

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Uncertainty quantification of reduced models

Permalink

<https://escholarship.org/uc/item/6t27294b>

Author

Liu, Zhenyuan

Publication Date

2020

Peer reviewed|Thesis/dissertation

Uncertainty quantification of reduced models

by

Zhenyuan Liu

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Engineering - Mechanical Engineering
and the Designated Emphasis

in

Computational and Data Science and Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael Frenklach, Chair
Professor Jyh-Yuan Chen
Professor Haiyan Huang

Summer 2020

Uncertainty quantification of reduced models

Copyright 2020
by
Zhenyuan Liu

Abstract

Uncertainty quantification of reduced models

by

Zhenyuan Liu

Doctor of Philosophy in Engineering - Mechanical Engineering

and the Designated Emphasis in

Computational and Data Science and Engineering

University of California, Berkeley

Professor Michael Frenklach, Chair

Detailed reaction models, such as detailed soot models, describing complex phenomena in combustion are typically computationally intensive. Reduced reaction models derived from a detailed, full-size reaction model are, thus, necessary in combustion simulations. The validation of a reduced model typically requires its predictions for selected quantities of interest (QOIs) to be close to those of the detailed model. Moreover, an accurate reduced model should be capable of reproducing faithfully the propagation of uncertainty by the detailed model.

Several reduced models for syngas combustion were developed by the “detailed reduction” method. This method was also adapted to develop several reduced models for a stochastic soot oxidation model, which are typically difficult to develop with other reduction methods. Different measures were developed to assess how uncertainties in the model parameters and in the model predictions behave for reduced models as compared to those for the detailed model. The uncertainty quantification (UQ) analysis was carried out through a numerically efficient, deterministic, and optimization-based framework of Bound-to-Bound Data Collaboration (B2BDC) and a Gibbs sampling algorithm adopted for B2BDC. The measures developed can be categorized into the sampling- and optimization-based measures.

The measures were applied to several reduced models of syngas combustion and several reduced models of soot oxidation in three different examples. The developed measures successfully quantified the propagation of uncertainty by the detailed and reduced models, and numerically demarcated the performance of different reduced models. The results demonstrated that assessment of the quality of a reduced model without considering parameter uncertainty may be misleading in that the deviation can be much larger when the uncer-

tainties in the model parameters are taken into account, highlighting the significance of UQ analysis in the validation of reduced models. The performed analysis demonstrated that when the experimental data are of bad quality, the posterior region of the model parameters (the feasible set) could have a very complex shape, posing a substantial challenge to the sampling-based measures. If no verifiably accurate experimental data exist, computer-generated data from the solution of the detailed model offer a reliable alternative, in which the desired level of reduced-model accuracy can be prescribed by specifying the accepted ranges of variations in prediction of training targets. If a feasible set has a very complex shape, uniform sampling of the feasible set could be very expensive. In such situations, the B2BDC framework offers a more practical alternative by quantifying the propagated uncertainty through numerically efficient computations of uncertainty intervals and their overlap, all with the added benefit of obtaining the uncertainty sensitivities.

To my family

Contents

Contents	ii
List of Figures	iv
List of Tables	vii
1 Introduction	1
1.1 Motivation	1
1.2 Background	3
1.3 Organization of dissertation	6
2 Methodology	7
2.1 Bound-to-Bound Data Collaboration (B2BDC)	7
2.2 “Detailed-reduction” method for reaction mechanisms	12
2.3 KMC-based soot oxidation model	13
3 B2BDC-based uncertainty quantification of reduced models	23
3.1 Comparison of predicted values	23
3.2 Comparison of feasible sets	25
3.3 Comparison of posterior distributions	25
3.4 Comparison of prediction intervals	27
3.5 Comparison of sensitivities	28
3.6 Summary	29
4 Reduced models of syngas combustion	31
4.1 Modeled system	31
4.2 A dataset with computer-generated ignition delay QOIs	37
4.3 A dataset with experimental ignition delay QOIs	54
4.4 Summary	64
5 Reduced models of soot oxidation	67
5.1 Modeled system	67
5.2 Parallel computing using MPI	73

5.3	A dataset with soot oxidation QOIs	74
5.4	Summary	82
6	Summary and Conclusions	84
A	Reduced models of syngas combustion	86
	Bibliography	197

List of Figures

2.1	Prior region, \mathcal{H} , and posterior region, \mathcal{F} , of the model parameters	10
2.2	Inner- and outer-bound prediction intervals of $[M_p, \overline{M_p}]$	12
2.3	Initial substrates for KMC-MMR simulations. The H atoms saturating the edge carbon atoms are not shown for clarity.	14
4.1	Absolute impact factors of all 55 parameters, computed with the detailed model for QOI 68.	42
4.2	Deviations, δ , $\delta^{\mathcal{H}}$, and $\delta^{\mathcal{F}}$, between the detailed (D) model and reduced (R25, R17, R12) models for 100 QOIs.	43
4.3	Histograms of deviations, δ , $\delta^{\mathcal{H}}$, and $\delta^{\mathcal{F}}$, between the detailed (D) and reduced (R25, R17, R12) models for 100 QOIs.	44
4.4	The relative volume of \mathcal{F}^{D} vs. the number of sampled points.	45
4.5	Bounds of active model parameters computed with the detailed model; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and vertical black lines are sample-based prediction intervals.	46
4.6	Histograms for selected QOIs of detailed (blue) and reduced (brown) model predictions over feasible sets.	47
4.7	Histograms of the Hellinger distance metrics computed for the prediction intervals (left) and the posterior distributions (right) of the reduced models.	48
4.8	QOI prediction intervals computed with the detailed, D, and reduced, R25, R17, and R12, models; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and horizontal dashed lines are prediction intervals of the detailed model.	50
4.9	Sensitivity of prediction interval for QOI 1 with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.	52
4.10	Sensitivity of prediction interval for QOI 92 with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.	53

4.11	Sensitivity of the Bhattacharyya coefficient of QOI 92 with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to reactions of a given model that are absent from the model displayed in the next panel below.	54
4.12	Deviations, δ , δ^H , and δ^F , between the detailed (D) model and reduced (R25, R17) models for 22 QOIs.	57
4.13	Bounds of active model parameters computed with the detailed model; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and vertical blue lines are sample-based prediction intervals.	59
4.14	Histograms for selected QOIs of detailed (blue) and reduced (brown) model predictions over feasible sets.	61
4.15	QOI prediction intervals computed with the detailed, D, and reduced, R25, and R17, models; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and horizontal dashed lines are prediction intervals of the detailed model.	62
4.16	Sensitivity of prediction interval for QOI 1 with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.	64
4.17	Sensitivity of prediction interval for QOI 18 with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.	65
5.1	Bounds of active model parameters computed with the detailed model; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and vertical blue lines are sample-based prediction intervals.	77
5.2	Histograms for QOIs 1 and 2 of detailed (blue) and reduced (brown) model predictions over feasible sets.	78
5.3	QOI prediction intervals computed with the detailed, D, and reduced, R1, and R5 models; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and horizontal dashed lines are prediction intervals of the detailed model.	79
5.4	Sensitivity of prediction interval for QOI 1 with respect to the uncertainty intervals of model parameters. Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.	81

5.5	Sensitivity of prediction interval for QOI 2 with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.	82
A.1	Histograms for 100 QOIs of detailed (blue) and reduced (brown) model predictions over feasible sets.	108
A.2	QOI prediction intervals computed with the detailed, D, and reduced, R25, and R17, models; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and horizontal dashed lines are prediction intervals of the detailed model.	122
A.3	Sensitivity of prediction interval for 100 QOIs with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.	173
A.4	Histograms for 22 QOIs of detailed (blue) and reduced (brown) model predictions over feasible sets.	180
A.5	QOI prediction intervals computed with the detailed, D, and reduced, R25, and R17, models; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and horizontal dashed lines are prediction intervals of the detailed model.	184
A.6	Sensitivity of prediction interval for 22 QOIs with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.	196

List of Tables

2.1	Surface reactions of the detailed soot oxidation model.	16
4.1	Reactions of the detailed model, reproduced from Slavinskaya et al. [112], and reduced models R25, R17 and R12.	32
4.2	Initial conditions of the Latin hypercube sampling (LHS) points used for building the reduced models R25 and R17 using the “detailed-reduction” method; ignition delay times τ^D are those computed with the detailed model, and the relative deviations for the reduced models are defined as $\frac{ \tau^R - \tau^D }{\tau^D} \times 100\%$	34
4.3	Number of active parameters of surrogate models for 100 QOIs with the detailed model and three reduced models, R25, R17, and R12.	38
4.4	Summary statistics of deviation measures (over 100 QOIs).	41
4.5	Summary statistics of Hellinger distances (over 100 QOIs).	49
4.6	Summary statistics of the relative lengths of sample-based prediction intervals (over 100 QOIs).	51
4.7	Initial conditions of 22 QOIs remained in the dataset for UQ analysis; ignition delay times τ^D are those computed with the detailed model and the relative deviations for the reduced models are defined as $\frac{ \tau^R - \tau^D }{\tau^D} \times 100\%$	55
4.8	Experimental bounds of the 22 QOIs remained in the dataset for UQ analysis.	55
4.9	Summary statistics of deviation measures (over 22 QOIs).	58
4.10	Summary statistics of Hellinger distances (over 22 QOIs).	61
4.11	Summary statistics of the relative lengths of sample-based prediction intervals to optimization-based prediction intervals (over 22 QOIs).	63
5.1	Gas-phase environment conditions for KMC oxidation simulation.	68
5.2	Reaction counts of all 107 reactions in the simulation with the detailed model, and identification of reactions in reduced models R1, R5, and R10.	68
5.3	QOIs calculated using the detailed and reduced models with nominal model parameters.	71
5.4	Groups of reactions.	71
5.5	Number of reactions and groups in the detailed and reduced models.	73
5.6	Number of active variables and fitting errors of surrogate models for QOIs 1 and 2.	75
5.7	Deviations, δ , δ^H , and δ^F , between the detailed (D) and reduced (R1, R5, R10) models for QOIs 1 and 2.	75

5.8	Hellinger distances, h^H and h^U , between posterior distributions of QOI predictions using the detailed (D) and reduced (R1, R5) models for QOIs 1 and 2.	78
5.9	Relative lengths of sample-based prediction intervals to optimization-based prediction intervals for QOIs 1 and 2.	80
A.1	Deviations between predicted values computed by the detailed and reduced models with the nominal parameter values, δ_e , and over the prior region, δ_e^H	86
A.2	Deviations between predicted values computed by the detailed and reduced models over the posterior region, δ_p^F	89
A.3	Hellinger distance calculated for the prediction intervals, h_p^U , and for the sampled histograms, h_p^H	109
A.4	Deviations between predicted values computed by the detailed and reduced models with the nominal parameter values, δ_e , over the prior region, δ_e^H , and over the posterior region, δ_p^F	174
A.5	Hellinger distance calculated for the prediction intervals, h_p^U , and for the sampled histograms, h_p^H	181

Acknowledgments

I'm extremely grateful to having met so many stellar, inspiring, and friendly people during my graduate studies. Working, studying, and interacting with all these amazing people have made this dissertation possible.

First, I would like to express my sincere gratitude to my advisors, Professor Michael Frenklach and Professor Andrew Packard. Michael has taught me how to better prioritize things, how to turn data into actionable insights, how to design computer experiments so as to maximize information gain, and how to communicate effectively in writing and other forms. His creativity in research, critical thinking, and exceptional time management also influenced me significantly. Andy has taught me the importance of criticizing one's own work, with the ultimate goal of improving that work. His rigorousness in research, his unparalleled dedications to work, and his tenacity in the face of cancer all inspired me. All these invaluable lessons I learned from my advisors will benefit me in the many years to come.

I would also like to thank my dissertation committee and qualifying exam committee members, Professor Michael Frenklach, Professor Jyh-Yuan Chen, Professor Haiyan Huang, Professor Carlos Fernandez-Pello, and Professor Tarek Zohdi for their time, comments, and insights. I enjoyed learning combustion and turbulent flows with Professor Chen and having multiple intriguing conversations about next-generation engines, batteries, electric vehicles, and fuel cell vehicles with him. I had the pleasure of taking an amazing statistics course with Professor Huang, in which I learned many powerful techniques. Moreover, this course helped me develop a statistical way of thinking that is indispensable in dealing with problems in this uncertain world.

My lab mates have been very friendly, supportive, and helpful. Ravi Singh helped me onboard in our group and taught me how to use the KMC code step by step, and his help continued even after he graduated. I am thankful to Wenyu Li for his help with the B2BDC package and numerous conversations ranging from satire talk shows to soccer, James Oreluk for his help with PrIME-Kinetics and lots of intriguing topics he brought about thanks to his tech-savviness, and Arun Hegde for his help with many issues related to optimization and statistics and many creative ideas he gave thanks to his vast knowledge base. I am also grateful to all the fellow graduate and undergraduate students, professors, lecturers, and staffs that helped me or inspired me during my graduate studies.

I would like to give thanks to the Mechanical Engineering department for the grants. My research also used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

Finally, I would like to express my deepest gratitude to my family. My parents and grandparents have always supported, encouraged, and loved me. They not only raised me as a child but also "raised me up to more than I can be".

Chapter 1

Introduction

1.1 Motivation

Declining costs, growing share of ESG (Environmental, Social, and Governance) investing, and favorable government policies, along with many other factors, have driven the continued growth of the renewable energy industry in the past few years [1–3]. Despite substantial efforts, renewable energy comprised only 11% of all energy consumption in the U.S. in 2019, on par with coal (11%), but much less than petroleum (37%) or natural gas (32%) [4]. In total, fossil fuels accounted for approximately 80% of U.S. energy consumption in 2019. The remaining 8%¹ of energy consumption was from nuclear energy, another form of clean energy. In its Annual Energy Outlook 2020 [5], U.S. Energy Information Administration (EIA) projected an increase in the use of fossil fuels through 2050.

Global energy consumption relies heavily on fossil fuels as well. According to the International Energy Outlook 2019 [6] published by EIA, petroleum, natural gas, coal, renewable energy, and nuclear energy comprised approximately about 32%, 22%, 26%, 15%, and 5% of global energy consumption in 2018. In total, fossil fuels accounted for about 80% of international energy consumption in 2018. The 26% share of coal, the least clean energy, in global energy consumption reflects a grimmer energy landscape in some developing countries. EIA’s projection [6] shows that the share of fossil fuels globally will decline modestly to 69% by 2050 from 80% in 2018. However, this decrease is eclipsed by the projected nearly 50% increase in global energy usage by 2050, led by growth in Asia [6].

Energy conversion of fossil fuels mainly takes the form of combustion. Carbon dioxide (CO₂) emissions from combustion of fossil fuels have been the main contributor of rising CO₂-concentration levels in the atmosphere [7]. Burning fossil fuels also leads to the emissions of pollutants such as carbon monoxide (CO), nitrogen oxides (NO_x), and soot among others [8, 9]. Soot is composed of carbonaceous particles produced in fuel-rich zones during combustion and they pose major threats to human health and the environment [8, 10–12]. Soot particles are a major type of particular matter (PM) particles [8], and they are regulated by the U.S.

¹Sum of components may not equal 100% because of independent rounding.

Environmental Protection Agency (EPA) [13]. Numerical methods, alongside experimental and theoretical methods, play an important role in designing combustion devices with a high efficiency and low emissions of CO₂ and pollutants [8]. Accurate predictive reaction models are the key to numerical methods.

Detailed reaction models, with the full set of equations (composed of the full set of species and reactions), to describe complex phenomena of combustion currently exist, but the computations with such models cannot be completed in a reasonable amount of time [8]. Reduced reaction models derived from a detailed, full-size reaction model are, thus, necessary in combustion simulations.

Reduced models derived from a detailed model have less species and reactions than the detailed model and, hence, have less model parameters. Accurate reduced models should be capable of reproducing the detailed models in computer simulations. Validation of a derived reduced model is typically assessed by comparing its predictions for selected quantities of interest (QOIs), such as ignition delays and soot oxidation rates, against the predictions of the same QOIs with the detailed model. Such a comparison is usually done with model parameters (rate coefficients, equilibrium constants, etc.) fixed at their respective nominal values. However, there exist uncertainties in the model parameters because the values of these parameters are unknown, cannot be known beyond a certain accuracy, or are merely approximations due to assumptions in the fundamental theory [14, 15]. Assessment of reduced models, accounting for uncertainties in the model parameters, has received more attention recently [16–19]. However, systematic treatments of uncertainties in the model parameters and in the model predictions and their interactions in the context of reduced models remain largely unavailable in the combustion community.

Detailed soot models can enhance the understanding of soot formation and oxidation. The latter is particularly important for soot emission control in combustion processes because soot oxidation counterbalances soot growth and, ultimately, is responsible for the removal of soot from exhaust [20]. However, detailed soot models require a substantial amount of computational time and cost [21]. Coupling a detailed soot model with a reactive flow model would be prohibitively expensive. There is a strong demand for reduced soot models that can reproduce faithfully a detailed soot model with less computing time. Such reduced soot models can also provide greater insight into the significance of different underlying mechanisms in soot processes. The validation of reduced soot models under uncertainties in the model parameters is a key step in developing accurate reduced soot models. This dissertation aims to develop quantitative measures of the difference between reduced and detailed models under uncertainties in the model parameters. Such measures would provide important information on whether a reduced model can reproduce the uncertainties of the detailed model and whether such a requirement can be used for the development of accurate reduced models.

1.2 Background

Three topics are essential to this dissertation: reduction methods for gas phase reaction models, soot modeling, and uncertainty quantification and its application in combustion. A brief review on these topics will be given in this section.

1.2.1 Reaction model reduction

Reaction model reduction is itself an active field of study, and there are numerous reduction methods. Often, a good reduced model, with a smaller set of equations, can obtain a close estimate of the solution of the full model with less computing time [8]. For a good reduced model, the benefit of lower computing effort outweighs the small deviations from the detailed model. Moreover, a reduced model provides increased insight into species interactions and the importance of certain mechanisms [8]. Several commonly used methods will be reviewed in this section.

Reaction rate analysis identifies unimportant reactions and species by checking the contributions of reactions on the formation and consumption of species at different time steps and locations. The idea of reaction rate analysis is adopted in many different reaction model reduction methods [16, 22–27]. The “detailed-reduction” method [22–24] removes non-contributing reactions with low heat release rates and reaction rates relative to some pre-selected rate-controlling reactions. Methods based on directed relation graph (DRG) [16, 25–27] also remove unimportant reactions, accounting for strongly coupled species groups. With the introduction of DRGs, graph algorithms, such as the Dijkstra’s algorithm and the AStar algorithm [28], can be leveraged to further reduce the size of the reduced model from conventional target search algorithms (TSAs) [27].

Quasi-steady-state assumption (QSSA) and partial equilibrium assumption (PEA) are commonly used in reaction model reduction [29–32]. They are either used by themselves or in combination with other methods [8]. QSSA states that the rate of production and the rate of consumption of a species is very close to each other such that the net consumption rate of that species is approximately zero. As a result, concentrations of such species can be calculated in terms of other known concentrations [29] so that the number of ordinary differential equations (ODEs) can be reduced as a result of the algebraic relations obtained [8]. PEA bears some similarities to QSSA but should not be confused with QSSA. PEA is established when the forward and backward reaction rates of some reactions are essentially equal to each other. This usually results from the very fast rates of both the forward and backward rates of these reactions at high temperatures [8]. Similarly, concentrations of some reactive species can be expressed in terms of more stable species [8]. One drawback with QSSA and PEA is that candidate species and reactions for these assumptions need to be known a priori [8]. Multiple algorithms exist for systematically identifying candidate species for QSSA and reactions for PEA [32].

CSP methods are based on the dynamical systems approach [33–37]. ODEs of combustion systems are usually stiff due to the very large differences in time scales of different reactions.

The presence of very fast reactions requires extremely small timescales for integration of the ODE system. While accounting for the bulk of the computation time, these fast reactions are of less interest compared to the slower rate-controlling ones [8]. CSP methods attempt to reduce the computation time dedicated to these fast reactions. These methods split the ODE system into fast and slow modes for a given time period of interest. The fast modes dissipate very fast, and they are exhausted after a short period of time and can be discarded thereafter. For the same time period of interest, the very slow modes can also be neglected due to their small contributions [34]. For each fast mode, CSP identifies species that can be accurately solved for from equations of state—algebraic equations like those obtained from QSSA and PEA, thus reducing the computation time. The main advantage of CSP-based methods is that it does not require identification of QSSA species or PEA reactions a priori. Intrinsic low-dimensional manifold-based methods [38, 39] simplify reaction models by identifying the fast time scale of the reaction systems. They are also based on the dynamical systems approach.

1.2.2 Soot modeling

An accurate predictive soot model is the key to the understanding of soot formation and oxidation and to the designing of environmentally friendlier combustion devices. In contrast to other well-specified species in reaction models, such as gas phase species, soot is not a clearly defined substance [11]. It is not feasible to definitively state what species and reactions are present in a soot reaction model because each soot particle is different and unique [40]. Modeling of soot is, thus, mathematically highly challenging [21]. To date, empirical and semi-empirical models have been widely used, but they are typically limited to specific conditions and are shown to have many deficiencies [21]. Detailed theoretical models containing all the components present in the soot formation with a high level of detailed chemical and physical processes also exist. However, such comprehensive models (detailed models) are, more often than not, computationally very expensive [21].

Soot formation is generally considered to include four steps: formation of soot precursors and the following nucleation of the soot particles, particle coagulation, particle surface reactions, and particle agglomeration [10]. Particle surface reactions refer to the heterogeneous reactions occurring at the particle surface, including growth and oxidation reactions [10]. Modeling of particle surface reactions is a key to the development of predictive soot models because the carbon mass accumulated in soot is determined mainly by surface reactions [10]. The detailed soot models developed by the Frenklach group [41, 42] were highly comprehensive and had been considered as the state of the art [11]. Recent efforts of the Frenklach group have been focused on soot modeling based on kinetic Monte-Carlo (KMC) simulations, which takes into account steric effects and substrate curvatures [43, 44]. A kinetic Monte-Carlo (KMC) simulation tracks the evolution of a graphene substrate in a flame-like environment in which the surface growth and oxidation reaction events are modeled as a Markovian process [45, 46]. At each time step of a given run, all applicable surface reactions are determined by the current status of the edge sites of the substrate. Among all applicable

surface reactions, one reaction is then selected using a stochastic algorithm [47, 48]. After that, the selected reaction is applied, the substrate is updated, the geometry is optimized, and this stochastic process is repeated until the end of simulation [43]. As is typical for Monte-Carlo methods, multiple runs with different seed numbers are performed in a KMC simulation, and the average over these runs is taken to derive the values of quantities of interest (QOIs). The Frenklach group has worked on many aspects of soot modeling using KMC simulations. They have investigated the graphene-edge growth chemistry [43, 49], edge oxidation by OH [40, 50], by O₂ [44, 51, 52], and by O [20, 44] in a series of studies. The latest soot model [20] will be taken as the detailed soot model in this dissertation.

1.2.3 Uncertainty quantification

Computer models are widely used to study a broad range of scientific phenomena because physical experimentation is too time consuming, too expensive, or even impossible for some physical processes [53]. A computer model takes some quantities as input variables and makes predictions for other quantities, which are termed output variables or responses of the model [54]. During the process of model-based predictions, uncertainties arise from a variety of sources, including uncertainties in the model parameters, model discrepancy (inadequacy) caused by the difference between the computer model and the underlying physics of the system, inherent randomness of the system, and algorithmic errors [55, 56]. Uncertainty quantification (UQ) can be thought of as the task of determining appropriate uncertainties during the process of model-based predictions [55]. UQ was also defined as “the end-to-end study of the reliability of scientific inferences” [57]. Both definitions of UQ highlight the breadth of UQ and the fact that UQ lacks a coherent general presentation [56]. Nonetheless, there are many well-defined problems in UQ. One common UQ problem is the so-called forward problem. In a forward problem, the uncertainties in the model parameters (input) are propagated into the uncertainties in the model predictions (output) [14, 58–62]. Another common UQ problem is the inverse problem, in which uncertainties in the model parameters are determined based on the uncertainties of the observations of the output variables [14, 58, 63–67]. Other UQ problems include the model calibration problem and the certification problem [55]. Forward and inverse problems in UQ will be addressed in this dissertation.

In combustion, particularly chemical kinetics, UQ has received increasing attention from the community [14, 15, 17–19, 58, 59, 67–76]. There are mainly two classes of methods, namely the probabilistic and the deterministic methods. In probabilistic methods, uncertainties are propagated forward or backward using Monte-Carlo methods [18, 19, 59, 67, 72, 73]. One deterministic method, namely the Bound-to-Bound Data Collaboration (B2BDC) method, combines response surfaces using solution mapping techniques with optimization techniques to solve forward and inverse UQ problems [14, 58, 63, 68–71, 74–78]. The Method of Uncertainty Minimization using Polynomial Chaos Expansions (MUM-PCE) has both flavors of the probabilistic and deterministic methods [17, 79–81]. It calculates credible regions in the model parameters and the associated uncertainties in the model predictions [15].

1.3 Organization of dissertation

This dissertation aims at application of UQ analysis based on the B2BDC deterministic framework to reduced models. Chapter 2 describes the B2BDC framework for uncertainty quantification, the “detailed-reduction” method for reaction model reduction, and the kinetic Monte-Carlo soot oxidation model. New developments are presented in Chapter 3, extending the formalism of B2BDC to assessing and comparing uncertainties of reduced and detailed models. The developed methodology is illustrated with a syngas combustion system in Chapter 4 and with a KMC soot oxidation system in Chapter 5. Finally, I will summarize the findings in Chapter 6.

Chapter 2

Methodology

This chapter ¹ describes the B2BDC framework for uncertainty quantification, the “detailed-reduction” method for reaction model reduction, and the kinetic Monte-Carlo soot oxidation model .

2.1 Bound-to-Bound Data Collaboration (B2BDC)

Bound-to-Bound Data Collaboration (B2BDC) is a deterministic uncertainty quantification framework for combining models and training data from multiple sources to explore their collective information content. The numerical methodology is built on specialized constrained-optimization algorithms; it makes predictions on a rigorously determined feasible set, transfers the uncertainties of both model parameters and training-set experiments directly into prediction, tests and quantifies consistency among data and models, explores sources of inconsistency, discriminates among differing models, and enables analysis of global sensitivities of uncertainty in prediction to the uncertainties in the data and model. The B2BDC framework was developed and applied in a series of studies [14, 58, 68–71, 74–78] and was shown to be congruent with the Bayesian sampling approach for uncertainty quantification [82].

2.1.1 Prior knowledge of uncertain parameters

Consider a model $f_e(x)$ used to predict a quantity of interest (QOI), where $x \in \mathbb{R}^n$ is the uncertain parameter vector, and the subscript e denotes the e -th QOI. In B2BDC, the prior knowledge of x is represented by its bounds and is refined by incorporating constraints on QOI predictions, typically through experimental data. This refinement process leads to posterior knowledge of x , represented by bounds and polynomial constraints. The refined

¹The material presented in Section 2.3 of this chapter is based on the paper published in Combustion and Flame [20].

set of parameters x is used to make predictions of QOIs, represented by uncertainty intervals and possibly reduced in size compared to those defined on the prior knowledge of x .

The prior knowledge of x is comprised of an n -dimensional hyperrectangle

$$\mathcal{H} := \{x \in \mathbb{R}^n : l_i \leq x_i \leq u_i, \quad i = 1, \dots, n\}, \quad (2.1)$$

where l_i and u_i are the lower and upper bounds of x_i , the i -th element of x , as shown in Fig. 2.1. These bounds are obtained from experimental data, literature recommendations, or assessed by domain experts. They represent the current “literature recommendations” [58].

2.1.2 Surrogate models

Surrogate models, sometimes called model emulators, are used broadly in many engineering fields because they are less expensive to evaluate than physics-based models [54, 60–62, 64, 83, 84]. In B2BDC, a surrogate model $M_e(x)$ is used to approximate $f_e(x)$, usually in a polynomial form. A single evaluation of $f_e(x)$ may (and typically does) require solving a system of differential equations and could be computationally expensive, whereas $M_e(x)$ can be evaluated much faster. More importantly, expressing $M_e(x)$ in a quadratic (or rational quadratic or other specific polynomial) form enables utilization of powerful optimization algorithms for uncertainty quantification [78, 85, 86].

Surrogate models are built using the solution mapping methodology [54, 85–87]. First, active model parameters are selected using sensitivity analysis. Not all model parameters in x “contribute” equally to the model response, and hence, to the fitting of surrogate models [54]. Oftentimes, only a small subset of model parameters are important for a given QOI. This phenomenon was termed effect sparsity by Box and Meyer [88]. The local sensitivity of the model prediction $f_e(x)$ with respect to the model parameter x_i is given by

$$S_i^e = \frac{\partial f_e(x)}{\partial x_i}. \quad (2.2)$$

However, the local sensitivity coefficient, S_i^e , does not take into account the uncertainty of x_i . A model parameter with a large local sensitivity but a very small uncertainty may contribute minimally to the variation in the model response. In contrast, a model parameter with a small local sensitivity but a very large uncertainty may lead to a large variation in the model response. The impact of a model parameter on the model response is determined by a quantity termed the impact factor

$$IF_i^e = \frac{\partial f_e(x)}{\partial x_i} \times (u_i - l_i). \quad (2.3)$$

The local sensitivity coefficient defined in Eq. (2.2) can be evaluated in a number of ways. In the so-called “brute force” method, two computer simulations are performed, with x_i set to two distinct values, e.g., l_i and u_i , whereas all the other model parameters are held at their nominal values [54]. The sensitivity $\frac{\partial f_e(x)}{\partial x_i}$ is then estimated from $\frac{\Delta f_e(x)}{\Delta x_i}$ with the results

from the two computer simulations. Other methods can be found in [54, 86]. The active parameters are then determined by the ranking of the impact factors. The selection criteria can be highly problem specific [78].

After that, Latin Hypercube Sampling (LHS) is used to generate a sample of design points, $\{x^{(i)}\}_{i=1}^m$, from the hyperrectangle \mathcal{H} as defined in Eq. (2.1). The number of design points m depends on the number of active model parameters. Then, the original model is evaluated at these design points, leading to a sample of model predictions, $\{f_e(x^{(i)})\}_{i=1}^m$. The next step is to fit a surrogate model using the sample of design points (input) and the sample of model predictions (output). Depending on the form of the surrogate model, the sample is split differently in the fitting process. If there are no hyperparameters in a surrogate model, then the sample is split into a training set and a test set. If there are hyperparameters in a surrogate model, e.g., a rational quadratic model, an extra validation set can be used to help select hyperparameters [89]. Alternatively, k-fold cross-validation [89] can be used to select hyperparameters without the use of an extra validation set; this scheme is adopted in B2BDC in fitting rational quadratic models [90]. The sample is simply split into a training set and a test set for a quadratic surrogate model because there are no hyperparameters. A quadratic surrogate model is fitted by minimizing the sum of squared errors using points from the training set, and then the mean absolute error is calculated using the test set:

$$\epsilon_e = \frac{1}{m_{\text{test}}} \sum_{i \in I_{\text{test}}} |M_e(x^{(i)}) - f_e(x^{(i)})|,$$

where the subscript e refers to the e -th initial condition, M_e is the fitted surrogate model, f_e is the original model, m_{test} is the number of points in the test set, and I_{test} is the set of indices of the design points in the test set. The error defined in this fashion is the average of the absolute error, which is suitable when models f_e and M_e are the natural logarithms of quantities of interest (QOIs). Other forms of “mean error” such as the root-mean-square deviation can be used to assess the accuracy of surrogate models as well. In practice, only surrogate models with a low fitting error are included in the B2BDC UQ analysis. The desired accuracy level of a surrogate model is problem dependent. Strategies like piece-wise fitting [91] and iterative fitting [92] have been developed to deal with large fitting errors.

2.1.3 Posterior knowledge of uncertain parameters

In the B2BDC framework, the e -th QOI, associated with an experimental observation, is represented by its lower bound, L_e , and upper bound, U_e in the form of an interval, $[L_e, U_e]$. This uncertainty interval, together with the corresponding surrogate model, define a dataset unit, (M_e, L_e, U_e) , and their collection, for $e = 1, \dots, N$, forms a dataset [58]. QOI bounds are applied to constrain the corresponding surrogate models, thus defining the feasible set

$$\mathcal{F} := \bigcap_{e \in \{1, \dots, N\}} \{x \in \mathcal{H} : L_e \leq M_e(x) \leq U_e\}, \quad (2.4)$$

where N is the number of QOIs in the dataset. For clarity of notations, all $M_e(x)$'s have x as the input. Yet x can mean different (small) subsets of active parameters for different surrogate models, as discussed in the previous subsection. The union of different subsets of active parameters forms the entire parameter set used in the B2BDC optimization. As illustrated in Fig. 2.1, the shaded region in the parameter space is the feasible set \mathcal{F} . The green point in the parameter space is in the feasible set \mathcal{F} because the associated values of the surrogate models are within the QOI bounds in the QOI space. In contrast, the red point is not in the feasible set \mathcal{F} .

The feasible set \mathcal{F} represents the posterior region of the uncertain parameter vector x . A dataset is referred to as consistent if \mathcal{F} is non-empty. In certain cases, the dataset is inconsistent, which can be caused by model inadequacies or measurement errors in experimental observations, etc. In such cases, analysis based on the scalar consistency measure (SCM) [68] or the vector consistency measure (VCM) [76] can be used to reveal the sources of inconsistencies.

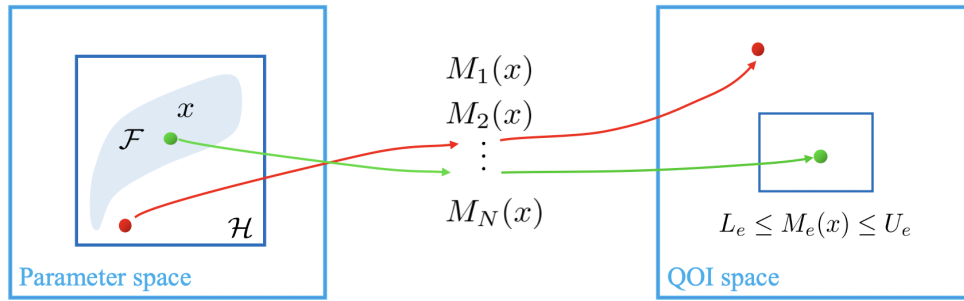


Figure 2.1: Prior region, \mathcal{H} , and posterior region, \mathcal{F} , of the model parameters

2.1.4 Model predictions

If \mathcal{F} is non-empty, the predicted interval of a QOI can be calculated as

$$\left[\min_{x \in \mathcal{F}} M_p(x), \max_{x \in \mathcal{F}} M_p(x) \right], \quad (2.5)$$

where $M_p(x)$ is the surrogate model of a prediction QOI (the p -th QOI), which is typically not used in the polynomial constraints in Eq. (2.4). The prediction interval captures the uncertainties in the model predictions. For simplicity, let \underline{M}_p and \overline{M}_p denote the two interval ends in problem (2.5), respectively. Problem (2.5) comprises two optimization problems:

$$\begin{aligned} \underline{M}_p &= \min_x M_p(x) & (2.6) \\ \text{s.t. } & l_i \leq x_i \leq u_i, \quad i = 1, \dots, n, \\ & L_e \leq M_e(x) \leq U_e, \quad e = 1, \dots, N, \end{aligned}$$

and

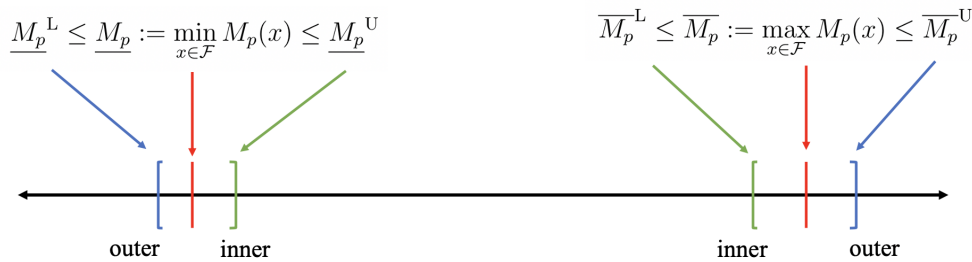
$$\begin{aligned} \overline{M}_p &= \max_x M_p(x) \\ \text{s.t. } & l_i \leq x_i \leq u_i, \quad i = 1, \dots, n, \\ & L_e \leq M_e(x) \leq U_e, \quad e = 1, \dots, N, \end{aligned} \quad (2.7)$$

Problems (2.6) and (2.7) are usually difficult to solve [78]. In a general case, when the optimization problem is not convex, it is computationally infeasible to find \underline{M}_p and \overline{M}_p . The quadratic (rational quadratic, or certain polynomial) form of the surrogates makes it possible to bracket \underline{M}_p or \overline{M}_p by two solvable optimization problems [78, 93]. Methods for efficiently bounding these prediction intervals were developed in [78] and are part of the B2BDC software package [90]. As shown in Fig. 2.2, a lower bound, \underline{M}_p^L , and an upper bound, \underline{M}_p^U , such that $\underline{M}_p^L \leq \underline{M}_p \leq \underline{M}_p^U$, can be obtained. The lower bound \underline{M}_p^L is obtained by relaxing the optimization problem in Eq. (2.6) to a semi-definite programming (SDP) problem. An SDP problem is a convex optimization problem [93] and can be solved efficiently with packages like SeDuMi [94]. The relaxation can either be done by the so-called \mathcal{S} -procedure [78, 93] or by dropping the rank-one constraint in formulating the problem as an optimization problem with linear matrix inequality (LMI) constraints [78]. The upper bound \underline{M}_p^U can be obtained by any feasible x . A global optimization algorithm is used to search for an upper bound \underline{M}_p^U as close to \underline{M}_p as possible. The MATLAB function *fmincon* [95] is used for this purpose as part of the B2BDC software package [90]. Similarly, \overline{M}_p^L and \overline{M}_p^U , such that $\overline{M}_p^L \leq \overline{M}_p \leq \overline{M}_p^U$, can be obtained. In optimization, an inner bound is an upper bound for the minimum or a lower bound for the maximum, whereas an outer bound is a lower bound for the minimum or an upper bound for the maximum. Therefore, $[\underline{M}_p^U, \overline{M}_p^L]$ is termed the inner-bound interval of $[\underline{M}_p, \overline{M}_p]$ and $[\underline{M}_p^L, \overline{M}_p^U]$ its outer-bound interval. The latter are conservative estimates, but they certify containment of the interval $[\underline{M}_p, \overline{M}_p]$. Both the inner- and outer-bound prediction intervals are used in assessing the prediction interval $[\underline{M}_p, \overline{M}_p]$. In many cases, inner- and outer-bound prediction intervals are very close to each other. To make the graphical and tabular presentation clear, the midpoints of the inner- and outer-bound prediction intervals will be used.

Replacing the objective function $M_p(x)$ in problem (2.5) with x_i for any $i = 1, \dots, n$ leads to a special case of problem (2.5),

$$\left[\min_{x \in \mathcal{F}} x_i, \max_{x \in \mathcal{F}} x_i \right]. \quad (2.8)$$

The optimization problem (2.8) solves for the prediction interval of x_i , i.e., the minimum and maximum of the projection of the feasible set \mathcal{F} onto the x_i -axis.

Figure 2.2: Inner- and outer-bound prediction intervals of $[M_p, \overline{M}_p]$

2.1.5 Sensitivities of uncertainties in model predictions

B2BDC also offers an efficient way to determine the sensitivities of uncertainties in model predictions with respect to uncertainties in both experimental observations and model parameters [69].

Methods used in B2BDC to solve problems (2.6) and (2.7) yield Lagrange multipliers that are partial derivatives of \overline{M}_p or M_p with respect to QOI bounds or parameter bounds [69]. These Lagrange multipliers, $\frac{\partial \overline{M}_p}{\partial U_e}$, $\frac{\partial \overline{M}_p}{\partial L_e}$, $\frac{\partial M_p}{\partial L_e}$, $\frac{\partial M_p}{\partial U_e}$, $\frac{\partial \overline{M}_p}{\partial u_i}$, $\frac{\partial \overline{M}_p}{\partial l_i}$, $\frac{\partial M_p}{\partial l_i}$, and $\frac{\partial M_p}{\partial u_i}$, provide important information on whether and by how much changing L_e , U_e , l_i or u_i would alter the prediction interval. These Lagrange multipliers are the by-products of the algorithms adopted in B2BDC and do not require any additional computational efforts, making B2BDC highly efficient in assessing sensitivities. Since the prediction interval is $\overline{M}_p - M_p$, the sensitivity of the p -th QOI prediction interval length with respect to the e -th QOI bounds is given by

$$S_{p,e} = \frac{1}{2} \left(\frac{\partial \overline{M}_p}{\partial U_e} - \frac{\partial \overline{M}_p}{\partial L_e} + \frac{\partial M_p}{\partial L_e} - \frac{\partial M_p}{\partial U_e} \right), \quad (2.9)$$

and its sensitivity with respect to the i -th parameter bounds by

$$S_{p,i} = \frac{1}{2} \left(\frac{\partial \overline{M}_p}{\partial u_i} - \frac{\partial \overline{M}_p}{\partial l_i} + \frac{\partial M_p}{\partial l_i} - \frac{\partial M_p}{\partial u_i} \right). \quad (2.10)$$

Sensitivities can be computed using either inner- or outer-bound prediction intervals.

2.2 “Detailed-reduction” method for reaction mechanisms

The “detailed-reduction” method [22–24] is a reaction mechanism reduction method that identifies non-contributing reactions by automated comparison of reaction and heat-release

rates to selected thresholds, namely,

$$\begin{aligned} r_{f,i} &< \epsilon_r \max_j(r_{f,j}), \quad j \in \mathcal{J}_{\text{ref}} \\ r_{r,i} &< \epsilon_r \max_j(r_{r,j}), \quad j \in \mathcal{J}_{\text{ref}} \\ |r_{\text{net},i} \Delta H_i| &< \epsilon_q \max_j(|r_{\text{net},j} \Delta H_j|), \quad j = 1, \dots, n, \end{aligned} \quad (2.11)$$

where $r_{f,i}$, $r_{r,i}$, and $r_{\text{net},i} = r_{f,i} - r_{r,i}$ are the forward, reverse, and net reaction rates of reaction i , respectively, ΔH_i is the enthalpy change of reaction i , n is the number of reactions in the detailed model, \mathcal{J}_{ref} is the set of indices of selected reference reactions, and ϵ_r and ϵ_q are two cutoff coefficients that are less than unity. The reference reactions are usually the main chain-controlling reactions of a reaction mechanism, e.g., $\text{H} + \text{O}_2 \rightleftharpoons \text{OH} + \text{O}$. If criteria (2.11) are satisfied during the phenomena of interest, e.g., over the entire induction period in ignition delay times, a reaction is identified as non-contributing. Repeating this analysis for selected values of ϵ_r and ϵ_q at a set of simulation conditions and removing the intersection of the identified non-contributing reactions from the initial, full-size reaction set produces a (ϵ_r, ϵ_q) -level reduced reaction model.

The “detailed-reduction” method is based on the postulate that if a reduced model can faithfully describe the dynamics of both the thermal and chain reaction processes of a detailed model in a simulation of reactive flows, then it will reproduce the chemical processes with approximately the same degree of accuracy [22]. What Eqs (2.11) imply is that reactions with low heat release rates and reaction rates can be safely removed without causing a significant change of the dynamics of the system. The “detailed-reduction” method is highly efficient in that it can remove reactions in batch and it is obtained alongside the solution of a simulation with the detailed model. Moreover, different reduced models with varying levels of accuracy (different (ϵ_r, ϵ_q) -levels) can be derived from one single simulation with the detailed model. Another benefit is that the “detailed-reduction” method is very general and can be adapted to reaction models beyond traditional mechanistic reaction models, e.g., the KMC-based soot model.

2.3 KMC-based soot oxidation model

The latest detailed model of soot oxidation developed by the Frenklach group employs a kinetic Monte-Carlo simulation with molecular mechanics relaxation (KMC-MMR) technique [20]. The modeling of soot oxidation followed the generally accepted supposition that carbon black and soot are structurally analogous to each other [96–98], soot particles are composed of polycyclic aromatic hydrocarbon (PAH) molecules [10, 99, 100], and, hence, soot particle oxidation was modeled as sequential oxidation of these PAH molecules, i.e., when one PAH molecule is finishing its oxidation, the oxidation continues with another PAH molecule.

The PAH substrates selected for the modeling of soot oxidation are shown in Fig. 2.3. Four of them—7, 19, 37, and 61 ring aromatics—are all of a similar shape, hexagon. The

25-ring structure is of a different shape, rhombus. The PAH selection for the soot oxidation model was employed to examine the influence of the substrate size and shape on the rate of oxidation and to cover their reported size [20]. All substrate edge sites were assumed to be saturated with H atoms.

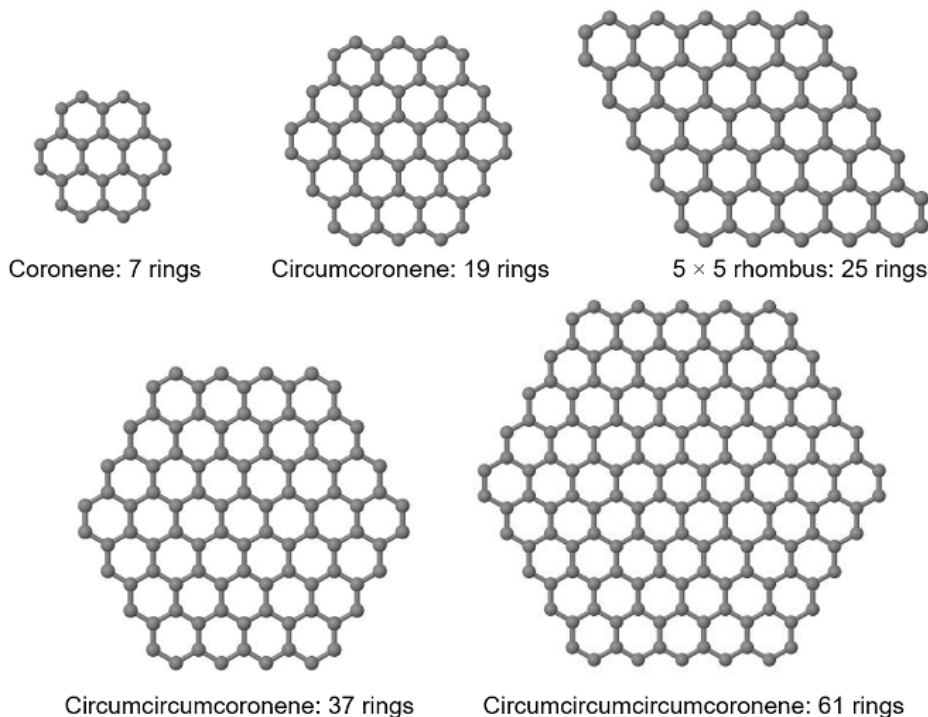


Figure 2.3: Initial substrates for KMC-MMR simulations. The H atoms saturating the edge carbon atoms are not shown for clarity.

The surface reaction model employed in the detailed soot oxidation model is reproduced in Table 2.1. The kinetics of the surface reactions, (S1) – (S107), was modeled using KMC-MMR, following the methodology established in the prior studies [43, 49, 51, 52]. Briefly, a simulation tracked a single aromatic “molecule” evolving in a specified gas-phase environment. A simulation ran as a sequence of Markov processes, i.e., each reaction event depending only on the current state of the molecular structure and not on the previous states. At each KMC time step, the reacting site and the reaction were selected stochastically, following the Gillespie algorithm [47, 48] adapted for surface processes [45, 46]. The algorithm can be summarized as follows. Given an instant of a current reaction event, t_n , the time of the next reaction event to occur at aromatic-edge site i , $t_{n+1,i}$, is

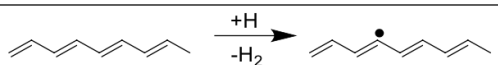
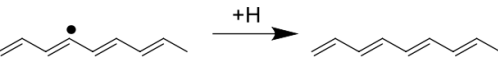
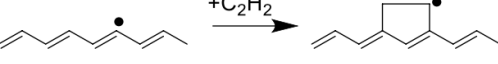
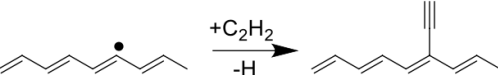
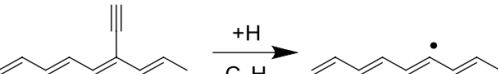
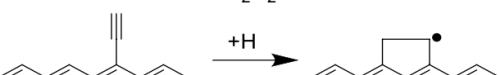
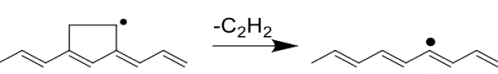
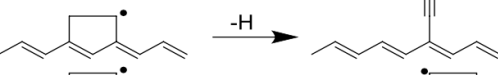
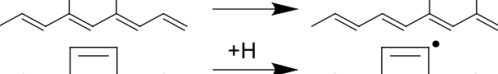
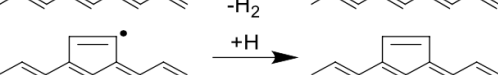
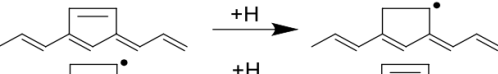
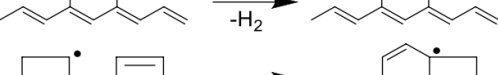
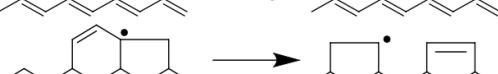
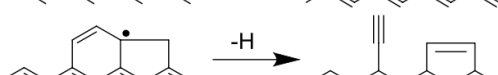


$$t_{n+1,i} = t_n - \frac{\ln(u)}{\sum_j k_{j,i}}, \quad (2.12)$$

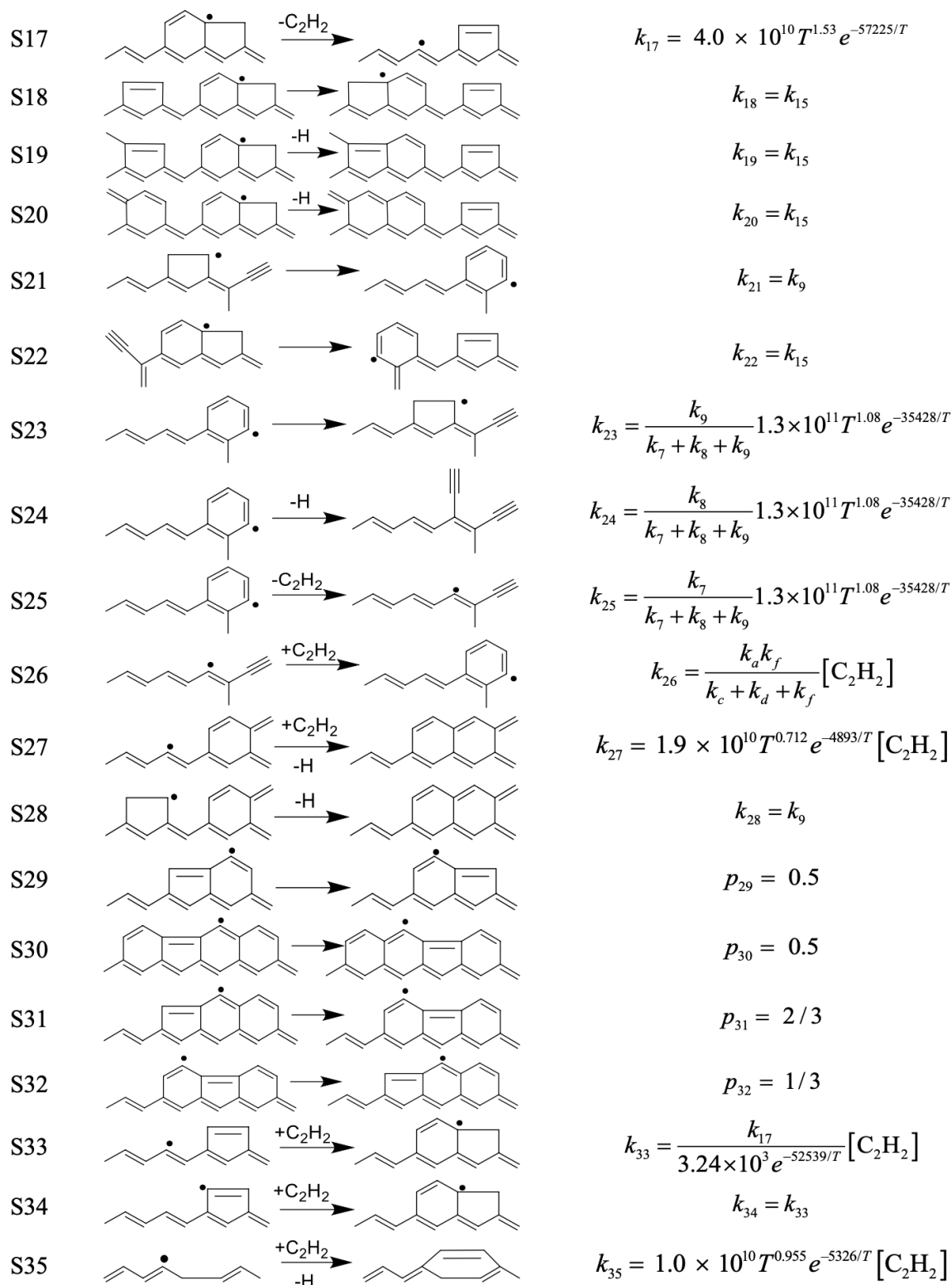
where u is a random number distributed uniformly within $[0, 1]$, $k_{j,i}$ is the per-site rate constant of reaction j at site i , and the sum $\sum_j k_{j,i}$ is taken over all reactions possible to occur at site i . The smallest among the $t_{n+1,i}$ values, computed for all surface reaction sites, becomes the time instant of the next reaction event, t_{n+1} , and the particular reaction to occur at that time is chosen according its probability $p_{j,i} = \frac{k_{j,i}}{\sum_j k_{j,i}}$ upon drawing another random number within $[0, 1]$. To account for the evolving curvature, at the end of the KMC step, the molecular structure is subjected to molecular-mechanics geometry optimization with the TINKER package [101] using the MM3 potential [102]. The process then repeats itself. Simulations for each set of initial conditions were repeated multiple times, each with a different starting random seed.

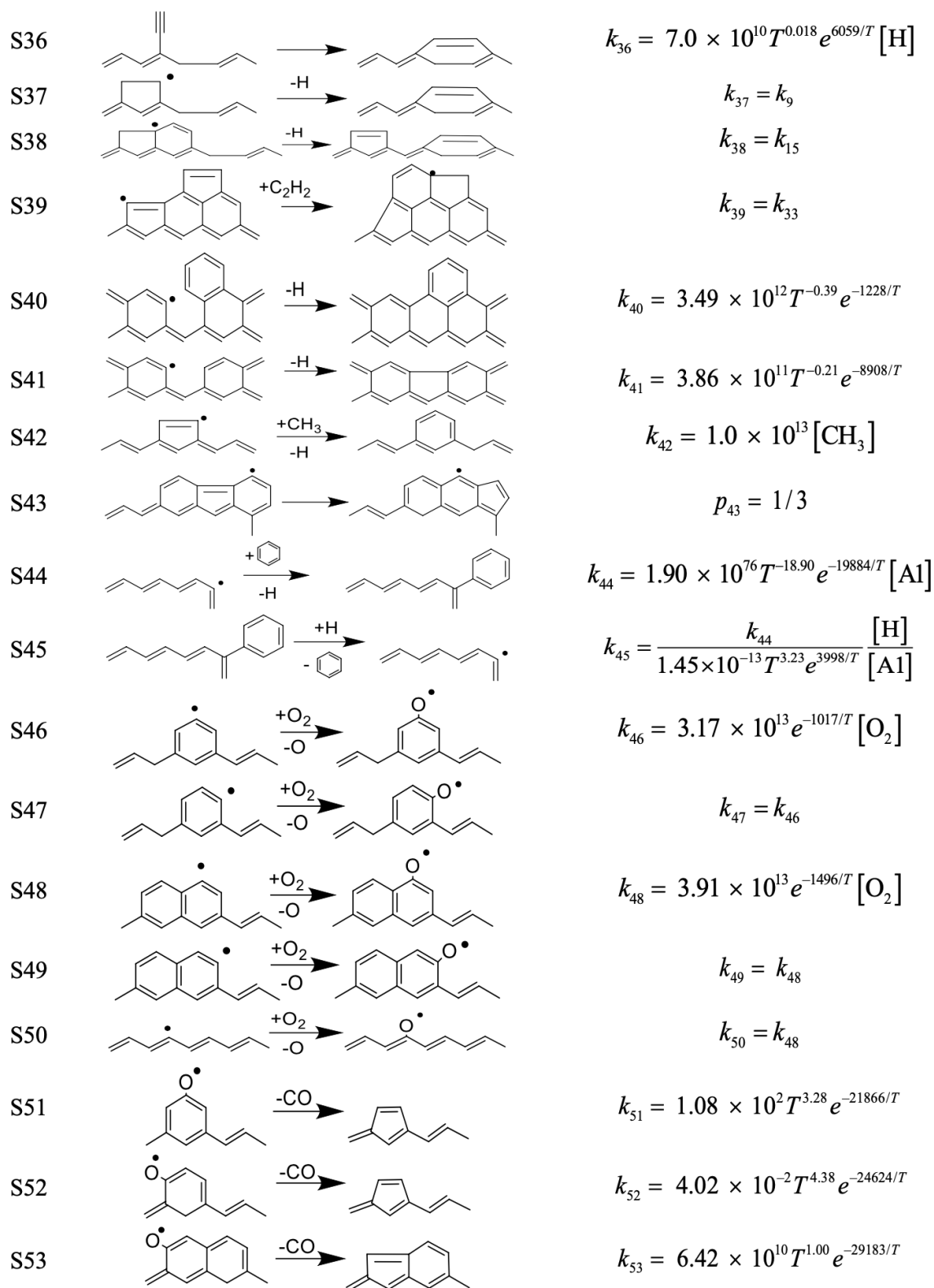
The algorithm requires that the $k_{j,i}$ values appearing in Eq. (2.12) are numerical constants [48]. The surface reactions are of two types: unimolecular “decompositions” of surface species, and bimolecular reactions between gaseous species and surface sites. The unimolecular reactions follow the first-order rate law and, hence, their rate coefficients are numerical constants by definition. The surface bimolecular reactions can be treated as pseudo-first-order processes, through multiplication of their gas-surface rate coefficients by the concentrations of the corresponding gaseous reactants. When the latter remain constant during the simulation, as in prior numerical studies [43, 49, 51, 52], the gas-surface KMC stochastic events can also be treated as first-order processes of the respective surface sites. The detailed soot oxidation model can work with changing gas-phase environments using the surface-gas coupling algorithm developed in [20]. It was validated against the high-temperature shock-tube experiments of Roth et al. [103]. The reaction model was able to reproduce the experimental results, but it required the coupling to particle nanostructure: partial oxidation of PAH molecules and the decrease in PAH initial sizes along the oxidation path.

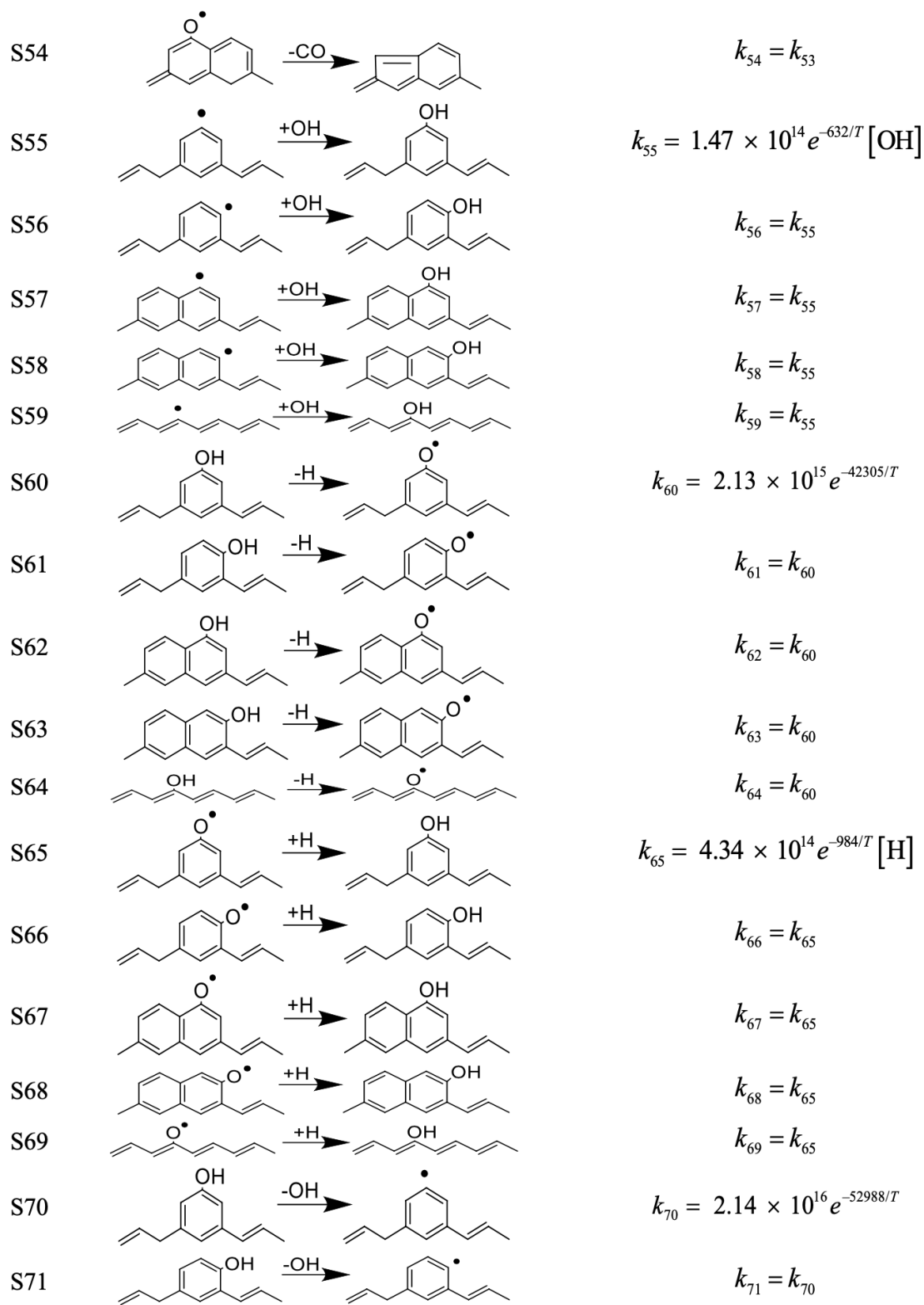
For this dissertation, constant gas-phase environments will be used to study model reduction with the detailed soot model, similar to the simulations in [43, 49, 51, 52]. The gas-phase environment is determined by the scenario variables, including the pressure, temperature, and mole fractions of species such as H, H₂, O₂, O, and OH. The simulation terminates after a preset simulation time. The counts of reactions in Table 2.1 can be used to identify non-contributing reactions and construct reduced models, similar to the “detailed-reduction” method introduced in Section 2.2.

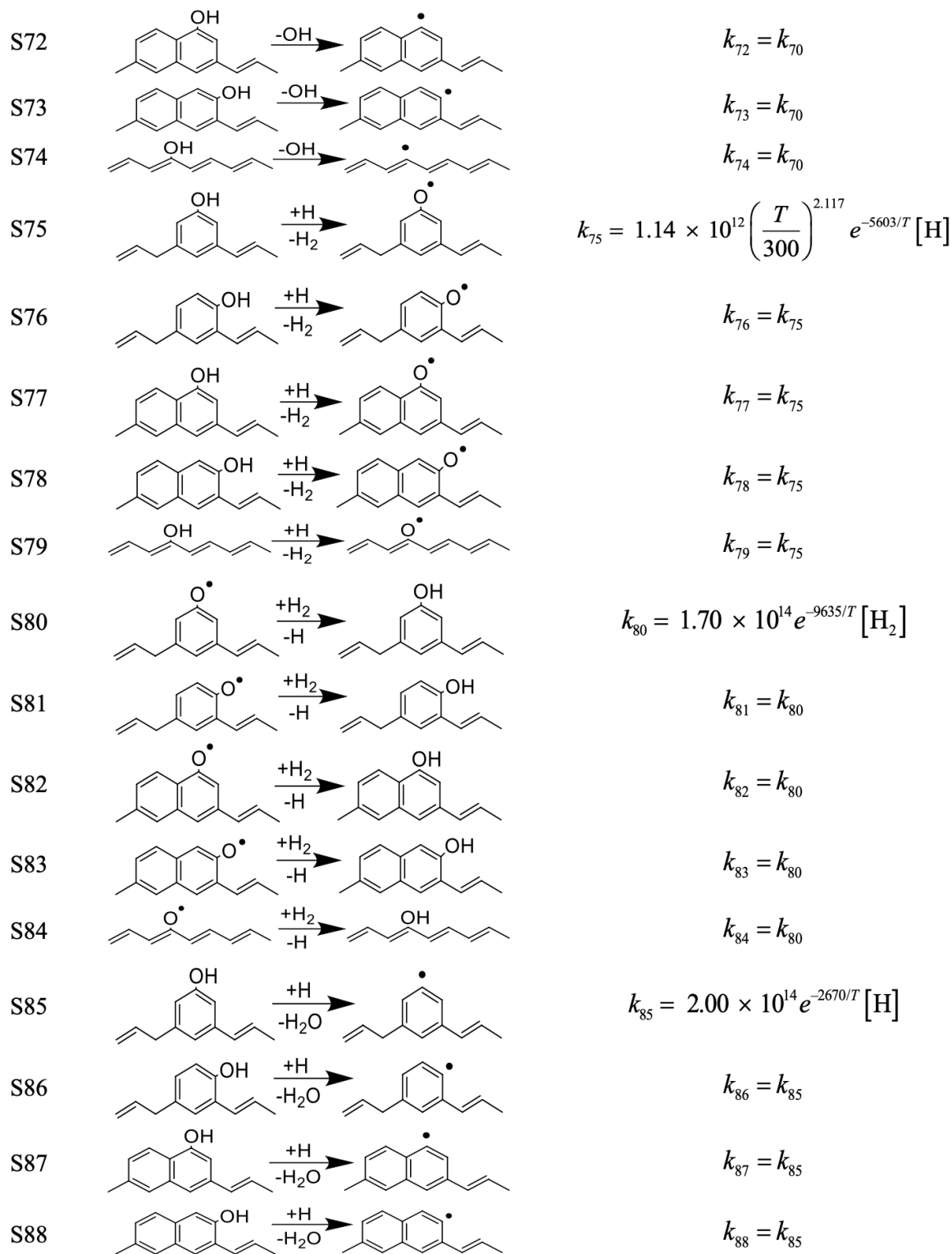
Table 2.1: Surface reactions of the detailed soot oxidation model.

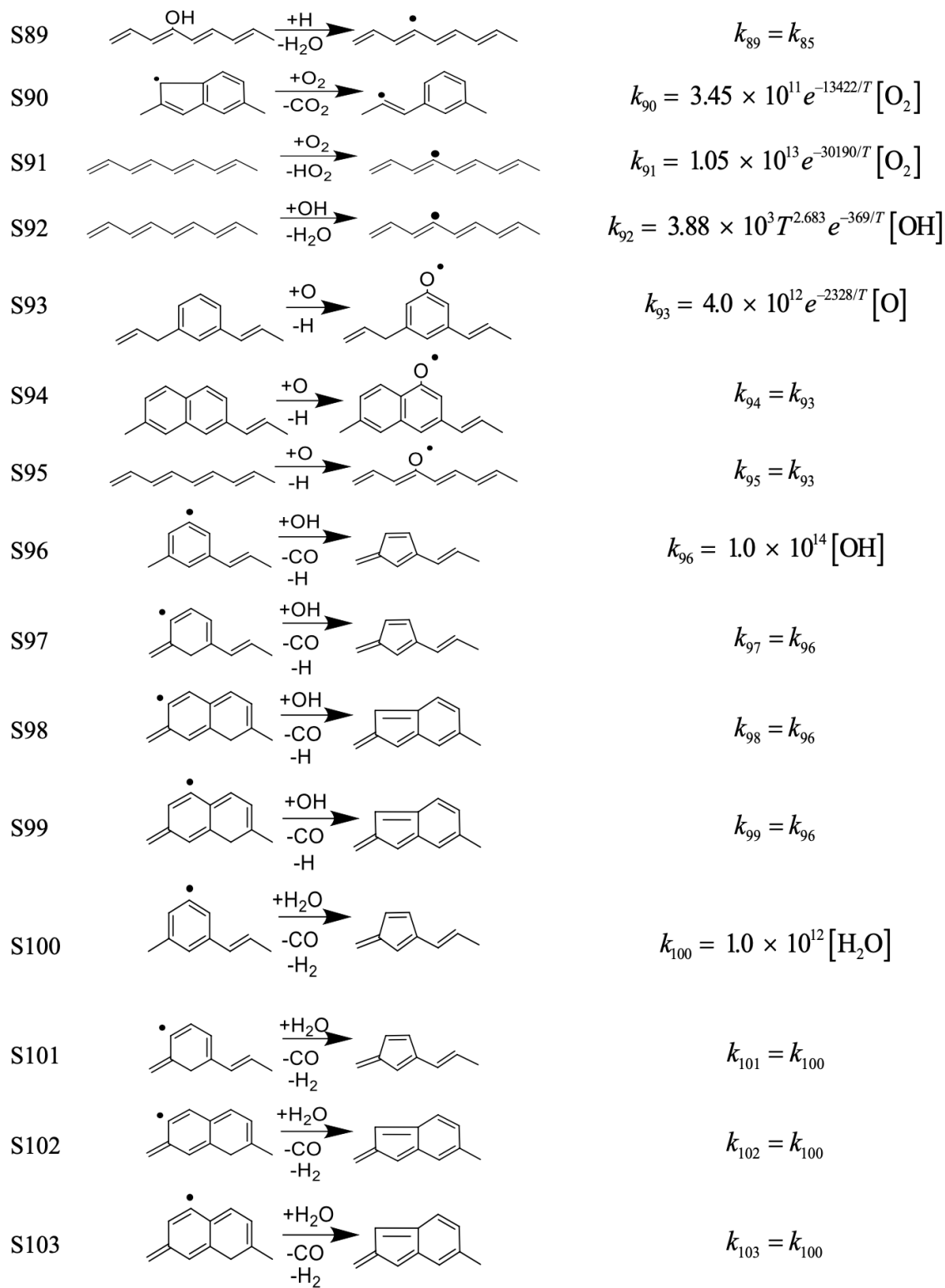
	reaction	per-site rate coefficient ^a (s, mol/cm ³ , K)
S1		$k_1 = 4.2 \times 10^{13} e^{-8052/T} [\text{H}]$
S2		$k_2 = 2 \times 10^{13} [\text{H}]$
S3		$k_3 = \frac{k_a k_b}{k_b + k_c + k_d} [\text{C}_2\text{H}_2]$
S4		$k_4 = \frac{k_a k_d}{k_b + k_c + k_d} [\text{C}_2\text{H}_2]$
S5		$k_5 = \frac{k_c k_e}{k_b + k_c + k_d} [\text{H}]$
S6		$k_6 = \frac{k_b k_e}{k_b + k_c + k_d} [\text{H}]$
S7		$k_7 = 3.1 \times 10^{11} T^{0.87} e^{-37403/T}$
S8		$k_8 = 6.7 \times 10^{11} T^{0.84} e^{-35625/T}$
S9		$k_9 = 1.3 \times 10^{11} T^{0.16} e^{-23099/T}$
S10		$k_{10} = 5.07 \times 10^7 T^{1.93} e^{-6518/T} [\text{H}]$
S11		$k_{11a} = 6.08 \times 10^{12} T^{0.27} e^{-141/T} [\text{H}]$
S12		$k_{12} = 5.40 \times 10^{11} T^{0.45} e^{-916/T} [\text{H}]$
S13		$k_{13a} = 2 \times 10^{12} [\text{H}]$
S14		$k_{14} = 8.9 \times 10^5 T^{2.28} e^{-30944/T}$
S15		$k_{15} = 2.1 \times 10^9 T^{1.14} e^{-41952/T}$
S16		$k_{16} = 3.8 \times 10^{10} T^{1.30} e^{-51929/T}$

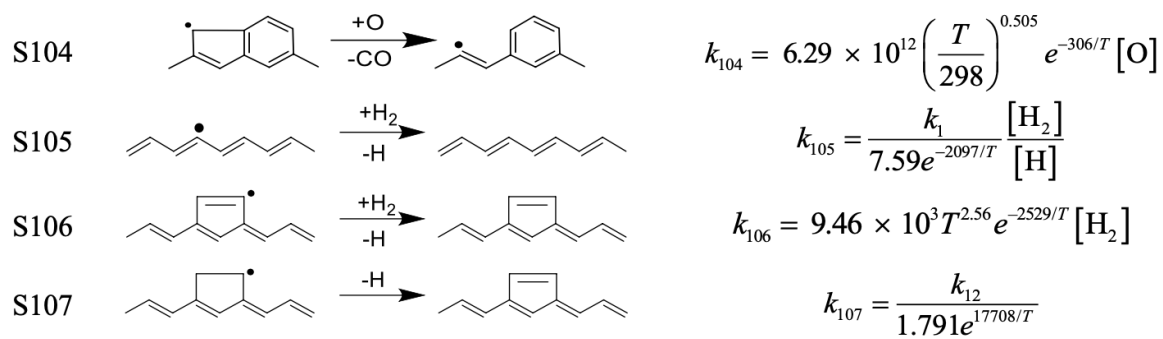












a For reactions 3–6 and 26: $k_a = 1.1 \times 10^7 T^{1.71} e^{-1960/T}$, $k_b = 6.8 \times 10^{11} e^{-11084/T}$,
 $k_c = 1.3 \times 10^{14} e^{-21025/T}$, $k_d = 4.8 \times 10^{12} e^{-16875/T}$, $k_e = 1.5 \times 10^{10} T^{0.85} e^{-601/T}$,
 $k_f = 2.5 \times 10^{12} T^{-0.13} e^{-7902/T}$

Chapter 3

B2BDC-based uncertainty quantification of reduced models

In this chapter, I introduce different measures to assess how uncertainties in the model parameters and in the model predictions behave for reduced models as compared to those for the detailed model. The methods introduced can be applied to models outside the combustion domain, although combustion models are used as examples in this dissertation. To illustrate the methods, a detailed model, $f_e^D(x)$, and a reduced model, $f_e^R(x)$ will be considered. Both models are used to predict the e -th quantity of interest (QOI) for $e = 1, \dots, N$, where N is the number of QOIs in a dataset. Superscripts D and R are used to distinguish the detailed and reduced models. Both models take as input x , the vector of uncertain parameters, with x_i being the i -th element of x . However, this does not mean that all parameters in x need to be active for either model. Some parameters in x may not be present at all in a reduced model. Typically, both the detailed and reduced models have different (small) subsets of active parameters. For simplicity of notation, x is used as the input for both models. Superscripts D and R are used for other variables as well, e.g., the surrogate models associated with $f_e^D(x)$ and $f_e^R(x)$ are $M_e^D(x)$ and $M_e^R(x)$, respectively.

3.1 Comparison of predicted values

Deviations between QOI predictions computed with the detailed and reduced models are examined here, and they can be expressed using three measures. The first measure is similar to the typical approach [16, 22, 23, 25–27, 29, 32, 34, 37–39, 104, 105]; it is a difference computed at a single parameter point and, thus, does not include parameter uncertainties,

$$\delta_e = \left| M_e^R(x^{\text{nominal}}) - M_e^D(x^{\text{nominal}}) \right|, \quad (3.1)$$

where M_e^R and M_e^D are the surrogate models associated with the reduced and detailed models, respectively, for the e -th QOI. Both surrogate models are evaluated at x^{nominal} , the nominal uncertain parameter vector.

The second measure expresses the deviations over the *prior region* of model parameters, \mathcal{H} ,

$$\delta_e^{\mathcal{H}} = \max_{x \in \mathcal{H}} |M_e^{\text{R}}(x) - M_e^{\text{D}}(x)|, \quad (3.2)$$

where $\mathcal{H} = \{x \in \mathbb{R}^n : l_i \leq x_i \leq u_i, \quad i = 1, \dots, n\}$, as discussed in Section 2.1.1. The objective function is the absolute value of the difference between the detailed and reduced models in the above optimization problem, which requires solving two optimization problems, $\min_{x \in \mathcal{H}} (M_e^{\text{R}}(x) - M_e^{\text{D}}(x))$ and $\max_{x \in \mathcal{H}} (M_e^{\text{R}}(x) - M_e^{\text{D}}(x))$. In this case, model parameters can take any value within their initially prescribed individual bounds, but no target predictability is included in the analysis. This prior-knowledge only measure is similar to the treatments of uncertainties in [17–19].

The third measure expresses the deviation over the *posterior region* of model parameters, \mathcal{F}^{D} ,

$$\begin{aligned} \delta_p^{\mathcal{F}} &= \max_x |M_p^{\text{R}}(x) - M_p^{\text{D}}(x)| \\ \text{s.t. } x \in \mathcal{F}_{-p}^{\text{D}} &= \bigcap_{\substack{e \in \{1, \dots, N\} \\ e \neq p}} \{x \in \mathcal{H} : L_e \leq M_e^{\text{D}}(x) \leq U_e\}, \end{aligned} \quad (3.3)$$

where L_e and U_e are the e -th QOI lower and upper bounds, usually obtained from experimental observations, $\mathcal{F}_{-p}^{\text{D}}$ is the feasible set established in the detailed-model setting that implements the leave-one-out scheme, in which the p -th QOI is excluded in the constraints; the calculations of Eq. (3.3) can be repeated for $p = 1, \dots, N$. The deviations in this case, $\delta_p^{\mathcal{F}}$, represent, in addition to the assumed bounds on model parameters, the target tolerances of the dataset. This measure treats the uncertainties in the model parameters in a similar fashion as the treatments of uncertainties in [67, 72, 73], in which posterior distributions of uncertain parameters were obtained using Bayesian analysis. The objective function is the absolute value of the difference between the detailed and reduced models in the above optimization problem, which also requires solving for $\min_x (M_p^{\text{R}}(x) - M_p^{\text{D}}(x))$ and $\max_x (M_p^{\text{R}}(x) - M_p^{\text{D}}(x))$ with the same constraints in Eq. (3.3).

The first measure, Eq. (3.1), can be computed directly at the nominal parameter values. The second and third ones, in Eqs. (3.2) and (3.3), can be obtained by solving respective optimization problems of the B2BDC framework, returning the inner and outer bounds of $\delta^{\mathcal{H}}$ or $\delta^{\mathcal{F}}$, as discussed in Section 2.1.4. The inner and outer bounds bracket the deviation measures and are typically very close to each other. To make the graphical and tabular presentation clear, the $\delta^{\mathcal{H}}$ and $\delta^{\mathcal{F}}$ values of individual QOIs are usually computed using midpoints of the respective inner and outer bounds; i.e., $\delta_e^{\mathcal{H}} = [\delta_e^{\mathcal{H}}(\text{inner}) + \delta_e^{\mathcal{H}}(\text{outer})] / 2$ and $\delta_e^{\mathcal{F}} = [\delta_e^{\mathcal{F}}(\text{inner}) + \delta_e^{\mathcal{F}}(\text{outer})] / 2$.

While the first measure, defined in Eq. (3.1), is overly loose, the second measure, defined in Eq. (3.2), could be overly strict. The third measure, defined in equation Eq. (3.3), overcomes the overly looseness in the first case by allowing x to take any values in some sets in order to account for the uncertainties in the model parameters, and it overcomes the overly

strictness in the second case by restricting the search domain of x to the feasible set. This is achieved by obtaining posterior knowledge of the uncertain parameters using QOI bounds in the form of polynomial constraints on surrogate models.

3.2 Comparison of feasible sets

In B2BDC, the feasible set \mathcal{F} , defined implicitly through Eq. (2.4), represents the posterior region of the uncertain parameter vector x . In other words, the feasible set characterizes the posterior uncertainties in the model parameters. Therefore, it is instructive to investigate the difference between the feasible set obtained for the detailed model, \mathcal{F}^D , and that obtained for a reduced model, \mathcal{F}^R , for the same collection of QOIs. Adding a superscript D or R to appropriate terms in Eq. (2.4) leads to

$$\mathcal{F}^D := \bigcap_{e \in \{1, \dots, N\}} \{x \in \mathcal{H} : L_e \leq M_e^D(x) \leq U_e\}, \quad (3.4)$$

and

$$\mathcal{F}^R := \bigcap_{e \in \{1, \dots, N\}} \{x \in \mathcal{H} : L_e \leq M_e^R(x) \leq U_e\}. \quad (3.5)$$

The relative volume of the feasible set, defined as the ratio of the feasible set volume to the volume of the associated hyperrectangle, $\frac{V(\mathcal{F})}{V(\mathcal{H})}$, is an important quantity of the feasible set. The relative volume of the feasible set for the detailed model, $\frac{V(\mathcal{F}^D)}{V(\mathcal{H})}$, and that for a reduced model, $\frac{V(\mathcal{F}^R)}{V(\mathcal{H})}$, can be compared. For this purpose, the hyper-rectangular region of the prior knowledge \mathcal{H} is sampled uniformly, and rejection sampling [89] is then used to estimate the relative volume of a feasible set by the fraction of points contained in \mathcal{H} that also fall within \mathcal{F}^D or \mathcal{F}^R . The number of points sampled in \mathcal{H} can be determined empirically to ensure the convergence of the estimated relative volume.

