

Teacher's Corner

Rethinking the Paper Helicopter: Combining Statistical and Engineering Knowledge

David H. ANNIS

Box's paper helicopter has been used to teach experimental design for more than a decade. It is simple, inexpensive, and provides real data for an involved, multifactor experiment. Unfortunately it can also further an all-too-common practice that Professor Box himself has repeatedly cautioned against, namely ignoring the fundamental science while rushing to solve problems that may not be sufficiently understood. Often this slighting of the science so as to get on with the statistics is justified by referring to Box's oft-quoted maxim that "All models are wrong, however some are useful." Nevertheless, what is equally true, to paraphrase both Professor Box and George Orwell, is that "All models are wrong, but some are more wrong than others." To experiment effectively it is necessary to understand the relevant science so as to distinguish between what is usefully wrong, and what is dangerously wrong.

This article presents an improved analysis of Box's helicopter problem relying on statistical *and* engineering knowledge and shows that this leads to an enhanced paper helicopter, requiring fewer experimental trials and achieving superior performance. In fact, of the 20 experimental trials run for validation—10 each of the proposed aerodynamic design and the conventional full factorial optimum—the longest 10 flight times all belong to the aerodynamic optimum, while the shortest 10 all belong to the conventional full factorial optimum. I further discuss how ancillary engineering knowledge can be incorporated into thinking about—and teaching—experimental design.

KEY WORDS: Aerodynamics; Engineering education; Experimental design; Factorial; Nonlinear; Quadratic.

1. PROBLEM DESCRIPTION

In his deservedly famous article, Box (1992) illustrates a fractional factorial design for studying the influences of eight factors on a paper helicopter's flight time. The goal is to find the factor setting that maximizes flight time. For our purposes, this article considers a lower-dimension variant of his experiment featuring only four factors: base length (B), base height (H), wing length (L) and wing width (W). Figure 1(a) shows the helicopter layout, and Figure 1(b) illustrates how the completed helicopter falls.

Note that the base is folded to a final length of one inch, irrespective of the initial base length. The design objective is to construct a helicopter that maximizes flight time when it is dropped from a fixed height. Because the helicopter is constrained to fit on a standard 8.5-inch $\times$ 11-inch sheet of paper, the design variables are restricted to the following ranges: $3 \leq B \leq 8$; $2 \leq H \leq 4$; $2 \leq L \leq 6$; and $.5 \leq W \leq 3$, where all dimensions are in inches.

The primary purpose of the paper helicopter exercise is to teach engineers about statistics—specifically designed, multifactor experiments. As statisticians, we tend (correctly) to think that engineers must know how to collect and analyze their data. "Cookbook statistics," we warn, can be very dangerous. However, if we are to model the data effectively, statisticians must understand the underlying physical processes, as well. Bisgaard (1991) made this point while discussing his use of the paper helicopter as a teaching aide. He said, "Rather than looking for engineering problems that fit into *our* perception of engineering statistics, we should turn the question around and ask the engineers what *they* do, and see how we can help them solve *their* (engineering) problems."

Unfortunately, however, many incarnations of the paper helicopter ignore this advice and proceed in the same fashion: introducing design variables, assuming ignorance of the physical process, and determining influential factors on the response (flight time) via a full or fractional factorial experiment. Little thought is given to understanding the mechanics of the system. I propose a different tack that incorporates engineering theory into the experimental design and subsequent analysis. Box and Liu (1999) hinted at such an approach, saying "If we were really in the business of making [paper] helicopters, there would be aerodynamicists and other engineers on the team; their help . . . would undoubtedly have produced better helicopters quicker."

This article shows that an integrated approach to data collection and analysis can result in improved experimental efficiency (i.e., fewer trials) and performance (i.e., increased flight times) compared to both the conventional factorial approach and deterministic engineering. Furthermore, in addition to teaching engineers about statistical techniques, when used in this manner, the paper helicopter also serves to teach statisticians about physics and the value of understanding the problem before attempting to solve it.

2. PROCEDURE

In many real-world applications of statistics, there is some a priori knowledge of the process to be modeled. Drawing on a

David H. Annis is Assistant Professor of Operations Research, Naval Postgraduate School, Monterey, CA 93943 (E-mail: annis@nps.edu). The author would like to thank two anonymous referees for their contributions.

