The GRL* Dot Gain Model

J A Stephen Viggiano**

ABSTRACT

The Plate/Press segment of any halftone printing process transforms halftone dots on film into halftone dots on paper, so that different shades of gray may be reproduced. During this process, the fractional dot area of these dots change. This phenenomenon is called dot gain. Its study is an important quality control function, and is essential for proper tone reproduction.

Because of their importance, dot gain curves have been modelled (put in equation form). This enables a parsimonious characterization of the entire dot gain curve by perhaps one or two parameters.

The GRL Dot Gain Model provides an excellent fit to actual data using only one or two parameters, and enables the calculation of the critical printing areas.

This paper discusses the basic model, its inverse relation, solution for the critical printing areas, determination of the parameters, and cascading two curves. The mathematical development and experimental verification of the model is presented.

*Graphic Research Labs
**Graduate Student, Rochester Institute of Technology

SYMBOLS AND EXPRESSIONS

af	-fractional dot area on film (0, 1)
ad	-fractional dot area on plate
ap	-fractional dot area on paper
Δ	-characteristic gain (in a single-transfer situation, this is maximum gain)
awf	-Lower critical dot area --- a small dot that just begins to print (area of white on film)
avf	-Upper critical dot area --- a large dot that just completely fills in (area of solid on film)
sqrt(*)	-Square root operator --- returns positive square root of (*)

INTRODUCTION

The study of dot gain is an integral part of a lithographic or relief quality control program. In addition, this information is necessary to obtain valid results from the several dot area - optical density models (viz, the Yule-Nielsen equation, the Murray-Davies equation, and the Neugebauer equations), because each assumes that the fractional dot area on the print is known. In other words, these equations require that the dot gain be added to the dot area on film before they are applied.

The level of gain is not constant throughout the scale, but starts out small, increases to a maximum (usually in the vicinity of the 0,50 dot area on film), and then tapers off.

Several models have been proposed for dot gain. One of these, the FOGRA model (developed by Karl Haller)[1], has found favor among researchers. Dot gain curves with characteristic gain values from -0,20 to 0,40 generated by this model are shown in Figure 1.

In most printing applications, there exist two critical printing areas --- the first, awf, is the smallest dot that overcomes dot sharpening and begins to print; the second, avf, is the dot area on film that produces a solid on the paper (i.e., where complete filling in begins). Typical values for these are 0,03 and 0,90, respectively.

These two dot areas are of particular interest because they determine (to a large degree) the requirements for halftone negatives. Knowing the values of the critical printing areas awf and avf, the processworker will know that the highlight dots may be no smaller than awf, and the shadow dots, in principle, may be no larger than avf. (In practice, the shadow dot is usually made larger than avf, because density will be continued to be added after complete filling in on the paper occurs.)

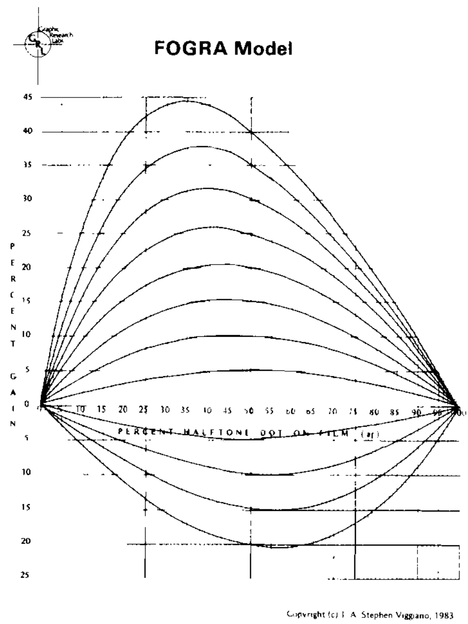

Figure 1.
FOGRA Dot Gain Model.

The lower left corner of the grid in Figure 1 is chopped off --- this boundary intersects the curve when af = awf. In other words, the lower left portion is bounded by an awf locator line. Similarly, the upper right corner is bounded by an avf locator line, which intersects the dot gain curve when af = avf. Note that in Figure 1 all curves have trivial awf's of 0 and trivial avf's of 1.

