ITNT - 2023
The IX International Conference
on Information Technology and
Nanotechnology

Method of Temporal Interpolation of the Corroding Gas Pipeline Wall Thickness Values Coordinated with a Physical Model

R.R. Gabbasov[1], R.A. Paringer[1,2]

1. Introduction

In many areas of human activity, in particular, in the context of time series analysis, relevant is also the problem of predicting (regressing) certain values. This paper is devoted to the problem of regression of pipe thicknesses tied to two wells of the same gas field. The piping of the wells considered in this paper are equipped with sensors that take readings of the physical characteristics of the passing condensate. Data from these sensors is used as input features for regression models. Thickness measurements are made in different places at different times on the well piping using ultrasonic diagnostics. We propose a method for interpolating thicknesses in the time domain that takes into account the physical characteristics of the of the passing condensate.

2. Dataset Description and Preparation

In this work, we used data from two wells (with the names "2-2" and "3-1") of the same gas field. For each of them, there were values of 17 time-varying parameters taken in 1-hour increments from sensors located on the well piping: pressure and temperature values taken at different piping locations (13 parameters), data on condensate flow, CO_2 content, and pH. In the process of machine learning, these parameters act as input features. The process of smoothing the outliers and filling in the gaps in these data was carried out using a sliding window. Also, for each well there were values of the target parameter – the results of measuring the wall thickness of the piping components at different piping locations at different times. Thickness values, in contrast to sensor readings, are widely spaced in time. In this paper, two methods of interpolation of this parameter were considered: a method using quadratic splines and **our proposed method of interpolation** coordinated with the **physical model**, i.e. produced in accordance with **physical parameters** of the transported gas condensate. For both methods of interpolation, two data sets were prepared for further experiments respectively.

2.A. Quadratic Spline Interpolation

The character of the intensity of the corrosion process is related to **the rate of change of the wall thickness**, so the models were trained to regress the values of this rate. Since the rate of change of the parameter is known to be the derivative of the parameter function, using linear interpolation would cause the rate of change of thickness between two known original timestamps to degenerate into a constant, so quadratic interpolation is used.

2.B. Interpolation coordinated with physical parameters

This interpolation method, which we propose, is based on the use of calculations according to the NORSOK M-506 standard.
This standard calculates the theoretical corrosion rate of a pipe (in mm/year) at a fixed point in time, based on the pressure, temperature, and pH of the flow, its CO_2 content, as well as the pipe diameter, wall roughness, and wall shear stress (the latter three parameters are known from technical documentation). With hourly pressure, temperature, CO_2, and pH data, we obtained hourly values for the theoretical wall thinning rate for each piping component. These values were converted to mm/h and averaged over the day, and then used as weights for the intensity of thinning between the two original measurements: the thinning value was brought into line with the available measurements, taking into account these weights as multiplication factors. Having obtained thickness values with a frequency of once a day in this way, at the final stage – in order to match the sampling rate of 17 features (1 hour) – we used quadratic spline interpolation, again guided by the ideas that were listed in the previous paragraph.
The graph shown in Figure 1 shows that between the measurement points, which are indicated by large diamonds, nonlinearity was added, which allows taking into account intermediate well operation modes to achieve greater detail.

