INCEFA-SCALE PROJECT

DATA MANAGEMENT AND DATA MINING WORKS

R. Cicero[1], A. McLennan[2], J. Mann[2], J. Beswick[2], L. Doremus[3], S. Cuvilliez[4]

[1] Inesco Ingenieros, Santander, Spain
[2] Jacobs, Warrington, UK
[3] Framatome, Le Creusot, France
[4] EDF, Lyon, France

INDEX

Index

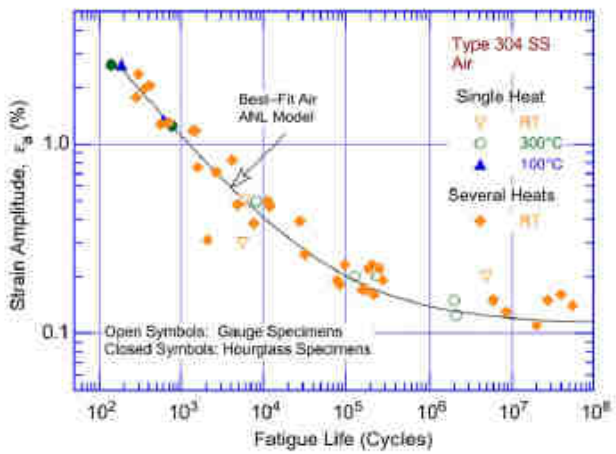
- INTRODUCTION
- DATABASE
- WEB APPLICATION
- DATA MINING
- CONCLUSIONS

INTRODUCTION

Background FATIGUE

- The structural integrity of a metal component may gradually degrade when it is subjected to cyclic loading.

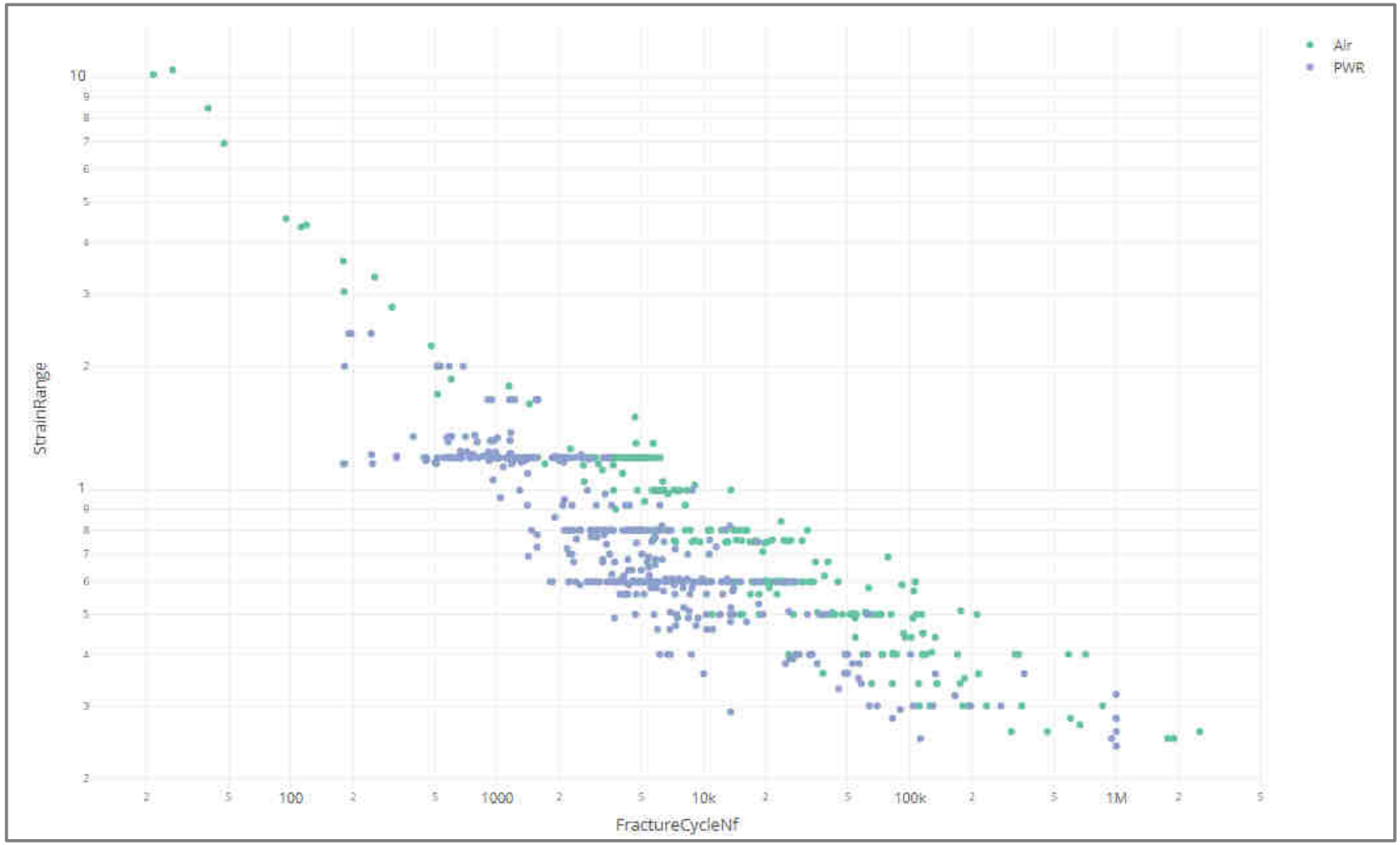
- Fatigue design curves define the allowable number of cycles as a function of applied stress amplitude. The fatigue design curves are based on strain–controlled tests of small polished specimens at room temperature in air.



