

A framework for the iterative dynamic optimisation of diesel engines: numerical methods, experimental setup, and first results

J. Asprion¹, G. Mancini², S. Zentner¹, Ch. H. Onder¹,
N. Cavina² & L. Guzzella¹

¹*Institute for Dynamic Systems and Control, ETH Zurich, Switzerland*

²*Department of Industrial Engineering, University of Bologna, Italy*

Abstract

Owing to the many degrees of freedom provided by current engine systems and to the need to consider transient operation, deriving an engine calibration has become a tedious task. Control trajectories resulting from the solution of an optimal-control problem provide a guideline to the calibration engineer, serve as a benchmark, and might be used for a partial automation of the calibration procedure. The common approach to dynamic optimisation consists of solving a single optimal-control problem. However, this approach requires the availability of models that are valid throughout the whole engine operating range and actuator ranges. In addition, the result of the optimisation is meaningful only if the model is very accurate. We propose a methodology to circumvent those demanding requirements: an iteration between transient measurements to refine a purpose-built model and a dynamic optimisation which is constrained to the model-validity region. This paper presents all numerical methods required to implement this procedure. The crucial steps are analysed in detail, and the most important caveats are indicated. Finally, an experimental validation demonstrates the applicability of the method and reveals the components that require further development.

Keywords: optimal control, engine optimisation, transient operation, mean-value model, turbocharger, EGR, VGT.

1 Introduction

Optimal control of Diesel engines becomes increasingly important. More stringent emission regulations [1, 2] require the exploitation of the remaining potential to



further reduce the pollutant emissions, especially during transient operation [3,4]. Simultaneously, the fuel consumption has to be minimised for economical and environmental reasons. Nowadays, deriving an engine calibration is a tedious and iterative task, owing to the many degrees of freedom provided by current Diesel engines and to the requirement of considering transient operation. Optimal control is a tool that can support the calibration engineer. The optimal control trajectories indicate directions for the calibration, disclose shortcomings of a chosen control structure and serve as a benchmark. If all relevant operating conditions are covered by the transient cycle, it may be even possible to derive the optimal static maps from the optimal control and state trajectories.

The common approach to dynamic optimisation is to use a large amount of stationary or transient measurement data to identify a globally valid model. Relying on this model, the optimal-control problem (OCP) is solved at once. This approach assumes that all influences on the relevant outputs are captured by the model and the selected set of control inputs. We propose to iterate between the dynamic optimisation, restricted to the current model-validity range, and the refinement of the model using specific transient measurements. This approach avoids measurements in irrelevant regions. Furthermore, refining the model using transient measurements on the driving cycle at hand allows to account for the effect of all influences that are missing in the model. This compensation mechanism allows for the use of a simple mean-value model for the air path as well as empirical setpoint-relative emission models. It is thus not required to develop sophisticated combustion models that have to provide reliable far-field extrapolation.

The drawback of the method is that multiple runs of the transient cycle have to be performed during each iteration. The procedure is thus best suited for compact test cycles containing all relevant transient patterns in condensed form. Alternatively, some recurring transient patterns such as load steps or accelerations can be optimised separately to gain insight into the structure of the optimal trajectories of the control inputs.

Similar approaches were already used in the early days of automated engine calibration [5,6]. However, due to limitations in computer hardware and software, a predefined structure of the control system had to be assumed. Furthermore, the transient measurements were used to directly construct gradient information for the optimisation, since the vast amount of model evaluations required to numerically solve a nonlinear OCP would have been prohibitive at that time. Nowadays, such computing power is readily available. Thus, we use a physics-based model for the air-path to capture the system dynamics, and embed time-variable models for the pollutant emissions and for the torque generation. The accuracy of the air-path model is increased by applying a “generalised Kalman filter”. All OCPs occurring within the methodology are solved numerically by transcribing the continuous-time problem into a finite-dimensional nonlinear program (NLP).

In this paper, we describe and analyse all models and numerical methods required to implement the iterative procedure. Furthermore, we present the



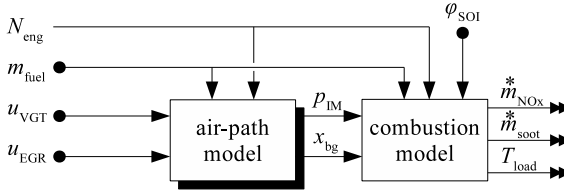


Figure 1: Structure of the engine model.

experimental setup and the testbench-software interfaces. Finally, one iteration of the procedure is executed on an engine testbench and the results are analysed and discussed.

2 Methods

We first describe the engine model. In the subsequent sections, the OCP is stated and its numerical solution is presented, the transient model refinement is described, and the implementation on the engine testbench is outlined. Finally, the iterative procedure combining all these methods is stated.

2.1 Engine model

The model can be subdivided into a dynamic and a static part as illustrated in Fig. 1. All dynamics are induced by the air path, i.e. the turbocharger inertia and the volumes. The full model is represented in state-space form as

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)), \quad (1a)$$

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t)), \quad (1b)$$

where the $n_x = 6$ state variables, the $n_u = 4$ control inputs, and the outputs are

$$\mathbf{x} = (p_{IM}, \vartheta_{IM}, x_{bg}, p_{EM}, p_{aTB}, \omega_{TC})^T, \\ \mathbf{u} = (u_{VGT}, u_{EGR}, \varphi_{SOI}, m_{fcc})^T, \quad \mathbf{y} = (m_{NOx}^*, m_{soot}^*, T_{load})^T.$$

The indices denote the intake and exhaust manifolds (IM/EM), the volume between the turbine and the exhaust-gas aftertreatment system (aTB), and the turbocharger (TC). The positions of the variable-geometry turbine (VGT) and the exhaust-gas recirculation valve (EGR) are pulse-width modulation signals with a range of $[0, 1]$. The crank angle of the start of the injection (SOI) is specified in degrees before top dead center. Finally, the burnt-gas fraction in the intake manifold is denoted by x_{bg} . The outputs of the air-path model, i.e. the dynamic inputs to the combustion model, are denoted by $\mathbf{v} := (p_{IM}, x_{bg})^T$, with

$$\mathbf{v}(t) = \mathbf{g}_{AP}(\mathbf{x}(t), \mathbf{u}(t)) = \mathbf{C}_{AP} \cdot \mathbf{x}(t) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \cdot \mathbf{x}(t). \quad (2)$$

2.1.1 Air-path model

The mean-value model for the air path is adopted from [7]. The thermal models for the walls of the exhaust and intake manifolds are omitted. Furthermore, the submodels for the EGR are added, and an additional restriction representing the exhaust-gas aftertreatment system (ATS) is included. Some other submodels are slightly adapted to the engine at hand and to the intended use. The new model parts are briefly described next.

