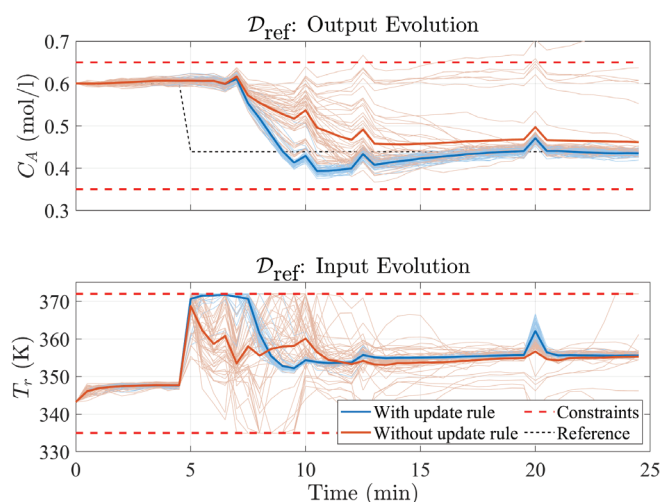


FIGURE 10 Influence of the update rule in Theorem 1, which permits inclusion of data point candidates only if they result in a decreasing value function. Outliers are generated at 7, 10, 12.5, and 20 min



be achieved by adding only a fraction of them. Hence, this shows not only that online learning can be achieved but also that it allows working with significantly smaller training data sets, which in turn result in lower computational costs.

After evaluating the influence of the parameters of Rule 1, we illustrate the influence of the update rule in Theorem 1, which guarantees a decrease of the value function. To this end, we continue with the D_{ref} case and additionally insert outliers into the output measurements in the course of the simulations. The effect of the update rule is shown in Figure 10. With it, the results are almost the same as before, except for the distortions due to the outliers, which however are compensated shortly after. All simulation outcomes are very similar in that case. Without the update rule, the resulting mean output sequence is different but not necessarily worse (smaller rise time, similar settling time, no overshoot) than the mean output sequence with the update rule. Some of the individual simulation outcomes perform even better, which is an indication that data points with valuable information are indeed discarded by the update rule as was also pointed out in Remark 7. On the other hand, the variability among the individual simulations is much larger. Several of the simulated output evolutions converge slower to the target and some do not converge at all until the end of the simulation. This is a direct result of the corresponding input sequences computed by the optimizer. In between 5 min and 11 min, the deviation of the mean input sequence from the optimal input sequence of the performance bound (oMPC, see Figures 4–6) is larger than in the case with the update rule. Furthermore, the individual input sequence outcomes vary considerably, even hitting the lower constraints. Due to the inclusion of every encountered data point candidate, the prediction model changes in some cases in an unfavorable way during the respective simulations, which leads to the depicted results. Note that qualitatively the same results (including not converging output sequences) are obtained, even without outliers. For instance, between the reference change at 5 min and the first outlier at 7 min, we observe that the input sequences already deviate considerably from the case with an active update rule, that is, the outlier is not the cause but usual noisy data

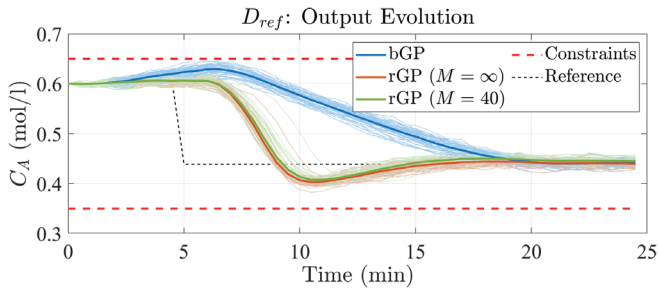


FIGURE 11 Influence of a limited number of training data points on the rGP-MPC with initial training data D_{ref}

