N95- 18980

1994

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER
THE UNIVERSITY OF ALABAMA

S3-20
30903
P. 6

DATA DRIVEN PROPULSION SYSTEM WEIGHT PREDICTION MODEL

Prepared By:	Richard J. Gerth, Ph.D.
Academic Rank:	Assistant Professor
Institution and Department:	The Ohio University Department of Industrial and Systems Engineering

NASA/MSFC:

Laboratory:	Propulsion
Division:	Motor Systems
Branch:	Performance Analysis

MSFC Colleague:	Dave Seymour John Cramer Richard Ryan Tom Byrd

INTRODUCTION

The objective of the research was to develop a method to predict the weight of paper engines, i.e., engines that are in the early stages of development. The impetus for the project was the Single Stage To Orbit (SSTO) project, where engineers need to evaluate alternative engine designs. Since the SSTO is a performance driven project the performance models for alternative designs were well understood. The next tradeoff is weight. Since it is known that engine weight varies with thrust levels, a model is required that would allow discrimination between engines that produce the same thrust. Above all, the model had to be rooted in data with assumptions that could be justified based on the data.

The general approach was to collect data on as many existing engines as possible and build a statistical model of the engines weight as a function of various component performance parameters. This was considered a reasonable level to begin the project because the data would be readily available, and it would be at the level of most paper engines, prior to detailed component design.

The modeling database consisted of 18 engines, 14 U.S. and 4 Russian. European and Japanese engines were not included because the data was not readily available. The engines ranged from 15,000 lb thrust to 1.5 million lb thrust. They included GG, expander, and staged combustion cycles. There were both booster and space engines that were fueled by kerosene, or storable propellants, or LOX/H2. They were all bi-propellant engines without annular nozzles, and made from metals, and not ceramics or composites.

The work is incomplete, and no final models were developed. However, a number of problems were encountered and approaches were attempted which will be described. A model was considered adequate and acceptable if:

a) it made sense, i.e., the variables in the model are conceptually related to the weight of the component, and the coefficients are of the correct sign;
b) the R^2 statistic is 0.85 or better;
c) the residuals are within 20% of the true weight; and
d) the model is able to predict other engines, such as paper engines, or engines not in the data base to within 20% of their observed weight.

All statistical analyses were performed in StatGraphics Version 7 by Manugistics. Best subset regression and step-wise regression were the primary modeling methods. Ridge regression and principle components regression were also explored to compensate for collinearity among the independent variables, but not further pursued because they were not appropriate given the lack of understanding of the component relationships.

TOTAL ENGINE MODEL

Since engine weight is strongly correlated with thrust, the following simple thrust model was developed. The model provides a minimum baseline for modeling accuracy, and points out some of the difficulties encountered in modeling. The regression results are presented in Table 1.

The correlation seems high and there is no time effect. However, examination of the residuals revealed two problems. First, the residuals are quite large: the average of their absolute values is 1,000 lb and the maximum is 3,987 lb for the RD170. Second, the residuals are not normally distributed: their pattern indicates a log-log transform into a power model may be more appropriate (see Table 2)

Table 1. Total Engine Thrust Model

Dependent variable: Total Engine Weight

Variable	coefficient	Std. Error	t-value	p
constant	-115.7	480..8	-0.2406	.8129
Thrust	0.0128	.00075	17.0476	0.0000

R^2=.945 Standard Error of Estimate = 1631 Durbin Watson = 1.36

Table 2. Total Engine Thrust Power Model

Dependent variable: LOG(Total Engine Weight)

Variable	coefficient	Std. Error	t-value	p
constant	-3.10	.591	-5.24	.0001
LOG(Thrust)	.897	.0484	18.62	0.0000

R^2=.953 Standard Error of Estimate = 0.255 Durbin Watson = 1.535

The power model has a better fit (higher R^2 value), and the residuals are normal, albeit still large. But, does it make sense? The LOG(thrust) coefficient is close to 1, which would be a linear model. It would seem that the residual structure and the particular form of this model is a function of the specific engines in the data base. This kind of problem, where the statistically better models did not necessarily make sense from an engineering view point, occurred frequently.

Since the total engine weight models were considered too imprecise, it was decided to model each component's weight separately as a function of component performance characteristics. The total engine weight was broken down into 8 major groups: thrust chamber ,including the injectors, main combustion chamber, and nozzle; the individual turbopumps; the gas generator or preburner; the lines, valves, and ducts; the engine mount; the igniter; other itemized weights; and unaccounted for weight. This latter category was the difference between the listed total engine weight and the sum of the other 7 categories. For most engines the unaccounted for weight was 0 or very small (<10%). The thrust chamber, turbopump, and ducts models will be presented.

