Accuracy and Precision in Color Characterization

Roger A. Siljander and Richard S. Fisch*

Keywords: Characterization, Method, Statistics, Variables, Variability

Abstract

Several press tests have been performed by industry organizations and private companies to characterize the colorimetric output of various printing processes. Since most attempts have significant difficulty achieving targets and results are difficult to reproduce, characterization data from these tests are usually surrounded by controversy and met with hesitant acceptance. Analysis of the need for characterization data shows that:

- Characterization data sets are critical for the implementation of digital imaging systems;
- The foundation of good characterization data is a press test in statistical control relative to a solidly fingerprinted press output condition; and,
- The quality of the characterization data itself is a function of the sampling plan which addresses test target configuration, test form configuration, press sheet selection and number of press tests.

This paper defines the relationship between fingerprinting, process control and characterization, presents a survey of press tests that demonstrates the effect of sampling on press test repeatability and reproducibility and proposes methodology that could be used to resolve many difficulties commonly associated with many characterization data sets.

* Imation Corporation

Introduction

In a time of change from analog to digital workflow, there are many process issues that need a very fine level of definition such that analog workflow can be transformed into digital workflow. The idea is to teach a computer to perform tasks normally performed by people. One of these process issues is describing to computers what the output of a press looks like. The data used to describe the color output of a press is conventionally called characterization data. Of course, the goodness of characterization data determines how well computers can perform the task. This paper addresses several variables that effect the goodness of press characterization data and presents methods to ensure quality in characterization data. A summary of industry activity relative to press characterization sets the stage.

To date, industry organizations and private companies have performed several press tests. The first organized industry effort took place in 1993 and was an attempt to characterize the publication printing process as defined by SWOP, Inc. This effort resulted in the standard ANSI CGATS.6 Type 1 Printing and the associated technical report TR 001. This pair of documents, CGATS.6 and TR 001, currently defines the process inputs and expected colorimetric output for SWOP or publication printing. While this first attempt has been criticized for several technical problems, the fact of the matter is that this work was on the leading edge of technology at its time and much can be learned by analyzing the difficulties that have been encountered in the application of CGATS.6 and TR 001.

There have been attempts by several other industry groups to characterize their particular printing processes. In 1996 work was initiated to characterize the newsprint printing process as defined by SNAP. And, in 1999 both GRACol and GAA initiated programs to characterize commercial and gravure printing processes. All of these attempts have met limited levels of success and all are riddled with the same technical issues as seen in the original work to characterize publication printing.

The underlying technical problems manifest themselves in an inability to consistently achieve process aims and inconsistent results between visual and numerical analysis. In an attempt to understand these problems, in 1999 the authors contracted a press test intended to replicate the work that resulted in CGATS.6 and TR 001. At the 2000 Annual TAGA Conference in Colorado Springs, the authors presented a comprehensive report that showed how the process aims and color characteristics of the press test were distributed throughout the entire test and showed what the effect of these distributions was on the characterization data.

While a great deal was learned about many of the difficulties associated with characterization data, this test also exhibited problems similar to other press tests. Since the authors were involved with several other characterization efforts, they reviewed data from various press tests and identified several common factors that would produce the difficulties common to all attempts. Assuming the presses were properly setup and maintained during the press test, most difficulties associated with characterization data sets could be resolved by modifying the press sheet sampling process and combining results from several press tests.

In the following sections, this paper

- defines the relationship between process control, fingerprinting and characterization,
- presents a summary of statistics necessary to ensure adequate information has been collected from a press test,
- summarizes data from several industry press tests that illustrates the confusion typically encountered, and,
- analyzes methodology relative to characterization data accuracy and precision.

Process Control, Fingerprinting and Characterization

In an analog process, jobs are typically processed for known output devices. Usually the person in prepress not only knows on which press the job will run, but also the press operator. Through internal control negotiated between prepress and the press room, jobs flow relatively smoothly. But with the digital revolution came a host of new problems. In the digital arena, workflow is rarely an internal proposition. Design and prepress are required to accommodate multiple printing locations, printing sometimes with different printing technology. To top it off, they are asked to accommodate the world wide web. In short, digital technology has wreaked havoc on traditional workflow. But steadily, digital technology becomes more widespread while digital technology and analog workflow are adjusted to accommodate each other. Through this adjustment process, the concept of color characterization has surfaced as an integral part of workflow in the graphic arts industry.

The goal of color characterization is to describe numerically the colorimetric properties of a printing process. While any press or any imaging system can be characterized at any condition, it is best to know what the condition is and how well it can be maintained in that condition – enter fingerprinting and process control. Furthermore, in the interest of maximizing press output, not only should the press condition be known, but also optimized. These terms are defined in ANSI literature as follows:

- Process Control is the use of process analysis for the purpose of keeping a press at a given operating condition. That is, process control keeps the press within known control limits.
- Fingerprinting is the use of process analysis to benchmark an operating condition, in other words, to establish the process aims and control limits. Fingerprinting results in a set of metrics and values that defines a process. The process control function uses these results to judge whether or not a press is compliant with the defined process.
- Optimization is the use of process analysis to determine the best operating condition by balancing process output and stability with customer expectations and economic factors.
- Characterization Data is the set of data determined from the measurement of samples that describes the relationship between input tone values and resulting output color.

Process analysis is at the foundation of all these concepts. Process analysis defines the strategies and methods used to quantify process parameters. For example, process control on a printing press involves process analysis in the form of the measurement of color bars and other specialized targets. This analysis provides information about the amount of ink being transferred to the paper and the quality of the transfer, e. g., that the solid ink densities are proper, there is no slippage, roller pressures are properly set, etc. This information is then used to decide whether or not a process change is required.

