Journal of the Korean Nuclear Society
Volume 19 Number 4, December 1987

A Study on the Method of Combining Empirical Data and Deterministic Model for Fuel Failure Prediction

Byeong Ho Cho*·Young Ku Yoon and Soon Heung Chang
Korea Advanced Institute of Science and Technology
(Received July 24, 1987)

핵연료 파손 예측을 위한 경험적 자료와 결정론적 모델의 접합 방법

조 병 호·윤 용 구·장 순 흥
한국과학기술원
(1987. 7. 24 접수)

Abstract

Difficulties are encountered when the behavior of complex systems (i.e., fuel failure probability) that have unreliable deterministic models is predicted. For more realistic prediction of the behavior of complex systems with limited observational data, the present study was undertaken to devise an approach of combining predictions from the deterministic model and actual observational data. Predictions by this method of combining are inferred to be of higher reliability than separate predictions made by either model taken independently. A systematic method of hierarchical pattern discovery based on the method developed in the SPEAR was used for systematic search of weighting factors and pattern boundaries for the present method. A sample calculation was performed for prediction of CANDU fuel failures that had occurred due to power ramp during refuelling process. It was demonstrated by this sample calculation that there exists a region of feature space in which fuel failure probability from the PROFIT model nearly agree with that from observational data.

요 약

본 연구는 제한된 수의 핵연료의 경험적 파손자료로부터 핵연료 파손 확률을 현실적으로 예측하기 위해 결정론적 모델로부터의 파손확률 예측치와 실제 경험적 자료로부터의 파손 확률 예측치를 접합하는 방법을 시도하였다. 이 접합 방법에 의한 파손 확률 예측치는 결정론적 모델 또는 경험적 파손 자료로부터의 독립적인 예측치보다 신뢰도가 높다. 본 연구에서는 핵연료 성능 예측코드인 SPEAR의 방법론을 응용한 핵연료 파손 패턴의 체계적 발견법(hierarchical pattern discovery)이 접합 모델에서의 결정론적 모델로부터의 예측치에 대한 가중치와 패턴 경계를 체계적으로 찾기 위해 고안되었다. 이 연구에서 개발된 접합 방법을 PROFIT 모델과 경험적 파손자료를 이용하여 CANDU형 핵연료 재장전중 출력 상승에 의해 수반되는 핵연료파손 예측에 적용시켜 보았다.

* Now, a member of KAERI

Nomenclature

bu burnup [MWh/kgU]

dbu burnup increase used in pattern boundary reassumption

ddelpo transient power increase used in pattern boundary reassumption

eps pattern description permission bound

fmb $\frac{1}{2}$(Train+Ftest)

Ftest observed failure probability of Test data

Ftrain observed failure probability of Train data

ni critical number of data point

Pi initial power level [kW/m]

Pof failure probability of input data calculated from deterministic model

Pr 1 realistic estimate of failure probability= $(1-w)\cdot$Ftrain$+w\cdot$Pof

Pr 2 realistic estimate of failure probability= $(1-w)\cdot$fmb$+w\cdot$Pof

Pvirt failure probability of Train data calculated from deterministic model

w weighting factor

I. Introduction

Difficulties are encountered when the behavior of a complex system (e.g. fuel failure) with limited observational data is predicted. If the deterministic model, either mechanistic or statistical, is not magic for predicting its behavior, there exist difficulties in treating the uncertainty of the deterministic model for prediction. For more realistic prediction of the behavior with limited observational data, an attempt is made to devise an approach of combining predictions from the deterministic model and actual observational data. This method comprises determination of weighting factors for predictions of the deterministic model relative to those of empirical behavior patterns found in real observational data.