In addition, the feasible-set overlap can be estimated. Uniformly-distributed points can be sampled from the feasible set using a Gibbs sampling algorithm [106] adopted to B2BDC [107]. The number of points to be sampled in the feasible set can be determined empirically by a convergence test [107]. After that, the fraction of points contained in \mathcal{F}^R that also fall within \mathcal{F}^D can be calculated, and vice versa. The feasible-set overlap provides, in addition to the volume, a direct measure of the difference between the posterior regions of uncertain parameters.

3.3 Comparison of posterior distributions

A general approach for examining the propagation of uncertainty in reduced models is by comparing QOI posterior distributions, which capture the posterior uncertainties in the model predictions. This analysis is usually preformed in the leave-one-out scheme, in which

the p -th QOI is excluded in the constraints in evaluation of $M_p^D(x)$, repeated for $p = 1, \dots, N$. For the detailed and reduced models, the corresponding feasible sets are

$$\mathcal{F}_{-p}^D = \bigcap_{\substack{e \in \{1, \dots, N\} \\ e \neq p}} \{x \in \mathcal{H} : L_e \leq M_e^D(x) \leq U_e\},$$

and

$$\mathcal{F}_{-p}^R = \bigcap_{\substack{e \in \{1, \dots, N\} \\ e \neq p}} \{x \in \mathcal{H} : L_e \leq M_e^R(x) \leq U_e\},$$

respectively. A uniform sample of x can be collected from each of the \mathcal{F}_{-p}^D or \mathcal{F}_{-p}^R for each $p = 1, \dots, N$, using the Gibbs sampling algorithm adopted to B2BDC [107]; the notation \mathcal{F}_{-p} is used to denote the feasible set for a dataset with the p -th QOI removed. Evaluating $M_p^D(x)$ or $M_p^R(x)$ with the points in the uniform sample collected from \mathcal{F}_{-p}^D or \mathcal{F}_{-p}^R leads to a sample of posterior QOI predictions. The two posterior distributions for the same QOI can then be compared, for example, by plotting the histograms of the two distributions.

In addition to the visual exhibition of the prediction histograms, their difference can be quantified numerically. For this purpose, the Hellinger distance measure can be employed [108],

$$h(P, Q) = \left[\frac{1}{2} \int_{\xi} \left(\sqrt{P(\xi)} - \sqrt{Q(\xi)} \right)^2 d\xi \right]^{1/2}, \quad (3.6)$$

which is designed to gauge the difference between two distributions, $P(\xi)$ and $Q(\xi)$. The Hellinger distance metric ranges from 0 to 1, with values closer to 0 meaning the two distributions are more similar to one another. Applied to the current case, this measure quantifies the difference between the prediction distributions of the detailed and reduced models and is expressed as

$$h_p^H(D, R) = \left[\frac{1}{2} \sum_{j=1}^{N_{\text{bin}}} \left(\sqrt{Z_{p,j}^D} - \sqrt{Z_{p,j}^R} \right)^2 \right]^{1/2}, \quad (3.7)$$

where N_{bin} is the number of histogram bins, and $Z_{p,j}$ is the normalized count of the sampled QOI predicted values that fall within the j -th bin, the subscript p refers to the p -th QOI, and the superscript H indicates that the Hellinger distance is calculated from histograms. In practice, Hellinger distances calculated using Eq. (3.7) are not stable in that they may change significantly as N_{bin} changes. This is a common issue with histograms in density estimation. To remedy this, kernel density estimation can be used because kernel density estimators are smoother and converge faster [109]. Specifically, the MATLAB function `ksdensity()` [95] was used for this purpose in this dissertation. After applying kernel density smoothing, a suitable N_{bin} can be selected empirically to ensure the convergence of the Hellinger distance.

3.4 Comparison of prediction intervals

Examination of posterior distributions, as discussed in the previous section, provides means to evaluate to what extent the reduced model reproduces propagation of uncertainty by the detailed model. However, such an approach based on direct sampling of x could become particularly expensive because the feasible set could be an extremely small region with a very complicated shape in a high-dimensional space. The underlying Gibbs sampling algorithm adopted to B2BDC [107] used to sample the feasible set uniformly could be substantially time consuming. The B2BDC framework offers an alternative by quantifying the propagated uncertainty through numerically efficient computations of intervals of prediction uncertainty.

For each $p = 1, \dots, N$, the prediction interval of the p -th QOI computed with the detailed model and that computed with the reduced model can be obtained by solving the following B2BDC optimization problems:

$$\left[\min_{x \in \mathcal{F}_p^D} M_p^D(x), \max_{x \in \mathcal{F}_p^D} M_p^D(x) \right] \quad (3.8)$$

where $\mathcal{F}_p^D = \bigcap_{\substack{e \in \{1, \dots, N\} \\ e \neq p}} \{x \in \mathcal{H} : L_e \leq M_e^D(x) \leq U_e\}$,

and

$$\left[\min_{x \in \mathcal{F}_p^R} M_p^R(x), \max_{x \in \mathcal{F}_p^R} M_p^R(x) \right] \quad (3.9)$$

where $\mathcal{F}_p^R = \bigcap_{\substack{e \in \{1, \dots, N\} \\ e \neq p}} \{x \in \mathcal{H} : L_e \leq M_e^R(x) \leq U_e\}$.

The calculations also use the leave-one-out scheme, where the p -th QOI is excluded in the constraints from the evaluation of the p -th prediction interval. The difference in the prediction intervals can be quantified by calculating the Hellinger distance measure defined in Eq. (3.6), assuming that the predicted QOI values are distributed uniformly within the predicted QOI intervals and the two prediction intervals are not disjoint. Under these assumptions, the Hellinger distance becomes

$$h_p^U(D, R) = \left[1 - \frac{\mathcal{L}_p^O}{(\mathcal{L}_p^D \mathcal{L}_p^R)^{1/2}} \right]^{1/2}, \quad (3.10)$$

where \mathcal{L}_p^D and \mathcal{L}_p^R are the lengths of the p -th prediction intervals of the detailed and reduced models, respectively, \mathcal{L}_p^O is the length of the overlapping region of the two intervals, and the superscript U indicates that the Hellinger distance is calculated from uniform samples on the prediction intervals, to be distinguished from the superscript H from the previous

section. The interval lengths are computed from the solutions of Eqs. (3.8) and (3.9), and are expressed as

$$\mathcal{L}_p^D = \overline{M}_p^D - \underline{M}_p^D \quad (3.11)$$

and

$$\mathcal{L}_p^R = \overline{M}_p^R - \underline{M}_p^R, \quad (3.12)$$

where $\overline{M}_p^D = \max_{x \in \mathcal{F}_p^D} M_p^D(x)$, $\overline{M}_p^R = \max_{x \in \mathcal{F}_p^R} M_p^R(x)$, $\underline{M}_p^D = \min_{x \in \mathcal{F}_p^D} M_p^D(x)$, and $\underline{M}_p^R = \min_{x \in \mathcal{F}_p^R} M_p^R(x)$, and the length of the overlap region as

$$\mathcal{L}_p^O = \min(\overline{M}_p^D, \overline{M}_p^R) - \max(\underline{M}_p^D, \underline{M}_p^R). \quad (3.13)$$

Note that it is assumed that the two prediction intervals are not disjoint. If they are indeed disjoint, \mathcal{L}_p^O would be a negative number following Eq. (3.13), leading to a Hellinger distance greater than one. To accommodate this, the Hellinger distance can be set to unity if the two prediction intervals indeed have no overlap.

3.5 Comparison of sensitivities

Addressing the capacity of a reduced reaction model to reproduce the uncertainty propagation of the associated detailed model, one may be interested in identifying the influence of specific model parameters on model reduction; several authors have tackled some of these aspects (e.g., [16, 19]). B2BDC offers an efficient way to determine the sensitivities of the uncertainties in model predictions with respect to the uncertainties in both experimental observations and model parameters [69]. Adding a superscript D to Eqs. (2.9) and (2.10) yields the sensitivity of the p -th QOI prediction interval length with respect to the e -th QOI bounds in the setting of the detailed model

$$S_{p,e}^D = \frac{1}{2} \left(\frac{\partial \overline{M}_p^D}{\partial U_e} - \frac{\partial \underline{M}_p^D}{\partial L_e} + \frac{\partial \overline{M}_p^D}{\partial L_e} - \frac{\partial \underline{M}_p^D}{\partial U_e} \right), \quad (3.14)$$

and its sensitivity with respect to the i -th parameter bounds

$$S_{p,i}^D = \frac{1}{2} \left(\frac{\partial \overline{M}_p^D}{\partial u_i} - \frac{\partial \underline{M}_p^D}{\partial l_i} + \frac{\partial \overline{M}_p^D}{\partial l_i} - \frac{\partial \underline{M}_p^D}{\partial u_i} \right), \quad (3.15)$$

where the partial derivatives appearing in these equations are obtained alongside the solution of the general optimization ansatz of B2BDC [69]. In other words, one does not need to perform additional calculations to obtain these derivatives; they become available with the solution of Eq. (3.8) for the interval bounds. Similarly, $S_{p,e}^R$ and $S_{p,i}^R$ for a reduced model are immediately available with the solution of Eq. (3.9),

$$S_{p,e}^R = \frac{1}{2} \left(\frac{\partial \overline{M}_p^R}{\partial U_e} - \frac{\partial \underline{M}_p^R}{\partial L_e} + \frac{\partial \overline{M}_p^R}{\partial L_e} - \frac{\partial \underline{M}_p^R}{\partial U_e} \right), \quad (3.16)$$

and its sensitivity with respect to the i -th parameter bounds

$$S_{p,i}^R = \frac{1}{2} \left(\frac{\partial \overline{M}_p^R}{\partial u_i} - \frac{\partial \overline{M}_p^R}{\partial l_i} + \frac{\partial \underline{M}_p^R}{\partial l_i} - \frac{\partial \underline{M}_p^R}{\partial u_i} \right). \quad (3.17)$$

Comparing $S_{p,e}^D$ with $S_{p,e}^R$ and $S_{p,i}^D$ with $S_{p,i}^R$ can reveal changes in the sensitivity spectra due to the removal of reactions in reduced models.

One can also examine the sensitivity of the overlap of prediction intervals of the detailed and reduced models. For this purpose, the Bhattacharyya coefficient [110] can be employed

$$\mathcal{B}(P, Q) = \int_{\xi} \sqrt{P(\xi)Q(\xi)} d\xi. \quad (3.18)$$

It is related to the Hellinger distance (3.6),

$$\mathcal{B}(P, Q) = 1 - h^2(P, Q), \quad (3.19)$$

and provides a measure of the amount of overlap between two statistical samples or distributions, P and Q .

Assuming uniform distribution for P and Q , as assumed in Section 3.4 in calculating the Hellinger measure, one can obtain the Bhattacharyya coefficient for the overlap of the prediction intervals of the detailed and reduced models:

$$\mathcal{B}_p^U(D, R) = \frac{\mathcal{L}_p^O}{(\mathcal{L}_p^D \mathcal{L}_p^R)^{1/2}}. \quad (3.20)$$

Differentiating Eq. (3.20) with respect to the uncertainty bounds, following [69], one obtains the sensitivity of the Bhattacharyya coefficient to the uncertainty bounds of QOIs,

$$\beta_{p,e}^R = \frac{S_{p,e}^O}{\mathcal{L}_p^O} - \frac{1}{2} \left(\frac{S_{p,e}^D}{\mathcal{L}_p^D} + \frac{S_{p,e}^R}{\mathcal{L}_p^R} \right), \quad (3.21)$$

and to the uncertainty bounds of the model parameters,

$$\beta_{p,i}^R = \frac{S_{p,i}^O}{\mathcal{L}_p^O} - \frac{1}{2} \left(\frac{S_{p,i}^D}{\mathcal{L}_p^D} + \frac{S_{p,i}^R}{\mathcal{L}_p^R} \right), \quad (3.22)$$

where $S_{p,e}$'s and $S_{p,i}$'s are from Eqs. (3.14) and (3.15), respectively.

3.6 Summary

Within the B2BDC framework, methods to assess how uncertainties in the model parameters and in the model predictions behave for reduced models as compared to those for the detailed

model, were developed. I first established measures to quantify the deviations between the predicted values of the detailed and reduced models when the model parameters are fixed at the nominal values (δ), constrained to the prior region ($\delta^{\mathcal{U}}$), or constrained to the posterior region ($\delta^{\mathcal{F}}$). Second, two measures to quantify the difference between feasible sets of the detailed and reduced models were developed. These two measures are the relative volumes of the feasible sets and the overlap between feasible sets computed with the detailed and reduced models. Third, the Hellinger distance measure was adopted to quantify the “distance” between posterior distributions of model predictions computed with the detailed and reduced models. Fourth, another measure, also a Hellinger distance measure, was constructed to quantify the “distance” between prediction intervals of the detailed and reduced models by assuming uniform distributions on the prediction intervals. Finally, sensitivities of the uncertainties in model predictions were established to measure the difference in influence of the model parameters on the model predictions between the detailed and reduced models. These methods will be illustrated with examples given in Chapters 4 and 5.

Chapter 4

Reduced models of syngas combustion

In this chapter, a syngas combustion system employed on a set of ignition delays is considered. The specifics of the selected full-size, detailed model and the reduced models developed from it are described in the following section. The computations of the ignition-delay times with the detailed and reduced models were performed using the ReactionLab code [111], assuming adiabatic, isochoric, homogeneous combustion at a set of initial conditions described below.

4.1 Modeled system

4.1.1 Detailed reaction model

The detailed reaction model used in this chapter is a syngas combustion model with 55 reactions and 14 reactive species developed recently by Slavinskaya et al. [112]. This model was obtained through uncertainty quantification analysis of an initial H_2/CO reaction model, assembled from 73 reactions and 14 reactive species. That initial model was optimized using 167 experimental targets. The optimized model demonstrated improved agreement with experiments compared to the initial model. This optimized model, specifically the one obtained by the least-squares minimization over the feasible set (the LS-F model), was the starting detailed model of the study in this chapter. The LS-F model was used as reported in Supporting Information of Ref. [112]; its 55-reaction set is reproduced in Table 4.1.

Table 4.1: Reactions of the detailed model, reproduced from Slavinskaya et al. [112], and reduced models R25, R17 and R12.

No.	Reaction	Multiplier λ		R25	R17	R12
		lower bound	upper bound			
1	$\text{H}_2 + \text{O}_2 = \text{OH} + \text{OH}$	0.18	17.86	✓	✓	—
2	$\text{H} + \text{H} + \text{M} = \text{H}_2 + \text{M}$	0.26	2.59	✓	✓	—
3	$\text{H}_2 + \text{O} = \text{OH} + \text{H}$	0.42	1.05	✓	✓	✓
4	$\text{OH} + \text{H}_2 = \text{H} + \text{H}_2\text{O}$	0.79	1.99	✓	✓	✓
5	$\text{OH} + \text{OH} = \text{H}_2\text{O}_2$	0.25	1.00	✓	✓	✓
6	$\text{H} + \text{O}_2 = \text{O} + \text{OH}$	0.83	1.31	✓	✓	✓
7	$\text{H} + \text{O}_2(+\text{M}) = \text{HO}_2(+\text{M})$	0.87	1.17	✓	✓	✓
8	$\text{O} + \text{O}(+\text{M}) = \text{O}_2(+\text{M})$	0.93	1.36	✓	—	—
9	$\text{OH} + \text{H}(+\text{M}) = \text{H}_2\text{O}(+\text{M})$	0.67	1.00	✓	✓	—
10	$\text{H} + \text{O} + \text{M} = \text{OH} + \text{M}$	0.04	1.00	✓	✓	—
11	$\text{H} + \text{HO}_2 = \text{O}_2 + \text{H}_2$	0.37	1.49	✓	✓	✓
12	$\text{H} + \text{HO}_2 = \text{OH} + \text{OH}$	0.95	1.89	✓	✓	—
13	$\text{H} + \text{HO}_2 = \text{O} + \text{H}_2\text{O}$	0.48	4.79	—	—	—
14	$\text{H} + \text{H}_2\text{O}_2 = \text{HO}_2 + \text{H}_2$	0.63	6.20	✓	✓	✓
15	$\text{H} + \text{H}_2\text{O}_2 = \text{OH} + \text{H}_2\text{O}$	0.32	1.27	—	—	—
16	$\text{OH} + \text{OH} = \text{O} + \text{H}_2\text{O}$	0.81	1.63	✓	—	—
17	$\text{OH} + \text{HO}_2 = \text{O}_2 + \text{H}_2\text{O}$	0.22	2.18	✓	—	—
18	$\text{OH} + \text{H}_2\text{O}_2 = \text{HO}_2 + \text{H}_2\text{O}$	0.71	2.86	✓	—	—
19	$\text{O} + \text{HO}_2 = \text{OH} + \text{O}_2$	0.28	2.80	✓	—	—
20	$\text{O} + \text{H}_2\text{O}_2 = \text{OH} + \text{HO}_2$	0.36	1.45	—	—	—
21	$\text{HO}_2 + \text{HO}_2 = \text{O}_2 + \text{H}_2\text{O}_2$	0.29	1.81	✓	—	✓
22	$\text{CO} + \text{O}_2 = \text{CO}_2 + \text{O}$	0.06	1.46	✓	✓	✓
23	$\text{CO} + \text{OH} = \text{CO}_2 + \text{H}$	0.96	1.52	✓	✓	✓
24	$\text{CO} + \text{O}(+\text{M}) = \text{CO}_2(+\text{M})$	1.00	4.00	✓	✓	✓
25	$\text{CO} + \text{HO}_2 = \text{CO}_2 + \text{OH}$	1.00	9.88	✓	—	✓
26	$\text{HCO} + \text{O}_2 = \text{CO} + \text{HO}_2$	0.40	2.50	✓	—	—
27	$\text{HCO} + \text{O}_2 = \text{OH} + \text{CO}_2$	0.17	1.07	—	—	—
28	$\text{H} + \text{HCO} = \text{H}_2 + \text{CO}$	0.25	1.01	✓	✓	—
29	$\text{O} + \text{HCO} = \text{OH} + \text{CO}$	0.94	3.77	—	—	—
30	$\text{O} + \text{HCO} = \text{H} + \text{CO}_2$	0.26	1.06	—	—	—
31	$\text{HCO} + \text{OH} = \text{CO} + \text{H}_2\text{O}$	0.26	1.05	—	—	—
32	$\text{HCO} + \text{HO}_2 = \text{OH} + \text{CO}_2 + \text{H}$	0.49	12.22	—	—	—
33	$\text{HCO} + \text{M} = \text{CO} + \text{H} + \text{M}$	1.00	4.00	✓	✓	—
34	$\text{O}_2 + \text{HCCO} = 2\text{CO} + \text{OH}$	0.04	1.01	—	—	—

35	<chem>HCCO+O2=HCO+CO+O</chem>	1.00	25.05	—	—	—
36	<chem>HCCO+O2=CO2+CO+H</chem>	0.18	4.47	—	—	—
37	<chem>HCCO+OH=2HCO</chem>	0.25	1.00	—	—	—
38	<chem>O+HCCO=2CO+H</chem>	0.25	1.01	—	—	—
39	<chem>CH+O2=CO+OH</chem>	0.22	2.14	—	—	—
40	<chem>O2+CH=CO2+H</chem>	0.17	1.72	—	—	—
41	<chem>CO+CH=HCCO</chem>	0.64	1.00	—	—	—
42	<chem>CH+CO2=HCO+CO</chem>	0.69	1.08	—	—	—
43	<chem>CH+O=CO+H</chem>	0.97	9.58	—	—	—
44	<chem>OH+CH=H+HCO</chem>	0.64	1.00	—	—	—
45	<chem>CH+O2=OH*+CO</chem>	0.98	3.92	—	—	—
46	<chem>O+H=OH*</chem>	0.26	1.06	—	—	—
47	<chem>OH*+Ar=OH+Ar</chem>	1.00	4.00	—	—	—
48	<chem>OH*+CO=OH+CO</chem>	1.00	4.00	—	—	—
49	<chem>OH*+CO2=OH+CO2</chem>	1.00	4.00	—	—	—
50	<chem>OH*+H=OH+H</chem>	1.00	4.00	—	—	—
51	<chem>OH*+H2=OH+H2</chem>	1.00	4.00	—	—	—
52	<chem>OH*+H2O=OH+H2O</chem>	0.26	1.02	—	—	—
53	<chem>OH*+O2=OH+O2</chem>	0.31	1.23	—	—	—
54	<chem>OH*+OH=OH+OH</chem>	0.93	3.70	—	—	—
55	<chem>OH*=OH</chem>	0.25	1.00	—	—	—

4.1.2 Reduced reaction models

Three reduced models were employed in the analysis of this chapter. Two of them were obtained from the detailed reaction model by applying the “detailed-reduction” method, as introduced in Section 2.2. Specifically, non-contributing reactions were identified by the following criteria,

$$\begin{aligned}
 r_{f,i} &< \epsilon_r \max(r_{f,6}, r_{f,7}), \\
 r_{r,i} &< \epsilon_r \max(r_{r,6}, r_{r,7}), \\
 |r_{\text{net},i} \Delta H_i| &< \epsilon_q \max_j (|r_{\text{net},j} \Delta H_j|), \quad j = 1, \dots, 55,
 \end{aligned} \tag{4.1}$$

where $r_{f,i}$, $r_{r,i}$, and $r_{\text{net},i} = r_{f,i} - r_{r,i}$ are the forward, reverse, and net reaction rates of reaction i , respectively, ΔH_i is the enthalpy change of reaction i , and ϵ_r and ϵ_q are two cutoff coefficients that are less than unity. The subscripts 6 and 7 refer to the selected reference reactions, H + O2 <=> OH + O and H + O2 (+ M) <=> HO2 (+ M) in Table 4.1, respectively, which are the main chain-controlling reactions of the modeled system. If criteria (4.1) are satisfied during the phenomena of interest, over the entire induction period in this case, a reaction is identified as non-contributing. Repeating this analysis for selected values of ϵ_r and ϵ_q at a set of initial conditions and removing the intersection of the identified non-

contributing reactions from the initial, full-size reaction set produces a (ϵ_r, ϵ_q) -level reduced reaction model.

The two reduced reaction models were developed with different *a priori* levels of accuracy by using $\epsilon_r = \epsilon_q = 0.02$ and $\epsilon_r = \epsilon_q = 0.5$. The computations for each of these levels were performed over the following ranges of initial conditions: temperature 1000–1800 K, pressure 0.5–2 atm, equivalence ratio 0.04–1.5, and CO/H₂ ratio 0–2. A set of 100 points within these ranges was created according to the Latin Hypercube design [87]; these initial conditions are reported in Table 4.2. The resulting sets contained 25 and 17 reactions, each containing 11 species, and, hence, will be referred to hereafter as R25 and R17 reduced reaction models, respectively; these reactions sets are reported in Table 4.1. The developed R25 and R17 reduced models reproduced the ignition delays computed by the full-size model for the 100-set of the initial conditions with an average relative error of 0.4% and 1.3%, respectively. Note that the ϵ 's and ranges were selected such that the obtained reduced models allowed for the exposure of the differences in their UQ properties.

Table 4.2: Initial conditions of the Latin hypercube sampling (LHS) points used for building the reduced models R25 and R17 using the “detailed-reduction” method; ignition delay times τ^D are those computed with the detailed model, and the relative deviations for the reduced models are defined as $\frac{|\tau^R - \tau^D|}{\tau^D} \times 100\%$.

No.	T (K)	P (atm)	Mole fraction				ϕ	τ^D (μ s)	Rel. dev. (%)		
			CO	H2	O2	N2			R25	R17	R12
1	1100	0.79	0.03	0.27	0.15	0.55	1.02	92.10	0.01	0.09	5.60
2	1067	1.71	0.12	0.08	0.17	0.64	0.57	80.04	0.06	1.17	28.17
3	1662	1.04	0.02	0.32	0.14	0.52	1.24	7.83	0.02	0.76	14.82
4	1064	1.40	0.03	0.23	0.16	0.58	0.85	75.44	0.07	0.39	20.19
5	1241	0.81	0.10	0.09	0.17	0.64	0.54	50.35	0.27	0.92	1.62
6	1688	1.49	0.06	0.31	0.13	0.49	1.43	5.24	0.21	0.12	12.78
7	1467	0.98	0.22	0.13	0.14	0.51	1.28	17.22	0.59	0.49	9.00
8	1293	1.56	0.17	0.10	0.15	0.58	0.87	20.50	0.17	0.42	0.96
9	1023	1.59	0.18	0.12	0.15	0.55	1.00	122.03	0.54	1.10	87.21
10	1006	1.63	0.07	0.21	0.15	0.57	0.92	143.62	0.40	1.15	332.13
11	1602	1.27	0.06	0.04	0.19	0.71	0.28	11.89	1.40	6.32	3.84
12	1003	0.63	0.09	0.08	0.18	0.66	0.46	289.22	0.19	0.31	14.80
13	1029	0.90	0.15	0.08	0.16	0.61	0.69	174.96	0.37	0.36	17.93
14	1338	1.90	0.07	0.05	0.18	0.69	0.34	17.04	0.94	1.34	4.11
15	1112	0.78	0.21	0.13	0.14	0.52	1.25	98.17	0.41	0.27	5.04
16	1630	1.79	0.05	0.05	0.19	0.71	0.26	7.59	0.01	5.47	5.25
17	1376	1.41	0.06	0.06	0.19	0.70	0.31	19.35	0.63	1.71	2.04

18	1058	2.00	0.05	0.03	0.19	0.73	0.20	142.41	0.05	3.76	49.53
19	1120	0.60	0.02	0.10	0.19	0.69	0.32	123.08	0.14	0.24	2.99
20	1016	1.25	0.05	0.16	0.17	0.62	0.64	134.09	0.11	0.73	41.26
21	1215	0.67	0.05	0.12	0.18	0.66	0.48	62.81	0.21	0.75	0.29
22	1750	1.98	0.17	0.20	0.13	0.50	1.40	3.66	0.59	0.45	15.29
23	1077	0.51	0.13	0.08	0.17	0.62	0.63	201.49	0.24	0.21	4.60
24	1043	1.64	0.06	0.03	0.19	0.72	0.23	161.27	0.56	2.81	44.19
25	1204	1.75	0.04	0.33	0.13	0.50	1.38	24.20	0.10	0.00	5.25
26	1138	1.82	0.18	0.21	0.13	0.49	1.48	36.61	0.16	0.16	11.93
27	1168	0.64	0.12	0.10	0.17	0.62	0.66	87.82	0.50	0.75	1.94
28	1385	1.15	0.03	0.03	0.20	0.74	0.15	30.99	0.09	2.65	4.02
29	1219	0.98	0.08	0.06	0.18	0.68	0.39	52.95	0.64	1.50	3.31
30	1583	1.74	0.11	0.09	0.17	0.63	0.60	7.14	1.25	2.44	2.10
31	1181	1.21	0.11	0.17	0.15	0.57	0.93	39.45	0.09	0.18	4.11
32	1785	1.89	0.08	0.08	0.18	0.67	0.43	4.24	0.03	4.72	1.08
33	1477	0.84	0.00	0.28	0.15	0.56	0.95	16.07	0.06	0.50	9.19
34	1744	1.00	0.21	0.11	0.14	0.54	1.10	9.15	0.06	1.32	15.45
35	1695	0.76	0.03	0.08	0.19	0.70	0.30	12.98	1.11	5.78	1.20
36	1096	0.73	0.01	0.05	0.20	0.74	0.17	153.00	0.37	0.00	6.23
37	1192	1.84	0.07	0.09	0.18	0.66	0.45	29.13	0.13	0.57	6.79
38	1449	1.46	0.06	0.12	0.17	0.65	0.50	11.53	0.16	0.88	2.01
39	1035	1.43	0.11	0.10	0.17	0.63	0.61	112.56	0.35	0.92	36.62
40	1759	0.75	0.11	0.15	0.16	0.59	0.83	10.09	1.10	2.36	24.31
41	1366	1.81	0.03	0.01	0.20	0.76	0.10	30.97	0.02	4.41	6.11
42	1391	1.50	0.10	0.06	0.18	0.67	0.44	17.37	0.00	2.08	1.60
43	1134	1.08	0.03	0.31	0.14	0.52	1.22	56.46	0.05	0.14	5.83
44	1119	1.38	0.02	0.01	0.21	0.77	0.06	199.51	0.60	1.39	11.68
45	1566	1.17	0.14	0.23	0.13	0.49	1.42	9.52	0.21	0.05	10.66
46	1330	0.58	0.16	0.17	0.14	0.53	1.16	43.35	0.37	0.91	4.93
47	1361	0.93	0.13	0.11	0.16	0.60	0.73	25.46	0.06	1.06	3.00
48	1230	0.70	0.18	0.10	0.15	0.57	0.91	61.17	0.70	1.20	0.51
49	1176	0.86	0.06	0.07	0.18	0.68	0.37	69.97	0.02	0.83	3.47
50	1150	0.94	0.15	0.08	0.16	0.61	0.68	72.90	0.44	0.64	5.08
51	1489	1.19	0.15	0.23	0.13	0.49	1.45	11.75	0.26	0.23	7.92
52	1197	0.55	0.16	0.11	0.15	0.58	0.87	88.00	0.48	0.82	0.30
53	1226	1.87	0.15	0.11	0.16	0.58	0.84	22.95	0.02	0.21	5.05
54	1589	1.94	0.07	0.06	0.18	0.69	0.35	6.99	0.97	4.69	3.79
55	1672	1.33	0.14	0.10	0.16	0.60	0.75	7.49	0.54	1.88	6.78
56	1322	1.32	0.18	0.12	0.15	0.55	1.01	20.51	0.37	0.97	0.96
57	1010	1.07	0.09	0.21	0.15	0.55	1.05	153.62	0.26	0.41	35.78
58	1431	1.86	0.04	0.16	0.17	0.63	0.59	8.85	0.26	0.01	1.45
59	1253	1.13	0.13	0.06	0.17	0.64	0.56	38.55	0.07	0.90	2.12

60	1541	0.69	0.11	0.28	0.13	0.48	1.49	17.25	0.65	0.53	11.46
61	1161	1.69	0.00	0.04	0.20	0.76	0.10	61.52	0.67	0.82	8.43
62	1507	1.14	0.19	0.16	0.14	0.51	1.30	12.31	0.30	1.17	8.96
63	1312	0.95	0.24	0.13	0.13	0.50	1.39	30.89	0.09	0.02	2.76
64	1185	1.18	0.20	0.14	0.14	0.52	1.20	42.70	0.22	0.02	3.88
65	1073	0.65	0.02	0.03	0.20	0.75	0.12	243.36	0.79	0.39	8.54
66	1021	0.72	0.06	0.09	0.18	0.67	0.42	215.99	0.09	0.18	13.47
67	1287	1.93	0.09	0.09	0.17	0.65	0.53	17.33	0.53	0.67	3.17
68	1639	1.96	0.20	0.11	0.15	0.54	1.07	5.42	0.10	1.35	7.38
69	1610	1.45	0.02	0.05	0.20	0.73	0.18	9.75	0.60	3.91	5.03
70	1499	0.57	0.21	0.12	0.14	0.53	1.15	28.40	0.07	0.27	14.35
71	1083	0.91	0.03	0.02	0.20	0.75	0.13	209.29	0.53	0.28	12.61
72	1518	1.09	0.14	0.17	0.15	0.55	1.05	12.05	0.28	0.01	9.35
73	1104	1.79	0.08	0.17	0.16	0.59	0.78	46.07	0.18	0.34	16.83
74	1534	0.52	0.16	0.17	0.14	0.53	1.19	25.10	1.08	1.45	14.78
75	1047	0.54	0.17	0.16	0.14	0.53	1.13	201.95	0.28	0.13	6.90
76	1420	0.82	0.07	0.29	0.13	0.50	1.34	20.53	0.36	0.50	6.95
77	1559	1.54	0.20	0.16	0.14	0.51	1.33	7.85	0.49	0.51	9.38
78	1526	1.61	0.06	0.12	0.17	0.65	0.52	8.16	0.62	0.89	2.31
79	1246	1.52	0.14	0.17	0.15	0.54	1.08	23.05	0.23	0.09	2.03
80	1344	1.28	0.05	0.05	0.19	0.71	0.26	26.74	0.41	1.35	2.58
81	1412	1.42	0.01	0.02	0.20	0.77	0.07	30.43	0.01	3.77	5.82
82	1091	1.01	0.13	0.20	0.14	0.53	1.18	81.04	0.15	0.09	9.00
83	1261	1.62	0.14	0.11	0.16	0.59	0.79	22.10	0.21	0.02	2.20
84	1285	0.89	0.09	0.14	0.16	0.61	0.71	33.28	0.30	0.49	1.33
85	1052	1.58	0.23	0.14	0.13	0.50	1.35	87.85	0.55	0.46	35.39
86	1458	0.85	0.19	0.10	0.15	0.56	0.96	21.32	0.67	2.83	7.14
87	1144	1.10	0.18	0.11	0.15	0.56	0.98	58.90	0.27	0.32	5.77
88	1155	1.51	0.15	0.20	0.14	0.51	1.27	37.89	0.08	0.19	7.23
89	1403	1.67	0.19	0.15	0.14	0.52	1.23	11.61	0.05	0.51	2.92
90	1274	0.62	0.05	0.22	0.15	0.57	0.90	47.25	0.07	0.19	2.76
91	1710	1.92	0.06	0.03	0.19	0.72	0.23	7.05	1.41	10.21	5.24
92	1726	1.67	0.09	0.05	0.18	0.68	0.39	6.40	0.08	7.36	1.13
93	1306	1.31	0.22	0.13	0.14	0.51	1.31	22.46	0.00	0.32	0.71
94	1127	1.23	0.01	0.01	0.21	0.77	0.05	174.07	0.49	1.70	9.96
95	1040	1.25	0.14	0.11	0.16	0.60	0.77	113.70	0.45	0.44	25.60
96	1647	1.03	0.00	0.07	0.19	0.73	0.20	11.10	0.11	4.57	4.37
97	1270	1.35	0.18	0.20	0.13	0.49	1.46	23.79	0.06	0.09	0.51
98	1354	1.36	0.13	0.19	0.14	0.54	1.11	16.06	0.21	0.28	2.05
99	1789	1.77	0.07	0.19	0.16	0.59	0.81	3.54	0.45	1.01	18.23
100	1439	1.73	0.07	0.16	0.16	0.61	0.70	9.29	0.13	0.70	2.85

A third reduced reaction model, was the one developed recently by Boivin et al. [105] for syngas ignition delays. It was composed of the 12 reactions suggested by Boivin et al., but all treated as reversible, because setting them irreversible, as in [105], led to much higher relative deviations as compared to the detailed model. These 12 reactions are identified in Table 4.1, and this reduced model will be referred to as R12. This reduced model reproduced the ignition delays computed by the full-size model for the set of the 100 initial conditions with an average relative error of 13%, significantly larger than those of R25 and R17.

4.1.3 Model parameters of reaction models

In this chapter, it is assumed that the uncertainty of the reaction models, including the detailed and three reduced models introduced in the previous subsections, resides in the reaction rate coefficients, and their uncertainties are represented by multipliers, λ_i 's, to the nominal parameter expressions [86],

$$k_i = \lambda_i k_i^{\text{nominal}}, \quad i = 1, \dots, 55, \quad (4.2)$$

where k_i is the rate coefficient of reaction i , ranging from 1 to 55 for the detailed model. The nominal rate coefficients in the present study are those of the LS-F optimized model of Slavinskaya et al. [112], as described in Section 4.1.1, i.e., $k_i^{\text{nominal}} = k_i^{\text{LS-F}}$. The bounds of the λ_i 's are taken from the same study [112]. Their values are reported in Table 4.1; these values, which are multipliers to $k^{\text{LS-F}}$, represent the posterior ranges of k 's that resulted from the UQ analysis in [112].

The surrogate models, described in Section 2.1.2, were then developed in transformed variables

$$x_i = \frac{\ln \lambda_i / \lambda_{i,\min}}{\frac{1}{2} \ln \lambda_{i,\max} / \lambda_{i,\min}} - 1, \quad i = 1, \dots, 55. \quad (4.3)$$

Defined in this manner, all model parameters have $[-1, 1]$ intervals, and the nominal set of rate coefficients is specified by $x^{\text{nominal}} = x(\lambda = 1)$. Because the set of reactions in a reduced model is a subset of the set of reactions in the detailed model, so is the set of uncertain model parameters of a reduced model a subset of the set of uncertain model parameters of the detailed model. There are 25, 17, and 12 uncertain model parameters for the reduced models R25, R17, and R12, respectively.

4.2 A dataset with computer-generated ignition delay QOIs

In this section, a dataset with 100 QOIs, namely ignition delay times under 100 different initial conditions from the LHS points in Table 4.2, is investigated. The results of applying the measures introduced in Chapter 3 to the three reduced models derived in Section 4.1 are discussed.

4.2.1 QOIs and surrogate models

The “observed” values of the QOIs were obtained from the solution of the detailed model with the nominal set of rate coefficients, denoted τ^D . The QOI uncertainty ranges, i.e., L_e ’s and U_e ’s in Section 2.1, were assumed to be $\pm 10\%$ intervals of τ^D ,

$$[0.9 \tau_e^D, 1.1 \tau_e^D], \quad e = 1, \dots, 100. \quad (4.4)$$

In application of B2BDC to UQ problems, as discussed in Section 2.1, the QOI ranges represent their respective uncertainties. In the present work, these QOI ranges prescribe the desired level of reduced-model accuracy by specifying the accepted ranges of variations in prediction of training targets.

For each of the 100 sampled initial conditions, a quadratic polynomial of $\ln \tau^D$ in x (defined in Eq. (4.3)) was built, as described in Section 2.1.2. Active parameters were selected according to the ranking of the impact factors. The absolute impact factors of all 55 parameters, computed with the detailed model for QOI 68, are plotted in Fig 4.1 as an example. The results show clearly that only a small subset of active parameters have a large impact on the QOI prediction. A parameter was selected as an active parameter if its absolute impact factor is either ≥ 0.01 or $\geq 5\%$ of the highest absolute impact factor among all parameters. In the case of QOI 68 and the detailed model, the number of active parameters is 19. In the setting of the detailed model, the number of active parameters ranges from 6 to 19 for the 100 QOIs listed in Table 4.2.

Similarly, quadratic polynomials for QOIs with the three reduced models were built. The number of active parameters for all QOI and model combinations is tabulated in Table 4.3. The union of active parameters of all 100 QOIs with the detailed model, R25, R17, and R12 have 32, 25, 17, and 12 parameters, respectively. The fitting errors of all these surrogate models were below 0.01. The surrogate models will be referred to hereafter as $M_e^D(x)$ and $M_e^R(x)$ for the detailed and reduced models, respectively, associated with the e -th QOI, where R is a place holder for R25, R17, or R12.

Table 4.3: Number of active parameters of surrogate models for 100 QOIs with the detailed model and three reduced models, R25, R17, and R12.

QOI No.	Number of active variables			
	detailed	R25	R17	R12
1	7	7	7	6
2	9	9	8	7
3	9	8	7	7
4	6	6	6	5
5	6	6	6	5
6	10	9	9	5
7	11	12	7	7

8	9	8	6	6
9	8	8	8	7
10	7	7	7	6
11	13	12	8	6
12	8	8	8	7
13	8	8	8	7
14	5	5	6	6
15	7	8	6	5
16	10	9	7	6
17	7	9	6	6
18	9	9	8	8
19	7	6	6	6
20	7	7	7	6
21	6	6	5	6
22	12	13	9	6
23	7	7	7	5
24	9	9	8	8
25	7	7	7	6
26	7	7	7	5
27	7	5	4	6
28	7	8	7	7
29	5	4	5	6
30	12	10	11	9
31	7	7	7	6
32	13	12	11	9
33	8	8	7	5
34	16	15	9	6
35	15	9	8	9
36	7	7	7	6
37	6	6	6	6
38	9	14	7	8
39	8	8	8	7
40	13	12	8	8
41	6	7	6	6
42	6	5	5	5
43	7	7	7	6
44	8	8	8	6
45	9	11	6	6
46	8	7	6	5
47	9	8	9	5
48	7	8	6	7
49	5	5	6	6

50	5	6	5	5
51	8	10	7	5
52	6	6	5	7
53	7	7	7	6
54	10	8	12	5
55	16	14	10	9
56	9	8	7	5
57	7	7	7	6
58	8	9	7	5
59	6	6	6	7
60	7	8	8	5
61	7	8	8	6
62	8	10	9	5
63	9	8	6	6
64	7	6	6	5
65	6	6	6	6
66	7	7	7	6
67	7	7	6	6
68	17	12	8	9
69	11	12	8	5
70	15	14	7	8
71	7	8	7	6
72	10	10	6	5
73	7	7	7	6
74	11	9	9	7
75	7	7	7	5
76	9	9	6	5
77	9	11	9	9
78	9	14	8	7
79	7	7	6	6
80	7	5	5	7
81	8	9	6	6
82	6	6	6	5
83	7	8	6	6
84	8	8	6	5
85	8	8	8	7
86	13	13	8	7
87	7	7	6	5
88	6	6	6	5
89	8	7	8	7
90	8	7	5	5
91	12	13	10	5

92	13	10	9	7
93	7	6	6	5
94	7	7	6	5
95	8	8	8	7
96	12	9	10	5
97	7	8	7	6
98	9	8	7	6
99	11	13	8	7
100	9	10	8	7

4.2.2 Comparison of predicted values

Three measures, δ , $\delta^{\mathcal{H}}$, and $\delta^{\mathcal{F}}$, as introduced in Section 3.1, were calculated for each of the 100 QOIs and each of the three reduced models. The resulting values are reported in Tables A.1 and A.2 of Appendix A. The inner and outer bounds bracket the deviation measures and, as can be observed by inspecting Tables A.1 and A.2 of Appendix A, they are very close to each other. To make the graphical and tabular presentation clear, the $\delta^{\mathcal{H}}$ values of individual QOIs were computed using midpoints of the respective inner and outer bounds; i.e., $\delta_e^{\mathcal{H}} = [\delta_e^{\mathcal{H}}(\text{inner}) + \delta_e^{\mathcal{H}}(\text{outer})] / 2$. Similarly for $\delta_p^{\mathcal{F}}$.

The three measures for each QOI and each reduced model are displayed in Fig. 4.2, and their summary statistics are listed in Table 4.4. Note that the y -axis limits are different for the three subplots in Fig. 4.2. Inspection of the computed first measure, δ , indicates that the deviations are small for both R25 and R17. Even in the case of R17, the deviations are below 0.05 for 95 QOIs and do not exceed 0.092 for all 100 QOIs. Given that $M(x)$'s are logarithms of ignition times and, hence, δ 's are logarithmic ratios of them, a 0.05 difference between $M_e^{\text{R}}(x)$ and $M_e^{\text{D}}(x)$ translates approximately to a 5% relative change in ignition delay times. Thus, considering the single-point evaluation, one may conclude that both R25 and R17 are relatively accurate reduced reaction models over the assumed ranges of the initial conditions.

Table 4.4: Summary statistics of deviation measures (over 100 QOIs).

	R25		R17		R12	
	Mean	Max.	Mean	Max.	Mean	Max.
δ	0.002	0.007	0.011	0.092	0.096	1.074
$\delta^{\mathcal{H}}$	0.013	0.071	0.053	0.248	0.267	4.360
$\delta^{\mathcal{F}}$	0.012	0.070	0.049	0.240	0.249	4.349

However, the deviation in these models may be much larger for model parameters that differ from their nominal values, for example, if one considers variation of parameters within their respective uncertainty ranges. This indeed is the case, as can be concluded by inspection

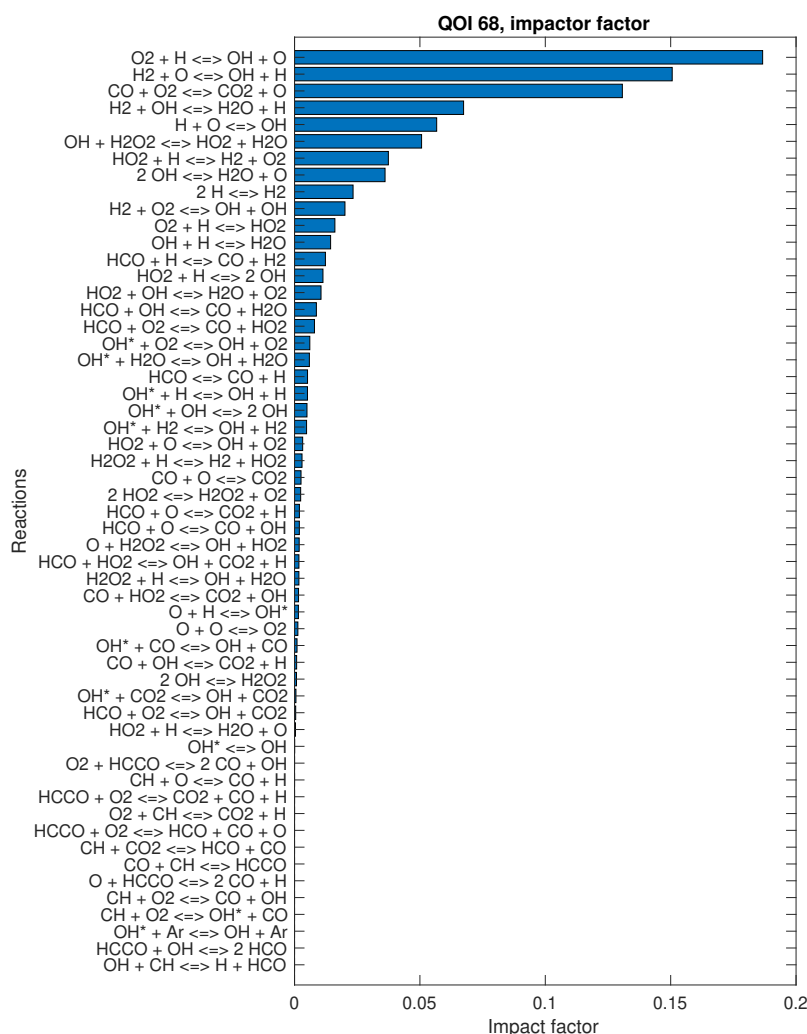


Figure 4.1: Absolute impact factors of all 55 parameters, computed with the detailed model for QOI 68.

of the $\delta^{\mathcal{H}}$ values computed by Eq. (3.2) and plotted in Fig. 4.2. For instance, the deviations between the predictions of R17 and those of D are quite large, with $\delta^{\mathcal{H}} \geq 0.05$ for 34 of the 100 QOIs and the maximum deviation reaching ~ 0.25 . These results indicate that R17 does not perform that well when parameter uncertainties are taken into account.

Yet the $\delta^{\mathcal{H}}$ measure is expected to overestimate the deviation because it is computed by considering variation of model parameters within their entire prior region, \mathcal{H} , without taking

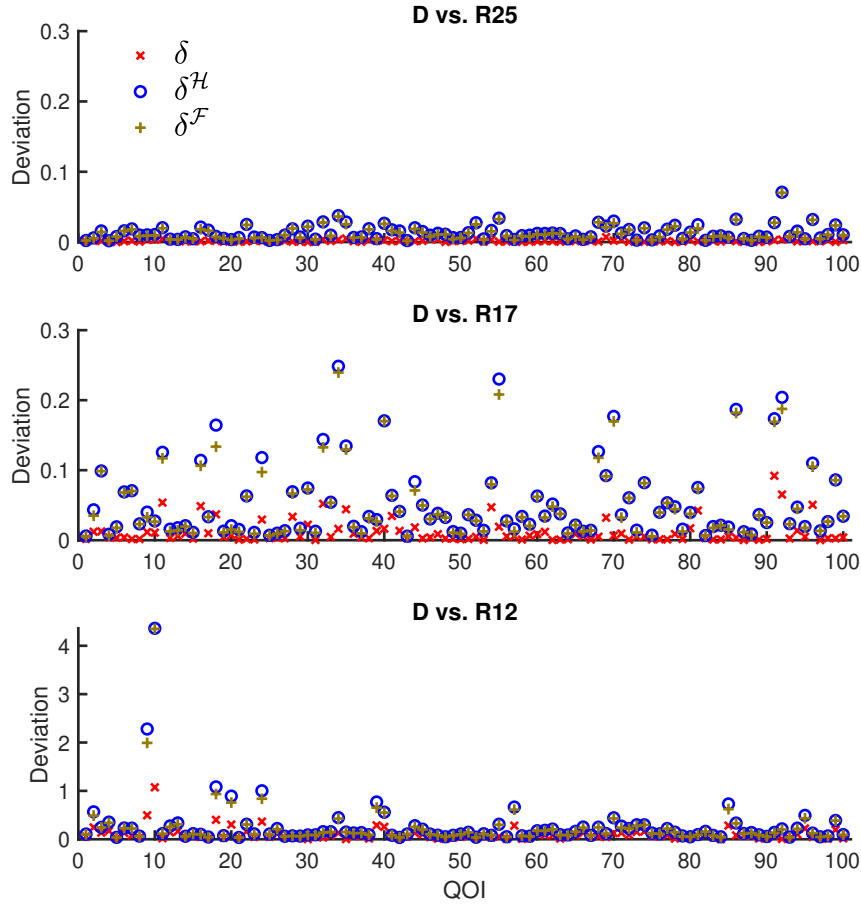


Figure 4.2: Deviations, δ , $\delta^{\mathcal{H}}$, and $\delta^{\mathcal{F}}$, between the detailed (D) model and reduced (R25, R17, R12) models for 100 QOIs.

into consideration parameter correlations. The latter originate from requiring the model to predict experimental data [14]. Including in the analysis the desired tolerance levels for QOI predictions by restricting variation of x to the feasible set, \mathcal{F} , gives a more realistic measure of reduced-model deviations, namely $\delta^{\mathcal{F}}$ of Eq. (3.3). The resulting deviations, as expected, were computed to be smaller than in the $x \in \mathcal{H}$ case of Eq. (3.2), yet by relatively small amounts, as shown in Fig. 4.2.

The third reduced model, R12, produced deviations, both δ and $\delta^{\mathcal{H}}$, substantially higher than those of R25 and R17, as can be seen from Fig. 4.2. This is not surprising because R12, with the nominal model parameters, reproduced the ignition delays computed by the full-size model for the set of the 100 initial conditions with an average relative error of 13%,

as reported in Section 4.1.2.

Fig. 4.2 highlights large differences between δ and $\delta^{\mathcal{H}}$ (or $\delta^{\mathcal{F}}$) for some QOIs in the case of R17 or R12. It is also of interest to view the overall performance of the reduced models, as shown by the histograms in Fig. 4.3. Note that the x -axis is in logarithmic scale due to the large values of deviations for R12. There is a clear trend of increasing deviations from R25 to R17 to R12.

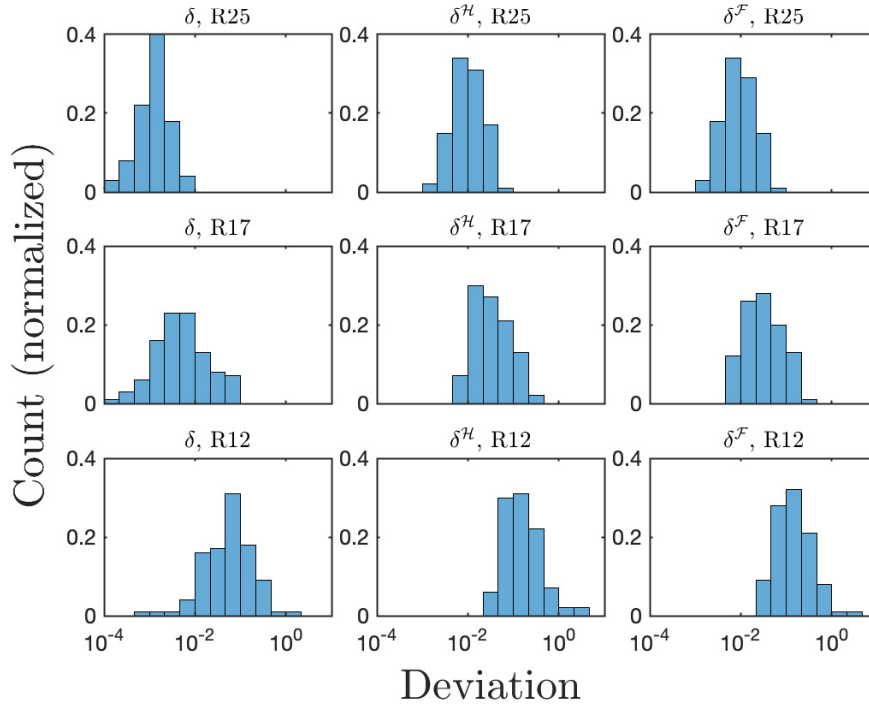


Figure 4.3: Histograms of deviations, δ , $\delta^{\mathcal{H}}$, and $\delta^{\mathcal{F}}$, between the detailed (D) and reduced (R25, R17, R12) models for 100 QOIs.

4.2.3 Comparison of feasible sets

Feasible sets \mathcal{F}^{R25} , \mathcal{F}^{R17} , and \mathcal{F}^{R12} were compared with \mathcal{F}^{D} in two different ways, as described in Section 3.2. First, the relative volumes of \mathcal{F}^{R25} , \mathcal{F}^{R17} , and \mathcal{F}^{R12} were compared with that of \mathcal{F}^{D} . The relative volume of the feasible set is defined as the ratio of the feasible set volume to the volume of the associated hyperrectangle, $\frac{V(\mathcal{F})}{V(\mathcal{H})}$, as described in Section 3.2. Then, the overlap between \mathcal{F}^{R25} , \mathcal{F}^{R17} , or \mathcal{F}^{R12} and \mathcal{F}^{D} was computed. Note that all 100 QOIs were used to obtain the feasible set for a given model, different from the leave-one-out scheme in Section 3.3. For each of the 100 QOIs, the other 99 QOIs would be used to obtain

the feasible set in the leave-one-out scheme, resulting in 100 different feasible sets for 100 different QOIs for a given model.

The hyper-rectangular region of the prior knowledge, \mathcal{H} , was sampled with 1×10^7 points, and the relative volume of the feasible set was estimated using rejection sampling [106] by the fraction of points contained in \mathcal{H} that also fall within the feasible set; these computations resulted in 0.072, 0.071, 0.060, and $< 1 \times 10^{-7}$ for \mathcal{F}^D , \mathcal{F}^{R25} , \mathcal{F}^{R17} , and \mathcal{F}^{R12} , respectively (not one of the 1×10^7 points fell into \mathcal{F}^{R12}). For the detailed, R25, and R17 models, the estimated relative volumes converged with 1×10^7 points, meaning that a further increase in the number of sampled points would not change the estimated relative volumes appreciably. The change of the relative volume of \mathcal{F}^D with the number of sampled points is illustrated as an example in Fig. 4.4. For R12, no more points were sampled because the current evidence reveals the large difference between the relative volumes of \mathcal{F}^{R12} and \mathcal{F}^D .

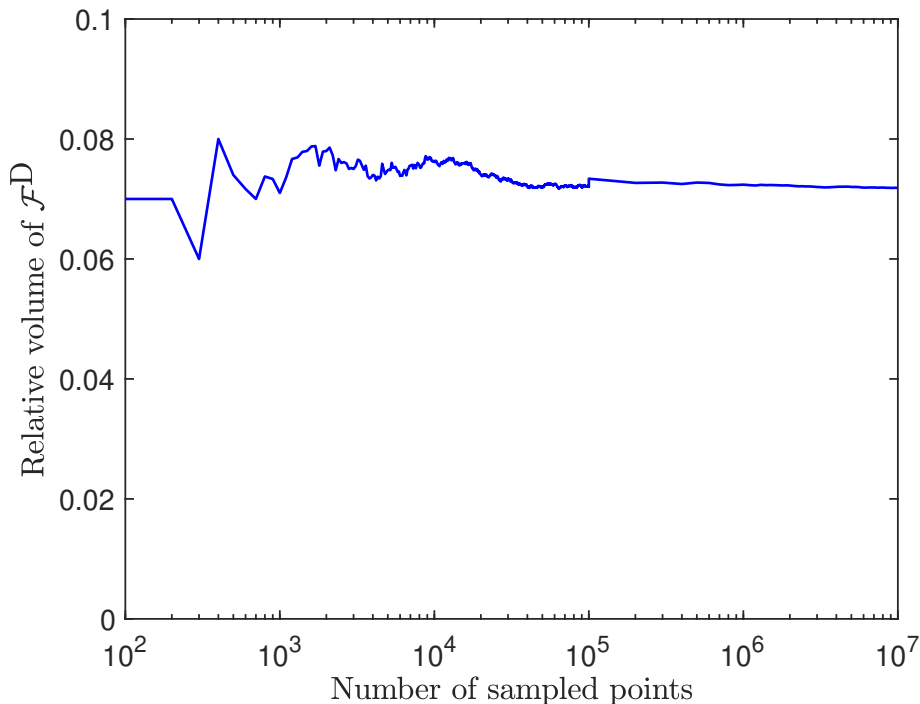


Figure 4.4: The relative volume of \mathcal{F}^D vs. the number of sampled points.

To estimate the feasible-set overlap, the feasible-set regions were sampled with 1.5×10^4 uniformly-distributed points using the Gibbs sampling algorithm [106] adopted to B2BDC [107]. Nearly all, 97% points of \mathcal{F}^{R25} fell within \mathcal{F}^D , and 96% points of \mathcal{F}^D fell within \mathcal{F}^{R25} , but only 75% points of \mathcal{F}^{R17} fell within \mathcal{F}^D , and 64% points of \mathcal{F}^D fell within \mathcal{F}^{R17} . These results indicate that \mathcal{F}^D and \mathcal{F}^{R25} are basically overlap, and \mathcal{F}^{R17} is moderately smaller in volume

than \mathcal{F}^D . Not a single point of \mathcal{F}^D fell within \mathcal{F}^{R12} and vice versa, thus showing that \mathcal{F}^{R12} is very small and located completely outside \mathcal{F}^D .

The uniform sample collected in the feasible set may have missed some regions in the feasible set, even if the test of uniformity based on the Gelman convergence test was passed [107]. As discussed in Section 4.2.1, the union of active parameters of all 100 QOIs has 32 parameters in the case of the detailed model. For each of the 32 active parameters, the interval formed by the minimum and maximum of that parameter among the sampled 1.5×10^4 points in the feasible set \mathcal{F}^D were computed, termed the sampled-based interval. The corresponding optimization-based interval for each active parameter was obtained by solving problem (2.8). The sample- and optimization-based intervals for all active parameters in the case of the detailed model are plotted in Fig. 4.5. For each active parameter, the vertical red line is the inner-bound interval, the vertical blue line is the outer-bound interval, and the vertical black line is the sample-based interval. The inner- and outer-bound prediction intervals for active parameters are highly close to each other for all active parameters; specifically, all of them are very close to the $[-1, 1]$ interval, which is the prior interval for all model parameters, as discussed in Section 4.1.3. This implies that the feasible set \mathcal{F}^D can either touch or is very close to all the sides of the initial hyper-rectangle \mathcal{H} . The sample-based interval for parameter 6 (associated with the reaction $\text{H} + \text{O}_2 \rightleftharpoons \text{OH} + \text{O}$) is slightly narrower than the optimization-based interval, as shown in Fig. 4.5. This indicates that the sampling algorithm did not visit some parts of the initial hyper-rectangle \mathcal{H} that are also in the feasible set \mathcal{F}^D .

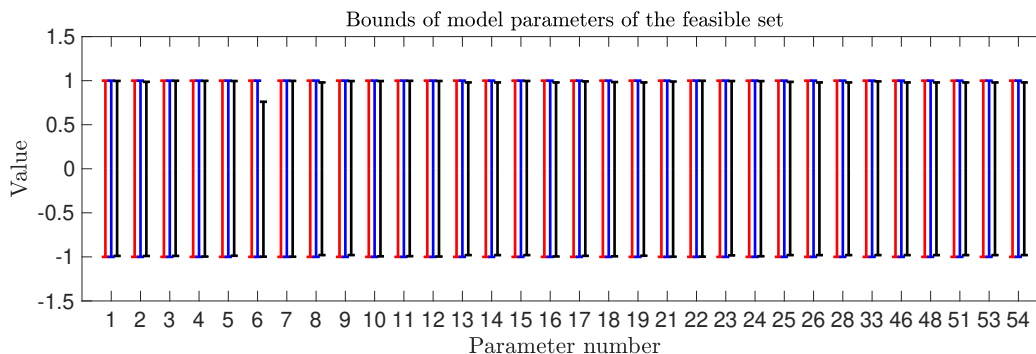


Figure 4.5: Bounds of active model parameters computed with the detailed model; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and vertical black lines are sample-based prediction intervals.

Overall, the sample-based intervals of active model parameters are very close to the optimization-based ones, implying that the uniform sample collected in the feasible set \mathcal{F}^D has visited most parts of the feasible set. The comparison of sample- and optimization-based

intervals of active model parameters can shed light on the difficulty in sampling all parts of the feasible set.

4.2.4 Comparison of posterior distributions

For each of the 100 QOIs, uniform samples were collected from the feasible sets. Specifically, a uniform sample of 1.5×10^4 points were generated in each of the \mathcal{F}_{-p}^D , \mathcal{F}_{-p}^{R25} , \mathcal{F}_{-p}^{R17} , and \mathcal{F}_{-p}^{R12} , for $p = 1, \dots, 100$, using the Gibbs sampling algorithm adopted to B2BDC [107]. The notation \mathcal{F}_{-p} denotes the feasible set for a dataset with the p -th QOI removed as a result of the leave-one-out scheme, as discussed in Section 3.3. The points of a uniform sample were then used to evaluate the corresponding surrogate model, and a sample of model predictions was subsequently obtained. The histograms of ignition delay times computed for two selected QOIs, demonstrating both good (QOI 2) and not-so-good (QOI 92) cases, are reported in Fig. 4.6. The histograms of all 100 QOIs can be found in Fig. A.1 of Appendix A.

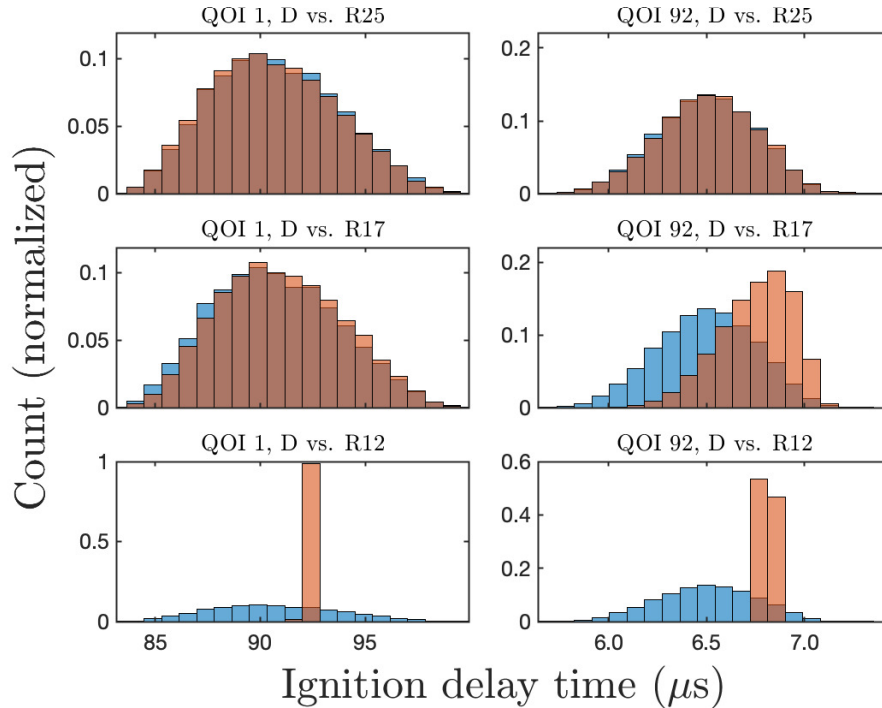


Figure 4.6: Histograms for selected QOIs of detailed (blue) and reduced (brown) model predictions over feasible sets.

Inspection of the histograms indicates that the reduced model R25 reproduces closely the posterior distributions of all QOIs computed by the detailed model, as demonstrated for

QOIs 1 and 92 in Fig. 4.6. The predictions of the reduced model R17 somewhat deviate from those of the detailed model, as can be observed from the histograms plotted for QOI 92, and those of R12 deviate substantially. The histograms of R12 also give a wrong impression that the posterior QOI distribution is narrower, thus falsely indicating a more accurate prediction of the QOI, but this illusion is created by the much smaller and distorted feasible set of R12. One should recall at this point that letting x vary outside \mathcal{F}^{R12} will necessarily violate the requirement for the reduced model to reproduce the QOIs within their assigned bounds.

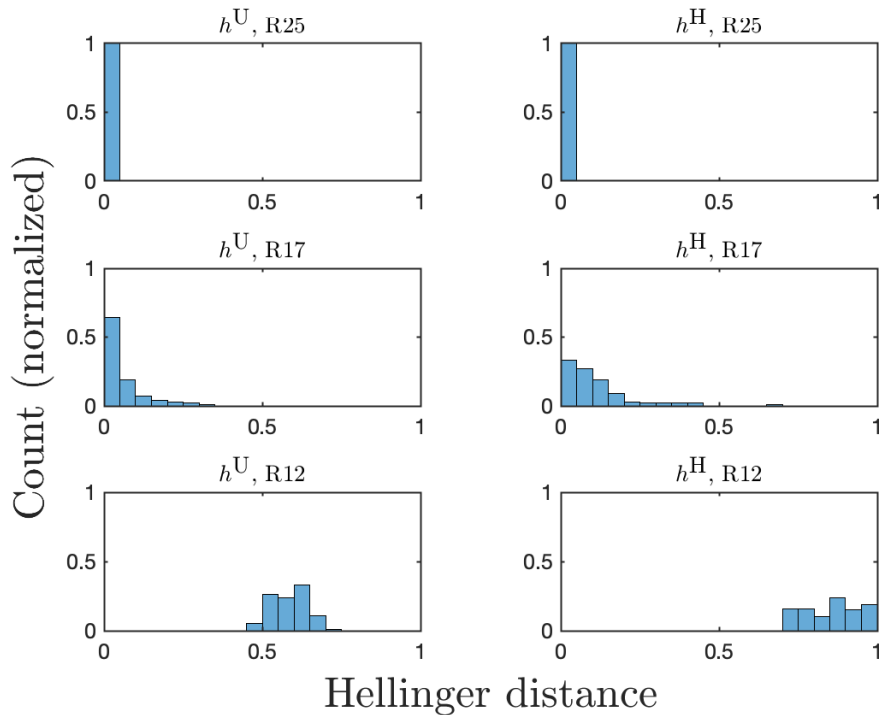


Figure 4.7: Histograms of the Hellinger distance metrics computed for the prediction intervals (left) and the posterior distributions (right) of the reduced models.

Hellinger distances h^{H} , as computed by Eq. (3.7), for all 100 QOIs with the three reduced models are reported in Table A.3 of Appendix A, the distributions of these values over the 100 QOIs are displayed in Fig 4.7, and their summary statistics are given in Table 4.5. The Hellinger distance metrics, displayed in Fig. 4.7 and listed in Table 4.5, capture the visual comparison among the distributions, illustrated in Fig. 4.6, and numerically demarcates the performance of the three reduced models.

Table 4.5: Summary statistics of Hellinger distances (over 100 QOIs).

	R25		R17		R12	
	Mean	Max.	Mean	Max.	Mean	Max.
h^H	0.010	0.027	0.111	0.654	0.859	1.000
h^U	0.007	0.031	0.061	0.346	0.588	0.715

4.2.5 Comparison of prediction intervals

For each $p = 1, \dots, 100$, prediction intervals of the detailed model and three reduced models were obtained by solving the corresponding optimization problems, as described in Section 3.4. Averaged over the 100 QOIs, the relative lengths of the inner-bound prediction intervals with respect to the outer-bound prediction intervals are 95.8, 95.6, 96.0, and 90.0% for the detailed, R25, R17, and R12 models, respectively. This indicates that inner- and outer-bound prediction intervals are indeed very close to each other for the QOIs in this dataset. The differences between inner- and outer-bound prediction intervals are discussed in Section 2.1.4. The prediction interval results of these computations for the two selected cases, QOIs 1 and 92, are shown in Fig. 4.8. The results for all 100 QOIs are presented in Figure A.2 of Appendix A.

As shown in Fig. 4.8, for each QOI, four pairs of intervals are plotted, corresponding to the detailed and reduced models. For each pair, colored in red is the inner-bound prediction interval and colored in blue is the outer-bound one. The horizontal dashed lines mark the prediction intervals obtained with the detailed model. In both cases, the prediction intervals of reduced model R25 are very close to those of the detailed model. In contrast, the prediction intervals of R17 are narrower than those using the detailed model for QOI 92. The same outcome was obtained for several other QOIs, thus indicating, that R17 underestimates the uncertainties in predictions, consistent with the assessment of the posterior distributions in the preceding subsection. The prediction intervals of R12 are shortened to a much larger extent and for all QOIs, again, consistent with the assessment by comparing the prediction histograms.

Hellinger distances h^U , as computed by Eq. (3.10), for all 100 QOIs with the three reduced models are reported in Table A.3 of Appendix A, the distributions of these values over the 100 QOIs are displayed in Fig 4.7 (left column), and their summary statistics are given in Table 4.5. Inspection of Fig. 4.7 and Table 4.5 indicates that the computed values of the numerical measure, $h_p^U(D, R)$, reflect the visual assessment of the intervals, illustrated in Fig. 4.8, and are consistent with the conclusions reached through the analysis of the posterior QOI distributions for the UQ performance of the reduced models.

The average h_p^U is smaller than the average h_p^H for all three reduced models in Table 4.5, also shown in Fig. 4.7. This may be explained by the assumption of uniform distributions on the prediction intervals when calculating h_p^U 's—all values within the prediction interval were treated equally. In contrast, posterior QOI prediction distributions contain richer

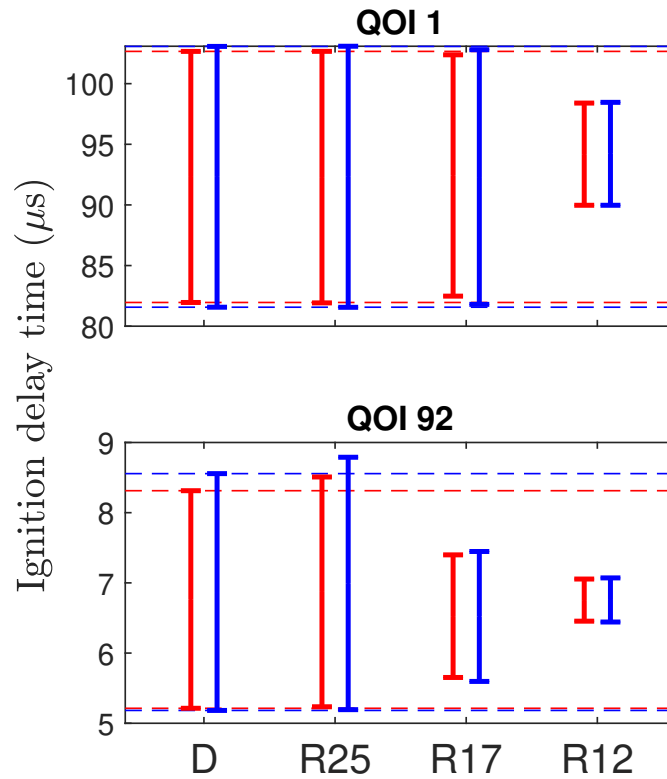


Figure 4.8: QOI prediction intervals computed with the detailed, D, and reduced, R25, R17, and R12, models; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and horizontal dashed lines are prediction intervals of the detailed model.

information in that they put different weights (probability density) on different values.

Despite possible loss of information, h_p^U 's can be computed much faster, as discussed in Section 2.1.4. In addition to the lower computational cost, using prediction intervals has another advantage—capturing extreme values in model predictions. For each QOI and model combination, the prediction interval is formed by the minimum and maximum of model predictions, as discussed in Section 3.4. In other words, the prediction interval is optimization-based. As a result, the sample of model predictions for the same QOI and model combination, as obtained in Section 4.2.4, should fall within the corresponding optimization-based prediction interval. For each QOI and model combination, the interval formed by the minimum and maximum of the sample of model predictions, termed the sample-based prediction interval, was computed. After that, the relative length of the sample-based prediction interval, defined as the ratio of the length of the sample-based prediction interval to the length of the optimization-based prediction interval, was computed. For each model, the

summary statistics of the relative lengths over the 100 QOIs are tabulated in Table 4.6. On average, the relative length is approximately 0.7 for the detailed, R25, and R17 models, with the minimum among the 100 QOIs reaching as low as 0.4626. This indicates that the sample of model predictions, i.e., the posterior distribution of model predictions in Section 4.2.4, only covers roughly 70% of all possible values of model predictions, determined by the optimization-based prediction interval.

Table 4.6: Summary statistics of the relative lengths of sample-based prediction intervals (over 100 QOIs).

	Mean	Min.	Max.
detailed	0.7147	0.4626	0.818
R25	0.7154	0.4760	0.8223
R17	0.7307	0.6184	0.8315
R12	0.1771	0.0600	0.2774

For each QOI and model combination, the sample-based prediction interval is shorter than the optimization-based prediction interval. This is because the uniform sample collected in the feasible set may have missed some parts of the feasible set, and points in these parts may lead to extreme values in model predictions. The comparison of sample- and optimization-based intervals of model parameters in the example with computer-generated QOIs in Section 4.2.3, as illustrated in Fig. 4.5, suggests that the sampling algorithm indeed missed some parts of the feasible set.

4.2.6 Comparison of sensitivities

Sensitivities of prediction interval with respect to the prediction intervals of other QOIs and with respect to the uncertainty intervals of model parameters were computed, as described in Section 3.5. The sensitivity coefficients computed for the selected QOIs 1 and 92 are shown in Figs. 4.9 and 4.10, respectively. The sensitivity coefficients for all 100 QOIs can be found in Figure A.3 of Appendix A.

Inspection of the results for QOI 1 shows that the sensitivity spectra are basically similar among the two reduced, R25 and R17, and detailed models, consistent with the other measures that indicated the close proximity of these models in the case of QOI 1. For QOI 92, the sensitivity spectra of R25 are still similar to the spectra of the detailed model, yet those of R17 are not, again, in accord with other indicators. It is interesting to note that the sensitivity spectra of R17 is missing three of the appreciable peaks present in the spectra of the R25 and detailed models—the peaks corresponding to reactions removed in the model reduction (namely, $\text{OH} + \text{OH} \rightleftharpoons \text{O} + \text{H}_2\text{O}$, $\text{OH} + \text{H}_2\text{O}_2 \rightleftharpoons \text{HO}_2 + \text{H}_2\text{O}$, and $\text{HO}_2 + \text{HO}_2 \rightleftharpoons \text{O}_2 + \text{H}_2\text{O}_2$)—and this removal has to be compensated by other reactions remaining in R17. The sensitivity spectra of R12, as expected from other indicators, are highly different from those of the detailed model.

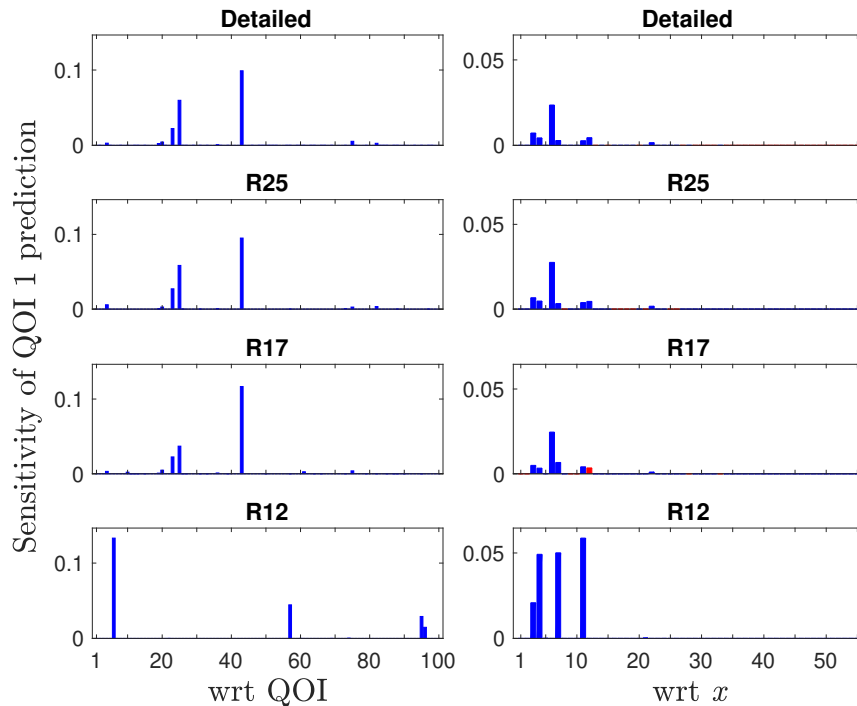


Figure 4.9: Sensitivity of prediction interval for QOI 1 with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.

The sensitivities of the Bhattacharyya coefficients, which measure the overlap, were computed as described in Section 3.5. The sensitivities of the Bhattacharyya coefficients for QOI 92 are displayed in Fig. 4.11. The first feature to notice in this figure is that both $\beta_{p,e}^R$'s (wrt QOIs) and $\beta_{p,i}^R$'s (wrt x) increase in their absolute values moving from R25 to R17 to R12. With R25 reproducing closely the prediction interval of the detailed model, the overlap between the two does not change appreciably with the uncertainty bounds. Reduction in model reproducibility from R25 to R17 increases the sensitivity values, as the predicted interval bounds of the reduced and detailed models begin to respond differently to the same variations in the uncertainty bounds. Further reduction of R12 leads to much larger values and more populated sensitivity spectra.

To help in understanding the sign change of the sensitivity values, one can consider a situation when the reduced-model interval is contained entirely within that of the detailed model, i.e., the overlap region in this case is that of the reduced model, $\mathcal{L}_p^O = \mathcal{L}_p^R$. The

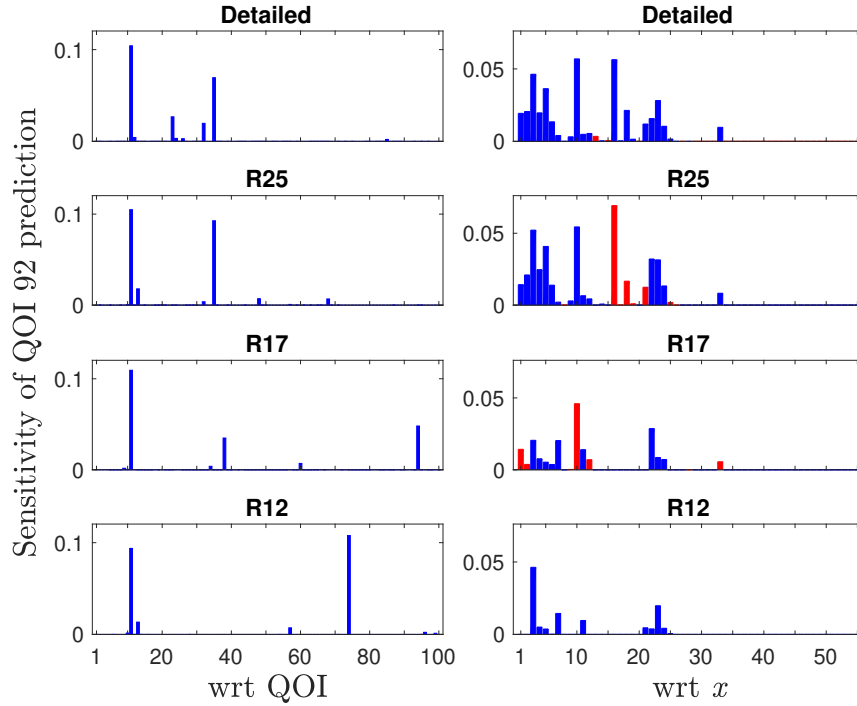


Figure 4.10: Sensitivity of prediction interval for QOI 92 with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.

Bhattacharyya coefficient (3.20) in this case takes the form

$$\mathcal{B}_p^U(\mathcal{D}, \mathcal{R}) = \left(\frac{\mathcal{L}_p^{\mathcal{R}}}{\mathcal{L}_p^{\mathcal{D}}} \right)^{1/2} \quad (4.5)$$

and its sensitivities become

$$\beta_p^{\mathcal{R}} = \frac{1}{2} \left(\frac{S_p^{\mathcal{R}}}{\mathcal{L}_p^{\mathcal{R}}} - \frac{S_p^{\mathcal{D}}}{\mathcal{L}_p^{\mathcal{D}}} \right). \quad (4.6)$$

For $S_p^{\mathcal{R}} \approx S_p^{\mathcal{D}}$ and $\mathcal{L}_p^{\mathcal{R}} \approx \mathcal{L}_p^{\mathcal{D}}$, $\beta_p^{\mathcal{R}} \approx 0$, as was discussed above for R25. With the increase in the level of reduction, the sensitivity values generally increase and the prediction interval shortens, thus leading to positive values of $\beta_p^{\mathcal{R}}$. However, as is seen in Figs. 4.9 and 4.10, some sensitivity peaks disappear for the increased-level reduced models, resulting in $S_p^{\mathcal{R}} = 0$ and correspondingly with $\beta_p^{\mathcal{R}}$ turning negative.

