Article

Model Predictive Supervisory Control for Integrated Emission Management of Diesel Engines

Johannes Ritzmann [1,*], Christian Peterhans [2], Oscar Chinellato [2], Manuel Gehlen [2] and Christopher Onder [1]

1   Department of Mechanical Engineering and Process Control, ETH Zürich, 8092 Zürich, Switzerland; onder@idsc.mavt.ethz.ch
2   FPT Motorenforschung AG, 9320 Arbon, Switzerland; christian.peterhans@ivecogroup.com (C.P.); oscar.chinellato@gmail.com (O.C.); manuel.gehlen@ivecogroup.com (M.G.)
*   Correspondence: jritzman@idsc.mavt.ethz.ch

Abstract: In this work, a predictive supervisory controller is presented that optimizes the interaction between a diesel engine and its aftertreatment system (ATS). The fuel consumption is minimized while respecting an upper bound on the emitted tailpipe NO_x mass. This is achieved by optimally balancing the fuel consumption, the engine-out NO_x emissions, and the ATS heating. The proposed predictive supervisory controller employs a two-layer model predictive control structure and solves the optimal control problem using a direct method. Through experimental validation, the resulting controller was shown to reduce the fuel consumption by 1.1% at equivalent tailpipe NO_x emissions for the nonroad transient cycle when compared to the operation with a fixed engine calibration. Further, the controller's robustness to different missions, initial ATS temperatures, NO_x limits, and mispredictions was demonstrated.

Keywords: integrated emission management; variable engine calibration; pollutant emissions; aftertreatment system; supervisory control; model predictive control

1. Introduction

Today, diesel combustion engines are used in a variety of applications. As a byproduct of combustion, they emit NO_x, which has adverse effects on human health and on the environment. Therefore, stringent pollutant legislation has been introduced in recent decades to limit the emission of NO_x [1]. At the same time, the fuel consumption must be minimized, in order to limit the operational cost as well as the CO_2 emissions. Unfortunately, these two goals are conflicting [2] and necessitate the careful control of the engine operation.

To extend this inherent trade-off, modern engine systems are equipped with an aftertreatment system (ATS) to combat NO_x and other pollutant emissions. For diesel engines, selective catalytic reduction (SCR) systems are typically used, whereby NO_x is reduced using ammonia [3]. For these systems to operate effectively, their temperature should surpass a lower limit of approximately 200 °C.

As the engine operation influences the operation of the ATS, a joint optimization of the engine and the ATS is required. Such an approach is known in the literature as integrated emission management (IEM) [4], whereby a supervisory controller is used to coordinate the operation of the engine and the ATS during a mission.

In a previous publication by the authors, offline optimization was used to investigate this problem [5]. It was shown that, in order to achieve optimal performance, the fuel consumption, the engine-out NO_x emissions, and the enthalpy provided to the ATS must be carefully managed during a mission.

In this paper, the previously obtained optimal input trajectories are applied at the testbench in a feedforward fashion. An error in the emitted tailpipe NO_x mass is observed and attributed to model mismatch. A predictive supervisory controller is therefore required

that optimizes the engine operation online and compensates for model mismatch and mispredictions using feedback control.

1.1. Literature Review

Early works neglected the ATS entirely [6,7]. As a result, good performance was achieved using the equivalent emission minimization strategy, whereby a feedback controller sets the equivalence factor used to determine the optimal trade-off between fuel consumption and NO_x emissions. The method is closely related to the equivalent consumption minimization strategy (ECMS) applied to the energy management problem of hybrid electric vehicles [8].

In later works [9,10], the NO_x reduction by the ATS was considered, but the effect of the engine operation on the ATS temperature was neglected. In [9], an ECMS-like causal controller is presented that adapts an equivalence factor for the tailpipe NO_x emissions. Subsequently, the equivalence factor for the engine-out NO_x emissions was determined, based on the measured NO_x-reduction efficiency of the ATS. In [10], a model predictive controller is used to set the operation of the engine and the ATS, whereby the effect of the control action on the ATS temperature is neglected. While these methods are straightforward to implement, the fact that they do not optimize the ATS temperature dynamics could result in a suboptimal operation, especially for missions with a low ATS temperature.

In other works [11–14], the benefit of heating the ATS is approximated offline. This approximation is then used by the supervisory controller when determining the optimal engine operation online. In [11], a feedforward controller based on Pontryagin's minimum principle was implemented, whereby a heuristic method was used to determine the equivalence factor for the ATS heating. In [12], this controller was extended to include feedback control based on the measured tailpipe NO_x emissions. In [13], dynamic programming was used to determine the optimal operation for several missions offline. Based on the results, a mapping was fitted that relates the equivalence factor for the ATS heating to the current value of the ATS temperature. This mapping was then used in an ECMS-like online controller in [14].

To the best of the authors' knowledge, no publication exists in which an online supervisory controller is developed that uses predictive information about the upcoming mission to minimize the fuel consumed during the mission, while respecting legislative limits on the tailpipe NO_x emissions and explicitly considering the effect of the engine operation on the ATS temperature.

1.2. Contribution

The contribution of this work is twofold. First, a real-time-capable optimization method to solve the IEM problem by solving a nonlinear program (NLP) is presented. Thereby, continuously differentiable look-up elements are used to describe a preoptimized engine operation, i.e., a Pareto front with respect to fuel consumption, NO_x emissions, and enthalpy provided to the ATS. Due to its computational efficiency, the resulting optimization can be performed in real time.

Second, a predictive supervisory controller is designed that employs this optimization method in combination with predictive information about the upcoming mission in a model predictive control framework. Thereby, the vehicle's fuel consumption is minimized, while a limit on the emitted tailpipe NO_x mass is respected. Based on the predictive information, the controller adapts the engine operation to the mission and optimally sets the trade-off between the fuel consumption, the NO_x emissions, and the ATS heating. The proposed controller was experimentally validated at an engine testbench.

2. System Description

A schematic of the engine system considered in this work is shown in Figure 1. It consists of a diesel engine equipped with an ATS. The main characteristics of the engine system are listed in Table 1.
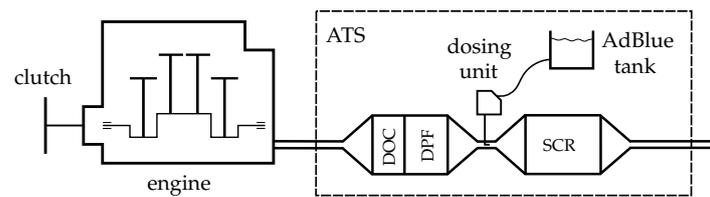
Figure 1. Schematic of engine system layout. The aftertreatment system (ATS) features a diesel oxidation catalyst (DOC), a diesel particulate filter (DPF), and a selective catalytic reduction (SCR) system.

Table 1. Main characteristics of the engine system.

Engine type	inline 4-cylinder diesel
Displacement	3 L
Max. power	130 kW
Homologation	EURO 6 step D

For the experiments, the engine was fitted to a dynamic testbench. An AVL KMA 4000 fuel flow sensor was used to measure the fuel consumption. The stock NO_x sensors were used to measure the engine-out NO_x concentration and the tailpipe NO_x concentration. The stock air mass flow sensor was used to measure the air mass flow. The temperature measured downstream of the ATS using the stock sensor was used to characterize the ATS temperature. The NO_x mass flow was calculated based the NO_x concentration and the exhaust mass flow, which was obtained as the sum of the air mass flow and the fuel mass flow. The NO_x-reduction efficiency of the ATS was calculated based the measured NO_x concentrations.

The engine considered in this work is overactuated. Consequently, some engine control inputs can be adapted online in order to achieve a certain objective. In this work, the five engine control inputs start of injection φ_{soi}, fuel rail pressure p_{rail}, variable geometry turbine actuator position u_{vgt}, exhaust gas recirculation valve position u_{egr}, and exhaust flap position u_{flap}, are considered as degrees of freedom that can be optimized while providing the requested torque T_e at the current speed ω_e. The resulting vector of engine control inputs is

$$\boldsymbol{u} = \begin{bmatrix} \varphi_{soi} \\ p_{rail} \\ u_{vgt} \\ u_{egr} \\ u_{flap} \end{bmatrix}. \tag{1}$$

As a simplification, only NO_x emissions were considered in this work. The engine operation was chosen such that the emitted carbon monoxide, unburned hydrocarbons, and particulate matter can be oxidized in the diesel oxidation catalyst (DOC) or trapped in the diesel particulate filter (DPF).

2.1. Engine Model

In this work, a reduced-order engine model resulting from a preoptimization of the engine operation was adopted from [5]. The benefit of the preoptimization is that the computational complexity of the subsequent online optimization is reduced. This is achieved by reducing the number of inputs from the five engine control inputs of (1) to two engine strategy inputs J_{Hd}^{rot} and $J_{NO_x}^{rot}$, by limiting the engine operation to the Pareto front defined by a multi-objective optimization. Furthermore, the preoptimization guarantees the feasibility of the engine operation according to a set of limits, which no longer need to be checked during the subsequent online optimization.

The objectives of the preoptimization are minimizing the fuel consumption, minimizing the engine-out NO_x emissions, and maximizing the enthalpy provided to the ATS. The latter is defined as

$$\overset{*}{H}_{\text{ATS}} = \overset{*}{m}_{\text{exh}} \, c_{\text{p}} \, (\vartheta_{\text{exh}} - \vartheta) \,, \tag{2}$$

where $\overset{*}{m}_{\text{exh}}$ is the exhaust mass flow, c_{p} is the specific heat capacity of the exhaust gas, ϑ_{exh} is the exhaust gas temperature, and ϑ is the ATS temperature.

The preoptimized engine model uses look-up elements to characterize the Pareto front. To achieve a numerically robust interpolation, a scaled and rotated version of the Pareto front is stored using the variables J_X^{rot}, where $X \in \{\text{Hd}, \text{NO}_x, \text{fuel}\}$. An example of such a Pareto front for a given speed, torque, and ATS temperature is shown on the left of Figure 2. The Pareto front is described by a five-dimensional look-up element for $J_{\text{fuel}}^{\text{rot}}$ with inputs $(\omega_{\text{e}}, T_{\text{e}}, \vartheta, J_{\text{Hd}}^{\text{rot}}, J_{\text{NO}_x}^{\text{rot}})$, while the bounds on $J_{\text{NO}_x}^{\text{rot}}$ are described by two four-dimensional look-up elements for $J_{\text{NO}_x,\text{min}}^{\text{rot}}$ and $J_{\text{NO}_x,\text{max}}^{\text{rot}}$ with inputs $(\omega_{\text{e}}, T_{\text{e}}, \vartheta, J_{\text{Hd}}^{\text{rot}})$.
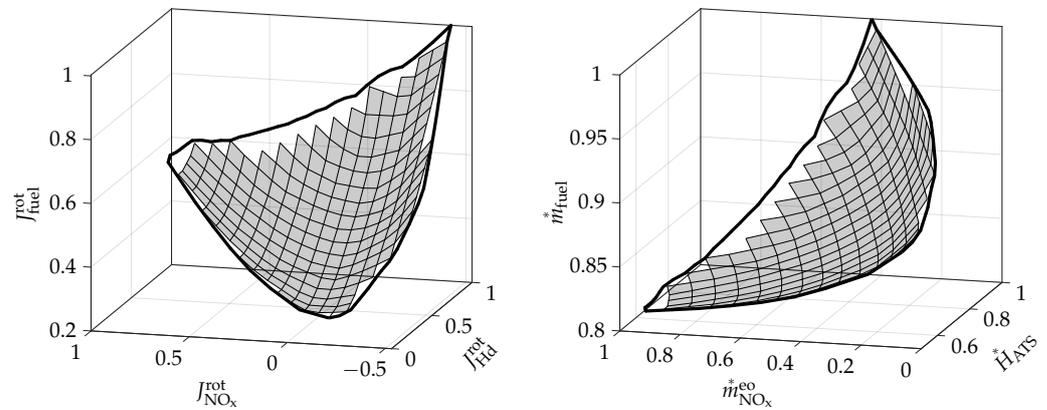
Figure 2. Exemplary Pareto front at an engine speed of 2000 rpm, an engine torque of 200 Nm, and an ATS temperature of 67 °C. The gray surface shows the identified Pareto front, while the solid black lines show the bounds on $J_{\text{NO}_x}^{\text{rot}}$. On the left, the Pareto front described by scaled and rotated values is shown. On the right, the Pareto front transferred back to physically meaningful values is shown. The values in the plot on the right were normalized.