Because the FOGRA model can only produce curves with awf's of zero and avf's of unity (100%), it cannot be used to assist in determination of these two fundamental parameters, and should not be used if such estimation is desired. Because these two critical dot areas are of such vital importance, a model should be used which permits the estimation of these areas.

A new dot gain model was developed at GRL which produced non-trivial awf's and avf's. This model is really a cascade of two or more very simple functions. Each function in this cascade represents a single transfer, or a group of closely related transfers of the same type. See Figure 2 for some curves generated by this basic function.

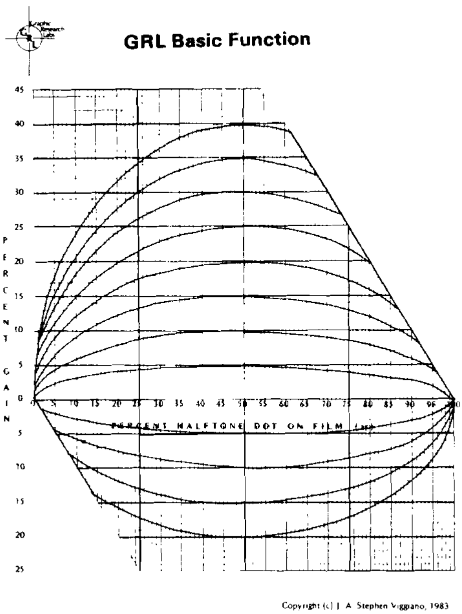

FIGURE 2.
The GRL Dot Gain Model.

THE GRL DOT GAIN MODEL

The Basic Function-
The GRL Model for a single transfer is:

$$gain = 2\Delta sqrt[af(1 - af)] \quad (1)$$

so that the dot area on paper as a function of the dot area on film may be expressed as:

$$ap = af + 2\Delta sqrt[af(1 - af)] \quad (2)$$

when a single transfer occurs. (sqrt(*) denotes the square root.) The mathematical development of this model will be discussed in a seperate section of this paper.

These equations are of course subject to the restriction that 0 <= ap => 1, because it makes little sense to consider dot areas less than zero or greater than unity. If the computed value of ap is less than zero, zero should be used instead. Similarly, if the computed value of ap is greater than unity, the value of ap should be set to one.

Note that the model contains a single parameter ---Δ--- which is the characteristic gain at af = 0,50. With one degree of freedom, this model allows either a non-trivial awf (if Δ is negative), or a non-trivial avf (if Δ is positive).

Before considering the model with both awf and avf non-trivial, some fundamental formulas will be presented.

The Inverse Relation-
Equation (2) maps a continuum of dot areas on film onto a similar continuum of dot areas on paper. In many instances, it is desirable to have the inverse relation --- that is, the dot area on film expressed as a function of the dot area on paper. This is:

$$\frac{2\Delta^2 + ap - 2\Delta \, sqrt[\Delta^2 + ap(1 - ap)]}{1 + 4\Delta^2} \qquad (3)$$

Solution for the Critical Printing Areas-
The critical printing areas awf and avf may be solved for by considering the special cases of ap = 0 and ap = 1, respectively.

$$awf = \begin{cases} 0, & \text{if } \Delta > 0 \\ 4\Delta^2/(1 + 4\Delta^2), & \text{if } \Delta < 0 \end{cases} \qquad (4)$$

$$avf = \begin{cases} 1/(1 + 4\Delta^2), & \text{if } \Delta > 0 \\ 1, & \text{if } \Delta < 0 \end{cases} \qquad (5)$$

Solutions for the Characteristic Gain-
One important advantage of dot gain models is their parsimonious description of a plate/press transfer function. Rather than having to specify every point on the

curve, these models make it possible to reconstruct the entire curve from one or two parameters.

