J Infrared Milli Terahz Waves (2013) 34:646–659
DOI 10.1007/s10762-013-0015-2

Prediction of the Thickness of a Thin Paint Film by Applying a Modified Partial-Least-Squares-1 Method to Data Obtained in Terahertz Reflectometry

Tetsuo Iwata · Shuji Yoshioka · Shota Nakamura · Yasuhiro Mizutani · Takeshi Yasui

Received: 27 June 2013 / Accepted: 31 July 2013 /
Published online: 14 August 2013

Abstract We applied a multivariate analysis method to time-domain (TD) data obtained in terahertz (THz) reflectometry for predicting the thickness of a single-layered paint film deposited on a metal substrate. For prediction purposes, we built a calibration model from TD-THz waveforms obtained from films of different thicknesses but the same kind. Because each TD-THz wave is approximate by the superposition of two echo pulses (one is reflected from the air–film boundary and the other from the film–substrate boundary), a difference in thickness gives a relative peak shift in time in the two echo pulses. Then, we predicted unknown thicknesses of the paint films by using the calibration model. Although any multivariate analysis method can be used, we proposed employing a modified partial-least-squares-1 (PLS1) method because it gives a superior calibration model in principle. The prediction procedure worked well for a moderately thin film (typically, several to several tens of micrometers) rather than a thicker one.

Keywords Terahertz · Reflectometry · Partial least squares · Paint film · Thickness

1 Introduction

Recently, terahertz (THz) electromagnetic radiation, covering the frequency range of around 0.1–10 THz, has been used for determining the thickness of a thin coating. The major reason for employing THz radiation in this application is that many coating materials are transparent in the THz frequency region but not in the visible and/or the mid-infrared wavelength regions. Another reason is based on safety considerations: x-ray transmission, for example, carries the risk of causing problems for human health. Although many promising

T. Iwata (✉) · S. Yoshioka · S. Nakamura · Y. Mizutani · T. Yasui
Institute of Technology and Science, University of Tokushima, 2-1, Minami-Josanjima-cho, Tokushima, Tokushima 770-8506, Japan
e-mail: iwata@tokushima-u.ac.jp

T. Yasui
Graduate School of Engineering Science, Osaka University, 1-3, Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan

applications of THz techniques have been proposed [1–4], one of the most practical applications in industry is to determine the thicknesses of paint films deposited on metal surfaces, such as car bodies and other objects [5–10], where a simple time-of-flight measurement technique in THz reflectometry is employed.

In this field, a crucial requirement is to lower the detection limit of the thickness measurement. To achieve this, various efforts has been made, for example, reducing the width of the radiated THz pulses [11], introducing numerical deconvolution techniques in data processing [5], and using a practical parameter-fitting procedure [12]. Although such techniques work well to some extent, it is still difficult to achieve a detection limit less than several tens of μm. The reason is that the two echo pulses (one is reflected from the air–film boundary and the other from the film–substrate boundary) obtainable from THz reflectometry are temporally overlapped with each other by an amount that is too large to allow them to be resolved. This difficulty is described in detail in the next section. If the sample is sandwiched between two high-refractive-index media, the frequency dip due to Fabry–Perot oscillations could be utilized, and this might allow a simpler approach to measure the thickness, as pointed out by Koch *et al*. [13].

In the present paper, we propose a completely different approach for solving the problem: we introduce a multivariate analysis technique in THz reflectometry. First, we gather a series of THz waveforms from many paint films whose thicknesses are known in advance and distributed in a wide range. From these waveforms, we build a calibration model, which is sometimes referred as a multivariate calibration (or working) curve. Then, by measuring the THz waveforms of samples whose thicknesses are unknown, we predict their thickness values by using the calibration model (or curve). In this scheme, the thickness of the paint film is a dependent variable (or objective variable), and discretized times in the TD-THz waveform become independent variables (or explanatory variables). Although various multivariate analysis methods are available, here we propose the use of a modified partial-least-squares-1 (PLS1) method because of the superiority of the calibration model constructed. One of the aims of the present paper is to show the superiority of the modified PLS1 method over the other methods when applied to THz reflectometry data.