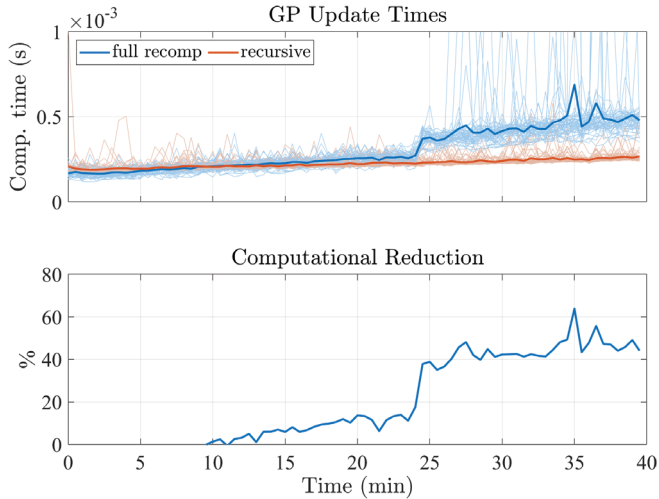


FIGURE 12 Comparison of computation times of the full recalculation of the Cholesky factor and the recursive update. The computational reduction that goes along with the recursive update increases with the amount of training data points

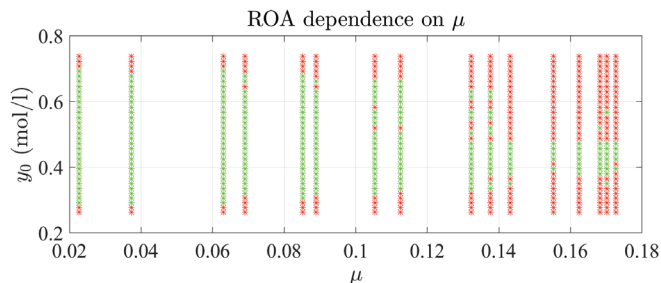
points. This illustrates the importance of the update rule in Theorem 1, not only for theoretical guarantees but also in terms of practical application.

Next, we consider the case that the number of training data points is limited by M . For the case of D_{ref} we set $M = 40$, which is the number of initially available training points, that is, the training data set cannot increase but old data points are exchanged with newer more informative ones. To this end, whenever a new point is added, the oldest data point is removed. In Figure 11 we compare the bGP (the initial training dataset is not updated at all), the rGP with $M = \infty$, $\bar{\epsilon} = \bar{\sigma}^2 = 0$ (every encountered data point is considered to be added), and the rGP with $M = 40$, $\bar{\epsilon} = 0.01$, $\bar{\sigma}^2 = 2 \cdot 10^{-5}$ (data points are only exchanged). The bGP result is the same as in Figure 5 and represents the worst case because the training data set is not updated at all. The $M = \infty$ case on the other hand represents the performance bound for this specific case because it includes the maximum of the incoming data points and does not remove any. As can be seen, the reaction of the limited case is a bit slower than the performance bound case but the resulting settling times are almost identical. Thus, with a training data set of only 40 points, where the points are exchanged during operation, almost the same performance can be achieved for the considered example as if every encountered point was included in the training dataset D .

Besides the computational cost reduction due to the possibility to work with smaller training data sets, we also illustrate the computational reduction due to the recursive update of the Cholesky factor. In Figure 12 we continue with the D_{ref} case, where we add every incoming point to the training data set and compare the computation times of the full and the recursive update of the Cholesky factor. The results show that the larger the training dataset becomes, the larger the absolute and relative computational reduction. At $t = 24$ min the full recomputation of the Cholesky factor increases significantly. Investigations point to the reason lying in the generation of the covariance matrix and the inner workings of Matlab's `chol` function to compute the Cholesky decomposition.