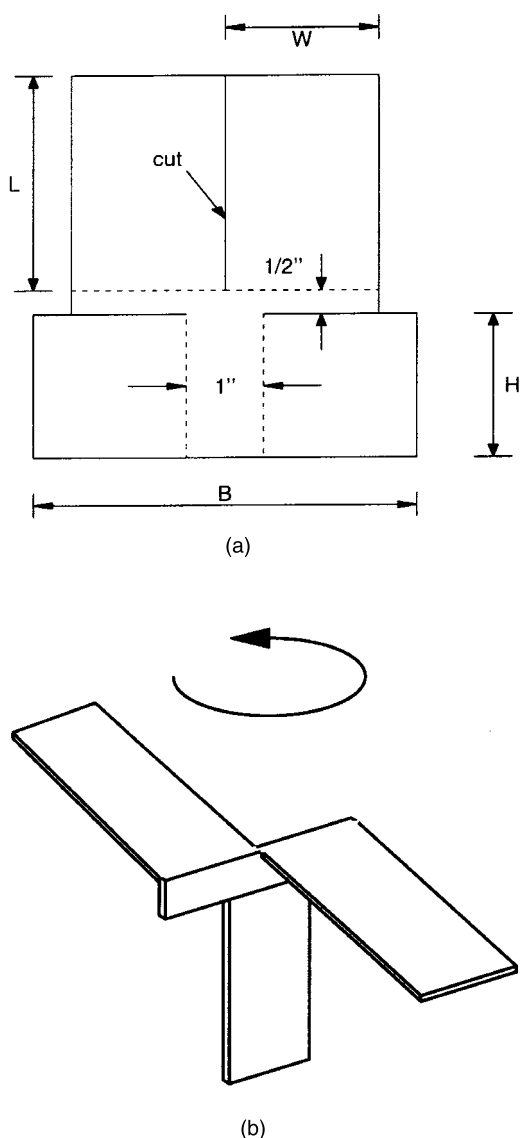


Figure 1. Layout and schematic of paper helicopter. (a) Paper helicopter dimensions are constrained so that the final design may be cut from an 8.5-inch $\times$ 11-inch sheet of paper. (b) The paper helicopter rotates as it falls.

century of aviation, the paper helicopter is no exception. Many statisticians tend to think that a priori information must manifest itself as a prior density or assumptions about the sampling distribution; seldom do they consider laws of physics as a priori knowledge. However, cursory understanding of physics and aerodynamic analysis can greatly improve the paper helicopter experiment. Three central ideas emerge: (1) you cannot effectively solve a problem you do not adequately understand; (2) real problems are seldom linear and even local approximations may be suspect; and (3) contrary to conventional wisdom, there are times when it is appropriate to include *only* higher-order terms in a model. The paper helicopter provides an excellent illustration of all of these. In this case, not only is the aerodynamic model nonlinear, it is solely a function of wing area, LW —even though neither the length nor width enters the model separately.

2.1 Aerodynamics

We begin by modeling the aerodynamics of the paper

helicopter's flight. (It is freely conceded that this analysis does not apply to conventional lift-producing helicopters.) To avoid possible confusion, we explicitly state that the subsequent use of ρ , whose units are mass per unit volume, will denote a material's density (in the engineering vernacular) rather than a correlation coefficient (as it commonly does in statistics).

As the helicopter falls, it is subjected to both a downward force (weight) and an upward force (drag) in the vertical direction. Horizontal forces do not affect the analysis explicitly, but they induce a couple which causes the helicopter's rotation. Figure 2 gives a free body diagram of the helicopter as it descends. As drawn, the figure allows for the possibility that the wings deflect upward as the helicopter falls. Although this is a familiar sight to anyone who has performed similar experiments, wing deflection may be omitted from the discussion for brevity without affecting the conclusions. Let θ denote the wing's half-dihedral angle as measured from vertical. Clearly, at equilibrium, the horizontal forces are in balance, $\cos(\theta)F_D = \cos(\theta)F_D$. Likewise, the vertical forces (which influence descent time) are in balance, $\sin(\theta)F_D = F_W$, and the paper helicopter reaches its terminal velocity. (Constant velocity implies no acceleration, and for constant-mass systems, such as the paper helicopter, force is proportional to acceleration, $F = ma$. Therefore, constant velocity implies no *net* force. That is, $\sin(\theta)F_D = F_W$.)

