POLITECNICO DI TORINO
Repository ISTITUZIONALE

Uncertainty Quantification and Computational Cost Reduction of a Chemical Reactor Network for a Hydrogen Combustor

Original
Uncertainty Quantification and Computational Cost Reduction of a Chemical Reactor Network for a Hydrogen Combustor / Madonia, V., Folcarelli, L., Ferrero, A., Masseni, F., Pastrone, D.. - ELETTRONICO. - (2026), pp. 1-41. (AIAA AVIATION 2026 Forum San Diego, CA (USA) 8-12 June 2026) [10.2514/6.2026-4632].

Availability:
This version is available at: 11583/3016390 since: 2026-10-07T09:20:33Z

Publisher:
American Institute of Aeronautics and Astronautics

Published
DOI:10.2514/6.2026-4632

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
AIAA preprint/submitted version e/o postprint/Author's Accepted Manuscript

(Article begins on next page)

19 November 2026

Uncertainty Quantification and Computational Cost Reduction of a Chemical Reactor Network for a Hydrogen Combustor

Vincenzo Madonia *, Lorenzo Folcarelli †, Andrea Ferrero‡, Filippo Masseni § and Dario Pastrone¶
Politecnico di Torino, Torino, Italy, 10129

This work presents an integrated framework to improve the predictive capability and computational efficiency of Chemical Reactor Network models for lean premixed swirl-stabilized hydrogen combustion, using the AHEAD burner as a reference test case. Six optimization algorithms benchmarked over twenty seeds under a common budget of about 2000 evaluations identify the Covariance Matrix Adaptation Evolution Strategy as the most reliable calibration method, with a 95% success rate. Skeletal mechanism reduction via the Directed Relation Graph with Error Propagation yields compact mechanisms with standalone speedups up to 2.06 and maximum nitrogen oxide errors below 0.71%. A Flow-Map In Situ Adaptive Tabulation scheme delivers per-equivalence-ratio speedups up to 24.88 with mean errors below $6 \cdot 10^{-4}$ %. Embedded in the calibration loop, the combined acceleration reduces the median wall-clock time of a twenty-seed campaign from about 70 to less than 5 hours, without significant loss in calibration quality. A Polynomial Chaos Expansion uncertainty analysis attributes about 87% of the joint variance to operating mass-flow perturbations. The resulting 95% uncertainty band brackets six of the eight experimental measurements, with the residual gap concentrated at the leanest points, where localized hot spots and ignition–reignition events not captured by the homogeneous network likely contribute to nitrogen oxide formation.

Nomenclature

Roman symbols

$\mathcal{A}_i$	set of multi-indices associated with basis functions depending only on ξ_i, [–]
$\mathcal{A}_i^T$	set of multi-indices for which ξ_i appears, alone or in interaction, [–]
$c_{\vec{\alpha}}$	PCE expansion coefficient, [–]
c_v	specific heat at constant volume, J/(kg·K)
c_1, c_2	PSO cognitive and social acceleration coefficients, [–]
C_A	total consumption rate of species A, kg/(m^3·s)
CR	DE crossover probability, [–]
$\mathrm{CI}_{95}^{\mathrm{lo}}, \mathrm{CI}_{95}^{\mathrm{hi}}$	lower and upper bounds of the 95% uncertainty interval, ppmv
D	dimensionality of the uncertain input vector, [–]
$\mathbb{E}[\cdot]$	expected value operator, [–]
$f_{\vec{\xi}}$	joint probability density function of $\vec{\xi}$, [–]
f_{ξ_j}	marginal probability density function of ξ_j, [–]
f_r	cumulative ISAT retrieve fraction, [–]
F	DE mutation factor, [–]
g	ADAM gradient vector, [–]
h	finite-difference step size, [–]
h_{forcing}	deterministic hash of forcing parameters, [–]
h_{in}	specific enthalpy at the reactor inlet, J/kg
H	Kruskal–Wallis test statistic, [–]
H_{kj}	regression matrix entry for PCE collocation, [–]

*PhD student, Department of Mechanical and Aerospace Engineering, Corso Duca degli Abruzzi 24, Torino, Italy, AIAA Member
†PhD candidate, Department of Mechanical and Aerospace Engineering, Corso Duca degli Abruzzi 24, Torino, Italy, AIAA Member
‡Associate Professor, Department of Mechanical and Aerospace Engineering, Corso Duca degli Abruzzi 24, Torino, Italy
§Assistant Professor, Department of Mechanical and Aerospace Engineering, Corso Duca degli Abruzzi 24, Torino, Italy, AIAA Member
¶Full Professor, Department of Mechanical and Aerospace Engineering, Corso Duca degli Abruzzi 24, Torino, Italy, AIAA Associate Fellow

J	objective (cost) function, [–]
K_1	number of retained calibration parameters after correlation analysis, [–]
K_2	number of retained calibration parameters after screening, [–]
L	ellipsoid-of-accuracy matrix in scaled state space, [–]
L_{AX}	axial length of the Z_{AX} reactor, mm
m	mass within the reactor, kg
$\dot{m}$	mass flow rate, kg/s
$\dot{m}_{k,\mathrm{gen}}$	generation rate of species k due to chemical reactions, kg/s
M_{tab}	ISAT table memory footprint, MB
$\mathcal{M}$	model mapping from uncertain input space to output space, [–]
n	dimension of the ISAT state vector, [–]
n_{cal}	number of calibration points (Eq. 5)
n_r	number of reactions in the mechanism, [–]
N	regression sample size per equivalence-ratio set point, [–]
N_{ext}	external validation sample size, [–]
$N_{\mathrm{scr}}, N_{\mathrm{refin}}$	screening and refinement sample sizes for Block C, [–]
N_ℓ	total number of ISAT leaves, [–]
N_ϕ	number of equivalence-ratio set points, [–]
N_s	number of chemical species in the mechanism, [–]
$N_{\mathrm{sp}}, N_{\mathrm{rx}}$	number of species and reactions of a skeletal mechanism, [–]
NO_x	nitrogen oxides
p	pressure, Pa
$p_{\mathrm{S}}, p_{\mathrm{P}}$	p-values for Spearman and Pearson correlation tests, [–]
p_{corr}	Bonferroni-corrected p-value, [–]
p_o	PCE truncation order, [–]
P	number of retained PCE basis functions, [–]
P_A	total production rate of species A, kg/(m^3·s)
q	retention threshold for variance-based screening, [–]
Q	integrated heat release, J
$Q_{\mathrm{skel}}/Q_{\mathrm{det}}$	skeletal-to-detailed heat-release ratio, [–]
Q^2_{LOO}	Stone–Geisser leave-one-out predictive coefficient, [–]
r_0	initial radius of the ISAT ellipsoid of accuracy, [–]
r_{AB}	DRGEP direct interaction coefficient between species A and B, [–]
r_{P}	Pearson product-moment correlation coefficient, [–]
R^2_{test}	coefficient of determination on the external test set, [–]
R_B	DRGEP overall interaction coefficient of species B, [–]
$\vec{R}$	classical ISAT reaction mapping, [–]
$\mathrm{RMSE}_{\mathrm{ext}}$	root-mean-square error on the external validation set, ppmv
$\vec{s}$	ISAT scaling vector, [–]
S	ISAT speedup, [–]
S_i	first-order Sobol index of input ξ_i, [–]
S_i^T	total-effect Sobol index of input ξ_i, [–]
$\overline{S}_k^T$	ϕ-averaged total-effect Sobol score for parameter k, [–]
$\langle S_T \rangle_\phi$	aggregated (mean over ϕ) total-effect Sobol score, [–]
S_{main}	first-order (main) Sobol index, [–]
S_n	swirl number at the mixing tube exit, [–]
S_L	laminar flame speed, m/s
t	time, s
T	temperature, K
u_k	specific internal energy of species k, J/kg
$\mathrm{Var}[\cdot]$	variance operator, [–]
V	reactor volume, m^3
w	PSO inertia weight, [–]

w_i	magnitude-dependent weight in the objective function, [–]
$\vec{X}$	concatenated thermochemical state vector of the CRN, [–]
$\tilde{\vec{X}}$	scaled thermochemical state vector, [–]
$X_{i,\mathrm{ref}}$	reference value for ISAT scaling of state component i, [–]
$X_{\mathrm{OH}}, X_{\mathrm{NO}}, X_{\mathrm{NO_2}}$	mole fractions of OH, NO, NO_2, [–]
Y_k	mass fraction of species k, [–]
$\vec{Y}_0$	ISAT mapped (output) state, [–]
$\vec{Y}_{\mathrm{exact}}$	direct-integration mapping used as ISAT reference, [–]
Y	scalar model response, ppmv
$\widetilde{Y}_p$	truncated PCE surrogate of order p, ppmv
$\overline{Y}$	sample mean of the model response, ppmv

Greek symbols

α	ADAM learning rate, [–]
α_{fw}	family-wise significance level, [–]
$\vec{\alpha}$	PCE multi-index, [–]
$\lvert\vec{\alpha}\rvert$	total order of the multi-index $\vec{\alpha}$, [–]
β_1, β_2	ADAM first- and second-moment exponential decay rates, [–]
γ	swirl deflection angle, rad
δ_a, δ_f	relative deviations of air and fuel mass-flow rates, [–]
$\delta_{\vec{\alpha}\vec{\beta}}$	Kronecker delta for multi-indices, [–]
$\delta_{B,i}$	species participation indicator in reaction i, [–]
$\delta\tilde{\vec{X}}$	displacement in scaled state space, [–]
$\Delta_{S^T}^{\max}$	maximum absolute Sobol drift between PCE orders p and $p-1$, [–]
Δt	time step (classical ISAT), s
ΔT_{OUT}	temperature deviation in the OUT reactor (ISAT vs. DI), K
ε	DRGEP threshold, [–]
$\varepsilon_{0\mathrm{th}}$	ISAT zeroth-order approximation error, [–]
$\varepsilon_{\mathrm{tol}}$	ISAT tolerance, [–]
$\varepsilon_{\mathrm{k}}^{LOO}$	leave-one-out residual for sample k, ppmv
$\overline{\varepsilon}_{\mathrm{NO}_x}$	mean relative NO_x error, %
$\varepsilon_{\mathrm{NO}_x}^{\max}$	maximum pointwise relative NO_x error, %
$\varepsilon_{\tau_{\mathrm{ign}}}$	relative error on ignition delay time, %
ε_{S_L}	relative error on laminar flame speed, %
$\varepsilon_{\tau_{\mathrm{ext}}}$	relative error on PSR extinction residence time, %
$\varepsilon_{X_{\mathrm{peak}}}$	relative error on peak species mole fractions, %
$\varepsilon_{\dot{\omega}_{\mathrm{peak}}}$	relative error on peak net production rates, %
$\theta_1, \ldots, \theta_{10}$	CRN topology calibration parameters, [–]
$\vec{\Theta}$	ten-dimensional CRN parameter vector, [–]
$\vec{\Theta}_0$	nominal (initial) CRN parameter vector, [–]
$\vec{\Theta}^*$	reference calibrated parameter vector, [–]
λ	CMA-ES offspring population size, [–]
μ_{NO_x}	PCE-predicted mean NO_x, ppmv
ν	Matérn kernel smoothness parameter, [–]
$\nu_{A,i}$	net stoichiometric coefficient of species A in reaction i, [–]
ξ_j	j-th uncertain input variable, [–]
$\vec{\xi}$	vector of uncertain input variables, [–]
ρ_{S}	Spearman rank correlation coefficient, [–]
σ	Beér–Chigier swirl-jet spreading coefficient, [–]
σ_0	CMA-ES initial step size, [–]
σ_{NO_x}	PCE-predicted standard deviation of NO_x, ppmv

$\sigma_O, \sigma_C, \sigma_G$	PCE-predicted standard deviations of NO_x for Blocks O, C, G, ppmv
τ	residence time, s
τ_{ign}	ignition delay time, s
τ_{ext}	PSR extinction residence time, s
ϕ	equivalence ratio, [–]
ϕ_1, ϕ_2	equivalence ratios of the two upstream reservoirs, [–]
ϕ_{ref}	equivalence-ratio set point for PCE propagation, [–]
$\vec{\phi}_0$	classical ISAT input composition vector, [–]
$\Phi_{\Delta t_{\text{tab}}}$	discrete flow map of the reactor network, [–]
$\Psi_{\vec{\alpha}}$	multivariate orthonormal polynomial basis function, [–]
ω_i	net rate of progress of reaction i, mol/(m^3·s)
$\dot{\omega}_{\text{peak}}$	peak net production rate, mol/(m^3·s)

Subscripts and superscripts

air	air-related quantity
calib	calibration point
det	detailed parent mechanism
exp	experimental quantity
ext	external validation
gen	chemical generation term
in	reactor inlet
k	species index
ℓ	equivalence-ratio set-point index
LOO	leave-one-out
out	reactor outlet
peak	peak value of a profile
ref	reference value
scr, refin	screening and refinement stages of Block C
sim	simulated quantity
skel	skeletal mechanism
test	test (external validation) set
tol	tolerance value
train	training set
T	total-effect Sobol index
$*$	reference (calibrated) value
0	nominal or tabulation point
(k)	k-th sample of the design of experiments
$(-k)$	surrogate refitted without sample k

Acronyms and abbreviations

ADAM	ADAptive Moment estimation
BO	Bayesian Optimization
CFD	Computational Fluid Dynamics
CMA-ES	Covariance Matrix Adaptation Evolution Strategy
CRECK	Chemical Reaction Engineering and Chemical Kinetics mechanism
CRN	Chemical Reactor Network
DE	Differential Evolution
DI	Direct Integration
DRG	Directed Relation Graph
DRGEP	Directed Relation Graph with Error Propagation
EARSM	Explicit Algebraic Reynolds Stress Model
EOA	Ellipsoid Of Accuracy

FL_{AX}	Axial flame zone reactor
FL_{SW}	Swirling flame zone reactor
GP	Gaussian Process
IQR	Interquartile Range
IRZ	Inner Recirculation Zone
ISAT	In Situ Adaptive Tabulation
L-BFGS-B	Limited-memory BFGS with Bound constraints
LHS	Latin Hypercube Sampling
LOO	Leave-One-Out
MAE	Mean Absolute Error
ODE	Ordinary Differential Equation
OC	Operating Condition
OH-PLIF	Hydroxyl Planar Laser-Induced Fluorescence
ORZ	Outer Recirculation Zone
OUT	Exit reactor
PCE	Polynomial Chaos Expansion
PDF	Probability Density Function
PIV	Particle Image Velocimetry
PFR	Plug Flow Reactor
PSO	Particle Swarm Optimization
PSR	Perfectly Stirred Reactor
PaSR	Partially Stirred Reactor
RMSE	Root Mean Square Error
$\mathrm{Res}_1, \mathrm{Res}_2$	Upstream reservoirs of the CRN
TU Berlin	Technische Universität Berlin (experimental dataset reference)
UQ	Uncertainty Quantification
YAML	YAML Ain't Markup Language (Cantera input format)
Z_{AX}	Axial flow zone reactor

Introduction

Decarbonization policies have promoted hydrogen as a key candidate for net-zero aviation emissions. The Paris Agreement and the European Union's Destination 2050 initiative have established ambitious targets for reducing aviation emissions, with hydrogen combustion emerging as a cornerstone technology for next-generation propulsion systems [1, 2].

Hydrogen combustion presents unique advantages and challenges that fundamentally distinguish it from conventional hydrocarbon fuels. Its exceptionally high specific energy content of approximately 120 MJ/kg, wide flammability range, and absence of carbon-based emissions make it attractive for aviation applications [3, 4]. However, elevated adiabatic flame temperatures create conditions conducive to thermal NO_x formation via the Zeldovich mechanism [5], while hydrogen's high flame speed poses flashback risks in premixed configurations [6, 7].

The prediction of NO_x emissions in hydrogen combustion systems requires capturing the complex interplay between fluid dynamics and detailed chemical kinetics. Scale-resolving Computational Fluid Dynamics (CFD) simulations, while high-fidelity, face prohibitive computational costs when coupled with comprehensive chemical mechanisms comprising hundreds of species and thousands of reactions [8]. The stiffness inherent in these systems, combined with the need to resolve thermal, NNH, and N_2O-intermediate nitrogen chemistry pathways, imposes severe limitations on parametric studies essential for combustor design optimization. These constraints have motivated the development of strategies that preserve essential physics while dramatically reducing computational expense [9, 10].

Chemical Reactor Networks (CRNs) have emerged as a promising framework for balancing predictive accuracy and computational efficiency. The CRN methodology discretizes combustion chambers into networks of interconnected ideal reactors, including perfectly stirred reactors, plug flow reactors, and partially stirred reactors, enabling direct integration of detailed chemical kinetic mechanisms while reducing computational time by three to four orders of magnitude compared to CFD [11–14]. Recent advances include CFD-guided reactor decomposition [11], Bayesian calibration frameworks for alternative fuels [15], and particle swarm optimization for parameter tuning [10, 16, 17].

Despite these advances, critical gaps persist in the robust calibration and practical deployment of Chemical Reactor Networks. While recent methodology successfully minimized reliance on extensive CFD simulations by approximating reactor volumes through flow physics and calibrating mass fluxes via heuristic algorithms [10, 18], the broader application of this framework requires addressing critical computational and statistical challenges. First, limited attention has been devoted to systematically comparing various optimization algorithms to identify the most reliable parameter tuning strategy for this specific approach. Second, computational efficiency remains hindered by the need to solve stiff ordinary differential equations for detailed chemical kinetics at each iteration. In particular, the integration of acceleration techniques, specifically skeletal mechanism reduction and In Situ Adaptive Tabulation (ISAT), remains largely unexplored within coupled-reactor CRN frameworks. Finally, there remains a lack of robust design frameworks that apply Polynomial Chaos Expansion (PCE) to systematically quantify how parametric uncertainties, ranging from boundary conditions to geometric closure assumptions, propagate into emission predictions within these newly developed networks.

The present work addresses these gaps through a comprehensive investigation of CRN optimization for a hydrogen combustor, with emphasis on reducing computational cost, improving temporal efficiency, and preserving prediction accuracy. The methodology integrates three complementary strategies. First, a comparative analysis of optimization algorithms, including gradient-based methods, differential evolution, and hybrid techniques, identifies optimal parameter calibration strategies. Second, computational time reduction is pursued through tabulation strategies and reduced chemical mechanisms derived via the Directed Relation Graph with Error Propagation. Third, comprehensive uncertainty quantification using non-intrusive Polynomial Chaos Expansion characterizes the sensitivity of NO_x predictions to parametric uncertainties [15, 19].

The remainder of this paper is organized as follows. The Methodology section (Sec. I) introduces the CRN formulation in Sec. I.A, the comparison of calibration algorithms in Sec. I.B, the computational acceleration based on tabulation and skeletal mechanisms in Sec. I.C, and the uncertainty quantification framework in Sec. I.D. Sec. II reports the corresponding results. The final section summarizes the main findings.

I. Methodology

This section presents the methodological framework developed to improve the predictive capability, computational efficiency, and robustness of Chemical Reactor Network models for emission prediction in hydrogen combustors. The framework is organized as follows.

First, Sec. I.A describes the Chemical Reactor Network formulation adopted throughout the study, including the reactor topology, governing equations, mass-flow redistribution strategy, and chemical kinetic mechanisms.

Sec. I.B then addresses the systematic comparison of optimization algorithms for CRN parameter calibration, covering the definition of the objective function, the selection and configuration of the six competing algorithms, and the benchmarking protocol used to assess their performance.

Sec. I.C focuses on the reduction of the computational cost associated with repeated CRN evaluations. Two complementary strategies are considered. The first is skeletal mechanism reduction, based on the Directed Relation Graph with Error Propagation methodology, which reduces the size and stiffness of the chemical kinetic mechanism while preserving the relevant NO_x formation pathways (Sec. I.C.1). The second is In Situ Adaptive Tabulation, which accelerates the time integration of the reactor network by reusing previously computed thermochemical mappings whenever the local interpolation error remains within a prescribed tolerance (Sec. I.C.2).

Finally, Sec. I.D introduces the uncertainty quantification framework used to assess the robustness of the calibrated CRN predictions. A non-intrusive Polynomial Chaos Expansion approach is employed to propagate uncertainties associated with operating conditions, calibration parameters, and geometric or closure assumptions. The resulting surrogate models are then used to estimate statistical moments, uncertainty bands, and variance-based sensitivity indices for the predicted NO_x emissions.

A. Chemical Reactor Network

The chemical reactor network employed in this work is built following the methodology described in Ref. [10] and Ref. [18] and is constructed to replicate the NO_x behavior observed experimentally on the AHEAD lean premixed swirl-stabilized hydrogen burner [20–22]. The open-source software package Cantera [23] is used to implement the reactor network and perform the thermochemical calculations. The three-dimensional combustor flow field is approximated by a reduced network of constant-volume ideal-gas reactors. The simulations are performed under the simplifying assumptions of adiabatic reactor walls, chemically inert wall surfaces, and constant reactor volumes during

each integration interval. Under these assumptions, the evolution of each reactor is governed by the mass balance, the species transport equations, and the energy balance, written as:

$$\frac{dm}{dt} = \sum_{\text{in}} \dot{m}_{\text{in}} - \sum_{\text{out}} \dot{m}_{\text{out}} \tag{1}$$

$$m\frac{dY_k}{dt} = \sum_{\text{in}} \dot{m}_{\text{in}}\left(Y_{k,\text{in}} - Y_k\right) + \dot{m}_{k,\text{gen}} \tag{2}$$

$$mc_v\frac{dT}{dt} = \sum_{\text{in}} \dot{m}_{\text{in}}\left(h_{\text{in}} - \sum_k u_k\, Y_{k,\text{in}}\right) - \frac{pV}{m}\sum_{\text{out}} \dot{m}_{\text{out}} - \sum_k \dot{m}_{k,\text{gen}}\, u_k \tag{3}$$

where m is the mass within the reactor, $\dot{m}_{\text{in}}$ and $\dot{m}_{\text{out}}$ represent mass flow rates through inlets and outlets, respectively. Y_k is the mass fraction of species k, $\dot{m}_{k,\text{gen}}$ is the generation rate of species k due to chemical reactions, c_v is the specific heat at constant volume, T is the temperature, p is the pressure, V is the reactor volume, h_{in} is the specific enthalpy at the inlet, and u_k is the specific internal energy of species k.

The spatial discretization adopted to define the CRN zones is shown in Fig. 1, where the main flow regions identified from the experimental velocity field are mapped onto the corresponding reactor domains.

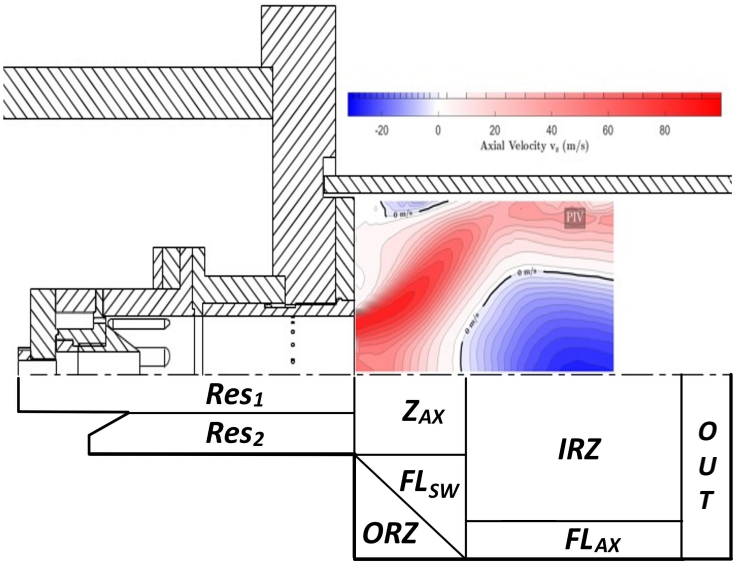


Fig. 1 Combustion chamber discretization employed in the CRN model. Top: experimental axial velocity field obtained from Particle Image Velocimetry (PIV) measurements, illustrating the characteristic flow structures of the swirl-stabilized combustor [20]. Bottom: corresponding CRN decomposition. The two upstream reservoirs model the mixing tube and the six downstream reactor zones discretize the combustion chamber.