At last we present simulations of the robust feasible set Ω^0 , also denoted region of attraction (ROA), and how it changes for different maximum prediction errors μ . We continue with the D_{ref} case with $\bar{\epsilon} = \bar{\sigma}^2 = 0$ such that every data point is considered as a candidate for inclusion. Furthermore, M is set to a large value such that no points are removed from D_k . Each initial condition $\mathbf{x}_0 = [y_0 \ y_0 \ y_0]^T$ is simulated 30 times. Different μ values are obtained by varying the measurement noise from $\sigma_n^2 = 0.003^2$ to $\sigma_n^2 = 0.012^2$, where μ is then the largest error of all simulation runs and time steps. The result in Figure 13 yields a clear tendency. The larger μ , the smaller Ω^0 .

FIGURE 13 Change of the region of attraction Ω^0 for different μ . Red stars denote infeasible initial conditions, green stars feasible initial conditions. An initial condition is marked as infeasible if at least one simulation resulted in a constraint violation.



6 | CONCLUSION

In this work, we outlined the use of a GP-based nonlinear autoregressive model with exogenous input for prediction in an output feedback model predictive control scheme. The approach allows for online learning, by means of updating the training data set, to account for limited a priori process knowledge and the possibility for adaptation during operation. To this end, the concept of evolving GPs was adapted together with a recursive formulation to update the Cholesky decomposition to minimize computational cost. It was shown that the resulting model predictive control scheme is input-to-state stable with respect to the prediction error, despite the time-varying nature of the GP prediction model. Notably, the theoretic guarantees are not limited to Gaussian processes. They are rather valid for all online learning methods that satisfy the presented conditions.

The approach was verified in simulations, which have shown that it is in general possible to start with limited a priori process knowledge and refining the model during operation. One important finding is that it is particularly beneficial to start with a model that captures at least the behavior at the target reference, which is fortunately an intrinsic necessity for all MPC schemes that use a terminal region, cost, and controller to guarantee recursive feasibility and stability. In the case of fixed hyperparameters during online operation, a further consequence is that the hyperparameters should be optimized offline for a dataset that captures the target reference. Furthermore, the presented formulation yields good closed-loop performance with few training data points, thereby efficiently reducing the computational load. This presents itself as a possible option for very fast processes, where hyperparameter optimization is not an option but some kind of online learning is desirable. Additionally, due to the output feedback scheme, this approach can be employed for processes, whose state cannot be measured or is difficult to be estimated.

Future work aims at implementing the presented approach in laboratory experiments, together with a combination of a deterministic base model and the GP prediction model. From a theoretical point of view, time-varying reference tracking instead of set-point changes would be interesting to investigate. For instance, what conditions does the initial training data set has to satisfy to achieve acceptable tracking results and how to automatically compute safe thresholds for the data inclusion approach. Another interesting question to investigate is how the approach performs for time-varying processes. A hypothesis would be to combine the squared exponential covariance function with a non-stationary one to account for time variance in the process model.

ACKNOWLEDGEMENTS

The authors would like to thank the reviewers for helpful suggestions and discussions. Open access funding enabled and organized by Projekt DEAL.

ORCID

Michael Maiworm  <https://orcid.org/0000-0001-8129-4132>

Daniel Limon  <https://orcid.org/0000-0001-9334-7289>

REFERENCES

1. Rawlings JB, Mayne DQ. *Model Predictive Control: Theory and Design*. Madison: Nob Hill Publishing; 2009.
2. Mayne DQ. Model predictive control: Recent developments and future promise. *Automatica*. 2014;50(12):2967-2986.
3. Lucia S, Kögel M, Zometa P, Quevedo DE, Findeisen R. Predictive control, embedded cyberphysical systems and systems of systems – a perspective. *Annu Rev Control*. 2016;41:193-207.
4. Scokaert POM, Mayne DQ. Min-max feedback model predictive control for constrained linear systems. *IEEE Trans Autom Control*. 1998;43(8):1136-1142.
5. Mayne DQ, Raković SV, Findeisen R, Allgöwer F. Robust output feedback model predictive control of constrained linear systems. *Automatica*. 2006;42(7):1217-1222.
6. Lucia S, Finkler T, Engell S. Multi-stage nonlinear model predictive control applied to a semi-batch polymerization reactor under uncertainty. *J Process Control*. 2013;23(9):1306-1319.