Paper weight is determined by the manufacturer and explicitly stated on the packaging. For these experiments 110-pound paper was used. (Paper weight is expressed in pounds per ream, which, when cut, results in 1,000 sheets each measuring 8.5 inches by 11 inches.) Assuming constant material properties (thickness, density, etc.) over the page, the weight of a helicopter is

$$F_W = W_0 \left(\frac{S}{S_0} \right), \quad (1)$$

where $W_0 = .11$ lb and $S_0 = 93.5$ in² are the weight and surface area, respectively, of a full sheet of paper and $S = BH +$

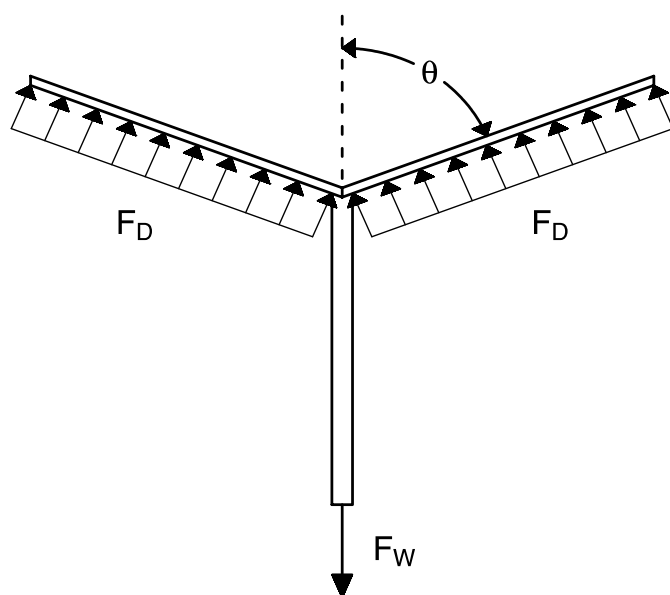


Figure 2. The free body diagram of the paper helicopter shows the forces due to weight and drag are in balance at equilibrium.

$(2L + 1)W$ is the surface area of the helicopter, which can be determined by summing rectangular areas in Figure 1(a).

Assuming that the flow of air over the helicopter is incompressible and laminar (i.e., slow-moving or having low Reynolds number), the force due to drag is given by

$$F_D = \frac{1}{2} \rho v^2 A c_D, \quad (2)$$

where ρ is the density of air, v is the velocity, A is the rotor area on which the drag force acts and c_D is the coefficient of drag (Avalone and Baumeister 1978, chap. 11). Combining (1) and (2), and solving for terminal velocity,

$$F_D = \frac{1}{2} \rho v^2 A c_D = W_0 \left(\frac{S}{S_0} \right) = F_W$$

$$v^2 = \left(\frac{2W_0}{\rho S_0 c_D} \right) \frac{S}{A} = \xi \frac{S}{A},$$

where $\xi = (2W_0)/(\rho S_0 c_D)$ is a function of physical constants and thus free of the design variables. If transient effects are

ignored, maximizing descent time is equivalent to minimizing the terminal velocity (or equivalently v^2), which is achieved when S/A is smallest.

However, while the surface area, S , is a known function of the design parameters, A , the area on which the drag force acts, is not. Because the paper helicopter resembles the winged seeds of the maple tree, one might suspect that, like the maple seeds, the helicopter spins quickly as it descends. In that case, the drag force would act on an area proportional to the circle traced by the wings as they rotate. Since the radius of this circle is the diagonal of the wing, $A = \sin(\theta) \pi r^2 = \sin(\theta) \pi (L^2 + W^2)$, and

$$v^2 \propto \frac{S}{A} = \frac{BH + (2L + 1)W}{\sin(\theta) \pi (L^2 + W^2)}.$$

Consequently, for moderate values of θ , the “best” helicopter would be one in which the base length (B), base height (H), and wing width (W) were at their minimum allowable values, while the wing length (L) was at its maximum.

Perhaps, however, the helicopter’s velocity is substantially slower than that of the maple seed. In this case, drag would act on the projection of the (stationary) rectangular rotor area and

$$v^2 \propto \frac{S}{A} = \frac{BH + (2L + 1)W}{2 \sin(\theta) LW}.$$

In this case, it remains desirable to minimize base length (B) and base height and maximize wing length (L). However, rather than minimizing wing width (W), it would be preferable to maximize it. Thus, depending on the engineering assumptions, different helicopters are called for. This underscores the necessity of collecting data and empirically validating theory, rather than relying exclusively on a “first principles” method of problem solving that is seldom realistic in practice.