As mentioned above, another point we emphasize in the present paper is that the thickness of the paint film is set as the dependent variable. The thickness is correlated directly with the time shift of the peak position of the two echo pulses. Therefore, information obtainable from the abscissa axis in the TD-THz waveform is set as the dependent variable. Such a setting is in contrast with the conventional multivariate-analysis approach used in spectroscopy. For example, in spectrochemical analysis, the concentration of the sample, which is related to the height of the peak (or the depth of the dip) in the spectrum, is set as the dependent variable, and information obtainable from the ordinate axis becomes the dependent variable. The reason why the prediction procedure in the proposed scheme works well in spite of the "unusual" setting is that the amount of the peak shift is approximately (inversely) proportional to the change in height of the overlapped area of the two echo pulses. Therefore, although it depends on the signal-to-noise ratio in the measurements, the prediction procedure works well only when the two echo pulses are overlapped by a moderate amount; without overlapping, it is obvious that the proposed scheme cannot work. As a result, the prediction procedure can be carried out more successfully for a moderately thin film rather than a thicker one.

One would think that adoption of such a setting or a scheme is obvious, but so far, we have not been able to find such a scheme in the literature, especially in the field of THz reflectometry. In the present paper, we focus on the application of this modified PLS1 method and show some numerical simulations and experimental results.

2 Principle of Operation

2.1 THz Reflectometry

Figure 1 shows the principle of measuring the thickness of the paint film on the metal substrate in THz reflectometry. As shown in Fig. 1(a), the THz pulse, $E_{in}(t)$, is incident on the surface of the single-layered paint film on the substrate with an incident angle $\theta = 0^{\circ}$, where E represents the electric field, and t time. The thickness and the group refractive index of the paint film are d and n_g, respectively. Then, as shown in Fig. 1(b), two THz echo pulses, $E_{1st}(t)$ and $E_{2nd}(t)$, reflected from the air–film boundary and the film–substrate boundary are superimposed on each other with a relative time delay of $\Delta t = t_1 - t_0$ because of the discontinuity of n_g. The synthesized waveform is therefore expressed by $E_{1st}(t) + E_{2nd}(t)$. The time delay, or difference, between the two echo pulses is given by $\Delta t = 2n_g d/c$. Hence, if we know the value of n_g, we can determine the geometric thickness d by measuring Δt. When Δt is large in comparison with the pulse width of the THz waveform, we can determine d easily. However, when d becomes small, precise measurement of Δt becomes difficult because of overlapping of the two pulses. In order to solve this difficulty, we employ the multivariate analysis method. In the multivariate analysis method, all uncertain factors, such as noise, multiple reflections (even if they occur), and changes in the physical and chemical conditions of the sample, are somehow modeled as the calibration curve, although such factors might degrade the quality of the calibration model. In this sense, the prediction ability or limitation of the method is finally determined by the goodness of the calibration model. Therefore, extension of the technique to the thickness measurement of a multilayered paint film is not difficult.

2.2 Modified PLS1 Method

Multivariate analysis methods can be classified into three categories: multiple linear regression (MLR), principal component regression (PCR), and partial least-squares (PLS). Furthermore, the PLS method can be divided into two sub-methods: the PLS1 method and the PLS2 method [14]. The difference between the PLS1 and PLS2 methods is the number of dependent (objective) variables: in the PLS1 method, one dependent variable is predicted, whereas in the PLS2 method, multiple variables are predicted at a time. Because the details of multivariate analysis methods have been described elsewhere [14–19], only a brief explanation is given here. The aim of the present paper is to propose a suitable scheme and to show some experimental results.

We propose the use of a PLS method. The reason is that the PLS method is free from the collinear problem suffered by the conventional MLR method [14]. In addition, the quality of

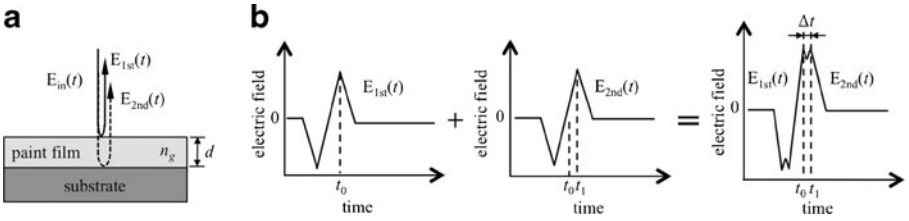

Fig. 1 (a) Principle of THz reflectometry and (b) the corresponding THz waveforms