To retrieve the desired engine outputs, the rotation is reversed to obtain the variables J_X and the scaling is reversed using

$$\overset{*}{H}_{\text{ATS}} = (1 - J_{\text{Hd}}) \, \overset{*}{H}_{\text{ATS,min}} + J_{\text{Hd}} \, \overset{*}{H}_{\text{ATS,max}} \tag{3}$$

$$\overset{*}{m}_{\text{NO}_x}^{\text{eo}} = (1 - J_{\text{NO}_x}) \, \overset{*}{m}_{\text{NO}_x,\text{min}}^{\text{eo}} + J_{\text{NO}_x} \, \overset{*}{m}_{\text{NO}_x,\text{max}}^{\text{eo}} \tag{4}$$

$$\overset{*}{m}_{\text{fuel}} = (1 - J_{\text{fuel}}) \, \overset{*}{m}_{\text{fuel,min}} + J_{\text{fuel}} \, \overset{*}{m}_{\text{fuel,max}} \,, \tag{5}$$

where the fuel mass flow is denoted $\overset{*}{m}_{\text{fuel}}$ and the engine-out NO_x mass flow is denoted $\overset{*}{m}_{\text{NO}_x}^{\text{eo}}$. The scaling values $\overset{*}{H}_{\text{ATS,min}}$, $\overset{*}{H}_{\text{ATS,max}}$, $\overset{*}{m}_{\text{NO}_x,\text{min}}^{\text{eo}}$, $\overset{*}{m}_{\text{NO}_x,\text{max}}^{\text{eo}}$, $\overset{*}{m}_{\text{fuel,min}}$, and $\overset{*}{m}_{\text{fuel,max}}$ are characterized by three-dimensional look-up elements with inputs $(\omega_{\text{e}}, T_{\text{e}}, \vartheta)$. In Figure 2, on the right, the Pareto front transferred back to physically meaningful values is shown.

To operate the engine, the engine control inputs must be determined. For this, a smoothed inverse mapping was developed that returns the engine control inputs as a function of the selected engine strategy inputs. An equivalent mapping is also developed for the exhaust mass flow, which is required to evaluate the ATS model. The mappings are stored in the form of look-up elements with inputs $(\omega_{\text{e}}, T_{\text{e}}, \vartheta, J_{\text{Hd}}^{\text{rot}}, J_{\text{NO}_x}^{\text{rot}})$.

2.2. Aftertreatment System Model

The ATS model used in this work was adopted from [5]. The thermal model consists of a single lumped brick temperature ϑ, whose dynamics are

$$\frac{\mathrm{d}}{\mathrm{d}t}\vartheta = \frac{\overset{*}{H}_{\mathrm{ATS}} - \overset{*}{Q}_{\mathrm{hl}}}{C_{\mathrm{ATS}}} \tag{6}$$

$$\overset{*}{Q}_{\mathrm{hl}} = \alpha\, S\, (\vartheta - \vartheta_{\mathrm{amb}})\,, \tag{7}$$

where C_{ATS} is the heat capacity of the ATS. The conductive heat loss to the environment is denoted $\overset{*}{Q}_{\mathrm{hl}}$, where α is the heat transfer coefficient, S is the outer surface area of the ATS, and ϑ_{amb} is the ambient temperature.

The chemical model consists of a static look-up element that describes the NO_x-reduction efficiency $\eta_{\mathrm{DeNO_x}}$ as a function of the ATS temperature and the exhaust mass flow. The resulting tailpipe NO_x mass flow is

$$\overset{*}{m}{}_{\mathrm{NO_x}}^{\mathrm{tp}} = \left(1 - \eta_{\mathrm{DeNO_x}}(\vartheta, \overset{*}{m}_{\mathrm{exh}})\right)\overset{*}{m}{}_{\mathrm{NO_x}}^{\mathrm{eo}}\,. \tag{8}$$

The dynamics of the emitted tailpipe NO_x mass are

$$\frac{\mathrm{d}}{\mathrm{d}t}m_{\mathrm{NO_x}}^{\mathrm{tp}} = \overset{*}{m}{}_{\mathrm{NO_x}}^{\mathrm{tp}}\,. \tag{9}$$

3. Real-Time-Capable Optimization Method

The objective of the control task is to minimize the fuel consumed during a given mission, while respecting an upper bound on the emitted tailpipe NO_x mass at the end of the mission. The resulting optimal control problem (OCP) is

$$\underset{J_{\mathrm{Hd}}^{\mathrm{rot}},\, J_{\mathrm{NO_x}}^{\mathrm{rot}}}{\text{minimize}} \quad \int_0^{t_{\mathrm{f}}} \overset{*}{m}_{\mathrm{fuel}}\left(\omega_{\mathrm{e}}, T_{\mathrm{e}}, \vartheta, J_{\mathrm{Hd}}^{\mathrm{rot}}, J_{\mathrm{NO_x}}^{\mathrm{rot}}\right)\, \mathrm{d}t \tag{10a}$$

$$\text{subject to} \quad \frac{\mathrm{d}}{\mathrm{d}t}\vartheta = f_{\vartheta}\left(\omega_{\mathrm{e}}, T_{\mathrm{e}}, \vartheta, J_{\mathrm{Hd}}^{\mathrm{rot}}, J_{\mathrm{NO_x}}^{\mathrm{rot}}\right) \tag{10b}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}m_{\mathrm{NO_x}}^{\mathrm{tp}} = \overset{*}{m}{}_{\mathrm{NO_x}}^{\mathrm{tp}}\left(\omega_{\mathrm{e}}, T_{\mathrm{e}}, \vartheta, J_{\mathrm{Hd}}^{\mathrm{rot}}, J_{\mathrm{NO_x}}^{\mathrm{rot}}\right) \tag{10c}$$

$$\vartheta(0) = \vartheta_0 \tag{10d}$$

$$m_{\mathrm{NO_x}}^{\mathrm{tp}}(0) = 0 \tag{10e}$$

$$m_{\mathrm{NO_x}}^{\mathrm{tp}}(t_{\mathrm{f}}) \leq \overline{m}_{\mathrm{NO_x}}^{\mathrm{tp}} \tag{10f}$$

$$J_{\mathrm{Hd}}^{\mathrm{rot}} \in [0, 1] \tag{10g}$$

$$J_{\mathrm{NO_x}}^{\mathrm{rot}} \geq J_{\mathrm{NO_x,min}}^{\mathrm{rot}}\left(\omega_{\mathrm{e}}, T_{\mathrm{e}}, \vartheta, J_{\mathrm{Hd}}^{\mathrm{rot}}\right) \tag{10h}$$

$$J_{\mathrm{NO_x}}^{\mathrm{rot}} \leq J_{\mathrm{NO_x,max}}^{\mathrm{rot}}\left(\omega_{\mathrm{e}}, T_{\mathrm{e}}, \vartheta, J_{\mathrm{Hd}}^{\mathrm{rot}}\right)\,, \tag{10i}$$

where t_{f} is the final time, the function f_{ϑ} represents (6), and $\overline{m}_{\mathrm{NO_x}}^{\mathrm{tp}}$ is the limit on the emitted tailpipe NO_x mass. The time dependency of all variables was omitted to improve readability.

In this work, the mission is assumed to be set by the driver and must be followed exactly. It is therefore not part of the optimization and consequently ω_{e} and T_{e} are exogenous inputs. The system states are ϑ and $m_{\mathrm{NO_x}}^{\mathrm{tp}}$. The inputs to be optimized are $J_{\mathrm{Hd}}^{\mathrm{rot}}$ and $J_{\mathrm{NO_x}}^{\mathrm{rot}}$.

In [5], dynamic programming (DP) was used to solve the OCP, which is a reasonable choice of method when a global optimality guarantee is important and computational time is not limited. However, its use in an online model predictive controller is severely limited by the constraint on the maximal computation time that must be met to allow for feedback control. In fact, examples from the literature show that the time required to solve the OCP with DP for a mission with a duration of roughly 30 min for similar problems is in the order

of a few hours to one week on a standard PC [15–18]. This clearly disqualifies DP as an online optimization method.

A more efficient way to solve the OCP is to employ a direct method, whereby the OCP is discretized and transformed to a finite-dimensional, constrained nonlinear optimization problem, i.e., an NLP [19]. Considering further that many established NLP solvers exist, this approach is well suited for the task at hand.

Common NLP solvers use first-order derivatives and approximate second-order derivatives using first-order derivatives. Hence, the model equations must be at least once continuously differentiable. Evaluating the look-up elements from Sections 2.1 and 2.2 using linear interpolation results in piecewise affine functions that are not continuously differentiable. A replacement for the look-up tables must therefore be found that can characterize the underlying data with the same accuracy, but is continuously differentiable.

In the following, continuously differentiable look-up elements are introduced and their application as a replacement for the look-up elements from Sections 2.1 and 2.2 is outlined. Next, OCP (10a–i) is discretized and an NLP is formulated. The result obtained by the real-time-capable NLP optimization is compared to a benchmark obtained using DP. Finally, the result from the NLP optimization is validated experimentally.

3.1. Continuously Differentiable Look-Up Elements

In this work, a piecewise quadratic interpolant with a continuous derivative is used. Using the terminology from [20], this corresponds to a Hermite quadratic interpolant. The resulting continuously differentiable look-up elements are referred to as cdElements. Their implementation is outlined in Appendix A.

The selected interpolation method is compared to the two predominant interpolation methods, namely linear interpolation and cubic spline interpolation, in Table 2. Compared to linear interpolation, Hermite quadratic interpolation has the advantage of continuous differentiability, whereas, compared to cubic spline interpolation, it has the advantage of less computations required for the evaluation.

Table 2. Comparison of interpolation methods.

Method	Continuously Differentiable	Evaluation Complexity
linear	no	low
Hermite quadratic	once	intermediate
cubic spline	twice	high

The model presented in Section 2.1 uses a five-dimensional look-up element to describe the preoptimized engine operation. According to OCP (10a–i), the model equations only have to be continuously differentiable with respect to ϑ, $J_{\mathrm{Hd}}^{\mathrm{rot}}$, and $J_{\mathrm{NO_x}}^{\mathrm{rot}}$, as only these can be influenced by the optimization. As ω_{e} and T_{e} are given and cannot be affected by the optimization, continuous differentiability in these directions is not required.