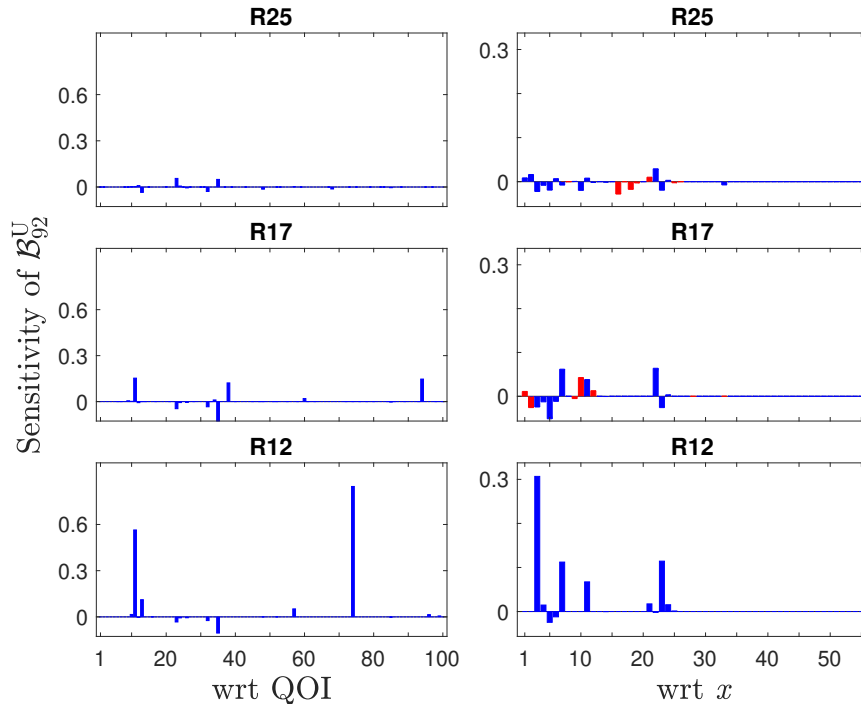


Figure 4.11: Sensitivity of the Bhattacharyya coefficient of QOI 92 with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to reactions of a given model that are absent from the model displayed in the next panel below.

4.3 A dataset with experimental ignition delay QOIs

The second example for syngas combustion is a dataset constructed with QOIs selected from the latest DLR-SynG dataset [113], which includes ignition delay times measured at 223 different initial conditions. In this section, R12 is not considered because of its large deviations from the detailed model. Measures introduced in Chapter 3 were applied to the two reduced models, R25 and R17.

4.3.1 QOIs and surrogate models

In contrast to the example in Section 4.2, the QOI bounds of the DLR-SynG dataset, i.e., L_e 's and U_e 's, were compiled from experimental studies [113]. The ranges of conditions of the 223 initial conditions in the DLR-SynG dataset are wider than those used in Section 4.1.2 to develop reduced models: temperature 1000–1800 K, pressure 0.5–2 atm, equivalence ratio

0.04–1.5, and CO/H₂ ratio 0–2. Because the reduced models were developed for such ranges, their applicability should be limited to such ranges—any extrapolation, if necessary, should be treated with extreme care. Only 26 of the 223 initial conditions fall into the abovementioned ranges of temperature, pressure, equivalence ratio, and CO/H₂ ratio simultaneously.

Table 4.7: Initial conditions of 22 QOIs remained in the dataset for UQ analysis; ignition delay times τ^D are those computed with the detailed model and the relative deviations for the reduced models are defined as $\frac{|\tau^R - \tau^D|}{\tau^D} \times 100\%$.

No.	T (K)	P (atm)	Mole fraction				ϕ	τ^D (μ s)	Rel. dev. (%)	
			CO	H ₂	O ₂	N ₂			R25	R17
1	1125	1.10	0.071	0.103	0.172	0.654	0.506	75.59	0.14	0.54
2	1169	1.00	0.104	0.070	0.174	0.651	0.500	73.39	0.20	0.39
3	1026	1.10	0.104	0.070	0.174	0.651	0.500	187.47	0.01	0.91
4	1162	1.00	0.104	0.070	0.174	0.651	0.500	75.93	0.57	0.09
5	1072	1.10	0.071	0.103	0.172	0.654	0.506	108.96	0.33	0.48
6	1175	1.00	0.071	0.103	0.172	0.654	0.506	61.80	0.28	0.58
7	1151	1.00	0.035	0.139	0.174	0.652	0.500	63.61	0.44	0.33
8	1161	1.50	0.012	0.028	0.020	0.000	1.000	303.50	0.35	0.29
9	1227	1.74	0.030	0.015	0.015	0.000	1.500	264.07	1.23	0.98
10	1053	1.89	0.030	0.015	0.015	0.000	1.500	724.53	0.09	0.85
11	1200	1.81	0.016	0.015	0.016	0.000	1.003	288.90	0.39	0.55
12	1064	1.79	0.016	0.015	0.016	0.000	1.003	680.68	0.49	0.10
13	1027	1.04	0.000	0.035	0.035	0.000	0.496	683.82	0.31	0.21
14	1061	1.01	0.000	0.059	0.029	0.000	0.996	524.56	0.27	0.31
15	1026	1.01	0.000	0.008	0.040	0.000	0.100	1213.13	0.09	0.10
16	1267	0.82	0.000	0.008	0.040	0.000	0.100	364.37	0.57	0.26
17	1154	0.69	0.000	0.040	0.202	0.000	0.100	146.07	0.45	0.54
18	1675	1.05	0.000	0.040	0.202	0.000	0.100	14.35	1.74	8.63
19	1146	1.60	0.004	0.004	0.089	0.000	0.050	346.22	0.19	1.97
20	1135	1.19	0.116	0.089	0.103	0.448	1.000	96.00	0.31	0.41
21	1077	2.00	0.000	0.300	0.148	0.552	1.014	55.47	0.02	0.54
22	1026	2.00	0.000	0.300	0.148	0.552	1.014	109.75	0.05	0.67

Table 4.8: Experimental bounds of the 22 QOIs remained in the dataset for UQ analysis.

No.	Lower bound (μ s)	Upper bound (μ s)
1	53.60	80.40

2	46.40	69.60
3	122.40	183.60
4	45.60	68.40
5	69.60	104.40
6	44.00	66.00
7	51.20	76.80
8	155.35	322.65
9	159.00	265.00
10	330.85	687.15
11	189.75	316.25
12	359.10	666.90
13	611.20	916.80
14	436.00	654.00
15	901.50	1502.50
16	223.20	334.80
17	108.80	163.20
18	12.60	23.40
19	334.16	620.58
20	66.40	99.60
21	29.09	60.41
22	87.91	163.25

For each of the 26 initial conditions selected from the DLR-SynG dataset, a quadratic polynomial of $\ln \tau^D$ in x was built, as described in Section 2.1.2. Similarly, quadratic polynomials for the reduced models were built. There were 24 QOIs with a fitting error, as defined in Section 2.1.2, less than or equal to 0.01 for the detailed, R25, and R17 models at the same time. Three datasets, each for the detailed, R25, and R17 models, with all of the 24 QOIs were built, and the dataset for R17 was shown to be inconsistent. In other words, the associated feasible set \mathcal{F} , as defined in Eq. (2.4), was an empty set. To resolve this issue, the vector consistency measure (VCM) [76] analysis was performed, and two QOIs were removed. The remaining 22 QOIs were used for UQ analysis in this section; the corresponding initial conditions are listed in Table 4.7, and the experimental bounds of these QOIs are listed in Table 4.8. At the nominal parameter values, the R25 and R17 reduced models reproduced the ignition delays computed by the full-size model with an average relative error of 0.4 and 0.9%, respectively, on par with the average relative errors for the 100 QOIs in Section 4.2. The union of active parameters of all 22 QOIs for the detailed, R25, and R17 models have 24, 21, and 15 parameters, respectively. The surrogate models will be referred to hereafter as $M_e^D(x)$ and $M_e^R(x)$ for the detailed and reduced models, respectively, associated with the e -th QOI.

4.3.2 Comparison of predicted values

Three measures, δ , $\delta^{\mathcal{H}}$, and $\delta^{\mathcal{F}}$, as defined in Section 3.1, were calculated for each of the 22 QOIs and each of the two reduced models. The resulting values are reported in Table A.4 of Appendix A. As in the example with computer-generated QOIs in Section 4.2.2, the $\delta^{\mathcal{H}}$ values of individual QOIs were computed using midpoints of the respective inner and outer bounds; i.e., $\delta_e^{\mathcal{H}} = [\delta_e^{\mathcal{H}}(\text{inner}) + \delta_e^{\mathcal{H}}(\text{outer})] / 2$. Similarly for $\delta_e^{\mathcal{F}}$.

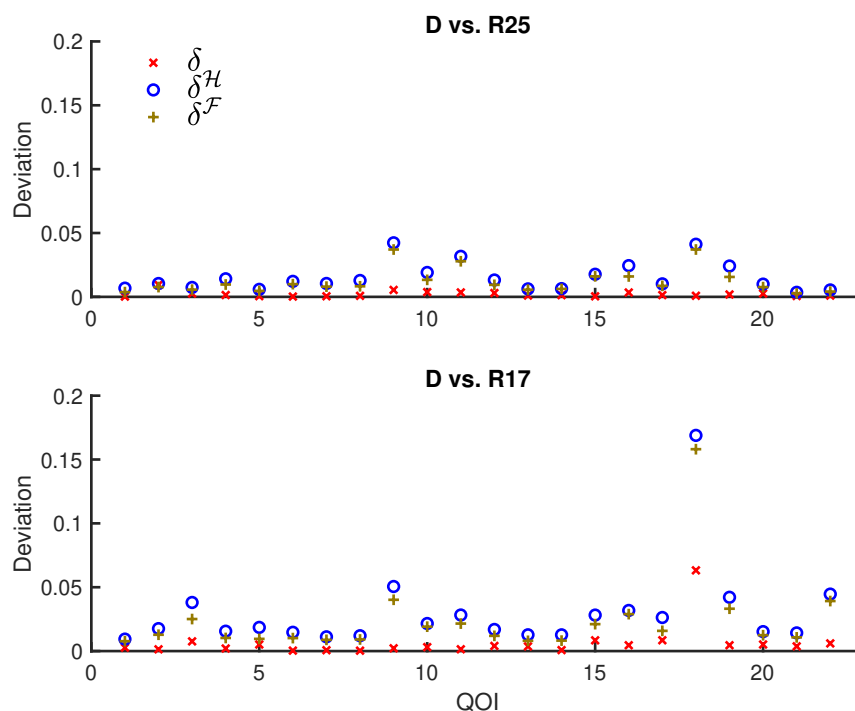


Figure 4.12: Deviations, δ , $\delta^{\mathcal{H}}$, and $\delta^{\mathcal{F}}$, between the detailed (D) model and reduced (R25, R17) models for 22 QOIs.

The three measures for each QOI and each reduced model are displayed in Fig. 4.12, and their summary statistics are listed in Table 4.9. The results show that δ 's are quite small for both R25 and R17. For R17, only one of the QOI has a δ above 0.05, namely 0.063 for QOI 18. As in the example with computer-generated QOIs in Section 4.2.2, one may conclude that both R25 and R17 are relatively accurate reduced reaction models over the assumed ranges of the initial conditions when the model parameters are fixed at the nominal values.

Table 4.9: Summary statistics of deviation measures (over 22 QOIs).

	R25		R17	
	Mean	Max.	Mean	Max.
δ	0.002	0.009	0.006	0.063
$\delta^{\mathcal{H}}$	0.015	0.042	0.030	0.169
$\delta^{\mathcal{F}}$	0.012	0.037	0.024	0.158

As in the example with computer-generated QOIs in Section 4.2.2, $\delta^{\mathcal{H}}$ values, computed by Eq. (3.2) and plotted in Fig. 4.12, are much larger than δ values for several QOIs. For instance, the deviation $\delta^{\mathcal{H}}$ between the detailed model and R17 reaches 0.169 for QOI 18. Again, these results indicate that R17 does not perform that well when the model parameters are different from the nominal values. For each QOI and reduced model, $\delta^{\mathcal{F}}$ of Eq. (3.3) is smaller than $\delta^{\mathcal{H}}$ of Eq. (3.2), yet by relatively small amounts, as shown in Fig. 4.12. The maximum value of $\delta^{\mathcal{H}}$ (0.042) and $\delta^{\mathcal{F}}$ (0.037) shows that R25 can accurately reproduce the detailed model in model predictions under uncertainties in the model parameters.

4.3.3 Comparison of feasible sets

Feasible sets \mathcal{F}^{R25} , and \mathcal{F}^{R17} were compared with \mathcal{F}^{D} in two different ways, as described in Section 3.2. First, the relative volumes of \mathcal{F}^{R25} and \mathcal{F}^{R17} were compared with that of \mathcal{F}^{D} . The relative volume of the feasible set is defined as the ratio of the feasible set volume to the volume of the associated hyperrectangle, $\frac{V(\mathcal{F})}{V(\mathcal{H})}$, as described in Section 3.2. Then, the overlap between \mathcal{F}^{R25} or \mathcal{F}^{R17} and \mathcal{F}^{D} was computed. As in the example with computer-generated QOIs in Section 4.2.3, all 22 QOIs were used to obtain the feasible set for a given model, different from the leave-one-out discussed scheme in Section 3.3.

The hyper-rectangular region of the prior knowledge, \mathcal{H} , was sampled with 1×10^8 points, and the relative volume of a feasible set was estimated using rejection sampling [106]. For the detailed, R25, and R17 models, the estimated relative volumes converged with 1×10^8 points. The convergence of the estimated relative volumes is discussed in the example with computer-generated QOIs in Section 4.2.3. The estimated relative volumes for \mathcal{F}^{D} , \mathcal{F}^{R25} , and \mathcal{F}^{R17} are 1.02×10^{-6} , 1.4×10^{-6} , and 7×10^{-7} , respectively. The relative volumes in this section are significantly smaller than those in the example in Section 4.2.3. In addition, the difference between the relative volume of \mathcal{F}^{D} and that of \mathcal{F}^{R25} is much larger than in the example in Section 4.2.3.

The feasible-set regions were sampled with 1.5×10^4 uniformly-distributed points using the Gibbs sampling algorithm [106] adopted to B2BDC [107] to estimate the feasible-set overlap. Compared with the overlap in the example with computer-generated QOIs in Section 4.2.3, the overlap between \mathcal{F}^{R25} and \mathcal{F}^{D} is much smaller. Specifically, 39.87% points of \mathcal{F}^{R25} fell within \mathcal{F}^{D} , and 50.64% points of \mathcal{F}^{D} fell within \mathcal{F}^{R25} . The overlap between \mathcal{F}^{R17} and \mathcal{F}^{D} is

even smaller—only 8.09% points of \mathcal{F}^{R17} fell within \mathcal{F}^{D} , and 8.48% points of \mathcal{F}^{D} fell within \mathcal{F}^{R17} .

As discussed in the example with computer-generated QOIs in Section 4.2.3, the uniform sample collected in the feasible set may have missed some parts of the feasible set, even if the test of uniformity based on the Gelman convergence test was passed [107]. The union of active parameters of all 22 QOIs in the current dataset has 24 active parameters in the case of the detailed model. For each of the 24 active parameters, the interval formed by the minimum and maximum of that parameter among the sampled 1.5×10^4 points in the feasible set \mathcal{F}^{D} was computed, termed the sampled-based interval. The corresponding optimization-based interval for each active parameter was obtained by solving problem (2.8). The sample- and optimization-based intervals for all active parameters in the case of the detailed model are plotted in Fig. 4.13. For each active parameter, the vertical red line is the inner-bound

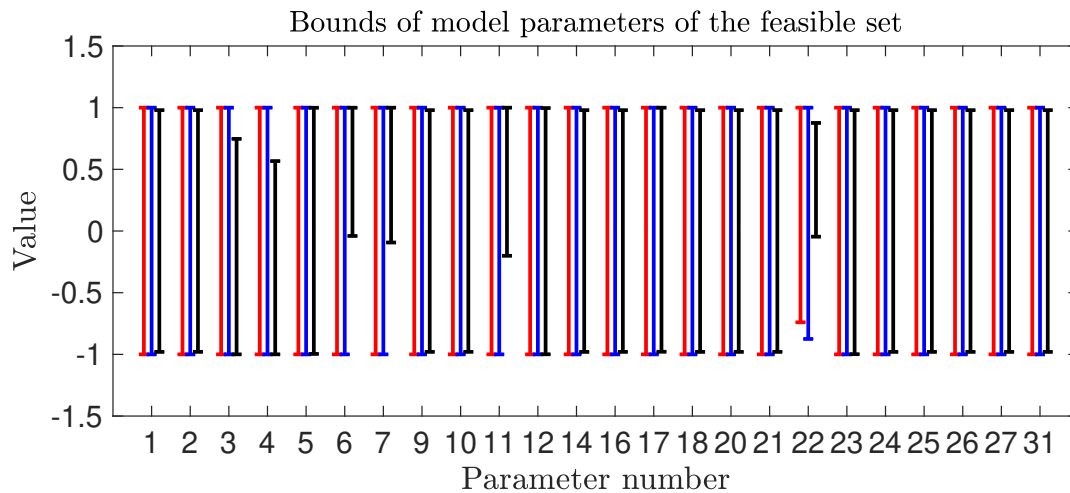


Figure 4.13: Bounds of active model parameters computed with the detailed model; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and vertical black lines are sample-based prediction intervals.

interval, the vertical blue line is the outer-bound interval, and the vertical black line is the sample-based interval. There are several implications from these results. First, the inner- and outer-bound prediction intervals of active parameters are highly close to each other for all but one parameter, namely parameter 22. But the difference is very small for parameter 22. Second, the inner- and outer-bound prediction intervals are very close to the $[-1, 1]$ interval for all active parameters except parameter 22. Recall that the uncertain parameter vector x is transformed so that all model parameters have $[-1, 1]$ intervals, as discussed in Section 4.1.3. This indicates that the feasible set \mathcal{F}^{D} can touch almost all the sides of the initial hyper-rectangle \mathcal{H} , even though its volume is very small relative to the volume of \mathcal{H} (1.02×10^{-6}).

This may pose a substantial challenge to the sampling algorithm—the sampling algorithm may miss some important parts of the feasible set. Third, the sample-based intervals for multiple active parameters are significantly narrower than the optimization-based intervals, as shown in Fig. 4.13. Compared with the results in the example with computer-generated QOIs in Section 4.2.3, the sample-based intervals deviate significantly from the optimization-based ones for multiple active model parameters.

The comparison of the sample- and optimization-based intervals of active model parameters implies that the sampling algorithm missed many parts of the feasible set \mathcal{F}^D , suggesting that the feasible set has a very complex shape. The complex shape of the feasible set may be caused by the poor quality of the data (L_e 's and U_e 's). Previous studies [112, 113] have shown disagreement between models and data in that many QOIs had to be removed to render a dataset consistent. It should be emphasized that the QOI bounds of the DLR-SynG dataset, i.e., L_e 's and U_e 's, were compiled from different sources of experimental studies [113]. The heterogeneous nature of the data could be problematic because data from different laboratories may have different biases and/or uncertainties, and some of them could even be erroneous. What makes the problem more challenging is that it is typically difficult to determine which subset of the data is of bad quality. As a result, a feasible set, defined by a set of polynomial constraints like $L_e \leq M_e(x) \leq U_e$, could have a very complex shape. In contrast, the data used in the example in Section 4.2 were computer generated, which avoids such problems.

4.3.4 Comparison of posterior distributions

For each QOI and model combination, a uniform sample was collected from the corresponding feasible set, as described in Section 3.3. Specifically, 1.5×10^4 points were generated in each of the \mathcal{F}_{-p}^D , \mathcal{F}_{-p}^{R25} , and \mathcal{F}_{-p}^{R17} , for $p = 1, \dots, 22$, using the Gibbs sampling algorithm adopted to B2BDC [107]. The notation \mathcal{F}_{-p} denotes the feasible set for a dataset with the p -th QOI removed in the constraints as a result of the leave-one-out scheme, as discussed in Section 3.3. The points of a uniform sample were then used to evaluate the corresponding surrogate model, and a sample of model predictions was subsequently obtained. The histograms of ignition delay times computed for two selected QOIs, QOI 1 and QOI 18, are shown in Fig. 4.14. The histograms of all 22 QOIs can be found in Fig. A.4 of Appendix A.

The top right subplot shows that the reduced model R25 reproduces closely the posterior distributions of QOI 18 computed with the detailed model. However, the posterior distribution of QOI 1 using R25 deviates appreciably from that computed with the detailed model, as shown in the top left subplot of Fig. 4.14. The predictions of R17 deviate significantly from those of the detailed model, particularly for QOI 18—the two histograms are almost disjoint.

Hellinger distances h^H , as computed by Eq. (3.7), for all 22 QOIs with the two reduced models are reported in Table A.5 of Appendix A, and their summary statistics are given in Table 4.10. In the case of R25, the maximum h^H among the 22 QOIs is much larger than the maximum h^H in the example with computer-generated QOIs in Section 4.2.4 (0.397 vs.

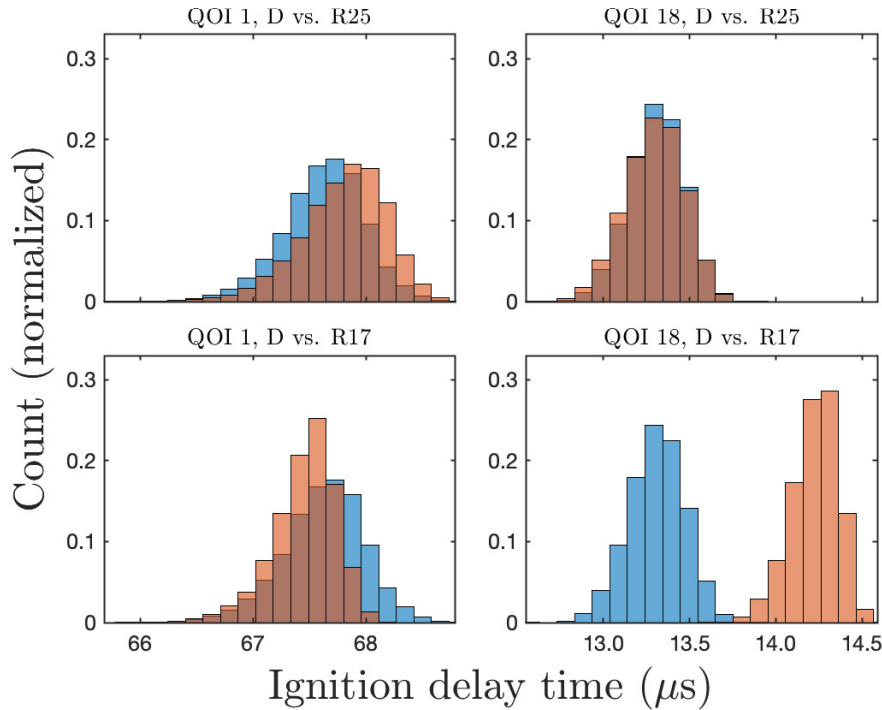


Figure 4.14: Histograms for selected QOIs of detailed (blue) and reduced (brown) model predictions over feasible sets.

0.027), as is the average h^H (0.137 vs. 0.010). As discussed in the previous section, the uniform samples may have missed some parts of the feasible sets and thus lead to larger deviations in the measures computed from the uniform samples.

Table 4.10: Summary statistics of Hellinger distances (over 22 QOIs).

	R25		R17	
	Mean	Max.	Mean	Max.
h^H	0.137	0.397	0.316	0.995
h^U	0.036	0.085	0.067	0.417

4.3.5 Comparison of prediction intervals

For each $p = 1, \dots, 22$, prediction intervals of the detailed and two reduced models were obtained by solving the corresponding optimization problems, as described in Section 3.4.

Averaged over the 22 QOIs, the relative lengths of the inner-bound prediction intervals with respect to the outer-bound prediction intervals are 92.1, 89.5, and 91.0% for the detailed, R25, and R17 models, respectively. This shows that the inner- and outer-bound prediction intervals are still close to each other for the QOIs in this dataset. The differences between inner- and outer-bound prediction intervals are discussed in Section 2.1.4. The prediction intervals for QOIs 1 and 18 are shown in Fig. 4.15. The results for all 22 QOIs are presented in Fig. A.5 of Appendix A.

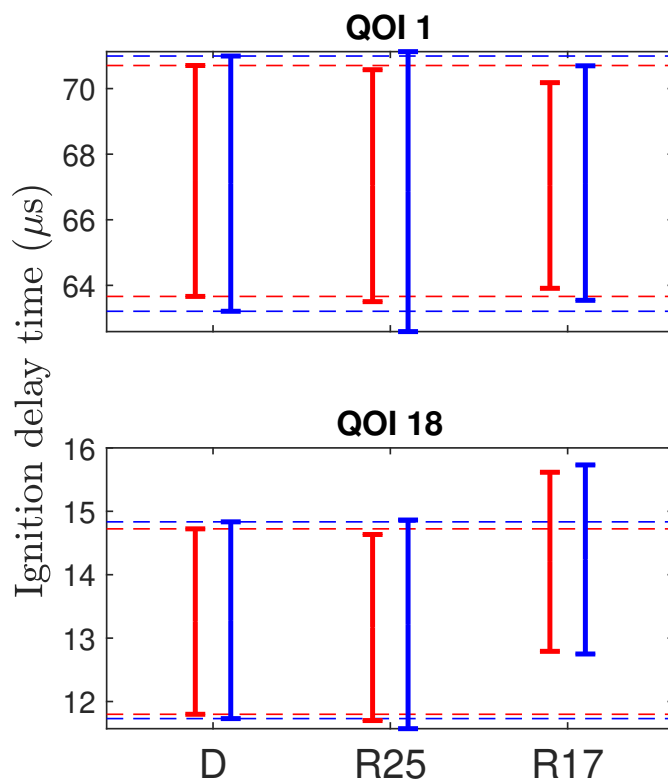


Figure 4.15: QOI prediction intervals computed with the detailed, D, and reduced, R25, and R17, models; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and horizontal dashed lines are prediction intervals of the detailed model.

As shown in Fig. 4.15, the inner-bound prediction intervals of the reduced model R25 are highly close to those of the detailed model, whereas the outer-bound prediction intervals of the reduced model R25 are somewhat wider. As for R17, whereas the prediction intervals are relatively close to those using the detailed model for QOI 1, the prediction intervals are shifted toward larger values for QOI 18.

Hellinger distances h^U , as computed by Eq. (3.10), for all 22 QOIs with R25 and R17 are reported in Table A.5 of Appendix A, and their summary statistics are given in Table 4.10. As in the example with computer-generated QOIs in Section 4.2.5, the average h_p^U is lower than the average h_p^H for both R25 and R17. However, the gap is much larger between h_p^U 's and h_p^H 's for the present system.

As discussed in Section 4.2.5, using h_p^U 's has an advantage—capturing extreme values in model predictions. In a high-dimensional feasible set, e.g., \mathcal{F}^D with 24 active parameters in the current dataset, a sampling algorithm will most likely miss feasible points in many regions. These points may lead to extreme values in model predictions. As in Section 4.2.5, for each QOI and model combination, the interval formed by the minimum and maximum of the sample of model predictions, termed the sample-based prediction interval, was computed. The relative length of the sample-based prediction interval, defined as the ratio of the length of the sample-based prediction interval to the length of the optimization-based prediction interval, was subsequently computed. For each of the detailed, R25, and R17 models, the summary statistics of the relative lengths over 22 QOIs are tabulated in Table 4.11. On average, the relative length is approximately 0.44 for the detailed, R25, and R17 models, with the minimum among the 22 QOIs reaching as low as 0.3051. These numbers are much smaller than those in Section 4.2.5, as listed in Table 4.6.

Table 4.11: Summary statistics of the relative lengths of sample-based prediction intervals to optimization-based prediction intervals (over 22 QOIs).

	Mean	Min.	Max.
detailed	0.4444	0.3623	0.6963
R25	0.4407	0.3239	0.6755
R17	0.4153	0.3051	0.7235

4.3.6 Comparison of sensitivities

Sensitivities of uncertainty of prediction with respect to the QOI bounds of other QOIs and with respect to the uncertainty intervals of model parameters were computed, as described in Section 3.5. The sensitivity coefficients computed for the selected QOIs 1 and 18 are shown in Figs. 4.16 and 4.17, respectively. The sensitivity coefficients for all 22 QOIs can be found in Figure A.6 of Appendix A. For both QOIs 1 and 18, the sensitivity spectra of sensitivities with respect to QOI bounds (left column) are basically similar for all three models, the detailed, R25, and R17 models. Only four to five of the other 21 QOIs seem to have an impact on the model prediction intervals of QOIs 1 and 18. The sensitivity spectra of sensitivities with respect to x (right column) are essentially similar for the detailed model and R25 for both QOIs 1 and 18. The sensitivity spectra of R17 is missing one or three of the appreciable peaks present in the spectra of the R25 and detailed models for QOIs 1 and 18, respectively.

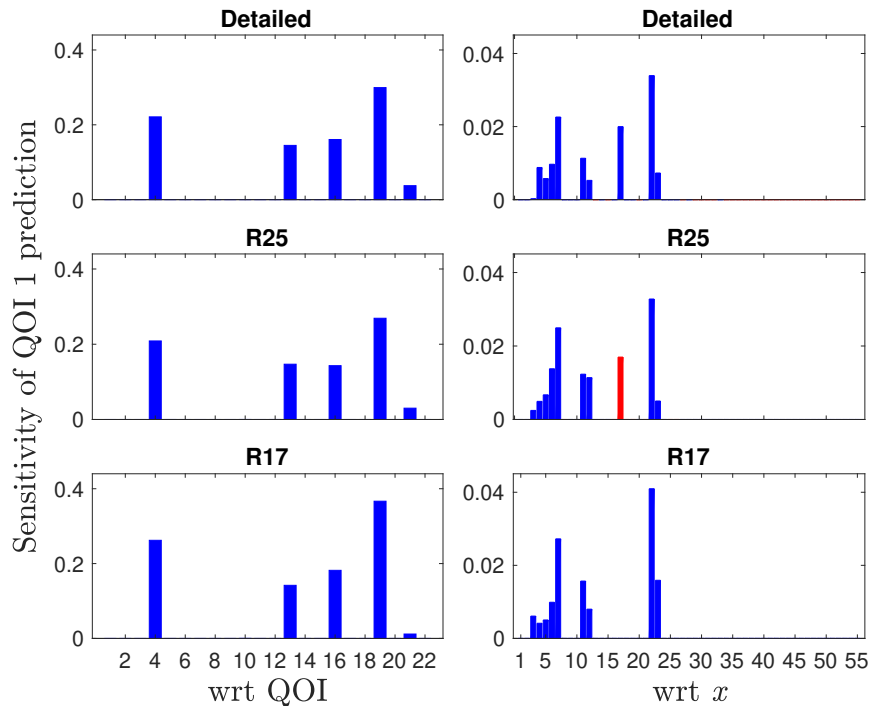


Figure 4.16: Sensitivity of prediction interval for QOI 1 with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.

4.4 Summary

In this chapter, two examples with ignition delay QOIs were investigated to assess how uncertainties in the model parameters and in the model predictions behave for reduced models as compared to those for the detailed model. In the example in Section 4.2, the QOI uncertainty ranges, i.e., L_e 's and U_e 's, were computer generated, whereas the QOI uncertainty ranges in the example in Section 4.3 were compiled from experimental studies.

In the example with computer-generated ignition delay QOIs in Section 4.2, three reduced models, R25, R17, and R12 were assessed under uncertainties in the model parameters. Predicted values of QOIs computed with R25 are very close to those with the detailed model (low $\delta^{\mathcal{H}}$'s and $\delta^{\mathcal{F}}$'s) even when model parameter uncertainties are taken into account. R25 also reproduces accurately the posterior uncertainties in the model parameters, and the associated feasible set, \mathcal{F}^{R25} , has a large overlap with the feasible set associated with the detailed model, \mathcal{F}^{D} , and their relative volumes are particularly close to each other as well. The prediction intervals (low h^U 's) and the posterior distributions of model predictions (low

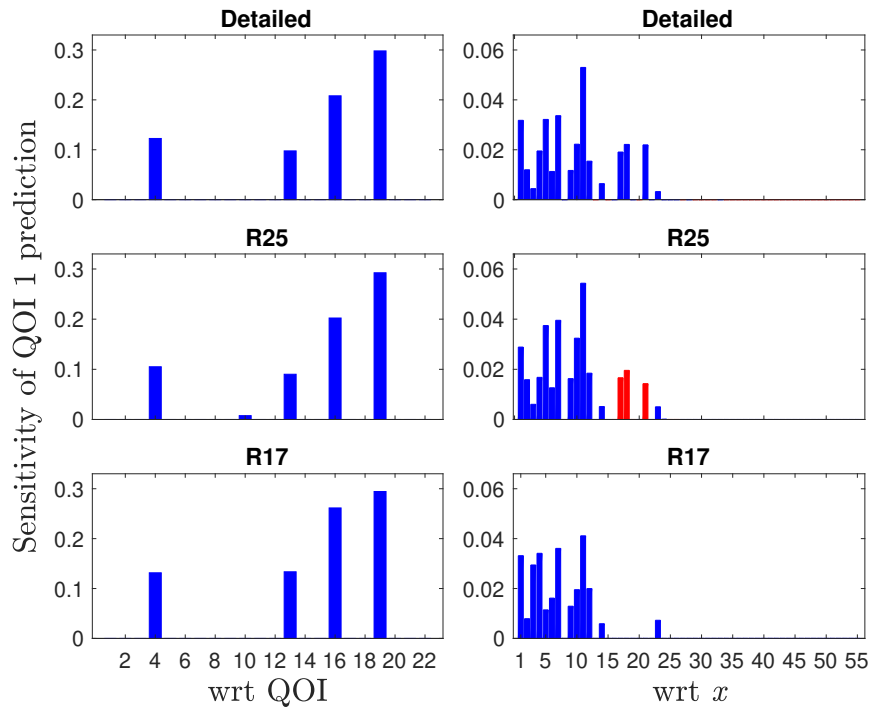


Figure 4.17: Sensitivity of prediction interval for QOI 18 with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.

h^H 's) computed with the detailed model are also reproduced faithfully by R25. In addition, R25 reproduces the spectra of sensitivities of uncertainty in model predictions of the detailed model. A less accurate model, R17, begins to exhibit deviations in all measures, and a much less accurate model, R12, substantially distorts the UQ properties of the detailed model. This example demonstrates several features. First, assessment of the quality of a reduced model without considering parameter uncertainty may be misleading, as is evidenced by the single-point δ test of R17. Inclusion of parameter variations over their uncertainty ranges is more indicative of the reduced-model reproduction of not only the target values but also of their uncertainties. Second, the reduced-model performance can be quantified by several measures of model deviations, and those tested in this example showed mutually consistent and qualitatively similar outcomes.

In the example with experimental ignition delay QOIs in Section 4.3, divergence between different measures occurs. Predicted values of QOIs computed with R25 are quite close to those computed with the detailed model (low δ^H 's and δ^F 's) under uncertainties. R25 also reproduces accurately the prediction intervals (low h^U 's) and the spectra of sensitivities of

uncertainty in model predictions of the detailed model. However, the feasible set \mathcal{F}^{R25} has a relatively small overlap with the feasible set \mathcal{F}^{D} , and their estimated relative volumes are significantly different. In addition, the posterior distributions of model predictions computed with R25 deviate appreciably from those computed with the detailed model (large h^{H} 's). This behavior is indicative of the quality of data rather than of the accuracy of R25. The data (L_e 's and U_e 's) used in this example were shown in previous studies [112, 113] to have disagreement with the model. The data were collected from different sources and they may have different biases and/or uncertainties and some of them could even be erroneous. As a result, the feasible set \mathcal{F}^{D} in this example has a very complex shape such that the sampling algorithm missed many parts of the feasible set. This is implied by the observation that the sample-based intervals for multiple active parameters are significantly narrower than the optimization-based intervals, as shown in Fig. 4.13. Therefore, measures computed from the uniform samples collected in the feasible sets tend to be larger than those in the example with computer-generated QOIs in Section 4.2. These results indicate that potentially bad data (L_e 's and U_e 's) can lead to challenges in UQ analysis. If no verifiably accurate experimental data exist, computer-generated data from the solution of the detailed model offer a reliable alternative, in which the desired level of reduced-model accuracy can be prescribed by specifying the accepted ranges of variations in prediction of training targets. The results also illustrate that measures based on optimization using the B2BDC methodology (δ^{H} 's, δ^{F} 's, and h^{U} 's) are free from issues caused by the complex shape of the feasible sets.

Chapter 5

Reduced models of soot oxidation

In this chapter, a KMC soot oxidation system is investigated. The specifics of the selected full-size, detailed model and the reduced models developed from it are described next.

5.1 Modeled system

5.1.1 Detailed KMC model

The detailed model used in this chapter is a KMC soot oxidation model with 107 surface reactions developed recently by Frenklach et al. [20, 44]. The model is introduced in Section 2.3, and the surface reactions are reproduced in Table 2.1. The model was validated against the high-temperature shock-tube experiments of Roth et al. [103]. In this chapter, a constant gas-phase environment will be used in running the soot oxidation model, similar to the simulations in [43, 49, 51, 52]. The choice of using a constant gas-phase environment was to focus on the KMC oxidation model, removing the effects of gas-phase environments on model predictions. For a given simulation, two quantities of interest (QOIs) will be used for analysis, the total number of CO molecules and the number of C₂H₂ molecules released.

5.1.2 Reduced KMC models

Reduced models were developed from the detailed model by identifying non-contributing reactions using the reaction counts from a simulation with the detailed model, similar to the idea of the “detailed-reduction” method, introduced in Section 2.2. The temperature and pressure used for this simulation were the same as in [20]. The concentrations (mole fractions) of the gases used were taken from the gas-phase profiles in the high-temperature simulation of [20] at the time instance of peak H-concentration. The temperature, pressure, and mole fractions of major gases are listed in Table 5.1. The other gases in the gas-phase environment, including C₂H₂, CH₃, A1, and H₂O, were assigned a value of mole fraction of zero. The simulation time period was 2 ms, and the substrate used in the simulation is the

19-ring PAH, as shown in Fig. 2.3. The rate parameters used are the nominal parameters of the detailed model, as reported in Table 2.1.

Table 5.1: Gas-phase environment conditions for KMC oxidation simulation.

T (K)	p (bar)	x_{H_2}	x_{H}	x_{O}	x_{OH}	x_{O_2}
1990	0.72	0.0050	0.0088	0.0027	0.0023	0.0028

Multiple KMC runs, specifically 100 runs with 100 different seeds, were performed. The reaction counts were averaged over these 100 runs. Three different reduced models were constructed by removing reactions whose reaction counts were less than 1, 5, and 10, respectively. These three reduced models will be referred to hereafter as R1, R5, and R10, where the number in the name is the cutoff for reaction counts. The resulting reduced models, R1, R5, and R10, include 62, 46, 38 reactions, respectively. The reactions retained in each of the reduced models are reported in Table 5.2.

Table 5.2: Reaction counts of all 107 reactions in the simulation with the detailed model, and identification of reactions in reduced models R1, R5, and R10.

Reaction No.	Reaction Counts	R1	R5	R10
1	485.78	✓	✓	✓
2	917.52	✓	✓	✓
3	0.00	—	—	—
4	0.00	—	—	—
5	78.14	✓	✓	✓
6	38.45	✓	✓	✓
7	19.50	✓	✓	✓
8	82.10	✓	✓	✓
9	11.12	✓	✓	✓
10	123.20	✓	✓	✓
11	121.93	✓	✓	✓
12	282.43	✓	✓	✓
13	210.52	✓	✓	✓
14	0.00	—	—	—
15	2.65	✓	—	—
16	4.12	✓	—	—
17	1.72	✓	—	—
18	0.00	—	—	—
19	0.00	—	—	—
20	0.00	—	—	—
21	4.27	✓	—	—

22	0.47	—	—	—
23	3.27	✓	—	—
24	16.27	✓	✓	✓
25	4.19	✓	—	—
26	0.00	—	—	—
27	0.00	—	—	—
28	0.02	—	—	—
29	0.12	—	—	—
30	0.02	—	—	—
31	0.22	—	—	—
32	0.01	—	—	—
33	0.00	—	—	—
34	0.00	—	—	—
35	0.00	—	—	—
36	0.00	—	—	—
37	0.00	—	—	—
38	0.00	—	—	—
39	0.00	—	—	—
40	0.00	—	—	—
41	0.00	—	—	—
42	0.00	—	—	—
43	0.00	—	—	—
44	0.00	—	—	—
45	0.26	—	—	—
46	1.87	✓	—	—
47	4.81	✓	—	—
48	5.81	✓	✓	—
49	6.37	✓	✓	—
50	28.93	✓	✓	✓
51	7.52	✓	✓	—
52	15.97	✓	✓	✓
53	26.23	✓	✓	✓
54	25.45	✓	✓	✓
55	10.93	✓	✓	✓
56	24.35	✓	✓	✓
57	30.06	✓	✓	✓
58	33.07	✓	✓	✓
59	141.67	✓	✓	✓
60	4.47	✓	—	—
61	10.11	✓	✓	✓
62	13.61	✓	✓	✓
63	13.59	✓	✓	✓

64	111.41	✓	✓	✓
65	0.61	—	—	—
66	3.23	✓	—	—
67	4.37	✓	—	—
68	5.22	✓	✓	—
69	171.48	✓	✓	✓
70	0.23	—	—	—
71	0.38	—	—	—
72	0.49	—	—	—
73	0.71	—	—	—
74	5.07	✓	✓	—
75	0.50	—	—	—
76	1.34	✓	—	—
77	1.59	✓	—	—
78	1.47	✓	—	—
79	13.87	✓	✓	✓
80	0.00	—	—	—
81	0.01	—	—	—
82	0.04	—	—	—
83	0.00	—	—	—
84	0.41	—	—	—
85	7.57	✓	✓	—
86	16.10	✓	✓	✓
87	20.71	✓	✓	✓
88	22.70	✓	✓	✓
89	178.06	✓	✓	✓
90	0.00	—	—	—
91	0.00	—	—	—
92	394.23	✓	✓	✓
93	4.11	✓	—	—
94	18.46	✓	✓	✓
95	18.46	✓	✓	✓
96	8.87	✓	✓	—
97	20.55	✓	✓	✓
98	28.94	✓	✓	✓
99	26.55	✓	✓	✓
100	0.00	—	—	—
101	0.00	—	—	—
102	0.00	—	—	—
103	0.00	—	—	—
104	90.99	✓	✓	✓
105	7.32	✓	✓	—

106	1.21	✓	—	—
107	0.09	—	—	—

The KMC simulations with the same environment were then repeated with the three developed reduced models. The two QOIs—the total number of CO molecules and the total number of C₂H₂ molecules released during the two-millisecond simulation—computed with the detailed and reduced models are listed in Table 5.3. Note that the relative errors for R1 and R5 are much larger than those in the case of ignition delay times of syngas combustion, as reported in Sections 4.2 and 4.3. Yet, this level of relative error is deemed acceptable in soot modeling due to larger model and experimental uncertainties [20]. R10 deviates significantly from the detailed model but will be included in the analysis for comparison. Nominal parameter values were used for all three reduced models.

Table 5.3: QOIs calculated using the detailed and reduced models with nominal model parameters.

QOI	detailed	R1	R5	R10
number of CO released	251.09	240.45	240.14	41.54
number of C ₂ H ₂ released	103.56	99.60	82.95	11.77

5.1.3 Model parameters of KMC models

Although there are 107 surface reactions in the detailed KMC soot oxidation model, as listed in Table 2.1, the number of model parameters is less than 107. This is because some surface reactions are related to each other. Surface reactions 1 and 105 are the forward and reverse directions of the same reaction. Surface reactions 55–59 are essentially the same reaction, except for the difference in the location of the active site. Surface reactions 29 and 30 are also coupled reactions in the detailed model [20]. As a result, these surface reactions were organized into 42 groups, each with its associated model parameters, as shown in Table 5.4.

Table 5.4: Groups of reactions.

Group No.	Reactions
1	[1, 105]
2	[2]
3	[3, 7]
4	[4, 5]
5	[6, 8]
6	[9, 21, 28, 37]
7	[10, 106]
8	[11]

9	[12, 107]
10	[13]
11	[14, 15, 18, 19, 20, 22, 38]
12	[16]
13	[17, 33, 34, 39]
14	[23]
15	[24]
16	[25, 26]
17	[27]
18	[29, 30]
19	[31, 32]
20	[35]
21	[36]
22	[40]
23	[41]
24	[42]
25	[43]
26	[44, 45]
27	[46, 47]
28	[48, 49, 50]
29	[51]
30	[52]
31	[53, 54]
32	[55, 56, 57, 58, 59, 70, 71, 72, 73, 74]
33	[60, 61, 62, 63, 64, 65, 66, 67, 68, 69]
34	[75, 76, 77, 78, 79, 80, 81, 82, 83, 84]
35	[85, 86, 87, 88, 89]
36	[90]
37	[91]
38	[92]
39	[93, 94, 95]
40	[96, 97, 98, 99]
41	[100, 101, 102, 103]
42	[104]

As in Section 4.1.3, it is assumed that the uncertainty of the soot oxidation models resides in the reaction rate coefficients, and their uncertainties are represented by multipliers to the nominal parameter expressions [86],

$$k_j = \lambda_i k_j^{\text{nominal}}, \quad j = 1, \dots, n_i, \quad (5.1)$$

where i is the group number (ranging from 1 to 42 in Table 5.4) and λ_i is the associated multiplier, k_j is the rate coefficient of reaction j , ranging from 1 to n_i (the number of reactions

in group i). The bounds of λ_i 's were assumed to be $[0.5, 2]$. The nominal rate coefficients in the present study are those reported in Table 2.1.

The surrogate models were developed in transformed variables

$$x_i = \frac{\ln \lambda_i / \lambda_{i,\min}}{\frac{1}{2} \ln \lambda_{i,\max} / \lambda_{i,\min}} - 1, \quad i = 1, \dots, 42. \quad (5.2)$$

Defined in this manner, all model parameters have $[-1, 1]$ intervals, i.e., $x \in [-1, 1]^{42}$, and the nominal set of rate coefficients is specified by $x^{\text{nominal}} = x(\lambda = 1)$. There are 42, 29, 23, 22 groups in the detailed, R1, R5, and R10 models, respectively, as listed in Table 5.5. In other words, there are that many uncertain parameters for each of the model. Note that a group retained for a reduced model may have less reactions than that group for the detailed model.

Table 5.5: Number of reactions and groups in the detailed and reduced models.

	detailed	R1	R5	R10
number of reactions	107	62	46	38
number of groups	42	29	23	22

5.2 Parallel computing using MPI

The detailed KMC soot oxidation model in [20] was written in Python 2 [114]. For this dissertation, the code was updated and adapted for Python 3 [115]. As described in Section 2.1.2, the original model, i.e., the detailed model in this case, has to be run for many design points in building a surrogate model for each QOI. Depending on the number of active parameters, the number of design points could vary between a few hundreds to a few thousands. For example, there were 816 design points when the simulation was run with the reduced model R10. Running one design point took approximately one hour. Therefore, it would have taken about 34 days to run all the design points for R10 (recall that there are three other models, the detailed, R1, and R5 models) when running on a single core. It would still have taken roughly 9 days on a high-end desktop with 4 physical cores.

To reduce the computation time, large-scale parallel computing was leveraged in running the code at the design points. In this specific problem, each design point is independent from other design points, making it a so-called embarrassingly parallel problem. The Python package of Message Passing Interface (MPI), MPI4PY [116–118], was used for this purpose. Note that MPI is capable of running much more complicated jobs that do have core-to-core communications. The specific computing resources used were Haswell compute nodes of the

Cori super computer system of NERSC ¹. Each Haswell compute node has 32 physical cores [119]. Therefore, 26 nodes were requested in running the job with 816 design points. The job took just about one hour, and there were no scaling issues because no communication is necessary between cores in a embarrassingly parallel problem.

5.3 A dataset with soot oxidation QOIs

In this section, I illustrate the developed methodology with a dataset with soot oxidation QOIs. The results of applying the measures introduced in Chapter 3 to the three reduced models, R1, R5, and R10 derived from the previous section, are discussed.

5.3.1 QOIs and surrogate models

Due to the high computation cost, only one simulation with one gas-phase environment was used in the analysis. The gas-phase environment was the same as in Section 5.1.2, as listed in Table 5.1. Two QOIs were derived from this simulation, namely the total number of CO molecules (QOI 1) and the total number of C₂H₂ molecules (QOI 2) released during a two-millisecond simulation.

The “observed” values of the QOIs were obtained from solution of the detailed model with the nominal set of rate coefficients, denoted y^D . The QOI uncertainty ranges, i.e., L_e ’s and U_e ’s defined in Section 2.1, were assumed to be $\pm 10\%$ intervals of y^D ,

$$[0.9 y_e^D, 1.1 y_e^D], \quad e = 1 \text{ or } 2. \quad (5.3)$$

For each of the two QOIs, a quadratic surrogate of $\ln y^D$ in x (defined by Eq. (5.2)) was built following the steps described in Section 2.1.2. Similarly, quadratic surrogate models for the reduced models were built. The number of active variables and the fitting errors of the surrogate models, as defined in Section 2.1.2, are listed in Table 5.6. For the detailed, R1, and R5 models, the fitting error was below 0.05, which is equivalent to a roughly 5% relative change, as discussed in Section 4.2.2. The fitting errors for R10 were somewhat larger, but R10 was known a priori to deviate significantly from the detailed model in that its predictions were off by a substantial amount, as shown in Table 5.3. However, R10 is included in the UQ analysis for comparison. The surrogate models will be referred to hereafter as $M_e^D(x)$ and $M_e^R(x)$ for the detailed and reduced models, respectively, associated with the e -th QOI, where R is a place holder for R1, R5, or R10.

¹The results in this chapter used resources of the National Energy Research Scientific Computing Center (NERSC) [119], a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

Table 5.6: Number of active variables and fitting errors of surrogate models for QOIs 1 and 2.

	detailed	R1	R5	R10
number of active variables	10	11	10	15
fitting error (QOI 1)	0.0346	0.0401	0.0403	0.0573
fitting error (QOI 2)	0.0360	0.0421	0.0429	0.0772

5.3.2 Comparison of predicted values

Three measures, δ , $\delta^{\mathcal{H}}$, and $\delta^{\mathcal{F}}$, as defined in Section 3.1, were calculated for QOIs 1 and 2 for each of the three reduced models. The resulting values are reported in Table 5.7. As before, the average of inner- and outer-bound values were taken for $\delta^{\mathcal{H}}$ and $\delta^{\mathcal{F}}$.

Table 5.7: Deviations, δ , $\delta^{\mathcal{H}}$, and $\delta^{\mathcal{F}}$, between the detailed (D) and reduced (R1, R5, R10) models for QOIs 1 and 2.

	R1	R5	R10
δ (QOI 1)	0.0589	0.1007	1.8091
$\delta^{\mathcal{H}}$ (QOI 1)	0.7463	0.9342	3.0054
$\delta^{\mathcal{F}}$ (QOI 1)	0.6175	0.8814	2.6573
δ (QOI 2)	0.0697	0.2846	2.2427
$\delta^{\mathcal{H}}$ (QOI 2)	0.8245	1.2009	3.6784
$\delta^{\mathcal{F}}$ (QOI 2)	0.6352	1.0237	3.3022

The deviation at the nominal parameter values, δ , is about 0.06 and 0.07 for QOIs 1 and 2 in the setting of the reduced model R1, which can be considered as small deviations in soot modeling. However, $\delta^{\mathcal{H}}$'s are much larger for both QOIs in the case of R1, namely 0.7463 and 0.8245, respectively. As expected, $\delta^{\mathcal{F}}$ is less than $\delta^{\mathcal{H}}$ for R1, although not by much. All three measures grow from R1 to R5 to R10. It is noteworthy that only reactions for which the reaction count is less than 1 were removed in constructing R1 in Section 5.1.2. Moreover, 28 of the 47 reactions removed in R1 had zero counts, as listed in Table 5.2. Yet, these reactions combined can lead to large values of $\delta^{\mathcal{H}}$. However, large values of $\delta^{\mathcal{H}}$ do not necessarily mean that R1 cannot reproduce with good accuracy the uncertainties in the model predictions of the detailed model, as will be discussed in Sections 5.3.4 and 5.3.5.

5.3.3 Comparison of feasible sets

Feasible sets \mathcal{F}^{R1} and \mathcal{F}^{R5} were compared with \mathcal{F}^{D} in two different ways, as described in Section 3.2. First, the relative volumes of \mathcal{F}^{R1} and \mathcal{F}^{R5} were compared with that of \mathcal{F}^{D} . The relative volume of the feasible set is defined as the ratio of the feasible set volume to

the volume of the associated hyperrectangle, $\frac{V(\mathcal{F})}{V(\mathcal{H})}$, as described in Section 3.2. Then, the overlap between \mathcal{F}^{R1} or \mathcal{F}^{R5} and \mathcal{F}^{D} was computed. The feasible set \mathcal{F}^{R10} is an empty set because the associated dataset was inconsistent, as discussed in Section 2.1.3. This is not surprising because the deviations at the nominal parameters are very large for R10, as shown in Table 5.3. As a result, not a single point exists in $[-1, 1]^{42}$ such that both $L_1 \leq M_1^{\text{R10}}(x) \leq U_1$ and $L_2 \leq M_2^{\text{R10}}(x) \leq U_2$ hold. Note that both QOIs 1 and 2 were used to define the feasible set for a given model, different from the feasible sets in the leave-one-out setting.

The hyper-rectangular region of the prior knowledge, \mathcal{H} , was sampled with 1×10^7 points, and the relative volume of a feasible set was estimated using rejection sampling [106]. For the detailed, R1, and R5 models, the estimated relative volumes converged with 1×10^7 points, as explained in Section 4.2.3. The estimated relative volumes for \mathcal{F}^{D} , \mathcal{F}^{R1} , and \mathcal{F}^{R5} are 0.3035, 0.342, and 0.1724, respectively. Whereas the relative volume of \mathcal{F}^{R1} is close to that of \mathcal{F}^{D} , the relative volume of \mathcal{F}^{R5} is much smaller than that of \mathcal{F}^{D} .

To estimate the feasible-set overlap, the feasible-set regions were sampled with 1.5×10^4 uniformly-distributed points using Gibbs sampling algorithm [106] adopted to B2BDC [107]. The majority, about 66% points of \mathcal{F}^{R1} fell within \mathcal{F}^{D} , and approximately 73% points of \mathcal{F}^{D} fell within \mathcal{F}^{R1} , but only about 15% points of \mathcal{F}^{R5} fell within \mathcal{F}^{D} , and roughly 8% points of \mathcal{F}^{D} fell within \mathcal{F}^{R5} . The results on the relative volumes and the overlap between feasible sets show that the feasible sets \mathcal{F}^{D} and \mathcal{F}^{R1} have large differences. However, these results do not imply that R1 cannot reproduce with good accuracy the uncertainties in the model predictions of the detailed model, as will be discussed in Sections 5.3.4 and 5.3.5.

The dataset constructed with the detailed model has 10 active parameters. For each of the 10 active parameters, the interval formed by the minimum and maximum of that parameter among the sampled 1.5×10^4 points in the feasible set \mathcal{F}^{D} was computed, termed the sampled-based interval. The corresponding optimization-based interval for each active parameter was obtained by solving problem (2.8). Note that both QOIs 1 and 2 were used as constraints to obtain the feasible set \mathcal{F}^{D} . The sample- and optimization-based intervals for the 10 active parameters in the case of the detailed model are plotted in Fig. 5.1. The vertical red lines (the inner-bound intervals), the vertical blue lines (the outer-bound intervals), and the vertical black lines (the sample-based intervals) are all highly close to the $[-1, 1]$ interval for all 10 active parameters. The results imply that the sampling algorithm was able to collect points in all parts of the feasible set.

5.3.4 Comparison of posterior distributions

For each QOI, uniform samples were collected from the corresponding feasible sets. Specifically, 1.5×10^4 points were generated in each of the $\mathcal{F}_{-p}^{\text{D}}$, $\mathcal{F}_{-p}^{\text{R1}}$, and $\mathcal{F}_{-p}^{\text{R5}}$, for $p = 1$ or 2 , using the Gibbs sampling algorithm adopted to B2BDC [107]. The notation \mathcal{F}_{-p} denotes the feasible set for a dataset with p -th QOI removed as a result of the leave-one-out scheme, as discussed in Section 3.3. The points of a uniform sample were then used to evaluate the cor-

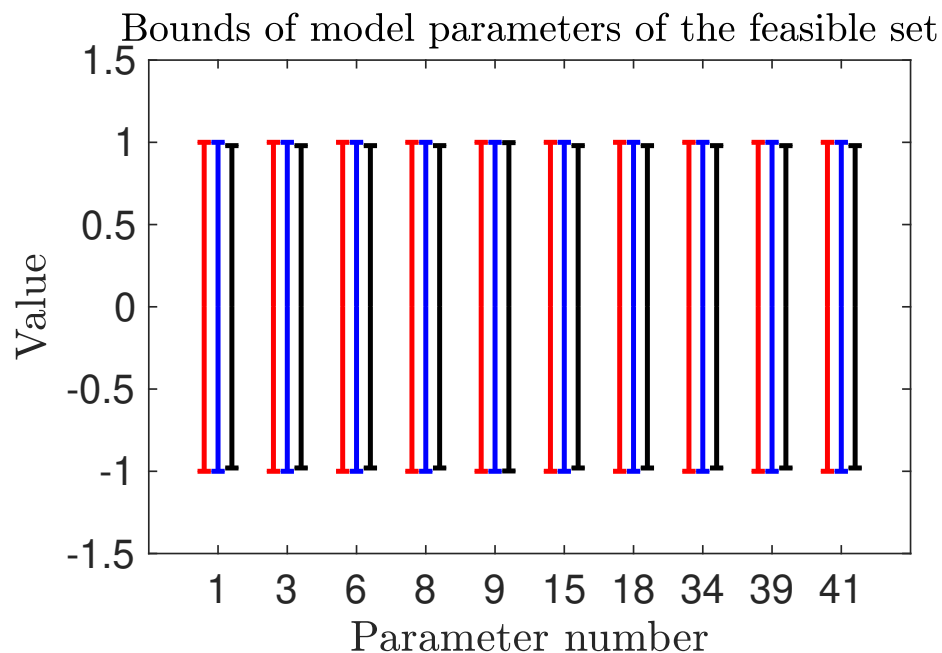


Figure 5.1: Bounds of active model parameters computed with the detailed model; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and vertical black lines are sample-based prediction intervals.

responding surrogate model, and a sample of model predictions was subsequently obtained. The histograms of predicted values computed for QOIs 1 and 2 are displayed in Fig. 5.2.

The reduced model R1 reproduced closely the posterior distributions of predictions of both QOIs computed with the detailed model, as shown in Fig 5.2. In contrast, the reduced model R5 led to posterior distributions of QOI predictions that are quite different from those of the detailed model. Whereas predictions of QOI 1 using R5 are higher, predictions of QOI 2 using R5 are lower. Recall that QOI 1 is the total number of CO molecules released in 2 ms and QOI 2 is the total number of C_2H_2 molecules released in 2 ms. Both CO and C_2H_2 are the results of soot oxidation, and they are produced from different competing mechanisms. The release of CO is due to oxyradical decomposition and oxidation by H_2O , OH, and O, namely reactions S51–S54, S90, S96–S99, and S100–S104 in Table 2.1, whereas the release of C_2H_2 is caused by thermal desorption, namely reactions S5, S7, S17, S25, and S45 in Table 2.1 [20]. Examination of the reactions in Table 5.2 shows that there is no difference in CO-related reactions for R1 and R5. Yet, more C_2H_2 -related reactions were removed for R5. Specifically, S17, S25, and S45 were removed for R5, whereas only S45 was removed for R1. Reactions S17 (1.7 reaction counts) and S25 (4.2 reaction counts) combined are responsible for the release of about 6 C_2H_2 molecules when the simulation was run with the detailed model, as shown in Table 5.2. However, removing these two reactions in R5 can cause the

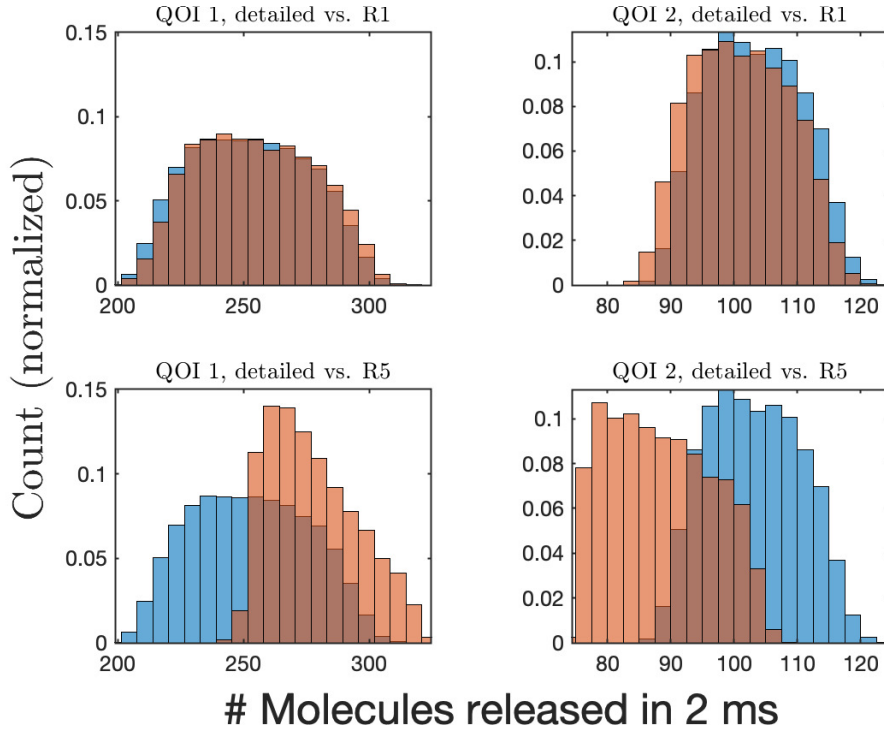


Figure 5.2: Histograms for QOIs 1 and 2 of detailed (blue) and reduced (brown) model predictions over feasible sets.

decrease in the number of C_2H_2 release to be significantly larger than 6, as shown in Fig 5.2.

Hellinger distances h^H , as computed by Eq. (3.7), for both QOIs with the two reduced models are reported in Table 5.8. The Hellinger distance metrics enable the quantification of the differences among the posterior distributions of model predictions, illustrated in Fig. 5.2.

Table 5.8: Hellinger distances, h^H and h^U , between posterior distributions of QOI predictions using the detailed (D) and reduced (R1, R5) models for QOIs 1 and 2.

	R1	R5
h^H (QOI 1)	0.0431	0.5086
h^U (QOI 1)	0.0467	0.3700
h^H (QOI 2)	0.1153	0.6489
h^U (QOI 2)	0.0841	0.3792

5.3.5 Comparison of prediction intervals

For both QOIs 1 and 2, prediction intervals computed with the detailed and two reduced models were obtained by solving the corresponding optimization problems, as described in Section 3.4. Averaged over the two QOIs, the relative length of the inner-bound prediction intervals with respect to the outer-bound prediction intervals is 89.6, 91.9, and 99.2% for the detailed, R1, and R5 models, respectively. This indicates that inner- and outer-bound prediction intervals are relatively close to each other for QOIs in this dataset, although not as close as in the example with the ignition delay QOIs in Section 4.2.5. The differences between the inner- and outer-bound prediction intervals are discussed in Section 2.1.4. The results of these computations for QOIs 1 and 2 are shown in Fig. 5.3.

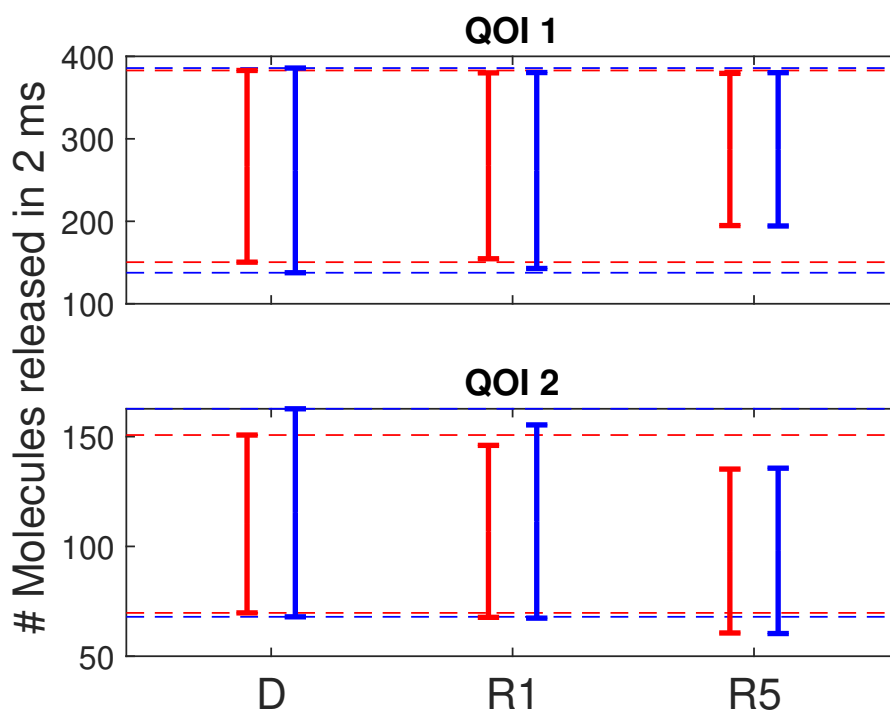


Figure 5.3: QOI prediction intervals computed with the detailed, D, and reduced, R1, and R5 models; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and horizontal dashed lines are prediction intervals of the detailed model.

As displayed in Fig. 5.3, for each case, three pairs of intervals are plotted, corresponding to the detailed and reduced models. For each pair, colored in red is the inner-bound prediction interval and colored in blue is the outer-bound one. The horizontal dashed lines mark the prediction intervals obtained with the detailed model. In both cases, the prediction intervals

of reduced model R1 are very close to those of the detailed model. In contrast, the prediction intervals of R5 are narrower than those using the detailed model for both QOIs. It occurs again that a smaller reduced model, R5 in this example, underestimates the uncertainties in model predictions. Hellinger distances, h^U 's, as computed by Eq. (3.10), for QOIs 1 and 2 with the two reduced models are reported in Table 5.8. Inspection of Fig. 5.3 and Table 5.8 indicates that the computed values of the numerical measure, $h_p^U(D, R)$, reflect the visual assessment of the intervals, illustrated in Fig. 5.8, and are consistent with the conclusions reached through the analysis of the posterior QOI distributions for the UQ performance of the reduced models.

Table 5.9: Relative lengths of sample-based prediction intervals to optimization-based prediction intervals for QOIs 1 and 2.

	detailed	R1	R5
Rel. length (QOI 1)	0.4702	0.5068	0.4615
Rel. length (QOI 2)	0.4425	0.4657	0.4624

As in Section 4.2.5, for each QOI and model combination, the interval formed by the minimum and maximum of the sample of model predictions, termed the sample-based prediction interval, was computed. The relative length of the sample-based prediction interval, defined as the ratio of the length of the sample-based prediction interval to the length of the optimization-based prediction interval, was subsequently computed. For each of the detailed model, R1, and R5, the relative lengths are tabulated in Table 5.9. The relative lengths vary from about 0.44 to around 0.50. This indicates that the posterior distributions of QOI predictions only cover roughly a half of all possible values of model predictions, determined by the prediction intervals. This shows again that uniform samples collected in the feasible sets miss certain regions in the high-dimensional feasible sets. In other words, sampling has a small chance of finding a point in the feasible set that maximizes/minimizes the surrogate model, which can be done very efficiently using optimization tools in B2BDC, as introduced in Section 2.1.4.

5.3.6 Comparison of sensitivities

Sensitivities of prediction interval with respect to the uncertainty intervals of model parameters were computed, as described in Section 3.5. Because the two QOIs in this dataset were derived from the same simulation, sensitivities of prediction interval with respect to the QOI bounds of other QOIs are expected to provide little information and are, thus, omitted. The sensitivity coefficients with respect to model parameters computed for both QOIs 1 and 2 are shown in Figs. 5.4 and 5.5, respectively.

Inspection of the results for both QOIs 1 and 2 shows that the sensitivity spectra are quite different even among the detailed model and R1. As shown in Fig. 5.4, the sensitivities of the

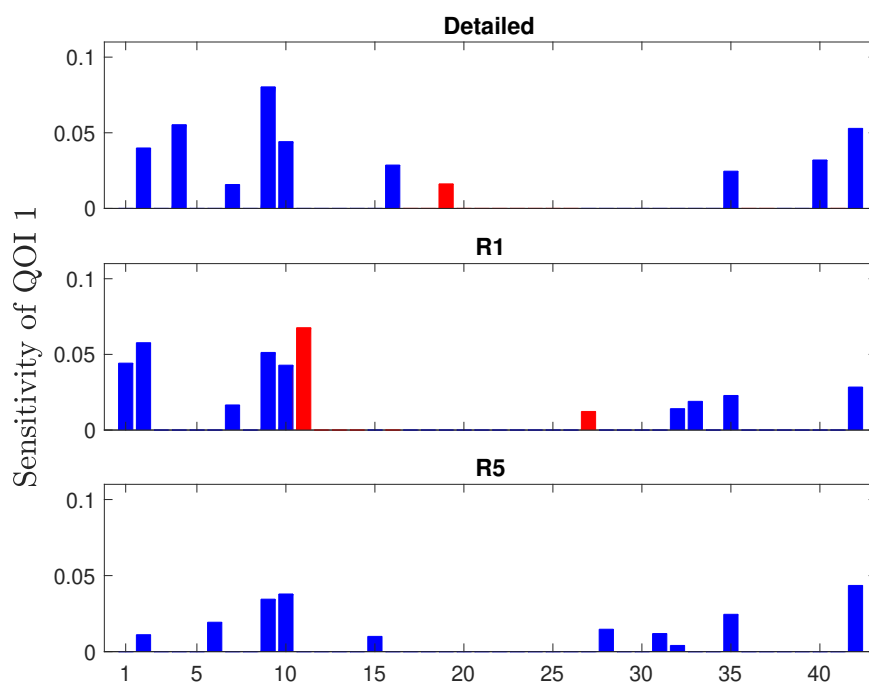


Figure 5.4: Sensitivity of prediction interval for QOI 1 with respect to the uncertainty intervals of model parameters. Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.

uncertainty in the prediction interval, $\overline{M}_p - \underline{M}_p$, with respect to parameters 4, 16, 19, and 40, corresponding to the multipliers of groups 4, 16, 19, and 40, have relatively high sensitivities in the case of the detailed model. Yet, the sensitivities with respect to these parameters are close to zero in the case of R1. However, the width of the prediction interval of QOI 1 using R1 is very close to that using the detailed model, as shown in Fig. 5.3. This can be explained by the change in sensitivities with respect to other parameters. Specifically, the sensitivities with respect to parameters 1, 11, 27, 32, and 33 are significant in the case of R1, whereas they are small in the case of the detailed model. This change in the sensitivity spectra implies a compensatory effect caused by the removal of some reactions. All sensitivities in the case of R5 are smaller, consistent with the narrower prediction interval of QOI 1 in Fig. 5.3. Similar patterns can be observed for QOI 2 in Fig. 5.5.

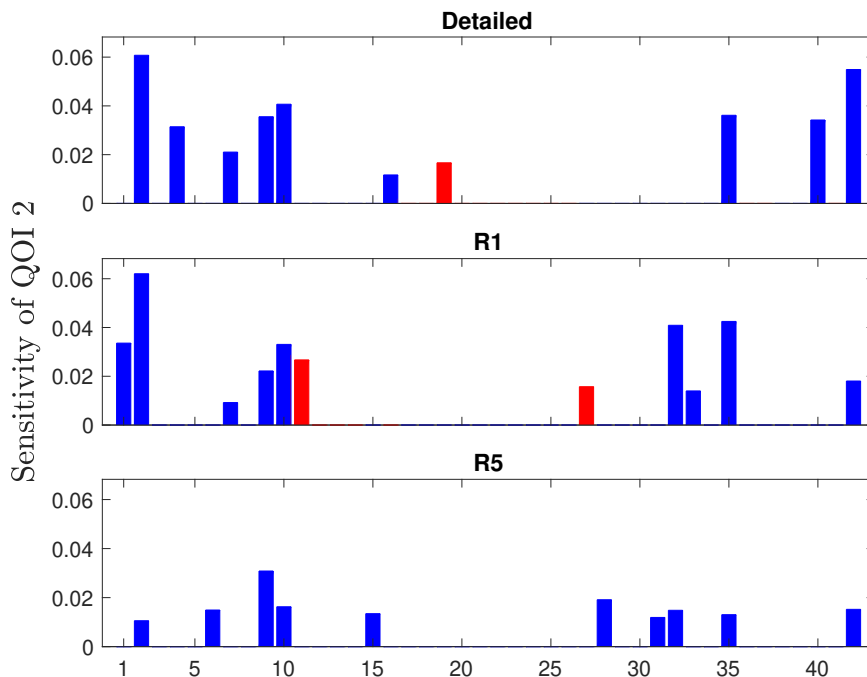


Figure 5.5: Sensitivity of prediction interval for QOI 2 with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.

5.4 Summary

In this chapter, an example with soot oxidation QOIs was investigated to assess how uncertainties in the model parameters and in the model predictions behave for reduced models as compared to those for the detailed model, a stochastic soot oxidation model. Large-scale parallel computing with MPI4PY was leveraged to expedite the computer simulations by multiple times.

Reduced models were developed from the detailed model, a stochastic KMC soot oxidation model, by identifying non-contributing reactions, adapted from the “detailed-reduction” method introduced in Section 2.2. Different reduction criteria resulted in three reduced models, R1, R5, and R10, in descending order of model size. The dataset built for the reduced model R10 was inconsistent. The reduced model R5 exhibits large deviations in all measures. The R1 reduced model reproduces accurately the prediction intervals (relatively low h^U 's) and the posterior distributions of model predictions (relatively low h^H 's) computed with the detailed model. Yet, predicted values computed with R1 have relatively large deviations from those with the detailed model under uncertainties (relatively large δ^H 's and δ^F 's). The

feasible set associated with R1 has a moderate overlap (around two thirds) with the feasible set associated with the detailed model, although their estimated relative volumes are quite close to each other. The spectra of sensitivities of uncertainty in model predictions are quite different for the detailed and R1 models. The results obtained for the different measures indicate that the shifted feasible set associated with R1 can compensate, to some extent, for the relatively large deviations between predicted values computed with the detailed and reduced models. This compensation can result in the mitigation of deviation measures in prediction intervals and posterior distributions of model predictions. Because the reproduction of model predictions and their uncertainties is what ultimately matters in many problems, the reduced model R1 may be accepted as sufficient for these purposes.

Chapter 6

Summary and Conclusions

In this dissertation, several measures were developed to assess how uncertainties in the model parameters and in the model predictions behave for reduced models as compared to those for the detailed model. First, deviations between the predicted values of the detailed and reduced models when the model parameters are fixed at the nominal values (δ), constrained to the prior region ($\delta^{\mathcal{H}}$), or constrained to the posterior region ($\delta^{\mathcal{F}}$), were established. The latter two can be obtained in a numerically efficient way using the B2BDC methodology. Second, two measures to quantify the difference between feasible sets of the detailed and reduced models were developed. These two measures are the relative volumes of the feasible sets and the overlap between feasible sets computed with the detailed and reduced models. The relative volumes of the feasible sets were estimated with a rejection sampling algorithm, while the overlap between feasible sets was estimated by uniform samples collected in the feasible sets using a Gibbs sampling algorithm adopted for B2BDC. Third, the Hellinger distance measure was adopted to quantify the “distance” between posterior distributions of model predictions computed with the detailed and reduced models. Fourth, another measure, also a Hellinger distance measure, was constructed to quantify the “distance” between prediction intervals of the detailed and reduced models by assuming uniform distributions on the prediction intervals. The prediction intervals represent the uncertainties in the model predictions and they can be computed in a numerically efficient way with the B2BDC methodology. Finally, sensitivities of the uncertainties in model predictions were established to measure the difference in influence of the model parameters on the model predictions between the detailed and reduced models. The sensitivity measures are the by-products of the B2BDC computations of the prediction intervals and do not require any extra computational efforts.

In the example with computer-generated ignition delay QOIs in Section 4.2, three reduced models, R25, R17, and R12 were tested. The results of this example provide several implications. First, assessment of the quality of a reduced model without considering parameter uncertainty may be misleading in that the deviation can be much larger when the uncertainties in the model parameters are taken into account, as is evidenced by the δ , $\delta^{\mathcal{H}}$, and $\delta^{\mathcal{F}}$ results of R17. Second, the reduced-model performance can be quantified by several measures of model deviations. Both Hellinger distance measures, based on posterior

distributions of model predictions (h^H 's) and based on prediction intervals (h^U 's), enable the assessment of how uncertainties in the model predictions behave for reduced models as compared to the detailed model. Comparison of the feasible sets, specifically the difference in their relative volumes and in the overlap between the feasible sets, provide means to assess how the posterior uncertainties in the model parameters behave for reduced models. The measures tested in this example showed mutually consistent and qualitatively similar outcomes.

In the example with experimental ignition delay QOIs in Section 4.3, a limitation of measures based on sampling was observed. The experimental data (L_e 's and U_e 's) used in this example were shown in previous studies [112, 113] to have disagreement with the model. As a result of the deficiency in the experimental data, the feasible set in this example has a very complex shape such that the sampling algorithm missed many parts of the feasible set. The results indicate that potentially bad data can lead to a substantial challenge for the sampling algorithm. If no verifiably accurate experimental data exist, computer-generated data from the solution of the detailed model offer a reliable alternative, in which the desired level of reduced-model accuracy can be prescribed by specifying the accepted ranges of variations in prediction of training targets. The results also illustrate that the measures based on optimization using the B2BDC methodology (δ^H 's, δ^F 's, and h^U 's) are free from sampling-related issues caused by the complex shape of the high-dimensional feasible sets. Generally, measures based on direct sampling could become limited by the computational cost when sampling uniformly in the feasible set is expensive. In such situations, the B2BDC framework offers a more practical alternative by quantifying the propagated uncertainty through numerically efficient computations of uncertainty intervals and their overlap, all with the added benefit of obtaining the uncertainty sensitivities.

In the soot oxidation example in Section 5.3, three reduced models, R1, R5, and R10 were tested. The reduction of a stochastic KMC soot oxidation model was accomplished by adapting the “detailed-reduction” method, which is otherwise difficult to achieve with other reduction methods introduced in Section 1.2.1. The R1 reduced model reproduced accurately the prediction intervals (relatively low h^U 's) and the posterior distributions of model predictions (relatively low h^H 's) computed with the detailed model. The accurate reproduction of model predictions and their uncertainties by the reduced model R1 implies that R1 is an acceptable reduced model. Yet, predicted values of model predictions computed with R1 have relatively large deviations from those with the detailed model under uncertainties (relatively large δ^H 's and δ^F 's). In addition, the feasible set associated with R1 has a moderate overlap (around two thirds) with the feasible set for the detailed model, although their estimated relative volumes are quite close to each other. The results obtained for different measures indicate that the shifted feasible set associated with R1 can compensate for the relatively large deviations in predicted values, and this compensation can result in the mitigation of deviation measures based on prediction intervals and posterior distributions of model predictions.