The EGR valve is modelled as a modified form of the simplified isenthalpic orifice [1, Sec. 2.3.2] with a variable effective cross-sectional area,

$$\dot{m}_{\text{EGR}}^* = (k_0 + k_1 \cdot u_{\text{EGR}}) \cdot p_{\text{EM}} \cdot \sqrt{\frac{p_{\text{IM}}}{p_{\text{EM}}} \cdot \left(1 - \frac{p_{\text{IM}}}{p_{\text{EM}}}\right)}. \quad (3)$$

For the temperature difference across the EGR cooler (EGC), a simple stationary energy balance is evaluated. The average of the inflow and outflow temperatures is used as the representative temperature, which yields

$$k_{\text{EGC}} \cdot \left(\frac{\vartheta_{\text{EM}} + \vartheta_{\text{EGR}}}{2} - \vartheta_{\text{clf}} \right) = \dot{m}_{\text{EGR}}^* \cdot (\vartheta_{\text{EM}} - \vartheta_{\text{EGR}}). \quad (4a)$$

The temperature of the cooling-fluid is assumed to be constant, i.e. $\vartheta_{\text{clf}} = 360\text{K}$, and the efficiency parameter k_{EGC} depends on the mass flow,

$$k_{\text{EGC}} = k_{\text{EGC},1} \cdot \dot{m}_{\text{EGR}}^* + k_{\text{EGC},2} \cdot \dot{m}_{\text{EGR}}^{*2}. \quad (4b)$$

Differential equations The change of the state variables, i.e. the equations defining (1a), represent balances for mass and energy. In the intake receiver, the mass balances for the air and the burnt gas as well as the energy balance is resolved for the state variables pressure, temperature and burnt-gas fraction $x_{\text{bg}} := \frac{m_{\text{bg}}}{m_{\text{air}} + m_{\text{bg}}}$.

$$f_1 = \frac{d}{dt} p_{\text{IM}} = \frac{R}{V_{\text{IM}}} \kappa \left(\dot{m}_{\text{CP}}^* \vartheta_{\text{IC}} + \dot{m}_{\text{EGR}}^* \vartheta_{\text{EGC}} - \dot{m}_{\text{cyl}}^* \vartheta_{\text{IM}} \right), \quad (5a)$$

$$f_2 = \frac{d}{dt} \vartheta_{\text{IM}} = \frac{R \vartheta_{\text{IM}}}{p_{\text{IM}} V_{\text{IM}} c_v} \left[c_p \left(\dot{m}_{\text{CP}}^* \vartheta_{\text{IC}} + \dot{m}_{\text{EGR}}^* \vartheta_{\text{EGC}} - \dot{m}_{\text{cyl}}^* \vartheta_{\text{IM}} \right) - c_v \vartheta_{\text{IM}} \left(\dot{m}_{\text{CP}}^* + \dot{m}_{\text{EGR}}^* - \dot{m}_{\text{cyl}}^* \right) \right], \quad (5b)$$

$$f_3 = \frac{d}{dt} x_{\text{bg}} = \frac{R \vartheta_{\text{IM}}}{p_{\text{IM}} V_{\text{IM}}} \left(\dot{m}_{\text{EGR}}^* x_{\text{bg,EM}} - (\dot{m}_{\text{EGR}}^* + \dot{m}_{\text{CP}}^*) \cdot x_{\text{bg}} \right). \quad (5c)$$

The indices CP and IC denote quantities corresponding to the compressor and the charge-air intercooler, respectively. The burnt-gas fraction in the exhaust gas is

$$x_{\text{bg,EM}} = \frac{x_{\text{bg}}^* \dot{m}_{\text{cyl}}^* + (1 + \sigma_0) \cdot \dot{m}_{\text{fuel}}^*}{\dot{m}_{\text{cyl}}^* + \dot{m}_{\text{fuel}}^*}.$$



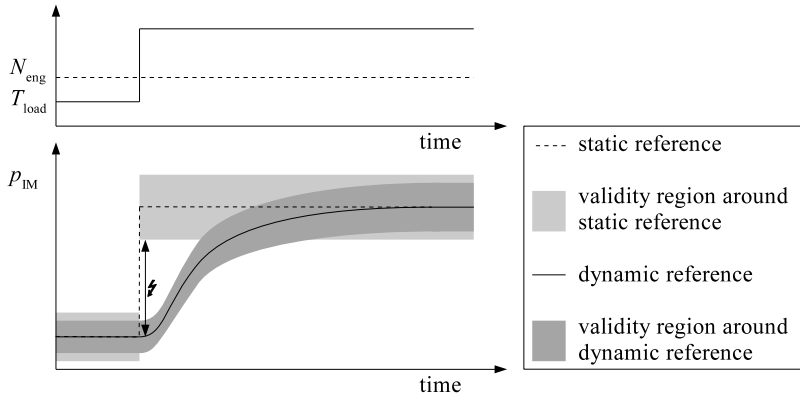


Figure 2: Exemplary static and dynamic references over a load step, and the validity regions of the corresponding setpoint-relative models.

The mass balances in the exhaust manifold and in the volume between turbine and ATS as well as the turbocharger momentum balance are provided in [7].

