The Dissertation Committee for William Ladd Gurecky
certifies that this is the approved version of the following dissertation:

A CFD-Informed Model for Subchannel Resolution Crud Prediction

Committee:

Derek Haas, Supervisor

Benjamin Leibowicz

Sheldon Landsberger

Kevin Clarno

Stuart Slattery

A CFD-Informed Model for Subchannel Resolution Crud Prediction

by

William Ladd Gurecky

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2018

To family and friends

Acknowledgments

I would like to acknowledge Dr. Kevin Clarno, Dr. Stuart Slattery, Dr. Derek Haas, Dr. Sheldon Landsberger, Dr. Ben Leibowicz and Dr. Robert Salko for their technical guidance and general advice throughout my graduate studies.

Special thanks to Dr. Erich Schneider who fostered many a students' interest in nuclear engineering & software development, including mine.

This research was supported by and performed in conjunction with the Consortium for Advanced Simulation of Light Water Reactors (http://www.casl.gov), an Energy Innovation Hub (http://www.energy.gov/hubs) for Modeling and Simulation of Nuclear Reactors under U.S. Department of Energy Contract No. DE-AC05-00OR22725.

A CFD-Informed Model for Subchannel Resolution Crud Prediction

William Ladd Gurecky, Ph.D.
The University of Texas at Austin, 2018

Supervisor: Derek Haas

A physics-directed, statistically based, surrogate model of the small scale flow features that impact Chalk River unidentified deposit (crud) growth is presented in this work. The objective of the surrogate is to provide additional details of the rod surface temperature, heat flux, and near-wall turbulent kinetic energy fields which cannot be explicitly captured by a subchannel code.

Operating as a mapping from the high fidelity computational fluid dynamics (CFD) data to the low fidelity subchannel grid (hi2lo), the model provides CFD-informed boundary conditions to the crud model executed on the subchannel pin surface mesh. The surface temperature, heat flux, and turbulent kinetic energy, henceforth referred to as the fields of interest (FOI), govern the growth rate of crud on the surface of the rod and the precipitation of boron in the porous crud layer. Therefore the model predicts the behavior of the FOIs as a function of position in the core and local thermal-hydraulic (TH) conditions.

The subchannel code produces an estimate for all crud-relevant TH quantities at a coarse spatial resolution everywhere in the core and executes substantially faster than CFD. In the hi2lo approach, the solution provided by the subchannel code is augmented by a predicted stochastic component of the FOI informed by CFD results to provide a more detailed description of the target FOIs than subchannel can provide alone. To this end, a novel method based on the marriage of copula and gradient boosting techniques is proposed. This methodology forgoes a spatial interpolation procedure for a statistically driven approach, which predicts the fractional area of a rod's surface in excess of some critical temperature but not precisely where such maxima occur on the rod surface. The resultant model retains the ability to account for the presence of hot and cold spots on the rod surface induced by turbulent flow downstream of spacer grids when producing crud estimates. Sklar's theorem is leveraged to decompose multivariate probability densities of the FOI into independent copula and marginal models. The free parameters within the copula model are predicted using a combination of supervised regression and classification machine learning techniques with training data sets supplied by a suite of precomputed CFD results spanning a typical pressurized water reactor TH envelope.

Results show that compared to the subchannel standalone case, the hi2lo method more accurately preserves the influence of spacer grids on the crud growth rate. Or more precisely, the hi2lo method recovers key statistical properties of the FOI which impact crud growth. Compared to gold standard high fidelity CFD/crud coupled results in a single assembly test case, the hi2lo model produced a relative total crud mass difference of -8.9% compared to the standalone subchannel relative crud mass difference of 192.1%.

Table of Contents

Acronyms

BHF	Boundary Heat Flux
CASL	Consortium for Advanced Simulation of LWRs
CDF	Cumulative Density Function
CFD	Computational Fluid Dynamics
CILC	Crud Induced Local Corrosion
CIPS	Crud Induced Power Shift
CRUD	Chalk River Unidentified Deposit (crud)
CTF	Coolant boiling in rod arrays–Two Fluid (COBRA-TF)
FOI	Field of Interest
GBRM	Gradient Boosted Regression Model
GBRT	Gradient Boosted Regression Tree
HTC	Convective Heat Transfer Coefficient
LANL	Los Alamos National Laboratory
LOO	Leave-one-out
LOOCV	Leave-one-out cross validation
LS	Least Squares
LWR	Light Water Reactor
ML	Maximum Likelihood
ORNL	Oak Ridge National Laboratory
PDF	Probability Density Function
PWR	Pressurized Water Reactor
ROM	Reduced Order Model
RV	Random Variable
TH	Thermal Hydraulic
TKE	Turbulent Kinetic Energy
VERA	Virtual Environment for Reactor Applications

Nomenclature

Functions & Maps

$c(\cdot)$	Copula density function
$C(\cdot)$	Copula cumulative density function
$\varphi(\cdot)$	Copula generator function
$f(\cdot)$	Marginal density function
$F(\cdot)$	Marginal cumulative density function
$\mathcal{G}(\cdot)$	Crud generator function
$h(\cdot)$	Joint density function
$H(\cdot)$	Joint cumulative density function
$\mathcal{F}(\cdot)_M$	Gradient boosted model
$\mathcal{R}$	Mapping from sample space to a location on the rod surface

Symbols

t	Time
T	Temperature
k	Turbulent kinetic energy
q''	Boundary heat flux
$\mathbf{p}$	Auxiliary predictive variables
Q_τ	Quantile function (inverse CDF)
q_τ	The τ^{th} quantile
ρ_τ	Kendall's tau
θ	Marginal distribution parameter
θ_c	Copula shape parameter
Θ_c	Archimedean copula family
$\mathbf{C}$	Crud state vector
C_m	Crud mass density
C_b	Crud boron density
C_t	Crud thickness
$\mathbf{X}$	Random vector
X	Random variable

List of Tables

List of Figures

1 | Introduction

The Consortium for Advanced Simulation of Light Water Reactors (CASL) selected several problems identified by industry partners as critical, inadequately understood, engineering-scale phenomena, which would provide financial and safety benefits to the nuclear power industry if resolved [1]. CASL supports technical challenges stemming from extending the operational lifetime of existing light water reactor plants with high performance neutronic, thermal hydraulic, and fuel performance software solutions. The problem of interest in this work is the prediction of Chalk River unidentified deposit (crud) growth rates. The growth of crud comes with neutronic and thermal hydraulic repercussions that are of interest to CASL's industry partners. In an effort to simulate the effects of crud on the power and burnup distribution, a code produced by a Los Alamos National Laboratory (LANL) and Oak Ridge National Laboratory (ORNL) collaboration under the name MAMBA was developed [2]. The development of the MPO Advanced Model for Boron Analysis (MAMBA) and other supporting Virtual Environment for Reactor Applications (VERA) tools provided a starting point for the high-to-low (hi2lo) methods at hand.

A phenomena known as crud-induced power shift (CIPS) is caused by the presence of elevated ^{10}B concentrations in the crud layer. Since crud is preferentially deposited on the fuel rods in hot regions of the core and ^{10}B is a strong neutron absorber, the crud buildup leads to a slight shift in power production toward the bottom of the core under steady-state operation. Crud induced power shift impacts the burnup distribution over a cycle, reduces shutdown margin, and is important to account for when computing thermal margins of the fuel [3]. The prediction of CIPS is especially important for older facilities seeking to uprate power output or extend their operational lifetime. Additionally, the presence of crud on the rod surface has been shown to exacerbate local oxide penetration rates of some zirconium alloys [4]. This is known as crud-induced local corrosion (CILC). Improvements in crud simulation techniques ultimately improve the ability to predict the CIPS and CILC phenomena for a given fuel loading pattern. If significant CIPS or CILC can be accurately predicted provided a candidate loading pattern, significant cost savings are possible by ensuring the target burnup is not missed due to the presence of excess crud in the core [3]. Loading patterns that would yield unfavorable crud buildup could be avoided provided an accurate and robust crud prediction capability is available for use in a production environment.

The Virtual Environment for Reactor Applications (VERA) is a key component of CASL's technical portfolio. The VERA meta-package integrates a variety of physics packages and multiphysics coupling options to form a robust reactor simulation capability. For multi-cycle depletion computations, VERA relies upon the Michigan Parallel Characteristics Based Transport (MPACT) code, a 2-D/1-D method of characteristics neutronics package, coupled with the subchannel thermal hydraulics code, Coolant Boiling in Rod Arrays–Two Fluid (CTF). An integrated crud modeling capability is provided by MAMBA to address the CIPS challenge problem.

To reduce computation times, the subchannel TH code discretizes the reactor domain into large, centimeter scale finite volumes. As a consequence of this discretization scheme, sub-centimeter scale thermal hydraulic effects of the spacer grids on crud are averaged over large regions on the fuel rods' surfaces. Though small scale phenomena are not explicitly modeled, they are approximately accounted for in a variety of empirically derived closure relations. In effect, a single constant estimate for the mean thermal hydraulic conditions is obtained in each finite volume.

Previous hi2Lo focused work in CASL focused on using experimental or computational fluid dynamics (CFD) data sets to improve heat transfer and turbulent mixing models in CTF. These studies focused on correcting biases in the bulk-average behavior of the flow (due to the previously neglected physics). Examples of such hi2lo models are given in chapter 2.

The traditional approach must be modified to accommodate the CILC and CIPS challenge problems. Here arises the need to retain not only the effect of fine-scale physics on the bulk, but also to predict if certain temperature or near-wall turbulent kinetic energy (TKE) thresholds are exceeded in a particular subchannel volume. Furthermore, for a complete characterization of thermal hydraulic impacts on crud growth, the scale-bridging model must describe the frequency distribution of extreme TH events above a given threshold.

1.1 Significance and Novelty

Crud growth is dominated by threshold physics [5]. Hot and cold spots present downstream of spacer grids must be accurately resolved by the hi2lo model to predict the maximum crud thickness and boron precipitation within the crud layer.

It is challenging to faithfully capture the peaks and valleys in rod surface temperature and TKE distribution by traditional interpolation techniques because such a model must guard against smearing out the sharp peaks present in the spatial distributions. In the present method we forgo a spatial shape function mapping strategy for a statistically driven approach that predicts the fractional area of a rod's surface in excess of some critical temperature but not precisely where such maxima occur.

In this approach, the method does not predict the fine scale flow and temperature field on the pin surface; rather, this approach estimates the joint temperature, TKE, and BHF probability density on coarse, centimeter scale patches on the rod surface. The size and position of the coarse patches is congruent with the coarse fidelity subchannel grid. The goal is retain the minimum amount of information required to predict the correct total amount of crud harbored in each coarse surface patch. The amount of crud deposition downstream of spacer grids is influenced by the presence of hot and cold spots present due to the turbulent flow induced by mixing vanes. Crud is highly sensitive to the rod surface temperature, particularly around the saturation point, and therefore it is important to account for these small scale flow features when providing boundary conditions to the crud simulation.

By capturing the action of local hot and cold spots on the crud deposition rate, the hi2lo method accounts for more physics when making predictions of the total integrated boron mass in the crud layer than a subchannel code could provide alone. An improvement in crud predictions in the immediate vicinity of mixing vanes results in an overall improvement in CIPS predictions since both the total integrated boron mass within the crud layer as well as the axial distribution of crud is of principle importance when predicting CIPS. Additionally, the ability to estimate the likelihood of extreme crud buildup events enables the hi2lo methods developed in this work to function as a CILC scoping tool. It is envisioned that such a tool will identify potential CILC hot spots where a significant amount of cladding is consumed by oxide ingress, resulting in potential fuel failure. The effectiveness of the model in this role is governed by the magnitude of propagated uncertainties through the hi2lo model.

Prior to this work, hi2lo efforts directed at improving subchannel thermal hydraulic predictions generally used CFD results as a data source to calibrate corrective or closure terms in the subchannel flow models, such as grid loss or mixing coefficients. Other efforts used the CFD data as a data source to construct spatial downscaling maps of key surface fields impacting crud growth. A statistically based CFD-informed subchannel downscaling implementation is novel, particularly as a means for improving crud predictions in a core simulator.

1.2 Crud Background

The buildup of crud results from the deposition of metal particulates and corrosion products entrained in the primary coolant loop of a light water reactor on the exterior surface of the fuel rods. These impurities arise from erosion and corrosion processes elsewhere in the loop. Of all the coolant impurities, the largest contributor the initial formation of a crud layer on the outer cladding surface is nickel ferrite. The initial build up of nickel ferrite may be described by the ordinary differential equation (ODE) shown in equation 1.1.

$$\frac{dN_{\mathrm{NiFe},c}}{dt} = (\alpha_{\mathrm{nb}} + \alpha_b q_b'')N_{\mathrm{NiFe,cool}} - \gamma_k k \tag{1.1}$$

Where $N_{\mathrm{NiFe},c}$ is the concentration of nickel ferrite in the crud within a small finite volume on the cladding surface. $N_{\mathrm{NiFe,cool}}$ is the concentration of nickel and iron impurities in the coolant. α_b and α_{nb} represent boiling and non-boiling rate constants respectively. The boiling component of the boundary heat flux (BHF) on the outer cladding surface is given by $q_b''[W/m^2]$. Note that q_b'' is only non-zero when $T > T_{sat}$. γ_k is an erosion multiplier and k is the near-wall local TKE. Crud typically forms where temperatures are high and where subcooled boiling occurs on the rod surface.

The primary porous matrix of crud is $NiFe_2O_4$; however, there are other constituents such as nickel oxide, Ni_2FeBO_5 and $Li_2B_4O_7$ compounds [6] [5]. In particular, the porous matrix of $NiFe_2O_4$ is filled in by precipitated $Li_2B_4O_7$ in regions that experience boiling, thus trapping boron inside the crud layer. The net result of the trapped boron in the crud layer is a shift in power toward the bottom of the core.

For the purpose of pressurized water reactor (PWR) core simulation crud is modeled at the core-wide scale. Typically TH boundary conditions are supplied to the crud simulation code by subchannel models in this application. Additionally, high fidelity CFD/crud coupling work has been conducted that predicted striping patterns, or high variations in azimuthal crud growth, downstream of spacer grids [7]. The coupled CFD/crud results were shown to be qualitatively consistent with the available experimental crud scrape data, which also shows high azimuthal variation downstream spacer grids [8]. In contrast, no such striping patterns are resolved by the subchannel model.

Three primary concerns were identified with the current state-of-the-art crud models used in core simulators in multi-cycle depletion applications. The first concerns passing incorrect boundary conditions to the crud model. Handing incorrect boundary conditions to the crud model will not produce the correct crud unless an a posteriori factor is applied to counteract the effects of poorly resolved boundary conditions supplied by the subchannel TH models. Errors resulting from poorly resolved boundary conditions is most severe downstream of spacer grids in situations where a subchannel code cannot resolve fine scale flow features that influence crud growth. The current work addresses this problem by improving the accuracy of the boundary conditions handed to the crud model by leveraging a suite of precomputed CFD results.

The second issue pertains to the physics models in the current crud model implementation. There are missing or incomplete models for the formation of nickel oxide in the crud layer, incorrect pore fill kinetics, and incorrect crud model parameters including parameters governing chimney heat transfer rates, Arrhenius rate constants, and species diffusion constants. These should be addressed via experiment and Bayesian model calibration which is beyond the scope of this work.

Finally, the source and rate at which primary loop impurities buildup over time has come into question. Different PWR designs of varying vintage have different metallurgy and components in the primary coolant loop circuit. These inconsistencies make it non-trivial to predict the release rate of nickel and iron impurities into the coolant loop in each of these plants. Determining the source term magnitude from these primary loop corrosion- and erosion-born impurities is an area of ongoing research.

1.3 Subchannel Background

It is helpful to review subchannel terminology before exploring CTF specific crud applications. The CTF theory manual provides a detailed explanation of the subchannel discretization and the geometric terms used in subchannel codes [9]. Figure 1.1 shows a top down view of four pins in a typical PWR lattice arrangement. The subchannel is filled with diagonally hashed lines. Each subchannel contacts four surrounding pins and the wetted surface formed between the pin and subchannel is referred to as a CTF face throughout this dissertation. In CTF, each pin's outer cladding surface is divided into four azimuthal segments.

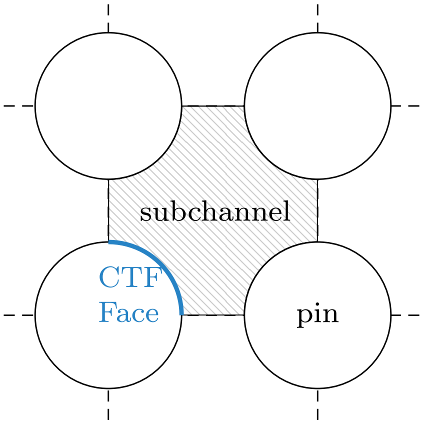


Figure 1.1: Top down view of the subchannel discretization of a PWR pin configuration.

For the typical PWR rods arrangements considered in this work, the rod is axially divided into approximately D2 centimeter segments. A 3-D depiction of the axial subchannel discretization is given in figure 1.2. Additionally in this figure, a CTF patch is highlighted in blue. A CTF patch and CTF face will be used interchangeably throughout this document as they both refer to a small centimeter-scale patch on the rod surface in contact with a neighboring subchannel.

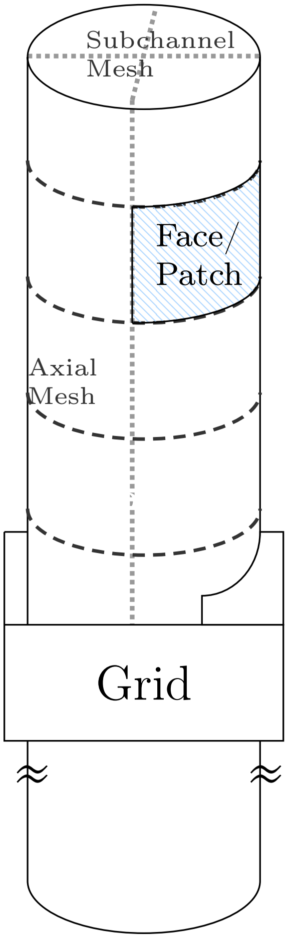


Figure 1.2: A subchannel discretization superimposed over a 3-D representation of a single pin.

1.4 Hi2lo Discussion

Hi2lo, or high to low modeling, implies that a source of high fidelity gold standard data produced by an expensive to evaluate physics model is used to downscale and augment a low fidelity model of the same physics. Provided sparsely available high fidelity data, this mapping must be possible even in the case where matching high and low fidelity results do not exist. Similarly, the hi2lo strategy put forward in this work may be viewed as a particular implementation of a statistical downscaling (SD) model, of which a large variety exist in the literature and some of which are described in chapter 2. One interesting challenge in this work that is atypical of SD models is the requirement of co-prediction of multiple correlated fields, sometimes this is referred to as multiple target regression.

It is assumed that the flow of information is unidirectional from the high fidelity data to the low fidelity model. Feedback between the disparate scale models is not included. This simplification is commonly made in the application of a statistical downscaling model. A tight coupling between multiscale transport models is the subject of dynamical downscaling and is beyond the scope of this document.

Generally, a surrogate model replaces expensive-to-evaluate physics with a quick-to-evaluate model that preserves some aspects of the physics. The hi2lo strategy seeks to capture the action of high fidelity CFD resolved flow phenomena on crud growth without having to run the CFD model outright. However, a key difference between the hi2lo model pursued in this work and a canonical dynamical system surrogate is that the hi2lo model does not seek to behave as a stand-in for a differential equation and should not be confused as such.

1.5 Document Layout

This dissertation is structured into five major sections, excluding the introduction. Chapter 2 begins with a review of statistical downscaling and Gaussian process regression procedures for making predictions between sparsely known data samples with uncertainty estimates. Chapter 2 also includes a discussion of previously conducted CFD-informed subchannel work.

The development of the hi2lo methodology for improving crud predictions is provided in chapter 3. The theory section covers copula, marginal density reconstruction from quantiles, and importance sampling. Chapter 3 also covers the application of the method to the time-dependent crud growth problem and an overview of the machine learning strategy used in this work, gradient boosting.

In chapter 4 the hi2lo method is applied to a synthetic-CFD single pin, single state point data set. The ability of the hi2lo method to recover key properties of the synthetic data set is demonstrated. In this chapter machine learning is absent since the target and the supplied synthetic data are co-located in TH state space. This section servers as an integration test for the copula and marginal density fitting routines and as a test bed for the Monte Carlo sampling routines.

The use of a gradient boosted machine as a means to make inferences about the jointly distributed fields on the rod surface given local core conditions supplied by CTF is discussed in chapter 5. Here, results from the machine learning model are presented alongside crud predictions. A small 5x5 pin assembly operating at nominal PWR conditions was modeled using a CFD package to generate the necessary training data.

Chapter 6 serves to draw conclusions from the results and to supply avenues for future work.

2 | Literature Review

Augmenting and bias-correcting coarse fidelity thermal hydraulic predictions provided by a quickly executing subchannel code by using higher fidelity CFD results can be viewed as a statistical downscaling problem. In this section previous efforts to tackle related problems in downscaling coarse fidelity data are considered. There is abundant usage of statistical downscaling techniques in the weather forecasting and geostatistics spaces and therefore these fields are responsible for developing and investigating a myriad of downscaling methods.

This section begins with an overview of statistical downscaling techniques followed by a pointed review of past hi2lo work directed at improving subchannel codes. Finally, past subchannel hi2lo efforts are connected with an interpolation procedure known as kriging. It is shown that kriging decomposes the hi2lo problem into mean-predicting and stochastic components. This general decomposition strategy will be slightly modified to accommodate the simultaneous prediction of correlated random fields and applied to the crud problem in the following chapters.

One commonality across all studied procedures is the presence of a high and low fidelity data source and a goal to make credible predictions of the target field between known coarsely resolved sample locations. The problem is one of data amalgamation, where the resultant downscaling model preserves some average aspects of the low fidelity model with the added benefits of uncertainty and spatial fidelity afforded by the finer scale data.

2.1 Statistical Downscaling

Statistical downscaling (SD) methods attempt to preserve statistical properties of historical, high fidelity data when making forecasts in time or space using a model. The forecasting model typically executes quickly and has low spatial and temporal resolution in order to reduce computation times. In general, the goal at runtime of the SD-enhanced low fidelity forecasting tool is to obtain mean and higher moment estimates of a random field at a fine resolution. This setup is analogous to the hi2lo problem at hand. The subchannel thermal hydraulics model is acting as the coarse fidelity model and a pre-computed set of high fidelity CFD computations is available to augment and improve the fidelity of the subchannel predictions.

In the climate community it is common to perform local bias-correction of coarsely resolved weather models so that the results retain some specified properties of past historical statistical trends [10] [11]. In climate studies low fidelity data is typically provided by a coarsely resolved global circulation model (GCM) and a secondary set of finely resolved local rain and wind field measurements are provided by local weather stations, satellite or radar sources [12]. In addition to the longitude and latitude of these measurements, the fine scale data may also be associated with auxiliary features at such as the terrain height.

Precipitation estimates provided by statistically downscaled climate models are used as a boundary condition to local hydrology models for runoff [13], flood [14], and aquifer replenishment studies. A strong parallel with the current crud simulation work may be drawn. Subchannel TH results are bias-corrected and augmented before being passed to a corrosion chemistry or crud simulation package. The problem is similar to the highly threshold sensitive crud problem because flood risk models require accurate predictions for the frequency and magnitudes of extreme rainfall events which are difficult to quantify with coarse scale GCMs alone. Similarly, crud prediction requires accurate prediction of extreme cladding surface temperatures occurring in coincidence with low local turbulent kinetic energies.

A particular class of SD methods known as bias-corrected spatial disaggregation (BCSD) rely on quantifying the biases between coarsely resolved model predictions and a secondary source of temporally and spatially fine scale data [13]. In this method, the spatially and temporally high resolution data is aggregated to the coarse scale GCM grid as a preprocessing step. Residuals between the coarse GCM predictions and the aggregated fine scale data sets are computed. Next, percentiles of the residual distributions in each coarse patch are computed. A mapping is established between the computed percentiles, taken as the output, the geographic location and the GCM coarse fidelity outputs.

Upon evaluation of the coarse fidelity model at some future desired forecast date, the established mapping function is inverted by supplying the desired geographic coordinates and interpolated coarse fidelity model results in order to obtain estimated percentiles. A bias-correction and spatial disaggregation step is then applied to obtain bias-corrected estimates on a fine grid. A multiplicative random cascade model which is statistically uniform on small length scales but exhibits high spatial volatility has been employed in the spatial disaggregation step [14].

The majority of BCSD literature does not consider the simultaneous prediction of multiple correlated random fields; however, simultaneously predicting correlated random fields has been addressed through the use of copula [15], though all studied implementations of copula enhanced SD employ parametric models for the marginal and copula distributions. Furthermore, resolving fine spatial detail of the temperature and TKE fields in a given CTF face isn't necessary for accurate crud prediction when using a single dimensional crud simulation code because no azimuthal or axial variation in these surface fields are utilized by the crud package. Therefore, the problem of finding the fractional area of a CTF face which exists above a threshold is a viable alternative to spatial disaggregation techniques in the current hi2lo crud application.

It is possible to nest a high fidelity simulation within a coarse fidelity weather simulation. Boundary conditions and constraints are supplied by the coarse fidelity model to the nested regional high resolution model. The practice of coupling regional weather models with coarse scale global models is sometimes referred to as dynamical downscaling [16], though, this modeling strategy can also be viewed as a particular implementation of a tightly coupled multiscale model. The construction of dynamical downscaling models are not the focus of the current hi2lo work and will not be discussed further.

2.2 Subchannel Hi2lo

The utilization of CFD data to improve subchannel thermal hydraulic models does not necessarily take on a statistical downscaling characteristic. Oftentimes the strategy by which one uses CFD data to improve a subchannel model can be developed using standard Bayesian inference techniques in which subchannel model parameters are inferred through comparing the low fidelity model to high fidelity experimental or CFD data. This typifies an inverse problem which oftentimes involves model calibration, model selection and experimental design aspects. A wide array of literature exists on each of these topics and will not be interrogated here. Instead, a pointed literature review of the latest CFD-informed subchannel work is considered.

M. Avramova developed CFD informed grid mixing models in CTF. Avramova's work leveraged CFD results to improve the grid-enhanced cross flow and turbulent mixing models in CTF [17]. The lateral momentum equations implemented in CTF are provided in equation 2.1.

$$\begin{aligned}
&\frac{\partial}{\partial t}(\alpha_l \rho_l \mathbf{U}_l) + \nabla \cdot (\alpha_l \rho_l \mathbf{U}_l \mathbf{U}_l^T) \\
&= \alpha_l \rho_l \mathbf{g} - \alpha_l \nabla P + \nabla \cdot (\alpha_l \boldsymbol{\tau}_l) \\
&+ M_l^L + M_l^d + M_l^T + M_l^{GDXF} \qquad (2.1)
\end{aligned}$$

Where l denotes the liquid phase and α is the volume fraction liquid, $\boldsymbol{\tau}$ represents the shear tensor, P is the static pressure, $\mathbf{U}$ is the velocity vector, ρ_l is the liquid phase density, and $\mathbf{g}$ is the gravitational acceleration vector. The terms $M^L, M^d, M^T, M_l^{GDXF}$ account for droplet or bubble entrainment, phase interfacial drag, turbulent mixing and grid directed cross flow respectively. Avramova devised a method to use CFD computations to obtain an accurate prediction of M_l^{GDXF} for a variety of grid designs. The grid directed cross flow momentum source term used in Avramova's model is defined by equation 2.2.

$$M_l^{GDXF} = f_{sg}^2(z) u_l^2 \rho_l A_g S_g \tag{2.2}$$

Where u_l is the axial liquid velocity, A_g is the subchannel gap area, S_g is a constant which takes on a value in $\{-1, 0, 1\}$ depending on the vane orientation. The cross flow factor, f_{sg}, is given by equation 2.3.

$$f_{sg}(z) = \frac{V_l^{CFD}(z - z_{in})}{U_{in}^{CFD}} \tag{2.3}$$

U_{in}^{CFD} is the subchannel average axial inlet velocity to the spacer grid under consideration and $V_l^{CFD}(z - z_{in})$ is the subchannel averaged CFD predicted lateral velocity downstream from the spacer grid.

The effectiveness of the grid enhanced cross flow model was determined by comparing exit bulk temperature profiles across a variety of assembly designs against experimental and CFD results. The results indicated a marked improvement in the rod-assembly outlet temperature distribution at little additional computational cost as compared to CTF without CFD informed grid enhanced cross flow corrections. Aramova's work succeeded in reproducing the correct bulk fluid behavior near spacer grids in CTF as compared to gold standard CFD results; however the goal was not to recover small scale flow features. A different approach is required to capture the influence of spacer grids on the crud deposition rate.

The next bodies of work are closer in alignment with traditional downscaling techniques. These hi2lo procedures are not statistical in nature, but rather seek to correct spatial biases in the field predictions made by a low fidelity subchannel code using a purely deterministic multiplier mapping procedure. The corrective multiplier maps are derived from either experimentally gathered or CFD sources.

S. Yao et al. developed an empirical model of the heat transfer coefficient downstream of spacer grids [18]. An empirical relationship between the Nusselt number ratio and the vane angle, ϕ, blockage ratio ϵ, dimensionless distance from the grid, x/D, and fraction of flow area impeded by the vanes, A, was produced. This relationship is provided in equation 2.4.

$$\frac{Nu}{Nu_0} = \left[1 + 5.55\epsilon^2 e^{-0.13(x/D)}\right] + \left[1 + A^2 \tan^2\phi e^{-0.034(x/D)}\right] \tag{2.4}$$

Where the first term accounts for the effect of grid flow restriction and the second term represents the contribution of vane induced swirl on the heat transfer. A graphical representation of Yao's model fit to experimentally determined Nusselt number data for a variety of grid designs is shown in figure 2.1.

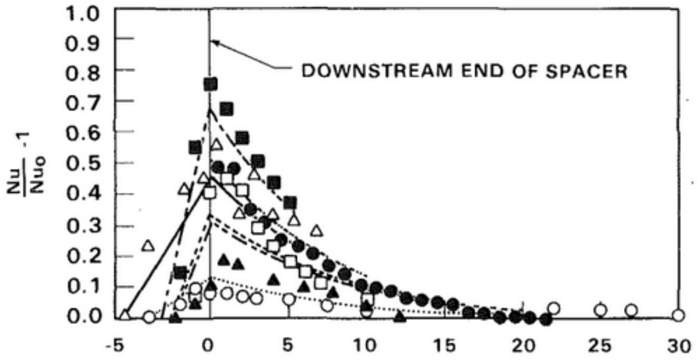


Figure 2.1: S. Yao empirical Nusselt number ratio vs. distance from upstream spacer grid plotted for a variety of grid designs [18].

This work is important because it represents an early approach to building experimentally informed hi2lo subchannel models. The Yao model is still employed by modern subchannel codes such as CTF to obtain more accurate rod surface temperature distributions near the spacer grids.

Similar to Yao's approach for capturing rod-enhanced heat transfer, B. Salko et al. developed a CFD-Informed hi2lo spatial remapping procedure for CILC/CIPS screening [19]. Rather than establishing a general empirical relationship between grid geometric features and the flow field, Salko developed grid specific maps. The developed multiplier maps are applicable only to the grid designed on which they are based. In contrast to Yao's approach, this approach enables the retention of much higher resolution flow field features in the multiplier maps. In addition to generating heat transfer multiplier maps, this method developed a TKE mapping procedure since both fields are required for accurate crud predictions. Both maps are applied in conjunction to a baseline CTF result to produce grid enhanced surface temperature and TKE distributions at runtime of the CTF model.

First, an intermediate coupling mesh is constructed on the rod surface with a resolution between the CFD mesh and the CTF grid. Next, the raw CFD surface fields are then mapped to the coupling mesh. In this approach crud is to be grown on the intermediate coupling grid. In theory, this grid can be refined to be equivalent to the CFD mesh size and indeed this would reduce interpolation error in the hi2lo procedure [19].

Shown in equation 2.5, the multiplier maps capture the ratio of the CFD predicted HTC and TKE surface distributions to the same surface distributions on a bare rod without spacer grids present. The bare rod heat transfer coefficient is denoted by h_0 and the grid-influenced heat transfer coefficient surface field is denoted by h_{cfd}.

$$\mathbf{m}_h = \frac{(Nu)_{cfd}}{(Nu)_0} = \frac{h_{cfd} L_{cfd} k_0}{h_0 k_{cfd} L_0} \tag{2.5}$$

Where Nu is the Nusselt number. Assuming equal length scales, L, and thermal conductivities, k, the Nusselt number ratio simplifies to equation 2.6.

$$\mathbf{m}_h = \frac{h_{cfd}}{h_0} = \frac{q''_{cfd}(T - T_\infty)_0}{q''_0(T - T_\infty)_{cfd}} \tag{2.6}$$

It is important to note that a uniform heat flux, q'', is used in both the bare and full gridded rod case so that $q''_{cfd}/q''_0 = 1$. The HTC remap is applied to the original CTF HTC by equation 2.7.

$$\hat{h}_l = \mathbf{m}_h h_{ctf} \tag{2.7}$$

Where $\hat{h}_l$ is the hi2lo remapped convective heat transfer coefficient. In CTF the wall heat transfer is split between phases:

$$q'' = q''_{conv} + q''_{boil} = (\hat{h}_l)(T - T_\infty) + q''_{boil}(T) \tag{2.8}$$

In order to compute augmented hi2lo surface temperatures several iterations are required to converge upon the correct surface temperature, $\hat{T}_s$, due to the surface boiling term as shown in algorithm 1.

Algorithm 1
Heat transfer coefficient map based hi2lo method for crud prediction (Salko. et. al.).

```
Initialization
Guess $T_s^{i=0} = T_0$. Maximum number, $N$ iterations.
for i in range($N$): do
    Evaluate effective multiphase CTF HTC: $h_{eff} = h_{ctf}(T_s^i, \hat{h}_l, q'')$
    Compute new hi2lo surface temperatures: $T_s = \frac{q''}{h_{eff}} + T_\infty$
    Under relax $T_s^{i+1} = \omega T_s + (1-\omega)T_s^i$; $\omega < 1$.
    break if: $|T_s^{i+1} - T_s^i| < tol$
end for
return: $\hat{T}_s = T_s^{i+1}$
```

Where $h_{ctf}(\cdot)$ is a callable CTF function that returns an effective multiphase HTC, h_{eff}. An under relaxation factor, ω, is supplied to aid convergence of the fixed point iterations at high heat fluxes since the function $h_{ctf}(\cdot)$ is nonlinear in surface boiling regimes approaching departure from nucleate boiling. Additional details on surface boiling heat transfer behavior are given in appendix D.

The TKE remap is constructed by evaluating the ratio given in equation 2.9 on all CTF faces.

$$\mathbf{m}_k = \frac{k_{cfd}}{k_0} \tag{2.9}$$

Where k_0 is the TKE distribution for a bare rod without spacer grids. The TKE multiplier map is applied in the same manner as the HTC map.

$$\hat{k} = \mathbf{m}_k k_{ctf} \tag{2.10}$$

Crud is grown on the coupling mesh using augmented temperature and TKE surface fields. By this method the integrated crud mass over a CTF face is given by equation 2.11.

$$C_m = \frac{1}{A}\sum_i^N a_i \mathcal{G}(\hat{T}_{s_i}, \hat{k}_i, q_i'') \tag{2.11}$$

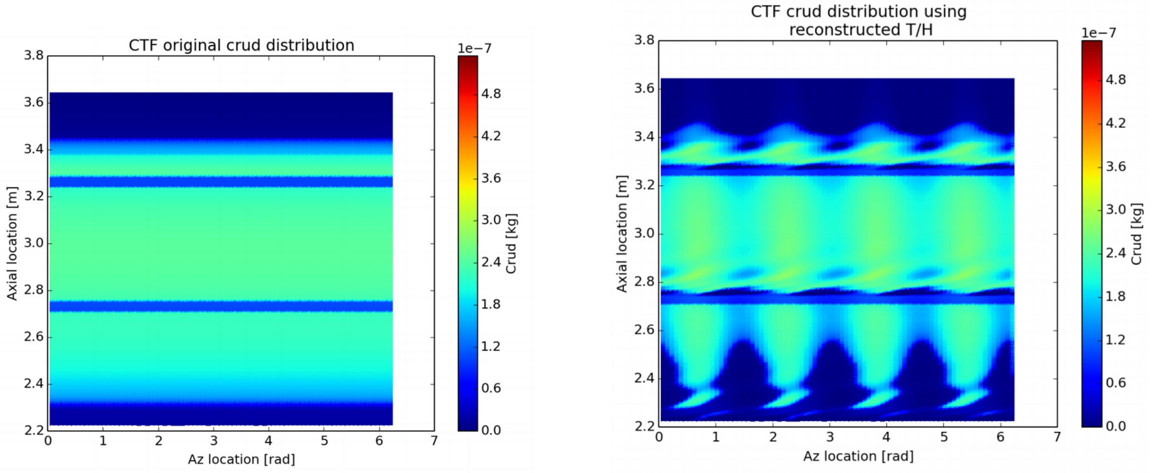
Where A is the area of the CTF face and a_i is the area of each cell face on the crud coupling mesh. The crud generation function is denoted by $\mathcal{G}$ and takes the surface temperature, TKE, and boundary heat flux as parameters. The impact of the simultaneous application of both the HTC and TKE maps on the crud distribution are shown for a single rod in figure 2.2. In the base case without the hi2lo maps applied, no azimuthal variation is observed in the crud distribution for this single quarter symmetric test case. However, when the hi2lo maps were employed the influence of the spacer grids on the crud distribution becomes visible.

A key assumption that the multiplier maps are insensitive to flow rate was made in the first implementation of this downscaling technique. However this assumption is not strictly true: The multiplier maps carry some dependence on the inlet flow conditions. An increase in flow rate changes the shape and extent of the wake region downstream of spacer grids which impacts the rod surface temperature and TKE fields.

An extension of the multiplier map hi2lo procedure could linearly interpolate between multiplier maps developed at high and low inlet flow rate conditions.

$$
\begin{aligned}
\mathbf{m}_k &= \alpha \mathbf{m}_k^h + (1-\alpha)\mathbf{m}_k^l \\
&= \alpha \frac{k_{cfd}^h}{k_0^h} + (1-\alpha)\frac{k_{cfd}^l}{k_0^l} \\
\alpha &= \frac{\dot{m}_i - \dot{m}_i^l}{\dot{m}_i^h - \dot{m}_i^l}
\end{aligned}
$$

Where $\dot{m}_i$ is the inlet mass flow rate. The superscript, $(\cdot)^l$, represents low flow conditions and $(\cdot)^h$ represent high flow conditions.



(a) CTF/MAMBA crud predictions without hi2lo remapping on a quarter symmetric pin.

(b) CTF/MAMBA crud predictions using hi2lo remapping on a quarter symmetric pin.

Figure 2.2: The impact of spatial HTC hi2lo remapping on CTF/MAMBA crud predictions [19].

Some simplifications are made in the application of this mapping. For a given assembly, the multiplier maps have been shown to have a high span to span repeatability. Therefore, a representative map is derived from a single span in a fully developed flow field. The representative map is then applied to all other spans in the model.

The multiplier map may not be transferable to other assemblies in the core due to geometric effects including the orientation of neighboring assemblies and TH/neutronic feedbacks. This represents a limitation to the spatial mapping procedure as unique maps must be generated for different assemblies in the core.

T. Blyth produced CFD informed grid enhanced heat transfer models for the advanced subchannel code, CTF [20] [21]. Blyth's work presented strategies for processing CFD data for use in generating enhanced heat transfer maps and for computing form loss coefficients across spacer grids. Blyth's work served as a precursor to Salko's CFD informed method for developing HTC and TKE maps. Blyth's grid enhanced heat transfer model followed the form given in equation 2.12 which was inspired by the approach taken by Yao and latter applied by Salko.

$$\mathbf{m}_h = \frac{h_{cfd}}{h_0} \tag{2.12}$$

Results from this work indicated that the a CFD driven hi2lo approach could capture more intricate details of the flow field when compared to the Yao heat transfer enhancement model. These intricate details were later found to be important to account for when modeling crud on the rods' surface [7]. This was expected because the spatial fidelity targeted by the approach of Blyth and Salko was fundamentally different than Yao's previous work. Furthermore, in contrast to the Yao model which can be tuned to accommodate different vane angles and blockage ratios, Blyth's approach requires CFD computations for each grid design of interest. As a consequence the hi2lo approaches developed by Blyth and Salko require a large up front computational cost driven by the necessary CFD computations for each grid design of interest.

2.3 Kriging

Taking Salko and Blyth's work as a starting point, one might consider developing an interpolating model built from a library of CFD computations which produces a hi2lo spatial map of the form indicated by equation 2.12. The method would allow interpolation of the hi2lo map between known geometric configurations and core states at which the upfront CFD computations were performed. The predicted hi2lo map from this procedure would also need to produce error bounds on the interpolated spatial HTC and TKE field maps. If the model's hi2lo field mapping errors follow a Gaussian-like distribution then kriging could be a suitable candidate to produce the desired geometry and flow dependent HTC maps. In this case the errors are defined as the model HTC prediction subtracted from the gold standard CFD HTC predictions.

Kriging was originally developed to address the problem of finding the most probable location of quality gold ore deposits given previous sparse samples of the surrounding ore body [22]. This interpolation method centered around modeling the spatial-autocorrelation of a random field in an effort to make credible predictions of the spatial distribution of gold ore concentrations given sparse, uncertain estimates [23]. The technique can be viewed as a special case of Gaussian process regression [24]. Kriging is related to Gaussian process regression since the underlying goal of both approaches is to model the spatial autocorrelation of a random field. This section will use the kriging nomenclature, however, the literature on Gaussian process regression can be useful in similar or identical contexts.

Since its inception, kriging has been employed to build surrogate models of complex physics where mechanistic models are unavailable. Notably, kriging approaches have been used to construct a surrogate model of aerospike nozzle performance to enable efficient optimization of many design parameters [25]. Similarly, kriging has been applied to airfoil design optimization [26]. In theses applications the kriging model fits into an optimization framework where the kriging model is used to build a response surface that is paired with a heuristic acquisition function to determine which parameter values are expected to yield the greatest design improvement. Kriging techniques have also been used to build spatial-temporal surrogates of rainfall for the assessment of flooding risks [27]. It is under the context of spatial interpolation where kriging is particularly relevant to the hi2lo problem at hand. Kriging is generally applicable when estimates of the mean and variance of a random field are desired in between sparse training data samples. Next, a brief introduction to the kriging procedure is given followed by an the application to a CFD and CTF data source.

Regression kriging (RK) decomposes interpolation problem into mean-predicting and bias-correcting residual models [28]. In an RK framework the spatial-autocorrelation in the residuals, computed by subtracting the mean from available fine scale field estimates, is captured by a covariance model. The mean response may be generated by a variety of regression strategies, with a common choice being an ordinary least squares model though works which investigate the use of random forests or more advanced machine learning strategies in this role are pervasive [29] [30]. In this application, the subchannel code, CTF, provides the mean thermal hydraulic predictions.

The general approach to the regression kriging problem is given in equation 2.13 where the surface temperature field, $T(z)$, is decomposed into a deterministic mean, $\mu_{T,\mathrm{ctf}}$, and a stochastic component, ϵ, where z represents the axial and azimuthal coordinates on the rod surface.

$$T(z) = \mu_{T,\mathrm{ctf}}(z) + \epsilon, \ \epsilon \sim \mathcal{GP}(0, K(\mathbf{z}_1, \mathbf{z}_2; \theta)) \tag{2.13}$$

Here ϵ is a zero mean *Gaussian process* [31]. Where $K(\mathbf{z}_1, \mathbf{z}_2; \theta)$ is a function with free parameters, $\theta = \{\theta_0, \theta_1\}$, that generates a valid covariance matrix. Assume that the mean temperature field, $\mu_{T,\mathrm{ctf}}$, is given by the subchannel code and that fine scale CFD temperature field data, T_{cfd}, is available at locations $\mathbf{z} = \{z_0, z_1, ..z_N\}$ where N is the number of CFD mesh elements on the rod surface. The residuals are given by: $\mathbf{e} = \mu_{T,\mathrm{ctf}}(\mathbf{z}) - T_{cfd}(\mathbf{z})$.

$K(\cdot)$ generates a matrix which describes the spatial autocorrelation present in the CFD field data. The commonly used squared-exponential covarience generation function is provided in equation 2.14 [31].

$$K_{ij}(\mathbf{z}_1, \mathbf{z}_2; \theta) = \theta_0^2 e^{-\frac{||z_{1,i} - z_{2,j}||^2}{\theta_1^2}} \tag{2.14}$$

Where i and j are indices of the vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ respectively. The parameters of the covariance function can be fit to the known residual vector, $\mathbf{e}$, using a maximum likelihood approach. The log likelihood function for the covariance model is given by equation 2.15 [32].

$$ln\mathcal{L}(\theta|\mathbf{e}) = -\frac{1}{2}\mathbf{e}^T K(\mathbf{z}, \mathbf{z}; \theta)^{-1}\mathbf{e} - \frac{1}{2}ln(\det(K(\mathbf{z}, \mathbf{z}; \theta))) - \frac{N}{2}ln(2\pi) \tag{2.15}$$

The optimal covariance function parameter values can be computed by solving the minimization problem in equation 2.16.

$$\hat{\theta} = \text{argmin}_\theta[-ln\mathcal{L}(\theta|\mathbf{e})] \tag{2.16}$$

The fitted kriging model can be queried for the mean temperatures at locations $\mathbf{z}_*$. The mean prediction vector is given in matrix form by equation 2.17.

$$\hat{T}(\mathbf{z}_*) = \mu_{T,\text{ctf}}(\mathbf{z}_*) + K(\mathbf{z}_*, \mathbf{z}; \hat{\theta})K(\mathbf{z}, \mathbf{z}; \hat{\theta})^{-1}\mathbf{e} \tag{2.17}$$

Efficient methods for computing the mean and drawing samples from the fitted kriging model can be found in *Gaussian processes in machine learning* (Rasmussen, 2004) [31]. This is not straightforward to do efficiently and the naive approach involves inverting a NxN matrix. The completed regression kriging model is visualized in figure 2.3. The high fidelity data serves to bias-correct the subchannel predictions.

Several difficulties preclude the application of kriging directly to the hi2lo problem at hand. The first issue involves minimizing the negative log likelihood function in equation 2.15 to fit the covarience function to the known CFD data. This requires solving a large linear system of size NxN where N can be on the order of several million points for a relatively small CFD computation.

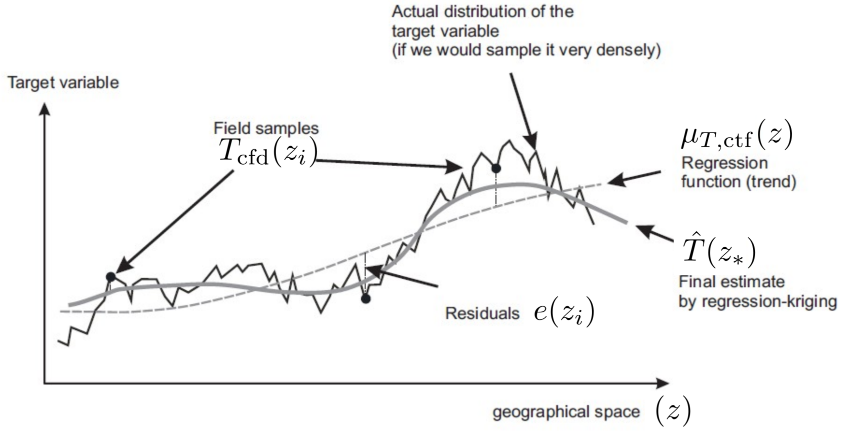

Actual distribution of the target variable (if we would sample it very densely)
Target variable
Field samples
$T_{\text{cfd}}(z_i)$
$\mu_{T,\text{ctf}}(z)$
Regression function (trend)
$\hat{T}(z_*)$
Final estimate by regression-kriging
Residuals $e(z_i)$
geographical space (z)

Figure 2.3: Regression kriging example [28]. T_{cfd} represents the fine scale CFD data samples, $\mu_{T,\text{ctf}}$ corresponds to the coarse fidelity model predictions and $\hat{T}$ is the RK model mean output.