The characteristic gain may be computed from a single (af, ap) pair. This enables the reconstruction of the entire plate/press characteristic curve from a single point on the curve. Of course, this assumes that Model (2) applies reasonably well to the transfer being considered.

By re-arrangement of Equation (2),

$$\Delta = \frac{ap - af}{2\,\mathrm{sqrt}[af(1 - af)]} \qquad (6)$$

However, a number of (af, ap) pairs are available when more than one gray patch is printed on each press sheet, in addition to the solid patch. This is usually the case. It is desirable to use all the patches on a press sheet, rather than a single patch, to minimize measurement error, bias from systematic error in the model, etc. The estimation of the characteristic gain, Δ, by least-squares is simple because Model (2) is linear in Δ and contains a single term:

$$\hat{\Delta} = \frac{\Sigma(ap - af)}{2\,\Sigma\,\mathrm{sqrt}[af(1 - af)]} \qquad (7)$$

Naturally, the solid patch (where both af and ap = 1) should not be included in this calculation. Similarly, no (af, ap) pair for which ap equals zero or unity should be excluded from the calculation of $\hat{\Delta}$. Otherwise, the estimate will be biased, producing estimates of Δ smaller than they should be.

Equation (7) assumes that the error is proportional to the square-root of the dot gain, as indicated by the experience of the author. (This is a comprimise between additive and multiplicative error. In addition, this automatically assigns extra weight to the dot areas at both ends of the scale. This is desirable, because it causes the points on the curve in the highlight and shadow regions to have the greatest weight in the determination of awf and avf.)

The variance of this estimate is

$$\hat{V}(\hat{\Delta}) = \frac{n\Sigma(ap - ap_{est})^2}{4(n-1)[\Sigma\sqrt{af(1-af)}]^2} \qquad (8)$$

which is useful for putting confidence limits, etc., on the estimate of Δ.

A CASCADE OF TWO TRANSFERS:

Equation (2) is a model for a single transfer, in which either dot gain or dot sharpening (negative gain) may occur. In practice, both of these phenomenon occur to some relative degree. Additionally, it "locks" the maximum dot gain to the 0,50 (50%) dot on film (see Figure 2). Naturally, the maximum gain may occur at other values. Because of this, it is recommended that, in general, two or more models of this type be cascaded, as follows:

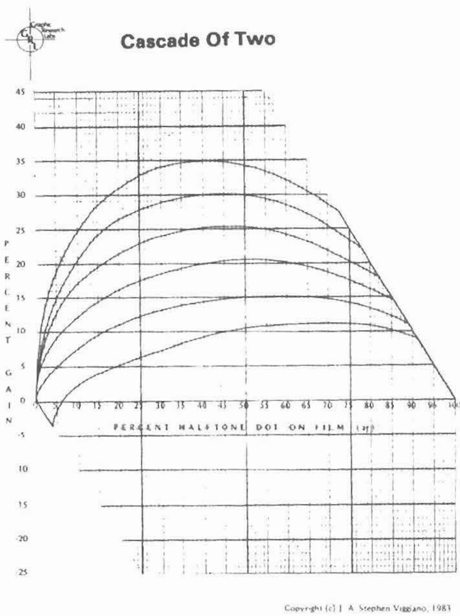

FIGURE 3.
Cascade of Two Transfers.
(Δd constant)

Consider the dot area on the plate, ad. Allow it to assume the role of ap in Equation (2). Then, we obtain

$$ad = af + 2\Delta d\ \mathrm{sqrt}[af(1 - af)] \qquad (9)$$

This gives the dot area on the plate as a function of the dot area on film. Naturally, the dot area on the plate is restricted to the interval from zero to 1. If an ad is calculated that is less than zero, the value of ad should