In the decision making process, the process control data is compared to fingerprinting data which is itself a result of process analysis. In the fingerprinting case, the same test strategies and methods are used to quantify the process parameters. While the process analysis is identical for both process control and fingerprinting, it is the use of the resultant data that differentiates process control and fingerprinting. Process control is the application of process analysis to guide a decision making process and fingerprinting is the application of process analysis to define a process.

While process control is the evaluation of a snapshot of a given process, the culmination of fingerprinting is a numerical description of the process center points and ranges taking into account within and between press run variations. As such, fingerprinting represents a large number of samples from several press runs, some of which may be process control data from live production run at the fingerprinted condition. The result of fingerprinting is a process definition. Examples of industry process definitions are SNAP, GRACoL, SWOP and FIRST. The SWOP and SNAP processes have also been incorporated into ANSI standards CGATS.6 and CGATS.16 although CGATS.16 has yet to be published.

While fingerprinting defines the aims and tolerances for process parameters like solid ink density and dot gain, color characterization defines the colorimetric output of a printing process. Because fingerprinting defines the process, characterization should only be attempted after a process is fingerprinted and the characterization press tests must be subject to process control to ensure that the inputs for the test are compliant with the fingerprint. There should always exist fingerprint / characterization pairs. For example, CGATS.6/TR 001 and CGATS.16/TR 002 are fingerprint / characterization pairs.

The relationship between fingerprinting, process control and characterization is somewhat intertwined and starting new processes can be challenging, but the important concepts to remember are that:

- fingerprinting defines the process control parameters,
- process control validates the press test, and
- characterization describes the expected colorimetric output.

Quantifying Accuracy and Precision

The value of any data set is correlated with its accuracy and precision. Historically, practitioners in the graphic arts industry dealt with these concepts intuitively through experience – the press operator had a "feel" for the press. When computers are asked to perform tasks once managed by intuition, however, chaos can result if the information once learned through intuition is not adequately translated into a language that can be understood by computers: numbers. This section describes the concepts of accuracy and precision and how to interpret measured values.

First, consider the statistics:

- sample size
- sample average
- population average
- sample standard deviation
- population standard deviation
- standard error

A population is defined as a large body of potential samples. Because it is so large, it cannot be exhausted by a sample of any size, and therefore, cannot be entirely sampled. The population is characterized by some center point (population average) and spread (population standard deviation). In order to gain information about this unknown population, a sample is drawn.

The size of the sample is described by an integer number of measurements, that is, the sample size. The sample average is defined as the arithmetic mean of the measured values and describes the center point of the distribution of measurements. The sample standard deviation describes the spread of measurements about the sample average.

Standard error describes the distribution of sample averages if samples of the same size are repeatedly drawn from the population. In other words, standard error describes the spread of averages if samples are replicated. Mathematically, standard error, $\sigma_{\bar{x}}$, is defined as sample standard deviation divided by the square root of the sample size as shown below.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}},$$

where σ is the sample standard deviation and n is the sample size.

In lay terms, standard error is a statement of precision. The concept of standard error is that a sample of size one will have an average value that with repeated sampling will vary identically to the parent population. However, as sample sizes increase, the variability of sample averages will be decreased proportionally to the size of the sample. For example, sample averages from samples of size n = 4 will have 1/2 the variation of the parent population while sample averages from samples of size n = 16 will have only 1/4th the variation of the parent population. An increase in the sample size corresponds to a decrease in standard error or, in other words, increased precision.

The concept of standard error is couched in the idea that the true population center point is unknown and an experiment is used to find the center point. Since samples drawn from the population vary, replicated sampling increases the precision of the average values measured. But since finding the exact value for the center point is very unlikely with any level of replication, standard error provides a method of determining the level of uncertainty between a given sample average and the true population average.

To put this back into the realm of printing process control and characterization, the data contained within a characterization data set based on a single measurement from a single press sheet will vary with the same magnitude as the parent population. As the number of samples increases, the precision of the average value also increases. While precision is increased with sample size, accuracy can only be improved by sampling more variables.

The concept of accuracy involves a reference value. Accuracy is the comparison of a measured value with a reference value. A highly accurate measurement is one for which the difference between the measured and reference values is small whereas an inaccurate measurement is one for which the difference is large. The difference between the measured and reference values is sometimes called bias. Common references in printing are the process aims defined by organizations like SNAP, GRACoL and SWOP.

The interpretation of a measured value, that is, the linkage between a measured value and its value in application, is best described by example. Consider two samples illustrated in figure 1 that are drawn from the same population where the first is very small (sample A) and the second is very large (sample B). The precision of sample A is very low due to the small sample size. That is, its distribution is very wide. The precision of sample B, on the other hand, is very high – its distribution is very narrow. Accuracy is quantified by subtracting the aim, or reference, from the measured value. For simplicity, both samples have the same accuracy in the case of this example.

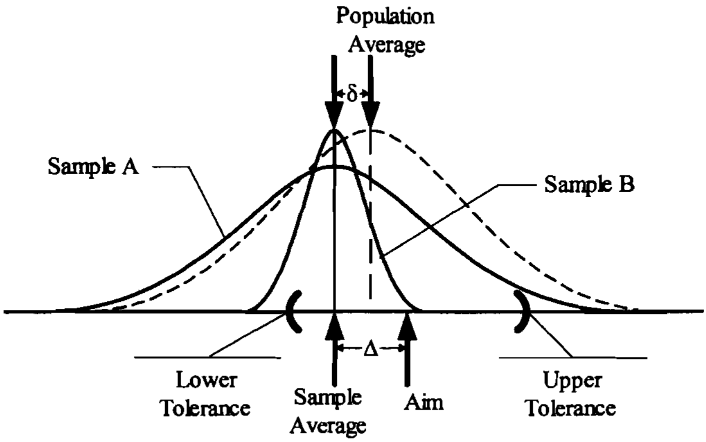


Figure 1. This figure shows the distributions of 2 samples with the same average value relative to an aim, or reference, value. Precision is described by the width of the distribution. Accuracy is described by the difference between the average and aim values. In this chart, accuracy and precision are represented by Δ and δ. A possible parent population is illustrated using a broken line, but keep in mind that the exact nature of the parent population is never really known.