In reality, the effective rotor area is somewhere in between the stationary wing area and the disk traced as the wings rotate. Figure 3(a) gives a schematic view looking down from above the falling helicopter, and Figure 3(b) shows that the effective rotor area is a function of wing area, LW , squared-wing radius, $L^2 + W^2$, and the effective angle of rotation, γ (expressed in radians). Specifically, $A = 2 \sin(\theta) [LW + \gamma(L^2 + W^2)]$, where $\sin(\theta)$ enters because the drag acts on the projection of the area perpendicular to the air flow.

One might expect γ to increase linearly with angular velocity, ω , that is, $\gamma = \omega \Delta t$. However, ω depends on the particular helicopter design. (For empirical validation, observe that smaller helicopters with minimal wing dimensions spin quickly while large helicopters with maximal wing dimensions spin slowly.) Again, we can apply basic physics to understand the effect that helicopter design has on rotation speed. Assuming homogenous paper density, Newton’s second law states that at equilibrium (i.e., constant angular velocity), the angular momentum (M) is constant and therefore, the angular velocity is inversely proportional to the body’s rotational (or polar) moment of inertia, I_p .

$$M \propto I_p \omega \Rightarrow \omega \propto \frac{1}{I_p}.$$

Approximating the entire helicopter’s polar moment of inertia by that of only its rotor, we see that $I_p = (2/3) LW (L^2 + W^2)$

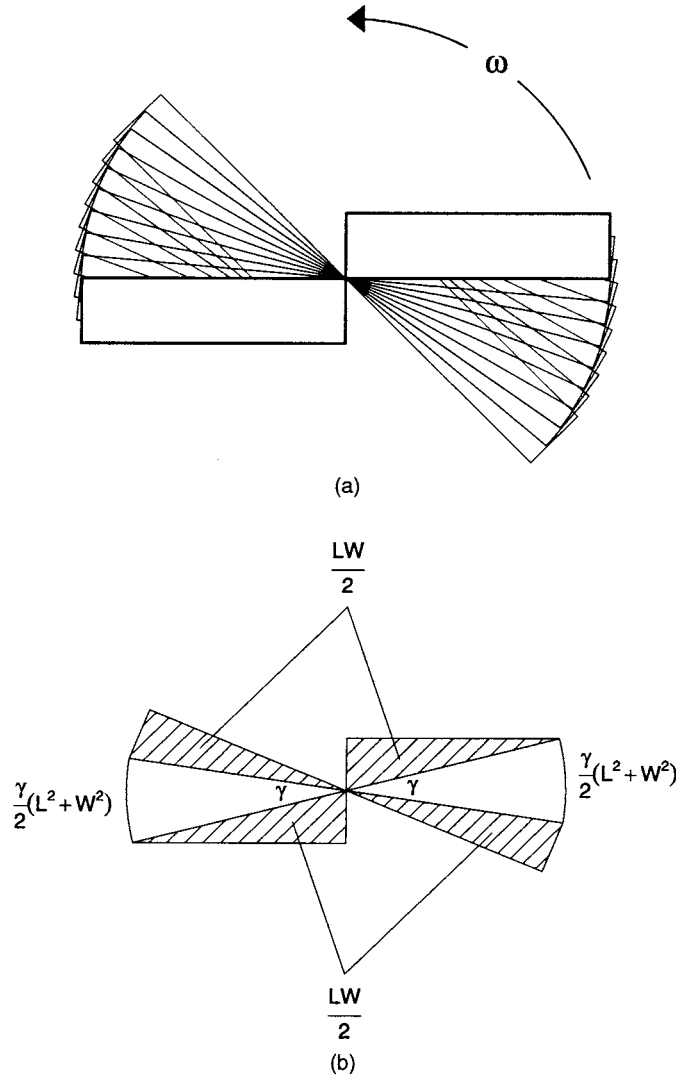


Figure 3. Effective rotor area depends on angular velocity, ω . (a) As the paper helicopter spins, its rotor traces a circle of radius $(L^2 + W^2)$. (b) The effective rotor area is proportional to the sum of the wing area, $2LW$, and a portion of the circle traced by the rotation of the wings, $\gamma(L^2 + W^2)$.

(Avallone and Baumeister 1978, chap. 3), and therefore

$$\omega \propto \frac{1}{I_p} \propto \frac{1}{LW(L^2 + W^2)}. \quad (3)$$

We are now ready to design the experiment and conduct the analysis.