the calibration model built by the PLS method is superior to that constructed by the PCR method. In particular, we use the PLS1 method, which targets one dependent variable, because it is suited for application to spectral datasets [15]. Furthermore, we introduce two techniques into the PLS1 method. One is called the uninformative-variable-elimination (UVE) technique [16], and the other is the uninformative-sample-elimination (USE) technique [17]. UVE is a procedure by which independent variables (discretized times) that cannot contribute to the calibration model are eliminated, and only those that can contribute are retained [18]. USE is a procedure by which calibration sample data (TD-THz waveforms themselves) that cannot contribute to the model are eliminated [19]. The word "uninformative" means "not relevant to building the calibration model". The reason for introducing these two techniques in the PLS1 method is to build a calibration model that is more robust against noise than a model that does not use these techniques [20]. We call the PLS1 method combined with the two techniques the "modified version of the PLS1 method". We tried two versions: UVE-PLS1 (a combination of UVE and the PLS1 method) and UVE-USE-PLS1 (a combination of UVE, USE, and the PLS1 method). In the UVE-PLS1 method, the UVE procedure is carried out before applying the PLS1 method. Similarly, in the UVE-USE-PLS1 method, the UVE and USE procedures are carried out before applying the PLS1 method. In the present paper, we show the results of comparing PLS1, UVE-PLS1, and UVE-USE-PLS1.

2.3 RMSEP Criterion

For the purpose of conducting a quantitative comparison of the three calibration models, we used a criterion known as the root-mean-square error of prediction (RMSEP), which is defined as [14, 15]

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{N} \left(\widehat{d_i} - d_i\right)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N} |\Delta d_i|^2}{N}}, \tag{1}$$

where N is the total number of calibration datasets (the total number of TD-THz waveforms or the total number of paint films whose thicknesses are known in advance), and d_i is the reference value of the i-th dependent variable (thickness of the i-th paint film). We measured the reference values with an eddy-current thickness meter (Coating Thickness Tester LZ-300C, Kett Electric Laboratory) and/or from microscope observations. $\widehat{d_i}$ is the i-th predicted thickness value obtained by using the calibration model built from N-1 calibration datasets, except for the i-th one. Therefore, $\Delta d_i = |\widehat{d_i} - d_i|$ gives an absolute value of the prediction error for the i-th sample. This prediction procedure is repeated N times from i = 1 to N. The RMSEP value obtained from such a "leave-one-out" procedure gives a sort of standard deviation in prediction and can be used as a measure of the goodness of the calibration model. The smaller the RMSEP value, the better the calibration model obtained.

The UVE procedure was carried out by adding random noise to the calibration dataset as independent variables. The number of noise variables was made equal to that of the original independent ones. Then, in the course of the conventional PLS1 procedure, based on the RMSEP criterion, we identified the independent variables that contributed to the model more rather than the noise ones. The USE procedure was carried out by judging the prediction error: the sample data that gave large prediction errors were eliminated in the course of the USE-PLS1 procedure [15–18].

3 Numerical Simulations

3.1 Highly-Overlapped Gaussian Waveforms

First, we have to verify whether the proposed scheme worked well. Specifically, we have to confirm whether the abscissa information could be extracted successfully from highly overlapped waveforms by using the PLS1 method. To this end, we carried out numerical simulations first by using a simple Gaussian waveform:

$$f(x) = A\exp\left\{-\frac{(x-p)^2}{2\sigma^2}\right\}, \tag{2}$$

where A is the peak height, p the peak position, and σ the standard deviation. The numerical simulation using a "slightly complicated" synthesized THz waveform is presented in the next section.

We prepared two datasets. One consisted of fifteen waveforms ($N = 15$), as shown in the upper part of Fig. 2(a), where p was varied from 730 to 870 with a step interval of 15 for fixed values $A = 10$ and $\sigma = 80$. The total number of data points was $M = 1{,}600$. Then, we superimposed uniformly distributed noise having an amplitude of 1.0 % with respect to the peak value of each waveform. The other dataset is shown in the upper part of Fig. 2(b), which is the same as Fig. 2(a) except that the peak height A was varied from $A = 10$ to 2/3 in steps of 2/3.

The results of applying the PLS1, UVE-PLS1, and USE-UVE-PLS1 methods to the dataset shown in Fig. 2(a) are summarized in Table 1(a), where $\hat{p}_i$ is the predicted value of the peak position of the i-th sample. The number of retained independent informative variables used to build each calibration model is also shown in Table 1(a) and is depicted

