Development of a Modified Kernel Regression Model for a Robust Signal Reconstruction

Ibrahim Ahmed & Gyunyoung Heo
Department of Nuclear Engineering, KHU, Yongin-si, Korea

October 26-28, 2016

Contents

Introduction (1/2)

- To achieve the desired operating configuration in any process, the system status needs to be accurately measured and monitored.

- To operate within the desired limits, it is important to know the reliability of plant measurements of the processes.

- The demand for robust and resilient performance has led to the use of online-monitoring techniques to monitor the process parameters.

- Many empirical models are used in online monitoring of the process parameters. One of those models is kernel regression (KR).

Introduction (2/2)

- KR is a nonparametric technique for estimating a regression curve without making strong assumptions about the shape of the true regression functions.

- Unlike parametric model, a nonparametric model is algorithmic estimation procedures which assumes no significant parameters.

- KR needs no training process after its development when new observations are prepared, this indeed good as long as a system characteristics are also changing due to ageing phenomenon.

- No need to understand the underlying physics of the system since KR is developed on the data from an operating system.

Problem Statement (1/2)

- KR, both the inferential and auto-associative models, has limitation in time-varying data that has several repetition of the same signal, particularly, if those signals are use to infer the other signals.

- Whereas many situations are related to variation and fluctuation during normal state, such as transient – start-up and shutdown mode.

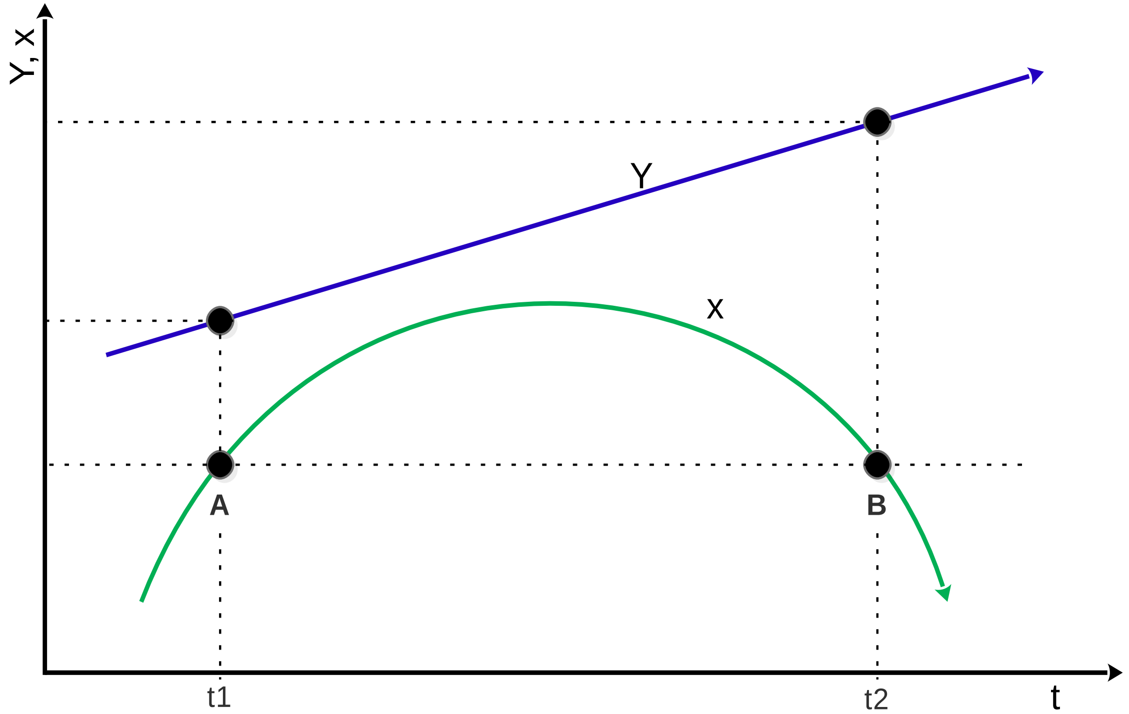
- The major problem of KR in such a condition is that, the values of dependent variable y in those points of the same value of the predictor variables x assume value of the average of those dependent variable values.

