Revue de Métallurgie **109**, 81–91 (2012)
© EDP Sciences, 2012
DOI: 10.1051/metal/2012002
www.revue-metallurgie.org

Revue de Métallurgie

Lubricant homogeneity of industrial rough metallic substrates: a multivariate statistical analysis of spectroscopic ellipsometry data

L. Abdessemed and M. Voué

Physique des Matériaux & Optique, Université de Mons, 20 Place du Parc, 7000 Mons, Belgium
e-mail: michel.voue@umons.ac.be

Key words:
Spectroscopic ellipsometry; outliers detection; multivariate analysis; lubricant; tinplated steel (TPS)

Received 3 May 2011
Accepted 18 January 2012

Abstract – In this paper, a non-destructive method is proposed for measuring the density of a very thin lubricant layer (weight and spatial) on an industrial surface. We considered spectroscopic ellipsometry measurements on rough tinplated steel substrates protected by a lubrication layer. The thickness of the coating was less than the roughness parameter characterizing the metallic surface. As the optical properties of the substrates could not be modelled in a conventional way due to the roughness and the complex structure of the metal, the variations of one of the ellipsometric angle (Δ) were evaluated as a function of the lubricant film surface density. After identification of the potential outliers using a multivariate analysis technique based on the Mahalanobis distance, we interpreted the data using the Drude's approximation for thin dielectric films. The values of Δ linearly decrease with the lubricant surface density, allowing us to evaluate locally the lubricant surface density and its point-to-point variations.

Tinplate steels (TPS) combine the physical strength and relatively low price of steel with the corrosion resistance of tin. They are manufactured by electroplating a tin layer over a mild steel substrate having low carbon content (<1%). Tin-plated substrates are protected against corrosion by electro-spraying of a very thin layer of lubricant (about 4 to 10 nm-thick) [1]. The structure of the material is rather complex (Fig. 1a): oxide layers are present at a low level at the steel-tin interface due to selective oxidation during the steel annealing process prior to tinning and an iron-tin alloy is expected to be formed during the tin melting process due to an interdiffusion mechanism between iron and tin. This operation is required for reducing the intrinsic steel roughness. As a consequence, iron can be detected at the surface of TPS. After being cooled, the TPS is protected by a passivation treatment and an oiling process, resulting in a temporary protection and lubrication of the surface.

Over-lubrication of the surface due to the impact of larger oil drops creates sometimes oil spots that do not level out with time and can induce disturbances in the further surface processing such as the dewetting of the varnish layer used at the inner surface of cans used in food industry. Identifying such an excess of lubricant remains a challenge: as a matter of a fact, standard optical techniques are not really adapted because the thickness of the lubricant layer is smaller than the roughness of the TPS substrate. Moreover, in industry, only the total amount of lubricant per m^2 can be obtained by burning the oil in oxygen and measuring the amount of CO_2 produced during the combustion. This latter method is of course destructive and there is need for non-destructive alternatives.

In this paper, we consider the analysis of spectroscopic ellipsometry (SE) data recorded between 1.5–2.5 eV to identify the presence of an excess of lubricant on industrial tin plated steel surfaces. Assuming multidimensional distribution of the ellipsometric parameters, we use robust estimation of the principal components to identify the multivariate outliers and the lubricant spots as explained hereafter.

Although well-defined for univariate data, the problem of identifying multivariate outliers is rather complex especially when

