# System for Assessing, Exploring and Monitoring Offset Print Quality

JENS LUNDSTRÖM
Intelligent Systems Laboratory
Halmstad University
Box 823, S 301 18 Halmstad
SWEDEN
jens.lundstrom@hh.se

ANTANAS VERIKAS
Intelligent Systems Laboratory
Halmstad University
Box 823, S 301 18 Halmstad
SWEDEN
antanas.verikas@hh.se

*Abstract:* Variations in offset print quality relate to numerous parameters of printing press and paper. To maintain constant quality of products, press operators need to assess, explore and monitor print quality. This paper presents a novel system for assessing and predicting values of print quality attributes, where the adopted, random forests (RF)-based, modeling approach also allows quantifying the influence of different parameters. In contrast to other print quality assessment systems, this system utilizes common print marks known as double grey-bars. A novel virtual sensor for assessing the miss-registration degree of printing plates using images of double grey-bars is presented. The inferred influence of paper and printing press parameters on print quality shows correlation with known print quality conditions.

*Key–Words:* Virtual sensor, Random Forests, Print quality assessment, Decision support system, Variable importance, Bar-code reader

## 1 Introduction

Offset lithography is the most widely used technique for newspaper printing. Four primary inks, cyan (C), magenta (M), yellow (Y) and black (K), are used to produce colour images in lithographic offset 4-colour printing. A roller clothed with a rubber blanket, a steel roller sprayed with fountain solution which mainly consists of water, a roller equipped with a printing plate and finally the paper, see Fig. 1, are the main elements in offset printing.

The surface properties of the plates allow the ink roller to transfer ink onto dry areas of the printing plates. The ink is then transferred to the rubber blanket, which in turn transfers the ink onto the paper. A specific balance of the amount of ink and fountain solution has to be kept to maintain even quality throughout the whole print job. Other than this balance, parameters as paper properties, paper web tension, ink recipe, air humidity and temperature, ink temperature, wear of printing plates also affect the print quality [6].

Today there exist commercial systems able to automatically control some specific print quality attributes. Such systems, however, are not capable of providing users with information about which process parameters influence a particular print quality attribute.

A system, able to predict correct values of print quality attributes given a set of parameter values char-
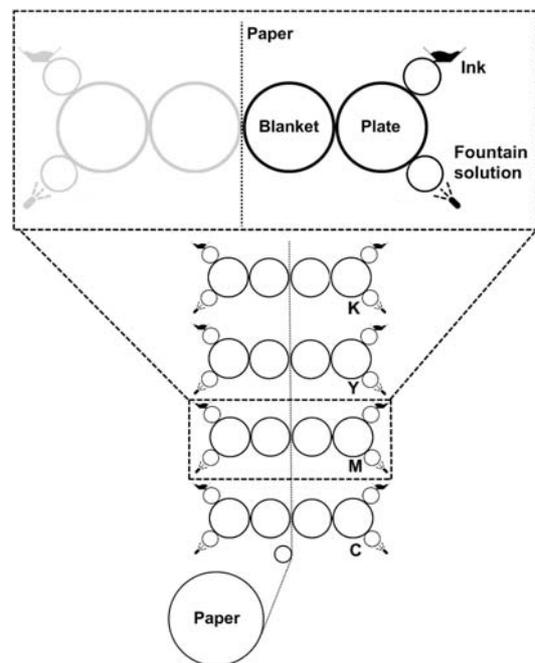


Figure 1: Key elements in a printing press.

acterizing papermaking and printing processes, would be of great value for manufacturers of paper and newspapers. It is believed that exploitation of such information may lead to higher quality of prints. To our knowledge, there have been only two attempts

to use some paper parameters in print quality modelling [4, 7].

There are numerous examples of virtual sensors applied in both papermaking and printing industries [12]. Trepanier et al. have demonstrated that image processing can be used to estimate paper surface characteristics such as unevenness of a paper sheet [11]. A system able to detect regions suitable for measuring colour registry i.e. misalignment of printing plates has also been proposed [10].