2.2 Experimental Design

Although the aerodynamic analysis allows us to minimize the base dimensions (which add weight to the helicopter without increasing drag), it does not provide much insight into the wing dimensions. First, depending on assumptions regarding angular velocity, ω , the wing width should be either maximized (small ω) or minimized (large ω). Second, even though the analysis suggests that the wing length should be maximized in either case, there are three reasons for leaving it in the model at this point:

1. The wing length appears in both S , the surface area, and A , the drag force area. Therefore, increasing it increases both weight and drag. The base dimensions, by contrast, appear only in S and can be more safely discarded.

2. Certain assumptions were made in the aerodynamic modeling. One in particular which may be of concern is the rigidity of the helicopter. The analysis assumes rigid wings. In reality, however, long wings (especially those which are narrow) will likely sag, slowing the spin and reducing the effective rotor area.

3. Potential transient effects were ignored. The analysis assumed that equilibrium conditions were reached immediately. In reality, the helicopter's rectilinear (downward) and angular velocities are zero at the time of the drop. Since L influences both linear and angular velocities, it is reasonable to examine its effects.

When the base dimensions are minimized at $B^- = 3$ inches and $H^- = 2$ inches, respectively, and transient effects are ignored, the physical model describing descent time (t) for a helicopter dropped a distance (d) from the ground is

$$t = \frac{d}{v} = \frac{d}{\xi^{1/2}} \left(\frac{A}{S} \right)^{1/2}, \quad (4)$$

where rotor area, A , and surface area, S , implicitly depend on the design variables, L and W . Because the model is far from linear, an appropriate experimental design involves at least three levels for each variable to capture the higher-order dependence of the

flight time on the predictors. The simplest design is a 3^2 design in which all possible combinations of design variables are tested at their extreme values and midpoints. Although this design is, in fact, D-optimal for the full quadratic design, we propose a slightly different design, given in Table 1. Our motivations for this new design are two-fold. First, from (4), it is clear that the model is not an additive quadratic one, so the associated optimality considerations are moot. Second, the minimum wing width (.5 inches) is quite narrow and results in large aspect ratios (ranging from 4:1 to 12:1) and possibly turbulent flight. Box (1992) observed this potential for aerodynamic instability in his discussion; and Box and Liu (1999) indicated that discussions with an engineer led them to consider the effect of aspect ratio. Given the relatively low structural strength of the paper, such aspect ratios can be expected to result in poor designs which might severely distort inference. If, however, smaller wing widths are associated with increased drop times, further experiments could be conducted to investigate times resulting from the absolute minimum width. This adaptive approach to optimization has been advocated by engineers (e.g., the "spiral of progress in quality" described by Juran and Gryna 1980) and statisticians alike (e.g., the "zig-zag" graph representing the iterative learning process given by Box, Hunter, and Hunter 1978).

Therefore, we choose to test wing width at $\{1, 2, 3\}$ rather than $\{.5, 1.75, 3\}$. Experimental results obtained by dropping helicopters in random order from $d = 15.5$ feet are given in Table 1.

2.3 Model Fitting

After developing the physical model (4) and collecting the data in Table 1, a response surface was fit to determine the joint influences of wing length (L) and width (W) on flight time. Given the multiplicative nature of the model and the strictly positive descent times, the response surface was fit to log-time rather than time itself. An additional benefit of this transformation is that by taking the logarithm, models based on time, time-squared, velocity or velocity-squared (all of which suggest themselves as possible candidate responses) become equivalent.

$$\begin{aligned} E[\log(t)] &= \log(d) - \frac{1}{2} \log(v^2) \\ &= \log(d) - \frac{1}{2} \log(\xi) - \frac{1}{2} \log(S) + \frac{1}{2} \log(A). \end{aligned}$$

Although S , the surface area, changes with each design, it is a known offset (i.e., S is unrelated to the regression parameters of interest), and modeling is simplified by considering $Y = \log(t) + \log(S)/2$ as the dependent variable.

$$\begin{aligned} E[Y] &= E[\log(t)] + \frac{1}{2} \log(S) \\ &= \left[\log(d) - \frac{1}{2} \log(\xi) \right] + \frac{1}{2} \log(A) \\ &= c_0 + \frac{1}{2} \log [c_1 \omega (L^2 + W^2) + 2LW] \\ &= c_0 + \frac{1}{2} \log \left[\frac{c_2}{LW} + 2LW \right], \end{aligned}$$

where the last line follows from assuming that the angular velocity, ω , is inversely proportional to the polar moment of inertia (3). The preceding arguments suggest a non-linear model of the