- Accurate estimation of the process signal can leads to the proper understanding of the equipment behaviors.

Problem Statement (2/2)

- Therefore, in this work, we proposed a modified KR model that is capable of resolving this setback of the conventional KR.

- This can improve the efficiency/robustness, usability, and applicability of the empirical modeling for process and equipment monitoring, and prognostics.



Motivation & Objective

- *Motivation* – the online monitoring and the process signal validation are crucial especially in case of an accident when the abnormal changes of the process together with possible severe damage of process sensors can occur.

- Preliminary investigations indicated that, developing a model to functions in wider range of application, especially, in time-varying signals will
  - ➢ Resolve the identified problem of KR,
  - ➢ lead to provisions of early warning information about the normal operational transient, &
  - ➢ Enhance the operational performance of the plant.

- Therefore, specifically, the main objective of this work is
  - ❖ To develop a robust model for signal reconstruction through the modification of KR
  - ❖ To improve the accuracy and applicability of the model for signal reconstruction in normal transient state and early warning alert for proper understanding of the process and equipment behaviors.

Kernel Regression Overview

Given a matrix of memory data of predictor variables X and the response variable y as

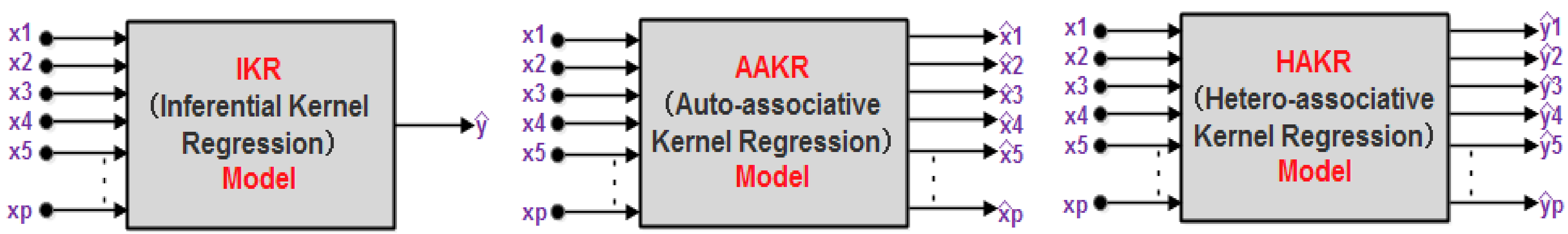
$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{bmatrix}$$

where p is the number of predictor variables, n is the number of memory vectors, and x(i,j) is the ith observ ation of the jth predictor variable.

For any observed query vector, $\mathrm{x}_q = \begin{bmatrix} x_1 & x_2 & \cdots & x_p \end{bmatrix}$

