DATA MINING TO PREDICT HYBRID LASER ARC WELDING IMPROVEMENTS IN
SHIP INTERIM PRODUCT ASSEMBLY

D Kolich, University of Rijeka, Croatia
Y L Yao, Columbia University, USA
R Neuberg, Columbia University, USA
R L Storch, University of Washington, USA
N Fafandjel, University of Rijeka, Croatia

SUMMARY

Most shipyards use gas metal arc welding (GMAW) or submerged arc welding (SAW) technologies in their steel panel line assembly processes. Since the welding work on panel assembly lines make up to 50 percent of total ship welds, improving the process by decreasing the duration time as well as energy use and filler materials will decrease production costs. The advent of lasers in production yields advantages of faster speed and smaller distortions. However, the shortcomings of using exclusively lasers to weld steel plates are the lack of filler material. Therefore hybrid laser arc welding (HLAW) takes advantage of the positive sides of both laser and gas metal arc welding to assemble panels with shorter duration times. Much data exists about assembling panels with conventional welding means. Therefore a data mining methodology using an interpretable regression method for accurately predicting the duration time outcomes of panels assembled with HLAW technology will allow shipyard management to evaluate better decisions to transform their assembly lines by integrating laser technology in order to become more competitive through the reduction of production costs.

NOMENCLATURE

GMAW Gas metal arc welding
HLAW Hybrid laser arc welding
LBW Laser beam welding
SAW Submerged arc welding
R² Coefficient of determination

1. INTRODUCTION

The standard welding processes in a panel assembly line consist of either gas metal arc welding (GMAW) or submerged arc welding (SAW). However with the increased use of lasers in various manufacturing industries, some shipyard management have become aware that panel assembly production can be significantly improved upon by integrating lasers with their standard welding processes [1,2]. These improvements include shorter duration time, reduced number of motions and fewer distortions of the plates as well as less rework, which is in compliance with lean manufacturing principles. This all translates to smaller production costs.

The integration of laser technology with its "clumpy" equipment requires capital investment as well as training workers and certifying a new welding procedure at its production facility. In order to convince management to invest in this new technology it is helpful to use and apply existing data about assembling panels, where the panel assembly duration time and workshop design characteristics such as length, width, number of plates, number of stiffeners, length of stiffeners, and plate thickness are known. Therefore the use of a data mining methodology to predict the assembly duration time after the laser technology has been integrated into the process will be helpful.

The traditional method for determining the duration time to assemble an interim product is to develop a detailed Gantt chart for each interim product from which it is possible to calculate the production times [3]. While very accurate, the drawback to developing a detailed Gantt chart for each interim product is that it would require an experienced production planner at least 4 hours per interim product. Since a typical commercial vessel consists of thousands of interim products this translates to 10000 or more engineering man-hours.

Data mining is useful in multiple industries for classification, prediction, clustering as well as association and concepts. In this paper, the duration time for the assembly of various panels and micropannels is used along with 11 predictors such as panel width, panel length, stiffener length, and number of stiffeners. The outcome of duration is listed for each of these interim products. In order for shipyard management to decide upon transforming the technology of their standard panel assembly lines, it is helpful to determine a methodology so that shipyards can predict the percentage improvement in applying the HLAW technology. Any type of change is naturally a drawback both for employees and management. The only incentive for change is when management is convinced that an additional investment will bring significant production improvements which translate to cost savings. For the purposes of this paper the data mining method known as tree regression is used to create a model that is interpretable and accurate as well as easy to use by most shipyard workers. A set of 229 different interim products consisting of large panels,
and three types of micro-panels were used to create a model, where 70 percent of the data is used for training and 30% for testing the accuracy of the model. The authors demonstrate a tree regression model that is accurate and easy to use.

2. BACKGROUND

The purpose of data mining is to take advantage of the plethora of data available by applying the proper algorithms that can analyse large amounts of data accurately and efficiently and make practical use of new information that we would like to uncover in the near future. This includes the creation of accurate predictive models, classification models, as well as for clustering, concept description and association [5]. The logarithmic increase in the use of computers in product design as well computer numerically controlled machines means that every day more data is made available in manufacturing databases [6]. Data mining in creating an accurate predictive model is used for the purposes of this paper. Being able to accurately predict the behaviour of a manufacturing system is important for any company [7]. Therefore, the same is true for any shipyard, especially in predicting the welding duration times with both arc welding and HILAW technologies.