2.1.2 Combustion model

Time-variable quadratic setpoint-deviation models are used for the emissions and for the combustion efficiency. The cross-terms of the full second-order Taylor expansion are omitted since also the $n_u(n_u - 1)/2$ cross variations would need to be performed to reliably identify the corresponding model coefficients. The model coefficients are identified around the current reference trajectory during each iteration. The vector $\mathbf{w} = (p_{IM}, x_{bg}, \varphi_{SOI}, m_{fcc})^T$ denotes all inputs to the combustion model. The model for each output thus reads

$$y_i(t) = y_{ref,i}(t) + \mathbf{k}_{lin}(t)^T \cdot \Delta \mathbf{w}(t) + \Delta \mathbf{w}(t)^T \cdot \mathbf{K}_{quad}(t) \cdot \Delta \mathbf{w}(t), \quad (6a)$$

with

$$\mathbf{K}_{quad}(t) = \text{diag}(k_{quad,1}(t), \dots, k_{quad,4}(t)), \quad (6b)$$

$$\Delta \mathbf{w}(t) = \mathbf{w}(t) - \mathbf{w}_{ref}(t). \quad (6c)$$

Such stepoint-relative models are often used for control and optimisation applications [8, 9]. However, usually the references for the inputs \mathbf{w} are stored as a static lookup map over engine speed and load. During transient operation such as a load step, the actual values of the dynamic inputs (p_{IM} and x_{bg} in the case at hand) can be far from the steady-state reference values. By using time-resolved reference values and correction factors, the validity range of the model is relocated to the actually relevant region, which is illustrated in Fig. 2.

2.2 Numerical optimal control

The Diesel-engine problem can be cast in the form of the following OCP.

$$\min_{\mathbf{u}(t), \mathbf{x}(t)} \left\{ m_{\text{fuel}} = \int_0^{t_f} \dot{m}_{\text{fuel}}^*(\mathbf{x}(t), \mathbf{u}(t), \tilde{n}_{\text{eng}}(t)) dt \right\} \quad (7a)$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t) - \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \tilde{n}_{\text{eng}}(t)) = \mathbf{0}, \quad (7b)$$

$$\int_0^{t_f} \dot{m}_{\text{em}}^*(\mathbf{x}(t), \mathbf{u}(t), \tilde{n}_{\text{eng}}(t)) dt - \hat{m}_{\text{em}} \leq \mathbf{0}, \quad (7c)$$

$$\tilde{T}(t) - T_{\text{load}}(\mathbf{x}(t), \mathbf{u}(t), \tilde{n}_{\text{eng}}(t)) \leq 0, \quad (7d)$$

$$\underline{\mathbf{u}}(t) \leq \mathbf{u}(t) \leq \overline{\mathbf{u}}(t), \quad \underline{\mathbf{x}}(t) \leq \mathbf{x}(t) \leq \overline{\mathbf{x}}(t). \quad (7e)$$

The instantaneous fuel consumption $\dot{m}_{\text{fuel}}^* = m_{\text{fcc}} \cdot \frac{\tilde{n}_{\text{eng}}}{120} \cdot N_{\text{cyl}}$ is integrated to yield the cumulative fuel consumption m_{fuel} . The dynamic constraints (7b) enforce the model equations and (7c) limit the cumulative emissions. The engine speed prescribed by the driving cycle, $\tilde{n}_{\text{eng}}(t)$, is considered a time-varying parameter. The desired load torque $\tilde{T}(t)$ is imposed as a time-local inequality constraint (7d).

2.2.1 Direct transcription

The continuous-time OCP (7) is transformed into a finite-dimensional mathematical program by direct transcription. An integration scheme is applied to approximate all continuous signals. Due to the stiffness of the system, the method of choice is the family of Radau collocation schemes [10]. These stiffly accurate schemes provide stiff decay and algebraic stability [11, Sec.3.5]. Their implicit nature is irrelevant when used in the context of direct transcription. The first-order representative, the well-known Euler-backward discretisation, is used for the solution of the OCPs in this study. It is found to be very robust and it provides the solution on a uniform discretisation grid, which can be consistently implemented on the testbench.

The set of continuous ordinary differential equations (7b) is transformed to

$$\mathbf{x}_{k+1} = \mathbf{x}_k + h \cdot \mathbf{f}(\mathbf{u}_{k+1}, \mathbf{x}_{k+1}, \tilde{n}_{\text{eng},k+1}), \quad (8)$$

adopting the notation $\mathbf{x}_k := \mathbf{x}(t_k)$. A uniform step size of $h = 0.1s$ is used, yielding a grid of N points $0 = t_0 < t_1 < \dots < t_{N-1} = t_f$. Accordingly, all integrals are approximated as a sum,

$$\int_0^{t_f} \dot{m}_{\bullet}^*(\mathbf{u}(t), \mathbf{x}(t), \tilde{n}_{\text{eng}}(t)) dt \approx h \cdot \sum_{k=0}^{N-2} \dot{m}_{\bullet}^*(\mathbf{u}_{k+1}, \mathbf{x}_{k+1}, \tilde{n}_{\text{eng},k+1}). \quad (9)$$

The bounds on the controls and on the state variables are imposed at the grid points.

This discretisation of the problem yields a sparse NLP. Sparsity signifies that only few entries in the Jacobian matrix, which contains the first partial derivatives

of the constraints, are non-zero. In fact, (8) reveals that only each two neighboring grid points are related. Furthermore, the control inputs and the state variables appear in nonlinear form only in $\mathbf{f}(\mathbf{u}_{k+1}, \mathbf{x}_{k+1}, \tilde{n}_{\text{eng},k+1})$ and in the output function \mathbf{g} . Thus, only the partial derivatives of the model functions \mathbf{f} and \mathbf{g} w.r.t. the control inputs and the state variables at each discretisation point have to be calculated to construct the Jacobian of the NLP. Assembling the first-derivative information of the NLP from the model derivatives corresponds to a perfect exploitation of the problem sparsity [12].