A systematic method of hierarchical pattern discovery is used as a pattern discovery algorithm. In applying this method, weighting factors for a determinstic model relative to empirical data set are searched systematically. The pattern boundary(boundary of a subset of feature* space; pattern discriminant) function for this method is hierarchical. Hence, the pattern boundary is parallel to each independent variable axis. Pattern discovery is continued until the whole feature space is partitioned. The feature space is defined as the multi-dimensional space, the axes of which are the independent variables or functions of the independent variables. With these patterns and weighting factors, realistic failure probabilities of certain input data are predicted by pattern matching calculation. The probability of a failure event within a pattern boundary from empirical data set corresponding to a set of independent feature variables and the weighting factor are used for realistic faieure prediction.

The present study is undertaken to demonstrate by a sample calculation based on power ramp-related fuel failures of the CANDU reactor [4] that fuel failure probabilities can be predicted by the method of combining a modified PROFIT model and observational failure data. The PROFIT model is used as the deterministic model and the probability corresponding to a set of independent variables is predicted by the method of combining a modified PROFIT model and observational data.

II. Methodology Description

II-1. Method of Combining Deterministic Model and Empirical Observational Data

* Feature means independent variable in pattern recognition jargon.

The objective in modeling any physical behavior is to determine a relationship between dependent variable D (i.e., fuel failure probability) and a set of independent variables i_1, $i_2, \cdots$, i_n. Namely, the equation on the form $D=f(i_1, i_2, \cdots i_n)$ is an example of deterministic model equations. In building fuel failure pattern files with use of empirical data set, the systematic method of hierarchical pattern discovery is used to partition the feature space into regions and to determine in each pattern boundary the weighting factors for the failure probability based on the deterministic model. The failure probability assigned to each pattern boundary is as follows:

$$P=(1-w)\cdot\text{Ftrain}+w\cdot\text{Pvirt} \quad (1)$$

where w is a weighting factor of a predetermined failure pattern, and Ftrain is a fuel failure frequency of the Train data of a predetermined failure pattern (a subset of feature space of fuel failure) and Pvirt is the failure probability of observational data that is calculated from the deterministic model.

As a first step in building failure patterns,

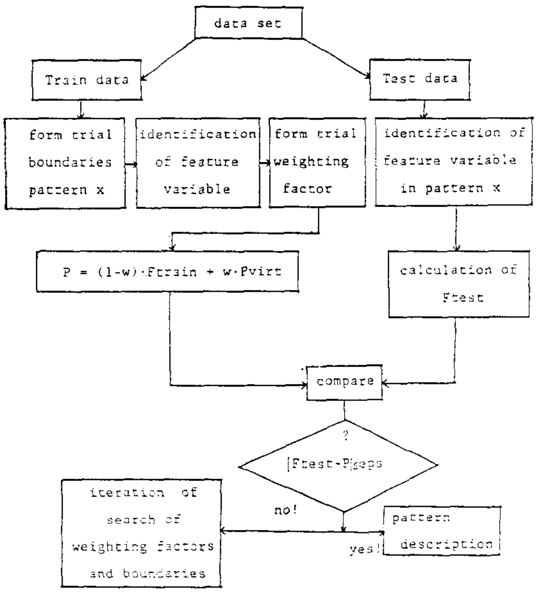

Fig. 1. Determination of Pattern Boundaries and Weighting Factors

the set of empirical observed data is split randomly into two equal-sized parts. Half the data is termed as Train data and the other half as Test data. This procedure is shown in Fig. 1. The 'Train' data set is used in finding the feature space partition and weighting factors. Weighting factor for calculated data is high in pattern boundaries with many data points whereas weighting factor is low in pattern boundaries with a few data points.

An $(n+1)$ dimensional space is formed using the depencent variable (i.e., fuel failure probability) and each of the n dimensional independent variables as each axis. The space is populated with each of the points $(D, i_1, i_2, \cdots i_n)$ taken from the data set. The failure pattern model is made by partitioning this space into a number of discrete patterns.

The next step in building the failure pattern is determination of the relative weighting factor of the observational failure probability and the calculated failure probability from the deterministic model. The subsequent problem is iden tifying those patterns(viz. failure pattern model) in the array of data bits forming the independent variables data file that reliably predict the patterns data bits forming the dependent variable data file.