This paper concerns a unique system developed to assess, explore, and monitor print quality in offset colour printing by using data from paper manufacturing and printing processes. The core of the system is a set of virtual sensors operating on images acquired on-line from the printing press. The outputs from these sensors are soft measures, or inferential calculations of print quality. The developed software of the system allows predicting values of these measures, called quality attributes (QA), using parameters characterizing the paper manufacturing and printing processes as input variables to the prediction model. Random Forests [1] is a core technique used to predict QAs. In addition to predicted values of QAs, random forests also provide an estimate of the variable importance—a measure of impact the variables have on model accuracy.

## 2 Approach

The proposed system is capable of assessing print quality attributes from images acquired on-line in printing press and linking these quality attributes with parameters of both paper manufacturing and printing processes. In contrast to previous studies relying on detection of suitable areas, where an image is acquired and evaluated, the proposed system uses designated halftone areas known as double grey-bars, shown in Fig 2. A typical grey-bar is of $8 \times 4$ mm size.
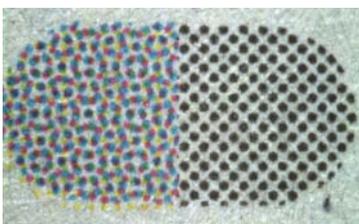


Figure 2: An example of a double grey-bar.

These grey-bars are in ordinary production used for manual inspection by operators eyes or using densitometers. Grey-bars are common in world lithographic newsprint and are typically printed at the edge of each page, and are, therefore, desirable to be used
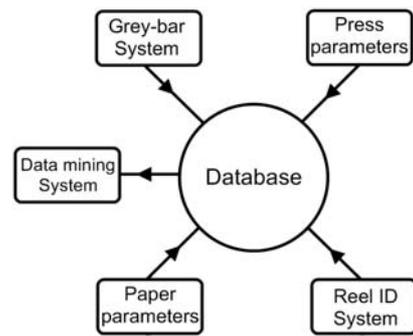


Figure 3: Flow of information from various sources.

as measuring areas for automatic print quality assessment. Use of well defined measuring areas, such as the grey-bars, is desirable when comparing different print jobs, detecting trends, and modelling print quality.

To model print quality, random forests [1] are used in this work. Random forest is a general data mining tool capable of accomplishing various data analysis tasks. Ability to provide the importance of input variables for model accuracy is a very useful function of the random forest software exploited in this work.

## 3 System Design

A database containing parameters of printing press and paper, and print quality attributes is a core of the developed system. A schematic view of the system is depicted in Fig 3. The quality attributes and parameters are accessed by the developed software for data mining.

Virtual sensors for computing print quality attributes, described in detail in section 5, operate on images of grey-bars obtained from an 8-bit CCD colour camera. The camera traverses over the web and acquires images of grey-bars of $1008 \times 1007$ pixels. One pixel corresponds to approximately 0.01 mm. In conjunction with the camera system a rotary encoder is mounted onto the black ink roller to synchronize grey-bar position with the camera and flash. Grey-bar images and IDs of the images are stored in the centralised database. In total 20 grey-bars are scanned along the printing cylinder at web speeds up to 15 m/s. It takes approximately two minutes to make the scan. A xenon flashlamp with a continuous line spectrum was used, to have low illumination influence on colour measurements.

A virtual paper reel identification sensor was developed for the press room. This enables linking grey-bar measurements with corresponding press and paper

parameters. Reels are identified by their bar-codes, processed by the developed image processing software. Reading bar-codes in a press room is a non-trivial task due to varying illumination conditions and occlusion of bar-codes.

Data linked to a specific paper reel are received upon request from the paper mill using a developed protocol. Paper reel parameters are assembled into a file in the XML format.

Software was developed for assessing, exploring, and monitoring print quality based on data available in the data base. There are two main software modes, *Exploring* and *Monitoring*. Predictive modelling is the main feature of the software working in the *Exploring* mode. In *Monitoring* mode, users can analyse print quality attributes in time domain and over the full web width. Fig 4 illustrates main software windows of the two modes.

The R implementation of random forests ported by Andy Liaw [8] is used for both prediction and exploring various process parameters. *R* functions are called from the software developed in C++ and the *Qt GUI library*. Plot rendering is handled by calling *Gnuplot* and *R* from the software.