Therefore, local continuously differentiable mappings with inputs $(\vartheta, J_{\mathrm{Hd}}^{\mathrm{rot}}, J_{\mathrm{NO_x}}^{\mathrm{rot}})$ are stored for a number of $(\omega_{\mathrm{e}}, T_{\mathrm{e}})$ points on a regular grid. When evaluating the global model, the local models are evaluated at the four $(\omega_{\mathrm{e}}, T_{\mathrm{e}})$ grid points surrounding the current point and bilinear interpolation is used to return the corresponding values for the current point. At each $(\omega_{\mathrm{e}}, T_{\mathrm{e}})$ grid point, two three-dimensional cdElements were used to characterize $J_{\mathrm{fuel}}^{\mathrm{rot}}$ and $\overset{*}{m}_{\mathrm{exh}}$, two two-dimensional cdElements were used to characterize the upper and lower bound on $J_{\mathrm{NO_x}}^{\mathrm{rot}}$, and six one-dimensional cdElements were used to characterize the normalization values. Further, a two-dimensional cdElement was used to characterize the NO$_x$-reduction efficiency map of the ATS used in (8).

3.2. NLP Formulation

In order to transform OCP (10a–i) to an NLP, it was discretized in time using a constant time step $\Delta t = 1\,\mathrm{s}$ and Euler forward integration. Multiple shooting was used for the ATS

temperature, as the state dynamics (10b) and (10c) depend on the ATS temperature in an nonlinear way [21]. For the emitted tailpipe NO_x mass state, on the other hand, single shooting was used, as the state dynamics do not depend on the emitted tailpipe NO_x mass state. The state dynamics (10c), the initial condition (10e), and the terminal constraint (10f) are combined to formulate (11d). The resulting NLP is

$$\underset{\vartheta, J_{Hd}^{rot}, J_{NO_x}^{rot}}{\text{minimize}} \quad \sum_{k=0}^{N} \overset{*}{m}_{fuel}(\cdot)\, \Delta t \tag{11a}$$

$$\text{subject to} \quad \vartheta[0] = \vartheta_0 \tag{11b}$$

$$\vartheta[k+1] - \vartheta[k] - f_\vartheta(\cdot)\, \Delta t = 0 \qquad k = 0, \ldots, N \tag{11c}$$

$$\sum_{k=0}^{N} \overset{*}{m}_{NO_x}^{tp}(\cdot)\, \Delta t \leq \overline{m}_{NO_x}^{tp} \qquad k = 0, \ldots, N \tag{11d}$$

$$J_{Hd}[k] \in [0,\, 1] \qquad k = 0, \ldots, N \tag{11e}$$

$$J_{NO_x}^{rot}[k] \geq J_{NO_x,min}^{rot}(\cdot) \qquad k = 0, \ldots, N \tag{11f}$$

$$J_{NO_x}^{rot}[k] \leq J_{NO_x,max}^{rot}(\cdot) \qquad k = 0, \ldots, N, \tag{11g}$$

where the arguments of $\overset{*}{m}_{fuel}$, f_ϑ, $\overset{*}{m}_{NO_x}^{tp}$, $J_{NO_x,min}^{rot}$, and $J_{NO_x,max}^{rot}$ are equivalent to those in OCP (10a–i) but have been omitted to improve readability. The number of considered time steps N is

$$N = \frac{t_f}{\Delta t} - 1. \tag{12}$$

In this work, the NLP solver IPOPT [22] was used to solve NLP (11a–g).

3.3. Comparison to DP Benchmark

A drawback of using a direct method is that there are no global optimality guarantees for nonconvex problems such as NLP (11a–g). To check the quality of the solution obtained from the NLP optimization, it was compared to a benchmark obtained using DP. As the goal of the DP optimization is to find a benchmark, a fine discretization was used. Specifically, 601 grid points were used for ϑ and 41 for J_{Hd} and $J_{NO_x}^{rot}$. An adapted version of the DP algorithm presented in [23] was used to solve the problem.

The comparison of the trajectories resulting from the DP and the NLP optimizations is shown in Figure 3. No discernible difference in the trajectories can be observed. For a closer analysis, the results listed in Table 3 are considered. The computation times reported were obtained on a PC with a 2.8 GHz quad-core processor and 32 GB of RAM. The results show that the NLP solution fulfills the constraint on the emitted tailpipe NO_x mass exactly. The consumed fuel mass was even 0.049% lower than that of the DP solution. This can be attributed to the fact that even though a fine discretization was used in the DP optimization, the error introduced by the discretization is still nonzero, while the NLP can set the inputs continuously and manages to find a superior solution for the considered case. While this example does not prove that the NLP solver converges to the global optimum in general, it indicates that the resulting solution is of high quality. Similar results were obtained for further missions.

The developed optimization method is a viable option for online optimization, as the time required to solve the NLP is significantly shorter than the duration of the cycle.

3.4. Experimental Validation of Solution

Next, the obtained solution is validated at the testbench by applying the obtained optimal input trajectories in a feedforward fashion. The reference trajectories and the measured trajectories shown in Figure 4 exhibit the same general trends, but it can be observed that the simple control-oriented model is not able to fully capture the behavior of the more complex real system. From the comparison, three main sources of model mismatch can be identified.
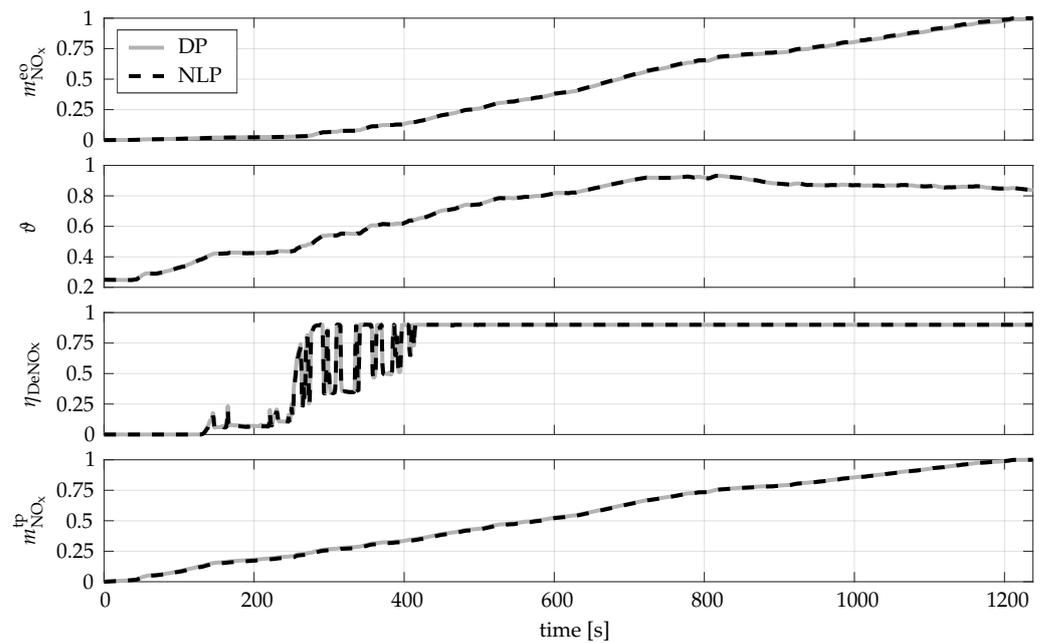
Figure 3. Comparison of the solution to the nonlinear program (NLP) to the dynamic programming (DP) benchmark for the nonroad transient cycle (NRTC) with $\vartheta(0) = 100\,°C$. Shown are the emitted engine-out NO_x mass $m_{NO_x}^{eo}$, the ATS temperature ϑ, the NO_x-reduction efficiency η_{DeNOx}, and the emitted tailpipe NO_x mass $m_{NO_x}^{tp}$. All values were normalized.

Table 3. Comparison of the DP and NLP optimizations for the NRTC with $\vartheta(0) = 100\,°C$. The consumed fuel mass is denoted m_{fuel} and the required computation time is denoted t_{calc}. The results of the DP optimization were used for the normalization.

Method	t_{calc}	$m_{fuel}\,[\%]$	$m_{NO_x}^{eo}\,[\%]$	$m_{NO_x}^{tp}\,[\%]$
DP	15.8 h	100	100	100
NLP	23.4 s	99.951	100.943	100.000

The first error source is the static engine model. The observed error in the emitted engine-out NO_x mass was -9.3%, while the error in the consumed fuel was $+2.9\%$.

The second error source is the simplified thermal model of the ATS that uses a single lumped temperature. Its effect is clearly visible during the initial 300 s of the considered mission. As the engine produces hot exhaust gases, the modeled ATS temperature rises. The temperature measured after the ATS, however, remains low, as there is a non-negligible axial temperature distribution in the real system that is not captured by the model. After this initial warm-up phase, the modeled temperature captures the actual dynamics reasonably well.

The third error source is the static map used to characterize the NO_x-reduction efficiency in the chemical model of the ATS. Neglecting the dynamics of the ammonia storage of the SCR, combined with the error in the ATS temperature, leads to a mismatch in η_{DeNOx}. The mismatch in the timing of the sharp rise in η_{DeNOx} at around 300 s is caused mainly by the error in the ATS temperature, while the mismatch at higher temperatures from 400 s onward is attributed mainly to the neglected ammonia storage dynamics.

The emitted tailpipe NO_x mass is affected by all three of the error sources, resulting in an overall error of $+27.9\%$. In conclusion, the simple control-oriented model captures the basic system behavior, but feedback control is required in order to compensate for model mismatch during operation.
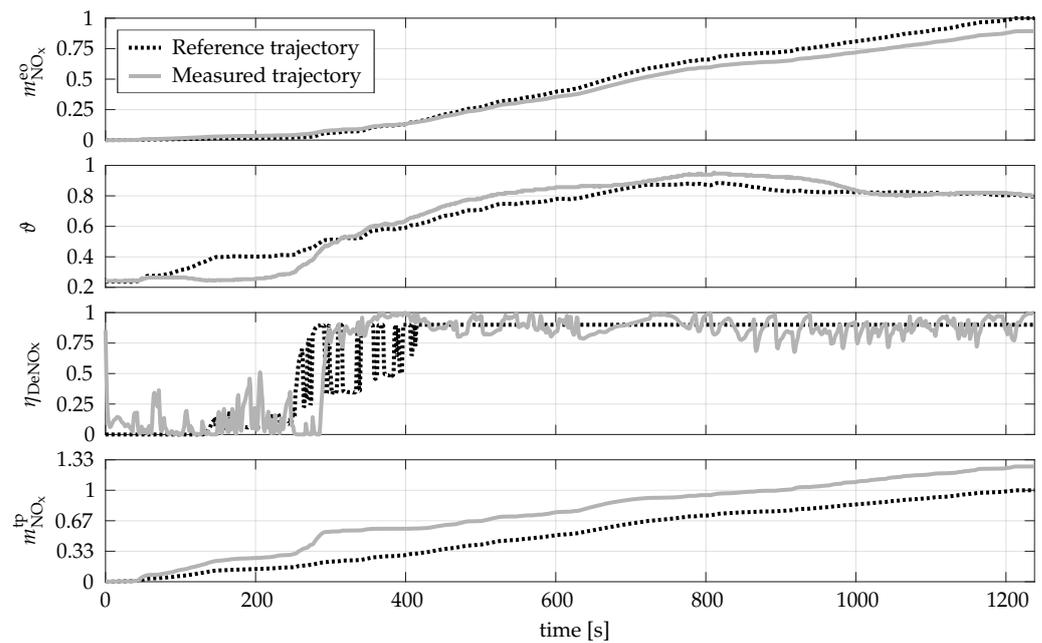
Figure 4. Comparison of the reference trajectories resulting from solving optimal control problem (10a–i) (dotted black) and the corresponding measured trajectories obtained by applying the input trajectories in a feedforward fashion (solid gray) for the NRTC with $\vartheta(0) = 100\,°C$. All values were normalized.