The authors Kaluzny, et al used a data mining application for estimating the total shipbuilding costs of naval vessels using characteristics such as ship type, range, total length, number of torpedo decoys, helicopters supported, etc. in order to develop a model tree that will help in predicting the cost of a new type of naval vessel [8]. However, the methodology developed does not treat the costs or duration times of intermid product assembly nor is it geared for commercial vessel production which is the aim of this paper.

Data mining clustering analysis has been performed on shipyard interim products in order to group them together in a technological way according to Kolich et. al [9]. Since data mining consists of prediction, classification and clustering, for the purposes of this paper the author’s were interested in developing a prediction methodology that is practical and easy for shipyard staff to use in estimating welding and assembly duration times.

3. WELDING METHODS

The traditional welding methods used during panel assembly include gas metal arc welding (GMAW) and submerged arc welding (SAW). Panel assembly includes joining and tack welding up to 5 steel plates at the first workstation. At the second workstation, the steel plates are welded first on one side, and then turned to the other side in order to be welded again. On the third workstation, the panel is traced and necessary holes are cut according to workshop drawings and documentation. Then grinding work of the welds and edges of the panel are also performed. At station four, stiffeners are added and hydraulically pressed and then welded to the steel plates using a double-sided welding technique.

Presently many shipyards still use the above mentioned standard arc welding procedures along the panel assembly line. However, since laser beam welding is used in many industries such as the automobile industry and the electronics industry, it has become necessary to analyse whether laser technology can be applied in a shipyard atmosphere. During the past decade some advanced shipyards have started integrating lasers into some of their production processes [2, 10].

3.1 GMAW

Gas metal arc welding (GMAW) has been used in the shipbuilding industry since the 1950s. The advantages are that the equipment is relatively cheap. There is a torch, which is connected to a wire feeder, shielding CO₂ gas and a power source. The three in conjunction create the arc, necessary to melt the metal which acts as filler material that bridges the gap between the two steel plates, while the CO₂ gas shields the weld from oxidation and ensures that pores do not develop either. As a result there is a good weld which will usually last the life cycle of the ship.

The disadvantages of GMAW are that it often requires high heat input in order to produce a good weld, which translates to a distorted steel panel, which afterwards needs to be levelled and therefore increases the amount of man-hours in the shipyard. Another drawback according to Katayama is shallow penetration of welds and slow welding speeds [11]. This means that the total assembly time is relatively long. Since shipyard management constantly strives to reduce all types of costs including production costs, it would be helpful to analyse how the welding process itself could be made more efficient.

3.2 LASER BEAM WELDING

This leads us to the analysis of pure laser beam welding (LBW), which has existed for the past two decades in various industries, usually for the welding of thin plates according to Reutz [12]. Its advantages are that in comparison to arc welding it is faster and uses less heat, which results in less distortion and therefore less rework of levelling of the material upon welding. The disadvantages of LBW are that since there are no filler materials, the weld metal dominates the properties according to Kristensen [13]. There is a “high degree of overmatching in strength” by a factor of 2. The major disadvantage of LBW is the result of flaws during the cooling or solidification process. These flaws are usually small and isolated but occur with “certain regularity along the weld.” [12]. Therefore, LBW while being faster than GMAW produces an inferior weld quality, and therefore will not be certified by the classification societies. Furthermore, pure LBW is not able to bridge
typical gaps between steel plates since it does not provide filler material. Likewise, “the focused energy of the laser beam results in a narrow heat affected zone (HAZ) that can lead to steep spatial and temporal thermal gradients that can result in brittle microstructures” [12].

Since laser beams are very precise, the problem with this is that only the seam itself and only a little space around it are heated, which results in a thin weld [15]. Since it is normal during the welding of thick steel plates to have wide gaps between the steel plates, the LBW will result in an undercut, and even in some of the laser beam passing through the gap instead of welding the thick structural steel plates together. Therefore, this would be a waste of heat and result in an unacceptable weld quality.