Appendix A

Reduced models of syngas combustion

Table A.1: Deviations between predicted values computed by the detailed and reduced models with the nominal parameter values, δ_e , and over the prior region, $\delta_e^{\mathcal{H}}$

e	δ_e			$\delta_e^{\mathcal{H}}$ (inner)			$\delta_e^{\mathcal{H}}$ (outer)		
	R25	R17	R12	R25	R17	R12	R25	R17	R12
1	0.000	0.002	0.054	0.002	0.006	0.107	0.002	0.006	0.107
2	0.002	0.012	0.245	0.006	0.043	0.567	0.006	0.043	0.567
3	0.002	0.013	0.137	0.016	0.099	0.245	0.016	0.099	0.246
4	0.001	0.004	0.173	0.002	0.009	0.354	0.002	0.009	0.354
5	0.000	0.004	0.016	0.008	0.019	0.039	0.008	0.019	0.039
6	0.001	0.004	0.133	0.016	0.069	0.233	0.016	0.069	0.233
7	0.002	0.002	0.079	0.018	0.071	0.229	0.020	0.071	0.229
8	0.001	0.002	0.030	0.010	0.023	0.065	0.010	0.024	0.065
9	0.003	0.012	0.496	0.010	0.040	2.279	0.011	0.040	2.279
10	0.003	0.011	1.074	0.011	0.027	4.360	0.011	0.028	4.360
11	0.005	0.054	0.020	0.020	0.126	0.114	0.021	0.126	0.116
12	0.002	0.003	0.142	0.004	0.016	0.275	0.004	0.016	0.275
13	0.002	0.006	0.164	0.004	0.018	0.338	0.004	0.018	0.338
14	0.001	0.010	0.037	0.007	0.021	0.061	0.007	0.021	0.062
15	0.001	0.002	0.058	0.005	0.011	0.115	0.005	0.011	0.115
16	0.001	0.049	0.046	0.020	0.114	0.101	0.022	0.114	0.104
17	0.002	0.010	0.013	0.017	0.034	0.046	0.017	0.034	0.047
18	0.002	0.037	0.402	0.008	0.164	1.082	0.008	0.164	1.082
19	0.001	0.004	0.030	0.005	0.013	0.075	0.005	0.013	0.075
20	0.001	0.007	0.304	0.004	0.021	0.892	0.004	0.021	0.892
21	0.000	0.001	0.014	0.006	0.015	0.037	0.006	0.015	0.037
22	0.001	0.002	0.161	0.023	0.063	0.310	0.026	0.064	0.310

23	0.002	0.001	0.047	0.007	0.011	0.111	0.007	0.011	0.111
24	0.001	0.029	0.365	0.006	0.118	1.002	0.007	0.118	1.002
25	0.001	0.003	0.050	0.002	0.007	0.108	0.002	0.007	0.108
26	0.001	0.003	0.114	0.003	0.010	0.222	0.004	0.010	0.222
27	0.002	0.003	0.028	0.010	0.014	0.062	0.010	0.014	0.062
28	0.001	0.034	0.039	0.019	0.069	0.071	0.020	0.069	0.071
29	0.001	0.005	0.031	0.007	0.017	0.065	0.007	0.017	0.066
30	0.000	0.022	0.001	0.022	0.074	0.090	0.023	0.075	0.090
31	0.000	0.000	0.050	0.004	0.013	0.092	0.004	0.013	0.092
32	0.004	0.052	0.036	0.027	0.144	0.154	0.031	0.144	0.158
33	0.001	0.005	0.077	0.010	0.054	0.157	0.010	0.054	0.157
34	0.004	0.016	0.130	0.036	0.248	0.446	0.039	0.249	0.446
35	0.005	0.044	0.004	0.028	0.134	0.149	0.030	0.135	0.149
36	0.001	0.009	0.063	0.006	0.020	0.130	0.007	0.020	0.130
37	0.000	0.004	0.070	0.007	0.012	0.138	0.007	0.012	0.138
38	0.003	0.003	0.015	0.018	0.033	0.094	0.020	0.035	0.094
39	0.002	0.013	0.288	0.005	0.029	0.771	0.005	0.030	0.771
40	0.003	0.016	0.257	0.027	0.170	0.560	0.027	0.171	0.560
41	0.001	0.035	0.055	0.017	0.063	0.082	0.017	0.065	0.083
42	0.002	0.013	0.016	0.016	0.041	0.037	0.016	0.041	0.038
43	0.001	0.003	0.053	0.002	0.006	0.108	0.002	0.006	0.108
44	0.001	0.018	0.131	0.020	0.083	0.280	0.021	0.084	0.280
45	0.002	0.003	0.103	0.015	0.050	0.207	0.015	0.050	0.207
46	0.004	0.004	0.043	0.010	0.030	0.107	0.010	0.030	0.107
47	0.004	0.009	0.012	0.012	0.038	0.084	0.013	0.038	0.084
48	0.000	0.003	0.010	0.011	0.032	0.052	0.011	0.033	0.052
49	0.001	0.002	0.033	0.006	0.012	0.082	0.006	0.013	0.082
50	0.002	0.001	0.052	0.006	0.010	0.109	0.007	0.011	0.109
51	0.002	0.002	0.068	0.013	0.036	0.144	0.013	0.037	0.145
52	0.005	0.005	0.013	0.028	0.028	0.041	0.028	0.029	0.043
53	0.001	0.001	0.062	0.004	0.014	0.114	0.004	0.014	0.114
54	0.003	0.047	0.035	0.016	0.082	0.081	0.017	0.082	0.082
55	0.001	0.019	0.055	0.032	0.230	0.310	0.036	0.230	0.310
56	0.001	0.005	0.008	0.009	0.028	0.052	0.009	0.028	0.052
57	0.002	0.006	0.282	0.004	0.016	0.665	0.004	0.016	0.665
58	0.000	0.001	0.009	0.009	0.034	0.072	0.009	0.034	0.073
59	0.001	0.006	0.022	0.010	0.022	0.063	0.010	0.023	0.063
60	0.001	0.009	0.121	0.012	0.063	0.176	0.013	0.063	0.176
61	0.001	0.012	0.087	0.012	0.034	0.185	0.013	0.034	0.185
62	0.001	0.000	0.081	0.012	0.052	0.210	0.013	0.052	0.210
63	0.001	0.001	0.013	0.012	0.037	0.095	0.012	0.038	0.095
64	0.002	0.001	0.045	0.004	0.010	0.090	0.005	0.010	0.090

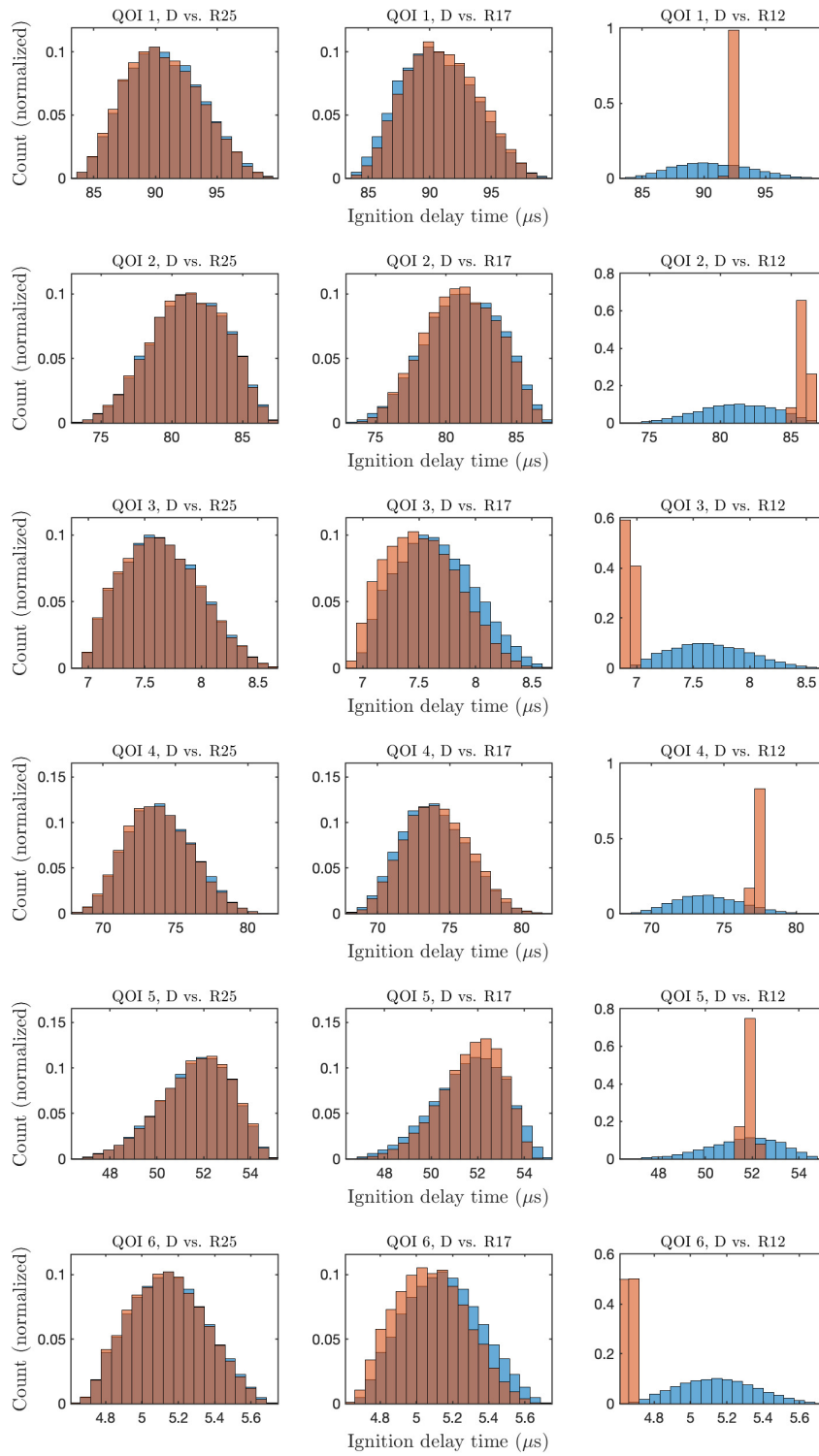
65	0.001	0.009	0.082	0.009	0.021	0.163	0.009	0.023	0.163
66	0.001	0.006	0.130	0.004	0.013	0.252	0.004	0.013	0.252
67	0.001	0.001	0.044	0.007	0.014	0.081	0.008	0.015	0.081
68	0.001	0.005	0.061	0.028	0.127	0.249	0.028	0.127	0.250
69	0.007	0.032	0.047	0.022	0.092	0.118	0.024	0.093	0.119
70	0.002	0.007	0.126	0.029	0.177	0.443	0.031	0.177	0.443
71	0.003	0.010	0.132	0.013	0.036	0.271	0.013	0.036	0.271
72	0.000	0.001	0.091	0.018	0.061	0.221	0.018	0.061	0.221
73	0.001	0.005	0.152	0.003	0.014	0.299	0.003	0.014	0.299
74	0.000	0.005	0.160	0.019	0.082	0.299	0.021	0.083	0.299
75	0.002	0.001	0.067	0.003	0.007	0.128	0.003	0.007	0.128
76	0.001	0.001	0.058	0.009	0.040	0.113	0.009	0.040	0.113
77	0.002	0.001	0.084	0.017	0.053	0.223	0.019	0.054	0.223
78	0.000	0.009	0.014	0.023	0.047	0.146	0.025	0.048	0.146
79	0.001	0.002	0.035	0.005	0.016	0.068	0.005	0.016	0.069
80	0.003	0.017	0.027	0.014	0.039	0.050	0.015	0.040	0.051
81	0.001	0.042	0.060	0.025	0.075	0.096	0.025	0.075	0.097
82	0.001	0.003	0.091	0.002	0.007	0.163	0.002	0.007	0.163
83	0.002	0.001	0.039	0.008	0.019	0.080	0.009	0.019	0.080
84	0.001	0.001	0.001	0.009	0.021	0.051	0.009	0.021	0.051
85	0.003	0.005	0.283	0.007	0.018	0.730	0.007	0.019	0.730
86	0.001	0.004	0.080	0.032	0.187	0.335	0.033	0.187	0.335
87	0.000	0.000	0.072	0.005	0.012	0.119	0.005	0.012	0.119
88	0.001	0.003	0.074	0.002	0.009	0.139	0.002	0.009	0.139
89	0.000	0.001	0.014	0.008	0.037	0.093	0.008	0.037	0.093
90	0.001	0.002	0.016	0.007	0.025	0.058	0.007	0.025	0.058
91	0.003	0.092	0.073	0.027	0.174	0.141	0.029	0.174	0.142
92	0.004	0.065	0.007	0.071	0.203	0.213	0.071	0.205	0.216
93	0.004	0.003	0.012	0.008	0.024	0.044	0.008	0.024	0.044
94	0.001	0.014	0.102	0.015	0.047	0.229	0.016	0.048	0.229
95	0.002	0.005	0.222	0.005	0.019	0.494	0.005	0.020	0.494
96	0.001	0.051	0.047	0.032	0.110	0.126	0.033	0.110	0.128
97	0.001	0.000	0.018	0.006	0.013	0.051	0.006	0.013	0.051
98	0.003	0.003	0.011	0.011	0.027	0.069	0.011	0.027	0.069
99	0.000	0.003	0.196	0.023	0.086	0.389	0.026	0.086	0.389
100	0.003	0.004	0.018	0.010	0.034	0.096	0.011	0.034	0.096

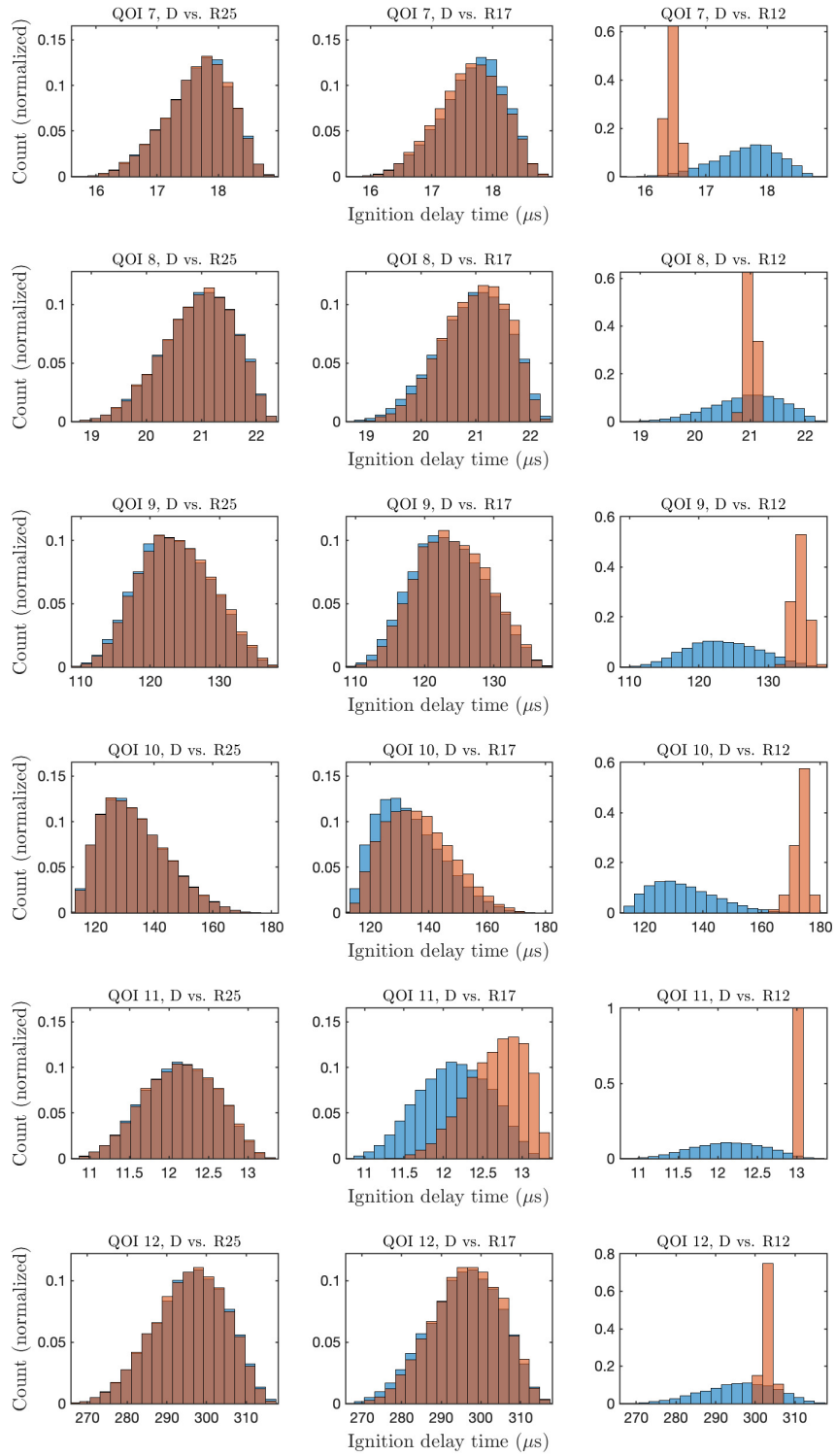
Table A.2: Deviations between predicted values computed by the detailed and reduced models over the posterior region, $\delta_p^{\mathcal{F}}$

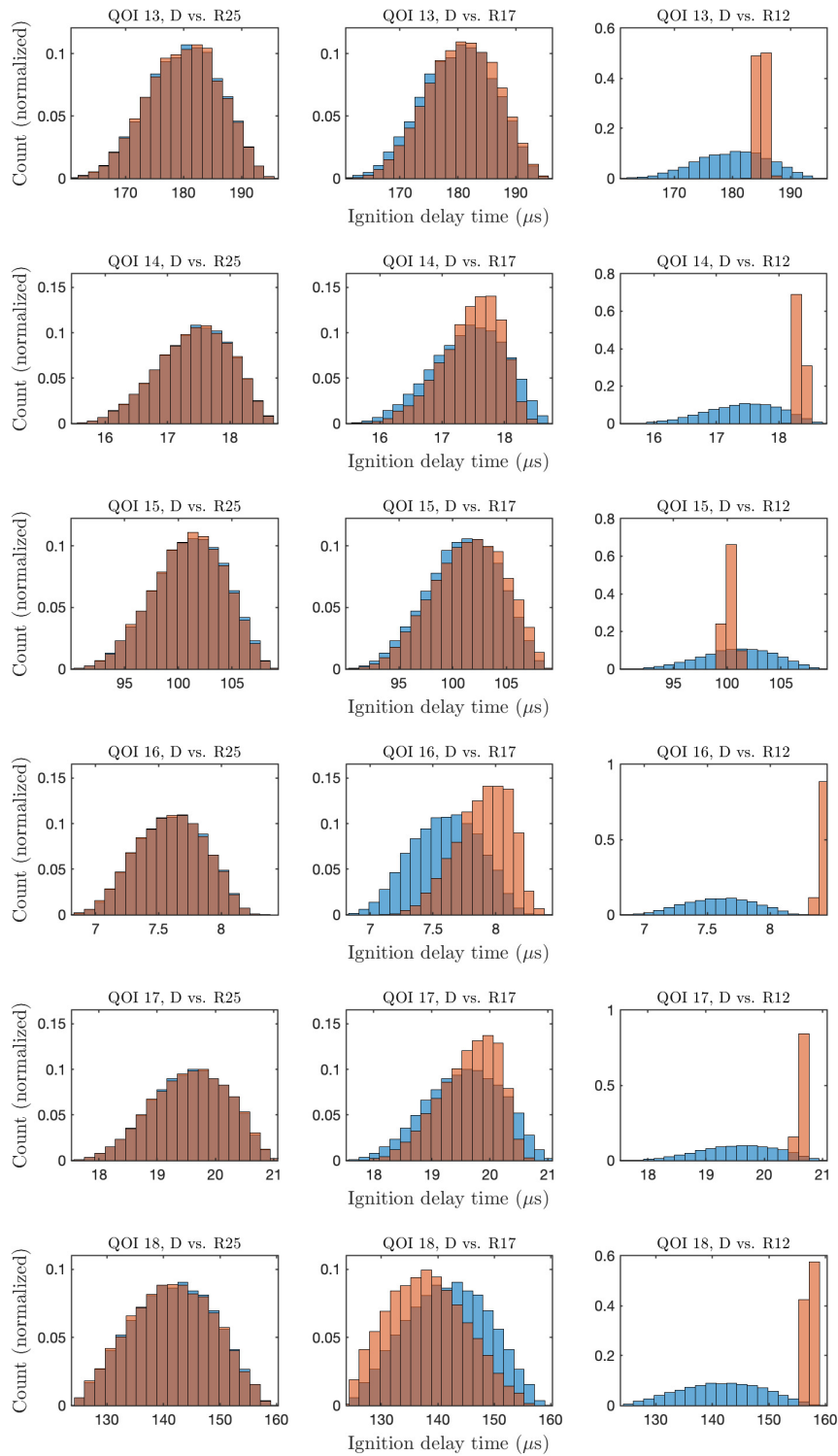
p	$\delta_p^{\mathcal{F}}$ (inner)			$\delta_p^{\mathcal{F}}$ (outer)		
	R25	R17	R12	R25	R17	R12
1	0.002	0.005	0.101	0.002	0.006	0.101
2	0.005	0.035	0.486	0.006	0.035	0.491
3	0.013	0.098	0.244	0.016	0.099	0.246
4	0.002	0.006	0.338	0.002	0.007	0.338
5	0.007	0.017	0.036	0.008	0.018	0.038
6	0.015	0.068	0.230	0.016	0.068	0.231
7	0.017	0.070	0.222	0.019	0.070	0.224
8	0.007	0.022	0.060	0.010	0.024	0.061
9	0.009	0.031	1.941	0.010	0.034	2.043
10	0.009	0.026	4.345	0.010	0.027	4.353
11	0.018	0.115	0.113	0.021	0.119	0.116
12	0.003	0.010	0.254	0.004	0.014	0.257
13	0.003	0.013	0.309	0.004	0.015	0.312
14	0.006	0.020	0.059	0.007	0.020	0.060
15	0.005	0.010	0.108	0.005	0.011	0.109
16	0.014	0.104	0.096	0.022	0.108	0.104
17	0.015	0.033	0.039	0.017	0.033	0.046
18	0.007	0.130	0.924	0.008	0.137	0.945
19	0.005	0.010	0.068	0.005	0.012	0.070
20	0.003	0.014	0.738	0.003	0.017	0.779
21	0.003	0.010	0.034	0.006	0.014	0.034
22	0.023	0.060	0.303	0.026	0.064	0.306
23	0.004	0.008	0.104	0.007	0.011	0.104
24	0.006	0.097	0.819	0.007	0.098	0.856
25	0.002	0.006	0.103	0.002	0.007	0.103
26	0.003	0.009	0.210	0.004	0.010	0.212
27	0.010	0.012	0.057	0.010	0.014	0.059
28	0.019	0.066	0.070	0.020	0.067	0.070
29	0.006	0.014	0.060	0.007	0.015	0.061
30	0.020	0.072	0.088	0.023	0.073	0.088
31	0.003	0.010	0.087	0.004	0.011	0.087
32	0.025	0.131	0.116	0.031	0.135	0.158
33	0.008	0.053	0.151	0.010	0.054	0.151
34	0.033	0.237	0.424	0.039	0.242	0.429
35	0.022	0.129	0.148	0.029	0.130	0.149

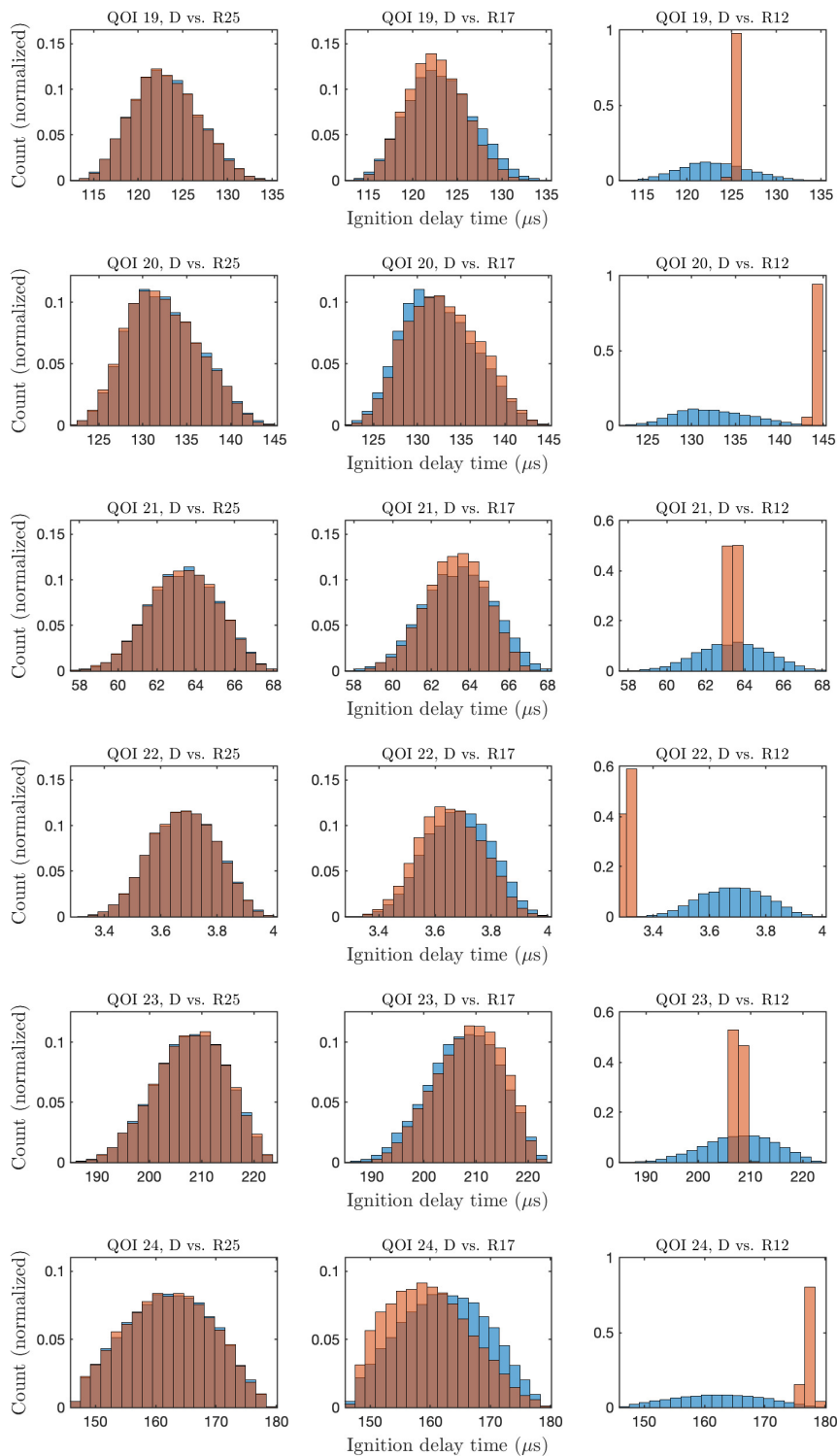
36	0.005	0.019	0.122	0.007	0.019	0.124
37	0.006	0.008	0.128	0.007	0.011	0.129
38	0.017	0.030	0.093	0.020	0.033	0.094
39	0.003	0.026	0.643	0.004	0.027	0.665
40	0.025	0.170	0.549	0.027	0.170	0.551
41	0.014	0.063	0.082	0.017	0.064	0.083
42	0.012	0.041	0.035	0.015	0.041	0.038
43	0.001	0.005	0.104	0.002	0.006	0.105
44	0.017	0.069	0.254	0.020	0.073	0.256
45	0.012	0.049	0.203	0.015	0.050	0.204
46	0.006	0.030	0.105	0.009	0.030	0.106
47	0.009	0.037	0.083	0.013	0.038	0.083
48	0.011	0.032	0.047	0.011	0.033	0.049
49	0.003	0.011	0.076	0.005	0.011	0.077
50	0.005	0.008	0.102	0.007	0.010	0.103
51	0.012	0.036	0.142	0.013	0.037	0.142
52	0.023	0.027	0.037	0.027	0.029	0.039
53	0.003	0.012	0.106	0.004	0.014	0.107
54	0.014	0.079	0.079	0.017	0.080	0.082
55	0.030	0.206	0.296	0.036	0.210	0.301
56	0.008	0.026	0.046	0.009	0.027	0.046
57	0.003	0.012	0.609	0.003	0.013	0.613
58	0.008	0.032	0.071	0.009	0.033	0.073
59	0.006	0.019	0.057	0.010	0.022	0.059
60	0.010	0.061	0.174	0.012	0.063	0.175
61	0.009	0.032	0.173	0.012	0.032	0.175
62	0.012	0.044	0.207	0.013	0.052	0.207
63	0.011	0.037	0.090	0.012	0.038	0.090
64	0.004	0.009	0.083	0.005	0.010	0.084
65	0.007	0.021	0.154	0.008	0.023	0.155
66	0.003	0.012	0.235	0.004	0.013	0.237
67	0.007	0.013	0.076	0.008	0.015	0.076
68	0.028	0.116	0.243	0.028	0.119	0.244
69	0.018	0.090	0.115	0.024	0.092	0.118
70	0.024	0.168	0.429	0.029	0.171	0.433
71	0.010	0.031	0.239	0.013	0.034	0.241
72	0.017	0.060	0.217	0.018	0.061	0.217
73	0.002	0.010	0.277	0.003	0.012	0.277
74	0.018	0.081	0.295	0.021	0.083	0.296
75	0.002	0.004	0.121	0.003	0.006	0.122
76	0.006	0.039	0.111	0.009	0.040	0.111
77	0.016	0.053	0.218	0.019	0.054	0.219

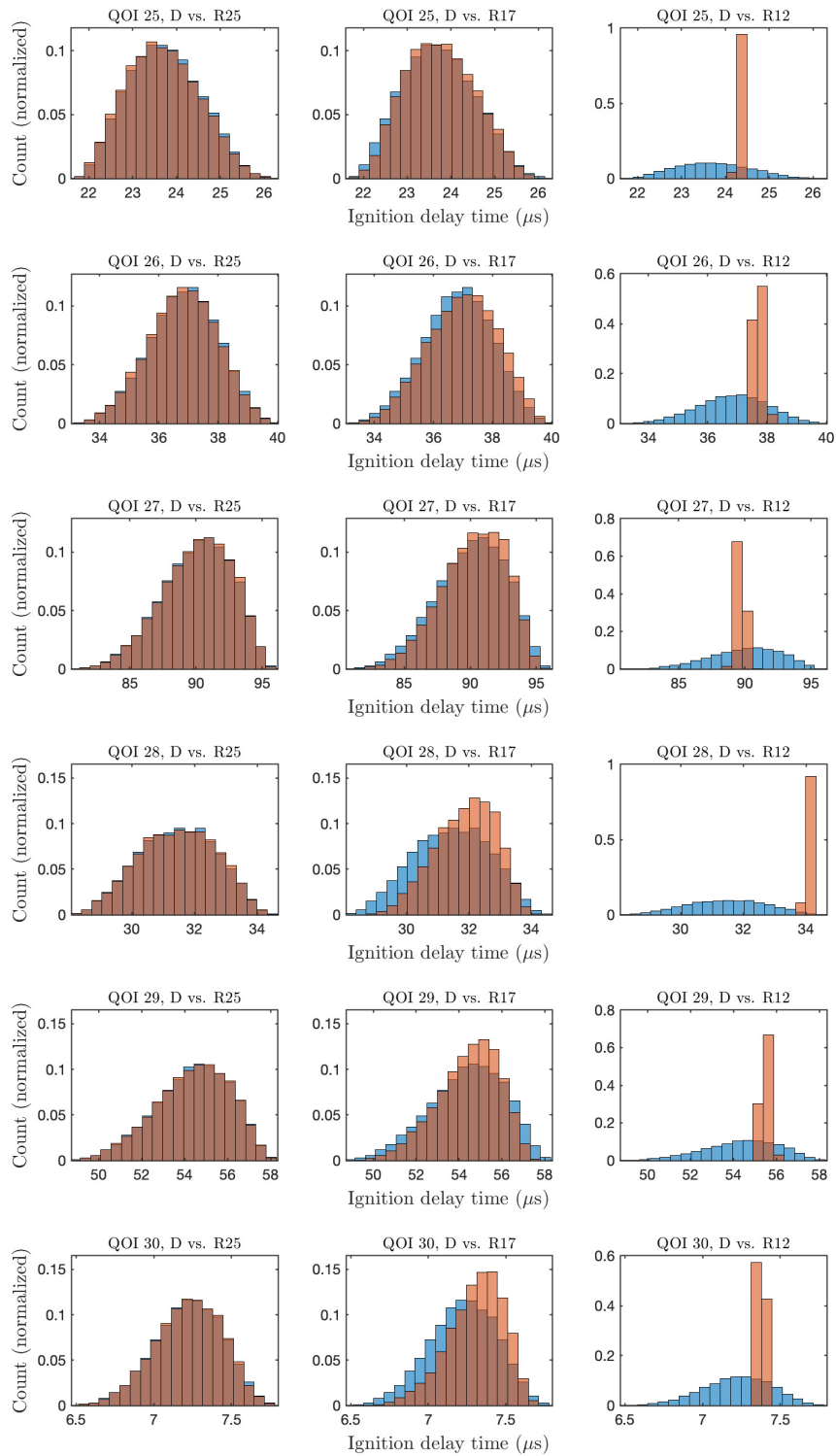
78	0.021	0.044	0.143	0.024	0.045	0.146
79	0.004	0.011	0.064	0.005	0.014	0.064
80	0.013	0.037	0.046	0.014	0.038	0.050
81	0.015	0.072	0.086	0.024	0.075	0.097
82	0.002	0.006	0.154	0.002	0.006	0.154
83	0.007	0.016	0.073	0.008	0.019	0.075
84	0.008	0.021	0.039	0.009	0.021	0.044
85	0.006	0.014	0.616	0.006	0.016	0.629
86	0.031	0.181	0.328	0.032	0.183	0.331
87	0.005	0.011	0.113	0.005	0.012	0.113
88	0.002	0.006	0.129	0.002	0.008	0.130
89	0.007	0.035	0.088	0.008	0.036	0.088
90	0.006	0.025	0.056	0.007	0.025	0.056
91	0.026	0.169	0.124	0.029	0.169	0.142
92	0.070	0.183	0.194	0.070	0.192	0.207
93	0.008	0.023	0.038	0.008	0.023	0.039
94	0.012	0.044	0.213	0.016	0.046	0.216
95	0.004	0.017	0.433	0.005	0.018	0.435
96	0.031	0.104	0.124	0.033	0.106	0.127
97	0.004	0.013	0.046	0.006	0.013	0.047
98	0.010	0.026	0.065	0.011	0.027	0.065
99	0.021	0.086	0.385	0.026	0.086	0.386
100	0.009	0.034	0.091	0.010	0.034	0.092

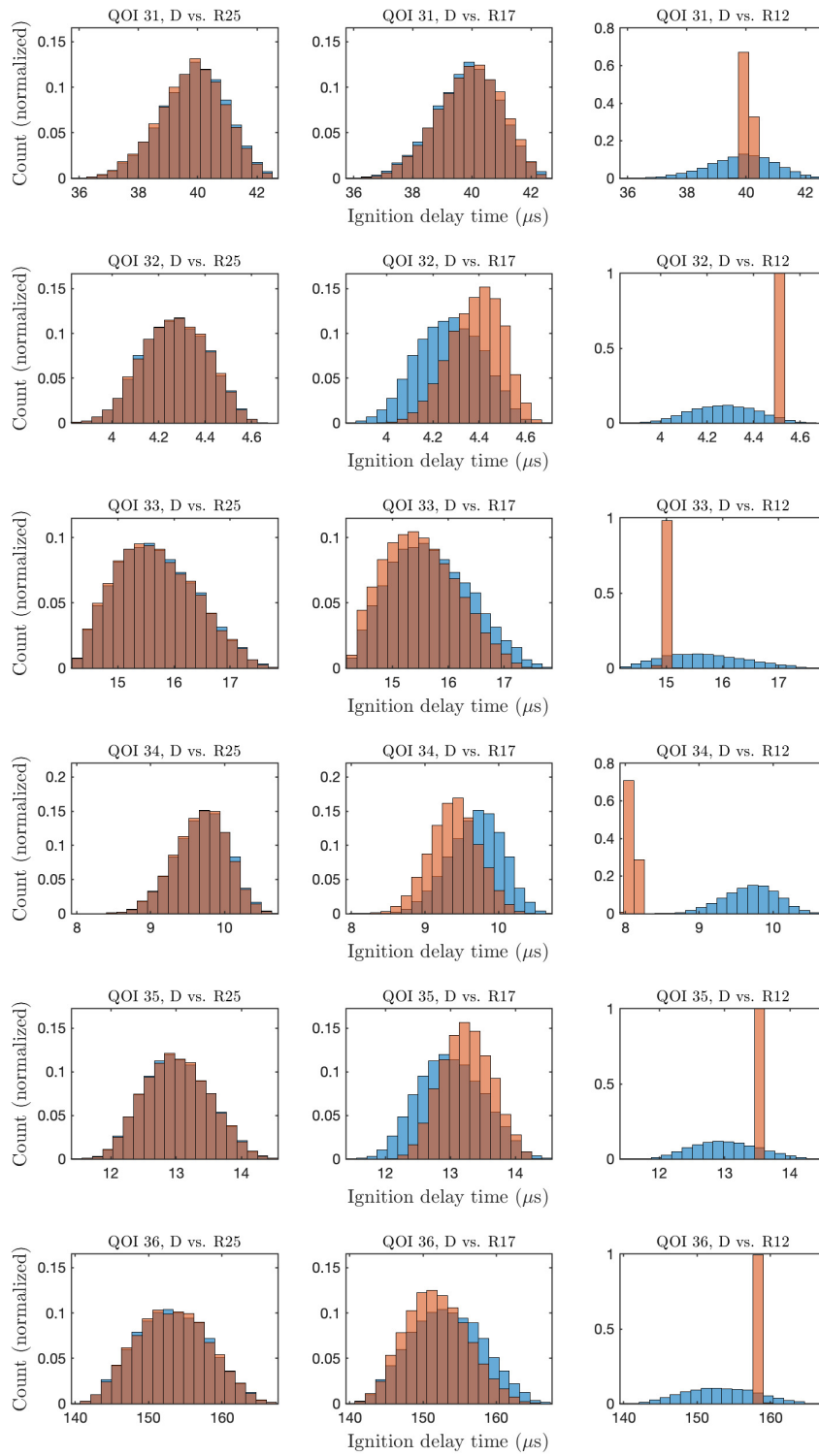


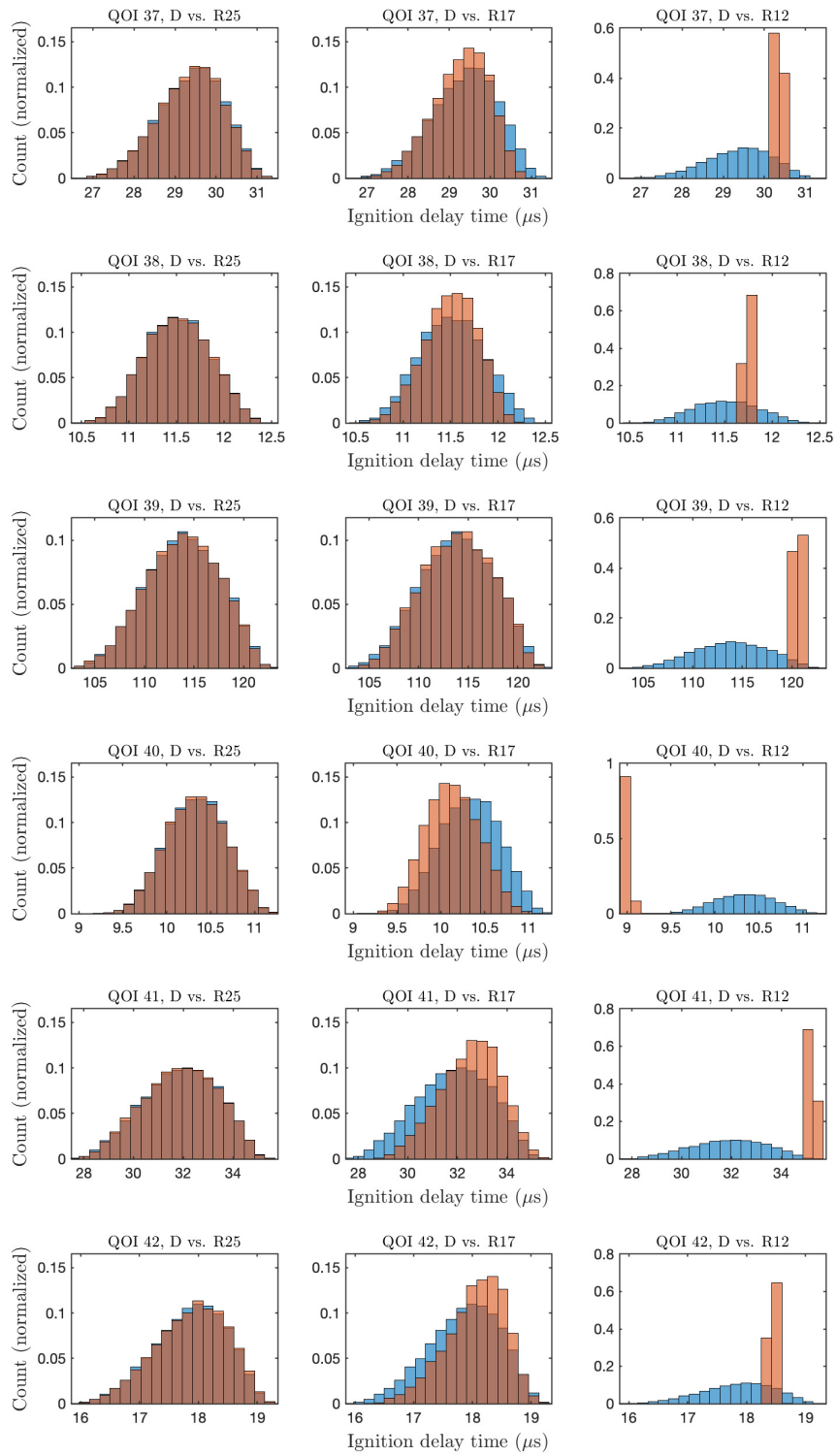


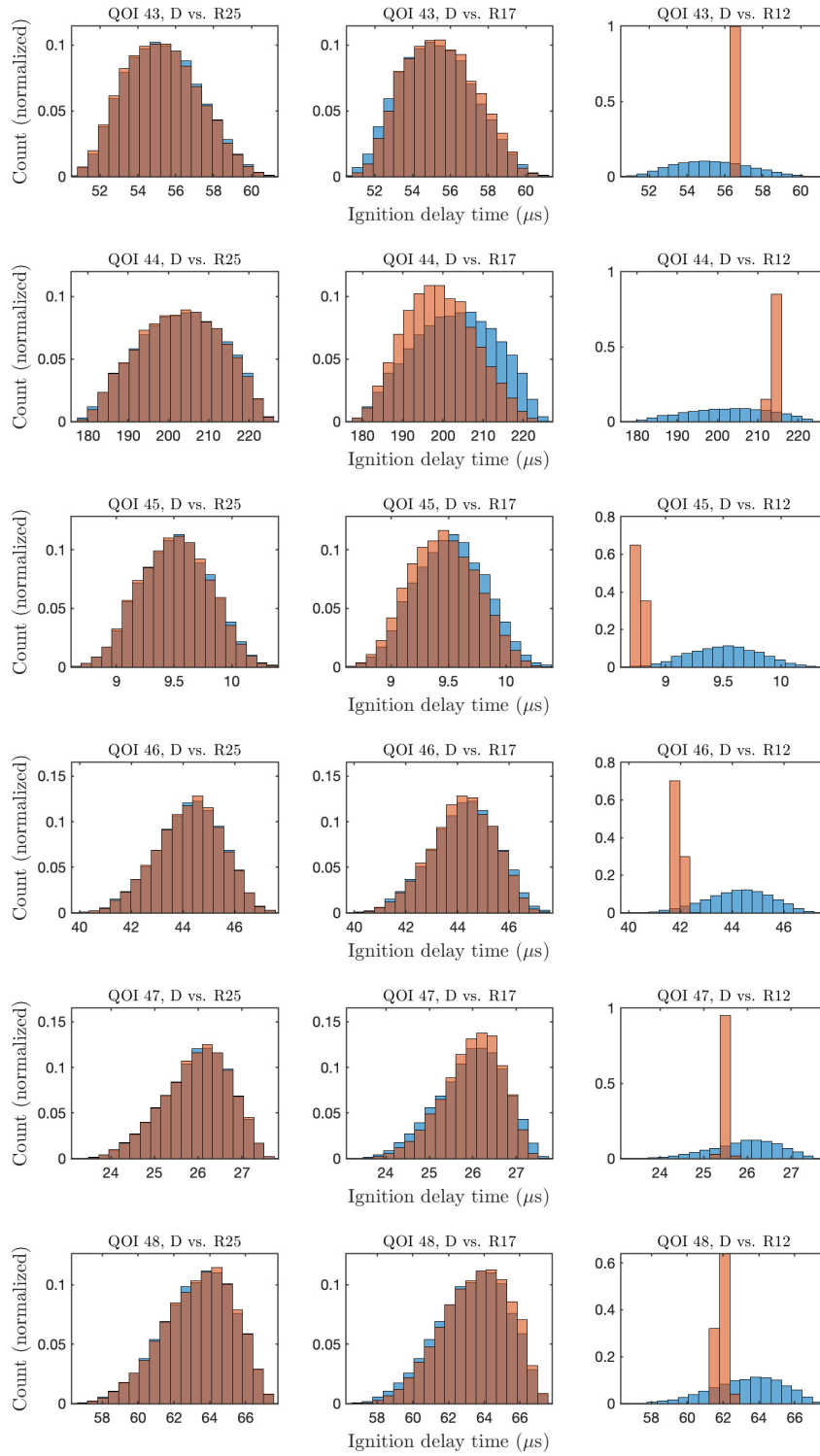


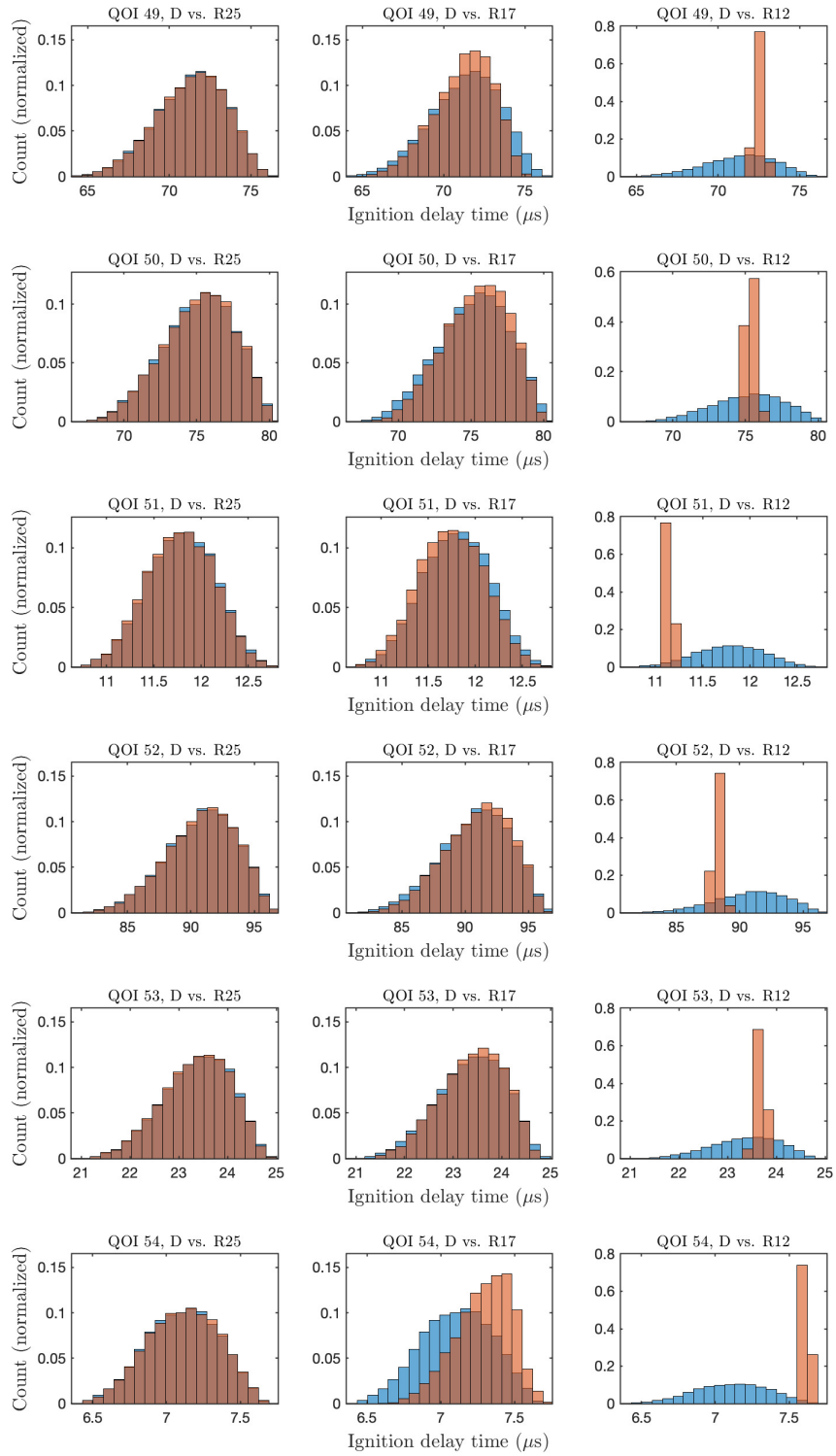


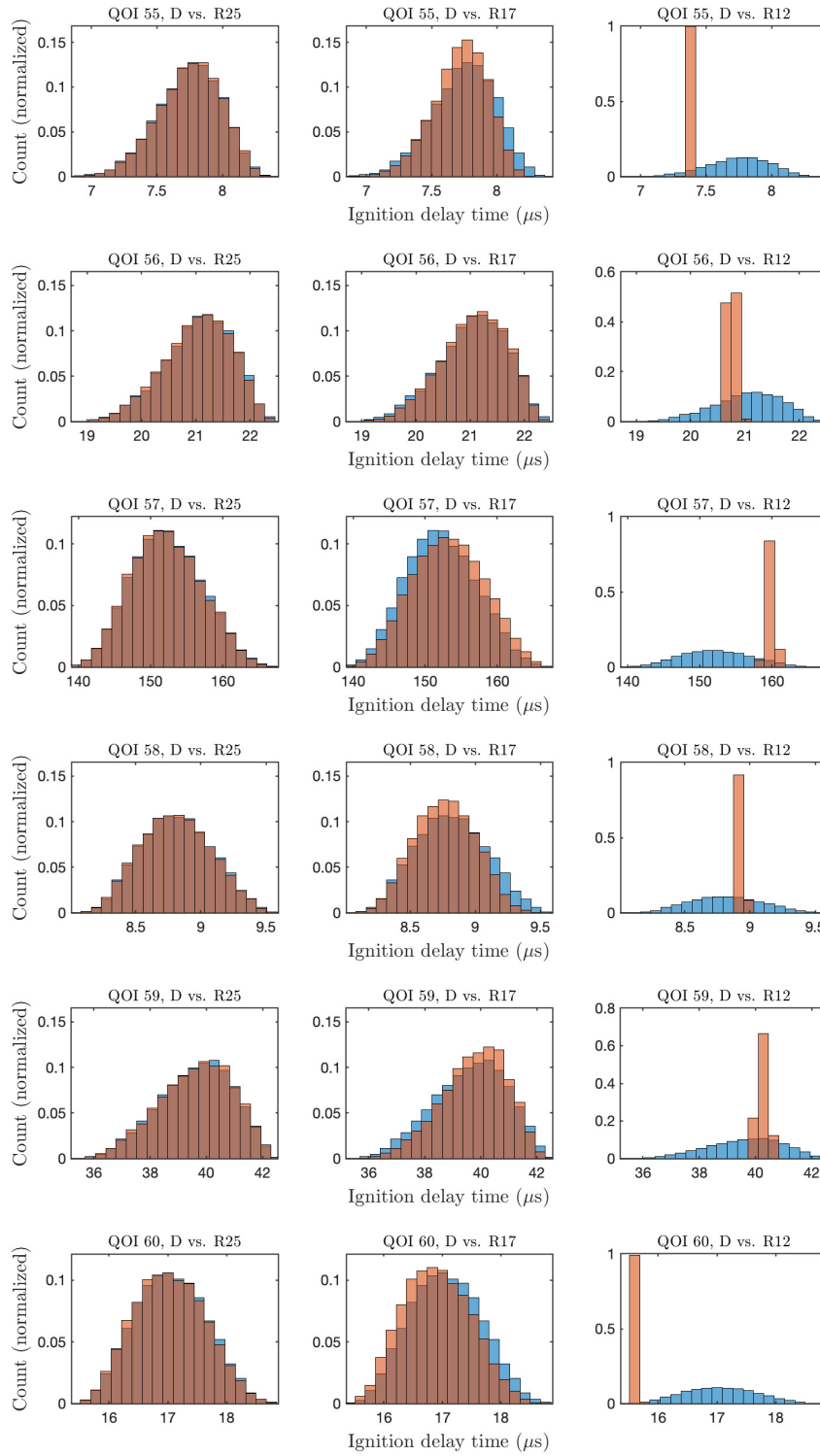


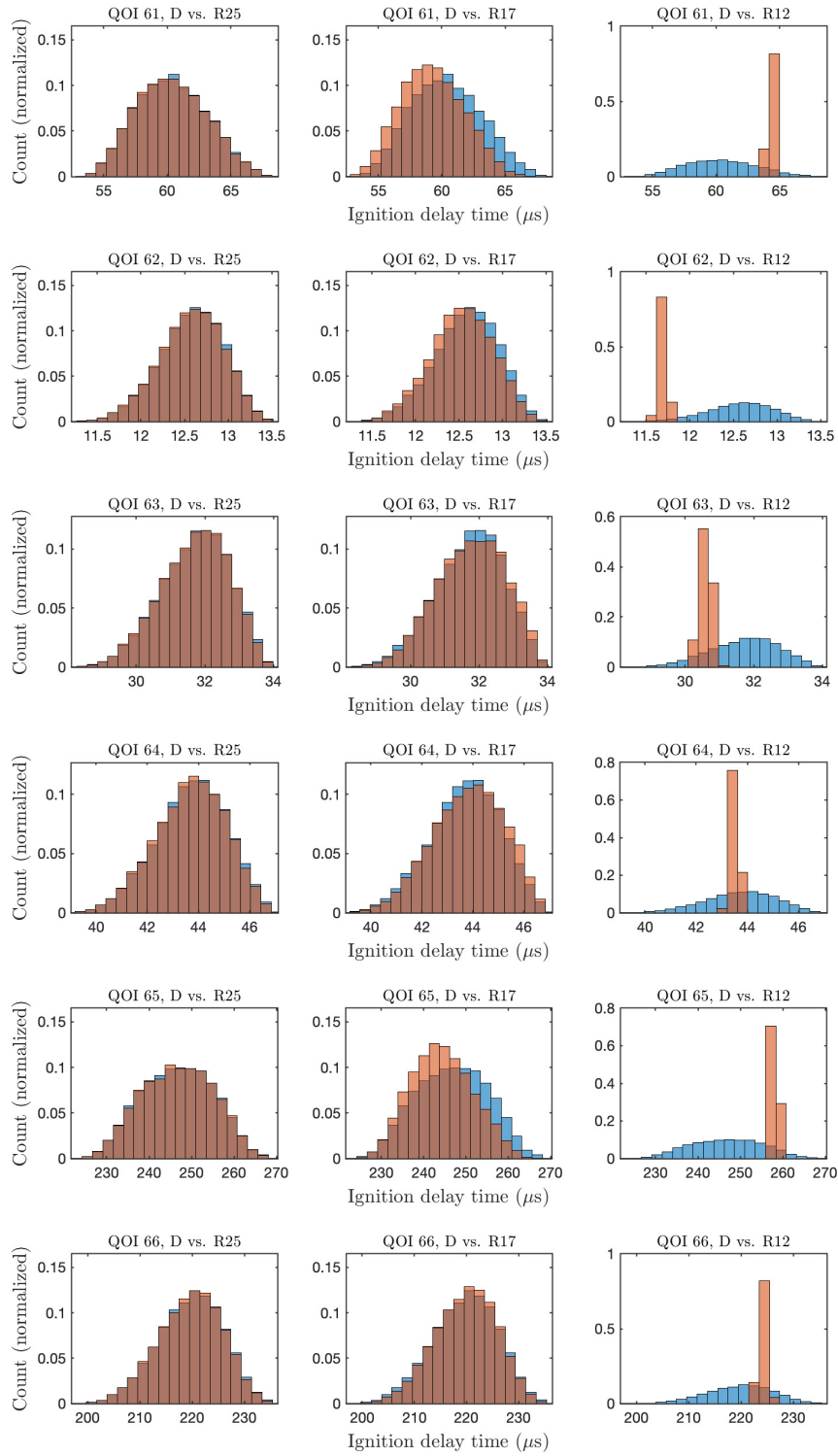


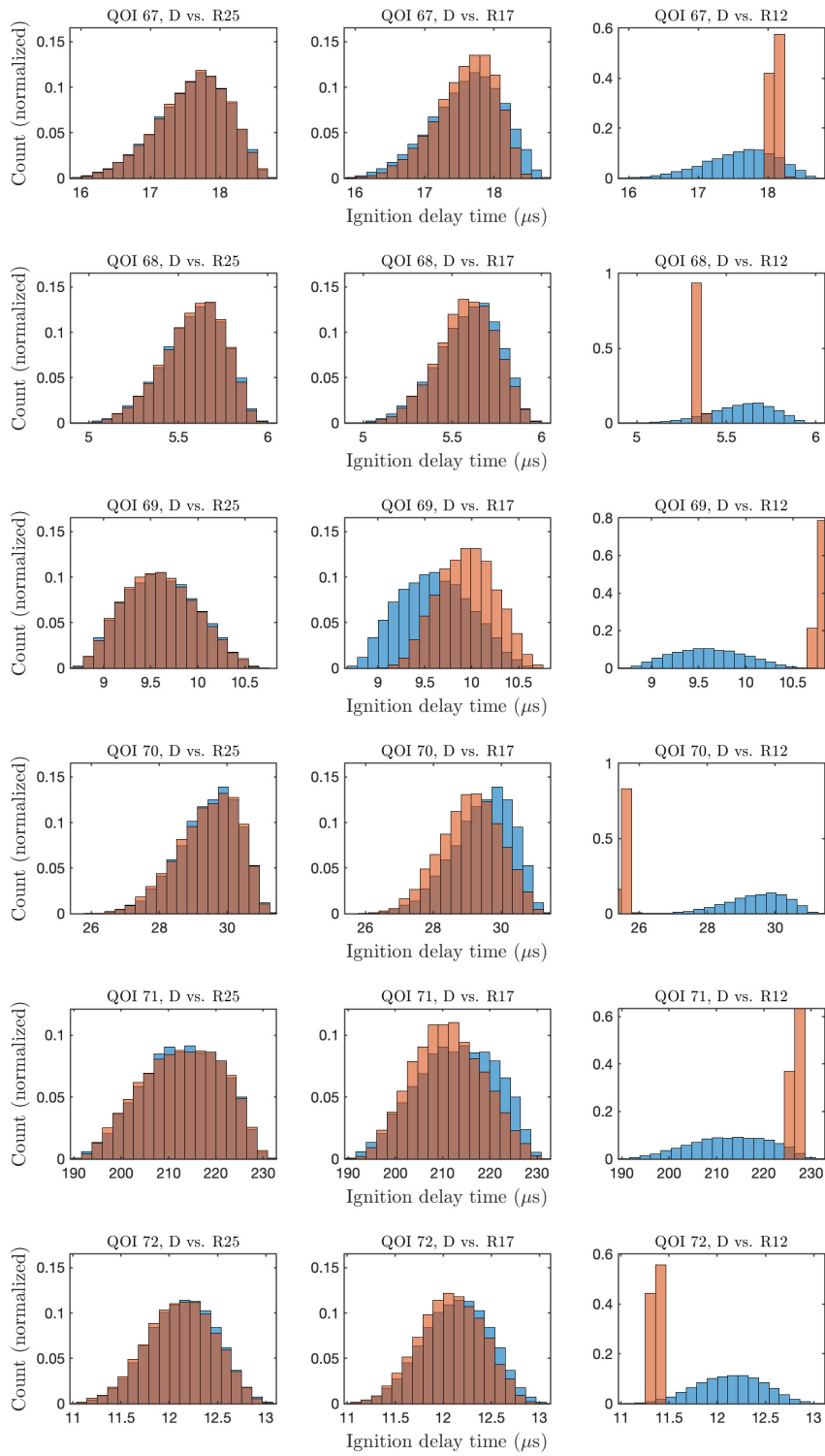


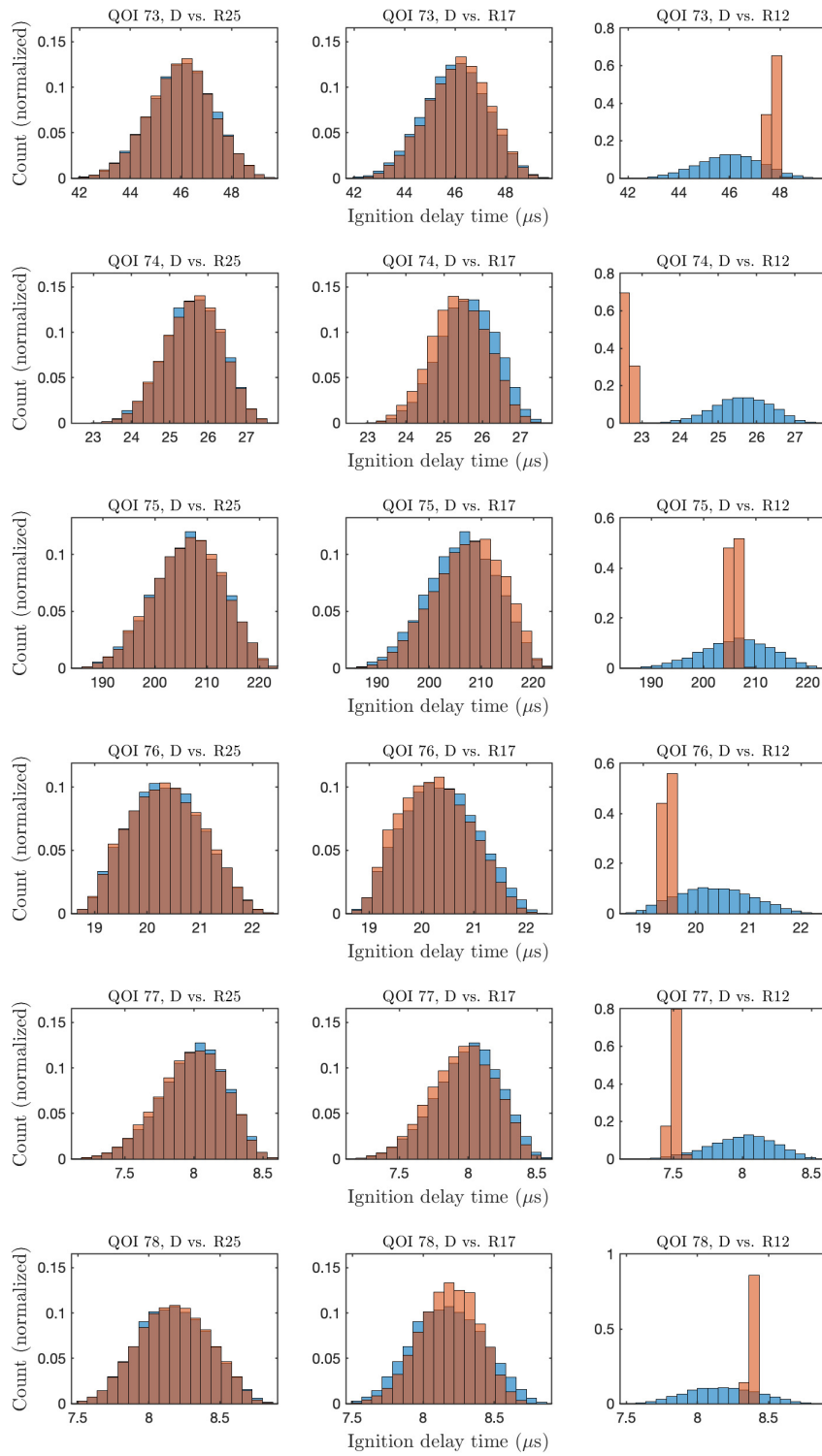


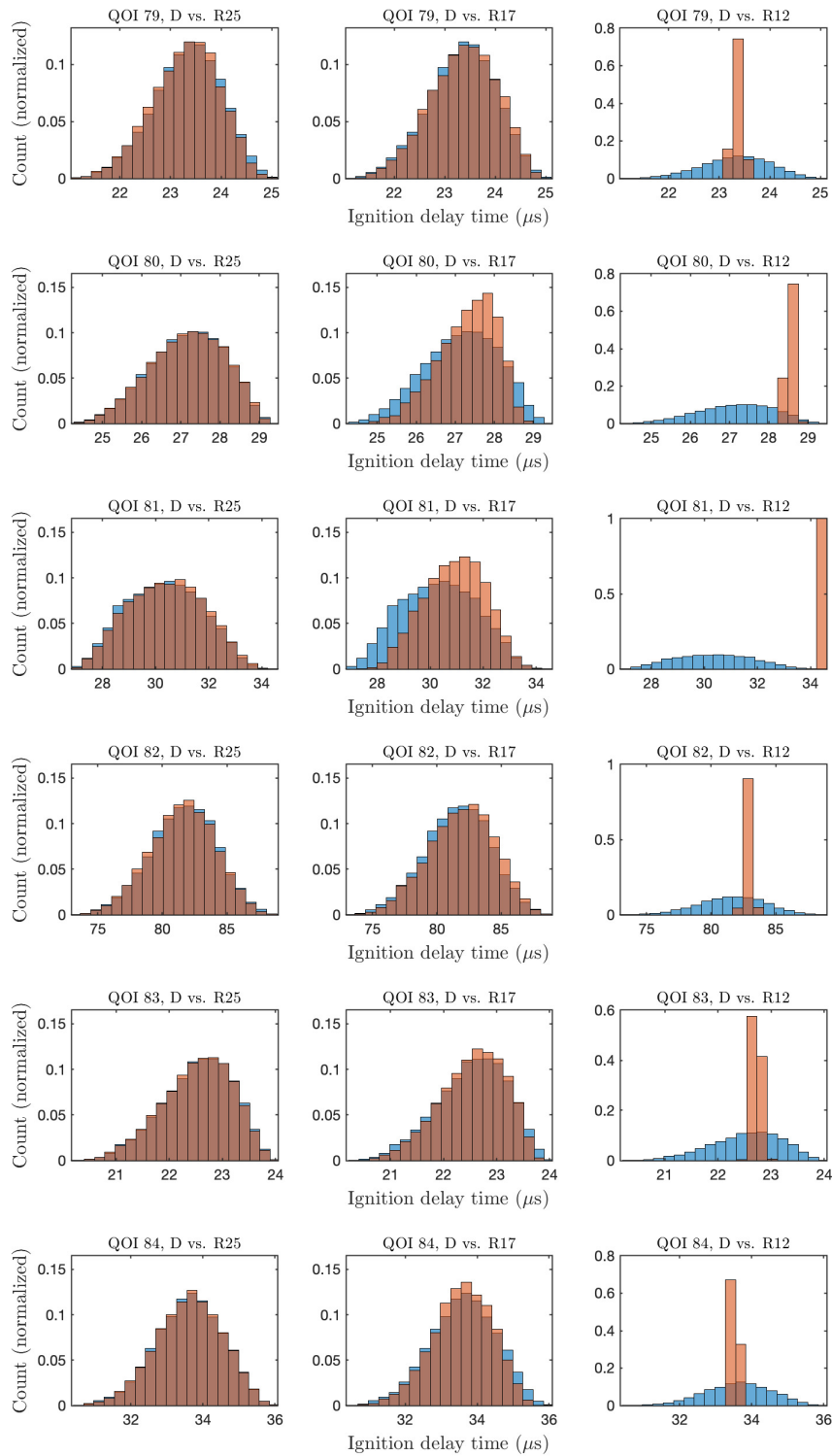


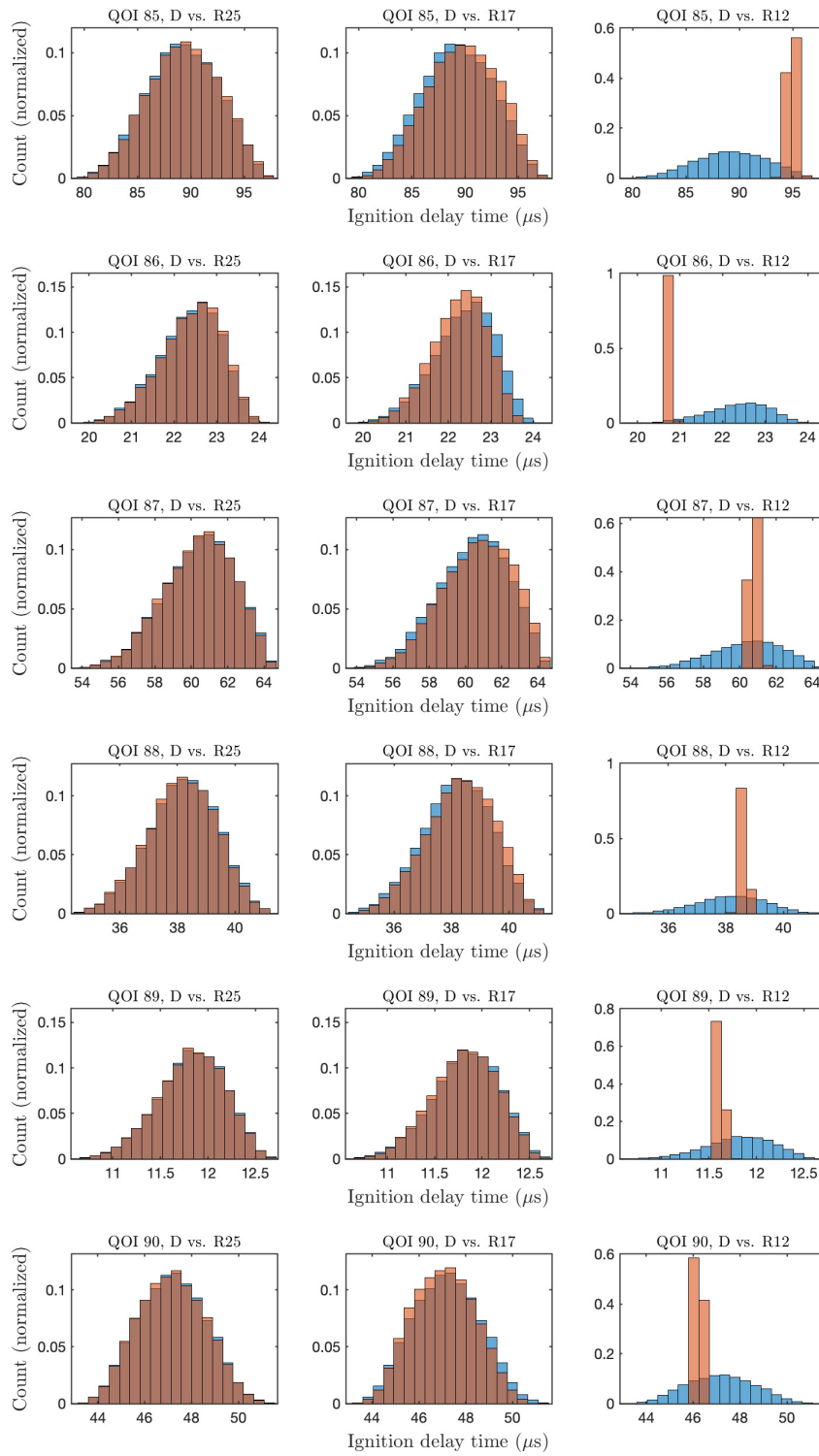


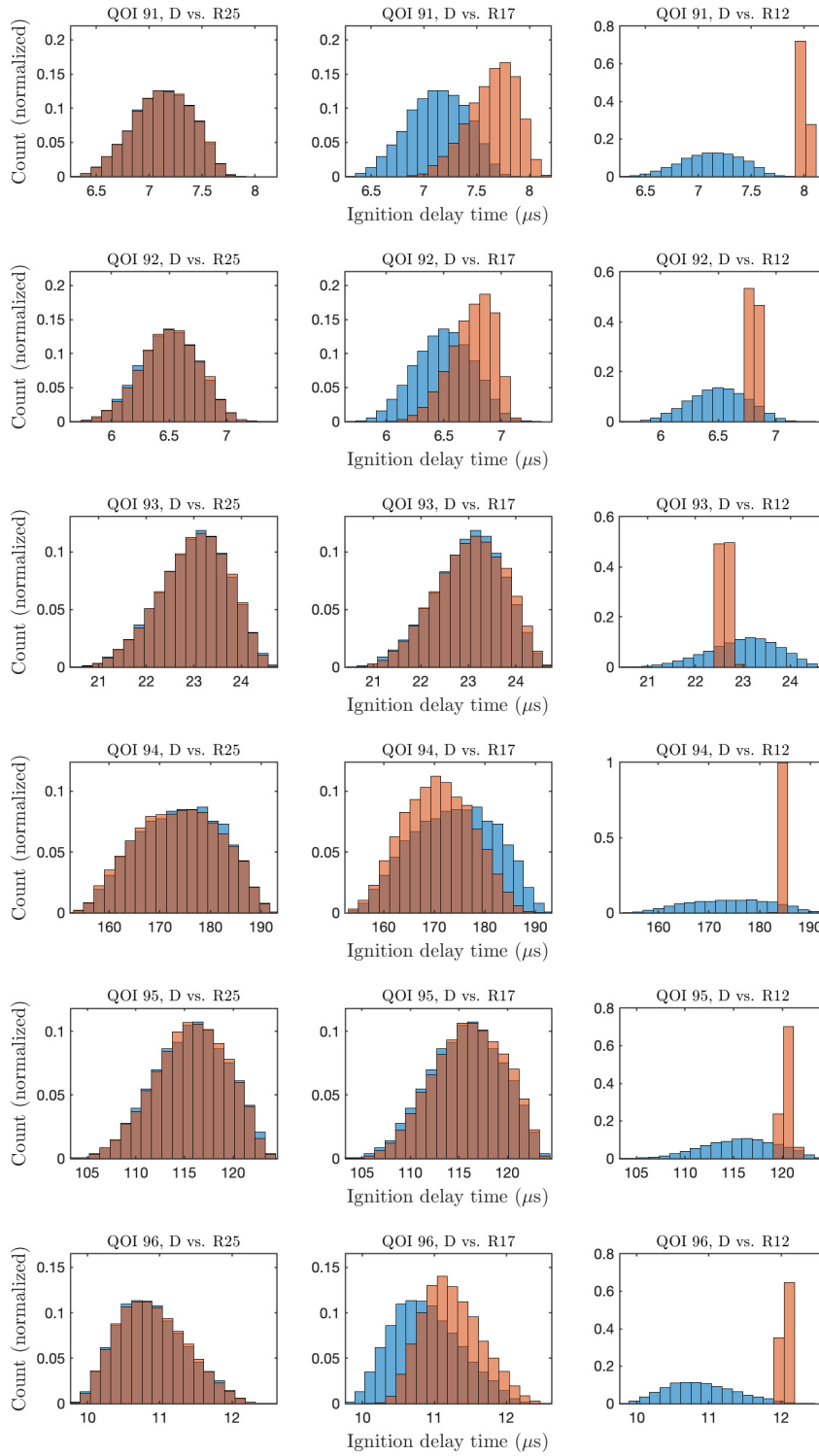












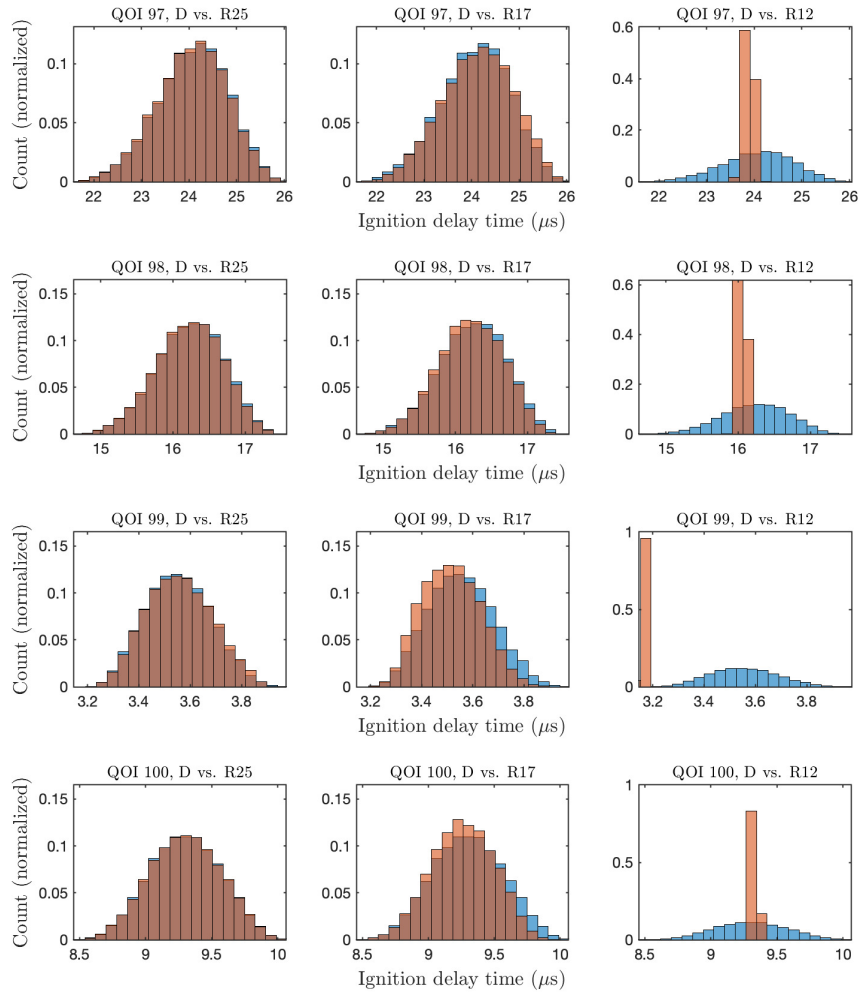


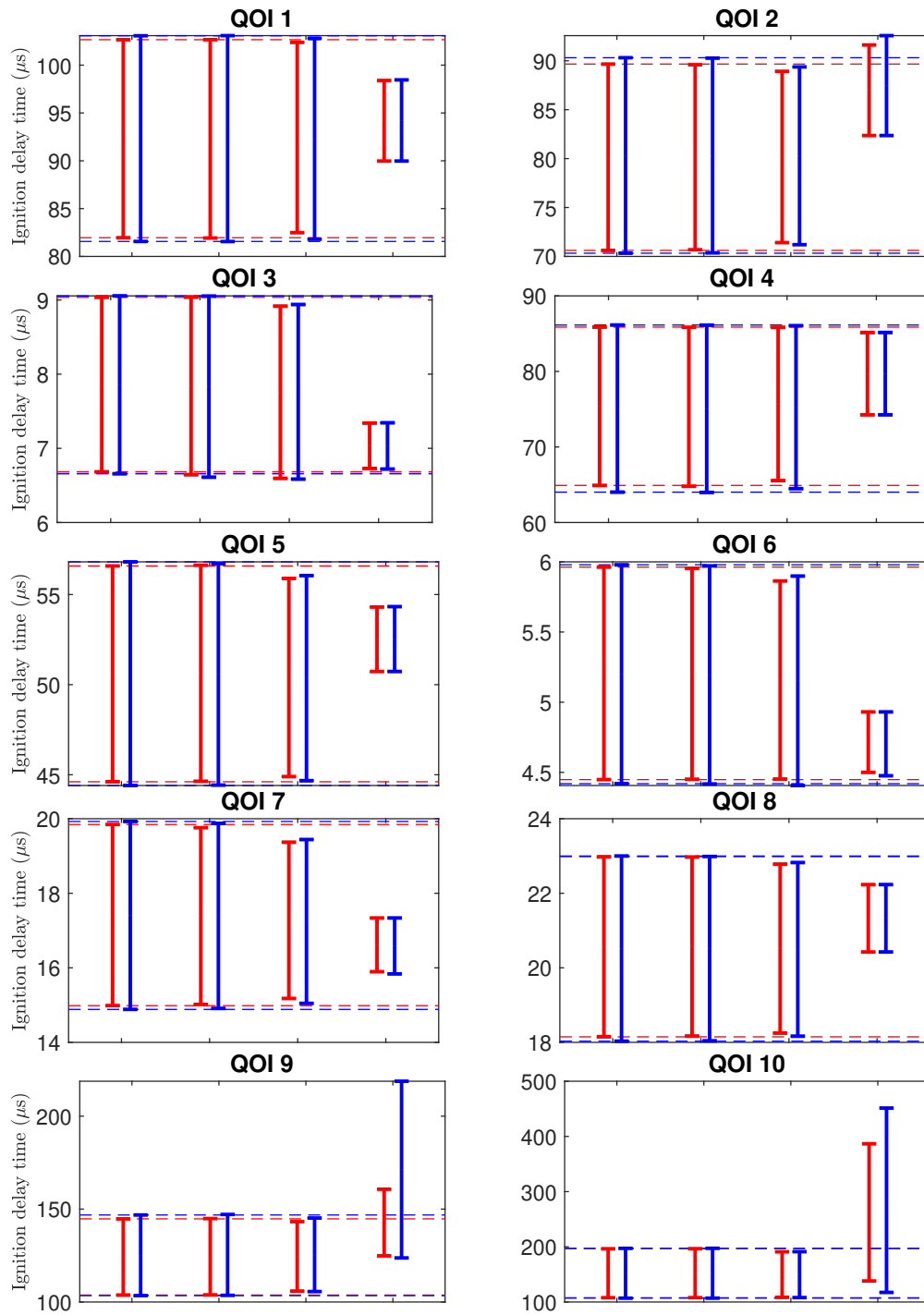
Figure A.1: Histograms for 100 QOIs of detailed (blue) and reduced (brown) model predictions over feasible sets.