Furthermore, variance estimates provided by the RK model assume that residuals are normally distributed which is not necessarily the case for residuals derived by subtracting CFD results from subchannel rod-surface fields. Additionally, the example here only considered the prediction of a single rod surface field, T. Crud prediction also requires estimates for the near-wall TKE and surface BHF. Each of these fields could be interpolated separately but care should be taken to preserve correlations between them because it is the action of hot locations on the rod surface occurring in coincidence with low local TKE which gives rise to the thickest crud deposits.

2.4 Copula

In contrast to multivariate Gaussian based approaches, non-Gaussian dependence structures between two or more correlated random variables can be represented by a copula. In particular, preserving the statistical relationship between the temperature and near-wall TKE on a small, localized patch on the rod surface is of great interest in this work. It cannot be assumed that the dependence structure between these fields follows a symmetric multivariate Gaussian.

Copula have seen historical use in the finance industry to predict correlated extreme value risks in credit portfolios [33]. Copula have received additional attention in financial and mathematics communities since simpler Gaussian based dependence modeling techniques were revealed to make erroneous expected CDO portfolio loss predictions under the market conditions present in the financial crisis of 2008-2009 [34], [35]. Despite the widespread adoption of copula models in financial risk assesment community, only recently have copula been applied to flood risk models [36], [37], and reliability analysis in nuclear plants [38]. The delayed adoption of the copula in the engineering realm is speculated to be due to a substantial increase in computational complexity required to construct and evaluate high dimensional copula over incumbent Bayesian network and multidimensional Gaussian based methods. Though higher dimensional copula do pose significant challenges in fitting and sampling, it is straightforward to fit low dimensional copula models to empirical data using a maximum likelihood or Markov Chain Monte Carlo approach [39]. A method for drawing correlated samples from a copula is provided in section 3.2.3.

3 | Theory

A review of the solution detail afforded by CFD and subchannel thermal hydraulic codes is provided to begin this section along with consequences of their respective spatial discretization schemes on crud growth. A simplified method for harnessing CFD results to improve expected-value computation of crud on a given CTF face is provided to introduce the hi2lo strategy. Next, copula and quantile regression are discussed as a means to model the joint distribution of temperature and TKE on the rod surface. This is followed by an introduction to the gradient boosting machine learning method used to predict the behavior of the joint distribution as a function of local core conditions. The Monte Carlo method for estimating the integral required to compute the expected crud value is given. A review of importance sampling is also provided to provide a means to increase the sampling efficiency of the Monte Carlo integration procedure. The section culminates in an integration of the copula, quantile regression, and importance sampling routines into a complete algorithm for time dependent crud prediction.

3.1 Model Approach

A fundamental difference between the CFD and CTF computations is the average size of the mesh cells. In the azimuthal coordinate, CTF decomposes a single rod surface into four patches. An example top down view of typical CFD and CTF meshes for a single pin are given in figure 3.1. Though both codes employ a finite volume spatial discretization, CFD can resolve the flow at much smaller length scales. Additionally, each code employs a different set of closure models to the underlying set of coupled energy, mass, and momentum balances. In practice these differences lead to large discrepancies in boiling, turbulent mixing, and rod surface temperature predictions between the two codes.

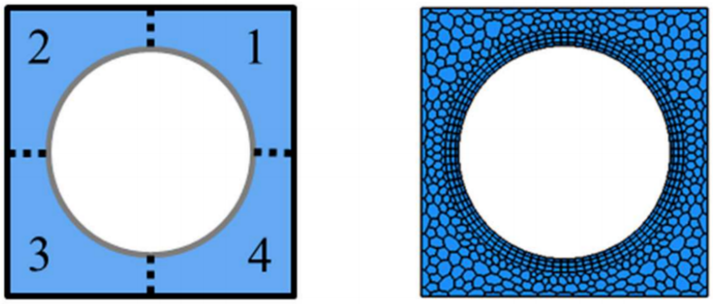


Figure 3.1: Top-down view of typical subchannel (left) and CFD mesh (right) for a single pin [9].

Shown in figure 3.2, on a given CTF rod surface patch, estimates for the surface temperature, TKE, and heat flux are provided as point estimates. The predicted CTF quantities are an estimate for the average thermal hydraulic conditions over that coarse patch. Consequently, CTF crud predictions deviate from reality since crud growth is highly sensitive to the presence of subcooled boiling on the rod surface; if CTF predicts a rod surface temperature less than the saturation point in a given patch little or no crud will form when in reality, a small portion of that rod surface could exist above the saturation point and thus harbor crud. Small localized mistakes in crud predictions compound throughout the core, leading to poor CIPS estimates.

In figure 3.2, f denotes a probability density function. Integration of this density function may be interpreted as computing a fractional area of the rod surface that exists within the specified integration limits.

CTF estimates mean TH conditions everywhere in the core at a low spatial resolution. The CFD informed model provides higher order moments about the mean.

$$\mathbf{X}(\mathbf{p},\mathbf{z}) = \underbrace{\boldsymbol{\mu}(\mathbf{p},\mathbf{z})}_{\text{CTF}} + \underbrace{\varepsilon(\theta(\boldsymbol{p},\mathbf{z})) + \boldsymbol{b}(\mathbf{p},\mathbf{z})}_{\text{CFD Informed}} \tag{3.1}$$

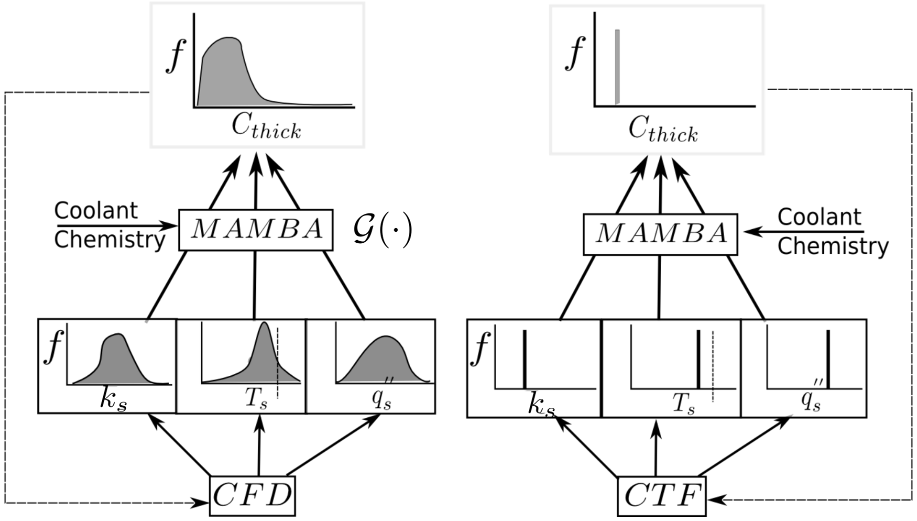


Figure 3.2: On a single coarse CTF patch: Differences in crud prediction between CFD and CTF models.

- $\mathbf{X}$ is a three component vector representing the cladding surface temperature, T, turbulent kinetic energy, k, and boundary heat flux, q''.

- $\mathbf{z}$ denotes spatial coordinates and $\mathbf{p}$ represents a set of auxiliary predictors. Auxiliary predictors are covariates that describe local core conditions and may be geometric or thermal hydraulic in nature. Specific auxiliary predictors are introduced in Chapter 5. Table 5.1 contains a detailed description of the auxiliary predictive features used in this work.

- ε is a three-component random vector comprised of temperature, turbulent kinetic energy and boundary heat flux fields. ε is distributed according to a CFD informed model with θ representing free model parameters which are determined from the CFD data.

- $\boldsymbol{b}$ is bias between the low and high fidelity models ($\boldsymbol{\mu}_{CTF} - \boldsymbol{\mu}_{CFD}$). Despite providing identical inlet boundary conditions to both codes bias exists due to algorithmic, closure model and meshing differences between the two codes.

- Field averages, $\boldsymbol{\mu}$, are piecewise constant over each CTF patch.

Consider a hypothetical case where the CFD results are normally distributed about the CTF results such that $\varepsilon \sim \mathcal{N}(0, \mathbf{\Sigma}(\mathbf{p}, \mathbf{z}))$, where $\mathbf{\Sigma}(\mathbf{p}, \mathbf{z})$ is a covariance matrix that depends on local core conditions. Shifting the distribution by a constant vector $\boldsymbol{c} = \boldsymbol{b} + \boldsymbol{\mu}_{ctf}$, results in a distribution denoted by h in equation 3.2.

$$\begin{aligned} h|_{(\boldsymbol{p},\boldsymbol{z})} &= \mathcal{N}(\boldsymbol{c}, \mathbf{\Sigma}(\mathbf{p}, \boldsymbol{z})) \\ &= \mathcal{N}\left(\begin{pmatrix} c_T \\ c_k \\ c_{q''} \end{pmatrix}, \begin{pmatrix} \Sigma_{TT} & \Sigma_{Tk} & \Sigma_{Tq''} \\ \Sigma_{kT} & \Sigma_{kk} & \Sigma_{kq''} \\ \Sigma_{q''T} & \Sigma_{q''k} & \Sigma_{q''q''} \end{pmatrix} \right) \Bigg|_{(\mathbf{p},\mathbf{z})} \end{aligned} \tag{3.2}$$

Where $\Sigma_{xx} = \sigma_x^2$ and $\Sigma_{xy} = cov(x, y)$.

Equation 3.3 estimates the expected crud mass C_m that accumulates on each CTF patch in time δt. Let the CTF face of interest have area A. Let $\mathbf{X} = \{T, k, q''\}$ denote a random vector of temperature, TKE, and BHF. $\mathbf{I}$ represents additional known crud parameters, $\mathbf{C}_o$ is the crud state at the start of the time step. Let the joint density function of $\mathbf{X}$ be denoted by h, and it's CDF be denoted by H. The crud model, $\mathcal{G}(\cdot)$, is common to all CTF faces. The joint cumulative density's parameters are predicted from the available high resolution CFD data in every CTF face. In the case of an assumed normal distribution model there are nine unknowns which require fitting: $\theta = \{\sigma_T^2, \sigma_k^2, \sigma_{q''}^2, \Sigma_{kT}, \Sigma_{q''T}, \Sigma_{q''k}, c_T, c_k, c_{q''}\}$. In the subsequent sections we will seek to relax the normality assumption of the CFD residuals about the CTF result.

$$\begin{aligned} C_m &= A\mu_m \\ &= A\,\mathbb{E}[\mathcal{G}(\mathbf{X}|\mathbf{C}_o, \mathbf{I}, \delta t)] \\ &= A \iiint \mathcal{G}(\mathbf{X}|\mathbf{C}_o, \mathbf{I}, \delta t) h(\mathbf{X}|\theta) d\mathbf{X} \end{aligned} \tag{3.3}$$

A strategy to compute the unknowns of the joint distribution on each CTF face is required. The current work proposes a data driven model, $\mathcal{F}_M$, to predict the unknowns provided a suite of pre-computed CFD results are used to train the machine learning model and local thermal hydraulic conditions provided by CTF at runtime are utilized to evaluate the model on each CTF face. Algorithm 2 is used to compute the total crud mass in each CTF face.

ALGORITHM 2
Generic hi2lo method for crud prediction.

```
 1: **Initialization**
 2: (1) Pre-process training set.
 3:     (1b) Fit the joint distribution parameters, $\theta$, to known CFD data.
 4:     (1c) **def:** $\theta \leftarrow \mathcal{F}_M(\mathbf{p}, \mathbf{z})$
 5: (2) Train model: $\hat{\mathcal{F}_M} = \text{argmin}_{\mathcal{F}}\ \mathbb{E}\left[L(\mathcal{F}_M(\mathbf{p}, \mathbf{z}), \theta)\right]$
 6: **for** CTF face, $j$ **do**
 7:     Evaluate ML model $\hat{\theta}_j \leftarrow \hat{\mathcal{F}_M}(\mathbf{p}_j, \mathbf{z}_j)$
 8:     Reconstruct $\hat{H}_j(\cdot|\hat{\theta}_j)$
 9:     Draw samples $\mathbf{X} \sim \hat{H}_j$
10:     Evaluate equation 3.3 via Monte Carlo approximation
11: **end for**
```

Where $L(\cdot)$ is a generic differentiable loss function. A discussion on the machine learning model and loss function follows in section 3.2.5. The reconstruction of the joint density function $\hat{H}$ from copula and univariate quantile functions is discussed in section 3.2.4. Monte Carlo and importance sampling are discussed in section 3.2.6.

Algorithm 2 may be broken down into five generic components. *Lines 2-5:* Establish a relationship between the local core geometry and thermal hydraulic conditions to the behavior of surface temperature, TKE and boundary heat flux distributions using a pre-computed suite of CFD results. *Line 7:* The prediction of joint distribution parameters. These parameters will take the form of conditional quantiles and parameters of a copula in the present work. *Line 8:* The reconstruction of the joint temperature, TKE, and boundary heat flux distribution in each CTF face given predictions provided by evaluating the data driven model $\mathcal{F}_M$ at the local core conditions adjacent to the CTF face under consideration. *Line 9:* Sampling from the reconstructed distribution. *Line 10:* Integration of the crud density by a Monte Carlo procedure.

3.2 Construction of the Hi2lo Map

Next, the multivariate Gaussian assumption made to capture the autocorrelation between the surface temperature, boundary heat flux and near wall TKE fields is relaxed. Flexibility is afforded by factoring the multivariate distribution into marginal distributions and a copula. The statistical parameters describing this multivariate distribution are still determined via a data driven model, as in algorithm 2. The result is a semi-parametric model of the conditional joint distribution of temperature, TKE, and boundary heat flux on the rod surface. The model is semi-parametric because the copula is selected from a library of parametric distribution families while the marginal models are constructed using conditional quantile prediction, where the number of quantiles used in the reconstruction is set at runtime. Quantile regression as applied in this work does not assume a priori the univariate distribution family governing the behavior of the surface temperature or near-wall TKE distributions. This is the case because no mechanistic model exists or could be identified which describes the distribution of finding a particular patch of the rod surface in excess of a given temperature. It is unlikely a physics based parametric model could be devised for this purpose due to the general complexity of the underlying governing equations, grid geometry, and since the flow regime of interest is turbulent. In lieu of such a mechanistic model non parametric distributions are adopted to represent the surface temperature and TKE distributions.

3.2.1 Capturing Dependence Between Random Variables

Since the outer cladding temperature, near-wall TKE and boundary heat flux are used as boundary conditions to a crud growth package, it is particularly important to understand and capture the relationship between these fields in the hi2lo model. The hi2lo model under consideration is not a dynamic model in the sense that it cannot be expressed as a coupled system of differential equations. Instead, in the purely data driven approach the relationships between the FOI are established through standard statistical correlation measures.

In a multiphysics simulation of a PWR core the coupled momentum, mass, and energy balances along with the appropriate closure models dictate the rod outer cladding surface temperature. The Dittus Boelter relationship is used to relate the surface heat transfer coefficient with the local Reynolds number (see appendix D). According to this relationship, larger Reynolds numbers corresponds to higher heat transfer coefficients. Newton's law of cooling states $T_s = q''/h + T_\infty$ and h may be computed via Dittus Boelter. Therefore, the surface temperature (T_s) is negatively correlated with the Reynolds number and where the local turbulent kinetic energy is large the rod surface temperature will be depressed if the local heat flux is held fixed. It is also apparent that the surface temperature is positively correlated with the the local boundary heat flux. In order to simplify the model only the dependency between the surface temperature and local turbulent kinetic energy is considered in hi2lo model.

A statistical relationship between the surface temperature, TKE, and boundary heat flux in each CTF face is sought. To this end vine copula provide a flexible framework to model high dimensional dependence structures [40]. Vine copula are hierarchical tree models in which the edges represent bivariate copula and the nodes are univariate distributions. The canonical vine (C-vine) copula shown in equation 3.4 may be used to express the trivariate ($n = 3$) distribution of temperature, TKE, and q'' on the rod surface.

$$h(T, k, q'') = f_T f_k, f_{q''} \prod_{m=1}^{n-1} \prod_{e \in E_m} c_{ij|D_e}(u_{i|D_e}, u_{j|D_e}) \tag{3.4}$$

Where m denotes the tree level in the vine and each bivariate copula model, c defined on the edge $\{e\}$ is known as a pair copula. The conditioning set at edge e, denoted D_e, is defined by a proximity condition [41]. E_m denotes the set of all edges at level m. The graphical representation of the vine is provided in figure 3.3. Simplifying independence assumptions can be made based on the heat transfer processes on a fuel rod that lead to certain copula in the vine to take the form of a uniform density distribution on the unit square.

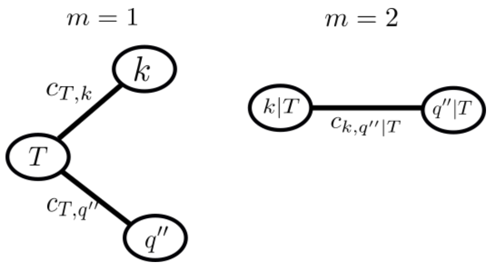


Figure 3.3: C-vine on 3 variables: $\{T, q'', k\}$.

In this work it is assumed that the cladding surface temperature and near-wall TKE are uncorrelated with the boundary heat flux. The justification is as follows.

Relative variations in boundary heat flux are very small over a CTF face (±5%) provided that a CTF face represents a small (approximately $1[cm^2]$) localized patch on the rod surface. Large absolute heat fluxes of approximately $80[W/cm^2]$ are possible on the cladding surface of fuel in a typical PWR. However, high axial and azimuthal gradients of boundary heat flux on the rod surface are not typical under standard operating conditions. High thermal gradients lead to thermal induced stresses on the cladding which can promote rod bowing or, extreme in conditions, fuel failure. Additionally, figures 4.12 to 4.14 show that the sensitivity of the crud growth rate to the boundary heat flux is small relative to the sensitivity of crud growth rates to surface temperature and local turbulent kinetic energy.

After applying the independence assumptions the original trivariate dependence model between the temperature, TKE, and boundary heat flux is reduced to a bivariate model in which the boundary heat flux is treated independently. Applying the simplifying assumptions results in: $c_{T,q''}(u_T, u_{q''}) = 1$ and $c_{k,q''|T}(u_{k|T}, u_{q''|T}) = 1$. The simplified joint density is given by equation 3.5:

$$h(T, k, q'') \approx f_T f_k, f_{q''} c_{T,k}(u_T, u_k) \cdot 1 \cdot 1 \tag{3.5}$$

Where $u_T = F(T)$ and $u_k = F(k)$. $F(\cdot)$ denotes the CDF. Following the independence assumption, the marginal density function of the outer cladding boundary heat flux is assumed to be a Dirac delta function centered on the value provided by CTF (VERA). We will apply this assumption in all sections which follow. Then in the j^{th} CTF face the boundary heat flux marginal distribution is always given by equation 3.6.

$$f_{j,q''} = \delta_{(q''_{j,\mathrm{ctf}})} \tag{3.6}$$

3.2.2 Copula

A copula is a function which relates marginal probability distributions to a multidimensional joint distribution. Copula provide a flexible alternative to multidimensional Gaussian based models. Copula are utilized in this work because of their ability to capture non-Gaussian dependence structure between two or more correlated random variables, for instance temperature and the TKE at a given point on a rod's surface. Furthermore, Sklar's theorem is used in this work in order to decompose joint distributions into a product of uni-variate marginal distributions and a copula function. In this section, Sklar's theorem is provided along with examples of copula functions and techniques to draw samples from them.

The product rule of probability is shown in equation 3.7. To clarify notation used in this section: The comma denotes the conjunction "and" and the bar, |, is read "given".

$$P(x, y) = P(x)P(y|x) \tag{3.7}$$

The marginal distribution of a bivariate joint distribution $f(x, y)$ is given by equation 3.8. The marginalization process is analogous to projecting the entire joint density onto a single axis.

$$f(x) = \int f(y)f(x|y)dy \tag{3.8}$$

The cumulative density function, F is defined as:

$$F = \mathbf{P}[X < x] = \int_{-\infty}^{x} f(x)dx$$

A joint d dimensional cumulative distribution is given by equation 3.9.

$$H(x_1, ...x_d) = \mathbf{P}[X_1 \leq x_1, ...X_d \leq x_d] \tag{3.9}$$

Where $X_1, ...X_d$ are random variables.

The process of decomposing a multivariate distribution into uni-variate marginal distributions and an object which describes their conditional dependence was formalized by Sklar [42]. Shown in equation 3.10, Sklar's Theorem defines a *copula* cumulative density function, C.

$$C(F_1(x_1), ...F_d(x_d)) = H(x_1, ...x_d) \tag{3.10}$$

If $F_1, ..F_d$ are continuous, then C is unique. Conversely, if C is a copula and $F_1, ..F_d$ are smooth cumulative destiny functions then the function H is a joint cumulative distribution with margins $F_1, ...F_d$. A proof is provided in Nelsen's introductory copula text [43].

Sklar also showed that the joint probability distribution, $h(x_1, ...x_d)$, can be computed from constituent marginalized univariate distributions and the copula density, c.

$$h(x_1, \ldots x_d) = c(F_1(x_1), \ldots F_d(x_d)) \cdot f_1(x_1) \cdot \cdots \cdot f_d(x_d) \tag{3.11}$$

For brevity, let $u_1, ..u_d$ represent samples from their CDFs as follows:

$$u_1 = F_1(x_1)$$

$$u_d = F_d(x_d)$$

$$u \in [0, 1]$$

Where the joint density of the copula, c, is given by equation 3.12:

$$c(u_1, ...u_d) = \frac{\partial C(u_1, ...u_d)}{\partial u_1 ... \partial u_d} \tag{3.12}$$

Sklar's theorem enables one to construct models for the margins separately from a model of the dependence structure. When combined, the margins and the copula specify a multivariate probability density function. Compared to rudimentary approaches based on covariance matrix dependence model, a copula based approach can treat skewed dependence structures in which the strength of dependence is allowed to vary depending on location in the parameter space.

Sampling Copula

For simplicity, this section demonstrates how to draw correlated samples from bivariate copula. Sampling from a bivariate copula is achieved by first defining a conditional distribution function and then applying the inverse probability integral transform. Let h represent the conditional distribution of u_1 given all other random variables $\{u_2, ...u_d\}$. In the two dimensional case h is given by equation 3.13 [43]:

$$h(u_1|u_2) = \frac{\partial C(u_1, u_2)}{\partial u_2} \tag{3.13}$$

If the distribution h is smooth and monotonic the inverse h^{-1} exists. For a bivariate Gaussian copula with a shape parameter $\theta_c = 0.7$ the conditional distribution functions are shown in figures 3.4 and 3.5 for several values of the conditioning variable (u_2).

Computing the inverse analytically is oftentimes not possible for some classes of copula and therefore, the more general method shown in equation 3.14 is used. A random vector of length N is drawn from the uniform distribution $\in [0, 1]$: $\{\mathbf{U}_2\}$. For each sample, u_{2_i} in $\{\mathbf{U}_2\}$ the 1-D line search problem given in equation 3.14 is solved. This produces a sample vector of length N: $\{\mathbf{U}_1\}$.

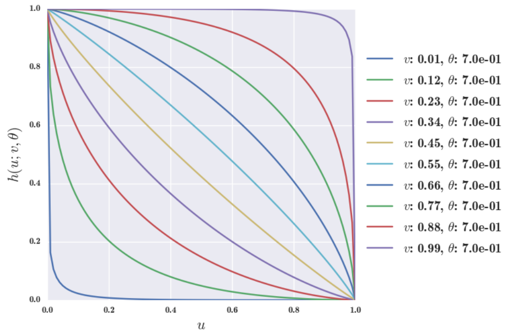


Figure 3.4: The conditional h function vs. value of the conditioning variable u_2 for a Gaussian copula with $\theta_c = 0.7$.

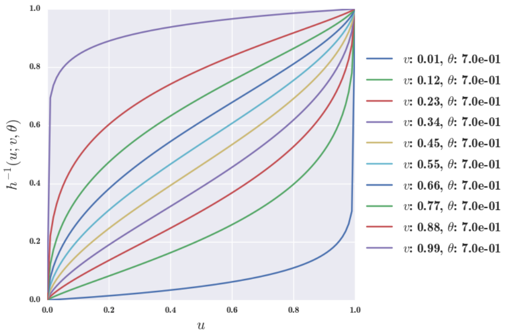


Figure 3.5: h^{-1} vs. value of the conditioning variable u_2 for a Gaussian copula with $\theta_c = 0.7$.

$$u_{1_i} = \text{argmin}_{u_1} \left[h(u_1 | u_{2_i}) - u_{2_i} \right], \text{ with } 0 < u_1 < 1 \tag{3.14}$$

The resulting correlated sample vectors $\{\mathbf{U}_1, \mathbf{U}_2\} \in [0, 1]^2$ are distributed according to the copula, c, and have uniform margins. An example of random samples drawn from a Gaussian copula are shown in figure 3.7. The smooth Gaussian copula PDF is provided in figure 3.6.

To apply arbitrary margins F_1 and F_2 we employ, again, the inverse probability transform. Correlated samples are then drawn according to:

$$\mathbf{X} = F_1^{-1}(\mathbf{U}_1) \tag{3.15}$$

$$\mathbf{Y} = F_2^{-1}(\mathbf{U}_2) \tag{3.16}$$

The sample vectors $\{\mathbf{X}, \mathbf{Y}\}$ are distributed according to the joint density, $C(F_1, F_2)$. An example bivariate sample set with exponentially distributed margins and a Gaussian copula is shown in figure 3.8. In the example figure both margins follow an exponential distribution given by $f(x) = \lambda e^{-\lambda x}$ with $\lambda = 2\text{E}{-3}$.

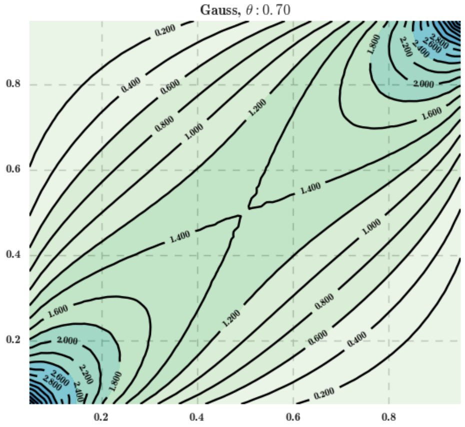


Figure 3.6: Gaussian copula density with $\theta_c = 0.7$.

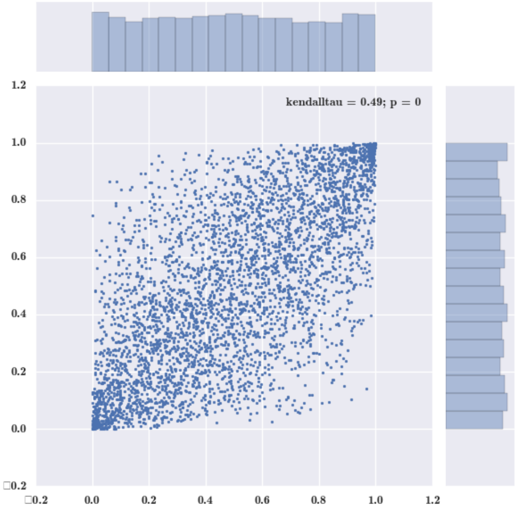


Figure 3.7: Samples drawn from Gaussian copula
with $\theta_c = 0.7$.

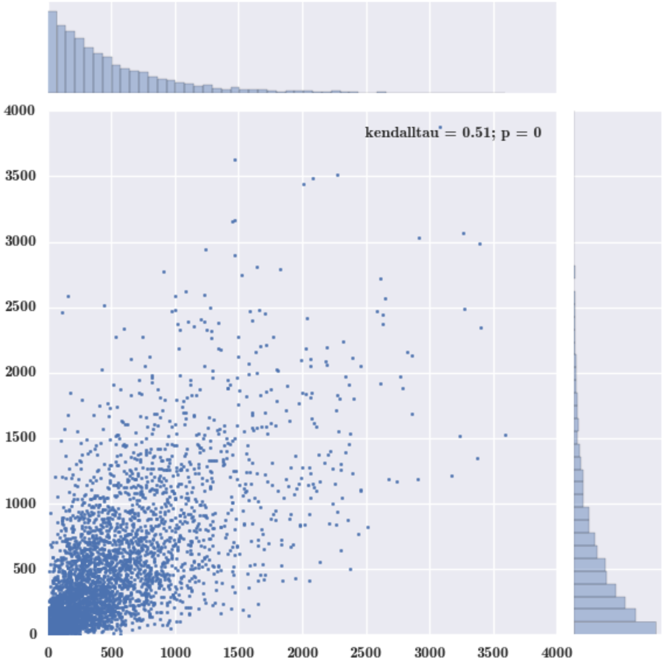


Figure 3.8: Samples drawn from Gaussian copula with exponential margins.

Copula Families

A wide range of copula functions are available in the literature. In order to satisfy the definition of a copula several criteria must be met:

1. Must integrate to one on $[0, 1]^n$

2. Must have uniform marginal distributions (as shown in figure 3.7).

3. When one argument to the joint copula CDF is zero, the CFD is zero:

$$C(u_1, u_2, ...0, ...u_d) = 0 \tag{3.17}$$

4. When one argument to the joint copula CDF is $u \in [0, 1]$ and all other arguments are one, the CFD takes a value equal to u:

$$C(1, 1, ...u, ...1) = u \tag{3.18}$$

Examples of valid copula are given in figure 3.9. A wide range of skewed dependence structures can be represented by considering only a few copula families. Each copula can be rotated to accommodate both positive or negative dependence.

3.2.3 Fitting Copula

Fitting copula to empirical data can be carried out by the method of maximum likelihood (ML). Consider the bivariate case where N sample pairs, $\{w_i, v_i\}_{i \in [1,N]}$ are known. The likelihood function for a copula is given by equation 3.19. The likelihood is a function of the distribution parameter θ_c with the data, $\{\mathbf{w}, \mathbf{v}\}$, held fixed. When integrated over all possible parameter values the integrated result does not necessarily take on a value of unity and therefore cannot be strictly interpreted as a probability density. Each constituent factor in the likelihood function can be interpreted as the relative likelihood that the sample pair $\{w_i, v_i\}$ arose from the copula density function with parameter θ_c.

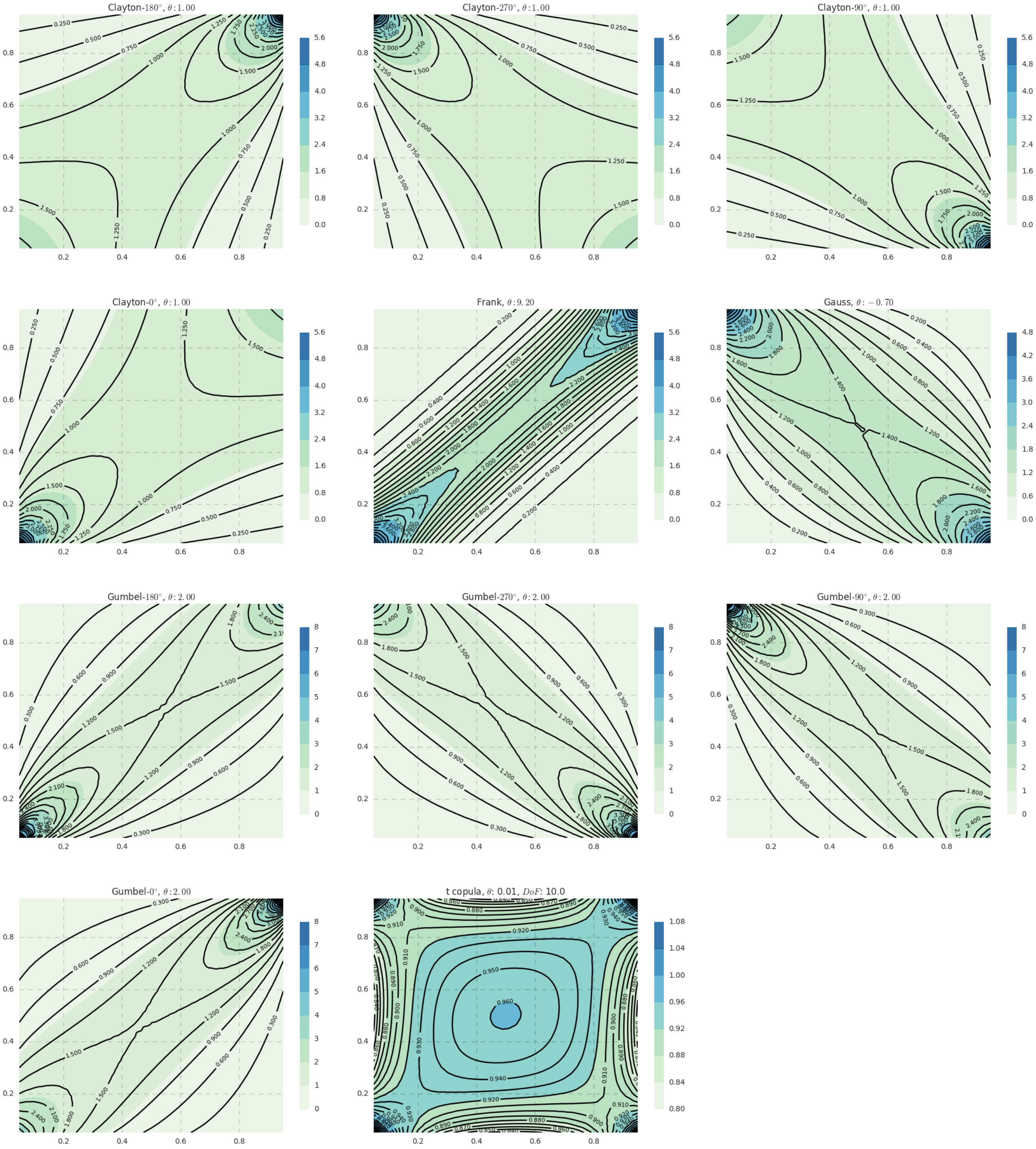


Figure 3.9: Examples of bivariate copula PDFs.

$$\mathcal{L}(\theta_c; \mathbf{w}, \mathbf{v}) = \prod_{i=1}^{N} c(w_i, v_i|\theta_c) \tag{3.19}$$

Where θ_c is the free copula shape parameter. Typically the negative log-likelihood, $-\ln\mathcal{L}$, is used in when performing ML estimation of a distribution parameter since the problem is typically cast in terms of a minimization problem.

$$\hat{\theta}_{c,ML} = \text{argmin}_{\theta_c}[-\ln\mathcal{L}(\theta_c; \mathbf{w}, \mathbf{v})] \tag{3.20}$$

To minimize the negative log likelihood in equation 3.20 one computes the partial derivative with respect to θ_c and finds the value $\hat{\theta}_{c,ML}$ for which this expression reaches zero. This can be carried out by Newton's method. If the partial derivatives of the copula's negative log likelihood are impossible to compute analytically one can estimate them by finite difference.

In order to determine which copula family best represents the data, an arsenal of statistical tests can be applied to select the copula which best fits the data. Here, we consider two methods: (1) Comparing Akaike information criterion (AIC) and (2) graphically comparing each fitted copula.

The AIC is computed by equation 3.21.

$$\text{AIC} = 2k - 2\ln(\mathcal{L}) \tag{3.21}$$

Where k is the number of free parameters in the model. The AIC penalizes models with larger numbers of parameters. Automated copula selection is achieved by selecting the copula that obtains the lowest AIC score.

A graphical method of copula selection was proposed by Barbe et. al. (1996) [44]. In this method each trial copula's Kendall's function, $K_c(t)$ is plotted against an empirical estimate of this function, $\hat{K}_c$. Given d random variables $\mathbf{U} = \{U_1, ...U_d\}$ distributed according to some d dimensional copula, C, Kendall's function is given by 3.22 [40].

$$K_c(t; C) = \mathrm{P}\left[C(\mathbf{U}) \leq t; \mathbf{U} \sim \ C\right] \tag{3.22}$$

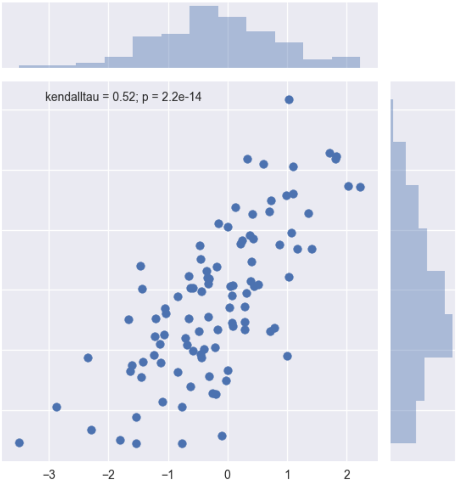


Figure 3.10: Ficticious bivariate data set.

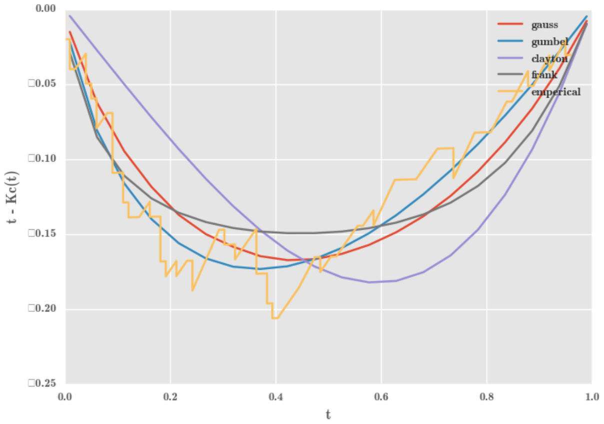


Figure 3.11: Graphical comparison of Kendall's distribution for several fitted copula.

For example, by graphical inspection of figure 3.11, the Gumbel copula is the best fit to the original data set. This visual process can be automated by computing and comparing $||\hat{K}_c(t) - K_c(t)||$ for each trial copula.

Kendall's Tau

Kendall's tau is a measure of concordance. Consider two correlated and uniformly distributed random variables, X, Y. Let (X_1, Y_1) and (X_2, Y_2) be identically distributed random vectors from some joint cumulative distribution so that individual samples from $X \ and \ Y$ will take on values in $[0, 1]$, then Kendall's tau is given by equation 3.23 [43].

$$\rho_\tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \tag{3.23}$$

In the case of Archimedean copula, ρ_τ is directly related to the copula's parameter, θ_c. Equation 3.24 relates an Archimedean copula's parameter to ρ_τ. This is useful since if one can estimate ρ_τ from the empirical data and the copula type is known, one can quickly compute the copula's shape parameter without resorting to the method of ML.

$$\rho_\tau = 1 + 4\int_0^1 \frac{\varphi(\theta_c, t)}{\varphi'(\theta_c, t)} dt \tag{3.24}$$

Where $\varphi(t)$ is the copula's generator function and $\varphi'(t)$ is the first derivative of the generator function with respect to t. A list of copula generator functions may be found in introductory copula texts [43].

3.2.4 Sample Quantiles

Here a non parametric representation of univariate distributions is given by utilizing a set of sample quantiles. Let the quantile function be represented by $Q = F^{-1}$ where F is a cumulative density function (CDF).

The τ^{th} quantile is $q_\tau = Q(\tau)$; where $F(t) = P[T \leq t]$. $\tau \in [0, 1]$. The quantile loss function is given by equation 3.25 and depicted in figure 3.12.

$$\mathcal{W}_\tau(u) = u \cdot (\tau - \mathbb{I}_{(u<0)}) \tag{3.25}$$

Where $\mathbb{I}$ is the indicator function which returns 1 if the argument is true and 0 otherwise. In order to estimate a sample quantile, $\hat{q}_\tau$, given the empirical CDF F, minimize: $\mathbb{E}[\mathcal{W}_\tau(T - q_\tau)]$ where T, the temperature, is treated as a random variable distributed according to F_T. Considering a sample set $\{T_0, T_i, \ldots T_N\}$ the desired quantile q_τ may be estimated by equation 3.26.

$$\begin{aligned}
\hat{q}_\tau &= argmin_{q_\tau}\, \mathbb{E}[\mathcal{W}_\tau(u)]; \quad u = T - q_\tau \\
&\approx argmin_{q_\tau} \frac{1}{N}\sum_i^N \mathcal{W}_\tau(u_i); \;\; u_i = T_i - q_\tau \\
&\approx argmin_{q_\tau} \left[(1-\tau)\sum_{T \leq q_\tau} (T_i - q_\tau) - \tau \sum_{T > q_\tau} (T_i - q_\tau) \right]
\end{aligned} \tag{3.26}$$

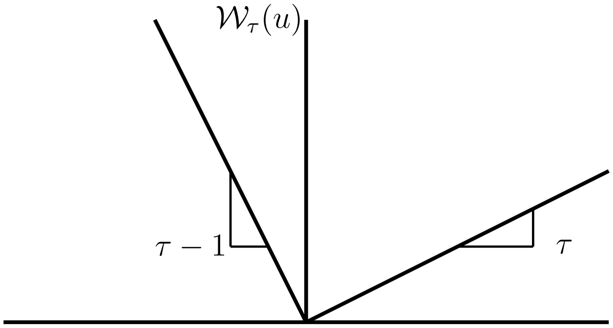


Figure 3.12: Quantile loss function.

In this work, a set of sample quantiles, $\hat{\theta}_\tau = \{\hat{q}_{\tau_0}, \ldots, \hat{q}_{\tau_i}, \ldots \hat{q}_{N_Q}\}$, are used to construct step-wise cumulative distributions, $\hat{F}$, for the cladding outer temperature and near-wall TKE surface fields. N_Q denotes the number of quantiles used in the CDF reconstruction and is a user set parameter. The stepwise quantile function with prescribed quantiles is given by equation 3.27 and depicted in figure 3.13.

$$\hat{F}_T = Q^{-1}(T; \{\hat{\theta}_\tau\}) \tag{3.27}$$

The reconstructed CDF can then in turn be used to build a histogram. In place of the stepwise representation, a monotone piecewise cubic hermite interpolating polynomial (PCHIP) may be fit to the stepwise conditional quantile distribution to generate a differentiable CDF. The PCHIP interpolation preserves monotonicity of the CDF if the provided quantiles are strictly monotone [45]. This condition is enforced in the software; any violation of the monotone restriction would indicate a software bug in the quantile regression code. Inverse transform sampling is used to draw samples from the univariate distributions.

Other strategies to approximate univariate density functions from a set of sample statistics are abundant. One alternative to quantile regression is to relate sample moments to the parameters of some parametric distribution. In this procedure a set of moment conditions is defined for the target parametric model. Then the method of mo-

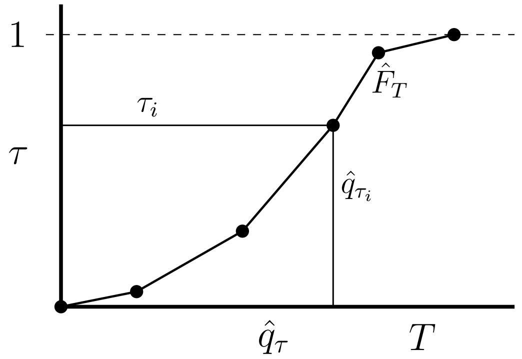


Figure 3.13: Piecewise linear CDF interpolated from a set of quantiles.

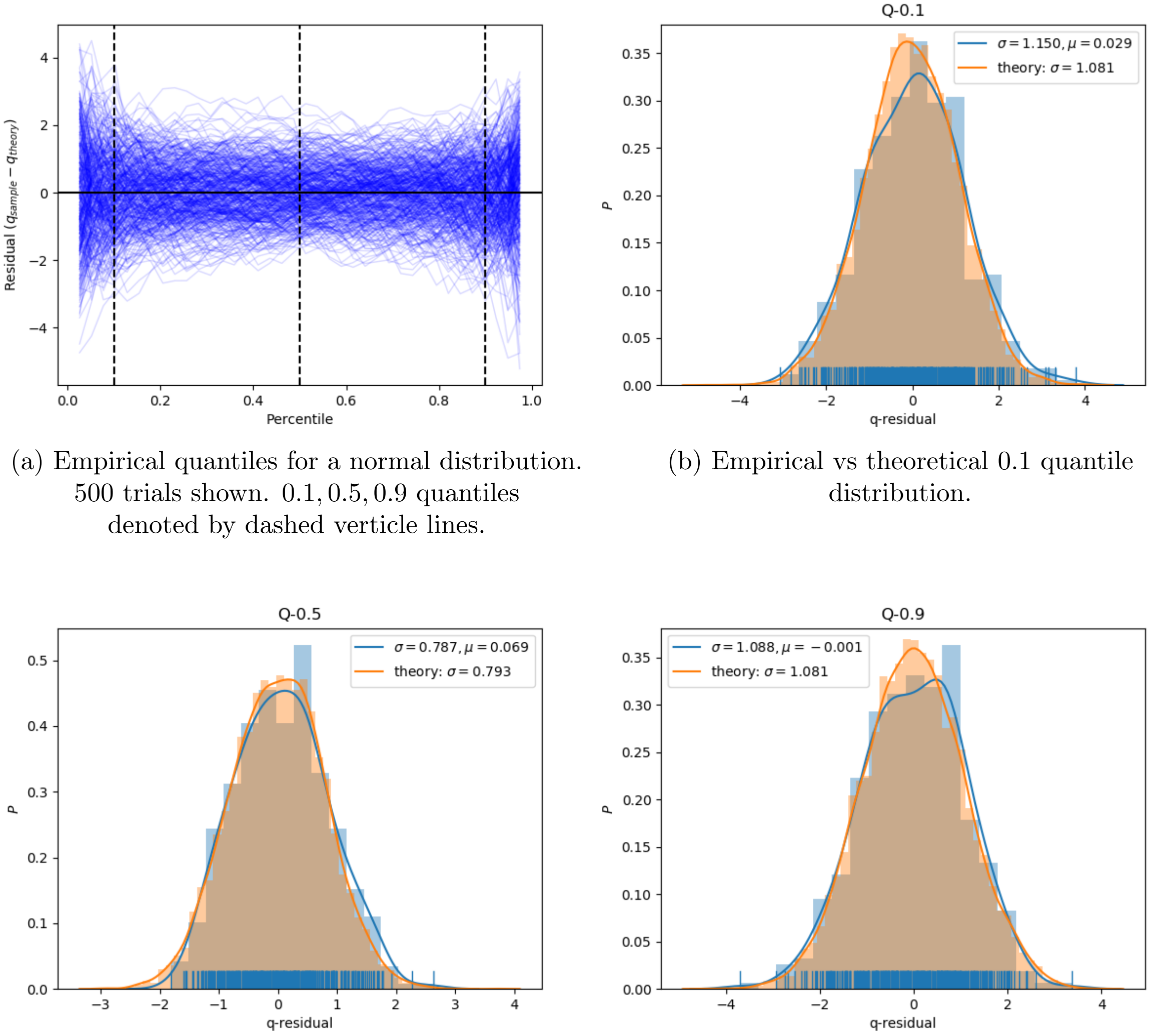
ments (MoM) can be used to determine the parameters of the parametric model which best satisfy the moment conditions in the L2 norm sense. The MoM is applicable when the number of defined moment conditions is greater than or equal to the number of free distribution parameters.

