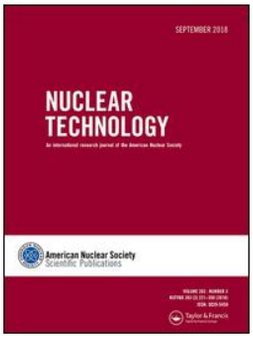

Nuclear Technology

ISSN: 0029-5450 (Print) 1943-7471 (Online) Journal homepage: https://www.tandfonline.com/loi/unct20

Machine-Learning Analysis of Moisture Carryover in Boiling Water Reactors

Haoyu Wang, Andrew Longman, J. Thomas Gruenwald, James Tusar & Richard Vilim

To cite this article: Haoyu Wang, Andrew Longman, J. Thomas Gruenwald, James Tusar & Richard Vilim (2019): Machine-Learning Analysis of Moisture Carryover in Boiling Water Reactors, Nuclear Technology

To link to this article: https://doi.org/10.1080/00295450.2019.1583957

Published online: 27 Mar 2019.

Submit your article to this journal

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=unct20

NUCLEAR TECHNOLOGY
© 2019 American Nuclear Society
DOI: https://doi.org/10.1080/00295450.2019.1583957

Machine-Learning Analysis of Moisture Carryover in Boiling Water Reactors

Haoyu Wang,[a] Andrew Longman,[b,c] J. Thomas Gruenwald,[b,c] James Tusar,[d] and Richard Vilim[a]*

[a]*Argonne National Laboratory, Nuclear Science and Engineering Division, 9700 S. Cass Avenue, Lemont, Illinois 60439*
[b]*Purdue University, Department of Physics and Astronomy, 525 Northwestern Avenue, West Lafayette, Indiana 47907*
[c]*Blue Wave AI Labs, 233 Goldenrain Drive, Suite 304, Celebration, Florida 34747*
[d]*Exelon Generation, Nuclear Fuels, 200 Exelon Way, Kennett Square, Pennsylvania 19348*

Received October 17, 2018
Accepted for Publication February 14, 2019

Abstract — *Moisture carryover (MCO) is modeled in the General Electric Type-4 boiling water reactor (BWR) using machine-learning methods and data from operating plants. Understanding MCO and the conditions that give rise to an elevated value is important since excessive MCO can damage critical turbine components, can result in elevated dose levels to on-site personnel, and can interfere with late-cycle power management. The analysis of MCO takes into account simplifying reactor symmetries and important geometric dependencies. The plant data are taken from several reactors and were collected over multiple years and multiple fuel cycles. A brief description of the origin of MCO in U.S. BWR plants is given. A machine-learning model is constructed from the data using applicable algorithms and data-reduction techniques. Matching model complexity with available data is one of the more challenging machine-learning tasks. Too many features and too little data will lead to overfitting. The data for each fuel cycle included over 6876 original features, 9 for each fuel bundle. Two approaches are used to reduce the data set into a manageable number of features. The first was an engineering analysis that resulted in the selection of steam quality Q and steam liquid phase velocity V_L as the main features driving MCO. Using a Q and a V_L for each fuel bundle gives 1528 Q and a V_L feature describing the reactor behavior. An analysis of different functional forms of these two variables led to the actual inputs to the neural network model. The second approach involved the use of statistical techniques such as Pearson's correlation and k-means analysis. The identified groupings of bundles behaved similarly. Treating each grouping as a single feature further reduced the input variable set to a manageable number. A model selection criterion is proposed, and results are presented along with a discussion of related issues.*

Keywords — *Boiling water reactor, moisture carryover, data analytics, machine learning, neural network.*

Note — *Some figures may be in color only in the electronic version.*

I. INTRODUCTION

The boiling water reactor (BWR) operates on a saturated steam cycle relying on mechanical equipment to remove liquid moisture from a two-phase mixture that exits the core. Excess moisture carryover (MCO) can lead to both intergranular stress corrosion cracking of turbine blades and an increase in exposure levels of maintenance personal to ^{60}Co that is carried over in the liquid droplets that leave the steam dryers.[1] While the BWR in the United States was originally designed to achieve MCO levels that are acceptable from a plant economic performance perspective, changes in recent years in the way these plants are operated and in the reactor core design efficiencies have led to increased MCO levels.[1] Specifically, new fuel management strategies and the use of increased core flow (>100%) to extend the cycle aimed at minimizing fuel costs have led to a greater fraction of liquid reaching the

*E-mail: rvilim@anl.gov

turbine as a consequence of the increased load on the steam separators and steam dryers.

The acceptable level of saturated water entering the power conversion system from the reactor vessel is primarily a problem in economics. While the new fuel management/core design strategies have reduced fuel costs, an associated increase in MCO raises operating costs as a result of greater personnel exposure and turbine damage.

In the United States, two technical solutions are presently being pursued for better managing MCO under these new operating regimes. The first is to replace the steam separators and steam dryers with higher-capacity units. This can make economic sense for a plant granted an operating license extension or power uprate, and hence, the opportunity to better offset the equipment and installation costs with a longer and increased revenue stream. In the case where license extension is uncertain, the operation of the reactor over the cycle is being examined with the goal of better optimizing operating conditions. This, however, requires an improved understanding of how MCO depends on operating conditions. That is the subject of this paper.

This paper describes the use of machine-learning and data analysis methods to characterize MCO as a function of operating condition. This data-driven approach to modeling takes advantage of operational data to learn the operating characteristics of the reactor. A data-driven model can potentially yield superior performance compared to a physics-based model as the latter invariably involves approximations. On the other hand, the data-driven approach of this work is limited primarily by its interpolation nature and is not valid for making process predictions outside the range of the training data set. For the particular application examined in this paper, the data set is in fact well posed for modeling MCO level over a wide range of reactor operation parameters. That is, the data set does provide a reasonable basis for learning MCO dependence in a way that is useful to better manage MCO through reactor operation and to utilize the information when performing core design to alter reactor conditions to manage MCO. This, in combination with power uprating and cycle extension techniques, provides a powerful basis upon which to improve economic and operational performance in the presence of license extension uncertainty.

II. DESCRIPTION OF THE BWR

Approximately a third of commercial nuclear reactors in the United States are BWRs. The main characteristic of a BWR compared to a pressurized water reactor is the presence of significant boiling within the reactor core with the steam going directly to the turbine after exiting the vessel. The structure, operation, and fuel management of a BWR described in this section can be found in Refs. 2, 3, and 4.

II.A. Nuclear Steam Supply System

A typical nuclear steam supply system in a BWR is shown in Fig. 1. A steel pressure vessel contains the core where pressurized water flowing over uranium-bearing fuel pins absorbs the heat of fission. The coolant is maintained at a pressure that allows controlled boiling to occur in the core. The resulting two-phase mixture that exits the core passes through mechanical equipment to remove the liquid phase, which is recirculated back to the inlet of the core, leaving saturated vapor to exit the vessel and enter the main steam line.

The removal of the liquid phase from the two-phase mixture exiting the core is accomplished by passing through two equipment units in succession. The steam separators located above the core outlet establish a vortex in the two-phase mixture flowing out of the core. The induced radial velocity component causes the liquid phase to separate from the vapor phase where the former is collected at the periphery of the separators, as shown in Fig. 2. The mixture exiting the separators is saturated vapor with suspended liquid droplets. It enters the steam dryers where the droplets are forced to make a series of rapid changes in direction while traversing the platelike structures resulting in separation from the mixture, as shown in Fig. 3. The combined steam separators and dryers are designed to yield a steam quality of greater than 99.9% for admission to the turbine.[2]

II.B. Principles of Operation

The at-power thermal-hydraulic and neutronic conditions in the core are actively managed by controlling four process variables: (1) coolant flow rate via jet pumps, (2) coolant inlet temperature via feedwater heaters, (3) pressure via feedwater pumps, and (4) external reactivity via control rods. When each of these variables is maintained at a set point, the core will stabilize at an equilibrium condition at which the reactivity components associated with structure, coolant, fuel, and control rods sum to zero. The thermal power and core exit steam quality at the full-power design condition are set by appropriately adjusting these control variables.

