

LINEAR REGRESSION MODELLING OF LOCAL PRINT DENSITY IN GRAVURE PRINTED SC PAPER

U. Hirn¹, M. Lechthaler², E. Wind³ and W. Bauer³

¹ Corresponding author. Graz University of Technology, Institute for Pulp- and Paper Technology, Kopernikusgasse 24, 8010 Graz, Austria. E-mail: ulrich.hirn@tugraz.at

² Andritz AG, Stattegger Strasse 18, 8045 Graz, Austria. E-mail: markus.lechthaler@andritz.com.

³ Graz University of Technology, Institute for Pulp- and Paper Technology
Email: elmar.wind@tugraz.at wolfgang.bauer@tugraz.at

Abstract. In our paper we examine the interrelation of local print density variations in gravure printed SC paper to local properties of the same paper sample. For this purpose we analyze maps of five local paper properties. These are local values for basis weight, surface topography, ink penetration, gloss and refractive index, which have been registered to maps of local print density. We apply multiple linear regression models in combination with a technique called *extra sum of squares*. This technique permits to quantify the contribution of the individual variables to the model. Specifically it is possible to analyze to which degree the information a variable provides is *redundant* to other measurements or not. We introduce a visualization technique that offers a condensed overview on relevance as well as redundancy of the individual model variables. Results for 20% and 70% tone value of gravure printed SC paper are presented.

Keywords: Image registration, Print mottle, Paper property maps, multiple linear regression, ANOVA

1 Introduction

Research on printability has increasingly focused on measurement methods that evaluate *local* paper properties, i.e. methods that deliver 2D paper property maps. These methods measure the local variations of paper properties and thus enable direct comparison to local inhomogeneity (i.e. print mottle) or defects in the print. A common approach is to *register*, i.e. to spatially align, an image of the printed paper to paper property maps of the printed region. Such aligned maps enable qualitative analysis by visual inspection but also quantitative analysis of interrelations between local paper properties and local print density, e.g. by point-wise correlation [1, 6, 10]. Local print density variations have been registered with maps of local basis weight (beta formation) and surface topography [1, 6, 10] or coating thickness and latex concentration [14]. Regions with missing ink have been linked to maps of surface topography [4] and formation [13].

In our paper we will model the interrelation of local print reflectance in gravure printed SC paper to local properties of the printed paper sample. For this purpose we

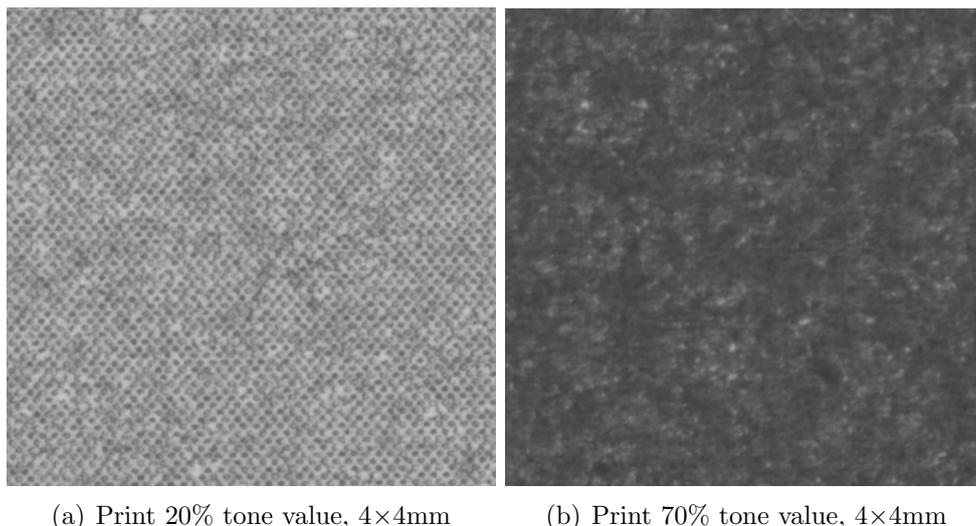


Figure 1. High resolution images of commercial SC-A paper G1 printed black with 20% (a) and 70% (b) tone value. The paper has been printed under industrial conditions in a commercial printing press.

analyze maps of five local paper properties. These are local values for basis weight, surface topography, ink penetration, gloss and refractive index, which have been registered to maps of local print reflectance.

We create statistical models using multiple linear regression. They model the local print reflectance values as a function of the five local paper property values p_1, p_2, \dots, p_5 . The key problem of multiple predictor modeling is to analyze the interrelation between the variables and to evaluate which variables contribute redundant information. We analyze and quantify the interrelation between the variables using the *extra sum of squares* technique which is a variance decomposition of differently configured linear multiple regression models. We will present the variance decomposition results in a form that permits the quantification of the *relevance* as well as the *degree of redundancy* of the individual predictors, i.e. local paper properties p_1, p_2, \dots, p_5 .

2 Materials and Methods

We analyzed 17 commercial SC-A papers from three different European paper producers. The papers have been printed in a full scale industrial gravure printing press. The print quality has been assessed in panel tests under controlled light and viewing distance. For modelling we chose two papers with very good print quality (named G1 and G2) and two with very bad print quality (named B1 and B2). For the models black printed fields with tone values of 20% and 70%, figure 1, have been analyzed.