4. Predictive Supervisory Controller

The controller structure used in this work is shown in Figure 5. It consists of a predictive supervisory controller that passes a strategy input vector s to a non-predictive low-level controller that sets the engine control input vector u. The vector of measured system states x is used as feedback. Here, bold symbols are used to denote multidimensional vectors.
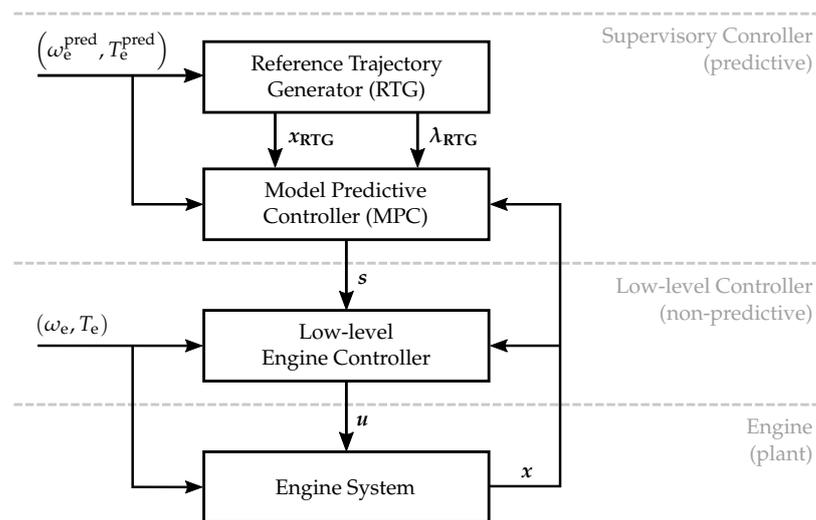
Figure 5. Controller structure consisting of a predictive supervisory controller and a non-predictive low-level engine controller.

The predictive supervisory controller, in turn, contains two layers. In both layers, an optimization problem related to OCP (10a–i) is solved to determine the optimal operation over the prediction horizon. The reference trajectory generator (RTG) considers the entire mission and is run once at the beginning of the mission. It provides reference trajectories that are tracked by the model predictive controller (MPC), which optimizes the engine

operation periodically for a shorter prediction horizon and considers feedback of the measured system states. The MPC's prediction horizon is chosen to be much shorter than that of the RTG, resulting in a shorter computation time and allowing for frequent updates.

At the testbench, the predictive supervisory controller was implemented in Matlab and ran on a PC with a 2.4 GHz dual-core processor and 16 GB of RAM, while the low-level engine controller ran on an ETAS ES910 rapid prototyping unit. INCA's Matlab integration package was used for the communication between the predictive supervisory controller and the low-level engine controller. The stock engine ECU was used to set the engine control inputs.

In the following, the elements of the controller are discussed in more detail.

4.1. Reference Trajectory Generator

The task of the RTG is to optimize the operation of the engine system over the entire mission and provide reference trajectories that the MPC will follow. To achieve this, the RTG solves OCP (10a–i) for the entire mission using the optimization method presented in Section 3. The resulting optimal trajectories for the states as well as the corresponding costates are passed to the MPC as

$$x_{\mathbf{RTG}} = \begin{bmatrix} \vartheta_{\mathrm{RTG}} \\ m^{\mathrm{tp}}_{\mathrm{NO_x,RTG}} \end{bmatrix} \quad \text{and} \quad \lambda_{\mathbf{RTG}} = \begin{bmatrix} \lambda_{\vartheta,\mathrm{RTG}} \\ \lambda_{\mathrm{m,RTG}} \end{bmatrix}. \tag{13}$$

In this work, the predictive information is considered to be of high quality, i.e., there is limited benefit in updating the reference trajectories during the mission. However, the RTG could be updated if a change in the mission or a significant offset of the current states from their respective reference is detected.

4.2. Model Predictive Controller

The task of the MPC is to correct for errors introduced by model mismatch and misprediction in an optimal fashion. To achieve this, an OCP similar to OCP (10a–i) is solved at every update time t_{u} and the resulting strategy inputs are passed to the low-level controller. The OCP is initialized with the current ATS temperature ϑ_0 and emitted tailpipe $\mathrm{NO_x}$ mass $m^{\mathrm{tp}}_{\mathrm{NO_x,0}}$. The terminal state constraints are chosen such that the emitted tailpipe $\mathrm{NO_x}$ mass does not exceed that of the reference and the ATS temperature is at least at the reference value at the end of the MPC's prediction horizon t_{p}.

The OCP solved by the MPC is

$$\underset{J^{\mathrm{rot}}_{\mathrm{Hd}},\, J^{\mathrm{rot}}_{\mathrm{NO_x}},\, \epsilon_\vartheta,\, \epsilon_{\mathrm{m}}}{\text{minimize}} \quad \int_{t_0}^{t_0+t_{\mathrm{p}}} \overset{*}{m}_{\mathrm{fuel}}\big(\omega_{\mathrm{e}}, T_{\mathrm{e}}, \vartheta, J_{\mathrm{Hd}}, J^{\mathrm{rot}}_{\mathrm{NO_x}}\big)\, \mathrm{d}t - \lambda_{\vartheta,\mathrm{RTG}}\, \epsilon_\vartheta + w_{\mathrm{m}}\, \epsilon_{\mathrm{m}} \tag{14a}$$

$$\text{subject to} \quad \frac{\mathrm{d}}{\mathrm{d}t}\vartheta = f_\vartheta\big(\omega_{\mathrm{e}}, T_{\mathrm{e}}, \vartheta, J_{\mathrm{Hd}}, J^{\mathrm{rot}}_{\mathrm{NO_x}}\big) \tag{14b}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} m^{\mathrm{tp}}_{\mathrm{NO_x}} = \overset{*}{m}{}^{\mathrm{tp}}_{\mathrm{NO_x}}\big(\omega_{\mathrm{e}}, T_{\mathrm{e}}, \vartheta, J_{\mathrm{Hd}}, J^{\mathrm{rot}}_{\mathrm{NO_x}}\big) \tag{14c}$$

$$\vartheta(t_0) = \vartheta_0 \tag{14d}$$

$$\vartheta(t_0 + t_{\mathrm{p}}) \geq \underline{\vartheta}_{\mathrm{RTG}} - \epsilon_\vartheta \tag{14e}$$

$$m^{\mathrm{tp}}_{\mathrm{NO_x}}(t_0) = m^{\mathrm{tp}}_{\mathrm{NO_x,0}} \tag{14f}$$

$$m^{\mathrm{tp}}_{\mathrm{NO_x}}(t_0 + t_{\mathrm{p}}) \leq \overline{m}^{\mathrm{tp}}_{\mathrm{NO_x,RTG}} + \epsilon_{\mathrm{m}} \tag{14g}$$

$$J_{\mathrm{Hd}} \in [0, 1] \tag{14h}$$

$$J^{\mathrm{rot}}_{\mathrm{NO_x}} \geq J^{\mathrm{rot}}_{\mathrm{NO_x,min}}\big(\omega_{\mathrm{e}}, T_{\mathrm{e}}, \vartheta, J_{\mathrm{Hd}}\big) \tag{14i}$$

$$J^{\mathrm{rot}}_{\mathrm{NO_x}} \leq J^{\mathrm{rot}}_{\mathrm{NO_x,max}}\big(\omega_{\mathrm{e}}, T_{\mathrm{e}}, \vartheta, J_{\mathrm{Hd}}\big) \tag{14j}$$

$$\epsilon_\vartheta, \epsilon_{\mathrm{m}} \geq 0. \tag{14k}$$

The time dependency of all variables was omitted to improve readability. OCP (14a–k) is solved by adapting the method presented in Section 3. The resulting MPC is a zone-MPC [24], whereby the fuel consumption is minimized, while an upper bound on the emitted tailpipe NO_x mass and a lower bound on the terminal ATS temperature are respected.

The terminal constraint on the emitted tailpipe NO_x mass is given by $\overline{m}_{NO_x,RTG}^{tp}$, which represents the emitted tailpipe NO_x mass of the RTG reference trajectory at the end of the MPC's prediction horizon. A soft constraint was formulated using the slack variable ϵ_m, in order to always allow the solver to find a solution to the OCP. The activation of the slack variable is penalized linearly using w_m. To avoid violation of the soft constraint when the problem is feasible, a large, constant weight w_m was used. This guarantees that constraint (14g) is fulfilled unless it is physically impossible to do so and that the emitted tailpipe NO_x mass is minimized when the non-slacked problem becomes infeasible.

When the entire mission is considered, there is no constraint on the terminal ATS temperature, as it does not influence the tailpipe NO_x emissions. However, the MPC's prediction horizon does generally not reach the end of the mission, meaning that the ATS temperature at the end of the MPC's prediction horizon will influence tailpipe NO_x emissions later on in the mission. It can therefore be beneficial to enforce a terminal constraint on the ATS temperature to shift the engine operation to utilize more fuel to heat up the ATS in the considered prediction horizon, as this can improve the NO_x-reduction efficiency for the remainder of the mission. To account for this, constraint (14e) is introduced, where $\underline{\vartheta}_{RTG}$ represents the ATS temperature of the RTG reference trajectory at the end of the MPC's prediction horizon. As for the tailpipe NO_x emissions, a soft constraint was formulated using the slack variable ϵ_ϑ. As constraint (14e) is only used to steer the MPC in the direction of the reference trajectory, a less aggressive penalty on ϵ_ϑ can be used. In [25], it was shown that the costate from the RTG can be used to quantify the sensitivity of the objective of OCP (10a–i) to a drift from the reference trajectory in a first-order approximation. The ATS temperature costate of the RTG reference at the end of the MPC's prediction horizon $\lambda_{\vartheta,RTG}$ is therefore used as the weight for the penalty on ϵ_ϑ.

4.2.1. MPC Tuning

The MPC's tuning parameters are the update time t_u and the prediction horizon t_p. They were selected by considering the controller performance for a grid of combinations thereof and ensuring that the average computation time required to solve OCP (14a–k) is well below the update time. In the following, $t_u = 20\,s$ and $t_p = 200\,s$ were used.

Note that the MPC's update time $t_u = 20\,s$ is larger than the sampling time $\Delta t = 1\,s$. Therefore, a trajectory of strategy inputs is passed to the low-level engine control rather than a scalar value.