The two reservoirs, Res_1 and Res_2, represent the two inlet streams entering the premixing section, each characterized by an individual equivalence ratio (ϕ_1, ϕ_2) and air mass-flow rate ($\dot{m}_{\text{air},1}$, $\dot{m}_{\text{air},2}$) derived from the global operating conditions through two splitting parameters: θ_1 and θ_2. Downstream, the combustion chamber is discretized into the following zones:

1) *Axial flow zone* (Z_{AX}): cylindrical region immediately downstream of the mixing tube exit, coinciding with the most probable flame-front location. Its volume is computed from the mixing tube cross-sectional area and an axial length of $L_{AX} = 51$ mm established from OH-PLIF imaging [18, 21].
2) *Swirling flame zone* (FL_{SW}): annular region surrounding Z_{AX} in which intense combustion occurs under significant tangential velocity. Its volume depends on the deflection angle $\gamma = \arctan(S_n)$, where S_n is the swirl number at the mixing tube exit, computed from the operating conditions following the simplified formulation of Beér and Chigier [24].

3) *Inner recirculation zone* (IRZ): central vortex region characterized by reverse flow that transports combustion products upstream, providing thermal feedback essential for flame stabilization. Its volume is obtained as the residual of the total combustion chamber volume after subtraction of all other zones.
4) *Outer recirculation zone* (ORZ): peripheral recirculation region adjacent to the chamber walls. Its volume is estimated by approximating the recirculation region as a solid of revolution with a triangular generating profile whose dimensions depend on γ and the chamber geometry [18].
5) *Axial flame zone* (FL_{AX}): downstream region bounded by the chamber wall and the IRZ, in which the flow is predominantly axial. Its volume is expressed as a fraction θ_9 of the total combustion chamber volume V_{CC} and is treated as a calibration parameter.
6) *Exit zone* (OUT): downstream reactor that collects the outflow from both the IRZ and the FL_{AX} zone before discharging to the exhaust. The extent of this region is set by the parameter θ_{10}, defined with respect to the total combustion chamber volume V_{CC}.

Figure 2 shows the adopted CRN topology and the mass-flow connections between the reservoirs and the reactor zones. The inter-reactor mass flow distribution is obtained by solving a system of mass balance equations at each node, closed by a control equation that enforces the equivalence ratio in the swirling flame zone FL_{SW} to match the global equivalence ratio ϕ of the operating condition [18]. This constraint, expressed in terms of elemental mass fractions of hydrogen and oxygen [25], yields a nonlinear equation in the mass flow rate $\dot{m}_A$ from Z_{AX} to FL_{SW}, which is solved via a hybrid grid-search and bisection algorithm. During the transient time integration, both the reactor volumes and the mass flow distribution are dynamically recomputed at regular intervals to account for the evolution of species compositions within each reactor, following the procedure detailed in Ref. [18].

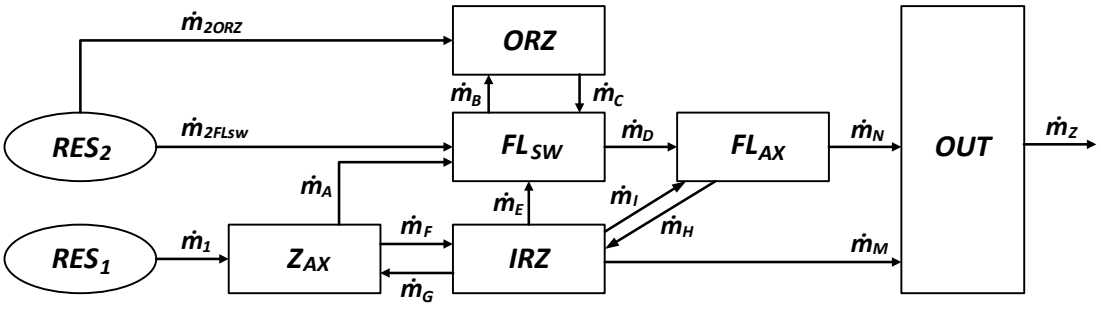


Fig. 2 Chemical Reactor Network topology adopted for the AHEAD combustor. The diagram reports the two inlet reservoirs, the reactor zones used to represent the combustion chamber, and the mass-flow paths employed in the network balance equations.

The network topology is fully defined by the ten-dimensional parameter vector:

$$\vec{\Theta} = \left[\theta_1,\ \theta_2,\ \theta_3,\ \theta_4,\ \theta_5,\ \theta_6,\ \theta_7,\ \theta_8,\ \theta_9,\ \theta_{10}\right] \tag{4}$$

The first two parameters control the upstream splitting: $\theta_1 = \phi_1/\phi_2$ sets the relative stoichiometry between the two inlet streams, and $\theta_2 = \dot{m}_{\mathrm{air},1}/\dot{m}_{\mathrm{air},2}$ determines their mass flow partitioning. The parameter θ_3 prescribes the fraction of the Res_2 outflow directed to the ORZ. Parameters θ_4 through θ_8 govern the internal mass flow redistribution among the combustion chamber reactors: θ_4 controls the FL_{SW} outflow split between ORZ and FL_{AX}. θ_5, θ_6, and θ_7 define a hierarchical redistribution of the IRZ outflow among its exit pathways (to Z_{AX}, FL_{SW}, and FL_{AX}, respectively). θ_8 quantifies the fraction of FL_{AX} outflow that recirculates to the IRZ. Finally, θ_9 and θ_{10} specify the volume fractions of FL_{AX} and OUT relative to the total chamber volume. The definition of the parameters is summarized in Table 2. This parameterization ensures that all splitting fractions remain strictly positive and bounded within the interval $(0, 1)$, while employing the minimum number of independent parameters necessary to fully specify the network topology. A comprehensive derivation of the mass flow algebra and the volume allocation strategy based on swirl number estimation is provided in Ref. [18].

The chemical kinetics are resolved using the Naik mechanism [26] comprising 21 species and 105 reactions, and the more comprehensive CRECK mechanism with 31 species and 203 reactions [27–29]. Both mechanisms include the species and reaction pathways necessary to capture the principal NO_x formation routes: the thermal mechanism, the N_2O-intermediate pathway, and the NNH route.

B. Comparison of Optimization Algorithms

The calibration of the CRN parameter vector $\vec{\Theta}$ constitutes an inverse problem in which ten model coefficients must be determined such that predicted NO_x emissions reproduce experimental observations within acceptable tolerances. In prior work by the present authors [10, 18], this calibration was performed exclusively via Particle Swarm Optimization. However, despite recent efforts based on Bayesian calibration of CRN models for pollutant emission prediction [15], systematic comparisons of optimization algorithms for this specific class of problems remain limited. The present section addresses this gap through a systematic benchmarking of six optimization algorithms spanning four distinct paradigmatic families: population-based metaheuristics, surrogate-assisted methods, gradient-based approaches, and hybrid global-local strategies. The primary objective is therefore to assess, for the AHEAD lean premixed burner and its corresponding CRN representation, which algorithmic paradigm provides the most favorable trade-off between solution quality and computational budget. The CRN is implemented within a standalone Python module that serves as the common computational kernel invoked by all six optimizers, ensuring that every algorithm operates on the same physical model.

1. Objective Function

A unified objective function is employed across all six optimizers. The CRN is calibrated against experimental NO_x measurements obtained by Reichel et al. for the AHEAD combustor at the reference inlet operating condition ($\dot{m}_{\text{air}} = 130$ kg/h, $T_{\text{air}} = 623$ K) [22, 30]. Four equivalence ratios spanning the full experimental range are selected as calibration points: $\phi = 0.40, 0.60, 0.87$, and 0.96. These values encompass both the low-emission lean regime and the high-emission near-stoichiometric regime, providing comprehensive coverage of the NO_x formation behavior.

The core metric is a weighted root-mean-square relative error:

$$J(\vec{\Theta}) = \sqrt{\sum_{i=1}^{n_{\text{cal}}} w_i \left(\frac{\text{NO}_{x,\text{sim},i} - \text{NO}_{x,\text{exp},i}}{\text{NO}_{x,\text{exp},i}} \right)^2} \tag{5}$$

where $n_{\text{cal}} = 4$ is the number of calibration points and the magnitude-dependent weights are defined as:

$$w_i = \frac{\text{NO}_{x,\text{exp},i}^{0.85}}{\sum_{j=1}^{n_{\text{cal}}} \text{NO}_{x,\text{exp},j}^{0.85}} \tag{6}$$

The exponent 0.85 gives greater weight to operating conditions characterized by higher absolute NO_x levels, while avoiding an excessive contribution from low-emission lean cases, where small absolute prediction errors can produce large relative deviations [18].

A candidate parameter vector is declared infeasible, and the objective function returns a large penalty value whenever the resulting CRN simulation produces non-physical outcomes. These include failure of any of the four calibration simulations to converge, or failure of the equivalence ratio control described in Ref. [18], which occurs when the available mass flow is insufficient to enforce $\phi_{FL_{SW}} = \phi$. This infeasibility handling is uniform across all six algorithms.

2. Optimization Algorithms

Six algorithms were selected to provide representative coverage of the main optimization paradigms. Each method follows a different strategy for navigating the search space, enabling assessment of which approach is best suited to the CRN calibration landscape. To ensure a fair comparison, all algorithms operate under a common evaluation budget of approximately 2000 objective function evaluations per run. The algorithm-specific settings governing how this budget is distributed are summarized in Table 1.

- Particle Swarm Optimization (PSO): PSO [16] provides the natural reference for this study, as it was employed in the original CRN calibration [18]. A swarm of 10 particles explores the parameter space, each updating its velocity based on its personal best position and the global best discovered by the swarm. Following the parameter-selection guidelines and convergence analysis of Trelea [31], an adaptive inertia weight decreasing linearly from 0.9 to 0.4 over 200 iterations is used to promote global exploration during the early stages and local exploitation toward the end of the search. Cognitive and social acceleration coefficients are set to $c_1 = c_2 = 1.5$, in accordance with standard PSO stability considerations [32]. The particle velocity is clamped to 20% of the bound range per dimension, and boundary violations are handled via reflection.

- Differential Evolution (DE): DE [33] generates trial solutions by adding weighted differences of randomly selected population members to a base vector. The best1bin strategy is adopted, in which the current best individual serves as the base vector. The mutation factor F is sampled uniformly in the range $[0.5, 1.0]$, and binomial crossover with probability $CR = 0.7$ determines which components are inherited from the mutant [34]. The initial population consists of 50 individuals, corresponding to five times the problem dimensionality, and is generated via Latin Hypercube Sampling. The nominal parameter vector is explicitly included as one of the initial candidates, so that the search starts from a population containing the same reference CRN configuration used across the benchmarking campaign.
- Covariance Matrix Adaptation Evolution Strategy (CMA-ES): CMA-ES [35, 36] adapts the full covariance matrix of a multivariate normal sampling distribution, effectively learning both the magnitude and orientation of the search ellipsoid from the ranking of candidate solutions. This adaptation renders the algorithm invariant to rigid transformations of the search space, a property that is particularly advantageous for problems with non-separable parameter interactions [37]. The offspring population size is set to $\lambda = 11$, and the initial step size is set to $\sigma_0 = 0.5$ in normalized coordinates, corresponding to an effective initial standard deviation of one quarter of the bound width per dimension. Bound constraints are handled through a sigmoid transformation, and infeasible evaluations are assigned a large finite penalty (10^8) rather than infinity to avoid corrupting the covariance matrix update.
- Bayesian Optimization (BO): BO [38, 39] constructs a Gaussian Process (GP) surrogate model of the objective function landscape and selects each successive evaluation point by maximizing an acquisition function that balances exploitation of regions with low predicted cost against exploration of regions with high predictive uncertainty. A Matérn kernel with smoothness parameter $\nu = 5/2$ is employed, while the gp_hedge acquisition portfolio strategy dynamically selects among expected improvement, lower confidence bound, and probability of improvement according to their past performance [40]. The initial GP training set comprises 50 points generated via Latin Hypercube Sampling, after which the remaining budget of 1950 evaluations is guided by the surrogate model. BO is fundamentally distinct from the population-based methods in that it constructs a global statistical model of the objective.
- ADAptive Moment estimation (ADAM): ADAM [41] is a first-order gradient-based optimization method that maintains adaptive per-parameter learning rates through exponential moving averages of the first and second moments of the gradient. Since the CRN does not provide analytical gradients, the gradients are estimated by central finite differences with a step size equal to 10^{-3} times the bound range in each dimension. The initial learning rate is set to $\alpha = 0.02$, with a decay factor of 0.001, while the moment parameters are $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Gradient norm clipping at $\|g\| \leq 10$ is applied to limit the effect of noisy finite-difference estimates in regions where the objective becomes discontinuous due to simulation failures. A multi-start strategy with 8 random restarts of 12 iterations each is employed to partially compensate for the inherently local nature of the method. Each iteration requires approximately $2 \cdot 10 + 1 = 21$ CRN evaluations, corresponding to two finite-difference perturbations per parameter plus one evaluation of the updated candidate, yielding a total budget consistent with the other algorithms.
- Hybrid Differential Evolution plus Limited-memory Broyden–Fletcher–Goldfarb–Shanno with Bound constraints (Hybrid DE + L-BFGS-B): This two-phase approach combines a global exploration stage with local gradient-based refinement [42]. In Phase 1, a short Differential Evolution run with a budget of 1000 evaluations, a population of 50 individuals, and the *best1bin* strategy is used to identify promising regions of the parameter space. In Phase 2, the top three unique solutions from Phase 1 serve as starting points for multi-start L-BFGS-B [43], a limited-memory quasi-Newton method that approximates the inverse Hessian from gradient information. Each local refinement is assigned a budget of 300 evaluations. This architecture is motivated by the complementary strengths of the two components: population-based methods excel at global exploration but converge slowly near optima, whereas quasi-Newton methods provide rapid local convergence but require initialization within the correct basin of attraction.

3. Benchmarking Design

The stochastic nature of the optimization algorithms considered in this study necessitates a benchmarking protocol that accounts for the variability induced by random initialization and algorithmic stochasticity. To this end, a design of experiments comprising twenty independent random seeds was constructed. Each seed defines a distinct random number generator state that governs all stochastic operations within a given algorithm, including initial population generation,

Table 1 Principal algorithm settings. All algorithms share the same parameter bounds (Table 2), objective function, and an approximate evaluation budget of 2000 CRN calls per run.

Algorithm	Setting	Value	Algorithm	Setting	Value
PSO	Particles	10	BO	GP kernel	Matérn ($\nu = 5/2$)
	Inertia w	$0.9 \rightarrow 0.4$ (linear)		Acquisition	gp_hedge
	c_1, c_2	1.5		Initial points (LHS)	50
	Max iterations	200		Total calls	2000
DE	Strategy	best1bin	ADAM	Learning rate α	0.02 (decay 10^{-3})
	F, CR	[0.5, 1.0], 0.7		β_1, β_2	0.9, 0.999
	Population	50		FD step h	$10^{-3}\cdot$ range
	Max generations	39		Restarts $\cdot$ iters	$8 \cdot 12$
CMA-ES	Population λ	11	Hybrid	Phase 1 (DE budget)	1000 evals
	σ_0	0.5		Phase 1 population	50
	Bound handling	BoundTransform		Phase 2 (L-BFGS-B)	3 starts
	Max generations	182		Phase 2 budget/start	300 evals

the stochastic coefficients of the velocity update in PSO, mutation sampling in DE, and the random exploration phase in BO. For every seed, all six algorithms are executed under identical conditions: the same parameter bounds (Table 2), the same initial nominal point $\vec{\Theta}_0$, and the same objective function implementation. This common-seed benchmarking protocol ensures that all algorithms are compared under reproducible and consistent conditions, so that observed performance differences can be attributed to the optimization strategy rather than to differences in problem formulation or initial conditions. The parameter bounds, reported in Table 2, restrict the search space to avoid non-physical CRN configurations.

Table 2 Parameter bounds for the CRN optimization. All splitting fractions ($\theta_3 - \theta_8$) are bounded to maintain strictly positive mass flow rates throughout the reactor network.

Parameter	Physical meaning	Lower	Upper
θ_1	ϕ_1/ϕ_2	0.2	5.0
θ_2	$\dot{m}_{\text{air},1}/\dot{m}_{\text{air},2}$	0.2	5.0
θ_3	$\dot{m}_{2,\text{ORZ}}/\dot{m}_2$	0.05	0.60
θ_4	$FL_{SW} \rightarrow$ ORZ split	0.05	0.60
θ_5	IRZ $\rightarrow Z_{AX}$ fraction	0.05	0.95
θ_6	IRZ $\rightarrow FL_{SW}$ fraction	0.05	0.95
θ_7	IRZ $\rightarrow FL_{AX}$ fraction	0.05	0.95
θ_8	$FL_{AX} \rightarrow$ IRZ recirculation	0.05	0.95
θ_9	$V_{FL_{AX}}/V_{CC}$	0.05	0.15
θ_{10}	V_{OUT}/V_{CC}	0.20	0.45

Therefore, a total of 120 simulations were conducted, consisting of twenty runs for each of the six algorithms. Each run produces a convergence trajectory recording the best objective function value as a function of the cumulative number of CRN evaluations and the elapsed wall-clock time. The resulting dataset of 120 final scores is then analyzed using non-parametric statistical tests to assess whether the observed performance differences among algorithms are statistically significant, as detailed in the results section.

C. Computational Cost Reduction

This section describes two complementary strategies for reducing the computational cost of CRN evaluations while preserving the accuracy of NO_x predictions. First, skeletal mechanism reduction via the Directed Relation Graph with Error Propagation (DRGEP) method systematically eliminates species and reactions with negligible influence on the target quantities, yielding smaller chemical systems with reduced stiffness and faster integration (Sec. I.C.1) [44]. Second, In Situ Adaptive Tabulation (ISAT) exploits the redundancy inherent in repeated thermochemical queries during reactor network calculations, replacing expensive direct integrations with rapid table lookups whenever a query falls within the accuracy region of a previously computed state (Sec. I.C.2) [45]. These strategies exhibit complementary characteristics: the reduced mechanism provides a baseline speedup applicable throughout all subsequent computations, and ISAT further accelerates repeated queries in similar thermochemical regions.

1. Skeletal Mechanism Reduction

Detailed hydrogen–air combustion mechanisms extended with comprehensive nitrogen chemistry typically involve tens of species and hundreds of reactions, introducing significant stiffness that limits integration time steps. The Naik mechanism [26] comprises 21 species and 105 reactions, while the more comprehensive CRECK mechanism [27–29] encompasses 31 species and 203 reactions. Skeletal mechanism reduction techniques aim to systematically eliminate species and reactions with minimal impact on the target quantities, here NO_x emissions and temperature profiles, thereby reducing the computational cost per CRN evaluation without appreciable loss in prediction accuracy.

The reduction is performed using the Directed Relation Graph with Error Propagation (DRGEP) method introduced by Pepiot-Desjardins and Pitsch [44]. DRGEP extends the earlier Directed Relation Graph (DRG) approach of Lu and Law [46] by accounting for the propagation of errors along multi-step coupling paths between species. Given a set of sampled thermochemical states, the method first computes, at each state, a direct interaction coefficient r_{AB} for every species pair (A, B). This coefficient measures the contribution of species B to the production or consumption of species A, normalized by the total production or consumption of A:

$$r_{AB} = \frac{\sum_{i=1}^{n_r} \left|\nu_{A,i}\,\omega_i\right| \delta_{B,i}}{\max\left(P_A,\ C_A\right)} \tag{7}$$

where $\nu_{A,i}$ is the net stoichiometric coefficient of species A in reaction i, ω_i is the net rate of progress of reaction i, $\delta_{B,i}$ is a species participation indicator equal to unity if species B participates in reaction i and zero otherwise, P_A is the total production rate of species A, and C_A is its total consumption rate. The element-wise maximum of r_{AB} over all sampled states yields a conservative estimate of the direct coupling strength between every species pair.

To quantify the overall importance of a species B relative to a set of user-specified target species, DRGEP employs a graph-theoretic approach. The species interaction network is interpreted as a weighted directed graph, and the overall interaction coefficient R_B is defined as the maximum-product path from any target species to B. This computation is efficiently performed using Dijkstra's shortest-path algorithm [47] in a transformed weight space ($w = -\ln r$), following the DRGEP formulation of Pepiot-Desjardins and Pitsch [44]. Species whose overall interaction coefficient R_B falls below a prescribed threshold ε are deemed removable. A sweep over multiple threshold values generates a family of skeletal mechanisms of varying sizes, from which the optimal trade-off between compactness and accuracy is identified.

These mechanisms are then written to Cantera-compatible YAML files. Each file contains only the retained species and reactions.

The thermochemical states used to evaluate the DRGEP interaction coefficients must adequately represent the conditions encountered during CRN operation. The sampling dataset is generated from two complementary simulation types at atmospheric pressure:

- *Constant-pressure autoignition*: batch reactor trajectories spanning $\phi \in [0.2,\ 2.0]$ and $T \in [300,\ 2700]$ K, for a total of 475 simulations.
- *Perfectly stirred reactor*: steady-state solutions at the same ϕ and T grid, with five residence times ($\tau = 10^{-5}$, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1} s) spanning the range relevant to the six CRN zones, yielding 2375 simulations.

The equivalence-ratio interval is intentionally extended beyond the nominal CRN operating envelope ($\phi = 0.30$–1.0) up to $\phi = 2.0$, while the temperature interval reaches 2700 K. This ensures that transient states encountered during CRN time integration, including locally rich zones and high-temperature recirculation products, are represented in the sampling dataset.

The target species for the DRGEP graph search are H_2, H_2O, N_2O, NO, NO_2, O_2, and OH. These species include the primary fuel and oxidizer, the main NO_x species, the N_2O-intermediate pathway, and the OH radical whose concentration governs flame structure and thermal NO_x formation kinetics. In addition, N_2, NNH, NH, and HNO are designated as protected species and are therefore never removed, regardless of their R_B value. This choice preserves the NNH and HNO pathways, which contribute to NO_x formation under lean conditions [48].

Each candidate skeletal mechanism is subjected to a comprehensive evaluation procedure. The mechanism is compared against its detailed parent across seven complementary metrics, summarized in Table 3. This protocol ensures that the retained skeletal mechanisms accurately reproduce both the global combustion behavior and the detailed nitrogen chemistry pathways essential for accurate NO_x prediction.

Table 3 Metrics used to evaluate the candidate skeletal mechanisms.

#	Metric
1	Ignition delay time
2	Laminar flame speed
3	Equilibrium NO_x concentration
4	PSR extinction residence time
5	Peak species mole fractions (NO, NO_2, OH)
6	Net production rates (NO, NO_2, OH)
7	Heat release ratio

As a final verification step, each skeletal mechanism is inserted into the full CRN model described in Sec. I.A. Simulations are performed over the range $\phi = 0.30$–1.00 at the reference inlet operating condition ($\dot{m}_{\mathrm{air}} = 130$ kg/h, $T_{\mathrm{air}} = 623$ K). For this comparison, the calibrated parameter vector $\vec{\Theta}$ obtained with the detailed parent mechanism is kept fixed, and no additional CRN recalibration is performed for the skeletal mechanism. The resulting NO_x predictions, species mole fractions, net production rates, and heat release rates are then compared against those obtained with the detailed mechanism. This comparison verifies that the accuracy observed in canonical reactor configurations is preserved in the interconnected multi-zone CRN environment.