2.1 Registering Maps of Local Paper Properties with Local Print Reflectance

A window of adhesive tape was placed on the print, the edges of the window are visible in all local paper property maps, figure 2. Print reflectance was measured with a desktop scanner and a local grammage map was captured using beta radiography [8]. Then the printing ink was removed using a nonpolar solvent to prevent fiber swelling and thus

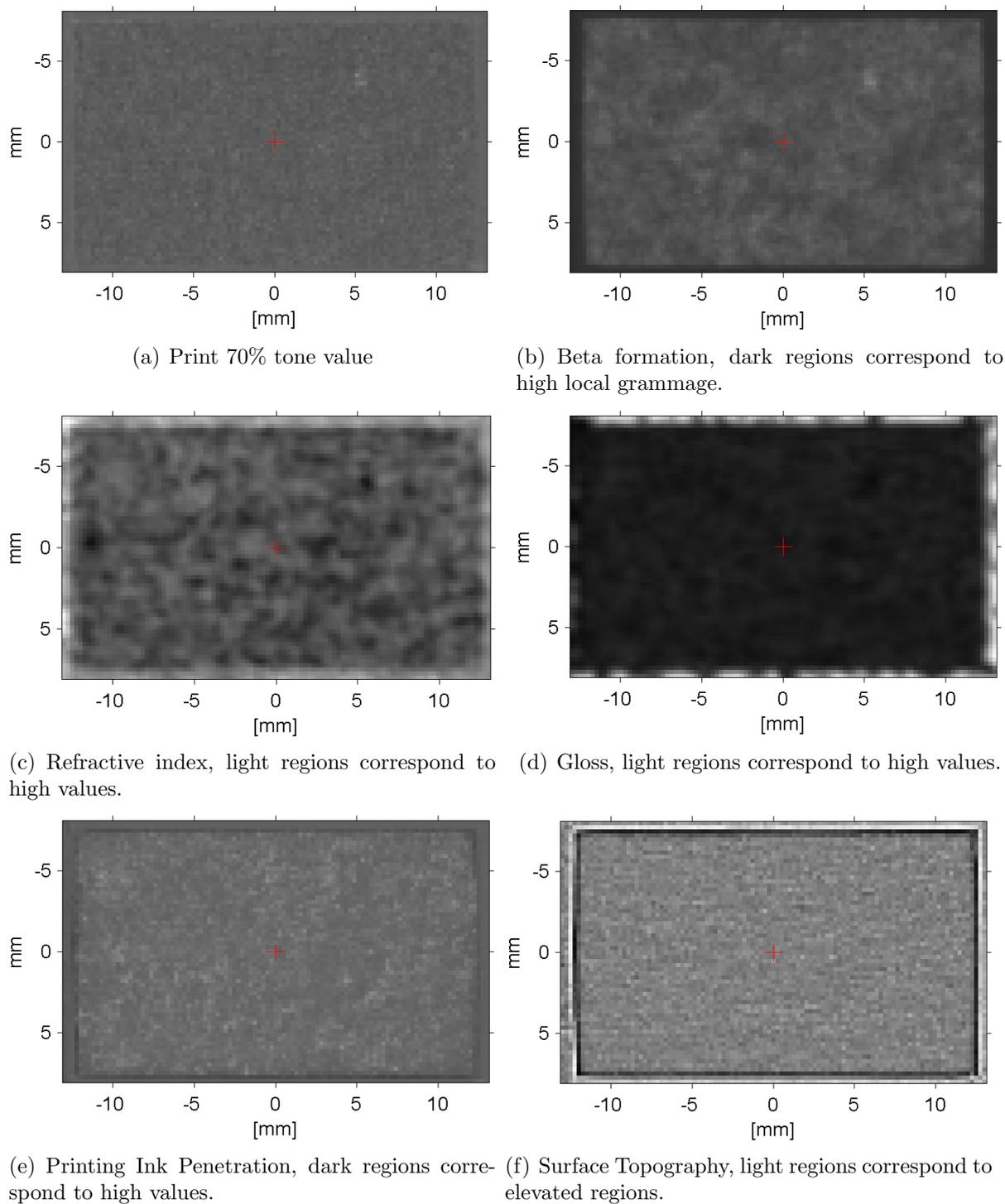


Figure 2. Registered maps of local print reflectance and local paper property maps as they have been used for the models in this paper. The resolution in these images is $250\mu\text{m}/\text{pixel}$ as it was used for the models. Please note that data acquisition and image registration has been performed at considerably higher resolution, the data has been rescaled to the final modeling resolution.

changes in the paper structure. Subsequently 2D maps of local surface topography [11], ink penetration [3], refractive index and gloss [2] were measured. These maps have all been measured with the highest possible resolution of the according measurement technique, pixel size was between $20\mu\text{m}/\text{pixel}$ and $100\mu\text{m}/\text{pixel}$. The data has been registered [5] using the edges of the tape window as markers, during registration they were rescaled to a pixel size of $250\mu\text{m}/\text{pixel}$.

A resulting dataset is shown in figure 2. The print and paper property maps have a congruent coordinate system, i.e. a given point coordinate x, y in all the maps refers to exactly the same position on the paper sample. Thus it is able to extract the local print reflectance together with the local paper properties for each point in the window. Performing this for a range of $x = \pm 11\text{mm}$ and $y = \pm 7\text{mm}$ leads to a sample region of 3.1cm^2 consisting of 5000 data points with local print reflectance and local paper properties. Each datapoint in this dataset represents an area of $250 \times 250\mu\text{m}^2$ - one pixel in figure 2 - with its print reflectance and paper property values.