## 4 Parameters

Paper parameters used in this work can be divided into four groups:

**Lab tests** are 13 parameters measured on a paper sample strip of a reel. **Online scanning** by the Quality Control System (QCS), located just before the paper is rolled into a jumbo-reel. To obtain the scans, a measurement head is traversed over the paper width, where each edge-to-edge scan takes 30 seconds. The head is equipped with sensors to measure moisture content, reflected light, thickness, and dry weight. **Pulp recipe** is acquired from the paper mill recipe system. **Paper machine parameters** such as press section speed and amount of additives. In total, values of 20 paper machine parameters are stored for each paper reel.

Several sub-systems are used to acquire relevant measures characterizing the printing process. The **grey-bar camera system** is located just after the black ink roller. For each scanned grey-bar 23 print quality attributes are computed using virtual sensors. Amount of ink, ink temperature, printing speed, press room humidity and air temperature, amount of dampening solution supplied by 8 nozzles for each ink, constitute the group of **printing press parameters**. Colour register is controlled, with sensitivity of 1/10 mm, by moving the ink rollers in the machine-direction (MD) and cross direction (CD).

Each measured grey-bar is mapped to 130 parameters characterizing paper and the printing press. However, when the *Exploring* mode is used, each observation vector corresponds to average and variance descriptors for a set of measurements within one paper reel.

## 5 Assessing Print Quality

Various parameters are used to characterize print quality [12]. The following print quality attributes are used in this work:

1. *Colour deviation*—usually measured as $\triangle E$ in the $L^*a^*b^*$ colour space:

$$\triangle E = [(\triangle L^*)^2 + (\triangle a^*)^2 + (\triangle b^*)^2]^{1/2} \quad (1)$$

where the difference is computed between "coloured" and "black" parts of the double grey-bar. We used $\triangle E$, $\triangle L^*$, $\triangle C = [(\triangle a^*)^2 + (\triangle b^*)^2]^{1/2}$, $a$, and $b$ to characterize colour deviation.

2. *Miss-registration* degree of printing plates in both $X$ (CD) and $Y$ (MD) directions for $C$, $M$, and $Y$ inks with respect to $K$. Pseudo-code for the algorithm is shown in Algorithm 1.

3. *Dot deformation*. Two measures were applied: standard deviation of the distance to the dot center $s_d$ and dot shape factor $f_s$:

$$f_s = \frac{p^2}{4\pi A} \quad (2)$$

where $p$ is dot perimeter and $A$ stands for dot area.

4. *Ink density estimate*. To obtain the estimate, we compute the average of $d(x, y)$—black ink density estimate at pixel $(x, y)$:

$$d(x, y) = -\log \frac{p(x, y) - n_{cam}}{pref - n_{cam}} \quad (3)$$

where $p(x, y)$ is pixel intensity, $n_{cam}$ is camera noise, and $\overline{pref}$ is paper reference average intensity.

5. *Dot gain*—the difference between the supposed average dot size and the actual measured average dot size—tone value. A tone value was measured in the "black" area of a double grey-bar.
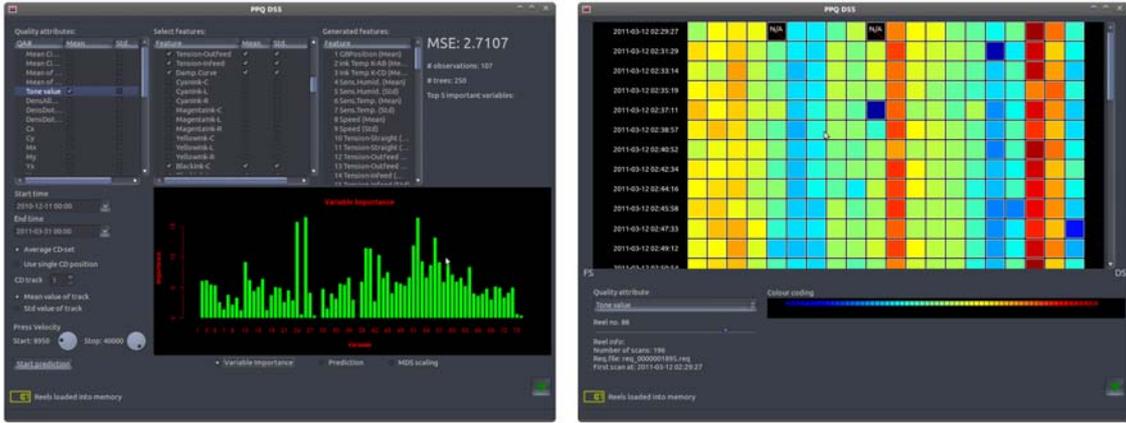
Figure 4: Left: *Exploring* mode. Right: *Monitoring* mode.