4.2.2. Offset Correction

An offset correction based on [26] was introduced to improve the MPC's tracking performance. Thereby, the MPC's terminal constraints are adapted using the current offset. For the emitted tailpipe NO_x mass, it results in

$$e_m = m_{NO_x,RTG}^{tp}(t_0) - m_{NO_x}^{tp}(t_0) \tag{15}$$

$$\overline{m}_{NO_x,RTG}^{tp} = m_{NO_x,RTG}^{tp}(t_0 + t_p) + e_m \, , \tag{16}$$

while, for the ATS temperature, this results in

$$e_\vartheta = \vartheta_{RTG}(t_0) - \vartheta(t_0) \tag{17}$$

$$\underline{\vartheta}_{RTG} = \vartheta_{RTG}(t_0 + t_p) + e_\vartheta \, . \tag{18}$$

The theory behind this correction guarantees offset-free tracking in steady state for linear MPC [21]. Even though the references never reach a steady state and nonlinear MPC is used here, the method was found to improve the tracking behavior of the MPC.

4.3. Low-Level Engine Controller

In this work, two realizations of the low-level engine controller are considered. The task of both realizations is to set the engine control input vector as stated in (1). However, the two realizations differ in the strategy input vector that they receive and in how the engine control inputs are determined. The two realizations are introduced in Sections 5.1 and 6.1, respectively.

5. Controller Performance without Misprediction

In this section, the performance of the predictive supervisory controller with perfect predictive information, i.e., without mispredictions, is analyzed experimentally. In order to obtain comparable and repeatable results, the ATS was preconditioned by emptying the ammonia storage and cooling it to $100\,^{\circ}\mathrm{C}$ before each experiment. As the initial ATS temperature is well below the temperature where the injection of AdBlue is enabled, the result is similar to a cold start.

5.1. Low-Level Engine Controller

The low-level engine controller used in this section evaluates the smoothed inverse mapping introduced in Section 2.1 to determine the engine control inputs. The maps are evaluated at the current engine speed and torque, the current ATS temperature, and the engine strategy inputs selected by the MPC. The strategy input vector becomes

$$s = \begin{bmatrix} J_{\mathrm{Hd}}^{\mathrm{rot}} \\ J_{\mathrm{NO_x}}^{\mathrm{rot}} \end{bmatrix}. \tag{19}$$

Note that the evaluation of the smoothed inverse mapping is not part of the optimization itself, but is carried out afterwards. Hence, continuous differentiability is not required and the smoothed input maps from Section 2.1 can be applied directly. This realization of the low-level engine controller is computationally inexpensive as finding the optimal engine control inputs is reduced to evaluating a mapping obtained offline.

5.2. Reference Tracking

First, the reference tracking capability of the controller is investigated; see Figure 6. For the first $300\,\mathrm{s}$, the operations with and without the MPC were equivalent. Due to the model error discussed in Section 3.4, the emitted tailpipe NO_x mass lies above the reference. During this time, the ATS temperature was lower than expected, resulting in a lower NO_x-reduction efficiency. Considering further that the operation without the MPC already produced minimal engine-out NO_x emissions, it becomes evident that a saturation is active, and even when feedback is introduced via the MPC, the emitted tailpipe NO_x mass cannot be reduced further in this part of the mission.

At $300\,\mathrm{s}$, the measured NO_x-reduction efficiency rose rapidly and the operation without the MPC moved to an engine operation with higher engine-out NO_x emissions. The operation with the MPC, on the other hand, continued to operate the engine with minimal engine-out NO_x emissions until the tailpipe NO_x reference was reached at around $500\,\mathrm{s}$. From there on, the operation with the MPC also switched to an engine operation with higher engine-out NO_x emissions and continued to track the tailpipe NO_x reference until the end of the mission.
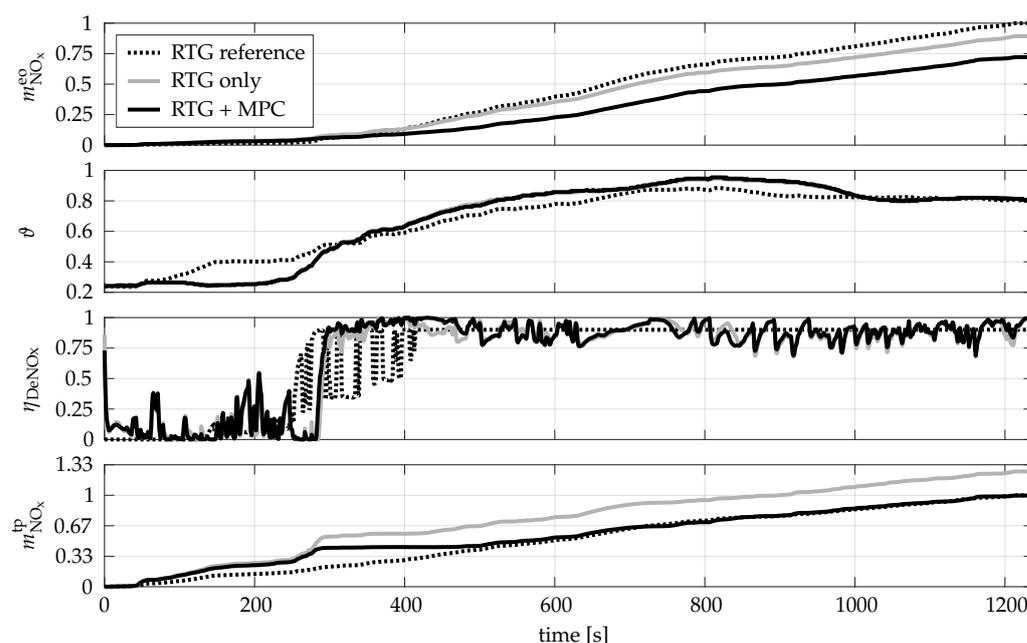
Figure 6. Resulting trajectories for the NRTC. The dotted gray trajectories were obtained by the RTG optimization. The solid gray trajectories show the measured operation with only the RTG, i.e., without feedback from the MPC. The solid black trajectories show the measured operation with the full predictive supervisory controller, i.e., with feedback from the MPC. All values were normalized.

The result shows that the effect of the model mismatch can be compensated for by the MPC when the operation is not limited by a saturation such as the lower limit on the engine-out NO_x emissions. When the reference trajectories can no longer be tracked exactly, the MPC follows them as well as the system allows. This can be observed during the initial 300 s of the mission, where the controller tolerated a violation of the terminal constraint on ϑ in order to get as close to the $m_{NO_x}^{tp}$ reference as possible.

5.3. Comparison to Fixed Engine Calibration

Next, the performance of the predictive supervisory controller is compared to that achieved using the fixed engine calibration of the stock ECU; see Figure 7. In the absence of a supervisory controller, the engine-out NO_x emissions depend solely on the mission. High tailpipe NO_x emissions result when the ATS is inactive (initial 300 s) and low tailpipe NO_x emissions result once the ATS has reached its operating temperature (from 300 s onward).

The tailpipe NO_x mass emitted by the fixed engine calibration was used as the limit for the predictive supervisory controller. During the initial 300 s of the cycle, the predictive supervisory controller adapted the engine operation to the low NO_x-reduction efficiency by reducing the engine-out NO_x emissions. At the end of this phase, the tailpipe NO_x mass emitted by the predictive supervisory controller was roughly half of that achieved with the fixed engine calibration. Once the ATS was fully operational, the predictive supervisory controller reacted by allowing higher engine-out NO_x emissions, thereby reducing the fuel consumption.

The final emitted tailpipe NO_x mass of the predictive supervisory controller was equivalent to that achieved with the fixed engine calibration, while the fuel consumption was reduced by 1.1%.
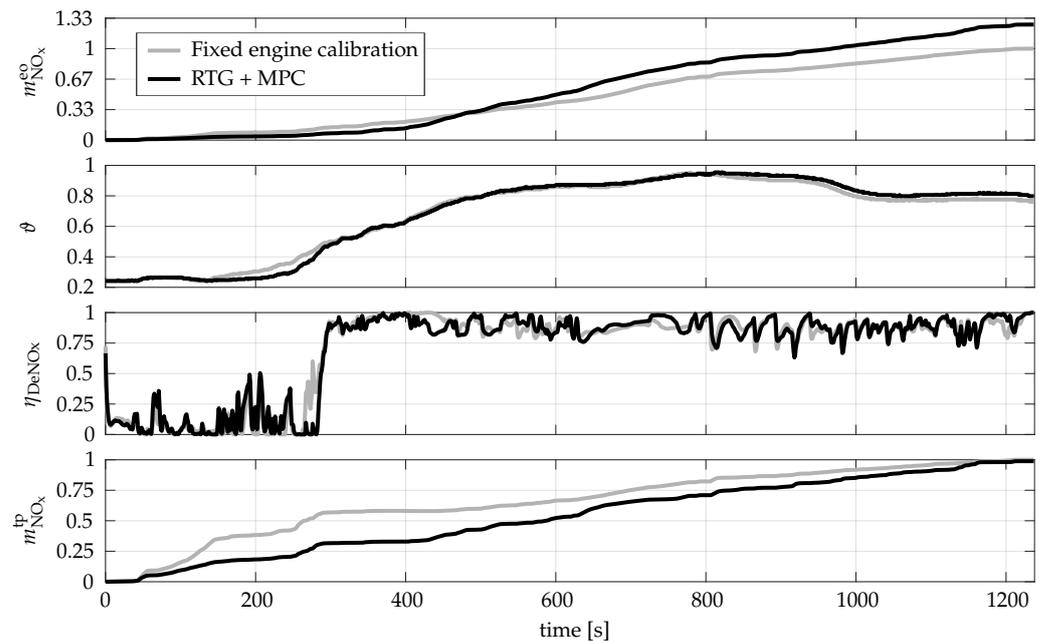
Figure 7. Comparison of the measurement obtained with the fixed engine calibration (gray) to that obtained with the predictive supervisory controller (black) for the NRTC. Using the predictive supervisory controller reduced the fuel consumption by 1.1%. All values were normalized.

5.4. Controllability Limits

On the left of Figure 8, the resulting operation for the nonroad transient cycle (NRTC) and different NO_x limits is shown. The difference between the RTG reference (triangles) and the measurements (circles) results from model mismatch. The results obtained without feedback (gray circles) always show a deviation from the tailpipe NO_x limit, which is reduced when feedback is introduced (black circles). The tailpipe NO_x limit was only respected and fully exploited for $\overline{m}_{NO_x}^{tp} = 0.66$ and $\overline{m}_{NO_x}^{tp} = 1$, while more stringent limits were violated and more relaxed limits were not fully exploited.

To investigate this behavior, the constant operating point cycle is considered. This mission consists of a constant engine speed of 1900 rpm and a constant engine torque of 200 Nm for the duration of 1800 s. Note that as the ATS is preconditioned to $\vartheta(0) = 100\,°C$, a dynamic operation results as the ATS warms up, even though the engine speed and torque are constant. The benefit of considering such a cycle is that any changes in the engine operation can be attributed to the predictive supervisory controller, rather than to a change in the engine speed or torque. The results for the constant operating point cycle are shown on the right of Figure 8.

As for the NRTC, intermediate tailpipe NO_x limits were respected and fully exploited by the predictive supervisory controller, whereas stringent limits were violated and relaxed limits were not fully exploited. Figure 9 shows the resulting trajectories for a representative of each of the three cases.