The derivatives of the model equations are calculated by forward finite differences. Therefore, one additional model evaluation is required for each partial derivative. The solver SNOPT 7.2 [13] is used to solve the sparse NLP. It approximates the second partial derivatives by iterative updates using the first-derivative information along the solution steps.

2.2.2 Regularisation

Singular arcs are time intervals during which the Hamiltonian (the combination of the objective and the appropriately weighted constraint violations) becomes affine in the controls. During such intervals, the second derivatives vanish, which results in spurious oscillations when applying direct transcription to solve the OCP. A more detailed analysis of this phenomenon is provided in [14, Sec. 4.5].

In the said thesis, a regularisation based on the “piecewise derivative variation of the control” is proposed as a countermeasure. The regularisation term, for a scalar control input u , is

$$L_{\text{reg}}(u_{\bullet}) := c_N \cdot \text{Var}_{t_N}^2(u_{\bullet}) = \frac{c_{\text{reg}}}{(N-3)(N-1)^2} \cdot \frac{1}{2} \sum_{l=3}^{N-1} |s_{l+1} - s_l|^2, \quad (10)$$

where $s_l = (u_l - u_{l-1}) / (t_l - t_{l-1})$ is the slope of the control in each discretisation interval l . The factor $(N-3)$ in c_N accounts for the number of summation terms, whereas $(N-1)^2$ is an approximation of the average step size. This formulation scales the regularisation term according to the chosen resolution of the approximation, such that the effect of the user-specified parameter c_{reg} is invariant. The regularisation term (10) is summed over all control inputs and added to the discretised form of the objective (7a).

2.3 Transient model refinement

This section describes the methods to refine the dynamic part of the model and to identify the combustion model at each iteration. The interaction between the two model portions is indicated at the appropriate locations in the text.

2.3.1 Refinement of the dynamic air-path model

A generalisation of a Kalman filter is applied to the dynamic part of the model [15]. In order to correct for systematic model errors, the model equations are augmented



by the dynamic and the static corrective variables $\tilde{\mathbf{x}}_f$ and $\tilde{\mathbf{x}}_g$,

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)) + \mathbf{K}_f \cdot \tilde{\mathbf{x}}_f(t), \quad (11a)$$

$$\mathbf{x}_{\text{corr}}(t) = \mathbf{x}(t) + \mathbf{K}_g \cdot \tilde{\mathbf{x}}_g(t). \quad (11b)$$

The matrices \mathbf{K}_f and \mathbf{K}_g define which dynamics are adjusted and for which state variables the absolute values are corrected.

The optimal-control framework described in Sec. 2.2 is used to derive the trajectories of the corrective variables. As objective, the integrated squared error in the corrected state variables is used, yielding a least-squares fit of these trajectories. The original state variables remain in the OCP, but the control inputs become time-varying parameters. Instead, the corrective variables $\tilde{\mathbf{x}}(t) = (\tilde{\mathbf{x}}_f^T(t), \tilde{\mathbf{x}}_g^T(t))^T$ are optimised in the OCP,

$$\min_{\mathbf{x}(t), \tilde{\mathbf{x}}(t)} \int_0^{t_f} \mathbf{b}^T \cdot \mathbf{K}_g^T \cdot \underbrace{(\mathbf{x}(t) + \mathbf{K}_g \cdot \tilde{\mathbf{x}}_g(t))}_{\mathbf{x}_{\text{corr}}(t)} - \hat{\mathbf{x}}(t))^2 dt \quad (12a)$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \hat{\mathbf{u}}(t)) + \mathbf{K}_f \cdot \tilde{\mathbf{x}}_f(t). \quad (12b)$$

Quantities with a hat, e.g. $\hat{\mathbf{x}}$, denote measured signals. The vector \mathbf{b} contains the weights to put more emphasis on the accuracy of some of the corrected state variables. The regularisation term introduced in Sec. 2.2.2 can be used to penalise fast changes of corrective variables. In fact, smooth trajectories are desirable since the model errors are assumed to be of a systematic nature.

If the data from a single measurement is used to identify the corrective variables by solving (12), a perfect match of the corrected state trajectories is obtained by adjusting $\tilde{\mathbf{x}}_g$ only. Therefore, multiple measurements need to be considered simultaneously. In the optimal-control framework, N_m instances of the model are stacked to yield a new system with $N_m \cdot n_x$ state variables. The error in the relevant outputs is cumulated over all measurements.

$$\min_{\mathbf{x}^{(1)}(t), \dots, \mathbf{x}^{(N_m)}(t), \tilde{\mathbf{x}}(t)} \sum_{k=1}^{N_m} \int_0^{t_f} \mathbf{b}^T \cdot \mathbf{K}_g^T \cdot \underbrace{(\mathbf{x}^{(k)}(t) + \mathbf{K}_g \cdot \tilde{\mathbf{x}}_g(t))}_{\mathbf{x}_{\text{corr}}^{(k)}(t)} - \hat{\mathbf{x}}^{(k)}(t))^2 dt \quad (13a)$$

$$\text{s.t.} \quad \begin{pmatrix} \dot{\mathbf{x}}^{(1)}(t) \\ \vdots \\ \dot{\mathbf{x}}^{(N_m)}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{f}(\mathbf{x}^{(1)}(t), \hat{\mathbf{u}}^{(1)}(t)) + \mathbf{K}_f \cdot \tilde{\mathbf{x}}_f(t) \\ \vdots \\ \mathbf{f}(\mathbf{x}^{(N_m)}(t), \hat{\mathbf{u}}^{(N_m)}(t)) + \mathbf{K}_f \cdot \tilde{\mathbf{x}}_f(t) \end{pmatrix}. \quad (13b)$$

2.3.2 Identification of the combustion model

The models for the emissions and the torque generation are re-identified at each iteration. In the final version of the algorithm, all data collected so far should be used during the identification to successively expand the model validity region over the iterations. It is an open point whether the model structure should be

adapted, i.e. its complexity increased according to the data available. Alternatively, the measurements could be weighted by their distance to the current $w(t)$ to yield a tradeoff between the representation of local and far-field trends. For now, the variations around the initial trajectory are considered only, and the quadratic setpoint-relative model described in Sec. 2.1.2 is used.