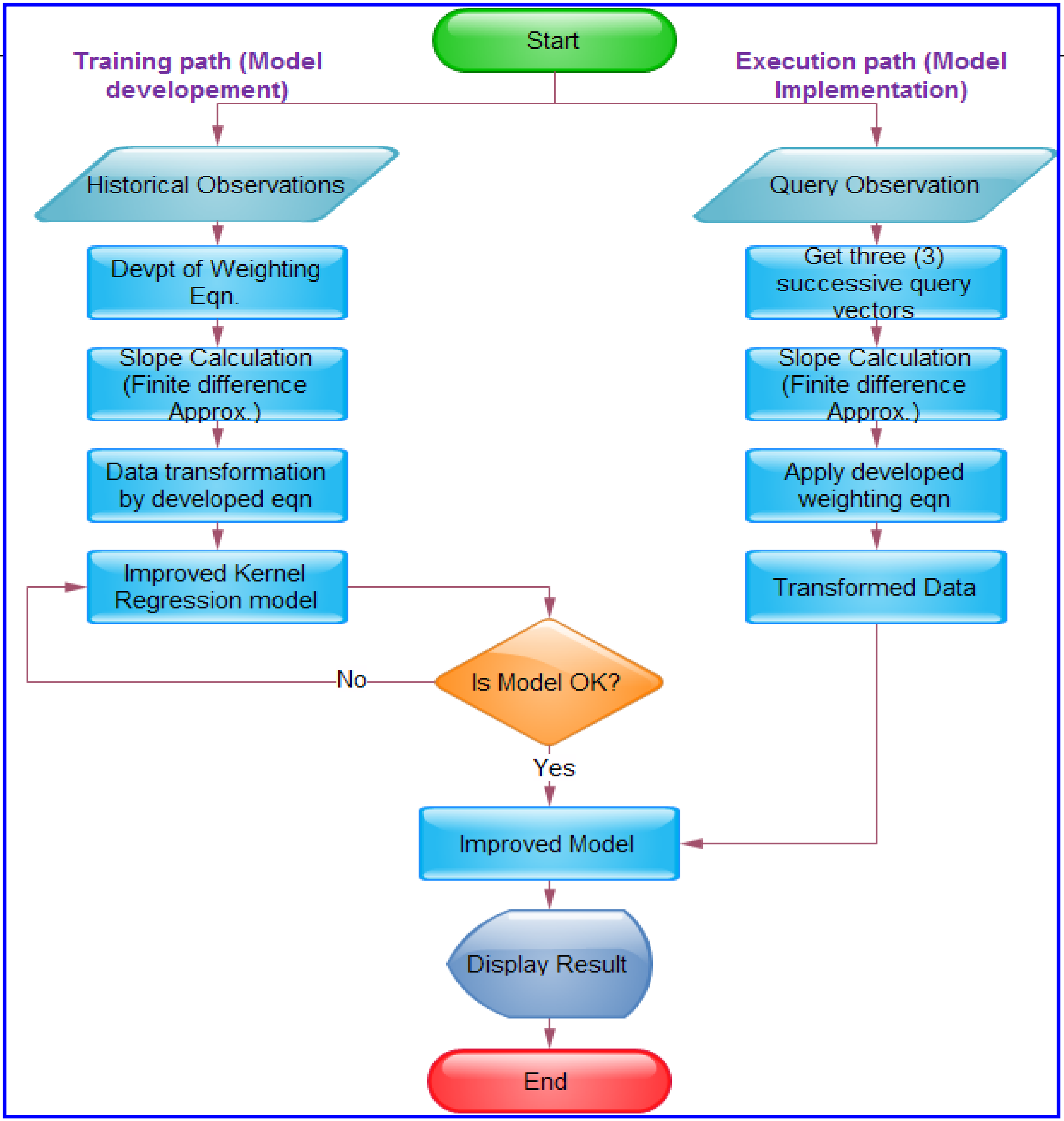
The Euclidean distance, d, the kernel weight, k (using Gaussian kernel), and the estimated response va riable , $\hat{y}$ are calculated as

- ***Euclidean distance,*** $\boldsymbol{d_i(X_i, x_q) = \sqrt{(x_{i,1} - x_1)^2 + (x_{i,2} - x_2)^2 + \cdots + (x_{i,p} - x_p)^2}}$
- ***Gaussian kernel weight,*** $\boldsymbol{K_i(X_i, x_q) = exp\left(\frac{-d_i^2}{2h^2}\right)}$
- ***Output Estimation,*** $\boldsymbol{\hat{y}_i(\, x_q) = \frac{\sum_{i=1}^{n} K(X_i, x_q).\, y_i}{\sum_{i=1}^{n} K(X_i, x_q)}}$



Methodology (1/4)

