DT-NMBP-08-2019
Real-time nano-characterisation technologies

NanoQI

Multimodal X-ray and Hyperspectral Thin-Film Nano-material Evaluation and Quality Imaging

Starting date of the project: 01/03/2020
Duration: 42 months

= Deliverable D6.3 =

Verification of performance of XRR and HSI as Quality control in industrial R2R vacuum coating: Set of surface Images of a roll 100 m x 0.6 m showing for each layer in a 4 layer stack the layer thickness and its deviation with sensitivity of ± 1 nm; after R2R processing at 2 m/min

Due date of deliverable: 31/08/2023
Actual submission date: 01/09/2023

WP and Lead Beneficiary: WP6, NORD
Version: V1.0

Dissemination level		
PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission Services)	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

AUTHOR

Author	Organization	Contact (e-mail, phone)
Dr. Matthias Fahland	FhG-FEP	matthias.fahland@fep.fraunhofer.de

DOCUMENT DATA

Keywords	R2R vacuum coating, HSI, validation
Point of Contact	Name: Dr. Matthias Fahland Partner: Fraunhofer FEP Address: Winterbergstraße 28, 01277 Dresden E-mail: matthias.fahland@fep.fraunhofer.de

DISTRIBUTION LIST

Date	Issue	Recipients
30/08/2023	V0.1	Partners for feedback
01/09/2023	V0.2	WP leader/PM for finalization
01/09/2023	V1.0	EC, consortium partners

REVIEW PROCESS

Document version	Date	Status/Change
V0.2	01/09/2023	Revised version sent to AMI
V1.0	01/09/2023	Final revision, formatting

VALIDATION PROCESS

Reviewers		Validation date
Work Package Leader	Fabiano Rimediotti (NORD)	30/08/2023
Project Manager	Martina Chopart (AMI)	01/09/2023
Project Coordinator	Matthias Fahland (FEP)	01/09/2023

Executive Summary

The results reported in this deliverable are related to the demonstration of the HSI system in vacuum roll-to-roll coating machines.

This demonstration was done at a pilot coating machine over a width of 0.49 m. This dimension puts the project partner into the position to clearly understand the potential as well as the challenges of the technology.

Various single layers and layer systems have been demonstrated by the technology. Generally, the hardware installation as well as the software demonstrated their capabilities to solve the measurement tasks.

Measurement signal stability and the handling of large amount of data, as well as special topics related to the uninterrupted substrate movement posed special problems to deal with. This is distinguishing this task from the work reported in the deliverables D6.1 and D6.2. The partners presented mitigating solutions for all issues. This makes the technology ready for usage in later follow-up projects.

Table of Contents

1. Introduction

The Deliverable D6.3 deals with the verification of the HSI as a valuable tool for roll-to-roll inline processes. This is demonstrated with the example of different single layers and one double layer system. Details are given in the corresponding chapters. Ground truth data came from different sources, among them also X ray Reflectometry. The general procedure of machine learning is presented in section 2.1.The examples for single layer systems and double layer systems are shown in the sections 2.3 and 2.2, respectively.

In the content of the deliverable various quantitative numbers are given:
1. a set of surface images of a roll need to be taken from a roll
2. a roll section of 100 m x 0.6 m
3. R2R processing at 2 m/min possible
4. 4-layer stack: individual layer thickness and its deviation with sensitivity of ± 1 nm.

Even though these four criteria had not been demonstrated together, the work for D6.3 dealt with all of them.

Images of the roll had been taken. Due to aperture restrictions at the coFlex 600, the maximum width of observation was 0.49 m. This value can easily be adapted to 0.6 m in a production machine. More than that, the scaling of resolution with observation width, achievable under normal circumstances can be evaluated based on the project results.

The maximum total length of 275 m was investigated with the HSI system in the roll-to-roll machine. This was done with the partner roll from NORD.

The HSI inspection was demonstrated at a maximum line speed of 6.4 m/min. The partners have developed a general understanding of the deformation of the field of view, in dependence of the interaction between the frame rate and line speed. This deformation of the field of view is not limiting the information provided by HSI. However, the understanding is necessary to identify the exact position which a measurement signal is related to.

Considering all the difficulties which had been faced during the installation, the four-layer stack seemed to be too complex for the evaluation approach of NanoQI. More details and a guideline for a modified approach in a possible follow up project are given in section 2.4.