7. Maiworm M, Bätthge T, Findeisen R. Scenario-based model predictive control: recursive feasibility and stability. Paper presented at: Proceedings of the 9th International Symposium on Advanced Control of Chemical Processes (ADCHEM). Whistler, Canada: IFAC; 2015:50-56.
8. Paulson JA, Streif S, Findeisen R, Braatz RD, Mesbah A. Fast stochastic model predictive control of end-to-end continuous pharmaceutical manufacturing. *Computer Aided Chemical Engineering*. Amsterdam, Niederlande: Elsevier; 2018:353-378.
9. Ljung L. *System Identification: Theory for the User*. 2nd ed. Upper Saddle River, NJ: Prentice Hall PTR; 1999.
10. Yang X, Maciejowski JM. Fault tolerant control using Gaussian processes and model predictive control. *Int J Appl Math Comput Sci*. 2015;25(1):133-148.
11. Ostafew CJ, Schoellig AP, Barfoot TD, Collier J. Learning-based nonlinear model predictive control to improve vision-based mobile robot path tracking. *J Field Robot*. 2016;33(1):133-152.
12. Hewing L, Liniger A, Zeilinger MN. Cautious NMPC with Gaussian process dynamics for autonomous miniature race cars. Paper presented at: Proceedings of the European Control Conference (ECC). Limassol, Cyprus; 2018:1341-1348.
13. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press; 2006.
14. Kocijan J. *Modelling and Control of Dynamic Systems using Gaussian Process Models*. New York, NY: Springer; 2016.
15. Kocijan J, Murray-Smith R, Rasmussen CE, Likar B. Predictive control with Gaussian process models. *The IEEE Region 8 EUROCON. Computer as a Tool*. Ljubljana, Slovenia: IEEE; 2003:352-356.
16. Klenske ED, Zeilinger MN, Schölkopf B, Hennig P. Gaussian process-based predictive control for periodic error correction. *IEEE Trans Control Syst Technol*. 2016;24(1):110-121.
17. Cao G, Lai EM-K, Alam F. Gaussian process model predictive control of an unmanned quadrotor. *J Intell Robot Syst*. 2017;88(1):147-162.
18. Maiworm M, Limón D, Manzano JM, Findeisen R. Stability of Gaussian process learning based output feedback model predictive control. Paper presented at: Proceedings of the Conference on Nonlinear Model Predictive Control (NMPC), Madison, Wis: IFAC; 2018:551-557.
19. Steinwart I, Christmann A. *Support Vector Machines*. Berlin, Germany: Springer Science & Business Media; 2008.
20. Snelson E, Ghahramani Z. Sparse Gaussian processes using pseudo-inputs. *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press; 2006:1257-1264.
21. Lazáro-Gredilla M, Quiñero-Candela J, Rasmussen CE, Figueiras-Vidal AR. Sparse spectrum Gaussian process regression. *J Mach Learn Res*. 2010;11:1865-1881.
22. Ostafew CJ, Schoellig AP, Barfoot TD. Learning-based nonlinear model predictive control to improve vision-based mobile robot path-tracking in challenging outdoor environments. Paper presented at: Proceedings of the International Conference on Robotics and Automation (ICRA). Hong Kong, China; 2014:4029-4036; IEEE.
23. Hewing L, Zeilinger MN. Cautious model predictive control using Gaussian process regression; 2017. arXiv preprint arXiv:1705.10702.