A slight averaging w.r.t. time is performed to suppress measurement noise and to provide a smooth model, which is advantageous in the context of optimisation. A windowed Gauss curve is used to weight the preceding and consecutive samples for the identification of the model at each time instant,

$$\theta(\Delta t) = \begin{cases} \exp\left(-\frac{\Delta t^2}{\sigma^2}\right), & \text{if } |\Delta t| < \sqrt{-\sigma^2 \cdot \log_e(\theta_{\text{cut}})}, \\ 0, & \text{else.} \end{cases} \quad (14)$$

The parameter θ_{cut} defines the value for θ at which the Gauss curve is cut. Figure 3 shows the curve for the parameter values $\sigma = 0.4s$ and $\theta_{\text{cut}} = 0.5\%$, which are found to be a reasonable choice.

For each output y_i , a weighted linear least-squares regression is used to identify the coefficients at each sampling point t_k . The $N_m - 1$ variations are used, which are again denoted by the superscript index in round brackets. The equation system

$$(\mathbf{X}^T \mathbf{W} \mathbf{X}) \cdot \mathbf{p} = \mathbf{X}^T \mathbf{W} \Delta \mathbf{y}_i, \quad (15a)$$

is solved for \mathbf{p} , where

$$\mathbf{p} = (k_{\text{lin},1}(t_k), \dots, k_{\text{lin},4}(t_k), k_{\text{quad},1}(t_k), \dots, k_{\text{quad},4}(t_k))^T, \quad (15b)$$

$$\Delta \mathbf{y}_i = \begin{pmatrix} \mathbf{y}_i(t_{-l}) - \mathbf{y}_{\text{ref},i}(t_{-l}) \\ \vdots \\ \mathbf{y}_i(t_l) - \mathbf{y}_{\text{ref},i}(t_l) \end{pmatrix}, \quad (15c)$$

$$\mathbf{X} = \begin{bmatrix} \Delta \mathbf{w}_1(t_{-l}) & \cdots & \Delta \mathbf{w}_4(t_{-l}) & \Delta \mathbf{w}_1(t_{-l})^2 & \cdots & \Delta \mathbf{w}_4(t_{-l})^2 \\ \vdots & & \vdots & \vdots & & \vdots \\ \Delta \mathbf{w}_1(t_l) & \cdots & \Delta \mathbf{w}_4(t_l) & \Delta \mathbf{w}_1(t_l)^2 & \cdots & \Delta \mathbf{w}_4(t_l)^2 \end{bmatrix}, \quad (15d)$$

$$\mathbf{W} = \text{diag}(\theta(t_{-l} - t_k) \cdot \mathbf{I}_{N_m-1}, \dots, \theta(t_l - t_k) \cdot \mathbf{I}_{N_m-1}). \quad (15e)$$

Here, l is the number of samples inside the window in both directions, and \mathbf{I}_{N_m-1} is the identity matrix. Each vector $\Delta \mathbf{w}_j := (w_j^{(1)} - w_{\text{ref},j}, \dots, w_j^{(N_m-1)} - w_{\text{ref},j})^T$ stacks all variations (similarly for \mathbf{y}_i). Here, $\Delta \mathbf{w}_j^2$ denotes element-wise squaring.

2.4 Engine and experimental setup

All measurements are performed on the Diesel engine specified in Table 1. The signals required to identify the air-path model are obtained from standard



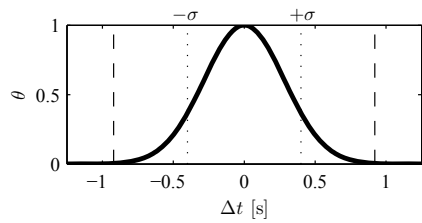


Figure 3: Windowed Gauss function used for the time-averaging of the combustion model. The dashed lines delineate the window.

ECU sensors installed on the engine as well as some additional sensors for the temperature and the pressure at specific locations. A fast measurement of the instantaneous emissions is required. The NO_x emissions are measured by a Combustion fNOx 400 Fast CLD and a UniNOx sensor. The accurate but slow signal from the latter is used to calibrate the fast but drifting signal of the fNOx 400. The soot emissions are recorded by means of an “AVL Micro Soot Sensor”. The dynamic response of this device is identified by performing several steps of the injected fuel mass at different operating points. A constant delay of 0.95s and a first-order element with a time constant of 0.25s are found to closely emulate the sensor dynamics.

The test bench is equipped with a dynamic brake which is controlled by a dSPACE rapid prototyping system. This system is able to prescribe arbitrary profiles for the engine speed and load. The production ECU of the engine can only be accessed for reading operations, thereby no changes or additions to the control structure can be realised. A higher degree of freedom for the engine control is provided by an ES910 rapid prototyping module. A schematic representation of its interaction with the ECU and the Host PC is shown in Fig. 4. CAN and LIN interfaces provide the connection of the ES910 module to the ECU bus and to the external CAN-modules (Input/Output operations). This setup provides complete control over all relevant control inputs of the engine.

INTECRIO is used to set up the rapid-prototyping models on the ES910, while the INCA add-on EIP provides online read and write access to all relevant labels. Any ECU function can be bypassed by means of the ECU-interface ETK. For the application at hand, this functionality is utilised to directly apply time-resolved control trajectories.

The iterative procedure relies on two modes of operating the engine.

Table 1: Main data of the engine used in this work.