Alternatively, one may construct a non-parametric univariate density given some set of predictive features by utilizing a set of predicted distribution cumulants. The estimated cumulants, produced by evaluating a trained machine learning model, may be used to build an Edgeworth series expansion [46]. It remains as future work to determine if cumulants or traditional moments behave in a predictable manner as a function of local core conditions.

It can be shown that the sample quantiles are distributed according to equation 3.28. This directly follows from the distribution of the order statistics [47]. The theoretical distribution of the quantiles is Gaussian in the large sample limit. The large sample behavior of the quantiles is demonstrated for a normal random variable in figure 3.14 and a beta random variable in figure 3.15. When benchmarking or validating quantile predictions it is essential to check that the predictions follow the expected Gaussian behavior. For the quantile regression procedures employed in this work, this check is performed in section 3.2.5. Deviance from the expected behavior would indicate additional model-induced quantile prediction variance or if skewness is observed, bias introduced by the quantile regression model. In the current application it is important to estimate uncertainties inherent to the sample quantile estimation procedure so such uncertainties may be propagated into the crud integration procedures.

$$
\begin{aligned}
q_p &\sim \mathcal{N}\left(F_T^{-1}(p), \sigma_{q_p}^2\right) \\
\sigma_{q_p}^2 &= \frac{p(1-p)}{n[f_T(F_T^{-1}(p))]^2}
\end{aligned}
\tag{3.28}
$$

In equation 3.28 the variance associated with a sample quantile depends on the CDF and PDF of the true distribution function of interest. The underlying CDF from which a sample population is drawn is typically unknown and the goal is often to infer properties of this distribution from the sample population. Therefore, it is not possible to employ equation 3.28 directly to compute the variance of a given sample quantile unless the underlying CDF is known. Note that this is indeed the case in section 3.2.5 where the large sample quantile theory is employed to check the residuals of a quantile regression procedure because the underlying distribution of the data was prespecified to be Gaussian. However, from more complex cases where the nature of the distribution functions is unknown, such as with raw CFD data, equation 3.28 is not directly useful but still can be used to aid interpretation of the quantile regression predictions.



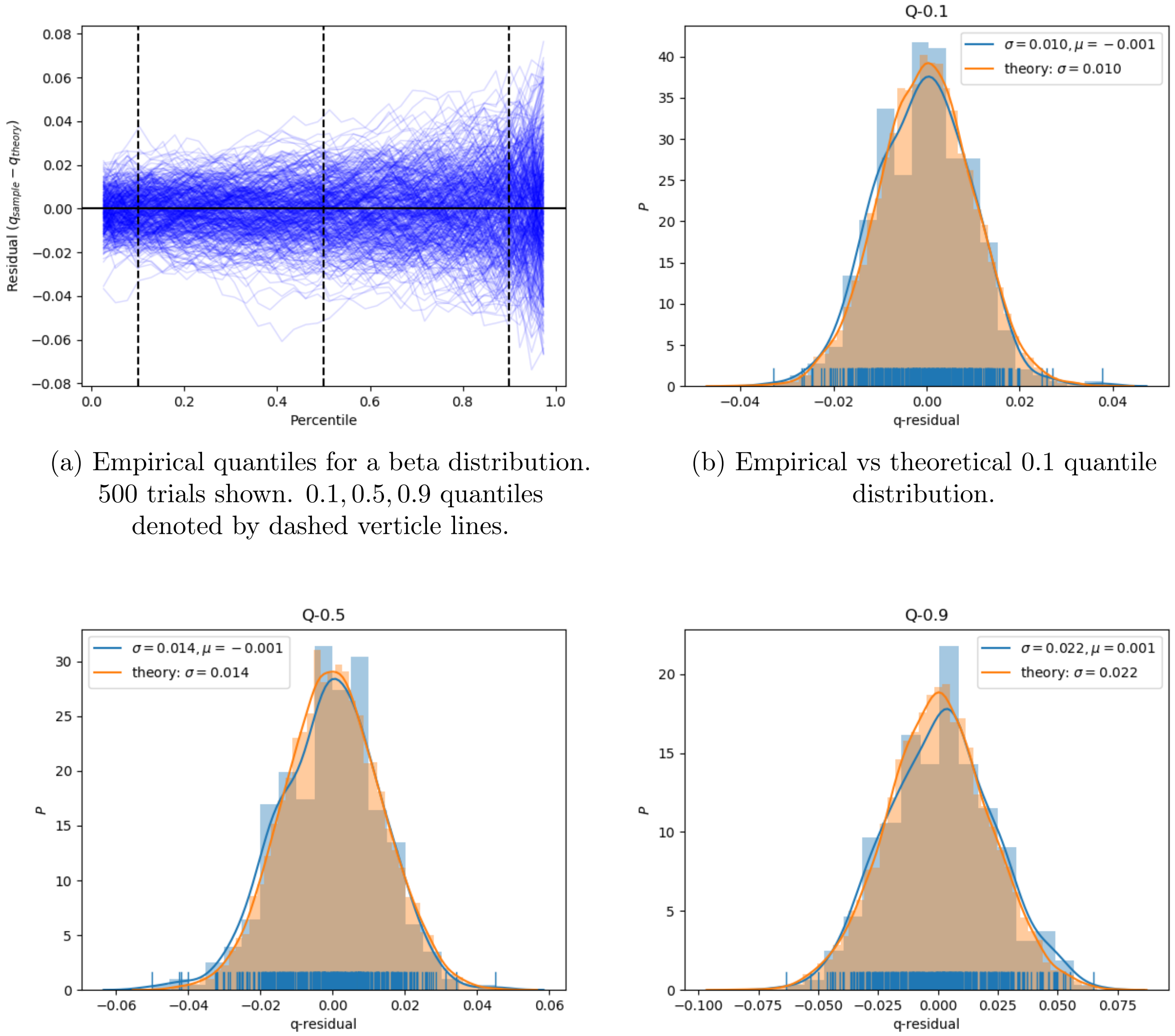
(a) Empirical quantiles for a normal distribution. 500 trials shown. $0.1, 0.5, 0.9$ quantiles denoted by dashed verticle lines.

(b) Empirical vs theoretical 0.1 quantile distribution.

(c) Empirical vs theoretical 0.5 quantile distribution.

(d) Empirical vs theoretical 0.9 quantile distribution.

Figure 3.14: Distribution of quantiles for a normally distributed RV: $X \sim \mathcal{N}(0, 10)$. Theoretical quantile standard deviation given by equation 3.28.



(a) Empirical quantiles for a beta distribution. 500 trials shown. $0.1, 0.5, 0.9$ quantiles denoted by dashed verticle lines.

(b) Empirical vs theoretical 0.1 quantile distribution.

(c) Empirical vs theoretical 0.5 quantile distribution.

(d) Empirical vs theoretical 0.9 quantile distribution.

Figure 3.15: Distribution of quantiles for a beta distributed RV: $X \sim \beta(2, 5)$

The sampled temperatures may be tallied over each CTF face to estimate the fractional area that exceeds some threshold temperature. The probability of exceeding a threshold temperature, T^*, is shown in equation 3.29.

$$p_e = Pr(T > T^*) = 1 - \int_0^{T^*} f_T dT \tag{3.29}$$

Let $q_{p_e} = F_T^{-1}(1 - p_e)$ denote the quantile associated with the threshold probability, p_e.

F_T^{-1} is the inverse CDF function and f_T is the probability density function of temperature on the patch.

The sample quantile corresponding to p_e is distributed according to:

$$
\begin{aligned}
q_{p_e} \quad &\sim \mathcal{N}\left(F_T^{-1}(p_e), \sigma^2_{q_{p_e}}\right) && (3.30)\\
\sigma^2_{q_{p_e}} \quad &= \tfrac{p_e(1-p_e)}{n[f_T(F_T^{-1}(p_e))]^2} && (3.31)
\end{aligned}
$$

The variance of upper tail probability mass estimate can be found by standard propagation of uncertainty principles:

$$
\sigma^2_{p_e} = \left(\frac{\partial p_e}{\partial q_{p_e}}\right)^2 \cdot \sigma^2_{q_{p_e}} + \ HOT. \tag{3.32}
$$

Where

$$
\begin{aligned}
\frac{\partial p_e}{\partial q_{p_e}} \quad &= \tfrac{\partial}{\partial q_{p_e}}\left(1 - \int_0^{q_{p_e}} f_T dT\right)\\
&= \tfrac{\partial}{\partial q_{p_e}}\left(-F_T(q_{p_e}) + F_T(0)\right)\\
&= -f_T(q_{p_e}) && (3.33)
\end{aligned}
$$

This dictates that estimates of extreme upper tail integrals carry large relative uncertainties.

3.2.5 Gradient Boosted Regression Trees

A data driven machine learning model is used to predict quantiles, $\{\hat{\theta}_\tau\}$, required to reconstruct cumulative density functions as shown in equation 3.27. The desired quantiles of the cladding surface temperature and TKE fields are conditioned upon the local core state denoted here by $\mathbf{p}$. The vector $\mathbf{p}$ contains the local thermal hydraulic conditions in the core and local geometric grid and pin factors. The purpose of this vector is to uniquely identify each CTF face in the core through a set of explanatory features. The problem is then of conditional quantile prediction and this is performed by gradient boosted regression trees in this work.

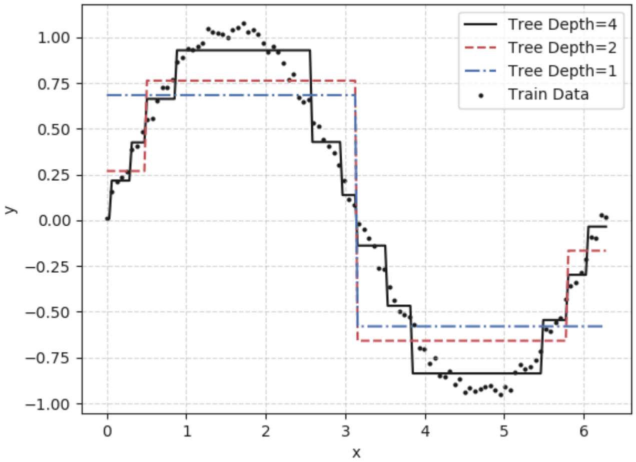


Figure 3.16: Single CART regression tree stump comparing a fit of depth 1 and 2 to the function $y = sin(x) + \varepsilon,\ x \in [0, 2\pi], \varepsilon \sim \mathcal{N}(0, 0.001)$.

The modern gradient boosting algorithm was developed by Friedman et. al. (1998) [48], [49]. The development of generalized boosting was significant because it re-envisioned previous boosting algorithms such as AdaBoost as special cases of gradient boosting with specific loss functions. Gradient boosting is a numerical optimization procedure carried out in function space with the goal of finding a function $\mathcal{F}_M$ that maps the inputs $\mathbf{p}$ to y where $\mathcal{F}_M$ is given by equation 3.34.

$$\mathcal{F}_M = \text{argmin}_{\mathcal{F}} \mathop{\mathbb{E}}_{y,\mathbf{p}} (L(y, \mathcal{F}(\mathbf{p}))) \tag{3.34}$$

Where $L(\cdot)$ is a differentiable loss function. $\mathbf{p}$ are known as predictive features and y is the response. The paired set $\{\mathbf{p}, y\}$ is referred to as the training data set. In the standard gradient boosting algorithm the functional form of $\mathcal{F}$ is chosen to be an additive model of the form 3.35:

$$\mathcal{F}_M(\mathbf{p}, \gamma, \mathbf{a}; \mathbf{b}) = \sum_{m=0}^{M} \gamma_m h_m(\mathbf{p}, a_m; b_m) \tag{3.35}$$

Where M is the number of constituent sub-models and $\{\gamma_m, a_m\}$ are coefficients and free parameters requiring fitting in each sub-model. b_m represent sub-model hyperparameters that are fixed at user set values. Rather than fitting all sub-models simultaneously, gradient boosting greedily fits the sub-models to the gradient of the loss function in a stage wise fashion as shown in algorithm 3. The loop depicted in line number 8-10 performs the computation of the loss function gradient at each boosted iteration m in the algorithm. In the literature the vector of gradients is sometimes referred to as the pseudo residuals vector since if the loss function is taken to be the L2 loss, the gradient of the loss is proportional to the residual vector as shown in equation 3.36.

$$\begin{aligned}(y_i - \mathcal{F}_m(x_i)) &\propto \frac{\partial[(y_i - \hat{y}_i)^2]}{\partial \hat{y}_i} \\ &\propto \frac{\partial(y_i^2 + \hat{y}_i^2 - 2y_i\hat{y}_i)}{\partial \hat{y}_i} = 2\hat{y}_i - 2y_i \end{aligned} \tag{3.36}$$

Each sub-model, h_m, is referred to as a weak learner and is defined to be a typical classification or regression tree (CART) depending on the problem context. Methods for fitting, pruning the decision trees and optimizing the numerical implementation of finding the best splits when fitting the tree to the pseudo-residuals are left to Friedman et. al. [50]. Provided CART trees are used for the weak learners, the free (sub-) model parameters a_m take the form of split locations and regional constants which are fitted to produce a piecewise constant predictor. In this case the user set hyperparameters, b_m, are taken to be the maximum tree depth allowed when fitting each weak learner to the pseudo-residuals.

Examples of single dimensional fitted CART trees with a depth of 1 and 2 are shown in figure 3.16. As the depth of a tree is increased the decision tree prediction converges onto the data; however, a single decision tree of large depth is prone to over fitting the data. Decision trees produce a piecewise constant prediction, and therefore since the final boosted model is a linear combination of decision tree models (weak learners), the fitted boosted model is piecewise constant.

Algorithm 3

Gradient boosting algorithm [50].

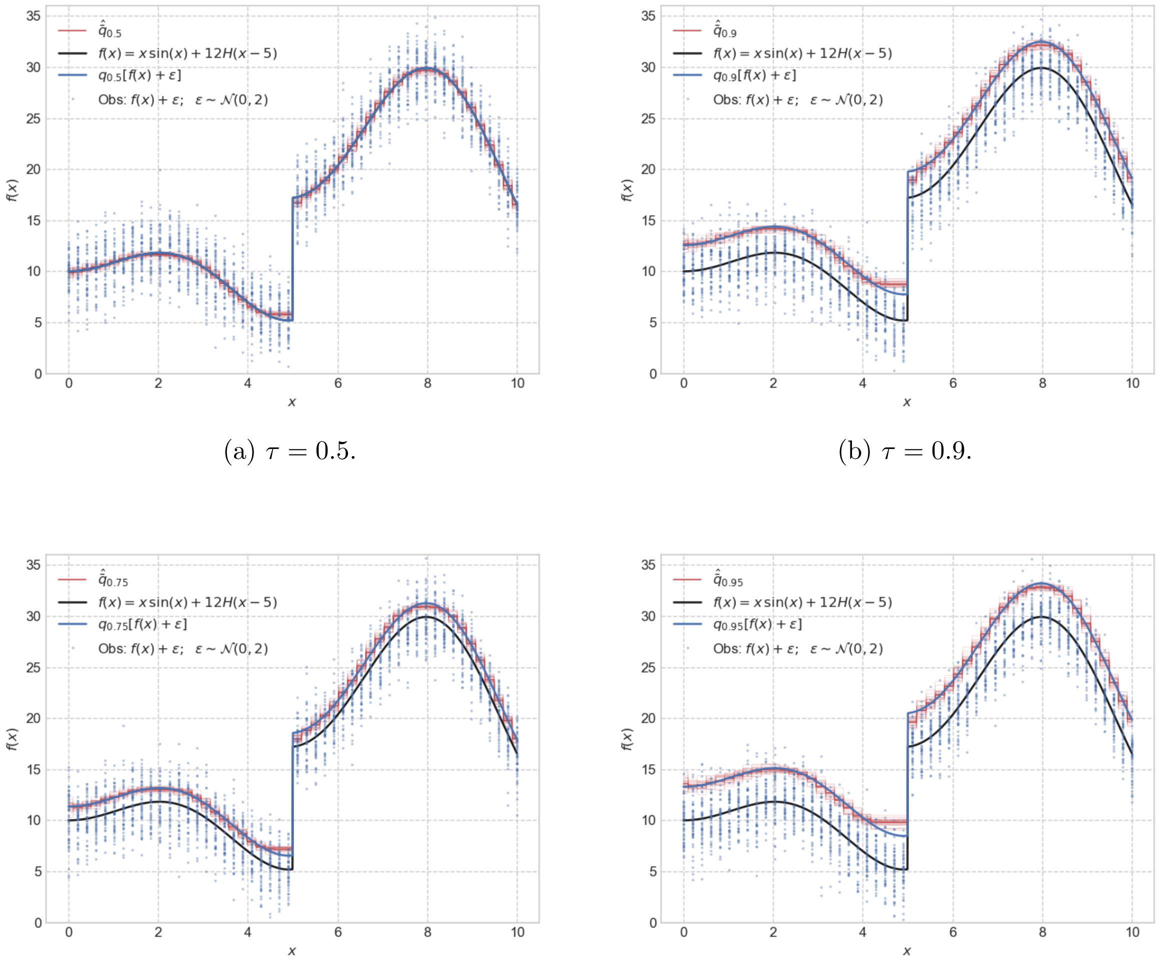
```
**Initialization**
(1) Training set $\{(p_i, y_i)\}_{i=1}^{n}$.
(2) Differentiable loss function $L(y, \mathcal{F}(p))$.
(3) Number of iterations $M$.
(4) Initialize model with a constant value: $\mathcal{F}_0(p) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$.
for $m = 1$ to $M$ do
    Compute the pseudo-residuals:
    for $i = 1, \ldots, n$ do
        $r_{im} = -\frac{\partial L(y_i, \mathcal{F}_{m-1}(p_i))}{\partial \mathcal{F}_{m-1}(p_i)}$
    end for
    Fit a weak learner $h_m(p; a_m)$ to pseudo-residuals, $r_m$:
        $h^* = \arg\min_{a_m}(||h_m(p; a_m) - r_m||)$
        Training data set is $\{(p_i, r_{im})\}_{i=1}^{n}$
    Compute multiplier $\gamma_m$ : $\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L\left(y_i, \mathcal{F}_{m-1}(p_i) + \gamma h_m^*(p_i)\right)$
    Update the model: $\mathcal{F}_m(p) = \mathcal{F}_{m-1}(p) + \nu \gamma_m h_m^*(p)$.
end for
Output $\mathcal{F}_M(p)$.
```

Where ν is a tunable constant in $(0, 1]$ called the learning rate. A value of $\nu < 1$ reduces the contribution of each weak learner in the final model. Reducing the learning rate increases resilience to over fitting but results in a proportional increase in the number of boosted iterations to achieve the same level of convergence of the boosted tree chain. Smaller values for the learning rate result in performing smaller steps in function space so we are less likely to overshoot the optimal function. Provided boosting produces a stage-wise additive model, overshoots are problematic since the final model carries memory of previous iterations. A typical learning rate is $\nu \leq 0.05$ but a balance between computation time and model prediction accuracy on a testing set should be considered when setting this value.

For gradient boosted quantile regression, the loss function given by equation 3.25 is substituted into algorithm 3.

$$L(y, \mathcal{F}(p); \tau) = \left[(1-\tau) \sum_{y \leq \mathcal{F}(p)} (y_i - \mathcal{F}(p))\right] - \left[\tau \sum_{y > \mathcal{F}(p)} (y_i - \mathcal{F}(p))\right] \tag{3.37}$$

Where τ is the user set quantile of interest. For instance, if the 95% percentile is desired then $\tau = 0.95$.



(a) $\tau = 0.5$.

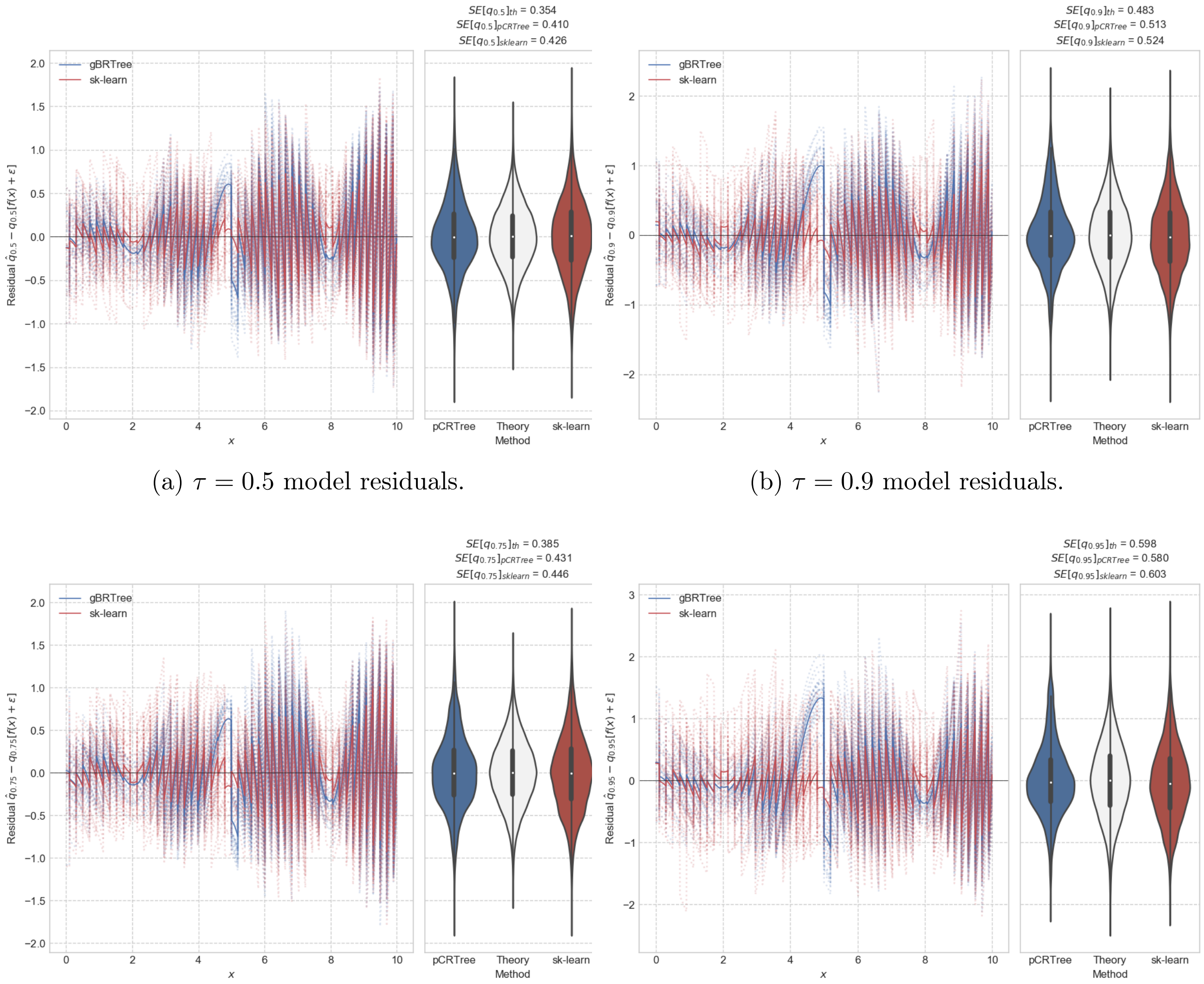
(b) $\tau = 0.9$.

(c) $\tau = 0.75$.

(d) $\tau = 0.95$.

Figure 3.17: Gradient boosted quantile regression example.

Two candidate gradient boosting implementations were evaluated for use in this work. A test problem was constructed to check that the standard gradient boosting regression technique is resilient to discontinuities in the data since the response fields are expected to abruptly change when moving across a spacer grid. The function $y = x\sin(x) + 12\mathcal{H}(x-5) + \varepsilon$ with $x \in [0, 10]$ was used for testing. $\mathcal{H}$ denotes the Heaviside function. Synthetic noise, $\varepsilon \sim \mathcal{N}(0, 2)$, was applied to the piecewise smooth test function. 5000 total samples were drawn from the test function. The test data was spatially aggregated to 100 axial levels to mimic CTF axial grid spacing. Four separate quantile regressors ($\tau = \{0.5, 0.75, 0.9, 0.95\}$) were then fit to the data by algorithm 3. The predicted quantiles, conditioned on x, were compared to the expected results, as shown in figure 3.17. For this problem results from the scikit-learn gradient boosting implementation are compared to a custom boosting implementation, named $pCRTree$, developed specifically to solve the classification and quantile regression problems in this work. The residuals provided figure 3.18 show that the custom implementation performs similarly to the well-tested scikit-learn implementation with both models agreeing with the theoretical large-sample quantile distributions provided by equation 3.28.



(a) $\tau = 0.5$ model residuals.

(b) $\tau = 0.9$ model residuals.

(c) $\tau = 0.75$ model residuals.

(d) $\tau = 0.95$ model residuals.

Figure 3.18: Gradient boosted quantile regression residual summary. The theoretical residual distribution were computed according to equation 3.28.

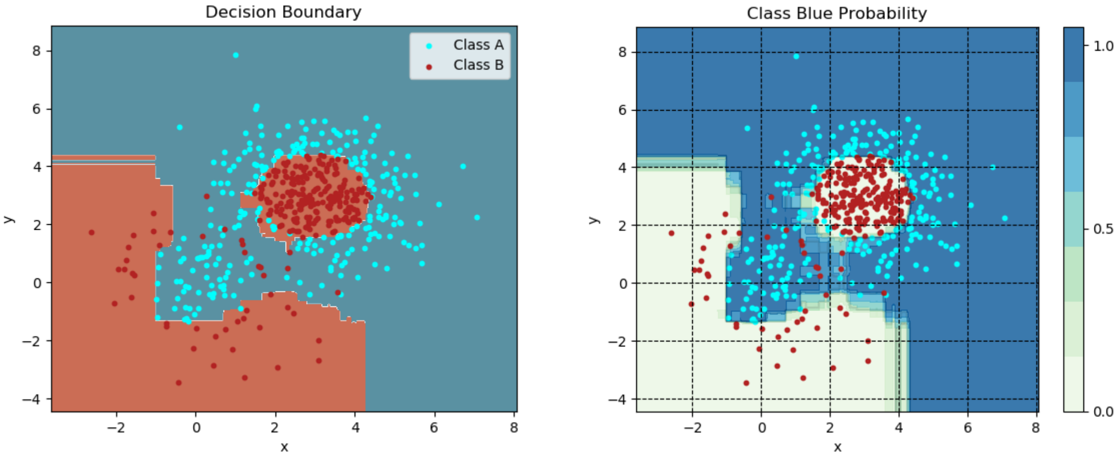
The parametric copula family which corresponds to a given set of local core conditions must also be determined in order to fully specify the copula required in the definition of the joint density function given in equation 3.5. This gives rise to a typical supervised classification problem which will be solved using the gradient boosting method. The family of copula, $\mathbf{\Theta}_c$, i.e. either Frank, Clayton, or Gumbel, should be determined by the classifier provided a local core thermal hydraulic state vector $\mathbf{p}$. The gradient boosted classification algorithm can be recovered by substituting the exponential loss function shown in equation 3.38 into algorithm 3.

$$
\begin{aligned}
L(y, F(\mathbf{p})) &= \mathbb{E}\left[e^{-yF(\mathbf{p})}\right] \\
&= \frac{1}{N}\sum_{i}^{N} e^{-y_i F(\mathbf{p}_i)}
\end{aligned} \tag{3.38}
$$

Where y takes on integer values in $\{-1, 1\}$ for the classification problem.

A two-class test case was devised to validate the custom boosting implementation against the scikit-learn implementation. This is meant as a demonstration for qualitative understanding of the boosted classification algorithm output. In this problem red and blue points were emplaced in a two dimensional space as shown in figure 3.19. These predefined points form the necessary training data for the test problem. The goal of this test is to segregate the input space, $\{x, y\}$, into regions which are most likely to contain only one color of points. Results from the test 2-D classification test problem are shown in the figure. In much the same way regression trees produce piecewise constant predictions, the boosted classifier partitions the input space by segregating the space along orthogonal splits.

Classifier predictions are made by tallying the weighted predictions of each constituent fitted classification weak learner (CART tree) in the boosted model. By tallying the predictions of all trees in the model one can obtain a probability mass function over all possible classes. The probability mass function is visualized in figure 3.19(b) wherein the predicted probability of the "blue" class obtained from the fitted boosted model is shown. The most likely class at each requested point is taken as the output of the classification model.



(a) Class predictions.

(b) Blue class probability.

Figure 3.19: Two-class gradient boosted classifier example.

Finally, gradient boosting can be used to estimate the relative explanatory power of each predictive variable included in the model. This is done by tallying the number of times a particular dimension is split upon in the CART fitting process and weighting these splits by a gain measure. The gain can be interpreted as the benefit of making a particular split in the decision tree as measured by improvement in R^2 for regression problems, or improvement in class purity as measured by the Gini impurity ratio for classification problems. Finally, a weighted sum of the split gains of each tree is performed over all boosted iterations to obtain relative variable importances. Friedman provides a detailed explanation of this relative variable importance computation [49]. This ability of gradient boosting to identify important predictors has been exploited in email filter and web page ranking applications [51], [52]. In the case of copula prediction, exogenous variables which are extraneous can be detected and eliminated from the model to save computation time. The ability to identify unimportant features is demonstrated in section 5.1.2 in figure 5.4.

3.2.6 Monte Carlo Crud Estimation

Before applying a sampling procedure to the temperature and TKE surface fields on a given CTF face, the the joint density function is reconstructed from copula and marginal distributions. On CTF face j, the gradient boosted model is queried for the conditional quantiles and the copula shape parameter:

$$\begin{aligned}\hat{\theta}_j \leftarrow & \mathcal{F}_M(\mathbf{p}_j, \mathbf{z}_j) \\ \hat{\theta}_j = & \{\hat{\theta}_{j,c}, \hat{\theta}_{j,\{T,k\}}\}\end{aligned} \tag{3.39}$$

The CTF face index is dropped in the remainder of this section to reduce subindex clutter.

Provided estimates for the conditional quantiles, $\hat{\theta}_{\{T,k\}}$ in each face, the margins are defined by their quantile functions as shown in equation 3.27. After separate reconstruction of the margins and applying the boundary heat flux independence assumption shown previously in equation 3.5, equation 3.40 defines the joint density.

$$h(T, k, q'') = f_T(T; \hat{\theta}_T) f_k(k; \hat{\theta}_k) f_{j,q''}[c_{T,k}(F_T(T; \hat{\theta}_T) F_k(k; \hat{\theta}_k); \hat{\theta}_c)] \tag{3.40}$$

After reconstruction of the joint temperature, boundary heat flux and turbulent kinetic energy distribution on each patch, the goal is to draw samples from the copula based distribution with arbitrary margins. This is accomplished through the methods outlined by section 3.2.2 in equation 3.14.

With the ability to draw samples from a multivariate distribution now established, the Monte Carlo approximation of integral 3.3 follows and is represented by equation 3.41. Let $X = \{T, k, q''\}$.

$$\mathbb{E}(\mathcal{G}(X)) \approx \frac{1}{N} \sum_i^N \mathcal{G}(X_i),\ \mathbf{X} \sim h \tag{3.41}$$

Rather than sampling from the density function h, one may draw samples from an alternate proposal density distribution denoted, $\tilde{h}$, and appropriately weight the samples by the probability ratio of the original density to the proposal density, $h/\tilde{h}$, so to avoid introducing bias in the approximation of the expected value. This leads to the importance sampling formulation of approximating integral 3.3 which is given in equation 3.42.

$$\mathbb{E}(\mathcal{G}(X)) \approx \frac{1}{N}\sum_{i}^{N}\mathcal{G}(X_i)\frac{h(X_i)}{\tilde{h}(X_i)},\ \mathbf{X} \sim \tilde{h} \tag{3.42}$$

In principal the proposal distribution, $\tilde{h}$, may be radically different from the target density, h. In practice computing the optimal choice for the proposal density distribution is non-trivial. The generic original and proposal densities can be written simply as shown in equation 3.43. A method for determining a near optimal proposal distribution is provided in section 4.1.2.

$$\begin{aligned} h(T,k,q'') &= c(F_T(T),F_k(k))f_T f_k f_{j,q''} \\ \tilde{h}(T,k,q'') &= \tilde{c}(\tilde{F}_T(T),\tilde{F}_k(k))\tilde{f}_T \tilde{f}_k f_{j,q''} \end{aligned} \tag{3.43}$$

$$\mathbb{E}(\mathcal{G}(X)) \approx \frac{1}{N}\sum_{i}^{N}\mathcal{G}(X_i)\omega(X_i),\ \mathbf{X} \sim \tilde{c}(\tilde{F}_T(T),\tilde{F}_k(k))\tilde{f}_T \tilde{f}_k f_{j,q''} \tag{3.44}$$

with the probability ratio of the target density to the proposal given by 3.45:

$$\omega_i = \frac{h_i}{\tilde{h}_i} = \frac{f_T(T_i)f_k(k_i)c(F_T(T),F_k(k))}{\tilde{f}_T(T_i)\tilde{f}_k(k_i)\tilde{c}(\tilde{F}_T(T),\tilde{F}_k(k))} \tag{3.45}$$

In some scenarios the proposal or target density is only known up to a constant, e.g. $h^* = ch(X)$. In this case the probability ratio is known up to a constant of proportionality and requires renormalization:

$$\mathbb{E}(\mathcal{G}(X)) \approx \frac{\sum_{i}^{N}\mathcal{G}_i(X)\omega_i(X)}{\sum_{i}^{N}\omega_i(x)},\ \mathbf{X} \sim \tilde{c}(\tilde{F}_T(T),\tilde{F}_k(k))\tilde{f}_T \tilde{f}_k f_{j,q''} \tag{3.46}$$

This is known as self normalizing importance sampling. Traditional importance sampling may be applied in this case since h and $\tilde{h}$ are properly normalized density functions in this hi2lo application which integrate to 1 over their respective support.

3.3 Propagating Crud Through Time

In the construction of the hi2lo method an assumption is made about the location of hot and cold spots on the rod surface as a function of time. The presence of hot and cold spots downstream spacer grids and the location on the rod surface these spots occupy are assumed to be principally governed by the geometry of the mixing vanes and geometric layout of the fuel and guide tubes. The influence of flow rate and core power on the relative location of the hot spots on the rod surface are assumed to be second order effects and are not explicitly captured by the hi2lo methodology at present.

3.3.1 Hot Spot Stationarity

The assumption that the location of eddy regions which develop downstream of spacer grids near the rod surface are principally governed by the grid geometry coupled with an assumption of steady state constant flow conditions leads to the notion of hot spot stationarity in time since the geometry of the core does not change throughout a cycle. To achieve a stable location of hot and cold spots on the rod surface through time, the hi2lo methodology establishes a mapping from the sample space containing all possible outcomes of surface temperature and TKE sample pairs to a location on the rod surface. To accomplish this, the order statistics of the joint temperate and turbulent kinetic energy distribution in each CTF patch are computed and the condition that the highest order statistic always falls on the same location within a given CTF patch is enforced.

Order statistics are well defined for a single dimensional random variable but in higher dimensions a consistent ordering is not possible. Therefore, as shown in equation 3.47 a convex combination of the surface fields with user specified weights is used to reduce the random vector of T, TKE, and q'' samples on any given CTF patch into a univariate random variable denoted by $\mathbf{m} = \{m_0, m_1, ...m_N\}$.

$$m_i = w_T \left(\frac{T_i - T_{min}}{T_{max} - T_{min}} \right) + w_k \left(\frac{k_i - k_{min}}{k_{max} - k_{min}} \right) + w_q^{''} \left(\frac{q_i^{''} - q_{min}^{''}}{q_{max} - q_{min}} \right) \tag{3.47}$$

Where the sample remapping coefficients $\{w_T, w_k, w_{q''}\}$ are set at runtime by the hi2lo model user and sum to 1:

$$w_T + w_k + w_{q''} = 1 \tag{3.48}$$

Next the order statistics of $\mathbf{m}$ are computed such that $\mathbf{m}' = \{m_{(0)} < m_{(1)} < ...m_{(N)}\}$. As shown in figure 3.20, the ordered samples are then emplaced on the CTF patch in an organized manner. The sample space is denoted by Ω and a specific temperature, TKE, and boundary heat flux sample is denoted by $\mathcal{F}_i$ as a single dot residing inside the sample space. The path taken on the patch surface is user controllable and is taken to be a simple serpentine left-to-right pattern in this work. Typical values for the remapping coefficients are $w_T = 0.6, w_k = 0.4, w_{q''} = 0$. With this setting, relatively high temperate and low TKE samples are likely to remain in the same location on the rod surface over multiple resampling events.

Since the crud simulation package utilized in this work is one-dimensional, the pattern chosen for sample emplacement has no influence on the integrated crud result over a patch. This would not be the case if the crud simulation package modeled a fully 3-D crud layer as the state of the neighboring crud nodes in that case would matter. Interestingly, since the hi2lo model user specifies the sample remapping pattern at run time, a physically realistic pattern could be prescribed on the rod surface - even prescribed as a function of local core conditions leading to a hybrid strategy between the current pure statistical hi2lo procedure and the spatial remapping procedure implemented by Salko et. al [19].

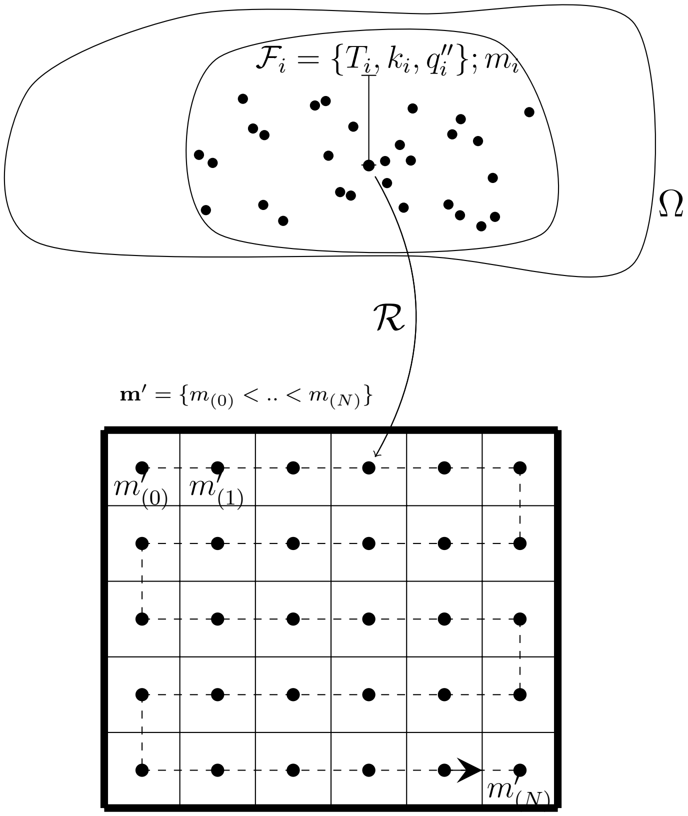

$\mathcal{F}_i = \{T_i, k_i, q_i''\}; m_i$
Ω
$\mathcal{R}$
$\mathbf{m}' = \{m_{(0)} < .. < m_{(N)}\}$
$m'_{(0)}$
$m'_{(1)}$
$m'_{(N)}$

Figure 3.20: Sample remapping. A single CTF face is shown as the bold square with a grid overlay partitioning the patch surface. In this figure the number of samples per CTF face is $N = 30$.

3.3.2 Time Stepping Scheme

Propagating the importance weights through discrete resampling steps requires special attention to ensure bias is not introduced into the crud results when moving through multiple time steps and VERA states. The key observation is that the importance weights computed at a given resampling event should neither be double-counted nor discarded in subsequent resampling steps. Instead, the importance weights are time averaged to achieve the correct importance weights used to evaluate the total crud mass on the rod surface at any given time step (resampling step).

Figure 3.21 depicts the time stepping scheme for a single patch. The patch index is not shown to reduce visual clutter. The VERA state-point index is denoted by ℓ. At the beginning of each VERA state-point the power distribution and thermal hydraulic conditions in the core change. Here it is assumed there is a constant power profile and constant flow conditions over a VERA depletion step.

In figure 3.21 dotted arrows represent sampling from a probability distribution. $\mathcal{R}$ is the spatial reordering map from sample space to a position on the rod surface. $\mathcal{G}$ represents the crud generation function provided by the crud simulation package which takes thermal hydraulic boundary conditions as input in addition to the previous crud state and produces a new crud state. Resampling events are performed at time intervals Δt_s and this interval is set at runtime by the user. The influence of the resampling interval size is discussed in section 4.1.4. The resample index is denoted by s. The subscript, $(\cdot)_s$, denotes evaluation at a resampling event. A prime-notated variable, $(\cdot)'$, represents a spatially re-mapped sample as in figure 3.20. At each resampling event the temperature, TKE and boundary heat flux samples are independently drawn from the reconstructed joint probability $\hat{H}^\ell$. Assume that the resample time step is constant and equal to Δt_s.

A dependency in time is introduced by the crud growth step because the current crud growth rate depends on the previous crud state (e.g. a severely thick crud deposit will

$$\begin{tikzcd}
 & \underline{\text{VERA STATE : } \ell = 0} & & & & \underline{\text{VERA STATE : } \ell = 1} & \\
 & \hat{H}^{\ell_0} \leftarrow \mathcal{F}_M(\mathbf{p}^{\ell_0}) \arrow[dl, dashed] \arrow[d, dashed] \arrow[dr, dashed] & & & & \hat{H}^{\ell_1} \leftarrow \mathcal{F}_M(\mathbf{p}^{\ell_1}) \arrow[dl, dashed] \arrow[d, dashed] \arrow[dr, dashed] & \\
\{\mathbf{X},\omega\}^{\ell_0}_{s_0} \arrow[d, "\mathcal{R}"] & \{\mathbf{X},\omega\}^{\ell_0}_{s_1} \arrow[d, "\mathcal{R}"] & \{...\}^{\ell_0}_{s_2} \arrow[d, "\mathcal{R}"] & . & \{\mathbf{X},\omega\}^{\ell_1}_{s_3} \arrow[d, "\mathcal{R}"] & \{\mathbf{X},\omega\}^{\ell_1}_{s_4} \arrow[d, "\mathcal{R}"] & \{...\}^{\ell_1}_{s_5} \arrow[d, "\mathcal{R}"] \\
\{\mathbf{X}',\bar{\omega}'\}^{\ell_0}_{s_0} \arrow[d, "\mathcal{G}(\mathbf{X}')"] & \{\mathbf{X}',\bar{\omega}'\}^{\ell_0}_{s_1} \arrow[d, "\mathcal{G}(\mathbf{X}'|\mathbf{C}_{s_0})"] & \{...\}^{\ell_0}_{s_2} \arrow[d, "\mathcal{G}(\mathbf{X}'|\mathbf{C}_{s_1})"] & . & \{\mathbf{X}',\bar{\omega}'\}^{\ell_1}_{s_3} \arrow[d, "\mathcal{G}(\mathbf{X}'|\mathbf{C}_{s_2})"] & \{\mathbf{X}',\bar{\omega}'\}^{\ell_1}_{s_4} \arrow[d, "\mathcal{G}(\mathbf{X}'|\mathbf{C}_{s_3})"] & \{...\}^{\ell_1}_{s_5} \arrow[d, "\mathcal{G}(\mathbf{X}'|\mathbf{C}_{s_4})"] \\
\{\mathbf{C},\bar{\omega}'\}^{\ell_0}_{s_0} \arrow[d, "\sum"] \arrow[ur] & \{\mathbf{C},\bar{\omega}'\}^{\ell_0}_{s_1} \arrow[d, "\sum"] \arrow[ur] & \{...\}^{\ell_0}_{s_2} \arrow[d, "\sum"] \arrow[urr] & . & \{\mathbf{C},\bar{\omega}'\}^{\ell_1}_{s_3} \arrow[d, "\sum"] \arrow[ur] & \{\mathbf{C},\bar{\omega}'\}^{\ell_1}_{s_4} \arrow[d, "\sum"] \arrow[ur] & \{...\}^{\ell_1}_{s_5} \arrow[d, "\sum"] \\
C^{\ell_0}_{m,s_0} \arrow[r, |-|, "\Delta t_s"'] & C^{\ell_0}_{m,s_1} & C^{\ell_0}_{m,s_2} & . & C^{\ell_1}_{m,s_3} & C^{\ell_1}_{m,s_4} & C^{\ell_1}_{m,s_5}
\end{tikzcd}$$

Figure 3.21: Multi-state point time stepping overview. Dashed arrows represent sampling from a distribution. Solid arrows represent a functional mapping or operation.

behave differently than a small crud deposit). This time dependence is denoted by the diagonal arrows linking the crud states together across resampling events in figure 3.21.

Since the resampling time step size is constant the time corresponding to any given sample event is computed by multiplying the resampling event index by the constant resampling step size:

$$t_{s_n} = n\Delta t_s \tag{3.49}$$

The time averaged sample weights are in general updated by equation 3.50 at each resampling step. Recall that the ratio of the target density and the sampling distribution density is denoted by ω and is given in equation 3.45.

$$\bar{\omega}'_{s_n,i} = \left(\frac{\sum_l^{n-1} \Delta t_{s_l}}{\sum_l^n \Delta t_{s_l}} \right) \bar{\omega}'_{n-1,i} + \left(\frac{\Delta t_{s_n}}{\sum_l^n \Delta t_{s_l}} \right) \omega'_{n,i} \tag{3.50}$$

Where i is the sample index within a single CTF face.

After applying the constant resample step size assumption the time averaged sample weights are updated at each resampling step according to equation 3.51.

$$
\begin{aligned}
\bar{\omega}'_{s_n,i} &= \left(\frac{(n-1)\Delta t_s}{n\Delta t_s}\right)\bar{\omega}'_{n-1,i} + \left(\frac{\Delta t_s}{n\Delta t_s}\right)\omega'_{n,i} \\
&= \left(\frac{(n-1)}{n}\right)\bar{\omega}'_{n-1,i} + \left(\frac{1}{n}\right)\omega'_{n,i} \qquad (3.51)
\end{aligned}
$$

After each resampling event the total crud mass, C_m on a CTF face, is computed by a wighted sum given in equation 3.52. The sampling and weighting procedure is depicted in figure 3.22.

$$
C_m = \left(\frac{A}{\sum_i^N \bar{\omega}'_i}\right)\sum_i^N C'_i\bar{\omega}'_i \qquad (3.52)
$$

Where N is the number of samples drawn per CTF face and A is the surface area of the j^{th} CTF face. The re-stepping index n and the patch index j are dropped to reduce clutter in equation 3.52.

The crud stepping strategy is explicit in time as the next crud samples are drawn from knowledge of the previous crud state and the current thermal hydraulic conditions. Feedback between the crud state and the underlying density functions are not captured in the hi2lo stepping scheme. The presence of crud influences the rod surface temperature distribution due to two competing effects: Providing favorable conditions for bubble nucleation and an increased thermal resistance. Though these effects are slight they impact not only the mean rod surface temperature but also higher moments about the mean of the probability densities of surface temperature. These higher order feedbacks are not currently considered, however it is possible to incorporate the current crud state into the explanatory variable set used by the machine learning model to resolve this feedback between the crud thickness and the higher order moments of temperature and TKE about the CTF prediction. This remains as future work as the impact of the crud layer on the flow field is hypothesized to only significantly matter if the crud deposits significantly shifts the locations of the onset of nucleate boiling within the core.