The question is how much confidence can be placed in either result. Both samples illustrated have the same accuracy, but comparing tolerances with the precision of the samples allows the quantification of the goodness, or confidence, of the measured value. In the case of sample B, examination of the figure shows that the true population average has a very good chance of falling into the region defined by the aim value and tolerances. On the other hand, since the width of sample B's distribution is so wide, there is only a medium chance that the sample average will actually predict that the population's average actually falls within the same region.

In the language of confidence intervals and using the proportion of this figure, it can be stated with 80% confidence that the value determined from sample A indicates the true population average satisfies the specification. Only a small portion of the left tail of this distribution falls outside of the specification. On the other hand, the level of confidence associated with the value determined using sample B is only about 50%.

This example illustrates several points. First that the true population average is only approximated by sampling and that the precision statement refers to the "goodness" of that approximation. Since the sample average is only an approximation of the population average, the level of certainty, or confidence, associated with the calculation of accuracy is governed by the precision in the sample.

The interpretation of sampled data may produce confusing results, particularly when small samples are used. For example, it is common that a sample average will indicate that a population average is compliant with a specification when in reality, the opposite is true. It is also possible that a measured value indicating non-compliance with a specification was derived from a compliant population. In the realm of press testing, the risks are that a good test could be labeled bad, or conversely, a bad test labeled good. In an analog workflow, human intuition relative to press performance and tolerances sufficed. But when a computer was introduced in the process and the information it was given was based on inadequate data, chaos prevailed. The concepts of accuracy, precision and interpretation translate the intuitive information used to guide an imaging system in an analog workflow into a set of data that can be understood by computers and facilitate the digitization of a process.

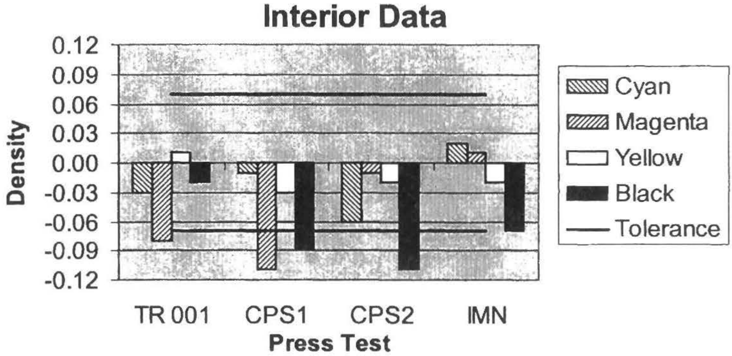
Analysis of Existing Press Test Data

The information presented in this section is intended to illustrate the level of variation in printing and how measurements are affected by the variability. Numerous press tests have been performed for various processes. Analysis of this data shows 3 things.

- First, variability within and between press sheets can be very large when measured numerically or visually.
- The probability of reproducing a press test within tolerances defined by the industry appears very low.
- And lastly, the end result of a press test is typically more questions than answers – confusion.

While the data may appear to paint a bleak picture, keep in mind that this presentation is not intended to diminish the value of the effort the individuals and organizations have invested in this work. On the contrary, this data is extraordinarily valuable for the industry to learn how to maximize the productivity of their changing workflow.

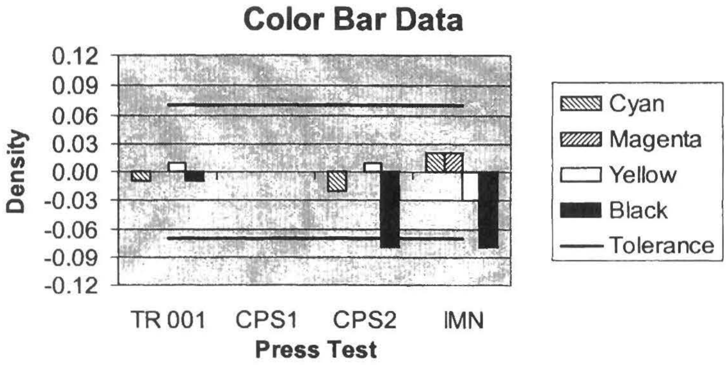
Consider, for example, that in the following graph the intended target for all 4 press tests was SWOP, but only the Imation press test narrowly satisfied the requirements.



Graph 1. Solid ink density measurements from IT8 targets of 4 press tests intended to meet SWOP specifications. The first, labeled TR 001 was the test from which the data in TR 001 was collected. The tests labeled CPS1 and CPS2 were press tests from which SWOP certified press sheets were produced. The last press test, labeled IMN, was the test contracted by Imation with the intent to replicate the data in TR 001.

Differences between aims, tolerances and press tests like the ones shown in graph 1 have created much controversy. For example, some question the validity of the characterization data in TR 001 that was derived from a press test that was not compliant with its process definition.

But, this is just one difficulty the industry has experienced. While graph 1 presented data from interior test form patches, graph 2 presents the data collected from press control bars.



Graph 2. Solid ink density measurements from press control bars of 4 press tests intended to meet SWOP specifications. Labels in this graph refer to the same tests as presented in graph 1. Control data was not available from CPS1.