- Existing fatigue ε–N data illustrate potentially significant effects of light water reactor (LWR) coolant environments on the fatigue resistance of carbon and low-alloy steels, wrought and cast austenitic SSs, and Ni-Cr-Fe alloys

INTRODUCTION

Background FATIGUE

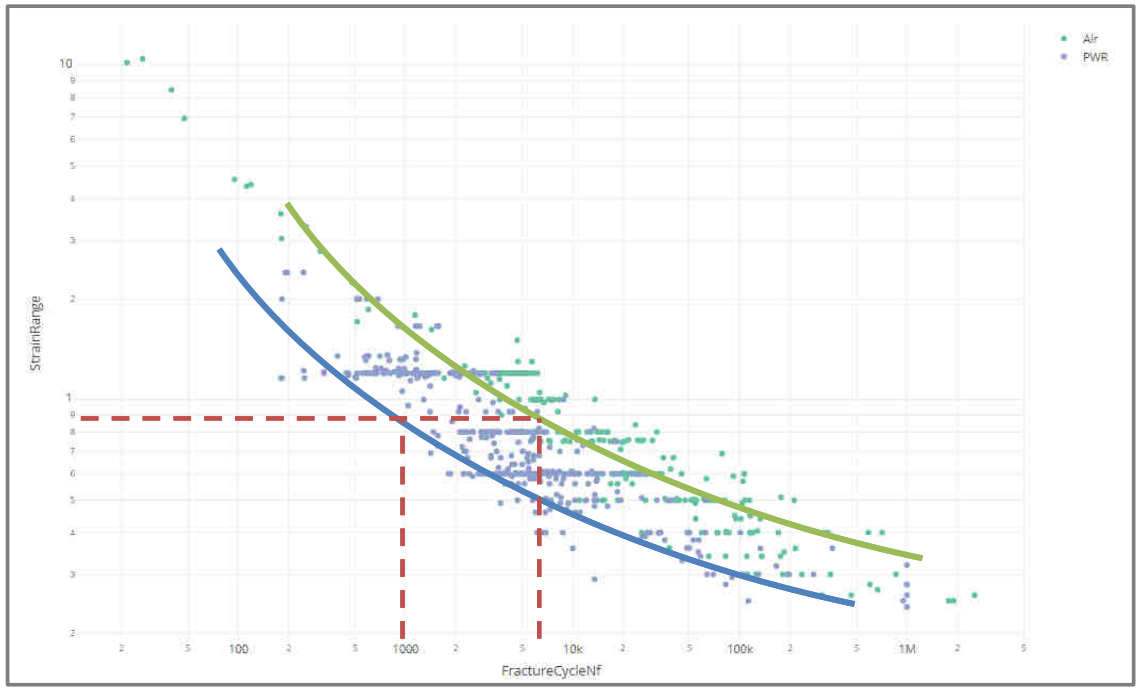


INTRODUCTION

Background FATIGUE

- The effects of water environments on fatigue lives were expressed in terms of an environmental factor **Fen**, which was defined as the ratio of life in air at room temperature, N_{RTair}, to that in water at the service temperature, N_{water}. Values of Fen were obtained from the ANL fatigue life models, where:

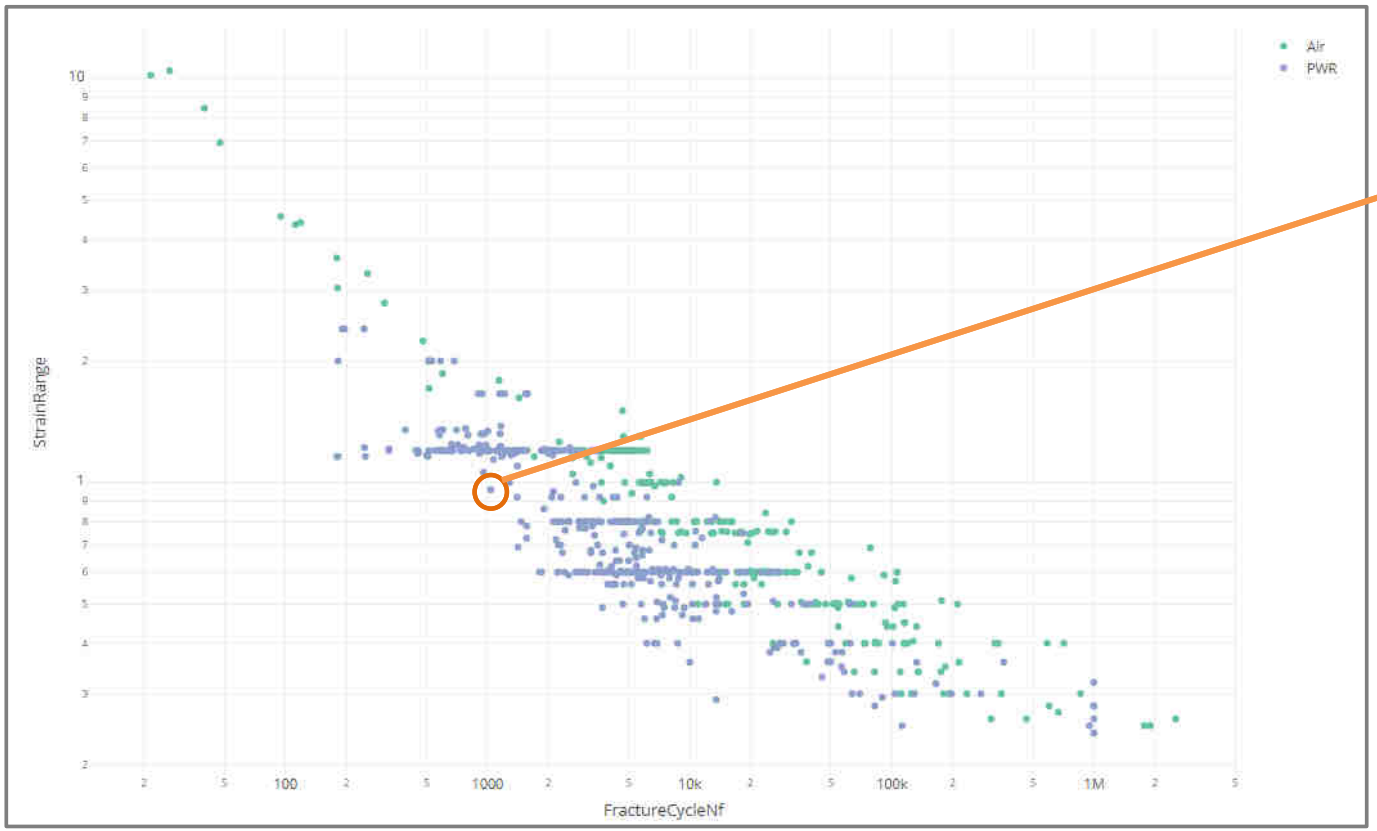
$$\ln(Fen) = \ln(N_{RTair}) - \ln(N_{water})$$