For the model obtained by this process to have statistical validity, the same relationship found in the Train data must hold in the Test data. The estimated true failure probability is then a weighted average of the observed failure probability Ftrain and the calculated failure probability Pvirt using weighting factor w. Next, the pattern discovery procedure is to find the best pattern description and associated pattern-by-pattern weighting factor. For pattern discovery Train data and Test data are plotted separately. The pattern discovery process then tries various combinations of weighting factors and pattern boundaries in the Train data space,

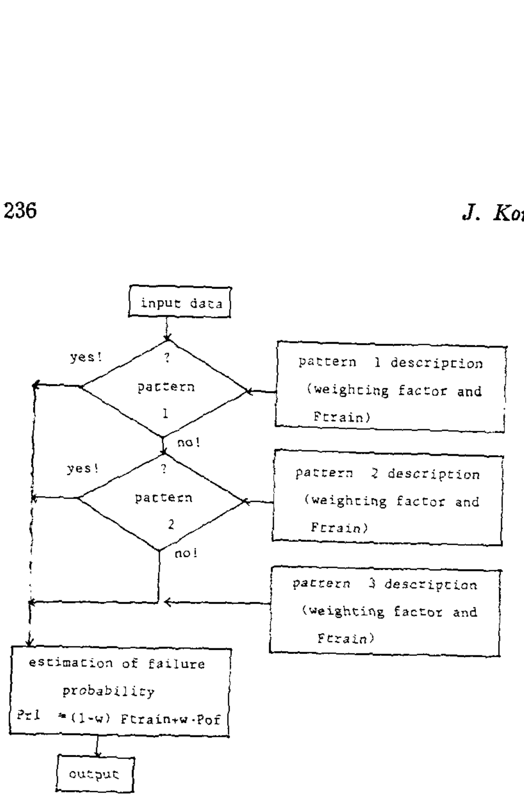

Fig. 2. Pattern Matching and Estimation of Failure Probability

seeking a set of conditions that best predicts the Test data set. This procedure is continued until the entire feature space is partitioned into a complete set of patterns. This procedure is shown in Fig. 2.

In general, high weighting factors result from sparse observational data and good agreement between calculated probability and observed failure probability. Conversely, low weighting factors result from large observational data or from poor agreement between the calculated probability from deterministic model and observational probability. The boundary of a region determines which point falls in each pattern. Hence the pattern boundary is the discriminant function in the pattern matching procedure. This pattern boundary function determines the belonging of the data point to a certain fuel failure pattern. For each fuel failure pattern, the failure probability is computed from the observational empirical data base.

To make a prediction using this failure pattern model, the dependent variables corresponding to the input data are computed through the deterministic model. These calculated failure probabilities with the n dimensional independent variables of input data fall into one of predetermined failure patterns. The pattern boundaries and pattern failure probabilities are actually constructed by considering both the observed data and the calculated data, i.e., the output from the deterministic model.

The process of failure probability prediction begins with the calculation of the failure probability for the input data. The features for a certain input data are passed through the pattern detection logic(pattern matching logic) that determines which of the patterns the particular input data matches. A failure probability from the Train data is assigned to each pattern. When it has been determined that a particular pattern is matched, for instance, pattern 1, certain input data is assigned the failure frequency from Train data and the weighting factor w associated with that pattern. The realistic failure probability prediction may be given by Eq. (2).