6. *Noise level assessed for C, M, Y, and K areas.* The noise level was assessed by applying the following measure:

$$\sigma_n^2 = \frac{1}{36(W-2)(H-2)} \sum_{x,y} (I(x,y) * N)^2 \tag{4}$$

where $W$ and $H$ is the image width and height, respectively, $N$ is a filter making the estimate less sensitive to structure in the image, and $*$ stands for convolution. The measure assesses the standard deviation of the noise and was originally suggested by Immerkaer [5]. The method is suitable for assessing noise in our grey-bar images due to low sensitivity to structure.

7. *Contrast assessed in "black" and "coloured" double grey-bar parts.* We assess contrast as the root mean square (RMS):

$$\text{RMS} = \sqrt{\frac{\sum_x \sum_y L(x,y)^2}{W \cdot H}} \tag{5}$$

where $L(x,y)$ is the $L$ image component in the $Lab$ colour space. This measure is intended to estimate the magnitude of the variation.

In total 23 print quality attributes were used. A unified measure of print quality has been studied in [9] using a similar set of print quality attributes.

## 6  Exploring Print Quality

Random forests (RF) [1] based modeling was used to predict values of print quality attributes using paper and press parameters as independent variables. Due to low computational complexity, RF can handle thousands of variables of different types with many missing values. For an RF tree grown on a bootstrap sample, the out-of-bag (OOB) data can be used as a test set for that tree. As the number of trees increases, RF provides an OOB data-based unbiased estimate of the test set error, estimate of variable importance, and the data proximity matrix.

We use the regression accuracy-based estimator of variable importance [2] in this work. The importance measure $\overline{D}_j$ for variable $x_j$ is given by

$$\overline{D}_j = \frac{1}{B} \sum_{b=1}^{B} (R_{b,j}^{oob} - R_b^{oob}) \tag{6}$$

where $B$ is the number of bootstrap samples (trees in the forest), $R_b^{oob}$ is the regression error for the OOB data by the tree $T_b$, and $R_{b,j}^{oob}$ is the regression error for the OOB data when the $x_j$ values in the OOB set were randomly permuted. Thus, variable importance is given by the decrease in regression accuracy.

## 7  Experimental Studies

To explore the influence of the paper and press parameters on black ink density variation, the *Exploring mode* was used. The data used for this study are values of 77 parameters collected from 107 reels. A random forest of 2000 trees was created using 8 randomly selected features to split a node. Two data sets were used. First, all grey-bars were used—even those sampled at low press speeds. Then, to create the second data set, the start-up sequence was filtered out by leaving aside grey-bars corresponding to press velocities below 5000 prints/hour. The variable importance values computed using the full and filtered data sets are shown in Fig 5 and Fig 6, respectively. In red shown are press parameters and in green–paper parameters.

1. Linearly transform an RGB image to $C$, $M$, and $Y$ images.

2. Compute binary images $B_C$, $B_M$, and $B_Y$.

3. Compute $B_K = C < t_C$ & $M < t_M$ & $C < t_Y$, where $t_C$, $t_M$, and $t_Y$ are the binarization thresholds.

**foreach** *C, M, and Y ink $i$* **do**

    4.1 Compute a 2D FFT of image $B_i \rightarrow F_i$.

    4.2 Remove frequencies from $F_i$ not expected in a halftone raster for ink $i \rightarrow Fl_i$.

    4.3 Inverse transform $Fl_i$ and multiply pixelwise by $B_K$ and $B_i \rightarrow Bh_i$.

    4.4 Build average vectors of rows and columns ($\mathbf{a}_r$ and $\mathbf{a}_c$) from $Bh_i$.

    4.5 Apply low-pass and derivative filters on $\mathbf{a}_r$ and $\mathbf{a}_c$ and threshold "edges" $\rightarrow \mathbf{af}_r$ and $\mathbf{af}_c$.