2.2 Linear Multiple Regression Modeling

In order to identify interrelations between local print reflectance (local print density) and local paper properties such datasets can be used to model local print reflectance d as a function $d = f(p_1, p_2, \dots, p_5)$ of the local paper properties p_1, p_2, \dots, p_5 . A straightforward approach [1, 6] is to employ a linear model between local print reflectance d and a local paper property p_1 using *linear least square regression*

$$d = \beta_1 p_1 + \beta_0. \quad (1)$$

This type of model fits a line with offset β_0 and slope β_1 to the data points. Usually the strength of the linear relationship between d and p_1 is quantified using the coefficient of determination r^2 . From an analysis of variance (ANOVA) point of view the definition for the coefficient of determination is

$$r^2 = \frac{SSR}{SSTO} \quad (2)$$

where SSR stands for the *sum of squares of regression* and $SSTO$ for *total sum of squares*. SSR is the variance explained by the model, equation 1, whereas $SSTO$ is the total variance in the independent variable d . The coefficient of variation r^2 thus gives the *proportion of total variance SSTO that is explained by the model*.

Having more than one predictor variable p single variable regression models raise some problems. Results are difficult to interpret if the variables are inter-correlated i.e. they contribute redundant information. The problem is to find out which variable contributes the most information, because a variable with a high r^2 might very well be redundant, i.e. its information is already contained in the other variables.

The key idea of our modeling approach is to apply *multiple linear regression*. Modeling local print reflectance by five local paper properties p_1, p_2, \dots, p_5 we employ the *combined* influence of all five variables as a linear combination

$$d = \beta_5 p_5 + \beta_4 p_4 + \beta_3 p_3 + \beta_2 p_2 + \beta_1 p_1 + \beta_0. \quad (3)$$

The definition of r^2 , equation 2, is still applicable for multiple linear regression, the

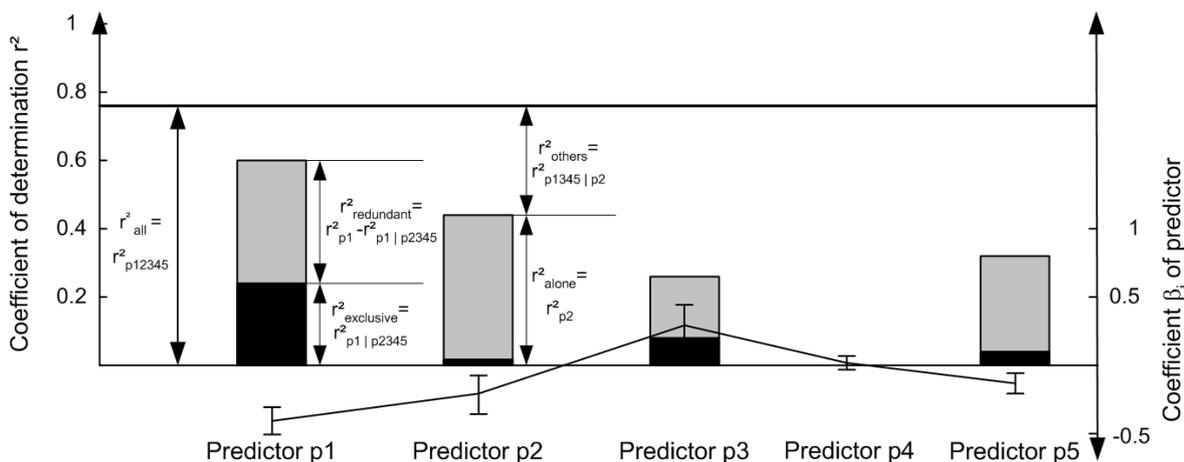


Figure 3. A sample plot of a five predictor linear multiple regression model.

only difference being that SSR is now the variance from the regression equation 3. The coefficient of determination from one predictor p_1 , model equation 1, will from now on be denoted $r_{p_1}^2$ and the combined coefficient of determination from the predictors p_1, p_2, \dots, p_5 , model equation 3, will be denoted $r_{p_12345}^2$.

A key feature of multiple linear regression is, that it can quantify the contribution of the individual variables to the model. This technique, called *extra sum of squares*, decomposes the model coefficient of determination $r_{p_12345}^2$ into so called *coefficients of partial determination* (COPD). A COPD, like a coefficient of determination, explains a part of the total variance in the data, compare equation 2. Precisely a coefficient of partial determination (COPD) gives the marginal increase of r^2 if a new variable is added to an existing model employing one or more other variables. For example consider a one predictor (p_1) model (equation 1). If we add a second predictor p_2 we can calculate the COPD and write this coefficient of partial determination $r_{p_2|p_1}^2$ speaking ‘... $r_{p_2|p_1}^2$ is the r^2 (coefficient of partial determination) of p_2 given p_1 ’. Thus the COPD $r_{p_2|p_1}^2$ is the improvement of the model which is achieved by adding predictor p_2 to the model consisting only of predictor of p_1 . The crucial thing is, that $r_{p_2|p_1}^2$ exactly quantifies the amount of new information introduced to the model, it is the non-redundant part of information added by predictor p_2 .

This concept can also be generalized for multi-variable systems, two cases are of special interest. First there is $r_{p_i|allothers}^2$, the marginal r^2 of a variable p_i given all other variables. In our case for example $r_{p_1|p_2345}^2$ gives the information that is introduced to the model by predictor p_1 if all other variables are already considered, it is thus the *exclusive* or *non-redundant* information provided by p_1 . Second there is $r_{allothers|p_i}^2$, the marginal r^2 of all other variables given variable p_i . For example $r_{p_2345|p_1}^2$ gives the new information of that *all other* variables introduce after p_1 has been considered. If this value is low, variable p_i is very important, it provides the bigger part of the information to the model.

The mathematical details of our approach are discussed in [9], the general concept of extra sum of squares is described in [12].