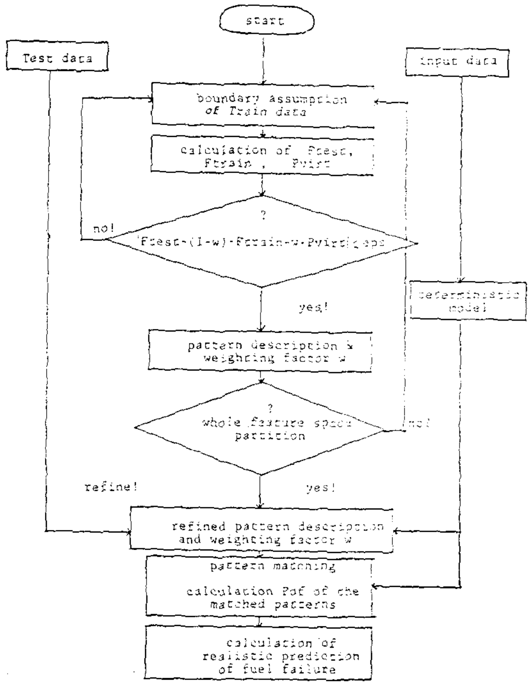

Fig. 3. Flow Chart of Overall Calculation

$$P=(1-w)\cdot \mathrm{Ftrain}+w\cdot \mathrm{Pof} \quad (2)$$

where w is a weighting factor of a predetermined failure pattern, Ftrain is a fuel failure frequency of the Train data of a predetermined failure pattern, and Pof is the failure probability of input data calculated from the deterministic model. Pattern matching and computation of failure probability are described in Fig. 2. The flow chart of the calculation of failure prediction by this method is described in Fig. 3.

II-2. Systematic Method of Hierarchical Pattern Discovery

In the systematic method of hierarchical pattern discovery the discriminant function [10] is composed of parallel boundary with feature axis. A failure probability of Train data in a pattern has a weighting factor with respect to that of the same data from the deterministic model. The systematic method of hierarchical pattern discovery is used for the systematic determination of relative weighting factors between the deterministic model and empirical data set. The pattern discovery procedure is as follows; First, a pattern boundary is assumed. Calculations of Ftrain, Ftest and Pvirt are performed with preassumed pattern boundary. Next, the following condition

$$\mathrm{Pvirt}>\mathrm{Ftrain} \quad (3)$$

is examined. If this condition is satisfied, the next procedure is performed. If the number of data points is below the critical, then systematic search of weighting factors begins with $w=0.51$ and the condition

$$|\mathrm{Ftest}-(1-w)\cdot \mathrm{Ftrain}-w\cdot \mathrm{Pvirt}|\leq \mathrm{eps} \quad (4)$$

is examined where eps is the permission bound for the pattern description. If this condition is satisfied, the weighting factor is 0.51 with this pattern boundary. If this condition is not satisfied, the condition

$$\frac{1}{2}(\mathrm{Ftrain}+\mathrm{Pvirt})<\mathrm{Ftest}<\mathrm{Pvirt} \quad (5)$$

is examined. For this case, the condition (4) is examined with $w=w+0.01$. If the condition (5) is not satisfied, the next condition

$$|\mathrm{Ftest}-\mathrm{Pvirt}|<\mathrm{eps} \quad (6)$$

is examined.

If conditions (4), (5) and (6) are not satisfied, another pattern boundary is reassumed and the procedure described above is performed.

If the number of data points is above the critical, then systematic search of weighting factors begins with $\omega=0.49$. Subsequently, the condition (4) is examined. If this condition is satisfied, the weighting factor is 0.49 with this boundary. If this condition is not satisfied, the next condition

$$\mathrm{Ftest}<\mathrm{Ftrain}<\frac{1}{2}\cdot(\mathrm{Ftrain}+\mathrm{Pvirt}) \quad (7)$$

is examined. For this case, the condition (4) is examined with $w=w-0.01$. If this condition is not satisfied, the following condition is examined;

$$|\mathrm{Ftest}-\mathrm{Ftrain}|<\mathrm{eps} \quad (8)$$

For this condition, the condition (4) is examined with $w=w-0.01$. If conditions (4), (7) and (8) are not satisfied, another pattern boundary is reassumed and the same procedure is repeated, starting from the condition (3). For the case of the complementary condition of (3)

$$\mathrm{Pvirt}\leq \mathrm{Ftrain}, \quad (9)$$

the entire procedure is the same as the case of condition (3) except conditions (5) and (7). For this case, conditions (5) and (7) are replaced with the following conditions;

$$\mathrm{Pvirt}<\mathrm{Ftest}<\frac{1}{2}(\mathrm{Ftrain}+\mathrm{Pvirt}) \quad (10)$$

and

$$\frac{1}{2}(\mathrm{Ftrain}+\mathrm{Pvirt})<\mathrm{Ftest}<\mathrm{Ftrain} \quad (11)$$

The pattern discovery procedure is continued until the whole feature space is partitioned.