<i>engine type:</i>	Diesel, V6	<i>displacement:</i>	2987 cm ³
<i>EGR:</i>	high pressure, cooled	<i>compr. ratio:</i>	15.5
<i>turbocharger:</i>	VGT, charge-air cooler	<i>bore/stroke:</i>	83/92 mm
<i>injection:</i>	CDI 4, max. 1600 bar	<i>valves/cyl.:</i>	4
<i>max. torque</i>	400 Nm (1400–3800 rpm)	<i>rated power:</i>	165 kW (3800 rpm)

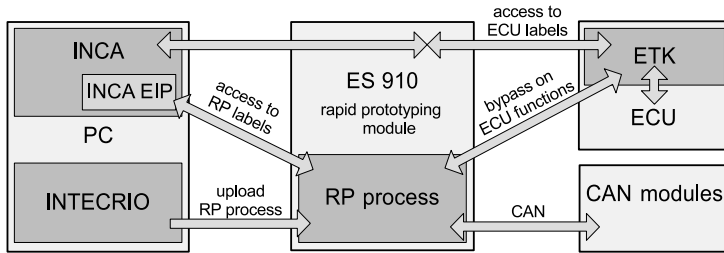


Figure 4: Hardware, software, and interfaces of the testbench setup.

- (A) The ECU with its standard calibration controls the engine. The testbench controller is used to follow the desired profiles of the engine speed (by controlling the brake torque) and the load torque (by controlling the fuelling). This mode is used for the initialisation of the iterative procedure. The resulting trajectories of the controls, including the injected fuel mass, are recorded.
- (B) Time-resolved trajectories are prescribed for all control inputs using the rapid-prototyping module and the bypasses of the ECU. The testbench controller is only used to follow the engine-speed profile. This mode is used for validation runs as well as for all variations.

2.5 Iterative procedure

The inputs to the procedure are a transient driving profile, consisting of engine speed and load torque trajectories, and any calibration of the ECU that is able to operate the engine along this profile.

1. Initialisation: drive the profile in mode A. Save the resulting trajectories of the control inputs, set them as the “current controls” $\mathbf{u}(t)$. Set the corrective variables $\tilde{\mathbf{x}}(t)$ to constant zero.
2. Perform $1 + 2 \cdot n_u$ testbench runs in mode B. Thereby, apply
 - (a) the current controls $\mathbf{u}(t)$, and
 - (b) isolated perturbations of all controls in both directions,
i.e. $u_i(t) \leftarrow u_i(t) \pm \Delta u_i$, for $i = 1, \dots, n_u$.
3. Use the measurement data from step 2 to identify the corrective variables $\tilde{\mathbf{x}}(t)$ by solving (13). Run simulations of the refined air-path model for all variations, and save the resulting corrected state trajectories $\mathbf{x}_{\text{corr}}(t)$.
4. Use the state trajectories of the air-path model from step 3 to identify the torque and emission models around the current references by solving (15).

5. Solve the control and state constrained OCP (7) to derive the improved control trajectories $\mathbf{u}^*(t)$. Thereby, set

$$\underline{\mathbf{u}}(t) = \mathbf{u}(t) - k_u \cdot \Delta \mathbf{u}, \quad \overline{\mathbf{u}}(t) = \mathbf{u}(t) + k_u \cdot \Delta \mathbf{u}, \quad (16a)$$

$$\underline{\mathbf{x}}(t) = k_x \cdot \min\{\mathbf{x}(t)^{(k)} \mid k = 1, \dots, 1 + 2 \cdot n_u\}, \quad (16b)$$

$$\overline{\mathbf{x}}(t) = k_x \cdot \max\{\mathbf{x}(t)^{(k)} \mid k = 1, \dots, 1 + 2 \cdot n_u\}. \quad (16c)$$

6. Set $\mathbf{u}(t) \leftarrow \mathbf{u}^*(t)$, repeat steps 2.-5. until the change in the controls is small.

Using the simulated state trajectories during the identification of the combustion model in step 4 ensures a consistent prediction of the emissions and of the torque inside the validity region. This fact is important since also the optimisation in step 5 is restricted to this region. This “trust region” can be slightly expanded by the factors k_u and k_x .

Since the identification of the models for the air path and for the combustion are identified separately, the physical causality is preserved. More exactly, there is no way that the combustion model corrects errors in the air-path model or vice-versa. Although a combined identification could yield a slightly higher accuracy, it would introduce unphysical cross corrections. Furthermore, the identification procedure would become more complex and non-convex.

3 Results and discussion

A short test cycle is used to evaluate the methods presented in the preceding sections. The engine is operated at a constant speed of 2500 rpm, and the desired load-torque profile is shown in Fig. 5. As variations, an additive offset of $\pm 5\%$ PWM is applied to the VGT and EGR positions, and $\pm 2^\circ$ to the SOI. A multiplicative factor of 1.05 defines the variation of the fuel mass. These variations are referred to as the *small* variations in the remainder of the text. To assess the accuracy of interpolation and extrapolation, *large* variations with offsets of $\pm 8\%$ for the VGT, $\pm 10\%$ for the EGR, $\pm 4^\circ$ for the SOI and a factor of 1.1 for the fuel mass are performed. Since the VGT and the EGR both dynamically affect the burnt-gas fraction and the pressure in the intake manifold, the four *cross variations* of these two actuators are additionally recorded.

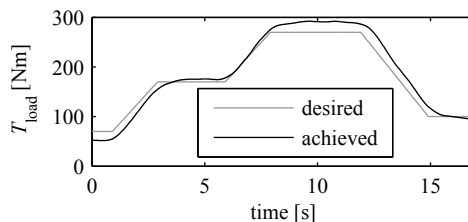


Figure 5: The desired load-torque profile and the trajectory achieved by the testbench controller.

For the identification of the combustion model, it would be desirable to have a constant offset in v . However, the EGR-VGT controller cannot perfectly follow reference trajectories for the burnt-gas fraction and the boost pressure. After initial tests, it has been found that applying constant offsets directly to the two dynamic controls is the best choice.

3.1 Transient air-path model refinement

The turbocharger speed and the pressure in the exhaust manifold represent the relevant dynamics in a turbocharged engine system with EGR. The boost pressure closely follows the turbocharger speed and the EGR mass-flow is defined, aside from the position of the EGR valve, by the pressure ratio between exhaust and intake manifolds. Therefore, it is sufficient to correct these two dynamics.