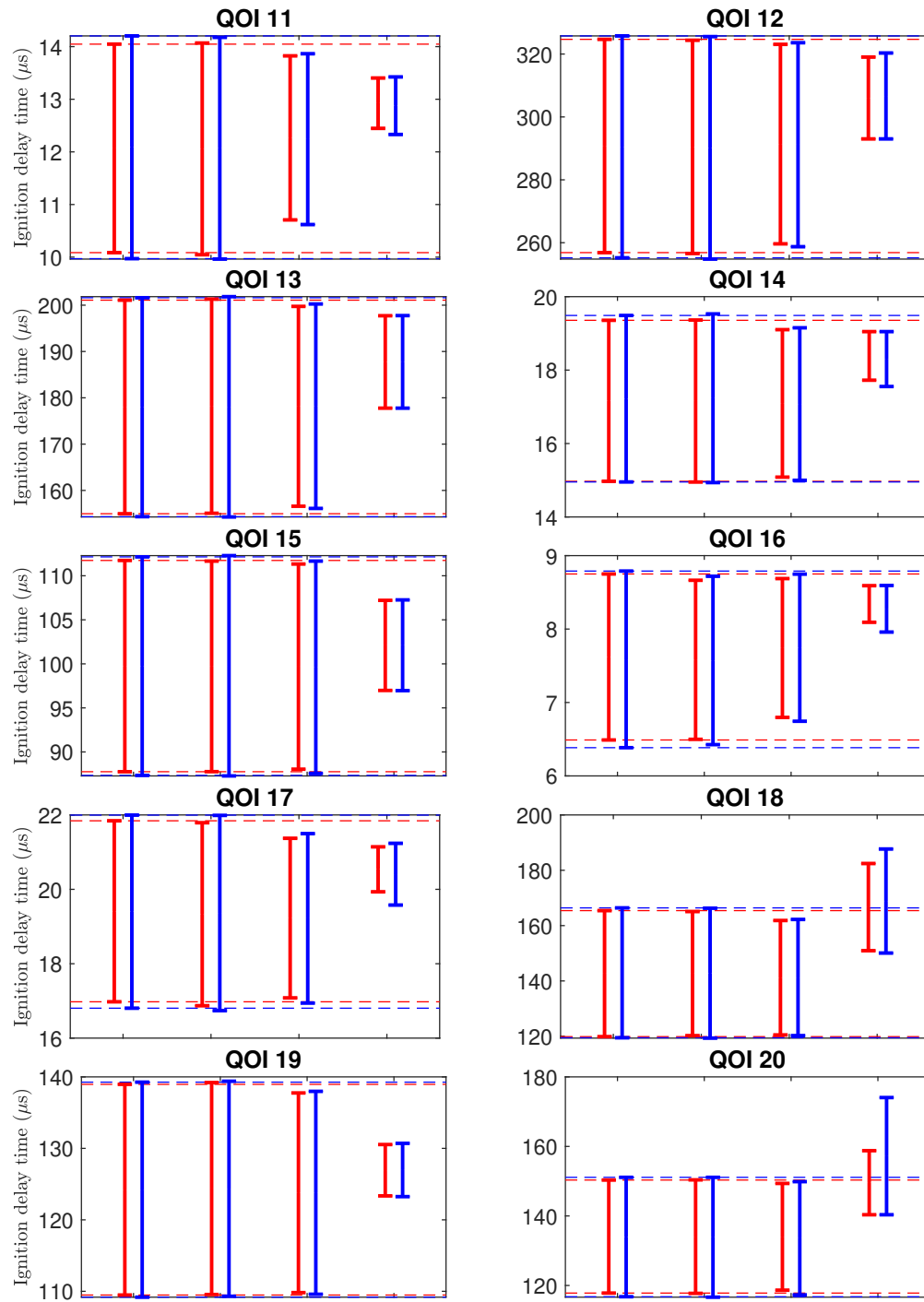
Table A.3: Hellinger distance calculated for the prediction intervals, h_p^U , and for the sampled histograms, h_p^H .

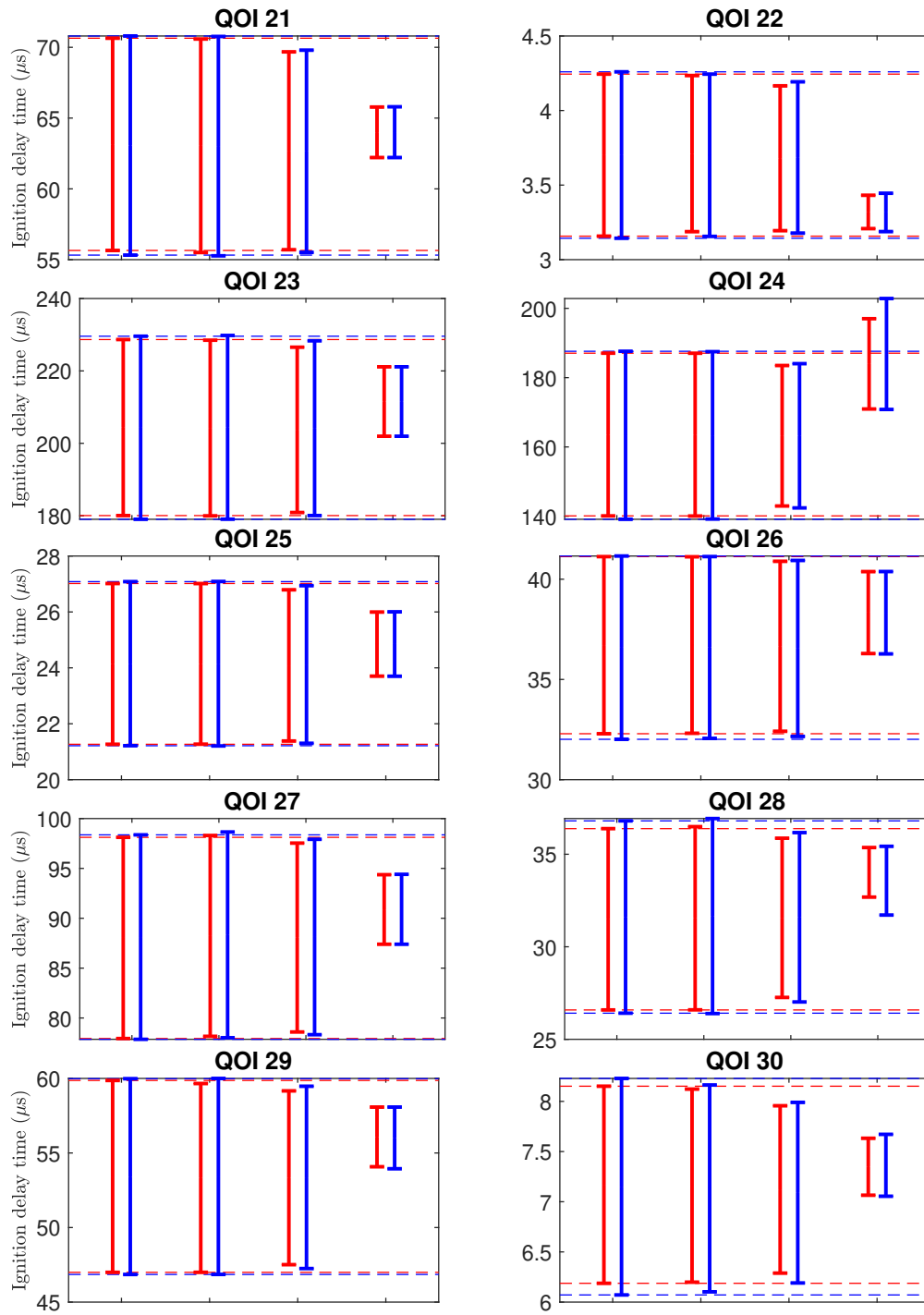
p	h_p^U (inner)			h_p^U (outer)			h_p^H		
	R25	R17	R12	R25	R17	R12	R25	R17	R12
1	0.001	0.017	0.499	0.000	0.009	0.509	0.010	0.040	0.860
2	0.003	0.042	0.584	0.002	0.050	0.575	0.007	0.033	0.869
3	0.012	0.038	0.635	0.013	0.034	0.630	0.006	0.128	0.971
4	0.003	0.021	0.469	0.001	0.014	0.486	0.013	0.034	0.890
5	0.002	0.036	0.598	0.004	0.037	0.604	0.010	0.087	0.747
6	0.003	0.036	0.605	0.002	0.028	0.597	0.013	0.111	0.971
7	0.009	0.067	0.576	0.006	0.063	0.575	0.006	0.034	0.906
8	0.003	0.024	0.536	0.002	0.025	0.546	0.004	0.043	0.725
9	0.004	0.055	0.563	0.006	0.055	0.664	0.017	0.037	0.881
10	0.002	0.054	0.617	0.001	0.058	0.562	0.005	0.099	0.971
11	0.007	0.175	0.659	0.003	0.173	0.640	0.009	0.430	0.947
12	0.003	0.031	0.549	0.004	0.040	0.549	0.009	0.036	0.764
13	0.003	0.027	0.515	0.002	0.029	0.520	0.006	0.041	0.760
14	0.003	0.036	0.622	0.005	0.045	0.592	0.004	0.132	0.858
15	0.001	0.011	0.484	0.003	0.011	0.495	0.010	0.052	0.737
16	0.019	0.134	0.692	0.018	0.152	0.653	0.008	0.417	0.998
17	0.015	0.063	0.652	0.008	0.062	0.590	0.006	0.170	0.921
18	0.009	0.077	0.673	0.003	0.088	0.667	0.006	0.179	0.940
19	0.004	0.023	0.641	0.003	0.024	0.636	0.005	0.087	0.879
20	0.002	0.034	0.658	0.002	0.037	0.716	0.011	0.045	0.979
21	0.006	0.037	0.644	0.002	0.037	0.649	0.007	0.091	0.803
22	0.019	0.047	0.669	0.009	0.038	0.649	0.005	0.083	0.990
23	0.002	0.025	0.519	0.002	0.018	0.531	0.005	0.052	0.737
24	0.001	0.090	0.627	0.002	0.094	0.642	0.004	0.123	0.929
25	0.000	0.023	0.511	0.001	0.015	0.516	0.012	0.034	0.868
26	0.002	0.015	0.482	0.003	0.015	0.494	0.011	0.053	0.750
27	0.008	0.026	0.558	0.008	0.018	0.561	0.007	0.052	0.739
28	0.005	0.062	0.640	0.005	0.057	0.565	0.005	0.184	0.953
29	0.008	0.040	0.594	0.001	0.028	0.590	0.004	0.115	0.759
30	0.008	0.073	0.601	0.017	0.083	0.605	0.009	0.165	0.864
31	0.013	0.014	0.522	0.006	0.015	0.531	0.015	0.028	0.729
32	0.017	0.145	0.687	0.025	0.170	0.682	0.007	0.320	0.938
33	0.004	0.053	0.541	0.002	0.060	0.541	0.006	0.100	0.845
34	0.008	0.161	0.635	0.004	0.138	0.547	0.013	0.298	1.000
35	0.002	0.257	0.729	0.005	0.224	0.702	0.005	0.245	0.898
36	0.006	0.010	0.633	0.008	0.016	0.606	0.005	0.124	0.901

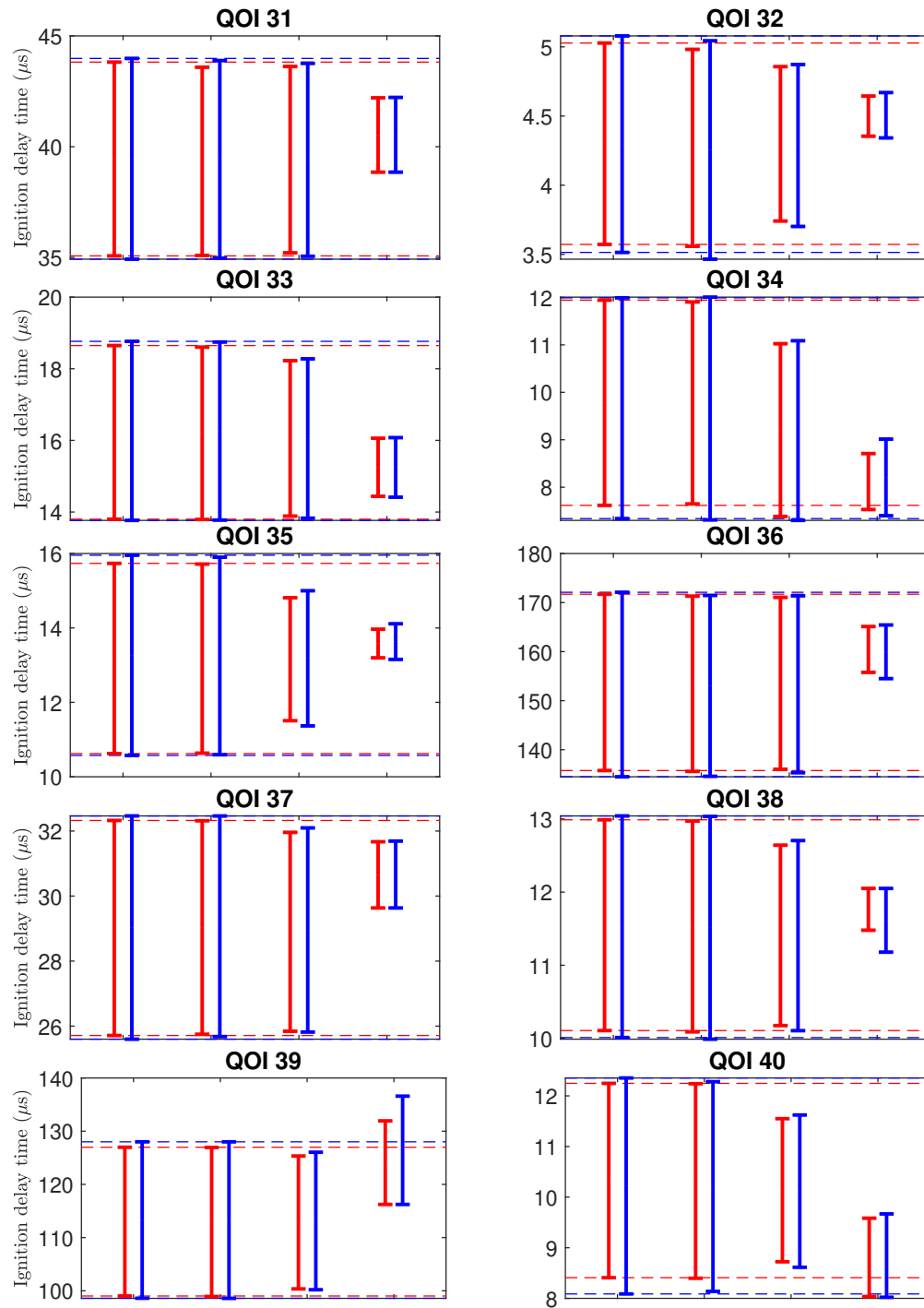
37	0.004	0.033	0.605	0.007	0.036	0.609	0.007	0.116	0.811
38	0.005	0.083	0.681	0.005	0.076	0.599	0.004	0.122	0.874
39	0.002	0.067	0.591	0.000	0.080	0.599	0.007	0.027	0.893
40	0.003	0.183	0.592	0.011	0.197	0.538	0.008	0.223	0.999
41	0.007	0.131	0.624	0.005	0.099	0.554	0.006	0.190	0.964
42	0.005	0.038	0.588	0.007	0.052	0.594	0.013	0.142	0.786
43	0.002	0.011	0.487	0.001	0.012	0.489	0.012	0.044	0.866
44	0.011	0.036	0.636	0.012	0.042	0.554	0.013	0.166	0.860
45	0.008	0.028	0.621	0.006	0.019	0.613	0.011	0.071	0.959
46	0.011	0.019	0.561	0.005	0.014	0.569	0.009	0.035	0.885
47	0.002	0.044	0.573	0.006	0.040	0.579	0.006	0.071	0.825
48	0.018	0.034	0.544	0.011	0.037	0.556	0.007	0.040	0.773
49	0.006	0.026	0.615	0.006	0.022	0.620	0.004	0.121	0.760
50	0.006	0.031	0.515	0.004	0.027	0.522	0.007	0.059	0.731
51	0.008	0.012	0.593	0.003	0.021	0.587	0.011	0.046	0.897
52	0.014	0.035	0.514	0.022	0.035	0.529	0.007	0.038	0.813
53	0.010	0.029	0.526	0.007	0.028	0.531	0.010	0.043	0.716
54	0.015	0.117	0.675	0.010	0.135	0.651	0.007	0.297	0.932
55	0.017	0.301	0.657	0.013	0.297	0.630	0.008	0.111	0.928
56	0.008	0.032	0.538	0.007	0.026	0.545	0.012	0.030	0.758
57	0.001	0.029	0.503	0.002	0.031	0.520	0.007	0.070	0.859
58	0.013	0.061	0.664	0.007	0.050	0.632	0.006	0.106	0.882
59	0.006	0.033	0.554	0.007	0.027	0.555	0.004	0.077	0.731
60	0.008	0.032	0.655	0.003	0.026	0.647	0.012	0.090	0.971
61	0.018	0.029	0.594	0.018	0.029	0.539	0.005	0.159	0.896
62	0.007	0.023	0.589	0.006	0.033	0.594	0.008	0.052	0.934
63	0.005	0.034	0.540	0.005	0.033	0.545	0.007	0.032	0.782
64	0.009	0.017	0.542	0.002	0.004	0.540	0.008	0.038	0.746
65	0.009	0.012	0.653	0.009	0.022	0.619	0.004	0.128	0.880
66	0.004	0.017	0.536	0.000	0.022	0.536	0.009	0.028	0.769
67	0.005	0.031	0.614	0.005	0.026	0.620	0.005	0.113	0.778
68	0.011	0.160	0.571	0.010	0.169	0.542	0.020	0.044	0.889
69	0.018	0.138	0.731	0.008	0.141	0.659	0.012	0.366	0.994
70	0.021	0.207	0.662	0.010	0.196	0.637	0.016	0.146	0.998
71	0.005	0.040	0.641	0.015	0.042	0.573	0.016	0.106	0.894
72	0.004	0.032	0.641	0.008	0.030	0.631	0.018	0.066	0.926
73	0.002	0.026	0.516	0.003	0.025	0.517	0.012	0.037	0.836
74	0.009	0.057	0.646	0.009	0.063	0.637	0.014	0.098	0.999
75	0.001	0.011	0.490	0.001	0.010	0.499	0.010	0.061	0.749
76	0.003	0.015	0.595	0.005	0.012	0.601	0.011	0.054	0.888
77	0.008	0.027	0.613	0.002	0.026	0.604	0.020	0.053	0.908
78	0.013	0.077	0.669	0.005	0.078	0.506	0.014	0.110	0.849

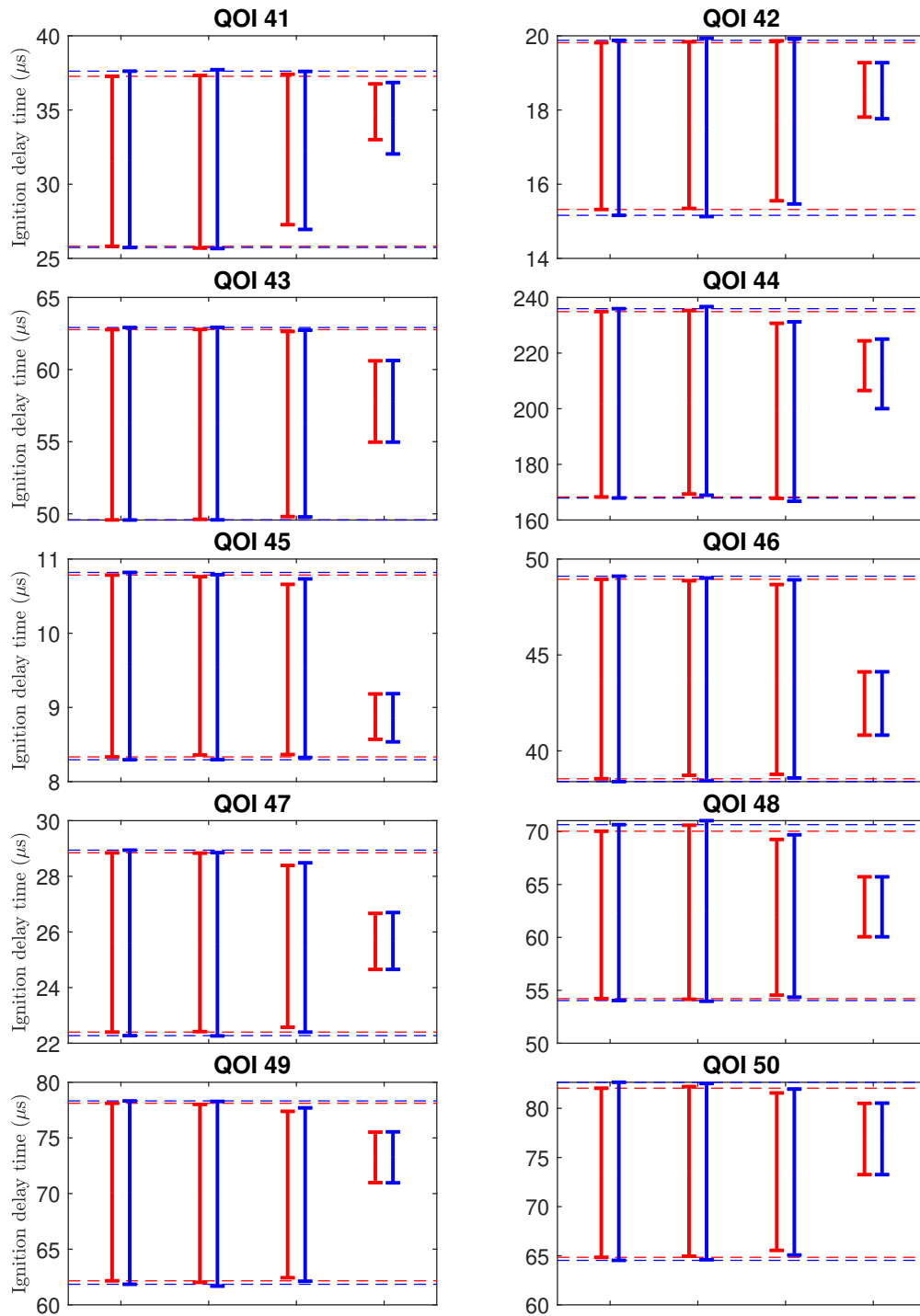
79	0.002	0.015	0.527	0.001	0.015	0.533	0.027	0.024	0.736
80	0.006	0.061	0.619	0.002	0.047	0.582	0.010	0.156	0.856
81	0.007	0.089	0.677	0.008	0.075	0.550	0.022	0.203	0.996
82	0.010	0.011	0.519	0.003	0.009	0.537	0.016	0.046	0.774
83	0.004	0.029	0.529	0.005	0.028	0.539	0.008	0.052	0.718
84	0.005	0.029	0.618	0.006	0.029	0.624	0.004	0.068	0.771
85	0.003	0.077	0.532	0.005	0.087	0.535	0.012	0.056	0.856
86	0.008	0.288	0.623	0.012	0.287	0.632	0.017	0.098	0.934
87	0.002	0.009	0.548	0.004	0.014	0.562	0.007	0.044	0.740
88	0.032	0.017	0.528	0.002	0.013	0.545	0.011	0.042	0.768
89	0.007	0.028	0.579	0.006	0.019	0.586	0.005	0.022	0.767
90	0.006	0.019	0.611	0.005	0.019	0.620	0.005	0.052	0.847
91	0.007	0.229	0.672	0.003	0.216	0.628	0.005	0.654	0.998
92	0.030	0.340	0.690	0.032	0.352	0.695	0.016	0.389	0.916
93	0.004	0.007	0.510	0.003	0.015	0.514	0.009	0.023	0.759
94	0.024	0.007	0.651	0.018	0.007	0.533	0.017	0.186	0.909
95	0.004	0.048	0.542	0.004	0.052	0.532	0.017	0.034	0.812
96	0.009	0.110	0.620	0.003	0.129	0.596	0.010	0.328	0.950
97	0.004	0.013	0.520	0.005	0.014	0.523	0.010	0.032	0.730
98	0.004	0.014	0.581	0.008	0.015	0.592	0.012	0.033	0.763
99	0.016	0.121	0.702	0.011	0.105	0.666	0.019	0.119	0.999
100	0.010	0.053	0.669	0.008	0.053	0.637	0.005	0.094	0.821

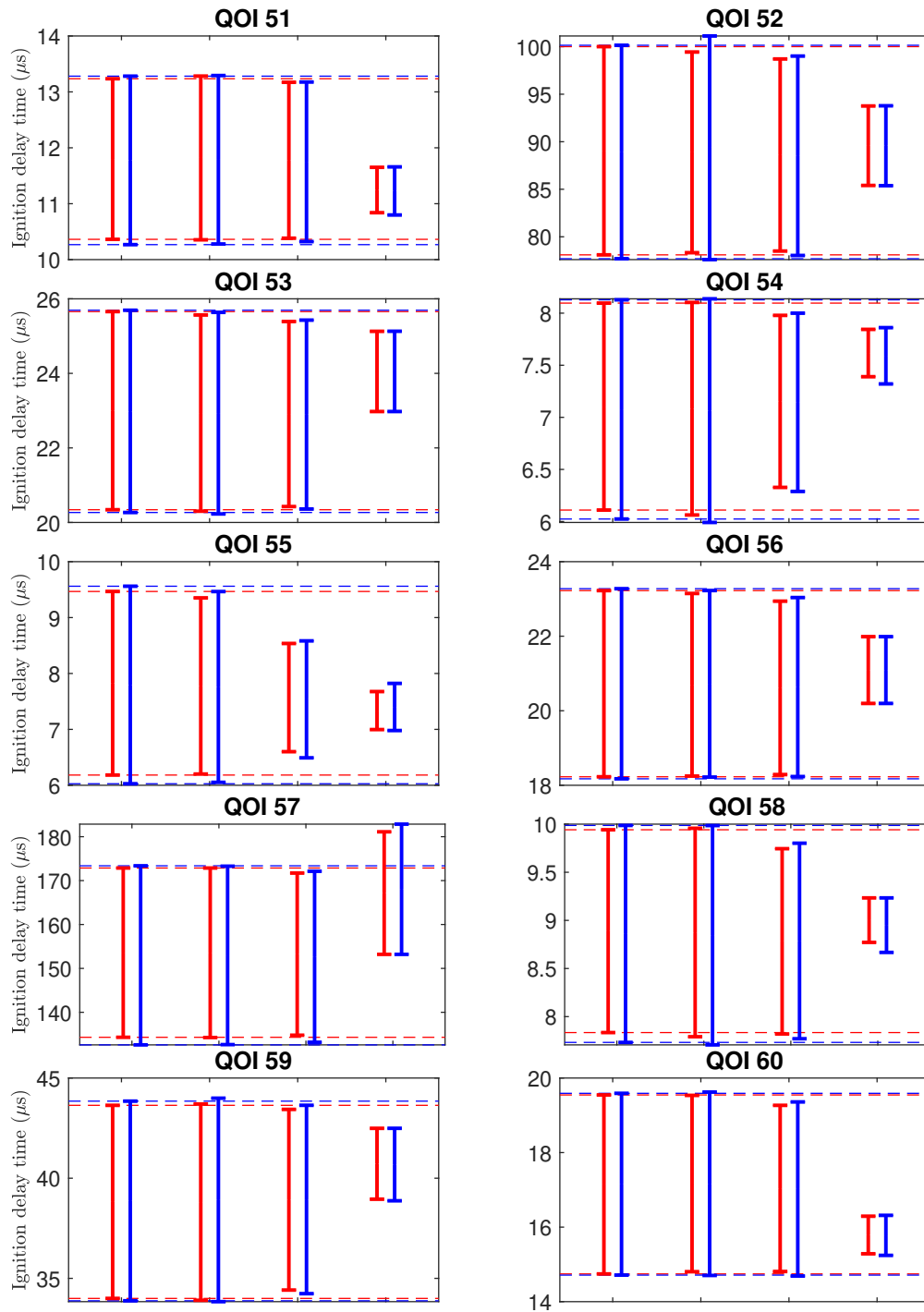


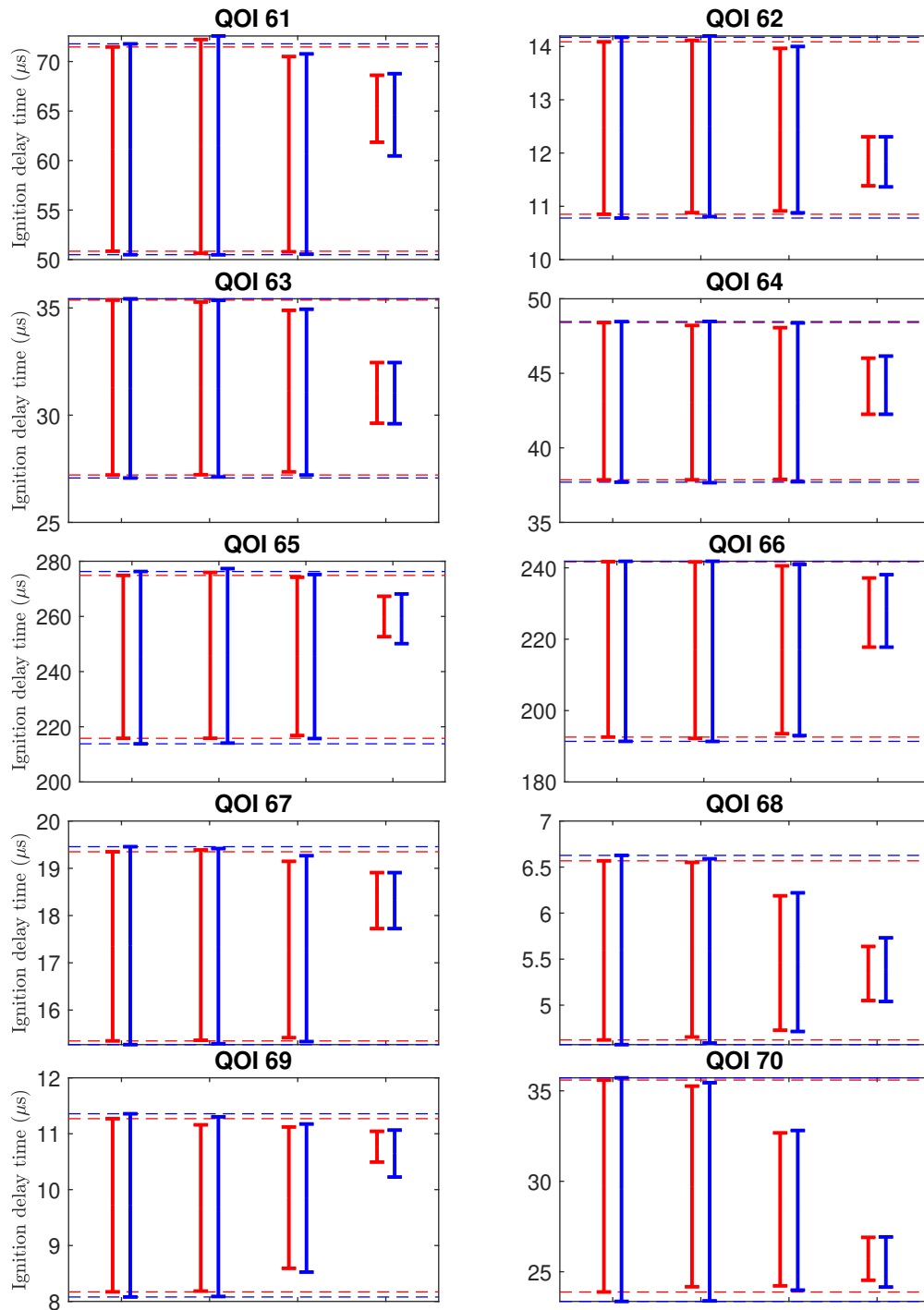


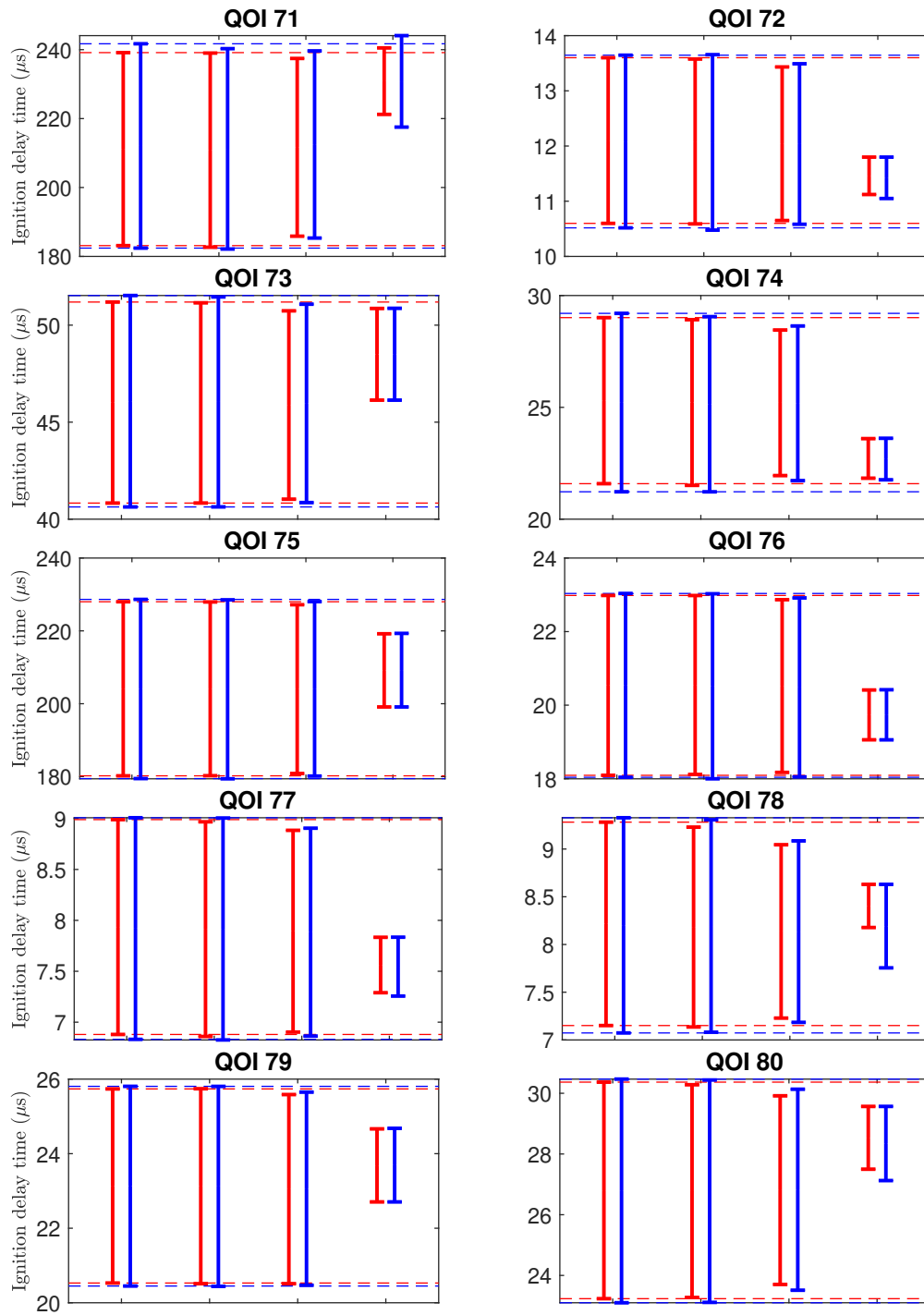


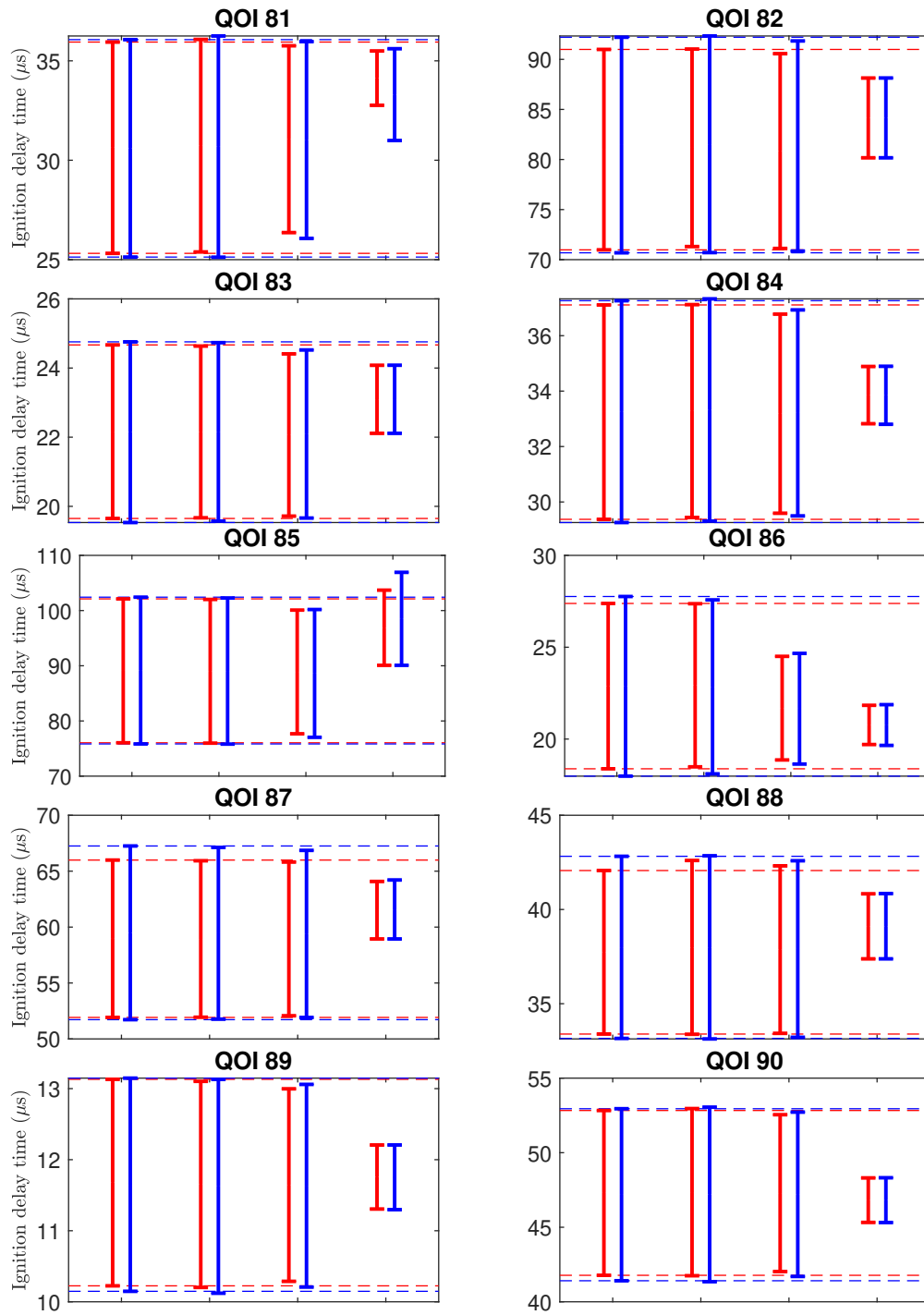












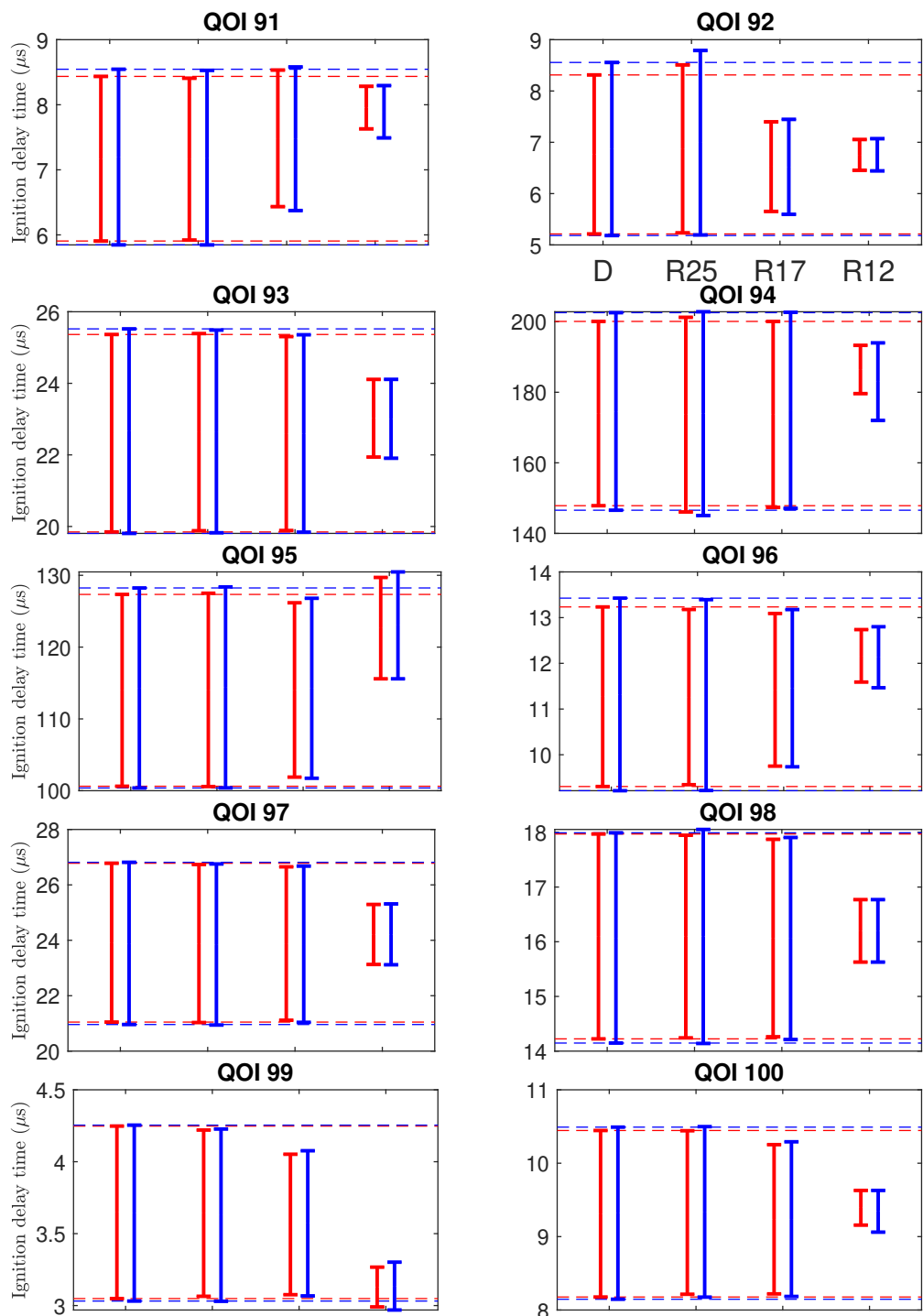
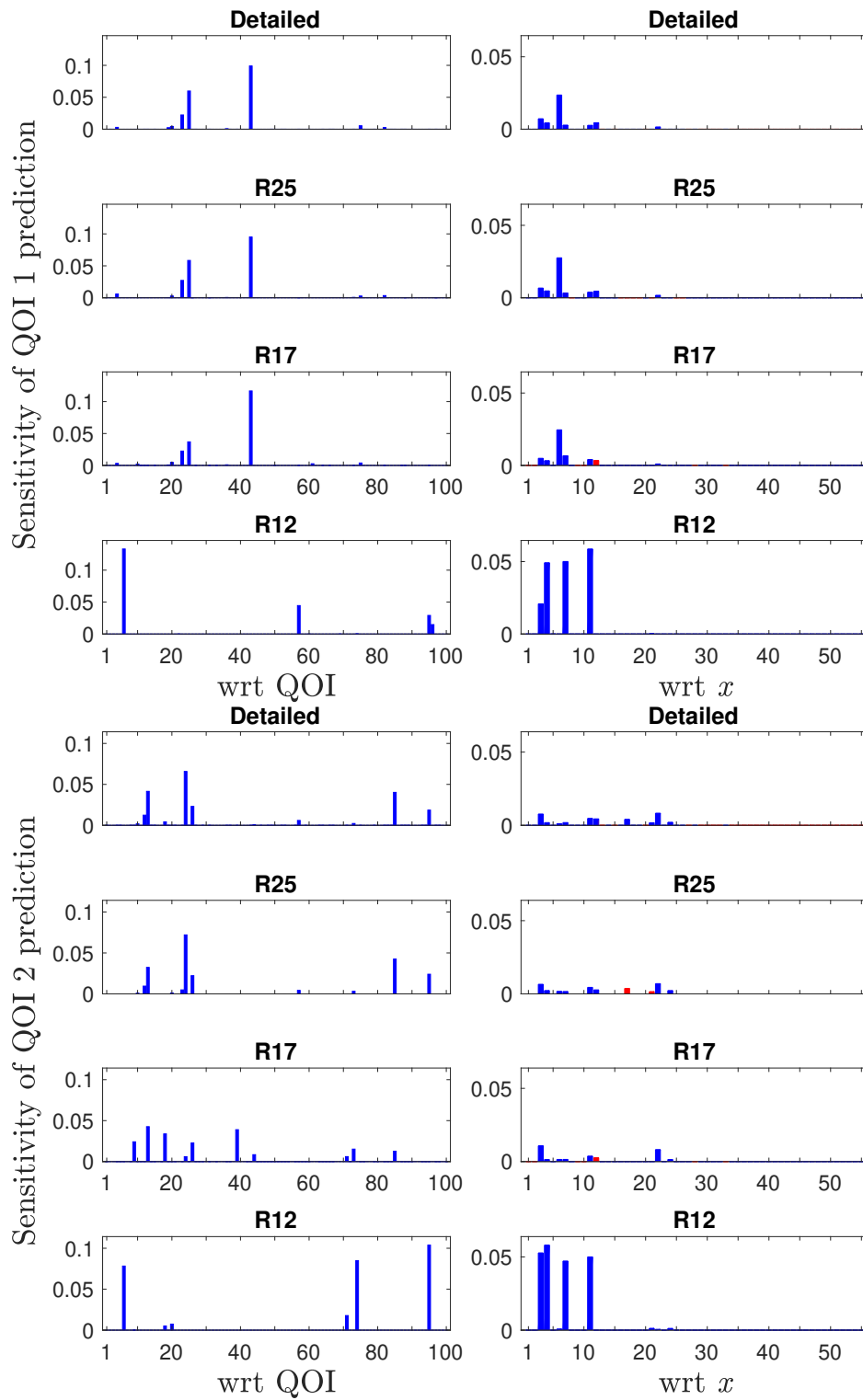
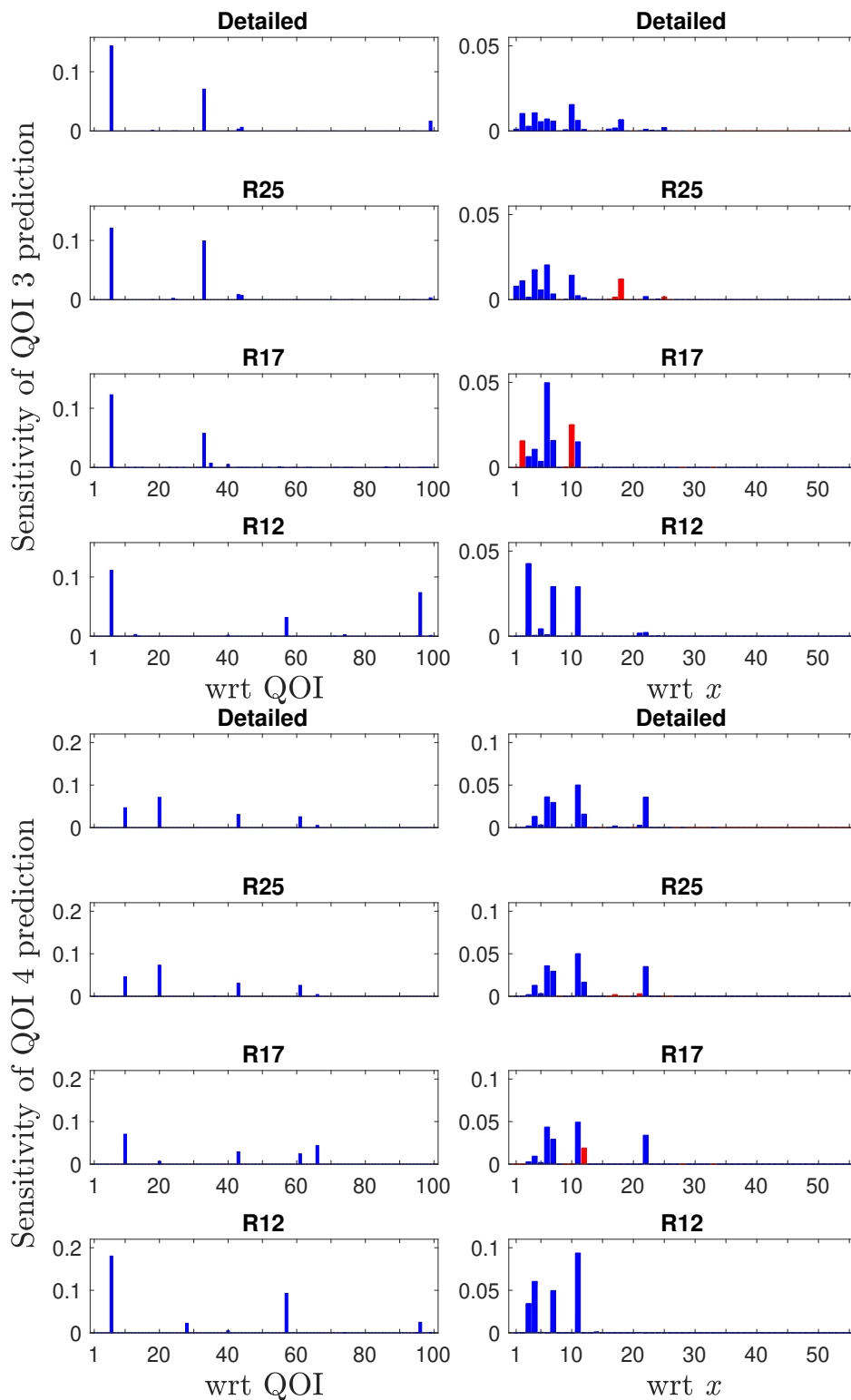
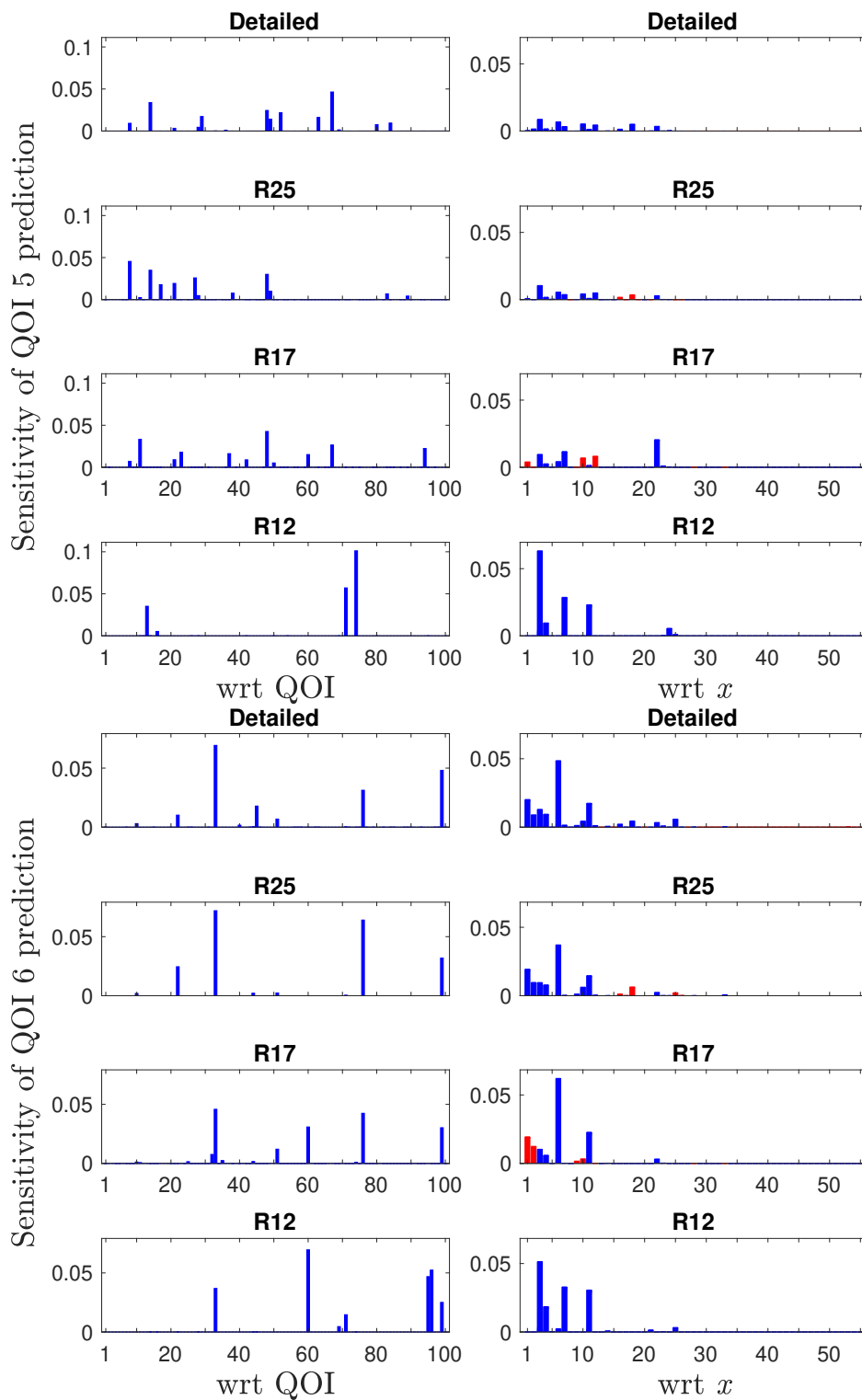
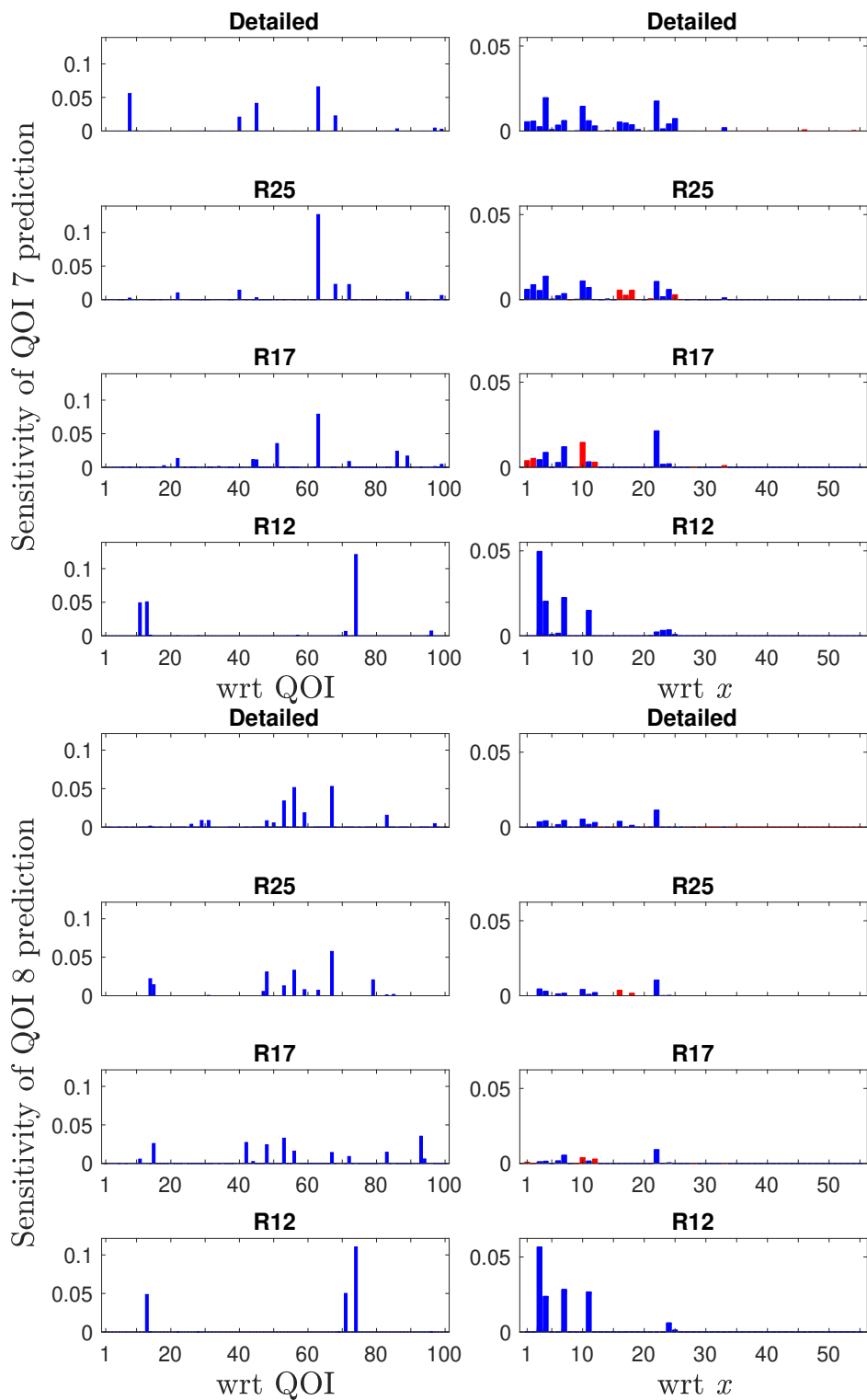


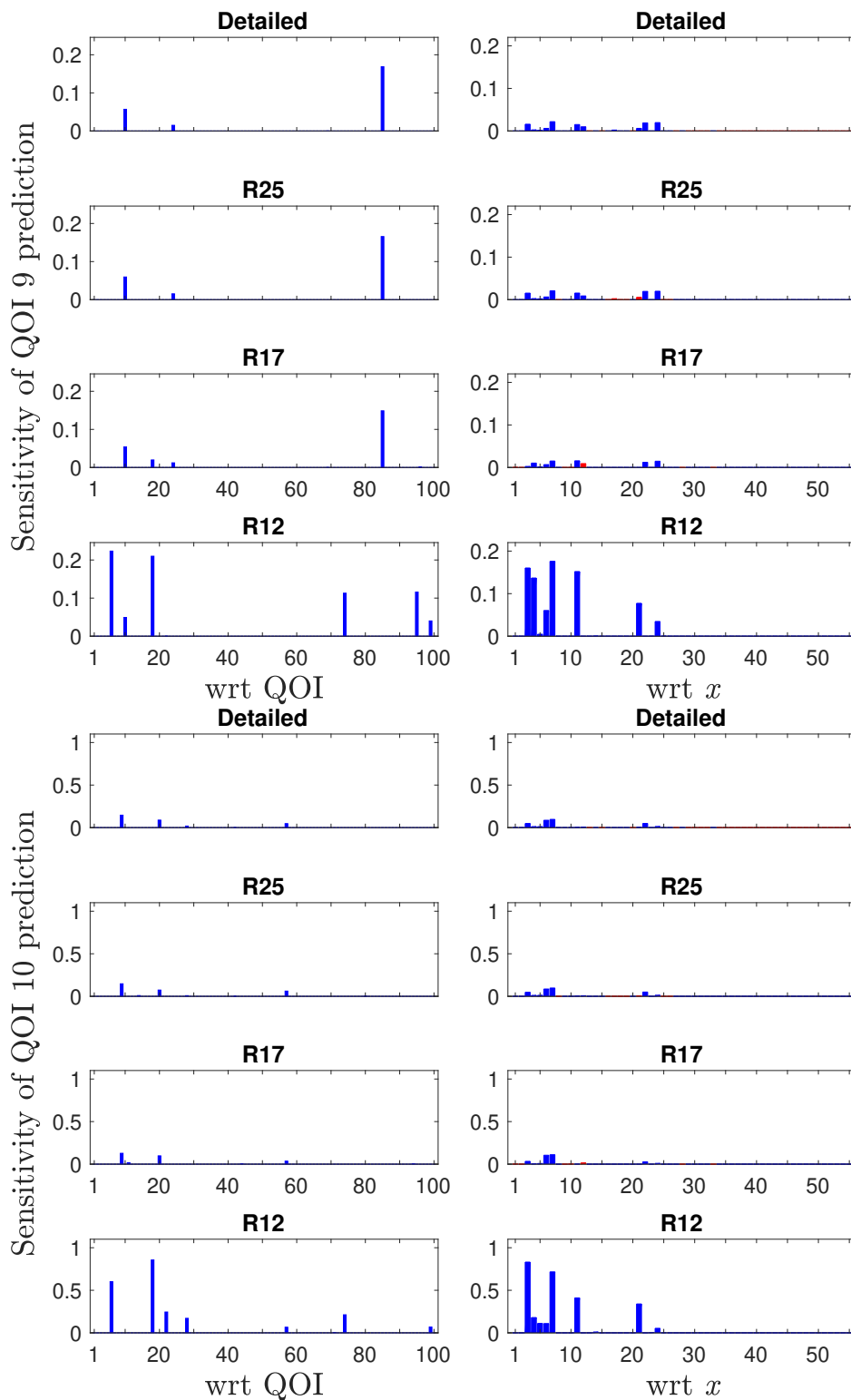
Figure A.2: QOI prediction intervals computed with the detailed, D, and reduced, R25, and R17, models; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and horizontal dashed lines are prediction intervals of the detailed model.