    4.6 Use prior knowledge of grey-bar height and width in conjunction with candidate edges in $\mathbf{af}_c$ and $\mathbf{af}_r$ to build membership functions of "plausible edges" fuzzy sets.

    4.7 Multiply values of membership functions to extract the most "plausible" edge candidates from $\mathbf{af}_c$ and $\mathbf{af}_r$.

**end**

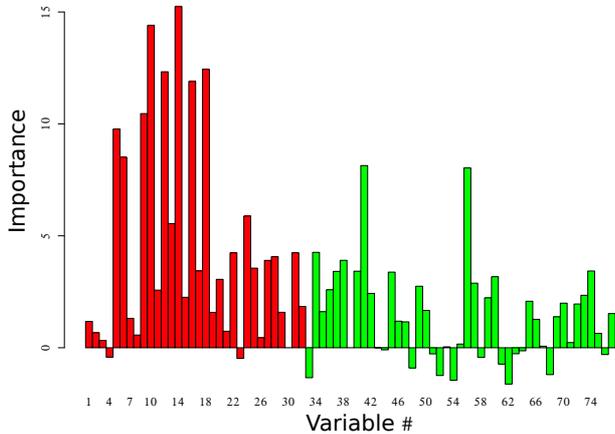    **Algorithm 1**: Assessing miss-registration.



Figure 5: Variable importance when modeling black ink density variation. Full range press velocity was used.

The figures indicate that press parameters are more important than the paper parameters when press start-ups and shut-down sequences are included. However, paper parameters are more important than the press parameters when the printing process stabilizes—when press start-ups and shut-downs are excluded. The five most important parameters for
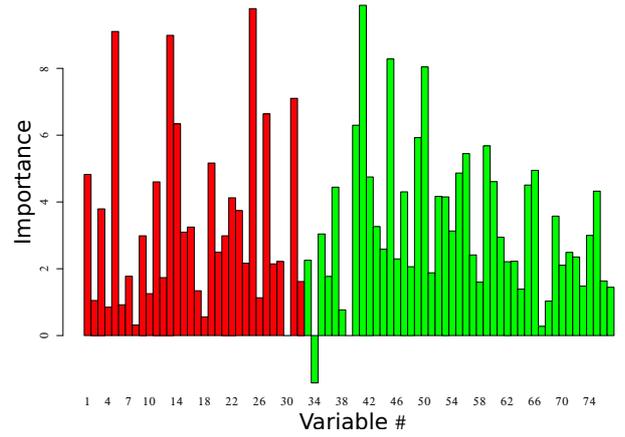


Figure 6: Variable importance when modeling black ink density variation. Low press velocities excluded.

Table 1: Top 5 parameters.

| N | With low velocities | Without low velocities |
|---|---|---|
| 1 | (PR) Variance of web tension (sensor 3). | (P) Surface roughness. |
| 2 | (PR) Variance of press speed. | (PR) Fountain solution nozzle setting. |
| 3 | (PR) Variance of fountain solution setting. | (PR) Mean humidity in a press room. |
| 4 | (PR) Variance of web tension (sensor 1). | (PR) Mean of web tension (sensor 3). |
| 5 | (PR) Variance of web tension (sensor 2). | (P) Paper permeability. |

each model identified in the experiment are shown in Table 1, where PR stands for printing press and P means paper. The first column of Table 1 lists parameters that are significant for the press startup. When the low press speeds where filtered out, paper parameters appeared, see the second column of Table 1. Surface roughness is a parameter, known for affecting ink transfer from blanket to paper [3, 6].

In the *Monitoring* mode, users can analyse quality attributes as matrices or images, where each row corresponds to a scan of grey-bars in the cross-direction and columns corresponds to the time dimension. We use colour to depict values of quality attributes in such type of data representation, see Fig 4 (right). Each scan is tagged with date and time. Measurement errors, e.g. black ink is missing, are presented in the matrix as "invalid measurement". Deviations in quality attribute vales are easily detected by the operator as colour of the matrix elements change.