Over a power-generating cycle the reactivity decreases and must be compensated for if power is to be maintained. There is a succession of operational phases that the reactor

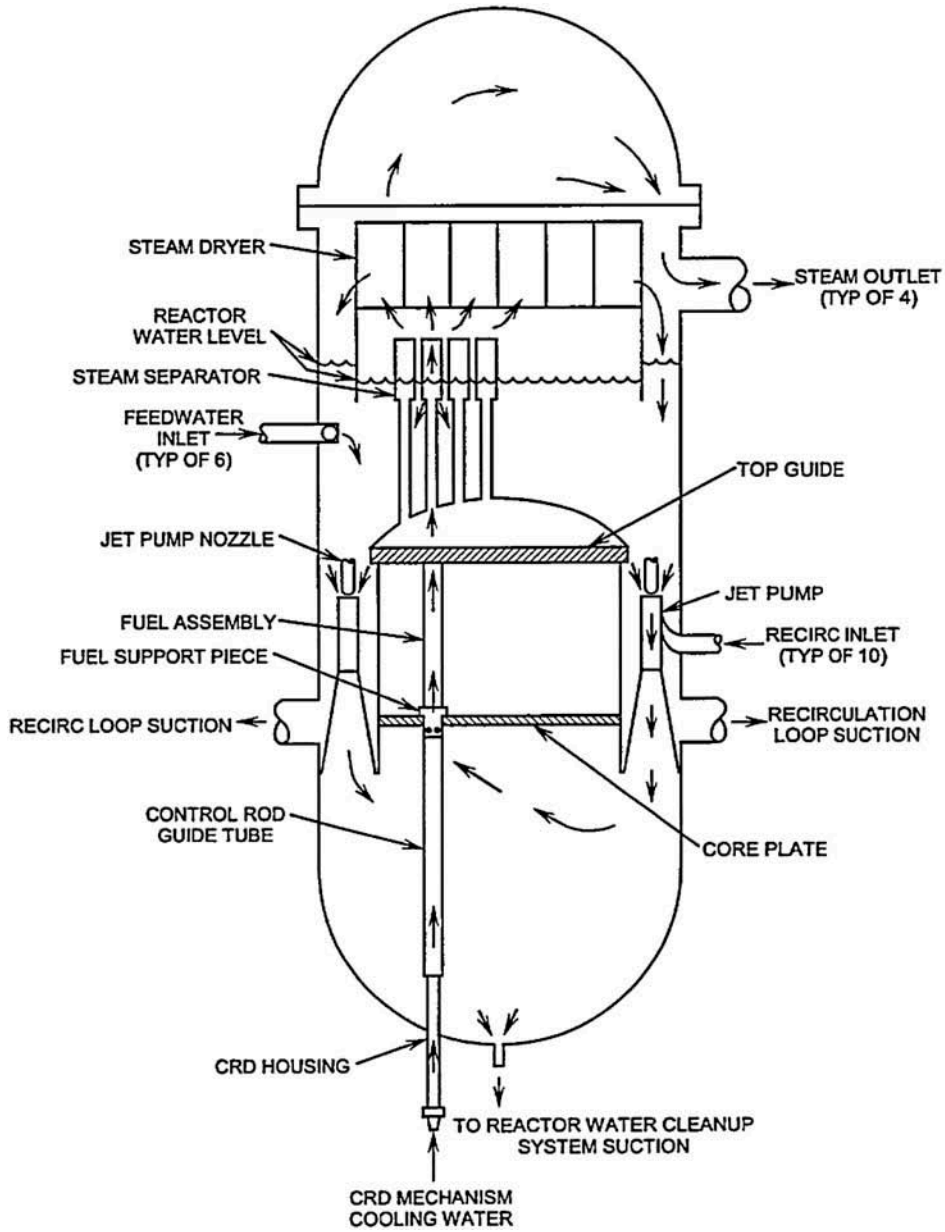

Fig. 1. Nuclear steam supply system in BWR.[2]

passes through to maintain power over a cycle. At initial fuel loading, excess reactivity is suppressed by gadolinium poison which is burnt out over the first part of the cycle. Second, control rods are slowly removed until completely withdrawn, and toward the end of this phase reactor flow is increased to reduce coolant voiding, which increases reactivity by increasing coolant density. Next, feedwater heaters are successively taken off line to lower core inlet temperature, which adds reactivity via coolant density. Finally, the reactor enters the coastdown phase where there are no further control actions and power decreases as fuel reactivity is lost with burnup.

II.C. Core Fuel Management and MCO

A current fuel management strategy involves managing the arrangement of fresh and burnt bundles in the core lattice to reduce neutron leakage at the periphery of the core. The reload strategy is to place twice-burnt bundles (with the lowest reactivity) in the outer rings of core lattice positions to serve as reflectors and thus reduce neutron leakage.[3,4] To maintain an acceptable steam exit quality in these low-power bundles, flow orificing is used to throttle the coolant flow. The power in these peripheral bundles can be as low as one-tenth of that of the interior bundles.

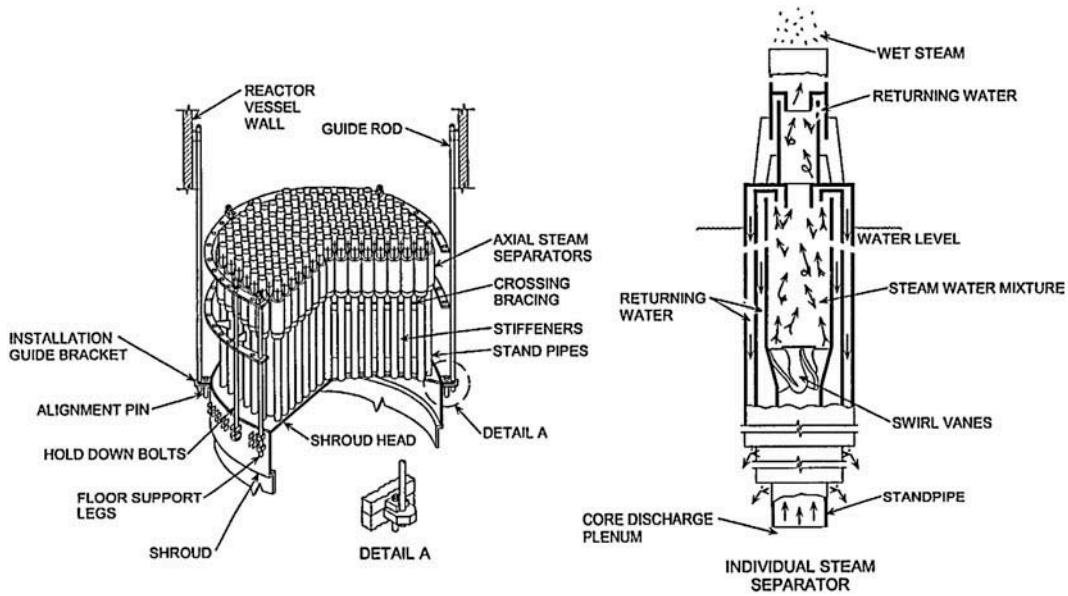

Fig. 2. Steam separator of BWR/4 reactor.[2]

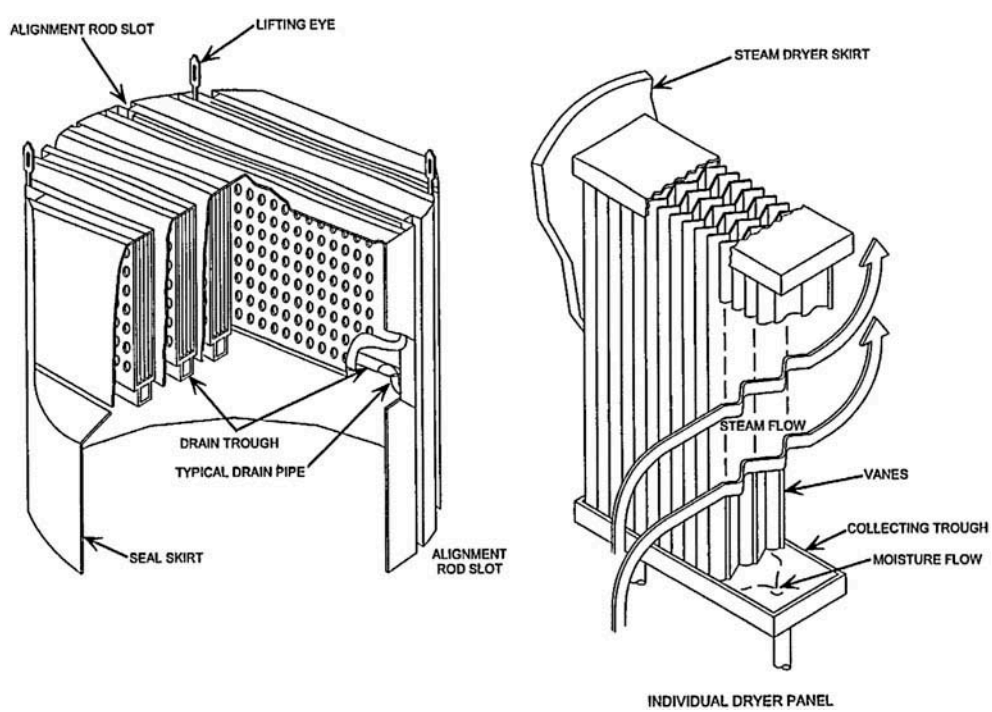

Fig. 3. Steam dryer of BWR/4 reactor.[2]

The inner region of the core is populated by a combination of fresh bundles and once-burnt bundles. A flat radial power distribution is maintained through a combination of control rod motion over the cycle and the burnup of gadolinium poison in selected fuel pins.

The problem of MCO becomes an operational issue with the introduction of new fuel management strategies. Twice-burnt bundles in the outer rings with a lower power-to-flow ratio than interior bundles have a lower steam exit quality, resulting in an elevated mass flow rate of saturated liquid leaving these bundles. This can overload the moisture separator and steam dryer units causing an elevation in the fraction of liquid water at the inlet to the turbine, or as referred to in the industry, elevated MCO. The problem is more pronounced if a group of low exit–quality bundles happen to be directly below a steam separator.