2. Results and discussion

2.1. General workflow for Machine Learning for HSI data

To understand the content of this work package, it is important to recall the general procedure of machine learning algorithm. This is given in the summary below:

- Get HSI data and corresponding ground truth data
- Load (and cut into smaller blocks if necessary) HSI data
- If necessary: do manual referencing (if not done in Imanto)
  - Standard referencing
  - "Blockwise referencing"
- Select spectra corresponding to ground truth data (e.g. measurement spot of XRR / ellipsometry)
  - Mean spectra of the area
  - Random spectra of the area
  - Spectrum most similar to the mean spectrum
  - All (or a certain number of) spectra
  - **Result:** data set of x data (spectra) and y data (ground truth)
- Train, validate and optimize regression/prediction model including pre-processing of the data (see Figure 1)
  - Train: finding rules to predict the ground truth based on the spectra
  - Validation: How good is the prediction of the trained model on unseen data?
    - Cross validation (different number of partitions, e.g. k-fold, leave-on-out)
    - Validation set (e.g. splitting the data set in train, validation & test set)
  - Optimization: finding the best hyperparameters for the prediction
    - Hyperparameters: set screws of the prediction algorithms
    - Second validation step (also called test step) necessary:
      - Estimate model performance with validation data
      - Train model with best hyperparameters on train & validation data
      - Predict test data → final model performance for unseen data
      - Methods:
        - Nested cross validation
        - Test set (e.g. splitting the data set in train, validation & test set)
    - Can also be used to compare different algorithms & data pre-processing methods
  - **Results:**
    - trained models ready to use for prediction
    - performance indicators (RMSE, R^2, accuracy, F1, …)
- Use model to predict new data
  - Same measurement conditions
  - Same pre-processing (can be part of the model e.g. using sklearn pipelines)

NanoQI
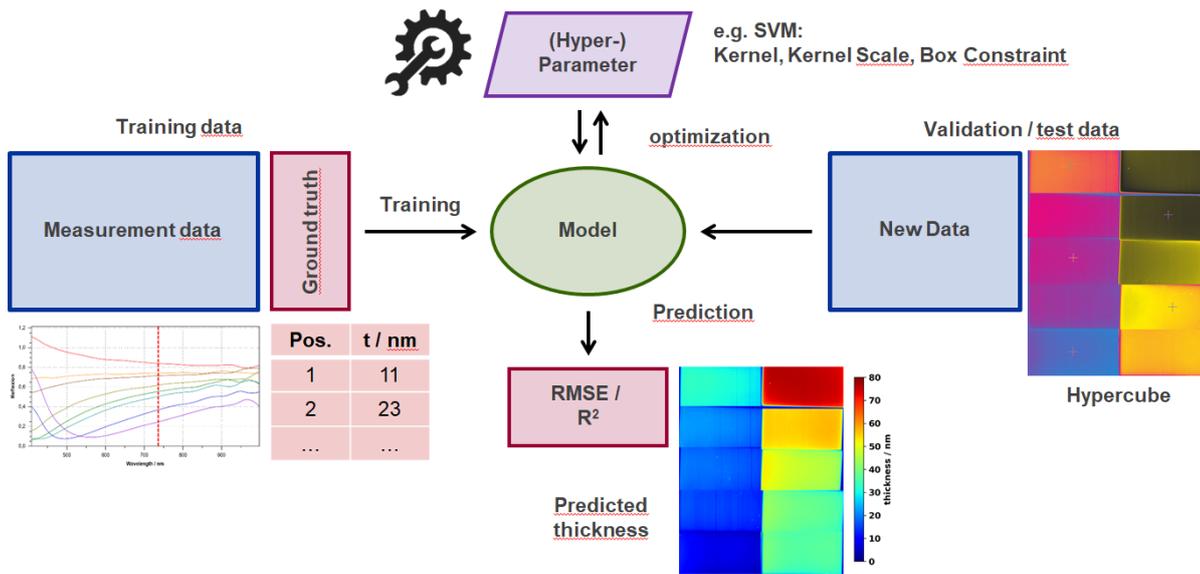
Figure 1: Schematic representation of the training of a ML prediction model.

Machine learning is a powerful technique that can be used to analyze and interpret HSI data. HSI data are images that contain spectral information for each pixel, which can be used to identify the material properties and composition of the sample. However, HSI data are often large and complex, and require a lot of processing and analysis to extract useful information. Therefore, we need to follow a general workflow that consists of several steps to apply machine learning for HSI data.

The first step is to get HSI data and corresponding ground truth data from measurements or experiments. The ground truth data are the reference values that we want to predict using the HSI data, such as the thickness or refractive index of a thin film. The ground truth data can be obtained from other measurement techniques, such as XRR or ellipsometry, or from simulations or calculations. The HSI data and the ground truth data should be acquired under the same conditions and for the same sample area.

The second step is to load the HSI data into a suitable format and cut it into smaller blocks if necessary for faster processing. The HSI data are usually stored as a three-dimensional array, where the first two dimensions are the spatial coordinates (x and y) and the third dimension is the spectral dimension (wavelength or energy). The HSI data can be very large, depending on the resolution and the number of spectral channels. Therefore, it may be necessary to cut the HSI data into smaller blocks that can fit into the memory and can be processed more efficiently.

The third step is to perform manual referencing if not done in Imanto, using either 'standard' referencing or 'blockwise' referencing method.