INTRODUCTION

Background FATIGUE

$$F_{en} = \exp(-T^* \dot{\varepsilon}^* O^*)$$



Environment,
Strain range
Strain rate
Temperature
Waveform
Surface roughness
DO
Grade
DH
….

Not all fatigue tests (points) were developed under same conditions

INTRODUCTION

Background FATIGUE

- The environmental factor (Fen) only considers a limit number of variables (after a detailed statistical analysis).
- The Fen is based on fatigue curves obtained in AIR and Water environment, where the ratio can be assessed.
- Both fatigue models are obtained based on a limit number of points (fatigue tests) where different variables are considered.
- The higher the number of fatigue tests (points) the higher the accuracy of the fatigue curves (and the Fen).

INTRODUCTION

Background INCEFA SCALE

- The purpose of INCEFA SCALE project is to improve the prediction of the durability of nuclear power plant components when subjected to environmental-assisted fatigue loading (EAF)

- In this project several fatigue tests are being (and will be) conducted in PWR environment and air under different environmental conditions, strain amplitude and strain rate, surface roughness, etc.

- All these data, along with other significant data shared from other international projects and organizations (from Europe, US, South Korea, and Japan), will be used to create the world's largest nuclear fatigue database.

- This database will be used to develop new predictive fatigue models able to determine the environmental factor and to qualitatively determine the influence on the fatigue life of the different parameters of analysis. *Does any variable not currently considered in the models have a significant influence on fatigue life*?

- A data mining tool was developed for that purpose, that is, to statistically analyze all data available to date.

DATABASE

Fatigue data available

SOURCE		TOTAL NUMBER OF TESTS	AIR	PWR	NUMBER OF MAIN VARIABLES RECORDED
NRC/EPRI		127	89	38	10
MATDB	INCEFA-PLUS	260	103	157	22
	MHI	146	0	146	10
	PE	69	54	15	10
	KAERI	93	72	21	10
VTT		16	0	16	10
		711	318	393	

The available fatigue data will be increased in the next years. New organizations and new projects (f.ex. INCEFA-SCALE) will add new data to the international fatigue database.

DATABASE

Variable of analysis

- There are 10 variables (highlighted in green) common to all available fatigue tests to date.
- In addition, there are some tests (conducted in the INCEFA-PLUS project) that contains 12 additional variables.

Variable	Common	Description	Units
Material grade		AISI 304, AISI 304L, AISI 316, AISI 316L, SA 182M F304L, X6 CrNiNb 18 10, X6 CrNiTi 18 10	
Diameter		*real value*	mm
Type of specimen		Full (bar), Hollow	
Environment		Air, PWR, VVER	
Temperature		*real value*	ºC
Waveform		Triangular, Sawtooth or Complex	
Strain ratio		*real value*	
Strain rate positive		*real value*	·/s
Strain rate negative		*real value*	·/s
Strain range		*real value*	%
Hold time		Yes/No	
Fracture Cycle Nf		*real value*	
Average Roughness		*real value*	µm
Max. Rough. Height		*real value*	µm
Surface finish		Ground, Honed, Lathed, Polished	
General Surf. finish		Rough or Smooth	
Env Pressure		*real value*	MPa
B content		*real value*	ppm
Li content		*real value*	ppm
Dissolved hydrogen		*inlet value*	cc/kg
Dissolved oxygen		Avg. value [inlet+outlet+1st/last cycle]	ppb
Conductivity		inlet value	mS/cm
pH		average value	

WEB APPLICATION

Data mining web app

- A specific web application was developed during the INCEFA SCALE project. The main objective of this web application is to analyze all current fatigue data available and future data to be generated in the following years.