Table 1. Proposed Experimental Design and Results

Design	L	W	Order	Time
1	2	1	1	1.93
2	2	2	4	2.67
3	2	3	2	3.01
4	4	1	3	2.39
5	4	2	6	3.48
6	4	3	8	3.06
7	6	1	7	2.61
8	6	2	9	3.00
9	6	3	5	2.98

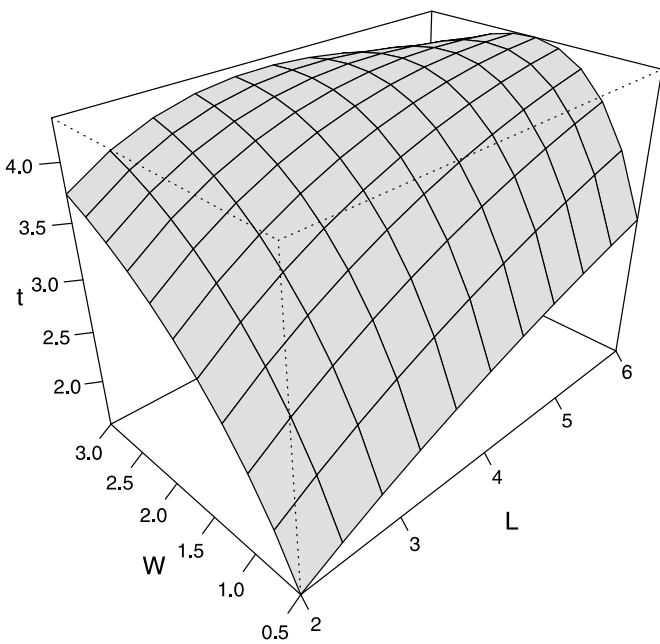


Figure 4. The response surface indicates that flight time is increasing in wing area throughout most of the design space, although it is decreasing in near the maximum.

form

$$E[Y] = \beta_0 + \beta_1 \log \left[\frac{\beta_2^2}{LW} + LW \right], \quad (5)$$

where β_2^2 is chosen so that the coefficient capturing the degree of rotation is forced to be nonnegative. Note that the factor of 2 in $(2LW)$ can be ignored. Although the β 's will change, the models are equivalent and therefore the fits are identical. The model (5) was fit using the nonlinear least squares `nls()` package in **R** (R Development Core Team 2005, www.r-project.org), resulting in the following parameter estimates (standard errors): $\hat{\beta}_0 = 6.147$ (.778), $\hat{\beta}_1 = -.790$ (.318) and $\beta_2 = 18.922$ (12.471). Figure 4 shows the resulting response surface. It is interesting to note that while increasing the wing dimensions increases flight time throughout most of the design space, there is a slight decrease in flight time when both dimensions are simultaneously at their maxima. This can be explained by the decreased rotation speed for such a design. Furthermore, such large wings might not be able to support their own weight. The parameter estimates and response surface suggest that the “optimal” helicopter is one for which the base dimensions are minimized ($B^- = 3$ inches and $H^- = 2$ inches), the wing length is maximized ($L = 6$ inches) and the wing width is $W = 1.81$ inches. Such a helicopter should take 4.34 seconds to fall 15.5 feet. It is interesting to note that the “optimal” wing width is neither $W_{\min} = .5$ inches nor $W_{\max} = 3$ inches—therefore either engineering assumption about rotational velocity (ω) would lead to an inferior design. This reinforces the need to validate theoretical arguments experimentally rather than apply them vacuously. It also provides a counter-point to those who discount “messy data” as ruining their “elegant theory,” namely, even if the theory is correct, analysis is only as good as the assumptions on which it is based. (And any nontrivial problem requires assumptions about the underlying physical behavior.) In this example, the

Table 2. Data Suggest that the “Dip” in the Response Surface at Maximal Wing Dimensions is Real

Trial	$L = 6; W = 1.81$	$L = 6; W = 3$
1	2.87	3.32
2	3.56	2.95
3	3.75	3.04
4	3.42	3.17
5	3.93	4.00
6	3.82	2.76
7	3.95	2.91
8	4.02	3.22
9	3.92	2.90
10	4.17	3.51
Mean	3.741	3.178
Std. dev	.376	.366

engineering theory and the statistical methodology complement each other and lead to a better result than either could achieve alone.