Data from two reels, A and B, acquired online in ordinary production were explored by analysing each grey-bar and applying the virtual sensor to assess the miss-registration degree. Fig 7 illustrates the miss-registration degree assessed for C, M, and Y inks at

different positions across the printing cylinder.

One can clearly see similar miss-registration patterns for C and M inks, but not Y. This is due to an effect known as *fan-out* (related to paper swelling). The effect can now be studied intuitively and graphically.
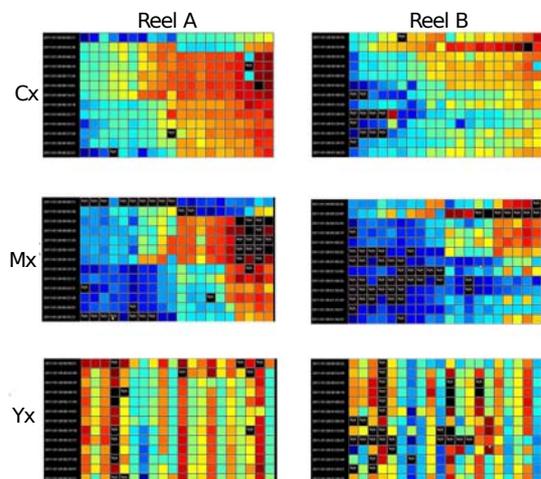


Figure 7: Matrices of miss-registration degree for cyan, magenta, and yellow inks measured by the virtual sensor.

# 8   Conclusions

This work has resulted in a novel system for assessing, monitoring and exploring print quality. An important part of the system is the set of virtual sensors used to assess print quality attributes. Quality attributes are linked to paper and press parameters and allow users, in intuitive fashion, to assess, model, and monitor print quality attributes using the *Monitoring* and *Exploring* modes of the software.

It was demonstrated experimentally that models are able to capture various effects of paper and print interaction such as influence of *surface roughness* on ink density, *fan-out*, and others.

Future work involves acquiring a large data set enabling studies of various "low frequency" effects on print quality. Hundreds or even thousands of reels are required to obtain a statistically significant result. Also, the monitoring mode can be extended to an online version, where operators in realtime can follow the quality matrices emerge as the print run continues. The work can be extended to obtain a proactive decision support system.

*References:*

[1] L. Brieman. Random forests. *Machine Learning*, 45, 2001.

[2] L. Brieman. Rftools for predicting and understanding data. *Tech. rep., Berkeley University, Berkeley, USA, http://oz.berkeley.edu/users/breiman/RandomForests /cc.papers.htm*, 2004.

[3] C. Fellers and B. Norman. *Pappersteknik*. Avdelningen för Pappersteknik, Kungliga Tekniska Högskolan, Stockholm, 1996.

[4] L. Guan and J. Lin. Study on the offset color reproduction control system based on fuzzy neural network. *Lecture Notes in Computer Science (LNCS)*, (5552):109 – 117, 2009.

[5] J. Immerkaer. Fast noise variance estimation. *Computer Vision and Image Understanding*, 64(2):300 – 302, 1996.

[6] H. Kipphan. *Handbook of Print Media*. Springer, 2001.

[7] L.G. Leloup. *Measurement and Prediction Procedures for Printability in Flexography*. PhD thesis, Swedish Royal Institute of Technology, June 2002.

[8] A. Liaw and M. Wiener. Classification and regression based on a forest of trees using random inputs. http://cran.r-project.org/web/packages/randomForest/index.html.

[9] J. Lundström and A. Verikas. Assessing print quality by machine in offset colour printing. *Submitted to Expert Systems With Applications (2011-07-18)*.

[10] C. Sodergard and R. Launonen. Inspection of colour printing quality. *International Journal of Pattern Recognition and Artificial Intelligence*, 10(2), 1996.

[11] R.J. Trepanier, B.D. Jordan, and N.G. Nguyen. Specific perimeter: a statistic for assessing formation and print quality by image analysis. *TAPPI Journal*, 81(10), 1998.

[12] A. Verikas, J. Lundström, M. Bacauskiene, and A. Gelzinis. Advances in computational intelligence-based print quality assessment and control in offset colour printing. *Expert Systems with Applications*, (38):13441 – 13447, 2011.