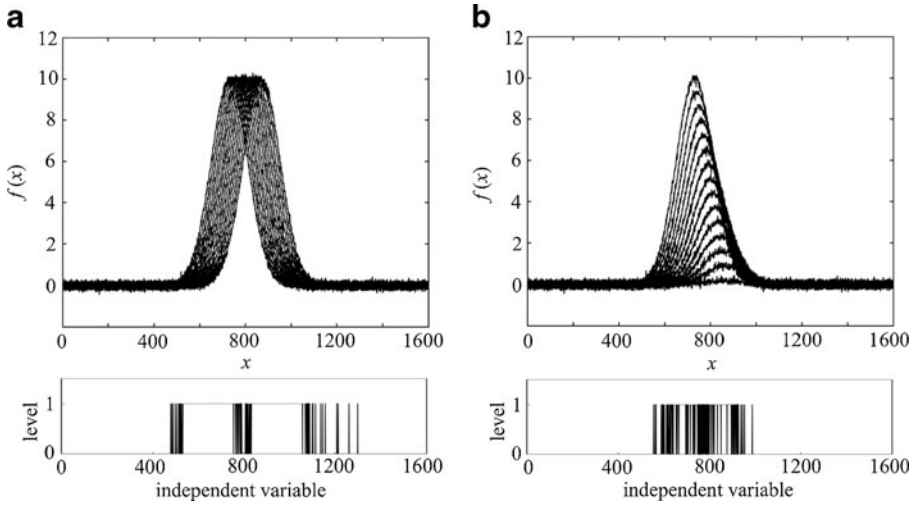

Fig. 2 (a) Calibration dataset ($N = 15$ and $M = 1{,}600$), in which each waveform was a single Gaussian waveform, where the peak position p was varied from 730 to 870 with a step of 10 for fixed values of $A = 10$ and $\sigma = 80$. Uniformly distributed noise with amplitude of 1.0 % with respect to the peak value was superimposed on each waveform. The lower part of the figure shows informative independent variables as level one and uninformative ones as level zero. (b) The same as (a), but the peak height was varied from $A = 10$ to 2/3 in steps of 2/3

Table 1 (a) Predicted values of $\widehat{p}_i$ for the i-th sample p_i, RMSEP values, and the number of retained variables for the PLS1, UVE-PLS1, and USE-UVE-PLS1 methods. The calibration dataset is shown in Fig. 2(a). (b) The same as (a) but showing those obtained from the calibration dataset shown in Fig. 2(b)

	Sample number i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	RMSEP	Number of retained variables
(a)	p_i	730	740	750	760	770	780	790	800	810	820	830	840	850	860	870		
	PLS1	737	740	748	758	767	777	788	799	812	823	833	841	852	860	864	3.05	1,600
	UVE-PLS1	723	741	750	759	770	782	793	800	810	820	831	840	850	856	876	2.54	960
	UVE-USE-PLS1	723	741	750	759	770	782	793	800	810	820	831	840	850	856	876	2.54	960
(b)	p_i	730	740	750	760	770	780	790	800	810	820	830	840	850	860	870		
	PLS1	729	738	749	760	771	781	791	802	812	822	831	840	850	858	865	1.81	1,600
	UVE-PLS1	729	738	750	760	771	781	791	800	811	821	830	841	850	859	867	1.00	213
	UVE-USE-PLS1	723	741	750	759	770	782	793	800	810	820	831	840	850	856	876	1.00	213

graphically in the lower part of Fig. 2(a). In the figure, the parts having level one are relevant for building the calibration model for the thickness extraction (informative), whereas the parts having level zero are not relevant (uninformative). The independent variables around the overlapped peak region in the waveforms were retained as informative variables. The number of retained variables with the USE-UVE-PLS1 method was the same as that with the UVE-PLS1 method. This means that no sample was eliminated in the USE-UVE-PLS1 method for this case. The prediction procedure, therefore, worked well even for the case of the UVE-PLS1 method.

The results of applying the methods to the dataset shown in Fig. 2(b) are summarized in Table 1(b). In this case, prediction errors and RMSEP values lower than those obtained in the previous case. The reason is that two kinds of information, peak height and peak position, were effectively used to build the calibration model. The number of retained independent informative variables used to build each calibration model is also shown in the lower part of Fig. 2(b). In general, the RMSEP value increased with decreasing degree of overlap of the calibration waveforms.

In summary, we found that the proposed scheme worked quite well for datasets with Gaussian waveforms.

3.2 Synthesized THz Waveforms

Next, we carried out numerical simulations by using synthesized THz waveforms. We made a calibration dataset, $E(t)$, from a reference waveform, $E_r(t)$, according to the equation $E(t)= E_r(t)+\Delta E_r(t+\Delta t)$, where $E_r(t)$ is a TD-THz pulse in free space. We used the actually measured waveform, $E_r(t)$, reflected from bare aluminum metal. To generate the synthesized THz waveforms, we set $\alpha = 1$ and varied Δt from 0.0125 to 0.5 ps in steps of 0.0125 ps. The reference values of the thicknesses, d, of the paint films were calculated from the equation $d= c\Delta t/(2n_g)$, with $n_g = 1.66$. For each synthesized waveform, we superimposed 1.0% uniformly distributed noise. Figure 3 shows the synthesized waveforms together with the waveform of $E_r(t)$, where $M = 3{,}960$. For the sake of clarity, we show only three waveforms (i = 13, 28, and 40) listed in Table 2. The synthesized waveforms were severely overlapped