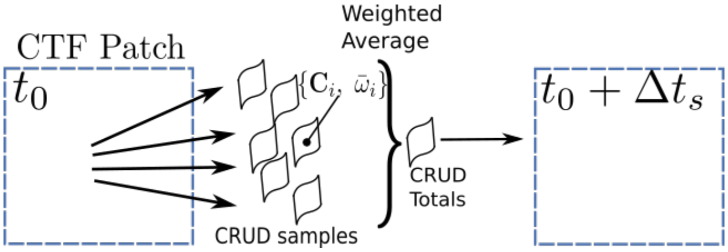


Figure 3.22: Time step procedure depicted for a single CTF face. A single resampling event is shown.

3.4 Method Summary

The time stepping, importance sampling and surface sample remapping strategies may be included into algorithm 2 to provide a detailed overview of the hi2lo procedure. The expanded version of the hi2lo model is given in algorithm 4.

Algorithm 4 begins with the same pre-processing and machine learning procedure outlined in the generic hi2lo algorithm 2. *Lines 2-5:* An overview of the CFD data pre-processing procedure is given in section 5.1.1 and the gradient boosted quantile regression model fitting routine is discussed in section 3.2.5. *Line 9:* Reconstructs the temperature and TKE CDFs from quantiles. The details of reconstructing a cumulative density function using a set of predicted quantiles was discussed in section 3.2.4. *Line 10:* The simplified treatment of the boundary heat flux distribution was discussed in section 3.2.1 and the application of this simplification resulted in a compact representation of the joint distribution of temperature, TKE, and BHF in each CTF face given in equation 3.5. *Lines 15-16:* Involves the construction of the importance sampling distributions. The details of this procedure are given in section 4.1.2. *Line 17-18:* Draws samples from the reconstructed joint distribution and follows the standard importance sampling procedure outlined in section 3.2.6. *Line 19:* Performs the spatial remapping procedure required to preserve hot spot stationarity. The details of this procedure may be found in section 3.3.1. *Line 20:* The importance weights are time averaged according to method provided in section 3.3.2 before the the crud distribution is integrated in each CTF face at each resampling step.

Algorithm 4

Statistically based hi2lo method for time dependent crud prediction.

```
Initialization
(1) Pre-process training set.
   (1b) Fit the joint distribution parameters to known CFD data: $\theta(\mathbf{p}, \mathbf{z})$.
   (1c) def: $\theta \leftarrow \mathcal{F}_M(\mathbf{p}, \mathbf{z})$
(2) Train model: $\hat{\mathcal{F}_M} = \text{argmin}_{\mathcal{F}} \ \mathbb{E}\left[L(\mathcal{F}_M(\mathbf{p}, \mathbf{z}), \theta(\mathbf{p}, \mathbf{z}))\right]$
for VERA State, $v$ do
   for CTF face, $j$ do
      Evaluate ML model $\hat{\theta}_j \leftarrow \hat{\mathcal{F}_M}(\mathbf{p}_j, \mathbf{z}_j)$
        $\hat{\theta}_j = \{\hat{\theta}_{j,c}, \hat{\theta}_{j,\{T,k\}}$
      Reconstruct margins (CDFs) from quantiles
        $\hat{F}_{j,T} = Q^{-1}(T; \{\hat{\theta}_{j,\{T,k\}}\})$ , $\hat{F}_{j,k} = Q^{-1}(k; \{\hat{\theta}_{j,\{T,k\}}\})$
      Def $q''$ margin $f_{j,q''} = \delta_{(q''_{\text{j,ctf}})}$
      Reconstruct joint distribution $\hat{h}_j(\cdot|\hat{\theta}_j) = \hat{f}_{j,T}\hat{f}_{j,k}f_{j,q''}c(\hat{F}_{j,T}, \hat{F}_{j,k}; \hat{\theta}_{j,c})$
   end for
   for Resample time step, $s$, $\Delta t_s$ do
      for CTF face, $j$ do
         Def importance mixture quantile functions by equation 4.8
           $\tilde{Q}_{j,k} = \lambda_{0,k}\hat{Q}_{j,k} + \lambda_{1,k}Q_{\beta_k}(k; \vartheta_k), \quad \tilde{Q}_{j,T} = \lambda_{0,T}\hat{Q}_{j,T} + \lambda_{1,T}Q_{\beta_T}(T; \vartheta_T);$
           $\sum_i \lambda_i = 1, \quad \tilde{F}_{j,T} = \tilde{Q}^{-1}_{j,T}, \quad \tilde{F}_{j,k} = \tilde{Q}^{-1}_{j,k}$
         Def importance sampling distribution $\tilde{h}_j = \tilde{f}_{j,T}\tilde{f}_{j,k}c(\tilde{F}_{j,T}, \tilde{F}_{j,k}; \hat{\theta}_{j,c})$
         Draw samples $\mathbf{x} \sim \tilde{h}_j$
         Compute importance weights $\omega = \hat{h}_j(\mathbf{x})/\tilde{h}_j(\mathbf{x})$
         Re-map samples $\mathbf{x}', \omega' \xleftarrow{\mathcal{R}} \mathbf{x}, \omega$
         Update importance weights by equation 3.51
         Evaluate equation 3.3 via importance sampling
           $\mathbf{C}_s = \mathcal{G}(x'_i; \mathbf{C}_{s-1}, \mathbf{I}, \Delta t_s)$
         Crud mass at step $s$ in patch $j$:   $C_{s,j,m} = \left(\frac{A_j}{\sum_i^N \bar{\omega}'_i}\right) \sum_i^N C_{s,i,j,m}\bar{\omega}'_i$
      end for
   end for
end for
```

All model variables which are set at runtime are provided in table 3.1. Recommended settings are also provided by each model parameter.

Finally, the simplifications and assumptions made in the construction of the hi2lo model are reviewed. The model requires that within a CTF face no spatial information is retained. The crud simulation package, MAMBA, used in this work is a one-dimensional code and therefore this simplification has no influence on the integrated crud results within each CTF face. The boundary heat flux is treated independently from the rod surface temperature or TKE. This simplifies the copula model used to reconstruct the joint distribution in each face. Hot spots are assumed to remain stationary on the rods' surface as a function of time. This assumption can be slightly relaxed with tuning of the sample remapping procedure; however, this assumption simplifies the treatment of the time dependent nature of crud build up in the core. The assumption takes into account that as time marches forward, the rate of crud growth depends on the previous crud state. The viability of these assumptions will be investigated in the following sections. The aim of the hi2lo model is to reproduce gold standard crud results provided by high fidelity CFD coupled crud simulations. The application of the simplifications should not impede the ability of the model to reproduce the correct axial and total crud deposition on the rod surface.

Table 3.1: Runtime hi2lo model parameters.

Sym	Default Value	Purpose
Δt_s	25 $[days]$	Resampling time step size.
N	400	Number of samples drawn per CTF face.
N_{Q_k}	20	Number of quantiles used in TKE marginal distribution reconstruction.
N_{Q_T}	20	Number of quantiles used in temperature marginal distribution reconstruction.
w_k	0.6	TKE weighting factor used to remap samples on the rod surface.
w_T	0.4	Temperature weighting factor used to remap samples on the rod surface.
ϑ_T	$\{1.0, 0.9\}$	Beta distribution parameters for temperature importance distribution.
ϑ_k	$\{1.1, 1.2\}$	Beta distribution parameters for TKE importance distribution.
$\lambda_{0,T}, \lambda_{1,T}$	$\{0.6, 0.4\}$	Mixture weighting parameters for temperature importance distribution.
$\lambda_{0,k}, \lambda_{1,k}$	$\{0.7, 0.3\}$	Mixture weighting parameters for TKE importance distribution.
b_m	4	Gradient boosting model parameter: Max CART tree depth. A larger CART tree depth value improves the quality of fit of each weak-learner but increases overfitting.
ν	0.01	Gradient boosting model parameter: Learning rate.
M	4000	Gradient boosting model parameter: Number of boosted iterations.

4 | Method Exploration Under a Synthetic Data Source

In this chapter the univariate distribution reconstruction from quantiles, copula parameter fitting, Monte Carlo and importance sampling strategies are exercised with a synthetic data set. Synthetic data offers advantages over CFD born data for the purposes of testing and evaluating the efficacy of the proposed models.

The synthetic data conforms to a known functional form with specified distribution and bias parameters. These properties are useful for benchmarking and validation investigations which seek to ensure that the fitted models retain key statistical properties from the synthetic data set. This chapter does not introduce machine learning components and does not explore forward model predictions. See chapter 5 for hi2lo model performance when used in a predictive capacity.

The availability of synthetic data alleviates the need to generate comparatively expensive CFD results to test the hi2lo strategy. Some aspects of CFD fields are preserved in the synthetic data, including expected biases between CFD and CTF results that arise due to discrepancies in wall heat transfer closure models, among other differences. Turbulent dispersion of the temperature and near wall TKE distributions around spacer grids are emulated by the synthetic data model. Additionally, the dependence structure between the cladding surface temperature, boundary heat flux, and near wall TKE may be enforced by the synthetic data generation tool. Accounting for spatial autocorrelation in the surface fields was not pursued. Consequently the synthetic data is not a direct replacement to CFD data but serves as data source for method interrogation and integration testing.

The joint distribution model comprised of a copula and quantile functions are fit to the synthetic data in each CTF face independently. The number of quantiles used to reconstruct the quantile functions is a user controllable quantity, and was set to 20 in this case. Samples are drawn from the fitted joint temperature, TKE and BHF density models on each patch using standard Monte Carlo methods or importance sampling. The surface samples are provided to a crud simulation package as cladding-surface boundary conditions. Additional required bulk coolant properties such as the bulk fluid temperature and bulk concentration of soluble boron are supplied by CTF.

Time dependent crud simulation is also discussed. The interaction between the sample surface remapping strategy and the integrated crud results are discussed. Furthermore, the process by which importance sample weights are averaged over discrete time steps is discussed.

Speedups afforded by importance sampling are presented and contrasted against standard Monte Carlo sampling results. The sampling distributions utilized in importance sampling are informed by the physics of crud growth.

4.1 Generating Synthetic Data

Synthetic data generation begins by running standalone CTF on a single quarter symmetric pin. The CTF result is then augmented with tailored noise. The augmented synthetic surface fields may be constructed by equation 4.1.

$$\begin{aligned}\boldsymbol{X} &= \boldsymbol{\mu}_{ctf} + \boldsymbol{b} + \boldsymbol{\varepsilon} \\ &= \begin{pmatrix} T \\ k \\ q'' \end{pmatrix} = \begin{pmatrix} \mu_T \\ \mu_k \\ \mu_{q''} \end{pmatrix}_{ctf} + \begin{pmatrix} b_T \\ b_k \\ b_{q''} \end{pmatrix} + \boldsymbol{\varepsilon}(\mathbf{z}; \boldsymbol{\theta}), \end{aligned} \tag{4.1}$$

Where $\boldsymbol{\varepsilon}(\mathbf{z}, \boldsymbol{\theta})$ is a user controlled spatially dependent residual random vector with a mean of 0. This residual is shifted by a bias vector $\mathbf{b}$ and where $\mathbf{z}$ is the axial and azimuthal location on the rod surface. $\boldsymbol{\theta}$ represents user specified distribution parameters.

Equation 4.1 represents three continuous random surface fields. In practice a large number of independent and identically distributed samples are drawn in each CTF face from the underlying random field. Individual surface samples may be specified by equation 4.2.

$$X_{ij} = \mu_{j,\mathrm{ctf}} + b_j + \varepsilon_{ij}; \quad \varepsilon_j \sim h_j \tag{4.2}$$

Where the index j represents the j^{th} CTF face on the rod, and the index i is the sample index within the j^{th} CTF face. The distribution parameters are constant over a given CTF face and are represented by $\boldsymbol{\theta} = \{\theta_c, \{\theta_x\}\}$ where θ_c is the copula parameter and $\{\theta_x\}$ are the set of marginal distribution parameters.

Shown in equation 4.3, according to Sklar's theorem the surface residual temperature and TKE joint distribution may be decomposed into copula an marginal models on each CTF face:

$$h_j = c_j(F_k(k), F_T(T); \theta_{c_j}) f_T(T; \theta_{T_j}) f_k(k; \theta_{k_j}) \tag{4.3}$$

Where the copula parameter θ_{c_j} and the marginal temperature and TKE distribution parameters θ_{T_j} and θ_{k_j} are set at runtime of the synthetic data generation tool.

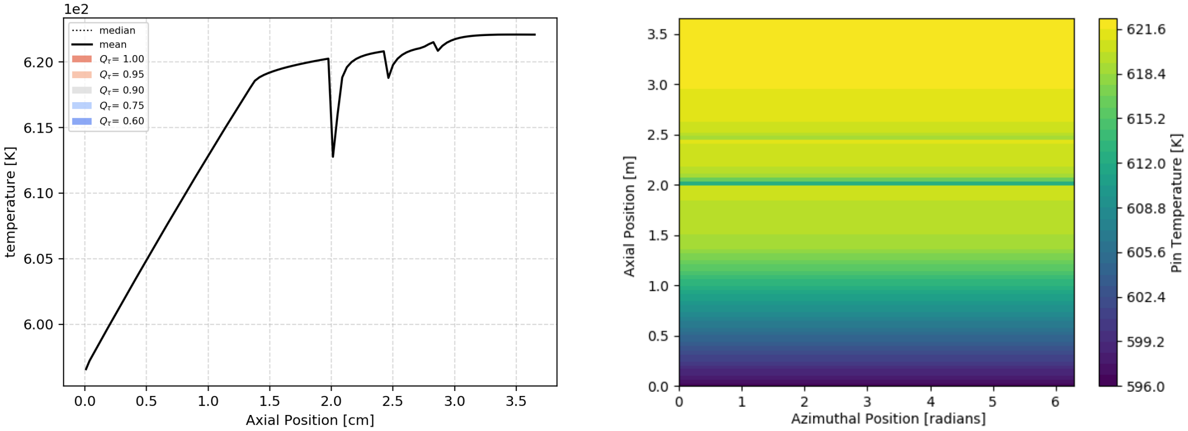
To allow for a great deal of flexibility in the synthetic data the copula family, Kendall's τ rank correlation coefficient and marginal distribution parameters are specified as a function of axial location and local TH conditions supplied by CTF. The copula's shape parameter, θ_c may be related to the rank correlation coefficient by equation 3.24 which is a one to one function for the Archimedean copula considered in this work.

4.1.1 Single Pin Synthetic Data Set

The original baseline CTF results are shown in figures 4.1 and 4.2. The CTF pin parameters are provided in table 4.1. The CTF result was produced from a quarter symmetric case, and therefore no azimuthal variation is observed.

Table 4.1: Single pin reference thermal hydraulic boundary conditions.

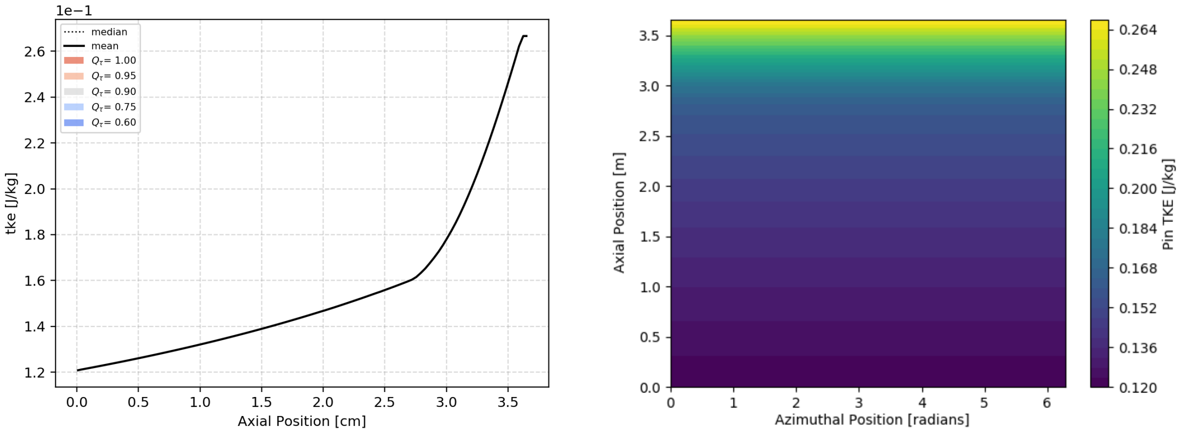
Setting	Value	Unit
Inlet Flow Rate	0.3	$[kg/s]$
Inlet Temperature	565	$[K]$
Pressure	2250	$[psia]$
Rod Outer Radius	0.425	$[cm]$
Pin Pitch	1.26	$[cm]$
Power Shape	constant	[]
Heat Flux	85.86	$[W/m^2]$
Rod Height	3.6275	$[m]$
Number of Grids	3	[]
Grid Locations	2.0, 2.4, 2.8	$[m]$



(a) Axial CTF cladding surface temperature result.

(b) 2-D rod map of CTF result.

Figure 4.1: Single pin CTF baseline temperature result. 160% nominal power conditions.



(a) Axial CTF cladding surface TKE result.

(b) 2-D rod map of CTF result.

Figure 4.2: Single pin CTF baseline TKE result. 160% nominal power conditions.

The boundary heat flux was uniform at $85.86[W/cm^2]$ which corresponds to approximately 160% nominal PWR power conditions.

Next synthetic noise was generated using copula and marginal distribution settings provided in table 4.2. The complete synthetic data generation input deck for this case along with references to the code are provided in appendix C.

Table 4.2: Per-span synthetic data generation settings.

Span 1	Node	z	Copula Settings	Margin Settings
N: 8000	1	0.0	Θ_c : Gaussian, θ_c : -0.6	$T \sim \beta(5.0, 5.0)$,$k \sim \mathcal{N}(0, 0.001)$
	2	2.0	Θ_c : Gaussian, θ_c : -0.6	$T \sim \beta(5.0, 5.0)$, $k \sim \mathcal{N}(0, 0.001)$
Span 2	Node	z	Copula Settings	Margin Settings
N: 8000	1	2.0	Θ_c : Clayton-90, θ_c : 2.0	$T \sim \beta(5.0, 2.7)$, $k \sim \beta(1.75, 5.0)$
	2	2.4	Θ_c : Frank-90, θ_c : 8.0	$T \sim \beta(5.0, 1.5)$, $k \sim \beta(1.75, 5.0)$
Span 3	Node	z	Copula Settings	Margin Settings
N: 8000	1	2.4	Θ_c : Clayton-90, θ_c : 2.0	$T \sim \beta(5.0, 2.7)$, $k \sim \beta(1.75, 5.0)$
	2	2.8	Θ_c : Frank-90, θ_c : 8.0	$T \sim \beta(5.0, 1.5)$, $k \sim \beta(1.75, 5.0)$
Span 4	Node	z	Copula Settings	Margin Settings
N: 8000	1	2.8	Θ_c : Clayton-90, θ_c : 2.0	$T \sim \beta(5.0, 2.7)$, $k \sim \beta(1.75, 5.0)$
	2	3.6	Θ_c : Frank-90, θ_c : 8.0	$T \sim \beta(5.0, 1.5)$, $k \sim \beta(1.75, 5.0)$

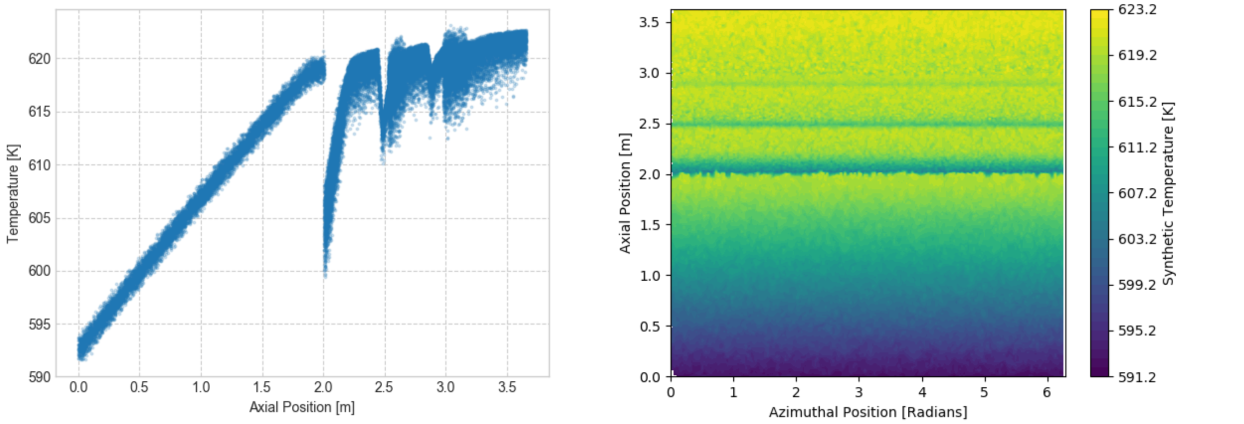
Samples are drawn with probability proportional to the inverse distance to the nearest specified node. Let the subscript $(\cdot)_u$ denote the location of the upstream span and $(\cdot)_d$ denote the downstream grid. d_{u_j} and d_{d_j} denote the distance from the centroid of the CTF face to the nearest upstream and downstream copula nodes respectively.

The mixture joint density model in any given CTF face is specified by equation 4.4.

$$h_j = \left(\frac{d_{u_j}}{d_d + d_u}\right) h_u + \left(\frac{d_{d_j}}{d_d + d_u}\right) h_d \tag{4.4}$$

Where h_u and h_d are the upstream and downstream joint density models respectively with parameters specified in table 4.2. For simplicity, two copula nodes were specified per span though more are possible for a finer grained control over the marginal and copula distributions. The copula nodes were located at the span extrema. This node specification pattern allows the synthetic data to mimic the expected sharp change in copula and marginal distributions when moving across spacer grids as seen in the raw CFD data presented in section 5.1.1 in figure 5.3.

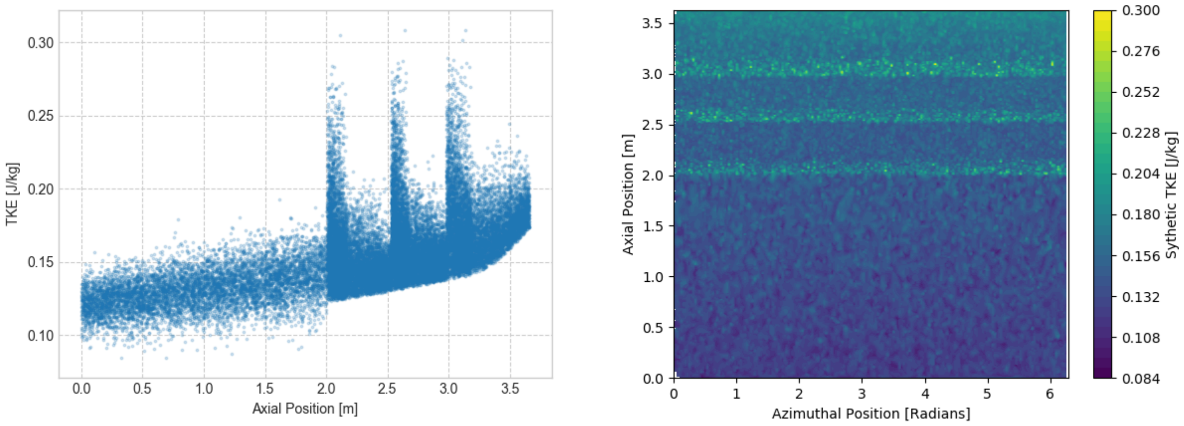
The copula models were sampled in each span, the original CTF result was augmented with the synthetically generated noise in accordance with equation 4.2.



(a) Spatial axial augmented CTF result.

(b) 2-D rod map of synthetically augmented CTF result.

Figure 4.3: Augmented CTF temperature result.



(a) Spatial axial augmented CTF result. (b) 2-D rod map of synthetically augmented CTF result.

Figure 4.4: Augmented CTF TKE result.

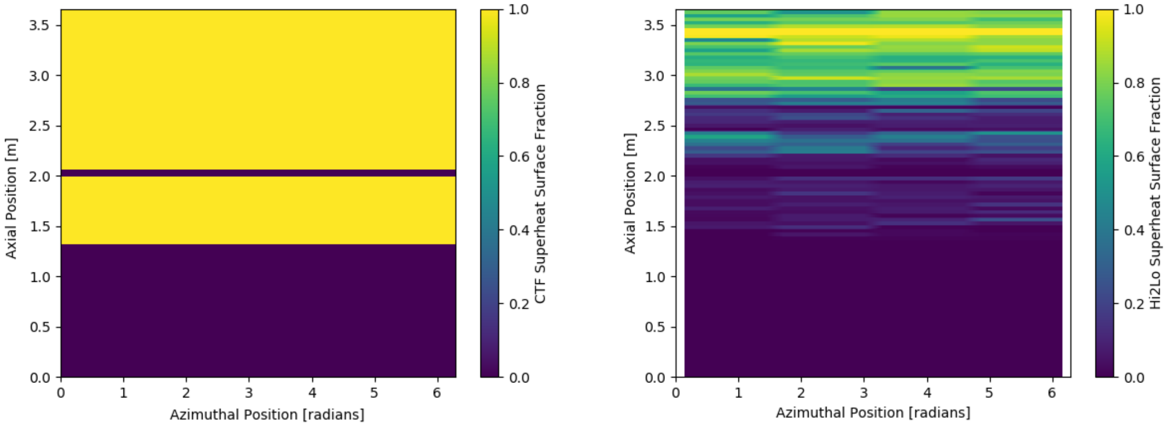
The augmented surface temperature and turbulent kinetic energy fields shown in figures 4.3 and 4.4 may be compared against the original CTF results provided in figures 4.1 and 4.2 respectively to qualitatively understand the fine scale surface variations introduced by the synthetic data model. No azimuthal variations are present in the augmented fields which would be present if a physics based model, such as CFD, were used. Additionally, no spatial autocorrelation in the temperature and TKE cladding surface fields are included in the synthetic data. Spatial autocorrelation could be captured with a kriging model in the future, however, the one dimensional nature of the crud simulation code used in this work dictates that the fine scale spatial detail in the surface fields are irrelevant when computing surface-integrated crud quantities.

4.1.2 Single Pin Reconstruction

The rod surface is subdivided into CTF faces before fitting and reconstructing the synthetic data. The location and extent of the CTF faces on the rod surface may be determined from a CTF output file.

In each face the empirical quantile distributions of temperature, turbulent kinetic energy, and boundary heat flux distributions were computed. The number of quantiles used to construct the empirical quantile distributions was set at 20 and a uniform spacing of quantiles was used. Copula were fit to the synthetic data based on maximum likelihood and the Akaike information criterion (AIC) in each CTF face. The maximum likelihood estimation procedure for copula parameters is described in section 3.2.3 and the AIC may be computed from equation 3.21.

For the synthetic single pin data the hi2lo predicted fractional surface area above a saturation temperature threshold ($T_{sat} = 618.1[K]$) in each CTF face is shown in figure 4.5.



(a) CTF predicted fractional area of each CTF face above the saturation point.

(b) Hi2lo predicted fractional area of each CTF face above the saturation point.

Figure 4.5: Fraction of each CTF face above the saturation point predicted by CTF (a) and the hi2lo model (b).

Provided that crud growth exhibits a temperature thresholding behavior about the saturation point it is important to predict the fractional area of the rod surface which exists above this critical temperature. This can be performed by evaluating equation 3.29 in each CTF face. Figure 4.5 shows a substantial difference in the fractional area predicted above the saturation point in each CTF face when utilizing the hi2lo model rather than the predictions generated from a CTF computation alone. The more significant the thresholding behavior of crud growth, the more important it becomes to accurately compute areas of the rod surface in excess of the saturation point.

Figure 4.6 examines the surface frequency distributions of temperature, crud boron mass and TKE for the CTF patch denoted by the red box in figure 4.7. A marked change in behavior of the crud boron mass vs cladding surface temperature scatter plot is exhibited at approximately $619[K]$. For samples which fell at or below this temperature, little crud was grown and thus the deposited crud boron mass is small. Past this temperature threshold, there is an approximately linear relationship between the surface temperature and the crud deposition rate.

An additional feature of note in figure 4.6 is the asymmetric dispersion of the TKE vs. surface temperature scatter plot. The dependence structure exhibits tighter coupling between these two fields at lower TKE values and less correlation at higher TKE. This behavior was imposed by specifying a Clayton copula in this location on the rod surface with the copula parameters given in table 4.2. The ability of the copula fitting routines to correctly preserve this skewed dependence structure is demonstrated in the figure and could not be achieved with multivariate Gaussian models.

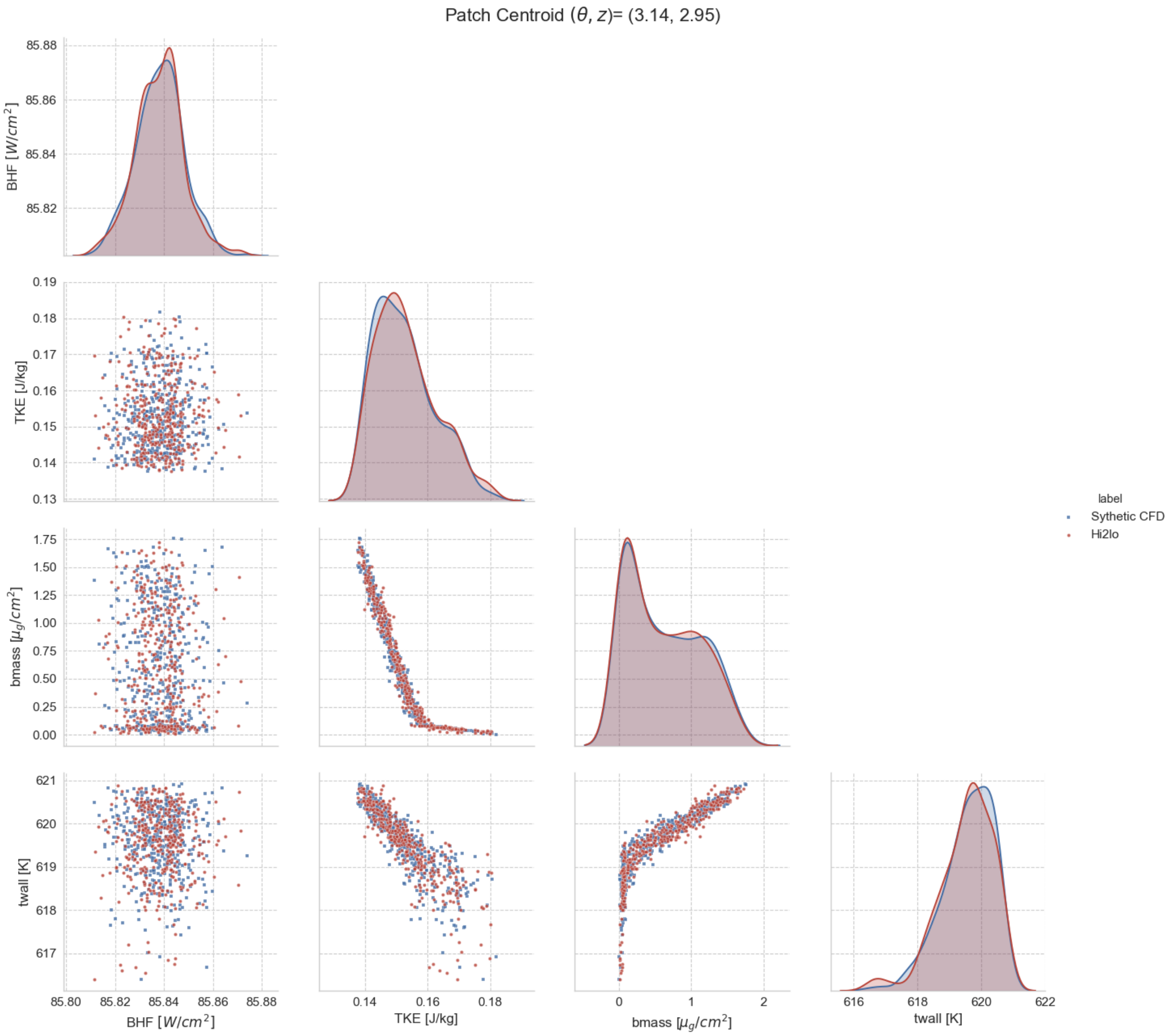
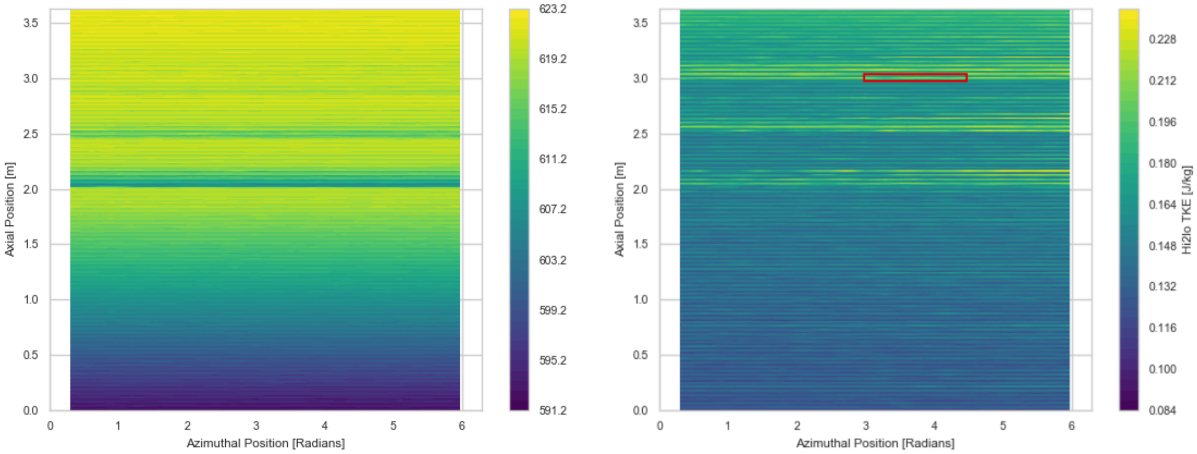


Figure 4.6: Single patch synthetic data vs hi2lo sampled data from patch centered on the rod at (3.14 $[rad]$, 2.85 $[m]$) at 300$[days]$. The location of this patch is outlined in red in figure 4.7(b).

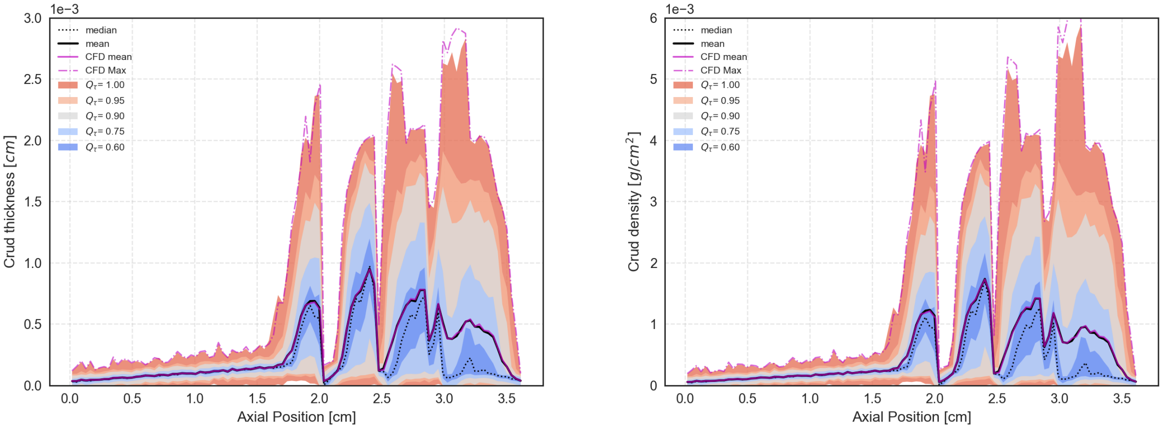


(a) Hi2lo temperature reconstruction.

(b) Hi2lo TKE reconstruction.

Figure 4.7: Hi2lo reconstruction of 2-D surface temperature and TKE fields from synthetic CFD data source. No azimuthal variation is observed in this quarter symmetric case.

After samples are independently drawn in each CTF face the temperature, TKE, and boundary heat flux samples were passed to a crud simulation packages as the cladding-side boundary conditions. The CTF bulk fluid properties were used as the coolant-side boundary conditions. The crud simulation was stepped forward for 300 days with a resample step size of $\Delta t_s = 50$ days. 400 samples per CTF face were drawn at each resampling event. The resultant crud distribution at 300 days is given in figure 4.8. Good agreement between the hi2lo predictions and the target synthetic data for the axial crud, temperature, and TKE distributions is exhibited.



(a) Hi2lo axial crud thickness.

(b) Hi2lo axial crud mass desnity.

Figure 4.8: Hi2lo axial crud results compared to synthetic CFD/crud results at 300 days simulation time. The CFD/crud result is shown in purple.

Table 4.3 summarizes the hi2lo crud predictions for the synthetic single pin data set. Good agreement is seen for the rod-integrated crud results. Again, the hi2lo model was not used in a predictive manner and this represents a best case scenario in which the copula and marginal distribution parameters were directly estimated from the known (synthetic) CFD and CTF data sets.

Table 4.3: Single pin crud totals at 300 days.

Copula, Θ_c	Crud Boron Total: C_B	Crud Mass Total: C_m
Synthetic CFD	2.79049E-4 [g]	5.34151E-1 [g]
Hi2lo Reconstruction	2.78012E-4 [g]	5.32146E-1 [g]
Rel Diff	0.374 [%]	0.377 [%]

Crud Copula Parameter Sensitivity

Here the sensitivity of the crud result to the copula parameters is investigated. Both the impact of the rank correlation coefficient, Kendall's τ, and the Archimedean copula family are investigated. The sensitivity results generated for the CTF face centered at $\{3.14[rad], 2.95[m]\}$ are shown in figure 4.9. There is noise present in the crud predictions due to the Monte Carlo integration of equation 3.3 over the patch. In this instance 2500 samples were used in the computation of the integral to reduce the magnitude of this noise. The crud is relatively insensitive to the choice of copula family, but the rank correlation coefficient is shown to have a significant influence on crud growth with an average boron deposition sensitivity of $\frac{\partial C_b}{\partial \rho_\tau}$ = -1.086e-7 $[g/cm^2/\tau]$ for this particular patch. Accurately predicting Kendall's τ provided local core conditions is important.

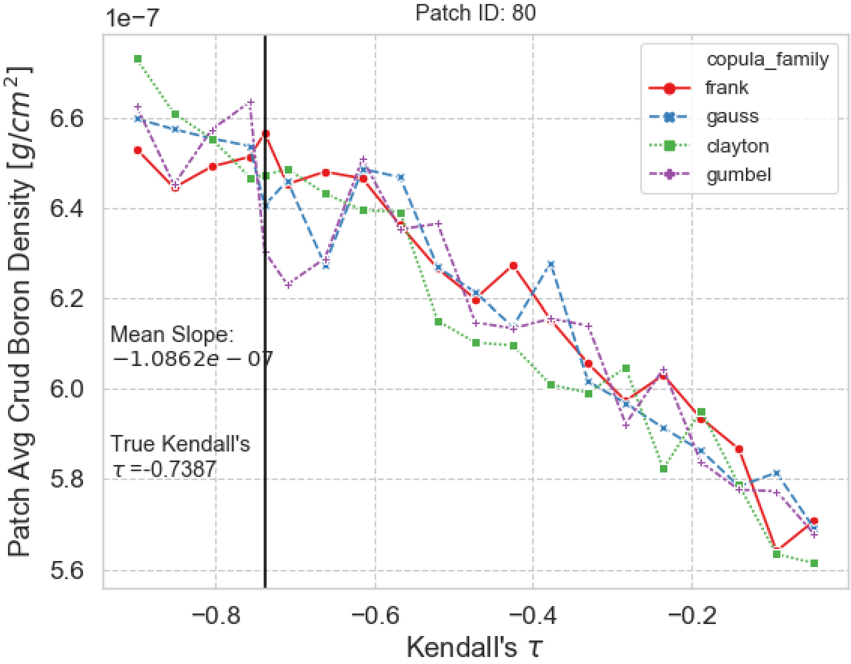
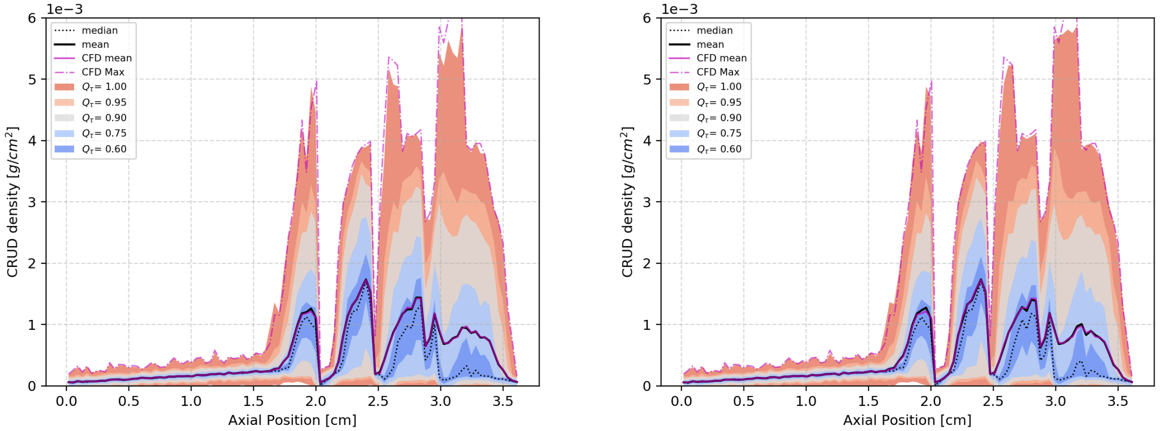


Figure 4.9: Single CTF face crud sensitivity to copula parameters.

Next, two full single pin scenarios were considered. In the first scenario, shown in figure 4.10a, the best-fit copula on each patch as determined by the AIC metric is applied on each CTF face. The second pin scenario enforces that a Gaussian copula model is used on every CTF face. The crud results from these scenarios were then compared. The data shows the choice of copula (between Gaussian, Frank, and Clayton) has a small overall impact on the total integrated rod boron mass. The total integrated crud mass and crud boron mass for these scenarios at 300 days simulation time are given in table

4.4.



(a) Best fit copula via AIC metric used in each CTF face.

(b) Gaussian copula used in each CTF face.

Figure 4.10: Influence of the choice parameters on the axial crud distribution. Synthetic CFD result shown in purple.

Table 4.4: Single pin crud totals at 300 days with different copula assumptions.

Copula, Θ_c	Crud Boron Total: C_B	Crud Mass Total: C_m
Best Fit	2.78953e-04 $[g]$	5.34015e-01 $[g]$
Gaussian	2.78301e-04 $[g]$	5.32769e-01 $[g]$
Rel Diff	0.017 [%]	0.234 [%]

The choice of the copula family, Θ_c, has a negligible impact on the integrated crud results over a pin. This result reduces the complexity of the hi2lo model by removing the need to predict the correct copula family on each CTF patch in the core. In section 5.1.1 it is shown that CFD data exhibits a complex relationship between the best fitting copula family and the axial position along the rod. This relationship proved difficult to model using standard classification techniques, though further testing with a larger quantity of training data is warranted to ascertain if the copula family describing the dependence between the temperature and TKE fields on the rod surface may be accurately predicted given local core conditions.

Crud Sample Size Study

The number of samples, N, used to estimate the integral given in equation 3.41 is a parameter set at runtime of the hi2lo method. Here, it is shown that the integrated crud variance is reduced by increasing the number of samples used per CTF face to estimate the integrated crud quantities of interest. Furthermore, section 4.1.2 demonstrates that improvements in sampling efficiency are possible by way of importance sampling.

Figure 4.11 shows the variance of the crud expectation value at 300 days of simulation time computed by the Monte Carlo approximation when using a sample sizes of 100, 400, and $800[\frac{N}{\mathrm{CTF}_{\mathrm{face}}}]$. 80 independent trials were conducted for each sample size to estimate the variance of the pin integrated crud results at 300 days.

To isolate the impact of increasing the sample size on the crud variance, a single 300 day time step was conducted without resampling the underlying density functions during this period. Importance sampling was not applied in this study.

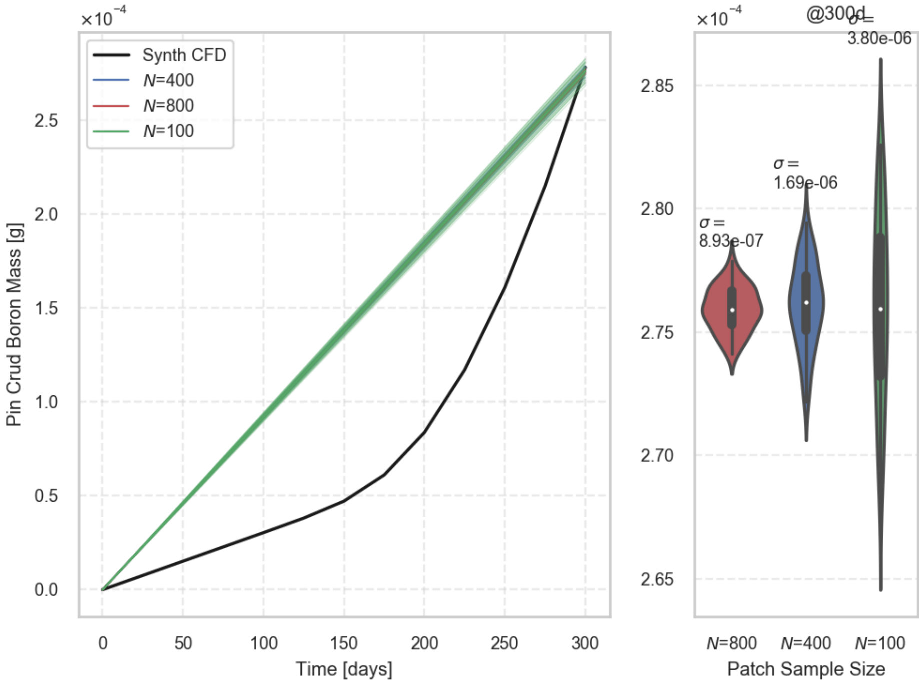


Figure 4.11: Effect of sample size on the integrated crud results.

The standard deviation of the rod integrated crud results at 300 days simulation time are summarized in table 4.5.

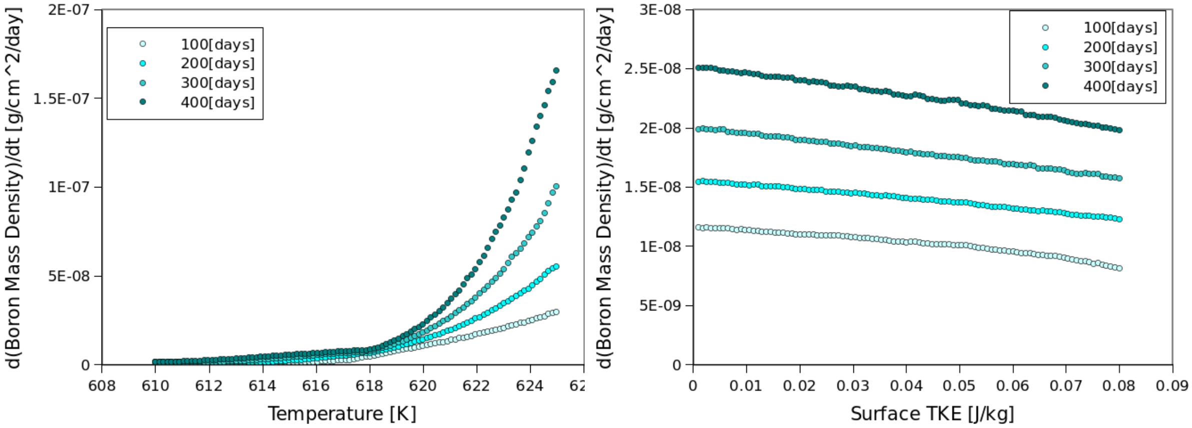
Table 4.5: Estimated sensitivity of the pin-integrated crud boron variance to the number of samples used per CTF face. Variance estimated using 80 independent trials.