Analyzing control bar data from these press tests shows that the control bar does not always predict the test form interior. In fact, for TR 001, the control bar indicates that the test is nearly perfectly compliant with the process definition, but the test form interior turns out to have a low magenta density. These difficulties are not unique to the web offset publication printing process. Table 1 shows the same problems in data collected from the newsprint process, SNAP.

Newsprint Solid Ink Densities from Press Control Bars and Test Form Interior

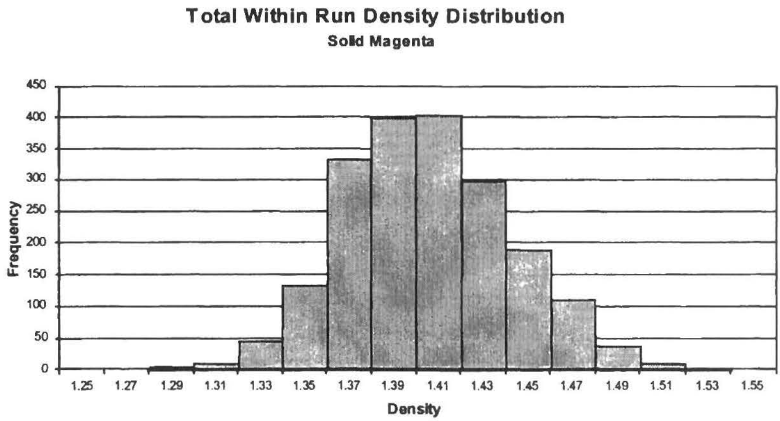
Ink	Aims	Tol	Color Bar	IT8.7/3
C	0.90	±0.05	0.91	0.90
M	0.90	±0.05	0.91	0.86
Y	0.85	±0.05	0.88	0.85
K	1.05	±0.05	1.01	0.96

Table 1. Solid ink density measurements from control bars and IT8 targets of a SNAP press test.

In table 1, control bar data indicates a press test in compliance with SNAP specifications, but solid ink densities extracted from the test form interior deviate from control bar values and are not compliant with the process definition. In particular, notice the deviation in black. As illustrated by the above graphs and table, it is clear that the printing industry has experienced serious difficulty consistently achieving process aims and reproducing press tests.

As alluded to in the introduction, the difficulty achieving process aims and reproducing press tests drove the authors to research the topic. A series of press tests was designed to analyze within and between press test variation such that it could be statistically analyzed. Since these tests were described in detail in a previous paper, only the pertinent results are discussed here.

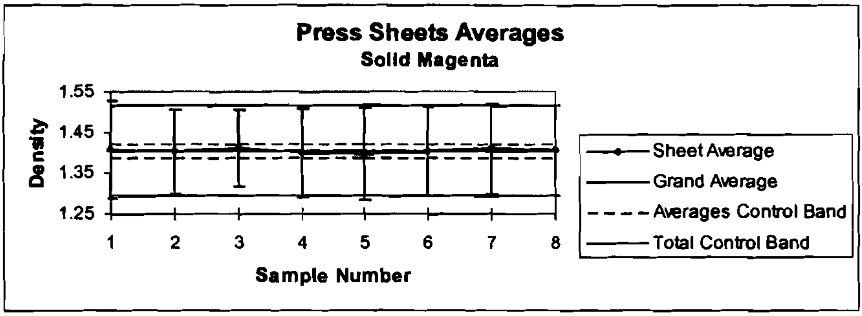
The first lesson from the Imation press tests was that within press sheet and within press run variations are large. The range of data is usually wider than the process specification. This is illustrated in the following graphs.



Graph 3. The distribution for solid magenta density from a press test run according to SWOP specifications. Data was collected from 247 randomly placed patches in 8 press sheets from an Imation press test.

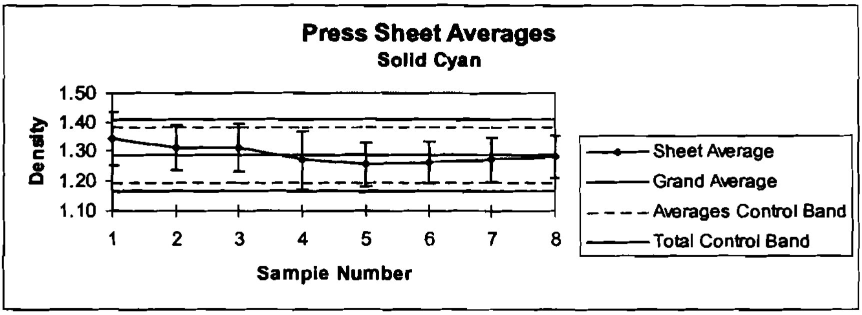
SWOP specifies a range of data from 1.33 – 1.47 for magenta. Even though this press test is nearly perfectly centered on the SWOP specification, the data ranges from 1.29 – 1.52. Five percent of the data could be considered out-of-spec. The width of this distribution was similar for all four process colors. While the specification for black is 1.52 – 1.66, the within run distribution runs from 1.47 – 1.73.

While the histogram above shows the total distribution, the variation can be resolved into within sheet and between sheet components. This is shown in graph 4.



Graph 4. Control chart of sheet average for solid magenta density from an Imation press test. Values for upper control limits (UCL) and lower control limits (LCL) are set at +/- 3 standard deviations. Error bars indicate the range of data within each sheet.