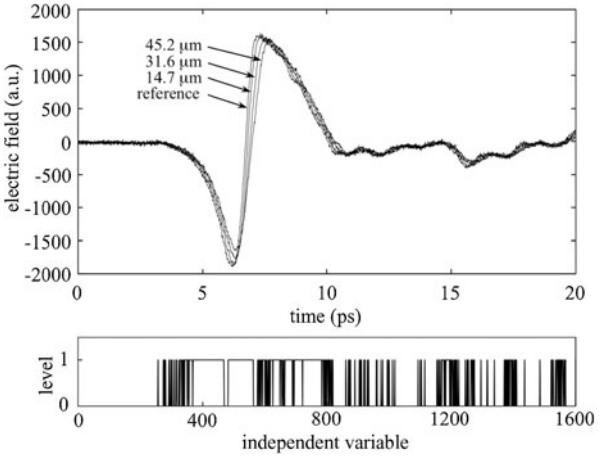

Fig. 3 Numerically synthesized THz waveforms (see text). For the sake of clarity, we show only three waveforms (i =13, 28, and 40, listed in Table 2) and the reference waveform $E_r(t)$. In the lower part of the figure, informative variables are shown as level one

Table 2 Predicted values of d_i for the i-th sample d_i, RMSEP values, and the number of retained variables for the PLS1, UVE-PLS1, and USE-UVE-PLS1 methods. The calibration dataset is shown in Fig. 3, where $N = 10$ and $M = 3{,}960$

Sample number i	1	4	7	10	13	16	19	22	25	28	31	34	37	40	RMSEP	Number of retained variables
d_i (μm)	1.1	4.5	7.9	11.3	14.7	18.1	21.5	24.9	28.3	31.6	35.0	38.4	41.8	45.2		
PLS1	1.9	4.6	7.8	11.0	14.2	17.7	21.4	24.9	28.4	31.8	35.3	38.5	41.6	44.4	0.33	3,960
UVE-PLS1	1.1	4.7	7.9	11.2	14.4	18.0	21.5	24.9	28.3	31.7	35.2	38.5	41.5	45.2	0.19	699
UVE-USE-PLS1	1.1	4.7	7.9	11.2	14.4	18.0	21.5	24.9	28.3	31.7	35.2	38.5	41.5	45.2	0.19	699

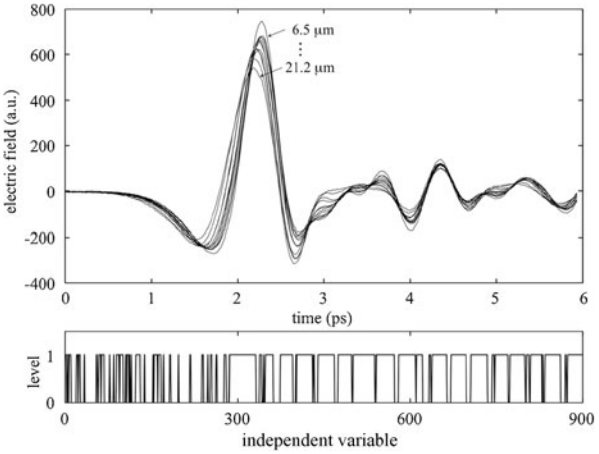

Fig. 4 Calibration waveforms obtained from various thickness of oil-based, acrylic, black paint films on an aluminum substrate, where $N = 12$ and $M = 900$. In the lower part of the figure, informative variables are shown as level one

in time when $d < 20$ μm. Predicted values of $\widehat{d}_i$ for the i-th sample and RMSEP values obtained from the PLS1, UVE-PLS1, and USE-UVE-PLS1 methods are listed in Table 2 for $i = 1, 3, 6, \cdots, 37$, and 40. In the lower part of Fig. 3, informative independent variables are shown. In spite of the high degree of overlapping of the waveforms, a film thickness of less than 10 μm could be predicted with a prediction error $\Delta d_i < 2\ \mu m$. In this numerical simulation, the error might be due to the artificially added noise.