According to Ribic et al, the LBW also has the potential to crack due to the presence of the brittle phase which is due to the lower cooling rate [16]. Likewise according to Katayama, the “fast cooling rates associated with lasers can lead to centreline cracking, hot cracking, , liquidation cracking, or formation of brittle and non-ductile solidification microstructures”. Likewise, there are also safety issues with eye safety of “Nd:YAG, fiber lasers and direct diode lasers” [11].

According to Katayama, the advantages of laser welding are “high quality, high precision, good flexibility and low deformation or distortion”. The drawbacks are high cost of equipment for laser welding, “small gap tolerance and easy formation of welding defects such as porosity in deeply penetrated weld fusion zones” [11].

During the 1970s, Professor William M. Steen from Imperial College London pioneered the technology that involved combining arc welding along with laser welding to solve the problem with LBW [13]. Initially since the first lasers used in the laboratory tests were 2-kW, increasing the strength of the lasers by other researchers resulted in the new idea to show promise during the last twenty years. The shipbuilding industry was the primary force in enticing more research in this field. Today laser hybrid welding is used by industry as a result.

3.3 HYBRID LASER ARC WELDING

In order to maintain the LBW benefits of faster speed and smaller thermal distortions, while removing the solidification flaws, the GMAW method is integrated into the welding process to form a HLV method. Therefore, panels with higher thicknesses can also be welded more efficiently.

Therefore the total advantages of HLV are:

- Deeper and wider root opening
- Necessary to weld on only one side
- Lower duration time
- Lower use of filler materials than GMAW
- Less distortion and less rework
- Less energy used than GMAW

According to Katayama, hybrid welding with “CO₂, YAG, diode, disk, or fiber laser and TIG, MIG, MAG, plasma or another arc heat source” has advantages of “deeper penetration, higher welding speeds, wider gap tolerance, better weld bead surface appearance and reduced welding defects leading to a smaller amount of porosity in addition to complements of the drawbacks of both individual processes” [11].

The disadvantages of HLV include higher initial cost to purchase and integrate the new technology in the assembly process and the necessity to invest in the training of workers who have no familiarity with the new technology. Therefore, developing a model to predict the welding duration times of using the HLV method would be very useful in helping shipyard management decide as to whether this investment is justifiable which would be based on the type of interim products that the shipyard produces.

According to Ribic et al, the laser beam has the exclusive affects on the weld pool depth, whereas the arc welding has the exclusive affect on the weld pool width and “gap bridgability” [16].

According to Kelly et al, HLV is 4 times faster than arc welding. Likewise, arc welding also uses 6 times more filler material, and adds about 4 times more heat and distortion. Therefore, it can be concluded that HLV is definitely advantageous over standard arc welding [17].

4. INTERIM PRODUCT DATA ANALYSIS

The interim products that have been analysed in this paper include semi-automatically assembled panels, robotically assembled micropanels, manually assembled panels and large panels (See Figure 1). During the downstream ship production process these above mentioned interim products are further assembled to create built-up panels which then when assembled with other regular panels make up the erection blocks [18].

Since each year a typical medium to large sized shipyard builds approximately 6 to 10 or even more ships yearly, and since each ship consists of thousands of interim products, no two which are exactly alike, there is a superabundance of information about ship interim products. This plethora of data concerning interim products can be harnessed in a database in order to take advantage of data mining techniques. No two interim products are exactly the same and each interim product is unique with regards to assembly time and man-hours. Therefore by creating and training a mathematical model based on the readily available database, it is possible to use it to predict outcomes such as assembly duration time for new interim products for new ships not yet built, but still in the design phase. It is necessary for shipyards to be able to accurately and quickly predict the duration times for each new interim product. Since drawing up
accurate Gantt charts would require thousands of engineering planning hours, the development of a mathematical model using data mining techniques to predict the duration times quickly and accurately would be a very handy tool. Likewise making use of all of the available characteristics of each interim product would make the model accurate and robust as well.