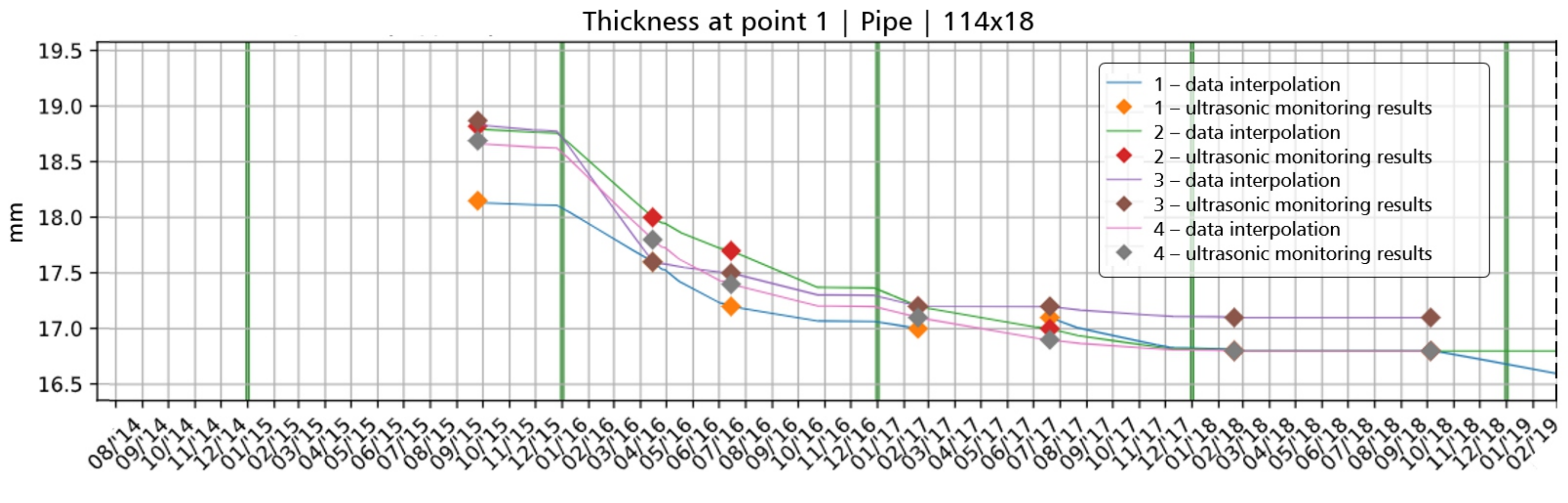

Fig. 1. Example of thickness values after using interpolation coordinated with the physical parameters of the condensate

3. Experiments Setup

Since the data under consideration is a time series, for each target value, a feature time window of size 3600 (3600 hours = 150 days = 5 months) was considered. As a regression model, this paper uses a linear regression model with the additional use of the RANSAC algorithm with the following hyperparameters: the minimum proportion of selected random elements is 0.1, the error function is quadratic. The Huber loss function regressor and Theil-Sen regressor models were also considered, but it was found that their training takes too much time relative to the training time of the regressor using RANSAC with comparable results in terms of accuracy. For each point on the pipe, its own regression model was trained.

4. Experiments Results

4.A. Models Similarity

Despite the fact that for each point on the well piping, its own regression model was trained, the fact that all these points belong to one single piping, through which the same gas condensate mixture passes, allows us to expect similarities in the behavior of models for different points on the piping. Therefore, the measure of the similarity of the models was further considered (regarding how they behave during regression). Due to the aforementioned facts, the more similar the behavior of the models will be, the greater the measure of correspondence of the trained models to physical reality will be.
The measure of similarity between the two models was calculated as follows. On the time interval between two neighboring ground-truth thickness measurements (from the training sample), the validation values of the rate of wall thinning were regressed on 17 features using both models. Between the two resulting vectors of values, the R2 metric was calculated, and in the case of obtaining a negative value, it was set to 0. The resulting number (from 0 to 1) was understood as a measure of the similarity of the models.
Figure 2 shows models similarity measure values for all models of the well "3-1" calculated using both quadratic spline interpolation and the interpolation coordinated with physical parameters.