N	Mean Pin Crud Boron Total [g]	Pin Integrated Crud Boron Std. Dev. [g]
100	2.78953e-04	3.80e-6
400	2.78301e-04	1.69e-6
800	2.78411e-04	8.93e-7

Importance Sampling

To obtain estimates for the efficiency gain offered by importance sampling to compute the expected crud value by equation 3.42, a singe patch was studied under synthetic TH data.

Here, the design of the sampling routines and the physics of crud growth is intertwined. To compute the integral 3.3 efficiently it is favorable to sample the TH distribution in regions which result in relatively large amounts of crud growth. To guide the design of the importance distributions the response surface of the crud simulation code is presented in figures 4.12 to 4.14. Larger surface temperatures result in a higher crud growth rate. Larger local TKE results in smaller crud growth rates due to the effects of erosion. Note the relatively small influence of the boundary heat flux on the crud growth rate. This is an important observation used to justify simplifications made in the joint density model previously provided in equation 3.5.

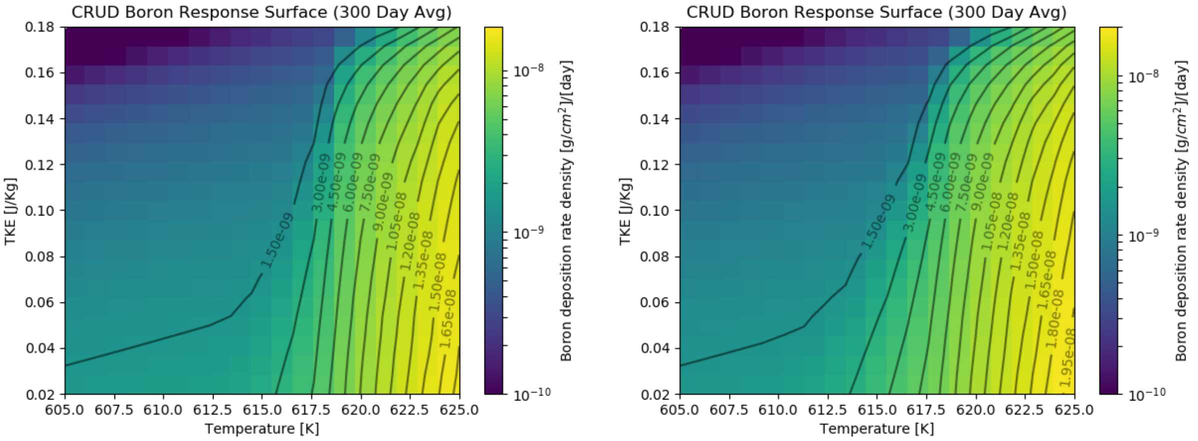


(a) Crud boron deposition sensitivity to temperature with TKE held fixed at 0.05 [J/kg].

(b) Crud boron deposition sensitivity to TKE with temperature held fixed at 620 [K].

Figure 4.12: Crud deposition rate sensitivity to varying temperature and TKE at different times.

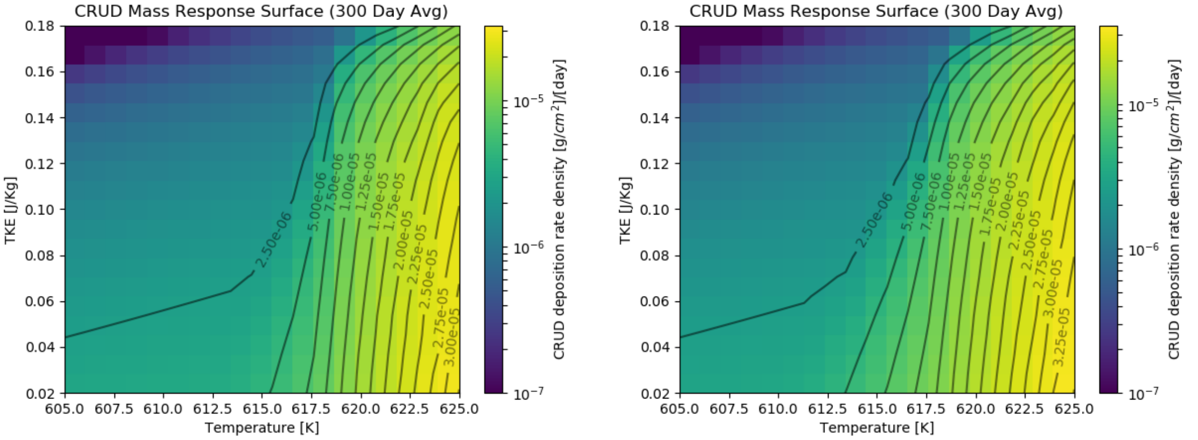
In figure 4.12 the crud growth rate response is depicted at 100 day increments. At temperatures exceeding the saturation temperature ($\approx 618[K]$) a marked change in crud growth rates is exhibited. When the saturation temperature is exceeded on the rod surface there is a rise in crud deposition rates and the rate at which boron is precipitated inside the crud layer.



(a) Crud boron deposition sensitivity with $q'' = 80[W/cm^2]$.

(b) Crud boron deposition sensitivity $q'' = 120[W/cm^2]$.

Figure 4.13: Crud boron response surface to varying temperature and TKE. The crud boron deposition rate is relatively insensitive to the boundary heat flux.



(a) Crud mass deposition sensitivity $q'' = 80[W/cm^2]$.

(b) Crud mass deposition sensitivity $q'' = [120W/cm^2]$.

Figure 4.14: Crud mass response surface to varying temperature and TKE. The crud mass deposition rate is relatively insensitive to the boundary heat flux.

Though an optimal importance distribution, $\tilde{h}^*$, can be found by equation 4.5,

$$\tilde{h}^* = \text{argmin}_{\tilde{h}} \text{Var} \left[\frac{\mathcal{G}(x) h(x)}{\tilde{h}(x)} \right] \tag{4.5}$$

the requisite minimization problem is not solved in this work and is left as an avenue for future investigation. It may be shown that the optimal importance distribution follows the form: $\tilde{h}^* \propto |\mathcal{G}(x)|h(x)$, [53], [54].

Although the theoretically optimal importance distribution is not achieved in this work, a locally adaptive importance function was adopted based on a distribution mixing approach. Through the mixture formulation, the importance distributions can be made to depend on the temperature and TKE marginal distributions in a particular CTF face. The temperature and TKE distributions in each CTF face are mixed with a beta distribution whose parameters are set at runtime of the hi2lo tool.

The importance mixture quantile functions are defined in equations 4.6 and 4.7.

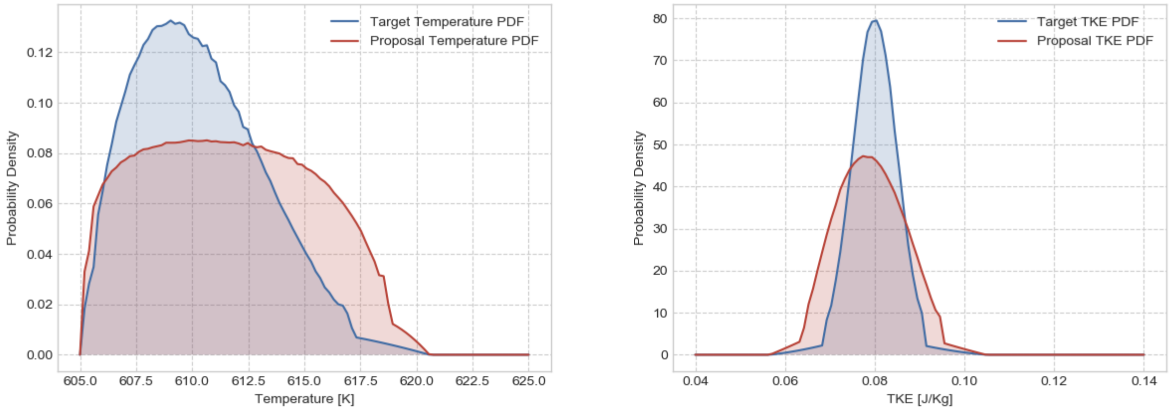
$$\tilde{Q}_{j,k} = \lambda_{0,k}\hat{Q}_{j,k} + \lambda_{1,k}Q_{\beta_k}(k;\vartheta_k), \tag{4.6}$$

$$\tilde{Q}_{j,T} = \lambda_{0,T}\hat{Q}_{j,T} + \lambda_{1,T}Q_{\beta_T}(T;\vartheta_T); \tag{4.7}$$

$$\sum_i \lambda_i = 1, \quad \tilde{F}_{j,T} = \tilde{Q}_{j,T}^{-1}, \quad \tilde{F}_{j,k} = \tilde{Q}_{j,k}^{-1} \tag{4.8}$$

Where $\tilde{Q}_{j,T}$ is the quantile function for the proposal temperature distribution in the j^{th} CTF face. λ_i are user set mixture weights. Standard Monte Carlo sampling can be recovered by setting $\lambda_{0,k} = 1,\ \lambda_{0,T} = 1$ and $\lambda_{1,k} = 0,\ \lambda_{1,T} = 0$.

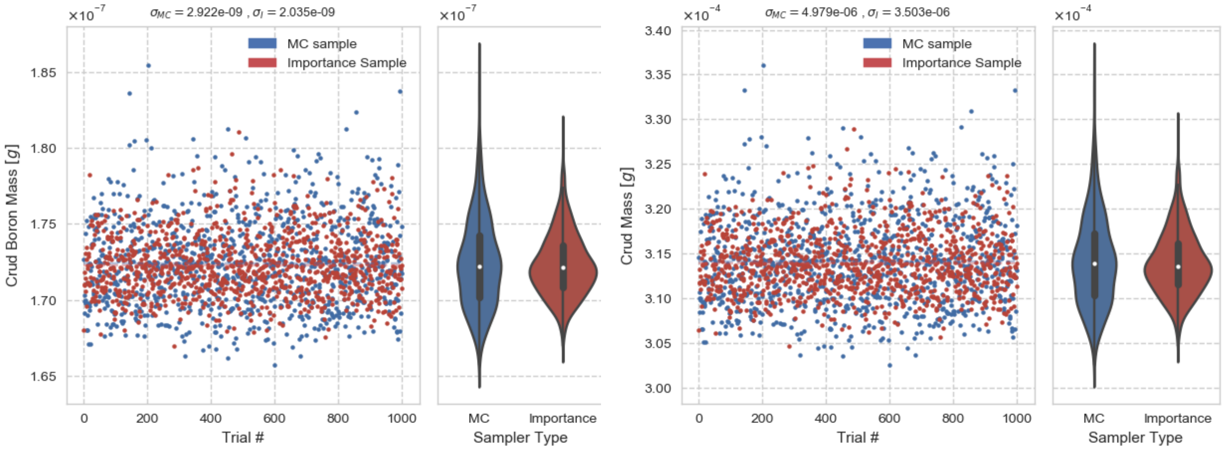
Beta distributions, with quantile functions Q_{β_k} and Q_{β_T}, with proscribed parameters, $\{\vartheta_k, \vartheta_T\}$, are used in mixture with the original target temperature and TKE density functions to produce a proposal density distribution for each patch. To provide additional flexibility in the design of proposal density, the mixture weights may be adjusted. By suitably tuning the parameters of the beta distributions and mixture weights, one can target the hot locations of the rod which occur in coincidence with low TKE. Mixture settings denoted in table 3.1 were adopted for this work. Shown in figure 4.15 with proper tuning of the mixture distribution parameters, the sampling distribution may be skewed towards higher temperatures and lower TKE.



(a) Temperature distributions.

(b) TKE distributions.

Figure 4.15: Proposal (red) vs. original (blue) marginal distributions for the (a) surface temperature and (b) TKE. Generated using importance distribution parameters: $\vartheta_T = \{1, 0.9\}$, $\vartheta_k = \{1.1, 1.2\}$, $\lambda_{0,T} = 0.6,\ \lambda_{1,T} = 0.4$ and $\lambda_{0,k} = 0.7,\ \lambda_{1,k} = 0.3$.



(a) Crud boron mass deposition.

(b) Crud mass deposition.

Figure 4.16: Importance sampling trial results on a single CTF face for (a) crud boron mass deposition and (b) Crud mass deposition at 300 days. Red denotes importance samples and blue denotes standard Monte Carlo samples. The sample population variance is shown above the figure for each case and the full sample distributions are given in the margins.

In figure 4.17 the relative importance weight is denoted by the size of each point in the scatter plot. Samples which have a small ratio $(h_i/\tilde{h}_i)$ appear as small points.

The sample weight is analogous to the rod surface area occupied by the sample. In comparison, figure 4.18 shows the same patch using a standard Monte Carlo sampling where each sample has the same weight. The number of samples drawn in the upper tail of the temperature distribution is greater when importance sampling is applied, though these samples carry expectedly small sample weights.

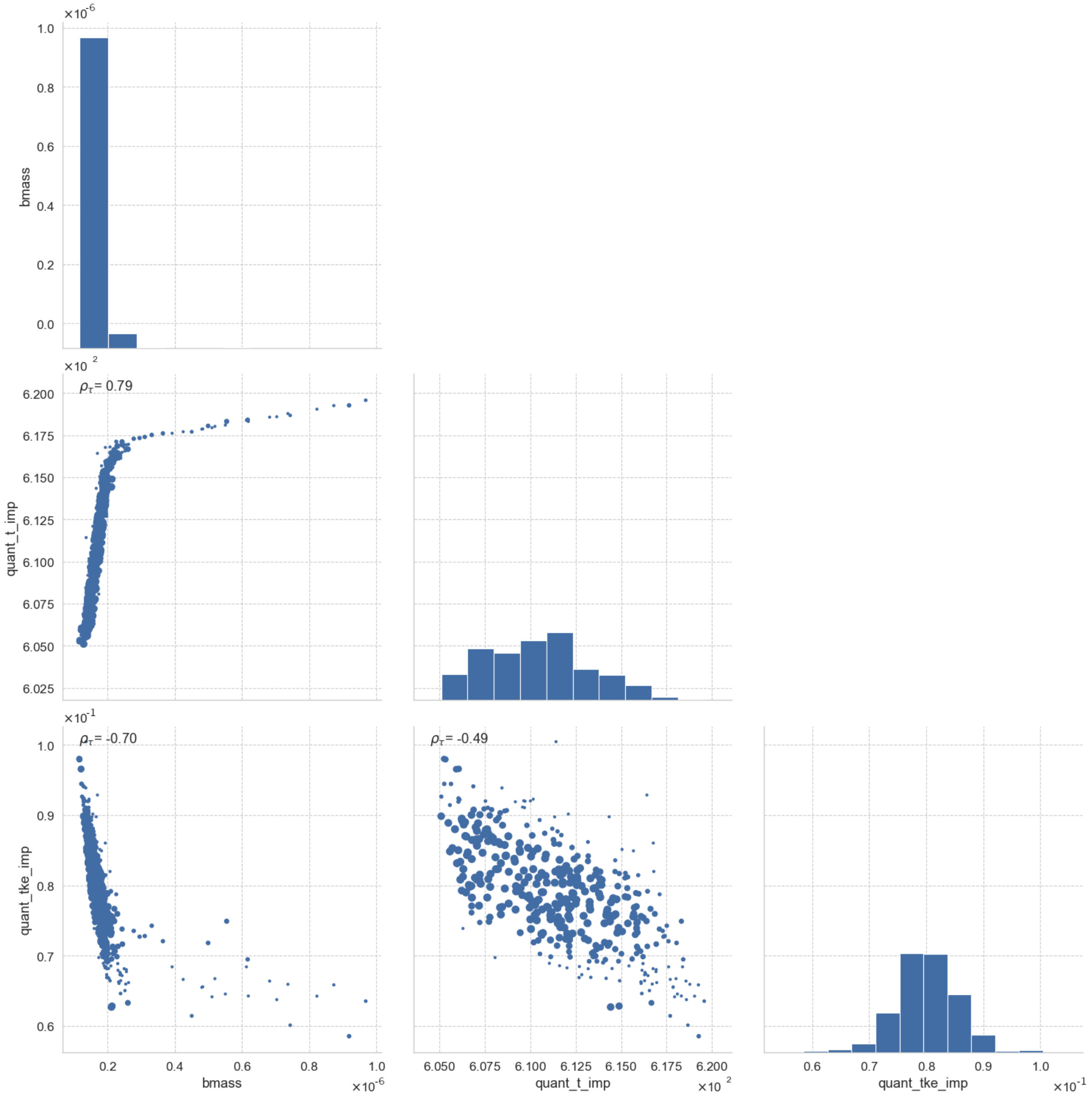


Figure 4.17: Importance sampled single patch crud result. Key: **quant_t_imp**: Importance samples from quantile-reconstructed patch temperature $[K]$ distribution with $N_{Q_T} = 20$, **quant_tke_imp**: Importance samples from quantile-reconstructed patch TKE $[J/kg]$ distribution with $N_{Q_k} = 20$, **bmass**: Resultant crud boron mass density samples in $[g/cm^2]$. Relative importance weights are denoted by the point size.

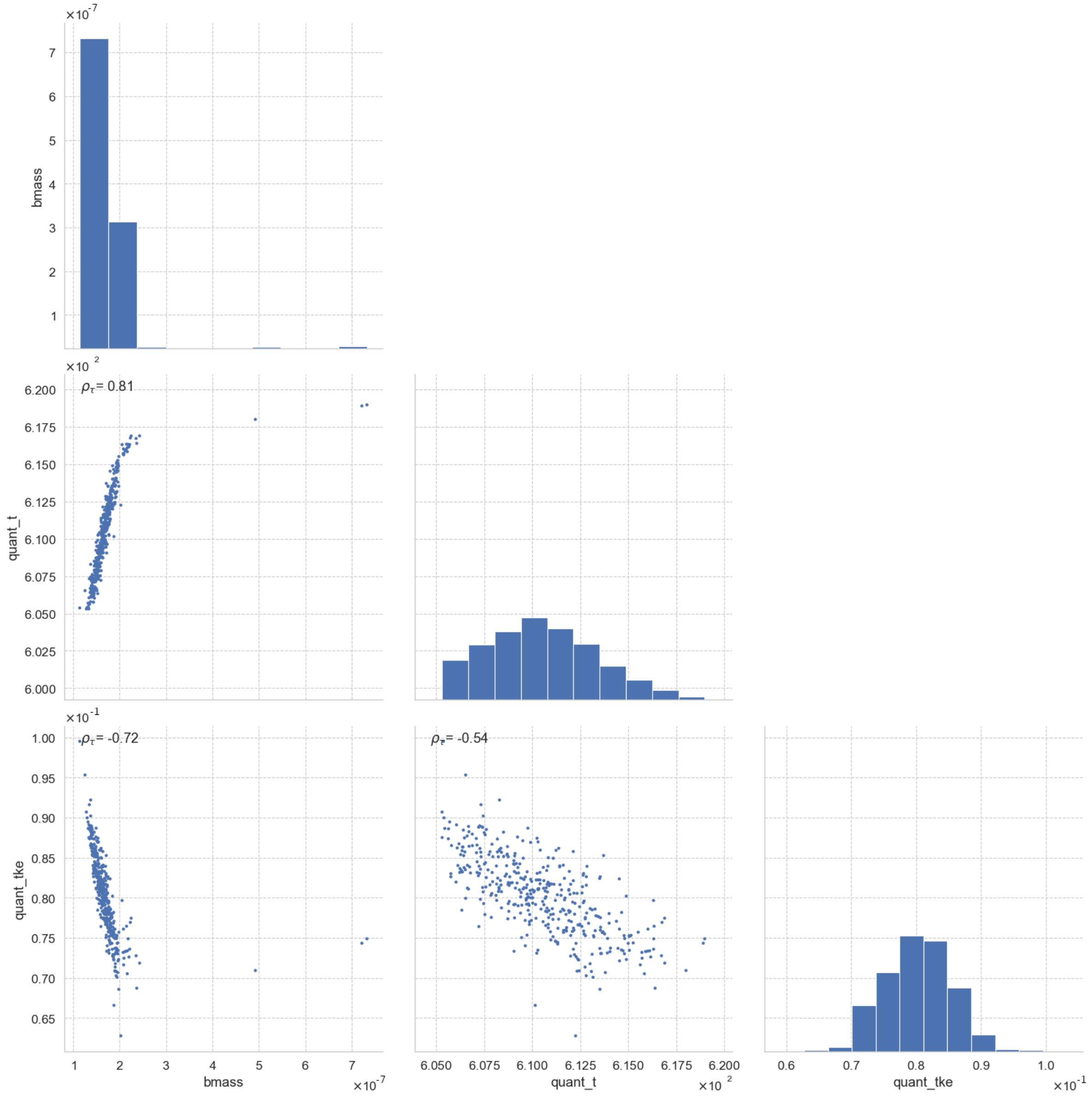


Figure 4.18: Standard Monte Carlo sampled patch crud result. Key: **quant_t**: Samples from quantile-reconstructed patch temperature $[K]$ distribution with $N_{Q_T} = 20$, **quant_tke**: Samples from quantile-reconstructed patch TKE $[J/kg]$ distribution with $N_{Q_k} = 20$, **bmass**: Resultant crud boron mass density samples in $[g/cm^2]$.

The importance sampling efficiency was estimated by computing the variance ratio: σ^2_{MC}/σ^2_I. The variance of the patch-integrated crud result for the Monte Carlo and importance sampling schemes are provided in figure 4.16. The variance estimates were computed by running 1000 independent trials in which crud was grown on the patch for

300 days. A total of 100 samples per patch per trial were used. For the case studied, the application of importance sampling reduced the crud mass and boron mass sample variance by a factor of 2.02. $\frac{\sigma_{MC}^2}{\sigma_I^2} \approx (4.979 \times 10^{-6})^2/(3.503 \times 10^{-6})^2 \approx 2.02$. The mean crud predictions did not significantly deviate between two sampling schemes indicating that importance sampling does not introduce any bias in the evaluation of the integral in equation 3.3.

The improvement in performance afforded by importance sampling may be attributed to expending a larger proportion of the total available samples in the upper tail of the temperature distribution as this is a region which strongly contributes to crud growth. Some samples are necessarily expended in cold regions of the rod surface but occur with less frequency when compared to a standard Monte Carlo sampling routine and the samples are appropriately weighted to avoid biasing the integrated result.

4.1.3 Single Pin with Time Stepping

Stepping the crud simulation forward under the application of hi2lo supplied boundary conditions demands careful treatment of hot spot stationarity assumptions. The time evolution of the crud simulation on the rod surface is strongly influenced by choices made both in the number of resampling steps taken as well as tunable constants which govern the sample remapping procedure and surface temperature mixing.

Spatial Remapping with Time Stepping

The influence of hot-spot stationarity assumptions on the overall integrated crud mass on the rod as a function of time can be seen in figures 4.19 and 4.20. When the surface temperature is allowed to randomly mix on each resampling event in each CTF face, the influence of the hot spots are smeared over the surface of the rod which leads to an overall under prediction in the total integrated crud mass. Reordering the samples in each CTF face by their temperature improves the ability of the hi2lo model to preserve the impact of stationary hot and cold spots on the rod surface. Good agreement with the original coupled synthetic CFD-Crud simulation data was achieved by tuning the constants introduced in equation 3.47 to values of $w_T = 0.4$, and of $w_k = 0.6$. This weighting seeks to preserve a heuristic thermal-hydraulic metric on the rod surface where the metric may be interpreted as some linear combination of cladding surface temperature and near-wall TKE.

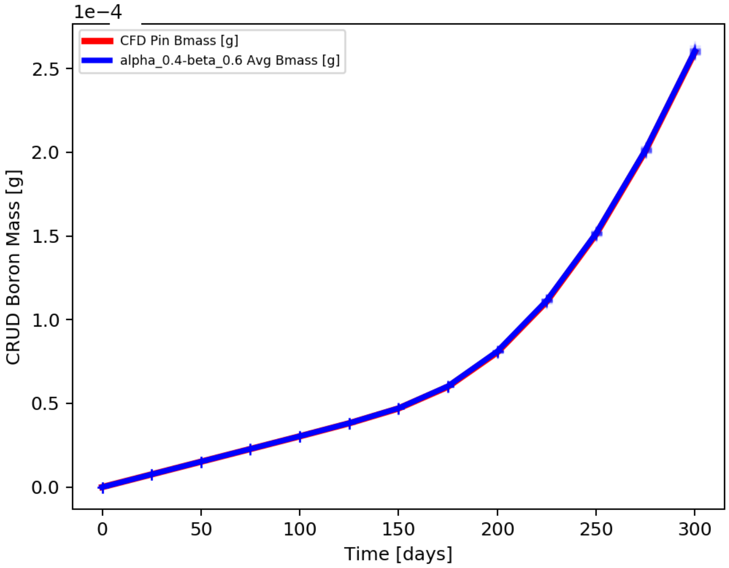


Figure 4.19: Total integrated crud boron mass vs. time using approximately optimal remapping weights ($w_T = 0.4, w_k = 0.6, w_{q''} = 0.0$).

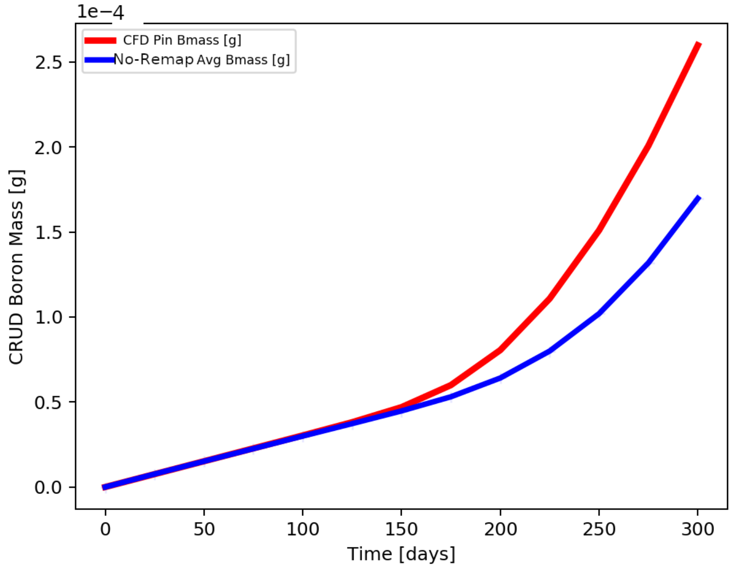
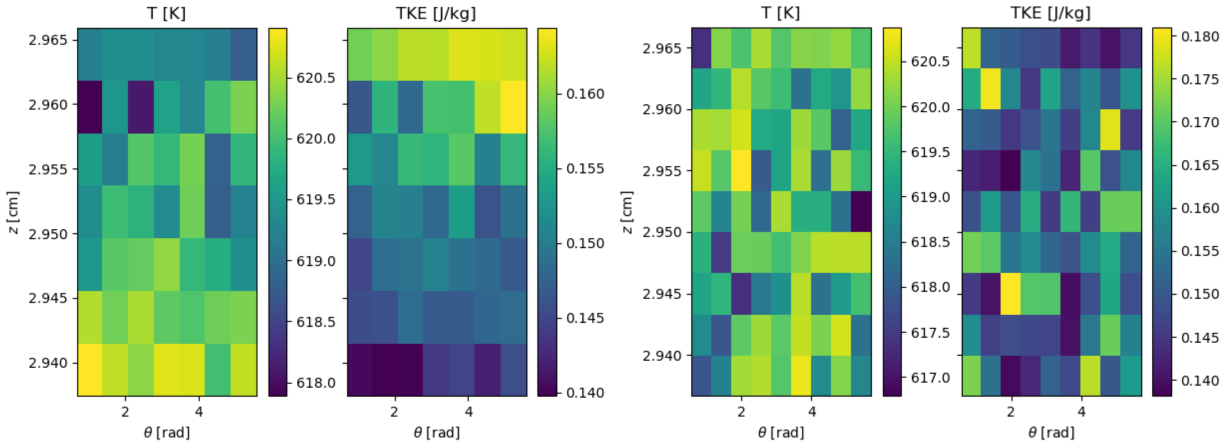


Figure 4.20: Total integrated crud boron mass vs. time without remapping samples.

A spatial representation of the samples pre and post-remapping are shown in figure 4.21. A visual representation of the remapping strategy presented is presented in figure 3.20. While the spatial distribution of the temperature and TKE fields are distinctly different after the remapping procedure is applied, the joint density distributions formed by the sample population over the patch are identical.



(a) Remapped surface samples.

(b) Non-remapped surface samples

Figure 4.21: Re-mapped (a) and non remapped (b) temperature and TKE surface samples for a single CTF face.

4.1.4 Resample Frequency

The frequency at which the distribution functions are sampled from on each CTF face influences the variance in the predicted integrated crud results. To investigate the behavior of the varience of the crud results as a function of resampling frequency, a parameter sweep was conducted in which crud was grown on the same pin with resampling time step sizes of 50, 100, and 300 $[days]$. 50 independent trials were conducted for each step size. Shown in figure 4.22 a smaller resampling steps size, Δt_s, results in a reduction in the variance of the rod integrated crud estimates at 300 days of simulation time. Additionally, the sampling induced rod integrated crud mass uncertainties were shown to be approximately distributed according to a normal distribution. It is also important to note that the variance of the rod integrated crud results increases as a function of time.

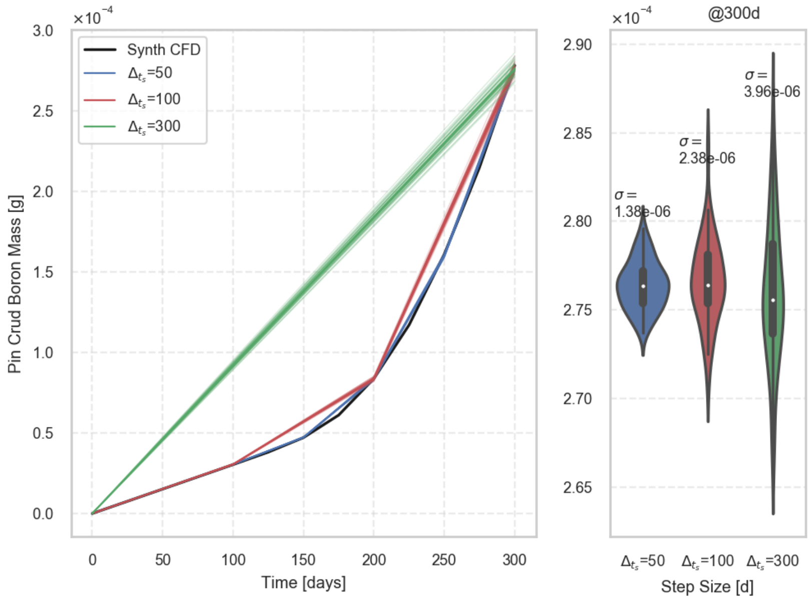


Figure 4.22: Influence of the resample frequency on the predicted rod integrated crud variance. Crud variance estimates at 300 $[days]$ are given in the right hand side of the figure from 50 independent trials for each of the step size cases.

Performing a larger number of resampling events per VERA state results in reduced variance at little additional computational effort. This is in part due to the minimal computational requirements of sampling the joint temperature and TKE distribution on each patch. Drawing samples from a bivariate copula density model is straight forward and can be done in parallel since each patch is treated as an independent sampling zone in this hi2lo approach. The crud computation, by comparison, is more expensive. Increasing the resampling frequency does not increase the total number of samples used per pin per time step, rather, this process only increases the number of (resample) steps per VERA state point. The reduction in variance stems from an improved sample density throughout time of the underlying random field. In time, the underlaying random field is fixed throughout a VERA state point. Repeatedly drawing samples from this field at small time steps rather than sampling the random field only once at the beginning of the VERA state point vastly increases the number of samples used to perform the time integration of the crud result on each CTF face.

4.2 Section Takeaways

- Good rod integrated crud agreement between the synthetic source data and the predicted results was observed. The rod integrated results were summarized in table 4.3. Fitting copula and marginal distributions to the sample data, resampling from these models, and then growing curd using these samples reproduces the correct total amount of crud at each time step and correctly reproduces the axial distribution of crud as shown in figure 4.8.

- Increasing the number of samples drawn per CTF patch decreases variance in total crud mass and total precipitated boron estimates. The number of samples used is an adjustable runtime parameter and can be increased depending on the available computational resources and accuracy desired.

- Increasing the number of resampling steps per VERA state point reduces variance in the final integrated crud results.

- After drawing samples from the joint density distribution a reordering of the samples on the rod surface is necessary to preserve hot spot stationarity. Demonstrated in figure 4.19 and 4.20, sample remapping with weights of $w_T = 0.4,\ w_k = 0.6$ was performed to achieve an optimal time marching sampling strategy.

- Importance sampling was shown to reduce the variance in the integrated crud results.

- A synthetic data generation tool allows absolute control over the properties of joint distribution of TH boundary conditions which are fed into the crud simulation code. Since the synthetic data has known statistical properties, this data serves as an important data source for benchmarking and validation operations.

5 | Model Performance Under a CFD Data Source

For deployment as an in-line statistically based downscaling tool which sits between a subchannel code and a crud simulation code in a core simulator such as VERA the model is required to perform the hi2lo mapping for all pins in the core at any operating condition. In other words, the model must be evaluable at any local core conditions typical of an operating PWR. Since the training data set cannot contain all possible pin geometries, loading configurations and operating conditions due to computational expense, the model produces a prediction for the copula and marginal distribution parameters between known states.

One might envision a table-lookup approach where high fidelity CFD flow field maps are precomputed and stored for a wide array of flow and power conditions. A nearest neighbor interpolation scheme could then be applied to extract the best-matching CFD map provided some local core state by VERA. This is not tractable since the number of CFD computations to build the data base would be prohibitively large. Instead of storing spatial CFD hi2lo maps, CFD data is distilled into a set of statistics tabulated as a function of local core state.

In this chapter the hi2lo methodology introduced in this work is exercised against a small CFD data set derived from a 5x5 fuel assembly operating at realistic PWR conditions. A leave-one-out cross validation strategy is used to assess the predictive performance of the model.

5.1 CFD Data Source

For the generation of high fidelity CFD data sets the Westinghouse 5x5 test stand shown in figures 5.1 and 5.2 was used to prepare the CAD geometry. The CFD mesh consisted of approximately 25 million cells and 1e5 surface elements per pin. A matching CTF input deck for the 5x5 assembly was also constructed with 100 axial zones. The CTF and CFD codes were then executed for a variety of flow conditions and power levels. StarCCM+ was utilized for the CFD simulations in this work.

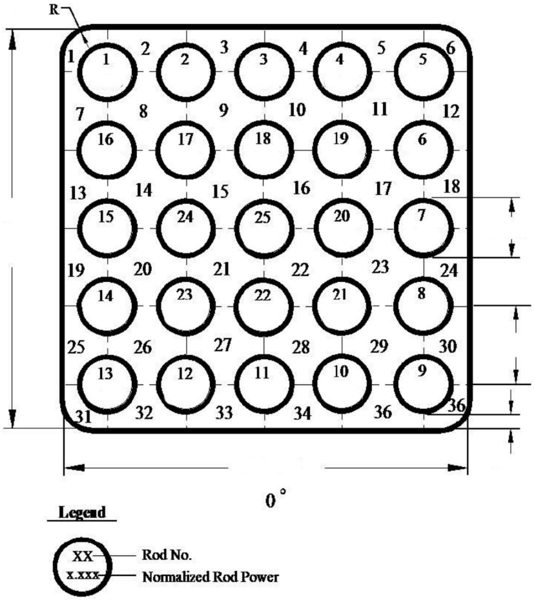


Figure 5.1: Top down view of 5x5 pin Westinghouse facility. Assembly dimensions and pin powers redacted.

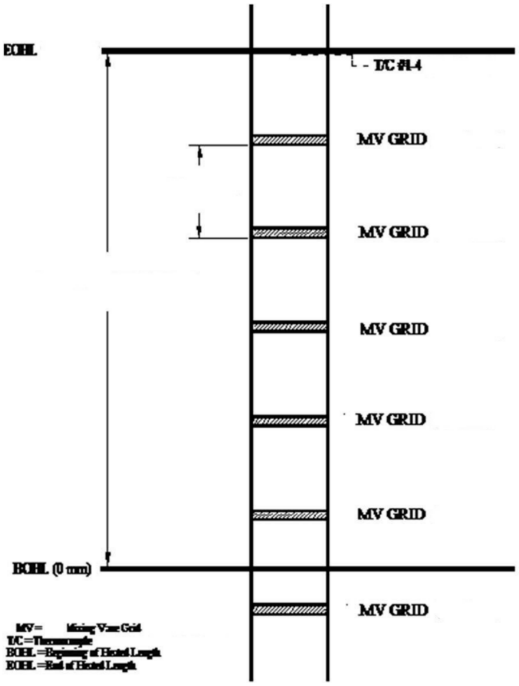


Figure 5.2: Side view of 5x5 pin Westinghouse facility. Pin dimensions redacted.

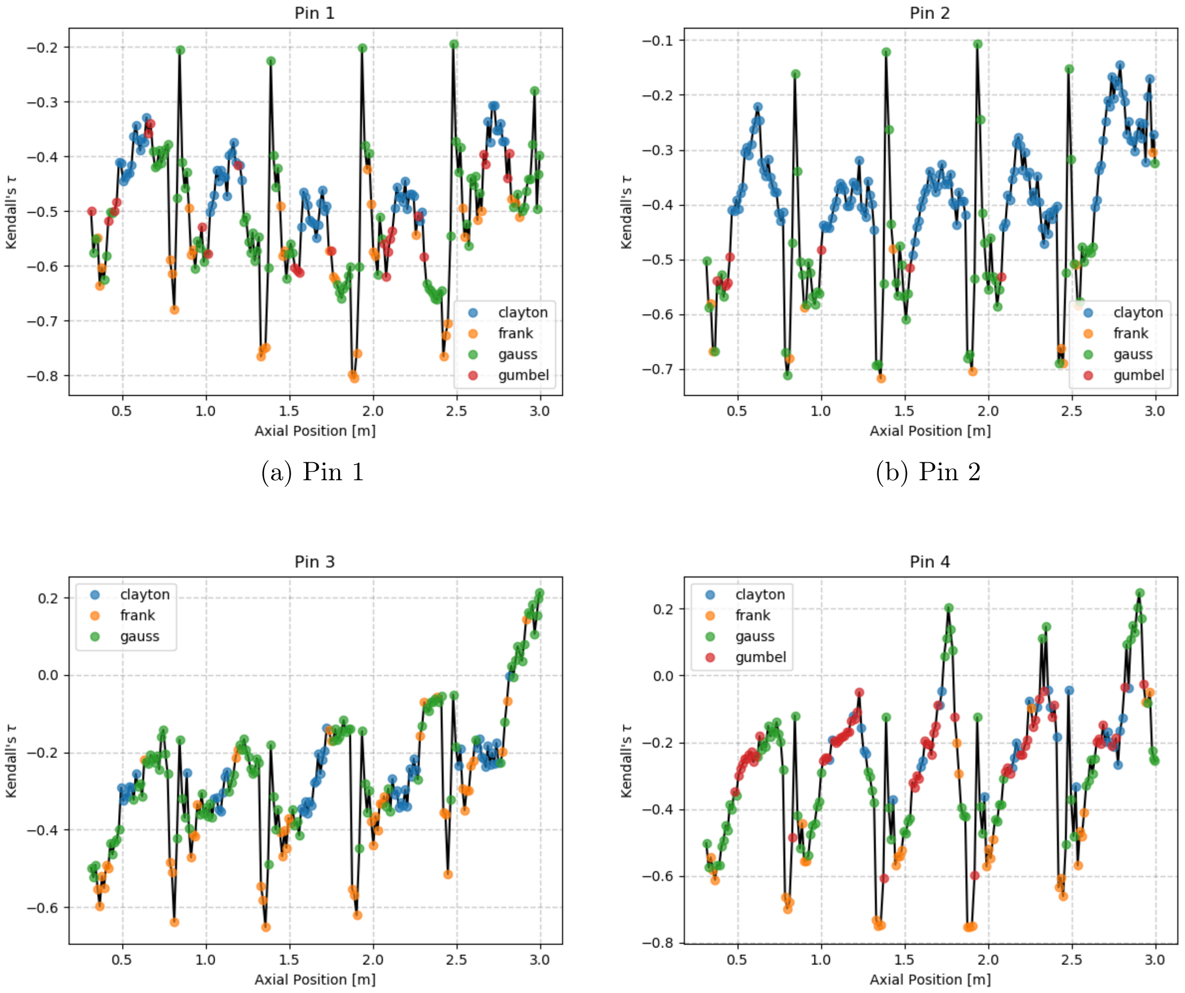
For this rod configuration, the axial pin power, total power and CFD simulation rod surface temperature distributions are available in external references. This information is purposely withheld from this document to protect the intellectual property of the Westinghouse electric company.

5.1.1 Preprocessing

Preprocessing requires paired CFD and CTF results for a given pin generated with consistent boundary conditions between the codes. This requires consistent geometry, inlet, outlet and power distributions between the codes.

The cladding surface temperature and near-wall TKE CFD fields are spatially aggregated onto the CTF face centers. The aggregation requires that the location and extent of each CTF face is known. These CTF face attributes are accessible from a CTF output file. In the aggregation procedure spatial information is discarded within each CTF patch as the spatial fields are agglomerated into sample distributions. In each CTF face each of the features given in table 5.1 are computed from the available CTF (or VERA) results. The aggregated CFD data fields are then associated with their corresponding feature set. In each face, the aggregated CFD field distributions are subtracted from the mean CTF predictions and the resultant (CFD-CTF) residual distributions are stored in a HDF5 table along with the associated predictive variables.

Next, correlation statistics are computed from the residual distributions. Copula fitting by the maximum likelihood method with AIC model selection is carried out on each CTF face. Additionally, the empirical Kendall's τ rank correlation coefficient is computed from the raw CFD data on each CTF face. Figure 5.3 shows the copula parameters estimated from the raw CFD data on each pin as a function of axial position for the first 4 pins in the 5x5 CFD model. There is a marked change in behavior of the copula between the pins. This was an unexpected find since the flow patterns were speculated to be reasonably consistent from pin to pin. Also, the influence of spacer grids on the correlation coefficient between the temperature and TKE fields is visible. Across spacer grids the correlation coefficient sharply falls indicating a tighter coupling between the TKE and temperature surface fields as the flow necks down when entering a grid. This is followed by a sharp change in Kendall's τ towards zero indicating the temperature and TKE surface fields become less correlated immediately following the mixing vanes. This change in correlation behavior is posited to be due to turbulent mixing effects. The computed copula parameters are also stored alongside the raw temperature, TKE, and boundary heat flux aggregated residual distribution data in the HDF5 table.



(a) Pin 1

(b) Pin 2

(c) Pin 3

(d) Pin 4

Figure 5.3: Best fitting copula determined by AIC model selection as a function of axial rod position.

After pre-processing, the HDF5 table includes a list of predictive scalar values, which are shown in table 5.1, and a list of associated response variables comprised of the copula parameters and residual sample distribution for $\{T, k, q''\}$ on each CTF face.

5.1.2 Feature Engineering

The objective of feature engineering is to select a predictive variable set that describes the behavior of the conditional quantiles and copula everywhere in the assembly.

Table 5.1: Features included in the gradient boosted models as exogenous variables.

Sym	Label	Feature	Unit
T	ctf_twall_avg	CTF Face surface temperature	$[K]$
R_T	ctf_twall_range	Surface temperature range in 4 adjacent faces	$[K]$
q''	ctf_bhf_avg	Local CTF face heat flux	$[W/m^2]$
$R_{q''}$	ctf_bhf_range	Heat flux range in 4 adjacent faces	$[W/m^2]$
u_z	w_bulk	CTF subchannel bulk Z Velocity	$[m/s]$
k	ctf_tke_avg	Local CTF face near wall TKE	$[J/kg]$
R_k	ctf_tke_range	CTF TKE range in 4 adjacent faces	$[J/kg]$
z	z	Global axial position	$[m]$
δz_g	dz_grid	Position relative to nearest spacer grid	$[m]$
N_g	n_upsteam_grid	Nearest upstream spacer grid ID	[]
T_∞	t_bulk	Subchannel bulk temperature	$[K]$

The predictive variables given in table 5.1 were selected based on two criteria: Availability and orthogonality. In order to evaluate the trained machine learning model at a TH state point each conditioning variable should be made available by VERA or must be computable from CTF results and supplied as input to the trained machine learning models. The exogenous variable set given in table 5.1 comprise the local core conditions at any given CTF face. The machine learning model uses the local core conditions as the exogenous feature set, thus these features must be supplied to the fitted gradient boosted regressors at runtime in order to evaluate the model.

At this juncture, the availability criteria precludes using some geometric information such as the orientation of a given spacer grid since it is not possible to extract or infer this information from the CTF output. Including additional geometric information into the exogenous variable set could potentially increase the ability of the machine learning models to distinguish unique CTF faces in the core though testing of this hypothesis is left to a future investigation. Additional software infrastructure is required to include and extract additional features from the CTF or VERA output files.

It is not useful to include features which are co-linear into the explanatory feature set. The bulk fluid density was not included in the predictive variable set as it strongly depends on the local temperature. Likewise the local static pressure was not used as a predictive variable since this would be approximately one-to-one with the axial position. The exclusion of this TH information is primarily done for computational saving when training the boosted models since, as opposed to other machine learning algorithms and statistical inference techniques, gradient boosting is robust to collinearity of features in the input space.

In the case of gradient boosting the inclusion of nuisance or collinear exogenous variables in the model will not necessarily reduce the model's ability to generalize to unseen data, only hamper computational efficiency. The resulting feature importance plot shown in figure 5.4 suggests that the relative axial position within a span does not provide predictive power since this information is redundant provided the absolute axial position and the nearest upstream spacer grid are included in the feature set.

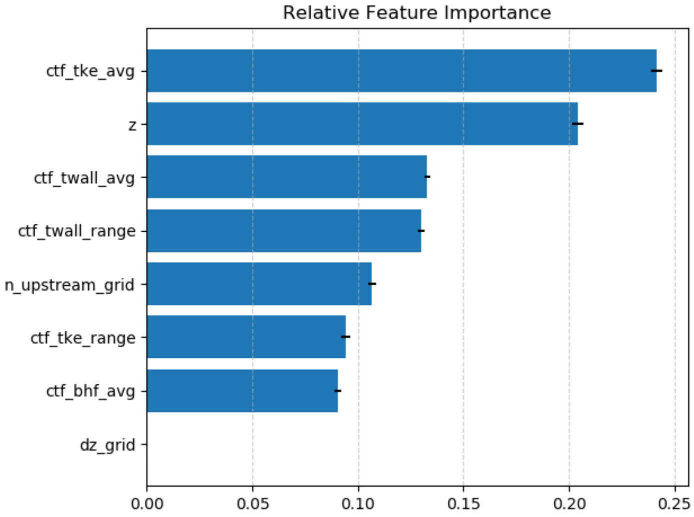


Figure 5.4: Relative feature importances on Kendall's τ.

Since the boosted regression (and classification) models are insensitive to multicollinearity in the feature space, the application of principal component analysis to the training data set was not pursued.