❑ *Model flowchart*



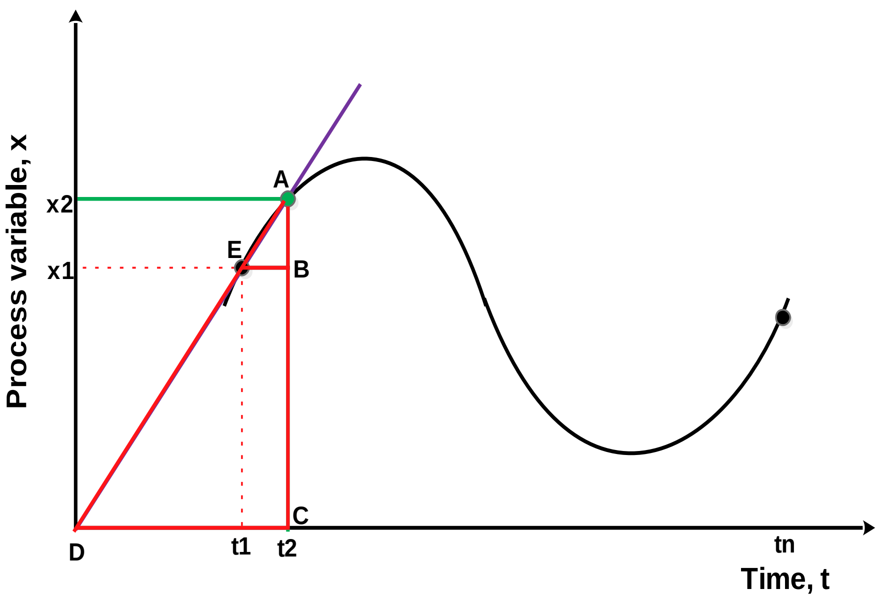
NOTE:

- **The modification is mainly on the improvement of measure of similarities used in KR.**

Methodology (2/4)

Weighting/Transformation Equation

- Consider the diagram below as a function of the process variable.
- Let's consider two data points x_1 and x_2 and their corresponding occurrence times t_1 and t_2 respectively.
- It is obvious from the diagram that, the two triangles ΔABE and ΔACD are similar.



- Hence, by applying the theorem of similar triangles, we have

$$\frac{AC}{AB} = \frac{DC}{EB}$$

- This implies that

$$\frac{x_2}{x_2 - x_1} = \frac{t_2}{t_2 - t_1} \Rightarrow x_2 = t_2\left(\frac{x_2 - x_1}{t_2 - t_1}\right)$$

$$x_2 = t_2\left(\frac{\Delta x}{\Delta t}\right) \approx t_2 * m$$

We called the resulted data point ω

- Thus $x_2 = t_2 * m_2 \Rightarrow \boxed{\omega_{i,j} = t_i * m_{i,j}}$

For multivariate

Methodology (3/4)

Gradient/slope approximation

- m is determines at each data point using finite difference derivative approximation.

Taylor Series Expansion ⟶ $$x(t+\Delta t) = x(t) + \Delta t\frac{dx(t)}{dt} + \frac{(\Delta t)^2}{2!}\frac{d^2x(t)}{dt^2} + \frac{(\Delta t)^3}{3!}\frac{d^3x(t)}{dt^3} + \cdots\cdots$$

Forward difference

$$\boxed{m_{i,j} = \frac{dx}{dt} = \frac{-x_{i+2,j} + 4x_{i+1,j} - 3x_{i,j}}{2\Delta t}}$$

Backward difference

$$\boxed{m_{i,j} = \frac{dx}{dt} = \frac{3x_{i,j} - 4x_{i-1,j} + x_{i-2,j}}{2\Delta t}}$$

Central difference

$$\boxed{m_{i,j} = \frac{dx}{dt} = \frac{x_{i+1,j} - x_{i-1,j}}{2\Delta t}}$$

Thus, we define a slope matrix, $M = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n,1} & m_{n,2} & \cdots & m_{n,p} \end{bmatrix}$

$$\boxed{\omega_{i,j} = t_i * m_{i,j}}$$

- In this equation, the information from previous input vectors is incorporated into the KR through the slope m and time t of the current input vector.
- This in fact will gives a more details representation of the estimations compare to that of the conventional KR that ignored any information leading to the current data point.