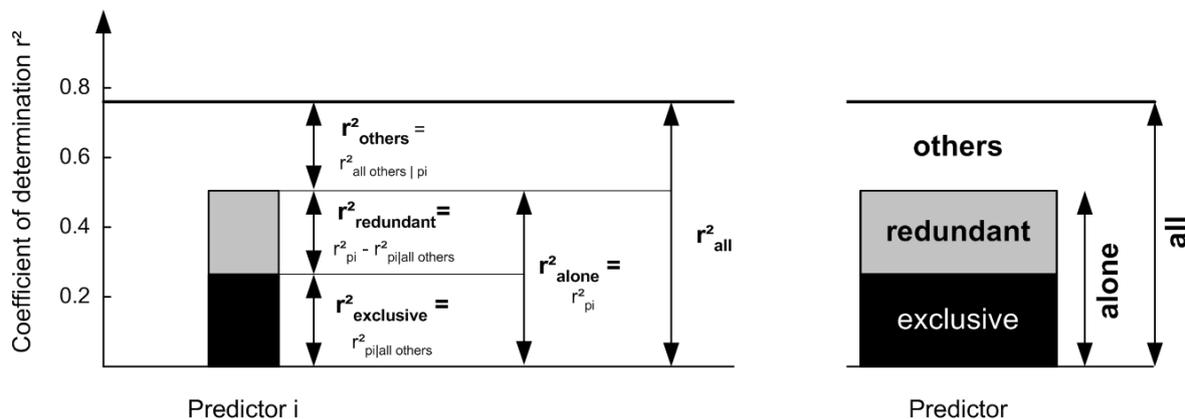


Figure 4. Visualization of predictor relevance in terms of r^2 values from extra sums of squares analysis of multiple linear regression models. For each predictor the amount of information it provides for the model is quantified, most important the *exclusive* or *non-redundant* part of the r^2 (black bar) and the *redundant* part of the r^2 (grey bar) is given. For a detailed description see sections 2.3 to 2.4.

2.3 Visualization of predictor relevance

The main benefit of the extra sum of squares approach is, that the contribution a predictor makes to the overall model can be expressed in terms of a coefficient of determination r^2 (COD) which is a widely accepted and well understood measure for model performance. We have developed a visualization technique that gives a condensed overview on the relevance of the individual model predictors, a schematic plot is shown in figure 3. Every predictor is represented by one bar in the barchart, bar height represents the coefficient of (partial) determination r^2 the predictor provides to the model. The interpretation of one bar is in illustrated in figure 4.

The line in the diagram, figure 4, gives the r^2 of a model containing all variables r^2_{all} . For each predictor bar four r^2 values are displayed: They refer to the variance fraction explained by the predictor alone r^2_{alone} , the non-redundant variance fraction $r^2_{exclusive}$ of the predictor, the redundant variance fraction $r^2_{redundant}$ and the variance fraction r^2_{others} the other variables have in addition to the predictor. The individual parts are now explained in detail.

- **Overall model** r^2_{all} . This is the coefficient of determination of the overall model using all predictor variables. In our case it is the five predictor model as given in equation 3, which models local print reflectance as a function of five local paper properties. The coefficient of determination for our overall model is denoted r^2_{12345} .
- **Predictor alone** r^2_{alone} . This is the coefficient of determination as it is often applied. It gives the variance proportion r^2_i explained by a linear one predictor model in the form of equation 1 using predictor p_i . Usually it is denoted r^2 , including the predictor name as an index the correlation for predictor 1 alone is denoted r^2_{p1} .
- **Non redundant contribution** $r^2_{exclusive}$. This is the most important score of the predictor, because it decides if the predictor provides any information to the model at all. If $r^2_{exclusive} = 0$ all the information of the variable is already contained

in other variables. Thus the variable can be dropped from the model without a reduction of overall model COD r_{all}^2 . $r_{exclusive}^2$ is defined as $r_{exclusive}^2 = r_{pi|allothers}^2$, for example the $r_{exclusive}^2$ for predictor p_2 in a five predictor model is $r_{p2|p1345}^2$.

- **Redundant contribution** $r_{redundant}^2$. This is the part of the predictor coefficient of variation r_i^2 that is also provided by other variables. It is defined as $r_{redundant}^2 = r_{alone}^2 - r_{exclusive}^2 = r_{pi}^2 - r_{pi|allothers}^2$. For example the $r_{redundant}^2$ for predictor p_2 in a five predictor model is $r_{p2}^2 - r_{p2|p1345}^2$.
- **Predictor importance** r_{others}^2 . This is the amount of information that must be provided by the other variables to reach the total model coefficient of determination r_{all}^2 . A low value of r_{others}^2 indicates that the predictor is important: Although some of its information might be redundant the predictor *alone* then provides nearly the same information as all other predictors included. It is defined $r_{others}^2 = r_{allothers|pi}^2$, for example the r_{others}^2 for predictor p_1 in a five predictor model is $r_{p2345|p1}^2$.

A schematic guideline for the interpretation of model chart gives the drawing on the right of figure 4.

Please note that only r_{alone}^2 and r_{all}^2 are coefficients of determination in a strict sense. The values for $r_{exclusive}^2$, $r_{redundant}^2$ and r_{others}^2 are technically coefficients of partial determination (COPDs), however they also express a fraction of the total variance *SSTO* as given in the generalized definition for a coefficient of determination r^2 , equation 2. Thus they are equivalent metrics and can be displayed in a joint axis.