The entire reduction and evaluation pipeline is mechanism-agnostic: it is applied to both the Naik and CRECK detailed mechanisms, producing separate families of validated skeletal mechanisms for each parent mechanism.

2. Flow-Map In Situ Adaptive Tabulation

The CRN simulations described in Sec. I.A require the repeated solution of a coupled system of stiff ODEs (Eqs. 1–3) until convergence is achieved for each equivalence ratio. In the optimization campaign described in Sec. I.B, each calibration evaluation involves four such simulations, one for each calibration point, and the full benchmarking protocol therefore requires thousands of CRN evaluations. The resulting computational cost, dominated by the direct integration (DI) of the chemical kinetics, motivates the introduction of an acceleration strategy based on In Situ Adaptive Tabulation (ISAT).

ISAT, originally proposed by Pope [45] for turbulent combustion calculations, constructs an adaptive look-up table of previously computed thermochemical mappings during the course of a simulation. When a new query composition falls within the accuracy region of an existing table entry, the result is approximated from the stored mapping, bypassing the expensive ODE integration. Queries that fall outside the accuracy regions of the existing entries are advanced by direct integration, and the resulting mappings are added to the table for future reuse. The method has been successfully applied to PDF calculations of turbulent flames [45], operator-split reacting flow simulations [49, 50], and networks of perfectly stirred reactors [51]. Algorithmic improvements, particularly regarding ellipsoid-of-accuracy management and binary-tree search strategies, were developed by Lu and Pope [52].

In the classical ISAT formulation, the tabulated quantity is the reaction mapping $\vec{R}(\vec{\phi}_0, \Delta t)$, which advances a single composition vector $\vec{\phi}_0$ by a time step Δt under pure chemical kinetics. The present implementation differs in a fundamental respect: the tabulated mapping is the *discrete flow map* of the entire reactor network,

$$\vec{X}(t + \Delta t_{\mathrm{tab}}) = \vec{\Phi}_{\Delta t_{\mathrm{tab}}}\big(\vec{X}(t)\big), \tag{8}$$

where $\vec{X}$ is the concatenated thermochemical state of all six CRN reactors and Δt_{tab} is a user-controlled macro-step.

This network-level tabulation preserves the bidirectional coupling among reactors, which would otherwise be lost if each reactor were tabulated independently.

The state vector $\vec{X}$ is defined as the ordered concatenation

$$\vec{X} = \left[\, T^{(1)},\, Y_1^{(1)}, \ldots, Y_{N_s}^{(1)},\, T^{(2)},\, Y_1^{(2)}, \ldots, Y_{N_s}^{(2)},\, \ldots,\, T^{(6)},\, Y_1^{(6)}, \ldots, Y_{N_s}^{(6)} \,\right]^{\top}, \tag{9}$$

where the superscript denotes the reactor index (ordered as Z_{AX}, ORZ, FL_{SW}, FL_{AX}, IRZ, OUT) and N_s is the number of chemical species. The total dimension of $\vec{X}$ is $n = 6\,(1 + N_s)$, yielding $n = 132$ for the Naik mechanism ($N_s = 21$) and $n = 192$ for the CRECK mechanism ($N_s = 31$). Pressure is not included in the state vector $\vec{X}$ because the AHEAD combustor is modeled at atmospheric pressure, and during ISAT retrieval the tabulated temperature and composition are restored at the prescribed operating pressure.

In addition to the thermochemical state, the network evolution over a macro-step depends on the inter-reactor mass flow rates and the reactor volumes, both of which are recomputed at each macro-step boundary by the dynamic update procedure described in Sec. I.A. These forcing parameters, together with the macro-step size Δt_{tab}, are separated from the thermochemical state and used to index distinct sub-tables. Specifically, the master ISAT table is organized as a dictionary of sub-tables keyed by the pair $\left(\Delta t_{\text{key}},\, h_{\text{forcing}}\right)$, where Δt_{key} is the macro-step rounded to three significant digits and h_{forcing} is a deterministic hash of the mass flow rates and reactor volumes rounded to four significant digits. This factorization enables each ISAT binary tree to operate in the pure thermochemical state space of dimension n, rather than in an augmented space that would include the forcing parameters and the time step as additional input dimensions. The coarse rounding of the hash arguments prevents the creation of spurious sub-tables caused by small floating-point variations in quantities that evolve slowly between consecutive macro-steps.

The components of the state vector $\vec{X}$ span vastly different magnitudes: temperatures are $O(10^3)$ K, major species mass fractions (N_2, O_2, H_2O) are $O(10^{-1})$, and trace species (NO, NO_2, radicals) are $O(10^{-4})$ or smaller. All distance computations and error assessments are performed in a scaled state space $\tilde{\vec{X}} = \vec{s} \odot \vec{X}$, where $\vec{s}$ is a componentwise scaling vector with entries $s_i = 1/X_{i,\text{ref}}$. The reference values are $T_{\text{ref}} = 2000\,\text{K}$ for temperature, $Y_{\text{ref}} = 0.1$ for major species, and $Y_{\text{ref}} = 10^{-4}$ for minor species, so that all scaled components are $O(1)$. This scaling is critical for the accuracy of the ellipsoid-of-accuracy criterion and for meaningful error control in the retrieval step [45, 52].

The ISAT table is organized as a binary tree [45], in which each leaf stores a tabulation record consisting of the tabulation point $\vec{X}_0$ (input state), the mapped state $\vec{Y}_0 = \vec{\Phi}_{\Delta t}(\vec{X}_0)$ (output state obtained by DI), and the ellipsoid of accuracy (EOA), represented by a matrix $\vec{L}$ such that the EOA is the set

$$\{\delta\tilde{\vec{X}} : \|\vec{L}\,\delta\tilde{\vec{X}}\|_2 \le 1\} \tag{10}$$

where $\delta\tilde{\vec{X}} = \vec{s} \odot (\vec{X} - \vec{X}_0)$ is the displacement in scaled space. The internal nodes of the tree store perpendicular-bisector hyperplanes in scaled space, enabling $O(\log N)$ traversal to the closest leaf for a given query [45]. When a query composition $\vec{X}_q$ is received, the algorithm proceeds as follows:

- *Retrieve*: if $\vec{X}_q$ lies within the EOA of the closest leaf ($\|\vec{L}\,\delta\tilde{\vec{X}}_q\|_2 \le 1$), the output is approximated by the stored mapped state via a zeroth-order approximation:

$$\vec{Y}_q \approx \vec{Y}_0. \tag{11}$$

  After retrieval, mass fraction positivity and normalization ($\sum_k Y_k = 1$) are explicitly enforced.
- *Grow*: if the query falls outside the EOA, a direct integration is performed to obtain the reference mapping $\vec{Y}_{\text{exact}}$. The local approximation error associated with the zeroth-order estimate is computed as

$$\varepsilon_{0\text{th}} = \left\| \vec{s} \odot \left(\vec{Y}_0 - \vec{Y}_{\text{exact}} \right) \right\|_2 . \tag{12}$$

  If $\varepsilon_{0\text{th}} \le \varepsilon_{\text{tol}}$, the EOA of the existing leaf is enlarged to include the query point via a rank-one update of $\vec{L}$, which extends the ellipsoid along the direction of $\delta\tilde{\vec{X}}_q$ while preserving its shape in the orthogonal directions, following Lu and Pope [52].
- *Add*: if $\varepsilon_{0\text{th}} > \varepsilon_{\text{tol}}$, a new leaf is inserted into the binary tree with the DI result as its tabulation record. The initial EOA is a hypersphere of radius r_0 in scaled space, set equal to ε_{tol}.

The use of zeroth-order retrieval is motivated by the structure of the present CRN flow-map tabulation problem. In contrast to classical single-reactor ISAT applications, the tabulated state corresponds to the concatenated thermochemical

state of the full six-reactor network. The mapping over one macro-step depends on strongly coupled inter-reactor mass flows and dynamically updated reactor volumes. A first-order approximation would therefore require an accurate evaluation of the full Jacobian of the coupled network mapping. Using the zeroth-order approximation as the baseline ISAT retrieval strategy avoids the evaluation, storage, and application of a mapping-gradient matrix during retrieval, thereby simplifying the implementation and reducing the computational overhead. As shown in Sec. II.B.3, the associated approximation error remains small for the operating conditions considered. These considerations motivate the adoption of zeroth-order retrieval in the present implementation.

The reference CRN solver described in Sec. I.A advances the reactor network with an adaptive time-integration procedure, in which the time-step size is selected internally by the integrator. In the ISAT-accelerated variant, this solution strategy is replaced by a two-phase methodology.

In the first phase, referred to as warm-up, a prescribed number of time steps, typically 200, is advanced using the standard integration procedure. This allows the ignition transient to develop, the integrator to resolve the stiff chemical time scales, and the dynamic update of reactor volumes and mass flow rates to be executed several times. No ISAT queries are issued during this phase, since the rapidly evolving transient states would populate the table with entries unlikely to be revisited.

In the second phase, referred to as macro-stepping, the simulation advances through user-controlled increments Δt_{tab}. These increments are selected from a time-dependent schedule consistent with the maximum time-step constraints adopted in the DI-based simulations. Small macro-steps ($\Delta t_{\text{tab}} \sim 10^{-6}$ s) are used during the post-ignition transient and are progressively increased up to $\Delta t_{\text{tab}} \sim 0.5$ s as the network approaches steady state. At each macro-step boundary, the dynamic update block, including the swirl number, reactor volumes, and mass-flow redistribution, is executed. The resulting forcing parameters are then frozen over the duration of the macro-step. The current network state is extracted and used to query the ISAT table. If the retrieval is successful, the tabulated output state is written directly into the six reactor states. If the retrieval is unsuccessful, direct integration is performed, and the resulting input-output pair is passed to the table for a grow or add operation.

The CRN topology, flow-system algebra, dynamic update logic, and equivalence-ratio control constraint remain identical to those of the reference solver described in Sec. I.A. The only modification concerns the time-advancement procedure: the underlying physical model is unchanged, while the ISAT wrapper introduces an approximation controlled by ε_{tol}.

D. Uncertainty Quantification of NOx Predictions

This section describes the framework employed to characterize the propagation of parametric uncertainty through the calibrated CRN of Sec. I.A. The objective is twofold: to quantify the statistical moments and the uncertainty band of the network NO_x response around the reference vector $\vec{\Theta}^*$, and to attribute the resulting variance to the individual sources of uncertainty acting on the model. Both objectives are addressed within a non-intrusive Polynomial Chaos Expansion (PCE) framework [53–55], organized as a block decomposition of heterogeneous uncertainty sources and complemented by a dedicated evaluation protocol.

Let $\vec{\xi} = (\xi_1, \ldots, \xi_D) \in \mathbb{R}^D$ denote the vector of D uncertain input variables, assumed mutually independent, with joint probability density function

$$f_{\vec{\xi}}(\vec{\xi}) = \prod_{j=1}^{D} f_{\xi_j}(\xi_j),$$

where f_{ξ_j} is the marginal density of the j-th input. The scalar model response is denoted by

$$Y = \mathcal{M}(\vec{\xi}),$$

where $\mathcal{M}$ is the model mapping from the uncertain input space to the output space. Assuming that the response has finite second moment, i.e. $\mathbb{E}[Y^2] < \infty$, the generalized Wiener-Askey polynomial chaos framework [55] states that Y can be represented, in the mean-square sense, as an expansion over a basis of multivariate orthonormal polynomials,

$$Y = \sum_{\vec{\alpha} \in \mathbb{N}^D} c_{\vec{\alpha}} \, \Psi_{\vec{\alpha}}(\vec{\xi}), \qquad \left\langle \Psi_{\vec{\alpha}}, \Psi_{\vec{\beta}} \right\rangle = \delta_{\vec{\alpha}\vec{\beta}}, \tag{13}$$

where $\vec{\alpha} = (\alpha_1, \ldots, \alpha_D)$ is a multi-index, $c_{\vec{\alpha}}$ are the expansion coefficients, and $\Psi_{\vec{\alpha}}(\vec{\xi})$ are multivariate polynomials orthonormal with respect to the joint density $f_{\vec{\xi}}$. The latter are obtained by tensor-product construction from the

corresponding univariate orthonormal polynomial families, selected through the Wiener–Askey correspondence (probabilist Hermite for Gaussian inputs, Legendre for uniform inputs). In practical applications, the infinite series of Eq. (13) is truncated to a finite total order p_o, so that only terms satisfying $|\vec{\alpha}| = \sum_{j=1}^{D} \alpha_j \leq p_o$ are retained. This yields the approximation

$$\widetilde{Y}_{p_o} = \sum_{|\vec{\alpha}| \leq p_o} c_{\vec{\alpha}} \, \Psi_{\vec{\alpha}}(\vec{\xi}),$$

whose number of retained basis functions is

$$P = \binom{D + p_o}{p_o}.$$

Two useful properties follow directly from the truncated expansion $\widetilde{Y}_{p_o}$. Owing to the orthonormality of the polynomial basis, the first two statistical moments of the surrogate are obtained in closed form: the mean is given by the coefficient associated with the constant basis term,

$$\mathbb{E}[\widetilde{Y}_{p_o}] = c_{\vec{0}},$$

where $\vec{0} = (0, \ldots, 0)$ denotes the null multi-index, while the variance is obtained by summing the squared coefficients of all non-constant terms,

$$\mathrm{Var}[\widetilde{Y}_{p_o}] = \sum_{\vec{\alpha} \neq \vec{0}} c_{\vec{\alpha}}^2.$$

A second important property is that the variance-based Sobol sensitivity indices [56] can also be evaluated directly from the PCE coefficients, following the formulation introduced for polynomial chaos surrogates in [57]. In particular, the first-order Sobol index S_i measures the fraction of the output variance associated with the isolated effect of the i-th input variable ξ_i, whereas the total-effect index S_i^T accounts for both its direct contribution and all interaction terms involving ξ_i. These quantities are written as

$$S_i = \frac{1}{\mathrm{Var}[\widetilde{Y}_{p_o}]} \sum_{\vec{\alpha} \in \mathcal{A}_i} c_{\vec{\alpha}}^2, \qquad S_i^T = \frac{1}{\mathrm{Var}[\widetilde{Y}_{p_o}]} \sum_{\vec{\alpha} \in \mathcal{A}_i^T} c_{\vec{\alpha}}^2, \tag{14}$$

where $\mathcal{A}_i$ denotes the set of multi-indices $\vec{\alpha}$ associated with basis functions depending only on ξ_i, and $\mathcal{A}_i^T$ the set of all multi-indices for which ξ_i appears, either alone or in interaction with other inputs [58]. Once the coefficients $\{c_{\vec{\alpha}}\}$ have been identified, both the statistical moments and the Sobol indices are therefore obtained at no additional model-evaluation cost, avoiding the direct Monte Carlo integration that would otherwise be required to estimate them from repeated sampling.

Three classes of parametric uncertainty are propagated through the calibrated CRN: the operating conditions imposed at the rig inlets, the topology calibration parameters described in Sec. I.A, and the volume-allocation and closure assumptions of the network. The framework is organized as a block decomposition: each uncertainty class is first propagated through a dedicated PCE of moderate dimension. The joint effect is then evaluated in a final block that collects the relevant parameters from the individual blocks.

- Block **O** ($D = 3$) propagates the inlet boundary conditions of the AHEAD test rig of Reichel et al. [22]: T_{air} is modeled as a Gaussian variable with mean 623 K and standard deviation 6.23 K, corresponding to an assumed 1% uncertainty on the overall temperature measurement chain, including sensor accuracy, acquisition, and installation effects. The relative deviations on the air and fuel mass-flow rates are represented as independent zero-mean Gaussian variables with standard deviation 0.015. Since the exact type and accuracy class of the mass-flow meters are not specified, this value is adopted as a conservative estimate of the corresponding measurement uncertainty.
- Block **C** ($D = 8$) propagates the local sensitivity of the network response to small perturbations of $\vec{\Theta}^*$ on the eight calibration parameters that govern the inter-reactor mass-flow distribution, θ_1 through θ_8. Each parameter in this class is modeled as a uniform variable over a narrow interval centered on its reference value Θ_k^*, with absolute half-width 0.025. As a preliminary diagnostic, the ensemble of acceptable calibration solutions is post-processed by evaluating both Spearman's rank correlation coefficient [59] and Pearson's product-moment correlation coefficient [60] for all parameter pairs. In the presence of strong correlations, a reduced parameterization is adopted in which one representative parameter is retained in the uncertain set while the others are fixed at their reference values. This step reduces the initial eight-parameter set to $K_1 \leq 8$ inputs for the subsequent stage. A Rosenblatt transform would provide a more rigorous, although more intrusive, alternative. The local PCE is

then constructed under the independence assumption. To further reduce the effective dimensionality, a screening procedure is applied to identify the subset of influential calibration parameters, leading to a reduced set of $K_2 \leq K_1$ variables that are retained for propagation in the final combined block.

- Block **G** ($D = 4$) propagates the residual uncertainty associated with the volume-allocation and closure methodology adopted in the CRN model. The two volume fractions $\theta_9 = V_{FL_{AX}}/V_{CC}$ and $\theta_{10} = V_{\mathrm{OUT}}/V_{CC}$, which by construction control the partitioning of the chamber volume among the downstream reactors, are modeled as uniform variables with absolute half-width 0.025 centered on Θ_9^* and Θ_{10}^*, in accordance with the local-perturbation treatment adopted in Block C. The axial length of the Z_{AX} reactor defined in Sec. I.A is assigned a relative half-width of 20% to account for the residual uncertainty between the OH-PLIF estimate and the cold-flow CFD reconstruction used to define its nominal value. Similarly, the Beér–Chigier swirl-jet spreading coefficient σ is modeled as a uniform variable ranging from 0.60 to 1.00 [24]. Variations in both Z_{AX} length and σ directly affect the reconstructed reactor volumes through the geometric relations of the CRN, following the volume-estimation procedure described in [18].
- Block **T** ($D = 3+K_2+4$) finally combines the operating, the influential calibration ($K_2 \leq 8$), and the volume/closure inputs into a single experimental design and is the source of the global predictive band on NO_x.

All blocks are propagated at the eight equivalence-ratio set points ϕ_{ref} that span the experimental sweep of [30]. The configuration of the four blocks is summarized in Table 4.

Table 4 Configuration of the PCE blocks: dimensionality D, total order p_o, basis cardinality P, regression sample size per equivalence-ratio set point N, and external-validation sample N_{ext}.

Block	D	p_o	P	N	N_{ext}
O (operating)	3	3	20	100	–
C (calibration, screening)	$K_1 \leq 8$	2	≤ 45	200	–
C (calibration, refinement)	$K_2 \leq K_1$	3	≤ 165	500	150
G (volume/closure)	4	3	35	150	–
T (joint)	$3+K_2+4$	2	≤ 136	500	150

For each uncertainty block and for each equivalence-ratio set point ϕ_{ref}, an experimental design of size N is generated by sampling the input vector $\vec{\xi}$ through the Sobol low-discrepancy sequence [61]. Each sampled realization $\vec{\xi}^{(k)}$ is then propagated through the CRN model to obtain the corresponding scalar response

$$Y^{(k)} = \mathcal{M}\left(\vec{\xi}^{(k)}\right),$$

so that the regression dataset is given by the pairs $\{(\vec{\xi}^{(k)}, Y^{(k)})\}_{k=1}^{N}$. For the moderate input dimensions considered in the present work, Sobol sampling was preferred to Latin Hypercube Sampling [62], since it provides a more uniform coverage of the parameter space and improves the conditioning of the regression problem, while remaining compatible with the mixture of Gaussian and uniform marginals used across the different blocks.

The coefficients of the truncated PCE are identified by point-collocation regression,

$$\widehat{\vec{c}} \;=\; \arg\min_{\vec{c}\in\mathbb{R}^P} \left\| \vec{Y} - \mathbf{H}\vec{c} \right\|_2^2, \qquad H_{kj} = \Psi_{\vec{\alpha}_j}(\vec{\xi}^{(k)}), \tag{15}$$

where H_{kj} is the value of the j-th multivariate polynomial basis function $\Psi_{\vec{\alpha}_j}$ evaluated at the k-th sample $\vec{\xi}^{(k)}$. The initial experimental design is chosen with $N > 3P$ to preserve an overdetermined regression system after infeasible samples are discarded. This conservative oversampling strategy accounts for the fact that some CRN evaluations may fall in infeasible regions of the parameter space and follows the general practice of using overdetermined regression designs to limit overfitting in PCE construction [63].

The regression problem of Eq. (15) is solved independently at each ϕ_{ref}. Once the surrogate has been identified, a Monte Carlo sampling step is performed directly on the PCE by drawing $2 \cdot 10^4$ additional realizations of $\vec{\xi}$. The empirical 2.5% and 97.5% quartiles of the resulting surrogate response are finally used to define the reported 95% uncertainty band.

For the calibration block, a two-stage adaptive workflow is implemented, in line with variance-based screening strategies commonly adopted in high-dimensional PCE in [57, 63]. In the screening stage, an order-two PCE is fitted on

the active set after the correlation diagnostic with $N_{\mathrm{scr}} = 200$ Sobol points per ϕ_{ref} (regression ratio $N_{\mathrm{scr}}/P \geq 4.4$). The total-effect indices $S_k^T(\phi_{\mathrm{ref}})$ are aggregated across the eight set points into a ϕ-averaged score

$$\overline{S_k^T} \;=\; \frac{1}{N_\phi}\sum_{\ell=1}^{N_\phi} \max\!\big(0,\; S_k^T(\phi_{\mathrm{ref}}^{(\ell)})\big) \tag{16}$$

The smallest subset whose cumulative score covers a fraction $q = 0.95$ of the total is retained. In the refinement stage, an order-three PCE is then fitted on the K_2 retained dimensions with $N_{\mathrm{refin}} = 500$ fresh Sobol points, thereby ensuring a regression ratio $N_{\mathrm{refin}}/P > 3$ over the entire range of admissible K.

Each surrogate is finally assessed through four complementary metrics:

- *Leave-one-out predictive coefficient.* Leave-one-out cross-validation [64, 65] provides the Stone–Geisser coefficient, commonly used in PCE validation [63],

$$Q^2 \;=\; 1 \;-\; \frac{\sum_{k=1}^{N}\big(\epsilon_k^{\mathrm{LOO}}\big)^2}{\sum_{k=1}^{N}\big(Y^{(k)} - \overline{Y}\big)^2}, \tag{17}$$

  where $\epsilon_k^{\mathrm{LOO}} = Y^{(k)} - \widetilde{Y}_{p_o}^{(-k)}\big(\vec{\xi}^{(k)}\big)$ is the residual obtained by refitting the surrogate without sample k.
- *Train–test predictive check.* An 80/20 train–test partition complements the leave-one-out analysis by exposing overfitting patterns that may not be apparent from the per-fold structure of the LOO estimate.
- *Order-convergence of the Sobol decomposition.* A PCE of order $p_o - 1$ is refitted on the same sample and the maximum absolute Sobol drift,

$$\Delta_{S_T}^{\max} = \max_i \big|S_i^T(p_o) - S_i^T(p_o - 1)\big|,$$

  is computed in order to verify that not only the surrogate response, but also the variance apportionment, has stabilized at the selected polynomial order.
- *External predictive assessment.* An independent Sobol design of N_{ext} points (Table 4), generated with a different seed, is propagated through the CRN and used to compare the surrogate predictions against previously unseen model evaluations.