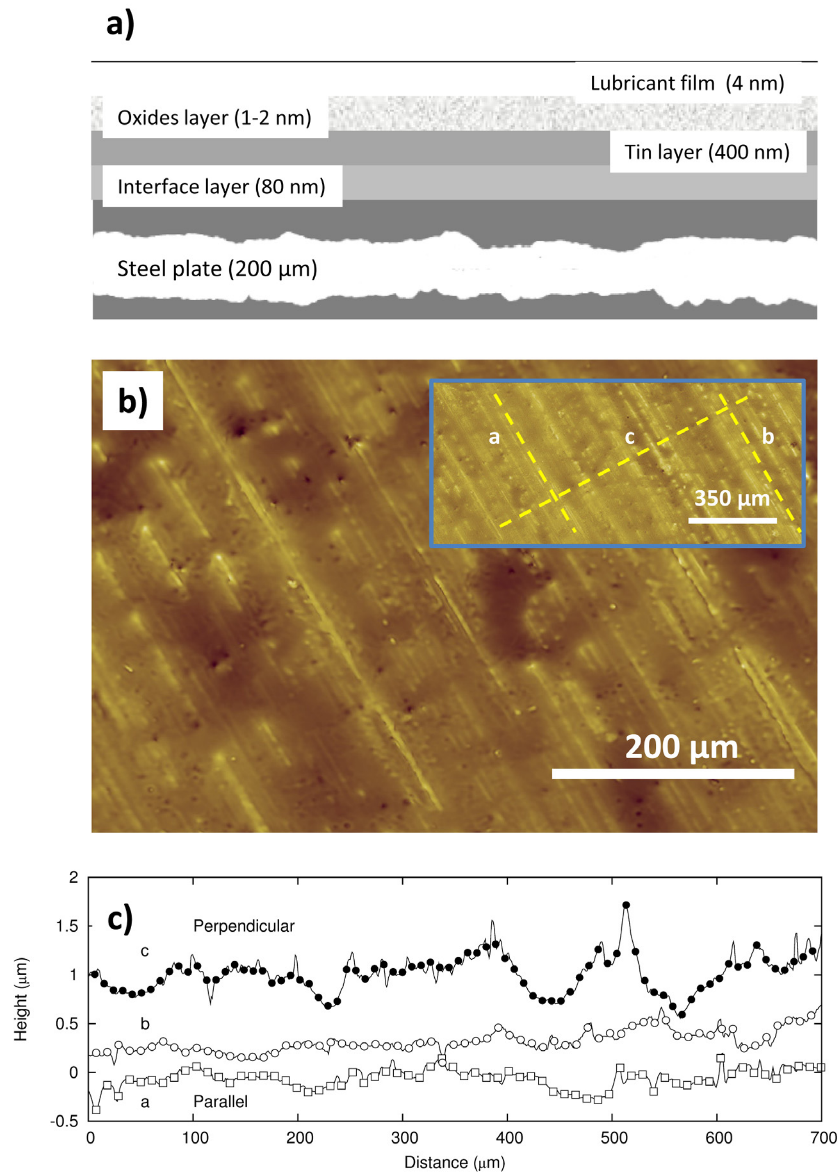

Fig. 1. (a) Schematic view of the cross section of a TPS sample (redrawn from [1], thickness of the layers not to scale); (b) topography of the TPS surface determined by optical profilometry (area: 600 μm × 450 μm, magnification 10.4 x, VSI mode) (Inset: global view of the surface and location of the profiles); (c) thickness profiles along the yellow dashed lines represented in the inset of Figure 1b.

both the number of data points and the number of variables are high. In the univariate case, many methods have been proposed and most of them are based on a (robust) estimation of the location and of the scatter of the data (see e.g. [2, 3], and the references therein). Nevertheless, the difficulty of identifying data with abnormal behaviour from the extremes of a regular distribution remains.

In multivariate data, the key feature is the Mahalanobis distance ($D_{M,i}$) defined for each data point x_i ($i = 1, n$) by

$$D_{M,i} = \left[(x_i - \mu)^T C^{-1} (x_i - \mu)\right]^{\frac{1}{2}} \quad (1)$$

where μ is the multivariate estimated location and C the estimated covariance matrix [4]. If the covariance matrix is the identity matrix, the Mahalanobis distance reduces to the Euclidian distance while in the

case of a diagonal covariance matrix, the $D_{M,i}$ is the normalized Euclidian distance to the location

$$D_{M,i}^{\text{diag}} = \left[\sum_i \frac{(x_i - \mu)^2}{\sigma_i^2}\right]^{\frac{1}{2}} \quad (2)$$

where σ_i is the standard deviation of the data.

The standard method for outlier detection in a multivariate data set is usually a two-steps procedure: (i) robust estimation of the Mahalanobis parameters μ and C and (ii) comparison of the D_M distribution with a χ_p^2 distribution with p degrees of freedom if p is the number of variables [5]. As introduced by Garret [6], a break in the tails of the distribution is an indication for outliers. In 2005, Filzmoser and co-workers proposed an adaptive outlier detection method in which the critical values for distinguishing between outliers and extremes values of the distribution was derived from simulation data [7].

Spectroscopic ellipsometry (SE) is a non-destructive optical analysis method which aims in the determination of the thickness of films and of their optical properties given by the complex refractive index [8]. In SE, a beam of polarized light is reflected by a planar surface or a planar layered system and the change of polarization is measured. More precisely, one measures the ellipticity ρ defined as

$$\rho = \frac{r_p}{r_s} = \tan\Psi e^{i\Delta} \quad (3)$$

where r_p and r_s are the complex reflectance coefficients of the $p-$ and the $s-$components of the electrical field associated to the light wave. Ψ and Δ are the ellipsometric angles defined as

$$\tan\Psi = \frac{|r_p|}{|r_s|} \quad \text{and} \quad \Delta = \delta_p - \delta_s \quad (4)$$

δ_p and δ_s being the phase difference of both components. The ellipticity is a function of the incidence angle, of the optical properties of the constitutive layers, of their thickness and of the wavelength of the incident radiation. The ellipsometric spectra are usually processed according to an optical model assuming the existence of flat and sharp interfaces between the layers. Although microscopic roughness (i.e. roughness less than the wavelength of the incident light) can be taken into account using effective media approximations [9], substrates used for SE analysis (e.g. silicon wafers, glass slides, electropolished steel plates) are usually of optical quality and can therefore be considered as ideally flat. On such surfaces, the sensitivity of SE measurements is of the order of several angstroems. As it will be shown hereafter, this is obviously not the case in the analysis of the TPS optical response and a statistical post processing of the data is required.

1 Experimental procedures

1.1 Materials

Disks of 5 cm in diameter were cut from TPS sheets. They were degreased by sonication in methyl ethyl ketone (MEK, Sigma-Aldrich, ≥99.0%) during 10 min and dried after rinsing in MEK under a nitrogen flow. To prepare samples with known lubricant weight, the degreased disks were dipped and withdrawn from a lubricant solution in MEK at controlled speed using a NIMA dip coater to achieve the required lubricant weights: 4, 9 and 13 mg/m^2. After coating the solvent was let to evaporate for 10 min at room temperature prior to SE analysis.