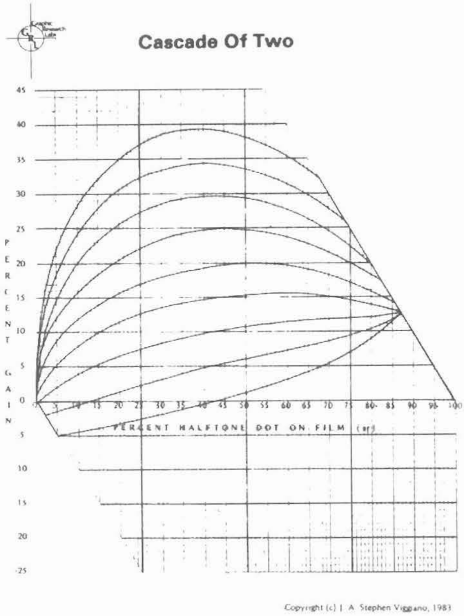

FIGURE 4.
Cascade of Two Transfers.
(Δp constant)

naturally be set to zero. Likewise, if an ad is calculated that is greater than one (100%), the value of ad should be set to one before proceeding. Otherwise, in the next equation, the square root of a negative number will be attempted, which can cause a computer run to seize.

Now, let ad play the role of af in Equation (2):

$$ap = ad + 2\Delta p \cdot sqrt[ad(1 - ad)] \qquad (10)$$

Again, the dot area on paper should be adjusted if it does not fall into the range from 0 to 1.

It should be noted that the characteristic gain on the plate (Δp) is ap - ad when <u>ad</u>, not af, equals 0,50 (50%). This distinction becomes particularly important when Δd is relatively large in magnitude.

These two equations, when used together, can be used to model two consecutive transfers, one in which dot sharpening occurs, and another in which dot gain occurs. Note that this model contains two parameters viz, Δd (dot gain on the plate), and Δp (dot gain on paper). When two transfers are "cascaded" this way, the curves in Figures 3 and 4 are generated.

The dot area on the plate is not usually measured, nor is the author suggesting that this task should be a standard quality control function. What the author is suggesting is that the dot sharpening be considered as occuring solely during the transfer between film and plate, and the gain be considered as occuring strictly on the press. While this is not entirely true, it does make the visualization of this cascade easier. That the dot area on the plate is not actually measured is not important.

As with one transfer, some fundamental relationships may be derived. Here, the first transfer will be considered to involve either gain or sharpening (Δd positive or negative), while the second transfer will be assumed to involve strict gain (Δp is positive).

The Inverse Relation- This is best accomplished by solving for ad in terms of ap, then af in terms of ad. This is accomplished via Equation (3).

Again, it should be kept in mind that dot areas less than zero or greater than unity may be computed. As before, if ad is not on this range, it should be set to the appropriate value (either 0 or 1) before computing the value of af.

Solution for the Critical Printing Areas- Here, neither awp nor avp is trivial so long as one transfer involves sharpening and the other gain. Again, this is best accomplished by solving for the critical printing areas on the plate (viz, awd and avd) as per Equations (4) and (5), substituting awd for awf, and avd for avf. The critical dot areas on the film (viz, awf and avf) may then be calculated as per Equation (3).

Solution for the Characteristic Gain Parameters- Unfortunitely, it is not possible to solve for both Δd and Δp with a single (af, ap) pair. While it may be possible to accomplish this knowing also ad, the dot area on the plate, the measurement of dot areas on the plate is not recommended. In addition, it was pointed out that the dot area on the plate is used mostly in a hueristic sense, as an expedient for visualization of the cascade of transfers.

At least two (af, ap) pairs, then, are necessary if both Δp and Δd are to be determined. A closed-form solution is possible only if one of the parameters (presumably Δd) is held fixed at some level. It therefore is recommended that either:

1) Both Δd and Δp be estimated through least squares; or

2) Δd is roughly estimated (from prior knowledge) and held fixed, while Δp is estimated as with a single transfer, calculating each ad as a function of each af, and allowing ad to play the role of af in Equation (6) or (7), as appropriate. This enables a closed-form solution, and may be estimated from a single (af, ap) pair.

If the first course is chosen, it is necessary to have an initial estimate of the parameter Δd, because it implicitly enters Equation (10) under a square root, and as a multiplier of itself. Consequently, non-linear least squares methods are required.