Methodology (4/4)

The resulted Distance

- Hence, by applying the transformation equation, the memory/train data X and the query/test data can be transformed as follows:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{bmatrix} \Rightarrow W = \begin{bmatrix} \omega_{1,1} & \omega_{1,2} & \cdots & \omega_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{n,1} & \omega_{n,2} & \cdots & \omega_{n,p} \end{bmatrix}$$

$$\mathrm{x}_q = \begin{bmatrix} x_{q,1} & x_{q,2} & \cdots & x_{q,p} \end{bmatrix} \Rightarrow Wq = \begin{bmatrix} \omega_{q,1} & \omega_{q,2} & \cdots & \omega_{q,p} \end{bmatrix}$$

- Thus, the Euclidian distance for any given query vector can be calculated by

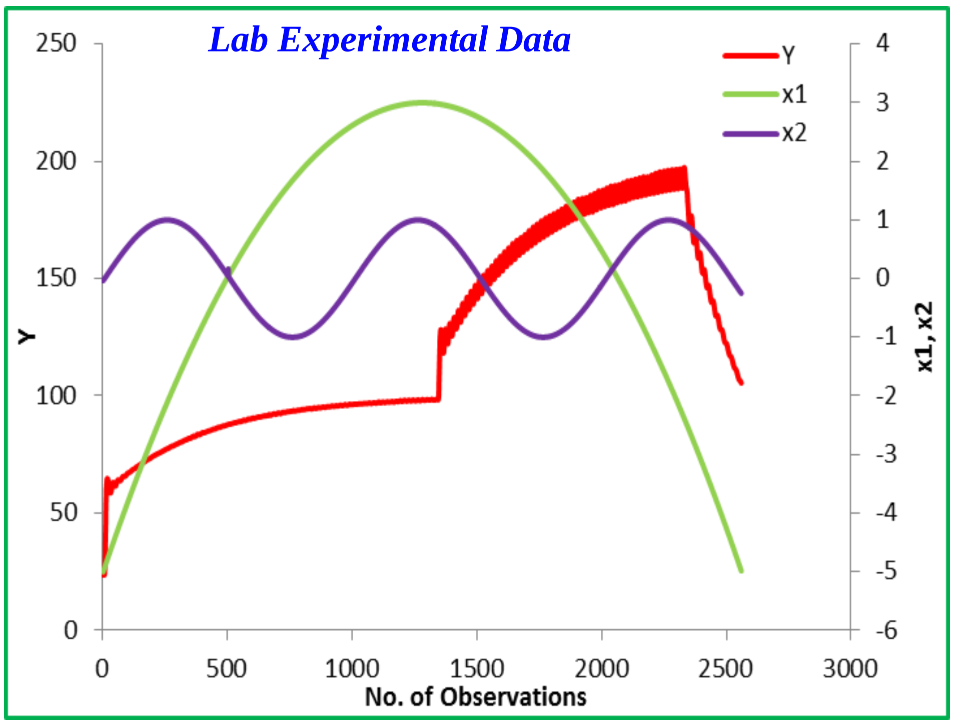
$$\boldsymbol{d_i(X_i, x_q) = \sqrt{(x_{i,1} - x_1)^2 + (x_{i,2} - x_2)^2 + \cdots + (x_{i,p} - x_p)^2}}$$