1.2 Optical profilometry

Topography of the degreased samples has been controlled by optical profilometry using a Wyko NT1100 optical profilometer in VSI mode (Magnification: 10.4x). The samples were characterized by their average roughness (R_a), their root-mean squared roughness (R_q) and their average peak-to-valley roughness (R_z) whose definitions are given hereafter

$$R_a = \frac{1}{N}\sum_i |z_i - \bar{z}| \quad (5)$$

$$R_q = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i - \bar{z})^2} \quad (6)$$

$$R_z = \frac{1}{10}\left[\sum_{i=1}^{10} H_i - \sum_{i=1}^{10} L_i\right] \quad (7)$$

where N is the number of pixels in the image, z_i is the height of the point i and $\bar{z}$ the average height of the surface. The H_i and L_i are the 10 highest and lowest points of the surface. An 11×11 region is excluded around each H or L point to avoid all peak or valley points emanating from one spike or hole. Unless otherwise stated, the imaged surfaces were 450 μm large and 600 μm long. The data were processed with the Vision 4.10 software (Veeco Instruments Inc., USA). Besides the global surface parameters, the line roughness parameters were determined for two orientations: parallel and perpendicular to the laminating rolls, as conventionally measured in metallurgy.

1.3 Spectroscopic ellipsometry

The ellipsometric spectra were recorded using a SOPRA GESP5 spectroscopic ellipsometer between 1.49 eV and 2.51 eV, with a fixed analyzer and a rotating polarizer (10 Hz). The ellipsometer was operated at an angle of incidence of 70 degrees. For each sample, 1941 data points were sampled on a rectangular grid superseded to the circular shape of the sample. The grid resolution was 1 mm in both X and Y directions. The light beam was focused on the sample surface to achieve a lateral resolution better than 1 mm and to ensure no overlap of consecutive spots in the direction perpendicular to the incidence plane. The light intensity was recorded using a 512 channels intensified photodiode array detector (IPDA). This method (i.e. scanning the samples over a grid and measuring at each place the ellipsometric spectra) is not suitable to in-line detection because of the time required to record the data (about 60 min). Nevertheless, the method for outlier detection combined to ellipsometry analysis is not related to the acquisition method and, as the technology evolves towards imaging ellipsometry, the acquisition will be drastically reduced to several seconds and could be applied to in-line detection in a near future.

1.4 Statistical analysis

The statistical analysis of the data was carried out using the R statistical software (version 2.12.1) [10]. The *robustbase* [11] and *mvoutlier* [12] packages were used for the identification of the outliers and the *akima* [13] package to simultaneously visualize the outlier spatial distribution and the ellipsometric data.

Table 1. Profile roughness parameters (in nm) of the TPS samples as measured by optical profilometry.

Profile [(*)]	Roughness parameters (nm)		
	Ra	*Rq*	*Rt*
a	74	96	580
b	88	113	640
c	148	176	1166

[(*)] Lines corresponding to the labels are given in Figure 1b.

2 Results and discussion

2.1 Surface topography

A typical image of the degreased TPS surface as seen by optical profilometry is given in Figure 1b. The marks left by the laminating rolls are clearly apparent and induce a preferential orientation for the optical analysis. The height profiles were measured along the dashed lines given in the inset of Figure 1c. For clarity, the profiles are shifted vertically by 0.5 μm from each other. Profiles labelled "*a*" and "*b*" correspond to profiles measured parallel to the laminating rolls marks. Profile labelled "*c*" is a profile measured perpendicularly to the marks. The roughness is obviously different between the "*a*" and "*b*" profiles and the "*c*" one (Tab. 1): R_a and R_t are about twice larger in the perpendicular direction than in the parallel one. In order to characterize the roughness parameters by global values, the surface roughness parameters were also measured. Their values are given in Table 2 for 6 different samples. R_a is about 153 ± 24 nm and R_q about 193 ± 28 nm. As expected, the value of R_z is much higher (1856 ± 211 nm) and close to the total roughness of the surface R_t (2341 ± 417 nm), showing the importance of the peaks and valleys left by the rolls on the surface (Tab. 2).

Going back to the electro-spraying of the lubricant in the industrial process, it becomes obvious that an explicit treatment of

Table 2. Surface roughness parameters.

	Roughness parameters (nm)			
	Ra	*Rq*	*Rt*	*Rz*
Sample 01	123	161	1981	1622
Sample 02	144	188	2146	1834
Sample 03	178	221	2499	2002
Sample 04	121	153	1803	1527
Sample 05	178	221	2499	2002
Sample 06	159	201	3071	1909
Mean	153	193	2341	1856
Stand. Dev.	24	28	417	211

the SE data is not possible due to the complex structure of the samples (Fig. 1a) and to its roughness, even when the sample is positioned in such a way that the lamination marks are parallel to the incidence plane. This experimental configuration is indeed the one for which the effect of the surface roughness has the least effect.