II. Results

This section presents the results of the proposed CRN enhancement framework. First, Sec. II.A reports the systematic comparison of six optimization algorithms for CRN parameter calibration, identifies the most effective algorithmic strategy, and defines the reference parameter vector used in the subsequent analyses. Then, Sec. II.B establishes the baseline performance of the calibrated CRN model and subsequently evaluates the computational cost reduction strategies, namely skeletal mechanism reduction and In Situ Adaptive Tabulation, by quantifying their accuracy-efficiency trade-offs with respect to this reference configuration.

A. Comparison of Optimization Algorithms

The benchmarking campaign described in Sec. I.B.3 totals 120 independent optimization runs, each operating under the same parameter bounds (Table 2), objective function (Eq. 5), and approximate evaluation budget of 2000 CRN calls. The runs were executed in parallel, with each optimizer assigned to a single CPU core, on a computing node equipped with Intel(R) Xeon(R) Gold 5318Y CPUs operating at 2.10 GHz. The Naik chemical kinetic mechanism was adopted throughout for all optimizers. The resulting dataset is analyzed below in terms of solution quality, convergence behavior, statistical significance, and parameter-space structure.

Table 5 summarizes the primary performance metrics. Overall, the results reveal a clear three-tier structure. *CMA-ES* provides the most consistent performance, achieving the lowest median score ($J = 0.0424$) and the tightest interquartile range ($[0.040,\ 0.044]$). This indicates high solution quality and reliability: 19 out of 20 seeds reach $J < 0.05$, corresponding to a 95% success rate, the highest among all algorithms. *PSO* attains the best individual score of the entire campaign ($J = 0.0360$, seed 17), but its median value ($J = 0.0447$) is slightly higher than that of CMA-ES and its interquartile range is substantially wider ($[0.041,\ 0.195]$). This larger spread highlights a stronger sensitivity to the initial conditions: only 60% of the seeds reach the $J < 0.05$ threshold. *DE* lies between these two behaviors: its median score is slightly higher ($J = 0.0485$), but the dispersion remains remarkably narrow ($[0.047,\ 0.055]$), and its success rate of 65% is comparable to that of PSO.

The remaining three algorithms form a distinctly lower-performance tier. The Hybrid DE+L-BFGS-B strategy reaches a median value of $J = 0.065$, with only one seed achieving $J < 0.05$. Bayesian Optimization stagnates at a median score of $J = 0.205$, and none of its 20 seeds falls below the $J < 0.05$ threshold. ADAM provides the poorest overall performance: its median score exceeds $J = 0.64$, five out of twenty seeds fail to identify any feasible CRN configuration across all eight restarts, and no seed approaches the $J < 0.05$ threshold.

Table 5 Performance summary across 20 seeds per algorithm. "Success" denotes the fraction of seeds reaching $J < 0.05$. "Feasibility" is the fraction of CRN evaluations producing a physically admissible solution.

Algorithm	Best J	Median J	IQR	Success	Feasibility
CMA-ES	0.0387	0.0424	[0.040, 0.044]	95%	76.6%
PSO	0.0360	0.0447	[0.041, 0.195]	60%	69.0%
DE	0.0440	0.0485	[0.047, 0.055]	65%	49.6%
Hybrid	0.0488	0.0651	[0.061, 0.076]	5%	38.5%
BO	0.0584	0.2047	[0.137, 0.269]	0%	85.8%
ADAM	0.0721	0.6473	[0.555, 1.000]	0%	75.8%

The convergence dynamics differ markedly across algorithms and provide insight into the mechanisms by which each strategy navigates the search space. Figure 3 displays the median best-so-far score as a function of cumulative CRN evaluations, with interquartile-range shading, for all six algorithms. Figure 4 presents the same data against wall-clock time.

CMA-ES exhibits the fastest initial descent and the narrowest uncertainty band throughout the optimization. Its median trajectory crosses the $J = 0.05$ threshold within approximately 1150 evaluations and continues improving steadily thereafter, reflecting the ability of the covariance-matrix adaptation to learn the local curvature of the objective landscape. PSO converges more slowly, its median does not cross $J = 0.05$ until approximately 1500 evaluations. Nevertheless, its median trajectory ultimately approaches performance levels close to those of CMA-ES. At the same time, it maintains a wide interquartile band until the end of the computational budget, as also observed in Table 5, indicating a pronounced sensitivity to the initial seed and a less consistent convergence behavior across runs. DE reaches the $J = 0.05$ level after approximately 1850 evaluations, but its dispersion remains extremely tight throughout the optimization, suggesting a highly repeatable search behavior despite the slower convergence rate. The Hybrid strategy behaves regularly during the initial DE stage of 1000 evaluations. In contrast, during the subsequent L-BFGS-B refinement, some runs occasionally drift into infeasible regions, leading to the premature termination of several seeds. As a result, the method does not systematically reach the 2000-evaluation budget, and its interquartile range does not fall below the $J = 0.05$ threshold by the end of the evaluation budget.

BO and ADAM converge to qualitatively different regimes. The BO median decreases monotonically but slowly, plateauing near $J \approx 0.2$ well before the evaluation budget is exhausted. This stagnation reflects the difficulty of constructing a faithful Gaussian Process surrogate in a ten-dimensional space with a substantial infeasible fraction that introduces discontinuities into the objective landscape, degrading the GP model quality. ADAM, instead, terminates markedly earlier than the other algorithms, with the median run ending after approximately 300 evaluations. As in the local L-BFGS-B stage of the Hybrid strategy, several seeds encounter infeasible regions. In this case, these points invalidate the finite-difference gradient estimates, so that the effective number of productive evaluations is considerably smaller than the nominal budget.

The wall-clock convergence reported in Figure 4 is consistent with the per-evaluation behavior of Figure 3. The total wall-clock cost of PSO, CMA-ES, and DE lies within the same order of magnitude. The residual differences among them are consistent with the fraction of evaluations spent in feasible regions reported in Table 5. DE is the fastest of the three but also exhibits the lowest feasibility (49.6%), since a substantial share of its evaluations is cheaply discarded through the infeasible-region branch. PSO occupies an intermediate position, with a feasibility rate of 69.0% and a corresponding wall-clock cost between those of DE and CMA-ES. CMA-ES is slightly more expensive in terms of total wall-clock time, but this is consistent with its higher feasibility rate (76.6%). BO, by contrast, incurs a substantially larger wall-clock time because each iteration requires an update of the Gaussian Process surrogate, whose cost grows with the accumulated dataset and dominates the per-call budget. As discussed above, both Hybrid and ADAM often terminate before exhausting the full evaluation budget due to infeasible-region encounters. This effect is particularly pronounced for ADAM, whose median run stops after approximately 300 evaluations, resulting in the lowest wall-clock

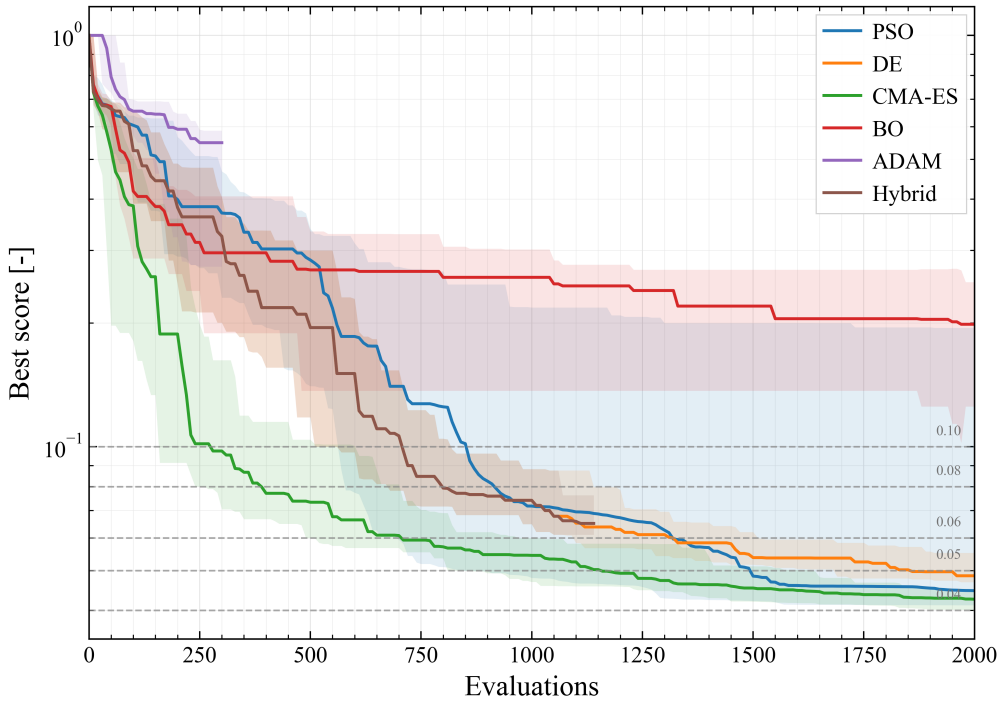


Fig. 3 Convergence comparison: median best-so-far score versus cumulative CRN evaluations (solid lines), with interquartile range (shaded bands). Horizontal dashed lines indicate reference score thresholds.

time among the algorithms.

A non-parametric Kruskal–Wallis test [66] was applied to the 120 final scores to assess whether the six algorithms exhibit statistically distinguishable performance. The test returns $H = 82.86$ with an associated p-value of $2.12 \cdot 10^{-16}$, where H measures the separation among the rank distributions of the algorithms and the p-value quantifies the probability of observing such a separation under the null hypothesis of equal performance. This very small probability therefore leads to rejection of the null hypothesis. Since the Kruskal–Wallis test only indicates the presence of significant differences among groups, post-hoc Mann–Whitney/Wilcoxon rank-sum comparisons [67] were then performed for all 15 algorithm pairs. A Bonferroni correction [68] was applied to control the family-wise significance level at $\alpha_{\mathrm{fw}} = 0.05$, leading to an adjusted significance threshold of $\alpha_{\mathrm{fw}}/15 = 0.0033$. The pairwise comparisons confirm a clear, although not strictly linear, hierarchy. CMA-ES, PSO, and DE form the leading group: PSO is not significantly different from either CMA-ES or DE, whereas CMA-ES significantly outperforms DE ($p_{\mathrm{corr}} = 1.46 \cdot 10^{-4}$), where p_{corr} denotes the multiplicity-adjusted p-value. The Hybrid strategy occupies an intermediate position, significantly outperforming BO and ADAM but remaining inferior to CMA-ES and DE, while BO is significantly better than ADAM. Overall, the statistical evidence identifies CMA-ES, PSO, and DE as the most competitive methods, places Hybrid below this top tier but above BO and ADAM, and indicates ADAM as the weakest optimizer in the present benchmark.

ADAM is included mainly as a diagnostic benchmark for the CRN calibration landscape. The results indicate that this objective is poorly suited to gradient-based optimization: five of the twenty seeds produce no feasible evaluation, and the best finite score ($J = 0.072$) remains above the level routinely reached by CMA-ES, PSO, and DE. The main reason is that finite-difference gradients become unreliable near infeasible regions, where the penalty discontinuity dominates the local sensitivity of NO_x to $\vec{\Theta}$. Combined with the presence of multiple attraction basins, this makes short first-order restarts ineffective. Overall, the landscape appears non-convex, locally rough, and interrupted by broad infeasible plateaus, so gradient information is of limited practical value.

The Hybrid method was intended to combine the global exploration of DE with the local efficiency of L-BFGS-B, but its performance remains clearly below standalone DE, with a median score of $J = 0.065$ versus $J = 0.049$, and only one seed reaching $J < 0.05$. This underperformance is mainly due to the reduced DE budget, limited to about 1000 evaluations, which leaves the global search in only moderately good regions ($J \approx 0.08$–0.10). The subsequent local refinement then inherits the same limitations already observed for gradient-based methods near infeasibility boundaries

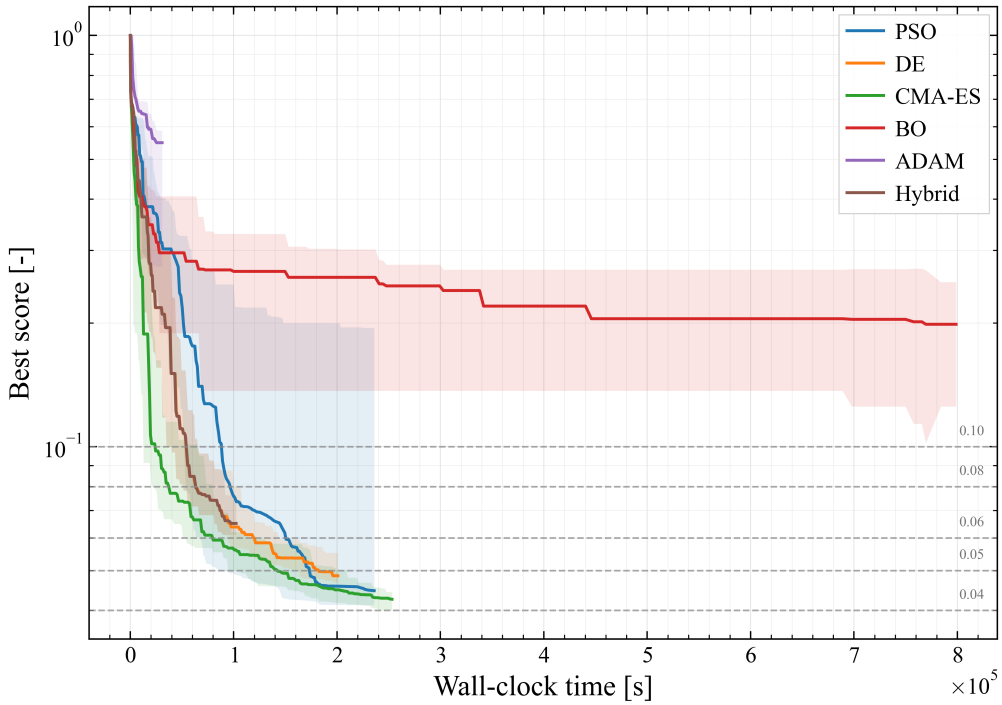


Fig. 4 Convergence comparison: median best-so-far score (solid lines) versus wall-clock time, with interquartile range (shaded bands).

and, in several cases, converges only to nearby local minima.

Figure 5 displays the distribution of the best-found parameter vectors across all 20 seeds, restricted to the three top-performing algorithms (PSO, CMA-ES, and DE) to ensure meaningful comparison.

Several parameters are tightly constrained across all three algorithms and all seeds. The inlet mass-flow ratio $\theta_2 = \dot{m}_{\text{air},1}/\dot{m}_{\text{air},2}$ converges consistently to $\theta_2 \approx 0.20$, indicating that nearly all of the air is routed through the second reservoir (Res_2) regardless of the optimization algorithm. The volume fractions θ_9 ($V_{FL_{AX}}/V_{CC}$) and θ_{10} (V_{OUT}/V_{CC}) similarly cluster near their lower bounds ($\theta_9 \approx 0.05$, $\theta_{10} \approx 0.20$), demonstrating that the best CRN configurations assign the minimum permitted volume to FL_{AX} and OUT, leaving the maximum residual volume to the inner recirculation zone. The IRZ-to-FL_{SW} recirculation fraction θ_6 converges to high values ($\theta_6 \approx 0.93$) for all three algorithms, reflecting the physical dominance of the recirculation feedback from IRZ to the swirling flame zone.

In contrast, the ORZ fraction (θ_3), the FL_{SW}-to-ORZ split (θ_4), the IRZ-to-Z_{AX} fraction (θ_5), and the IRZ-to-FL_{AX} fraction (θ_7) exhibit substantial variability both within and across algorithms. This dispersion suggests that multiple distinct configurations produce comparably low objective values, reflecting compensating effects among the internal mass-flow redistribution pathways.

The best solutions from each algorithm are compared in Table 6, which reports the achieved NO_x concentrations at the four calibration points.

The preceding analysis identifies CMA-ES as the most suitable algorithm for CRN parameter calibration. It achieves the lowest median score, the tightest interquartile range, and the highest success rate (95%). Although PSO attains the best individual calibration score, its substantially wider dispersion and lower success rate make it less reliable for routine use. The parameter vector $\vec{\Theta}^*$ obtained from the best CMA-ES run (seed 4, $J = 0.0387$) is therefore retained as the reference calibration for all subsequent analyses (Table 7).

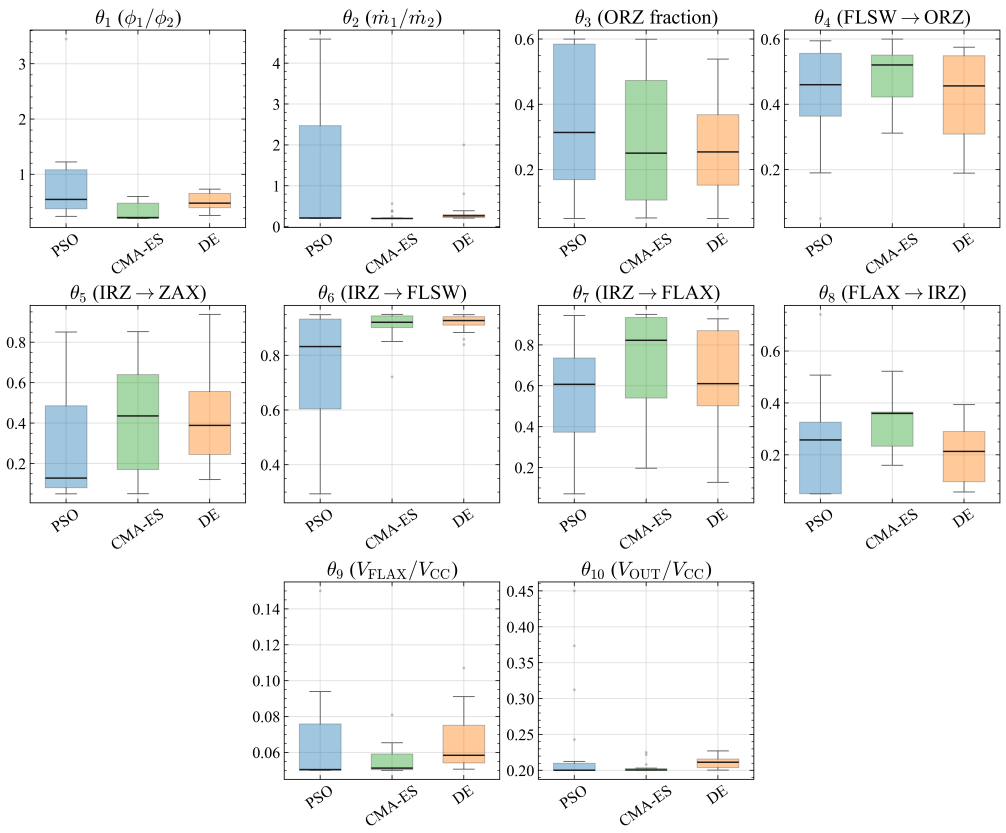


Fig. 5 Distribution of best-found parameter vectors across 20 seeds for the three top-performing algorithms (PSO, CMA-ES, DE).

B. Computational Cost Reduction

1. Validated CRN Baseline

Using the parameter vector $\vec{\Theta}^*$ identified in Sec. II.A, the CRN is evaluated with both the Naik (21 species, 105 reactions) and CRECK (31 species, 203 reactions) mechanisms. For the CRECK mechanism, the corresponding optimal parameter vector was determined through a local CMA-ES optimization in the neighborhood of the Naik-based solution, in accordance with the approach adopted in [18].

The calibration was performed at the reference inlet operating condition ($\dot{m}_{\mathrm{air}}$ = 130 kg/h, T_{air} = 623 K). Therefore, the network must be assessed at withheld equivalence ratios under the same inlet conditions, as well as under entirely different inlet configurations that were not included in the calibration. A detailed analysis of the CRN predictive capability is provided in [18]. The present discussion focuses on establishing the quantitative error baseline and the computational cost that motivate the acceleration strategies of the next subsection.

Figure 6 compares the predicted NO_x emissions (ppmv, dry, at 15% O_2) against the experimental data of Reichel et al. [22, 30] at four operating conditions. At the reference condition (Fig. 6a), both mechanisms reproduce the characteristic exponential increase of NO_x with equivalence ratio. The Naik mechanism achieves relative errors between 0.2% and 24% at the four calibration points, with the largest deviation observed at the leanest calibration point (ϕ = 0.40), where the very low absolute NO_x emissions amplify the corresponding relative error. The CRECK mechanism exhibits larger errors at low equivalence ratios (up to 62% at ϕ = 0.40) but matches the calibration points at higher equivalence ratio within 4%. At the four validation equivalence ratios for this inlet configuration (ϕ = 0.30, 0.50, 0.70, 0.80), the Naik mechanism maintains good agreement. The point at ϕ = 0.80 deviates from the trend exhibited by the remaining data, possibly due to measurement anomalies or localized combustion instabilities during acquisition. This interpretation is consistent with a companion multi-fidelity analysis of the same AHEAD dataset [69], which shows that this operating point is systematically underpredicted not only by the present CRN, but also by the corresponding 2D

Table 6 Experimental NO_x targets and best-run predictions for each optimization algorithm at the four calibration points. NO_x values are reported in ppmv on a dry basis, corrected to 15% O_2.

	J	ϕ=0.40	ϕ=0.60	ϕ=0.87	ϕ=0.96
Exp. targets	–	5.28	26.13	212.07	305.54
PSO	0.0360	4.11	26.24	218.05	300.28
CMA-ES	0.0387	4.01	26.08	218.25	299.87
DE	0.0440	3.85	26.26	220.40	301.28
Hybrid	0.0488	3.85	26.78	218.81	294.31
BO	0.0584	3.64	24.47	211.36	289.21
ADAM	0.0721	2.98	22.96	223.09	305.68

Table 7 Reference parameter vector $\vec{\Theta}^*$ adopted for subsequent analyses using the Naik mechanism, corresponding to the best CMA-ES run (seed 4, J = 0.0387).

Parameter	Optimized value	Parameter	Optimized value
θ_1	0.223	θ_6	0.931
θ_2	0.201	θ_7	0.555
θ_3	0.286	θ_8	0.362
θ_4	0.467	θ_9	0.051
θ_5	0.056	θ_{10}	0.202

RANS and 3D RANS–EARSM simulations. Across the different formulations, the predicted values remain in the range 127–146 ppmv, compared with the measured value of 182.33 ppmv. This indicates that the discrepancy is unlikely to be specific to the present CRN topology or calibration strategy. Excluding this point, the absolute errors remain below approximately 6 ppmv for the Naik mechanism and below approximately 10 ppmv for the CRECK mechanism.

Both mechanisms underpredict NO_x at the leanest equivalence ratios. The same trend is reported in the companion CFD analysis of the AHEAD combustor [69], where both the 2D RANS and the 3D RANS–EARSM simulations underestimate the two leanest measured points at the reference inlet condition. This indicates that the low-ϕ discrepancy is not only a consequence of the lumped, homogeneous CRN description, but reflects a broader difficulty in representing the very lean, low-emission regime with steady, time-averaged models. A plausible explanation is the presence of localized temperature and composition inhomogeneities in the real combustor, which may persist under partially premixed lean hydrogen operation [70] and are smoothed by averaged modeling approaches. Table 8 reports the complete per-point comparison under the reference inlet conditions.