24. Ortmann L, Shi D, Dassau E, Doyle FJ, Leonhardt S, Misgeld BJE. Gaussian process-based model predictive control of blood glucose for patients with type 1 diabetes mellitus. Paper presented at: Proceedings of the Asian Control Conference (ASCC). Gold Coast, Australia; 2017:1092-1097; IEEE.
25. Murray-Smith R, Sbarbaro D, Rasmussen CE, Girard A. Adaptive, cautious, predictive control with Gaussian process priors. *IFAC Proc Vol*. 2003;36(16):1155-1160.
26. Kocijan J, Murray-Smith R. Nonlinear predictive control with a Gaussian process model. *Lect Notes Comput Sci*. 2005;3355:185-200.
27. Ažman K, Kocijan J. Non-linear model predictive control for models with local information and uncertainties. *Trans Inst Meas Control*. 2008;30(5):371-396.
28. Berkenkamp F, Moriconi R, Schoellig AP, Krause A. Safe learning of regions of attraction for uncertain, nonlinear systems with Gaussian processes. Paper presented at: Proceedings of the Conference on Decision and Control (CDC). Las Vegas, NV; 2016:4661-4666; IEEE.
29. Vinogradska J, Bischoff B, Nguyen-Tuong D, Romer A, Schmidt H, Peters J. Stability of controllers for Gaussian process forward models. Paper presented at: Proceedings of the International Conference on Machine Learning, PMLR. New York, NY; 2016:545-554.
30. Akametalu AK, Kaynama S, Fisac JF, Zeilinger MN, Gillula JH, Tomlin CJ. Reachability-based safe learning with Gaussian processes. Paper presented at: Proceedings of the Conference on Decision and Control (CDC). Los Angeles, CA; 2014:1424-1431; IEEE.
31. Koller T, Berkenkamp F, Turchetta M, Krause A. Learning-based model predictive control for safe exploration and reinforcement learning; 2018. arXiv preprint arXiv:1803.08287.
32. Fisac JF, Akametalu AK, Zeilinger MN, Kaynama S, Gillula J, Tomlin CJ. A general safety framework for learning-based control in uncertain robotic systems. *IEEE Trans Autom Control*. 2018;64(7):2737-2752.
33. Wabersich KP, Zeilinger MN. Scalable synthesis of safety certificates from data with application to learning-based control. Paper presented at: Proceedings of the European Control Conference (ECC). Limassol, Cyprus; 2018:1691-1697; IEEE.
34. Aswani A, Gonzalez H, Sastry SS, Tomlin C. Provably safe and robust learning-based model predictive control. *Automatica*. 2013;49(5):1216-1226.
35. Bethge J, Morabito B, Matschek J, Findeisen R. Multi-mode learning supported model predictive control with guarantees. Paper presented at: Proceedings of the Conference on Nonlinear Model Predictive Control (NMPC), Madison, Wis: IFAC; 2018:616-621.
36. Soloperto R, Müller MA, Trimpe S, Allgöwer F. Learning-based robust model predictive control with state-dependent uncertainty. Paper presented at: Proceedings of the Conference on Nonlinear Model Predictive Control (NMPC), Madison, Wis: IFAC; 2018:538-543.
37. Bastani O. Safe planning via model predictive shielding; 2019. arXiv preprint arXiv:1905.10691.
38. Limón D, Alamo T, Raimondo DM, et al. Input-to-state stability: a unifying framework for robust model predictive control. *Nonlinear Model Predictive Control*. New York, NY: Springer; 2009:1-26.
39. Petelin D, Kocijan J. Control system with evolving Gaussian process models. Paper presented at: Proceedings of the Workshop on Evolving and Adaptive Intelligent Systems (EAIS). Paris, France; 2011:178-184.