III. MOISTURE CARRYOVER

In a BWR a small amount of liquid water is entrained in the steam flow to the turbine giving rise to MCO. MCO is typically cited as the mass percentage of liquid phase in the steam flow entering the turbine. For U.S. commercial BWRs the MCO value during normal full-power operation is to be maintained below 0.1% (Ref. 5). The means for measuring MCO are described in Refs. 3 and 6. Briefly, the coolant samples are first obtained from both the coolant condensate system and the reactor vessel steam/water interface at a steady operating condition, with the interval between sampling times minimized. The ^{24}Na activity of both samples are then measured using a high purity germanium gamma-ray detector system, from which the concentration of ^{24}Na in both samples are calculated after the decay correction. The MCO is finally calculated as the ratio of two ^{24}Na concentration values.

The models and experimental data used originally to design current-generation BWR moisture separators and dryers are proprietary and date back to the 1970s. The fluid-flow phenomena in these two components, while being multidimensional and multiphysics, were represented using one-dimensional models with supporting experiments performed to characterize separate effects and to provide integral validation. Undoubtedly, some excess capacity was introduced in these units to cover uncertainties. But given the cost premium on in-reactor vessel components, it is unlikely there was capacity added beyond this. Importantly, in this paper the machine-learning approach places no dependence on this original design work. Instead, in its place are recent measurement data from multiple operating plants and an associated computer code described in Sec. IV.

III.A. Causes of MCO

While the steam separators and steam dryers are designed to provide name-plate MCO efficiency at the full-power condition, there are events that can increase MCO. A core flow imbalance caused by a jet pump failing to provide rated flow can skew bundle flow across the core and result in an MCO increase. A depression in local power caused by a control blade will decrease bundle quality and increase MCO if a steam separator is located above. Or during maintenance on a control rod hydraulic control unit, the blade may be fully inserted again creating local low-quality conditions.

But more consequential than the above events, which tend to be infrequent, are the effects of recent changes in how some U.S. utilities are operating their current fleet of BWRs compared to when these plants first came online. Improved fuel management/core design strategies have resulted in increased discharge burnups and have caused high MCO. Fuel loading plans that place high-burnup and low-power bundles in the core periphery result in lower quality steam in these bundles; see Fig. 4. The challenge is to map out MCO behavior in this new regime and adjust operating conditions so that the MCO level is acceptable while maintaining power.

III.B. Impact

Wet steam at elevated temperature can cause corrosion of turbine blades necessitating more frequent maintenance.

Further, the moisture droplets transport ^{60}Co from the reactor vessel to the turbine, which results in higher dose rates in the power conversion system. The ^{60}Co originates from the corrosion of elemental cobalt at the jet pumps that becomes ^{60}Co after irradiation by neutrons and subsequent neutron capture.[1] Increased exposure for maintenance personnel has an economic cost and human risk making a plant less competitive.

III.C. Management

The objective is to understand quantitatively how MCO correlates with reactor operating conditions so as to be able to perform an economic trade-off and risk analysis for various core-loading patterns and control rod patterns with the associated fuel costs, personnel exposure, and asset management of the turbine with increased MCO. With knowledge of the relationship one can perform an economic optimization, yielding a fuel loading plan and a reactor operating plan for a proposed power-generating cycle. Additionally, over the course of that cycle, continuous refinements can be made to the operating plan based on the feedback obtained from periodic MCO measurements. This real-time plant data can be used to further refine the model for MCO as a function of operating condition for use in future run plans.

IV. PLANT DATA

During the power-generating cycle, the BWR core process variables are calculated on a daily basis via General Electric's (GE's) proprietary three-dimensional BWR core simulation package PANACEA. The PANACEA package reads in the measured values of the core forcing functions, including control rod pattern, thermal power level, coolant

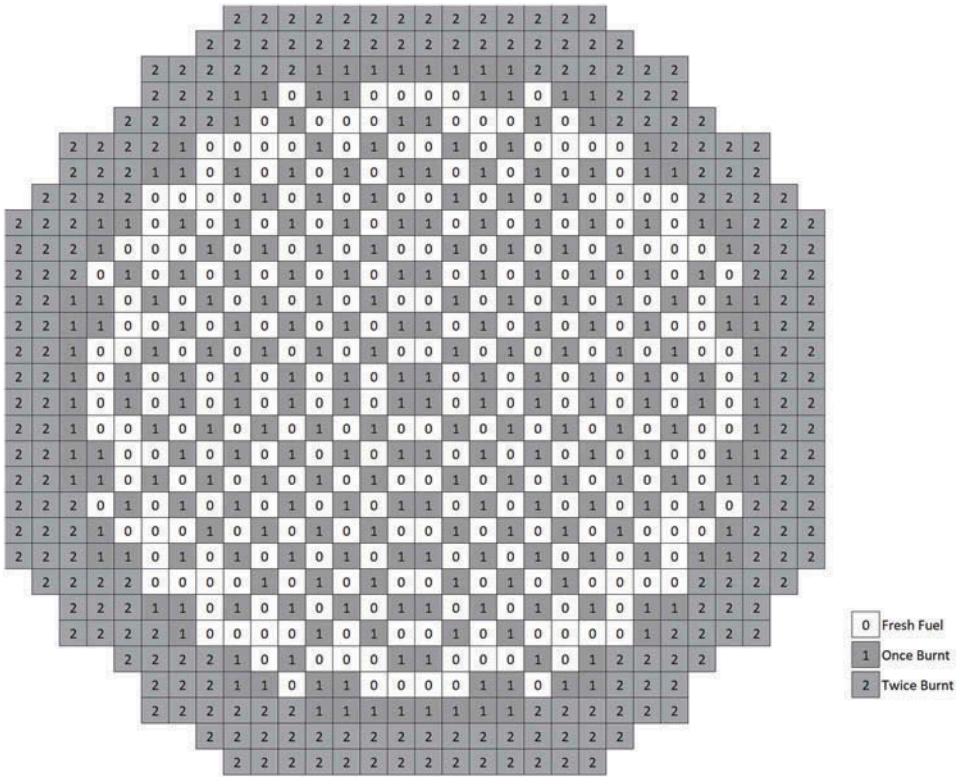

Fig. 4. Core loading plan with twice-burnt fuel in the periphery.[3]

flow rate and pressure, and feedwater temperature from the sensors indicated in Fig. 5. It then calculates the process variables for each fuel bundle including steam quality, void fraction, power distribution, and coolant flow rate distribution, as shown in Fig. 6.

There are nine bundle parameters as shown in Table I. There are 764 bundles for each reactor, thus totaling 6876 variables for each fuel cycle of a reactor. In addition, the core status over the planned cycle is calculated as part of the reactor run plan using the planned values for the core forcing functions of pressure, temperature, flow rate, and rod reactivity. The reactor run plan thus details how the reactor will be operated over a planned power-generating cycle.

Moisture carryover is a complicated and expensive radiochemistry measurement and so is limited to being performed on a weekly basis at the Limerick Generating Station. Thus, only a fraction of the daily PANACEA runs has an accompanying MCO measurement for use in supervised learning. The data used in this paper include 540 MCO measurements.

The current practice of taking MCO measurements on a periodic basis combined with the fact that MCO tends to increase late in the cycle, as seen in Fig. 5, results in a preponderance of MCO data points below the 0.1% high threshold. But from a machine-learning standpoint one prefers the opposite, that is, a greater fraction of the data points at high MCO values as this is the region where model prediction accuracy is most valuable.

V. FEATURE SPACE

In engineering applications, machine learning can be used to encapsulate data from a physical system in a way that captures the underlying processes in model parametric form enabling system behavior to be represented. The input-output relationships of such a model are useful for predicting system behavior as driven by system input variables that are to be manipulated to achieve a system performance objective. In the MCO problem, the performance objective is to optimize the core design to minimize fuel costs while also managing MCO (the output) by appropriately regulating the nuclear and thermal-hydraulic processes through plant actuators and fuel loading (the inputs).

In machine learning, the groupings of inputs deemed to capture relevance are referred to as "features," while the collection of candidate features is referred to as "feature space." The model developer

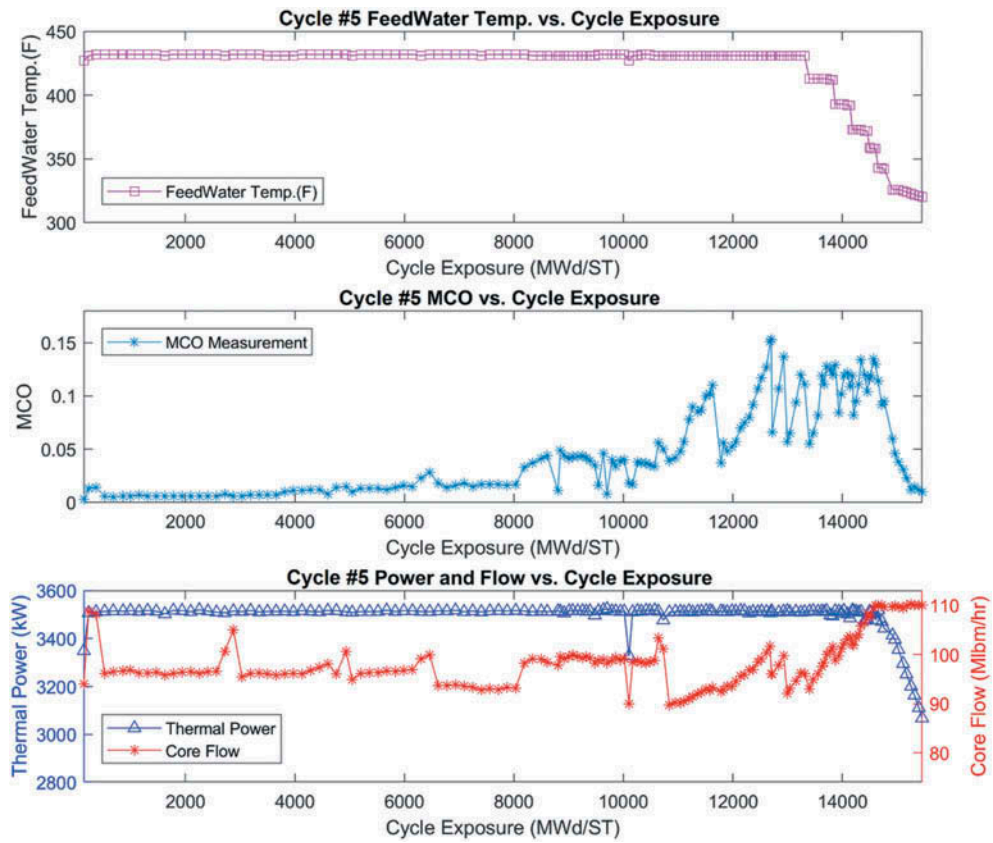

Fig. 5. Typical forcing functions and MCO values during a cycle.