THRUST CHAMBER MODEL

The thrust chamber was the first component to be modeled, and is probably the most promising, i.e., the component model most able to meet the 4 evaluation criteria. A major problem with this and the turbopump model is that many of the independent variables, such as chamber diameter, exit area, expansion ratio, cycle, L*, etc. are not really independent, but rather collinear. Geometric variables, such as throat area and L* tend to o-vary with thrust levels. Thrust was such a pronounced factor, that if thrust was used as an independent variable, nothing else was significant. Thus, the chamber weight per unit thrust (nweight) became the dependent variable.

The next difficulty was determining which variables were significant, and what the appropriate functional form was for the model. Dimensional analysis was pursued, but did not result in a satisfactory model. A engineering analysis based on wall thickness and chamber volume indicated that the chamber weight per unit thrust was proportional to L*. The expansion ratio and nozzle cooling method are important nozzle variables. Additional variables that were examined were propellant types and engine cycles. The results of a forward step-wise regression are presented in Table 3.

Table 3. Forward Stepwise Normalized Thrust Chamber Model

Dependent variable: COMBUST. nweight

Variable	coefficient	Std. Error	t-value	p
constant	9.759E-6	0.0000594	0.0164	0.9871
ER	.000053	5.09E-6	10.3155	0.0000
1/Pc	2.370	0.31368	7.557	0.0000
ablative/thrust	108.775	50.4	2.16	0.0502

R^2=0.8753 Standard Error of Estimate = 0.000745 Durbin Watson = 2.694

The residuals and other statistical elements looked fine, except the J-2 and A-7 were very large outliers, but not high leverage points. The R^2 value is acceptable, but L* proved non-significant. Also there are very few ablative nozzles (3), and thus, their significance must be approached with caution.

The backwards step-wise regression resulted in a model that also included L* and a kerosene propellant effect. However, the residuals in the model were not as well behaved, and the large coefficient values and large errors on the coefficients indicate possible collinearity problems. However, these problems could be overcome with ridge regression, for example, if it was warranted. What is really needed is a subject matter expert who can see how the variables are entering and leaving a particular model, and make value judgments as to the sign and magnitude of coefficients. This is the second type of problem that plagued the modeling process. Numerous models can be constructed, but only a component designer has the expertise to make judgments between them and guide the model building process.

TURBOPUMP MODEL

Total turbomachinery weight is strongly correlated with thrust. But, the thrust model for the turbopumps is complicated by the variety of pump configurations: single turbines driving a single pump or multiple pumps, gear driven or single shaft pumps, boost pumps, etc. Thus, a linear model on thrust, multiple, boost, and gear was attempted. Although, they were significant, the residuals were quite large, often larger than the weight of the pumps. Alternative models were investigated revealing interactions between thrust and the other variables, leading to models of turbopump weight per unit thrust. Constructing these models was difficult because of the many collinear variables, and the many models with high R^2 values (.85 to .95). The initial models attempted only to model the multiple pump weights per unit thrust. That was not a problem, until one attempted to model the single pumps. The various configurations, and in particular the boost pumps, could not rationally be divided by the thrust. Thus, given the complexity of the configurations, alternative paths were pursued.

A subcomponent model correlating the weight of impellers, housing, and volutes to their sizes was attempted. The housings would be correlated with the impeller sizes, so they did not need to be modeled separately. The volutes would be a function of the volumetric flowrates, if the volutes were external. If they were internal, they could be ignored. This left the impeller size, which is a function of the number of impellers, the diameter of the impeller, and its thickness. I had not found a way to account for the distance between the turbine and the pumps, which on some pumps was large. Using dimensional analysis for compressible flow it can be shown that the impeller diameter, D, is a function of the pump pressure, P, mass flowrate, m-dot, and pump speed, N. However, attempts to validate the relationship from know diameters were inconsistent and it was concluded the relationship was either non-linear, or the approach was not appropriate.

Principle components was attempted to eliminate the collinearity structure. However, the interpretation of the components was beyond the analyst's capability, and thus the principle components regression model is not presented. It is a statistically demanding procedure, and requires extensive component related subject expertise.

Returning to multiple regression approaches, it was recommended to model weight based on configuration parameters: boost, multiple, gear, cycle, and propellant types, as well as on headrise, total volumetric flowrate, and turbine horsepower. I believe this approach is the most sensible Please note that the turbine horsepower is, in effect, an interaction (product) between the volumetric flow rates and the headrise. Thus, it is likely to be collinear with either of the two (especially flowrate), and it may not be appropriate to model the weight with both horsepower and flowrate. The data thus far indicates that flowrate and headrise correlate with weight better than horsepower. Thus, the models reported here have flowrate and not horsepower as independent variables.