Although large in magnitude, the significance of β_2 could be questioned. The `nls()` package in **R** gives a one-sided p value of .09, and when β_2 is removed from the model, the optimal design is $(L, W) = (6, 3)$. This maximal-wing configuration was tested and its flight times were significantly inferior to the $(L, W) = (6, 1.81)$ design. Ten prototypes of each helicopter were dropped in random order. The results are given in Table 2. For these data, a one-sided Wilcoxon rank sum test returns a p value of .006. Likewise, a parametric t test gives a p value of .002. Therefore, the “dip” in the response surface near $L = 6$ and $W = 3$ is physically real, albeit statistically insignificant based on a Wald-type test of $\beta_2 = 0$. It is perhaps interesting that while the data validate the dip in the response surface, the drop times for the proposed design were all less than the predicted optimum of 4.34 seconds.

3. COMPARISON TO FULL FACTORIAL

A full factorial experimental design is used as a baseline to assess the merits of the previously outlined experimental design.

Table 3. Design and Results of 2^4 Full Factorial Design

	B	H	L	W	Run order	Time (s)
1	–	–	–	–	15	1.72
2	–	–	–	+	16	3.22
3	–	–	+	–	11	1.75
4	–	–	+	+	12	2.75
5	–	+	–	–	4	1.90
6	–	+	–	+	3	2.20
7	–	+	+	–	5	1.81
8	–	+	+	+	9	3.17
9	+	–	–	–	7	1.43
10	+	–	–	+	8	2.36
11	+	–	+	–	6	1.91
12	+	–	+	+	1	3.55
13	+	+	–	–	2	1.10
14	+	+	–	+	14	2.09
15	+	+	+	–	13	1.40
16	+	+	+	+	10	2.48

Table 4. Estimated Effects for Full Factorial Design

	Contrast	Effect	Sum sq.
B	-2.20	-.14	.30
H	-2.54	-.16	.40
L	2.80	.17	.49
W	8.80	.55	4.84
BH	-1.82	-.11	.21
BL	1.92	.12	.23
BW	.48	.03	.01
HL	.34	.02	.01
HW	-1.34	-.08	.11
LW	1.36	.08	.12

For such a design, each of the four factors is set to either its minimum (−) or maximum (+) value. All $2^4 = 16$ combinations were run. Table 3 gives the design layout as well as observed descent times for randomized trials at $d = 15.5$ feet.

Assuming only main effects and two-factor interactions are significant, the estimated factorial effects and their associated sums of squares are given in Table 4. Note that only one variable, wing width, is significant at the $\alpha = .10$ level.

Based on the factorial analysis, one would maximize wing width and ignore other design variables. In this case the “best” helicopter is not unique. In the comparison that follows, the full factorial optimal design is assumed to be one where the three insignificant design variables are set to their midpoints and the wing width is maximized.

3.1 Comparison of “Optimal” Helicopters

Our analysis results in a helicopter which differs noticeably from the one proposed by the factorial experiment. Ten nominally identical prototypes of each “optimal” helicopter configuration (aerodynamic or full factorial) were constructed and dropped determine which design is, in fact, optimal. Drop order was assigned randomly subject to the constraint that for all $n = 2, 4, \dots, 20$, the number of each type dropped after n total trials was $n/2$. (By doing so, any linear temporal influence was mitigated.)

Of 20 trials, the longest 10 flight times all were helicopters of the aerodynamic optimal prototype, while the shortest 10 all belong to the full factorial optimum. Complete results are given

Table 5. The Proposed Method Produces Substantial Improvement over the Conventional Factorial Optimum

	Aero	Factorial
1	3.63	2.93
2	3.99	2.91
3	4.26	2.91
4	3.74	2.98
5	3.92	2.72
6	4.15	2.74
7	3.92	3.15
8	3.75	2.89
9	3.96	2.99
10	4.15	3.04
11	3.95	2.93
Mean	3.947	2.926
Std. Dev	.191	.122

in Table 5; the aerodynamic and factorial optima are labeled “Aero” and “Factorial”, respectively. A nonparametric rank-sum test gives a p value of 5.41×10^{-6} for testing the one-sided alternative that the aerodynamic design is superior to the full factorial one. Similarly, a parametric t test returns a p value of 3.34×10^{-10} . Indeed for evidence this compelling, it is difficult to conceive of any reasonable method which would not detect a significant difference between flight times.