For an intermediate NO_x limit, i.e., $\overline{m}_{NO_x}^{tp} = 1$, the RTG set the engine-out NO_x emissions to the minimum during the initial 300 s of the cycle. Due to model mismatch, these low engine-out NO_x emissions could not be achieved at the testbench, resulting in the emitted tailpipe NO_x mass rising faster than the reference. The predictive supervisory controller reacted by keeping the engine-out NO_x emissions at the minimum until the tailpipe NO_x reference was reached at around 600 s. From there on, the tailpipe NO_x reference was tracked well, even allowing the engine-out NO_x emissions to exceed those of the RTG.
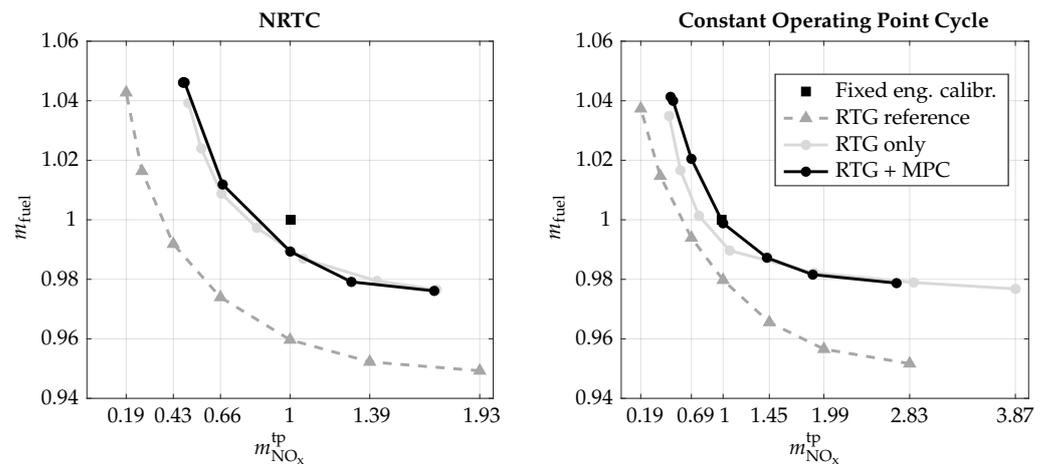
Figure 8. Pareto fronts obtained by setting different limits on the emitted tailpipe NO_x mass. The black square marker shows the operation with the fixed engine calibration. The gray triangular markers show the results from the RTG optimization for different NO_x limits. The gray circular markers show the corresponding measured operation with the RTG result used in a purely feedforward fashion, i.e., without feedback from the MPC. The black circular markers show the corresponding measured operation with the full predictive supervisory controller, i.e., with feedback from the MPC. The plot on the left shows the results for the NRTC, while the plot on the right shows the results for a mission consisting of a constant speed of 1900 rpm and a constant torque of 200 Nm for the duration of 1800 s. All values were normalized.
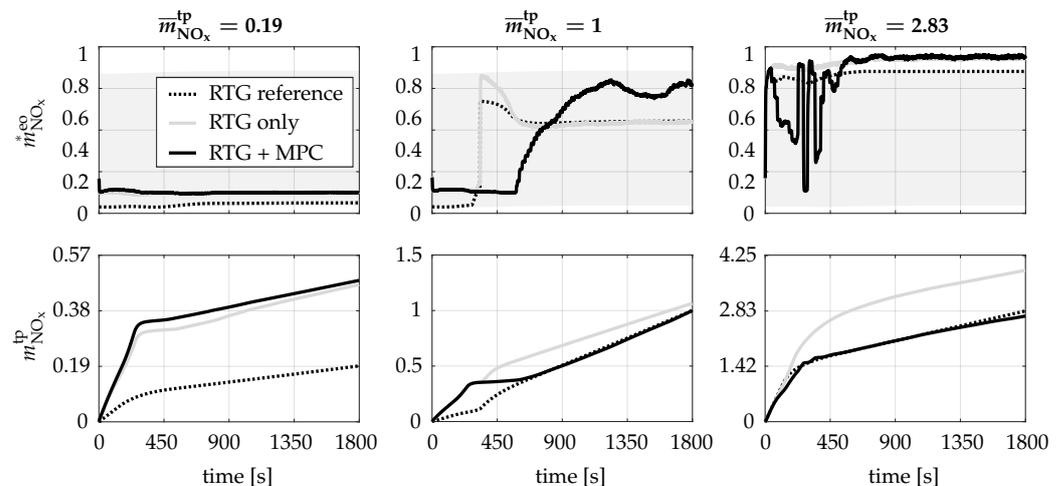
Figure 9. Resulting trajectories for the constant operating point cycle for different NO_x limits. The gray regions in the upper row of plots show the modeled feasible range for the engine-out NO_x mass flow. The dotted black trajectories were obtained by the RTG optimization. The solid gray trajectories show the measured operation with only the RTG, i.e., without feedback from the MPC. The solid black trajectories show the measured operation with the full predictive supervisory controller, i.e., with feedback from the MPC. All values were normalized.

For minimal tailpipe NO_x emissions, i.e., $\overline{m}_{NO_x}^{tp} = 0.19$, the RTG operated the engine at minimal engine-out NO_x emissions throughout the cycle. As previously observed, the actual engine-out NO_x emissions were significantly higher. As a consequence, the emitted tailpipe NO_x mass exceeded the final limit within the first 300 s of the cycle and the predictive supervisory controller continued to keep NO_x emissions as low as possible. The observed violation of the NO_x limit must therefore be attributed to a lower limit on the engine-out NO_x emissions that is higher than the minimal engine-out NO_x emissions

of the model. The MPC is confronted with a lower saturation that limits its reference tracking ability.

For maximal tailpipe NO_x emissions, i.e., $\overline{m}_{NO_x}^{tp} = 2.83$, the RTG operated the engine at maximal engine-out NO_x emissions throughout the cycle. The resulting operation of the predictive supervisory controller can be divided into two phases. For the initial 500 s, the MPC controller adapted the engine-out NO_x emissions in such a way that the tailpipe NO_x reference was tracked. From 500 s onward, the engine-out NO_x emissions were at the maximum. Due to model mismatch, the actual engine-out NO_x emissions exceeded the value predicted by the model. However, as the NO_x-reduction efficiency was also higher than modeled, the resulting tailpipe NO_x emissions still fell below the reference. While the MPC tried to compensate for this mismatch, in this case, it is confronted with an upper saturation that limits its reference tracking ability.

In summary, the predictive supervisory controller showed good reference tracking behavior when the system bounds allowed for sufficient controllability. When the reference obtained from the RTG could not be achieved with the real system due to saturation, the predictive supervisory controller continued to track the reference as well as the real system allowed.

5.5. Benefit of Online Control

Finally, the potential of the predictive supervisory controller to adapt to different conditions and tailpipe NO_x limits is analyzed; see Figure 10. The comparison on the left corresponds to the one analyzed in Section 5.3. The measured operation obtained with the predictive supervisory controller shows that the tailpipe NO_x limit was respected and that the fuel consumption was reduced by 1.1%.
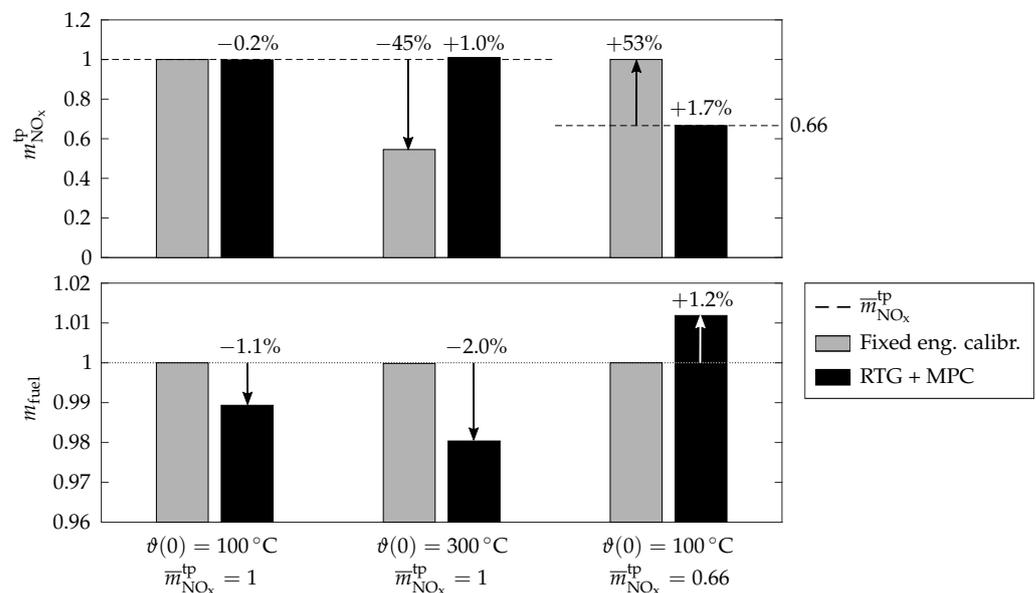
Figure 10. Comparison of the measured operation for the NRTC for different initial ATS temperatures and tailpipe NO_x limits. The results obtained with the fixed engine calibration are shown in gray, while the results obtained with the predictive supervisory controller are shown in black. The results on the left and the right correspond to points on the Pareto front on the left of Figure 8. All values were normalized with respect to the operation with the fixed engine calibration for $\vartheta(0) = 100\,^{\circ}C$ and $\overline{m}_{NO_x}^{tp} = 1$.

For the comparison in the middle, the same tailpipe NO_x limit as before was set, but the ATS was preconditioned to a temperature of 300 °C and the corresponding steady-state value for the ammonia storage. As a result, the average NO_x-reduction efficiency was much higher. The operation obtained with the fixed engine calibration resulted in a final emitted

tailpipe NO$_x$ mass 45% below the limit. The fixed engine calibration is too conservative for this mission and fuel could have been saved by adapting it. This was achieved by the predictive supervisory controller, which exploited the fact that the tailpipe NO$_x$ limit was less restrictive, resulting in a fuel reduction of 2.0% while exceeding the tailpipe NO$_x$ limit by 1.0%.

For the comparison on the right, the initial ATS temperature was $100\,^\circ$C but the tailpipe NO$_x$ limit was reduced to 0.66. The operation obtained with the fixed engine calibration reached a final emitted tailpipe NO$_x$ mass 53% above the limit. The supervisory controller, on the other hand, adapted to the new tailpipe NO$_x$ limit and only exceeded it by 1.7%. This was achieved at the cost of a 1.2% increase in fuel consumption.

In summary, the predictive supervisory controller was shown to adapt well to changes in the initial ATS temperature and the tailpipe NO$_x$ limit, adapting to minimize the fuel consumption, while never exceeding the tailpipe NO$_x$ limit by more than 1.7%.

6. Controller Performance with Misprediction

The drawback of the low-level engine controller used in the previous section is that it is not robust against mispredictions. The task of the supervisory controller is to select the optimal trade-off between the individual engine objectives. This is achieved by selecting the engine strategy inputs $J_{\text{Hd}}^{\text{rot}}$ and $J_{\text{NO}_x}^{\text{rot}}$, which describe a relative position on the Pareto front. However, when mispredictions occur, the form of the Pareto front and the trade-off described by $J_{\text{Hd}}^{\text{rot}}$ and $J_{\text{NO}_x}^{\text{rot}}$ changes, resulting in an engine operation that will be suboptimal for the current mission. Furthermore, as the bounds on $J_{\text{NO}_x}^{\text{rot}}$ change with the engine speed and torque, it is possible that the selected point lies outside the bounds. As a consequence, the engine strategy inputs $J_{\text{Hd}}^{\text{rot}}$ and $J_{\text{NO}_x}^{\text{rot}}$ are not suited to an operation with misprediction. An alternative low-level controller, suited to deal with misprediction, is introduced and validated in this section.