III. Sample Calculation

III-1. Deterministic Model for Power Ramp-Related Failure of CANDU Reactor Fuels and Observational Data [4]

As a sample calculation, the power ramp-related failure of CANDU reactor fuels is treated. In CANDU reactors with on-power refuelling the great majority of defects(i.e., 71%) have been of the power ramp-related type [8, 9]. During on-power refueling process, fuel bundles that have operated at low power in early life are moved to higher power locations concurrent with the introduction of new fuel bundles. The movement of the fuel bundles is processed in discrete stages because of operational modes of the fuelling machines. Thus, fuel-bundles destined for moderate power positions reside transiently in high power positions. The major cause of defects is power ramping and the defect mechanism is considered on the basis of a large weight of circumstantial evidence to be fission-product induced stress corrosion cracking of Zircaloy fuel cladding.

The PROFIT model [5] (failure model for estimating the probability of Failure In Transient increase in power) model is adopted as the deterministic model. The literature [5] is referred to for more detailed understanding of the PROFIT model.

III-2. Assumptions

The random split procedure described in the preceding methodology description is performed. In this sample calculation a random number is assigned to an observational data point in a random number sequence. If a random number is below the half of the interval, then the data point to which that random number is assigned is the family of Train(or Test) data. The random split of data set is performed by this method. ΔP failure boundary as a function of burnup is shown in Eq. (12).

$$P_c=6+1460/\mathrm{bu} \tag{12}$$

The PROFIT model is modified to take the observational propensity of failure data into consideration. Failure probability from the deterministic model is assumed as follows:

If $\Delta P<6+1460/\mathrm{bu}$,

then Pof(or Pvirt)=0.00

Initial power levels of observational data are not known. Hence Pi is asumed to be 20.0kw/m due to the fact that the mean value of data set used in the PROFIT model is 6.1kw/ft ($\sim$20.0 kw/m) and fuel failure is less sensitive to initial power level than to transient power increase. The dimension of the feature space is reduced by this assumption but unreliability is increased.

The SEAF ratio is assumed to be 1.1. due to the closeness of probabilities from the deterministic medel and from observational data, where SEAF is strain energy absorption to failure at a specific straining rate in the Zircaloy fuel cladding corresponding to an observed power ramping rate. Pattern discovery is performed for the following ranges.

$30\leq \mathrm{bu}\leq 170,\ 0\leq \Delta P\leq 50$

The overall calculational procedure is shown in Fig. 3.

IV. Results and Discussion

The factor that influences most both failure patterns and weighting factors is the critical number of data points that determines the starting point of systematic search of weighting factors. Examination of the propensity of power ramp-related falure data by a large number of the pattern discovery procedure yielded the critical number to be approximately 10 in this sample calculation. Good agreement between Ftrain and Ftest in a pattern is required for an ideal random split. A large data set yields better agreement between Ftrain and Ftest in a pattern. In case of no sufficient data for a pattern, there is the possibility of a skewed-split of observational data set. The skewed-split patterns resulted from a random split are attributed to sparsity of data for the pattern. The algorithm for an ideal split of observational data in a

Table 1. Variations of Estimated Failure Probabilities with Different Permission Bounds of Boundary Reassumption

	0.01	0.03	0.05	0.07	0.09	0.11
Case 1	0.102	0.194	0.194	0.194	0.194	0.194
Case 2	0.228	0.175	0.178	0.220	0.210	0.228

	burnup (MWh/kg U)	Power increase (kW/m)
Case 1	50	38
Case 2	120	21

Table 2. Variations of Estimated Fialure Probabilities with Different Unit Intervals of Power Increase for Pattern Boundary Reassumptions

	2	4	5	8	10
Case 1	0.166	0.191	0.194	0.200	0.203
Case 2	0.207	0.170	0.178	0.155	0.181

pattern is needed to overcome the skewness of a random split.