There are many interim products where the welding times are less than 100 minutes. These would be the micropanels since they are smaller. There are also some welding times at around 200 minutes which represents the manually assembled micropanels. Finally over to the right, the large panels take over 600 minutes welding time because they are longer and have more butt and fillet seams to weld. See Figures 1 and 2.

![Diagram](image1)

**Figure 1:** a) semi-automatically assembled micropanel; b) robotically assembled micropanel; c) manually assembled micropanel; d) semi-automatically assembled panel

The mathematical models analyzed in this paper include linear regression and tree regression. Linear regression is the most common method used for predictions. However, linear regression is not as accurate as other regression methods such as tree regression. Likewise tree regression develops a model which is accurate, and practical and easy to follow.

In order to reap the benefits of the multiple characteristics of each interim product the entire list of characteristics is included. This includes a set of data with 11 predictors: panel mass, plate thickness, panel length, panel width, stiffener height, stiffener thickness, stiffener length, stiffener number of types, stiffener number, stiffener tensile strength, and plate tensile strength and one outcome, duration time. This was done to a total of 229 interim products from an actual ship which consists of large panels and micropanels assembled in four different processes. Developing a histogram of values of the welding duration time is illustrated in Figure 2.

![Histogram](image2)

**Figure 2:** Histogram of arc welding times of the interim products

In comparison the histogram showing the welding times for the HLAW interim products is in Figure 3 below. As expected, the higher duration times are smaller in comparison to the arc welding times of Figure 2. For instance, the higher duration times are about 600 minutes for panels welded with arc technology, whereas the duration times for HLAW are around 80 minutes. Likewise, the spread of the duration times for all the interim products has significantly decreased. For instance the duration times of HLAW of Figure 3 are between 5 and 80 minutes, as opposed to the higher 50 to 600 minute spread for arc welding technology in Figure 2.

![Histogram](image3)

**Figure 3:** Histogram of HLAW welding times of the interim products

Moving forward to a clear comparison, the following figure 4 is generated, where the y-axis represents the HLAW welding time and the x-axis represents the arc...
welding time. There is a diagonal line and data points which show the value for both arc and HILAW welding. All the points in Figure 4 lie below the diagonal line. This means that the arc welding time is greater. There is more discrepancy between the arc and the laser as we move to the right of the figure because the interim products get bigger, and the efficiency of HILAW comes in place. Whereas the micropanels are smaller and usually consist of one steel plate with different configurations of stiffeners, the large panels consist of multiple steel plates. These steel plates need to be welded on both sides before the stiffeners are welded using arc welding technology. However, with HILAW technology, welding on one side suffices.

Figure 4: Arc welding time vs. HILAW time in minutes on the same plot

The Figure 5 shows the ratio of the welding time HILAW over the welding time using arc technology. The first set of points on the left which represent the smaller micro-panels are about 25% of the duration time of HILAW as compared to arc welding. As we move to the right of Figure 5, the data points show an even smaller percentage of the HILAW times at around 12 percent. This makes sense because during the use of arc welding technology for the assembly of large panels, the steel plates need to be turned over and welded on the second side in order to have a satisfactory assembly weld. However, with the use of HILAW it is not necessary to turn the steel plates to the second side, because one-sided welding is one of the advantages of HILAW over GMAW.

Therefore, the welding time ratio HILAW to ARC is 25% which corresponds to a 400% improvement in duration time for the set of data points to the left representing the micro-panel assembly. However, as we move to the right where there are big panels, the improvement is actually an 800% improvement because the panel does not need to be turned over and welded on the other side (See Figure 5).

Figure 5: Relative perspective

5. LINEAR REGRESSION VS. TREE REGRESSION MODELS

Linear regression was performed on all 11 predictors listed earlier for each interim product. These data sets can be called tuples. There are a total of 229 interim products or tuples. The linear model is trained with the tuples of the training set which are 70% of the tuples or 160, whereas 30% on the training set or 69 tuples. Then backwards regression elimination is done using the criteria of eliminating the predictor with the highest p-value and testing the performance of the model with R^2 comparisons. The best linear model is found to be six predictors that we keep, whereas the others were eliminated using individual t-tests (See equation 1). Panel length, stiffener length, stiffener thickness, stiffener number, stiffener tensile strength are the most important as judged by the p-values. See Table 2 below. Notice how using the p-value criteria, the number of predictors has decreased from 11 to 6. The rest are not necessary. The R^2 value is 0.921, which means that we have a very accurate model [19]. Equation 1 was developed from using the intercept and the coefficients for each predictor generated during the training of 70% of the tuples seen in Table 3 below.