40. Levin AU, Narendra KS. *Identification of Nonlinear Dynamical Systems Using Neural Networks*. Amsterdam, Niederlande: Elsevier; 1997:129-160.
41. McKinnon CD, Schoellig AP. Learning multimodal models for robot dynamics online with a mixture of Gaussian process experts. Paper presented at: Proceedings of the International Conference on Robotics and Automation (ICRA). Marina Bay Sands, Singapore; 2017:322-328.
42. Berkenkamp F, Turchetta M, Schoellig A, Krause A. Safe model-based reinforcement learning with stability guarantees. Paper presented at: Proceedings of the International Conference on Neural Information Processing Systems, Curran Associates Inc. Long Beach, CA; 2017:908-918.
43. Snelson E, Ghahramani Z, Rasmussen CE. Warped Gaussian processes. *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press; 2004:337-344.
44. Maiworm M, Wagner C, Temirov R, Tautz FS, Findeisen R. Two-degree-of-freedom control combining machine learning and extremum seeking for fast scanning quantum dot microscopy. Paper presented at: Proceedings of the American Control Conference (ACC). Milwaukee, Wis; 2018:4360-4366; IEEE.
45. Gregorčič G, Lightbody G. Gaussian process approach for modelling of nonlinear systems. *Eng Appl Artif Intell*. 2009;22(4-5):522-533.
46. Roberts S, Osborne M, Ebden M, Reece S, Gibson N, Aigrain S. Gaussian processes for time-series modelling. *Philos Trans Royal Soc Lond A Math Phys Eng Sci*. 2013;371(1984):20110550.
47. Williams Christopher K.I., Rasmussen Carl Edward. Gaussian processes for regression. In: *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press; 1996:514-520.
48. Kocijan J, Girard A, Banko B, Murray-Smith R. Dynamic systems identification with Gaussian processes. *Math Comput Model Dyn Syst*. 2005;11(4):411-424.
49. Ackermann ER, De Villiers JP, Cilliers PJ. Nonlinear dynamic systems modeling using Gaussian processes: predicting ionospheric total electron content over South Africa. *J Geophys Res Space Physics*. 2011;116(A10).
50. Smola AJ, Bartlett PL. Sparse greedy Gaussian process regression. *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press; 2001:619-625.
51. Seeger M, Williams C, Lawrence N. Fast forward selection to speed up sparse Gaussian process regression. *Artif Intell Stat*. 2003;9. <https://infoscience.epfl.ch/record/161318>.
52. Carron A, Todescato M, Carli R, Schenato L, Pillonetto G. Machine learning meets Kalman filtering. Paper presented at: Proceedings of the Conference on Decision and Control (CDC). Las Vegas, NV; 2016:4594-4599; IEEE.
53. Neal RM. Monte Carlo implementation of Gaussian process models for bayesian regression and classification; 1997. arXiv preprint physics/9701026.
54. Osborne MA. Bayesian Gaussian processes for sequential prediction, optimisation and quadrature [Ph. D thesis]. Oxford, UK: Oxford University; 2010.
55. Huber MF. Recursive Gaussian process: On-line regression and learning. *Pattern Recogn Lett*. 2014;45:85-91.
56. Van Vaerenbergh S, Lázaro-Gredilla M, Santamaria I. Kernel recursive least-squares tracker for time-varying regression. *IEEE Trans Neural Netw Learn Syst*. 2012;23(8):1313-1326.
57. Pérez-Cruz F, Van Vaerenbergh S, Murillo-Fuentes JJ, Lázaro-Gredilla M, Santamaria I. Gaussian processes for nonlinear signal processing: an overview of recent advances. *IEEE Signal Process Mag*. 2013;30(4):40-50.
58. Mayne DQ, Rawlings JB, Rao CV, Scokaert Pierre OM. Constrained model predictive control: stability and optimality. *Automatica*. 2000;36(6):789-814.
59. Limón D, Alamo T, Salas F, Camacho EF. On the stability of constrained MPC without terminal constraint. *IEEE Trans Automat Control*. 2006;51(5):832-836.
60. Raković SV, Levine WS. *Handbook of Model Predictive Control*. New York, NY: Springer; 2019.
61. Abrahamsen P. *A Review of Gaussian Random Fields and Correlation Functions*. Oslo, Norway: Norsk Regnesentral/Norwegian Computing Center Oslo; 1997.
62. Park J, Sandberg IW. Universal approximation using radial-basis-function networks. *Neural Comput*. 1991;3(2):246-257.
63. De Nicolao G, Pillonetto G. A new kernel-based approach for system identification. Paper presented at: Proceedings of the American Control Conference (ACC). Seattle, WA; 2008:4510-4516.
64. Pillonetto G, Chiuso A. Gaussian processes for Wiener-Hammerstein system identification. *IFAC Proc Vol*. 2009;42(10):838-843.
65. Engel Y. Algorithms and Representations for Reinforcement Learning [PhD thesis]. Jerusalem, Israel: Hebrew University of Jerusalem; 2005.
66. Srinivas N, Krause A, Kakade SM, Seeger MW. Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Trans Inf Theory*. 2012;58(5):3250-3265.
67. Seborg DE, Edgar TF, Mellichamp DA. *Process Dynamics and Control*. New York, NY: Wiley; 1989.