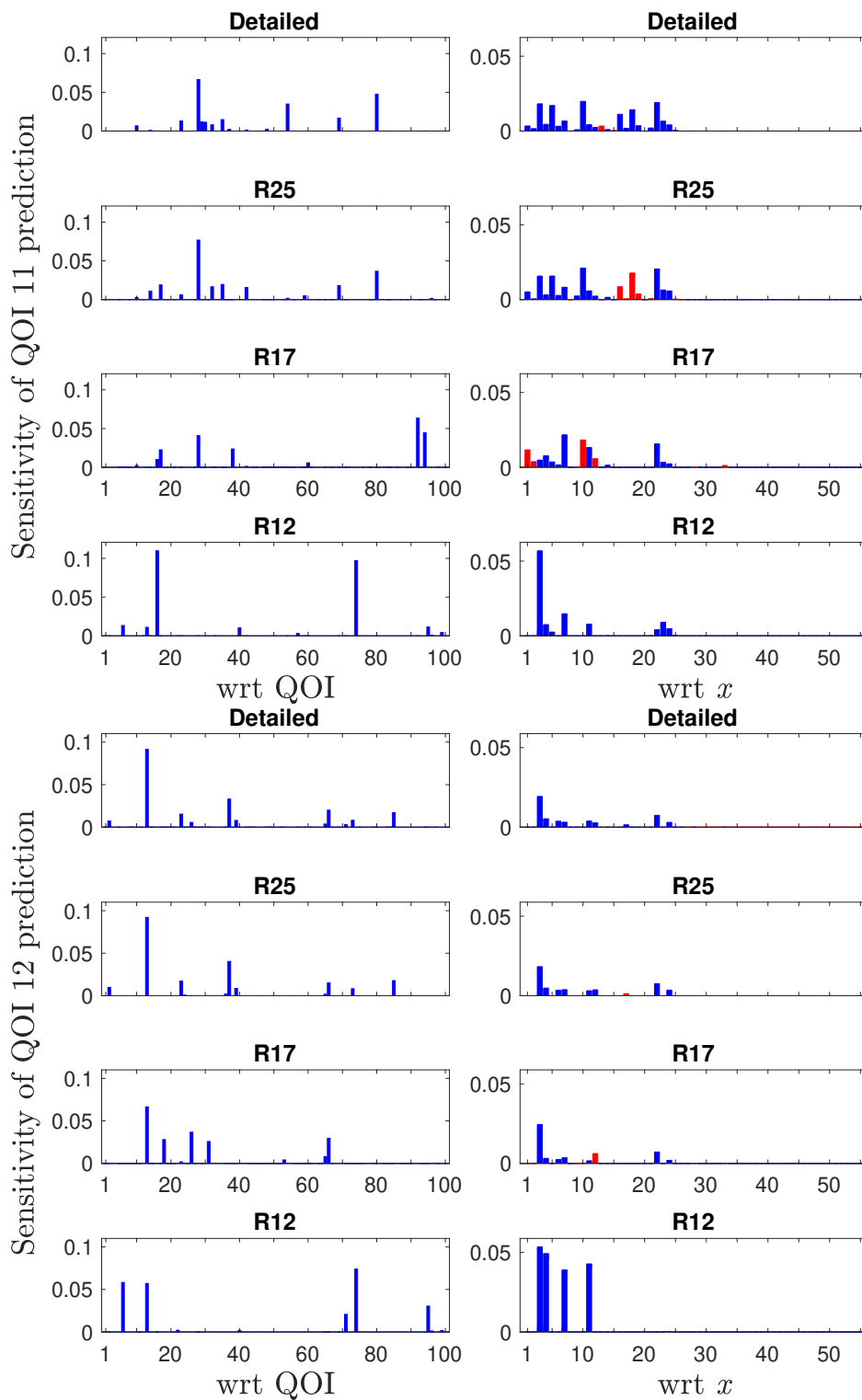


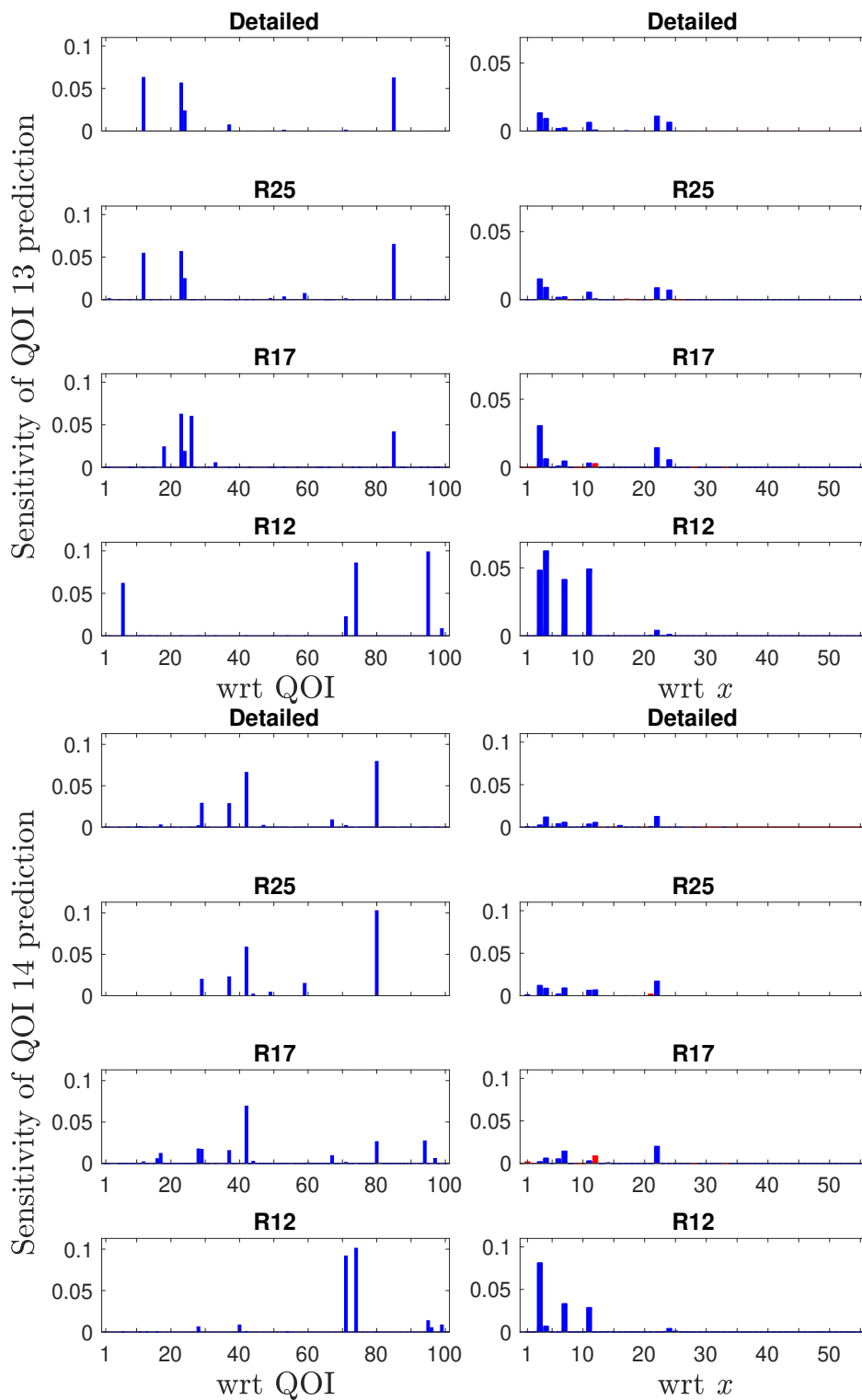


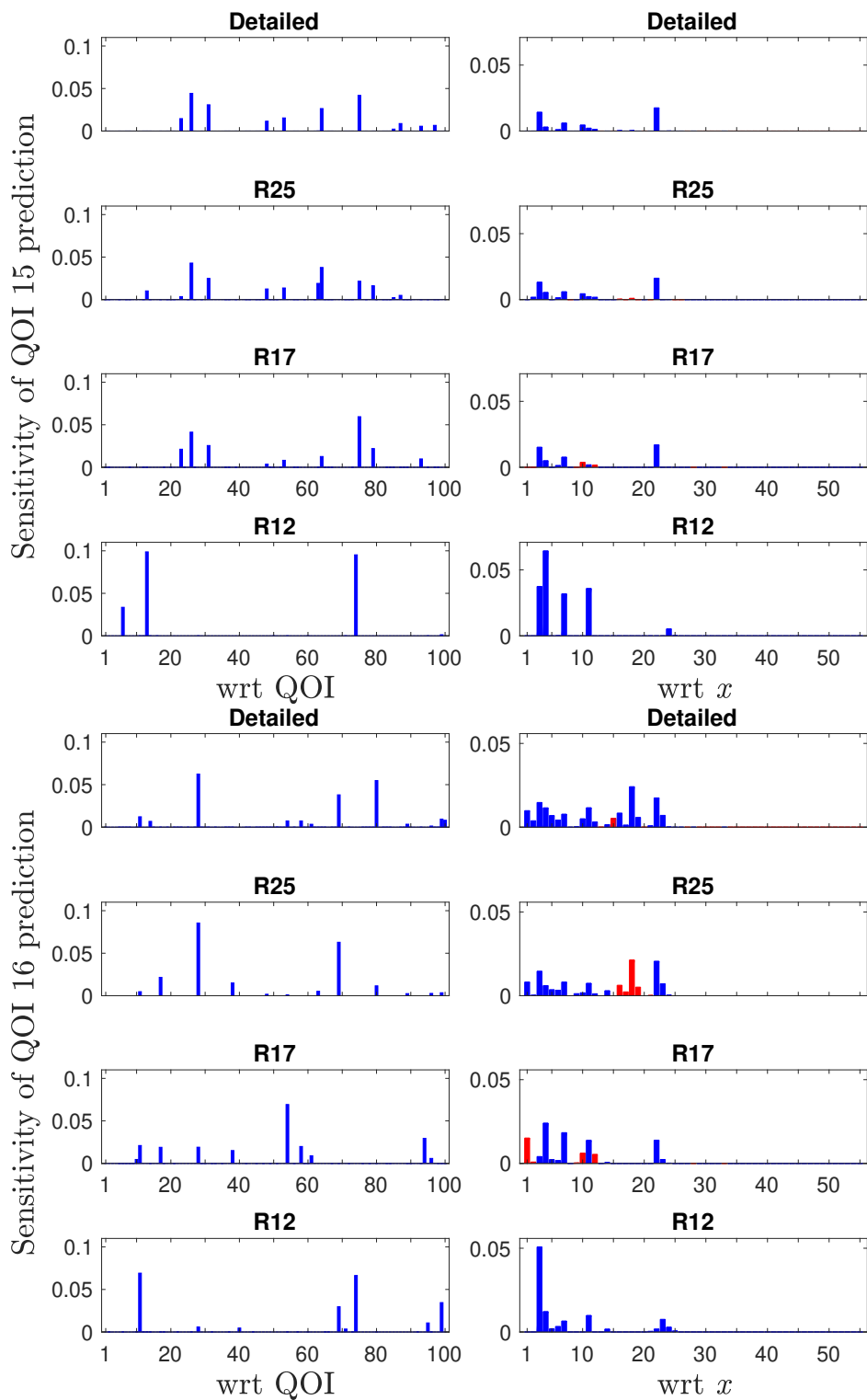


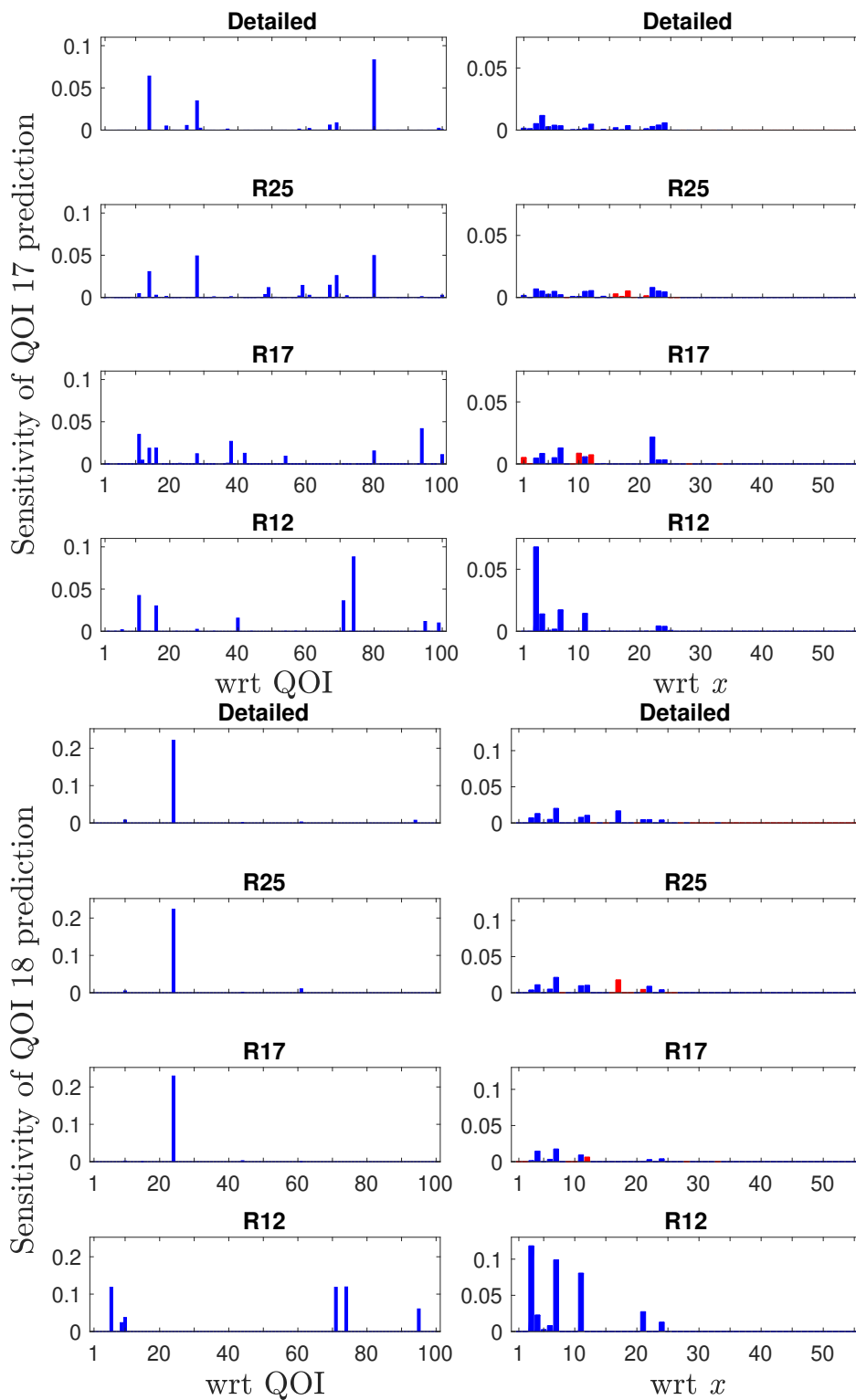


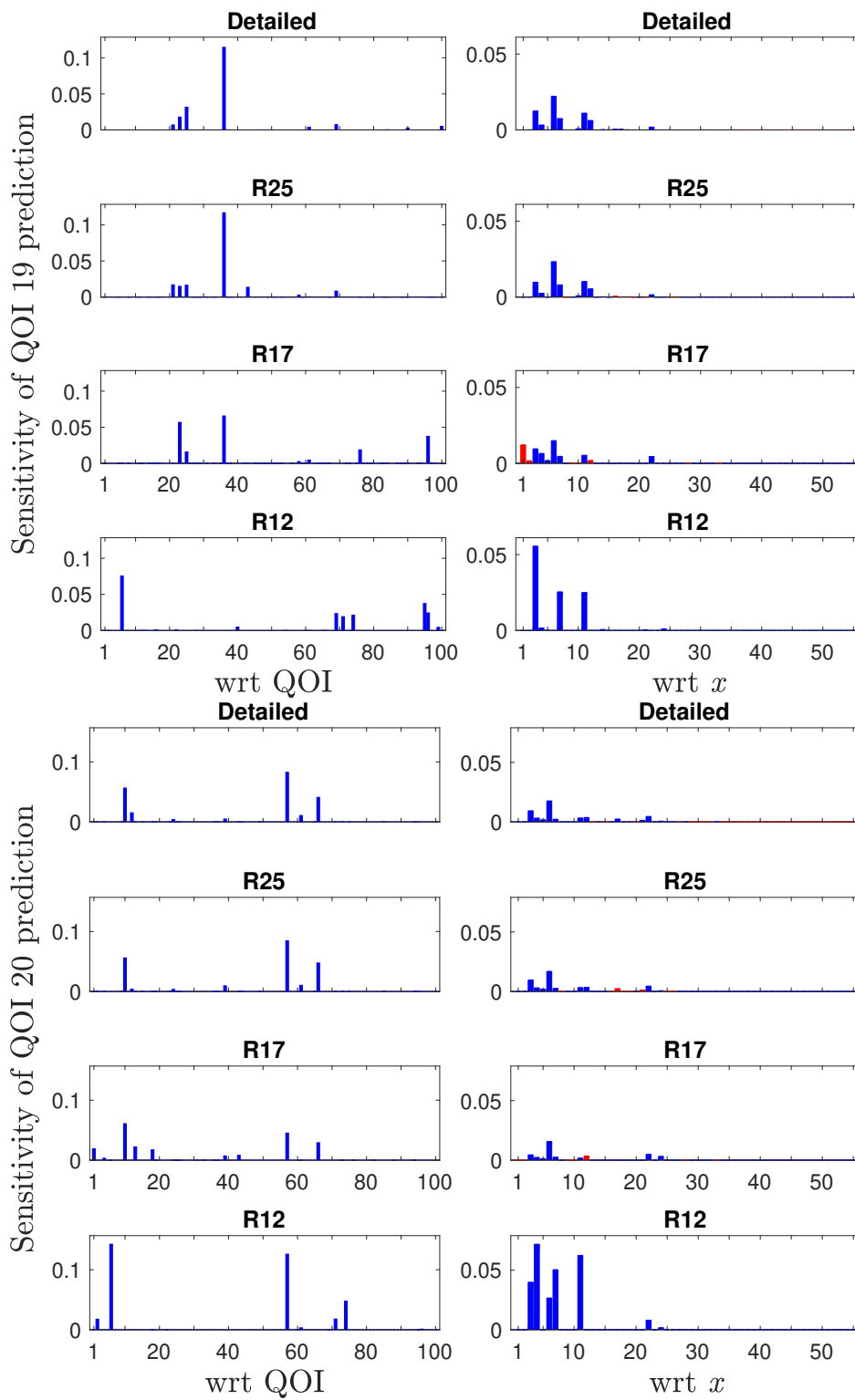


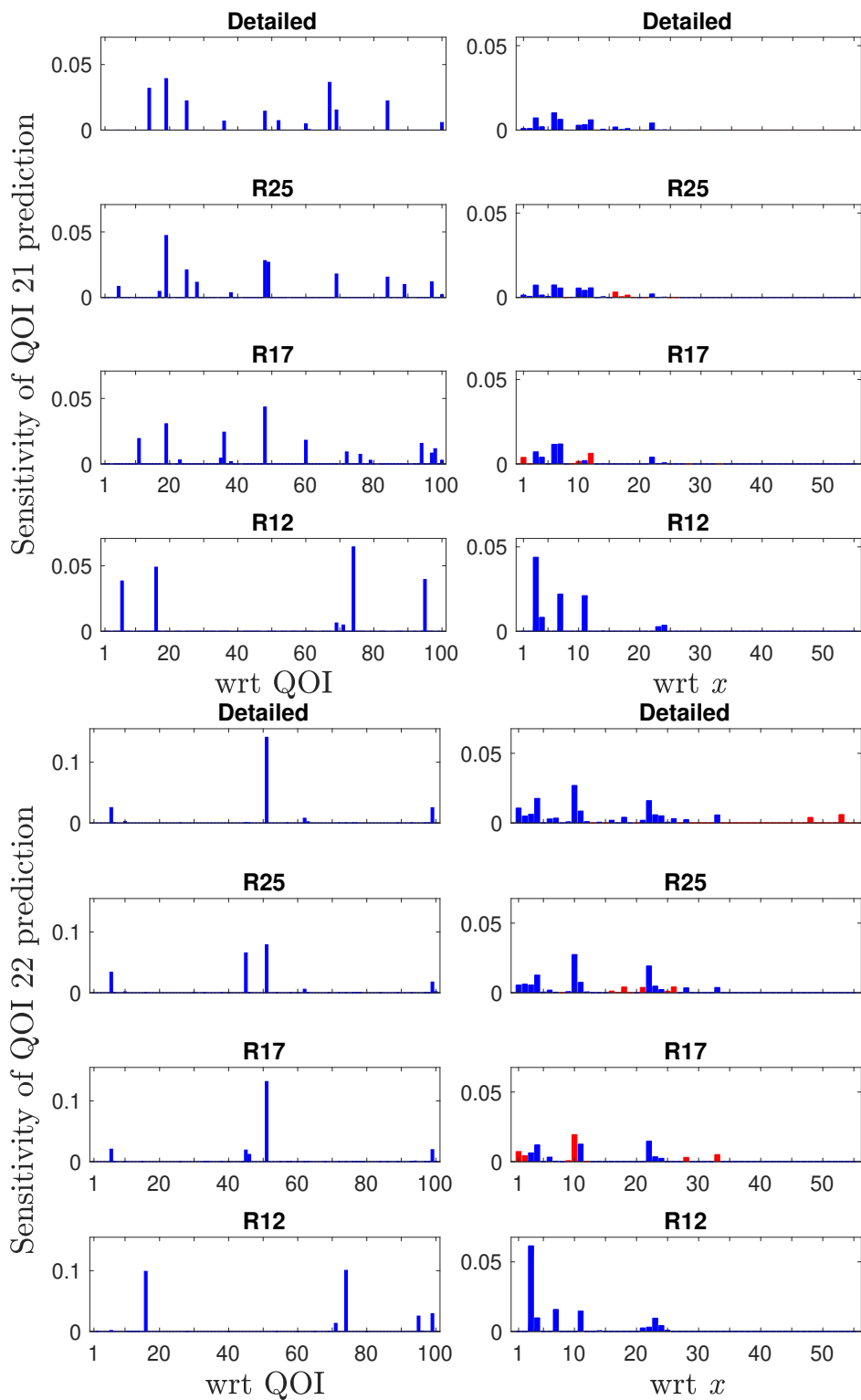


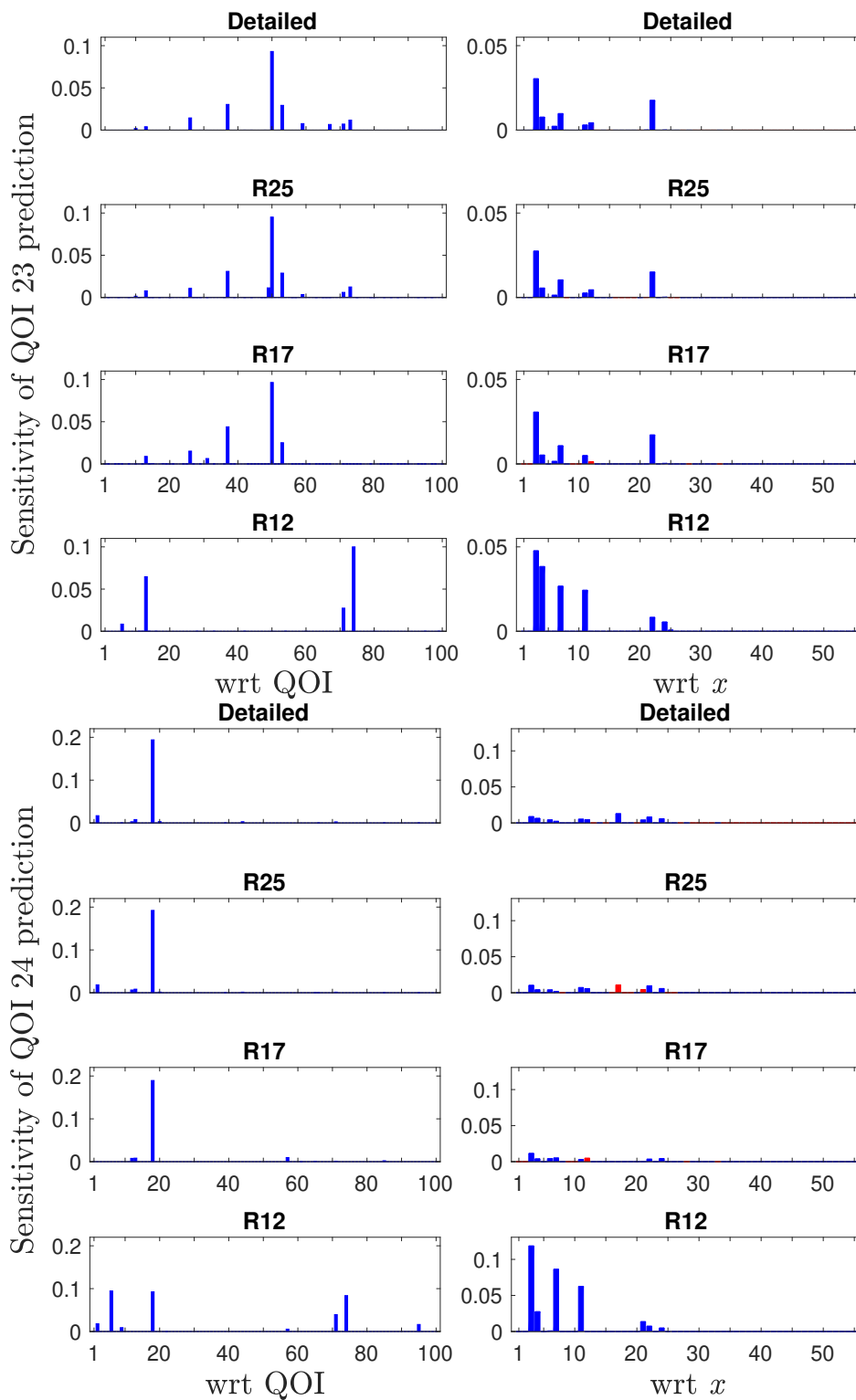


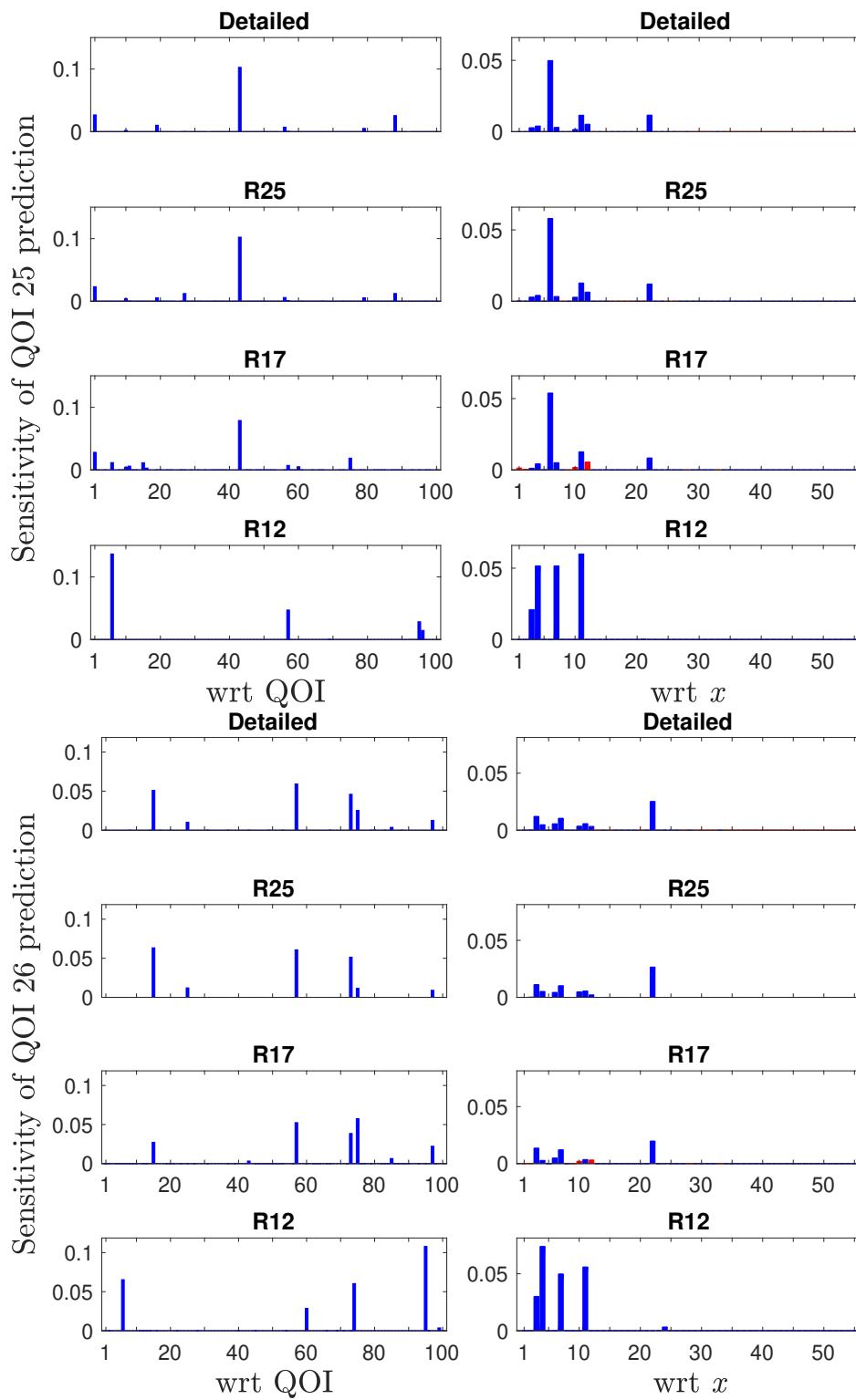


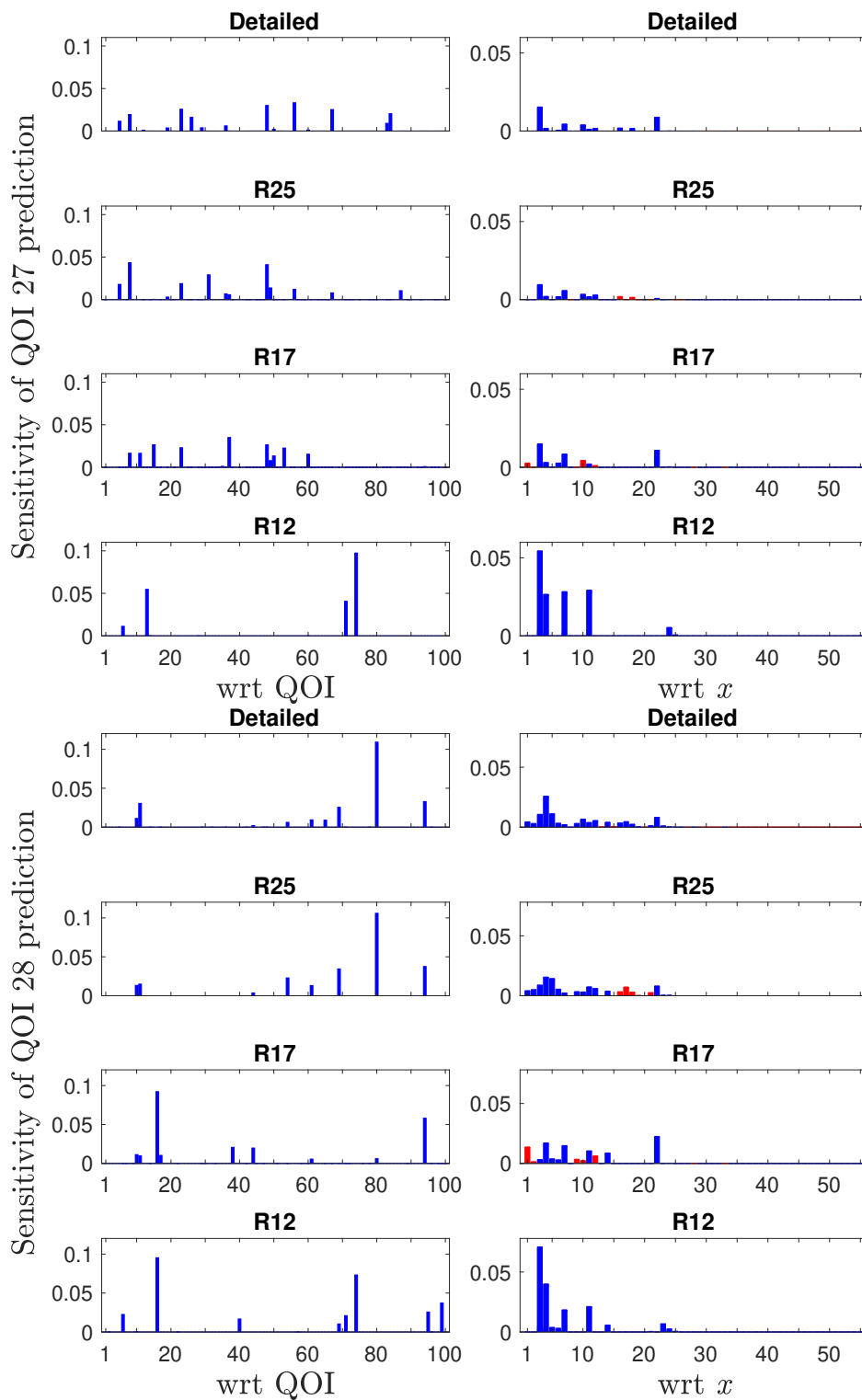


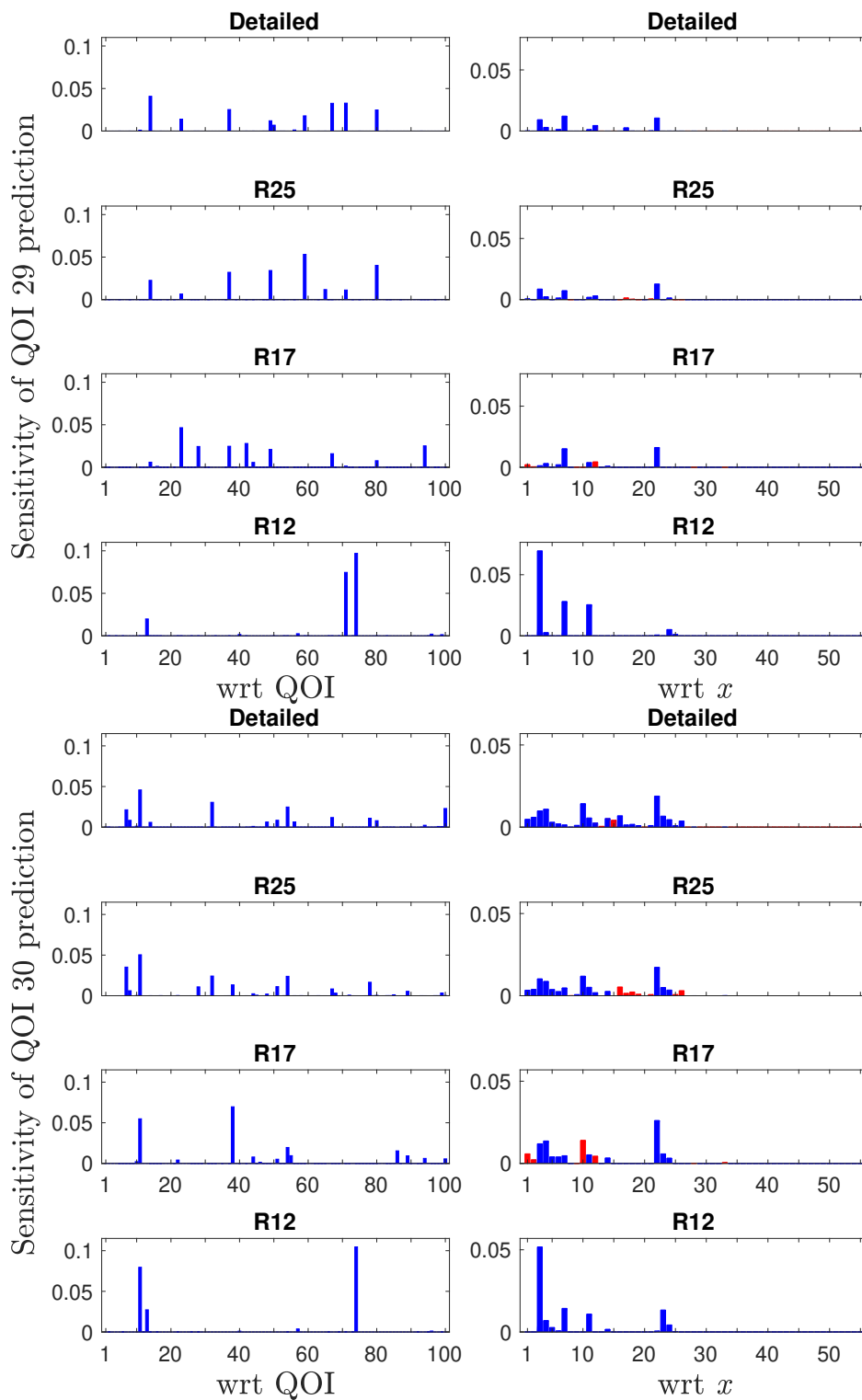


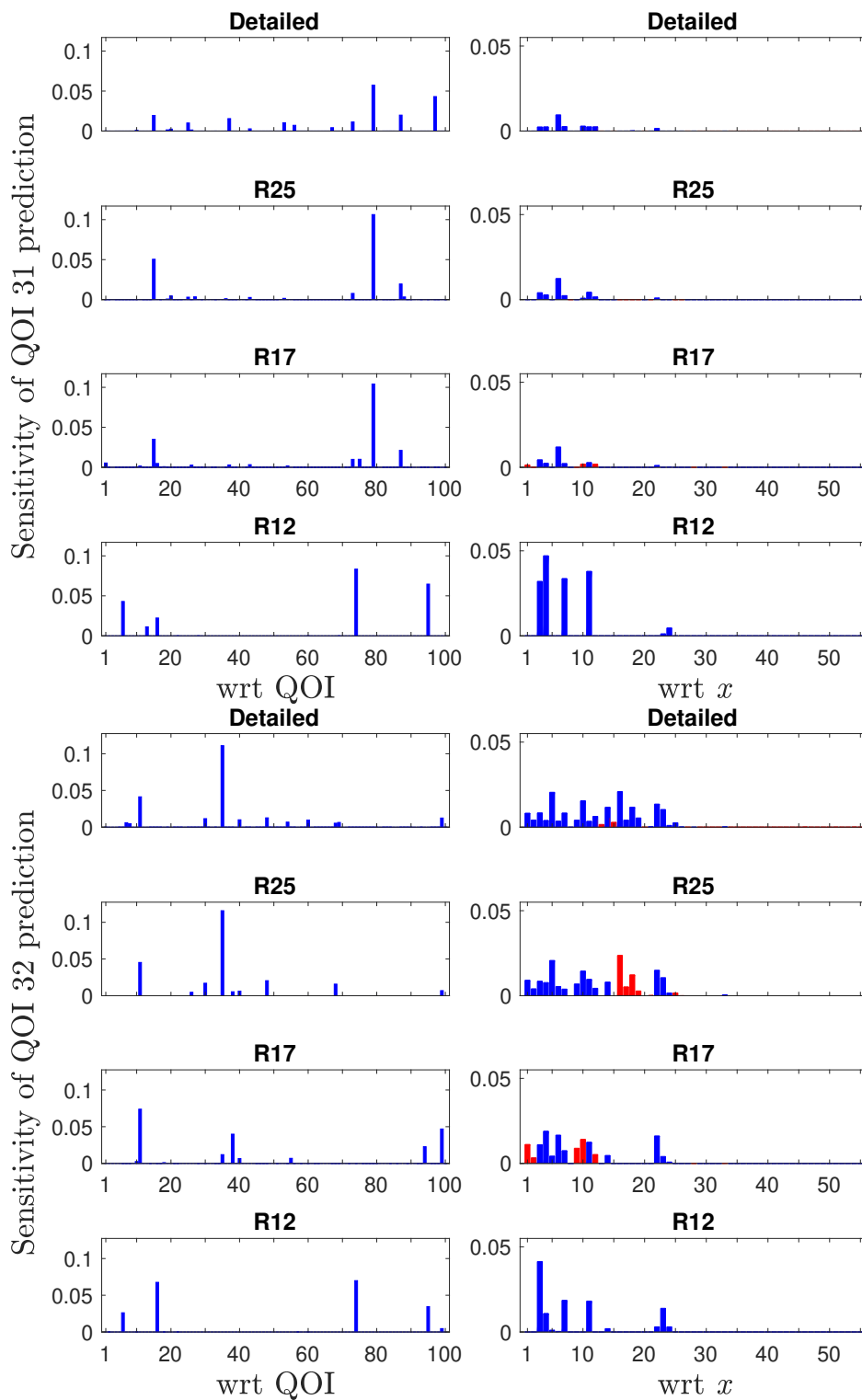


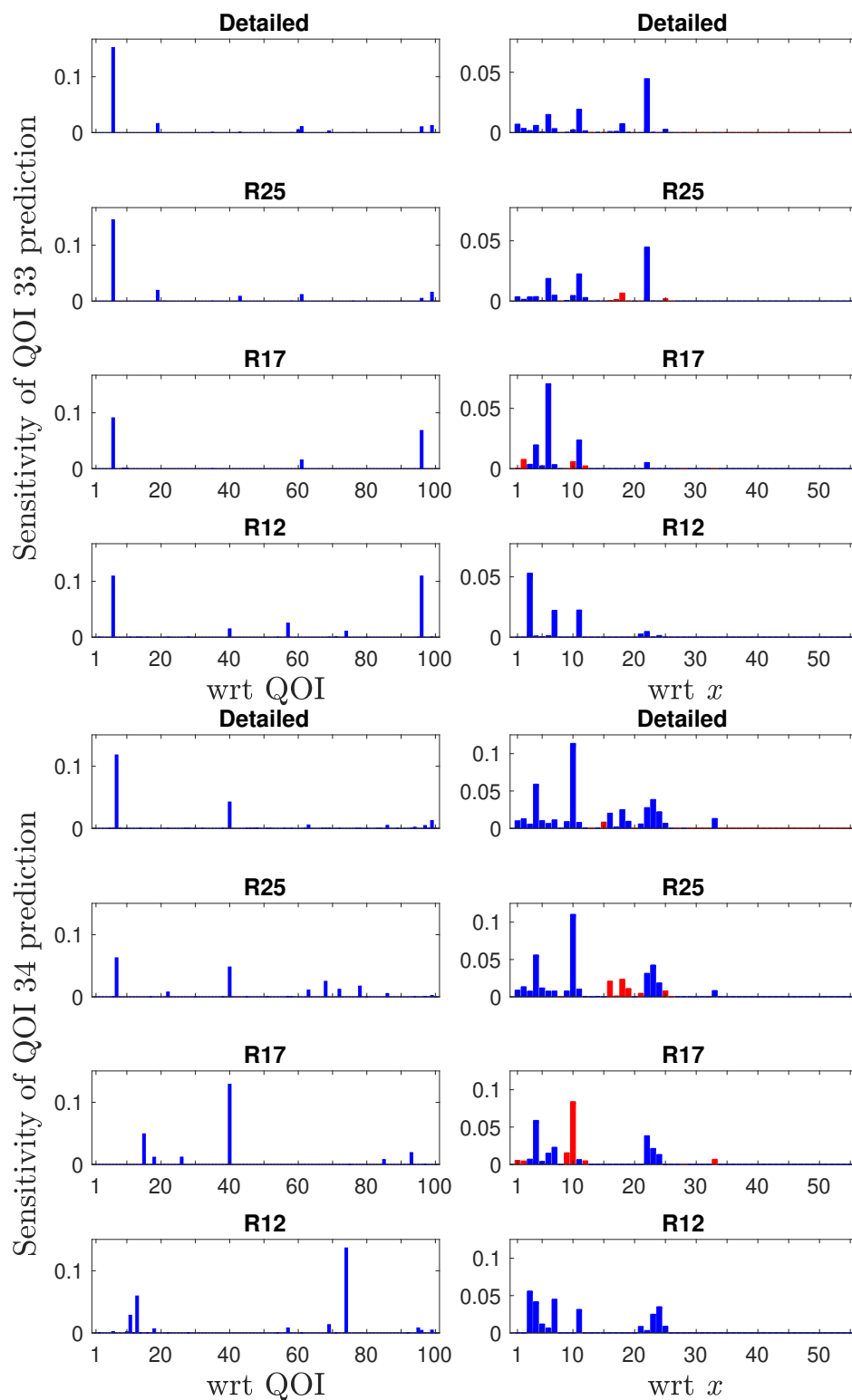


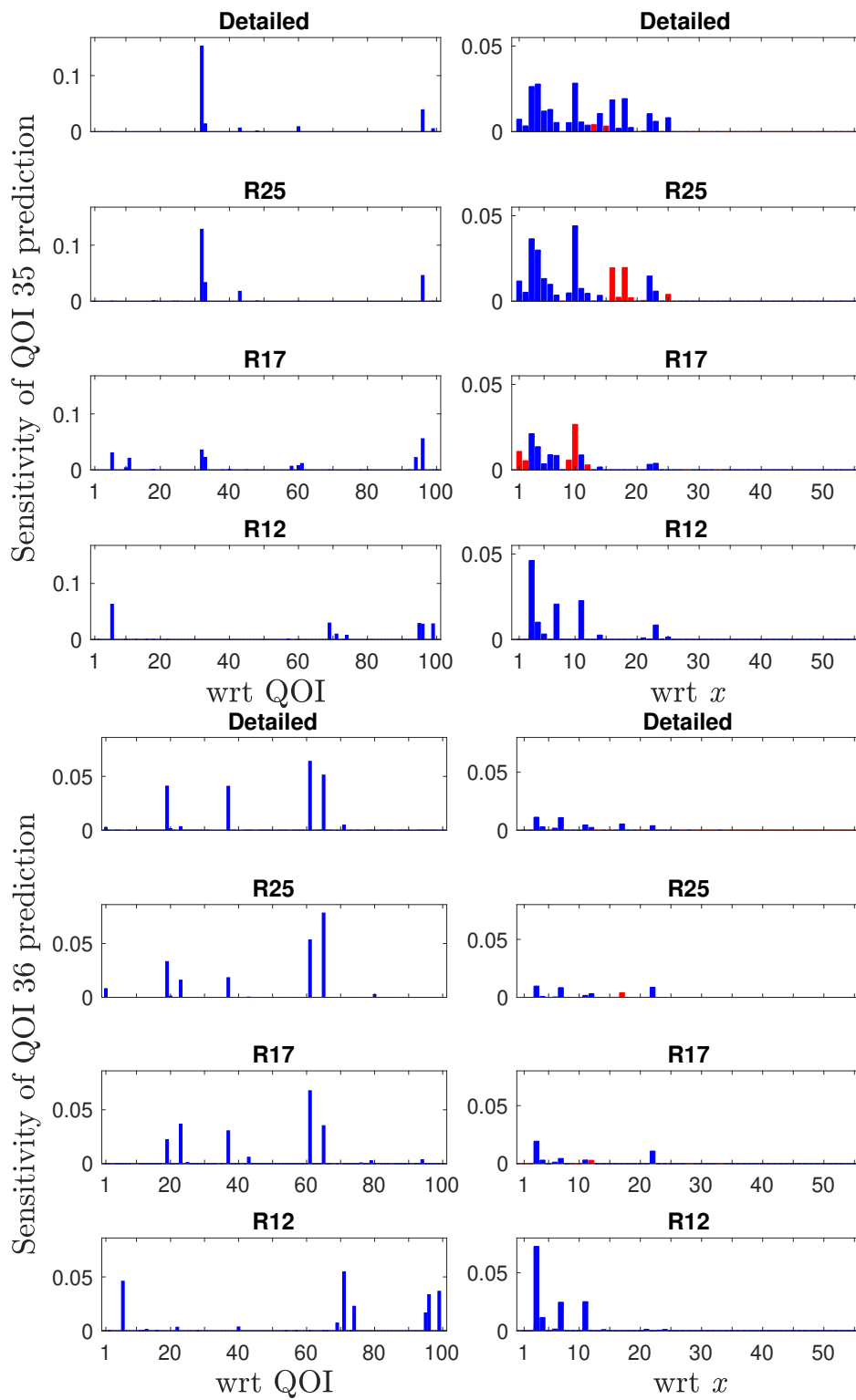


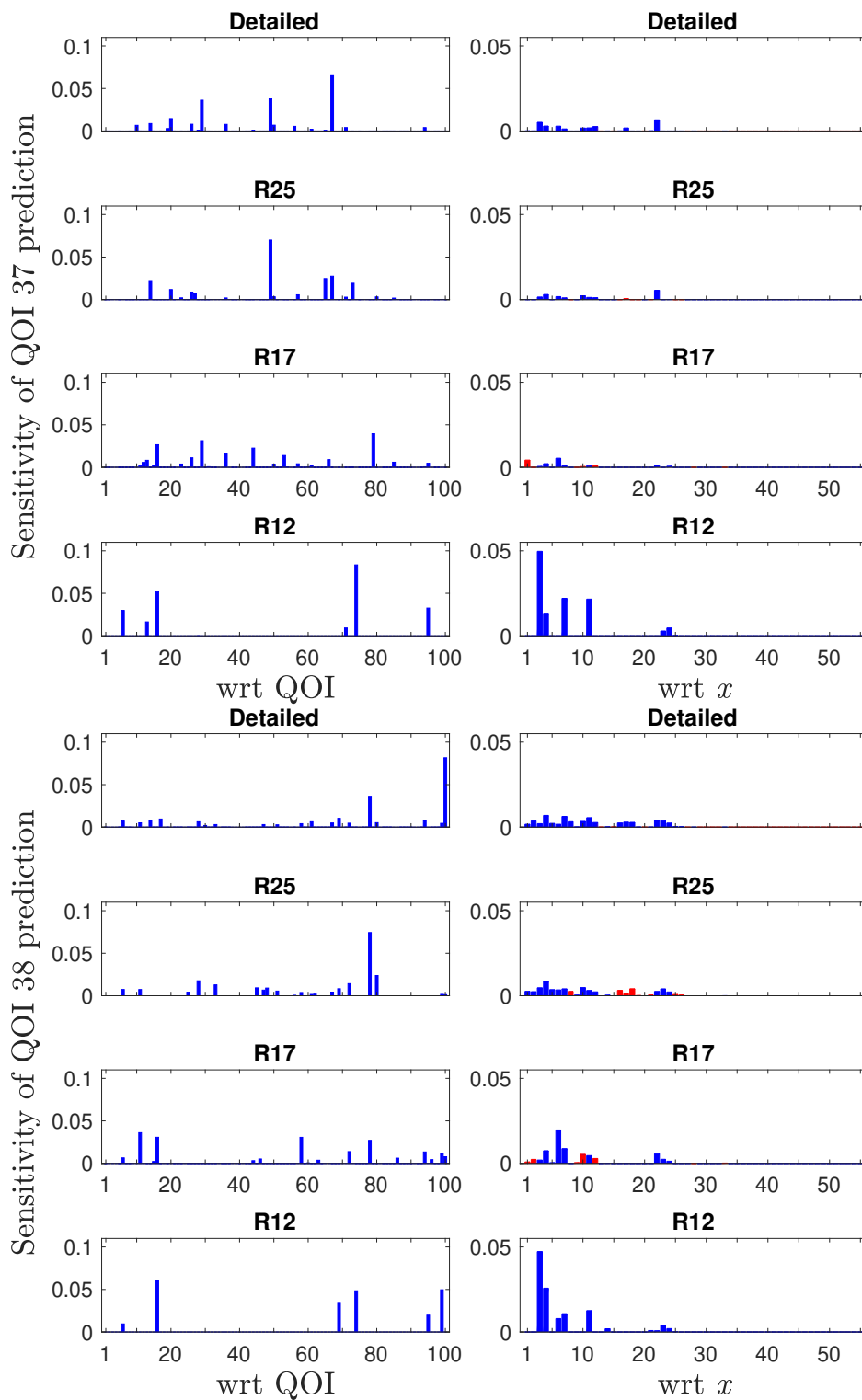


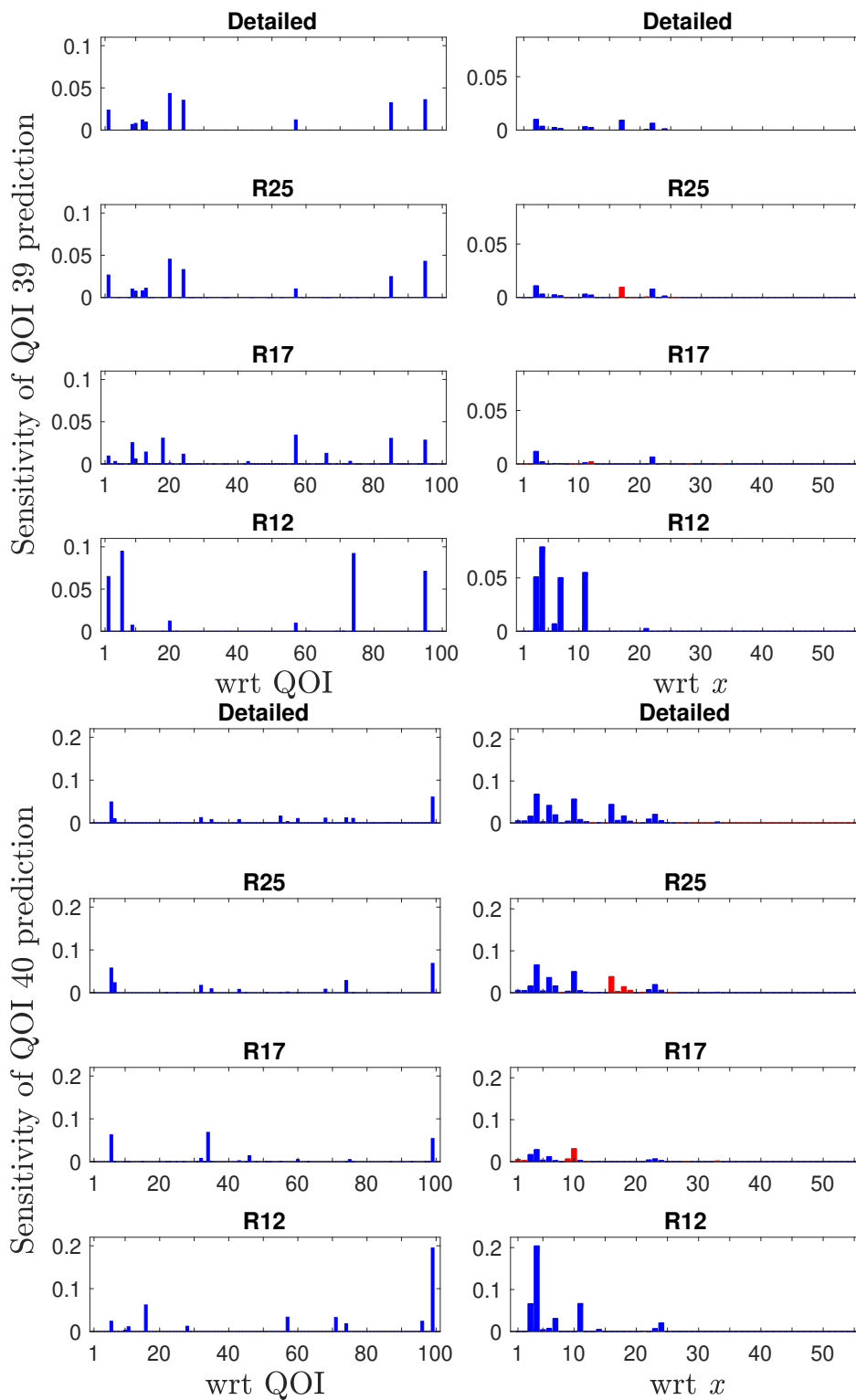


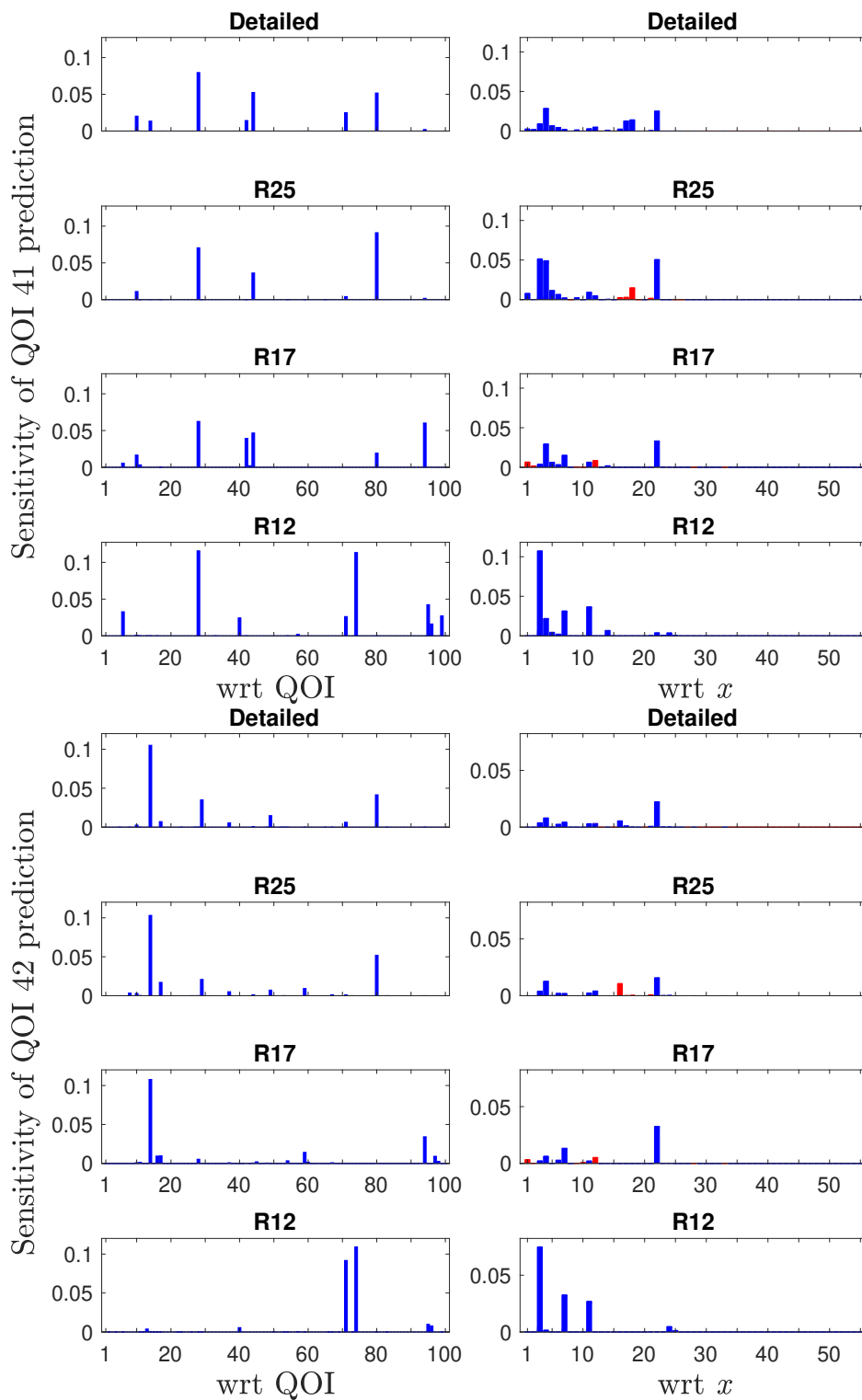


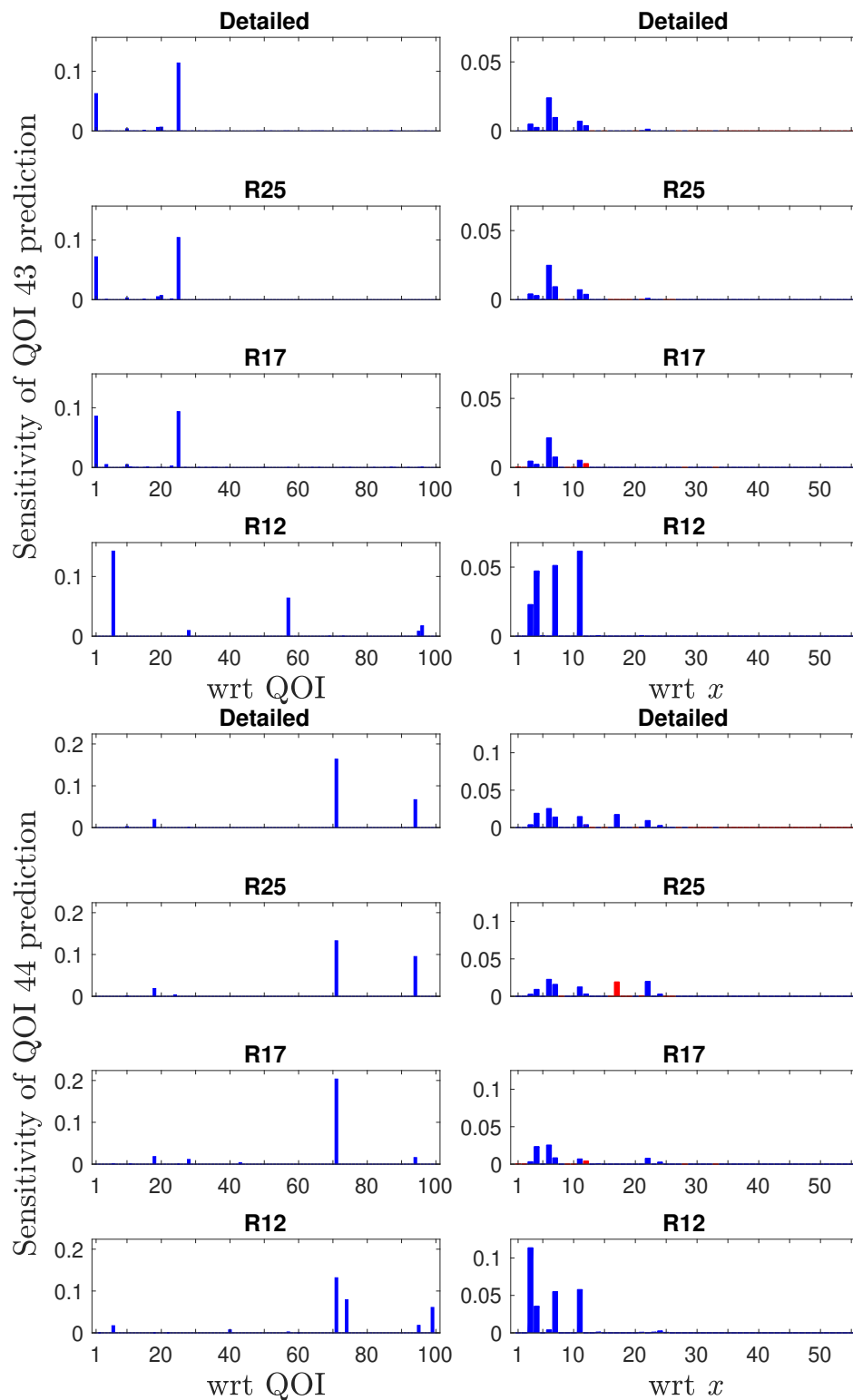


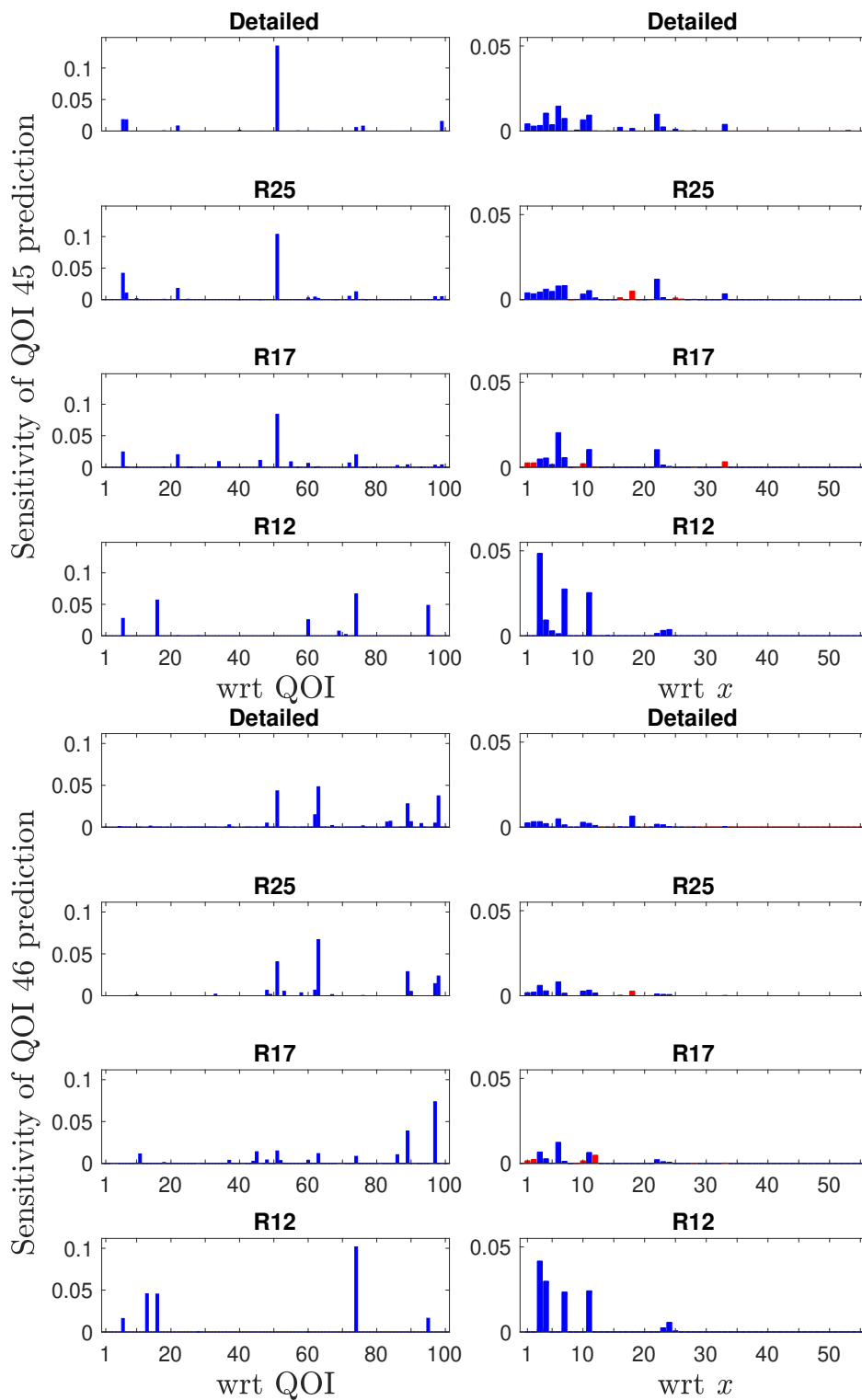


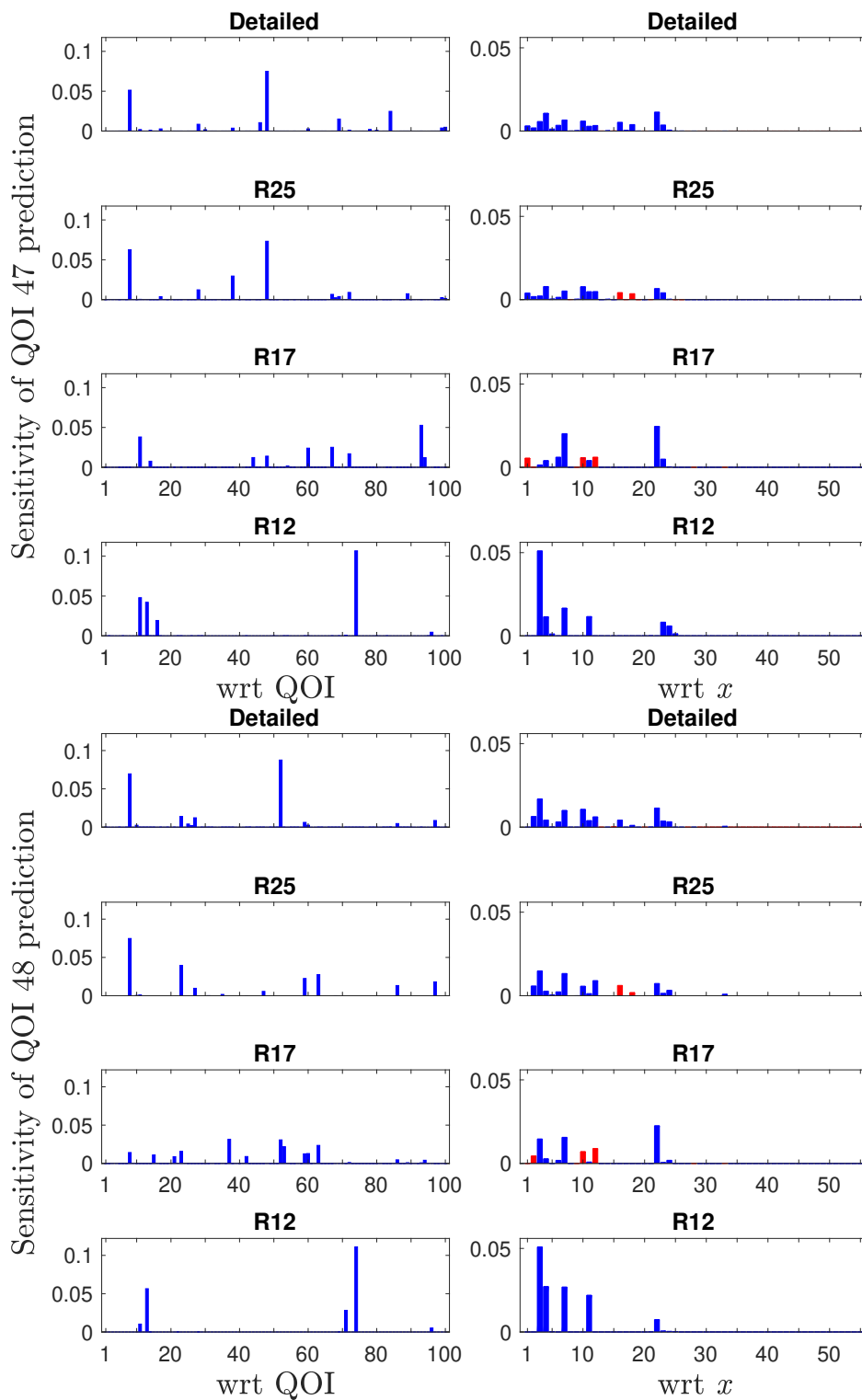


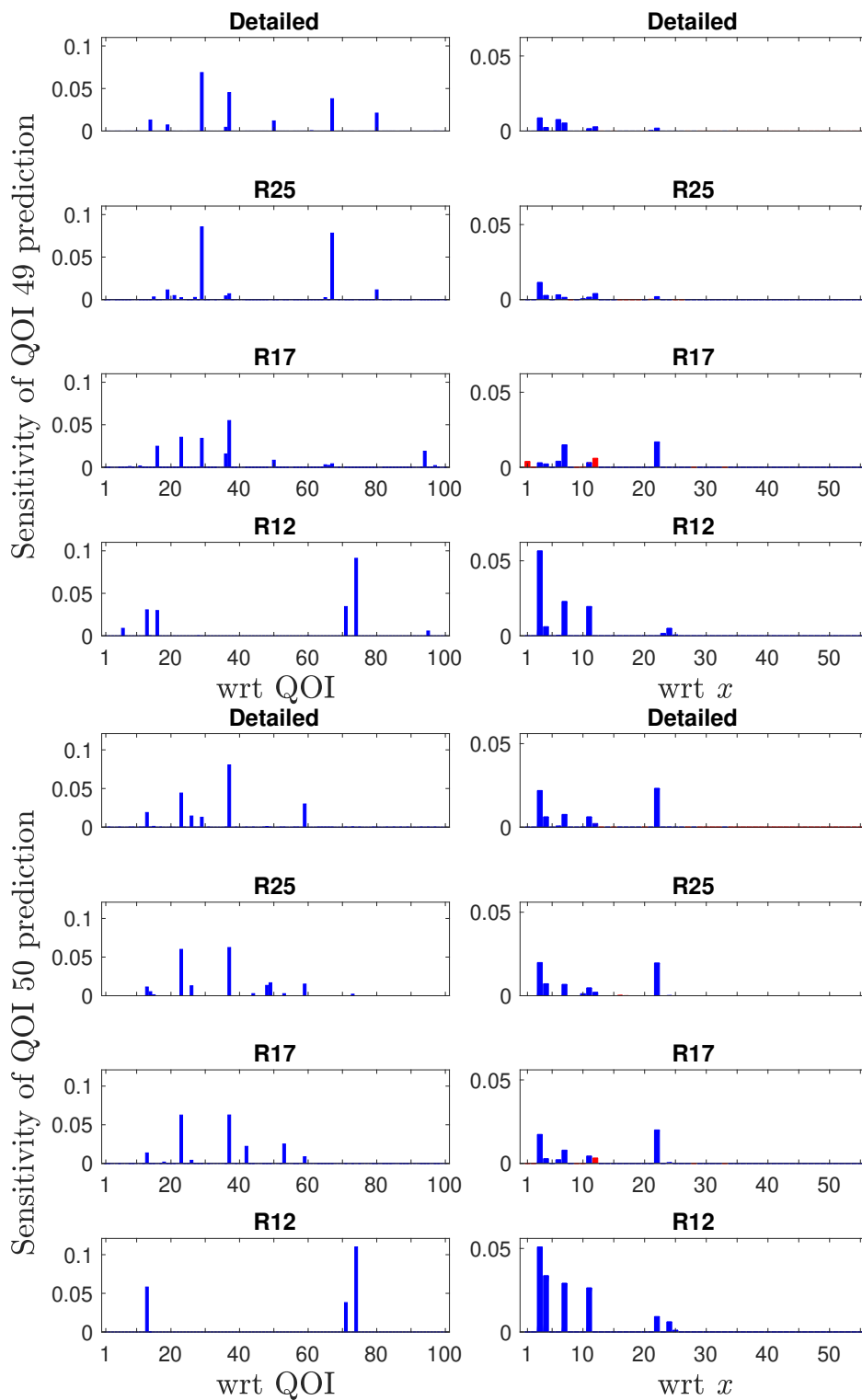


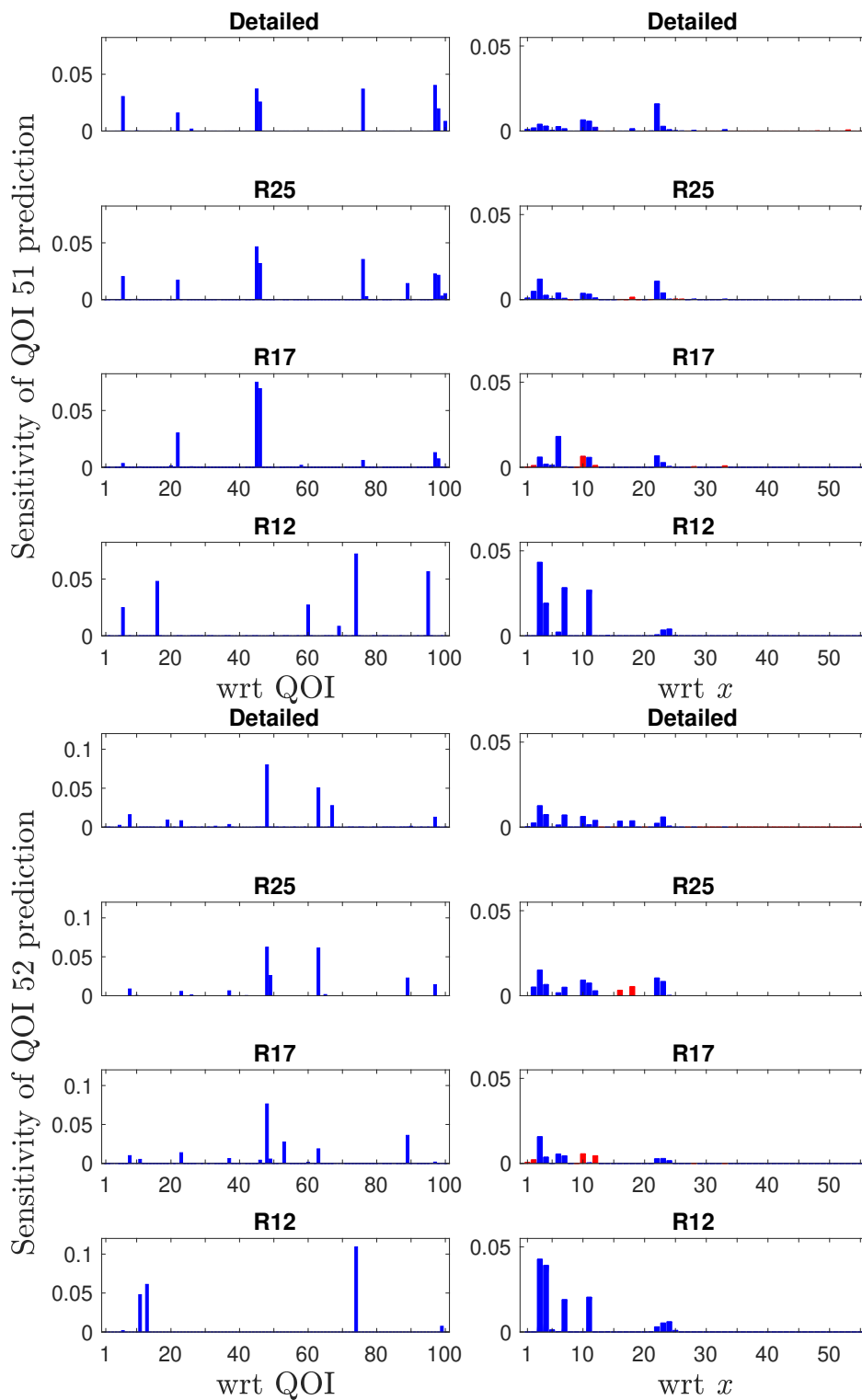


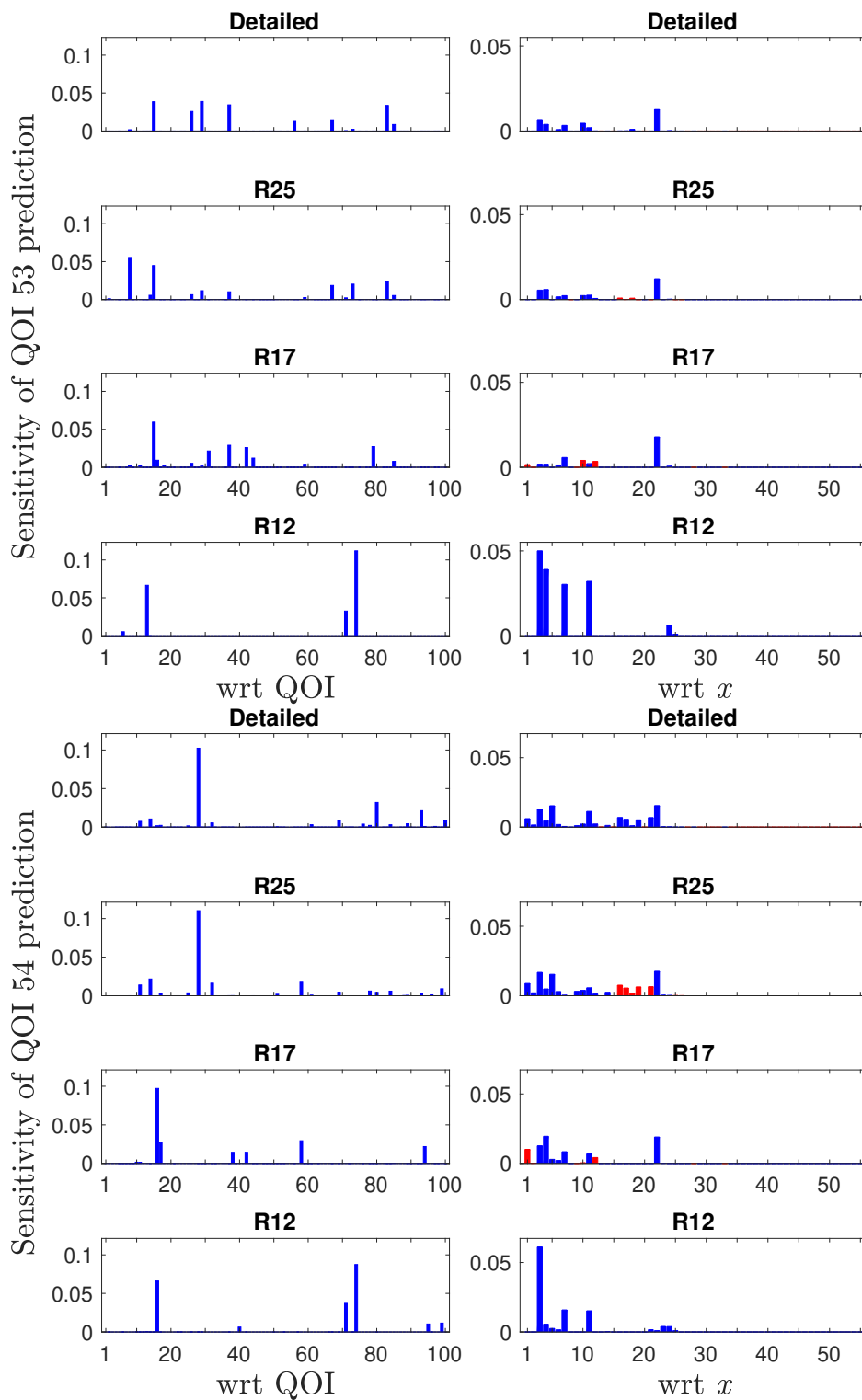


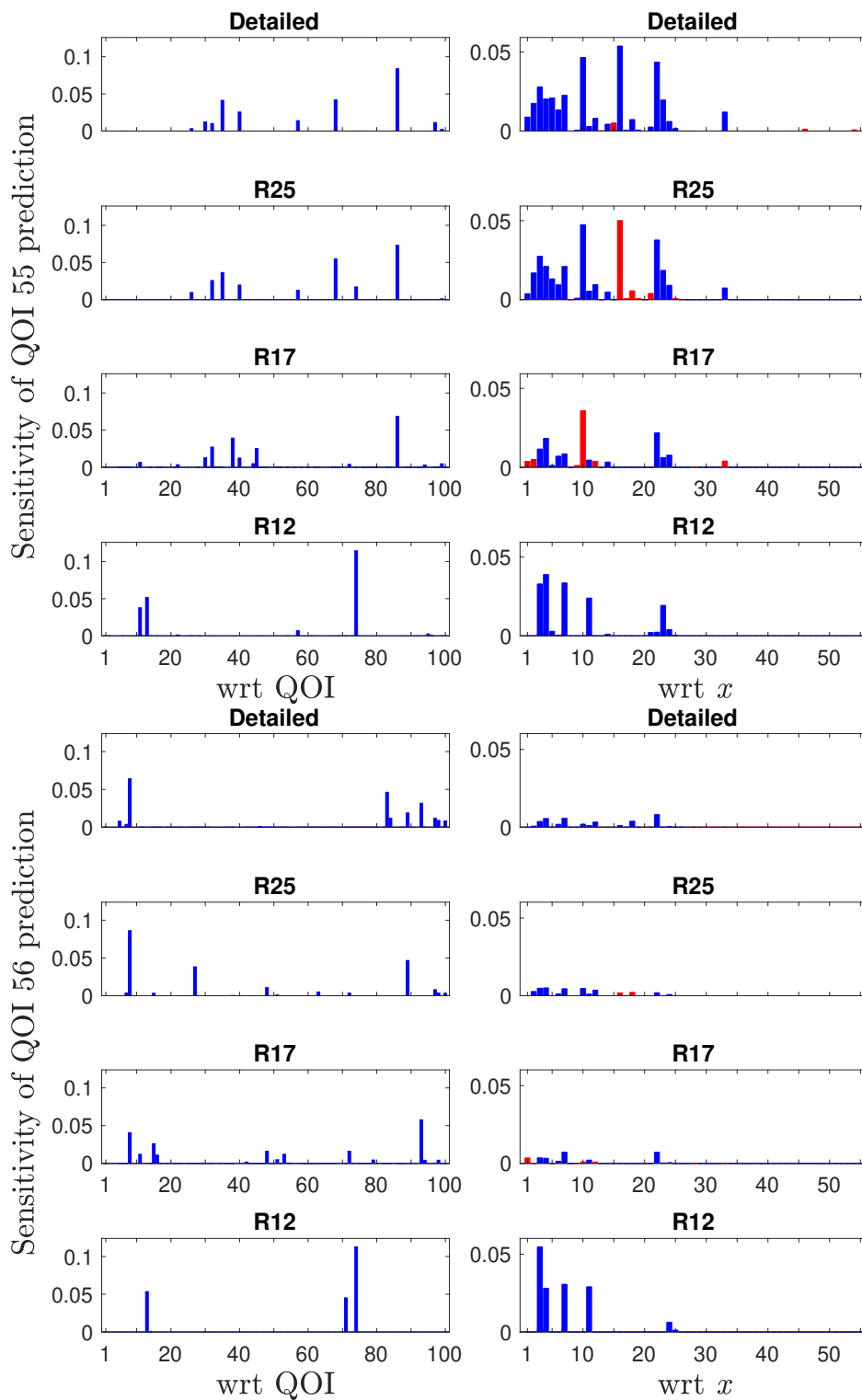


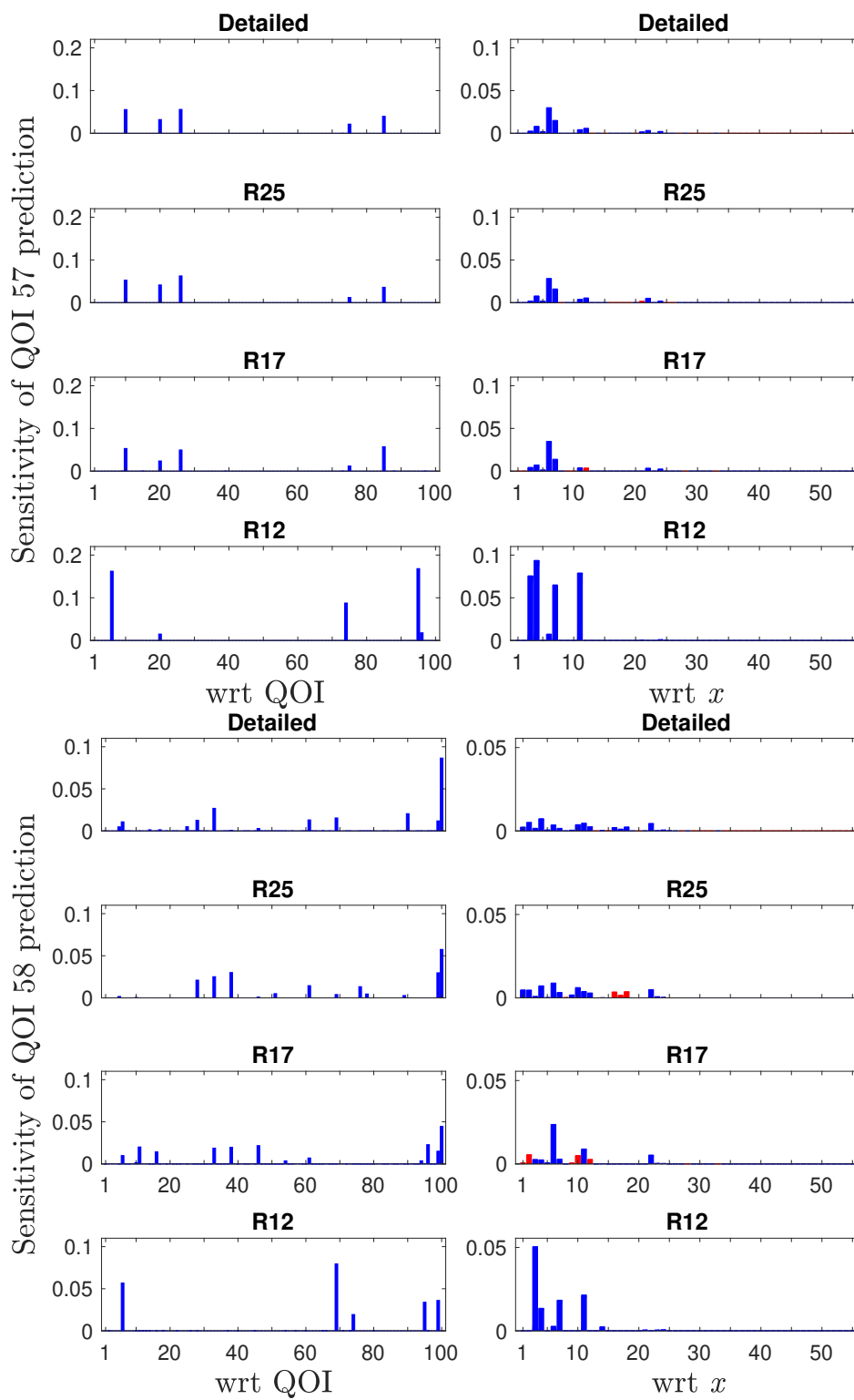


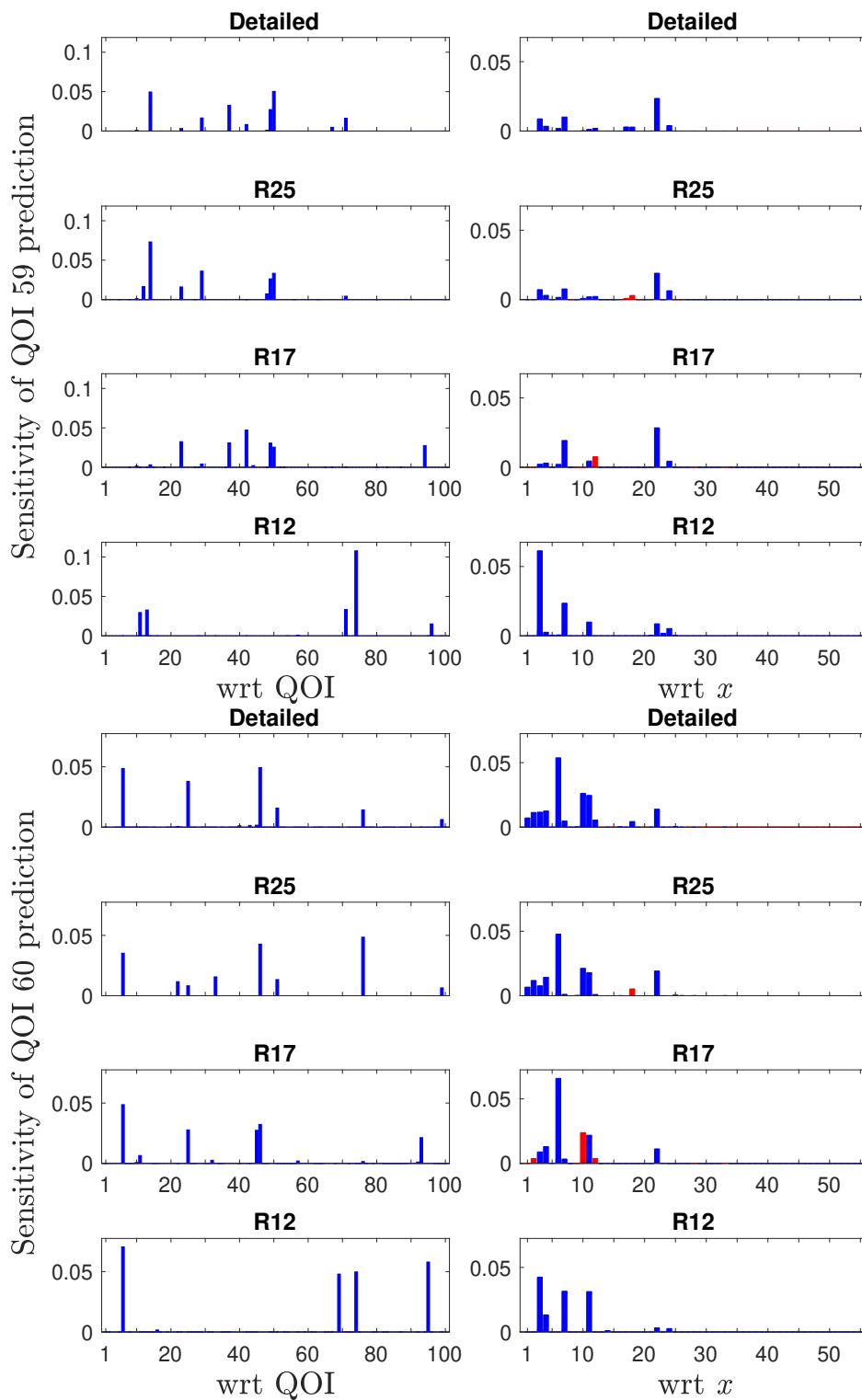