It was originally belived that a good estimate for the strictly nonlinear parameter, Δd, is

$$\hat{\Delta} d = 1/2 - \widetilde{af} \qquad (11)$$

where $\widetilde{af}$ is the dot area on film that produces maximum gain. This value could be interpolated from the step producing the maximum gain, and the two steps that bracket this step. However, it has been found that this does not always produce a satisfactory estimate.

Because of the conditional linearity of the parameter Δp, no initial estimate is necessary.

MATHEMATICAL DEVELOPMENT OF MODEL

There are essentially two contrasting views on the mechanism of dot gain --- the first, called the perimeter theory, holds that the gain is proportional to the perimeter of a dot, see Figure 5.

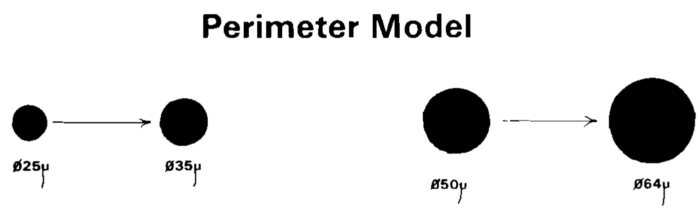

FIGURE 5.
Dot Growth under the Perimeter Model.

The second theory, called the IsoKonturen model, holds that all dots, regardless of their size, undergo a constant increase in diameter. See Figure 6.

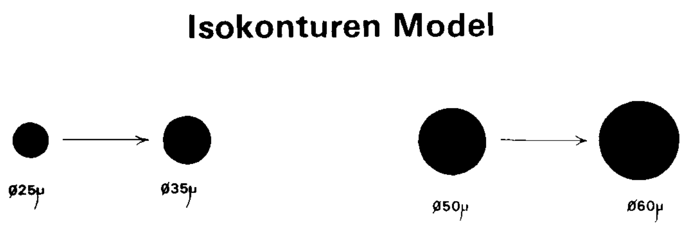

FIGURE 6.
Dot Growth under the IsoKonturen Model.

The GRL Dot Gain Model was empirically derived to fit the general shape of typical dot gain curves. Nevertheless, it has theoretical basis in that it represents a reconciliation of these two opposing theories. (There is really no doubt that both of these models contain some element of the true mechanism.)

The dot gain will be explicitly solved for under both of these models, for dots of different shapes (viz, round, elliptical, square, and rhombic). The GRL Dot Gain Model, given in Equation (1), will be shown to be a compromise solution between these two theories.

The Perimeter Model-
When the dots are small, the perimeter will be the distance around the dot itself. Conversely, when the dots are large (near 100%), the perimeter of the dot will be the distance around the clear hole. For expediency, dots on film with less than 0,50 fractional dot area will be referred to as "highlight dots," while those dots on film with greater than 0,50 fractional dot area will be referred to as "shadow dots."

In addition, it will be assumed here tentatively that the corners of the dots do not overlap. While this assumption is violated in practice for dots in the middletone range (approx. 50% dots), we shall, for the moment, confine the region of interest to the areas where the dot corners do not join.

Solution for Gain for Round and Elliptical Dots- While one would expect to find so-called "elliptical" halftone dots that were not true geometric ellipses, this assumption, however naive, is useful here because it defines the general, if not exact, dot boundaries.

Also, circles are really special ellipses (i.e., with both axes equal in length). Circular halftone dots will be treated as a special case of elliptical dot.