2.2 Spectroscopic ellipsometric analysis

2.2.1 Influence of the roughness and of the lubricant film thickness on the raw ellipsometric data

It is well known that the roughness of the sample considerably influences the results of SE analysis: for macroscopic roughness, only a part of the probed surface contributes to the ellipsometric signal due to the local character of the laws of reflection. On the other hand, for microscopic roughness (i.e. when the roughness parameters are of the order of the wavelength of the light), reflected light contains information relevant to the surface topography. This phenomenon is usually accounted for by considering the rough surface as a flat one with some porosity or void inclusions. This consideration is at the basis of the effective medium approximations (EMA) [9]. The effect of roughness is therefore more important in the UV part of the spectrum than in the IR one. For that reason, the first part of the presentation of our results will be focused on the SE data recorded at 1.49 eV, i.e. at the limit between the visible and the near-infrared ranges

Figure 2 presents the spatial distribution of the Cos Δ values for the degreased surface (a) and for three different lubricant weight (b to d): 4, 9 and 13 mg/m^2. Except for the sample c that will be discussed later, the data are relatively homogeneous. The Cos Δ values increase with the weight of the lubricant layer but stay negative, showing that the relative phase shift between the $p-$ and the $s-$components is greater than 90 degrees. From these spatial distributions, average of the Cos Δ for each wavelength in the 1.5–2.5 eV energy range can be calculated (Fig. 3a), as well as their standard deviations (Fig. 3b). At a given lubricant weight, the Cos Δ values increase with energy and at a given energy, they also increase with the lubricant weight, i.e. with the thickness of the lubrication layer, as already evidenced by Figure 2. More interesting is the fact that the scattering of the data, measured by their standard deviations, also increases with energy and film thickness. The increase is about 0.01 unit of Cos Δ over the probed energy range and is relatively more important for the thinner films or in the absence of lubricant. It should be kept in mind that two mechanisms are possibly the cause of the statistical scattering of the data: the roughness of the sample and the inhomogeneity of the lubricant film.

2.2.2 Influence of the outliers on the optical response

For each sample, the data matrix consisted in 1941 spectra of 2×103 data points. Prior to the identification of the outliers, it was important to evaluate the spectral range over which the identification could be carried out. Starting from the data represented in Figure 3b, the limit between multivariate data to be considered and the ones to be kept aside was not clear. To that purpose, we ran the outlier detection algorithm over increasing spectral ranges and monitored the number of outliers (N_{outliers}). We expected that number to progressively increase when extending the spectral range to the data recorded in the near-UV. The number of outliers is represented in a typical case in Figure 4. Increasing the data range from 1.5 eV to 2.16 eV only leads to fluctuations of the number of outliers around 230, i.e. about 12% of the data. But increasing the energy of the photons above 2.16 eV led to a step in the variations

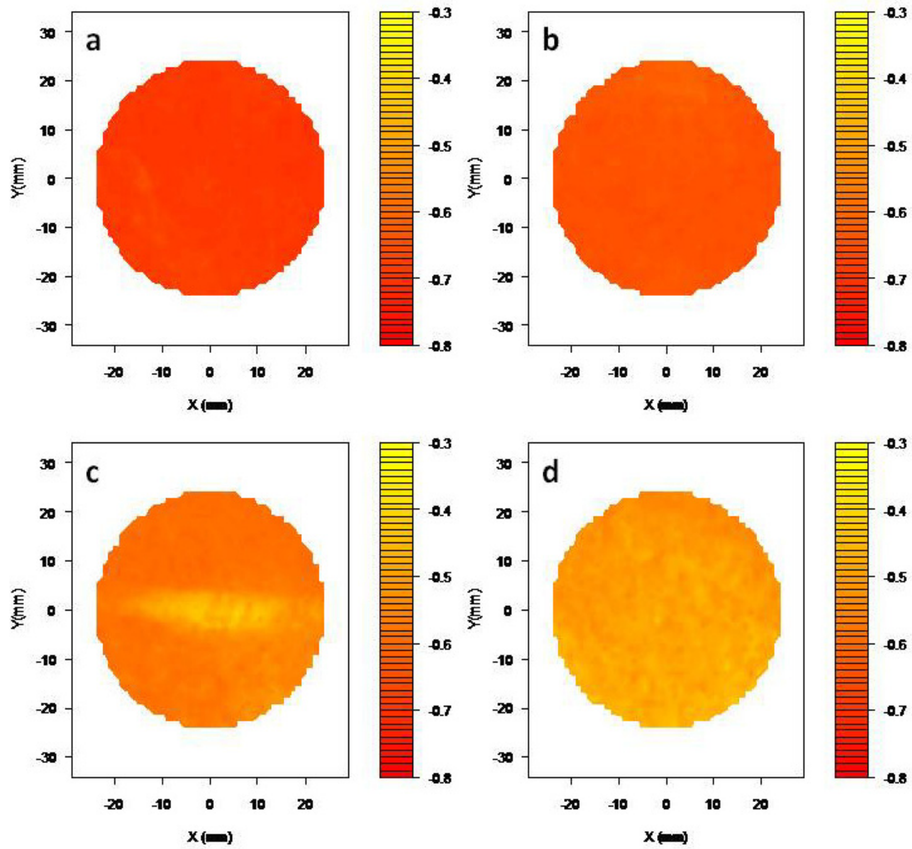

Fig. 2. Maps of the Cos Δ values at 1.49 eV as a function of the lubricant weight (a: 0 mg/m^2; b: 4 mg/m^2; c: 9 mg/m^2; d: 13 mg/m^2).