Table 2: Testing of the predictors

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.684</td>
<td>5.103e+00</td>
<td>0.330</td>
<td>0.742</td>
</tr>
<tr>
<td>Panel length</td>
<td>9.326e-04</td>
<td>1.995e-04</td>
<td>4.760</td>
<td>4.44e-06</td>
</tr>
<tr>
<td>Stiffener length</td>
<td>5.642e-03</td>
<td>4.022e-02</td>
<td>-2.224</td>
<td>0.0276</td>
</tr>
<tr>
<td>Stiffener thickness</td>
<td>4.282e-01</td>
<td>2.657e-01</td>
<td>-1.612</td>
<td>0.1529</td>
</tr>
<tr>
<td>Stiffener length</td>
<td>3.172e-04</td>
<td>2.407e-05</td>
<td>13.17</td>
<td>2e-16**</td>
</tr>
<tr>
<td>Stiffener no.</td>
<td>8.527e+00</td>
<td>1.562e+01</td>
<td>5.460</td>
<td>1.89e-16**</td>
</tr>
<tr>
<td>Stiffener ten_str</td>
<td>1.503e-02</td>
<td>9.344e-03</td>
<td>1.609</td>
<td>0.110</td>
</tr>
</tbody>
</table>

© 2015: The Royal Institution of Naval Architects
Table 3: Linear model coefficients and interim product check

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>Interim product 1</th>
<th>Interim product 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Panel (214)</td>
<td>Micro-panel (44)</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.684</td>
<td>11050</td>
<td>4800</td>
</tr>
<tr>
<td>Panel length $L_1$</td>
<td>9.326e-04</td>
<td>370</td>
<td>150</td>
</tr>
<tr>
<td>Stiffener mt $S_m$</td>
<td>5.642e-03</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Stiffener thick $S_t$</td>
<td>-4.282e-01</td>
<td>154700</td>
<td>8766</td>
</tr>
<tr>
<td>Stiffener length $S_l$</td>
<td>3.172e-04</td>
<td>8.527e-00</td>
<td>12</td>
</tr>
<tr>
<td>Total time from linear regression modeled</td>
<td>74.4</td>
<td>14.2</td>
<td></td>
</tr>
<tr>
<td>Total time actual</td>
<td>68.6</td>
<td>13.9</td>
<td></td>
</tr>
<tr>
<td>% difference</td>
<td>5.3%</td>
<td>2.2%</td>
<td></td>
</tr>
</tbody>
</table>

Equation one uses the coefficients from column 1 of Table 3 above with each corresponding predictor or interim product characteristic. The two columns to the right test out the equation for accuracy which is demonstrated as 5.3% off from the actual value for one of the large panels, and 2.2% off from one of the micro-panels. This shows that the linear model is accurate. Likewise, the $R^2$ value or coefficient of determination of the linear model is 0.92. Model accuracy is judged based upon how close the $R^2$ value is to one.

$D_{HLAW} = 1.684 + 0.00093P_1 - 0.0056S_m - 0.43S_t + 0.000317S_l + 0.85S_m + 0.015S_t$

5.1 Arc tree regression model

The advantages of tree regression is that it results in a model that is both accurate and very interpretable which translates to something that can be used by much of the shipyard staff to predict welding duration times of new not yet assembled interim products, regardless of experience (See Figure 6).