2.4 Interpretation of the result plots

In order to interpret a result plot as it is introduced in this paper we want to give an example, figure 3. The bars must be read on the left axis, they give coefficient of determination, r^2 , values. In figure 3 the total model $r_{all}^2 = r_{p12345}^2$ is somewhat below 0.8, all predictors together explain nearly 80% of the total variance in the data. Variable p_1 is clearly the strongest predictor: Not only does it have the highest amount of information that it provides exclusively ($r_{exclusive}^2 = r_{p1|p2345}^2 \approx 0.25$) also r_{others}^2 is low - the predictor alone $r_{alone}^2 = r_{p1}^2 = 0.6$ is already close to the total model $r_{all}^2 = 0.8$. Predictor p_2 is a weak predictor. Although its $r_{alone}^2 = r_{p2}^2 = 0.45$ is considerable its non redundant contribution $r_{exclusive}^2 = r_{p2|p1345}^2$ is low. Nearly all of the information in p_2 is redundant to the other variables thus it does not provide usable information and can be dropped from the model. Predictor p_3 is not very strong, still it provides some non-redundant information $r_{exclusive}^2$, thus it does have some effect. Predictor p_4 is irrelevant as it will be explained below. Finally predictor p_5 provides a little non redundant information $r_{exclusive}^2$, it causes only minor variations in local print reflectance, if any at all.

The right axis of figure 3 gives the value of the predictor coefficients β_i in equation 3. The actual *value* of the coefficients β_i is irrelevant because it changes with rescaling of the predictor data. However the *sign* of the coefficient is crucial, because it determines the nature of the interrelation between the local paper property (predictor) and the local print reflectance. A positive sign indicates that a higher value of the predictor leads to a higher value of local print reflectance (i.e. a brighter region of the print). In figure 3 predictors p_1, p_2 and p_5 have a negative sign, high values of these local paper properties lead to locally brighter regions in the print. Locally higher values of predictor p_5 lead

to darker regions in the print. All predictor coefficients β_i are plotted with their 95% confidence intervals. If the confidence interval intersects with the x-axis, as it is the case for predictor p_4 , it is unclear if the sign of the coefficient is positive or negative. In this case we can not make a statement about the effect of the predictor - does it produce locally darker or brighter print? - thus the predictor is considered irrelevant and it is dropped from the model.

3 Results and Discussion

First we will give some results for modeling the print reflectance variations of 20% and 70% tone value black print, figure 1. In these models the influence of the local paper properties on local variations of print reflectance is quantified. Furthermore stability of the modeling results is examined regarding overall reproducibility and repeated image registration.

3.1 Models for 20% and 70% Tone Value Black Print

The key aim of the research presented in this paper is to quantify the influence of local paper property variations on print unevenness of SC Paper. For this purpose we model local print reflectance d as a linear combination of five paper properties (predictor variables) p_1, p_2, \dots, p_5 . The paper properties measured are local basis weight (formation), local ink penetration, surface topography, local refractive index and local gloss. For analysis we chose black print in two different tone values, 20% and 70% (figure 1). The reason for this was, that different paper properties might be responsible for print unevenness in different tone values. Low tone values might be more prone to variations in surface topography while high tone values might be more affected by variations in ink penetration. Result plots for the models are given in figure 5, interpretation of these plots has been described in section 2.3 and 2.4.

For the 20% tone clearly formation is the most important predictor. It provides the highest direct correlation (r_{alone}^2) as well as the highest non-redundant contribution $r_{exclusive}^2$. Furthermore formation gives an r_{alone}^2 which is nearly as high as r_{all}^2 including all other variables. Surprisingly ink penetration also seems to play a role at 20% tone value. Over all four papers there is a low but stable irredundant contribution to the model. For paper G2 the predictor ink penetration has even the same importance as formation. The three other paper properties deliver unstable and/or redundant information. Surface topography could not explain any of the print variations, the non-redundant contribution is low and the model coefficients are very unstable: the coefficient β has a negative sign for G1, a positive sign for G2 and is undefined (intersects with the x-axis) for B1 and B2. Local gloss is also unstable regarding the sign of the predictor coefficient, finally refractive index provides only fully redundant information. The total model r_{all}^2 is rather low for all samples, $r_{all}^2 \approx 0.3$. Only 30% of the print unevenness can be explained by the five local paper properties examined.

Results for the 70% tone value are somewhat different. First of all the over all model can explain a larger fraction of the variation in print reflectance, $r_{all}^2 \approx 0.6$. The most important predictor now is ink penetration. For all four papers it provides the highest r_{alone}^2 , the highest $r_{exclusive}^2$ and a low $r_{others}^2 \approx 0.1$. Still formation plays some role, however apart from paper B2 its non-redundant contribution $r_{exclusive}^2$ is rather small. In