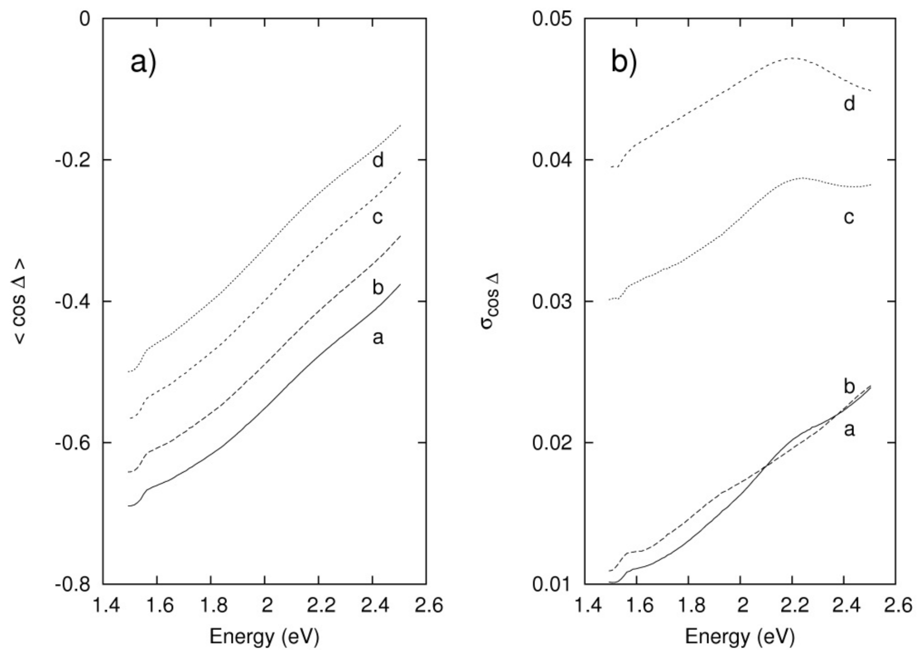

Fig. 3. (a) Mean of the Cos Δ values as a function of the wavelength; (b) standard deviation of the Cos Δ values (a : 0 mg/m^2; b: 4 mg/m^2; c: 9 mg/m^2; d: 13 mg/m^2).

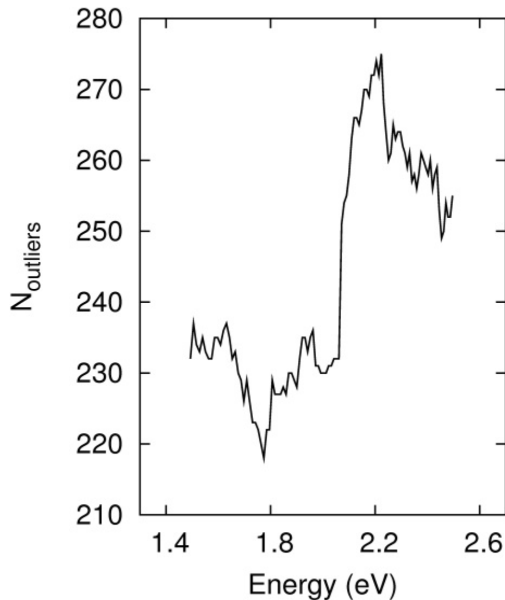

Fig. 4. Dependence of the number of outliers on the spectral domain for the degreased case.

of N_{outliers}. This corresponds in only considering 60 data over the 103 recorded for each spectrum. This value was considered as an empirical threshold for the rest of our study.

Samples were prepared in duplicate for the following lubricant density: 0, 4, 9 and 13 mg/m^2 and their optical response recorded at 70 degrees of incidence at 1941 locations over the 1.5–2.5 eV spectral range (103 energies). The potential outliers were identified using the method proposed by Filzmoser et al. [7] and histograms of the optical responses were build, differentiating the potential outliers from the regular data. An example of results is presented in Figure 5.

This sample was prepared in a slightly different way from the others: during the dipping process, a soon as the contact line of the liquid reached the middle of the sample, the withdrawal speed was increased during a couple of seconds. As the speed increased, the thickness of the entrained liquid film also increased according to the Landau-Levich equation [14,15] and after evaporation of the solvent, the lubricant film was expected to be higher on a band than on the remaining part of the TPS disk.

Figure 5a shows the spatial distribution of the Cos Δ values at 1.49 eV and the position of the potential outliers (blue circles). The correlation between both types of information is completely satisfactory. Regular points as well as outliers could be extracted from the raw data: the last ones correspond to the more negative values of Cos Δ which span from –0.55 to –0.37 (Figs. 5b and 5c).

Let us now briefly come back to the influence of the energy on the scattering of the data and to the advantage of considering multi- instead of univariate data. Figure 6 presents the distribution of the Cos Δ values at four different energies (1.49 eV, 1.69 eV, 2.19 eV and 2.50 eV). For clarity, the values are randomly scattered in the horizontal direction. To overcome the differences between the data ranges of the different variables, the data were first centred and scaled to highlight the influence of the energy on the outlier detection. The data are shown for the degreased sample (top) and for the sample already presented in Figure 5 (bottom). The potential outliers appear in red in the graphs while the regular data appear in green. The two first energies are below the empirical threshold value discussed at the beginning of this section, while the two last are greater than the threshold value. It can be seen for both samples that although the outliers determined from the multivariate data correspond to the extremes data of the distributions at low energy, both red and green clusters of points strongly interpenetrate each other at higher energies. Considering only data in this energy range would lead to an underestimation of the number of potential outliers.