- The web application is open to the public:

https://inesco.shinyapps.io/incefascale/

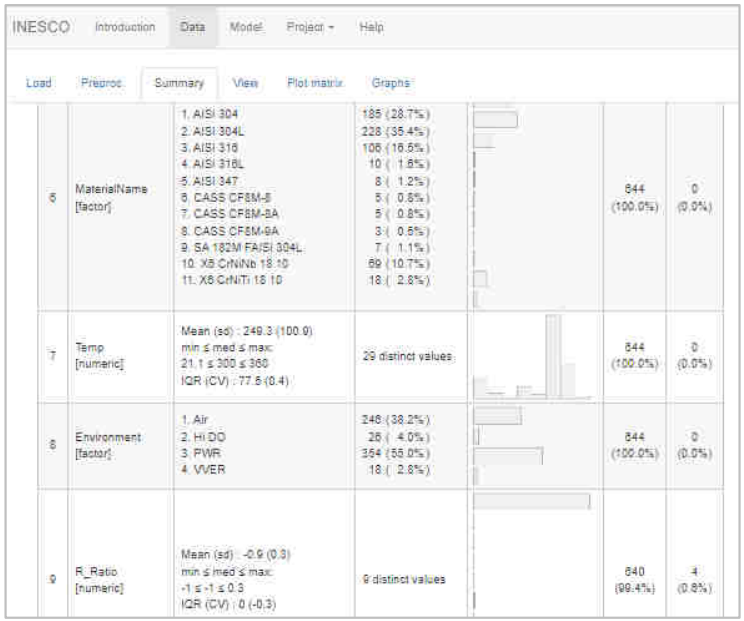
- All users can analyze their fatigue data without restriction, but the input files must be structured in a specific way explained in the web application.

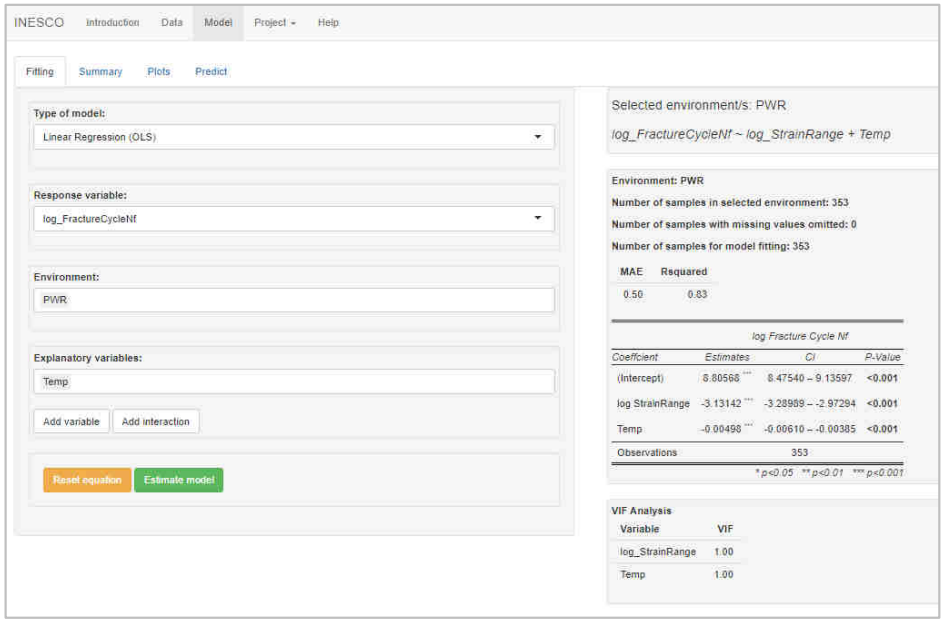
- The web application works with two types of models: linear regression and neural network (using three neurons).

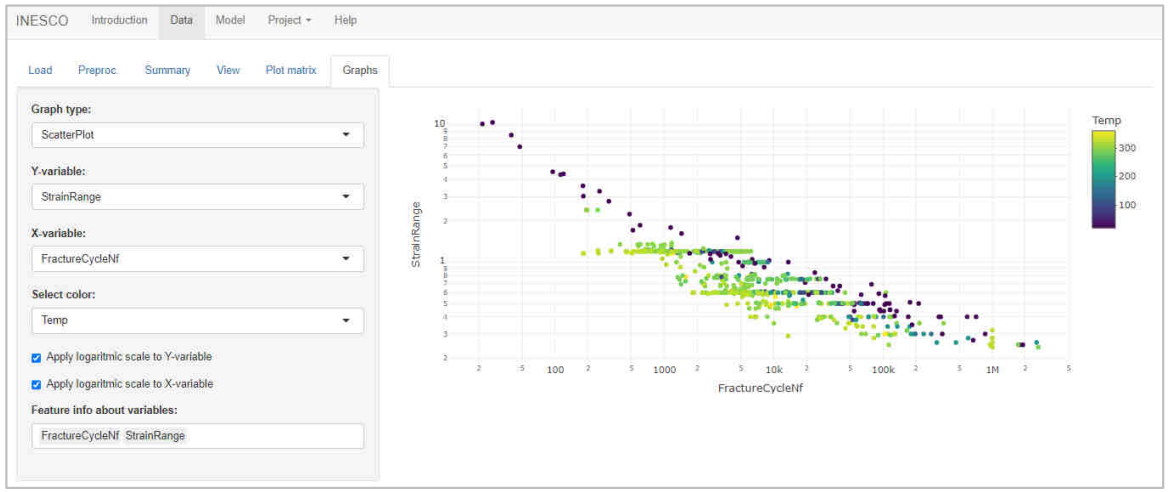
- The application allows analyzing all the fatigue parameters that have or may have influence on the fatigue life of the steels (only the variable “Fracture Cycle Nf” can be selected as target variable).