5.1.3 Cross Validation

An estimate of the per pin crud prediction errors incurred when evaluating the trained models at unknown CFD states were made by performing a leave one out (LOO) cross validation study. Cross validation is used to estimate how well the machine learning models employed in this work generalized to previously unseen local core conditions; i.e. core conditions that are not included in the training data set.

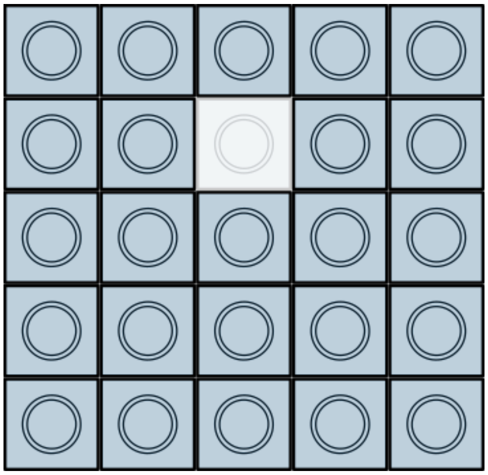


Figure 5.5: Example pin layout for leave-one-out cross validation procedure. The gradient boosted models are trained on CFD and CTF data extracted from the blue pins. Crud predictions are made on the missing pin.

The LOO cross validation procedure is depicted in figure 5.5. In this procedure a single CFD-CTF pin pair is removed from the database and then the model is trained on remaining data. Following data culling and training, the machine learning model is evaluated and crud predictions are made at the missing pin's TH conditions. The predicted crud results are compared against crud results generated using the original CFD data for the missing pin. This process is repeated for each pin in the 5x5 assembly.

This cross validation technique ascertains crud prediction errors within the TH envelope enclosed by the original full 25 pin training set. The resulting crud prediction error estimates cannot be extrapolated to core conditions that lay outside of the thermal hydraulic envelope formed by the training set. For a robust crud prediction error analysis, a much larger training data set is required which would essentially span all possible TH conditions encountered in an operational PWR. This will require large scale CFD runs and is left as an avenue for future uncertainty quantification work. A larger training set would also increase the viability of other multi-fold cross validation techniques which require permutations of stratified chunks to be excised from the training pool in their application. This would involve removing multiple pins from the training set.

5.1.4 Quantile Regressors

A principal goal of the machine learning model is to predict the conditional quantiles of the temperature and TKE distributions as a function of local core conditions. In this light, the trained quantile regression models are compared against the left-out CFD data set on each pin. The accuracy of both the TKE and temperature quantile regressors is assessed using both quantile-quantile plots and quantile vs axial rod position comparisons.

Quantile-quantile (Q-Q) plots of the temperature and TKE residual distributions are used to elucidate bias introduced by the machine learning model in the conditional quantiles at a variety of axial positions and local core conditions. Estimated quantiles are obtained for the left-out pin by evaluating the trained reduced LOO model and are compared to the expected CFD result. A subset of the TKE residual quantile regression results are given in figures 5.6 to 5.8. A complete set of quantile regression results are provide in appendix A. The Q-Q plots summarize the biases in the conditional quantile distributions when compared to the target golden standard CFD data. The maximum and average Kolmogorov–Smirnov (KS) statistic is provided in the Q-Q figures for each pin. The KS statistic is given by equation 5.1.

$$KS = \sup(\{\hat{F}(q_\tau) - F(q_\tau)\}) \tag{5.1}$$

Where $\sup(\cdot)$ is the supremum of the set of distances between the predicted and empirical cumulative densities. The cumulative densities are supported at the specified quantile levels: $\{\tau\} = \{0.000, 0.0526, 0.1052, ...1.000\}$ since the number of quantiles used in the reconstruction of the marginal temperature and TKE distributions was set to be 20 and were evenly spaced. The KS statistic was computed at each axial level on the CTF grid.

A two sample KS test was performed on the temperature and TKE distribution reconstructions from predicted quantiles on each CTF axial edit for every pin in the assembly. The null hypothesis is that the predicted and empirical (CFD derived) distributions are the same on a given CTF axial zone. To reject the null hypothesis the KS distance must satisfy the inequality 5.2.

$$KS_D^* > c(\alpha)\sqrt{\frac{n+m}{nm}} \tag{5.2}$$

Where $n = 20$ in this case since the predicted CDFs are supported at 20 locations. The number of CFD surface samples available to construct the empirical distribution, m, on each axial edit was approximately 800, though this varied slightly from zone to zone and is dependent on the CFD mesh density on the rod surface. In general $c(\alpha) = \sqrt{-\frac{1}{2}\text{ln}\alpha}$, therefore at $\alpha = 0.1$, $KS_D^* \approx 0.243$. A summary of the KS distances and test results are provided in table 5.3 for the temperature quantiles and in table 5.2 for the predicted TKE quantiles.

By inspecting the KS test results presented in tables 5.2 and 5.3 it can be concluded that the prediction of the conditional temperature distribution on each CTF axial edit was far more difficult than predicting the conditional TKE distribution. A large maximum KS temperature distribution distance was seen for the majority of the pins in the assembly. The worst performing pins in this respect were pin 4, pin 9, and pin 20. This is due to the aforementioned high span-to-span and pin-to-pin repeatability of the TKE distributions and conversely the low repeatability of the temperature distribution. Since the maximum KS distance may occur in CTF axial edit which do not contain temperatures in excess of the saturation point, the maximum KS distance is not an indicator of poor crud predictive

Table 5.2: TKE distribution KS statistic summary. Values in bold result in rejection of the null hypothesis at significance level $\alpha = 0.1$.

Pin	KS_{μ}	KS_{max}
1	2.949e-02	**3.7027e-01**
2	8.487e-02	**3.7030e-01**
3	2.999e-02	1.9407e-01
4	1.111e-01	**5.5811e-01**
5	3.175e-02	1.8512e-01
6	2.891e-02	2.3469e-01
7	4.466e-02	2.2317e-01
8	5.642e-02	**3.2113e-01**
9	1.943e-02	1.3405e-01
10	5.650e-02	**2.8362e-01**
11	3.155e-02	1.5097e-01
12	6.701e-02	**4.3280e-01**
13	3.847e-02	2.3050e-01
14	4.220e-02	**3.5174e-01**
15	5.040e-02	**2.7588e-01**
16	9.001e-02	**4.1846e-01**
17	2.129e-02	2.0583e-01
18	2.702e-02	1.7656e-01
19	2.959e-02	1.5939e-01
20	3.804e-02	1.8140e-01
21	3.092e-02	2.3890e-01
22	2.263e-02	1.6278e-01
23	2.181e-02	1.4727e-01
24	3.227e-02	1.3258e-01
25	2.728e-02	1.5842e-01

performance. The presented KS tests only serve to quantify the ability of the gradient boosted quantile regressors to reproduce the expected distributions.

Pin-average KS distances, KS_{μ}, indicate that the the null hypothesis was not rejected in the majority of CTF axial zones, however. This indicates that the TKE distributions predicted by the gradient boosted quantile regressors were, on average, statistically indistinguishable from the empirical CFD distributions.

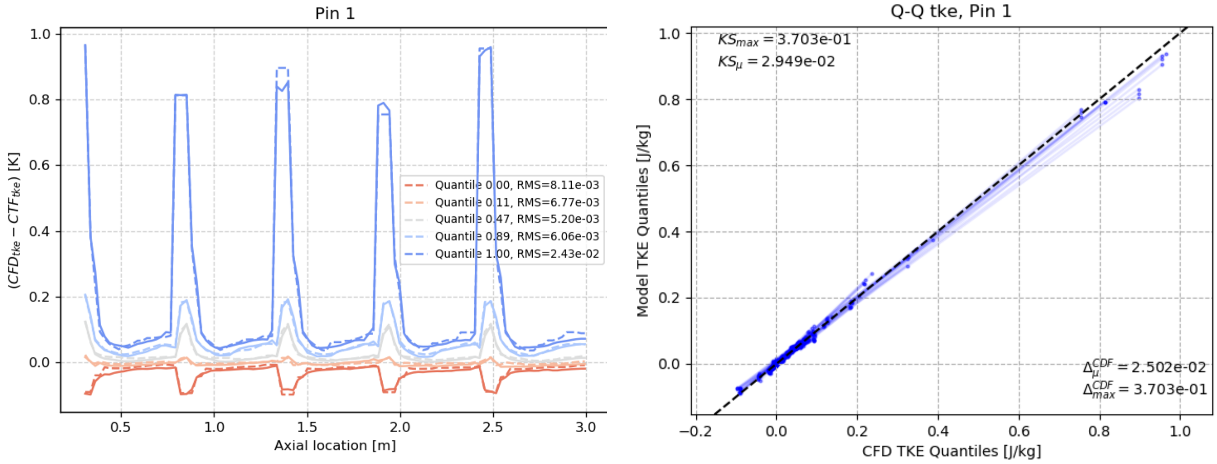
Caution should be observed when drawing conclusions from this goodness-of-fit study.

Table 5.3: Temperature distribution KS statistic summary. Values in bold result in rejection of the null hypothesis at significance level $\alpha = 0.1$.

Pin	KS_μ	KS_{max}
1	4.329e-02	**3.2094e-01**
2	8.477e-02	**3.9065e-01**
3	6.725e-02	**4.4959e-01**
4	1.509e-01	**6.0957e-01**
5	4.483e-02	**2.7359e-01**
6	8.718e-02	**4.4920e-01**
7	8.848e-02	**3.2938e-01**
8	5.668e-02	**4.2560e-01**
9	7.273e-02	**6.3407e-01**
10	1.292e-01	**4.6647e-01**
11	6.765e-02	**3.5969e-01**
12	1.124e-01	**4.9874e-01**
13	5.064e-02	**3.2152e-01**
14	6.728e-02	**4.7099e-01**
15	1.429e-01	**5.4111e-01**
16	7.260e-02	**3.9793e-01**
17	3.975e-02	2.3933e-01
18	3.209e-02	2.5577e-01
19	6.852e-02	**3.7904e-01**
20	1.273e-01	**8.3754e-01**
21	3.059e-02	**3.8896e-01**
22	1.147e-01	**6.6852e-01**
23	6.098e-02	**4.6728e-01**
24	6.525e-02	**4.7225e-01**
25	4.811e-02	**3.3235e-01**

The two sample KS test is generally regarded as a statistically weak, requiring a relatively large number of samples and high KS distance to reject the null hypothesis [55] [56]. The statistical power of a hypothesis test is defined as the probability of avoiding a type II error. In the currently considered case there is high probability of committing type II errors, or in other words, failing to reject the null hypothesis. For this reason and provided only 20 quantiles available for use in the KS test there is insufficient evidence to conclude the gradient boosted quantile regression models properly reproduced the expected temperature and TKE distributions on each face. In future work, a larger number of CFD data points and larger number of quantile regressors should be used to improve the ability of the KS test to identify incongruence between the model predictions and the expected distributions.

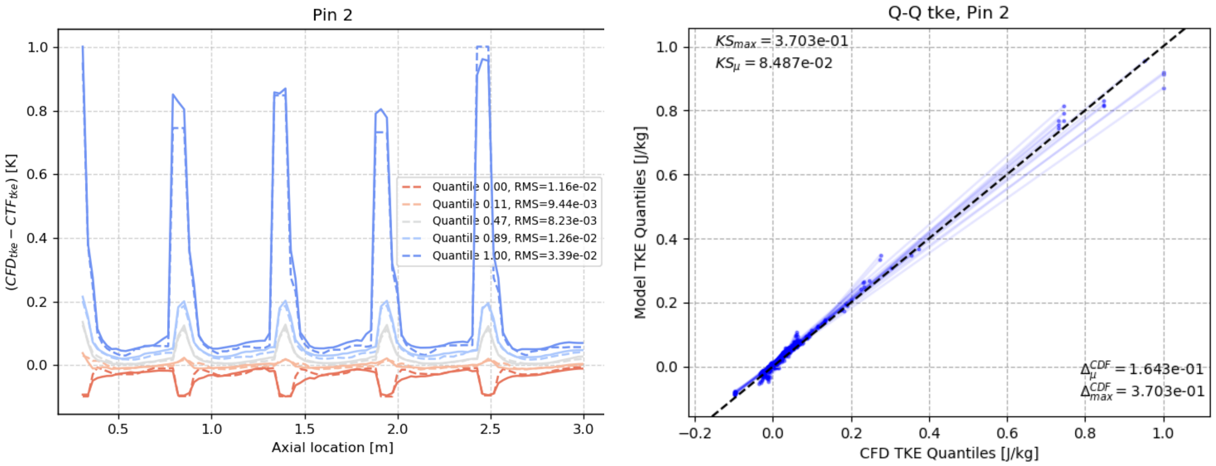
Note that since a LOO CV technique was used for comparing the predicted distributions to the empirical CFD distributions complications in the KS test which arise when the parameters of the predicted distribution are estimated from the target empirical data set were avoided [57]. Under these circumstances the KS test would no longer be valid though methods based on bootstrap resampling have been proposed to resolve this specific limitation of the traditional KS test [57].



(a) TKE quantile regression results. CFD in dashed line. Predicted values as solid. Azimuthally integrated.

(b) Q-Q plot of TKE quantile regression predictions from LOO cross validation study

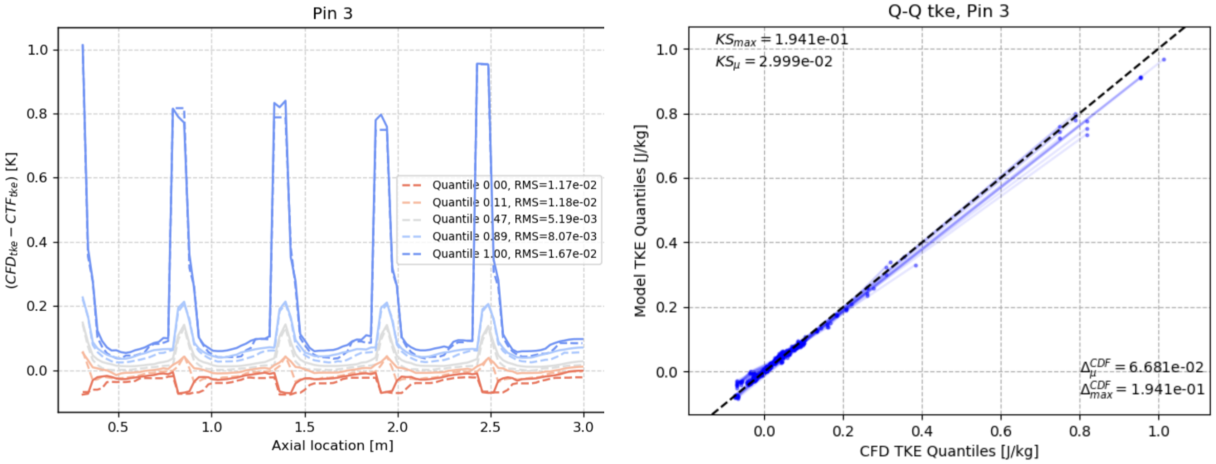
Figure 5.6: Pin 1 TKE quantile regression predictions from LOO cross validation study.



(a) TKE quantile regression results. CFD in dashed line. Predicted values as solid. Azimuthally integrated.

(b) Q-Q plot of TKE quantile regression predictions from LOO cross validation study

Figure 5.7: Pin 2 TKE quantile regression predictions from LOO cross validation study.



(a) TKE quantile regression results. CFD in dashed line. Predicted values as solid. Azimuthally integrated.

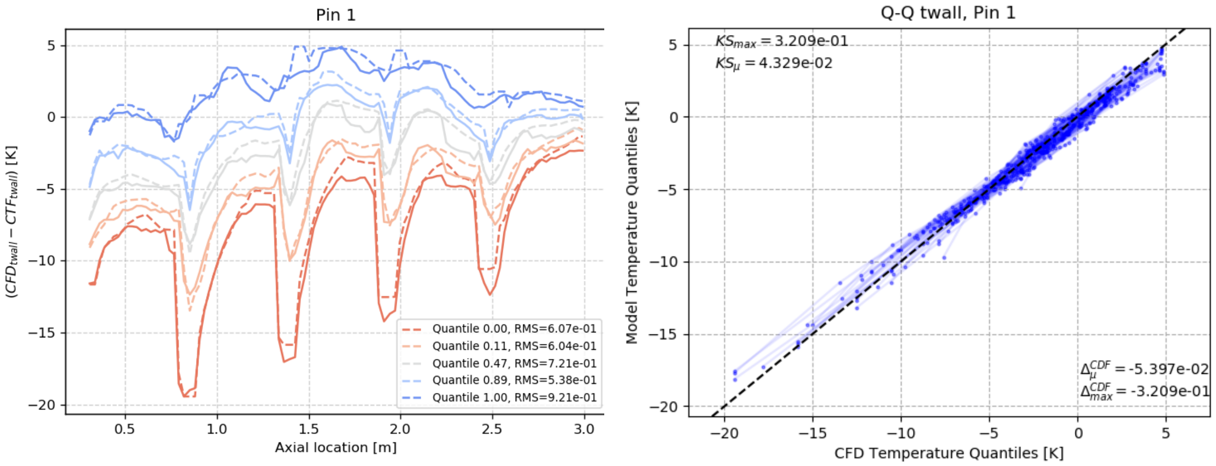
(b) Q-Q plot of TKE quantile regression predictions from LOO cross validation study

Figure 5.8: Pin 3 TKE quantile regression predictions from LOO cross validation study.

Shown in the axial plots in figures 5.6 to 5.8, the TKE distribution is drastically influenced by spacer grids. The maximum near-wall TKE sharply increases following a spacer grid followed by a decay towards a more orderly flow. The location of minimum predicted near wall TKE also immediately follows the spacer grids. In addition to increasing the net turbulent kinetic energy of the flow, mixing vanes also produce eddy regions of stagnant flow thus giving rise to regions of low near wall TKE. The hi2lo model retains both of these properties of the flow field resolved by CFD.

Good overall performance of the TKE quantile regression models may be attributed to high pin-to-pin and span-to-span similarities of the surface TKE distributions. The observation of high span-to-span repeatability of the TKE distributions is consistent with those found in other hi2lo studies by Salko et. al [19].

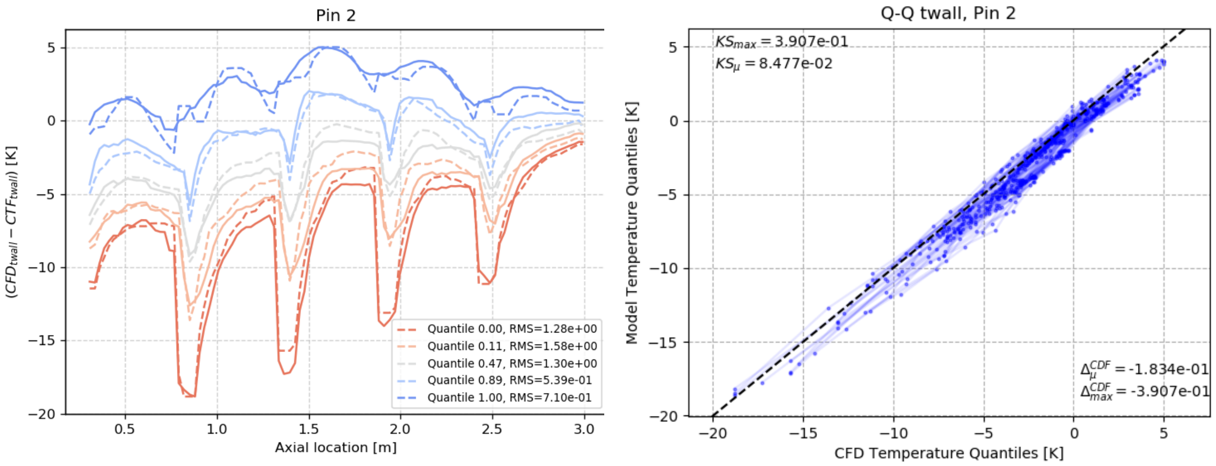
Temperature residual quantile regression results are given in figures 5.9 to 5.11. Similar to the TKE conditional quantiles, the conditional temperature distribution exhibits sharp changes in behavior across the spacer grids. Unlike the TKE residual distribution the surface temperature distributions do not exhibit the same degree of similarity from span to span or from pin to pin. The presence of discontinuities in the temperature distributions enforced the choice of the gradient boosted tree machine learning algorithm which is resilient to steep gradients in the response surface.



(a) Temperature quantile regression results. CFD in dashed line. Predicted values as solid. Azimuthally integrated.

(b) Q-Q plot of Temperature quantile regression predictions from LOO cross validation study

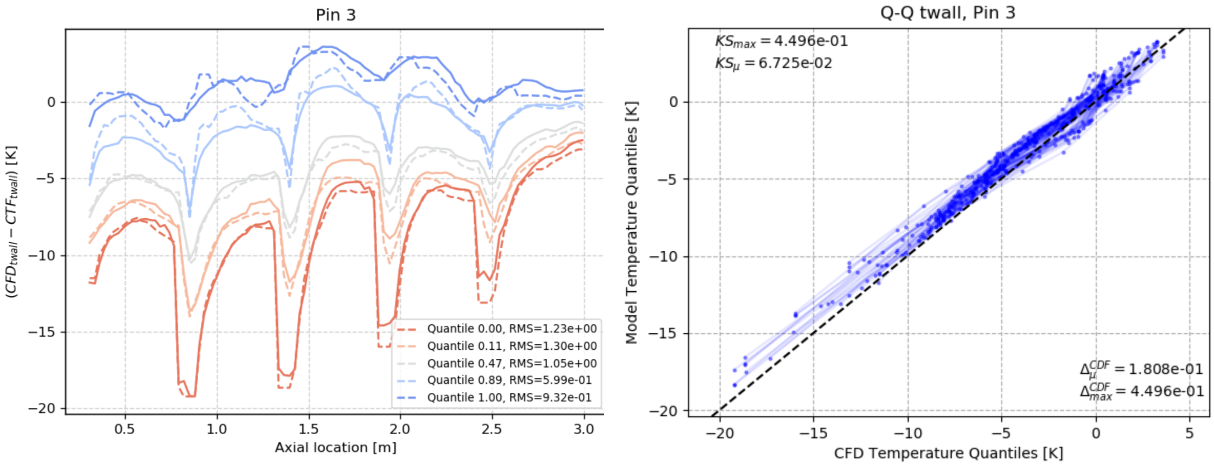
Figure 5.9: Pin 1 Temperature quantile regression predictions from LOO cross validation study.



(a) Temperature quantile regression results. CFD in dashed line. Predicted values as solid. Azimuthally integrated.

(b) Q-Q plot of Temperature quantile regression predictions from LOO cross validation study

Figure 5.10: Pin 2 Temperature quantile regression predictions from LOO cross validation study.



(a) Temperature quantile regression results. CFD in dashed line. Predicted values as solid. Azimuthally integrated.

(b) Q-Q plot of Temperature quantile regression predictions from LOO cross validation study

Figure 5.11: Pin 3 Temperature quantile regression predictions from LOO cross validation study.

Inspecting the axial quantile difference figures 5.9a to 5.11a, on average the extreme quantiles exhibit the largest axial root-mean-squared (RMS) prediction errors. The magnitude of the extreme quantile prediction errors may also gauged by inspecting the upper and lower regions in the Q-Q plots.

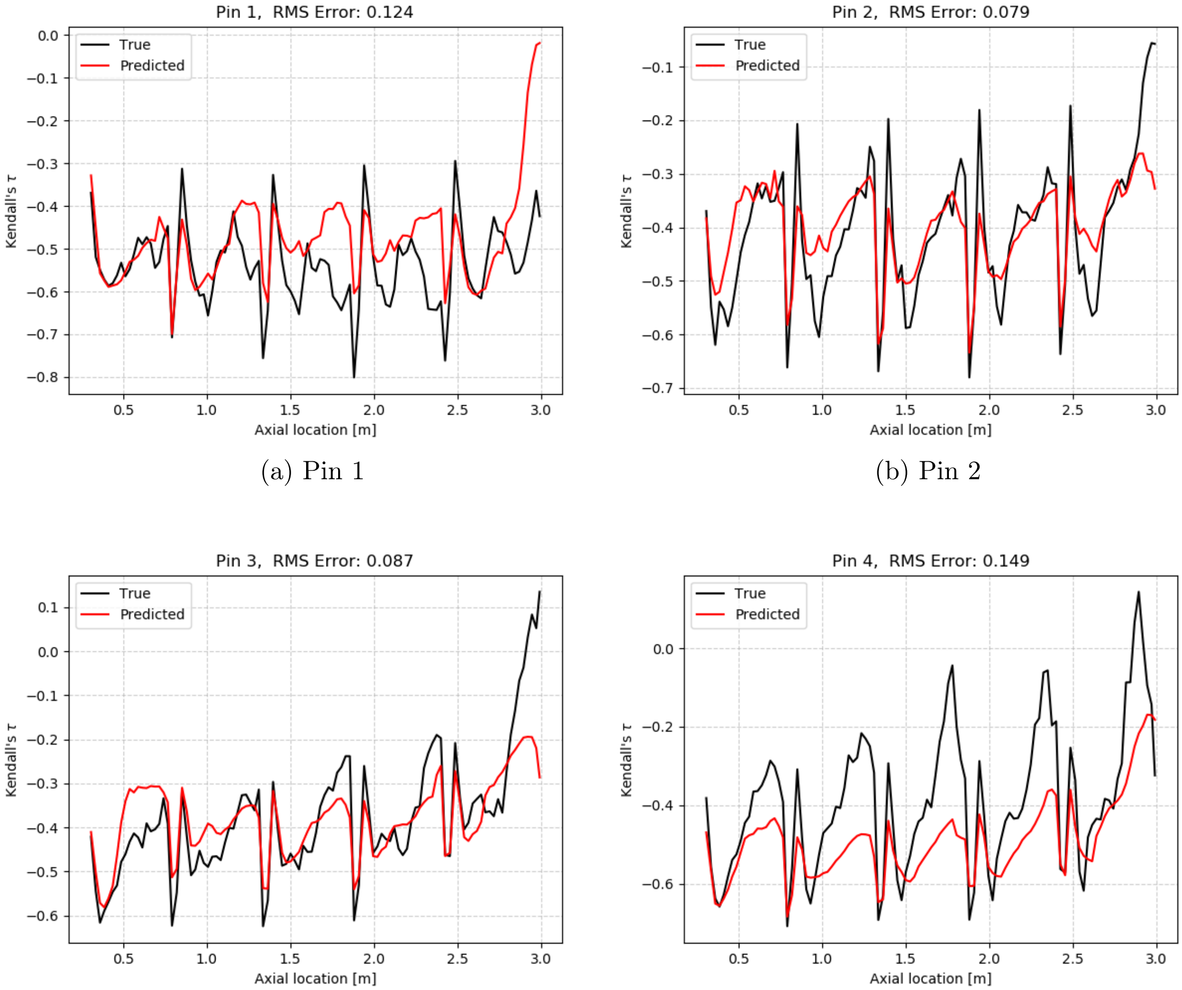
Estimates of the extreme quantiles from a sample population are naturally fraught with high variance as described by equation 3.28. Recall that this fact was also experimentally demonstrated using a simple test quantile regression problem in section 3.2.5. In both cases the distribution of the residuals between the gradient boosted quantile predictions and the empirical sample quantiles increased in variance when estimating the more extreme conditional quantiles.

5.1.5 Kendall's τ Regression

The rank correlation coefficient, Kendall's τ (ρ_{τ}), is used to quantify the strength of correlation between the temperature and TKE on the rod surface in each CTF face. A separate gradient boosted regression model was tasked with predicting this statistic as a function of local core conditions. The growth rate of crud was shown to be sensitive to ρ_τ in section 4.1.2, figure 4.9. It is therefore important to understand the error and uncertainty carried by the predicted $\hat{\rho}_\tau$ values in each CTF face.

A subset of the 5x5 assembly's Kendall's τ regression results are given in figure 5.12 and the complete 5x5 ρ_τ LOO cross validation results are given in figure A.5. There is a marked change in behavior of the rank correlation coefficient as a function of axial position in the core from pin to pin. The influence of Kendall's τ on the CTF face-integrated crud results was discussed in section 4.1.2, and it was shown to be an important parameter to accurately predict via the machine learning model. Pins with large relative errors for Kendall's τ are expected to produce anomalously poor crud predictions.

The worst performing pin with respect to $\hat{\rho}_\tau$ prediction was pin 8, as indicated in figure A.5. Interestingly, this pin exhibited relatively good agreement between the predicted crud distribution and the expected CFD crud distribution as indicated in table 5.4 and figure B.1. This pin, was relatively cold in comparison to the others in the fuel bundle which resulted growing only 5.9e-2 $[g]$ of crud in 300 days when the hottest rods grew $\approx$ 1.4e0 $[g]$ in the same time. In the case of pin 8, since the majority of the rod surface exists below the saturation point the crud result was not sensitive to the shape of the joint temperature and TKE distributions, and thus, even with poor ρ_τ predictions the axial and integrated crud results agree with the original CFD result.



(a) Pin 1

(b) Pin 2

(c) Pin 3

(d) Pin 4

Figure 5.12: Azimuthally integrated Kendall's τ regression results from LOO cross validation study.

To improve the performance of the Kendall's τ regressors, a larger training set could be generated in future work. For this limited 25 pin data set, it is hypothesized that each pin has a substantially unique flow field when compared to the other 24 pins. Expelling a pin from the training data set for cross validation causes the predictive performance of the model to suffer since the remaining pins in the training set do not provide the requisite information about the local core conditions vs. Kendall's τ relationship for the missing pin.

5.1.6 Copula Classifier

In addition to the rank correlation coefficient, Kendall's τ, the copula family is also required to recover the copula density function on each CTF face. To this end a gradient boosted classifier was trained on the available CFD data. Copula information extracted from the raw CFD results is shown in figure 5.3.

Figure 5.13 summarizes the LOO cross validation results of the copula classifier as a confusion matrix. The diagonal entries of the confusion matrix represent the correctly labeled copula predictions made by the reduced LOO trained classifier for each copula family average over the entire 5x5 assembly. It is shown that on average the classifier predicts an incorrect result more often than not.

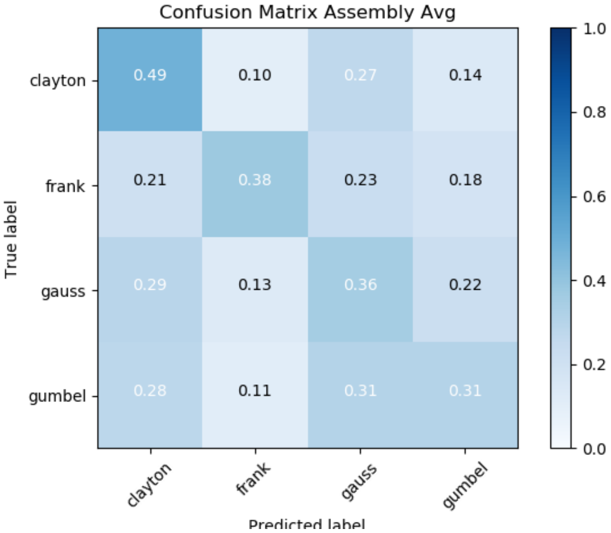


Figure 5.13: Copula classifier confusion matrix.

It is clear that the copula classifier struggles to predict the correct copula class given the local TH conditions. As previously indicated in figure 5.3, the behavior of the copula as a function of axial rod position is erratic and inconsistent from pin to pin. This erratic behavior proved too difficult to capture provided the limited training data set. It is not possible to conclude that the copula are well-described by the local thermal hydraulic conditions and axial position. It remains as future work to investigate if including additional geometric pin and grid attributes could improve the classification results. Additional software infrastructure would be required to both write geometric pin and grid features from the CTF code and to utilize these geometric features in the current model.

Future work could include performing a transformation of the input space so that the copula family labels are separable in the transformed space. A potential candidate for building this transformation is the UMAP manifold learning algorithm [58].

Further improvements in prediction accuracy are possible by applying an ensemble machine learning technique known as stacking. Stacking combines the predictions of multiple classifiers using a meta-classifier. Stacking increases model complexity since each classifier in the ensemble contains hyper-parameters which require tuning. Since machine learning model tuning and performance is not a focus of this work, the application of this technique to improve copula classification results is left as future work.

Though improvements are possible, it should also be noted that section 4.1.2 and table 4.4 show that the copula family does not substantially influence pin integrated crud results. Due to this, gains in the copula classifier accuracy will not necessarily translate to a large improvement in crud prediction accuracy.

5.2 Crud Results

The presented case considered the 5x5 array operating at a single state with fixed power profile and flow conditions for 300 days. The hi2lo model was marched forward in time using a resampling step size of 50 days. A sample size of 400 was used to estimate the crud distribution in each CTF face. The importance sampling distribution parameters were set to values given in table 3.1. The default remapping weights of $w_T = 0.4,\ w_k = 0.6$ were used in this case.

The comparisons presented are the result of the LOO cross validation study. Ergo the hi2lo model was used in a properly predictive manner since distribution parameters had to be inferred from the machine learning model at local core conditions outside of the training set.

The error estimates provided by the present LOO cross validation study may be viewed as conservative. The LOO strategy expunged an entire pin from an already limited training pool of only 25 pins. This is not representative a production-ready training data set. A training set to be used in a production environment will have all possible pin geometries represented within it; that is all possible pin configurations within a bundle. Not all combinations of inlet and power conditions will be simulated by CFD due to computational time limitations. Interpolation error can be expected even if provided a geometrically rich training set.

5.2.1 Single Pin Comparisons

A single pin was selected from the 25 pin array for detailed comparison of the CFD, CTF, and hi2lo models. For this pin, axial crud distribution comparisons were made at 300 [*days*] of simulation time are shown in figures 5.14 and 5.15. Axial crud distributions of all pins are provided in appendix B. The CTF standalone case generally predicts a greater amount of crud at all axial positions. Since the CTF model did not include any grid-enhanced heat transfer model it is to be expected that surface temperature downstream spacer grids would be over-predicted since the influence of the mixing vanes on the rod surface temperature distributions are partially neglected. The Hi2lo model preserves the influence of the spacer grids on the crud distributions predicted by CFD computations.

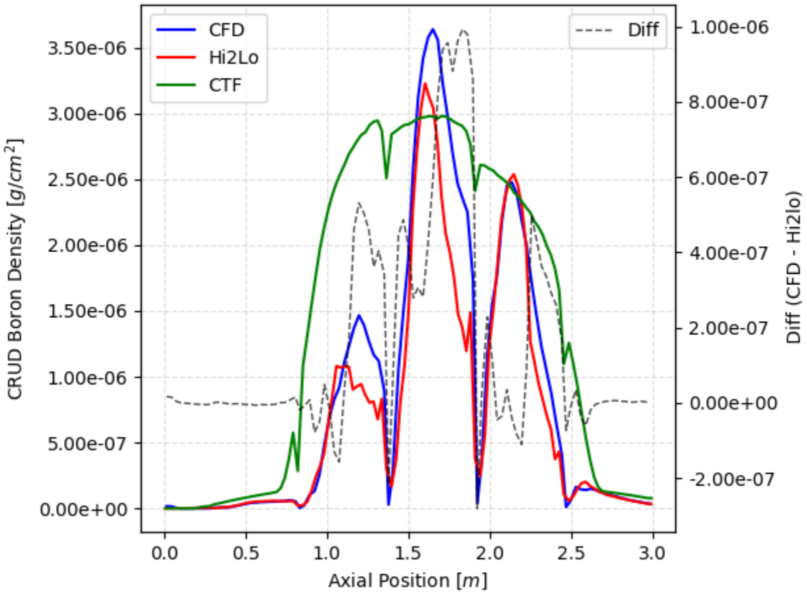


Figure 5.14: Pin 1 CTF vs CFD vs Hi2lo axial crud boron mass distribution at 300 days.

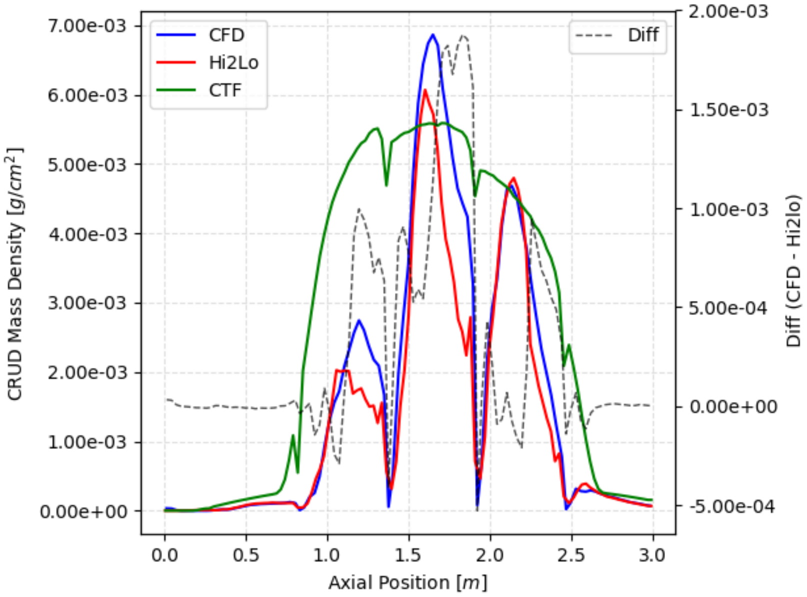


Figure 5.15: Pin 1 CTF vs CFD vs Hi2lo axial crud mass distribution at 300 days.

The total crud mass and total boron hideout mass were computed at each resampling step and presented in figures 5.16 and 5.17. The time evolution of the crud total mass for all pins is given in appendix B. The hi2lo model under predicted the crud mass on pin 1 when compared to the CFD model.

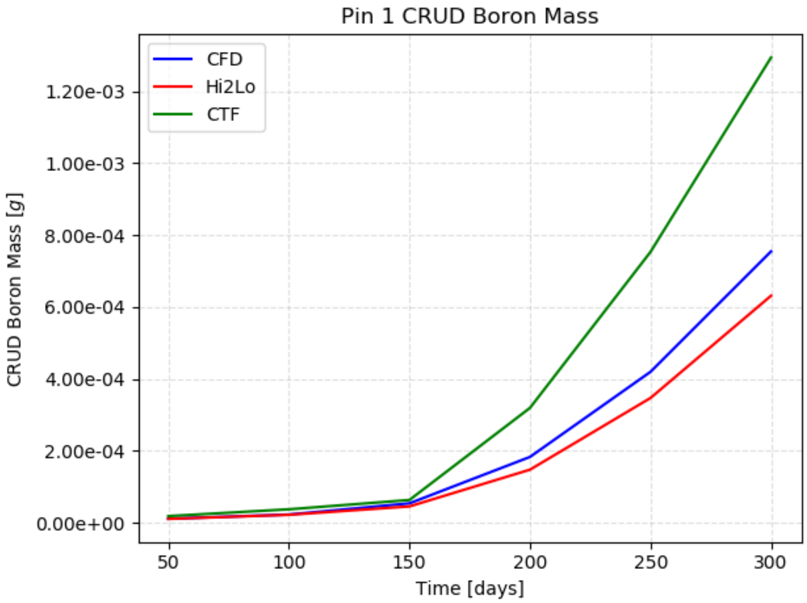


Figure 5.16: Pin 1 CTF vs CFD vs Hi2lo integrated crud boron mass distribution as a function of time.

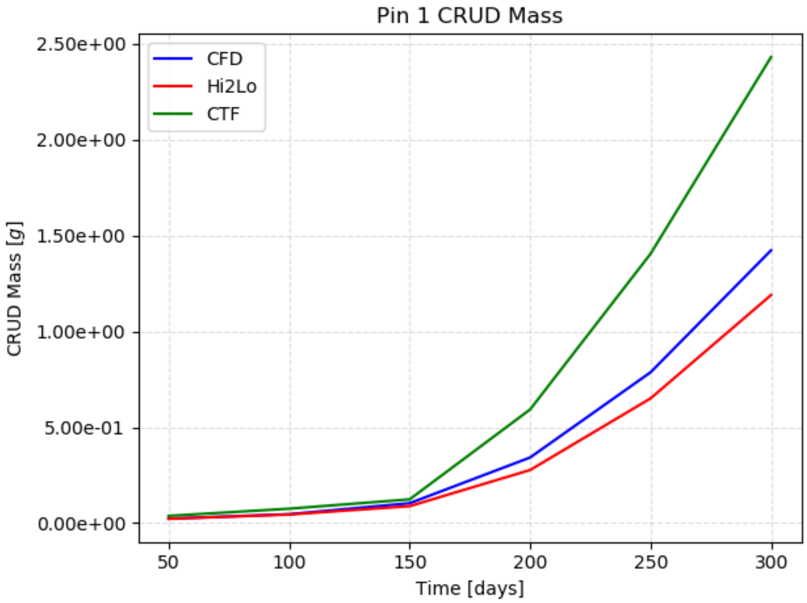


Figure 5.17: Pin 1 CTF vs CFD vs Hi2lo integrated crud mass distribution as a function of time.

Figures 5.18 and 5.19 show the hi2lo predicted crud surface distributions at 300 days. The result of re-ordering samples onto each CTF face to preserve hot spot stationarity in time is visible. The stripped patterns are non-physical and are an artifact of the remapping procedure. Recall that the overarching goal is not to reproduce the detailed intra-CTF face spatial crud distributions rather the model specifically attempts to reproduce the correct average crud behavior on each CTF face, even in regions near spacer grids, and estimate the frequency of extreme crud events so to be relevant for CILC risk estimates. Note that the crud surface field results are left in the area-normalized form with units of $[g/cm^2]$ which is the natural result from the 1-D crud growth package.

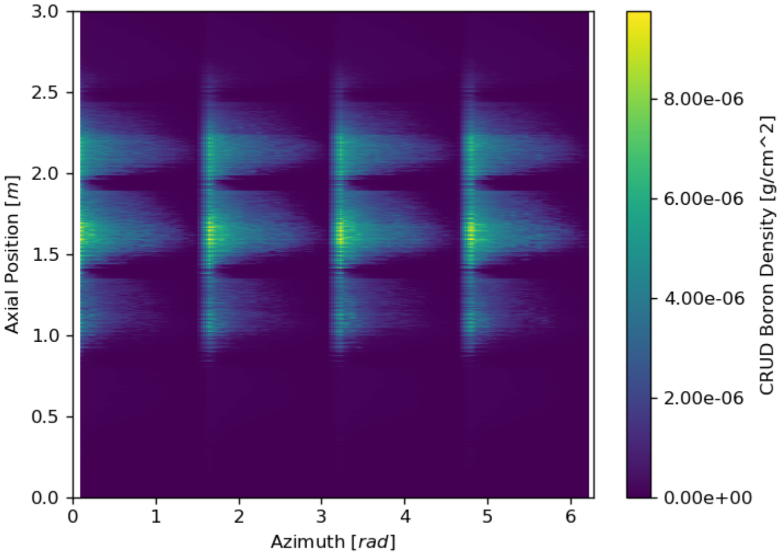


Figure 5.18: Pin 1 hi2lo 2-D surface map of crud boron mass density.

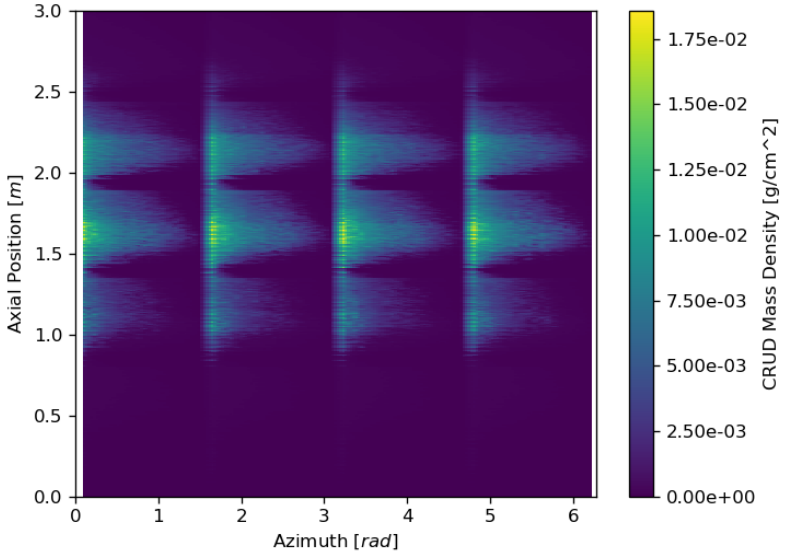
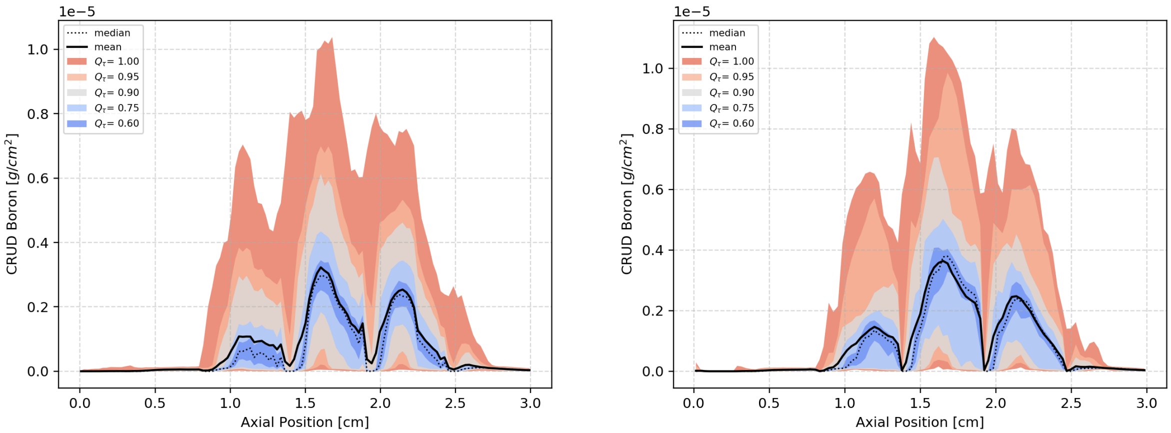


Figure 5.19: Pin 1 hi2lo 2-D surface map of crud mass density.

The average crud behavior as a function of axial position along the rod is given in figures 5.14 and 5.15. The axial crud root-mean-squared error is given in table 5.4 alongside other pins in the assembly. Pin 1 exhibits good agreement between the hi2lo model's crud predictions and the CFD results for the axial crud distribution when compared to other pins in the assembly. The rod integrated crud mass is also consistent between the two.

The crud density distributions predicted by the Hi2lo procedure are approximately consistent with the gold-standard CFD result as shown in figures 5.20 and 5.21. Some difficulty in capturing the extreme quantiles of the crud distributions as a function of axial position along the pin is shown in the figures.