4 Experimental Results and Discussion

4.1 Dry Paint Film

We prepared twelve calibration samples ($N = 12$) with different thicknesses. The samples were oil-based, acrylic, black dry paint films deposited on aluminum substrates. The experimental setup was the same as that described in the previous paper [6, 12], where typical THz time-of-flight measurement in a reflection geometry was used. The group refractive index of the film in the THz wavelength region was $n_g = 1.66$, which was obtained

Table 3 Predicted values of d_i for the i-th sample d_i, RMSEP values, and the number of retained variables for the PLS1, UVE-PLS1, and USE-UVE-PLS1 methods. The calibration dataset ($N = 12$ and $M = 900$) is shown in Fig. 4

Sample number i	1	2	3	4	5	6	7	8	9	10	11	12	RMSEP	Number of retained variables
d_i (μm)	6.5	6.8	9.4	11.0	12.2	14.3	15.2	15.5	18.2	18.9	21.0	21.2		
PLS1	5.6	8.9	9.0	10.2	11.9	12.9	14.4	14.1	17.8	20.1	22.3	20.4	1.11	900
UVE-PLS1	6.6	9.3	8.8	10.3	12.2	14.1	14.1	14.5	18.4	19.9	21.7	20.4	0.99	134
UVE-USE-PLS1	7.3	-	9.7	11.0	12.6	14.6	14.3	14.4	17.6	19.3	21.4	20.3	0.71	588

Table 4 Predicted values of d_i for the i-th sample d_i, RMSEP values, and the number of retained variables for the PLS1, UVE-PLS1, and USE-UVE-PLS1 methods. The calibration dataset ($N = 15$ and $M = 900$, not shown) was obtained from white polyester resin paint films deposited on galvanized steel substrates

Sample number i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	RMSEP	Number of retained variables
d_i (μm)	2.5	5.0	6.0	6.5	7.0	8.0	9.0	9.5	11.0	13.0	13.5	16.0	17.0	18.0	20.0		
PLS1	7.4	6.0	7.1	7.0	7.6	6.3	7.6	9.4	11.1	12.5	11.7	17.6	14.2	19.1	17.9	1.85	900
UVE-PLS1	5.8	6.2	7.4	6.4	7.9	8.0	8.7	9.9	10.6	11.6	11.0	16.8	14.6	19.4	19.4	1.47	699
UVE-USE-PLS1	5.8	6.2	7.4	6.4	7.9	8.0	8.7	9.9	10.6	11.6	11.0	16.8	14.6	19.4	19.4	1.47	699

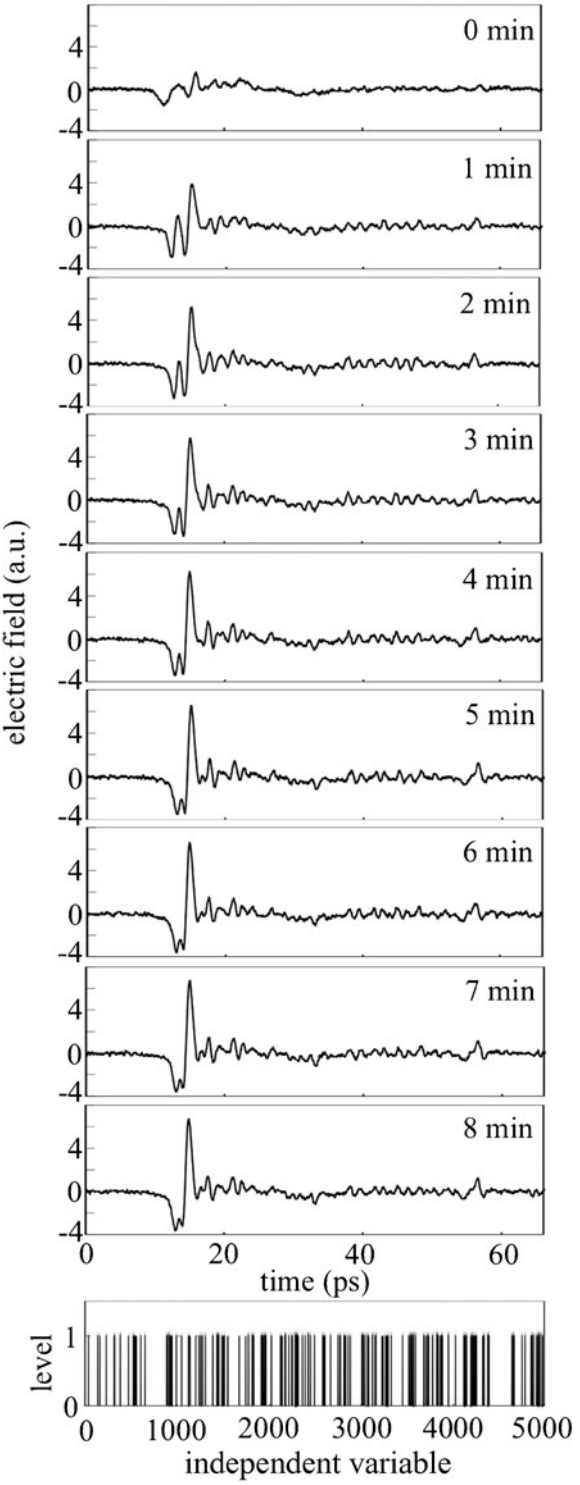

Fig. 5 Rescaled and reconstructed THz waveforms obtained from a wet paint film for eight elapsed times from 0 to 8 minutes with a step interval of one minute after painting. In the lower part of the figure, informative variables are shown as level one