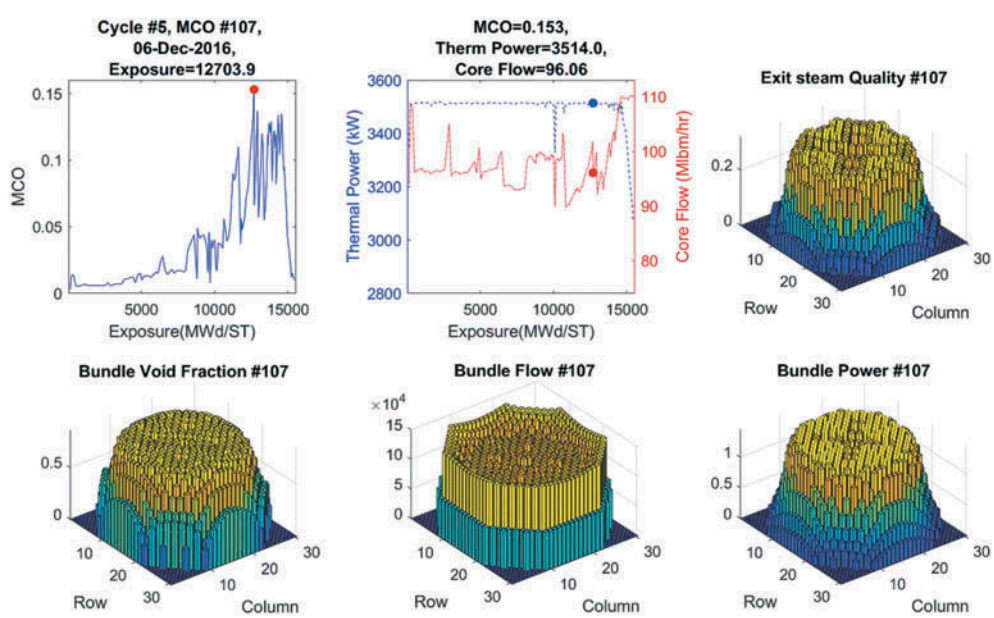

Fig. 6. Some output parameters from each fuel bundle calculated by PANACEA package.

must choose a set of features that is appropriate to the problem, with the features that force the physical system being the desired choice. However, not all individual engineering variables should necessarily be included as inputs as described below. The data at hand include 6876 distinct initial engineering variables. We discuss our approach to reducing the data set to a manageable number of features. We show results

TABLE I

Bundle Parameters Simulated in PANACEA Package

Name	Description	Unit
Bundle power	Integrated thermal power of each bundle	Normalized
Bundle flow	Coolant mass flow rate through each bundle	Pounds of steam mixture/h
Exit steam quality	Steam quality (mass flow rate fraction of vapor to steam mixture) when exiting each bundle from the top	1
Bundle void fraction	Steam void fraction (area fraction occupied by vapor) when exiting each bundle from the top	1
Average *k*-infinity	Average *k*-infinity value over the total length of each fuel bundle	1
Control rod pattern	Length of each control rod withdrawn from the reactor core	Notches
Top 1 slice nodal power	Power distribution in the first top layer of each fuel bundle divided into 25 virtual vertical layers	Normalized
Top 2 slice nodal power	Power distribution in the second top layer of each fuel bundle divided into 25 virtual vertical layers	Normalized
Top 3 slice nodal power	Power distribution in the third top layer of each fuel bundle divided into 25 virtual vertical layers	Normalized

using steam quality Q and mixture liquid phase velocity V_L and functions of these two variables to reduce the data features to a manageable number. The variables Q and V_L are features of each fuel bundle, so using this feature set initially reduces the number of variables to 1528. It should be noted that while at first blush the use of Q and V_L appears to be an over simplification, these quantities by nature carry spatial information about the reactor operation and thus can closely characterize its operation. This avoids models that are too complex for the available amount of data and saves computational effort.

In a mathematical sense, feature space contains the collection of inputs that force the system. In a physical sense there may be several distinct feature spaces. Table II identifies three such feature spaces for the MCO problem, each of which influences MCO and reactor power. Feature space FS1 contains the process variables that exit the core and are presented to the steam separators. Feature space FS2 contains process variables internal to the core that in turn manifest themselves as the quantities in FS1. Feature space FS3 contains the raw inputs such as actuators and fuel loading that influence FS2. This paper is focused primarily on FS1 in terms of model creation but does examine FS2 for the purpose of understanding dependencies in the former.

V.A. Feature Selection

The selection of features in engineering problems warrants special attention given that the data tend to be sparse, unlike problems found in the financial industry. This is the case for the MCO problem, which is described later. The challenge is to find the important inputs that affect MCO so as to restrict the number of inputs to avoid overfitting in the presence of limited-size data sets. This necessarily involves a trade-off between good model generalization capability and good model predictive accuracy. The appropriate trade-off in this work is analyzed through simulation studies where the choice of features is varied and the model predictive performance is evaluated. Sensitivity studies are performed to identify the degree to which feature-space variables affect the output. It is also important to identify and eliminate inputs that are linearly dependent. Such inputs unnecessarily increase the number of model parameters and, in analogy with a poorly conditioned matrix in linear algebra, increase the sensitivity of the model prediction to noise and error in the data.

V.A.1. Engineering Analysis

The engineering analysis approach to feature selection puts to use expert knowledge about the system and what is forcing it. This information, combined with modeling studies, is used to appropriately limit the number of inputs selected from feature space so that overfitting is avoided. This process, which is described conceptually here and put into practice below, makes evident the inherent limitations on what is achievable in terms of the quality and accuracy of the predictive power of the model. It also reveals how predictive performance is inextricably linked to the richness and diversity of the plant data set.

TABLE II

Three Candidate Feature Spaces for Modeling Moisture Separator Problem

Feature Space	Description	Feature
FS1: core outlet process variables	Bundlewise coolant conditions exiting the core	Bundle outlet coolant phase velocities V_L *and* V_G; void fraction α; quality Q; and densities
FS2: core internal process variables	Coupled nodalized thermal-hydraulic and neutronic conditions in the core	Bundle *k*-eff and control rod position; nodalized bundle power; coolant flowrate; void fraction; coolant density and quality
FS3: forcing functions and loading plan	Variables used to manage the reactor thermal performance and MCO	Control rod position; coolant flow rate; coolant inlet temperature; coolant pressure; fuel bundle loading

In the MCO problem, expert nuclear engineering knowledge was used to identify a candidate set of features. The variables leaving the core at the bundle level are related through the conservation of mass:

$$w \cdot \left(\frac{1-Q}{1-\alpha}\right) = \rho_L \cdot A \cdot V_L \,, \tag{1}$$

where

w = total mass flowrate

Q = flowing quality

α = void fraction

ρ_L = liquid density

A = cross-sectional flow area

V_L = liquid velocity.

It is assumed pressure is constant, as was verified by the utility for the data sets provided, and so thermophysical properties are constant.

There is an additional candidate variable in feature space FS1, vapor velocity V_G, and it is expressible in terms of the variables above through the relationship

$$Q = \frac{\rho_G \cdot A_G \cdot V_G}{\rho_G \cdot A_G \cdot V_G + \rho_L \cdot A_L \cdot V_L} \,, \tag{2}$$

where

$$A_L = (1-\alpha) \cdot A = A - A_G \,. \tag{3}$$

It is assumed that the MCO analysis is limited to a single plant type, in which case the geometry and dimensions of components are fixed. In this case the area variables above have constant values.

In order to find a set of features of limited dimension to mitigate overfitting issues, a set of physical arguments are introduced to reduce the list of above variables from w, Q, α, V_L, and V_G to a smaller number. The reasoning exercised is approximate. With data-driven systems, the lack of a physics-based model requires some judgement on the part of the analyst aided by analytical insight into the structure of the data.