6.1. Low-Level Engine Controller

The realization of the low-level engine controller used in this section determines the engine control inputs by solving a static optimization problem online. The optimization performed by the predictive supervisory controller is equivalent to that of the previous section, but the engine strategy inputs provided to the low-level engine controller change to

$$ \boldsymbol{s} = \begin{bmatrix} w_{\text{Hd}} \\ w_{\text{NO}_x}^{\text{tp}} \end{bmatrix}, \tag{20} $$

where w_{Hd} and $w_{\text{NO}_x}^{\text{tp}}$ represent equivalence factors for the enthalpy provided to the ATS and the tailpipe NO$_x$ emissions, respectively. Their interpretation is analogous to that of the equivalence factor used to penalize the use of the battery in the energy management of HEVs.

The values for w_{Hd} and $w_{\text{NO}_x}^{\text{tp}}$ are calculated from the costates of the MPC optimization as

$$ w_{\text{Hd}} = -\frac{\lambda_\vartheta}{C_{\text{ATS}}} \tag{21} $$

$$ w_{\text{NOx}}^{\text{tp}} = \lambda_{\text{NOx}}^{\text{tp}} . \tag{22} $$

The corresponding derivation is shown in Appendix B. The equivalence factor for tailpipe NO$_x$ emissions is then converted to an equivalence factor for engine-out NO$_x$ emissions using the current NO$_x$-reduction efficiency η_{DeNOx}, resulting in

$$ w_{\text{NOx}}^{\text{eo}} = (1 - \eta_{\text{DeNOx}}) \, w_{\text{NOx}}^{\text{tp}} . \tag{23} $$

Finally, the low-level engine controller determines the engine control inputs by minimizing the objective

$$L = \overset{*}{m}_{\text{fuel}}(\boldsymbol{u}) + w_{\text{NOx}}^{\text{eo}} \, \overset{*}{m}_{\text{NOx}}^{\text{eo}}(\boldsymbol{u}) - w_{\text{Hd}} \, \overset{*}{H}_{\text{ATS}}(\boldsymbol{u}) \,, \tag{24}$$

while considering all limits on the engine operation. The objective L represents a weighted sum of multiple objectives, where the equivalence factor $w_{\text{NOx}}^{\text{eo}}$ is used to penalize engine-out NO$_\text{x}$ emissions and the equivalence factor w_{Hd} is used to encourage ATS heating. The optimal engine control inputs are determined using a grid search.

With this realization, the optimal engine control inputs are determined for the actual mission rather than for the predicted one, ensuring robustness toward mispredictions, at the cost of an increased computational demand.

6.2. Robustness to Misprediction

In this section, the performance of the predictive supervisory controller is analyzed for two misprediction cases. To obtain comparable results, the ATS is preconditioned as described in Section 5.

Figure 11 shows the prediction cycles considered in this section. The first misprediction case is shown in the middle. It was obtained by filtering the NRTC using a low-pass filter with a cut-off frequency of 0.01 Hz and represents the case where only averaged information about the upcoming missions is available. The second misprediction case is shown on the right. It was obtained by rearranging the individual segments of the NRTC and represents the case where the timing of the predictive information is wrong. The cycles with misprediction were used by the predictive supervisory controller, while the NRTC was followed during all experiments.
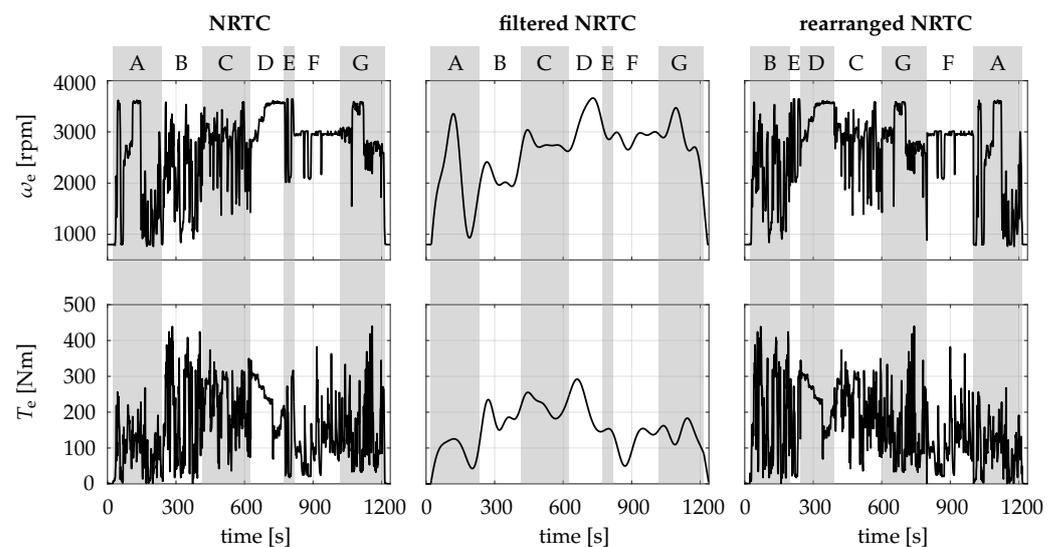
Figure 11. Considered misprediction cases. The NRTC is shown on the left. It consists of seven segments that describe the operation of different nonroad applications [27]. These segments are visualized by the gray areas and assigned a capital letter to illustrate their order. The filtered NRTC is shown in the middle. The rearranged NRTC is shown on the right.

The corresponding results are shown in Figure 12. The upper plot shows the error in the emitted tailpipe NO$_\text{x}$ mass. The effect of misprediction can be analyzed by first considering the operation without feedback from the MPC. For the case without misprediction (left), the error of 10.5% stems solely from model mismatch. For the cycles with misprediction, the error increased to 13.0% for the filtered NRTC (middle) and to 20.3% for the rearranged NRTC (right).
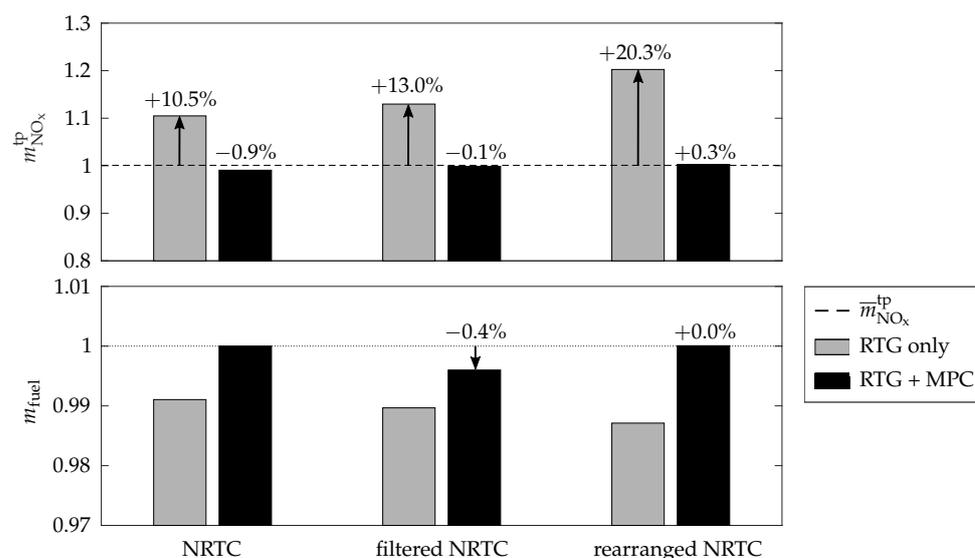
Figure 12. Comparison of the measured operation for different misprediction cases. The results obtained without feedback from the MPC are shown in gray, while the results obtained with the full predictive supervisory controller are shown in black. The emitted tailpipe NO_x mass is normalized with respect to the limit, while the fuel mass is normalized with respect to the operation with the full predictive supervisory controller for the case without misprediction.

When feedback was introduced via the MPC, the predictive supervisory controller never exceeded the tailpipe NO_x limit by more than 0.3% for all considered cases. This result indicates that the predictive supervisory controller is robust towards mispredictions. Furthermore, the increased feedback control action did not result in an increase in fuel consumption.

In fact, the fuel consumption for the filtered NRTC is lower than that for the case without misprediction. A possible explanation for this is that the misprediction counteracts the model mismatch, resulting in a reference trajectory that results in better performance on the real system. This suggests that using a filtered prediction could improve the performance of the predictive supervisory controller.

7. Conclusions

In this work, a predictive supervisory controller for the integrated emissions management of a diesel engine and its aftertreatment system was developed. First, an efficient optimization method was developed to solve the optimal control problem of operating the engine system over a mission. The considered objective was the minimization of the fuel consumption, while respecting an upper bound on the emitted tailpipe NO_x mass. Next, the resulting optimization method was used in a predictive supervisory controller and the controller performance was evaluated experimentally.

The results show that the predictive supervisory controller successfully adapts the engine operation to the specific mission. At equivalent tailpipe NO_x emissions, the resulting controller was shown to reduce the fuel consumption by 1.1% for the nonroad transient cycle, compared to the operation with a fixed engine calibration. Further, the controller's robustness to different missions, initial ATS temperatures, and NO_x limits was demonstrated.

Finally, an adaptation of the low-level engine controller was presented that allows the controller to deal with mispredictions. The results for two misprediction cases showed that the predictive supervisory controller never exceeded the tailpipe NO_x limit by more than 0.3% and did not require more fuel than for the case without misprediction.

Based on this work, there are three avenues for future research. First, the developed methodology could be investigated further. For example, the type of predictive information

required by the predictive supervisory controller could be analyzed and further experiments with filtered predictive information could be performed. Further, improvements to the model and the optimization method could be considered to lower the computation time or achieve a more accurate result.

Second, the presented methodology could be used for the development of a predictive supervisory controller for series production vehicles. The focus should lie on the integration of the presented optimization method into embedded systems and on obtaining predictive information.

Third, the methodology could be used to obtain an experimental benchmark for the development of simpler, potentially non-predictive supervisory controllers. During the development of such controllers, their performance can be evaluated by comparing it to that of the method developed in this work.

Appendix A. Implementation of cdElements

The basic principle of the continuously differentiable look-up elements (cdElements) is illustrated using the example shown in Figure A1.
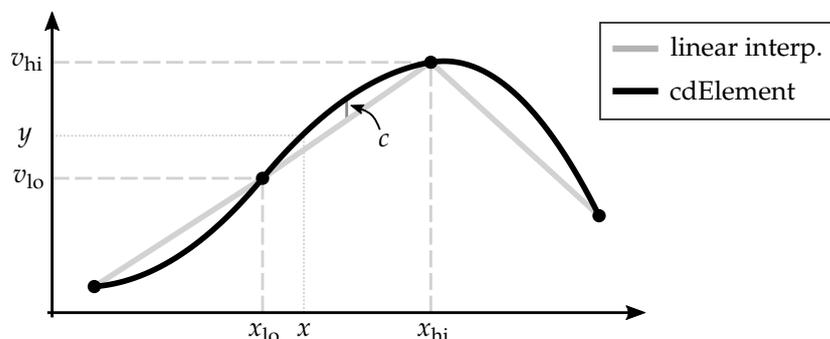
Figure A1. Illustration of a one-dimensional continuously differentiable look-up element.