Table 5 Predicted values of d_i for the i-th sample d_i, RMSEP values, and the number of retained variables for the PLS1, UVE-PLS1, and USE-UVE-PLS1 methods. The calibration dataset ($N = 9$ and $M = 5{,}000$) is shown in Fig. 5

Sample number i	1	2	3	4	5	6	7	8	9	RMSEP	Number of retained variables
Elapsed time (min)	0	1	2	3	4	5	6	7	8		
d_i (μm)	143	115	98.0	88.0	83.0	79.0	78.0	77.0	75.0		
PLS1	120	112	99.0	87.8	80.9	84.0	77.2	75.9	78.6	8.12	5,000
UVE-PLS1	128	114	99.2	87.5	80.8	82.5	77.1	75.9	78.8	5.34	2,298
UVE-USE-PLS1	-	-	95.7	88.9	83.3	80.2	76.7	76.1	77.1	1.43	444

from the literature [6, 12]. The reference thickness values of the paint films were obtained using an eddy-current meter. The precision of the thickness measurements at 20 μm was around 3.0 % according to the manufacturer's data sheet. The relative standard deviation of repetitive measurements of the same film was about 10 %. The relatively large deviation might be due to the spatial non-uniformity of the film thickness.

The upper part of Fig. 4 shows the twelve THz waveforms used as the calibration dataset ($M = 900$). Table 3 summarizes predicted values of $\widehat{d}_i$ and RMSEP values obtained from the PLS1, UVE-PLS1, and USE-UVE-PLS1 methods. With the USE-UVE-PLS1 method, sample #2 was eliminated as uninformative, resulting in a smaller value of RMSEP. From this experimental result, the USE-UVE-PLS1 method produced a better calibration model than the UVE-PLS1 or PLS1 method. For reference, we show informative variables in the lower part of Fig. 4. Thus, the USE-UVE-PLS1 method enables us to predict the film thickness with a prediction error of $\Delta d_i < 0.9$ μm for $d_i < 20$ μm.

Next, we measured the thickness of a white polyester resin paint film on a galvanized steel substrate, which is used as a pre-painted steel sheet in many steel products. We prepared fifteen calibration samples ($N = 15$ and $M = 900$) whose thickness were from 2.5 to 18.0 μm (waveforms not shown). Table 4 summarizes predicted values of $\widehat{d}_i$ for the i-th sample and RMSEP values obtained from the PLS1, UVE-PLS1, and USE-UVE-PLS1 methods. The USE-UVE-PLS1 method gave a relative prediction error of around 10 %, except for $d_1 = 2.5$ μm, whose prediction error was the lower detection limit in this case.

4.2 Wet Paint Film

One of the advantages of the THz paint meter over the conventional one is the ability to monitor the drying process of a wet film. Such an ability was already confirmed in the previous paper by using a parameter fitting method [12]. In the present paper, in order to demonstrate the usefulness of the proposed scheme, we applied a modified PLS1 method to the same dataset (Fig. 8 in Ref. 12). The THz waveforms corresponding to elapsed time from 0 to 8 minutes with a step interval of one minute after painting (0 min) are shown in Fig. 5, where $M = 5{,}000$. As reference thickness values for building the calibration model, we directly used estimated values of d_i obtained from the fitting method (Table II in Ref. 13). The predicted values of $\widehat{d}_i$ for the i-th sample obtained by the proposed method are summarized in Table 5. As was discussed in the previous paper [12], we found that the thickness of the film reduced with the elapsed time. From this we can judge whether the paint is dried, although the change in refractive index (and absorption coefficient) during the drying must affect the obtained thicknesses. The informative variables are shown in the

lower part of Fig. 5. Again, the USE-UVE-PLS1 method gave the best calibration model for prediction. In this case, sample #1 and #2 were eliminated as uninformative in building the model.