To proceed, note that Eq. (1) implies an implicit dependence among the five variables that appear there. As such, any four of these can be regarded as independent variables that affect the physical system. The fifth should not be included in the input as its value is constrained by Eq. (1). This fifth variable can be thought of as representing overdetermined information or, in analogy with linear algebra, a colinearity. If it were to be included in the network at all, it should be treated as an output.

Further reduction in input feature space is achieved by examining the plant data and finding that for a very large fraction of bundles α is always contained in the range 0.3 and 0.4. Additionally, α trends with Q whose range of values is similarly limited. For small changes such as these, the dependence between these two variables will tend to be linear. And so, we elected to remove α from the input feature space. Similar arguments apply for V_L and V_G. We choose to remove V_G from the input feature space.

We are left with variables Q and V_L as feature-space inputs. If indeed the other three variables are important for learning MCO dependence, then a model trained on Q and V_L alone should have inferior generalization performance compared to a model expanded to include any of these three other inputs. This in fact was found not to be case; most MCO behavior can be described with Q and V_L with associated spatial and clustering information. For example, when V_G was additionally included in input feature space the model generalization performance was not as good as with Q and V_L

alone. Thus, the above physical and mathematical arguments for limiting feature space were born out in actual model training/testing trials.

Consider now the problem of how Q and V_L should appear at the input layer. It is noted that suitably transforming an input variable by input data manipulation is superior to leaving this job to the model trainer where data are limited and the type of network is a single hidden layer, feedforward. That is, it is prudent for the engineer to determine if there is a more natural functional form of the variable rather than simply the variable itself so as to minimize the burden placed on the model to modify features in automated fashion. To those ends we note that MCO should increase with a decrease in steam quality exiting the bundles, that is,

$$\text{MCO} \sim \frac{1}{Q^m}, (m > 0) \ , \tag{4}$$

so the preferred input to the model is the right side of Eq. (4).

Similarly, MCO may increase with a decrease in the velocity of the liquid component at the bundle outlet as the steam separators will be less effective, that is,

$$\text{MCO} \sim \frac{1}{V_L^n}, (n > 0) \ , \tag{5}$$

so the preferred input to the model is the right side of Eq. (5).

The above principles guided the treatment of feature space in the modeling studies reported in Sec. VII.

V.A.2. Correlation Analysis

A validation of the physics-based arguments that were introduced to arrive at the reduced dimension inputs described above was accomplished by performing a correlation analysis. A correlation analysis provides a cross-check means for identifying the colinearities that are hypothesized to exist in feature space FS1. The analysis was performed using the Pearson correlation method.[7] Briefly, a correlation coefficient of +1 or −1 indicates perfect positive or negative linear correlation between two variables, and a correlation coefficient of 0 indicates no obvious linear correlation.

Taking advantage of the pronounced octagonal symmetry in both the reactor core lattice and fuel loading pattern seen in Fig. 4, the core is split into eight sectors each covering 45-deg angle. The core process variables (w, Q, α, V_L, and V_G) distributed over the entire core are similarly split into the same eight sectors for the correlation study. This partitioning increases the number of training data points by a factor of 8 while reducing the number of unique bundle positions by a factor of 8, both of which work toward reducing overfitting.

As seen in Fig. 7a, the correlation coefficients between V_L and V_G are close to unity for the majority of fuel bundles, which means that these two process variables are highly correlated. A high correlation is also observed between Q and α, as seen in Fig. 7b. Such a high level of correlation means the corresponding variables are redundant, which agrees with the conclusion drawn from engineering analysis. Importantly, Q and V_L have a lower level of correlation as shown in Fig. 7c, indicating that these two variables are more independent from each other and therefore important for capturing MCO dependence. Note that the variable w is not shown here as it was eliminated as a dependent feature-space variable in Sec. V.A.1.

Having identified the number and types of degrees of freedom in feature space through a correlation analysis, the correlation between MCO and each of the presumed process inputs $1/Q$ and $1/V_L$ is calculated along with the correlation for related and presumed redundant parameters $1/\alpha$ and $1/V_G$. The resulting correlation coefficients are plotted in Fig. 8 as a function of bundle radial distance from the core center over all cycles. It is clear that the $1/Q$ and $1/V_L$ of peripheral bundles have high positive correlation with MCO and that $1/\alpha$ and $1/V_G$, respectively, replicate that and therefore do not contain significant new information. This agrees with the conclusion in Sec. III.A.

The correlation study and remarks are shown in Table III. In summary, both an engineering analysis and correlation analysis indicate that the input space should be limited to $1/Q$ and $1/V_L$.

V.B. Data Reduction

As discussed previously, to build a robust model for prediction, the number of input variables should be a canonical set to avoid modeling the noise in MCO data, which can result in overfitting. In addition to rejecting the redundant quantities $1/\alpha$ and $1/V_G$ as discussed above, the remaining $1/Q$ and $1/V_L$ of each bundle in the one-eighth sector contribute approximately 200 variables. This number is too high for the total 540 MCO points considering that a single hidden layer one-neuron feedforward model will use up all the degrees of freedom. The "effective" number of points may well be lower as about 70% of the data points are clustered together with similar conditions and low MCO values.

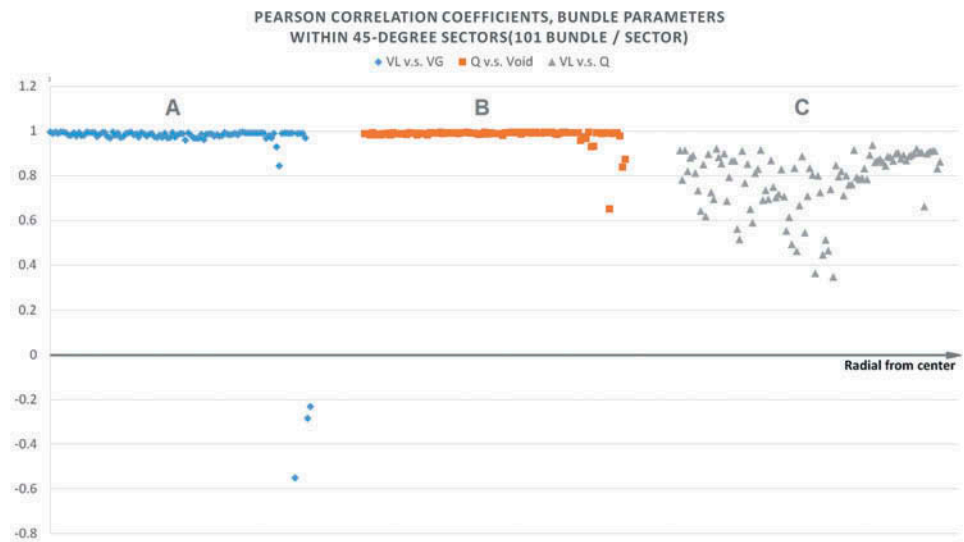

Fig. 7. Correlation between three different pairs of process variables with each shown as a function of distance from center of core.

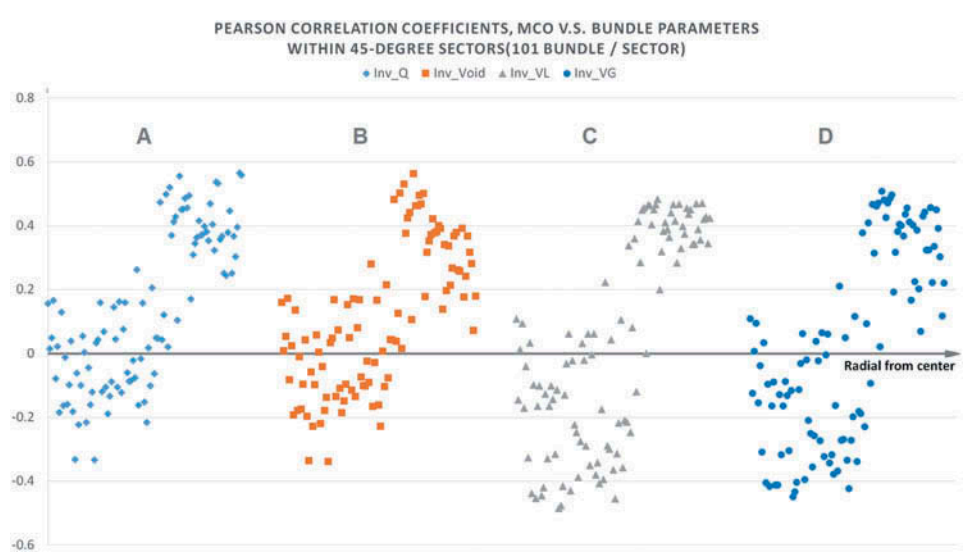

Fig. 8. Correlation between MCO and inverse process variables.

To discover the similarities between the fuel bundles, a k-means auto-clustering analysis was performed for each of the eight sectors. By analyzing the similarities of bundles in the $1/Q$ and $1/V_L$ domain, it is found that the bundle division stabilizes at three groups, as shown in Fig. 9. To capture additional physics behind the MCO related to steam separator spatial location,[5] the bundles were grouped more finely. Incorporating information on separator spatial location resulted in one of the best-performing grouping schemes, which is shown in Fig. 10. Thus, we are left with 14 features with which to describe the problem.