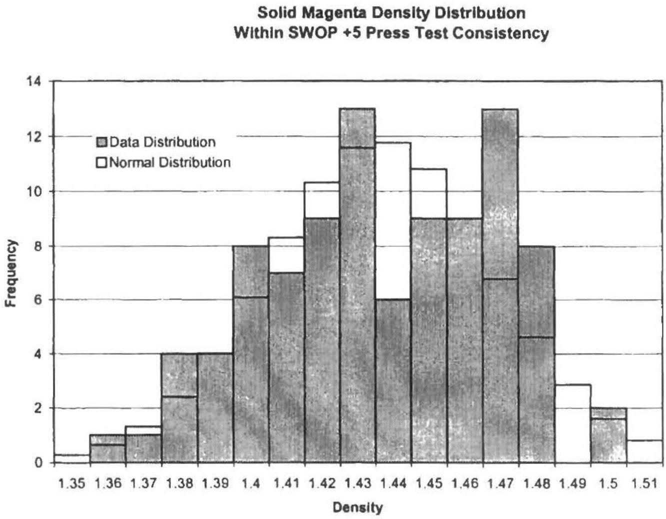
As this example shows, within sheet variability is very large while between sheet variability is quite small. But, this is not always the case. Graph 5 shows the case for cyan solid ink density from the same press test.



Graph 5. Control chart of sheet average for solid cyan density from an Imation press test. Values for upper control limits (UCL) and lower control limits (LCL) are set at +/- 3 standard deviations. Error bars indicate the range of data within each sheet.

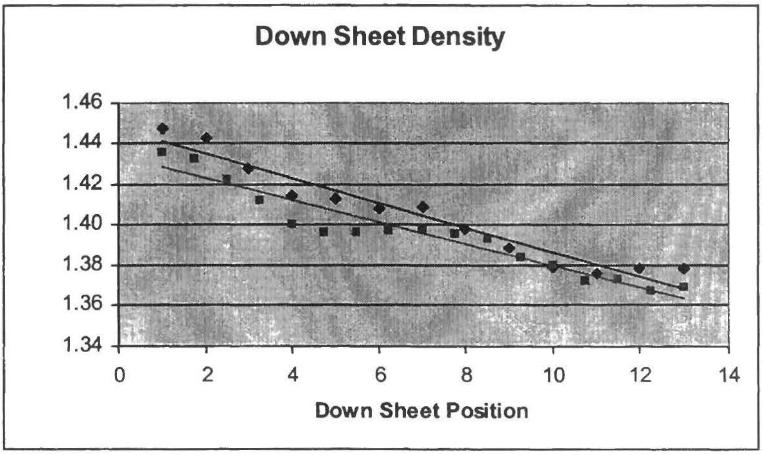
In the case of solid cyan, these control charts illustrate sheet to sheet deviations that are about the same as within sheet variations. Notice that total density variability is approximately the same for solid magenta and solid cyan. The net result is that the size of the total distribution is approximately the same. For cyan, the data ranges from 1.17 – 1.42.

Again, the size of distributions are not specific to any particular press test or process as shown in graph 6.



Graph 6. The distribution for solid magenta density from a press test run to a GRACoL specification. Note that this distribution is tri-modal with modes at 1.40, 1.43 and 1.47 corresponding to ink keys.

As the caption for graph 6 alludes, these distributions do not arise solely from random effects. In the case of the distribution shown in graph 3, 0.07 of the width is caused by ink starvation. Graph 7 illustrates the effect of ink starvation.



Graph 7. Down sheet variations caused by ink starvation for solid magenta from 2 Imation press tests.

The points illustrated in graphs 6 and 7 are that there are many factors that contribute to total variability. Two have been presented here, but there are a multitude of others related to spatial and temporal variables.

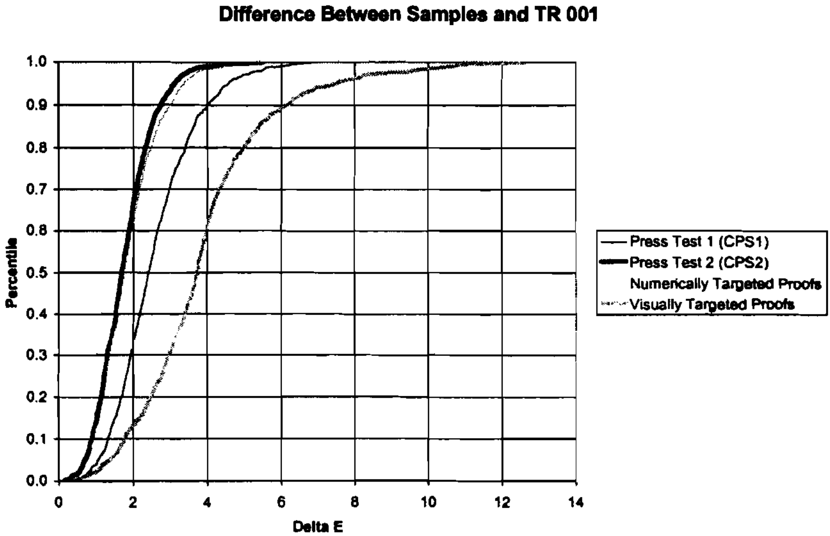
While several numerical variations in press test data have been identified, the real question is whether or not these variations have practical significance relative to visual impact. In an attempt to answer this question, the authors once again turned to the only characterization data widely available to the industry: TR 001.

While analyzing the visual impact of printed output is a more challenging proposition than analyzing the colorimetry data, Imation designed an experiment with the intent to evaluate the visual impact of the variations described above. The purpose of this experiment was to determine whether or not the numerical differences were visually significant.