Once the outliers identified, SE data could be averaged on the regular points. The influence of the lubricant weight on the Cos Δ values at 1.49 eV is presented in Figure 7a. Those data are linearly correlated and the comparison between the linear regression results of the raw (uncorrected) and of the corrected data (plain and dashed lines, respectively) shows that the identification of the outliers contributes to a statistically significant decrease of the slope: 0.0135 ± 0.0009 for the raw data and 0.0126 ± 0.0008 for the corrected ones. Converting the corrected data to Δ values gives us also a straight line with negative slope (Fig. 7b): −0.9113 ± 0.0871. This linear

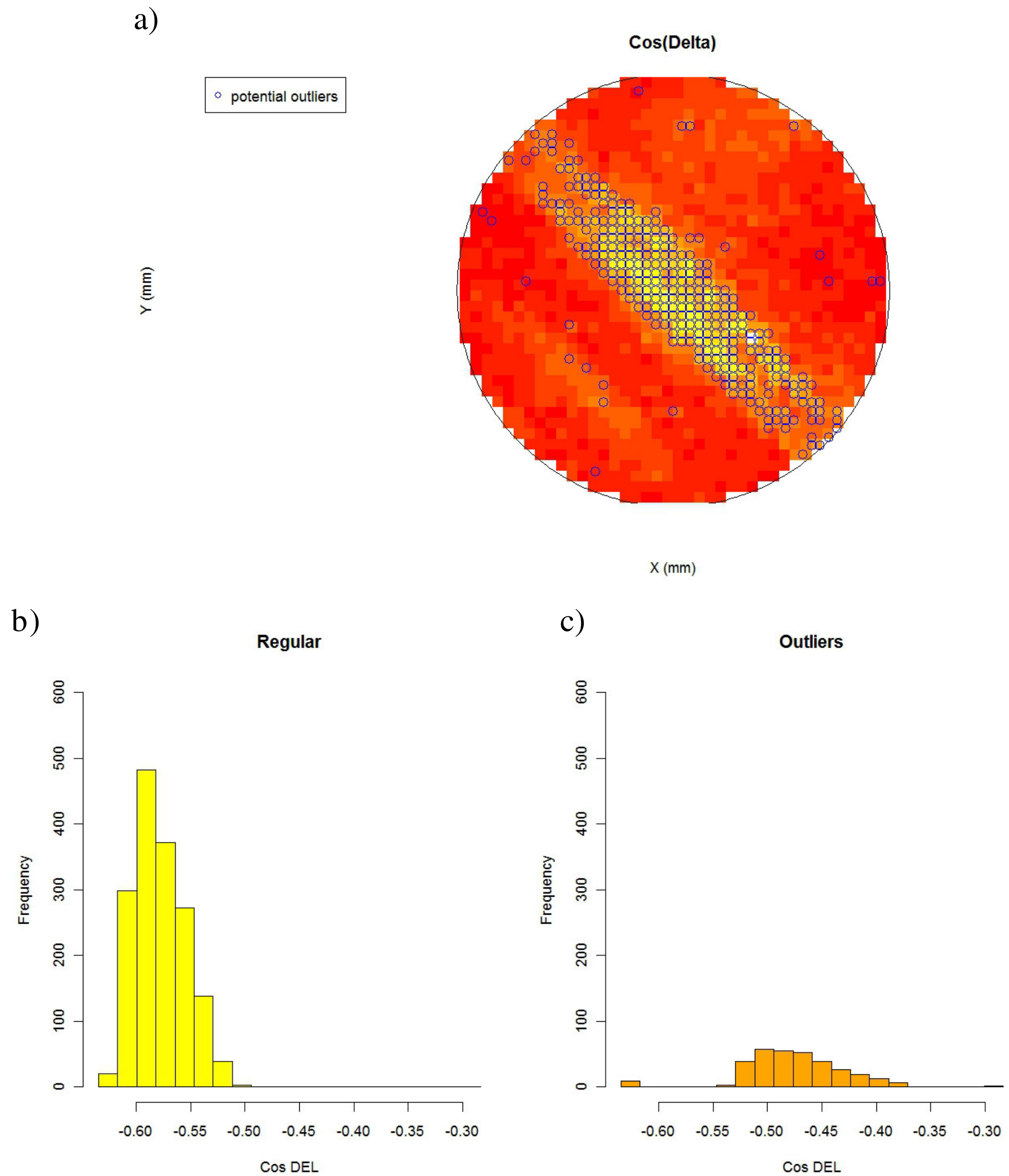

Fig. 5. Spatial repartition of the outliers and histograms of the Cos Δ values at 1.53 eV and for a lubricant weight of 4 mg/m^2.

variation can be understood in terms of the Drude's approximation for thin films [9,16]. This approximation allows the calculation of the thickness of the layer once its optical properties are known as explained hereafter. Fresnel's equations applied to optical stack with one substrate and one dielectric film were solved to the first order, yielding a linear relation between Δ and the thickness of the film d:

$$\Delta = \Delta_0 - C_X d \qquad (8)$$

where Δ and Δ_0 are the ellipsometric angles measured on the optical stack and on the film-free substrate, respectively and where C_X is a parameter which only depends on the optical properties of the film and of the substrate. The approximations of Drude and Saxena are valid for optically flat dielectric films whose thickness is less than 4 nm. The data presented in this study trend to show that the thickness range can be slightly