Table 1 shows variations of estimtated failure probabilities for two cases of feature variables set with different permission bounds for pattern boundary reassumption. It can be seen from this Table 1 that there is a firm tendency of about 20% probability of fuel failure in the case 1 (burnup of 50MWh/kg U and power increase of 38kW/m) whereas there is some variation in estimated failure probabilities with different permission bounds in the case 2 (burnup of 120MWh/kg U and power increase of 21kW/m). Table 2 shows variations of estimated failure probabilities with different unit intervals of power increase for pattern boundary reassumption. It can be seen from this Table 2 that there is the propensity of increasing failure probability with increasing unit interval of power increase for pattern boundary reassumption in the case 1 whereas there is some variation in estimated failure probabilities with increasing unit interval in the case 2. If many bundles with the same value for the independent feature variable are tested by sipping, computed failure probabilities represent estimates of the bundle failure frequency.

Realistic failure probabilities of input data

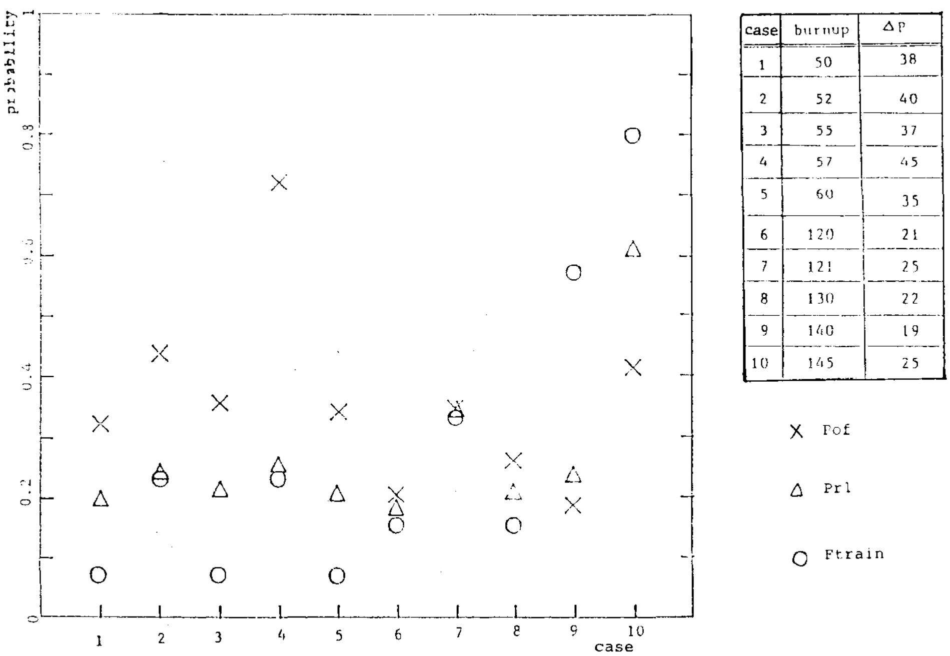

Fig. 4. Comparison between 10 Characteristic Cases

Table 3. Output from Pattern Discovery Procedure

no.	burnup range (MWh/kg U)		power increase range (kW/m)		no. of data	Ftrain	Ftest	Pvirt	w
1	30	65	0	40	26	0.071	0.068	0.118	0.49
2	30	65	40	50	10	0.248	0.215	0.554	0.05
3	65	100	0	40	6	0.125	0.154	0.092	0.51
4	65	100	40	50	0	0.000	0.154	0.000	0.50
5	100	135	0	25	30	0.156	0.144	0.062	0.49
6	100	135	25	50	1	0.333	0.179	0.350	0.50
7	135	170	0	25	5	0.571	0.177	0.140	0.80
8	135	170	25	50	3	0.800	0.177	0.497	0.50

eps=0.05, ni=10 delpol=0 ddelpo=5 bul=30 dbu=35

corresponding to 10 characteristic cases of the CANDU refuelling process are shown in Fig. 4. It can be seen from Fig. 4 that the PROFIT model is overestimated in the low burnup range and underestimated in the high burnup range from the viewpoint of observational data.