The cdElement includes a center c between any two vertices v_{lo} and v_{hi} and a quadratic function is evaluated as

$$\xi = \frac{x - x_{lo}}{x_{hi} - x_{lo}} \tag{A1}$$

$$y = (1 - \xi)\, v_{lo} + \xi v_{hi} + 4\,\xi\,(1 - \xi)\, c = v_{lo} + (v_{hi} - v_{lo} + 4\,c)\,\xi - 4\,c\,\xi^2 \,. \tag{A2}$$

where ξ is the relative position of x with respect to v_{lo} and v_{hi}, and the center c is the distance between the quadratic function and the line segment connecting the vertices x_{lo} and x_{hi} at $\xi = 0.5$.

During the fitting process of cdElements, continuous differentiability is enforced using respective constraints at each vertex. The example in Figure A1 features four grid points and four vertices. Additionally, the cdElement features three centers: one for each of the three segments. To enforce continuous differentiability, two constraints are considered

during the fitting process. The first constraint enforces that the gradient of the left segment and the gradient of the middle segment must be equal at x_{lo}. The second constraint enforces that the gradient of the middle segment and the gradient of the right segment must be equal at x_{hi}. The gradient at the outer bounds of the outermost segments is free.

The resulting evaluation of the cdElement is shown by the solid black line in Figure A1 and is smooth. The first and second derivative of cdElements are computed as

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial \xi} \frac{\partial \xi}{\partial x} = \frac{1}{x_{\mathrm{hi}} - x_{\mathrm{lo}}} \left(v_{\mathrm{hi}} - v_{\mathrm{lo}} + 4\, c - 8\, c\, \xi \right) \tag{A3}$$

$$\frac{\partial^2 y}{\partial x^2} = \frac{\partial}{\partial \xi} \left(\frac{\partial y}{\partial x} \right) \frac{\partial \xi}{\partial x} = \frac{-8\, c}{(x_{\mathrm{hi}} - x_{\mathrm{lo}})^2} \,. \tag{A4}$$

The evaluation of the derivatives is computationally inexpensive. The quadratic function between two grid points results in a piecewise affine first derivative and a piecewise constant second derivative. As a result, the cdElements are once continuously differentiable.

For the sake of brevity, the working principle of cdElements is only shown for a one-dimensional look-up element. Nevertheless, the method is extendable to higher-dimensional look-up elements. For a two-dimensional cdElement, the line segment between two vertices is replaced by a surface spanned by four vertices. Four edge centers and one face center are added. Continuous differentiability must be enforced along all four edges. In three-dimensional cdElements, cubes spanned by eight vertices must be considered. Twelve edge centers, six face centers, and one cube center must be added. Continuous differentiability must be enforced along all six faces.

While the method is applicable to multi-dimensional look-up elements, the scalability with dimensionality limits its application. In this work, one-, two-, and three-dimensional cdElements were used.

Appendix B. Derivation of Equivalence Factors

The Hamiltonian of OCP (14a–k) and the objective of the low-level engine controller presented in Section 6.1 rewritten to consider tailpipe NO$_\mathrm{x}$ emissions are

$$H_{\mathrm{MPC}} = \overset{*}{m}_{\mathrm{fuel}} + \lambda_{\mathrm{NO_x}}^{\mathrm{tp}} \frac{\mathrm{d}}{\mathrm{d}t} m_{\mathrm{NO_x}}^{\mathrm{tp}} + \lambda_{\vartheta} \frac{\mathrm{d}}{\mathrm{d}t} \vartheta \tag{A5}$$

$$L = \overset{*}{m}_{\mathrm{fuel}} + w_{\mathrm{NO_x}}^{\mathrm{tp}}\, \overset{*}{m}_{\mathrm{NO_x}}^{\mathrm{tp}} - w_{\mathrm{Hd}}\, \overset{*}{H}_{\mathrm{ATS}} \,. \tag{A6}$$

The equivalence factors $w_{\mathrm{NO_x}}^{\mathrm{tp}}$ and w_{Hd} are found by ensuring that L represents the same cost function as H_{MPC}. The first term is already equivalent. Comparing the second term and considering that $\frac{\mathrm{d}}{\mathrm{d}t} m_{\mathrm{NO_x}}^{\mathrm{tp}} = \overset{*}{m}_{\mathrm{NO_x}}^{\mathrm{tp}}$, this results in

$$w_{\mathrm{NO_x}}^{\mathrm{tp}} = \lambda_{\mathrm{NO_x}}^{\mathrm{tp}} \,. \tag{A7}$$

Comparing the third term yields

$$-w_{\mathrm{Hd}}\, \overset{*}{H}_{\mathrm{ATS}} = \lambda_{\vartheta} \frac{\mathrm{d}}{\mathrm{d}t} \vartheta = \frac{\lambda_{\vartheta}}{C_{\mathrm{ATS}}} \left(\overset{*}{H}_{\mathrm{ATS}} - \alpha\, S\, (\vartheta - \vartheta_{\mathrm{amb}}) \right) \tag{A8}$$

$$w_{\mathrm{Hd}} = -\frac{\lambda_{\vartheta}}{C_{\mathrm{ATS}}} \left(1 - \frac{\alpha\, S}{\overset{*}{H}_{\mathrm{ATS}}} (\vartheta - \vartheta_{\mathrm{amb}}) \right) , \tag{A9}$$

which is approximated by neglecting the convective heat loss to the environment, resulting in

$$w_{\mathrm{Hd}} \approx -\frac{\lambda_{\vartheta}}{C_{\mathrm{ATS}}} \,. \tag{A10}$$

References

1. Emission Standards. Available online: https://dieselnet.com/standards/ (accessed on 24 February 2022).
2. Guzzella, L.; Onder, C. *Introduction to Modeling and Control of Internal Combustion Engine Systems*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009.
3. Nova, I.; Tronconi, E. *Urea-SCR Technology for deNOx after Treatment of Diesel Exhausts*; Springer: New York, NY, USA, 2014.
4. Cloudt, R.; Willems, F. Integrated emission management strategy for cost-optimal engine-aftertreatment operation. *SAE Int. J. Engines* **2011**, *4*, 1784–1797. [CrossRef]
5. Ritzmann, J.; Chinellato, O.; Hutter, R.; Onder, C. Optimal integrated emission management through variable engine calibration. *Energies* **2021**, *14*, 7606. [CrossRef]
6. Zentner, S.; Asprion, J.; Onder, C.; Guzzella, L. An equivalent emission minimization strategy for causal optimal control of diesel engines. *Energies* **2014**, *7*, 1230–1250. [CrossRef]
7. Elbert, P.; Amstutz, A.; Onder, C. Adaptive control for the real driving emissions of diesel engines. *MTZ Worldw.* **2017**, *78*, 68–74. [CrossRef]
8. Guzzella, L.; Sciarretta, A. *Vehicle Propulsion Systems*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 1.
9. van Dooren, S.; Amstutz, A.; Onder, C.H. A causal supervisory control strategy for optimal control of a heavy-duty diesel engine with SCR aftertreatment. *Control Eng. Pract.* **2022**, *119*, 104982. [CrossRef]
10. Chen, P.; Wang, J. Nonlinear model predictive control of integrated diesel engine and selective catalytic reduction system for simultaneous fuel economy improvement and emissions reduction. *J. Dyn. Syst. Meas. Control* **2015**, *137*, 081008. [CrossRef]
11. Mentink, P.; Willems, F.; Kupper, F.; den Eijnden, E.V. *Experimental Demonstration of a Model-Based Control Design and Calibration Method for Cost Optimal Euro-VI Engine-Aftertreatment Operation*; SAE Technical Paper Series; Society of Automotive Engineers: Warrendale, PA, USA, 2013.
12. Ramachandran, S.N.; Hommen, G.; Mentink, P.; Seykens, X.; Willems, F.; Kupper, F. Robust, cost-optimal and compliant engine and aftertreatment operation using air-path control and tailpipe emission feedback. *SAE Int. J. Engines* **2016**, *9*, 1662–1673. [CrossRef]
13. Maamria, D.; Chaplais, F.; Petit, N.; Sciarretta, A. Numerical optimal control as a method to evaluate the benefit of thermal management in hybrid electric vehicles. *IFAC Proc. Vol.* **2014**, *47*, 4807–4812. [CrossRef]
14. Maamria, D.; Sciarretta, A.; Chaplais, F.; Petit, N. Online energy management system (EMS) including engine and catalyst temperatures for a parallel HEV. *IFAC-PapersOnLine* **2017**, *50*, 8913–8920. [CrossRef]
15. van Schijndel, J.; Donkers, M.C.F.; Willems, F.P.T.; Heemels, W. Dynamic programming for integrated emission management in diesel engines. *IFAC Proc. Vol.* **2014**, *47*, 11860–11865. [CrossRef]
16. Simon, A.; Nelson-Gruel, D.; Charlet, A.; Jaine, T.; Nouillant, C.; Chamaillard, Y. Optimal supervisory control of a diesel HEV taking into account both DOC and SCR efficiencies. *IFAC-PapersOnLine* **2018**, *51*, 323–328. [CrossRef]
17. Ritzmann, J.; Christon, A.; Salazar, M.; Onder, C. *Fuel-Optimal Power Split and Gear Selection Strategies for a Hybrid Electric Vehicle*; SAE Technical Paper; Society of Automotive Engineers: Warrendale, PA, USA, 2019.
18. Ritzmann, J.; Lins, G.; Onder, C. Optimization method for the energy and emissions management of a hybrid electric vehicle with an exhaust aftertreatment system. *IFAC-PapersOnLine* **2020**, *53*, 13797–13804. [CrossRef]
19. Asprion, J. Optimal Control of Diesel Engines. Ph.D. Thesis, ETHZ, Zürich, Switzerland, 2013.
20. Kahaner, D.; Moler, C.; Nash, S. *Numerical Methods and Software*; Prentice-Hall, Inc.: Hoboken, NJ, USA, 1989.
21. Albin Rajasingham, T. *Nonlinear Model Predictive Control of Combustion Engines*; Springer: Cham, Switzerland, 2021.
22. Wächter, A.W. *Interior Point Methods for Large-Scale Nonlinear Programming with Applications in Process Systems Engineering*; Carnegie Mellon University: Pittsburgh, PA, USA, 2002.
23. Sundström, O.; Guzzella, L. A generic dynamic programming Matlab function. In Proceedings of the Control Applications, (CCA) & Intelligent Control, (ISIC), St. Petersburg, Russia, 8–10 July 2009; pp. 1625–1630.
24. Maciejowski, J.M. *Predictive Control: With Constraints*; Pearson Education: London, UK, 2002.
25. Salazar, M.; Balerna, C.; Elbert, P.; Grando, F.P.; Onder, C.H. Real-time control algorithms for a hybrid electric race car using a two-level model predictive control scheme. *IEEE Trans. Veh. Technol.* **2017**, *66*, 10911–10922. [CrossRef]
26. Rawlings, J.B. Tutorial overview of model predictive control. *IEEE Control Syst. Mag.* **2000**, *20*, 38–52.
27. Ettl, J.; Bernhardt, H.; Thuneke, K.; Emberger, P.; Remmele, E. Transient emission and fuel consumption measurements on plant oil tractors. *Landtechnik* **2016**, *71*, 44–54.