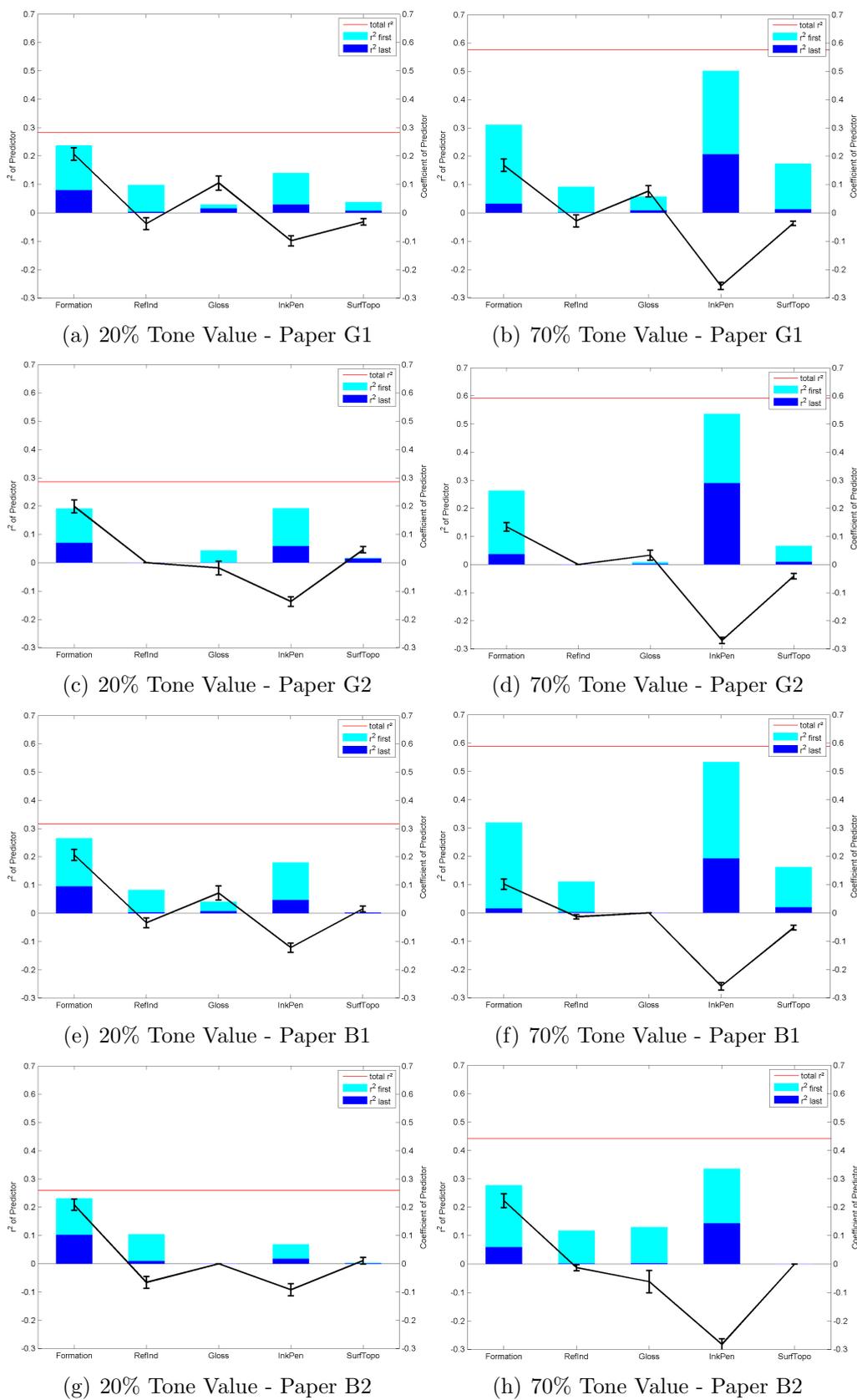


Figure 5. Models for 20% (left column a, c, e, g) and 70% tone value (right column b, d, f, h) of two papers with good print (G1, G2) and two papers with bad print (B1, B2).