The aspect ratio of an ellipse (the ratio of the "length" to the "width") of an ellipse is h. (In the case of a circle, h = 1). For highlight dots, the gain under the perimeter model will be:

$$\text{gain} = g \cdot \text{sqrt}[af\ (1 + h^2)/h] \qquad (12)$$

where g is a constant of proportionality. For shadow dots, the gain will be:

$$\text{gain} = g \cdot \text{sqrt}[(1 - af)\ (1 + h^2)/h] \qquad (13)$$

Gain for Rhombic and Square Dots- The measure of the acute angle of a rhombus formed by two sides, q, is 90 degrees in the case of a square; less than 90 degrees in the case of a rhombus. For highlight dots, the gain under the perimeter model will be:

$$\text{Gain} = g \cdot \text{sqrt}[af/\sin(q)] \qquad (14)$$

and for shadow dots:

$$gain = g \cdot sqrt[(1 - af)/sin(q)] \qquad (15)$$

The IsoKonturen Model-

Under this model, all dots, regardless of size, undergo a constant increase in dimensions. In offset processes (e.g., letterset, offset lithography) this manifests itself in one direction more than the other in the form of slur. However, here the radial dot growth will be assumed to be constant in all directions. This assumption is of course violated; nevertheless, useful approximations can be derived under it.

Gain for Round and Elliptical Dots-

For highlight dots,

$$gain = g(1 + h) \cdot sqrt[af/h] + c \qquad (16)$$

with c a constant; and for shadow dots,

$$gain = g(1 + h) \cdot sqrt[(1 - af)/h] + c \qquad (17)$$

Gain for Rhombic and Square Dots-

For highlight dots,

$$gain = g \cdot sqrt[af\ sin(q)] + c \qquad (18)$$

and for shadow dots,

$$gain = g \cdot sqrt[(1 - af)\ sin(q)] + c \qquad (19)$$

Equation (1) as an Approximation to Equations (12) - (19)-

For highlight dots, the dot gain for round, elliptical, rhombic, and square halftone dots, under both models, was of the form:

$$gain = g \cdot sqrt(af) + c \qquad (20)$$

while, for shadow dots, we had:

$$gain = g \cdot sqrt(1 - af) + c \qquad (21)$$

where, in some cases, $c = 0$. Keeping in mind that when af is close to zero, $(1 - af)$ is close to one, we have as an approximation for highlight dots:

$$\text{gain} = g \cdot \text{sqrt}[af\ (1 - af)] + c \qquad (22)$$

(This approximation seems to account for dot join.) As an approximation for the shadow dots:

$$\text{gain} = g \cdot \text{sqrt}[af\ (1 - af)] + c \qquad (23)$$

Note that (22) and (23) are identical. Thus, we have a single approximation for both highlight and shadow dots, of several shapes.

The "intercept" term in Equation (23), c, has been found, by experimentation, to be not significantly different from zero. It would appear that Equation (23) is not only an approximation to Equations (12) - (19), but it actually represents a reconciliation of the models from which these equations are derived.

If we stipulate that the "intercept" term, c, in Equation (23) is, in fact, zero, we have Equation (1) (the GRL Dot Gain Model) with $g = 2\Delta$. In fact, it is desirable to use Δ as a parameter, rather than g, because it has particular significance (i.e., it represents the gain at af = 0,50).

EXPERIMENTAL VERIFICATION OF THE MODEL:

On-press testing was used to verify the GRL Dot Gain model involving two transfers, given in Equations (9) and (10).

Test Target- A test target was prepared for these tests. A hard-dot halftone image was produced by contacting a step-wise exposed halftone film onto contact lith film. This hard-dot target contained 9 steps, plus clear and solid steps.

The halftone dot areas of these steps were measured with a Tobias PCT Dot Area Meter. These dot areas are recorded in Table 1.

Also included was a solid patch, a gray bar, a GATF Star Target, and the first 10 steps of a 21-step transmission guide.

On-Press Testing- The test form was printed on a Heidelberg KORA lithographic offset press. Sixteen sheets were selected at random intervals from the run. The inking level was purposely varied throughout the run to produce different dot gain curves for each sheet.

Step	Dot Area	
1	0,89	(shadow)
2	0,81	
3	0,72	
4	0,64	
5	0,51	
6	0,34	
7	0,23	
8	0,13	
9	0,05	(highlight)

TABLE 1.
Dot Areas of the Test Target.