Table 8 NO_x predictions [ppmv] versus experiment at the reference condition ($\dot{m}_{\mathrm{air}}$ = 130 kg/h, T_{air} = 623 K). Relative errors are reported in parentheses.

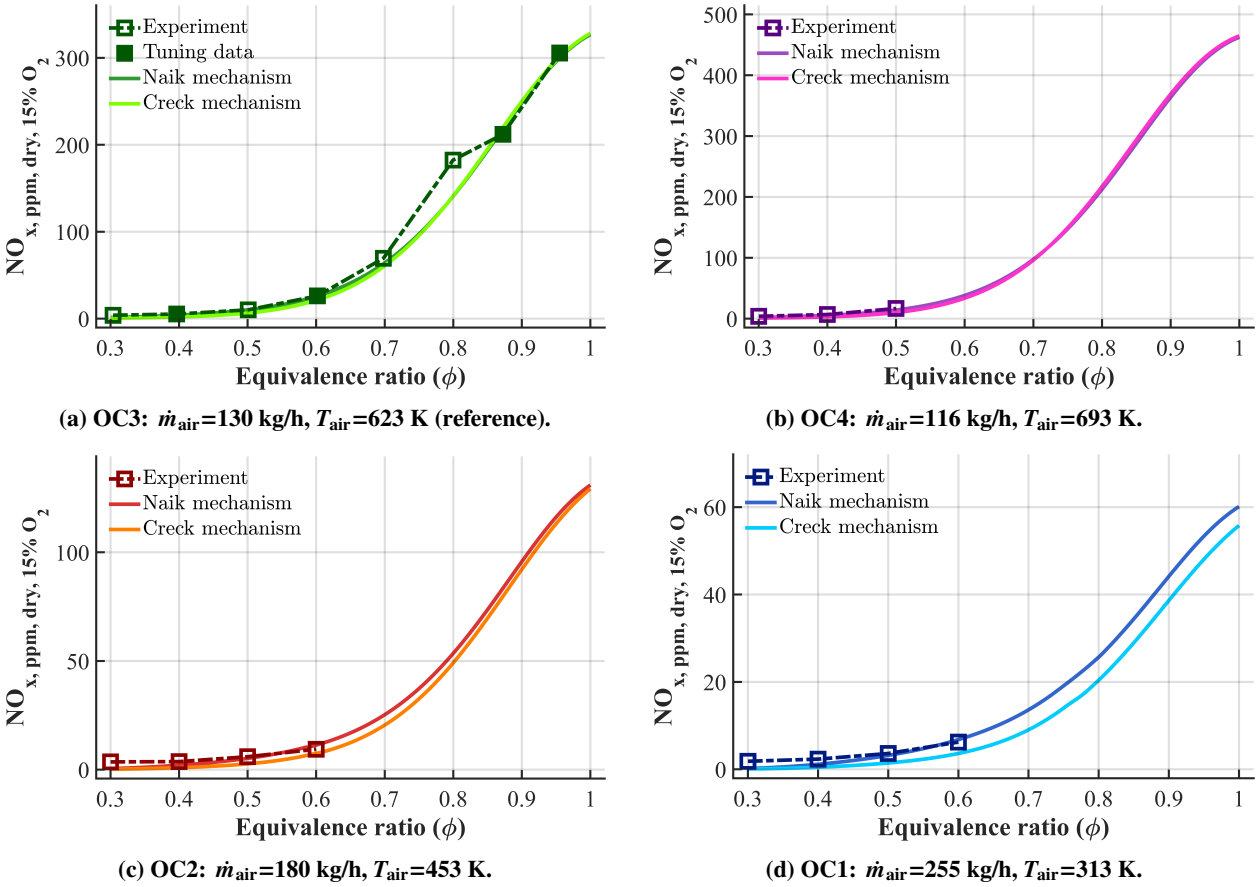
	ϕ	Exp.	Naik	CRECK
Calib.	0.40	5.28	4.01 (24.2%)	2.00 (62.1%)
	0.60	26.13	26.08 (0.2%)	21.47 (17.8%)
	0.87	212.07	218.24 (2.9%)	220.04 (3.8%)
	0.96	305.54	299.86 (1.9%)	301.10 (1.5%)
Valid.	0.30	3.92	1.37 (65.0%)	0.61 (84.4%)
	0.50	10.12	10.16 (0.5%)	6.52 (35.6%)
	0.70	69.40	63.08 (9.1%)	59.77 (13.9%)
	0.80	182.33	140.68 (22.9%)	141.04 (22.7%)

To assess whether the calibrated CRN generalizes beyond the reference inlet conditions, the network is evaluated,

without re-calibration, at three additional operating inlet configurations from the experimental dataset [30]:

- $\dot{m}_{\text{air}}$ = 255 kg/h, T_{air} = 313 K (Fig. 6d);
- $\dot{m}_{\text{air}}$ = 180 kg/h, T_{air} = 453 K (Fig. 6c);
- $\dot{m}_{\text{air}}$ = 116 kg/h, T_{air} = 693 K (Fig. 6b).

Under these operating conditions, the model still underpredicts NO_x emissions at the lowest equivalence ratios (ϕ = 0.30 and ϕ = 0.40), although the absolute errors remain limited, below 3 ppmv for Naik and below 6 ppmv for CRECK. No experimental data are available in these cases at equivalence ratios closer to stoichiometric conditions. Nevertheless, the predicted trends remain physically consistent, as discussed in [18].



(a) OC3: $\dot{m}_{\text{air}}$=130 kg/h, T_{air}=623 K (reference).

(b) OC4: $\dot{m}_{\text{air}}$=116 kg/h, T_{air}=693 K.

(c) OC2: $\dot{m}_{\text{air}}$=180 kg/h, T_{air}=453 K.

(d) OC1: $\dot{m}_{\text{air}}$=255 kg/h, T_{air}=313 K.

Fig. 6 NO_x versus equivalence ratio at four operating conditions. In (a), filled markers denote calibration points and open markers denote validation points. In (b)–(d), the CRN uses the same $\vec{\Theta}^*$ calibrated at OC3 without re-calibration.

The computational cost of this validated baseline provides the reference denominator for all speedup factors reported in Sec. II.B.2 and Sec. II.B.3.

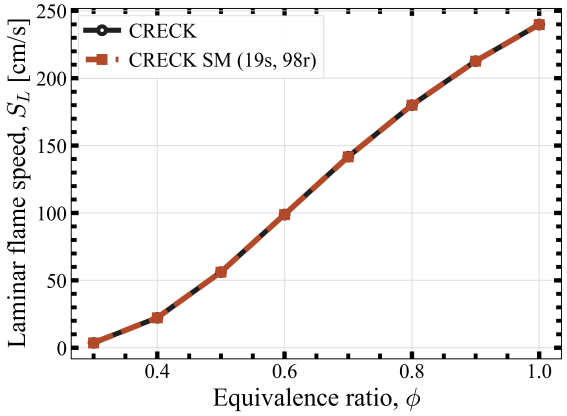
2. Skeletal Mechanism Reduction

The detailed Naik (21 species, 105 reactions) and CRECK (31 species, 203 reactions) mechanisms underlying the validated baseline of Sec. II.B.1 are here reduced through the DRGEP pipeline described in Sec. I.C.1, with the aim of retaining the accuracy of the detailed mechanisms while lowering the per-evaluation cost.

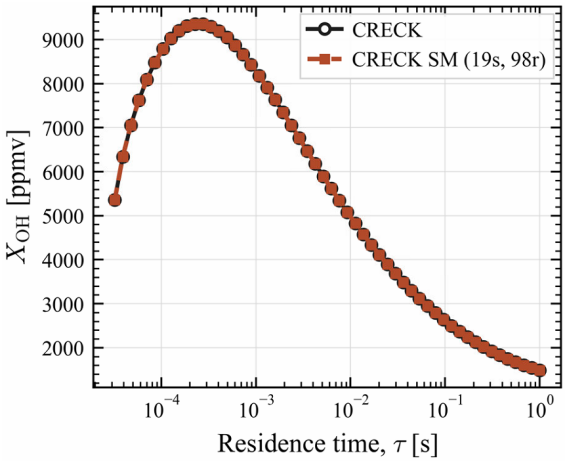
The sampling pool used to populate the interaction matrix was generated from a structured set of reactor calculations, consisting of 475 constant-pressure autoignition trajectories and 2375 PSR steady-state solutions computed over five residence times. The conditions span $\phi \in [0.2, 2.0]$ and $T \in [300, 2700]$ K at atmospheric pressure, thereby covering both the nominal operating envelope of the combustor and the additional thermochemical states that may arise during the transient time integration.

Starting from the thermochemical database described above, the DRGEP procedure was applied by progressively increasing the threshold ε, so as to generate a controlled sequence of increasingly compact mechanisms. This threshold sweep induces an almost species-by-species removal from the detailed parent down to mechanisms small enough to break the thermal and N_2O-intermediate pathways. The resulting reduction staircase consists of twelve candidate skeletal CRECK mechanisms, ranging from 29 down to 16 species, and six candidate skeletal Naik mechanisms, ranging from 20 down to 14 species. For the CRECK family, the first two species eliminated are the inert bath gases Ar and He, whose removal has no kinetic consequence. At subsequent thresholds, the reduction successively removes $HONO_2$, NO_3, HNO_2, N_2H_4, NH_2OH, HNOH, H_2NO, HONO, N_2H_2, and N_2H_3. The reduction reaches its most compact admissible configuration at the 19-species, 98-reaction skeletal mechanism. Further increasing ε removes H_2O_2, which causes an abrupt decrease of the heat-release ratio to $Q_{\text{skel}}/Q_{\text{det}} = 0.45$. These mechanisms were therefore discarded and not used in the subsequent analyses.

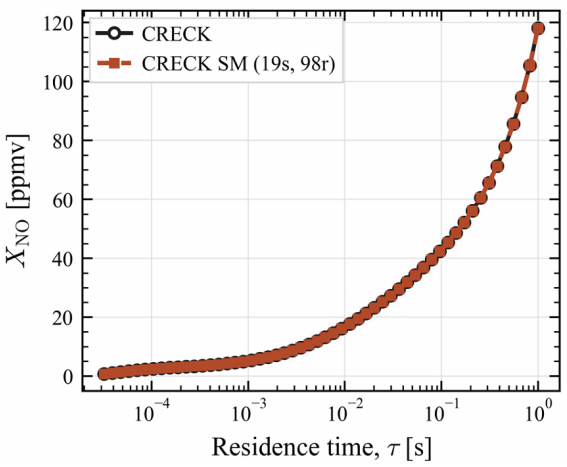
An example of the evaluation metrics used in this selection process is reported in Fig. 7. This compares the detailed CRECK mechanism with the selected 19-species skeletal mechanism in terms of laminar flame speed as a function of equivalence ratio (Fig. 7a) and steady-state PSR outlet mole fractions of OH, NO, and NO_2 as a function of residence time at $\phi = 0.60$ (Figs. 7b–7d). The near-perfect overlap of the curves shows that the selected skeletal mechanism accurately reproduces the parent mechanism with regard to quantities directly related to flame structure and NO_x formation.



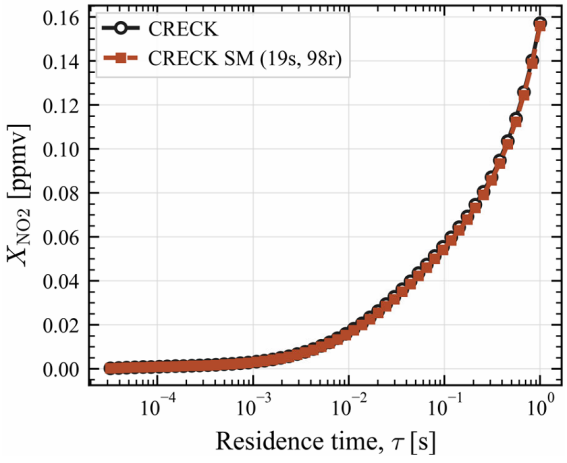
(a) Laminar flame speed as a function of equivalence ratio.



(b) PSR OH mole-fraction profile at $\phi = 0.60$.



(c) PSR NO mole-fraction profile at $\phi = 0.60$.



(d) PSR NO_2 mole-fraction profile at $\phi = 0.60$.

Fig. 7 Comparison between the detailed CRECK mechanism and the selected 19-species, 98-reaction skeletal mechanism for representative validation quantities.

For the Naik family, the reduction limit is reached at a less aggressive threshold, as expected from the already compact size of the parent mechanism. The 20-species mechanism removes $HONO_2$ with negligible impact on the evaluation metrics. The 18-species skeletal mechanism further removes NH_2 and NH_3 without degrading the target quantities. In contrast, the next reduction step requires the removal of atomic nitrogen N, a key intermediate in the extended Zeldovich thermal-NO pathway [5]. This interrupts the corresponding NO-formation branch and produces a 93.7% error on the peak NO mole fraction, together with a non-physical heat-release ratio.

Overall, the threshold sweep generates 18 candidate skeletal mechanisms. Of these, 11 exhibit high accuracy across the validation metrics reported in Table 3: nine CRECK skeletons with 19–29 species and two Naik skeletons with 18 and 20 species. The evaluation metrics of these mechanisms are summarized in Table 9. Across the accepted subset, ignition delay times (τ_{ign}), PSR extinction residence time (τ_{ext}), and equilibrium NO_x concentrations (NO_x^{eq}) are matched exactly. The laminar flame speed error (ε_{S_L}) remains below 2.85% for the Naik skeletons and below 7.16% for the CRECK skeletons, with the largest discrepancy occurring at the leanest condition ($\phi = 0.3$), where the absolute flame speed is smallest. Peak species mole-fraction and production-rate errors ($\varepsilon_{X_{\text{peak}}}$ and $\varepsilon_{\dot{\omega}_{\text{peak}}}$ respectively) remain below 6% in all cases, and the integrated heat release (Q) is preserved within 0.06% throughout.

Table 9 Values of the validation metrics defined in Table 3 for all skeletal mechanisms retained as valid. N_{sp} and N_{rx} denote the number of species and reactions, respectively. Errors are reported as maximum relative deviations with respect to the detailed parent. The heat-release ratio is reported as $Q_{\text{skel}}/Q_{\text{det}}$ across all autoignition and PSR validation conditions.

Skeleton	N_{sp}	N_{rx}	$\varepsilon_{\tau_{\text{ign}}}$ [%]	ε_{S_L} [%]	$\varepsilon_{\tau_{\text{ext}}}$ [%]	$\varepsilon_{NO_x^{\text{eq}}}$ [%]	$\varepsilon_{X_{\text{peak}}}$ [%]	$\varepsilon_{\dot{\omega}_{\text{peak}}}$ [%]	Q_{ratio} [–]
Naik_SM_20s_101r	20	101	0.00	2.85	0.00	0.00	0.07	0.41	0.9997
Naik_SM_18s_79r	18	79	0.00	2.85	0.00	0.00	2.79	4.08	0.9998
CRECK_SM_29s_203r	29	203	0.00	7.16	0.00	0.00	0.07	0.55	1.0000
CRECK_SM_27s_192r	27	192	0.00	7.16	0.00	0.00	0.06	1.15	0.9999
CRECK_SM_26s_180r	26	180	0.00	7.16	0.00	0.00	0.27	1.46	0.9996
CRECK_SM_25s_170r	25	170	0.00	7.16	0.00	0.00	0.27	1.79	0.9999
CRECK_SM_24s_156r	24	156	0.00	7.16	0.00	0.00	0.27	2.14	0.9997
CRECK_SM_23s_143r	23	143	0.00	7.16	0.00	0.00	0.27	2.49	0.9994
CRECK_SM_21s_119r	21	119	0.00	7.16	0.00	0.00	5.58	4.74	0.9954
CRECK_SM_20s_110r	20	110	0.00	0.03	0.00	0.00	5.58	4.73	0.9997
CRECK_SM_19s_98r	19	98	0.00	0.17	0.00	0.00	5.58	4.74	0.9999

The previous analysis verifies the behavior of the skeletal mechanisms in isolated canonical configurations, but an additional check is required before using them in the calibrated CRN, where the chemistry is exposed to coupled reactor interactions and recirculation-driven transient states. Each accepted skeleton was therefore inserted into the full CRN with the reference parameter vector $\vec{\Theta}^*$, without recalibration, and evaluated over the equivalence-ratio range $\phi \in [0.30, 1.00]$ at the reference inlet condition, $\dot{m}_{\text{air}} = 130$ kg/h and $T_{\text{air}} = 623$ K.

The error statistics for the skeletal mechanisms are reported in Table 10. For the CRECK family, the full hierarchy of nine accepted skeletons reproduces the NO_x prediction of the 31-species parent with a maximum pointwise relative error of 0.05% and a maximum absolute deviation below 0.02 ppmv across the whole range of equivalence ratio. The 19-species skeleton, located at the edge of the reduction cliff, matches the parent mechanism with the same accuracy as the larger 29-species skeleton. For the Naik family, the 20-species skeleton reproduces the parent mechanism within 0.005% and 0.002 ppmv. The more aggressive 18-species skeleton shows a slightly larger deviation, with a maximum relative NO_x error of 0.71%. This discrepancy is concentrated at the leanest point, $\phi = 0.30$, where the low absolute NO_x levels amplify the relative error. Across both mechanism families, the pointwise maxima of the species and production-rate errors for NO, NO_2, and OH evaluated at the OUT reactor remain below 7% for all retained skeletons, while deviations in the volumetric heat-release ratio remain below 10^{-4}. The agreement observed within the full CRN therefore confirms that the retained mechanisms preserve both the relevant nitrogen chemistry and the overall heat-release behavior under coupled network conditions.

The computational benefit of the reduction was quantified by timing the 40-point ϕ sweep with the detailed parent

and with each accepted skeleton on a single core of a machine equipped with a 13th Gen Intel(R) Core(TM) i7-13700 processor operating at 2.10 GHz. All simulations were performed using the same CRN module, initial conditions, and simulation settings. The resulting wall-clock times and speedup factors are collected in Table 10.

For the Naik family, the 20-species skeleton reduces the sweep runtime from 708 s to 530 s (1.34 speedup), while the 18-species skeleton drops it further to 438 s (1.62 speedup). The modest gain reflects the already compact size of the detailed parent mechanism. For the CRECK family, the speedup increases from ~ 1.0 at 29 species down to 2.06 at 19 species, where the sweep time drops from 1112 s (detailed) to 540 s.

Overall, the most aggressively reduced mechanisms: the 19-species, 98-reaction CRECK skeleton and the 18-species, 79-reaction Naik skeleton, provide the largest cost reduction while keeping the maximum CRN-level NO_x error below 0.05% and 0.8%, respectively.

Table 10 NO_x errors and computational cost of the accepted skeletal mechanisms relative to the detailed parent. Wall-clock times refer to a 40-point ϕ sweep over [0.30, 1.00]. $\varepsilon_{NO_x}^{max}$ and $\overline{\varepsilon}_{NO_x}$ denote the maximum and mean pointwise relative NO_x errors across the sweep.

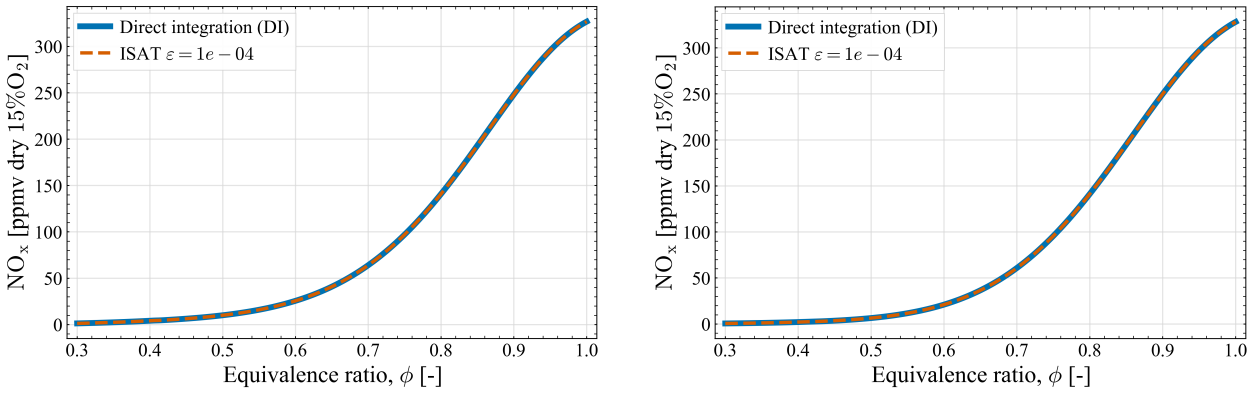
Skeleton	N_{sp}	N_{rx}	$\varepsilon_{NO_x}^{max}$ [%]	$\overline{\varepsilon}_{NO_x}$ [%]	Runtime [s]	Speedup [–]
Naik (parent)	21	105	–	–	708	1.00
Naik_SM_20s_101r	20	101	0.005	0.000	530	1.34
Naik_SM_18s_79r	18	79	0.706	0.421	438	1.62
CRECK (parent)	31	203	–	–	1112	1.00
CRECK_SM_29s_203r	29	203	0.003	0.000	1115	1.00
CRECK_SM_27s_192r	27	192	0.004	0.000	879	1.27
CRECK_SM_26s_180r	26	180	0.004	0.001	796	1.40
CRECK_SM_25s_170r	25	170	0.004	0.000	870	1.28
CRECK_SM_24s_156r	24	156	0.004	0.000	810	1.37
CRECK_SM_23s_143r	23	143	0.007	0.001	769	1.45
CRECK_SM_21s_119r	21	119	0.047	0.009	654	1.70
CRECK_SM_20s_110r	20	110	0.047	0.009	593	1.88
CRECK_SM_19s_98r	19	98	0.047	0.009	540	2.06

3. Flow-Map In Situ Adaptive Tabulation

The Flow-Map ISAT scheme introduced in Sec. I.C.2 is assessed on the same calibrated CRN configuration used for the preceding analysis, retaining the reference inlet condition ($\dot{m}_{air}$ = 130 kg/h, T_{air} = 623 K) and the reference parameter vector $\vec{\Theta}^*$ of Table 7. The ISAT simulations were executed on a single CPU core of a machine equipped with a 13th Gen Intel(R) Core(TM) i7-13700 processor operating at 2.10 GHz.

For this assessment, 100 simulations are performed over the equivalence-ratio range $\phi \in [0.30, 1.00]$. For each mechanism the benchmark is repeated at five error tolerances $\varepsilon_{tol} \in \{5 \cdot 10^{-3}, 10^{-3}, 5 \cdot 10^{-4}, 10^{-4}, 10^{-5}\}$ for a total of five ISAT cases per mechanism and one thousand individual simulations. Each case is run with a 200-step direct-integration warm-up phase, after which the macro-stepping schedule described in Sec. I.C.2 takes over. A single ISAT table is shared across the entire ϕ sweep so that the structure populated at one equivalence ratio remains available at the next.

Figure 8 reports the direct comparison between the DI predictions and the ISAT-accelerated predictions at the reference tolerance $\varepsilon_{tol} = 10^{-4}$. For both mechanisms, the dashed ISAT curves closely overlap the DI reference over the entire ϕ range. Quantitatively, for the Naik mechanism, the mean relative NO_x error is $2.0 \cdot 10^{-4}$ %, and the maximum relative error over any individual ϕ is $3.0 \cdot 10^{-3}$ %. The absolute temperature error in the OUT reactor, evaluated with respect to the corresponding DI-based CRN prediction, never exceeds $8 \cdot 10^{-3}$ K. For the CRECK sweep, the errors are of the same order: a mean relative error of $4.4 \cdot 10^{-4}$ %, a maximum relative error of $4.0 \cdot 10^{-3}$ %, and $|\Delta T_{OUT}| \leq 7.5 \cdot 10^{-3}$ K. These results show that, at this tolerance, the ISAT formulation reproduces the DI-based CRN predictions without appreciable loss of accuracy.