Tree regression works by making use of all of the predictors and sets if then statements until a set of logic rules are created that can easily be utilized once completed. The trees can handle many types of predictors without needing to pre-process. The model begins parsing the entire set of data and each predictor until it finds a predictor and "split value" that splits the "data into two groups ($S_1$ and $S_2$) such that the overall sums of squares error are minimized." See equation 2. Within each of the "groups S1 and S2", the method continues searching for the "predictor and split value that best minimizes the SSE. "Tree regression is also known as recursive partitioning because of the repetitive splitting nature."

$$SSE = \sum_{i=1}^{n} (y_i - \bar{y}_1)^2 + \sum_{i=1}^{n} (y_i - \bar{y}_2)^2$$

Figure 6: Tree for determining Arc welding times

5.2 HLAW tree regression model

If we compare the results of table 3 above with the results of table 4 below, we find that the tree regression values are very close to the actual values for interim products 1 and 2. For instance interim product 1 using tree regression of Figure 6 yields 3.1% difference (Table 4) from the actual welding duration time, compared to 5.3% from linear regression in Table 3. For interim product two in Table 4, the difference from the actual is 7.8% as opposed to 2.2% in Table 2. This goes to show that even though the tree regression model is a little bit more accurate on average due to a higher coefficient of determination, the linear regression model may be more accurate for certain interim products. Even so, the differences are relatively small so that the advantages of using the tree regression outweigh that of the linear regression model.

Table 4: Comparison of HLAW assembled interim product duration times of using both linear regression and tree regression

<table>
<thead>
<tr>
<th>Method</th>
<th>Interim product 1 (213)</th>
<th>Interim product 2 (44)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration time HLAW</td>
<td>68.6</td>
<td>13.9</td>
</tr>
<tr>
<td>Duration time HLAW from linear regression</td>
<td>74.4</td>
<td>14.2</td>
</tr>
<tr>
<td>Duration time HLAW from tree regression</td>
<td>70.74</td>
<td>14.99</td>
</tr>
<tr>
<td>% difference HLAW from tree regression</td>
<td>3.1%</td>
<td>7.9%</td>
</tr>
</tbody>
</table>

© 2015: The Royal Institution of Naval Architects
The tree regression computes an HLAW duration based on the outcome values of the 229 tuples for the HLAW duration times. The R² value is 0.93 which means that the regression tree is more accurate than the linear regression model. Likewise, another advantage is that there is no pre-processing necessary and the tree takes advantage of all of the predictors (See Figure 7). The HLAW regression tree also results in a model which can be applied by most shipyard staff members.

Figure 7: HLAW regression tree

6. CONCLUSIONS

In conclusion, determining the duration time using arc welding as compared to HLAW can be performed relatively quickly once a data mining model has been developed. The linear model is quick and similar in accuracy to the tree model. However, the regression tree model offers more interpretability and is easier to use once developed. Therefore, based upon these developed models, shipyard management is able to predict welding duration times for new interim products both for the traditional arc welded interim products as well as the HLAW interim products. The general conclusion is that when a shipyard has many large interim products such as large panels, it is definitely wise to invest in the new technology, since the welding duration time will decrease by as much as eight times in comparison to arc welding duration time.

7. ACKNOWLEDGEMENTS

The above paper was completed with the financial support of scientific research to Advancing the ship production process design methodology within the framework of support from the University of Rijeka for 2014, along with joint research activities performed at Columbia University’s Advanced Manufacturing Laboratory of the Department of Mechanical Engineering.

REFERENCES


9. **AUTHORS BIOGRAPHY**

Damir Kolich is a Senior Assistant (PhD) in the Technology and Organisation (Naval Architecture and Ocean Engineering Department) at the University of Rijeka, Croatia.

Y. Lawrence Yao is a Professor (PhD) of the Department of Mechanical Engineering at Columbia University, where he directs the Advanced Manufacturing Laboratory.

Richard Neuberg is a PhD candidate of Statistics at Columbia University where his interests are predictive modelling, machine learning, risk management and decision analysis.

Richard Lee Storch is a Professor Emeritus (PhD) and former Chair of the Industrial and Systems Engineering Department at the University of Washington, Seattle, WA, USA.

Niksa Fafandjel is a Professor and Chair (PhD) of the Technology and Organisation (Naval Architecture and Ocean Engineering Department) at the University of Rijeka, Croatia.