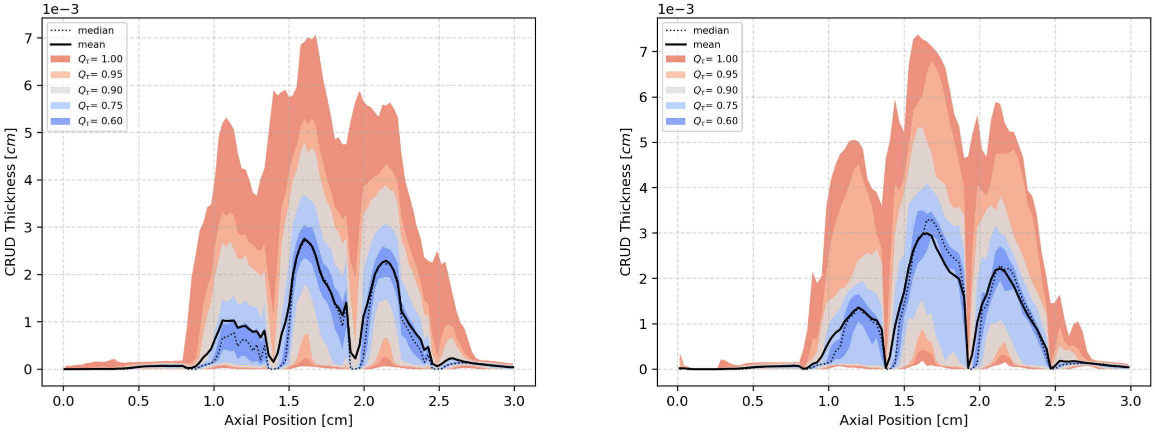
The ability of the hi2lo model to accurately predict the fraction of the rod surface which experiences extreme crud thickness, a precursor quantity to CILC risk estimation, is hampered by limitations of the quantile regression and the relative sparsity of the available training data. Recall that a given large-sample quantile follows a Gaussian distribution according to equation 3.28. By the propagation of uncertainty to upper tail integrals of the probability density detailed in equations 3.29 and 3.32, estimates for extreme crud distribution quantiles (i.e. estimates of how much of the rod surface experiences crud with a thickness exceeding some critical CILC crud threshold) will have high variance. Difficulty in predicting extreme quantiles by standard quantile regression reflects basic facts about the large sample limit of extreme quantiles. Circumventing these difficulties is a non-trivial undertaking. Without making assumptions for the functional form of the surface temperature distribution, thereby adopting a parametric model, it is difficult to estimate the likelihood of extreme crud events.



(a) Hi2lo pin boron mass.

(b) CFD pin boron mass.

Figure 5.20: Pin 1 crud boron mass density results at 300 days. Select crud quantiles are indicated via colored bands. The agreement of the mean axial crud boron density distribution between the hi2lo vs CFD models is better than in the upper quantiles.



(a) Hi2lo pin crud thickness.

(b) CFD pin crud thickness.

Figure 5.21: Pin 1 crud thickness results at 300 days. Select crud quantiles are indicated via colored bands. The maximum crud thickness predicted by the hi2lo model is approximately 70 microns at 300 days. Likewise the maximum crud thickness predicted by coupled CFD/crud computations was approximately 72 microns. Additionally note that the the mean crud thickness deviates from the median indicating asymmetry in the crud thickness density distribution.

5.2.2 Multi Pin Comparisons

Results for each pin in the LOO cross validation study are presented here. There was random variation in the prediction accuracy of the model across the 5x5 assembly with no apparent spatial bias in the model prediction errors towards the edge of the assembly, as one may expect. This would indicate that some pins in the 5x5 assembly are, in a sense, more unique with respect to thermal hydraulic flow conditions than others. Some pins, especially pin 9, show a small difference between the hi2lo model predictions and the gold-standard CFD result. The thermal hydraulic conditions surrounding these high performing pins are well represented in the training set.

In table 5.4 and 5.5 rod integrated crud results for each pin are given at 300 days of simulation time. The worst performing pin with respect to boron deposition prediction was pin 4 by relative percent difference between the hi2lo result and the CFD driven crud result. The boosted regressor produced large Kendall's τ prediction errors for this pin as shown in table 5.6. Incorrect predictions made for Kendall's τ acted in concert with a net under-prediction of the temperature quantiles, as shown in figure A.3, which gave rise to a significant ($\approx -48\%$) net under prediction of the total crud mass on pin 4. This highlights the importance of correctly predicting the conditional quantiles of the temperature distribution on the rod surface as a function of local core conditions since crud growth is highly sensitive to the outer cladding temperature.

Table 5.4: Crud boron mass hi2lo LOO result summary at 300 days.

Pin	CTF Bmass	CFD Bmass	CTF-CFD	Hi2lo Bmass	Hi2lo-CFD	Hi2lo-CFD
	$[g]$	$[g]$	Rel Δ%	$[g]$	$[g]$	Rel Δ%
1	1.2940e-03	7.5489e-04	71.4	6.3163e-04	-1.2326e-04	-16.3
2	1.1458e-03	5.1953e-04	120.5	3.1487e-04	-2.0466e-04	-39.4
3	1.0265e-03	3.2678e-04	214.1	3.4192e-04	1.5140e-05	4.6
4	1.0111e-03	5.6847e-04	77.9	2.9848e-04	-2.6999e-04	**−47.5**
5	9.5319e-04	1.8974e-04	402.4	2.2723e-04	3.7490e-05	19.8
6	4.0505e-04	9.1686e-05	341.8	1.0065e-04	8.9640e-06	9.8
7	8.0907e-05	4.0210e-05	101.2	3.7515e-05	-2.6950e-06	-6.7
8	6.5705e-05	2.9861e-05	120.0	3.3475e-05	3.6140e-06	12.1
9	6.7204e-05	3.3324e-05	101.7	3.3063e-05	-2.6100e-07	-0.8
10	6.6850e-05	3.5892e-05	86.3	2.8171e-05	-7.7210e-06	-21.5
11	8.7449e-05	4.0316e-05	116.9	3.7932e-05	-2.3840e-06	-5.9
12	4.6693e-04	1.1722e-04	298.3	8.2669e-05	-3.4551e-05	-29.5
13	1.0616e-03	2.2909e-04	363.4	2.6878e-04	3.9690e-05	17.3
14	1.0617e-03	2.7471e-04	286.5	3.9548e-04	1.2077e-04	44.0
15	1.0366e-03	4.5067e-04	130.0	3.3163e-04	-1.1904e-04	-26.4
16	1.1594e-03	2.9707e-04	290.3	4.3385e-04	1.3678e-04	46.0
17	9.1090e-04	2.6739e-04	240.7	2.6569e-04	-1.7000e-06	-0.6
18	7.1616e-04	2.3468e-04	205.2	2.0092e-04	-3.3760e-05	-14.4
19	6.1329e-04	1.1289e-04	443.3	1.4341e-04	3.0520e-05	27.0
20	1.6370e-04	7.2277e-05	126.5	4.8991e-05	-2.3286e-05	-32.2
21	1.2524e-04	4.5454e-05	175.5	4.4092e-05	-1.3620e-06	-3.0
22	1.7007e-04	4.4576e-05	281.5	6.1729e-05	1.7153e-05	38.5
23	6.2594e-04	1.5092e-04	314.7	1.1521e-04	-3.5710e-05	-23.7
24	7.1144e-04	1.8479e-04	285.0	2.1561e-04	3.0820e-05	16.7
25	4.1668e-04	1.3290e-04	213.5	8.6106e-05	-4.6794e-05	-35.2
Totals	1.5443e-02	5.2453e-03	194.4	4.7791e-03	-4.6623e-04	-8.88

Table 5.5 also displays the relative percent difference between the CFD/crud mass estimates and the standalone CTF/crud mass estimates at 300 days are provided in the fourth column. Averaged over the assembly the standalone CTF/crud calculation produced a relative difference of 192.1%. After application of the hi2lo model, the assembly

averaged crud mass relative error dropped to -8.9%. Both the bias and variance of the relative differences were reduced by the application of the hi2lo method.

Table 5.5: Crud mass hi2lo LOO result summary at 300 days.

Pin	CTF Cmass	CFD Cmass	CTF-CFD	Hi2lo Cmass	Hi2lo-CFD	Hi2lo-CFD
	[g]	[g]	Rel Δ%	[g]	[g]	Rel Δ%
1	2.4316e+00	1.4232e+00	70.9	1.1899e+00	-2.3330e-01	-16.4
2	2.1537e+00	9.8041e-01	119.7	5.9479e-01	-3.8562e-01	-39.3
3	1.9302e+00	6.1962e-01	211.5	6.4702e-01	2.7400e-02	4.4
4	1.9040e+00	1.0737e+00	77.3	5.6557e-01	-5.0813e-01	**−47.3**
5	1.7982e+00	3.6318e-01	395.1	4.3346e-01	7.0280e-02	19.4
6	7.6893e-01	1.7713e-01	334.1	1.9431e-01	1.7180e-02	9.7
7	1.5980e-01	7.9597e-02	100.8	7.4350e-02	-5.2470e-03	-6.6
8	1.3032e-01	5.9122e-02	120.4	6.6394e-02	7.2720e-03	12.3
9	1.3329e-01	6.6094e-02	101.7	6.5574e-02	-5.2000e-04	-0.8
10	1.3259e-01	7.1162e-02	86.3	5.5866e-02	-1.5296e-02	-21.5
11	1.7213e-01	7.9487e-02	116.6	7.5059e-02	-4.4280e-03	-5.6
12	8.8404e-01	2.2533e-01	292.3	1.5989e-01	-6.5440e-02	-29.0
13	2.0002e+00	4.3636e-01	358.4	5.1048e-01	7.4120e-02	17.0
14	1.9970e+00	5.2130e-01	283.1	7.4719e-01	2.2589e-01	43.3
15	1.9474e+00	8.4999e-01	129.1	6.2756e-01	-2.2243e-01	-26.2
16	2.1787e+00	5.6215e-01	287.6	8.1922e-01	2.5707e-01	45.7
17	1.7124e+00	5.0727e-01	237.6	5.0361e-01	-3.6600e-03	-0.7
18	1.3477e+00	4.4664e-01	201.7	3.8244e-01	-6.4200e-02	-14.4
19	1.1573e+00	2.1674e-01	434.0	2.7412e-01	5.7380e-02	26.5
20	3.1488e-01	1.4063e-01	123.9	9.6251e-02	-4.4379e-02	-31.6
21	2.4285e-01	8.9271e-02	172.0	8.6669e-02	-2.6020e-03	-2.9
22	3.2634e-01	8.7701e-02	272.1	1.2055e-01	3.2849e-02	37.5
23	1.1797e+00	2.8810e-01	309.5	2.2075e-01	-6.7350e-02	-23.4
24	1.3379e+00	3.5204e-01	280.0	4.0989e-01	5.7850e-02	16.4
25	7.8681e-01	2.5506e-01	208.5	1.6677e-01	-8.8290e-02	-34.6
Totals	2.9128e+01	9.9713e+00	192.1	9.0877e+00	-8.8360e-01	-8.9

At each resample step the crud mass was summed over all pins in the assembly. The time dependent assembly crud mass is presented in figure 5.22. The total assembly crud mass predicted by CFD and the hi2lo model are in in close agreement. At 300 days simulation time the relative difference in the crud mass results between the hi2lo and the CFD model was -8.9%, as shown in table 5.5. In general the expected error produced by the hi2lo model depends on the quantity and quality of the training data available. A study of the relative crud error as a function of training data set size is left to future work as a study of this nature would require performing a significantly larger number of

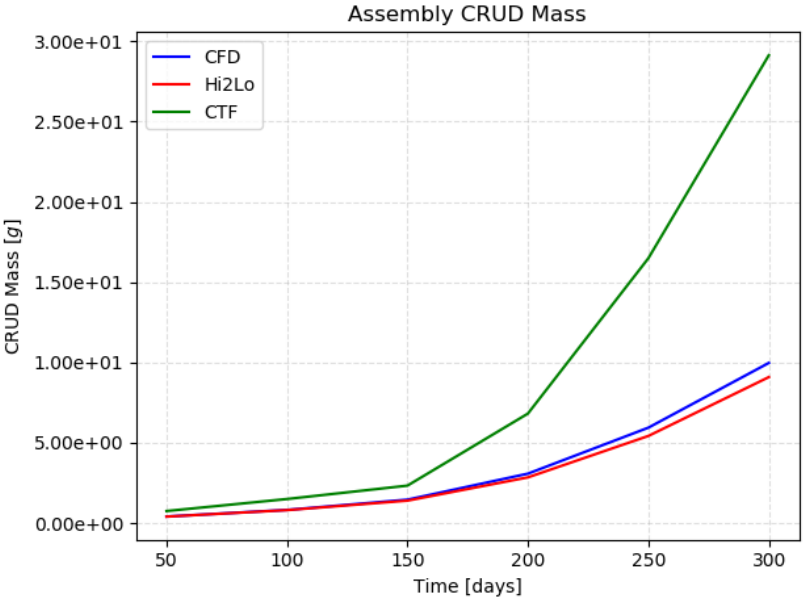


Figure 5.22: Assembly integrated CTF vs CFD vs Hi2lo crud mass as a function of time.

CFD computations than was performed here.

The RMS axial crud distribution errors are summarized in table 5.6. To examine the hi2lo model predictions for geometric biases across the assembly, top-down views of the RMS boron mass and crud mass error distributions were generated and presented in figure 5.23 and 5.24 respectively. The pins which reside on the edge of the assembly did not exhibit any increase in crud RMS error on average. The root mean squared error is given by: $RMS = \sqrt{\frac{1}{J}\sum_j^J(\text{Hi2lo}_j - \text{CFD}_j)^2}$ where j is the CTF face index on a given pin and J is the total number of CTF faces the pin. A large RMS error corresponds to a large mismatch in the axial crud distribution between the hi2lo model and the CFD/crud model predictions.

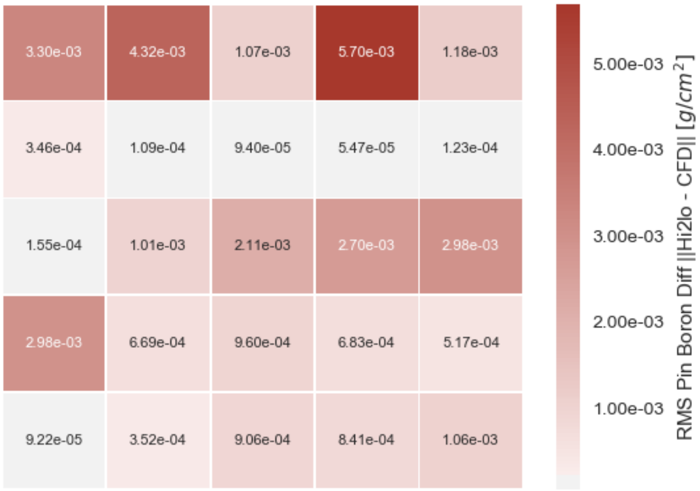


Figure 5.23: 5x5 average axial RMS crud boron mass error distribution. Top down bundle view.

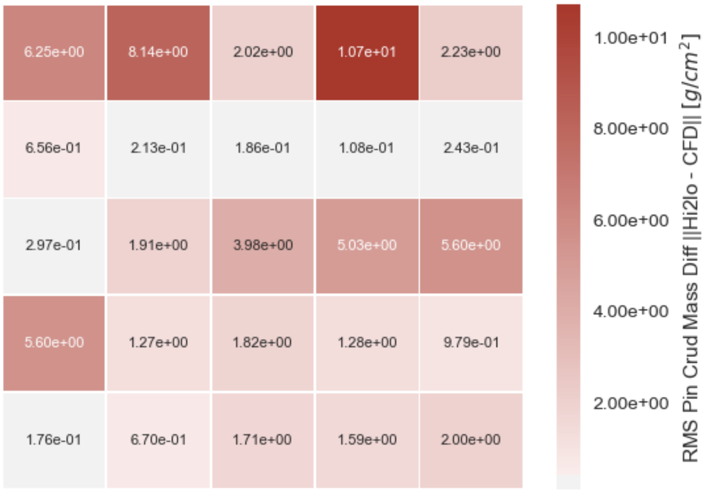


Figure 5.24: 5x5 average axial RMS crud mass error distribution. Top down bundle view.

Consistent under prediction of crud across the assembly was not observed. The top down assembly view of the crud prediction relative error distribution provided in figures

5.25 and 5.28 do not exhibit a tilt or other regular geometric pattern which would be indicative of location specific bias. The hi2lo model over predicts the total amount of crud for some pins in the assembly, particularly pins 14 and 16 but strongly under predicts the total crud in pin 4. This high variance in the prediction errors across the assembly can be partially attributed to small training sample size; however, a more in depth cross validation should be conducted as part of a future study in which a larger CFD training pool is available.

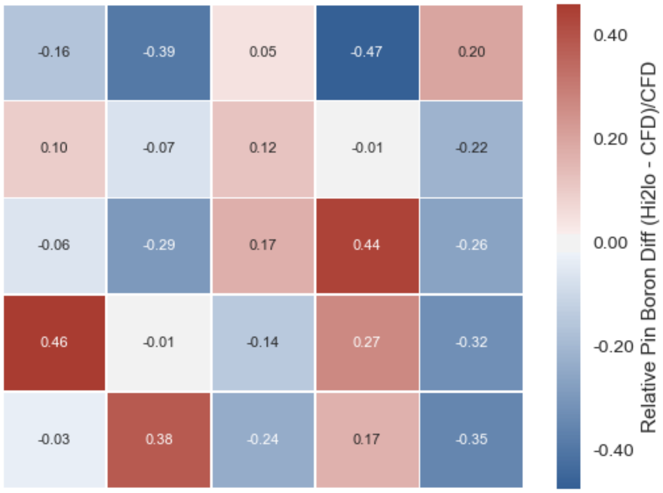


Figure 5.25: 5x5 integrated crud boron mass relative error distribution. Top down bundle view.

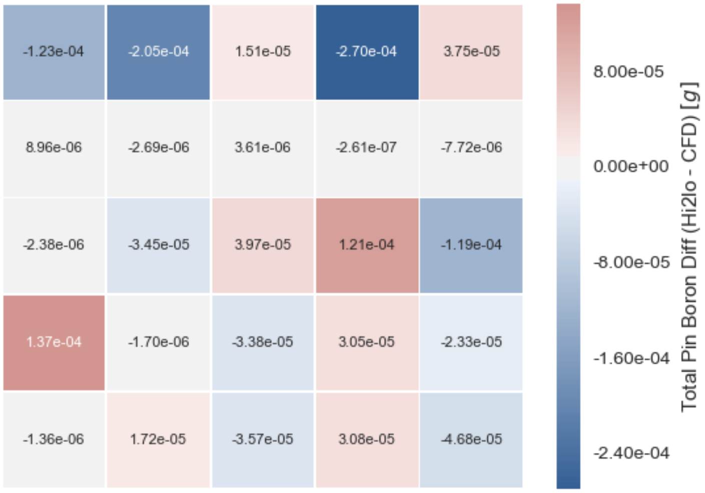


Figure 5.26: 5x5 integrated crud boron mass absolute error distribution. Top down bundle view.

The per pin absolute crud errors are given in figures 5.26 and 5.27. The absolute difference maps reiterate the that the crud errors do not exhibit a geometric bias or radial tilt.

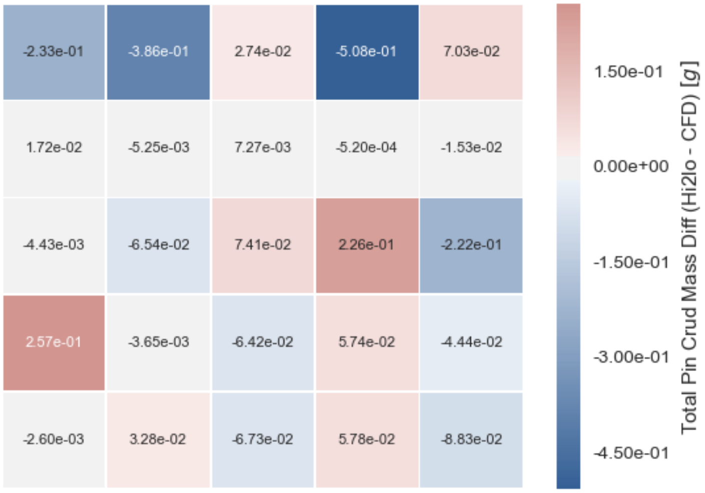


Figure 5.27: 5x5 integrated crud mass absolute error distribution. Top down bundle view.

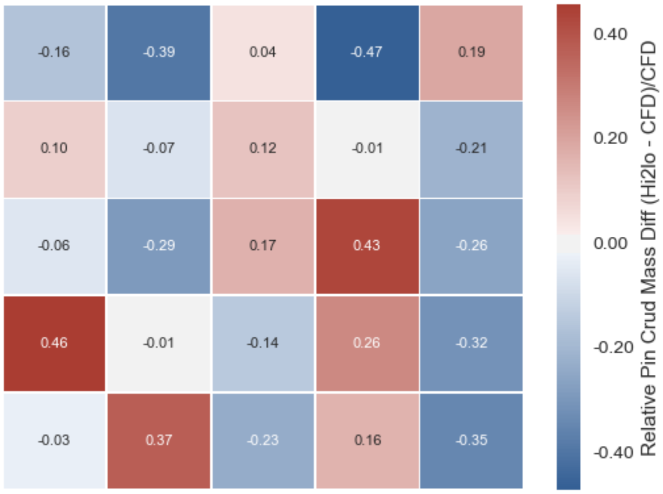


Figure 5.28: 5x5 integrated crud mass relative error distribution. Top down bundle view.

Correlations between crud prediction errors and errors committed by the quantile regressors were investigated in an attempt to establish performance metrics. Understanding the correlation between the machine learning model prediction accuracy and the crud prediction errors is helpful if one wished to estimate the expect crud growth errors before employing the hi2lo model in a production setting.

Provided sensitivities of the crud results to the machine learning prediction errors one may estimate the expected accuracy of the crud predictions obtained via the hi2lo model by standard propagation of error procedure shown in equation 5.3.

$$E_{c_j} \approx \sqrt{\sum_l \left(\frac{\partial E_{c,j}}{\partial E_{l,j}} \right)^2 E_{l,j}^2} \tag{5.3}$$

Where $E_{c,j}$ is the j^{th} pin crud mass error and E_l is the error associated with the l^{th} component of the ML model predictions. $E_{l,j}$ can be computed at training time since this quantity does not depend on the crud growth rate. Figure 5.29 displays estimates of the partial derivatives, $\frac{\partial E_{c,j}}{\partial E_{l,j}}$. This information is required in order for a user of the hi2lo model to detect problems with the trained quantile regression and copula models prior to employing the model to make crud predictions.

A Student-T test was conducted on the slope of each fitted linear trend lines. The results of the Student-T tests are shown in the upper-triangle of figure 5.29. The null hypothesis was taken to be a slope of zero. The standard deviation of the computed sensitivities is high when using a small sample size making it difficult to rigorously conclude that errors made by the machine learning models correspond to errors in the crud predictions.

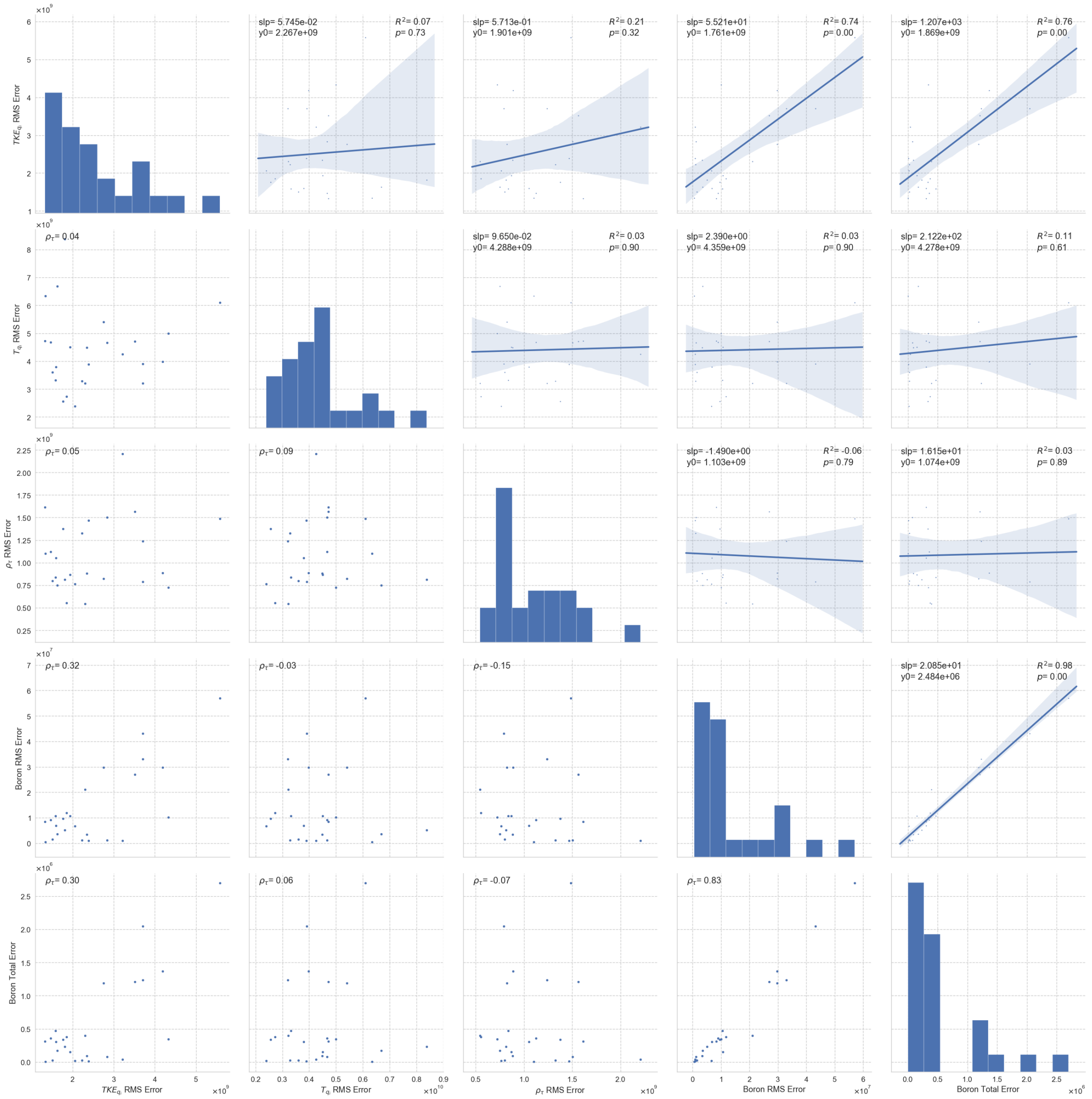


Figure 5.29: Correlation of ML errors with crud prediction errors.

Statistically significant trends were found between the RMS error committed by the TKE quantile regressors and the crud boron and mass distribution errors, as measured by root-mean-squared error. This suggests there is a link between being able to accurately predict the conditional quantiles of the TKE distributions and obtaining accurate crud estimates.

A strong positive correlation was observed between the root-mean-squared crud errors and the total crud mass prediction errors. This is a trivial result since one expects pins which exhibited large axial crud distribution RMS errors would also be likely to experience a large total integrated crud mass error unless, by happenstance, there was a cancellation of errors.

5.3 Section Takeaways

- Pre-processing the data set requires generating paired CFD and subchannel results with congruent geometries and inlet boundary conditions. The fine CFD data is first aggregated onto the subchannel grid. Explanatory features are extracted from the available subchannel results and paired with statistical properties of the residual distribution of the CFD result about the subchannel result in each subchannel face. The paired explanatory feature set and distribution properties are written to an HDF5 file for use as a training data set.

- Crud predictions made by the hi2lo model were compared against CFD/crud coupled results and CTF/crud results. The axial and integrated crud results produced by the hi2lo model compared favorably to the CFD results. The impact of spacer grids on the crud distribution was captured by the hi2lo model. Shown in figure 5.22, the assembly integrated crud mass results for the 5x5 assembly differed from the gold-standard CFD assembly integrated results by -8.8360e-01 $[g]$ for a relative difference of -8.9%.

- A leave-one-out cross validation strategy was utilized to estimate the predictive performance of the model.

- The prediction accuracy of the temperature and TKE quantile regression models was summarized through Q-Q plots for each pin in the LOO cross validation study.

- The prediction accuracy of the Kendall's τ regression model was assessed using the root-mean-square error for each pin in the LOO cross validation study.

- Correlations between the errors committed by the machine learning models and the crud prediction errors were computed. High uncertainty associated with these correlation measures did not permit a statistically significant link between poor Kendall's τ predictive performance and poor crud predictions to be drawn.

- The copula classifier performed poorly given the current set of considered explanatory variables and limited size of the training data set. A Gaussian copula was assumed on each CTF face in place of the poorly predicted copula family from the classifier. Recalling the results shown in section 4.1.2, this is not expected to reduce crud prediction accuracy.

Table 5.6: Hi2lo vs CFD crud RMS summary.

Pin	Axial RMS Error Crud Mass [g/cm^2]	Axial RMS Error Crud Boron Mass [g/cm^2]
1	6.2482e-04	3.3009e-07
2	8.1392e-04	4.3183e-07
3	2.0189e-04	1.0731e-07
4	**1.0737e−03**	5.7008e-07
5	2.2335e-04	1.1847e-07
6	6.5623e-05	3.4608e-08
7	2.1295e-05	1.0898e-08
8	1.8572e-05	9.4019e-09
9	1.0839e-05	5.4659e-09
10	2.4348e-05	1.2298e-08
11	2.9659e-05	1.5478e-08
12	1.9073e-04	1.0111e-07
13	3.9831e-04	2.1135e-07
14	5.0327e-04	2.6950e-07
15	5.5978e-04	2.9813e-07
16	5.6026e-04	2.9830e-07
17	1.2664e-04	6.6854e-08
18	1.8205e-04	9.5976e-08
19	1.2838e-04	6.8271e-08
20	9.7915e-05	5.1674e-08
21	1.7636e-05	9.2180e-09
22	6.7042e-05	3.5199e-08
23	1.7096e-04	9.0595e-08
24	1.5885e-04	8.4133e-08
25	1.9995e-04	1.0617e-07

6 Conclusion

This work joined gradient boosted quantile regression with copula to enable the joint temperature and TKE probability density on each CTF face to be predicted as a function of local core conditions. This enabled the evaluation of the total crud deposited on each CTF face via Monte Carlo integration. Additionally, the impact of hot spot stationarity assumptions were investigated. From this study it was shown that reordering samples within a CTF face at each sub-sampling step was an effective method to preserve hot spot stationarity. In addition to increasing the sample size, the ability to take many sub-sampling steps per VERA state point was shown to reduce Monte Carlo sampling uncertainties in the time stepping scheme. Finally, importance sampling was shown to be an effective means to increase sample efficiency.

6.1 Discussion

The application of copula allowed for independent treatment of the temperature, TKE, and BHF rod surface fields and their corresponding dependence structure. Treatment of the BHF as an independent random variable allowed further simplification of the copula model. The decomposition of the joint density into marginal models and a copula allowed for the marginal densities to be reconstructed via multiple quantile regression. Gradient boosted quantile regressions were employed in this role.

Employing a gradient boosted machine learning model in the role of quantile regression affords a great deal of flexibility in the selection of predictive features used in the model's construction. A suite of exogenous variables that could be obtained through

VERA or CTF results were identified; however, additional geometric pin and grid features could be included in the exogenous variable set in the future. The boosted models were shown to be robust to discontinuities in the response variables, which is a required property in the current application given the sharp jumps seen in the temperature, TKE, and rank correlation coefficient behavior across spacer grids.

An investigation of the 1-D crud code's response to varying surface temperature, TKE, and boundary heat flux led to the development of tailored proposal density distributions for use in importance sampling. The proposal distributions target the upper tail of the temperature distribution and lower tail of the TKE distribution so that a larger proportion of the available samples are expended on regions of the rods' surface that are more likely to harbor crud.

The ability to produce a hi2lo mapping in the case of missing CFD data at the desired TH state point was demonstrated through a leave one out cross validation study. This is a capability any candidate hi2lo tool must posses because not all geometric and TH conditions can be simulated upfront, so inevitably the a hi2lo mapping must be producible in cases where precisely matching CFD data does not exist in the training pool. However, no method was in place to detect if the boosted models were queried outside of their training envelope. Although the present model still produces predictions when extrapolated, the extrapolated predictions are not credible. Boosted trees are particularly sensitive to extrapolation since boosted tree models produce piecewise constant predictions.

The overarching strategy sought to estimate the expected value of the crud mass (and crud boron mass) in each CTF face by first reconstructing the joint density distribution of key surface field quantities on each face and then carrying out the expected value calculation through Monte Carlo integration. This approach opens up tremendous flexibility in the methods chosen to reconstruct the joint densities - though certain required model characteristics were identified, such as robustness to discontinuities. In future studies, deep learning strategies or stacked machine learning models could be employed in place of the gradient boosted models demonstrated in this work. It should be cautioned that incremental improvements in prediction accuracy possible through more complex machine learning strategies should be weighted against the benefits of simply increasing the size of the CFD training data set. A detailed investigation quantifying the crud prediction accuracy as a function of the training data set size should be conducted in a future study.

Applying the hi2lo technique presented in this work preserves more information about the flow field around spacer grids in the computation of the expected value of crud on each CTF face when compared to subchannel/crud standalone estimates. The presence of localized hot and cold spots on the rod surface were implicitly accounted for through the reconstruction of the joint density distributions of these fields in each CTF face. Forgoing the prediction of fine scale spatial details of the surface fields is justified because the crud models employed in this work are one dimensional in nature and do not require intra-CTF face resolved surface fields.

Employment of a Monte Carlo–based crud estimation procedure allows for physical intuitions to be built between the sample weights and the physical area represented by a crud sample on the rod surface. This intuition is especially helpful in the interpretation of the importance weights, which result from the application of the importance sampling variance reduction technique. Additional improvements in sampling efficiency are possible through other variance reduction techniques beyond importance sampling. To the author's knowledge, this work demonstrates a first-of-its-kind core-simulator scale Monte Carlo–based crud estimation procedure.

One strength of the Monte Carlo–based crud procedure is that it is straight forward to propagate hi2lo model uncertainties through time. Additionally, the Monte Carlo–based approach enables extreme value crud event estimation; although, these estimates are expected to carry large uncertainty because extreme upper and lower sample quantiles are plagued by high variance. Overall, difficulty predicting the upper conditional quantiles of the temperature distributions diminishes the applicability of this method to CILC.

It is not trivial to incorporate feedback between the crud layer and the hi2lo mapping because this would involve making the conditional surface temperature distribution depend on the current crud state. This crud/TH feedback was missed in the current implementation.

6.2 Future Work

A starting point for future hi2lo efforts should be a data scalability study in which the model's predictive performance is characterized as a function of the available training data set size. Additionally, the hi2lo model's performance under extrapolation was not investigated in the present work. The boosted models should not be employed in an extrapolation mode, so it is of interest to identify core conditions which would result in evaluating the trained machine learning models outside of the training data envelope at runtime. It is envisioned that a warning should be raised notifying the user that additional CFD data is required when attempting to evaluate the model outside of the TH zone of applicability.

A complete uncertainty quantification effort should precede efforts to perform a forward uncertainty propagation through the hi2lo model into the crud estimates. The following sources of uncertainty should be categorized: (1) Those arising due to the data or measurement based uncertainties and (2) those inherent in the Monte Carlo sampling procedures.

Some uncertainties are obvious, such as the uncertainties that arise due to Monte Carlo–based integration procedures, but other model uncertainties can be more difficult to identify. The CFD data itself presents one source of presently unquantified uncertainty. This type of CFD born uncertainty arises from the inability to exactly specify the coefficients used in the TH closure models, given some experimental flow data. The conditional quantile predictions made by the current boosted machine learning model take the form of point estimates where in reality there is some uncertainty in precisely where these quantiles lie. Future studies could consider employing Bayesian additive regression trees for estimating the variance associated with each conditional quantile prediction [59].

Once the sources of uncertainty in the training data and models are accounted for, they must be propagated from the machine learning model predictions through the distribution reconstruction procedures and finally into the crud estimates. For a credible CILC risk assessment, estimates for the maximum expected crud thickness in addition to the uncertainty in this value are required. Uncertainty quantification of thin nature is prerequisite for adoption of this methodology in a production setting.

Future feature engineering efforts should be directed at improving predictive performance under a wide variety of pin orientations and local core conditions. It is possible to use geometric features, such as rod position inside a bundle and bundle position inside the core, in addition to the presently considered local TH conditions. Careful selection of additional predictive features would improve the ability of the hi2lo model to generalize to previously unseen TH core conditions or unique pin configurations.

Since application of the present hi2lo model results in an estimate for the crud density distribution on each CTF face, it is natural to extend the model toward quantifying CILC risk. The first objective of a future CILC risk assessment study requires one to derive a CILC risk metric from the hi2lo crud result. Such a CILC cladding failure probability model could be described by $\mathcal{P}_f \propto Pr(C_t(x) > C_t^*)$, where C_t^* is some critical crud thickness and $\mathcal{P}_f$ is a cladding failure probability. This would be difficult to quantify with CTF/MAMBA alone and requires either a hi2lo approach or detailed investigation of at-risk pins with high fidelity CFD computations. A significant challenge is computing an estimate for $Var(\mathcal{P}_f) = \mathbb{E}[(\mathcal{P}_f - \mathbb{E}(\mathcal{P}_f))^2]$, or similarly, the variance in the expected amount of crud over a given threshold. This quantity is necessary to conduct credible CILC risk assessment.

6.3 Concluding Remarks

The primary goal of this work was to account for the fine scale flow features encountered downstream from spacer grids on the crud growth rate. A standalone subchannel code coupled to a crud simulation package could not achieve this feat alone. The hi2lo model developed in this work successfully demonstrated the ability to retain the influence of spacer grids on the axial crud growth profile and total crud mass in a small assembly test problem. These quantities are important in the evaluation of CIPS risk. The prediction accuracy of the trained hi2lo model was estimated using a LOO cross validation study.

The potential to increase predictive crud performance by expanding the size of the CFD training data set is a strength of the developed statistically based CFD-informed hi2lo model. This is possible because the hi2lo model employs a data driven, supervised machine learning strategy to predict key statistics governing the joint temperature, TKE, and BHF distributions on the rod surface. It is hypothesized that the prediction accuracy of the method will improve provided a larger number of CFD training examples than considered in this work. Gradient boosting was adopted as the machine learning algorithm of choice in the present work; however, the ability to interchange a more sophisticated ensemble technique or a deep neural network is a strength of the hi2lo framework developed in this dissertation.

Appendices

A | 5x5 Machine Learning Results

A.1 5x5 Leave-One-Out Machine Learning Results

Gradient boosted quantile regression model results are presented in figures A.1 and A.2. For each pin, the predictions are made for the left-out-pin and compared to the original CFD training data. The gradient boosted results are shown as solid lines and the original CFD-CTF data is shown as broken lines.

The presented quantile regression results are shown as a function of axial position along the rod for the residual surface temperature and TKE distributions, e.g $\hat{q}_\tau(z) = \mathbf{b}(z) + \varepsilon(z)$, where $\mathbf{b}(z) = \mu_{\text{cfd}} - \mu_{\text{ctf}}$. The results were averaged over the 4 azimuthal CTF faces at each axial level in the CTF grid. The root-mean-square (RMS) error of select quantiles prediction vs axial location are given in each figure.

Figures A.3 to A.4 show quantile-quantile (Q-Q plots) for each pin in the 5x5 LOO results. Each Q-Q plot summarizes the overall prediction quality afforded by the quantile regression averaged over the entire pin length. At each CTF axial grid level, the Kolmogorov–Smirnov (KS) statistic was computed to quantify the goodness-of-fit of the quantile distribution reconstruction to the original, empirical CFD-CTF distribution. The average and maximum KS statistic encountered is recorded in each figure.

Additionally, the predicted rank correlation coefficient as a function of axial position made by the LOO-trained models are compared to the expected result in A.5. The RMS error between the predicted $\hat{\rho}_\tau(z)$ and the CFD computed $\rho_\tau(z)$ is shown in each figure.

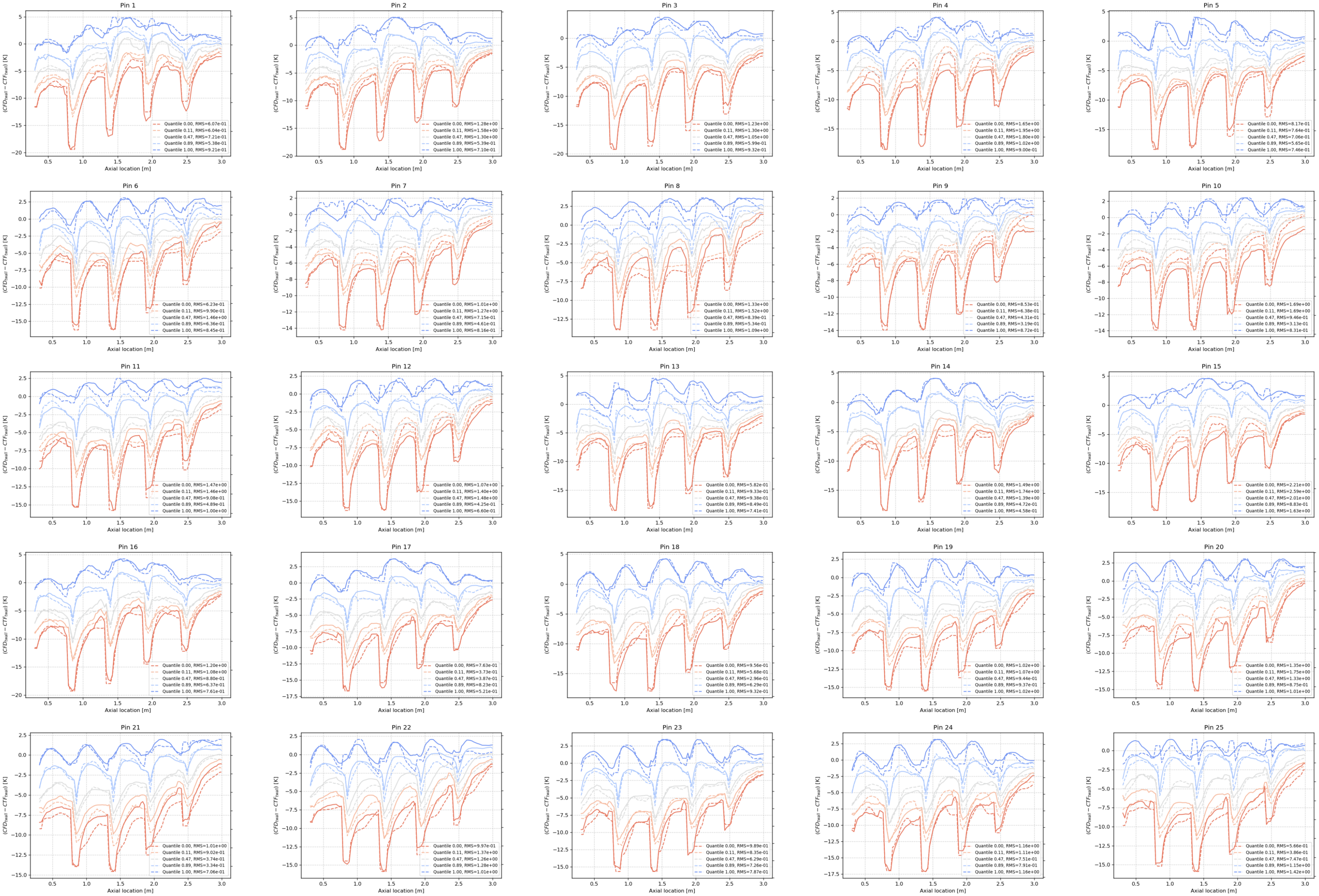


Figure A.1: 5x5 Axial surface temperature residual (CFD-CTF) quantile predictions.

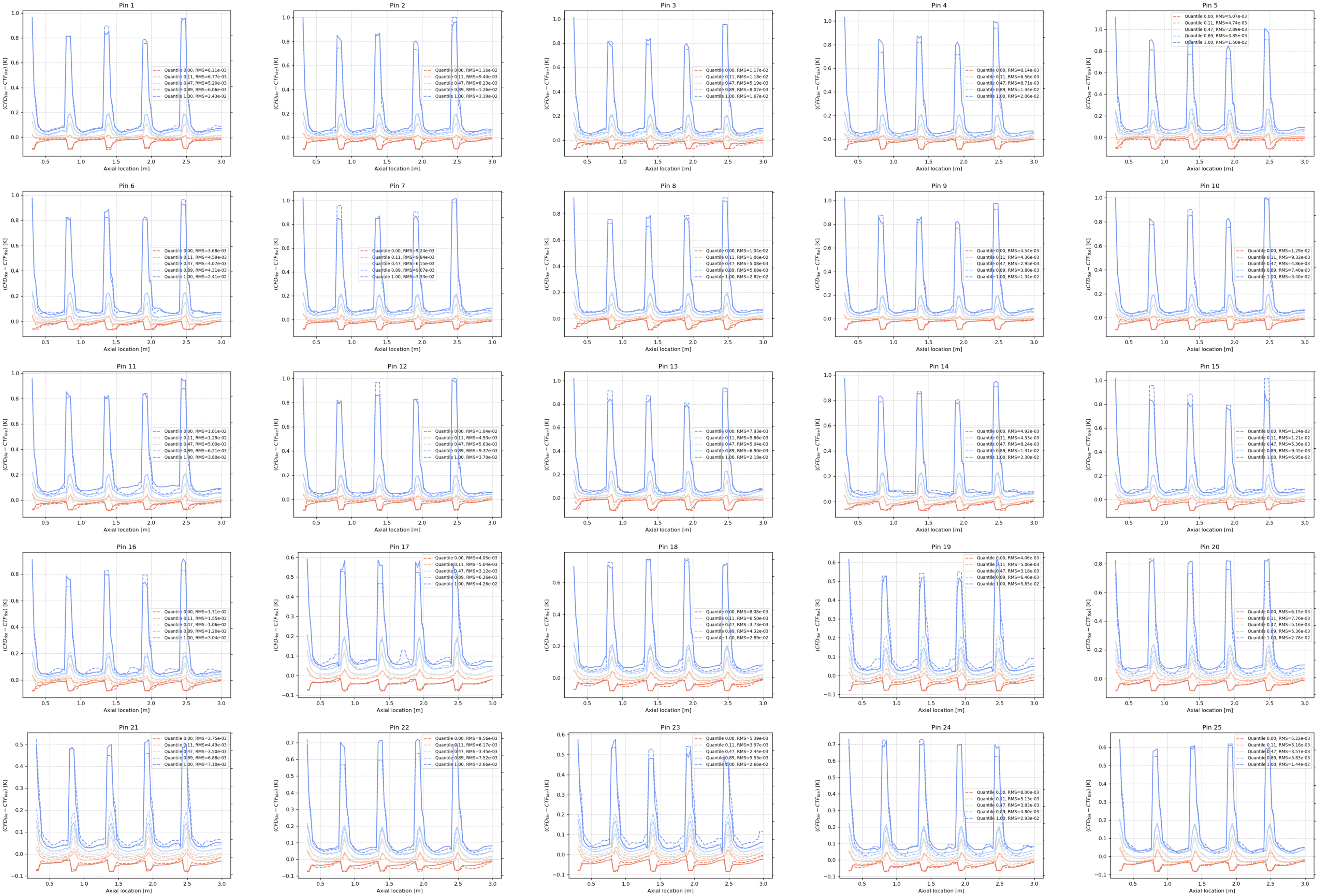


Figure A.2: 5x5 Axial TKE residual (CFD-CTF) quantile predictions.