(a) Naik mechanism. **(b) CRECK mechanism.**

Fig. 8 NO_x emission as a function of equivalence ratio at $\dot{m}_{air}$ = 130 kg/h and T_{air} = 623 K, predicted by direct integration and by the Flow-Map ISAT formulation at $\varepsilon_{tol} = 10^{-4}$.

The very small errors obtained with zeroth-order retrieval justify its adoption for the present flow-map ISAT implementation. This choice simplifies the algorithm substantially, since the mapping-gradient matrix does not need to be computed, stored, or applied during table lookup. At the same time, eliminating the gradient evaluation provides an additional computational advantage with respect to a first-order ISAT formulation, while preserving the accuracy required for the NO_x predictions considered in this work.

The analysis is then extended by comparing all the ISAT tolerances considered in this work. As expected, relaxing ε_{tol} increases the mean speedup, since a larger fraction of the queried states can be recovered from the ISAT table. The mean relative NO_x error increases only slightly, or remains nearly constant as the tolerance becomes less restrictive, although it remains very small over the whole range investigated. This behavior is illustrated in Fig. 9, which reports the results obtained with the CRECK mechanism: the left panel shows the mean relative NO_x error, while the right panel gives the mean speedup with respect to direct integration. The complete set of results for both the Naik and CRECK mechanisms is reported in Table 11.

For the Naik mechanism, the mean relative NO_x error remains nearly constant as the ISAT tolerance becomes less restrictive, remaining on the order of $2.0 \cdot 10^{-4}$ % for all tolerances. In contrast, over the tolerance range, the mean speedup increases markedly, from 7.12 to 17.08, with an average ISAT simulation time of approximately 0.92 s at the least restrictive tolerance, $\varepsilon_{tol} = 5 \cdot 10^{-3}$. The maximum pointwise NO_x error remains below $3.0 \cdot 10^{-3}$ % for all tolerances. For the CRECK mechanism, the mean relative NO_x error shows a slight increase as the tolerance is relaxed, rising from $3.7 \cdot 10^{-4}$ % at $\varepsilon_{tol} = 10^{-5}$ to $5.7 \cdot 10^{-4}$ % at $\varepsilon_{tol} = 5 \cdot 10^{-3}$. The corresponding mean speedup increases from 7.00 to 24.88, with an average ISAT simulation time of approximately 1.58 s at the least restrictive tolerance, $\varepsilon_{tol} = 5 \cdot 10^{-3}$. The maximum pointwise NO_x error remains below $5.9 \cdot 10^{-3}$ % throughout the sweep. The retrieve fraction and the number of ISAT leaves follow the expected trend with ε_{tol}. For both mechanisms, relaxing the tolerance increases the cumulative retrieve fraction, from 0.35 to 0.87 for Naik and from 0.33 to 0.87 for CRECK, indicating a progressively larger use of table lookups instead of direct integration. At the same time, the number of leaves decreases markedly: from 60009 to 11758 for Naik and from 63800 to 12008 for CRECK. Thus, looser tolerances produce more compact tables and higher retrieval rates, which explains the increase in speedup reported in Table 11.

The weak dependence of the NO_x error on ε_{tol} may be explained by the structure of the present analysis. Only the equivalence ratio is varied, while the inlet condition and the calibrated parameter vector $\vec{\Theta}^*$ are kept fixed. Consequently, after the direct-integration warm-up phase, the CRN is likely to visit recurrent and relatively regular thermochemical regions. In this situation, relaxing the tolerance mainly increases the number of successful table retrievals within regions that are already well represented by the populated ISAT table, rather than introducing substantially different thermochemical mappings. This explains why the computational speedup increases while the NO_x error remains nearly unchanged.

The consistently higher speedup of the CRECK family, at a fixed tolerance and retrieve fraction, is a direct consequence of the larger chemical system: a retrieve that bypasses a CRECK integration saves more wall-clock time than one that bypasses a Naik integration. On the other hand, the cost of the binary-tree traversal and the ellipsoid-of-accuracy test is primarily influenced by the network state dimension $n = 6(1 + N_s)$, which only increases

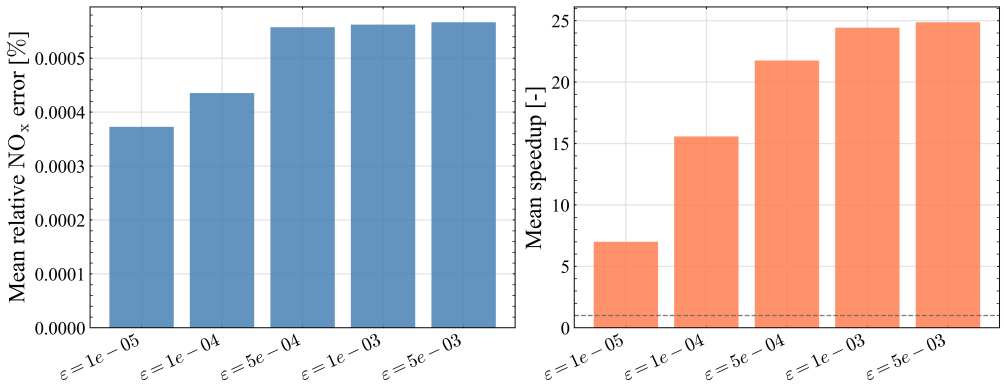


Fig. 9 Effect of the ISAT tolerance ε_{tol} on NO$_x$ accuracy and computational speedup for the CRECK mechanism. The left panel reports the mean relative NO$_x$ error, while the right panel shows the mean speedup with respect to direct integration.

Table 11 Flow-Map ISAT performance over 100 equivalence-ratio points in the range $\phi \in [0.30, 1.00]$. $\bar{\varepsilon}_{\text{NO}_x}$ and $\varepsilon^{\text{max}}_{\text{NO}_x}$ are the mean and maximum relative NO$_x$ errors, $\overline{S}$ is the mean per-ϕ speedup relative to DI, f_{r} is the cumulative retrieve fraction at the end of the calculation, N_ℓ is the total number of leaves populated across all sub-tables, and M_{tab} is the final ISAT table memory footprint.

Mechanism	ε_{tol}	$\bar{\varepsilon}_{\text{NO}_x}$ [%]	$\varepsilon^{\text{max}}_{\text{NO}_x}$ [%]	$\overline{S}$ [–]	f_{r} [–]	N_ℓ [–]	M_{tab} [MB]
Naik	$5 \cdot 10^{-3}$	$2.0 \cdot 10^{-4}$	$3.0 \cdot 10^{-3}$	17.08	0.87	11758	1597.4
Naik	10^{-3}	$2.0 \cdot 10^{-4}$	$3.0 \cdot 10^{-3}$	16.66	0.84	14678	1994.4
Naik	$5 \cdot 10^{-4}$	$1.9 \cdot 10^{-4}$	$3.0 \cdot 10^{-3}$	14.62	0.76	22263	3025.7
Naik	10^{-4}	$2.0 \cdot 10^{-4}$	$3.0 \cdot 10^{-3}$	11.07	0.62	35689	4851.2
Naik	10^{-5}	$2.3 \cdot 10^{-4}$	$3.0 \cdot 10^{-3}$	7.12	0.35	60009	8157.8
CRECK	$5 \cdot 10^{-3}$	$5.7 \cdot 10^{-4}$	$5.9 \cdot 10^{-3}$	24.88	0.87	12008	3428.3
CRECK	10^{-3}	$5.6 \cdot 10^{-4}$	$5.9 \cdot 10^{-3}$	24.44	0.86	12497	3568.0
CRECK	$5 \cdot 10^{-4}$	$5.6 \cdot 10^{-4}$	$5.9 \cdot 10^{-3}$	21.75	0.65	20200	5768.4
CRECK	10^{-4}	$4.4 \cdot 10^{-4}$	$4.0 \cdot 10^{-3}$	15.57	0.48	37995	10851.5
CRECK	10^{-5}	$3.7 \cdot 10^{-4}$	$4.5 \cdot 10^{-3}$	7.00	0.33	63800	18222.8

mildly between $n = 132$ and $n = 192$.

The memory footprint of the ISAT table was also monitored for each tolerance level. In all cases, the table was shared across the full set of 100 equivalence-ratio points and was not reset between consecutive simulations. The reported memory therefore corresponds to the final table state, when the tabulation is at its maximum size. As expected, the memory requirement increases as ε_{tol} becomes more restrictive, following the growth in the number of populated leaves. For the Naik mechanism, the memory footprint ranges from 1597.4 MB at $\varepsilon_{\text{tol}} = 5 \cdot 10^{-3}$ to 8157.8 MB at $\varepsilon_{\text{tol}} = 10^{-5}$. For CRECK, the corresponding range is larger, from 3428.3 MB to 18222.8 MB, reflecting the larger thermochemical state vector and the higher cost of storing each tabulated flow-map entry. The complete memory data are included in Table 11.

4. Combined Acceleration: CMA-ES Calibration with Skeletal Mechanism and Flow-Map ISAT

The previous sections showed that both skeletal mechanism reduction and Flow-Map ISAT can reduce the cost of CRN evaluations while preserving the predicted NO$_x$ emissions with respect to the direct-integration reference. The next step is to verify whether this accuracy–efficiency trade-off is maintained during the most computationally demanding CRN workflow: the CMA-ES calibration of the parameter vector $\vec{\Theta}$.

This assessment is carried out using the Naik mechanism only, since the corresponding detailed, direct-integration CMA-ES campaign is available from Sec. II.A and provides the baseline against which the accelerated variants can be compared. The 20-seed CMA-ES campaign is therefore repeated in three additional configurations, changing only the chemical-kinetics evaluation through mechanism reduction and/or tabulation. The objective function (Eq. 5), parameter bounds (Table 2), CMA-ES settings (Table 1), and random seeds are kept unchanged.

The four campaigns are denoted as follows:

- Baseline: using the detailed Naik mechanism with direct integration.
- Skeletal: using the skeletal Naik mechanism with 18 species and 79 reactions with direct integration.
- ISAT: using the detailed Naik mechanism with zeroth-order Flow-Map ISAT at $\varepsilon_{\mathrm{tol}} = 10^{-4}$.
- ISAT + Skeletal: using the skeletal Naik mechanism with 18 species and 79 reactions with the same ISAT configuration.

All campaigns are executed using one CPU core per seed on a computing node equipped with Intel(R) Xeon(R) Gold 5318Y CPUs operating at 2.10 GHz. This paired design isolates the effect of the chemical-kinetics evaluation, so that differences in calibration quality or wall-clock cost can be attributed to the reduced chemistry and tabulation strategies rather than to changes in the optimization setup.

Figure 10 reports the median best-so-far score as a function of the number of CRN evaluations for the four campaigns, together with the corresponding interquartile ranges. The convergence histories are similar when measured in evaluation count: by the end of the common budget of 2000 evaluations, the median best score is comparable for all four configurations and remains close to $J \simeq 0.042$. The interquartile range is also narrow in all cases and, at the end of the optimization budget, remains entirely below the $J = 0.05$ threshold for each configuration. Interestingly, the accelerated configurations reach the target threshold slightly earlier than the baseline in terms of median trajectory. The baseline campaign crosses $J = 0.05$ after approximately 1150 evaluations, whereas the Skeletal and ISAT campaigns reach the same level after approximately 1050 and 1000 evaluations, respectively. The combined ISAT + Skeletal configuration shows the earliest crossing, at about 770 evaluations.

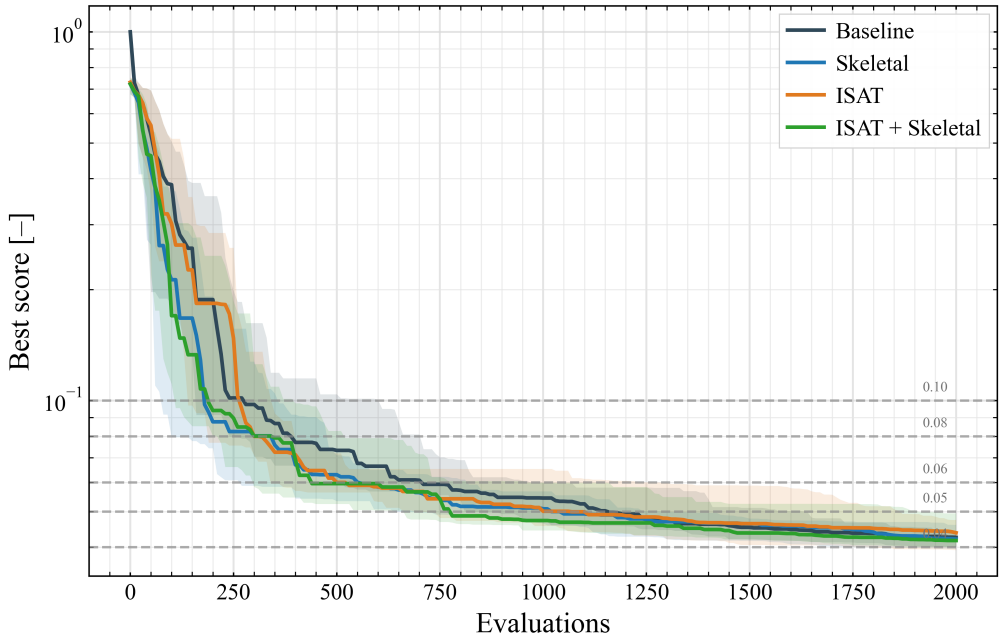


Fig. 10 Median best-so-far score (solid lines) as a function of the number of CRN evaluations for the four CMA-ES calibration campaigns. Shaded bands denote interquartile ranges over 20 seeds.

The statistical analysis confirms this equivalence. A Kruskal–Wallis test on the four 20-seed score distributions yields $H = 0.028$ with $p = 0.999$, and no pairwise Mann–Whitney comparison with Bonferroni correction detects any significant difference ($p_{\mathrm{corr}} = 1.00$ for all six pairs). In terms of solution quality, the three accelerated campaigns are therefore statistically equivalent to the detailed-mechanism, direct-integration baseline.

The main difference emerges when the same convergence histories of Fig. 10 are replotted against wall-clock time,

as shown in Fig. 11. In this representation, the computational advantage of the accelerated configurations becomes evident. The baseline campaign requires a median time of $2.53 \cdot 10^5$ s, corresponding to 70.37 h, to complete the full evaluation budget. The Skeletal configuration reduces this time to $1.18 \cdot 10^5$ s (32.85 h), which is less than half of the baseline cost. When ISAT is introduced, the computational time decreases by approximately one order of magnitude, reaching $2.13 \cdot 10^4$ s (5.90 h) for the ISAT campaign. The lowest cost is obtained by combining ISAT with the skeletal mechanism, for which the median wall-clock time is $1.73 \cdot 10^4$ s, corresponding to 4.79 h.

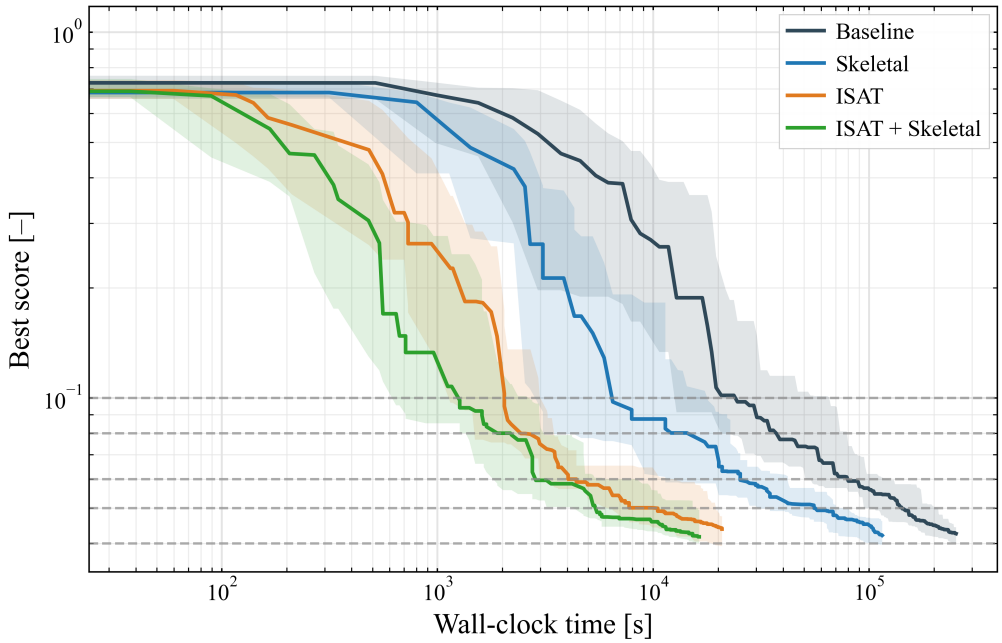


Fig. 11 Median best-so-far score (solid lines) as a function of wall-clock time for the four CMA-ES calibration campaigns. Shaded bands denote interquartile ranges over 20 seeds.

The corresponding speedup distributions with respect to the baseline are reported in Fig. 12. For each accelerated configuration, the boxes denote the interquartile range of the speedup, while the horizontal bars indicate the corresponding median values. Each marker represents the speedup obtained for a single seed with respect to the corresponding baseline run with the same seed. The Skeletal configuration provides a median speedup of approximately 2.1, while ISAT increases the median speedup to 11.6. The combined ISAT + Skeletal configuration reaches the highest acceleration, with a median speedup of 15.1. This combined gain is lower than the product of the two individual speedups because the two methods act partly on the same computational bottleneck. In particular, once the skeletal mechanism has reduced the cost of each direct integration, every ISAT retrieval avoids a cheaper calculation than in the detailed-mechanism case. At the same time, the fixed overhead associated with ISAT, including macro-stepping, table search, and accuracy checks, becomes relatively more important. Even so, the two strategies remain complementary: adding the skeletal mechanism on top of ISAT still increases the median seed-paired speedup by a factor of 1.30, while preserving comparable median calibration accuracy.

The primary performance metrics are summarized in Table 12. The success rate, defined as the fraction of seeds reaching $J < 0.05$, decreases slightly from 95% for the baseline to 90% for the Skeletal campaign, and then to 85% for the ISAT campaign. The combined ISAT + Skeletal configuration exhibits a more pronounced reduction, with a success rate of 75%. This trend suggests that the accelerated formulations may slightly alter the local structure of the objective function. Small changes in local minima, mild irregularities, or weak numerical perturbations can make some optimization trajectories less favorable within the fixed evaluation budget. This effect becomes more evident when the two acceleration strategies are combined, leading to the lower success rate observed for the ISAT + Skeletal campaign.

C. Uncertainty Quantification of NOx Predictions

The block-based PCE pipeline of Sec. I.D was executed on the calibrated CRN at the reference inlet operating condition, using the reference vector $\vec{\Theta}^*$ and the Naik mechanism. The analysis was performed at the experimental

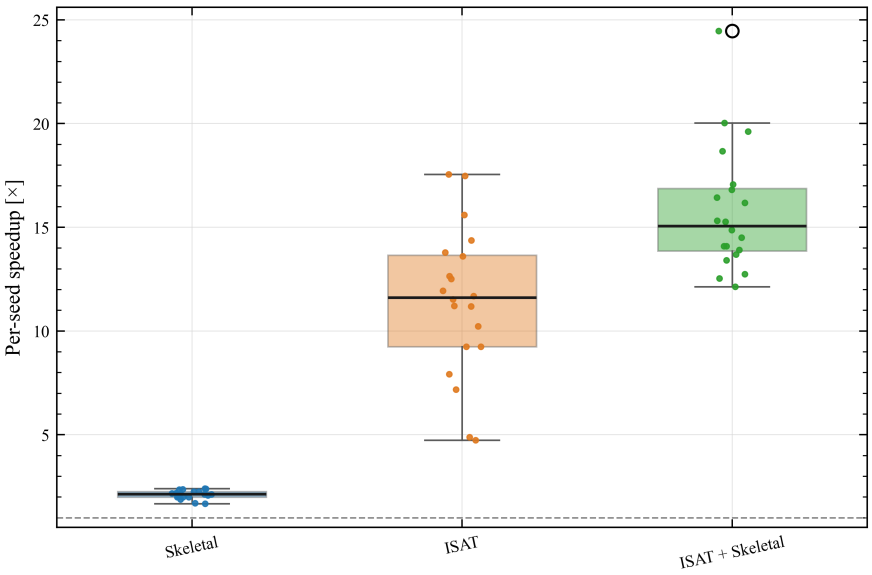


Fig. 12 Seed-paired speedup distributions of the accelerated CMA-ES campaigns relative to the direct-integration baseline. Boxes denote interquartile ranges, horizontal bars denote median values, and markers represent individual seeds.

Table 12 Primary performance metrics of the four CMA-ES calibration campaigns. The speedup column reports the median seed-paired ratio $T_{\text{Baseline}}/T_{\text{campaign}}$.

Campaign	Median J	IQR J	Success	Median time [s]	Speedup
Baseline	0.0424	[0.0400, 0.0443]	95%	253345	1.00 (ref.)
Skeletal	0.0421	[0.0397, 0.0460]	90%	118262	2.13
ISAT	0.0438	[0.0393, 0.0474]	85%	21252	11.61
ISAT + Skeletal	0.0417	[0.0401, 0.0495]	75%	17253	15.07

equivalence-ratio set points $\phi_{\text{ref}} \in \{0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.87, 0.96\}$. The full campaign comprises 14000 direct CRN evaluations, distributed across the four uncertainty blocks and the external validation sets associated with Blocks C and T. The simulations were parallelized over 48 CPU cores on a computing node equipped with Intel(R) Xeon(R) Gold 5318Y CPUs operating at 2.10 GHz, and were completed in approximately 3.5 h.

Block O isolates the propagation of operating uncertainty through the calibrated network. The PCE mean reproduces the deterministic CRN curve of Fig. 6a within 0.5 ppmv at every ϕ_{ref}, while the standard deviation grows with the equivalence ratio up to a peak of $\sigma_O = 22.9$ ppmv at $\phi = 0.87$, before decreasing to $\sigma_O = 19.4$ ppmv at $\phi = 0.96$ (Fig. 14a). This non-monotonic behavior may reflect the saturation of NO_x emissions at near-stoichiometric conditions, where the response to small flow-rate perturbations becomes weaker. The Sobol decomposition attributes the variance almost exclusively to the two flow-rate perturbations: $\langle S_T \rangle_\phi$ is 0.537 for δ_a, 0.420 for δ_f, and 0.047 for T_{air}, where δ_a and δ_f denote the relative deviations of the air and fuel mass-flow rates, respectively. Their dominant contribution is expected, since both perturbations act directly on the equivalence ratio. The T_{air} contribution exhibits a clear U-shape, with $S_T(T_{\text{air}}) = 0.081$ at $\phi = 0.30$, dropping to a minimum of 0.028 at $\phi = 0.70$, and rising back to 0.084 at $\phi = 0.96$. The non-monotonic contribution of T_{air} indicates that the sensitivity to inlet-temperature uncertainty depends on the local operating point, although its overall contribution remains secondary compared with the flow-rate perturbations. Across the full sweep, S_{main} and S_T for each input agree to within the third decimal, indicating that interaction effects among the operating perturbations are negligible and that the network response is essentially additive. The internal validation protocol of Sec. I.D is satisfied for Block O at all set points. The leave-one-out predictive coefficient remains above $Q^2_{\text{LOO}} = 0.994$, indicating strong out-of-sample consistency, while the order-three Sobol drift stays below $3.3 \cdot 10^{-3}$, confirming the stability of the sensitivity indices with respect to the lower-order expansion.