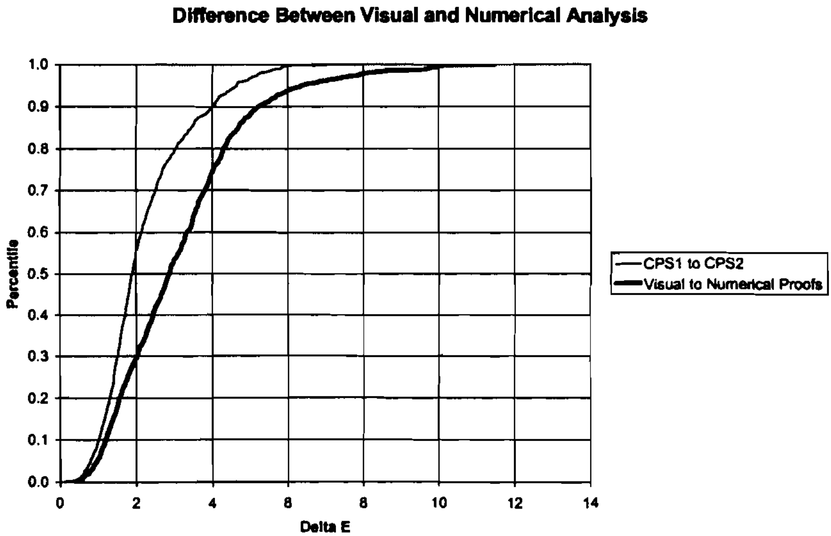
In short, the experiment involved the production of two sets of digital off-press proofs which were compared to SWOP certified press sheets. One set of proofs was aimed using the data contained in TR 001 and the other aimed at the best visual match of SWOP certified press sheets from the first set of certified sheets. A panel of individuals with various backgrounds in the graphic arts industry were asked to comment about the color match between both sets of proofs and the first set of certified press sheets. Additionally, the panel was asked to comment on the color match between the first and second set of certified press sheets. Recall that the two sets of press sheets were produced from 2 different press runs and are referred to as CPS1 and CPS2.

Results of the visual analysis indicated significant visual differences between all sets of images. In particular, the largest differences were found between the CPS1 press sheets and the proofs targeted to TR 001 data. Some classified the differences as objectionable. Differences between the CPS1 press sheets and proofs visually aimed at the CPS1 press sheets were identified, but generally, the color match was acceptable. Interestingly, differences between CPS1 and CPS2 press sheets were also mentioned. The most common comment was that the CPS1 set had better color while the CPS2 set was too muddy.

To quantify the visual results, IT8.7/3 targets were imaged with both sets of proofs. These targets, as well as the IT8 targets included with the press sheets were measured colorimetrically and compared with TR 001 and each other. Graph 8 presents the comparison with TR 001 and graph 9, with each other.



Graph 8. The cumulative distribution of delta E for the difference between SWOP certified press sheets and proofs with TR 001.



Graph 9. The cumulative distribution of delta E for the difference between SWOP certified press sheets and the difference between proofs targeted to the press sheets and targeted to TR 001.

Graph 8 shows that CPS2 and the numerically targeted proofs are closest to TR 001 with a distance at the 50th percentile of about 1.5 delta E. CPS1 is next with a 50th percentile value of about 2.5 delta E. The visually targeted proofs lie farthest from TR 001 at a distance of 3.5 delta E at the 50th percentile. Graph 9 helps quantify the visually identified differences. The distance between CPS1 and CPS2 is 2 delta E at the 50th percentile while the difference between visually and numerically targeted proof is almost 3 delta E at the same percentile.

This experiment confirmed the controversy originating with the observation that control bar data and test form data sometimes have large differences. This analysis shows that the printing process is generally a variable process. Significant variations are not only observed between press tests, but also within press tests and sheets. None of these tests is truly compliant with their specifications – on average or point by point. Only SWOP offers the opportunity to investigate the difference between visual and numerical representations of their process and significant visual differences support the numerical findings.

In light of the discussion about the relationship between fingerprinting, process control and characterization, this analysis raises several questions. Are the processes adequately defined? Is the process control adequate? Do specifications need to be adjusted? How is a good characterization prepared? How much confidence can be placed in the characterization? This analysis illustrates the high level of variability within press tests and relatively poor level of reproducibility between press tests as gauged by industry specifications. The following section analyzes the practices of press testing relative to variability and improving reproducibility.

Characterization Practices

The strategies and methods used to prepare a characterization govern the precision and accuracy of the resultant data. This section analyzes the practices of characterization relative to accuracy and precision and illustrates where and how the statistical concepts apply to fingerprinting, process control and characterization. Traditional practice has resulted in the controversy discussed above, but the work performed in the early and mid '90s was performed without the quantified knowledge of the variability within and between press tests. As more is learned about the variability in printing, methods have begun to change and improvements can be observed.

Traditional characterization practice has been to target the press test as close as possible to some defined condition. Then, a number of impressions were run. From these impressions, sheets were sampled to find the "sweet spot" where process control parameters best fit the process control requirements. From the sweet spot, characterization samples were drawn and measured. The test form typically consisted of the necessary process control elements, a single IT8.7/3 element in the default order and several pictorials. In short, the idea was to qualify a few sheets as close as possible to process aims and from those sheets, measure and document the colorimetric attributes.

The first difficulty with the traditional approach was finding sheets within the test run where solid ink density and dot gain are simultaneously within specification for all 4 primary colors. Data collected from several industry press tests indicates that compliance with process definitions typically runs at about 80 – 90 percent for each process control parameter. The probability of the combined event that all metrics simultaneously satisfy fingerprinting requirements is very low. With this level of compliance, only 10% of the sheets will have all metrics simultaneously within specification if all metrics are perfectly centered on their respective aims. As shown above, centering a press test on the aim is a rare event and the analysis of existing press testing data shows that in practice only about 1% of sheets will simultaneously satisfy all process control requirements. Considering that a typical press test consisting of 2000 impressions will only produce 20 qualifiable sheets relative to existing process definitions, finding sheets for characterization is an extraordinarily difficult task. Because this level of sampling is nearly impossible, sheets for characterization sometimes did not satisfy process control requirements.