Since for the combustion model only the *relative* accuracy is of interest, it is not necessary to require $v(t)$ to be accurate in terms of absolute values. Rather, the state variables that are critical for a safe engine operation should be matched to the measured trajectories to enable an accurate limitation of these quantities in the OCP. Since no limits on any state variables are included in the OCP as of yet, we choose the pressures in the exhaust and intake manifolds to illustrate the methodology. The choice of the dynamics and the state variables to be corrected is represented by

$$\mathbf{K}_f^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{K}_g^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}. \quad (17)$$

The errors in the two pressures are equally weighted, i.e. $\mathbf{b}^T = (1 \ 1)$.

The top plot in Fig. 6 shows the measured trajectories of the intake pressure along with the model output before and after the refinement. The refinement is performed using only the small variations. The resulting good match for the large variations indicates that the model errors in fact are systematic. For example, the too fast speedup of the turbocharger predicted by the model might be caused by the omission of the thermal models. The heat losses to the manifold walls are neglected and consequently, the enthalpy available to the turbine is overestimated.

The bottom plot shows the difference of the variations to the reference trajectory. Obviously, the refinement has no influence on the predicted difference. Therefore, the refinement of the air-path model is not critical for the combustion model but rather a tool that allows for an accurate limitation of any state variable, e.g. maximum turbocharger speed, exhaust-manifold pressure and temperature, etc.

In the context of the generalised Kalman filter, the regularisation described in Sec. 2.2.2 may be used to enforce smooth trajectories of the corrective variables \tilde{x} . The demand for smooth trajectories is justified by assuming the model errors to be of systematic nature and thus not to exhibit a stochastic or arbitrarily fast changing behaviour. For the results presented here, a value of $c_{\text{reg}} = 50$ is used.



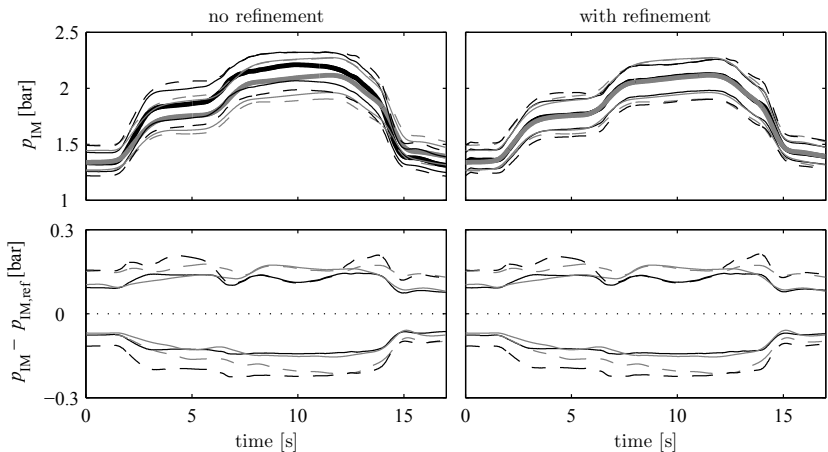


Figure 6: Effects of the dynamic model refinement, VGT variations. Measurement data (grey) versus model outputs (black). Line styles: reference (bold), small variations (solid), large variations (dashed).

3.2 Static combustion model

The results presented here are all derived using the time averaging introduced in Sec. 2.3.2. Interpolation and extrapolation is not influenced by the time averaging. Table 2 shows the errors of the combustion model when identified using three different data sets, namely the small variations, the small and the cross variations, and the large variations. The figures indicate that the model identified using the small variations is able to accurately predict the cross variations, except for the soot emissions. Furthermore, interpolation is significantly more reliable than extrapolation. This fact encourages the use of rather large variations and small factors k_{u} and k_{x} instead of the reliance on small variations and extrapolation.

3.3 Optimal control

For the OCP, the regularisation was set to $c_{\text{reg}} = 100$. This choice successfully suppresses oscillating solutions while not affecting the parts that exhibit smooth

Table 2: Static combustion model: average magnitude of the relative error in % for the instantaneous NO_x and soot emissions, and the torque T_{load} .

ident. data:	small variations			small & cross vars.			large variations		
	NO_x	soot	T_{load}	NO_x	soot	T_{load}	NO_x	soot	T_{load}
small vars.:	0.14	0.89	0.09	0.70	4.45	0.25	1.14	7.70	0.75
cross vars.:	2.85	15.84	0.82	1.60	12.41	0.27	2.60	18.92	0.73
large vars.:	3.31	20.56	2.36	3.46	21.17	2.11	0.19	1.10	0.09



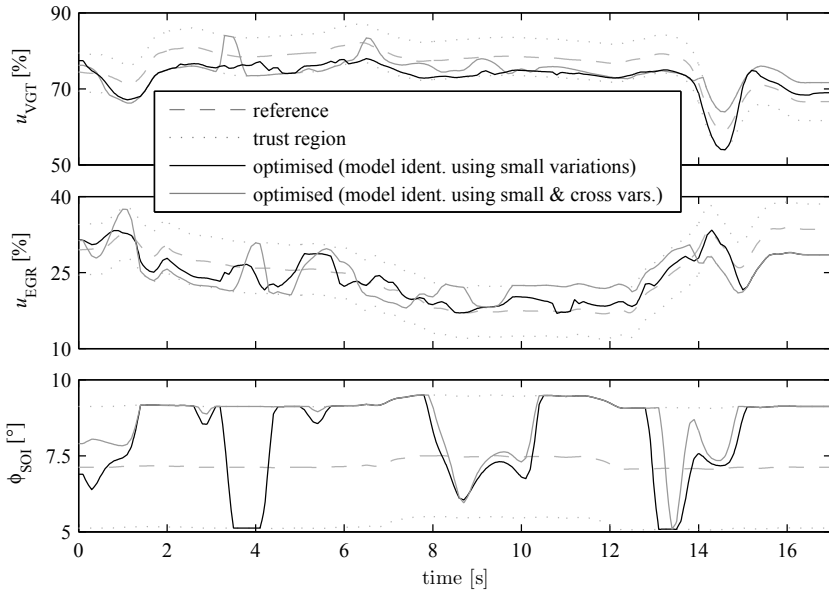


Figure 7: Controls resulting from the solution of the constrained OCP.

trajectories anyways. Especially for the SOI, which has no influence on the air-path dynamics, fast oscillations result when no regularisation is applied. The time averaging during the identification of the combustion model has a similar effect as the regularisation. It is left to be analysed in more detail which approach provides the most plausible solutions and the fastest convergence of the iterative procedure.