VI. MACHINE-LEARNING MODEL

VI.A. Network Architecture

The neural network architecture used is shown in Fig. 11. It is composed of one input layer that serves to normalize the multidimensional data, one hidden layer with multiple neuron nodes and their associated nonlinear activation functions, and a linear output layer. The neuron will manipulate the value coming from each input by first multiplying by a weight W and

TABLE III

Correlation Among Variable Pairs

Pair	Variables	Figure	Remarks
Input-Input	V_L and V_G	Fig. 7a	Highly correlated
	Q and α	Fig. 7b	Highly correlated
	V_L and Q	Fig. 7c	Weaker correlation
Input-Output	$1/Q$ and MCO	Fig. 8a	Accept as a canonical pair representing a degree of freedom
	1/α and MCO	Fig. 8b	Dependency similar to $1/Q$ and MCO; reject as representing a new degree of freedom
	$1/V_L$ and MCO	Fig. 8c	Accept as a canonical pair representing a degree of freedom
	$1/V_G$ and MCO	Fig. 8d	Dependency similar to $1/V_L$ and MCO; reject as representing a new degree of freedom

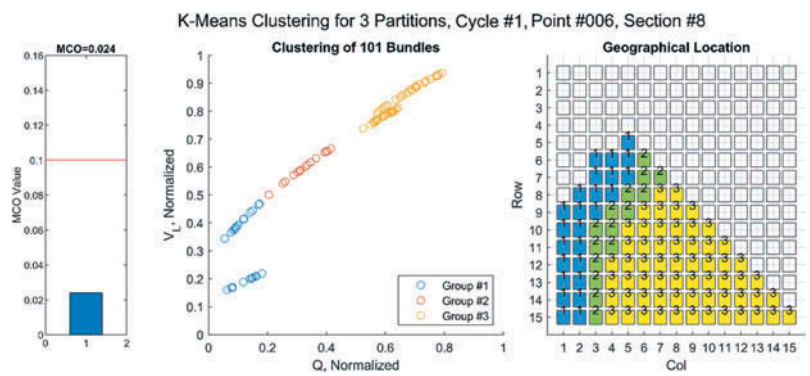

Fig. 9. *k*-means clustering for bundle parametric similarities.

then adding an offset b and feeding the result to the nonlinear activation function to approximate the nonlinearity. The weight and offset in the hidden layer will be optimized automatically by optimizing the cost function, which is also known as the "learning" process. We used a Levenberg-Marquardt backpropagation algorithm to adjust the ultimate neural network weights used in the final model.[8,9]

Certain elements of the neural network architecture are well suited to capture the nature of the MCO problem. For example, the MCO is a single physical quantity at the reactor vessel outlet measured by means of a radiochemistry experiment, and its value is the result of the performance of individual steam separators whose individual contributions sum in some nonlinear fashion. Presenting at the input spatial regions that reflect individual steam separators combined with automatic adjustment of the weights provides a nonlinear fitting capability suited for modeling the main phenomena providing that there is no measurement that could determine the contribution of MCO values from each core region.

VI.B. Data Division

The MCO data set was randomly divided into four logical subsets comprising training (70%), validation (10%), valid-2 (10%), and test (10%) for use as indicated by these names. Given these logical partitions and associated fraction of the MCO data set, a so-called training/validation/valid-2/test data set was created by populating these logical subsets with data from the MCO data set. The populating procedure involved selecting at random from the MCO data set. This process of creating the training/validation/valid-2/test data set was repeated again and again to create 2048 such data sets. A sensitivity study was performed by varying the data division fraction. The cost function (see Sec. VI.C) behavior was essentially unchanged for data divisions that varied the training data component from 43% to 88%. This indicates that the results we report for the 70–10–10–10 split described above are not sensitive to this particular split.

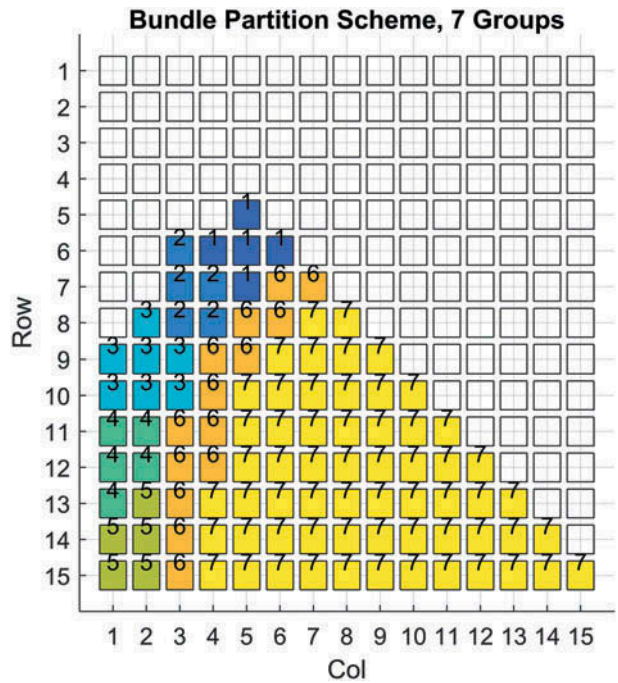

Fig. 10. One of the best performing bundle grouping schemes: 7 groups.

VI.C. Cost Function

A cost function, the mean square error (MSE), is defined for later use as

$$\mathrm{MSE} = \sum_{i=1}^{N} \frac{1}{N} \left(y_{MODEL,i} - y_{TARGET,i} \right)^2 , \qquad (6)$$

where

y_{MODEL} = output of the neural network model

y_{TARGET} = target value for the output

subscript i = i'th exemplar in a data subset.

VI.D. Training

For the neural network model described in Sec. VI.A, a trained instance was created for each of the 2048 training/validation/valid-2/test data sets. Each instance began first with network weights being randomly assigned. Thus the 2048 neural network models were created from a random data division of the MCO data and random initial guesses for model weights.

The training procedure used a cost function defined as the difference between the neural network model output and the target output for the supervised learning. The minimization of the cost function is the objective of the model training process. Given that training the neural network model amounts to a regression process, the MSE is an intuitive and standard practice choice for the cost function.

The MSE is defined as

$$\mathrm{MSE}_{TRAINING} = \sum_{i=1}^{N} \frac{1}{N} \left(y_{MODEL,i} - y_{TRAINING,i} \right)^2 , \qquad (7)$$

where index i is over the exemplars in the training data set.

However, the minimization of MSE during the training process to a near-zero residual may result in the model being overfit to the data. Essentially the model begins to fit the noise within the training data resulting in poor generalization for data that were never seen by the model. This problem is typically addressed as follows. The training set is used to optimize the neural network by adjusting its parameters so as to optimize the cost function, and the smaller validation set is used to stop the training when the cost function on the validation set stops decreasing, thus avoiding overfitting.[10]

VI.E. Model Selection

A model selection procedure was invoked to choose a "best" model from among the 2048 trained neural networks. This procedure defines the best model as that neural network model that has the lowest value of a figure of merit given by

$$F = \sqrt{\left(\mathrm{MSE}_{TRAINING} + \mathrm{MSE}_{VALID-2}\right)^2 + \left(\mathrm{MSE}_{TRAINING} - \mathrm{MSE}_{VALID-2}\right)^2} , \qquad (8)$$

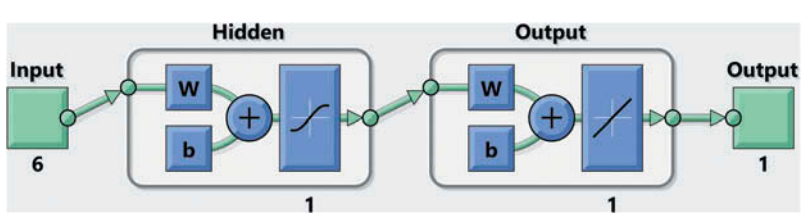

Fig. 11. Example of one hidden layer neural network.

where

$$\mathrm{MSE}_{VALID-2} = \sum_{i=1}^{N} \frac{1}{N} \left(y_{MODEL,i} - y_{VALID-2,i} \right)^2 . \qquad (9)$$

This criterion reflects an imposed requirement that a best trained neural network should satisfy both the smallness and closeness of the cost functions within the training and valid-2 data sets.

VII. RESULTS

VII.A. Feature Analysis

It is essential to show how the selected input features, exiting steam quality Q and liquid velocity V_L, might be optimally combined before input to any machine-learning algorithm. Intuitively, the MCO was fit into a logarithm function shown in Eq. (10) and a similar power function shown in Eq. (11):

$$\log(\mathrm{MCO}) = a \cdot \log(\langle Q \rangle) + b \cdot \log(\langle V_L \rangle) + c \qquad (10)$$

and

$$\mathrm{MCO} = c \cdot \langle Q \rangle^a \cdot \langle V_L \rangle^b + d \, , \qquad (11)$$

where the independent variables $\langle Q \rangle$ and $\langle V_L \rangle$ are the steam quality and liquid component velocity averaged over all the fuel bundles.

Both trivial curve fittings returned MSE values no less than 8×10^{-2}, which is two orders of magnitude larger than the MCO variance due to measurement uncertainty.[3] Such a poor quality of fit demonstrated the need for finer feature tuning and nonlinear mapping capability such as is obtainable through neural network.