In general, the larger number of data points yields better agreement between Ftrain and Ftest. If Ftrain, Ftest and Pof are in good agreement with one another in a pattern, it can be inferred that failure probabilities from the PROFIT model and from observational data are in good agreement with each other. Hence a region of the feature space in which failure probabilities from the PROFIT model are in good agreement with those from observational data was identified from the output of the pattern discovery procedure given in Table 3, namely, the burnup range of 100~135 MWh/kg U and the power increase range of 0~25 kW/m. It can be seen from this result that the PROFIT model is effective in predicting realistic failure probability of an independent featuer set corresponding to this region of pattern. It is, therefore, recommended to use the combined model to predict realistic failure probability of the independent feature set for other patterns except for the above-mentioned region.

As opposed to the present systematic method of hierarchical pattern discovery, the method of Entropy Minimax pattern discovery [1] on which the SPEAR-BETA fuel performance code is based has been reported as a pattern discovery algorithm for prediction of realistic fuel failure probabilities by the combining method in the case of situations with high dimensional mechanistic model and observational data.

V. Conclusions

The systematic method of hierarchical pattern discovery can be used as an effective pattern discovery algorithm in the case of two or three dimensional feature space models with the critical number of data points that determines the starting point for searching weighting factors.

The strategy for an ideal split of observational data for a pattern is required to overcome the skewness of a random split.

This systematic searching method of weighting factors can be used for any type of pattern boundaries on the basis of the critical number of data points that provides a starting keystone for the assumption of weighting factors.

A sample calculation based on the PROFIT model and power ramp-related fuel failure data from a CANDU reactor showed that realistic fuel failure probability can be predicted by the method of combining the deterministic model and empirical observational data.

References

1. R. Christensen, "SPEAR-BETA Fuel Performance Code System," Volume 1; General Description, EPRI Final Report NP-2291, April, 1982.
2. R. Christensen, 'SPEAR-BETA Fuel Performance Code System," Volume 1; General Description, EPRI Interim Report, NP-1378, March, 1980.
3. Gary S. Was, R. Christensen, Chang Park, R. W. Smith, "Statistical Patterns of Fuel Failure in Stainless Steel Clad Light Water Reactor Fuel Rod," *Nuclear Technology*, *71*, 445-457 (1985).
4. W.J. Penn, R.K. Lo, J.C. Wood, "CANDU Fuel-Power Ramp Performance Criteria," *Nuclear Technology*, *34*, 249-268 (1977).
5. P.J. Pankaskies, "Mechanistic Considerations Used in the Development of the PROFIT PCI Fuel Failure Model," NUREG CR-1462, PML-3386 R3.
6. J.T. Roberts, F.E. Gelhause, H. Ocken, N. Hoppe, S.T. Oldberg, G.R. Thomas, "LWR Fuel Performance Program: Progress in 1978," NP-1024-SR, Electric Power Research Institute (1979).
7. Trivedi, Probability and Statistics with Reliability, Queuing, and Computer Science Application, 54-70, Prentice-Hall, New York (1982).
8. J.C. Wood, "CANDU Fuel Performance and the Derivation of 'FUELOGRAM,' CRNL-2494-6, AECL (1983).
9. R.D. MacDonald, J.J. Lipsett, B.J. Lewis, "The In-Reactor Behavior of Defective Zircaloy Sheathed-UO_2 Fuel Elements," CRNL-2494-9, AECL (1983).
10. T. Tou, R.C. Gonzalez, Pattern Recognition Principles, 39-74, Addison-Wesley Publishing Company, New York (1974).