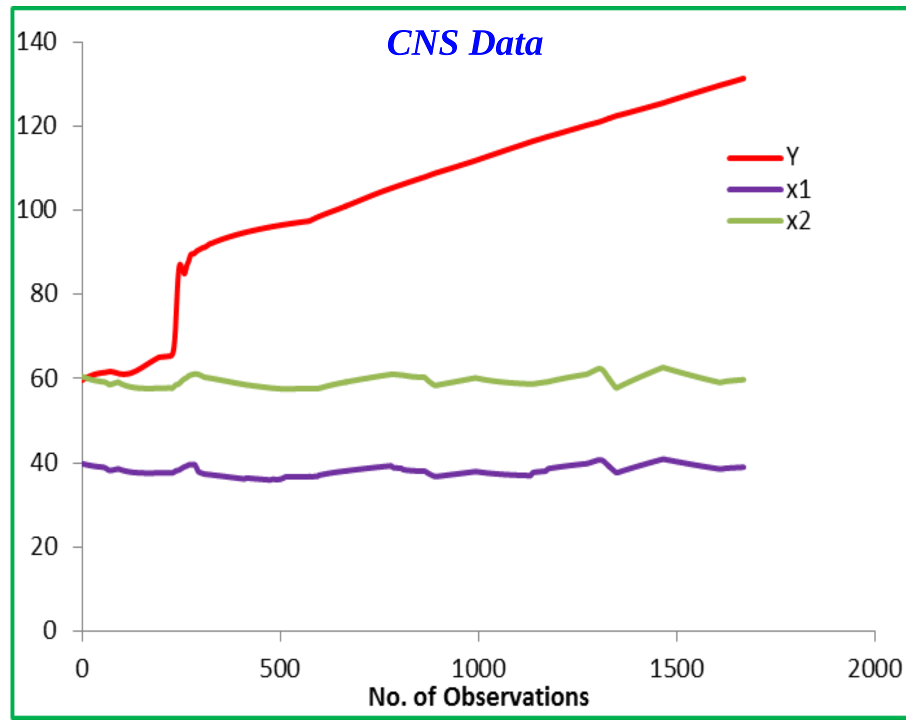
$$\boxed{\boldsymbol{d_i(X_i, x_q) = \sqrt{\frac{(\omega_{i,1} - \omega_{q,1})^2}{\sigma_1^2} + \frac{(\omega_{i,2} - \omega_{q,2})^2}{\sigma_2^2} + \cdots + \frac{(\omega_{i,p} - \omega_{q,p})^2}{\sigma_p^2}}}}$$

Where σ_i^2 = the variance of the transformed predictor variable i.

Test & Results (1/4)

- To test the performance of the developed model, two different kinds of datasets are used.
- Lab experimental data generated from heat conduction experiment to evaluate the performance of the model and compare the results with the conventional KR.
- Real time simulation data of heating and cooling from compact nuclear simulator (CNS) provided by KAERI.





Test & Results (2/4)

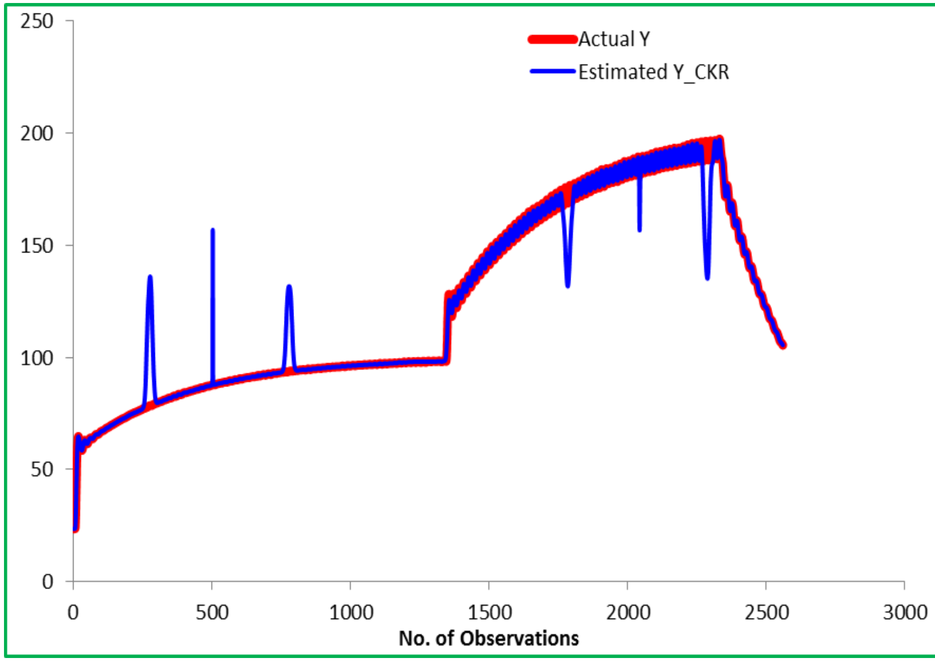
- x_1 and x_2 are independent variables that are used to estimate/predict the value of Y.
- In the lab expt. data, Y is a temperature data, while x_1 and x_2 are the artificially generated data.
- In the CNS data, all the variables are the real plant simulation data

- The measure of performance of the model used is mean square error (MSE).
- MSE is used as a measure of accuracy of the prediction of the models.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_{iest} - y_i)^2$$

- where y_{iest} is the estimated value and y_i is the actual value. N is the number of data points/observations used as test data.