The multinode neural network is able to capture the nonlinearities in the input variables when the training data sets are diverse and large in size. However, the small size of the training data sets in this work limited the network to at most two hidden nodes and consequently limited the power of the network to capture all nonlinearities in generalization. Given this limited mapping capability, efforts were undertaken to move into feature-space nonlinear operations that might otherwise challenge the fitting capability of a couple of hidden nodes.

Many training sessions were conducted to identify a nonlinear combination of the inputs represented by Eqs. (4) and (5) that resulted in the best fit to the data. These quantities were superposed in various functional combinations to find an optimized combination and associated exponential parameters m and n that yielded the smallest and most stable MSE for the test set with the least number of neurons in the network.

After a coarse search that limited the exponents to integer values, several candidate combinations were found that yielded a robust and minimum MSE for all random training and test set division, using two neurons in the neural network. Among these candidates, the best input identified was

$$x_i = \sum_{j \in Group_i} \frac{1}{Q_j^2 \cdot V_{L_j}^2} , i \in [1, 7]$$

and

$$x_{i+7} = \sum_{j \in Group_i} \frac{1}{V_{L_j}^1} , i \in [1, 7] \ . \qquad (12)$$

VII.B. Selecting Number of Neurons

A parametric study was performed to identify the optimum number of neurons in the hidden layer of the neural network. The cost function (MSE) distribution for each of the training, validation, valid-2, and test sets were plotted as a function of number of neurons. As shown in Fig. 12, the MSE for the training set decreases with more neurons in the network, but the MSE for the test set stops decreasing with more than three neurons, which indicates overfitting. Indeed, the overfitting appears at three neurons, where the suddenly increased error bar indicates that the model becomes less robust on the test set. To this end, a two-neuron model was selected to be the optimal point that balances both the model robustness and flexibility.

One advantage of a neural network algorithm in a machine-learning approach lies in part with the nonlinear mapping capability afforded by the neural network. Classic models for learning data associations assume a linear-type regression approach. The performance advantage was investigated by replacing the neural network with a linear regression model that for consistency used the same features shown in Eq. (12). The MSE result obtained for this linearly regressed model appears in Fig. 12 as the "0 neuron" case. It is observed that the nonlinear activation function in the neural network provides superior mapping capability. The MSE of the test set for the two-neuron case is 2.9×10^{-4} compared to a value of 6.0×10^{-4} for the linear regression model, a reduction of 52%. Thus the

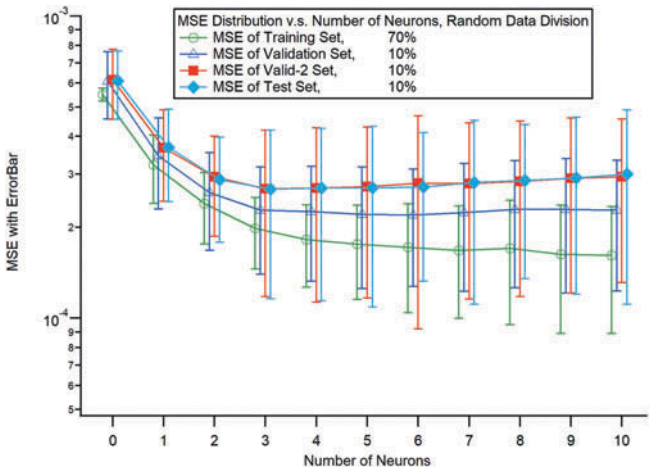

Fig. 12. Cost function distribution versus number of neurons.

neural network is better able to capture the dependency of MCO on core parameters.

VII.C. Model Selection

A selection of a "best" neural network model from among the 2048 trained models was made. The best is defined to be that model that yields the minimum F in Eq. (8). In Fig. 13, each of the 2048 models trained according to the training-validation procedure is represented by a circle. The x-axis is the value of the $MSE_{TRAINING}$ cost function of Eq. (7) and the y-axis is the $MSE_{VALID-2}$ cost function of Eq. (9). The best model appears as the green point and has the identifier "Model 937."

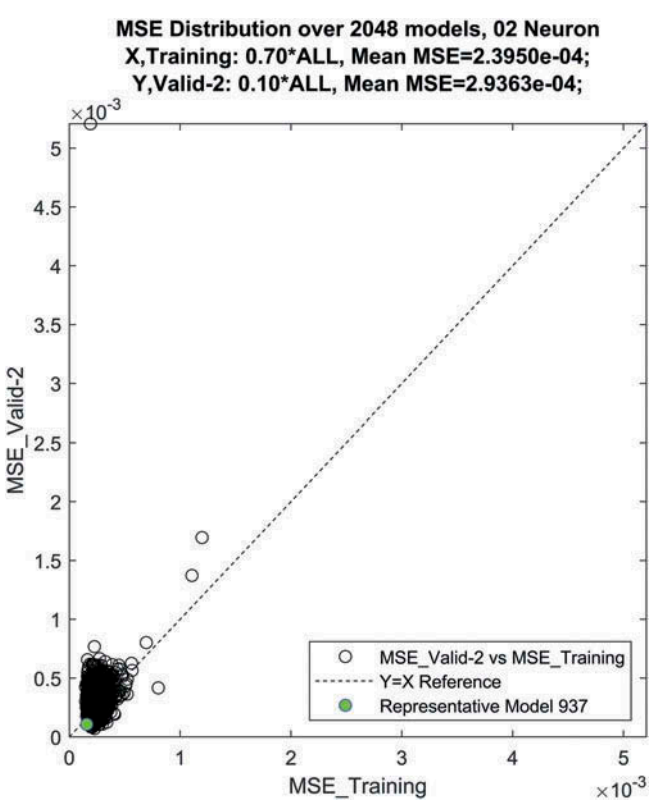

Fig. 13. Selecting the neural network minimizing F.

A comparison of output from the selected model, Model 937, with measured MCO is shown in Fig. 14. Recall that the measured MCO is the target the model was trained against. Shown in Fig. 14 is the output from Model 937 as driven by the validation and the valid-2 data from the training/validation/valid-2/test data set. The closeness of the validation and valid-2 set results to measured MCO suggest this model might generalize well. However, only when we see it actually generalize well (test set) can we say, "Model 937 does a good job predicting MCO," and at that point it is proper to say "predicting."

Table IV reports the performance of the best model, Model 937, on a cycle-by-cycle basis. That is the R^2 and MSE values for Model 937 when presented with each of the four different data set types: training, validation, valid-2, and test.

VII.D. Performance of Selected Model on Test Data Set

The litmus test for model performance is how well the trained model predicts data not yet seen before. Such data are those of the test data set. The test data set is not involved in modeling. It is not involved in training, validation, or model selection. In Fig. 15 the MCO measurements are shown for comparison with Model 937 output as driven with its associated test data. The model prediction performance is quite good indicating the model training and model selection process were successful in generating a model for MCO prediction that generalizes well.

VII.E. Discussion

Model testing is an integral component of machine learning and is performed to ensure the robustness of model predictions. The testing process ensures that the training data have been learned to the degree that the model has predictive or generalization power, which implies that excursions due to idiosyncratic features in the data have been limited. Such features can include second-order physical phenomena that are not fully expressed in the data set or measurement error. Essentially the degrees of freedom in the model must be aligned with the main physics features present in the data to avoid the classic problem of overfitting.

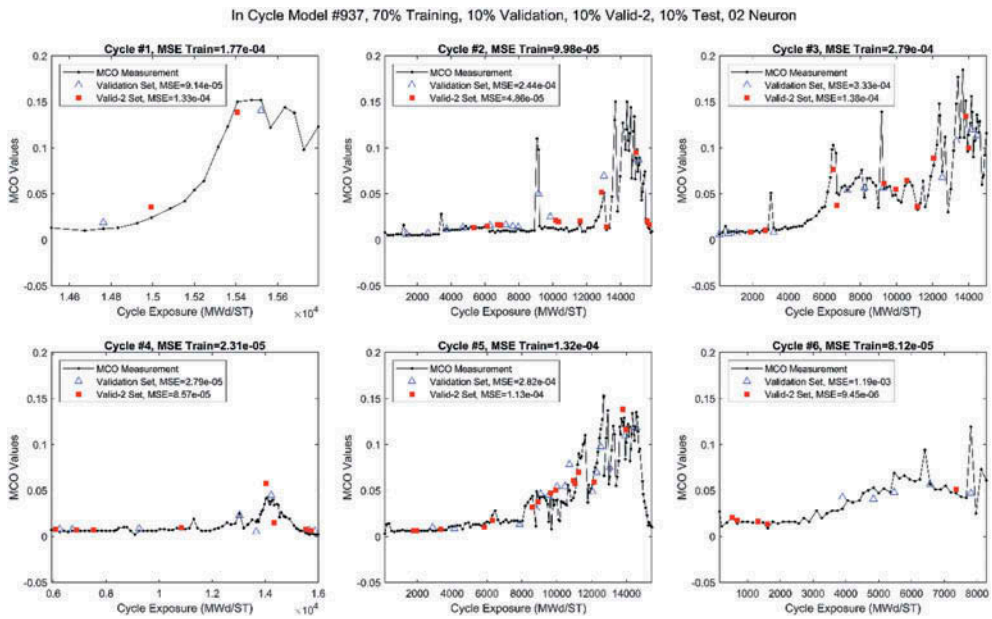

Fig. 14. Comparison of MCO measurements with MCO output by Model 937 for both validation and valid-2 data sets.