Block C addresses the local sensitivity to the calibration parameters. As anticipated in Sec. I.D, a preliminary correlation analysis was performed on the ensemble of 19 acceptable CMA-ES seeds ($J < 0.05$) using Spearman's rank coefficient and Pearson's product-moment coefficient [59, 60]. Spearman's coefficient, ρ_S, measures monotonic association between ranked variables, whereas Pearson's coefficient, r_P, measures linear correlation between the parameter values. The associated p-values, p_S and p_P, test the null hypothesis of no correlation. The analysis identified a single strongly significant pair, (θ_1, θ_8), with $\rho_S = -0.78$ ($p_S = 9.6 \cdot 10^{-5}$) and $r_P = -0.91$ ($p_P = 7.1 \cdot 10^{-8}$). Following the mitigation strategy declared in Sec. I.D, θ_8 was held at its reference value θ_8^* and the local PCE was constructed under the independence assumption on the remaining seven calibration parameters $(\theta_1, \ldots, \theta_7)$. The maximum residual Spearman correlation among the active set drops to $\rho_S < 0.50$. Given the local nature of the present ± 0.025 perturbation window, this residual correlation was considered sufficiently moderate to retain the independence assumption.

The screening PCE on the seven-dimensional active set (order $p_o = 2$, $N = 200$ Sobol samples, basis size 36) produces the parameter ranking summarized in Fig. 13. The aggregated total-effect score is dominated by θ_2 ($\langle S_T \rangle_\phi = 0.362$) and θ_6 (0.336), followed by θ_1 (0.153), θ_5 (0.101), and θ_4 (0.038). The remaining two parameters, θ_3 and θ_7, contribute 0.031 and 0.002, respectively. Together, the first five parameters account for 96.8 % of the aggregated $\langle S_T \rangle_\phi$, exceeding the 95 % retention threshold. They are therefore propagated to the order-three refinement stage (basis size 56, $N = 500$). This hierarchy is consistent with the CMA-ES calibration results discussed in Sec. II.A. Excluding θ_9 and θ_{10}, which are treated separately in Block G, the most tightly constrained parameters in the CMA-ES optima, θ_1, θ_2, and θ_6, are also those with the largest Sobol contributions in Block C. Conversely, the parameters that exhibited greater variability among the best seeds during calibration show lower Sobol indices, indicating a weaker impact on the propagated NO_x uncertainty.

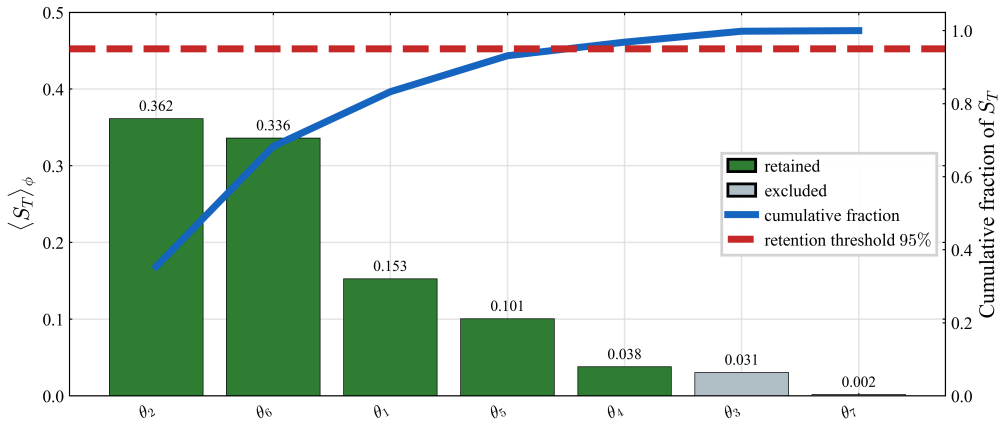
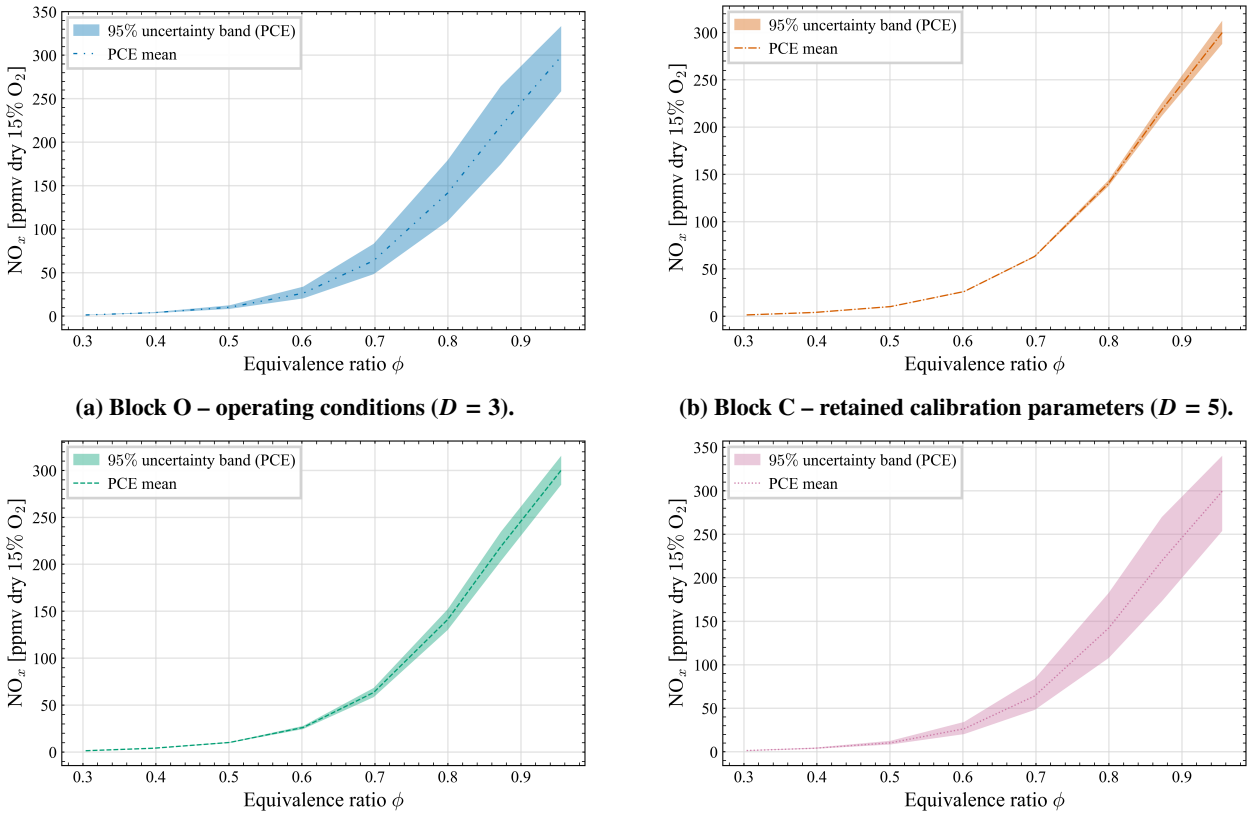


Fig. 13 Block C calibration retention diagram. Bars: aggregated total-effect Sobol score $\langle S_T \rangle_\phi$ obtained from the screening PCE. Cumulative fraction (right axis) crossing the 95 % threshold (dashed line) defines the retained subset (green bars).

The refined Block C surrogate concentrates its standard deviation in a narrow envelope around the deterministic curve, with σ growing from 0.008 ppmv at $\phi = 0.30$ to 6.71 ppmv at $\phi = 0.96$. Compared with Block O, the dispersion induced by the calibration parameters is one to two orders of magnitude smaller in the lean and intermediate regimes (Fig. 14b). This gap decreases as the equivalence ratio approaches stoichiometric conditions, where σ_C and σ_O differ by less than a factor of three. Thus, within the local ± 0.025 perturbation range around $\vec{\Theta}^*$, the CRN response is much less sensitive to calibration-parameter uncertainty than to operating uncertainty, although the calibration contribution becomes increasingly relevant near $\phi \simeq 1$. The refined Sobol decomposition shows that the dominant calibration parameter depends strongly on the equivalence ratio. In the lean and intermediate regimes, the variance is mainly controlled by θ_6, which regulates the IRZ→FL_{SW} recirculation. This is consistent with the role of hot-product feedback from the inner recirculation zone in sustaining the swirl flame at low and moderate ϕ. At higher equivalence ratios, the dominant contribution shifts to θ_2, which sets the inlet air mass-flow partition. Its total-effect index increases from $S_T = 0.47$ at $\phi = 0.70$ to $S_T = 0.79$ at $\phi = 0.96$, while θ_1 remains the main secondary contributor with $S_T \simeq 0.21$. This transition indicates that, as the global equivalence ratio approaches stoichiometric conditions, the calibration sensitivity becomes increasingly governed by the inlet partitioning and the resulting local stoichiometry in the swirl flame zone, rather than by recirculation feedback alone. For Block C, the order-three refinement confirms good surrogate accuracy



(a) Block O – operating conditions ($D = 3$).

(b) Block C – retained calibration parameters ($D = 5$).

(c) Block G – geometric and closure inputs ($D = 4$).

(d) Block T – joint propagation ($D = 12$).

Fig. 14 Block-by-block PCE mean and 95 % uncertainty band on NO_x over the equivalence ratio range $\phi \in [0.30, 0.96]$.

at six of the eight set points, with $Q^2_{\mathrm{LOO}} > 0.999$. The only exceptions are the low-equivalence-ratio cases $\phi = 0.40$ and $\phi = 0.50$, where $Q^2_{\mathrm{LOO}} = 0.805$ and 0.808, respectively. The external validation on 150 fresh Sobol samples shows the same trend: $R^2_{\mathrm{test}} \geq 0.999$ at the other six set points, while the two low-ϕ cases retain lower but acceptable values of 0.856 and 0.901. Although the predictive coefficients are lower at these two points, the surrogate still reproduces the NO_x response with very small absolute errors, with an external RMSE of the order of 10^{-2} ppmv.

Block G isolates the contribution of the geometric and closure inputs. These include the OH-PLIF/CFD residual uncertainty on the axial length L_{AX}, modeled with a ±20 % range, the Beér–Chigier swirl coefficient σ [24], assumed uniformly distributed in the interval [0.60, 1.00], and the ±0.025 half-width assigned to the volume fractions θ_9 and θ_{10}. The block is propagated with a third-order PCE and $N = 150$ Sobol samples per equivalence-ratio set point. The surrogate converges within numerical precision at all eight set points, with the predicted standard deviation σ_G growing monotonically from 0.017 ppmv at $\phi = 0.30$ to 7.92 ppmv at $\phi = 0.96$ (Fig. 14c). The geometric dispersion is therefore comparable in magnitude to the Block C calibration dispersion, and it remains throughout the sweep below the operating contribution of Block O. The ϕ-averaged Sobol attribution is dominated by the uncertainty on L_{AX} ($\langle S_T \rangle_\phi = 0.564$), followed by the volume fraction θ_9 (0.196), the swirl coefficient σ (0.130), and θ_{10} (0.118). The relative importance of these inputs varies with equivalence ratio. The swirl coefficient σ is dominant at $\phi = 0.40$ ($S_T = 0.45$), where changes in the swirl-induced redistribution of the FL_{SW} volume affect the early NO_x rise. From $\phi = 0.50$ onward, L_{AX} becomes the leading contributor ($S_T = 0.41$–0.78), reflecting the influence of the upstream Z_{AX} residence time on NO_x formation before the flow enters the swirl flame zone. Near stoichiometric conditions, θ_9 and θ_{10} become comparable to L_{AX}, with a combined S_T above 0.55 at $\phi = 0.96$, indicating that the downstream volume partitioning also contributes to the NO_x saturation. The agreement between S_{main} and S_T to within the third decimal further indicates weak interactions among the geometric and closure inputs within the prescribed ranges.

Block T corresponds to the joint propagation step and provides the global predictive band for NO_x. The original

fifteen-dimensional input set is reduced to twelve independent variables by holding θ_8 fixed at its nominal value, following the correlation analysis, and by excluding θ_3 and θ_7 after the screening stage of Block C. The final set therefore comprises the three operating perturbations of Block O, the five retained calibration parameters of Block C, and the four geometric/closure inputs of Block G. It is sampled with a Sobol design of $N = 500$ points per ϕ_{ref} using an order-two PCE with basis size 91. A fresh Sobol sample of $N_{\mathrm{ext}} = 150$ points per ϕ_{ref} is used for external validation. The resulting moments, 95 % uncertainty band, and external test metrics are summarized in Table 13 and Fig. 14d.

The external validation confirms the accuracy of the joint PCE surrogate. At the most demanding set point, $\phi_{\mathrm{ref}} = 0.96$, the test-set RMSE remains below 0.42 ppmv, corresponding to 0.14 % of the NO_x baseline. Moreover, $R^2_{\mathrm{test}} \geq 0.999$ at all set points except $\phi = 0.40$, where $R^2_{\mathrm{test}} = 0.994$ still indicates excellent predictive accuracy.

Table 13 Block T joint PCE statistics over the equivalence-ratio range. The mean, standard deviation, and 95 % uncertainty band of NO_x are obtained by Monte Carlo sampling of the order-two surrogate. $\mathrm{RMSE}_{\mathrm{ext}}$ and R^2_{test} are evaluated on an independent Sobol validation set with $N_{\mathrm{ext}} = 150$ samples per ϕ.

ϕ_{ref}	μ_{NO_x} [ppmv]	σ_{NO_x} [ppmv]	$\mathrm{CI}^{\mathrm{lo}}_{95}$ [ppmv]	$\mathrm{CI}^{\mathrm{hi}}_{95}$ [ppmv]	$\mathrm{RMSE}_{\mathrm{ext}}$ [ppmv]	R^2_{test} [–]
0.30	1.38	0.13	1.14	1.66	0.001	0.999
0.40	3.99	0.34	3.37	4.70	0.026	0.994
0.50	10.26	1.09	8.35	12.58	0.024	0.999
0.60	26.43	3.55	20.30	34.29	0.056	0.999
0.70	64.11	9.34	48.10	83.79	0.138	0.999
0.80	142.23	19.23	107.75	182.58	0.225	0.999
0.87	219.45	24.63	173.00	269.82	0.318	0.999
0.96	299.47	21.85	253.87	340.28	0.420	0.999

The aggregated main and total Sobol indices of Block T over the twelve independent inputs, averaged across the eight ϕ_{ref} set points, are reported in Fig. 15 and clarify which uncertainty source drives the predictive band. The air-mass flow and fuel-mass flow perturbations δ_a and δ_f jointly account for approximately 87 % of the joint variance ($\langle S_T \rangle_\phi = 0.49$ and 0.38). The next contributions are T_{air} and L_{AX}, at about 4 % each, followed by θ_9, θ_{10}, and θ_2 in the 1–1.5 % range. The remaining calibration and closure inputs lie below 0.5 % on average. Main and total indices are visually indistinguishable for every input, indicating a near-additive CRN response with negligible interactions among the uncertainty sources. When the attribution is evaluated at each ϕ_{ref} and grouped by uncertainty category, the same results expose a clear ϕ-dependence. The operating block covers 98.2 % of $\sum_i S_{T,i}$ at $\phi = 0.30$, declines to 90.3 % at $\phi = 0.80$, and reaches 78.4 % at $\phi = 0.96$. The geometric and closure inputs grow from 1.7 % to 12.9 % over the same range, with L_{AX} as the largest single contributor and θ_9 and θ_{10} becoming comparable to L_{AX} near stoichiometric. The retained calibration parameters contribute below 1 % at most lean and intermediate set points and rise to 9.3 % at $\phi = 0.96$, driven mainly by θ_2 and θ_1 in the near-stoichiometric regime. Overall, the calibrated CRN is robust against small perturbations of its calibration and geometric parameters. Throughout the lean and intermediate regimes, the residual predictive uncertainty on NO_x over the AHEAD operating envelope is dominated by operating measurement uncertainty, while calibration and geometric contributions become appreciable only at high equivalence ratios.

The cross-block comparison of the 95 % uncertainty bands in the NO_x–ϕ plane is shown in Fig. 16, together with the experimental data of Reichel et al. [22, 30]. The Block O band provides the widest envelope over the entire equivalence-ratio range, and the joint Block T band remains almost coincident with it. In contrast, the Block C and Block G bands are very narrow in the lean and intermediate regimes, becoming appreciable only as the equivalence ratio approaches stoichiometric conditions. The PCE mean reproduces the deterministic CRN curve of Fig. 6a within 1.1 ppmv at every ϕ_{ref}. The Block T band brackets the experimental measurements at six of the eight set points. The point at $\phi = 0.80$ lies at the upper edge of the predicted interval, with $\mathrm{NO}_{x,\mathrm{exp}} = 182.33$ ppmv and an uncertainty band of [107.75, 182.58] ppmv. As discussed in Sec. II.B.1, this point deviates from the trend of the remaining data and may be affected by measurement anomalies or localized combustion instabilities during acquisition.

The two leanest set points are not bracketed by the uncertainty band. This behavior is consistent with the lean-limit limitations already discussed in Sec. II.B.1: under very lean partially premixed hydrogen operation, localized hot spots and ignition–reignition events may contribute to NO_x formation in the real combustor, while these effects are inherently

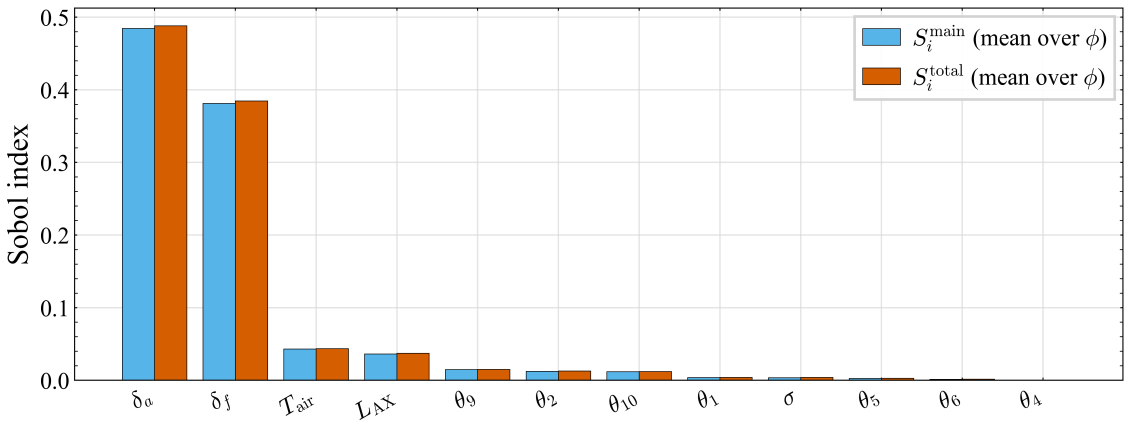


Fig. 15 Block T aggregated main and total Sobol indices over the twelve independent inputs, averaged over the eight ϕ_{ref} set points.

averaged within the CRN representation [70].

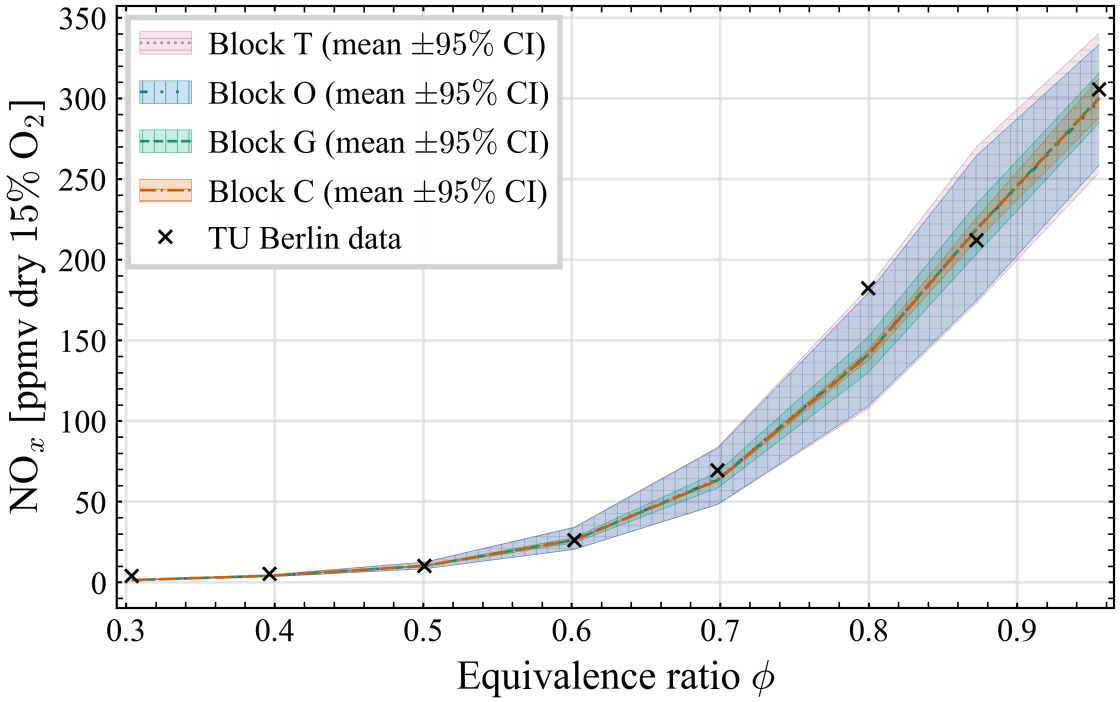


Fig. 16 95 % NO_x uncertainty bands of the four UQ blocks overlaid on the same axes. The TU Berlin reference data of Reichel et al. [22, 30] are shown as black crosses.

Conclusions

This work developed an integrated framework to improve both the predictive reliability and the computational efficiency of Chemical Reactor Network models for lean premixed hydrogen combustion. The AHEAD swirl-stabilized burner of Reichel et al. was used as a reference configuration, and the analysis combined three main elements: a systematic benchmark of optimization algorithms for CRN calibration, two acceleration strategies based on skeletal mechanism reduction and Flow-Map In Situ Adaptive Tabulation, and a block-structured uncertainty quantification study based on non-intrusive Polynomial Chaos Expansion.

The optimization benchmark showed that the calibration problem is better handled by population-based global strategies than by surrogate-assisted or gradient-based methods. Among the six algorithms tested over twenty independent seeds and under a common budget of approximately 2000 CRN evaluations, CMA-ES, PSO, and DE clearly formed the best-performing group. CMA-ES provided the most consistent results, with the lowest median objective value ($J = 0.0424$), the narrowest interquartile range, and a success rate of 95 %. Although PSO reached a comparable best-case performance, its larger dispersion made it less reliable for routine calibration. In contrast, Bayesian Optimization, ADAM, and the Hybrid DE + L-BFGS-B approach were less effective, mainly because the objective landscape contains broad infeasible regions together with non-convex and locally irregular features. The CMA-ES reference vector $\vec{\Theta}^*$ was therefore retained for the subsequent analyses. With this calibration, the CRN reproduced the experimental NO_x trend at the four calibration points with relative errors between 0.2 % and 24 %, the largest discrepancy occurring at the leanest condition, where the very low absolute NO_x level makes the relative error particularly sensitive to small deviations.