WEB APPLICATION

Data mining web app







DATA MINING

Analysis

- The web tool will be used to compare fatigue models based on the data available.
- Different models will be developed based on the data used:

	Data	Environment
Case 1.1	NUREG/CR-6909	AIR
Case 1.2	ALL	AIR
Case 2.1	NUREG/CR-6909	PWR
Case 2.2	ALL	PWR

- The first objective is to determine solid predictive models. The MAE parameter will determine the accuracy of each model (see next slide).
- The second objective is to determine the influence of the (10) different variables of analysis (see slide 10) have on the number of cycles to failure.

DATA MINING

Quality of the models

- To measure the **quality of the model** the **MAE** parameter is used**:**
  - **MAE**: Mean absolute error. Average error in the model prediction.

    NOTE*: All errors are measured in ln(FC), being FC= Fracture Cycles estimated by the model*

- A **MAE of 0.3** means, **on average**, that the model is **off** in $\pm$**0.3 units of ln(FC)**.

  On average, the real values will be $\ln(FC) \approx \ln\left(\widehat{FC}\right) \pm 0.3$

- Translating this into **real cycles**:
  - $FC \approx \mathrm{e}^{\ln(\widehat{FC})+0.3} = \widehat{FC} \cdot e^{0.3} = \boldsymbol{\widehat{FC} \cdot 1.349}$
  - $FC \approx \mathrm{e}^{\ln(\widehat{FC})-0.3} = \widehat{FC} \cdot e^{-0.3} = \widehat{FC}/1.648 = \boldsymbol{\widehat{FC} \cdot 0.741}$

F.ex: $\underline{\widehat{FC} = 10000\ cycles \quad \gg \quad 7410 < \mathrm{FC} < 13490}$

NOTE: *R squared (**R²**) is also available in the web tool*

DATA MINING

CASE 1 – DATA NUREG/CR-6909 IN AIR

	CASE 1 (89 TESTS)	MAE
MODEL 1.1	**Strain Range** $\ln(N) = 8.917 - 2.982 \cdot \ln \varepsilon$	0.49
MODEL 1.2	**Strain Range + Temp** $\ln(N) = 9.146 - 2.989 \cdot \ln \varepsilon - 0.001926 \cdot T$	0.48

** The other variables are not present in the model due to their negligible influence on the model's accuracy (p-value <0.001)*

DATA MINING

CASE 2 – ALL DATA AVAILABLE IN AIR

	CASE 2 (318 TESTS)	MAE
MODEL 2.1	**Strain Range** $\ln(N) = 8.826 - 2.907 \cdot \ln \varepsilon$	0.40
MODEL 2.2	**Strain Range + Temp** $\ln(N) = 9.112 - 2.938 \cdot \ln \varepsilon - 0.001618 \cdot T$	0.39

** The other variables are not present in the model due to their negligible influence on the model's accuracy (p-value <0.001)*

It can be observed that the inclusion of more fatigue data from other sources (Case 2) leads to a reduction in the Mean Absolute Error

DATA MINING

CASE 3 – DATA NUREG/CR-6909 IN PWR

	CASE 3 (38 TESTS)	MAE
MODEL 3.1	**Strain Range** $\ln(N) = 8.826 - 2.907 \cdot \ln \varepsilon$	0.73
MODEL 3.2	**Strain Range + Temp** $\ln(N) = 10.20 - 3.109 \cdot \ln \varepsilon - 0.00932 \cdot T$	0.48
MODEL 3.3	**Strain Range + Temp + Strain rate** $\ln(N) = 9.198 - 2.840 \cdot \ln \varepsilon - 0.006813 \cdot T + 210.2 \cdot \varepsilon'$	0.39

** The other variables are not present in the model due to their negligible influence on the model's accuracy (p-value <0.001)*

DATA MINING

CASE 4 – ALL DATA AVAILABLE IN PWR

	CASE 4 (393 TESTS)	MAE
MODEL 4.1	**Strain Range** $\ln(N) = 7.368 - 3.104 \cdot \ln \varepsilon$	0.57
MODEL 4.2	**Strain Range + Temp** $\ln(N) = 8.819 - 3.114 \cdot \ln \varepsilon - 0.00501 \cdot T$	0.52
MODEL 4.3	**Strain Range + Temp + Strain rate** $\ln(N) = 8.267 - 2.859 \cdot \ln \varepsilon - 0.003650 \cdot T + 189.3 \cdot \varepsilon'$	0.45

** The other variables are not present in the model due to their negligible influence on the model's accuracy (p-value <0.001)*

DATA MINING

CASE 3 VS CASE 4

If the most accurate models from both databases (Model 3.3 vs Model 4.3) are compared, it can be observed that the inclusion of more fatigue data from other sources does not lead to a reduction in the MAE.