How to cite this article: Maiworm M, Limon D, Findeisen R. Online learning-based model predictive control with Gaussian process models and stability guarantees. *Int J Robust Nonlinear Control*. 2021;31:8785-8812. <https://doi.org/10.1002/rnc.5361>

APPENDIX A. RECURSIVE CHOLESKY FACTOR UPDATE

According to Osborne,⁵⁴ the Cholesky factor can be updated recursively as presented in the following. Regarding the case of including a new data point, consider the covariance matrix \mathbf{K} , represented in block form as

$$\begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{13} \\ \mathbf{K}_{13}^T & \mathbf{K}_{33} \end{bmatrix},$$

and its Cholesky factor

$$\begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{13} \\ 0 & \mathbf{R}_{33} \end{bmatrix}.$$

Now, given an updated covariance matrix

$$\begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \mathbf{K}_{13} \\ \mathbf{K}_{12}^T & \mathbf{K}_{22} & \mathbf{K}_{23} \\ \mathbf{K}_{13}^T & \mathbf{K}_{23}^T & \mathbf{K}_{33} \end{bmatrix},$$

that differs from the previous by insertion of a new row and column, the updated Cholesky factor

$$\begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \mathbf{S}_{13} \\ 0 & \mathbf{S}_{22} & \mathbf{S}_{23} \\ 0 & 0 & \mathbf{S}_{33} \end{bmatrix}.$$

can be computed via

$$\begin{aligned} \mathbf{S}_{11} &= \mathbf{R}_{11} & \mathbf{S}_{22} &= \text{chol}(\mathbf{K}_{22} - \mathbf{S}_{12}^T \mathbf{S}_{12}) \\ \mathbf{S}_{12} &= \mathbf{R}_{11}^T \setminus \mathbf{K}_{12} & \mathbf{S}_{23} &= \mathbf{S}_{22}^T \setminus (\mathbf{K}_{23} - \mathbf{S}_{12}^T \mathbf{S}_{13}) \\ \mathbf{S}_{13} &= \mathbf{R}_{13} & \mathbf{S}_{33} &= \text{chol}(\mathbf{R}_{33}^T \mathbf{R}_{33} - \mathbf{S}_{23}^T \mathbf{S}_{23}). \end{aligned} \quad (\text{A1})$$