The DRGEP reduction procedure was applied to both parent mechanisms considered in this study: the Naik mechanism, with 21 species and 105 reactions, and the CRECK mechanism, with 31 species and 203 reactions. The threshold sweep produced eighteen candidate skeletal mechanisms, eleven of which were retained, as they preserved the evaluation criteria based on ignition delay, laminar flame speed, equilibrium NO_x, PSR extinction residence time, peak species mole fractions, net production rates, and integrated heat release. The most compact accepted skeletons were the 18-species, 79-reaction Naik mechanism and the 19-species, 98-reaction CRECK mechanism. When used in the calibrated CRN without further recalibration, these mechanisms preserved the NO_x predictions of the parent mechanisms with maximum pointwise relative errors of 0.71 % and 0.05 %, respectively. The corresponding standalone speedups over a 40-point equivalence-ratio sweep reached 1.62 for Naik and 2.06 for CRECK, with the smaller gain for Naik reflecting the already limited size of the original mechanism.

The Flow-Map ISAT formulation introduced here proved to be a more substantial source of acceleration. Instead of tabulating the reaction mapping of an individual reactor, the method tabulates the discrete flow map of the full six-reactor network, thereby retaining the coupling imposed by the inter-reactor mass-flow redistribution. With sub-table indexing based on the rounded macro-step and a deterministic hash of the forcing parameters, and using zeroth-order retrieval in scaled state space, the ISAT predictions remained almost indistinguishable from direct integration. Across the tolerance range investigated, the mean relative NO_x error stayed below $5.7 \cdot 10^{-4}$ %. At the same time, as the imposed tolerance ε_{tol} was relaxed, the mean per-ϕ speedup increased from 7.12 to 17.08 for the Naik mechanism and from 7.00 to 24.88 for CRECK. At the least restrictive tolerance, this corresponds to average simulation times of approximately 0.92 s for Naik and 1.58 s for CRECK.

When the two acceleration strategies were inserted into the full CMA-ES calibration loop for the Naik mechanism, the computational gains observed in isolated sweeps were largely preserved. Skeletal reduction alone gave a median speedup of 2.13, ISAT alone a speedup of 11.61, and the combined ISAT + Skeletal configuration a speedup of 15.07. As a result, across the 20-seed calibration benchmark, the median per-seed wall-clock time decreased from approximately 70 hours to less than 5 hours. The Kruskal–Wallis test performed on the four score distributions ($H = 0.028$, $p = 0.999$), together with the Bonferroni-corrected pairwise Mann–Whitney comparisons, did not indicate statistically significant differences in calibration quality between the accelerated campaigns and the detailed direct-integration baseline. The only noticeable effect was a moderate reduction in success rate when both acceleration strategies were combined, suggesting that the joint approximation may introduce a mild local perturbation of the objective landscape.

The uncertainty quantification analysis showed that the residual dispersion of the calibrated CRN is governed mainly by the operating conditions. In Block O, the variance was driven almost entirely by the air- and fuel-mass-flow perturbations, with only a minor contribution from the inlet temperature. In Block C, the strong correlation between θ_1 and θ_8 first led to fixing θ_8 at its reference value. The subsequent screening phase then identified a reduced five-parameter active set, θ_2, θ_6, θ_1, θ_5, and θ_4, which accounted for 96.8 % of the aggregated total-effect score. The dominant calibration contribution moved from θ_6, related to the IRZ→FL_{SW} recirculation, in the lean and intermediate regimes, to θ_2, controlling the inlet air-mass partition, close to stoichiometric conditions. In Block G, the largest contribution came

from the axial length L_{AX}. The resulting dispersion was comparable to that of Block C, but smaller than the operating contribution. The joint Block T propagation, built from the twelve retained inputs, led to the same conclusion: the PCE surrogate assigned about 87 % of the total variance to the air- and fuel-mass-flow uncertainties. The corresponding 95 % uncertainty band enclosed six of the eight experimental NO_x measurements, with the remaining discrepancy located at the two leanest points, where the homogeneous CRN representation is expected to be less effective in capturing localized hot spots and ignition–reignition events. The close agreement between main and total Sobol indices also indicates that, within the prescribed uncertainty ranges, the CRN response is essentially additive.

Overall, the calibrated CRN showed limited sensitivity to local variations in the optimized calibration and geometric parameters. For most of the AHEAD operating envelope, the residual uncertainty in NO_x prediction is governed primarily by inlet measurement uncertainty, while calibration and geometric effects become appreciable mainly near stoichiometric conditions. Coupled with the proposed acceleration strategies, the model provides a computationally efficient tool for hydrogen combustor emission analysis and preliminary design optimization: the fastest configuration reduces the median wall-clock time by approximately a factor of 15 relative to the detailed-mechanism direct-integration baseline and remains several orders of magnitude less expensive than typical CFD-based analyses.

References

[1] Cadman, T., "The United Nations Framework Convention on Climate Change," *The Palgrave Handbook of Contemporary International Political Economy*, edited by T. M. Shaw, L. C. Mahrenbach, R. Modi, and X. Yi-chong, Palgrave Macmillan UK, London, 2019, pp. 359–375. https://doi.org/10.1057/978-1-137-45443-0_23.

[2] Van der Sman, E., Peerlings, B., Kos, J., Lieshout, R., and Boonekamp, T., "Destination 2050 – A Route to Net Zero European Aviation," Tech. rep., Royal NLR – Netherlands Aerospace Centre, 2021.

[3] Adler, E. J., and Martins, J. R., "Hydrogen-powered aircraft: Fundamental concepts, key technologies, and environmental impacts," *Progress in Aerospace Sciences*, Vol. 141, 2023, p. 100922. https://doi.org/10.1016/j.paerosci.2023.100922, Special Issue on Green Aviation.

[4] Folcarelli, L., Gedda, G., Zumbo, A., Ferrero, A., Masseni, F., and Pastrone, D. G., "Integrated Numerical Modeling of a Hydrogen Turbofan Engine and Fuel System With Alternative Thermodynamic Cycles," *SCITECH 2025 Forum*, AIAA, 2025. https://doi.org/10.2514/6.2025-0328.

[5] Miller, J. A., and Bowman, C. T., "Mechanism and modeling of nitrogen chemistry in combustion," *Progress in Energy and Combustion Science*, Vol. 15, No. 4, 1989, pp. 287–338. https://doi.org/10.1016/0360-1285(89)90017-8.

[6] Vance, F., de Goey, L., and van Oijen, J., "Development of a flashback correlation for burner-stabilized hydrogen-air premixed flames," *Combustion and Flame*, Vol. 243, 2022, p. 112045. https://doi.org/10.1016/j.combustflame.2022.112045.

[7] Reichel, T. G., and Paschereit, C. O., "Interaction mechanisms of fuel momentum with flashback limits in lean-premixed combustion of hydrogen," *International Journal of Hydrogen Energy*, Vol. 42, No. 7, 2017, pp. 4518–4529. https://doi.org/10.1016/j.ijhydene.2016.11.018.

[8] Young, T., and Boris, J., "A numerical technique for solving stiff ordinary differential equations associated with the chemical kinetics of reactive-flow problems," *The Journal of Physical Chemistry*, Vol. 81, No. 25, 1977, pp. 2424–2427. https://doi.org/10.1021/j100540a018.

[9] Folcarelli, L., Spagnolo, A., Dicech, F., Ferrero, A., Masseni, F., and Pastrone, D. G., "Multi-Fidelity Modeling of a Lean Premixed Swirl-Stabilized Hydrogen Burner With Axial Air Injection," *SCITECH 2025 Forum*, AIAA, 2025. https://doi.org/10.2514/6.2025-0941.

[10] Folcarelli, L., Concetti, R., Cozzi, F., Ferrero, A., Masseni, F., Pastrone, D. G., Hasslberger, J., and Klein, M., "A Chemical Reactor Network methodology for estimating NOx emissions under non-premixed lean hydrogen combustion with water injection," *International Journal of Hydrogen Energy*, Vol. 168, 2025, p. 150561. https://doi.org/10.1016/j.ijhydene.2025.150561.

[11] Fichet, V., Kanniche, M., Plion, P., and Gicquel, O., "A reactor network model for predicting NOx emissions in gas turbines," *Fuel*, Vol. 89, No. 9, 2010, pp. 2202–2210. https://doi.org/10.1016/j.fuel.2010.02.010.

[12] Trespi, S., Nicolai, H., Debiagi, P., Janicka, J., Dreizler, A., Hasse, C., and Faravelli, T., "Development and application of an efficient chemical reactor network model for oxy-fuel combustion," *Energy & Fuels*, Vol. 35, No. 9, 2021, pp. 7121–7132. https://doi.org/10.1021/acs.energyfuels.0c03560.

[13] Perpignan, A. A., Sampat, R., and Gangoli Rao, A., "Modeling pollutant emissions of flameless combustion with a joint CFD and chemical reactor network approach," *Frontiers in Mechanical Engineering*, Vol. 5, 2019, p. 63. https://doi.org/10.3389/fmech.2019.00063.

[14] Innocenti, A., Andreini, A., Bertini, D., Facchini, B., and Motta, M., "Turbulent flow-field effects in a hybrid CFD-CRN model for the prediction of NOx and CO emissions in aero-engine combustors," *Fuel*, Vol. 215, 2018, pp. 853–864. https://doi.org/10.1016/j.fuel.2017.11.097.

[15] Savarese, M., Giuntini, L., Malpica Galassi, R., Iavarone, S., Galletti, C., De Paepe, W., and Parente, A., "Model-to-model Bayesian calibration of a Chemical Reactor Network for pollutant emission predictions of an ammonia-fuelled multistage combustor," *International Journal of Hydrogen Energy*, Vol. 49, 2024, pp. 586–601. https://doi.org/10.1016/j.ijhydene.2023.08.275.

[16] Kennedy, J., and Eberhart, R., "Particle swarm optimization," *Proceedings of ICNN'95 - International Conference on Neural Networks*, Vol. 4, 1995, pp. 1942–1948 vol.4. https://doi.org/10.1109/ICNN.1995.488968.

[17] Folcarelli, L., Masseni, F., and Pastrone, D. G., "Chemical Reactor Network for Hybrid Rocket Engines Optimization," *SCITECH 2026 Forum*, AIAA, 2026. https://doi.org/10.2514/6.2026-1574.

[18] Madonia, V., Folcarelli, L., Ferrero, A., Pastrone, D. G., and Masseni, F., "NOx Emission Prediction of a Lean Premixed Hydrogen Combustor via Reactor Network Analysis," *SCITECH 2026 Forum*, AIAA, 2026. https://doi.org/10.2514/6.2026-1186.

[19] Yousefian, S., Bourque, G., and Monaghan, R. F. D., "Review of Hybrid Emissions Prediction Tools and Uncertainty Quantification Methods for Gas Turbine Combustion Systems," *Proceedings of the ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition*, ASME, 2017, p. V04BT04A005. https://doi.org/10.1115/GT2017-64271.

[20] Reichel, T. G., Terhaar, S., and Paschereit, O., "Increasing Flashback Resistance in Lean Premixed Swirl-Stabilized Hydrogen Combustion by Axial Air Injection," *Journal of Engineering for Gas Turbines and Power*, Vol. 137, No. 7, 2015, p. 071503. https://doi.org/10.1115/1.4029119.

[21] Reichel, T. G., Goeckeler, K., and Paschereit, O., "Investigation of Lean Premixed Swirl-Stabilized Hydrogen Burner With Axial Air Injection Using OH-PLIF Imaging," *Journal of Engineering for Gas Turbines and Power*, Vol. 137, No. 11, 2015, p. 111513. https://doi.org/10.1115/1.4031181.

[22] Reichel, T. G., Terhaar, S., and Paschereit, C. O., "Flashback Resistance and Fuel–Air Mixing in Lean Premixed Hydrogen Combustion," *Journal of Propulsion and Power*, Vol. 34, No. 3, 2018, pp. 690–701. https://doi.org/10.2514/1.B36646.

[23] Goodwin, D. G., Moffat, H. K., Schoegl, I., Speth, R. L., and Weber, B. W., "Cantera: An Object-Oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes," https://www.cantera.org, 2024. https://doi.org/10.5281/zenodo.14455267, version 3.1.0.

[24] Beér, J. M., and Chigier, N. A., *Combustion aerodynamics*, Applied Science Publishers Limited, 1972.

[25] Bilger, R., "The structure of turbulent nonpremixed flames," *Symposium (International) on Combustion*, Vol. 22, No. 1, 1989, pp. 475–488. https://doi.org/10.1016/S0082-0784(89)80054-2.

[26] Naik, C. V., "Detailed Mechanism for Hydrogen Combustion with NOx," Ansys, Inc., 2009. Originally developed by Reaction Design, now part of Ansys.

[27] Song, Y., Marrodán, L., Vin, N., Herbinet, O., Assaf, E., Fittschen, C., Stagni, A., Faravelli, T., Alzueta, M. U., and Battin-Leclerc, F., "The sensitizing effects of NO2 and NO on methane low temperature oxidation in a jet stirred reactor," *Proceedings of the Combustion Institute*, Vol. 37, No. 1, 2019, pp. 667–675. https://doi.org/10.1016/j.proci.2018.06.115.

[28] Stagni, A., Cavallotti, C., Arunthanayothin, S., Song, Y., Herbinet, O., Battin-Leclerc, F., and Faravelli, T., "An experimental, theoretical and kinetic-modeling study of the gas-phase oxidation of ammonia," *Reaction Chemistry & Engineering*, Vol. 5, No. 4, 2020, pp. 696–711. https://doi.org/10.1039/C9RE00429G.

[29] Stagni, A., Arunthanayothin, S., Dehue, M., Herbinet, O., Battin-Leclerc, F., Bréquigny, P., Mounaïm-Rousselle, C., and Faravelli, T., "Low- and intermediate-temperature ammonia/hydrogen oxidation in a flow reactor: Experiments and a wide-range kinetic modeling," *Chemical Engineering Journal*, Vol. 471, 2023, p. 144577. https://doi.org/10.1016/j.cej.2023.144577.

[30] TNF Workshop Team, "AHEAD Lean-Premixed Swirl-Stabilized Hydrogen Burner TU Berlin Dataset," https://tnfworkshop.org/ahead-lean-premixed-swirl-stabilized-hydrogen-burnertu-berlin/, 2023. Accessed: May 21, 2025.

[31] Trelea, I. C., "The particle swarm optimization algorithm: convergence analysis and parameter selection," *Information Processing Letters*, Vol. 85, No. 6, 2003, pp. 317–325. https://doi.org/10.1016/S0020-0190(02)00447-7.

[32] Clerc, M., and Kennedy, J., "The Particle Swarm – Explosion, Stability, and Convergence in a Multidimensional Complex Space," *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 1, 2002, pp. 58–73. https://doi.org/10.1109/4235.985692.

[33] Storn, R., and Price, K., "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," *Journal of Global Optimization*, Vol. 11, No. 4, 1997, pp. 341–359. https://doi.org/10.1023/A:1008202821328.

[34] Price, K. V., Storn, R. M., and Lampinen, J. A., *Differential Evolution: A Practical Approach to Global Optimization*, Springer, Berlin, Heidelberg, 2005. https://doi.org/10.1007/3-540-31306-0.

[35] Hansen, N., and Ostermeier, A., "Completely Derandomized Self-Adaptation in Evolution Strategies," *Evolutionary Computation*, Vol. 9, No. 2, 2001, pp. 159–195. https://doi.org/10.1162/106365601750190398.

[36] Hansen, N., Müller, S. D., and Koumoutsakos, P., "Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES)," *Evolutionary Computation*, Vol. 11, No. 1, 2003, pp. 1–18. https://doi.org/10.1162/106365603321828970.

[37] Igel, C., Hansen, N., and Roth, S., "Covariance Matrix Adaptation for Multi-objective Optimization," *Evolutionary Computation*, Vol. 15, No. 1, 2007, pp. 1–28. https://doi.org/10.1162/evco.2007.15.1.1.

[38] Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. 13, No. 4, 1998, pp. 455–492. https://doi.org/10.1023/A:1008306431147.

[39] Snoek, J., Larochelle, H., and Adams, R., "Practical Bayesian Optimization of Machine Learning Algorithms," *Advances in Neural Information Processing Systems*, Vol. 25, edited by F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Curran Associates, Inc., 2012, pp. 2951–2959. URL https://proceedings.neurips.cc/paper_files/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf.

[40] Hoffman, M. W., Brochu, E., and de Freitas, N., "Portfolio Allocation for Bayesian Optimization," *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, 2011, pp. 327–336.

[41] Kingma, D. P., and Ba, J. L., "Adam: A Method for Stochastic Optimization," *International Conference on Learning Representations (ICLR)*, 2015. https://doi.org/10.48550/arXiv.1412.6980.

[42] Balsa-Canto, E., Peifer, M., Banga, J. R., Timmer, J., and Fleck, C., "Hybrid Optimization Method with General Switching Strategy for Parameter Estimation," *BMC Systems Biology*, Vol. 2, 2008, p. 26. https://doi.org/10.1186/1752-0509-2-26.

[43] Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J., "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Softw.*, Vol. 23, No. 4, 1997, p. 550–560. https://doi.org/10.1145/279232.279236.

[44] Pepiot-Desjardins, P., and Pitsch, H., "An efficient error-propagation-based reduction method for large chemical kinetic mechanisms," *Combustion and Flame*, Vol. 154, No. 1, 2008, pp. 67–81. https://doi.org/10.1016/j.combustflame.2007.10.020.

[45] Pope, S., "Computationally efficient implementation of combustion chemistry using in situ adaptive tabulation," *Combustion Theory and Modelling*, Vol. 1, No. 1, 1997, pp. 41–63. https://doi.org/10.1080/713665229.

[46] Lu, T., and Law, C. K., "A directed relation graph method for mechanism reduction," *Proceedings of the Combustion Institute*, Vol. 30, No. 1, 2005, pp. 1333–1341. https://doi.org/10.1016/j.proci.2004.08.145.

[47] Dijkstra, E. W., "A Note on Two Problems in Connexion with Graphs," *Numerische Mathematik*, Vol. 1, 1959, pp. 269–271. https://doi.org/10.1007/BF01386390.

[48] Glarborg, P., Miller, J. A., Ruscic, B., and Klippenstein, S. J., "Modeling nitrogen chemistry in combustion," *Progress in Energy and Combustion Science*, Vol. 67, 2018, pp. 31–68. https://doi.org/10.1016/j.pecs.2018.01.002.

[49] Singer, M. A., and Pope, S. B., "Exploiting ISAT to solve the reaction–diffusion equation," *Combustion Theory and Modelling*, Vol. 8, No. 2, 2004, pp. 361–383. https://doi.org/10.1088/1364-7830/8/2/009.

[50] Singer, M. A., Pope, S. B., and Najm, H. N., "Modeling unsteady reacting flow with operator splitting and ISAT," *Combustion and Flame*, Vol. 147, No. 1, 2006, pp. 150–162. https://doi.org/10.1016/j.combustflame.2006.06.007.

[51] Adhikari, S., Sayre, A., and Chandy, A. J., "In situ adaptive tabulation (ISAT) for combustion chemistry in a network of perfectly stirred reactors (PSRs)," *Computers & Chemical Engineering*, Vol. 97, 2017, pp. 124–134. https://doi.org/10.1016/j.compchemeng.2016.11.023.

[52] Lu, L., and Pope, S. B., "An improved algorithm for in situ adaptive tabulation," *Journal of Computational Physics*, Vol. 228, No. 2, 2009, pp. 361–386. https://doi.org/10.1016/j.jcp.2008.09.015.

[53] Wiener, N., "The Homogeneous Chaos," *American Journal of Mathematics*, Vol. 60, No. 4, 1938, pp. 897–936. URL http://www.jstor.org/stable/2371268.

[54] Ghanem, R. G., and Spanos, P. D., *Stochastic finite elements: a spectral approach*, Courier Corporation, 2003.

[55] Xiu, D., and Karniadakis, G. E., "The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations," *SIAM Journal on Scientific Computing*, Vol. 24, No. 2, 2002, pp. 619–644. https://doi.org/10.1137/S1064827501387826.

[56] Sobol', I., "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Mathematics and Computers in Simulation*, Vol. 55, No. 1, 2001, pp. 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6, the Second IMACS Seminar on Monte Carlo Methods.

[57] Sudret, B., "Global sensitivity analysis using polynomial chaos expansions," *Reliability Engineering & System Safety*, Vol. 93, No. 7, 2008, pp. 964–979. https://doi.org/10.1016/j.ress.2007.04.002, bayesian Networks in Dependability.

[58] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., *Variance-Based Methods*, John Wiley & Sons, Ltd, 2007, Chap. 4, pp. 155–182. https://doi.org/10.1002/9780470725184.ch4.

[59] Spearman, C., "The Proof and Measurement of Association between Two Things," *The American Journal of Psychology*, Vol. 15, No. 1, 1904, pp. 72–101. https://doi.org/10.1037/11491-005.

[60] Pearson, K., "VII. Note on regression and inheritance in the case of two parents," *Proceedings of the Royal Society of London*, Vol. 58, No. 347-352, 1895, pp. 240–242. https://doi.org/10.1098/rspl.1895.0041.

[61] Sobol', I., "On the distribution of points in a cube and the approximate evaluation of integrals," *USSR Computational Mathematics and Mathematical Physics*, Vol. 7, No. 4, 1967, pp. 86–112. https://doi.org/10.1016/0041-5553(67)90144-9.

[62] McKay, M. D., Beckman, R. J., and Conover, W. J., "Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, Vol. 21, No. 2, 1979, pp. 239–245. https://doi.org/10.1080/00401706.1979.10489755.

[63] Blatman, G., and Sudret, B., "Adaptive sparse polynomial chaos expansion based on least angle regression," *Journal of Computational Physics*, Vol. 230, No. 6, 2011, pp. 2345–2367. https://doi.org/10.1016/j.jcp.2010.12.021.

[64] Stone, M., "Cross-Validatory Choice and Assessment of Statistical Predictions," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 36, No. 2, 1974, pp. 111–133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x.

[65] Geisser, S., "The Predictive Sample Reuse Method with Applications," *Journal of the American Statistical Association*, Vol. 70, No. 350, 1975, pp. 320–328. https://doi.org/10.1080/01621459.1975.10479865.

[66] Kruskal, W. H., and Wallis, W. A., "Use of Ranks in One-Criterion Variance Analysis," *Journal of the American Statistical Association*, Vol. 47, No. 260, 1952, pp. 583–621. https://doi.org/10.1080/01621459.1952.10483441.

[67] Mann, H. B., and Whitney, D. R., "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *The Annals of Mathematical Statistics*, Vol. 18, No. 1, 1947, pp. 50–60. https://doi.org/10.1214/aoms/1177730491.

[68] Dunn, O. J., "Multiple Comparisons among Means," *Journal of the American Statistical Association*, Vol. 56, No. 293, 1961, pp. 52–64. https://doi.org/10.1080/01621459.1961.10482090.

[69] Folcarelli, L., Marcì, S. G. I., Madonia, V., Ferrero, A., Masseni, F., and Pastrone, D. G., "Numerical Simulation of NOx Emissions in a Lean Premixed Hydrogen Combustor Using Different Levels of Fidelity," *Aviation 2026 Forum*, AIAA, 2026. To appear.

[70] Schmidt, N., Müller, M., Preuster, P., Zigan, L., Wasserscheid, P., and Will, S., "Development and characterization of a low-NOx partially premixed hydrogen burner using numerical simulation and flame diagnostics," *International Journal of Hydrogen Energy*, Vol. 48, No. 41, 2023, pp. 15709–15721. https://doi.org/10.1016/j.ijhydene.2023.01.012.