		MAE
MODEL 3.3	**Strain Range + Temp + Strain rate** $\ln(N) = 9.198 - 2.840 \cdot \ln \varepsilon - 0.006813 \cdot T + 210.2 \cdot \varepsilon'$	0.39
MODEL 4.3	**Strain Range + Temp + Strain rate** $\ln(N) = 8.267 - 2.859 \cdot \ln \varepsilon - 0.003650 \cdot T + 189.3 \cdot \varepsilon'$	0.45

This fact may lead to thinking that the model with more data is less accurate, that is, the error is higher.

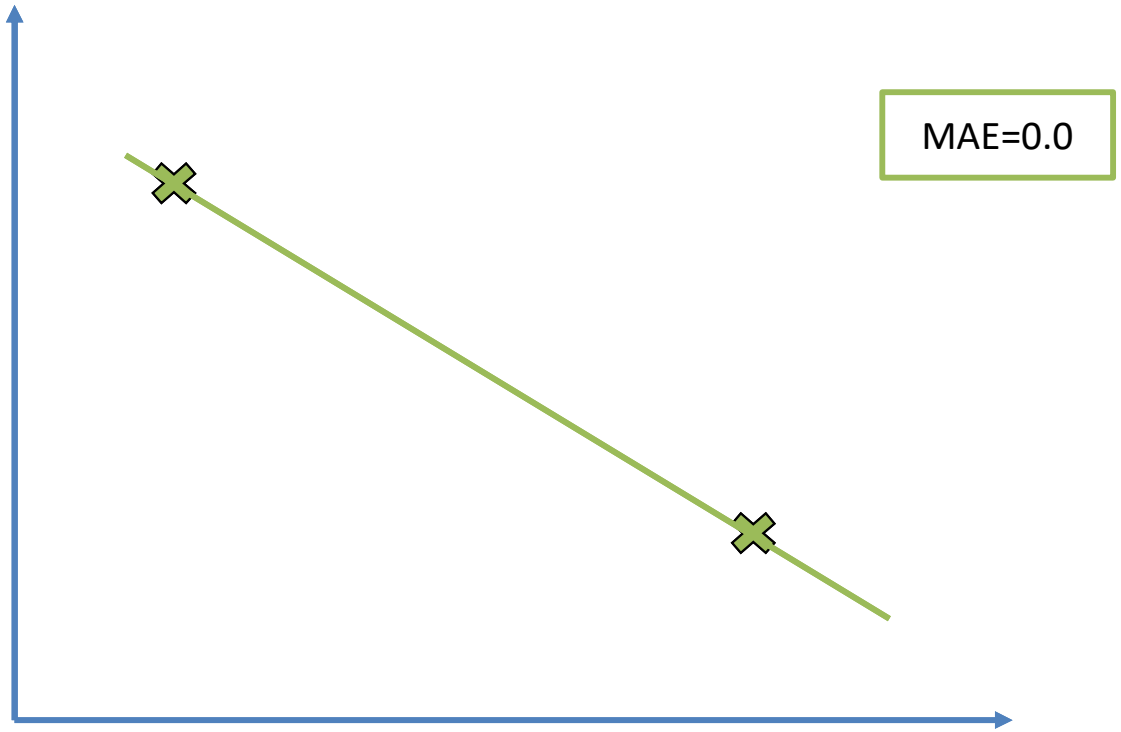
- Firstly, including a more significant number of points means that, although the MAE "is" less precise, the predictive coefficients of the model's equation are more solid.
- Secondly, the comparison between the models should be done by calculating the MAE when all the fatigue data (from all sources) are evaluated using the equation obtained for Model 3.3. This specific analysis determines a MAE of 0.465, higher than the MAE obtained for Model 4.3.