TABLE IV

Cycle-by-Cycle Performance Results for Model 937

Criteria_2	Training (70%)		Validation (10%)		Valid-2 (10%)		Test (10%)	
Model 937	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE
6 Cycles	0.908	1.57E–04	0.825	3.58E–04	0.927	1.07E–04	0.798	4.50E–04
Cycle 1	0.952	1.77E–04	N/A[a]	9.14E–05	N/A[a]	1.33E–04	N/A[a]	1.04E–04
Cycle 2	0.943	9.98E–05	0.717	2.44E–04	0.966	4.86E–05	0.870	2.80E–04
Cycle 3	0.856	2.79E–04	0.965	3.33E–04	0.934	1.38E–04	0.852	4.41E–04
Cycle 4	0.812	2.31E–05	0.815	2.79E–05	0.639	8.57E–05	0.851	4.32E–05
Cycle 5	0.930	1.32E–04	0.819	2.82E–04	0.917	1.13E–04	0.674	7.05E–04
Cycle 6	0.807	8.12E–05	−0.282[b]	1.19E–03	0.979	9.45E–06	0.185	7.10E–04

[a]Only two points exists in the validation, valid-2, and test set contributed by Cycle 1. R^2 is meaningless for a linear fitting of only two points due to lack of degree of freedom.
[b]A negative R^2 value indicates a large variance between the model output value and the target value in which the residual from the optimal linear fit is larger compared to the residual from a simple average.

The work reported here marks the completion of the first phase of a two-phase approach to the MCO problem using machine-learning methods. This work demonstrates that, having used random data division, a model was created which can accurately predict MCO at any other arbitrarily chosen historical state points inside that modeled fuel cycle.

Having achieved this indicates that MCO in general is a predictable phenomenon when data quantity is sufficient. The high accuracy under the in-cycle conditions gives us confidence that pursuing out-of-cycle prediction will be fruitful.

In this first phase a model training procedure was used to select a best model, and this model was then tested against data not seen before from the same six cycles.

The second phase will involve testing the model against data from a fuel cycle not contained in the fuel cycles used for training. From a physical data point of view, each fuel cycle has its own characteristics that introduce variation and error into the results. First, each cycle begins with a different fuel loading pattern. The number and positions of twice-burnt bundles change as do the number and positions of the new and once-burnt bundles. Perhaps even more importantly, the positions of the control rods change over time in a different

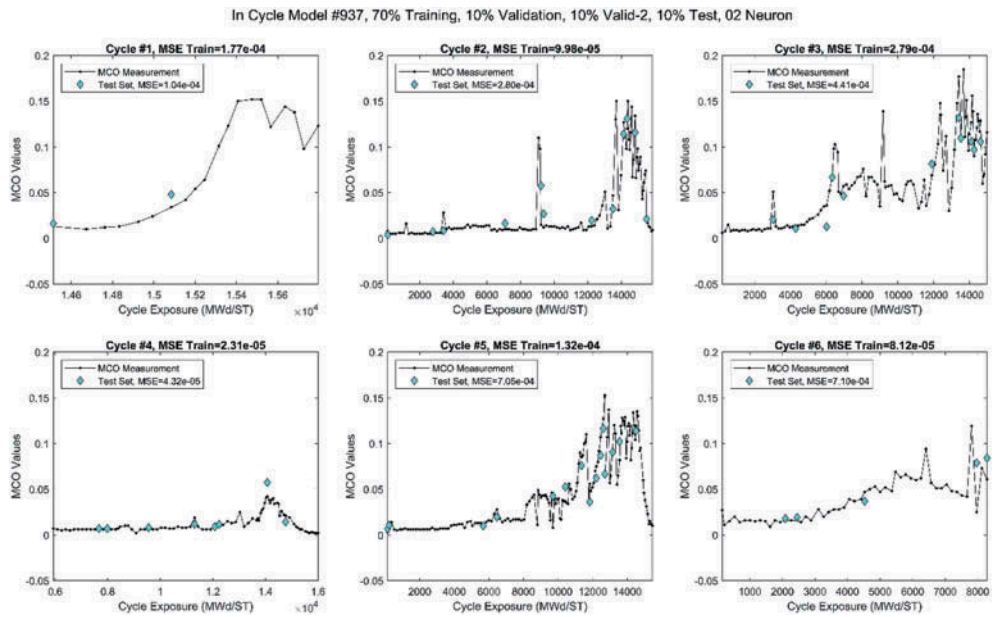

Fig. 15. Comparison of MCO measurements with MCO output by Model 937 for test data set.

way during each cycle. These changes can introduce asymmetries and local "hot" or "cold" spots which in turn affect MCO to various degrees depending on geometry.

By its nature the machine-learning model should be able to deal with these variations in cycle behavior. The ability to do this depends critically on the amount of data and the variety of conditions the data represent. In the case studied in this paper, there is a modest amount of data and not yet enough to model the entire spectrum of possible fuel cycle conditions. Using the inherent symmetry of the reactor system and reducing the feature set to functions of Q and V_L has certainly helped. However, we do know that asymmetric conditions exist, and in a subsequent paper we will analyze their impact.

Future enhancements to this analysis will include using interpolations of the MCO values. While reactor-state measurements occur every day, MCO is measured once per week limiting the number of data points. Also, further refining the feature set will improve the model's predictive power.

Finally, there is the issue of the intrinsic error in the measured MCO values and PANACEA output. This will be analyzed and accounted for in future efforts as well. Currently, there is limited knowledge of these errors, so as a first step we have examined models assuming errors in reported values to be negligible.

VIII. CONCLUSION

The prediction of MCO values as reactor burnup progresses is an important operational issue with significant economic impacts to power generation. It depends on a very complex set of operating conditions, and the electric utility has limited ways to control and manage MCO while keeping power at planned levels. Having a useful predictive model of MCO is thus of high importance. Given the large amount of data collected over the years of nuclear reactor operation, machine learning presents itself as a very promising way to model MCO. Machine learning is based on rigorous mathematical theory and, in simple terms, posits that the function governing the system's evolution is in the data and can be approximated if enough data are available. Of course, the data also contain noise and so appropriate care was taken in this paper to avoid fitting the noise.

We have used data from two GE Type-4 BWR reactors with a combined six fuel cycles to model MCO as a function of various reactor variables. Given the large number of features per fuel cycle, we used a physics-guided approach to reduce those variables to a small set of independent variables. Additionally, we took advantage of the apparent symmetry of the reactor layout to virtually increase our data set by a factor of eight. Also, the k-means technique was used to cluster variables and so find the Q and V_L natural groupings.

In this paper we have demonstrated that a model of MCO can be constructed for random data division within multiple fuel cycles and will generalize by predicting randomly selected points in the space that

are not part of the training of the model. We have also taken advantage of the random seeding inherent in neural networks to explore the space of valid models which differ by their random starting point. We then picked a model that had best valid-2 set performance. Thus, we are encouraged that this approach will allow even more useful models to be constructed. Future work will develop models of separate fuel cycles given a set of completed cycles from existing reactors. In addition, we will develop models of future MCO values within a given fuel cycle.

Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Nuclear Energy. The submitted manuscript was created by UChicago Argonne, LLC, operator of Argonne National Laboratory. Argonne, a DOE Office of Science laboratory, is operated under contract DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan at http://energy.gov/downloads/doe-public-access-plan.

ORCID

Richard Vilim http://orcid.org/0000-0003-1426-0738

References

1. *Scoping Study of Moisture Carryover in Boiling Water Reactors*, p. 1021185, Electric Power Research Institute, Palo Alto, California (2010).
2. *BWR/4 Technology Manual (R-104B)*, U.S. Nuclear Regulatory Commission Technical Training Center (2002).
3. Private communication with Limerick Generating Station for BWR geometry, fuel loading plan, and MCO measurement procedure (2017).
4. R. G. COCHRAN et al., "The Nuclear Fuel Cycle: Analysis and Management," *Am. Nucl. Soc.*, p. 181 (1990).
5. E. STUMBRIS, "Study of Moisture Carryover in Boiling Water Reactors," Baccalaureate Thesis, Pennsylvania State University (2010).
6. R. FOURNIER, M. THIBODEAU, and C. T. FRENCH, "Measurement of Steam Generator or Reactor Vessel Moisture Carryover Using a Non-Radioactive Chemical Tracer," *17th Int. Conf. on Nuclear Engineering*, p. 459, American Society of Mechanical Engineers (2009); https://doi.org/10.1177/1753193408101466.
7. J. BENESTY et al., "Noise Reduction in Speech Processing," Vol. 2, Springer Science & Business Media (2009).
8. K. LEVENBERG, "A Method for the Solution of Certain Non-Linear Problems in Least Squares," *Q. Appl. Math.*, **2**, *2*, 164 (1944); https://doi.org/10.1090/qam/1944-02-02.
9. D. W. MARQUARDT, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *J. Soc. Ind. Appl. Math.*, **11**, *2*, 431 (1963); https://doi.org/10.1137/0111030.
10. R. GENCAY and M. QI, "Pricing and Hedging Derivative Securities with Neural Networks: Bayesian Regularization, Early Stopping, and Bagging," *IEEE Trans. Neural Networks*, **12**, *4*, 726 (2001); https://doi.org/10.1109/72.935086.