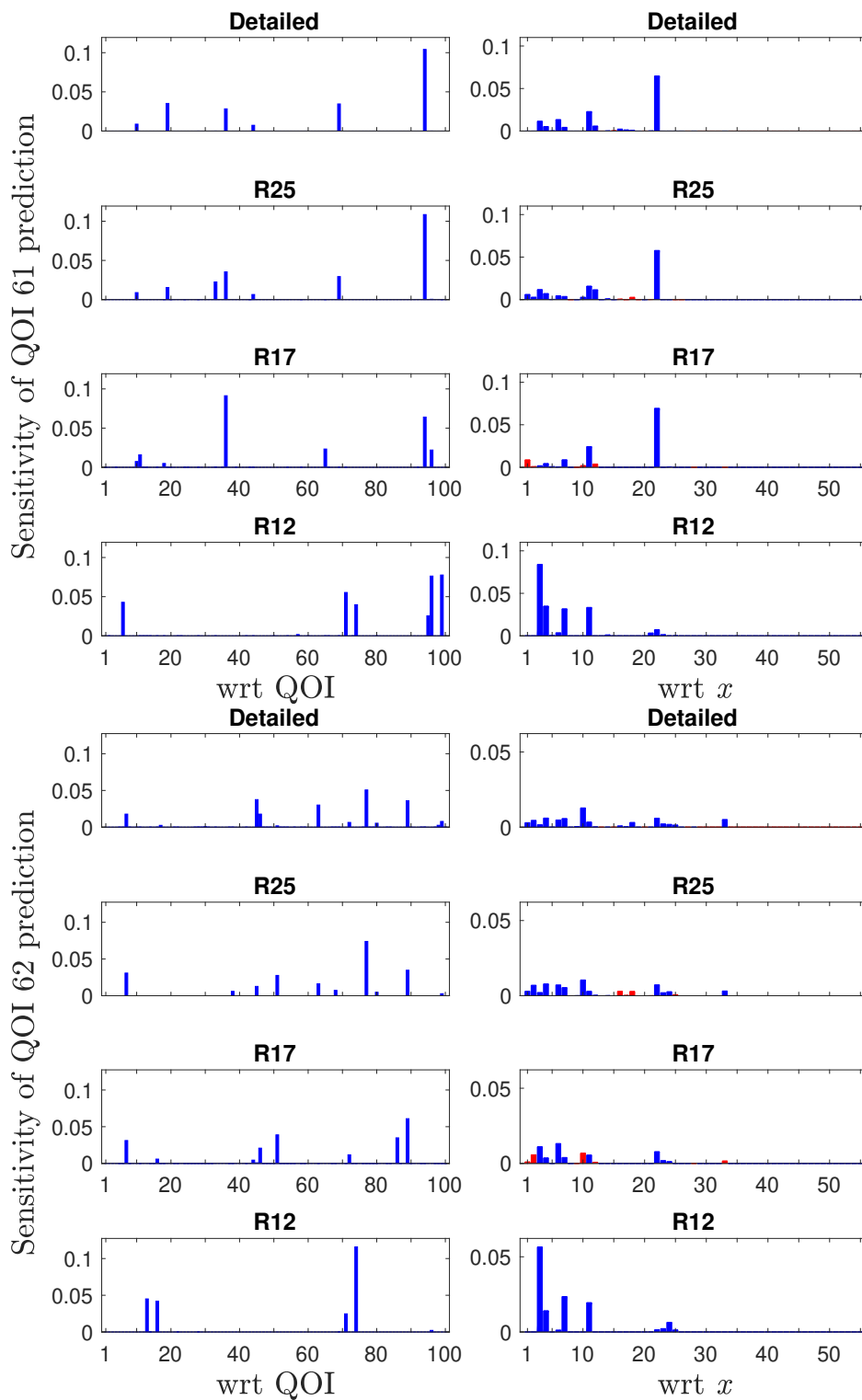


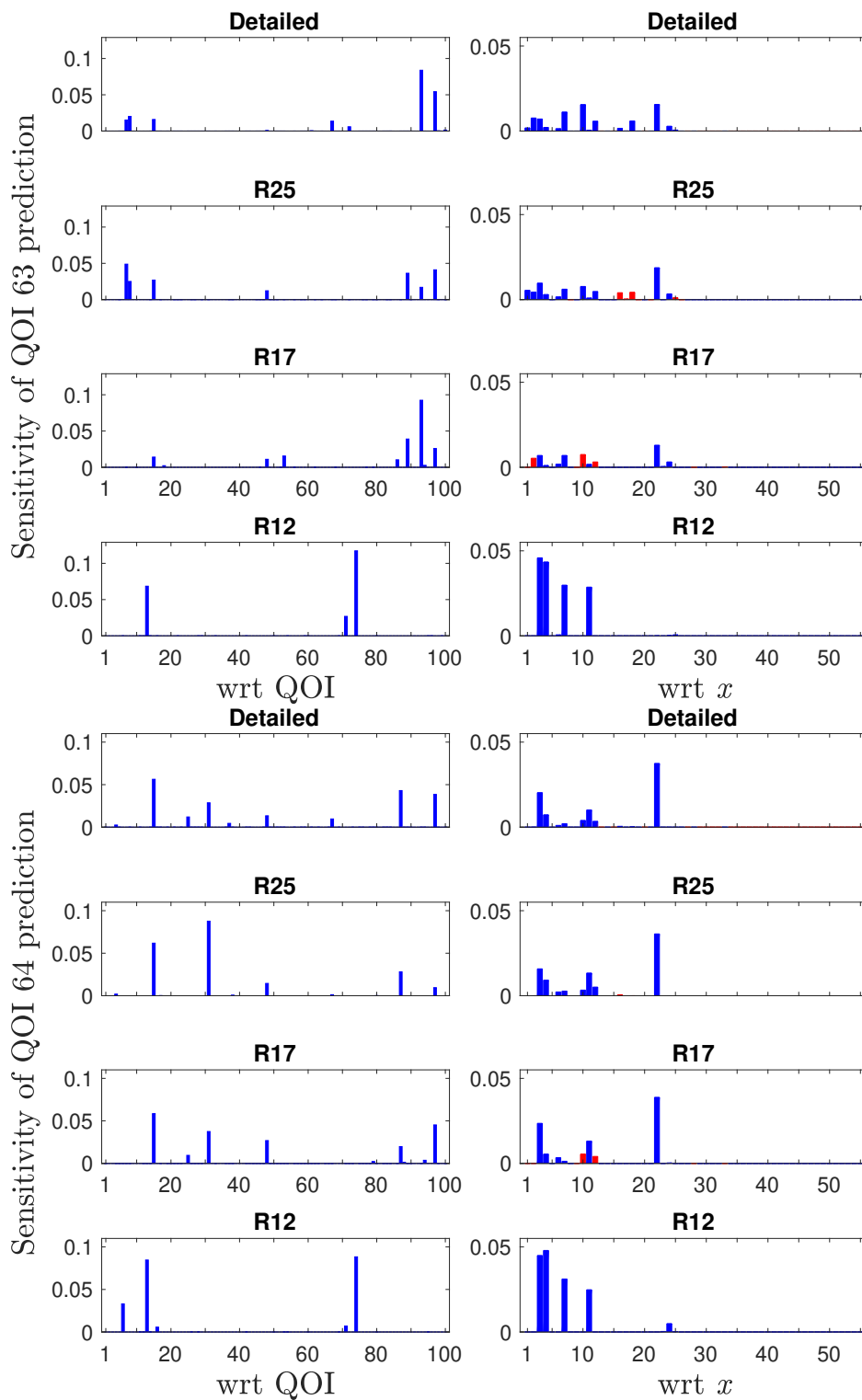


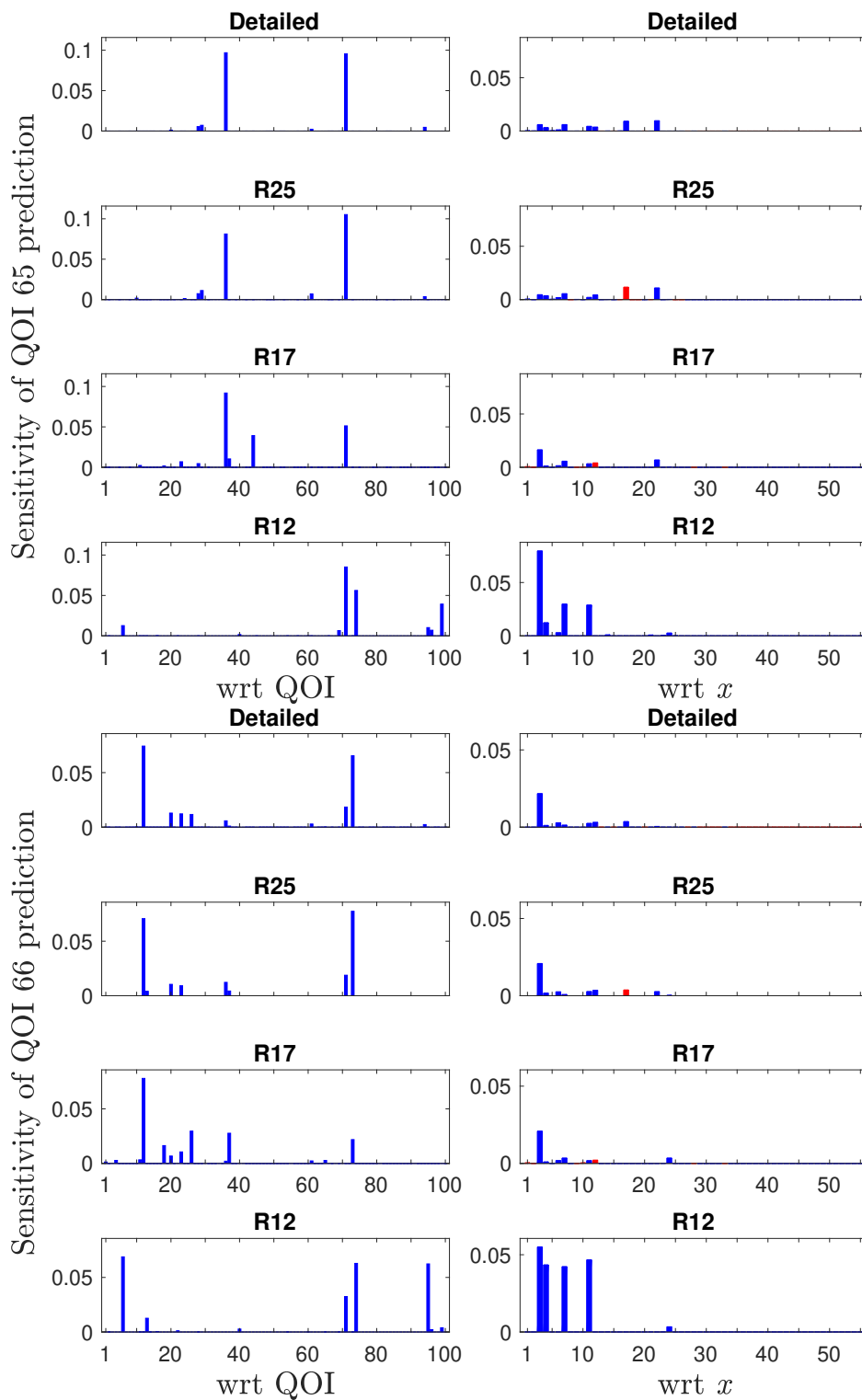


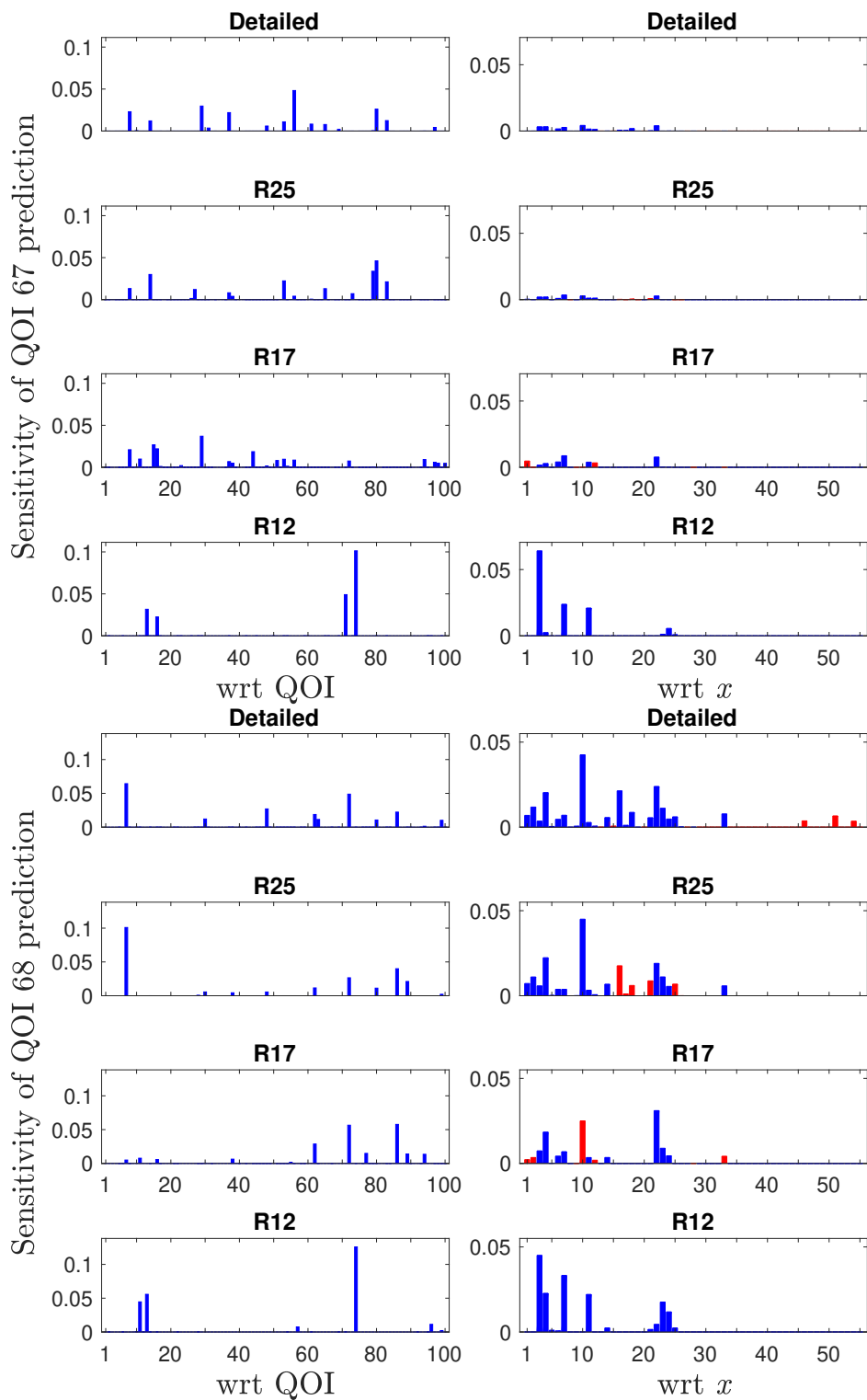


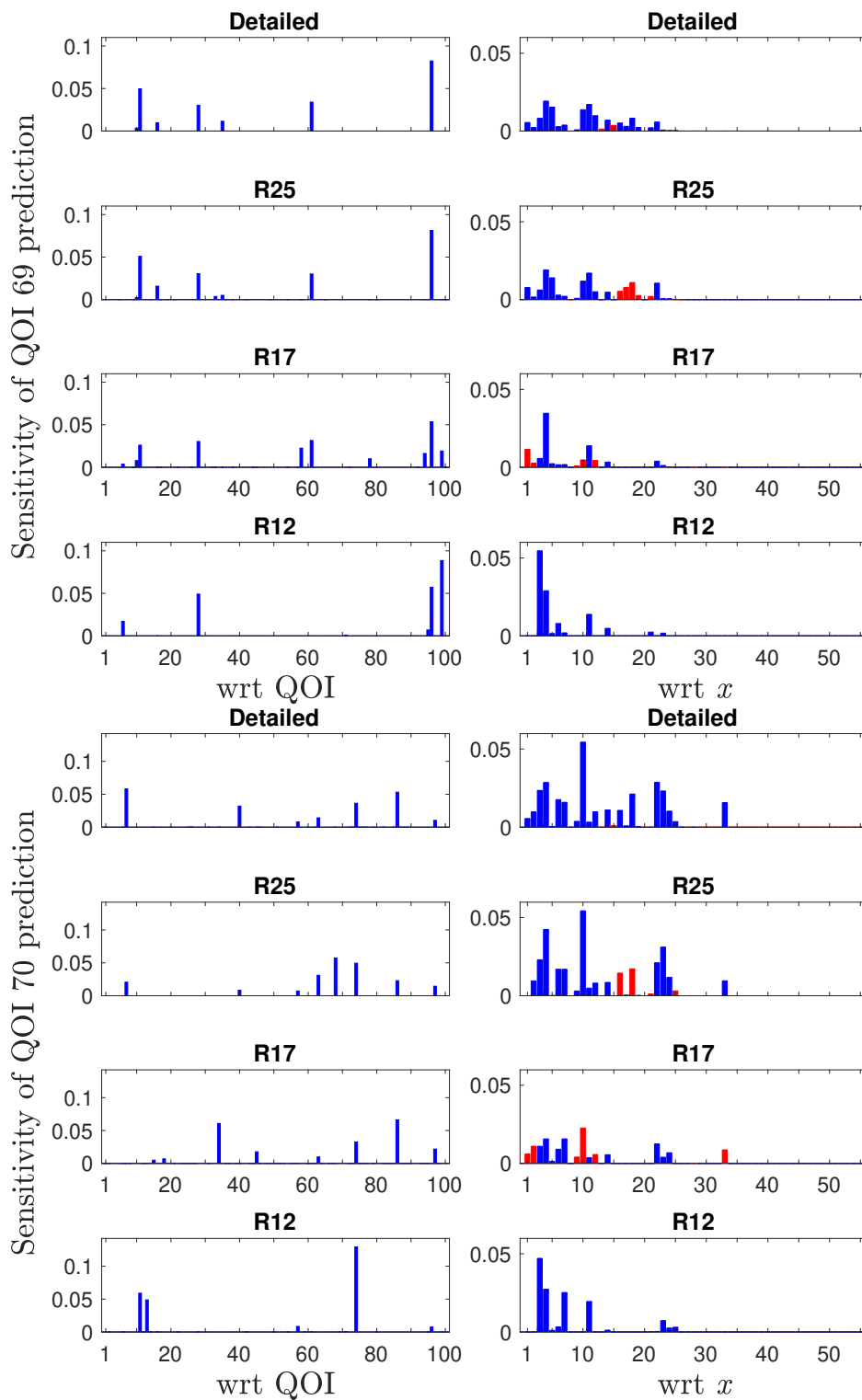


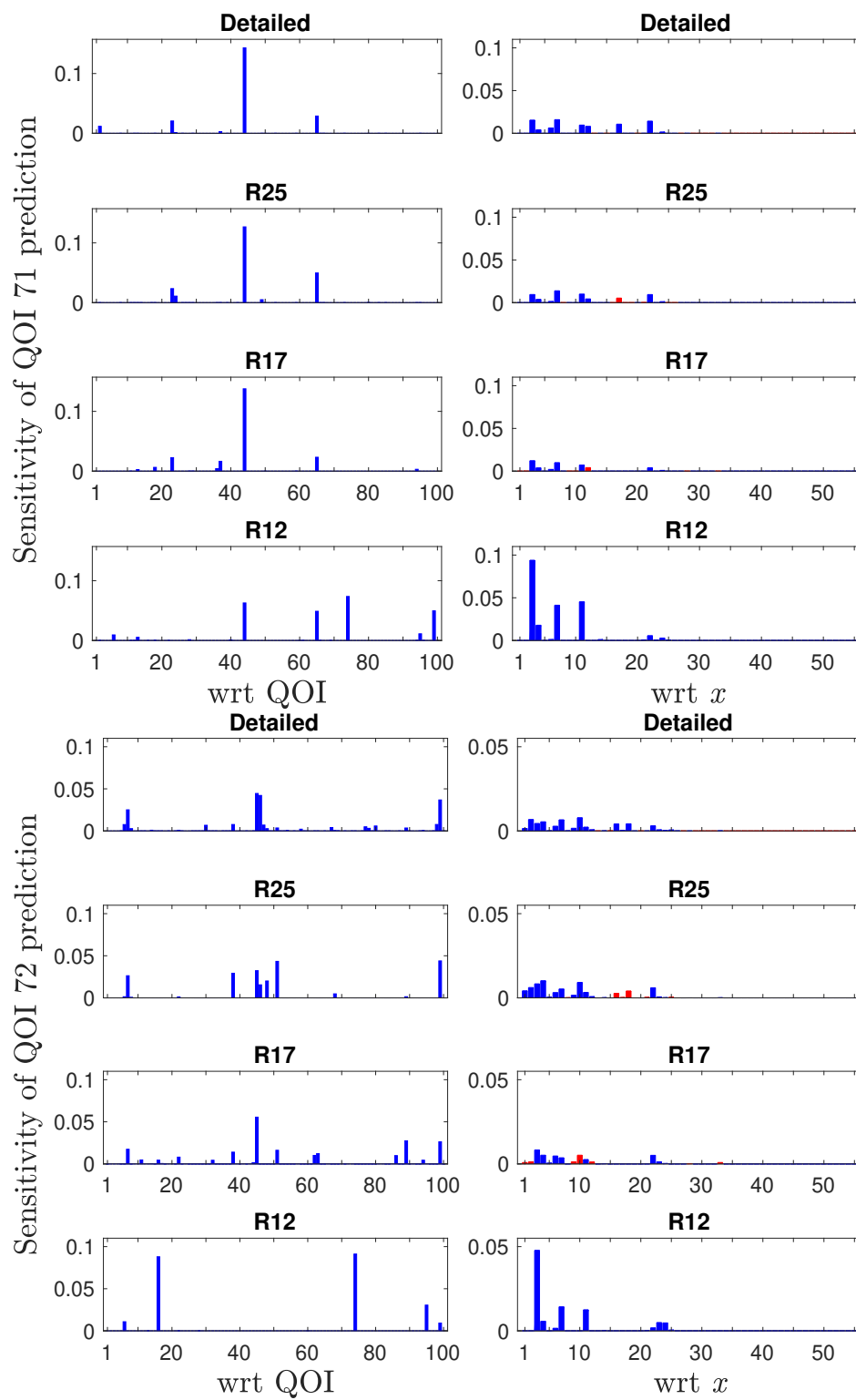


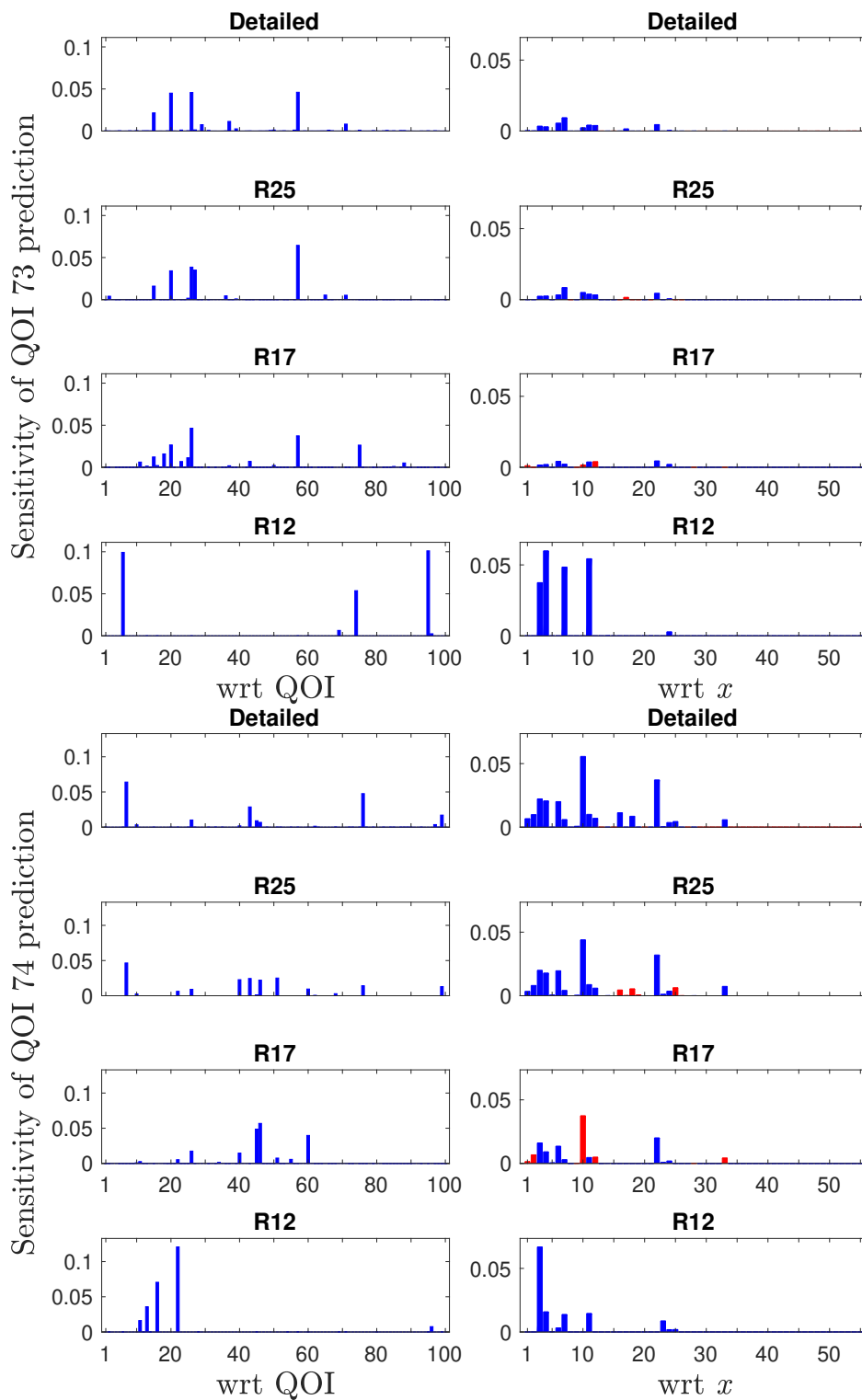


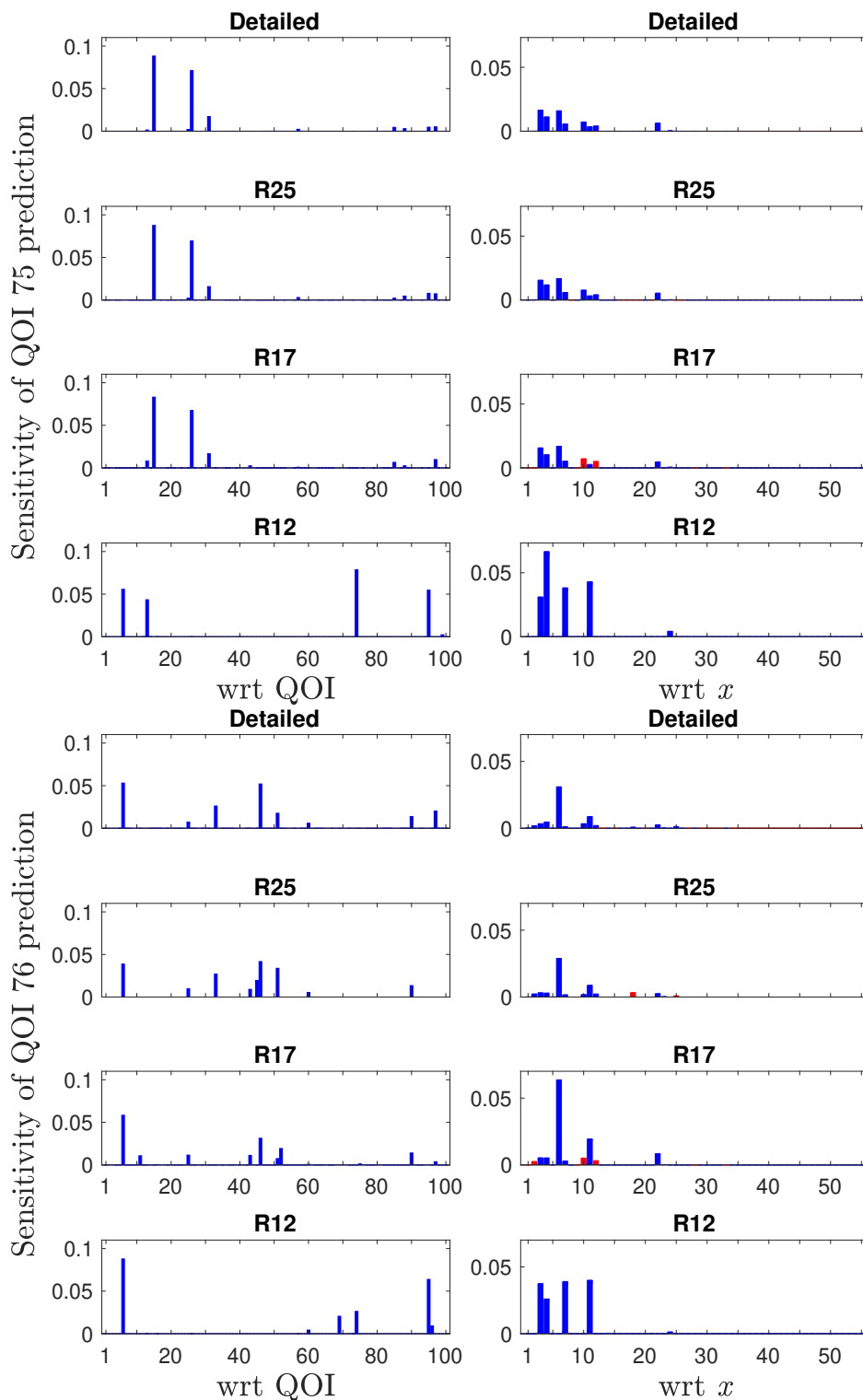


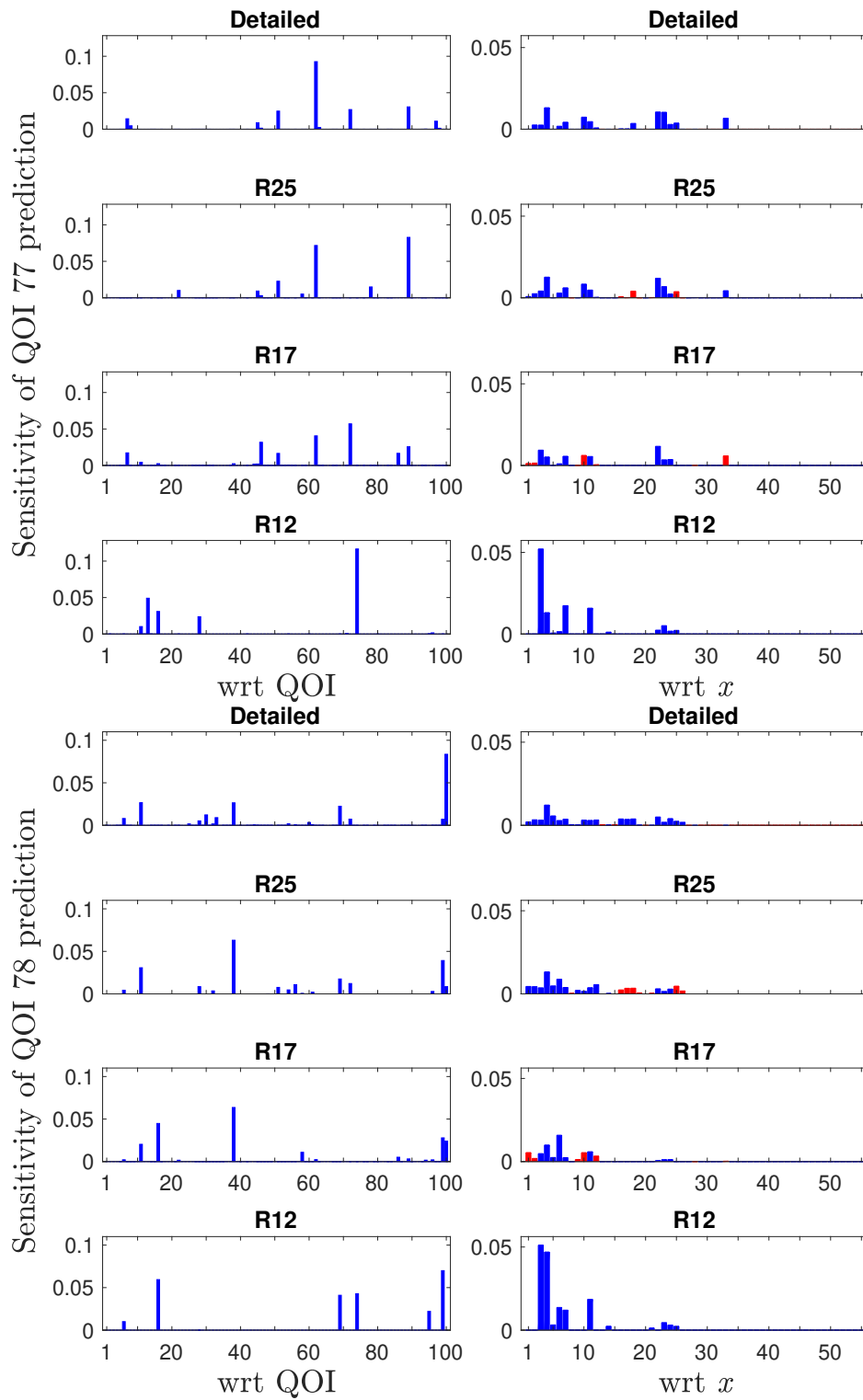


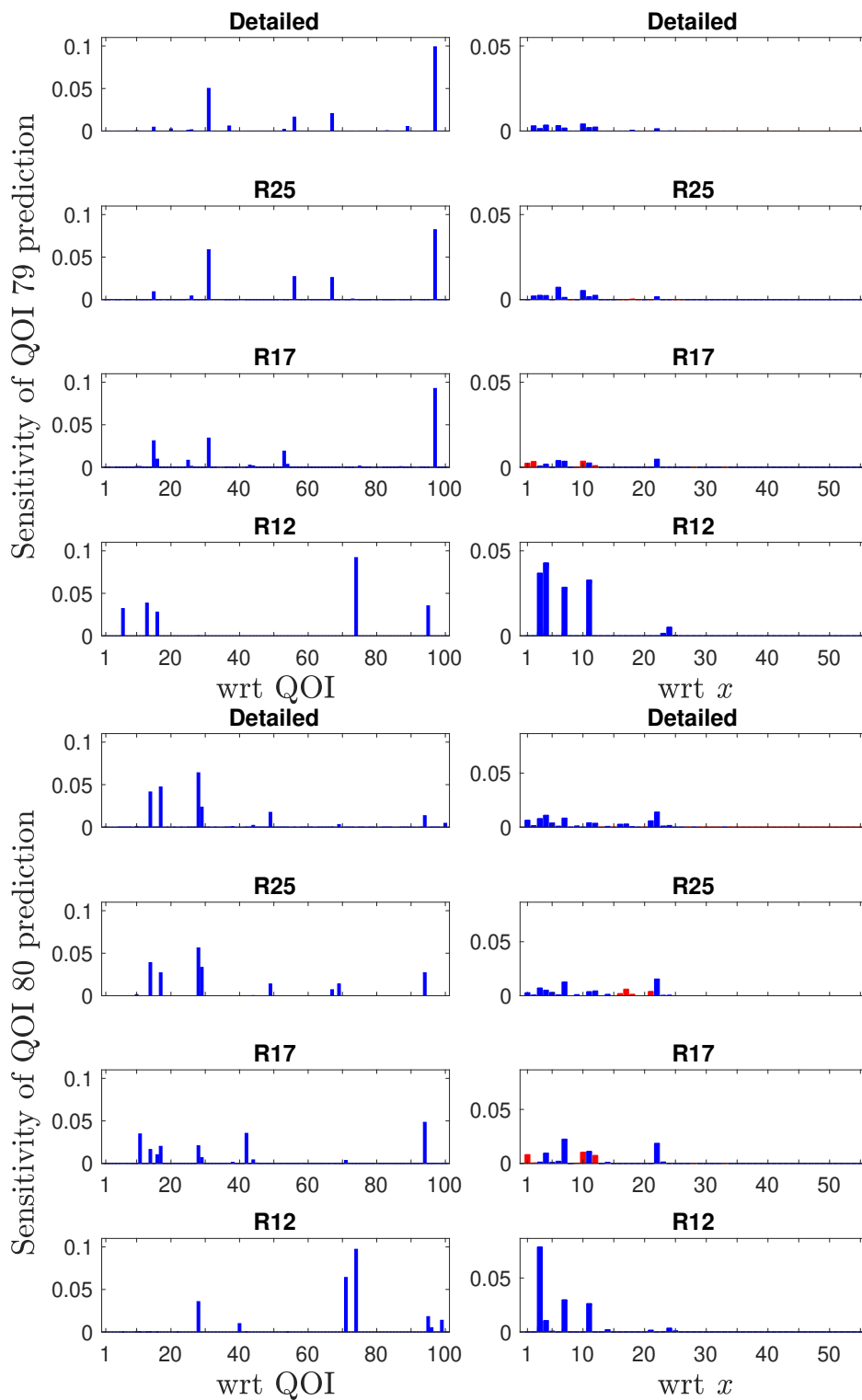


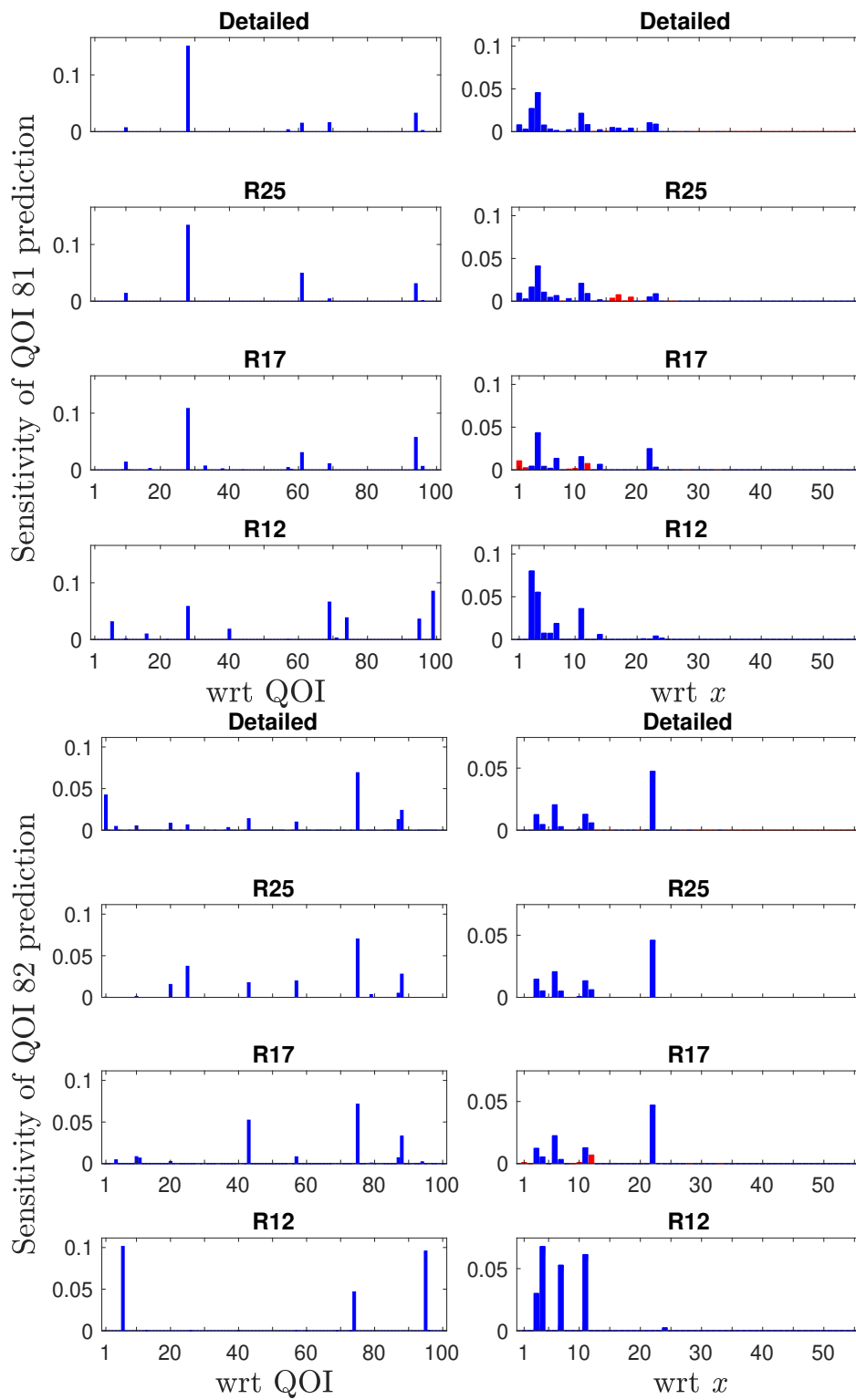


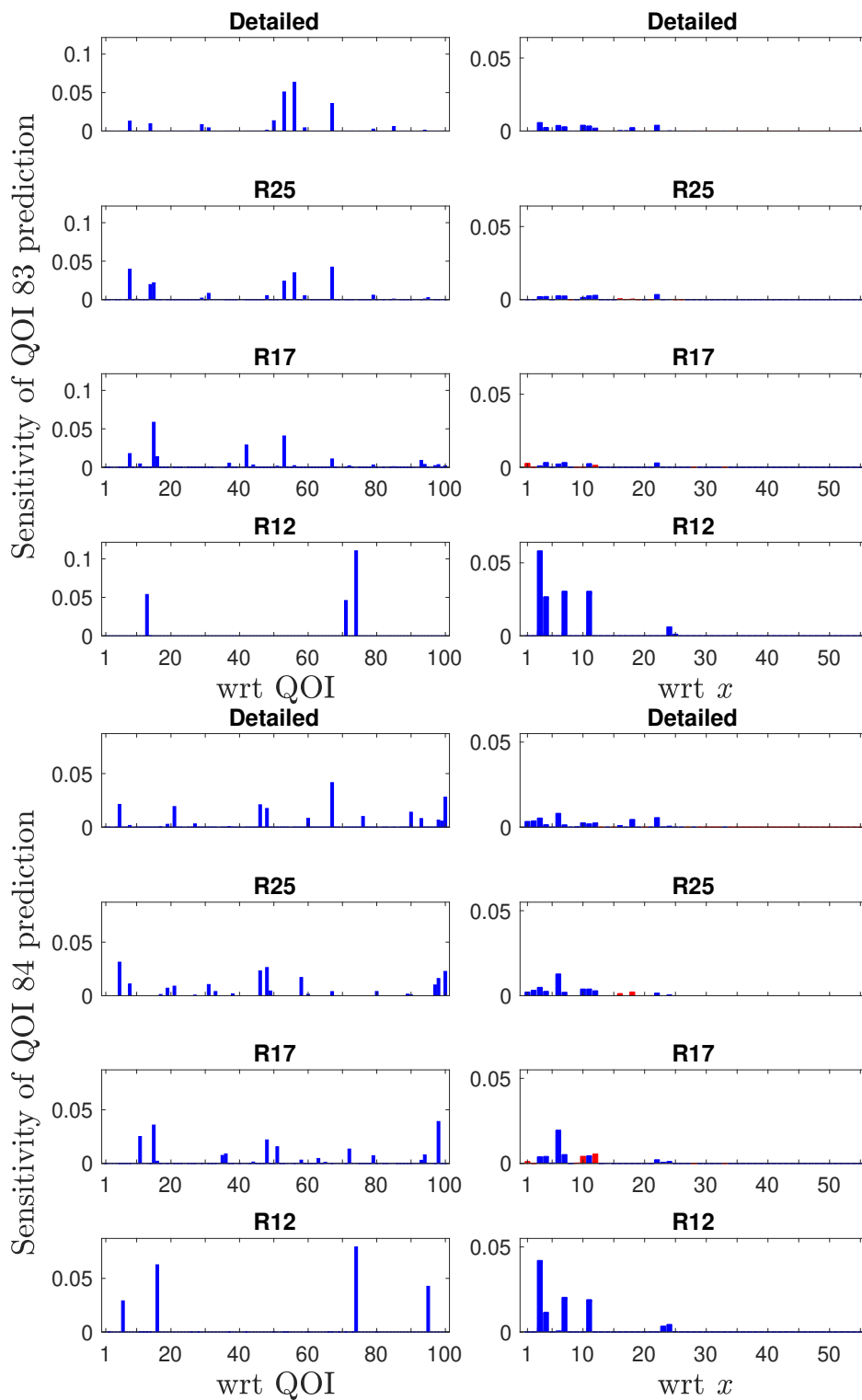


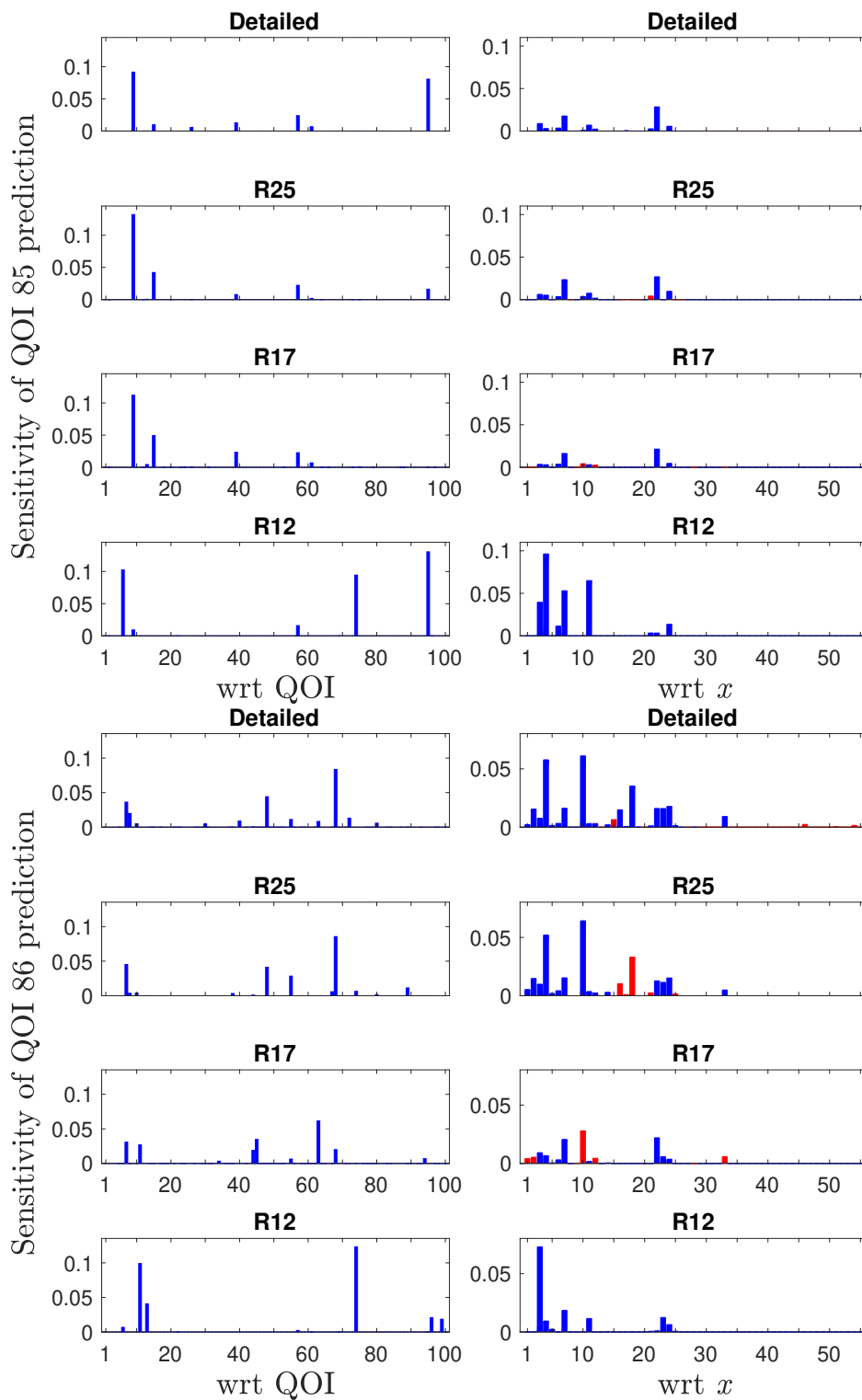


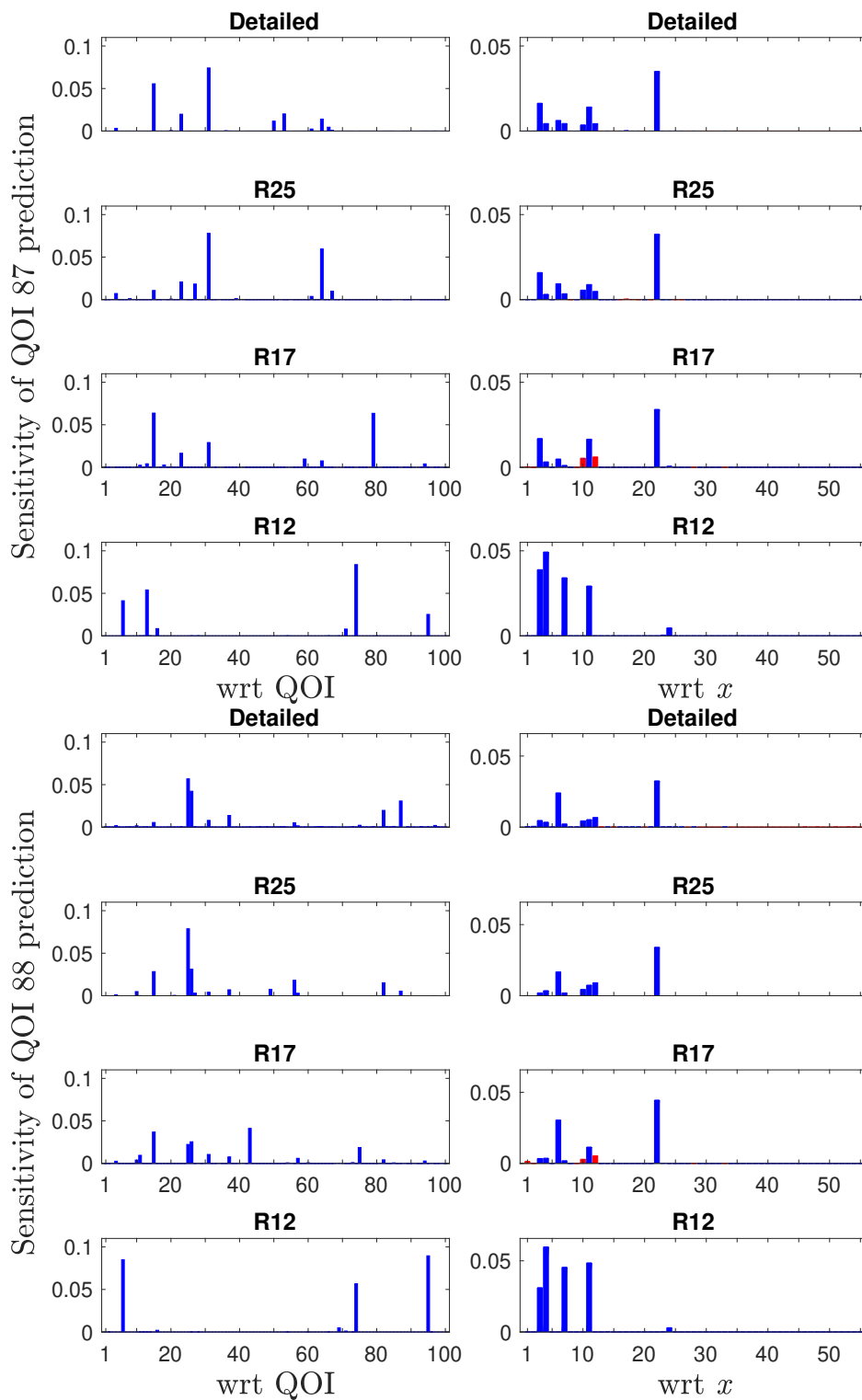


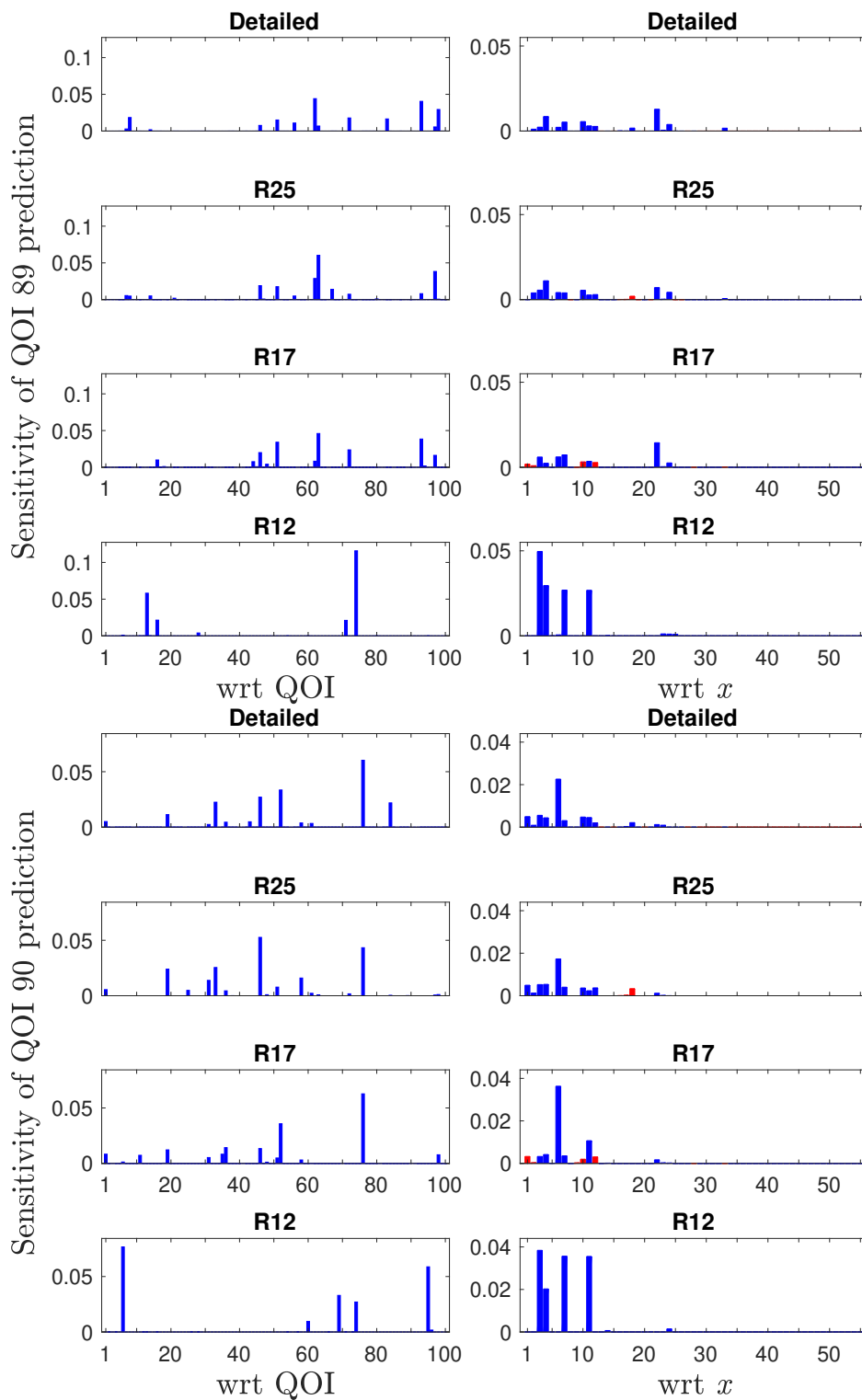


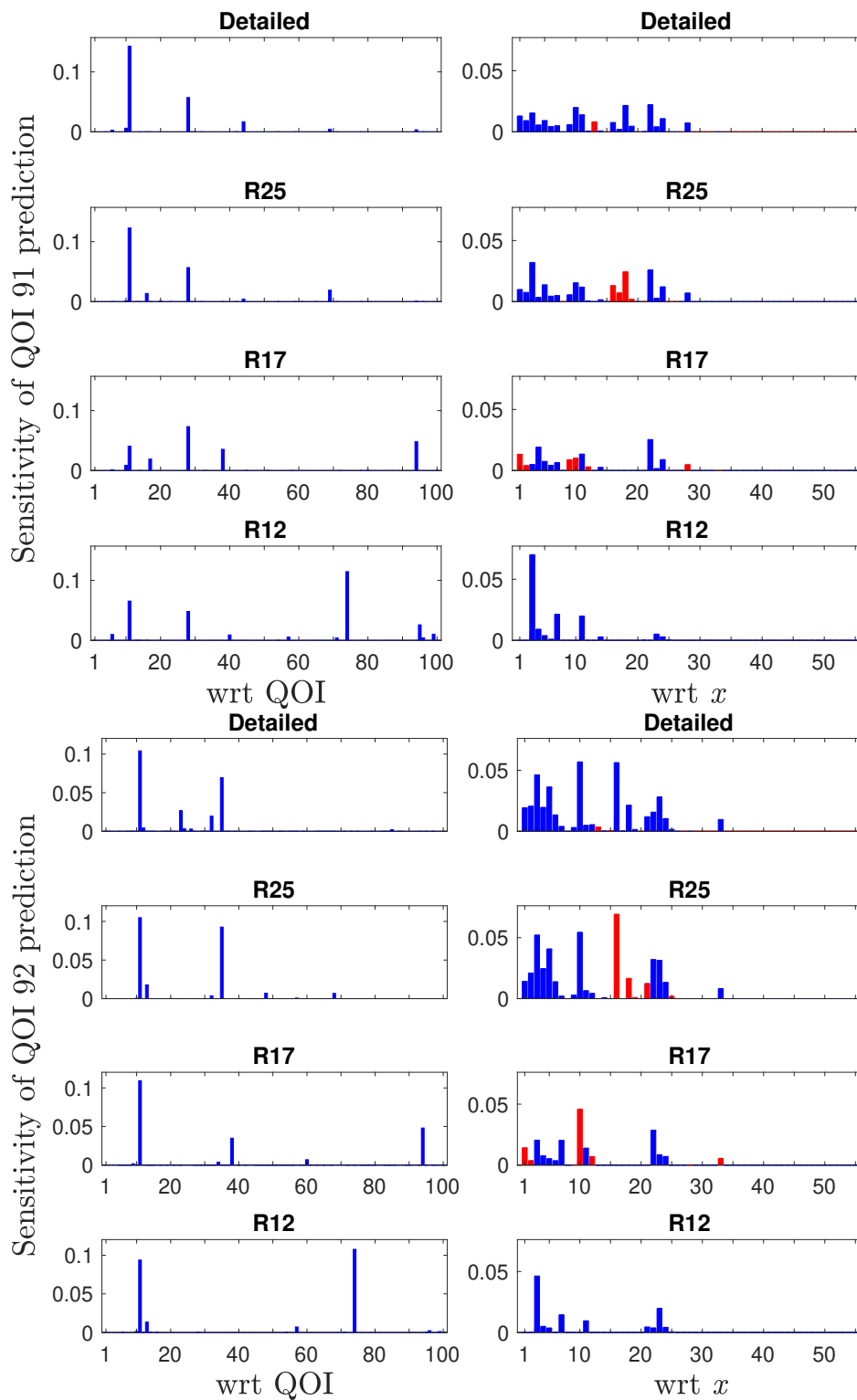


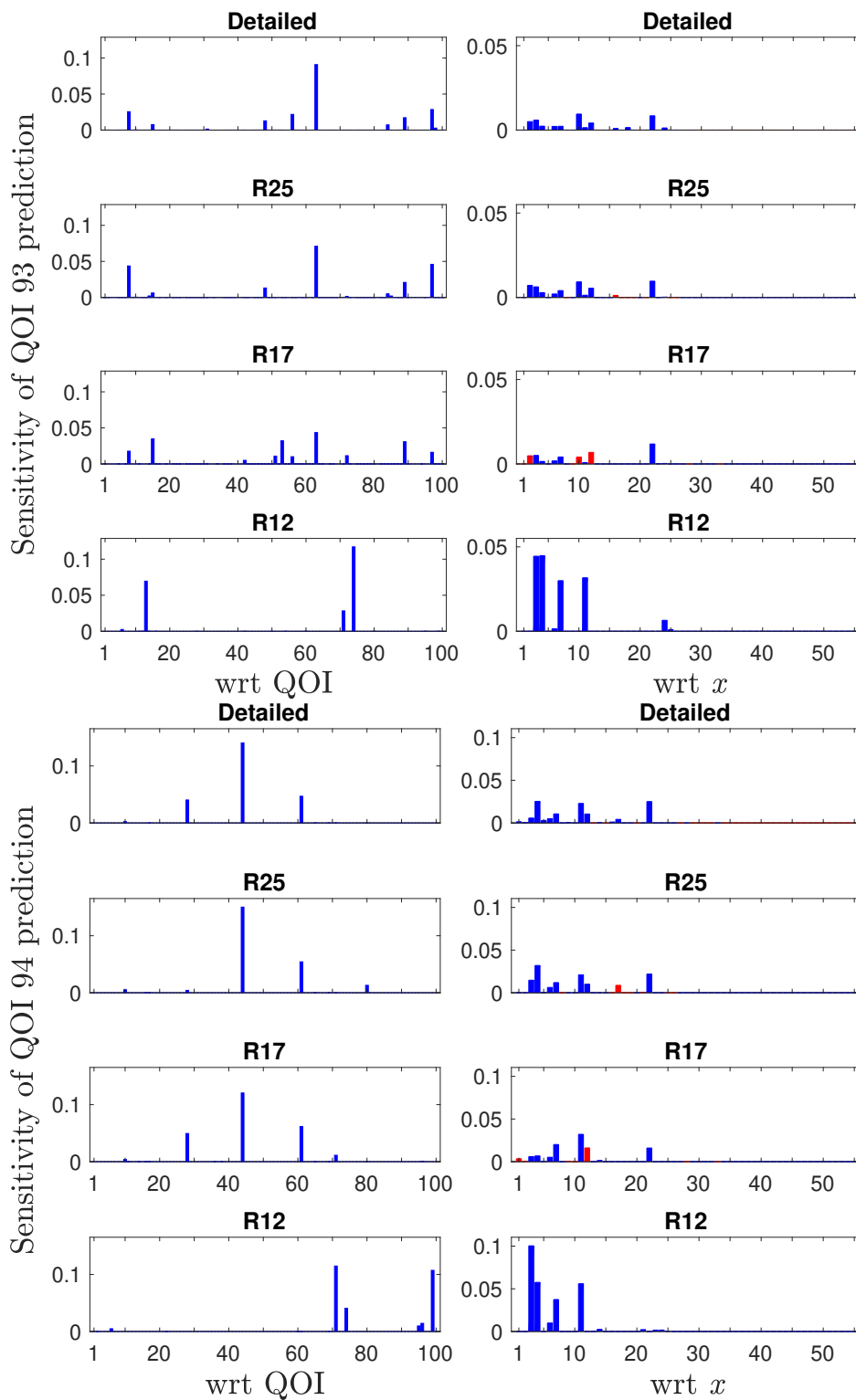


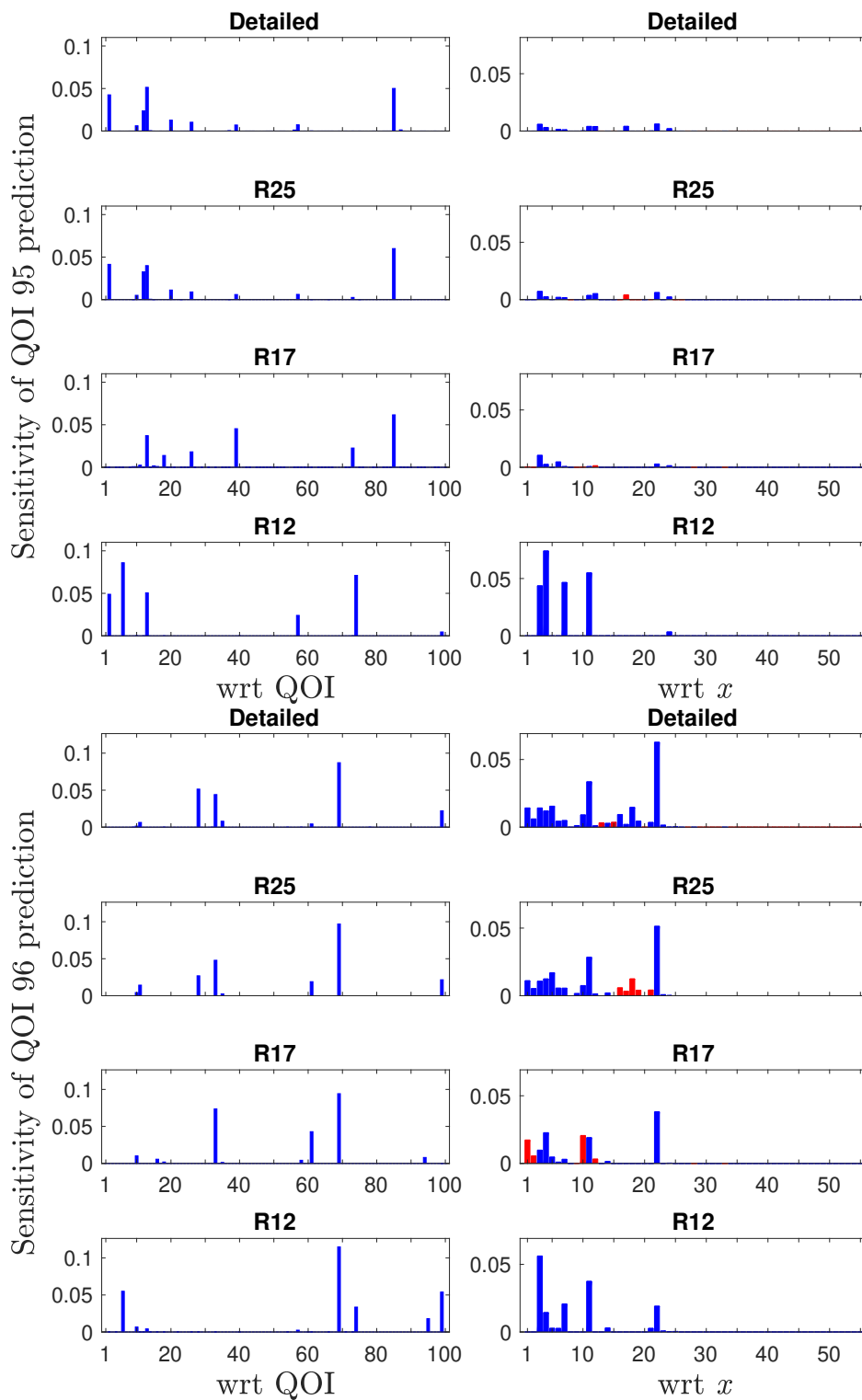


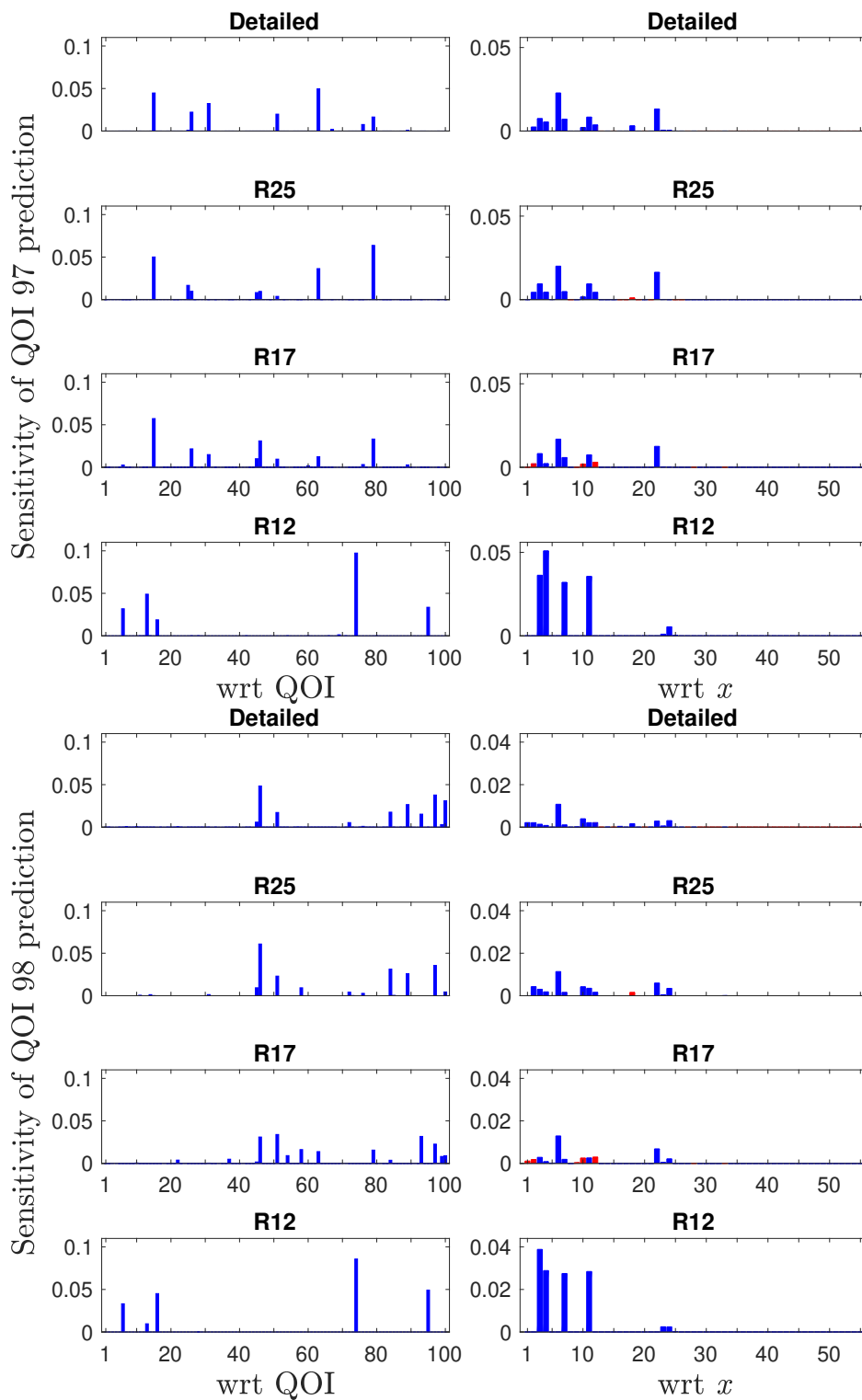












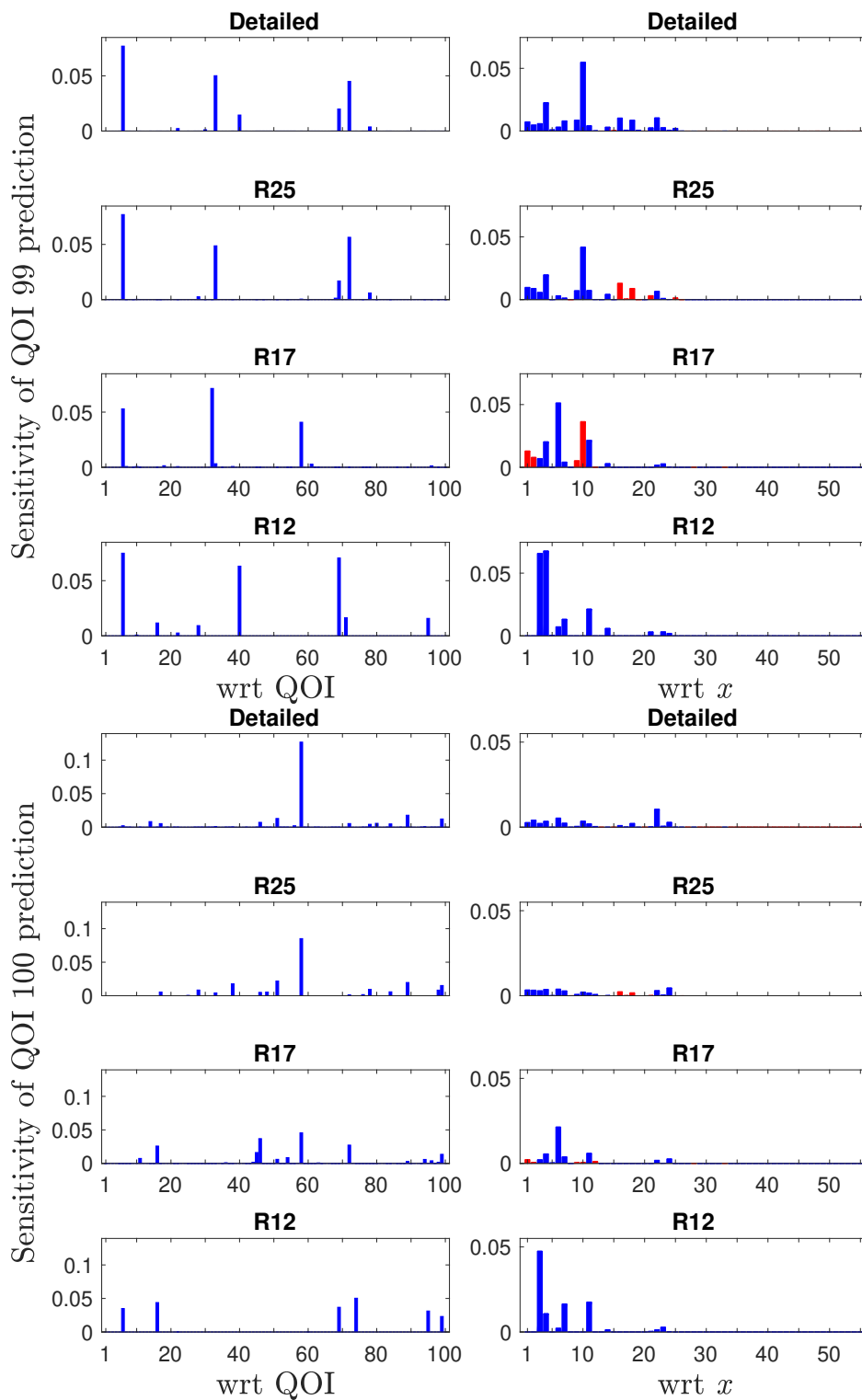
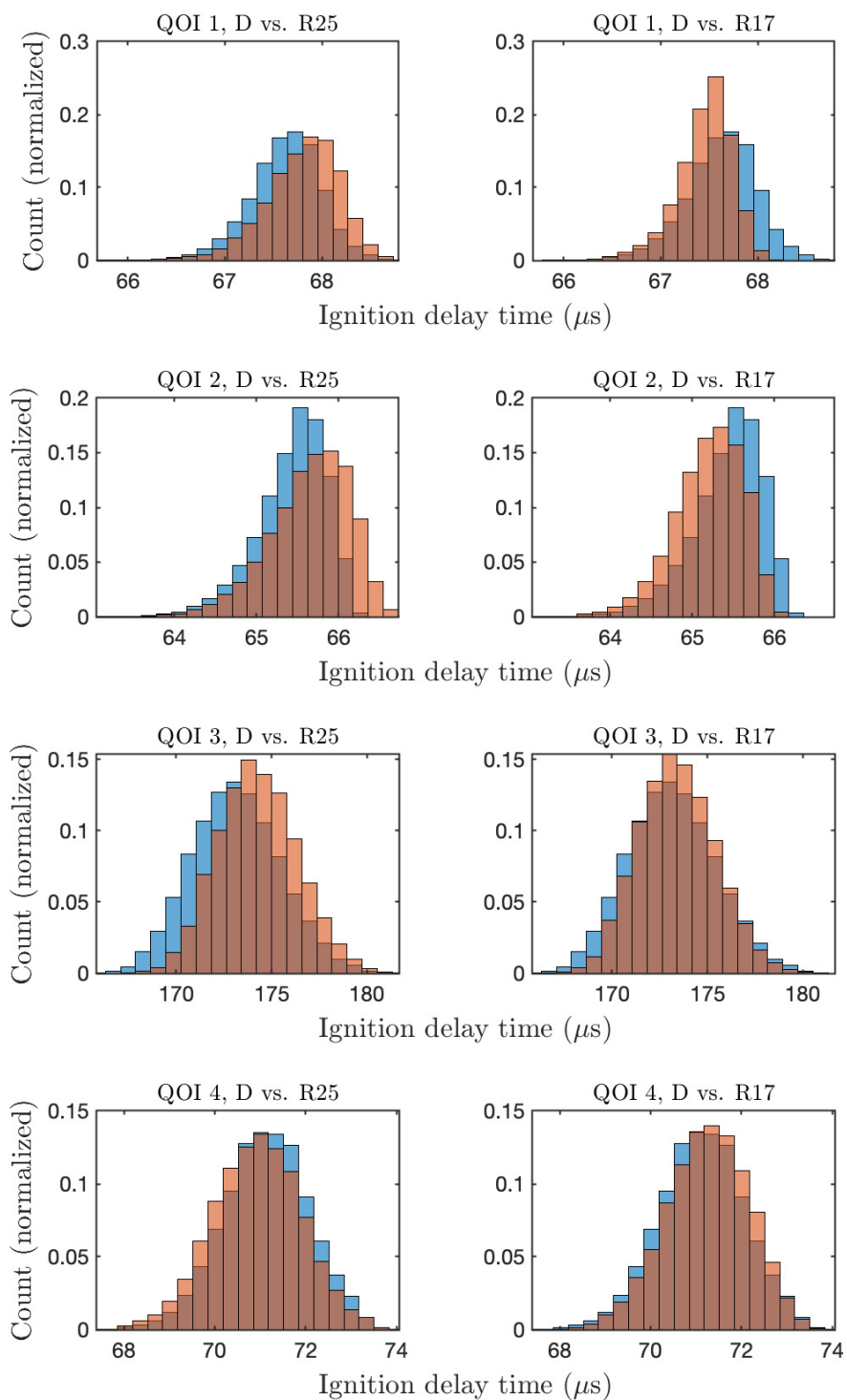
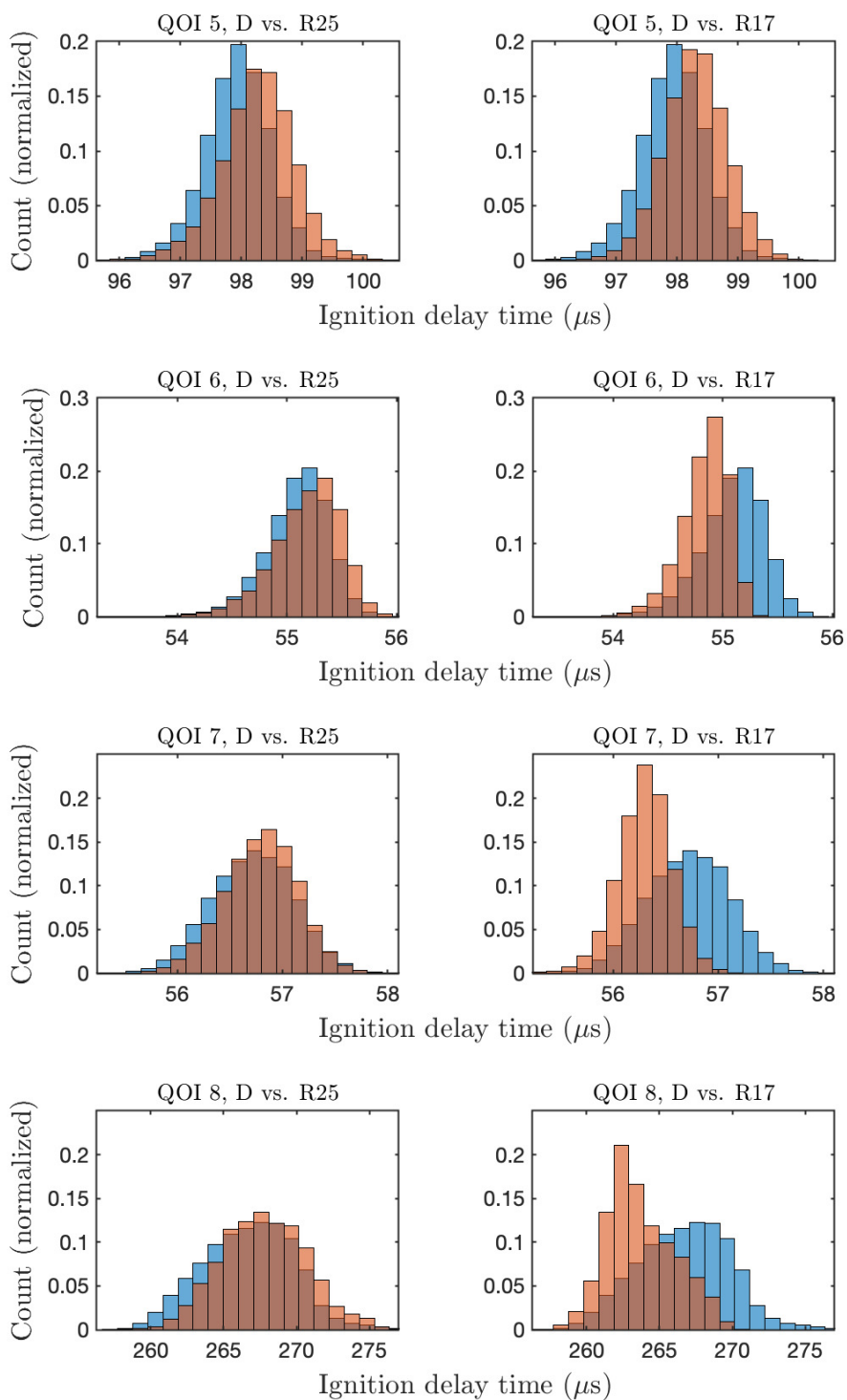


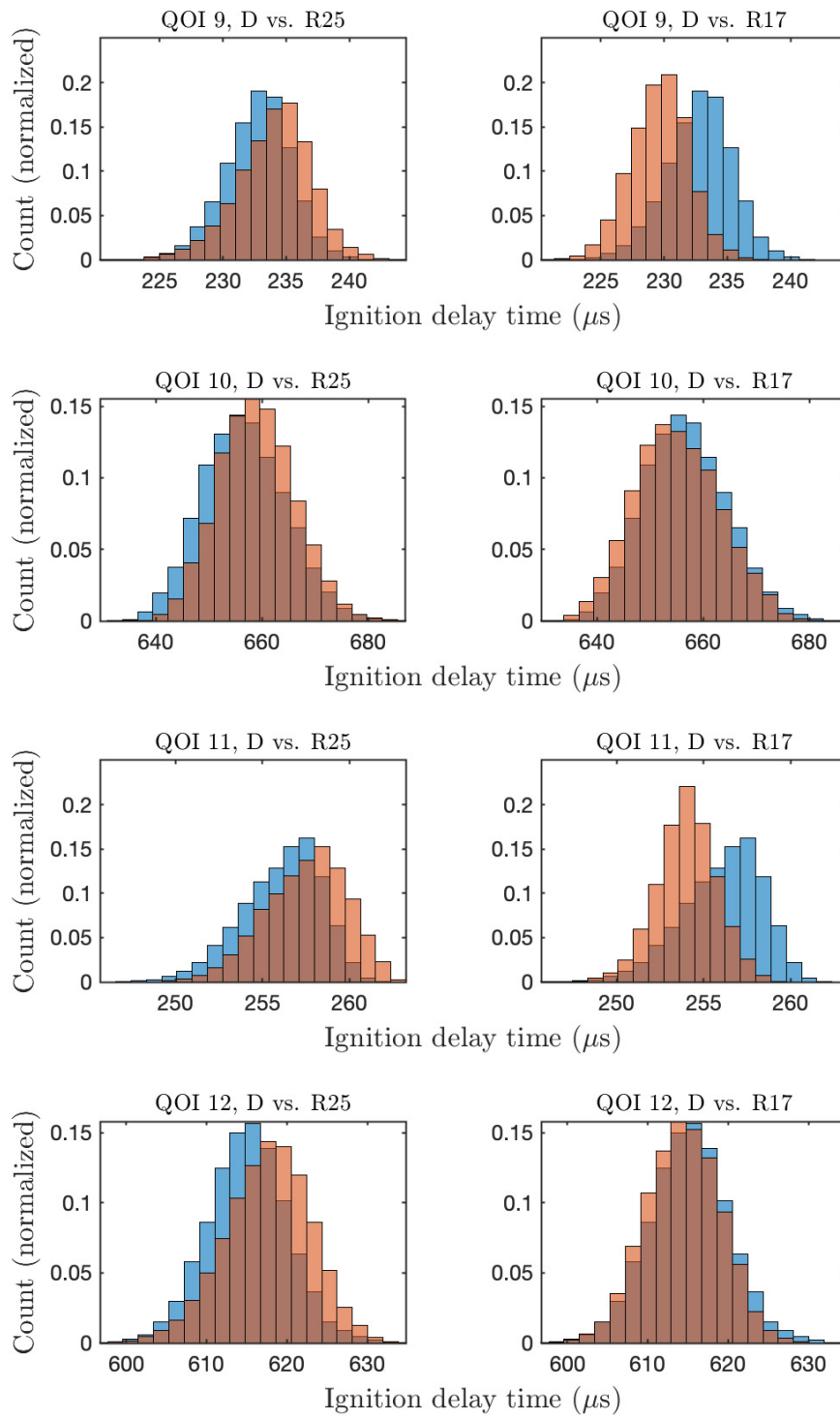
Figure A.3: Sensitivity of prediction interval for 100 QOIs with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.