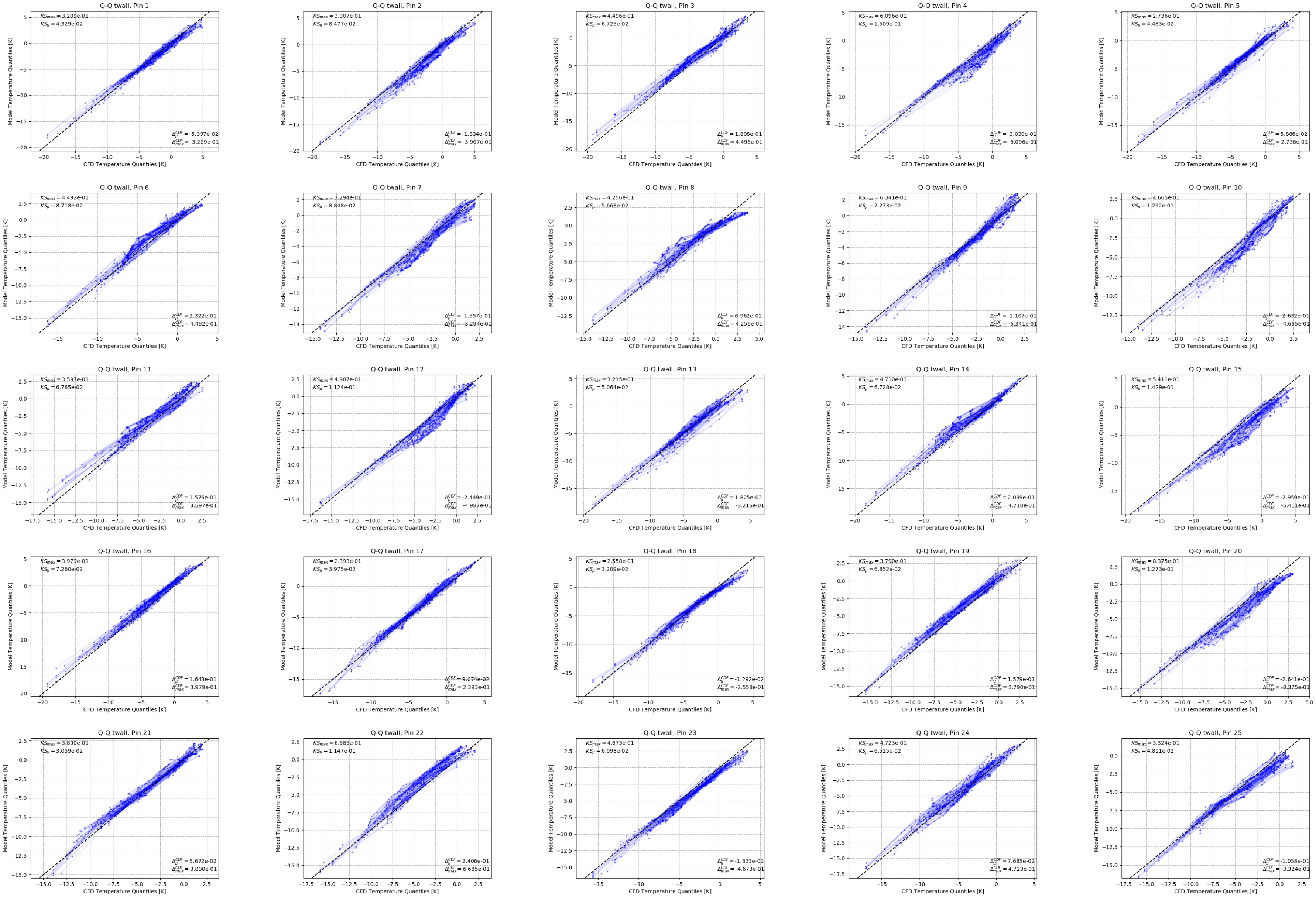


Figure A.3: 5x5 surface temperature quantile predictions Q-Q goodness-of-fit summary.

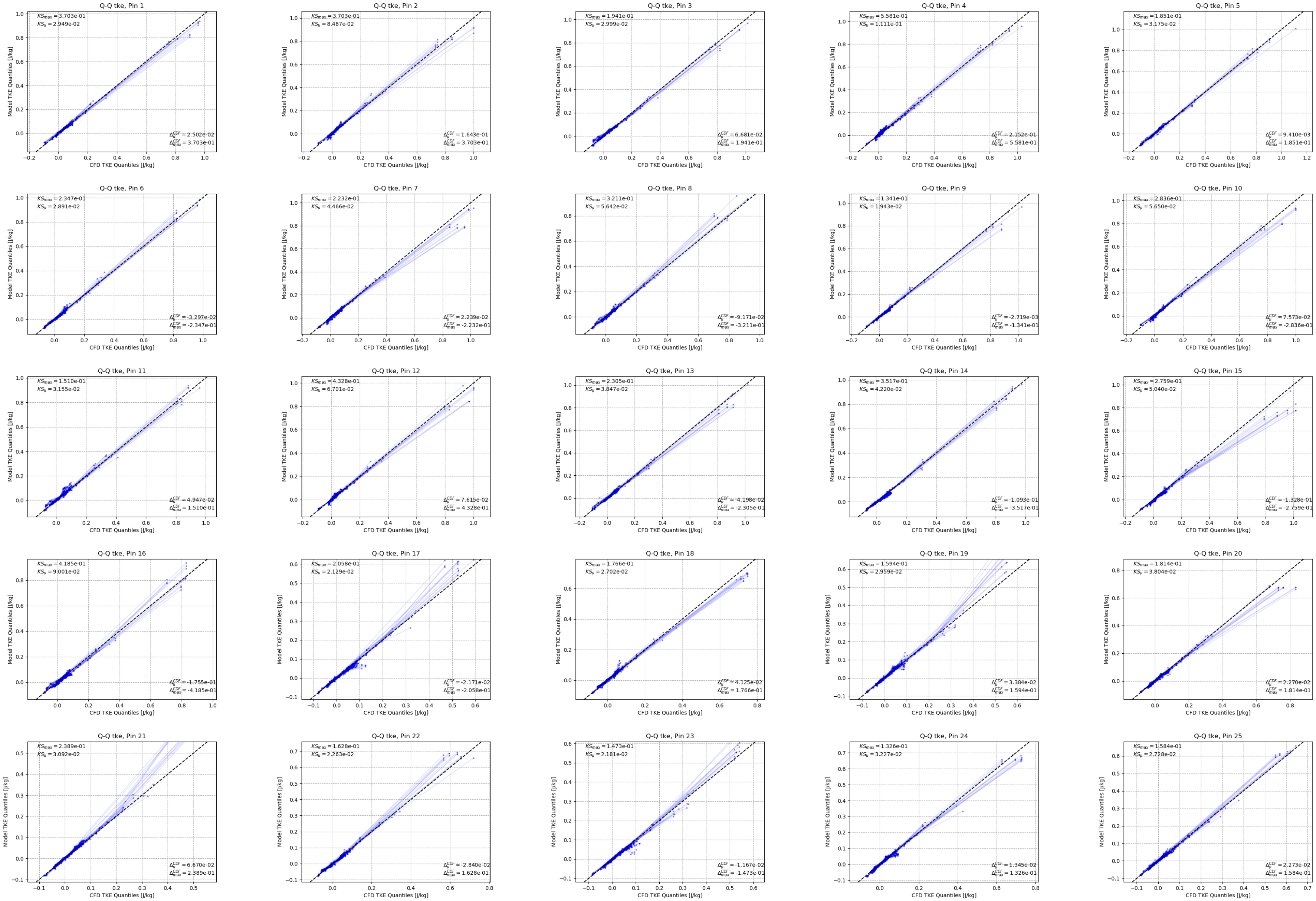


Figure A.4: 5x5 TKE quantile predictions Q-Q goodness-of-fit summary.

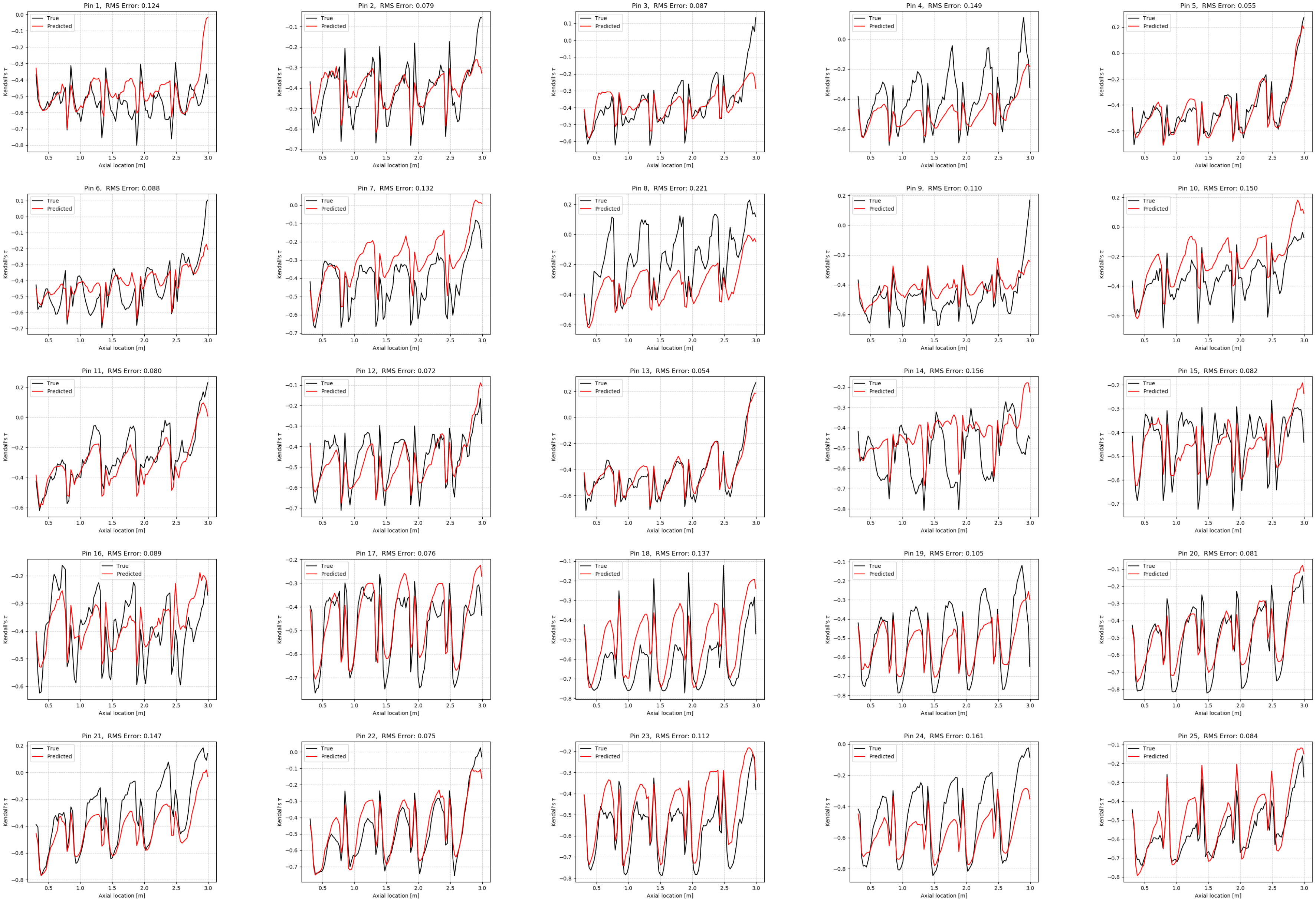


Figure A.5: 5x5 Kendall's τ vs axial position predictions.

B | 5x5 Crud Results

B.1 5x5 Results

All axial crud results for the 5x5 model are shown in figures B.1 and B.2. The axial distributions are shown at a simulated time of 300 days. Pin-integrated crud results plotted as a function of time are provided in figures B.3 and B.4.

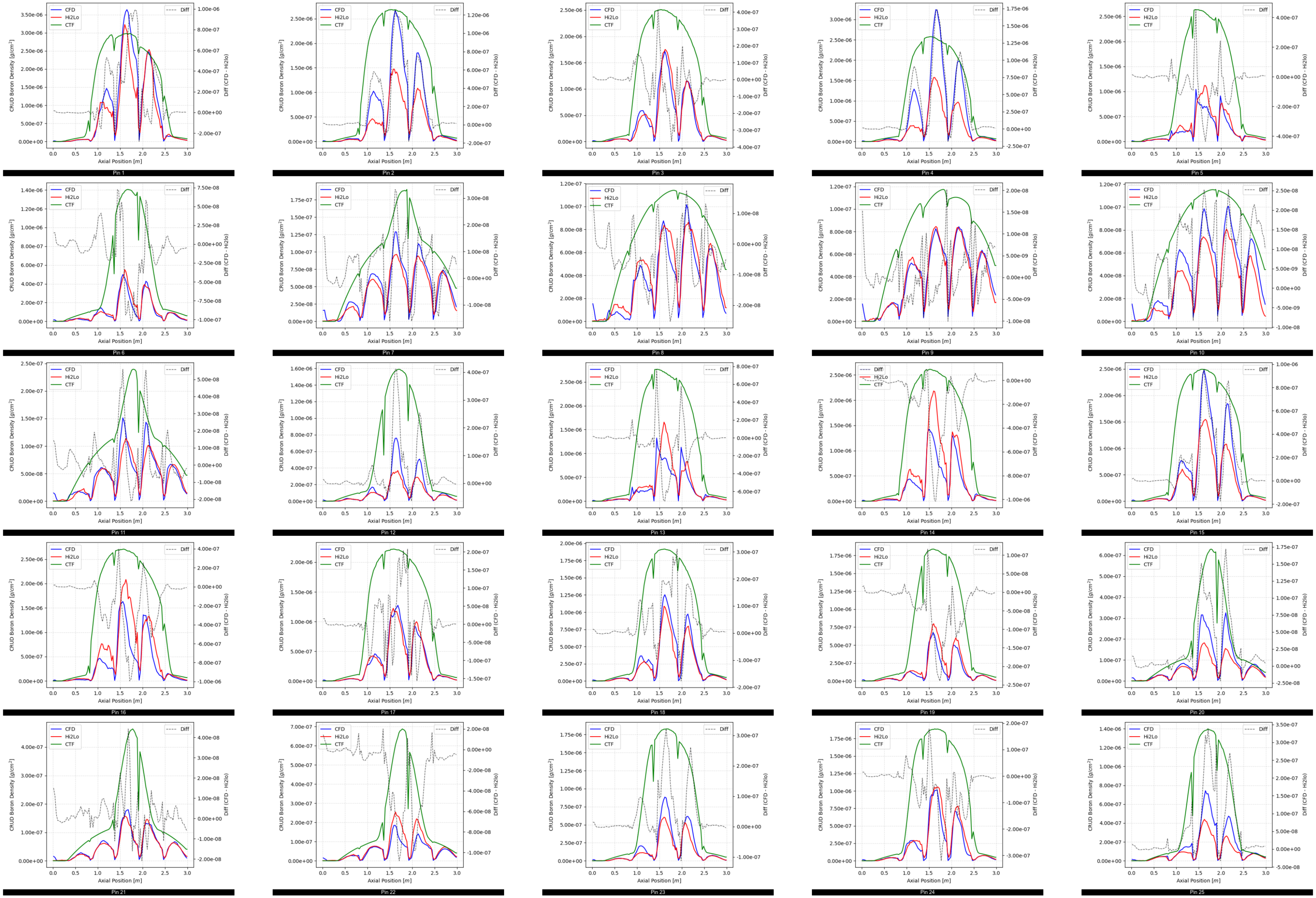


Figure B.1: 5x5 axial crud boron mass results at 300 days.

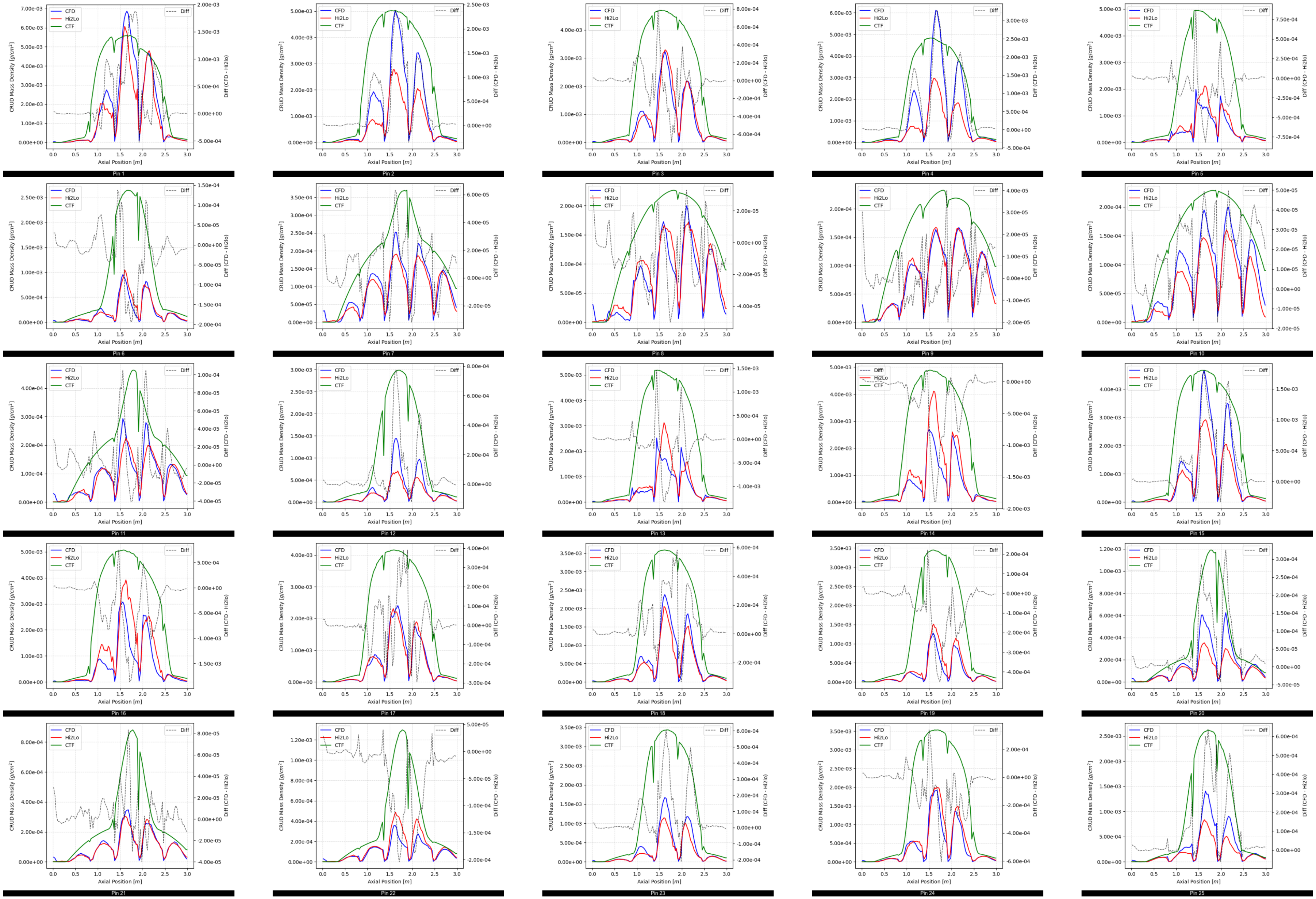
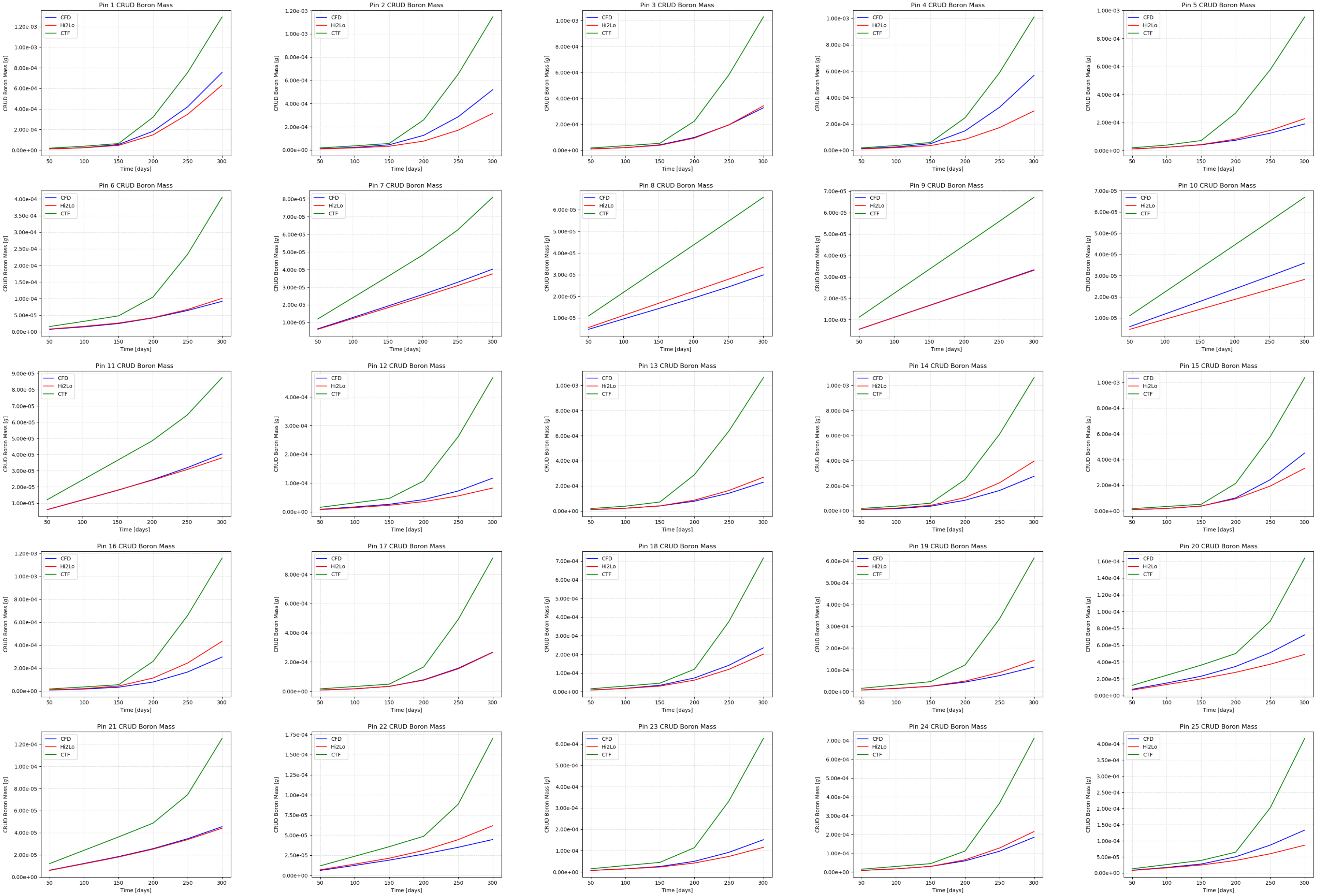


Figure B.2: 5x5 axial crud mass results at 300 days.



Pin 1 CRUD Boron Mass

Pin 2 CRUD Boron Mass

Pin 3 CRUD Boron Mass

Pin 4 CRUD Boron Mass

Pin 5 CRUD Boron Mass

Pin 6 CRUD Boron Mass

Pin 7 CRUD Boron Mass

Pin 8 CRUD Boron Mass

Pin 9 CRUD Boron Mass

Pin 10 CRUD Boron Mass

Pin 11 CRUD Boron Mass

Pin 12 CRUD Boron Mass

Pin 13 CRUD Boron Mass

Pin 14 CRUD Boron Mass

Pin 15 CRUD Boron Mass

Pin 16 CRUD Boron Mass

Pin 17 CRUD Boron Mass

Pin 18 CRUD Boron Mass

Pin 19 CRUD Boron Mass

Pin 20 CRUD Boron Mass

Pin 21 CRUD Boron Mass

Pin 22 CRUD Boron Mass

Pin 23 CRUD Boron Mass

Pin 24 CRUD Boron Mass

Pin 25 CRUD Boron Mass

Figure B.3: 5x5 rod integrated crud boron mass vs time.

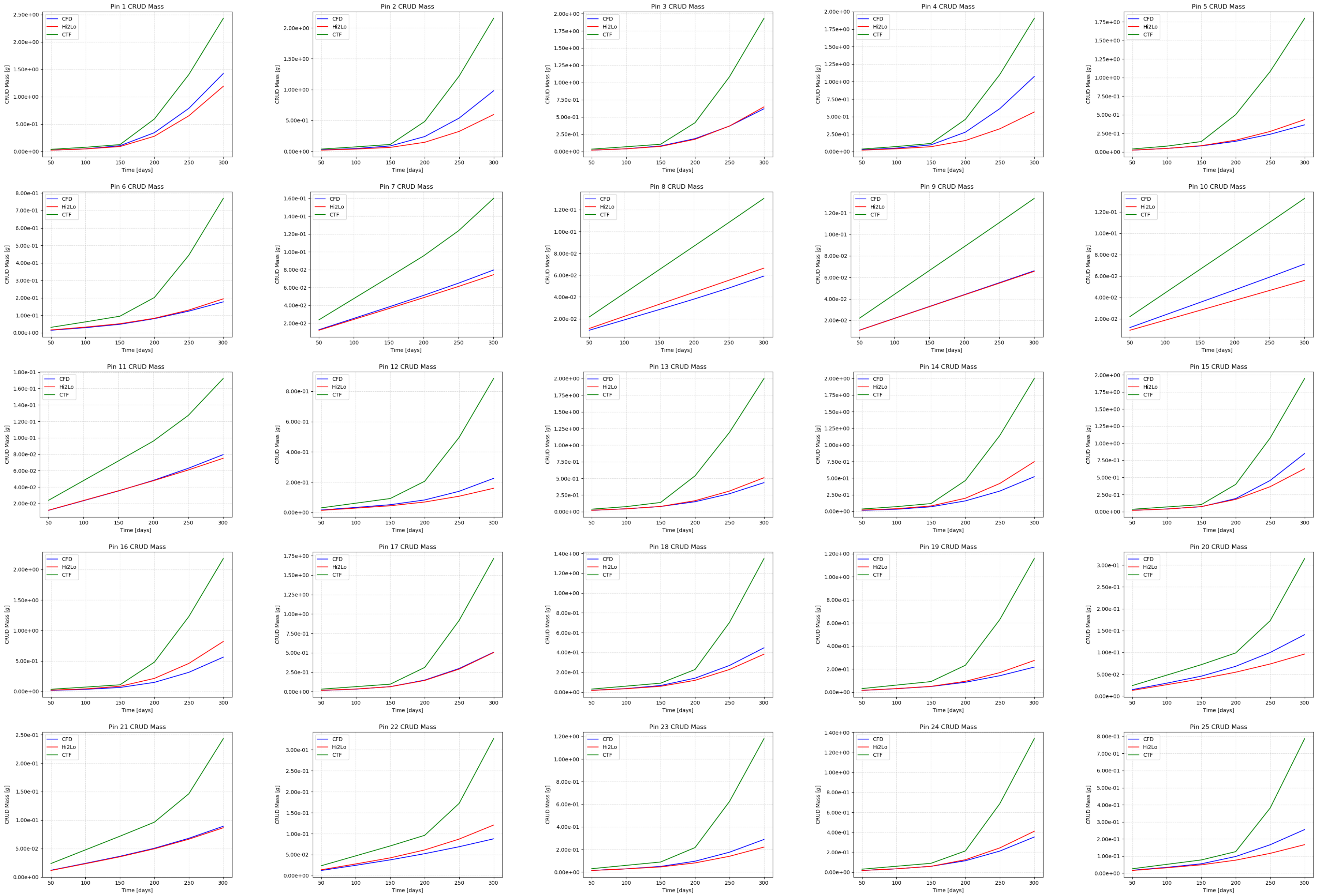


Figure B.4: 5x5 rod integrated crud mass vs time.

C | Software

C.1 Gradient Boosting Toolkit

A gradient boosting library was developed in the python programming language to support the hi2lo work. This package provides an easily extensible loss function class that a user can use to implement arbitrary loss functions in the gradient boosting framework. As required by the hi2lo work, both quantile and least squares loss functions are included. The package is applicable to both regression and classification problems. CART tree construction controls are also provided allowing fine grained control over the weak learners. The library interface was constructed to be similar to Scikit-learn's gradient boosting API so that the newly developed boosting algorithms can stand as drop in replacements for those available in Scikit-learn.

The gradient boosting package *pCRTree* is available at `https://github.com/wgurecky/pCRTree.git`.

C.2 Copula Toolkit

For copula simulation, the CDvine toolkit (GPLv3 licensed) is available for the R programming language. This packages does not implement all rotations of copula making it burdensome to handle negative dependence structures out-of-the-box. Furthermore, the maximum likelihood fitting method included in CDVine does not allow the user to specify sample weights, a key feature for the CFD data under consideration since the CFD mesh cells vary in size.

To circumvent these deficiencies and potential license compatibility issues with VERA, a new copula toolkit was developed in python and is BSD3 licensed. Careful attention was paid to develop a flexible abstract copula class which enables custom copula functions to be specified. Importantly, all copula rotations are supported by default allowing one to model positive and negative dependence structures without duplication of code. Canonical vine-copula construction and sampling algorithms are included in this package to handle the decomposition of arbitrary joint density functions of any dimension. Copula parameters can be determined by a weighted maximum likelihood fit to empirically supplied data with included sample weights or by specifying a rank correlation coefficient in the case of Archimedean copula. In the current hi2lo work, both capabilities are leveraged.

The *StarVine* copula software package and documentation is available at `https://github.com/wgurecky/StarVine.git`.

C.3 Python Interfaces to Crud Codes

As part of this work, python interfaces were developed for both the legacy CASL crud tool known as MAMBA1D and the state-of-the art crud package, Mamba. The python wrappers to these Fortran codes facilitate rapid prototyping of hi2lo procedures which provide boundary conditions to the crud codes. Additionally, the high level interface simplifies the process of orchestrating large crud sensitivity studies.

The python wrappers are available in the Virtual Environment for Reactor Analysis (VERA) developed by CASL `https://www.casl.gov`.

C.4 Hi2lo Code

A package that leverages all the aforementioned tools to produce estimates of crud growth rates was developed. This high level package is the primary user facing result of the current work. It should be noted this package is heavily dependent on crud simulation, copula construction, and gradient boosting technologies. This package orchestrates the construction and evaluation of gradient boosted regression trees which provide the copula and marginal distribution parameters as a function of local core conditions. Currently, multi pin, multi state point simulation is implemented with future work focused on parallelization, training data acquisition, and improvements to the machine learning model implementation.

The hi2lo crud growth package and documentation is available at `https://github.com/wgurecky/crudBoost.git`.

C.5 Synthetic Training Data Generation

A toolkit to overlay custom noise atop a CTF solution was developed to provide a secondary source of training data sets aside from running a CFD code. The synthetic data generation tool provides training data sets with lower computational cost than CFD calculations. Some properties of a true CFD solution field are preserved by the tool, namely that the shape of the marginal and copula distributions change as a function of position and local thermal hydraulic conditions in the core. The synthetic data is not to be viewed as a complete substitute for CFD data since it lacks the ability to capture spatial autocorrelation in the predicted spatial fields that arise naturally from the governing PDEs. Neighboring points on the rod surface do not exchange any TH information in this tool. Despite the unphysical nature of the synthetic data, the tool provides a means to verify that known relationships between the explanatory variables and the copula parameters are recovered by the gradient boosted regression model. This is possible because the user specifies these relationships up-front as inputs to the surface field sampling routines.

An excerpt of an input to generate a synthetic single pin data set is given below:

```
{
    "pinID": 1,
    "chanID": 1,
    "averageHeatFlux": 1.2e6,
    "spans": {
                "0.0": {"model": "lower", "samples": 1000},
                "2.01": {"model": "upper", "samples": 4000},
                "2.53": {"model": "upper", "samples": 4000},
                "2.98": {"model": "upper", "samples": 4000}
    },
    "upper": {
            "0.0": {"copula":  {"family": "gauss", "params": [-0.5], "rot": 0},
                "tke": {"type": "gauss", "params": [0.001, 0.02]},
                "temp": {"type": "beta", "params": [5.0, 2.7], "loc": -9.2, "scale": 12.0},
                "bhf": {"type": "gauss", "params": [0.001, 2.6e4]}
                },
            "0.3": {"copula":  {"family": "gauss", "params": [-0.6], "rot": 0},
                "tke": {"type": "gauss", "params": [0.01, 0.008]},
                "temp": {"type": "beta", "params": [5.0, 1.7], "loc": -7.0, "scale": 8.0},
                "bhf": {"type": "gauss", "params": [0.01, 1.1e4]}
                },
            "1.0": {"copula":  {"family": "frank", "params": [4.0], "rot": 1},
                "tke": {"type": "gauss", "params": [0.01, 0.005]},
                "temp": {"type": "beta", "params": [5.0, 1.5], "loc": -4.0, "scale": 5.0},
                "bhf": {"type": "gauss", "params": [0.01, 0.9e4]}
                }
            },
    "lower": {
            "0.0": {"copula":  {"family": "gauss", "params": [-0.6]},
                "tke": {"type": "gauss", "params": [0.001, 0.0001]},
                "temp": {"type": "beta", "params": ["5.0*(t)/600.0", 5.0], "loc": -2.0, "scale": 4.0},
                "bhf": {"type": "gauss", "params": [0.01, 1.0e3]}
                },
            "1.0": {"copula":  {"family": "gauss", "params": [-0.6]},
                "tke": {"type": "gauss", "params": [0.001, 0.0002]},
                "temp": {"type": "beta", "params": [5.0, 5.0], "loc": -2.0, "scale": 4.0},
                "bhf": {"type": "gauss", "params": [0.01, 1.0e3]}
                }
            }
}
```

The synthetic data generation tool is available for download at `https://github.com/wgurecky/ctfpurt.git`

D | Rod Surface Heat Transfer

D.1 Subcooled Boiling and DNB

The relationship between the surface temperature of an internally heated object and the heat flux from the surface into the surrounding fluid is shown in figure D.1.

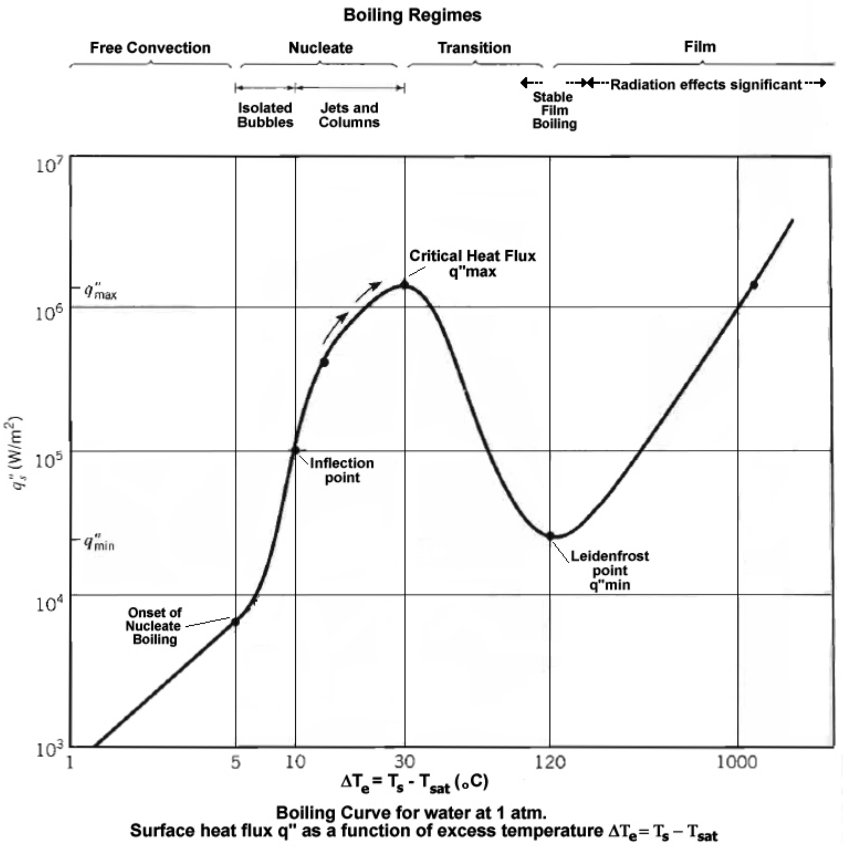


Figure D.1: Boiling curve.

The curve can be approximated by equation D.1. Note that surface temperature T_s is equivalent to the wall temperature T_w in the equations which follow. The critical heat flux (CHF) is the point at which film boiling begins to dominate and is accompanied by a precipitous drop in the heat transfer and a rise in the surface temperature. This condition is known as departure from nucleate boiling (DNB) and must be avoided when operating a PWR.

$$q''(T_w) = \begin{cases} h(T_w - T_\infty), & \text{if } T_w < T_{sat} \\ h(T_w - T_\infty) + q''_{nb}, & \text{if } T_{sat} \leq T_w < T_{CHF} \end{cases} \tag{D.1}$$

Where h is the single phase convective heat transfer coefficient which is in turn a function of the Nusselt number given in equations D.3 and D.4. The contribution of nucleate boiling to the heat transfer can be approximated by the Rohsenow model given in D.2 [60].

$$q''_{nb} = \mu_L h_{fg} \left[\frac{g\,(\rho_L - \rho_v)}{\sigma} \right]^{1/2} \left[\frac{c_{pL}\,(T_w - T_{sat})}{C_{sf} h_{fg} Pr_L^n} \right]^3 \tag{D.2}$$

Where h_{fg} is the latent heat of vaporization, μ_L is the liquid viscosity, ρ_v, ρ_L are the vapor and liquid phase densities, c_{pL} is the specific heat of the liquid phase, and C_{sf} is a tunable empirical constant.

$$h = \frac{k_l \mathrm{Nu}}{L} = \frac{q''}{T_w - T_\infty} \tag{D.3}$$

Where k_l is the thermal conductivity of the liquid, L is the characteristic length scale, and Nu is the Nusselt number. For non-boiling flows over a flat vertical surface, the Nusselt number can be approximated by the Dittus-Boelter equation:

$$\mathrm{Nu} = 0.023\,\mathrm{Re}^{4/5}\,\mathrm{Pr}^n \tag{D.4}$$

Where Re is the Reynolds number and Pr is the Prandtl number. n is an empirically derived constant and is typically 0.4 for a heated flow.

Bibliography

[1] P. Turinsky and D. Kothe. Update on capabilities development at CASL. In *International Congress on Advances in Nuclear Power Plants*. American Nuclear Society, May 2015.

[2] B. Collins, R.K. Salko, S. Stimpson, K. Clarno, and A. Godfrey. Simulation of Crud Induced Power Shift using the VERA Core Simulator and MAMBA. In *PHYSOR 2016*, 2016.

[3] T. Lange. *Methodology for an Advanced Risk Assessment of Crud Induced Power Shift using Coupled Multi-Physics Simulations and a Monte Carlo Scenario Analysis of the Potential Financial Benefits*. PhD thesis, University of Tennessee, Knoxville, 2017.

[4] R. Adamson, B. Cox, A. Strasser, and P. Rudling. Corrosion mechanisms in zirconium alloys. Technical report, Advanced Nuclear Technology International, 2007.

[5] B. Collins and J. Galloway. *Mongoose Methods and Theory*. Consortium for Advance Simulation of Light Water Reactors.

[6] J. Henshaw, J.C. McGurk, H.E. Sims, A. Tuson, S. Dickinson, and J. Deshon. A model of chemistry and thermal hydraulics in PWR fuel crud deposits. *Journal of Nuclear Materials*, 353:1–11, 07 2006.

[7] S. Slattery and W. Gurecky. Support for CILC L1 Milestone Using STAR-CCM+. Technical Report CASL-U-2016-1237-000, L3:PHI.CMD.P12.02, Consortium for Advanced Simulation of Light Water Reactors, 2016.

[8] B. Kendrick, V. Petrov, D. Walker, and A. Manera. CILC studies with comparative analysis to existing plants. Technical Report CASL-U-2013-0224-000, 2013.

[9] R.K. Salko and M.N. Avramova. *CTF Theory Manual*. The Pennsylvania State University.

[10] R.L. Wilby, T.M. Wigley, D. Conway, P.D. Jones, B. Hewitson, J. Main, and D.S. Wilks. Statistical downscaling of general circulation model output: A comparison of methods. *Water Resources Research*, 34, 1998.

[11] A. Werner and A. Cannon. Hydrological extremes—an inter comparison of multiple gridded statistical downscaling methods. *Hydrology and Earth System Sciences*, 20, 2016.

[12] A. Goly, R. Teegavarapu, and A. Mondal. Development and evaluation of statistical downscaling models for monthly precipitation. *Earth Interactions*, 18, 2014.

[13] A. Wood, A. Kumar, and D.P. Lettenmaier. Long range experimental hydrologic forecasting for the eastern U.S. *Journal of Geophysical Research*, 107, 2002.

[14] D. Sharma, A. Das Gupta, and M.S. Babel. Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping river basin, thailand. *Hydrology and Earth System Sciences*, 11(4):1373–1390, 2007.

[15] M.A. Ben Alaya, F. Chebana, and T.B.M.J. Ouarda. Probabilistic gaussian copula regression model for multisite and multivariable downscaling. *Journal of Climate*, 27(9):3331–3347, 2014.

[16] H.-N.S. Chin P. Caldwell, D.C. Bader, and G. Bala. Evaluation of a wrf dynamical downscaling simulation over California. *Climatic Change*, 95(3):499–521, August 2009.

[17] M. Avramova. *Development of an Innovative Spacer Grid Model Utilizing Computational Fluid Dynamics Within a Subchannel Analysis Tool.* PhD thesis, The Pennsylvania State University, 2007.

[18] S.C. Yao, L.E. Hochriter, and W.J. Leech. Heat-transfer augmentation in rod bundles near spacer grids. *Journal of Heat Transfer*, 1982.

[19] R.K. Salko, W. Gurecky, S. Slattery, K. Clarno, D. Pointer, D. Walker, and V. Petrov. Implementation of a Grid Heat Transfer and Turbulent Kinetic Energy Hi2Lo Remapping Capability into CTF in support of the CIPS Challenge Problem. Technical Report CASL-U-2017-1322-000, Consortium for Advanced Simulation of Light Water Reactors, 2017.

[20] T. Blyth. CFD-informed spacer grid model implementation in cobra-tf. Technical Report CASL-U-2014-0131-000, Consortium for Advanced Simulation of Light Water Reactors, 2014.

[21] T. Blyth. *Development and Implementation of CFD-Informed Models for the Advanced Subchannel Code CTF.* PhD thesis, The Pennsylvania State University, 2017.

[22] D.G. Krige. A statistical approach to some mine valuations and allied problems at the Witwatersrand. Master's thesis, University of Witwatersrand, 1951.

[23] D.G. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52:119–139, 1951.

[24] C. Williams and C. Rasmussen. Gaussian processes for regression. *Advances in neural information processing systems*, pages 514–520, 1996.

[25] Timothy W Simpson, Timothy M Mauery, John J Korte, and Farrokh Mistree. Kriging models for global approximation in simulation-based multidisciplinary design optimization. *AIAA journal*, 39(12):2233–2241, 2001.

[26] S. Jeong, M. Murayama, and K. Yamamoto. Efficient optimization design method using kriging model. *Journal of aircraft*, 42(2):413–420, 2005.

[27] W. Hsu, P. Huang, C. Chang, C. Chen, D. Hung, and W. Chiang. An integrated flood risk assessment model for property insurance industry in taiwan. *Natural Hazards*, 58(3):1295–1309, 2011.

[28] T. Hengal, B. Heuvelink, and B. Rossiter. About regression-kriging: From equations to case studies. *Computers and Geosciences*, 33, 2007.

[29] J. Li, A.D. Heap, A. Potter, and J.J. Daniell. Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling and Software*, 26(12):1647–1659, 2011.

[30] J. Li, B. Alvarez, J. Siwabessy, M. Tran, Z. Huang, R. Przeslawski, L. Radke, F. Howard, and S. Nichol. Application of random forest, generalised linear model and their hybrid methods with geostatistical techniques to count data: Predicting sponge species richness. *Environmental Modelling and Software*, 97:112–129, 2017.

[31] C. Rasmussen. Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer, 2004.

[32] E. Snelson. Tutorial: Gaussian process models for machine learning. *Gatsby Computational Neuroscience Unit, UCL*, 2006.

[33] M. Geidosch, M. Fischer. Application of vine copulas to credit portfolio risk modeling. *Risk and Financial Management*, 9, 2016.

[34] T. MacKenzie, D. Spears. The formula that killed Wall Street: The Gaussian copula and modeling practices in investment banking. *Social Studies of Science*, 44, 2014.

[35] X. Li. On default correlation: A copula function approach. *Journal of Fixed Income*, 9, 2000.

[36] J. Dupuis. Using copulas in hydrology: Benefits, cautions, and issues. *Journal of Hydraulic Engineering*, 12, 2007.

[37] M. Ganguli, P. Reddy. Probabilistic assessment of flood risks using trivariate copulas. *Theoretical and Applied Climatology*, 111, 2012.

[38] D. Kelly. Using copulas to model dependence in simulation risk assessment. In *Proceedings from International Mechanical Engineering Congress and Exposition in Seattle, WA, USA*. ASME, November 2007.

[39] R. Jouini, M. Clemen. Copula models for aggregating expert opinions. *Operations Research*, 44, 1996.

[40] H. Joe. *Dependence Modeling with Copulas.* CRC Press, 2015.

[41] T. Bedford and R.M. Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32, 2001.

[42] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8, 1959.

[43] R. Nelsen. *An Introduction to Copulas.* Springer, 2006.

[44] P. Barbe, C. Genest, K. Ghoudi, and B. Remillard. On Kendall's process. *Journal of Multivariate Analysis*, 58, 1996.

[45] F. Fritsch and R. Carlson. Monotone piecewise cubic interpolation. *Numerical Analysis*, 1980.

[46] P. Hall. *The Bootstrap and Edgeworth Expansion.* Springer, 1997.

[47] F. Mosteller. On some useful inefficient statistics. *Ann. Math. Statist.*, 17(4):377–408, 12 1946.

[48] J.H. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: A statistical view of boosting. *Annals of Statistics*, 28, August 2000.

[49] J.H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189, October 2001.

[50] J.H. Friedman. Stochastic gradient boosting. *Computational Statistics and Data Analysis*, 38:367, Febuary 2002.

[51] O. Chapelle and Y. Chang. Learning to rank challenge overview. *Machine Learning Research, Workshop and Conference Proceedings*, 14, 2011.

[52] S. Tyree, K. Weinberger, and K. Agrawal. Parallel boosted regression trees for web search ranking. In *Proceedings of the 20th International conference on World Wide Web*, January 2011.

[53] R.Y. Rubinstein and D.P. Kroese. *Simulation and the Monte Carlo Method.* 2011.

[54] A.B. Owen. *Monte Carlo Theory, Methods and Examples.* 2013.

[55] B.M. Boyerinas. Determining the statistical power of the Kolmogorov–Smirnov and tests via Monte Carlo simulation. Technical report, Center for Naval Analyses, 2016.

[56] H.J. Khamis. The delta-corrected Kolmogorov–Smirnov test for goodness of fit. *Journal of Statistical Planning and Inference*, 24(3):317–335, 1990.

[57] G.J. Babu and E.D. Feigelson. Astrostatistics: Goodness-of-Fit and All That! In C. Gabriel, C. Arviset, D. Ponz, and S. Enrique, editors, *Astronomical Data Analysis Software and Systems XV*, volume 351 of *Astronomical Society of the Pacific Conference Series*, page 127, July 2006.

[58] L. McInnes and J. Healy. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints*, February 2018.

[59] H.A. Chipman, E.I. George, and R.E. McCulloch. Bart: Bayesian additive regression trees. *Ann. Appl. Stat.*, 4(1):266–298, 03 2010.

[60] W.M. Rohsenow. A method of correlating heat transfer data for surface boiling of liquids. Technical report, Office of Naval Research, 1951.