DATA MINING

CASE 3 VS CASE 4

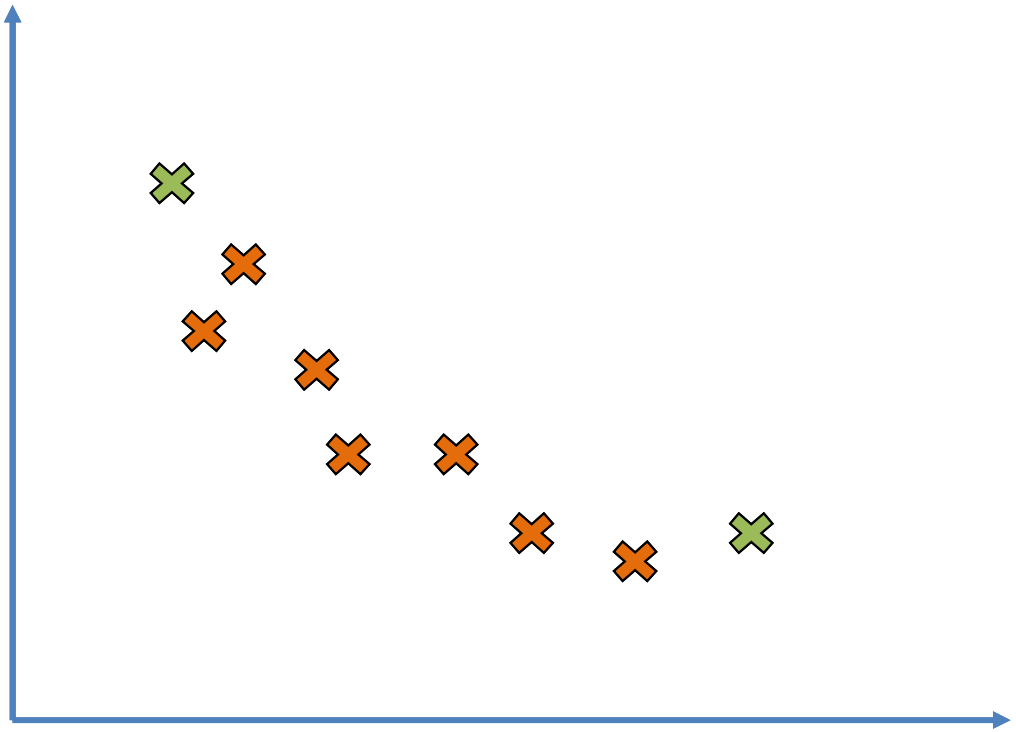
DATA MINING

CASE 3 VS CASE 4



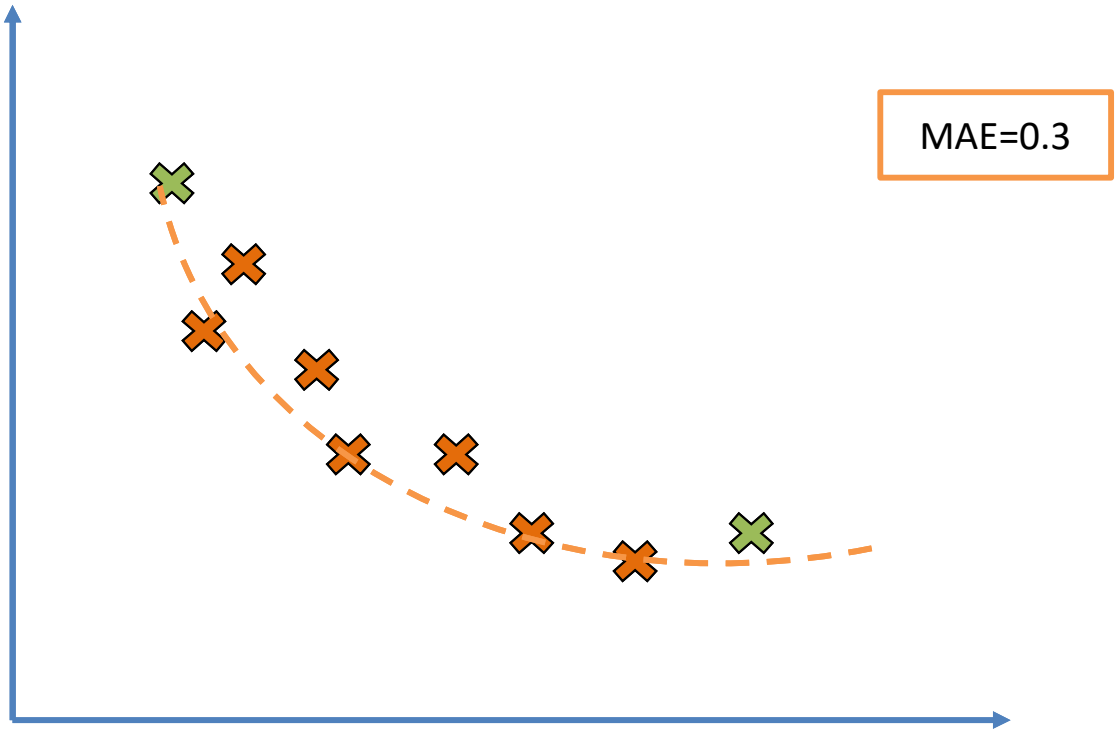
DATA MINING

CASE 3 VS CASE 4



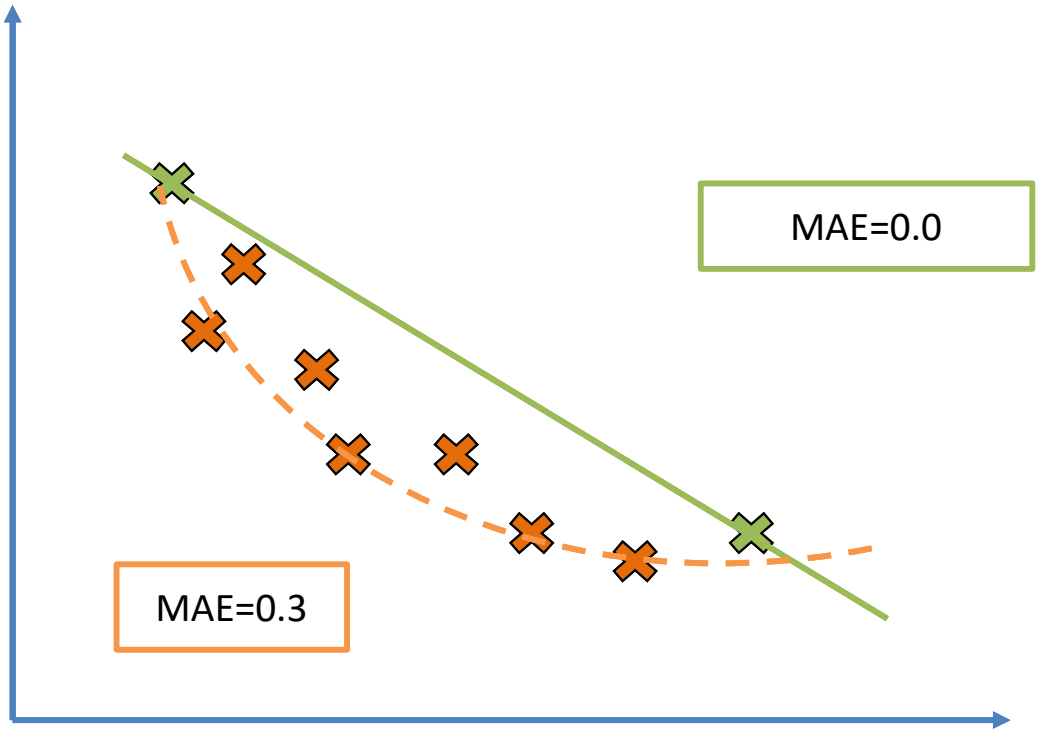
DATA MINING

CASE 3 VS CASE 4



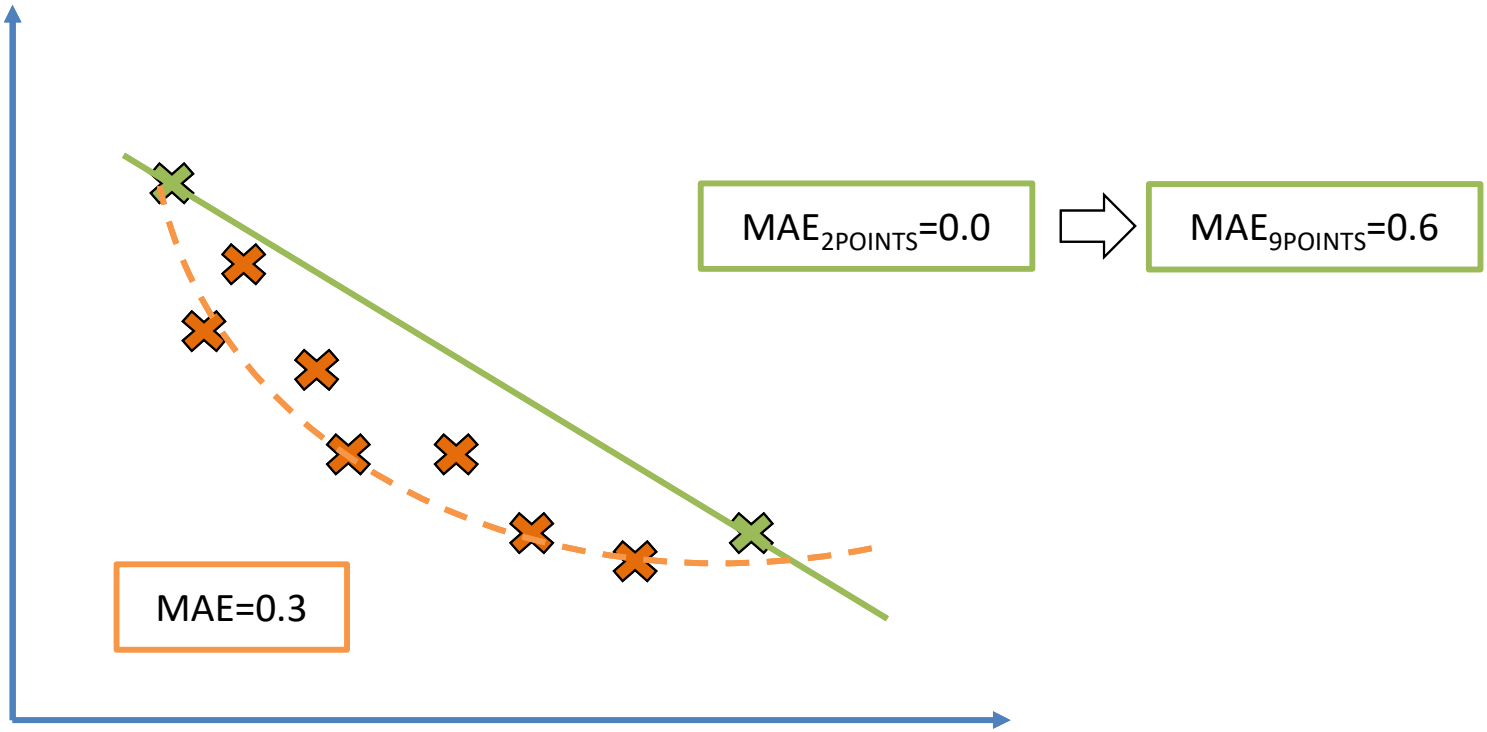
DATA MINING

CASE 3 VS CASE 4



DATA MINING

CASE 3 VS CASE 4



CONCLUSIONS

LESSONS LEARNED DURING INCEFA SCALE PROJECT

- In recent years, significant efforts have been made to eliminate some knowledge gaps about the fatigue behavior of steels in NPPs.

- The increase in the number of tests and fatigue data allows for the generation of more solid and reliable predictive models.

- The new models in PWR environment have not yet been better explained by any additional variable not currently considered in existing models. In the case of models in Air, the calculation accuracy is increased if temperature is included as an explanatory variable (Model 2.2 vs Model 2.1).

- The INCEFA-SCALE project is in progress. New fatigue data will be obtained to eliminate some knowledge gaps (additional variables not currently considered?), thus increasing the reliability of predictive models.

- IFD will be a great tool for improving the models and increasing the accuracy.

- A new evaluation tool is available for future evaluation of new models based on new fatigue data that may be generated.