Several models were constructed leading to the conclusion that the flowrates and whether the pumps were single or multiple were the two most important variables. Further investigation showed the interaction to be more significant than the main effects. Although, it is reasonable to expect both pump types to be dependent on flow with different slopes, it is unacceptable that the model be driven by the interaction effect alone because this would exclude all single pumps. This lead to attempting to build two models, one for multiple pumps, and one for single pumps.

The multiple pump model is presented in Table 4. The residuals are not that well behaved, but are reasonably small with an error of 30% of the observed value or less for 11 of the 16 observations. Two of the five high percentage outliers are small pumps. The remaining three had errors of 43% to 71%.

Table 4. Multiple Pump Flowrate Model

Dependent variable: Turbopump Weight select (multiple = 1)

Variable	coefficient	Std. Error	t-value	p
constant	41.605	94.00	0.438	0.6681
Ttl_volflow	43.219	2.77	15.61	0.0000

$R^2 = 0.942$ Standard Error of Estimate = 310.0962 Durbin Watson = 2.381

The single pumps are very difficult to model because they include boost pumps and main pumps across different cycles. The single pumps are the 8 pumps of the J-2, SSME, and D170 boost pumps. It is particularly here that the headrise may play a significant role for the staged combustion cycles since low Pc pumps are typically flowrate driven, whereas we would expect to see high Pc engines to have a pressure component. This would be need to be evaluated in future models.

The most recent hypotheses that could not be verified or included in the model due to lack of time are that the headrise will not show a significant effect for low Pc engines, but will play a significant role in high Pc engines, i.e., staged combustion engines. Thus, Pc or an interaction between headrise and staged combustion may improve the model. Another possibility would be to normalize on volumetric flowrate similar to the way the combustion chamber was normalized on thrust: divide the turbopump weight by the volumetric flowrate.

LINES MODEL

Lines and ducts are hypothesized on volumetric flowrate (diameter), pressure (wall thickness), and basic engine size (thrust). In other words, for small engines there is a minimum weight in ducts that must exist. It is likely that their wall thickness is not pressure driven but structural so that is can withstand handling and assembly. The individual flowrates were summed to obtain a total flowrate and to eliminate collinearity between fuel and oxidizer flowrates.

Examination of the correlation structure among the variables indicates that thrust and the flowrates are correlated as is the chamber pressure and the staged combustion cycle. This makes sense since flowrates scale well with thrust, and the staged combustion cycles typically have much higher Pc's. Thus, neither thrust and flowrate nor staged combustion and Pc should be simultaneously in the same model. There are 4 high leverage engines (heavy duct weights): SSME, F-1, RD0120, and the RD170. Thus, these four engines are likely to drive the coefficient values. The power model results are presented in Table 6.

Table 6. Lines and Ducts Flowrate Power Model

Dependent variable: log(ducts)

Variable	coefficient	Std. Error	t-value	p
constant	3.53	0.1725	20.443	0.0000
log(loxvolflow + fvolflow)	0.586	0.081	7.20	0.0000
MCC Pc	0.000446	0.000132	3.389	0.0044
Russian	0.6126	0.260	2.359	0.0334

$R^2 = 0.931$ Standard Error of Estimate = 0.335 Durbin Watson = 2.083

This model is believed to be correct. The R^2 is much better and the intercept is believable. The residual structure is excellent, with no outliers. The variables that were selected were the same as was expected from the initial hypothesis. And, the coefficient on the total flow is very close to 0.5 indicating that the duct weight is proportional to the square root of the total flow, which in turn would be proportional to the diameter. The only variable of concern is Russian, since it is based on so few data points. Of all the models, this is the one in which I have the most faith.

CONCLUSION

A factor that was not considered originally was the effect of "generations" of the same engine, such as the RL10-3-3, RL10-3-3A, and the RL10A-4. In initial models, they were all included in the database to increase the number of data points. This is however false, for two reasons. First it artificially weights characteristics particular to those engines which have multiple generations in the data base versus those that do not, thereby inflating the statistical significance of those characteristics. Secondly, it increases the variation within that engine family. Thus, the earlier generations were eliminated since the most recent generation is more representative of what can be accomplished today.

In retrospect, this logic may be faulty, since this would compare 3rd and 4th generation engines with other 1st generation engines. Thus, it would probably be better to compare first generation engines only. If this were done, a time effect may become evident that would need to be considered. Should sufficient data exist, a separate study involving generations of engines and their evolution may be possible.

From the analysis to date, it appears that there is too much variation between engines to obtain an accurate model at the level that would meet the objective, i.e., within ±10%. If an accurate weight prediction model is to be created from past data, much more detailed weight and engine design information will be needed.