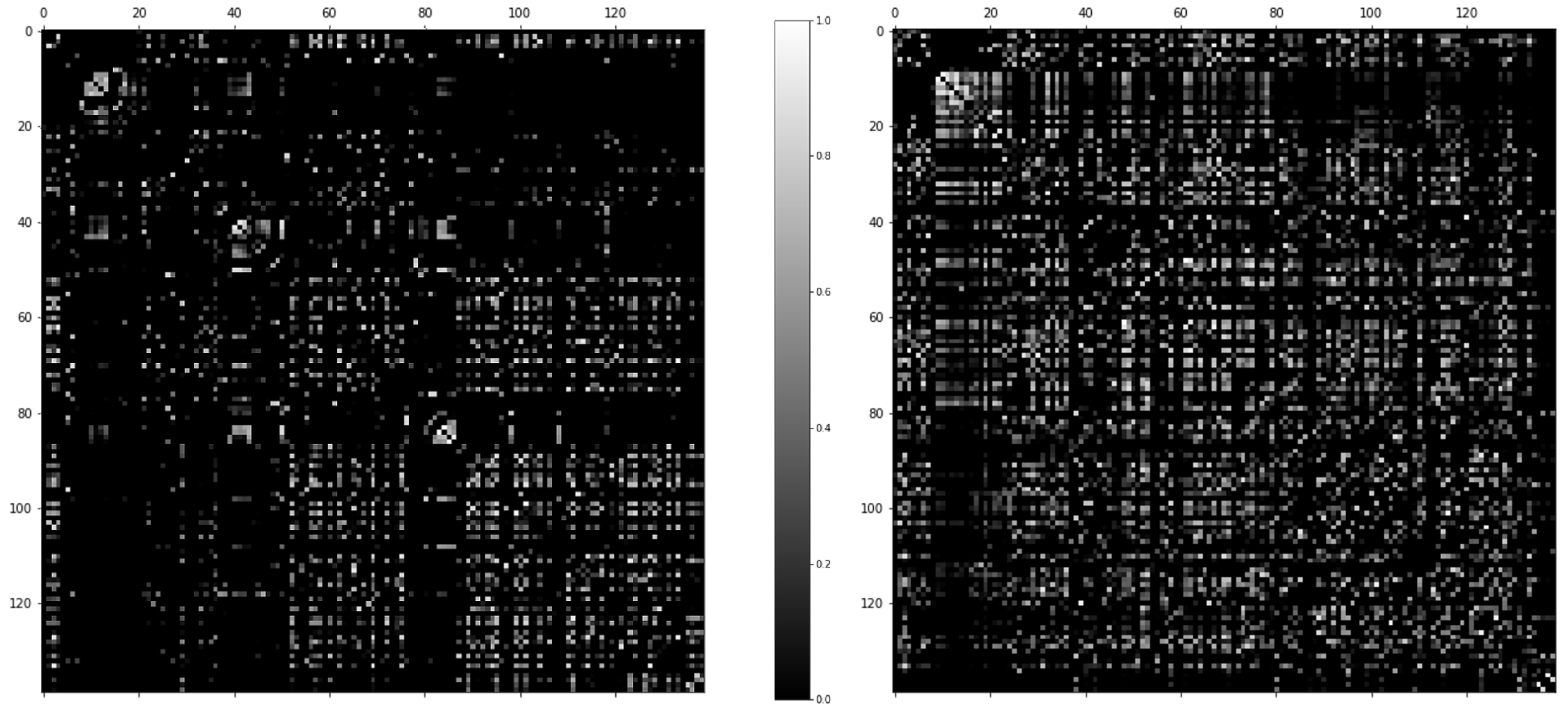

Fig. 2. Models similarity measure values for all models of the well "3-1" calculated using quadratic spline interpolation (on the left) and the interpolation coordinated with physical parameters (on the right)

Next, a binary relation of similarity was introduced, which included pairs of models with a measure of similarity between themselves greater than 0.9 (i.e. such models were considered similar), and then, according to this binary relation, the models were divided into groups (groups of similar models).
The first metric of correspondence of the trained models to physical reality (M_1) was defined as the average size of the groups of models.
As a result of the experiment, it was found that the use of the interpolation method coordinated with the physical parameters leads to higher similarity values, which, in turn, contributes to a more intensive grouping of models. Table I shows the resulting values of the M_1 metric for two wells considered in the paper ("2-2" and "3-1") for two methods of interpolation of the target parameter.

Table I. M_1 values

Well	2-2	3-1
Quadratic spline interpolation	6.87	4.32
Interpolation coordinated with physical parameters	10.23	10.57
Ratio	*1.49*	*2.45*

4.B. Relative Importance of Features

The relative importance of the features used in the models for the regression of the wall thinning at points in various sections of the piping (near the wellhead, before the choke, after the choke) was assessed. The initial features were formed into 6 sets (F1-F6) depending on the location of the corresponding sensors on the well piping. Then, a set of 6 regression models was trained for each measurement point on the piping, and each model was trained on the corresponding set of features F1-F6.
Then the models were divided into three classes: near the wellhead, before the choke, and after the choke – depending on the location of the corresponding measurement point on the piping. Then, for each class, for each set of features, the average value of the L_1 metric for the corresponding models was calculated between the values regressed by the models and the initial values of the target parameter on the test sample. The obtained values were normalized and inverted so that the smallest value of L_1 corresponded to 1, and the largest values of L_1 corresponded to numbers less than 1. The obtained numbers were understood as the significance of the features.
The results of measuring the feature importance for models trained using two different methods of interpolation of the target value are presented in Figures 3 and 4.