The densities above paper of each patch from each sheet sampled were recorded, as well as the Solid Ink Density (SID). These densities were converted into dot areas on the paper via the Yule-Nielsen equation, with an n-value of 1,7, which is recommended for general conditions.[2]

The dot gain for each step was calculated for each step by subtracting the target dot areas on film from the dot areas on paper. The parameter for gain on the plate, Δd, was estimated for each sheet sampled by determining the dot area on film that produced the greatest amount of gain (by quadratic interpolation), and applying Equation (11).

The weighted-least-squares solution for the gain on the paper, $\hat{\Delta} p$, was estimated for each sheet sampled as per Equation (7). These estimates for each sheet sampled are given in Table 2. The ANOVA Summary is given in Table 3. Note that the overall value of r-squared, over 96%, indicates an excellent fit of the model to the data.

A problem arose with the estimation of the critical printing area avf. The model predicted values lower than those predicted by the Yule-Nielsen Equation. Photomicrographs reveal that the tints filled in before the Solid Ink Density was reached. This effect has been attributed to the spread of the ink film on the paper, producing a greater dot area, but a thinner ink film in tinted areas. In other words, the density of the dots on paper which made up the tints was lower than the density of the solid.

Sheet #	SID	$\hat{\Delta}d$	$\hat{\Delta}p$
143	1,04	0,17	0,09
155	1,12	-0,01	0,26
166	1,14	-0,01	0,27
178	1,16	-0,01	0,27
190	1,16	0,17	0,13
202	1,18	0,17	0,15
213	1,14	0,16	0,18
220	1,14	0,16	0,21
226	1,14	0,17	0,16
234	1,11	0,17	0,13
241	1,09	-0,01	0,30
246	1,11	0,17	0,14
252	1,09	0,17	0,15
258	1,14	0,17	0,12
264	1,14	0,17	0,11
269	1,03	0,16	0,19

TABLE 2.
Solid Ink Densities and Parameter Estimates

Source	S.S.	DF	M.S.
Total	3,7935	127	
Δp, 143	0,0560	1	
Δp, 155	0,4580	1	
Δp, 178	0,4698	1	
Δp, 190	0,1122	1	
Δp, 202	0,1516	1	
Δp, 213	0,2183	1	
Δp, 220	0,2563	1	
Δp, 234	0,1065	1	
Δp, 241	0,6061	1	
Δp, 246	0,1197	1	
Δp, 252	0,0854	1	
Δp, 258	0,0934	1	
Δp, 264	0,0761	1	
Δp, 269	0,2271	1	
Regression	3,6640	16	0,2290
Residual	0,1295	111	0,00117

R-Square = 0,965

TABLE 3.
ANOVA Summary.

FOR FURTHER INVESTIGATION

1. A better estimate for the non-linear parameter, Δd, is needed. The approximation given in this paper seems to work fairly well in most cases. However, it seems that a much more precise estimate is possible.

2. Through non-linear regression, it is possible to obtain least-squares estimates of the Yule-Nielsen n-value, as well as the two gain parameters. An extensive model is being developed for the Plate/Press Characteristic, with solid ink density and screen ruling, in addition to the dot area on film, as inputs to the equation.

ACKNOWLEDGMENTS

The author would like to thank Mr. John A. Pagluica for his ongoing assistance on this project, Mr. Irving Pobboravsky for his proofreading and criticism, and Mr. Philip Tobias of Tobias Associates, for loan of the PCT Dot Area Meter.

This work was started while under fellowships from the Navigators of New York City and TAGA/3M. The author would like to extend his thanks to these groups for their support.

LITERATURE CITED

1. Haller, Karl, "Mathematical Models for Screen Dot Shapes and for Transfer Characteristic Curves," in Advances in Printing Science and Technology - Proceedings of the 15th Conference of Printing Research Institutes, W H Banks, Ed. London: Pentech Press, 1979, p. 85 - 103.

2. Pearson, Milton, "n Value for General Conditions," 1980 TAGA Proceedings, p. 415 - 425.