In the proposed scheme, a major factor that determines the precision in predicting the thickness of the film is the uncertainty in detecting the zero-point position of time in the THz waveforms. The uncertainty is mainly due to the poor reproducibility of the mechanical error in attaching the sample: which might be around a few micrometers. In such a case, a systematic error is introduced when building the calibration model. In addition, for thicker paint films, the leading edge of the first echo pulse appears earlier in time, and the trailing edge of the second echo pulse appears later, even if the zero position could be determined precisely. Taking these facts into account, we defined the zero position as a 10%-constant-fraction point on the leading edge of the first echo pulse. Because only the relative time shift is important in building the calibration model, this procedure worked well. Another factor that affects the precision and accuracy in prediction is errors due to the reference thickness values used for building the calibration model. However, the overall precision and accuracy in the proposed scheme might be improved if the number of calibration samples is increased, thanks to employing the multivariate analysis method.

5 Conclusions

We demonstrated that the application of a multivariate analysis method to a dataset obtained in THz reflectometry is effective for predicting the thickness of a single-layered paint film. Also, we found that the modified PLS1 method is useful. The prediction procedure worked well, provided that an accurate calibration model is built in advance, even for a thickness of less than 10 μm. Although we applied the method to a single-layered film in the present paper, the proposed scheme is applicable to data obtained from multilayered paint films merely by increasing the number of the dependent variables.

Acknowledgements This work was supported by Grants-in-Aid for Scientific Research 21650111 and 23656265 from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. We also gratefully acknowledge financial support from the Renovation Center of Instruments for Science Education and Technology at Osaka University, Japan. The authors thank Mr. Hiroyasu Furukawa of Nippon Steel and Sumitomo Metal Corporation, Japan, for preparation of the paint samples and Drs. Hideyuki Ohtake and Jun Takayanagi of AISIN SEIKI Co. Ltd., Japan, for conducting the THz reflectometry measurements.

References

1. D. M. Mittleman, Sensing with THz radiation, 1st edn. (Springer, 2003).
2. M. Tonouchi, Nature 1, 97 (2007).
3. P. U. Jepsen, D. G. Cooke and M. Koch, Laser Photon. Rev. 5, 124 (2011)
4. E. P. Parrott, S. M. Sy, T. Bllu, V. P. Wallace and E. Pickwell-Macpherson, J. Biomed. Opt. 16, 106010 (2011).
5. D. M. Mittleman, S. Hunsche, L. Boivin and M. C. Nuss, Opt. Lett. 22, 904 (1997).
6. T. Yasui, T. Yasuda, K. Sawanaka and T. Araki, Appl. Opt. 44, 6849 (2005).
7. T. Yasuda, T. Yasui, T. Araki and E. Abraham, Opt. Comm. 267, 128 (2006).
8. D. J. Cook, S. J. Sharpe, S. Lee and M. G. Allen, "Terahertz time domain measurements of marine paint thickness," in *Optical Terahertz Science and Technology*, 2007 OSA Technical Digest Series (Optical Society of America, 2007), paper TuB5.
9. Y. Izutani, M. Akagi, and K. Kitagishi, "Measurements of paint thickness of automobiles by using THz time-domain spectroscopy," in *Proceedings of 37th International Conference on Infrared, Millimeter, and Terahertz Waves*, Fri-A-3-4.

10. J. B. Jackson, J. Bowen, G. Walker, J. Labaune, G. Mourou, M. Menu and K. Fukunaga, IEEE Trans. Terahertz Sci. Tech. 1, 220 (2011).
11. J. Takayanagi, H. Jinno, S. Ichino, K. Suizu, M. Yamashita, T. Ouchi, S. Kasai, H. Ohtake, H. Uchida, N. Nishizawa and K. Kawase, Opt. Express 17, 7533 (2009).
12. T. Yasuda, T. Iwata, T. Araki and T. Yasui, Appl. Opt. 46, 7518 (2007).
13. M. Schwerdtfeger E. Castro-Camus, K. Krügener, W. Viöl, and M. Koch, Appl. Opt. 52, 375 (2013).
14. P. Geladi and B. R. Kowalski, Anal. Chim. Acta. 185, 1 (1986).
15. D. M. Haaland and E. V. Thomas, Anal. Chem. 60, 1193 (1988).
16. J. Koshoubu, T. Iwata and S. Minami, Appl. Spectrosc. 54, 148 (2000).
17. J. Koshoubu, T. Iwata and S. Minami, Anal. Sci. 17, 319–322 (2001).
18. V. Centner, D. L. Massart, O. E. de Noord, S. de Jong, B. M. Vandeginste and C. Sterna, Anal. Chem. 68, 3851 (1996).
19. J. Ferre and F. X. Rius, Anal. Chem. 68, 1565 (1996).
20. P. J. Rousseeuw and A. M. Leroy, Robust regression and outlier detection, 1st edn. (John Wiley & Sons, New York, 1987).