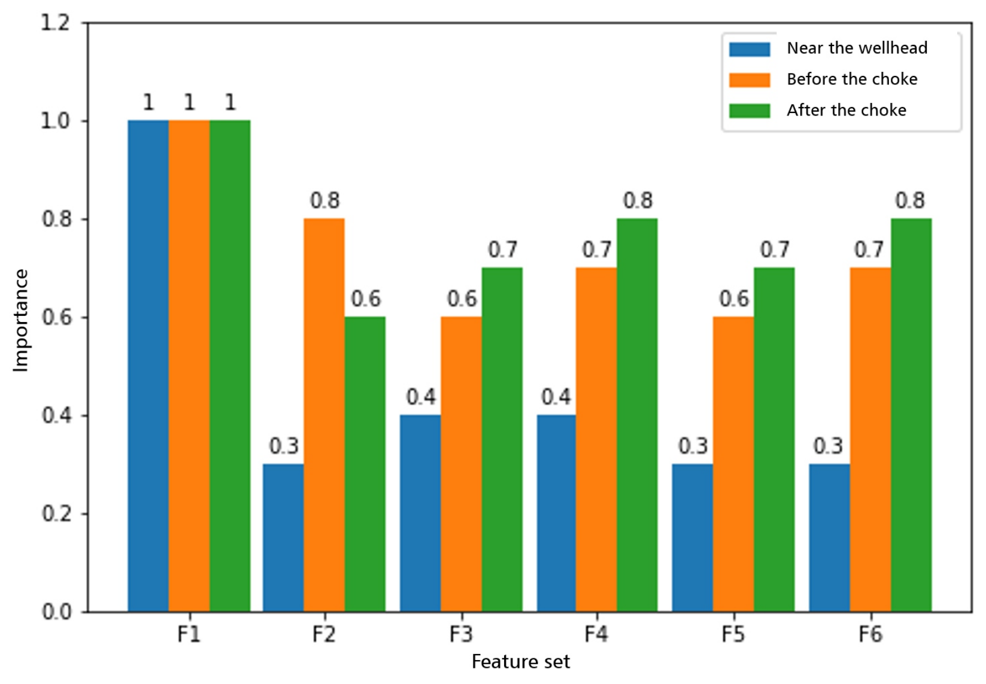

Fig. 3. Feature importance plot for models of the well "2-2" trained using quadratic spline interpolation

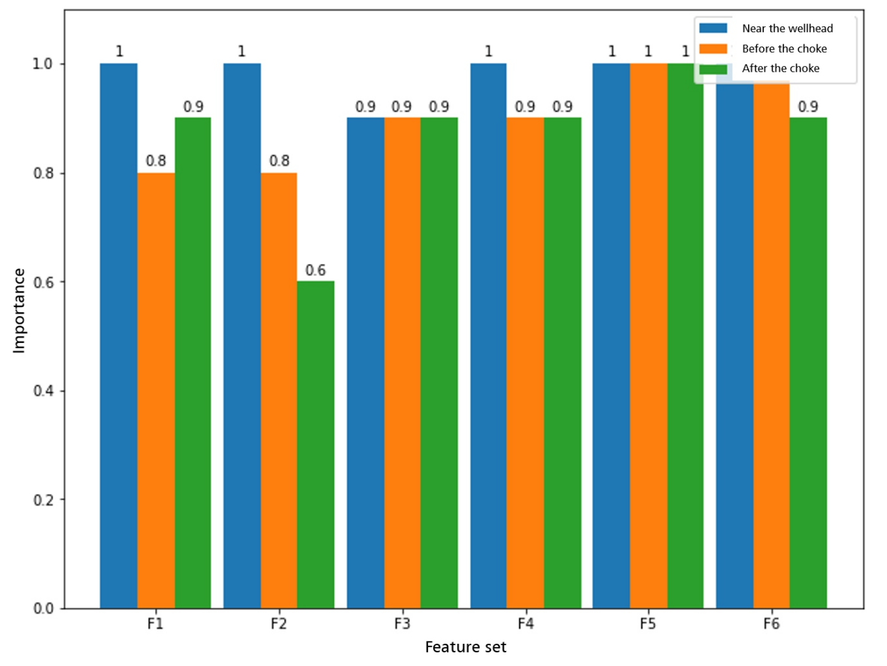

Fig. 4. Feature importance plot for models of the well "2-2" trained using interpolation coordinated with physical parameters

The second metric of correspondence of the trained models to physical reality (M_2) was defined on the basis of the obtained feature importance values as follows:
$M_2 = (L_1(\boldsymbol{i}, \boldsymbol{i'}))^{-1}$,
where $L_1(*,*)$ is the L_1 metric,
$\boldsymbol{I} \in \mathbb{R}^{18}$ is the feature importance vector for different classes (6 feature sets × 3 model classes),
$\boldsymbol{i'} \in \mathbb{R}^{18}$ is the vector, consisting of the same value, which is the median of the vector i.
Table II shows the resulting values of the M_2 metric for two wells considered in the paper ("2-2" and "3-1") for two methods of interpolation of the target parameter.

Table II. M_2 values

Well	2-2	3-1
Quadratic spline interpolation	0.31	0.24
Interpolation coordinated with physical parameters	0.77	0.83
Ratio	*2.48*	*3.46*

Conclusion

The result obtained allows us to conclude that the use of the proposed method of interpolation of the target parameter instead of more simple quadratic spline interpolation leads to an increase in the values of the metrics of correspondence of the trained models to physical reality: the value of the measure M_1 increased by 1.49 and 2.45 times for wells 2-2 and 3-1, respectively, and for M_2, the corresponding increases were 2.48 and 3.46 times. It should be noted that the consistency of the results for both proposed measures indicates a high level of their reliability.

САМАРСКИЙ УНИВЕРСИТЕТ
SAMARA UNIVERSITY

[1] SAMARA NATIONAL RESEARCH UNIVERSITY
34, Moskovskoye shosse, Samara, Russia (www.ssau.ru)

ИСОИ РАН

[2] IMAGE PROCESSING SYSTEMS INSTITUTE
151, Molodogvardeyskaya st., Samara, Russia (www.ipsi.smr.ru)