4. DISCUSSION

Statisticians often smile knowingly when they see a familiar statistical tool obviously misapplied because its underlying assumptions were violated (e.g., ordinary least squares applied to a binary response). It is clear in such cases that the culprit was a user who did not know, or chose to ignore, the procedure’s tenets. Unfortunately, engineers can tell you all their engineering assumptions but seldom any of the premises on which their statistical model is based. But ask the statistician what *engineering* assumptions support his statistical model and you are likely to receive a blank look. We, as statisticians, could enhance the quality and credibility of our analyses by taking time to understand the underlying scientific phenomena that influence the outcome of a statistical procedure. Otherwise, our omission is equally glaring to the engineer as his oversight is to us.

The paper helicopter has been used for decades to introduce engineers to experimental design. It can also be quite powerful for introducing statisticians to physics. Furthermore, it emphasizes the value of applying both problem-specific scientific judgement *and* sound statistical techniques, rather than one or the other. The following benefits are realized by incorporating engineering knowledge in the design and analysis: (1) a reduction in the design space, (2) an improved statistical model, and (3) a better design optimum. These benefits cannot be overstated. Although the cost of conducting a large paper helicopter experiment is relatively low, in many real-world applications (especially those involving destructive evaluation), specimens are at a premium, so a nearly 50% reduction in the experimental budget can be enormous.

Most importantly, head-to-head comparisons between helicopters suggested by the aerodynamic and the conventional full factorial analyses indicate that by considering the process generating the data, one can improve the quality of the analysis. Not only did *all* trials of the aerodynamic optimum outperform *all* trials of the full factorial optimum, they outperformed *every* full factorial observation as well. Thus, even a full factorial, which “saturates the design space” (as opposed to a fractional factorial which leaves some high/low combinations untested), cannot detect an extremum on the interior of the design hypercube. The model-based aerodynamic design does not share this limitation.

From the instructor’s perspective, the paper helicopter experiment, as presented, offers a jumping-off point for a number of discussions. First, it provides a concrete illustration of how subject-specific knowledge can be brought to bear on the problem at hand. Second, while the factorial method is often an excellent tool for variable screening *in the absence of physical knowledge*, when such knowledge exists, it should be incorporated as much as possible in the experimental design and analysis. As discussed, rote application of the factorial experimental design is unsuitable because it assumes that all effects are lin-

ear, or at least sufficiently approximated as such over the design space. Examining the physics of the paper helicopter reveals a model that is nonlinear and contains *only* a second-order predictor. Third, while it is a reasonable rule of thumb to include lower-order predictors in a model containing higher-order ones, it is not sacrosanct. Even in simple physical examples, responses are often related to areas or volumes, in which case quadratic or even cubic terms should be estimated even if the linear term is insignificant. Likewise, if the response is related to a predictor by a linear differential equation and a known boundary condition, the regression equation is a second-order polynomial with no linear effect. Fourth, in many cases, transformation of the response variable or the predictors is appropriate—either due to physical reasoning, mathematical convenience, or both. In this case, not only did taking the logarithm preclude negative fitted values for the time, it rendered moot the decision to model time or velocity (or powers thereof). Similarly, more sensible predictors of flight time were wing area and aspect ratio rather than wing length and width. All of these discussions emphasize that real problems seldom come prepackaged and ready for textbook analysis.

Any discussion of the proposed method would be incomplete without a few caveats. This approach is well-suited for use in a mature field of study (such as fluid dynamics). Although many applied problems present the analyst with similar theoretical

knowledge, all do not. In those cases where there is little understanding of the problem, the applicability of this approach is limited. Furthermore, engineering knowledge cannot and should not *replace* statistical tools. Rather, it should supplement them. Neither the deterministic engineering approach nor full-factorial statistical methods result in satisfactory answers. Only by incorporating both is an optimum achieved.

[Received February 2005. Revised June 2005.]

REFERENCES

- Avallone, E. A., and Baumeister, III, T. (eds.) (1978), *Marks' Standard Handbook for Mechanical Engineers*, Boston: McGraw-Hill.
- Bisgaard, S. (1991), "Teaching Statistics to Engineers," *The American Statistician*, 45, 274–283.
- Box, G. E. P. (1992), "Teaching Engineers Experimental Design with a Paper Helicopter," *Quality Engineering*, 4, 453–459.
- Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters*, New York: Wiley.
- Box, G. E. P., and Liu, P. Y. T. (1999), "Statistics as a Catalyst to Learning by Scientific Method Part I—An Example," *Journal of Quality Technology*, 31, 1–15.
- Juran, J. M., and Gryna, Jr., F. M. (1980), *Quality Planning and Analysis: From Product Development Through Use*, New York: McGraw-Hill.
- R Development Core Team (2005), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.