Figure 7 shows the optimal control trajectories obtained when using the model identified by the small variations only, or by the small and the cross variations. In both cases, no extrapolation is performed, i.e. $k_{\text{u}} = 1$ and $k_{\text{x}} = 1$. The limits for the cumulative emissions, \hat{m}_{em} , are chosen such that the brake-specific emissions of the reference are maintained. The torque profile resulting from the reference control trajectories is imposed by (7d).

The optimal control trajectories are experimentally validated on the engine testbench. To account for possible deviations in the resulting torque, the emissions and the fuel consumption are related to the integrated engine power. Figure 8 summarises the measured cumulative emissions and the fuel consumption.

Five runs are recorded for each set of trajectories. The average as well as the minimum and maximum values are shown in the plots. The emissions remain within the measurement uncertainty, while the fuel consumption is reproducibly reduced. The soot measurement exhibits a large variability, which is also present in the identification data. Therefore, the corresponding model is not reliable and possibly hinders a more effective reduction of the fuel consumption. Furthermore, due to this inconsistency, the model quality seems not to improve when including the cross variations in the identification data.

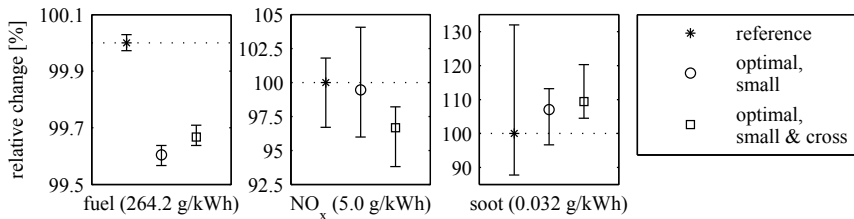


Figure 8: Experimental results of the first iteration. The error bars indicate the range (minimal to maximal values) of the five measurements.

4 Conclusion

We present the numerical methods and the testbench setup required to perform an iterative dynamic optimisation of Diesel engines over prescribed driving profiles. One exemplary iteration of the methodology is performed and experimentally validated. Significant progress towards lower fuel consumption while maintaining the emission levels is observed.

Further development of the methodology will focus on the replacement of the fast but manually calibrated, failure-prone and high-maintenance emission-measurement devices by standard instrumentation, and on a fully automated implementation of the iterative procedure. The availability of a fully automatic and reliable setup allows to perform parametric studies, such as which set of variations to perform during each iteration and how the models can make use of data recorded during previous iterations. Furthermore, following the desired load-torque profile more closely should be included as an additional objective of the iterative procedure.

References

- [1] Guzzella, L. & Onder, C.H., *Introduction to Modeling and Control of Internal Combustion Engine Systems*. Springer Berlin Heidelberg, 2nd edition, 2010.
- [2] Rakopoulos, C. & Giakoumis, E., *Diesel Engine Transient Operation*. Springer London, 2009.
- [3] Lindgren, M. & Hansson, P.A., Effects of transient conditions on exhaust emissions from two non-road diesel engines. *Biosyst Eng*, **87**, pp. 57–66, 2004.
- [4] Hagen, J.R., Filipi, Z.S. & Assanis, D.N., Transient diesel emissions: Analysis of engine operation during a tip-in. *SAE technical paper*, **2006-01-1151**, 2006.
- [5] Cassidy, J.F., A computerized on-line approach to calculating optimum engine calibrations. *SAE technical paper*, **770078**, 1977.
- [6] Dohner, A.R., Optimal control solution of the automotive emission-constrained minimum fuel problem. *Automatica*, **17**, pp. 441–458, 1981.



- [7] Asprion, J., Chinellato, O. & Guzzella, L., Optimisation-oriented modelling of the NO_x emissions of a Diesel engine. *Energ Convers Manage*, **75**, pp. 61–73, 2013.
- [8] Brahma, I., Sharp, M.C. & Frazier, T.R., Empirical modeling of transient emissions and transient response for transient optimization. *SAE technical paper*, **2009-01-1508**, 2009.
- [9] Benz, M., Hehn, M., Onder, C. & Guzzella, L., Model-based actuator trajectories optimization for a diesel engine using a direct method. *J Eng Gas Turbines Power*, **133(3)**, pp. 032806–1–11, 2010.
- [10] Hairer, E. & Wanner, G., Stiff differential equations solved by Radau methods. *J Comput Appl Math*, **111(1–2)**, pp. 93–111, 1999.
- [11] Butcher, J., *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons, Winchester, 2003.
- [12] Patterson, M.A. & Rao, A.V., Exploiting sparsity in direct collocation pseudospectral methods for solving optimal control problems. *J Spacecraft Rockets*, **49**, pp. 364–77, 2012.
- [13] Gill, P.E., Murray, W. & Saunders, M.A., SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Rev*, **47**, pp. 99–131, 2005.
- [14] Schwartz, A.L., *Theory and Implementation of Numerical Methods Based on Runge-Kutta Integration for Solving Optimal Control Problems*. Ph.D. thesis, University of California, Berkeley, 1996.
- [15] Houska, B., Logist, F., Diehl, M. & Impe, J.V., A tutorial on numerical methods for state and parameter estimation in nonlinear dynamic systems. *Identification for Automotive Systems*, eds. D. Alberer, H. Hjalmarsson & L. del Re, Springer London, 2012.