The next difficulty encountered with traditional practice was that while several sheets were sampled, only a single characterization element was contained in the test form. In other words, within sheet variables were minimally sampled. Because so few within sheet variables were sampled, accuracy suffered. In other words, bias was maximized! This can be observed by comparing data presented in graphs 1 and 2. This difficulty can manifest itself in differences between control bar and characterization data.

For example, recall from graph 1 that the TR 001 characterization appears biased in magenta; that is, magenta solid ink density fell 0.08 below the specified value. The data in TR 001 was derived from the measurement of six test forms each of which only contained a single IT8 target. In terms of this discussion, the sheet to sheet variable was sampled six times while the within sheet variable was sampled once.

To determine the accuracy and precision of this particular metric requires knowledge of the components of variability and the use of standard error. The standard deviation for magenta density in this case is 0.03 for the within sheet variables and 0.01 for between sheet variables. These standard deviations can be confirmed by examining the control chart in graph 4. The within sheet precision using the concept of standard error for a single sample is identical to the parent population, or in this case, 0.03.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.03}{\sqrt{1}} = 0.03.$$

The precision of the between sheet variable sampled six times is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.01}{\sqrt{6}} = 0.004.$$

The combined precision, or expected range of data, using a 99% confidence interval is 0.18.

Because the process that produced these press sheets is characterized by large within sheet variations and small between sheet variations for solid magenta, it is not surprising that the sampling did very little to improve precision; notice that the 99% range of data in graph 1 is 0.20 and the standard error predicts an expected range of averages in a range of 0.18. Additionally, the data is biased toward the particular conditions under which the magenta patch was printed. A test form that included replicated magenta patches would improve both precision and accuracy. Precision would be improved by minimizing variability in results. Accuracy would be improved by reducing the bias associated with unsampled variables.

Because there is significant controversy surrounding existing press test results, an alternative approach to press test sampling was researched by the authors. In this approach, within sheet, between sheet and between press test samples are collected and accuracy and precision are observed to correlate with both the number of samples drawn and the number of variables sampled. While it is true that more individual data points fall out of specification than the traditional practice, some are low and some are high. The net effect is that in the long run, results converge and less bias is observed.

In order to maximize precision, as many variables should be sampled with as many replications as possible. This paper has identified within and between sheet variations, but the causes of these variations has not been explicitly assigned. If the sources of variation are unknown, samples should be drawn at random from within each press sheet and from within the press run. The only time samples should not be drawn at random is when the source of variation is known and the sampling plan is designed to detect the effect of the known or hypothesized source of variation. Random sampling necessitates variation of the test target in the test form and for this reason, Imation and others have introduced randomized test targets usually based on the IT8 input data set.

The intention of randomized test targets is to sample within sheet variables. Some of the causes of within sheet variations are known. For example, it is known that ink starvation is a factor that causes reduced down sheet / web ink densities in offset printing. Further, it is known that the amount of ink take-off in the early portions of the press sheet affects the magnitude of the ink starvation effect. But, because ink starvation is only one source of variation and other variations are observed, randomized targets are used to sample as many known and unknown causes as possible.

If the placement of a single color patch in a test form is considered relative to all the other patches, a replicated randomization has the effect of sampling the color in various printing conditions. In terms of ink starvation, patches near the top of the test form sample the condition where little ink take-off has occurred while patches near the bottom sample the condition where significant ink take-off has occurred. Ink key balance may be considered as another example. Patches located in various positions across the test form sample the various ink key conditions. This same discussion holds for unknown causes of variation. Since the nature of these factors is unknown, the randomization allows their sampling without in depth knowledge. Because randomization facilitates the sampling of all variables, randomization improves accuracy. Replication on the other hand improves precision.

Determining test form design without any knowledge of within sheet variability is challenging, but existing press test data provides a guide. Based on the knowledge of ink starvation and ink key balance, patches should be distributed both across and down the test form to accommodate these affects. But to accommodate unknown affects, the patches should distributed randomly within the sheet. To accommodate these seemingly conflicting objectives, a randomized block architecture satisfies the need to ensure that each color is sampled down and across the sheet while randomization satisfies the need to sample unknown factors.

A randomized block design divides the test form into down and across sheet areas, or blocks, into which randomized test elements may be placed. The number of blocks needed is determined by evaluating the level of accuracy and precision desired. In the case of magenta with the distribution described above and using standard error, the number of within sheet replications needed to achieve a precision within one half the SWOP specification of 0.07 with 99% certainty is 7. This number is determined by solving the equation of standard error for the sample size. Or,

$$n = \left(\frac{\sigma}{\sigma_{\bar{x}}} \right)^2 = \left(\frac{0.0300}{0.0117} \right)^2 = 7.$$

Using normal statistics theory which states that the 99% confidence interval corresponds approximately to a range of $\pm 3\sigma$, $\sigma_{\bar{x}}$ is determined by the relationship that the desired range, 0.035 is equal to 3 times the standard error. Mathematically, $0.035 = 3\sigma_{\bar{x}}$. Solving for $\sigma_{\bar{x}}$ yields 0.0117. In other words, the within sheet sample size is equal to the square of the within sheet standard deviation divided by the desired level of precision.