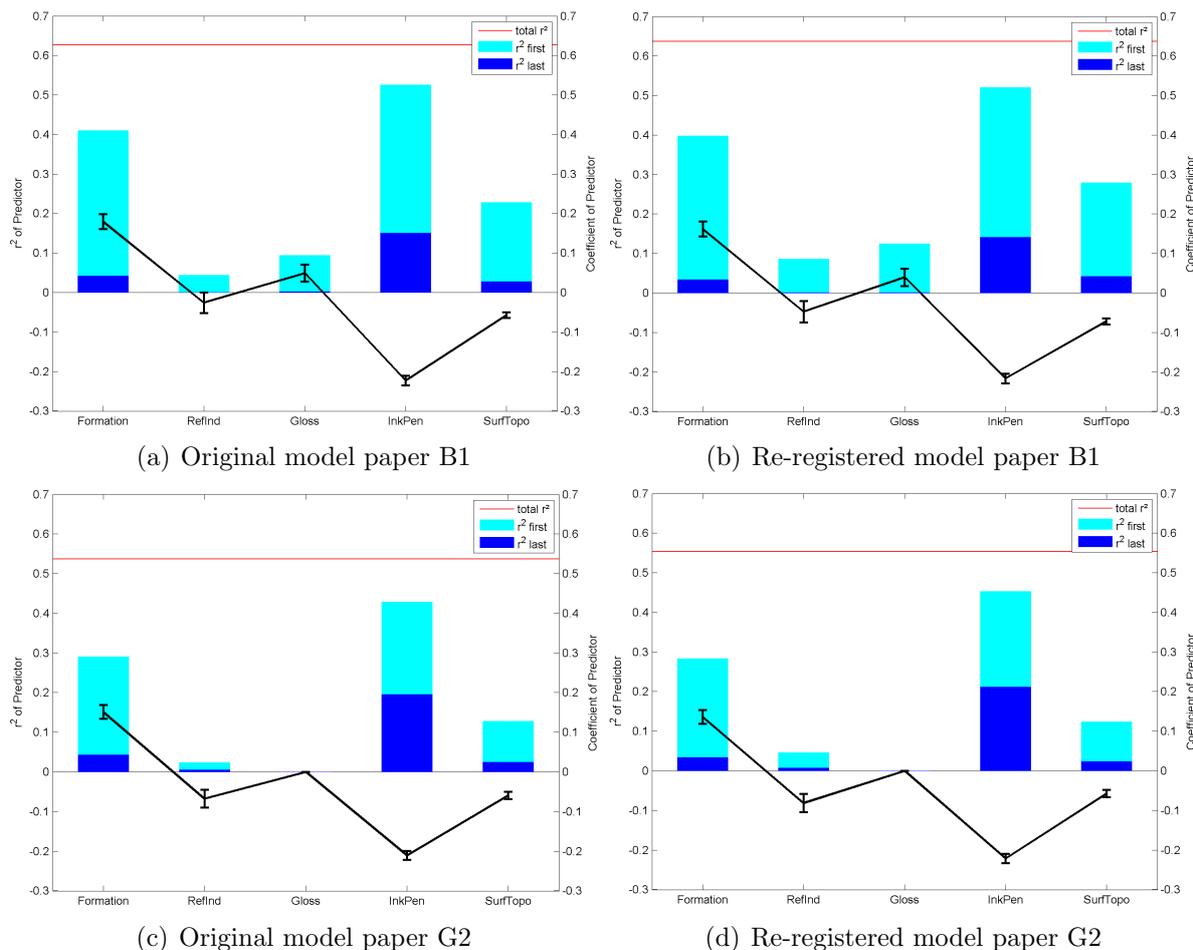


Figure 6. Influence of the registration process on the models. The 70% tone value data of the papers B1 and G2 has been re-registered. Registration does not affect the model outcome, compare (a,b) respectively (c,d).

three of the four papers surface topography has a very small but irredundant contribution, the sign of the coefficient is stable. Local refractive index and local gloss exhibit unstable coefficient signs and are completely redundant to the other measurements and can thus be ruled out as relevant influence factors.

In conclusion an impact of local grammage variations was observed for both, 20% and 70% tone value. High grammage regions tend to have darker print, this has already previously observed for offset print [1, 7]. Local ink penetration proved to be an even more valuable predictor, especially in the 70% print. As it would be expected regions with stronger printing ink penetration exhibit locally brighter print. Surface topography showed a small influence in 70% tone value, the model coefficient suggests that lower regions of the paper surface exhibit lighter print. Measurement of local refractive index and local gloss did not show any interrelation with local print reflectance.

3.2 The influence of registration on the resulting models

For registration [5] of the paper property maps - image analytical combination of the individual measurement maps to a joint dataset - the adhesive tape window placed

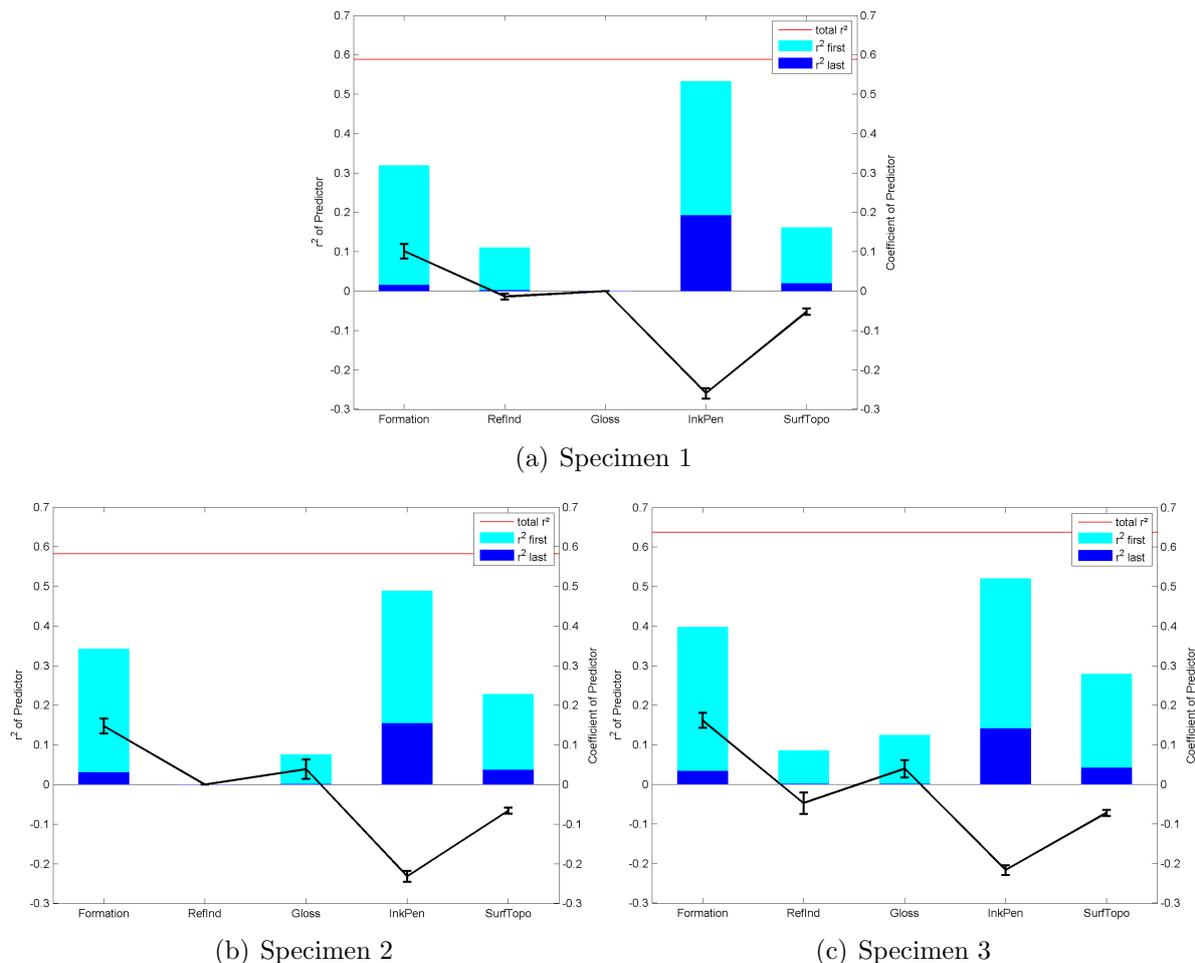


Figure 7. Repeatability of the analysis method. Three different specimen of paper B1, 70% tone value have been measured, registered and modeled. Although the results are not exactly equivalent the three relevant predictors formation, ink penetration and topography give the nearly the same results. The predictors refractive index and gloss are always redundant and - if at all - marginally significant.