On the other hand, if the current covariance matrix in block form

$$\begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \mathbf{K}_{13} \\ \mathbf{K}_{12}^T & \mathbf{K}_{22} & \mathbf{K}_{23} \\ \mathbf{K}_{13}^T & \mathbf{K}_{23}^T & \mathbf{K}_{33} \end{bmatrix},$$

with Cholesky factor

$$\begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{R}_{13} \\ 0 & \mathbf{R}_{22} & \mathbf{R}_{23} \\ 0 & 0 & \mathbf{R}_{33} \end{bmatrix},$$

is reduced by one row and column, such that we obtain

$$\begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{13} \\ \mathbf{K}_{13}^T & \mathbf{K}_{33} \end{bmatrix},$$

the downdated Cholesky factor

$$\begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{13} \\ 0 & \mathbf{S}_{33} \end{bmatrix}$$

can be computed via

$$\begin{aligned} \mathbf{S}_{11} &= \mathbf{R}_{11} \\ \mathbf{S}_{13} &= \mathbf{R}_{13} \\ \mathbf{S}_{33} &= \text{chol}(\mathbf{R}_{23}^T \mathbf{R}_{23} + \mathbf{R}_{33}^T \mathbf{R}_{33}). \end{aligned} \quad (\text{A2})$$

APPENDIX B. POSTERIOR MEAN GRADIENT

The optimal control problem (12) requires a terminal cost function, which can be based on a linearized version of the prediction model in Section 5. To this end we require the gradient of the GP posterior mean function

$$\nabla m_+(\mathbf{w}) = \frac{\partial m_+(\mathbf{w})}{\partial \mathbf{w}} = \left[\frac{\partial m_+(\mathbf{w})}{\partial w_1} \quad \dots \quad \frac{\partial m_+(\mathbf{w})}{\partial w_{n_w}} \right]^T$$

w.r.t. to its regressor $\mathbf{w} = [w_1, \dots, w_{n_w}]$, where we omit the dependence on the training data \mathcal{D} for the sake of brevity.

Assuming a constant prior mean in (4a) we obtain

$$\nabla m_+(\mathbf{w}) = \frac{\partial k(\mathbf{w}, \mathbf{w})}{\partial \mathbf{w}} \mathbf{K}^{-1}(\mathbf{z} - m(\mathbf{w})),$$

with

$$\frac{\partial k(\mathbf{w}, \mathbf{w})}{\partial \mathbf{w}} = \left[\frac{\partial k(\mathbf{w}, \mathbf{w}_1)}{\partial \mathbf{w}} \quad \dots \quad \frac{\partial k(\mathbf{w}, \mathbf{w}_n)}{\partial \mathbf{w}} \right],$$

TABLE B1 CSTR parameters

Parameters	Explanation	Value
q_0	Reactive input flow	10 l/min
V	Liquid volume in the tank	1501
k_0	Frequency constant	$6 \cdot 10^{10}$ 1/min
E/R	Arrhenius constant	9750 K
ΔH_r	Reaction enthalpy	10,000 J/mol
UA	Heat transfer coefficient	70,000 J/(min K)
ρ	Density	1100 g/l
C_p	Specific heat	0.3 J/(g K)
τ	Time constant	1.5 min
C_{Af}	C_A in the input flow	1 mol/l
T_f	Input flow temperature	370 K

TABLE B2 Hyperparameters

	c	l_1	l_2	l_3	l_4	σ_f^2
D_0	0.64	0.07	0.29	0.14	9.93	0.06
D_{ref}	0.36	0.20	11.7	0.64	5.07	0.13
D_{comb}	0.43	0.42	2.09	1.01	2.83	0.26

	D_0	D_{ref}	D_{comb}
oMPC	59.5	59.5	59.5
bGP-MPC	71.3	95.3	66.2
rGP-MPC	64.5	63.6	66.7

TABLE B3 Model predictive control performance computed by (18)

where $\mathbf{w}_1, \dots, \mathbf{w}_n$ is the corresponding regressor of each of the n measured training data points in D .

For the covariance function (6), we obtain

$$\frac{\partial k(\mathbf{w}, \mathbf{w}')}{\partial \mathbf{w}} = k^*(\mathbf{w}, \mathbf{w}')\Lambda(\mathbf{w}' - \mathbf{w}),$$

where $k^*(\mathbf{w}, \mathbf{w}')$ is (6) without the noise term, i.e., $\sigma_n^2 = 0$.