Whether or not randomizing the test element is completely necessary has not been proven. However, the data and statistical theory presented here shows that sampling a single test target is not sufficient to achieve the level of accuracy and precision required in characterization data. Randomizing and decoding these targets can be complicated and resource intensive. A less intensive method of increasing the number of within sheet variables is to use a standard test target, say the default IT8, replicated with various rotations. The net effect would be to sample more within sheet variables, but because the sampling is restricted to single target configuration, several unknown effects may be missed. For example, some patches will always be located on the perimeter and some patches will always be located near the center of the test element. If the magnitudes of these spatial effects is large, the data will be biased. However, if it can be shown that the size of the effects are small, then economic and ease of use factors should be considered when designing the test form. At this time, though, the difference between randomized targets and a single target replicated with different rotations is unknown and experimentation should be done such that an educated decision about where each type of sampling is best can be made.

With respect to sheet selection for measurement and ultimately the derivation of a characterization data set, a similar discussion relative to known and unknown causes of variation apply to the sheet selection plan. Because the causes of sheet to sheet variation are treated as an unknown, sheets should be selected randomly from the press test and the number of sheets drawn should be sufficient to achieve an adequate level of precision. More sheets than needed should be selected to accommodate experimental mortality or the rejection of sheets based on physical or other defects.

More succinctly, the number of sheets needed can be determined before the press test with knowledge of process variability and desired precision by solving the equation for standard error presented above for the sample size. Using the data presented in graph 5 for solid cyan which exhibits a 99% confidence range of 0.19, the equation for standard error shows that 8 sheets are needed to ensure that the resultant precision is within one half the SWOP specification of the true center point of the press test. More realistically, though, twice that many sheets should be sampled to verify press test stability and accommodate the rejection of sheets due unexpected problems.

In order to determine the level of within and between sheet sampling requires knowledge of the process before the press test is run. This shows the importance of fingerprinting and process control. These activities provide knowledge of the process and should guide the development of the sampling plan for characterization efforts.

Sheets selected for characterization should be verified compliant with process aims and control limits, but no further selection criteria should be applied. The logic here is that while the sampling plan should accommodate known factors, it should not exclude unknown factors. The risk of excluding sheets based on their distance from the intended target is that the resultant data may be biased because unknown factors were excluded. Defining anything more than a random sheet selection criteria should be governed by the knowledge of the affect of some factor. If selection criteria is defined without this knowledge, the effects of unknown causes will not be accounted for and bias will be introduced.

Lastly, results from more than one press test should be combined to form the final data set whenever possible. While the number of press tests per defined process is small, the results shown in graphs 1 and 2 illustrate that the same theoretical concepts hold for between press test factors. The data presented in these tables show that individually, each press test exhibits significant deviation from their intended process aims in some metric. But, if the collection of tests are considered enmass, the resultant data set appears to converge on values that simultaneously satisfy the process specification for all process parameters. Table 2 shows the combined Solid Ink Density results of process data from 4 press tests run to SWOP aims.

Solid Ink Density from the Combination of 4 Press Tests

Ink	Aim	Tolerance	Data
C	1.30	±0.07	1.30
M	1.40	±0.07	1.40
Y	1.00	±0.07	1.00
K	1.59	±0.07	1.53

Table 2. Solid ink density measurements combined from 4 press tests intended to meet SWOP specifications. Notice that all solid ink densities are compliant with the SWOP specifications.

The intention of the sampling plan for all variables, within sheet, between sheet and between press test, is not necessarily to quantify variability, but more to account for the effects of all causes of variability and ensure precise and accurate characterization data.

Conclusion

After all the theoretical discussion and data analysis, what has been learned?

First, that press output is typically very noisy and that the size of the variations are large relative to process specifications and visually perceivable levels. Using standard error, we can quantify the level of accuracy and precision in press test results. Significantly, standard error may be used to determine the sample size before a press test is run. Given knowledge of process variability and desired levels of accuracy and precision, standard error facilitates the development of sampling plans.

We have learned that traditional methods did not sample all printing variables and that results from traditional tests show predictable results given the methodology. That is, they are typically imprecise and impaired by bias in one or more process parameters. Again, this is pointed out not to diminish the value of the work, but to learn from experience. Press tests run using sampling plans that incorporate randomized elements demonstrate better accuracy (less bias). Press tests run using sampling plans that incorporate replication show significantly higher levels of precision. These observations are true for all within sheet, between sheet and between test variables.

The goal of press testing is to first, determine the process definition and then, to characterize the color output. Using traditional methods, press test reproducibility was larger than the process specifications. Therefore, controversy surrounding these results was serious enough to impede implementation. Analyzing the traditional work shows that modified methods will improve reproducibility and hopefully ease the implementation of digital technology.

Selected Bibliography

Fisch, Richard and Bartels, Sharon

1999 "Characterizing an Ink on Paper Four Color Print Process," TAGA Proceedings 1999, Rochester, NY, pp. 433 – 451.

Graphic Communication Association

1999 "General Requirements for Applications in Color Offset Lithography, Version 3.0," GCA, Alexandria, VA.

Hutcheson, Don

2000 "Concept to Output," Imation Corporation, Oakdale, MN.

NPES

1995 ANSI/CGATS.6 , "Graphic Technology – Specifications for graphic arts printing – Type 1," New York, NY.

1995 ANSI/CGATS TR 001, "Graphic Technology – Color Characterization Data for Type 1 Printing," New York, NY.

Siljander, Roger, Fisch, Richard and Bartels, Sharon

2000 "Characterization Data Requirements for Color Management," TAGA Proceedings 2000, Rochester, NY, pp. 521 – 556.

2001 "Introduction to Fingerprinting and Characterization," ANSI/CGATS/STF2 N 027, NPES, New York, NY.

SWOP

1997 "SWOP, Eighth Edition," SWOP Inc., New York, NY.

1998 SWOP 5/1998, "Digital Specifications and Requirements," SWOP Inc, New York, NY.