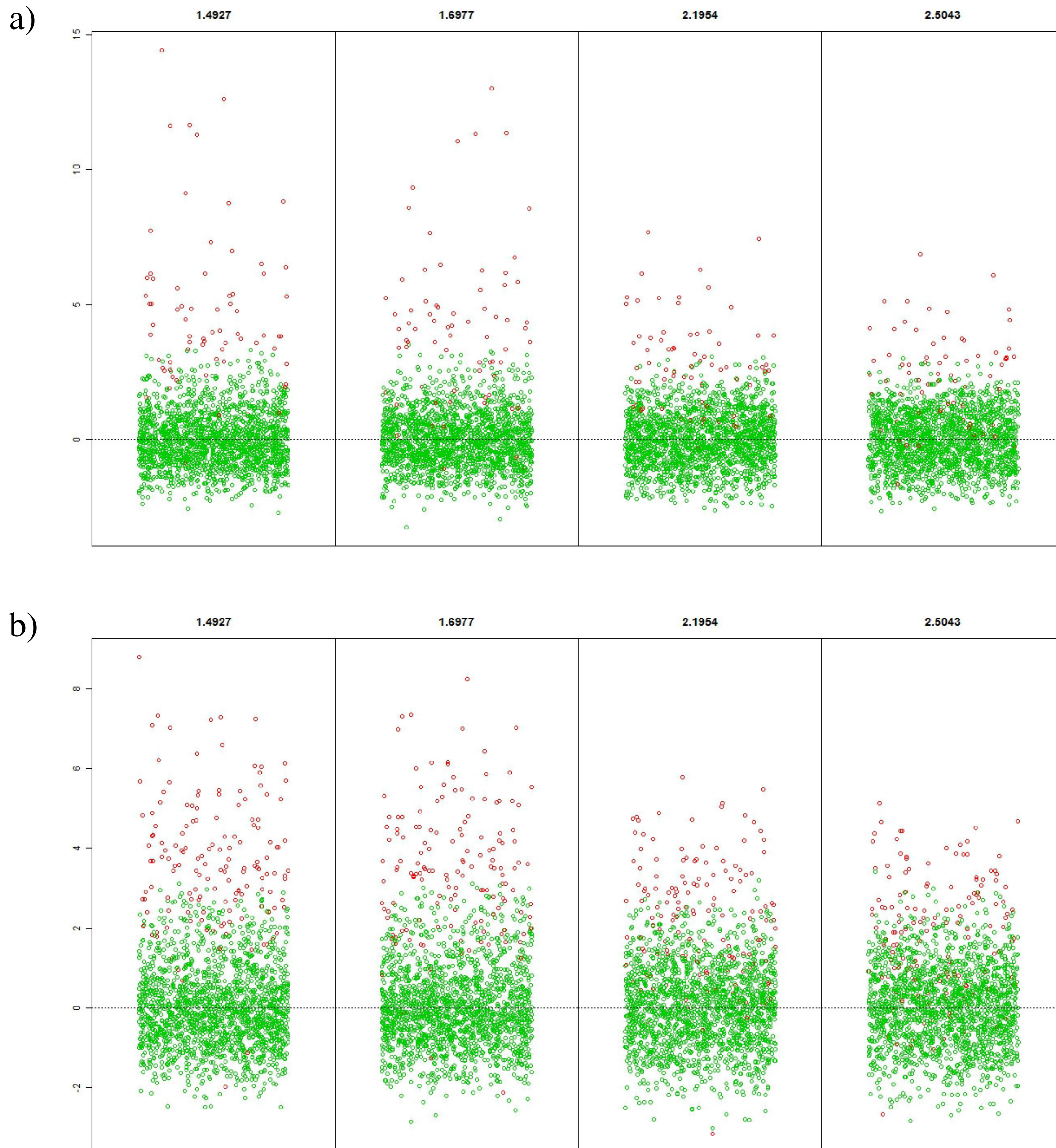

Fig. 6. Scatterplots of the Cos Δ values (after centring and scaling) as a function of the energy of the incident light (top: degreased sample; bottom: sample of Fig. 5). Green circles: regular data – Red circles: potential outliers.

extended to 10 nm and to substrates exhibiting some roughness.

3 Conclusion

In this study, we considered SE measurements on rough metallic substrates coated with a dielectric layer whose thickness is less than the roughness parameter characterizing the metallic surface. As the optical properties of the substrates could not be modelled in a conventional way due the roughness and the complex structure of the metal, the variations of one of the ellipsometric angle (Δ) were evaluated as a function of the lubricant film surface density expressed in mg/m^2. After identification of the potential