on the paper sample ist used. The edges of the tape mark are visible in all images, figure 2. These edges are manually marked by the operator, the edge lines form a coordinate system which is used for image registration. In order to evaluate the effect of image registration on the results an existing dataset has been re-registered by different operators. The results for two specimen, B1 and G2 70% tone value, are shown in figure 6. The model results are nearly identical, compare paper B1 (a,b) and G2 (c,d). This demonstrates, that the image registration procedure which is used to create the datasets for modeling, does not affect the results. Image registration is sufficiently precise to produce highly repeatable results.

3.3 Model stability and reproducibility

In order to examine the stability of the analysis method reproducibility trials were carried out. Three different prints from paper B1 were analyzed. From each print a specimen from the 70% tone value field was extracted and the local paper properties as well as

the local print reflectance was measured. All three datasets were registered and multiple linear regression modelling was performed as described above. The final model plots are displayed in figure 7. The results are well reproducible for the relevant paper properties delivering non-redundant information for the model, i.e. ink penetration, formation and surface topography. For the fully redundant predictors local gloss and local refractive index the results are instable, however in all three specimen these predictors turned out to be irrelevant. Similar results have been found in further reproducibility tests.

Considering the fact that the modeled sample area of one specimen is 3.1cm^2 these results are remarkably stable.

4 Conclusions

The presented analysis technique seems to be a useful tool to examine the interrelationship between local variations in print reflectance (density) and local variations of paper properties. For gravure printed SC paper local variations of grammage and printing ink penetration were identified as major influence factors for print unevenness. Surface topography seems to play a minor role, measurements of local refractive index and local paper gloss did not reveal any interrelation with local print reflectance variations.

The presented statistical modeling approach provides a structured method to quantify *relevance* as well as *redundancy* of the examined variables. A new visualization technique gives a condensed overview on the modeling results and permits intuitive yet precise interpretation of the model performance and impact of the individual variables.

References

- [1] Dickson, A.: Evaluating the relationship between grammage, topography and print properties in newsprint. In: 60th Appita Annual Conference, Melbourne, pp. 19–23 (2006)
- [2] Elton, N., Preston, J.: Polarized light reflectometry for studies of paper coating structure part 1 + 2. *Tappi J.* **5**(7;8), 8–16;10–16 (2006)
- [3] Enzi, S., Bauer, W., Hirn, U.: Einfluss der Penetration auf das Bedruckbarkeitsergebnis im Tiefdruck. (In German). In: CTP-PTS-Symposium Papier und Bedruckbarkeit, PTS Manuskript CP 410, pp. 15.1 – 15.17 (2004)
- [4] Hansson, P., Johansson, P.A.: A new method for the simultaneous measurement of surface topography and ink distribution on prints. *Nord. Pulp Paper Res. J.* **14**(4), 315–319 (1999)
- [5] Hirn, U., Lechthaler, M., Bauer, W.: Registration and correlation of local paper property maps. *Nordic Pulp and Paper Res. J.* **23**(4), 374–381 (2008)
- [6] Kajanto, I.: The effect of formation on print quality with woodfree offset papers. *Nordic Pulp and Paper Res. J.* **4**(1), 8–15 (1989)
- [7] Kajanto, I.: Pointwise dependence between local grammage and local print density. *Paperi Puu* **73**(4), 338–345 (1991)

-
- [8] Keller, D.S., Kellomäki, M., Pawlak, J., Hägglund, J.E., N., J.: Three storage phosphor systems for beta-radiographic imaging paper. *Nordic Pulp and Paper Res. J.* **19**, 170–175 (2004)
- [9] Lechthaler, M.: Bewertung des ortsabhängigen Verformungsverhaltens von SC Papier (in German). Ph.D. thesis, Graz University of Technology (2007)
- [10] Mettänen, M., Ihalainen, H., Ritala, R.: Alignment and statistical analysis of 2-d small scale paper property maps. In: *Proceedings of the 2007 International Paper Physics Conference*, Gold Coast, Australia, pp. 624–632 (2007)
- [11] M.Wanske, H.Großmann, S.Scherer: Messtechnische Bewertung von Tissue Produkten mit dem optischen Messsystem InfiniteFocus (in German). *Wochenblatt der Papierfabrikation* **136**(9), 473–477 (2008)
- [12] Neter, J., Kutner, M., Nachtsheim, C., Wassermann, W.: *Applied Linear Statistical Models*, 4th edn. McGraw Hill (1996)
- [13] Praast, H., Göttching, L.: Analyse und Ursachen von Missing Dots im Tiefdruck (in German). *Papier* **47**(1), 12–19 (1993)
- [14] Vyörykä, J., Juvonen, K., Bousfield, D., Vuorinen, T.: Raman microscopy in lateral mapping of chemical and physical composition of paper coating. *Tappi J.* **3**(9), 19–24 (2004)