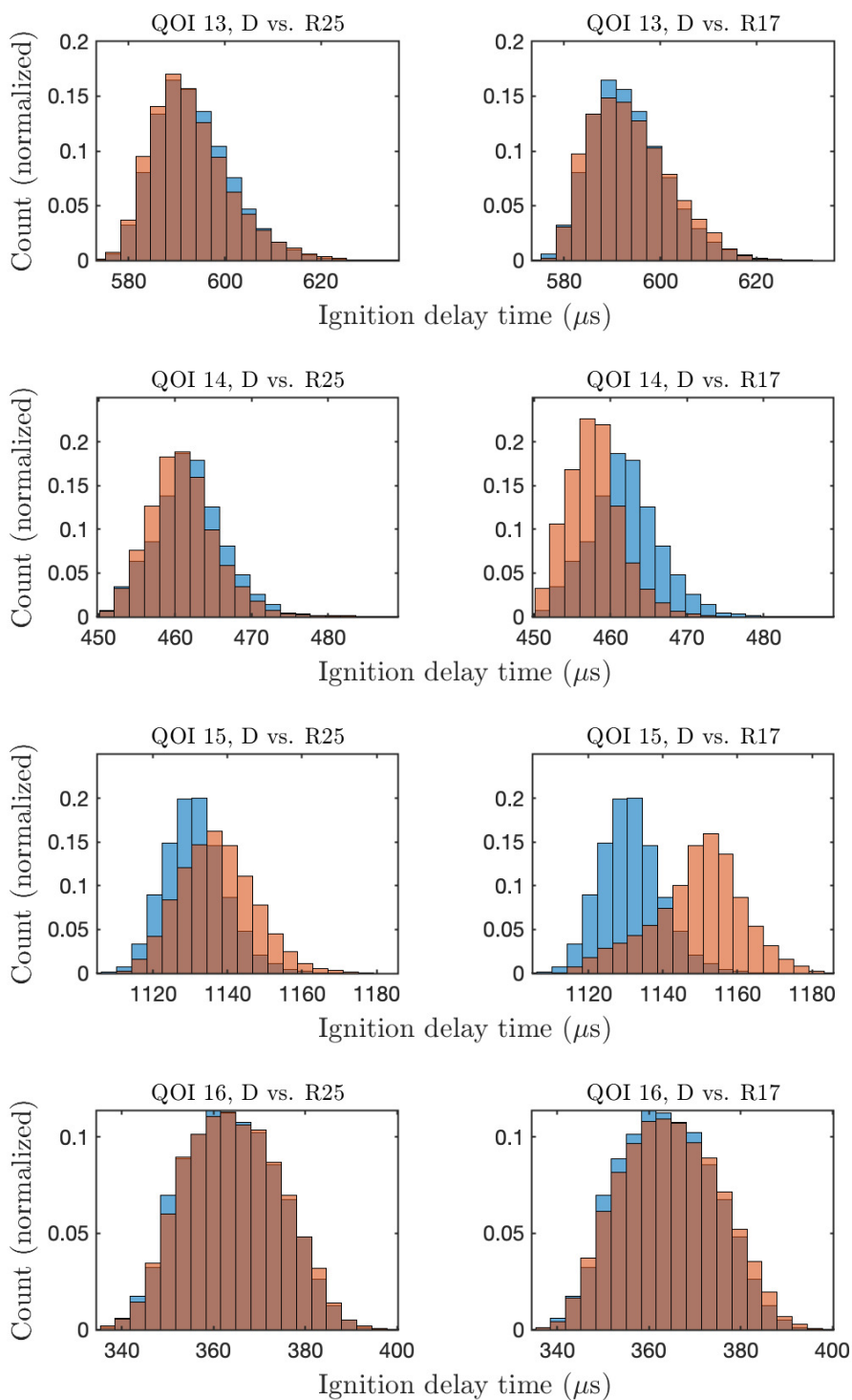
Table A.4: Deviations between predicted values computed by the detailed and reduced models with the nominal parameter values, δ_e , over the prior region, $\delta_e^{\mathcal{H}}$, and over the posterior region, $\delta_p^{\mathcal{F}}$

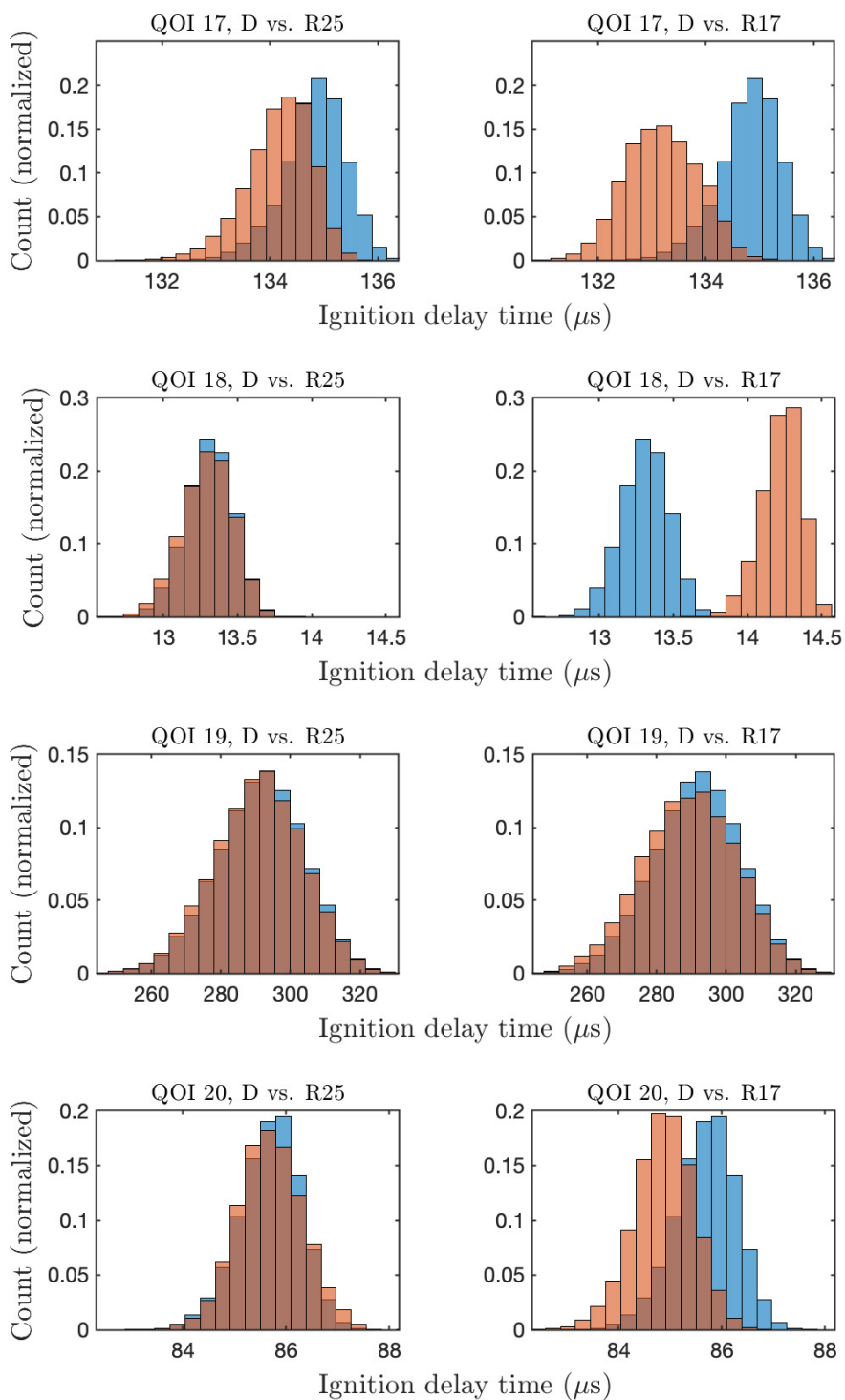
$e p$	δ_e		$\delta_e^{\mathcal{H}}$ (inner)		$\delta_e^{\mathcal{H}}$ (outer)		$\delta_p^{\mathcal{F}}$ (inner)		$\delta_p^{\mathcal{F}}$ (outer)	
	R25	R17	R25	R17	R25	R17	R25	R17	R25	R17
1	0.000	0.003	0.007	0.009	0.007	0.010	0.002	0.007	0.005	0.009
2	0.009	0.001	0.010	0.017	0.010	0.018	0.006	0.010	0.010	0.016
3	0.003	0.008	0.007	0.038	0.008	0.038	0.005	0.023	0.006	0.027
4	0.001	0.002	0.014	0.016	0.014	0.016	0.007	0.008	0.013	0.013
5	0.001	0.005	0.006	0.019	0.006	0.019	0.004	0.004	0.006	0.015
6	0.000	0.000	0.012	0.015	0.013	0.015	0.009	0.008	0.011	0.012
7	0.000	0.001	0.010	0.011	0.011	0.011	0.006	0.009	0.010	0.009
8	0.001	0.000	0.013	0.012	0.013	0.012	0.007	0.008	0.010	0.010
9	0.005	0.002	0.042	0.049	0.043	0.052	0.032	0.030	0.042	0.050
10	0.004	0.003	0.019	0.021	0.019	0.022	0.011	0.017	0.016	0.021
11	0.003	0.001	0.030	0.028	0.033	0.029	0.024	0.016	0.031	0.027
12	0.003	0.004	0.013	0.016	0.013	0.018	0.006	0.009	0.013	0.015
13	0.001	0.004	0.006	0.013	0.006	0.013	0.005	0.007	0.006	0.009
14	0.001	0.001	0.006	0.013	0.006	0.013	0.006	0.007	0.006	0.010
15	0.000	0.008	0.017	0.028	0.019	0.028	0.014	0.018	0.018	0.024
16	0.003	0.005	0.024	0.032	0.024	0.032	0.013	0.026	0.019	0.031
17	0.001	0.008	0.009	0.026	0.011	0.026	0.007	0.009	0.010	0.022
18	0.001	0.063	0.041	0.167	0.041	0.170	0.035	0.152	0.039	0.164
19	0.002	0.005	0.024	0.041	0.024	0.043	0.011	0.032	0.020	0.034
20	0.003	0.005	0.010	0.015	0.010	0.016	0.007	0.011	0.009	0.014
21	0.001	0.004	0.004	0.014	0.004	0.014	0.002	0.010	0.003	0.012
22	0.001	0.006	0.005	0.045	0.005	0.045	0.004	0.035	0.004	0.043











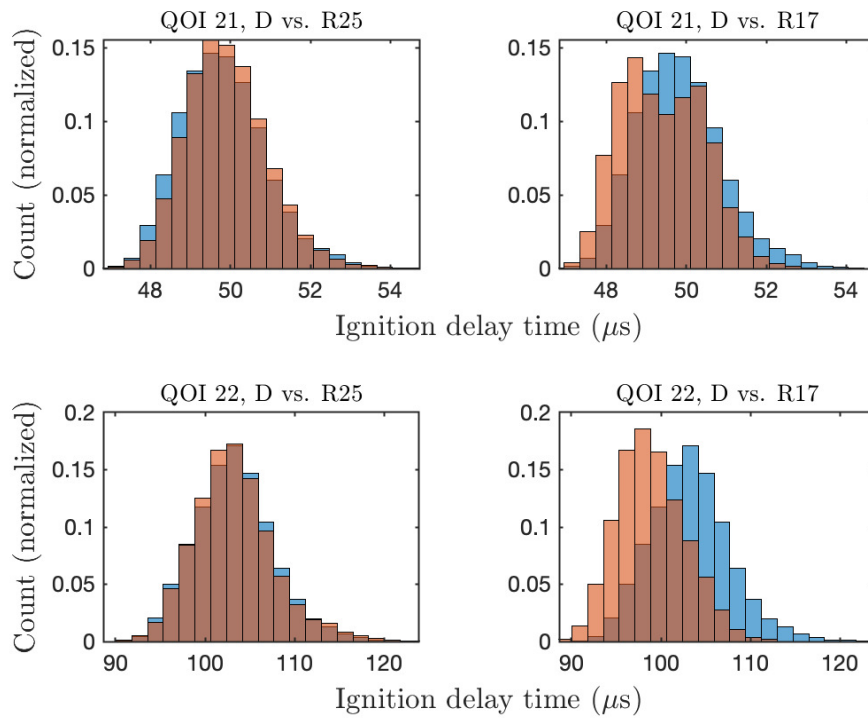
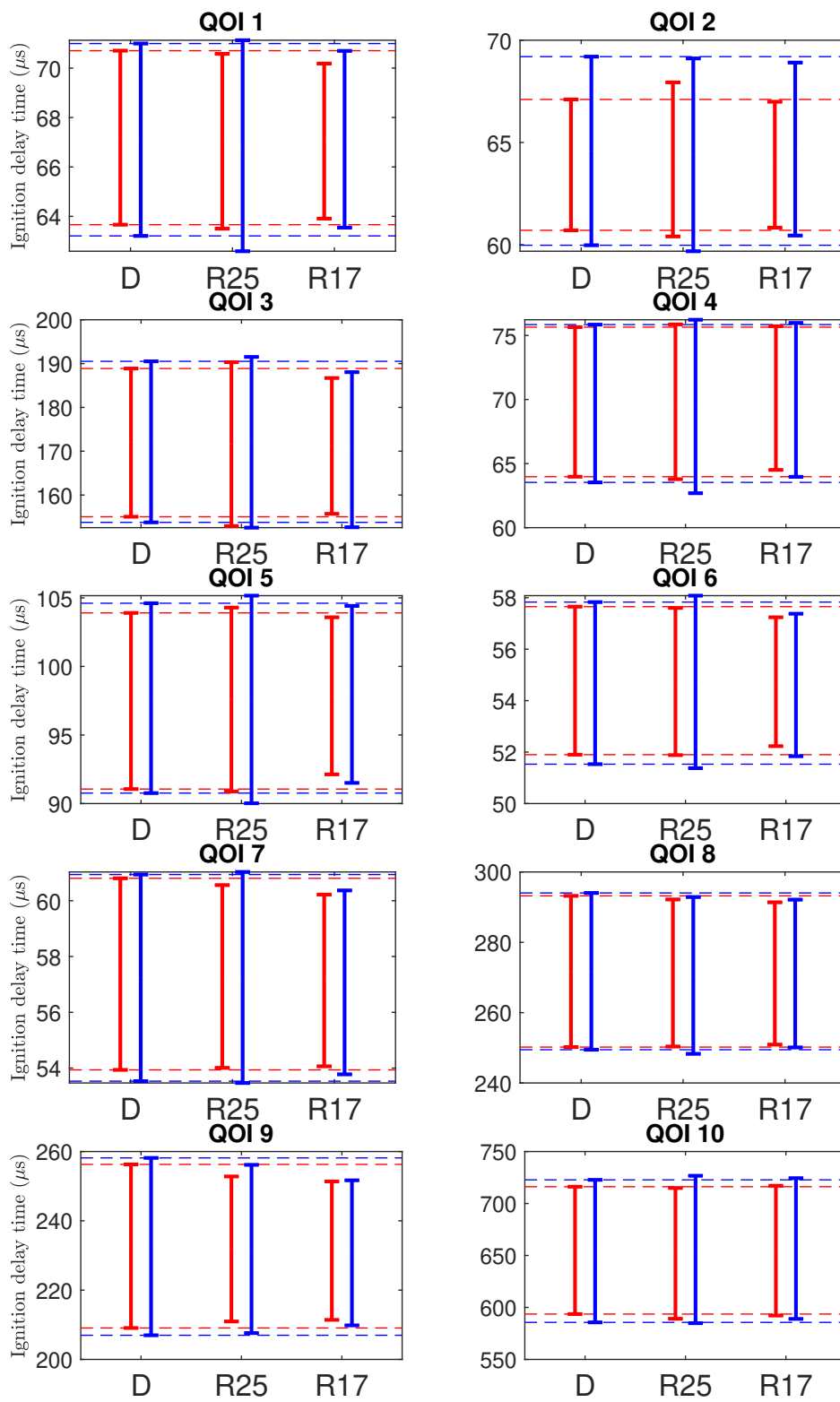
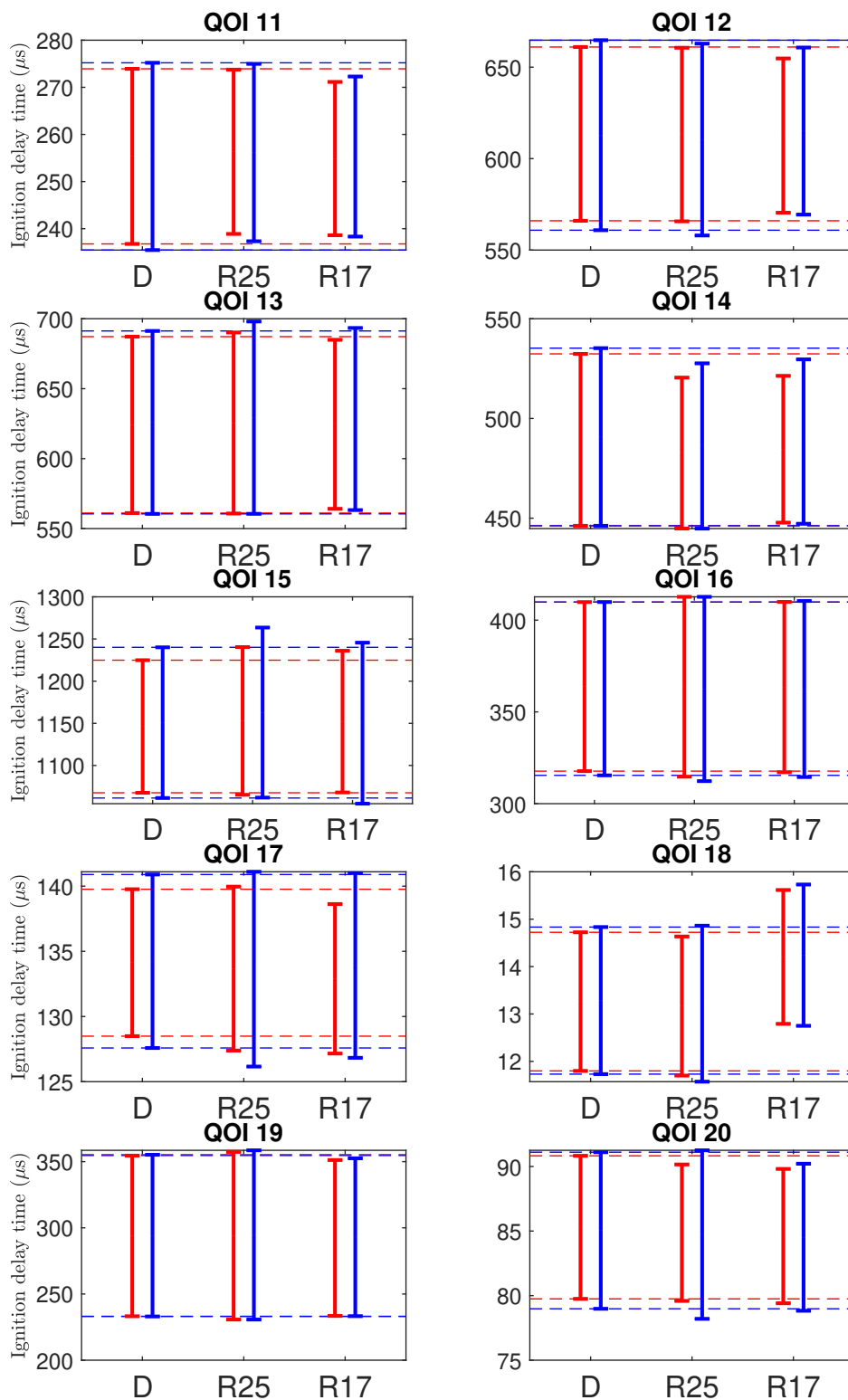


Figure A.4: Histograms for 22 QOIs of detailed (blue) and reduced (brown) model predictions over feasible sets.

Table A.5: Hellinger distance calculated for the prediction intervals, h_p^U , and for the sampled histograms, h_p^H .

p	h_p^U (inner)		h_p^U (outer)		h_p^H	
	R25	R17	R25	R17	R25	R17
1	0.001	0.058	0.058	0.051	0.184	0.269
2	0.003	0.024	0.023	0.049	0.258	0.219
3	0.012	0.053	0.026	0.054	0.193	0.092
4	0.003	0.038	0.052	0.036	0.074	0.054
5	0.002	0.076	0.042	0.053	0.221	0.244
6	0.003	0.066	0.027	0.072	0.138	0.442
7	0.009	0.062	0.008	0.061	0.093	0.469
8	0.003	0.028	0.025	0.028	0.138	0.392
9	0.004	0.085	0.024	0.108	0.184	0.437
10	0.002	0.009	0.016	0.020	0.125	0.075
11	0.007	0.063	0.035	0.077	0.247	0.402
12	0.003	0.055	0.022	0.082	0.180	0.068
13	0.003	0.020	0.030	0.019	0.036	0.048
14	0.003	0.101	0.061	0.041	0.091	0.344
15	0.001	0.050	0.088	0.035	0.229	0.564
16	0.019	0.005	0.027	0.008	0.018	0.036
17	0.015	0.127	0.081	0.049	0.397	0.748
18	0.009	0.422	0.040	0.412	0.039	0.994
19	0.004	0.016	0.019	0.012	0.019	0.060
20	0.002	0.071	0.046	0.055	0.064	0.461
21	0.006	0.113	0.027	0.059	0.047	0.162
22	0.019	0.014	0.003	0.015	0.030	0.361





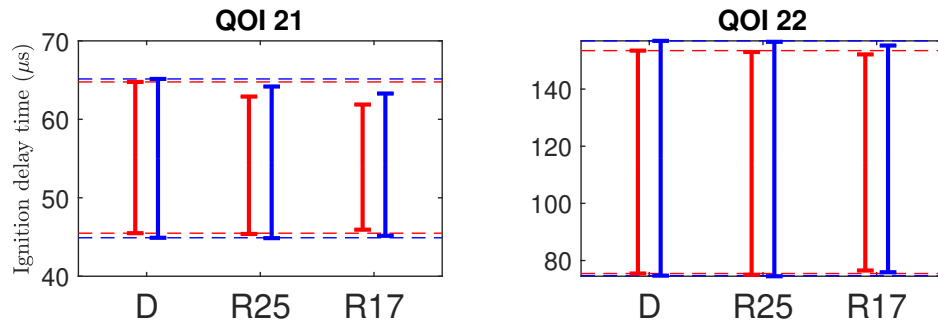
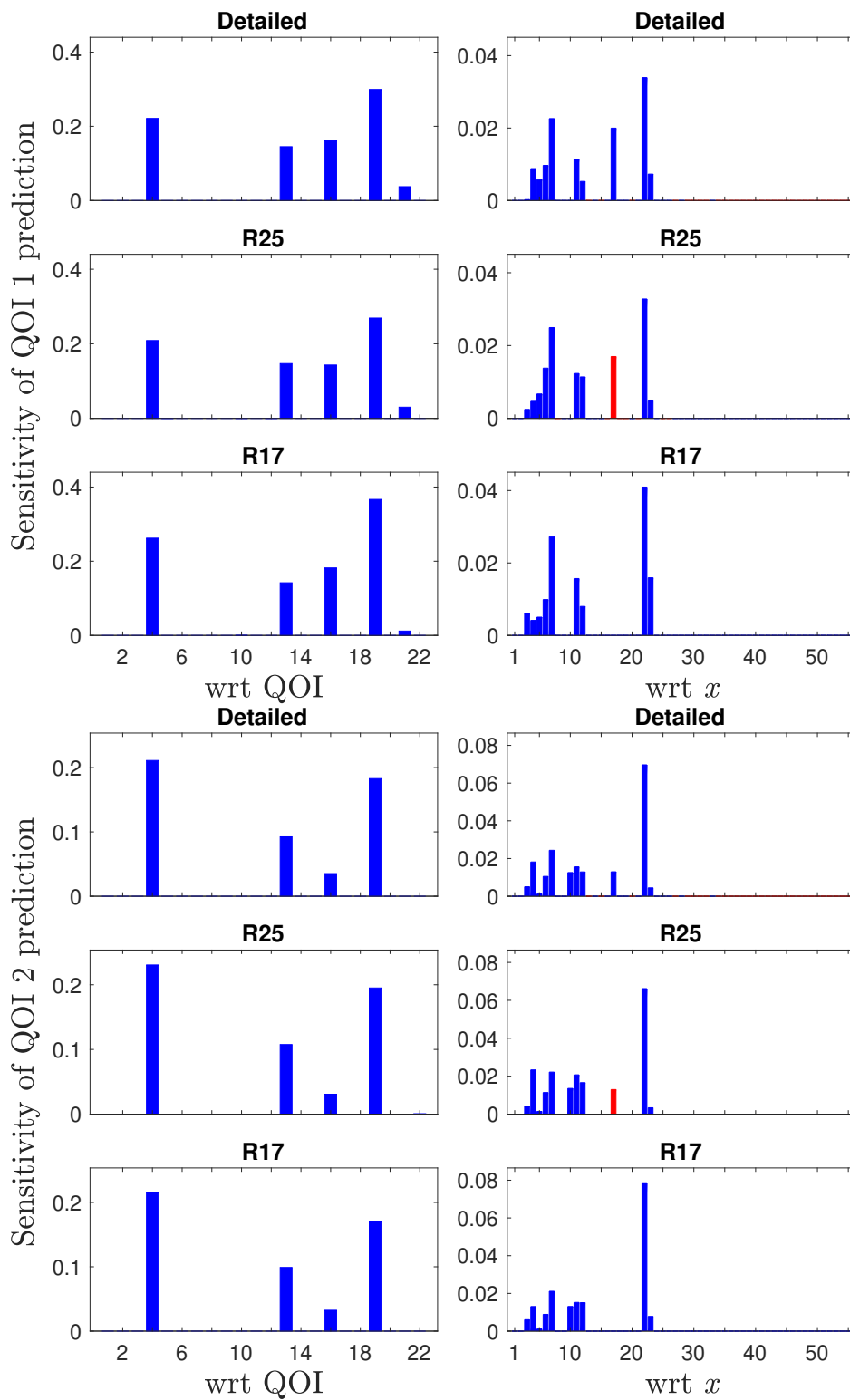
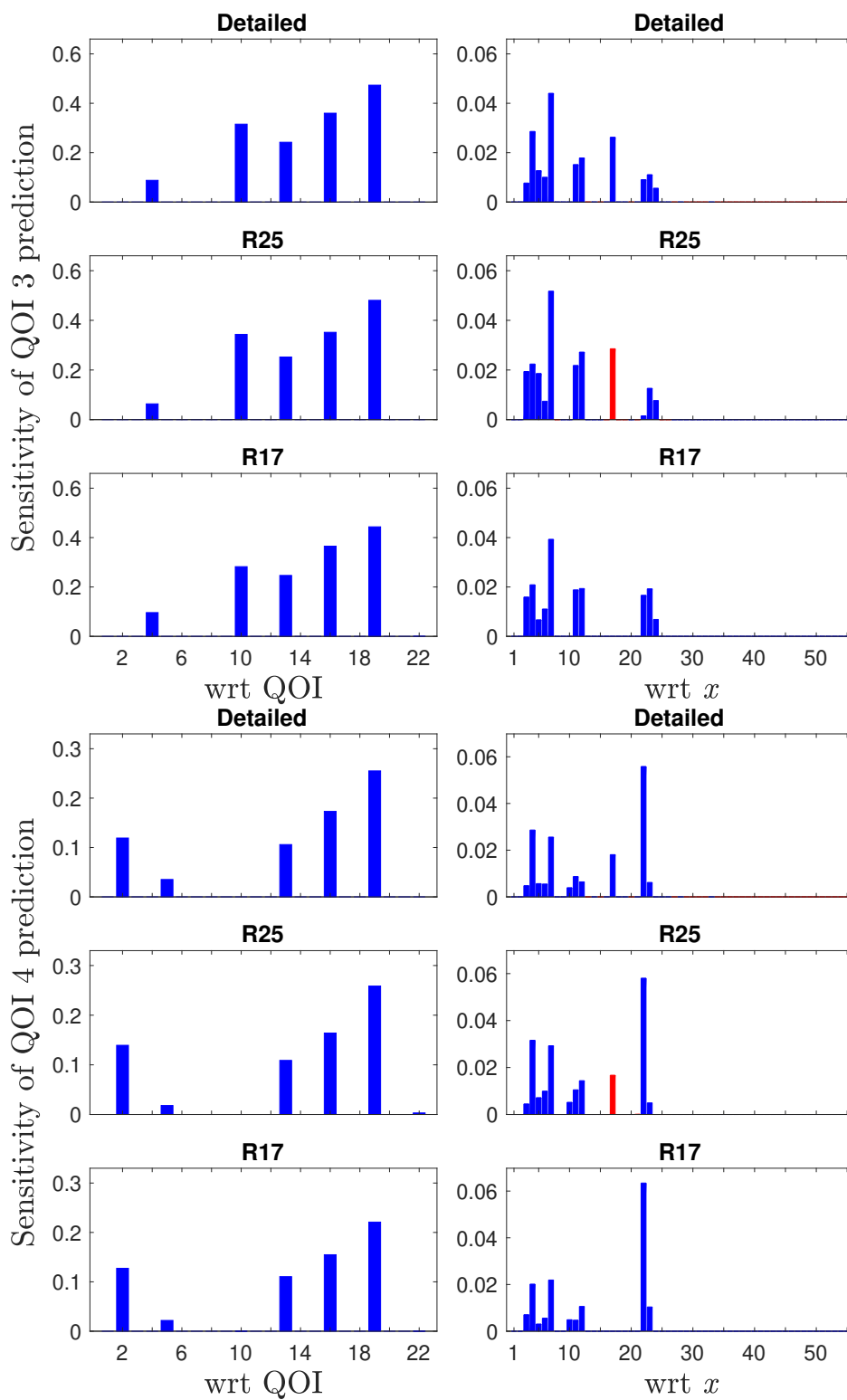
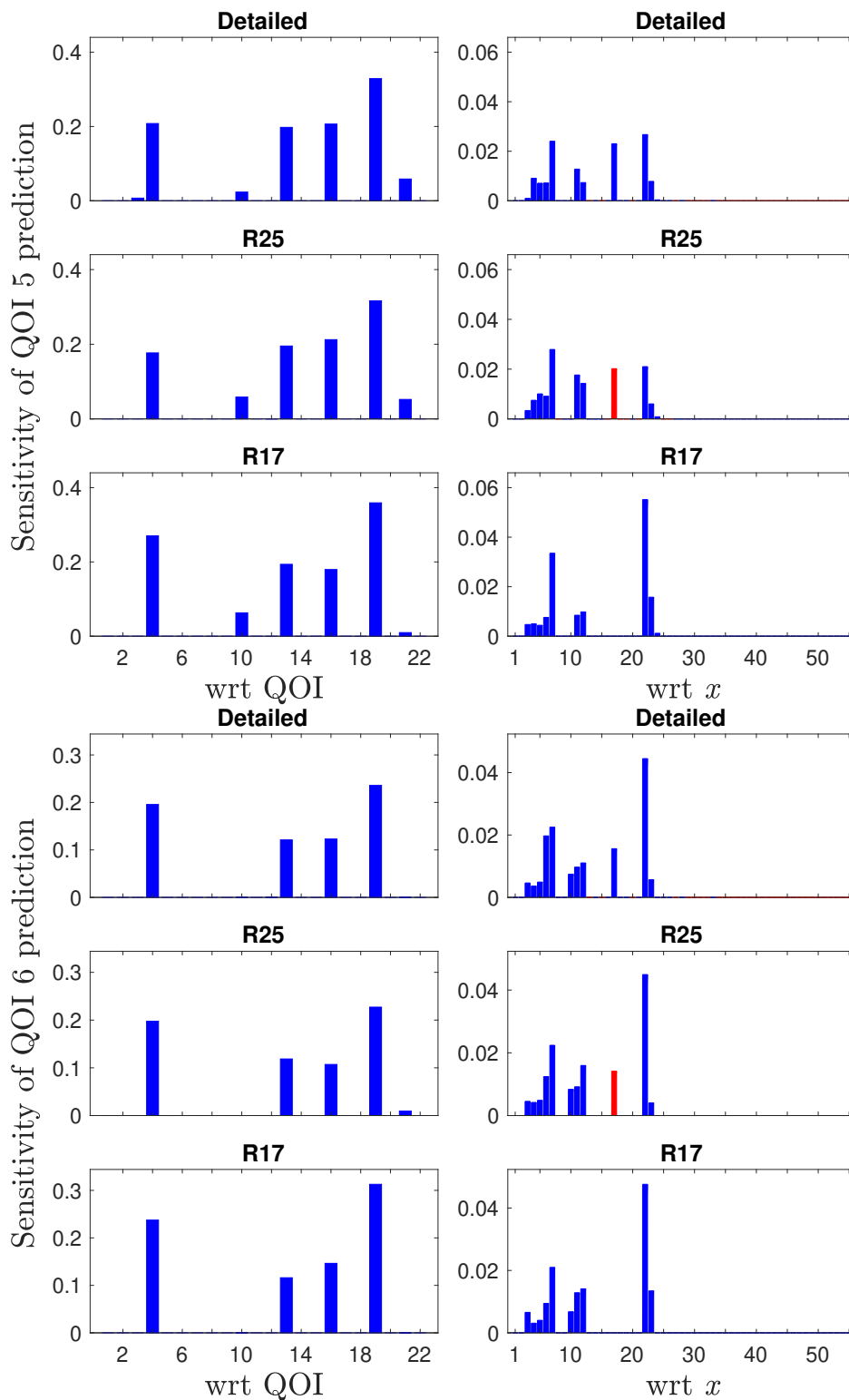
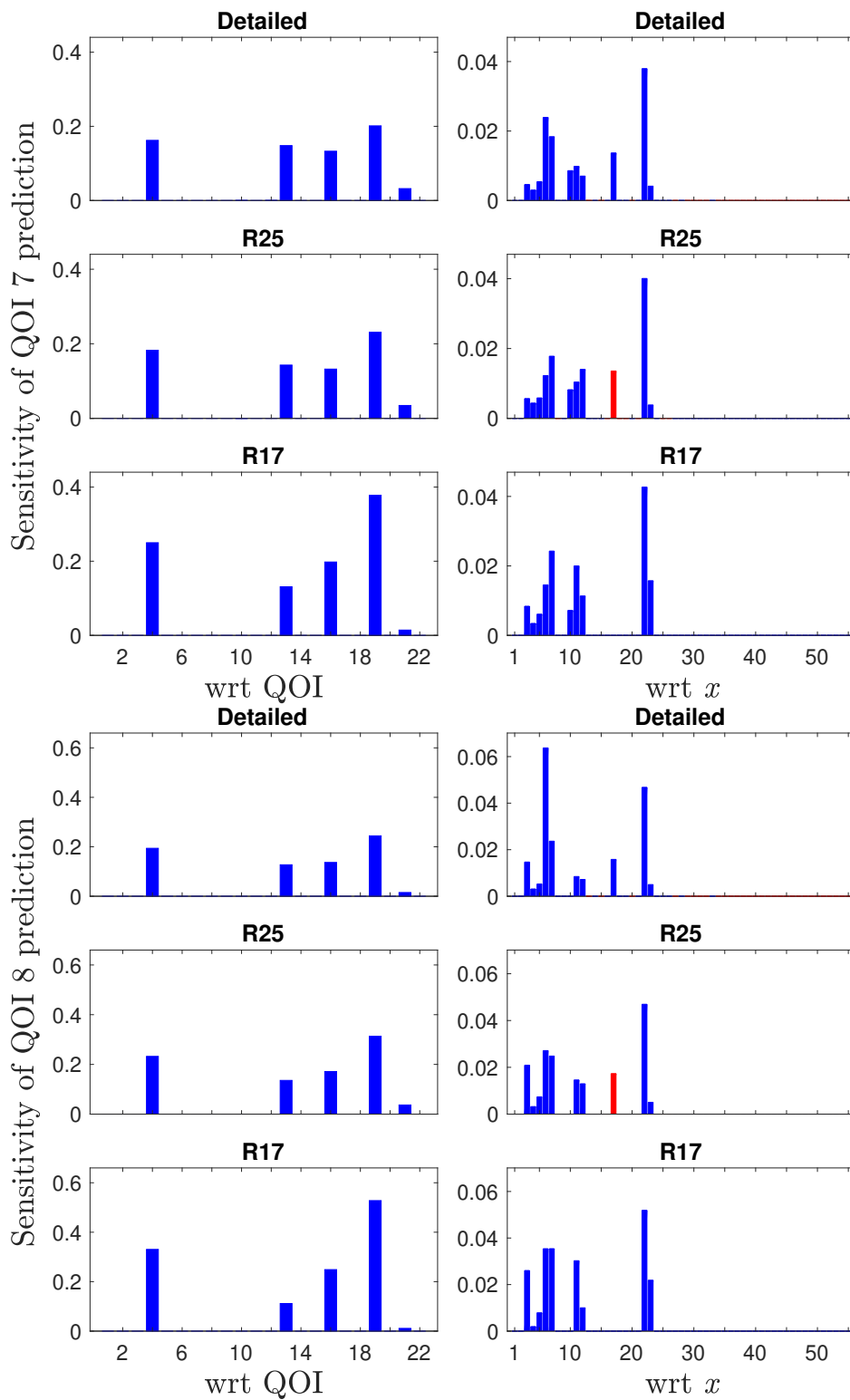


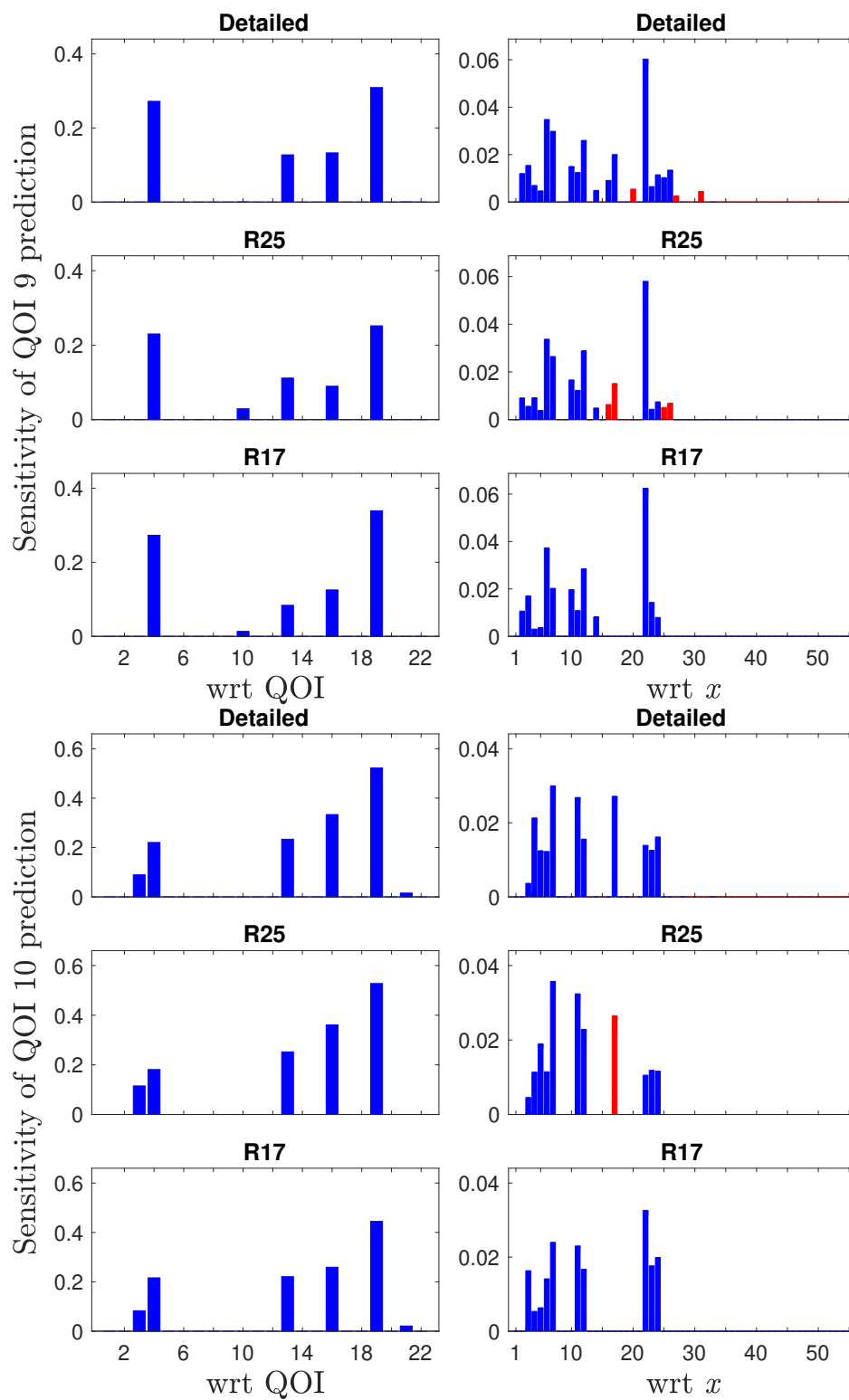
Figure A.5: QOI prediction intervals computed with the detailed, D, and reduced, R25, and R17, models; vertical red lines are inner-bound prediction intervals, vertical blue lines are outer-bound prediction intervals, and horizontal dashed lines are prediction intervals of the detailed model.

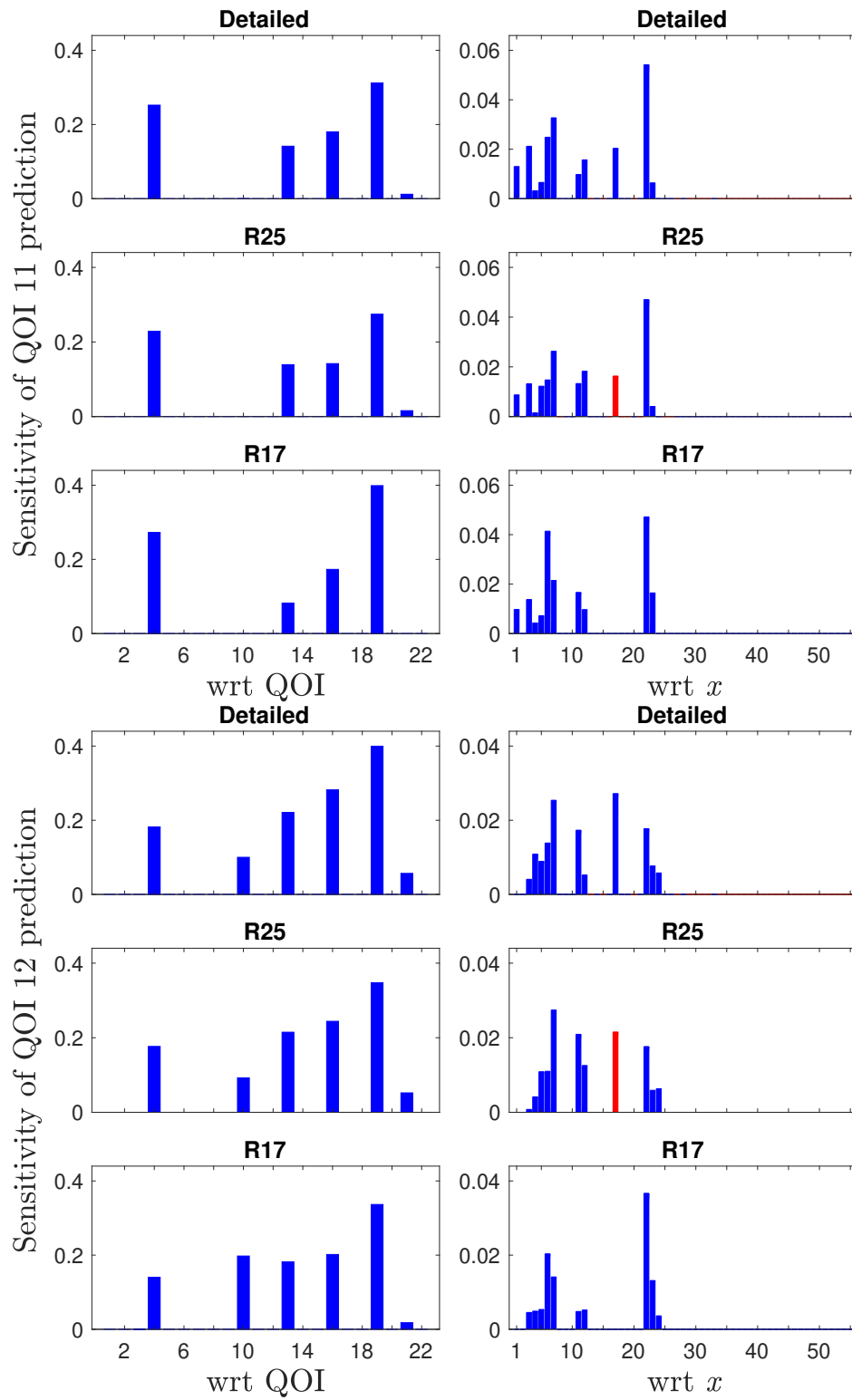


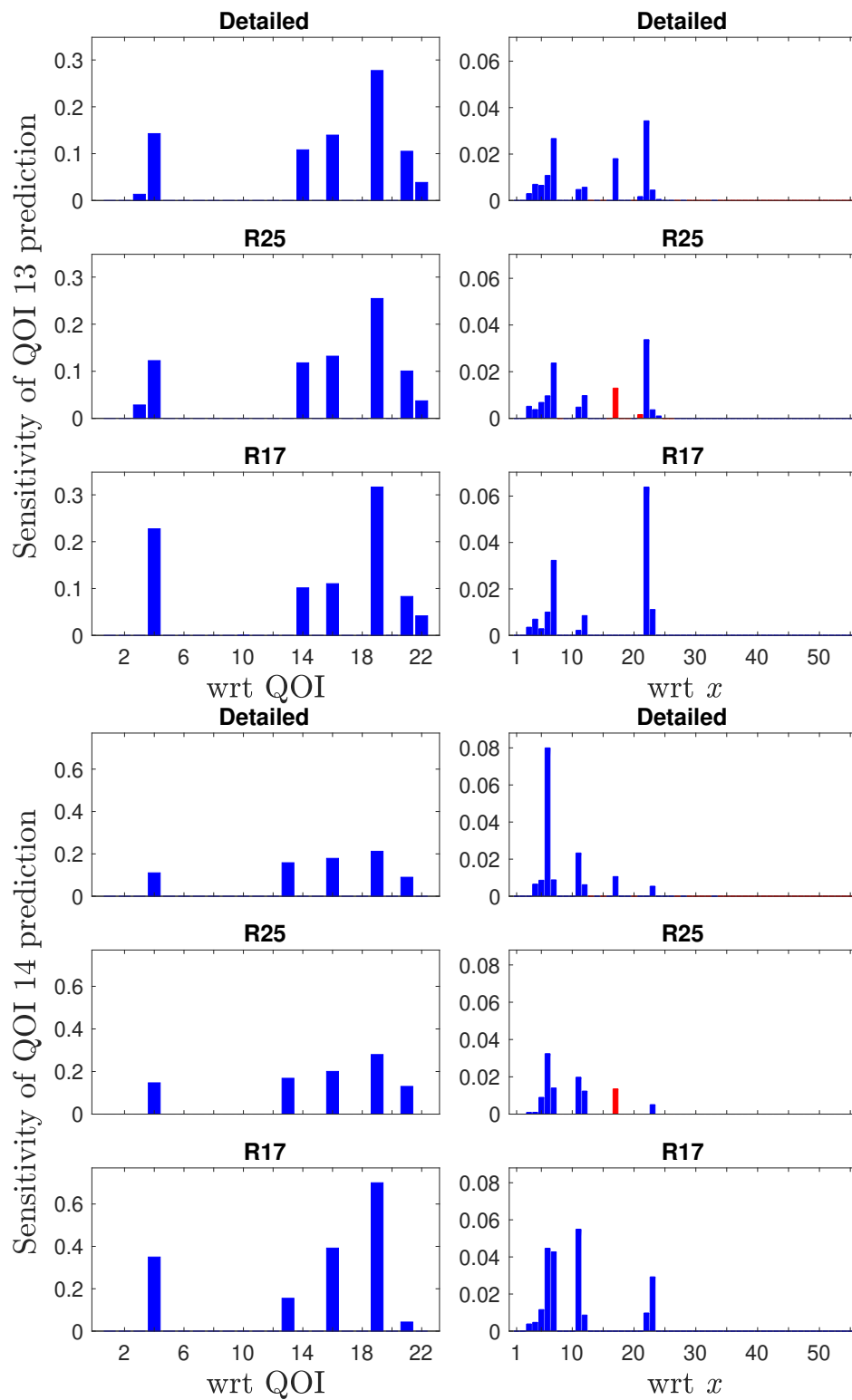


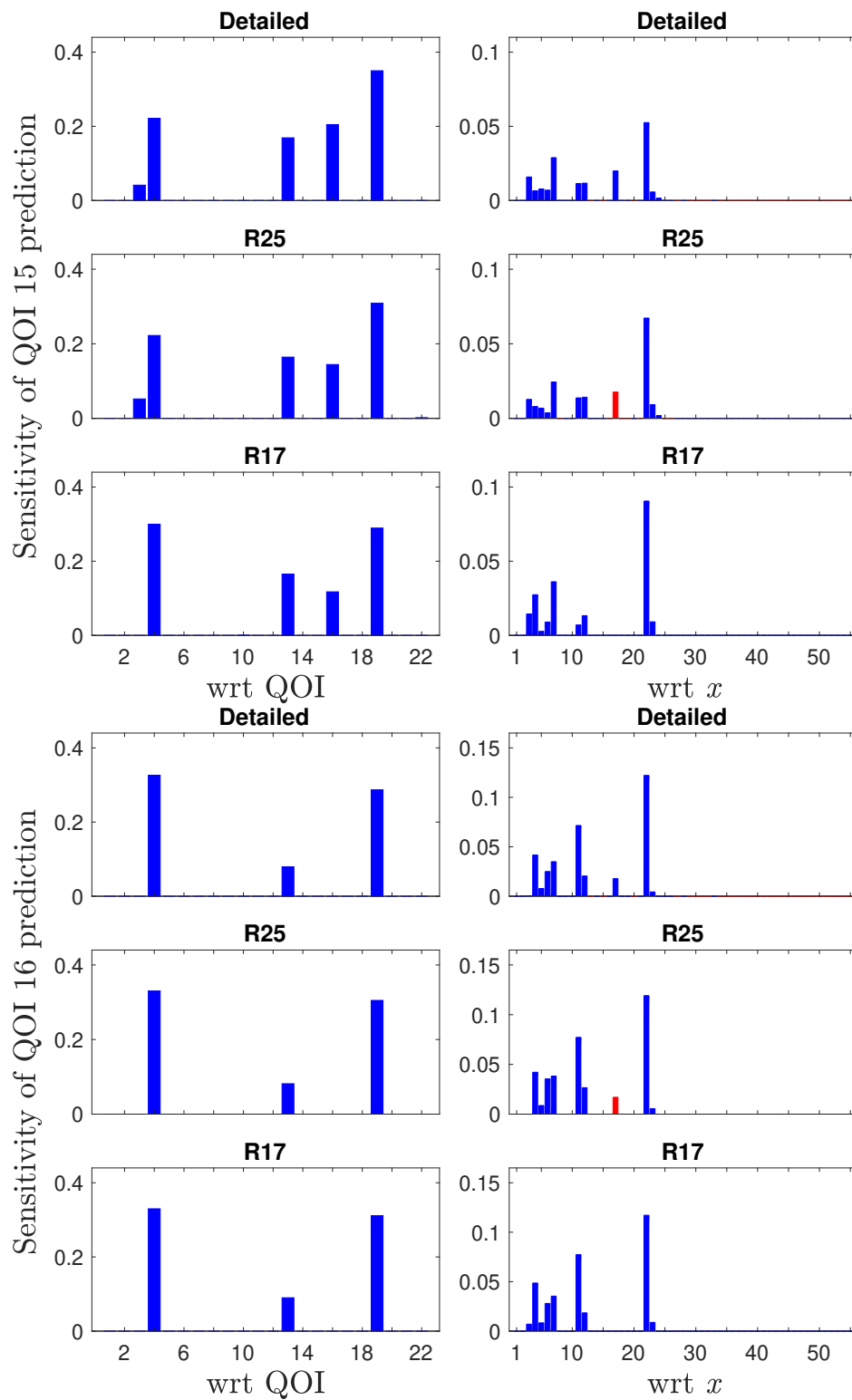


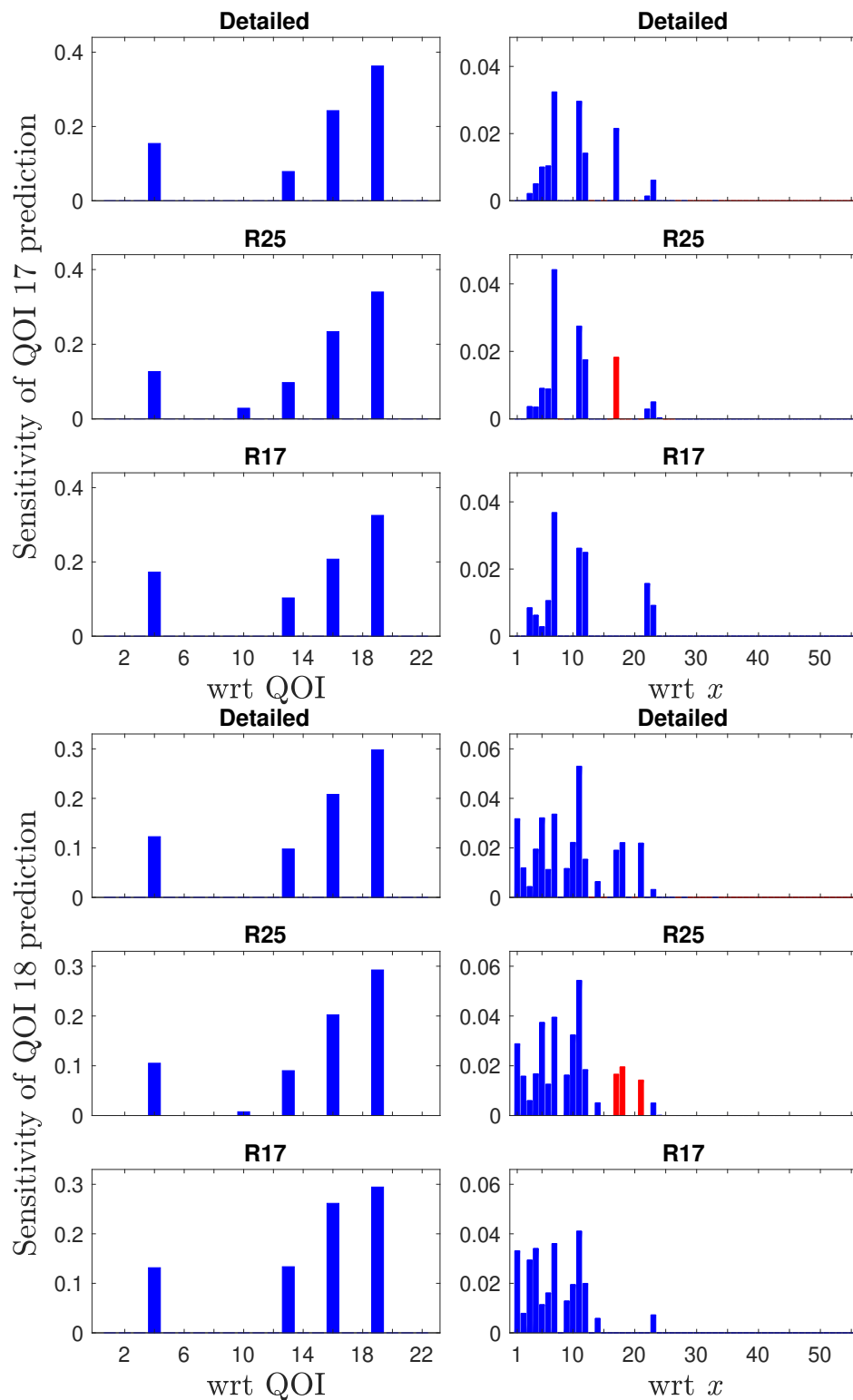


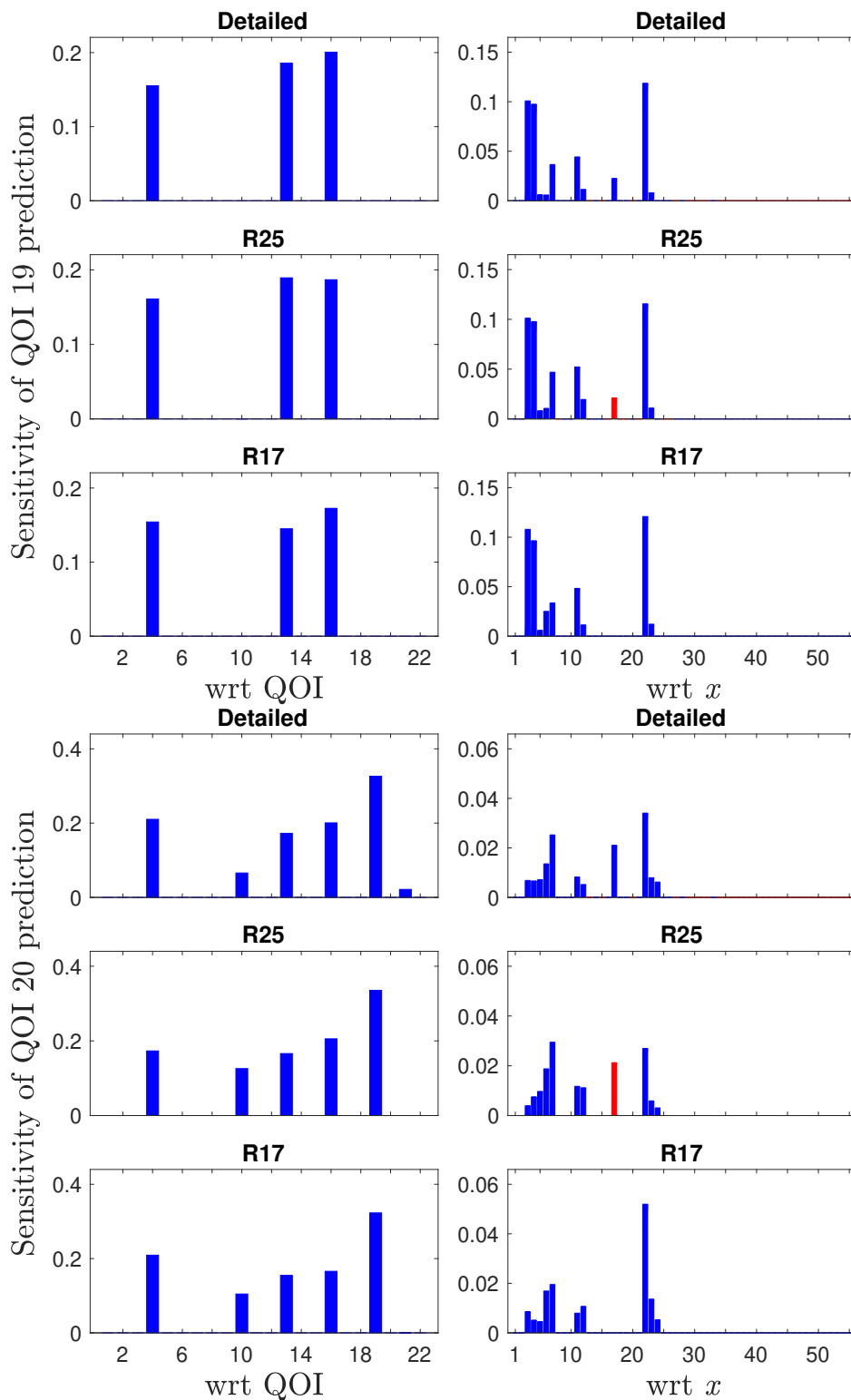












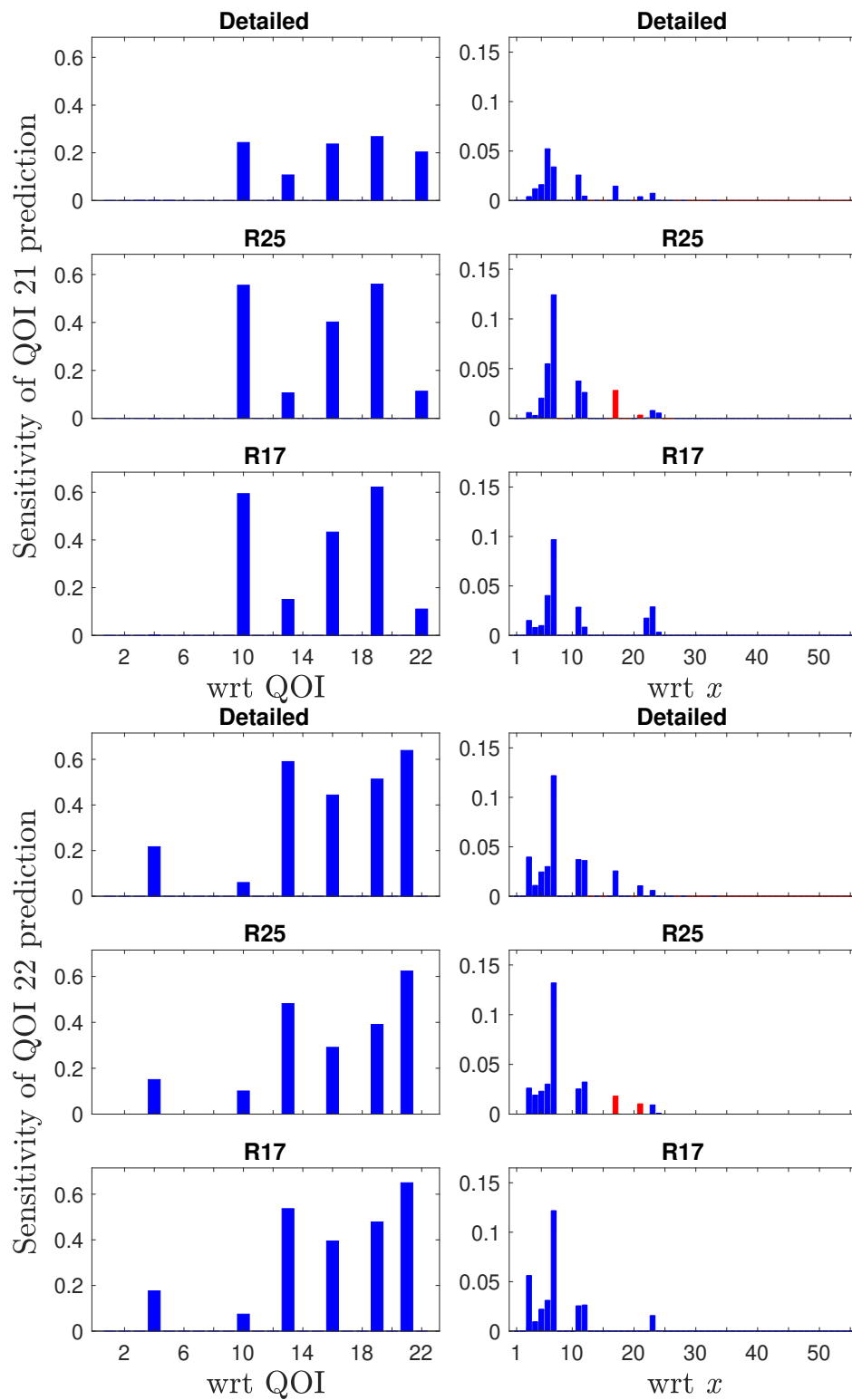


Figure A.6: Sensitivity of prediction interval for 22 QOIs with respect to the prediction intervals of the rest of the QOIs (left panels) and with respect to the uncertainty intervals of model parameters (right panels). Colored in red are the sensitivities corresponding to the model parameters that are absent from the model displayed in the next panel below.

Bibliography

- [1] Deloitte Touche Tohmatsu Limited. *2020 Renewable Energy Industry Outlook*. Last accessed June 10, 2020. 2020. URL: <https://www2.deloitte.com/us/en/pages/energy-and-resources/articles/renewable-energy-outlook.html>.
- [2] American Council on Renewable Energy. *ESG 2.0: How to Improve ESG Scoring to Better Reflect Renewable Energy Use and Investment*. Last accessed June 11, 2020. 2019. URL: <https://acore.org/esg-2-0-acore-releases-recommendations-on-esg-scoring/>.
- [3] U.S. Energy Information Administration. *Renewable energy explained*. Last accessed June 10, 2020. 2020. URL: <https://www.eia.gov/energyexplained/renewable-sources/>.
- [4] U.S. Energy Information Administration. *U.S. energy facts explained*. Last accessed June 10, 2020. 2020. URL: <https://www.eia.gov/energyexplained/us-energy-facts/>.
- [5] U.S. Energy Information Administration. *ANNUAL ENERGY OUTLOOK 2020*. Last accessed June 10, 2020. 2020. URL: <https://www.eia.gov/outlooks/aeo/>.
- [6] U.S. Energy Information Administration. *INTERNATIONAL ENERGY OUTLOOK 2019*. Last accessed June 10, 2020. 2020. URL: <https://www.eia.gov/outlooks/ieo/>.
- [7] Intergovernmental Panel on Climate Change. *Climate Change 2014, Mitigation of Climate Change, Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Last accessed June 11, 2020. 2014. URL: https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_chapter5.pdf.
- [8] K. Kuo. *Principles of Combustion*. Hoboken, NJ: Wiley, 2005.
- [9] S.R. Turns. *An Introduction to Combustion: Concepts and Applications*. McGraw-Hill series in mechanical engineering. McGraw-Hill, 2012.
- [10] Michael Frenklach. Reaction mechanism of soot formation in flames. *Phys. Chem. Chem. Phys.* 4 (11 2002), pp. 2028–2037.
- [11] Dale R. Tree and Kenth I. Svensson. Soot processes in compression ignition engines. *Progress in Energy and Combustion Science* 33.3 (2007), pp. 272–309.

- [12] Hai Wang. Formation of nascent soot and other condensed-phase materials in flames. *Proceedings of the Combustion Institute* 33.1 (2011), pp. 41–67.
- [13] United States Environmental Protection Agency. *Particulate Matter (PM) Pollution*. Last accessed June 19, 2020. 2020. URL: <https://www.epa.gov/pm-pollution>.
- [14] M. Frenklach. Transforming data into knowledge—Process Informatics for combustion chemistry. *Proc. Combust. Inst.* 31 (2007), pp. 125–140.
- [15] H. Wang and D. A. Sheen. Combustion kinetic model uncertainty quantification, propagation and minimization. *Prog. Energy Combust. Sci.* 47 (2015), pp. 1–31.
- [16] P. Pepiot-Desjardins and H. Pitsch. An efficient error-propagation-based reduction method for large chemical kinetic mechanisms. *Combust. Flame* 154 (2008), pp. 67–81.
- [17] D. A. Sheen, X. You, H. Wang, and T. Løvås. Spectral uncertainty quantification, propagation and optimization of a detailed kinetic model for ethylene combustion. *Proc. Combust. Inst.* 32 (2009), pp. 535–542.
- [18] R. Malpica Galassi, M. Valorani, H. N. Najm, C. Safta, M. Khalil, and P. P. Ciottoli. Chemical model reduction under uncertainty. *Combust. Flame* 179 (2017), pp. 242–252.
- [19] Y. Chang, M. Jia, B. Niu, Y. Zhang, M. Xie, and Y. Li. Construction and assessment of reduced oxidation mechanisms using global sensitivity analysis and uncertainty analysis. *Proc. Combust. Inst.* 37 (2019), pp. 751–761.
- [20] Michael Frenklach, Zhenyuan Liu, Ravi I. Singh, Galiya R. Galimova, Valeriy N. Azyazov, and Alexander M. Mebel. Detailed, sterically-resolved modeling of soot oxidation: Role of O atoms, interplay with particle nanostructure, and emergence of inner particle burning. *Combustion and Flame* 188 (2018), pp. 284–306.
- [21] Hamid Omidvarborna, Ashok Kumar, and Dong-Shik Kim. Recent studies on soot modeling for diesel combustion. *Renewable and Sustainable Energy Reviews* 48 (2015), pp. 635–647.
- [22] M Frenklach, K Kailasanath, and E. S. Oran. Systematic development of reduced reaction mechanisms for dynamic modeling. *Prog. Astronaut. Aeronaut.* 105 (1986), pp. 365–376.
- [23] H. Wang and M. Frenklach. Detailed reduction of reaction mechanisms for flame modeling. *Combust. Flame* 87 (1991), pp. 365–370.
- [24] M. Frenklach. Reduction of chemical reaction models. *Numerical Approaches to Combustion Modeling*. Ed. by E. S. Oran and J. P. Boris. Washington, D.C.: American Institute of Aeronautics and Astronautics, 1991, pp. 129–154.
- [25] T. Lu and C. K. Law. A directed relation graph method for mechanism reduction. *Proc. Combust. Inst.* 30 (1 2005), pp. 1333–1341.

- [26] W. Sun, Z. Chen, X. Gou, and Y. Ju. A path flux analysis method for the reduction of detailed chemical kinetic mechanisms. *Combust. Flame* 157 (7 2010), pp. 1298–1307.
- [27] Y. Chen and J.-Y. Chen. Application of Jacobian defined direct interaction coefficient in DRGEP-based chemical mechanism reduction methods using different graph search algorithms. *Combust. Flame* 174 (2016), pp. 77–84.
- [28] Sanjoy Dasgupta, Christos H. Papadimitriou, and Umesh Vazirani. *Algorithms*. 1st ed. USA: McGraw-Hill, Inc., 2006.
- [29] J.-Y. Chen. A General procedure for constructing reduced reaction mechanisms with given independent relations. *Combust. Sci. Technol.* 57.1-3 (1988), pp. 89–94.
- [30] Simon J. Fraser. The steady state and equilibrium approximations: A geometrical picture. *The Journal of Chemical Physics* 88.8 (1988), pp. 4732–4738.
- [31] T. Turanyi, A. S. Tomlin, and M. J. Pilling. On the error of the quasi-steady-state approximation. *The Journal of Physical Chemistry* 97.1 (1993), pp. 163–172.
- [32] D. A. Goussis. Quasi steady state and partial equilibrium approximations: their relation and their validity. *Combust. Theory Model.* 16 (2012), pp. 869–926.
- [33] S. H. Lam. Singular Perturbation for Stiff Equations Using Numerical Methods. *Recent Advances in the Aerospace Sciences: In Honor of Luigi Crocco on His Seventy-fifth Birthday*. Ed. by Corrado Casci and Claudio Bruno. Boston, MA: Springer US, 1985, pp. 3–19.
- [34] S. H. Lam and D. A. Goussis. Understanding complex chemical kinetics with computational singular perturbation. *Symp. (Int.) Combust.* 22 (1989), pp. 931–941.
- [35] S. H. Lam and D. A. Coussis. Conventional asymptotics and computational singular perturbation for simplified kinetics modelling. *Reduced Kinetic Mechanisms and Asymptotic Approximations for Methane-Air Flames: A Topical Volume*. Ed. by Mitchell D. Smooke. Berlin, Heidelberg: Springer Berlin Heidelberg, 1991, pp. 227–242.
- [36] D.A. Goussis and S.H. Lam. A study of homogeneous methanol oxidation kinetics using CSP. *Symposium (International) on Combustion* 24.1 (1992). Twenty-Fourth Symposium on Combustion, pp. 113–120.
- [37] S. H. Lam and D. A. Goussis. The CSP method for simplifying kinetics. *Int. J. Chem. Kinet.* 26 (4 1994), pp. 461–486.
- [38] U. Maas and S. B. Pope. Simplifying chemical kinetics: Intrinsic low-dimensional manifolds in composition space. *Combust. Flame* 88 (1992), pp. 239–264.
- [39] U. Maas. Efficient calculation of intrinsic low-dimensional manifolds for the simplification of chemical kinetics. *Comput. Visual. Sci.* 1 (1998), pp. 69–81.
- [40] David Edwards. *Elementary Kinetics of Soot Oxidation by OH*. PhD thesis. UC Berkeley, 2014.

- [41] Michael Frenklach and Hai Wang. Detailed Mechanism and Modeling of Soot Particle Formation. *Soot Formation in Combustion: Mechanisms and Models*. Ed. by Henning Bockhorn. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 165–192.
- [42] Jörg Appel, Henning Bockhorn, and Michael Frenklach. Kinetic modeling of soot formation with detailed chemistry and physics: laminar premixed flames of C2 hydrocarbons. *Combustion and Flame* 121.1 (2000), pp. 122–136.
- [43] Russell Whitesides and Michael Frenklach. Detailed Kinetic Monte Carlo Simulations of Graphene-Edge Growth. *The Journal of Physical Chemistry A* 114.2 (2010), pp. 689–703.
- [44] Ravi Singh. *Development and Implementation of Detailed Soot Surface Oxidation Model into Kinetic Monte Carlo Simulations of Graphene-Edge Oxidation*. PhD thesis. UC Berkeley, 2016.
- [45] M. Frenklach. Monte Carlo simulation of diamond growth by methyl and acetylene reactions. *J. Chem. Phys.* 97 (1992), pp. 5794–5802.
- [46] M. Frenklach. Monte Carlo simulation of hydrogen reactions with the diamond surface. *Phys. Rev. B* 45 (1992), pp. 9455–9458.
- [47] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* 81.25 (1977), pp. 2340–2361.
- [48] D. T. Gillespie. *Markov processes: An introduction for physical scientists*. San Diego, CA: Academic Press, 1992.
- [49] R. Whitesides and M. Frenklach. Effect of reaction kinetics on graphene-edge morphology and composition. *Zeitschrift für Physikalische Chemie* 229.4 (2015), pp. 597–614.
- [50] David E. Edwards, Dmitry Yu. Zubarev, William A. Lester, and Michael Frenklach. Pathways to Soot Oxidation: Reaction of OH with Phenanthrene Radicals. *The Journal of Physical Chemistry A* 118.37 (2014), pp. 8606–8613.
- [51] Ravi I. Singh, Alexander M. Mebel, and Michael Frenklach. Oxidation of Graphene-Edge Six- and Five-Member Rings by Molecular Oxygen. *The Journal of Physical Chemistry A* 119.28 (2015), pp. 7528–7547.
- [52] Ravi Singh and Michael Frenklach. A mechanistic study of the influence of graphene curvature on the rate of high-temperature oxidation by molecular oxygen. *Carbon* 101 (2016), pp. 203–212.
- [53] Jerome Sacks, William J. Welch, Toby J. Mitchell, and Henry P. Wynn. Design and Analysis of Computer Experiments. *Statist. Sci.* 4.4 (Nov. 1989), pp. 409–423.
- [54] M. Frenklach. Modeling. *Combustion Chemistry*. Ed. by W. C. Gardiner, Jr. New York: Springer-Verlag, 1984, pp. 423–453.
- [55] T. J. Sullivan. *Introduction to Uncertainty Quantification*. Cham, Switzerland: Springer, Cham, 2015.

- [56] R. Ghanem, D. Higdon, and H. Owhadi. *Handbook of Uncertainty Quantification*. Cham, Switzerland: Springer, Cham, 2017.
- [57] U.S. Department of Energy. *Scientific Grand Challenges for National Security: The Role of Computing at the Extreme Scale*. 2009.
- [58] M. Frenklach, A. Packard, P. Seiler, and R. Feeley. Collaborative data processing in developing predictive models of complex reaction systems. *Int. J. Chem. Kinet.* 36 (2004), pp. 57–66.
- [59] M. T. Reagan, H. N. Najm, P. P. Pébay, O. M. Knio, and R. G. Ghanem. Quantifying uncertainty in chemical systems modeling. *Int. J. Chem. Kinet.* 37 (2005), pp. 368–382.
- [60] Ilias Billionis, Nicholas Zabaras, Bledar A. Konomi, and Guang Lin. Multi-output separable Gaussian process: Towards an efficient, fully Bayesian paradigm for uncertainty quantification. *Journal of Computational Physics* 241 (2013), pp. 212–239.
- [61] Rohit Tripathy, Ilias Billionis, and Marcial Gonzalez. Gaussian processes with built-in dimensionality reduction: Applications to high-dimensional uncertainty propagation. *Journal of Computational Physics* 321 (2016), pp. 191–223.
- [62] Khachik Sargsyan. Surrogate Models for Uncertainty Propagation and Sensitivity Analysis. *Handbook of Uncertainty Quantification*. Ed. by R. Ghanem, D. Higdon, and H. Owhadi. Cham, Switzerland: Springer, Cham, 2017, pp. 673–698.
- [63] David Miller and Michael Frenklach. Sensitivity Analysis and Parameter Estimation of Dynamic Modeling of Chemical Kinetics. *International Journal of Chemical Kinetics* 15 (1983), pp. 677–696.
- [64] Marc Kennedy and Anthony O’Hagan. Bayesian calibration of computer models. *J. R. Statist. Soc. B* 63 (2001), pp. 425–464.
- [65] Youssef M. Marzouk, Habib N. Najm, and Larry A. Rahn. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics* 224.2 (2007), pp. 560–586.
- [66] P. Stark and L. Tenorio. A primer of Frequentist and Bayesian inference in inverse problems. *Large-Scale Inverse Problems and Quantification of Uncertainty* (2010), pp. 9–32.
- [67] J. Bell, M. Day, J. Goodman, R. Grout, and M. Morzfeld. A Bayesian approach to calibrating hydrogen flame kinetics using many experiments and parameters. *Combust. Flame* 205 (2019), pp. 305–315.
- [68] R. Feeley, P. Seiler, A. Packard, and M. Frenklach. Consistency of a reaction dataset. *J. Phys. Chem. A* 108.44 (2004), pp. 9573–9583.
- [69] T. Russi, A. Packard, R. Feeley, and M. Frenklach. Sensitivity analysis of uncertainty in model prediction. *J. Phys. Chem. A* 112 (2008), pp. 2579–2588.

- [70] T. Russi, A. Packard, and M. Frenklach. Uncertainty quantification: Making predictions of complex reaction systems reliable. *Chem. Phys. Lett.* 499 (2010), pp. 1–8.
- [71] X. You, T. Russi, A. Packard, and M. Frenklach. Optimization of combustion kinetic models on a feasible set. *Proc. Combust. Inst.* 33 (2011), pp. 509–516.
- [72] K. Braman, T. A. Oliver, and V. Raman. Bayesian analysis of syngas chemistry models. *Combust. Theory Model.* 17 (2013), pp. 858–887.
- [73] Jens Prager, Habib N. Najm, and Judit Zádor. Uncertainty quantification in the ab initio rate-coefficient calculation for the $\text{CH}_3\text{CH}(\text{OH})\text{CH}_3 + \text{OHCH}_3\text{C}(\text{OH})\text{CH}_3 + \text{H}_2\text{O}$ reaction. *Proceedings of the Combustion Institute* 34.1 (2013), pp. 583–590.
- [74] D. E. Edwards, D. Yu. Zubarev, A. Packard, W. A. Lester, and M. Frenklach. Interval prediction of molecular properties in parametrized quantum chemistry. *Phys. Rev. Lett.* 112 (25 2014), p. 253003.
- [75] D. R. Yeates, W. Li, P. R. Westmoreland, W. Speight, T. Russi, A. Packard, and M. Frenklach. Integrated data-model analysis facilitated by an Instrumental Model. *Proc. Combust. Inst.* 35 (2015), pp. 597–605.
- [76] A. Hegde, W. Li, J. Oreluk, A. Packard, and M. Frenklach. Consistency analysis for massively inconsistent datasets in Bound-to-Bound Data Collaboration. *SIAM/ASA J. Uncertainty Quantif.* 6 (2018), pp. 429–456.
- [77] M. Frenklach, A. Packard, and P. Seiler. “Prediction uncertainty from models and data”. *Proceedings of the 2002 American Control Conference (IEEE Cat. No. CH37301)*. Vol. 5. May 2002, pp. 4135–4140.
- [78] P. Seiler, M. Frenklach, A. Packard, and R. Feeley. Numerical approaches for collaborative data processing. *Optim. Eng.* 7 (2006), pp. 459–478.
- [79] David A. Sheen and Hai Wang. The method of uncertainty quantification and minimization using polynomial chaos expansions. *Combustion and Flame* 158.12 (2011), pp. 2358–2374.
- [80] David A. Sheen and Hai Wang. Combustion kinetic modeling using multispecies time histories in shock-tube oxidation of heptane. *Combustion and Flame* 158.4 (2011). Special Issue on Kinetics, pp. 645–656.
- [81] David A. Sheen, Claudette M. Rosado-Reyes, and Wing Tsang. Kinetics of H atom attack on unsaturated hydrocarbons using spectral uncertainty propagation and minimization techniques. *Proceedings of the Combustion Institute* 34.1 (2013), pp. 527–536.
- [82] M. Frenklach, A. Packard, G. Garcia-Donato, R. Paulo, and J. Sacks. Comparison of statistical and deterministic frameworks of uncertainty quantification. *SIAM/ASA J. Uncertainty Quantif.* 4 (2016), pp. 875–901.

- [83] Nestor V. Queipo, Raphael T. Haftka, Wei Shyy, Tushar Goel, Rajkumar Vaidyanathan, and P. Kevin Tucker. Surrogate-based analysis and optimization. *Progress in Aerospace Sciences* 41.1 (2005), pp. 1–28.
- [84] Alexander I.J. Forrester and Andy J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences* 45.1 (2009), pp. 50–79.
- [85] Michael Frenklach, Hai Wang, and Martin J. Rabinowitz. Optimization and analysis of large chemical kinetic mechanisms using the solution mapping method—combustion of methane. *Progress in Energy and Combustion Science* 18.1 (1992), pp. 47–73.
- [86] M. Frenklach, A. Packard, and R. Feeley. Optimization of reaction models with Solution Mapping. *Modeling of Chemical Reactions*. Ed. by R. W. Carr. Amsterdam: Elsevier, 2007, pp. 243–291.
- [87] R. H. Myers and D. C Montgomery. *Response Surface Methodology*. Wiley Series in Probability and Statistics. New York, NY, USA: Wiley, 2002.
- [88] George Box and Daniel Meyer. Some New Ideas in the Analysis of Screening Designs. *Journal of Research of the National Bureau of Standards* 90 (1985), pp. 495–500.
- [89] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014.
- [90] W. Li, A. Hegde, J. Oreluk, A. Packard, and M. Frenklach. *Bound-to-Bound Data Collaboration*. Last accessed March 20, 2020. 2016. URL: <https://github.com/B2BDC/B2BDC>.
- [91] James Oreluk. *Role of experimental data in validating and quantifying uncertainties in complex physical systems*. PhD thesis. UC Berkeley, 2019.
- [92] Arun Hegde. *Quantifying the agreement between computational models and experimental data under uncertainty*. PhD thesis. UC Berkeley, 2020.
- [93] S. Boyd and L. Vandenberghe. *Convex Optimization*. New York, NY, USA: Cambridge University, 2004.
- [94] Jos F. Sturm. Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software* 11.1-4 (1999), pp. 625–653.
- [95] MATLAB. *version 9.8.0 (R2020a)*. Natick, Massachusetts: The MathWorks Inc., 2020.
- [96] H. B. Palmer and C. F. Cullis. The formation of carbon from gases. *Chemistry and Physics of Carbon* (1965), pp. 265–325.
- [97] J. Lahaye and G. Prado. Morphology and Internal Structure of Soot and Carbon Blacks. *Particulate Carbon: Formation During Combustion*. Ed. by Donald C. Siegl and George W. Smith. Boston, MA: Springer US, 1981, pp. 33–55.

- [98] Donald Rivin and Avrom I. Medalia. A Comparative Study of Soot and Carbon Black. *Soot in Combustion Systems and Its Toxic Properties*. Ed. by J. Lahaye and G. Prado. Boston, MA: Springer US, 1983, pp. 25–35.
- [99] B S Haynes and H G Wagner. Soot formation. *Prog. Energy Combust. Sci.* 7 (1981), pp. 229–273.
- [100] Michael Frenklach and Lawrence B. Ebert. Comment on the proposed role of spheroidal carbon clusters in soot formation. *The Journal of Physical Chemistry* 92.2 (1988), pp. 561–563.
- [101] J.W. Ponder. *TINKER: Software tools for molecular design*. Washington University School of Medicine: Saint Louis, MO, 2004.
- [102] Norman L. Allinger, Young H. Yuh, and Jenn Huei Lii. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *Journal of the American Chemical Society* 111.23 (1989), pp. 8551–8566.
- [103] P. Roth, O. Brandt, and S. Von Gersum. High temperature oxidation of suspended soot particles verified by CO and CO₂ measurements. *Symposium (International) on Combustion* 23.1 (1991), pp. 1485–1491.
- [104] P. Boivin, C. Jiménez, A. L. Sánchez, and F. A. Williams. An explicit reduced mechanism for H₂-air combustion. *Proc. Combust. Inst.* 33 (2011), pp. 517–523.
- [105] P. Boivin, A. L. Sánchez, and F. A. Williams. Analytical prediction of syngas induction times. *Combust. Flame* 176 (2017), pp. 489–499.
- [106] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian Data Analysis*. Boca Raton, FL.: CRC, 2013.
- [107] W. Li, A. Hegde, A. Packard, and M. Frenklach. Uniform sampling of a feasible set (in preparation).
- [108] F. Liese and Klaus-J. Miescke. *Statistical Decision Theory: Estimation, Testing, and Selection*. New York, NY, USA: Springer-Verlag New York, 2008.
- [109] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Publishing Company, Incorporated, 2010.
- [110] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.* 35 (1943), pp. 99–109.
- [111] M. Frenklach. *ReactionLab*. Last accessed March 20, 2020. 1999. URL: <https://github.com/PrimeKinetics/ReactionLab>.
- [112] N. A. Slavinskaya, M. Abbasi, J. H. Starcke, R. Whitside, A. Mirzayeva, U. Riedel, W. Li, J. Oreluk, A. Hegde, A. Packard, M. Frenklach, G. Gerasimov, and O. Shatalov. Development of an uncertainty quantification predictive chemical reaction model for syngas combustion. *Energy & Fuels* 31.3 (2017), pp. 2274–2297.

- [113] A. Mirzayeva, N. Slavinskaya, U. Riedel, M. Frenklach, A. Packard, W. Li, J. Oreluk, and A. Hegde. Investigation of Dataset construction parameters and their impact on reaction model optimization using PrIME. *2018 AIAA Aerospace Sciences Meeting*. AIAA SciTech Forum. American Institute of Aeronautics and Astronautics, Jan. 2018.
- [114] Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [115] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.
- [116] Lisandro Dalcín, Rodrigo Paz, and Mario Storti. MPI for Python. *Journal of Parallel and Distributed Computing* 65.9 (2005), pp. 1108–1115.
- [117] Lisandro Dalcín, Rodrigo Paz, Mario Storti, and Jorge D’Elía. MPI for Python: Performance improvements and MPI-2 extensions. *Journal of Parallel and Distributed Computing* 68.5 (2008), pp. 655–662.
- [118] Lisandro D. Dalcin, Rodrigo R. Paz, Pablo A. Kler, and Alejandro Cosimo. Parallel distributed computing using Python. *Advances in Water Resources* 34.9 (2011). New Computational Methods and Software Tools, pp. 1124–1139.
- [119] *The National Energy Research Scientific Computing Center*. Lawrence Berkeley National Laboratory, Building 59, 1 Cyclotron Rd, Berkeley, CA 94720. URL: <https://www.nersc.gov/>.