Test & Results (3/4)

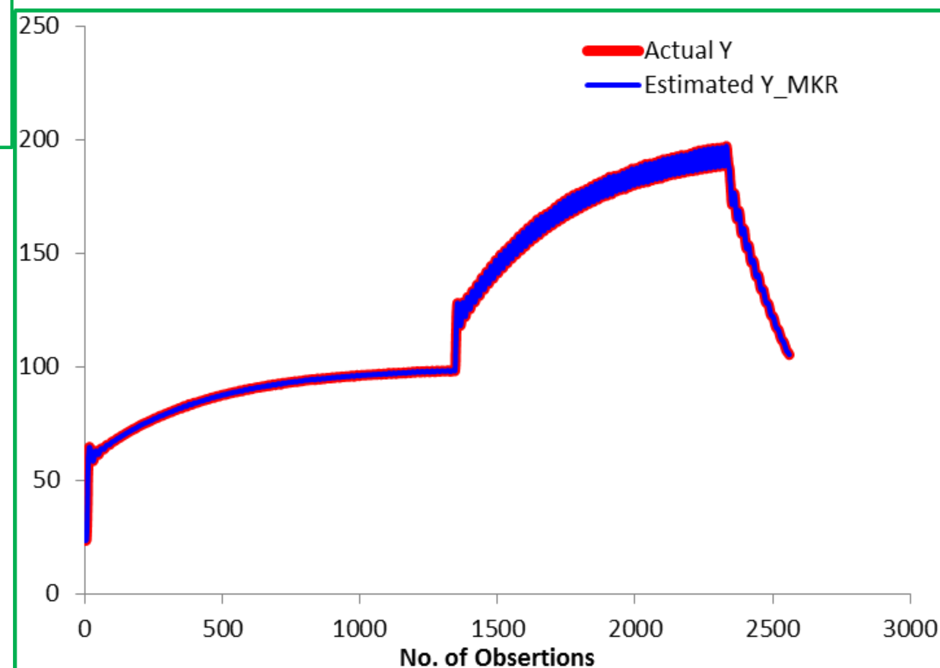


Conventional KR model

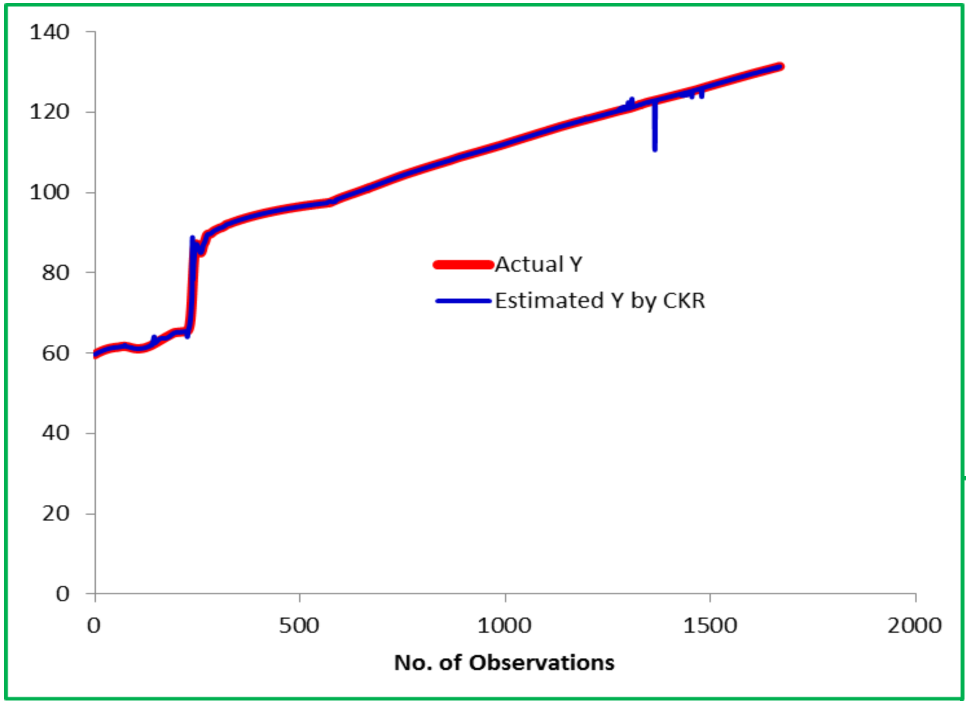
Results of the Lab Experimental Data

MODELS	LAB EXPERIEMENT	CNS DATA
Conventional KR **MSE (%)**	**63.26**	14.43
Modified KR **MSE (%)**	**0.086**	0.009

Modified KR model



Test & Results (4/4)

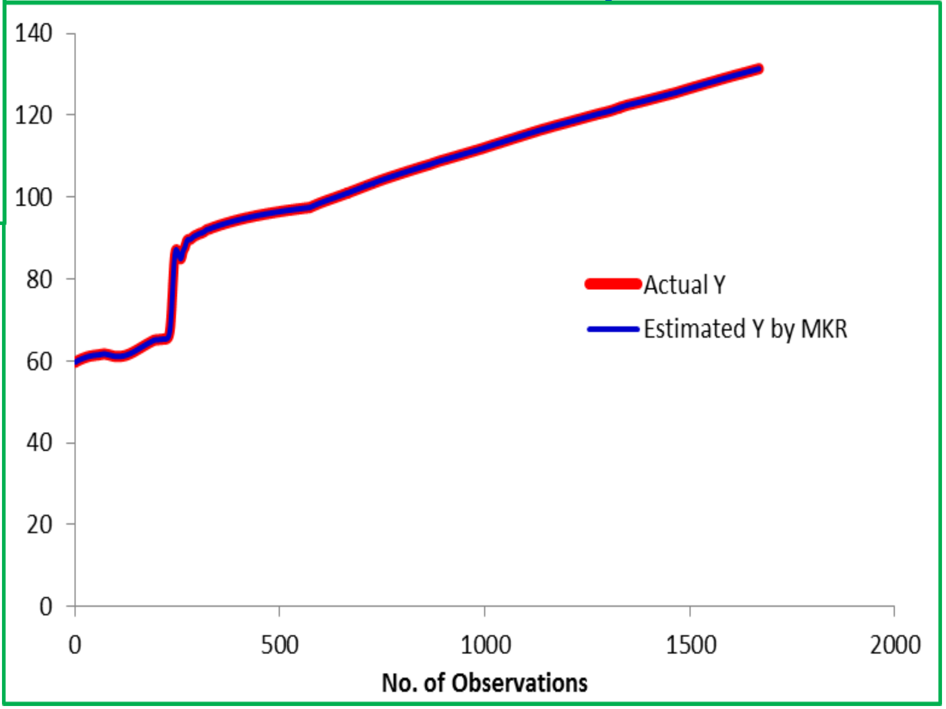


Conventional KR model

Results of the CNS Data

MODELS	LAB EXPERIEMENT	CNS DATA
Conventional KR **MSE (%)**	63.26	**14.43**
Modified KR **MSE (%)**	0.086	**0.009**

Modified KR model



Conclusions

- The conventional KR has limitation in correctly estimating the signals when time-varying data with repeated values are used to estimate the dependent variable especially in signal reconstruction, validation and monitoring.
- we presented here in this work a modified KR that can resolve this issue in time domain.
- A time-dependent equation was developed first to transform the data into another space prior to the Euclidian distance calculation considering their slopes/changes with respect to time.
- The performance of the developed model is evaluated and compared with that of conventional KR using both the lab experimental data and the real time plant data of CNS provided by KAERI.
- The results show that the proposed model, having demonstrated high performance accuracy than that of conventional KR in signal reconstruction, is capable of resolving the identified limitation with conventional KR, and can be used to improve process and equipment monitoring applications.

Thank You