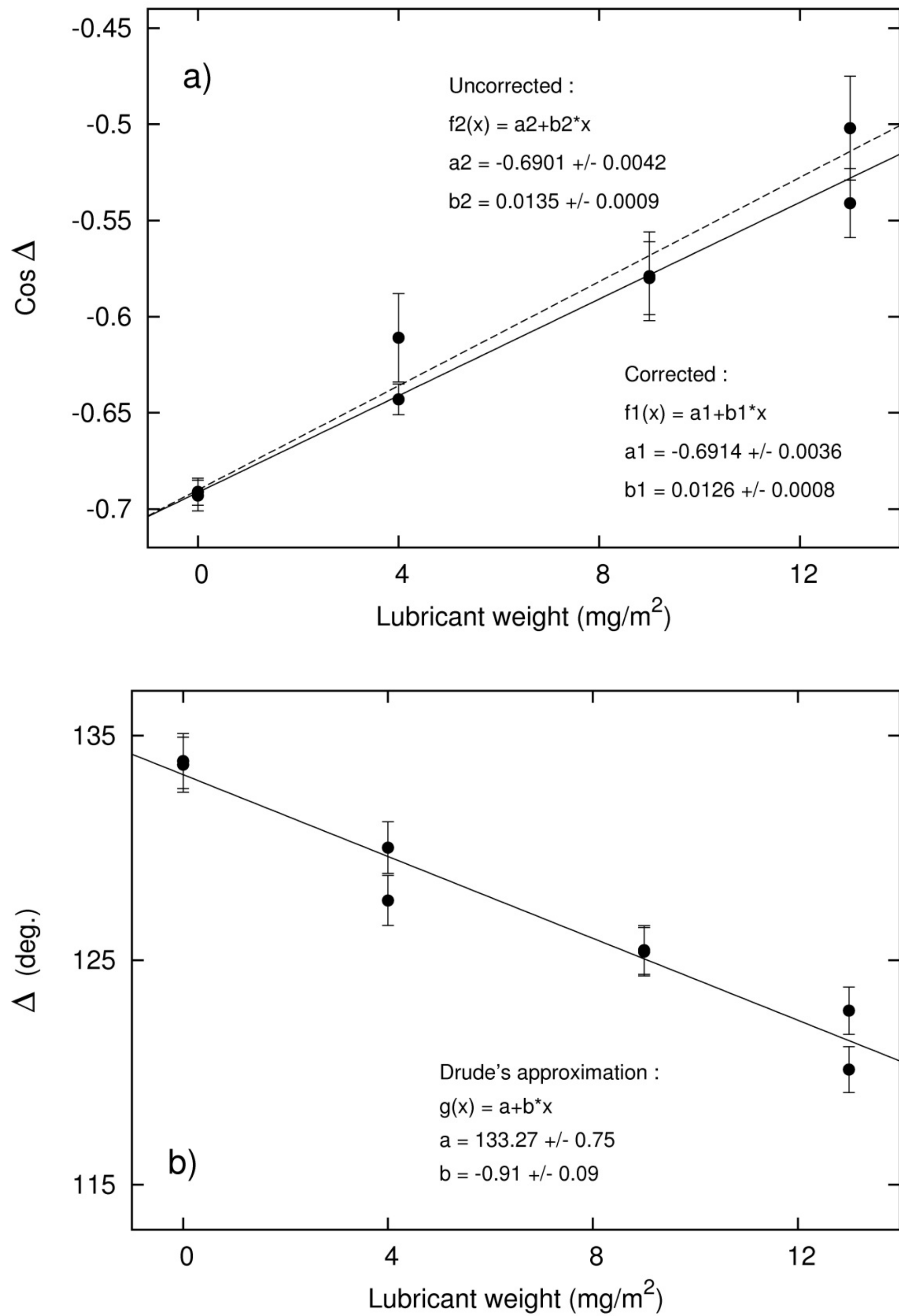

Fig. 7. (a) Weighted regression on the Cos Δ values at 1.49 eV (mean ± standard deviation). Plain line: regular values only. Dashed line: complete data set. (b) Drude's approximation of the data (Eq. (8)): linear variation of Δ values with negative slope as a function of the lubricant weight.

outliers using a multivariate analysis technique based on the Mahalanobis distance, we interpreted the data using the Drude's approximation for thin dielectric films. The values of Δ decrease with the lubricant surface density, allowing us to evaluate locally the lubricant surface density and its point-to-point variations.

References

[1] J.-L. Vignes, G. André, D. Fousse, *Bulletin de l'Union des Physiciens* **88** (1994) 627-652

[2] P.J. Rousseeuw, M. Hubert, *WIREs Data Mining Knowl Discov* **1** (2011) 73-79

[3] X. Su, C.-L. Tsai, *WIREs Data Mining Knowl Discov* **1** (2011) 261-268

[4] P.C. Mahalanobis, *Proceedings of the National Inst. Sci. India* **2** (1936) 49-55

[5] P.J. Rousseeuw, B.C. Van Zomeren, *J. Am. Stat. Assoc.* **85** (1990) 633-651

[6] R.G. Garrett, *J. Geochem. Explor.* **32** (1989) 319-341

[7] P. Filzmoser, R.G. Garrett, C. Reimann, *Computers & Geosciences* **31** (2005) 579-587

[8] H.G. Tompkins, E.A. Irene, Handbook of Ellipsometry, Springer, Heidelberg, 2005

[9] H.G. Tompkins, W.A. McGahan, Spectroscopic Ellipsometry and Reflectometry, A user's guide, John Wiley & Sons, New-York, 1999

[10] R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

[11] P. Rousseeuw, C. Croux, V. Todorov, A. Ruckstuhl, M. Salibian-Barrera, T. Verbeke, M. Maechler, robustbase: Basic Robust Statistics. R package version 0.5-0-1, 2009 http://CRAN.R-project.org/package=robustbase

[12] M. Gschwandtner, P. Filzmoser, mvoutlier: Multivariate outlier detection based on robust methods. R package version 1.5, 2010 http://CRAN.R-project.org/package=mvoutlier

[13] A. Gebhardt, T. Petzoldt, M. Maechler, akima: Interpolation of irregularly spaced data. R package version 0.5-4, 2009 http://CRAN.R-project.org/package=akima

[14] L. Landau, B. Levich, *Acta Physicochimica URSS* **17** (1942) 42-54

[15] V. Levich, Physicochemical Hydrodynamics, Prentice-Hall, Englewood Cliffs, 1962

[16] A.N. Saxena, *J. Optical Soc. Am.* **55** (1965) 1061-1067