The fourth step is to select spectra that correspond to the ground truth data, such as the measurement spot of XRR or ellipsometry. Different selection strategies can be used, depending on the quality and variability of the spectra and the ground truth data. Some possible selection strategies are:

- Taking the mean spectrum of the area, which reduces noise and outliers but may lose some spectral features.
- Taking random spectra of the area, which preserves some spectral variability but may introduce noise and outliers.
- Taking the spectrum most similar to the mean spectrum, which balances noise reduction and feature preservation but may be sensitive to small differences.
- Taking all or a certain number of spectra, which maximizes spectral variability but may increase noise and outliers.

The result of this step is a data set of x data (spectra) and y data (ground truth), which can be used for machine learning.

The fifth step is to train, validate and optimize a regression or prediction model based on the data set. This involves finding rules to predict the ground truth based on the spectra (training), evaluating how well the model performs on unseen data (validation), and finding the best hyperparameters for the prediction algorithm (optimization). Different methods can be used for validation and optimization, such as cross validation, validation set, test set, or nested cross validation. These methods can also be used to compare different algorithms and data pre-processing methods. The results of this step are trained models that are ready to use for prediction and performance indicators such as RMSE, R2, accuracy, F1, etc.

The final step is to use the model to predict new data under the same measurement conditions and with the same pre-processing steps. The new data can be from different samples or different regions of interest within the same sample. The pre-processing steps can include normalization, smoothing, filtering, feature extraction, dimensionality reduction, etc. The pre-processing steps can be part of the model using sklearn pipelines or other tools.
This is an example of a general workflow for machine learning for HSI data. Depending on the specific application and problem, some steps may vary or be skipped.

2.2. Double layer measurement

The following text describes the data evaluation of the "double layer sample" in detail and can be used for example for the preparation of the deliverables of WP6.

Description of the sample

The double layer sample is a sample with a total length of 110 m produced on the CoFlex system of FhG FEP. On a PET film, first a ZTO layer with stepwise varied layer thickness was applied and then an ITO layer was applied on each of the ZTO steps, also stepwise. Figure 2 shows a schematic representation of the sample. The nominal ZTO film thickness was varied from 40 nm to 130 nm and the nominal ITO film thickness from 30 nm to 120 nm in 10 nm steps.
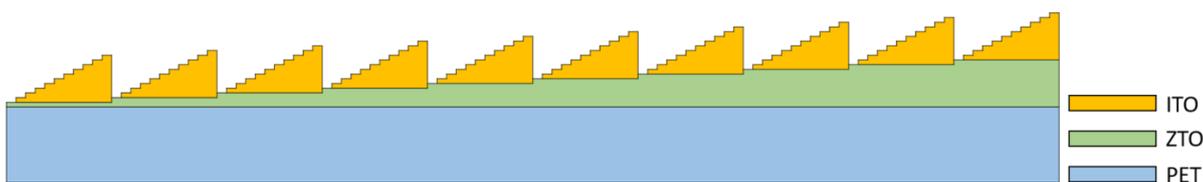
Figure 2: Schematic representation of the double layer sample.

After fabrication, the sample was measured using a NEO VNIR HSI camera (HySpex VNIR-1800) in the spectral range from 400 to 1000 nm. The measurement was done only with dark current correction. Therefore 200 frames were measured with the shutter of the camera closed. For the white referencing, a not coated part of the PET foil was also measured.

The FOV of the camera was 40 cm and the sample was measured several times with different roll speeds and a constant frame rate of 160 Hz. The minimum measurement speed was 0.2 m/min (non-distorted measurement of the sample) and the maximum measurement speed was 6.4 m/min (thirty-twofold compression of the measurement in rolling direction). For the evaluation, the measurement with a speed of 0.8 m/min (fourfold compression of the measurement in rolling direction) was used. This forms a compromise between information content and size of the HSI data set to be evaluated. The size of the obtained file is 163 GB which corresponds to a size of the HSI hypercube of about 240000x1800x186 pixels (length in rolling direction x length perpendicular to rolling direction / FOV x wavelengths).

Generation of ground truth data

To generate the ground truth data, a piece of the sample was cut out for each layer thickness combination and the real layer thickness of the ZTO and the ITO layer was determined using XRR and/or ellipsometry. For further evaluation, it was assumed that the layer thicknesses obtained in this way are representative for the entire range of the respective layer thickness combination.

Data selection and preparation

Since the complete data set with a size of 162 GB cannot be processed properly, it was divided in a first step into partial data sets with a size of 3000x1800x186 pixels and a resulting file size of about 2 GB. A Python tool developed by NEO was used for this purpose.

The next step was the referencing of the measurements. Two approaches were chosen: 'standard' referencing and 'blockwise' referencing. For both methods, the measurement of the uncoated PET film was used. In 'standard' referencing, the average of this uncoated part was taken in the direction of roll movement (around 600 frames), resulting in a dataset of 1x1800x186 pixel. Subsequently, each line of the sample data set was divided by the white reference obtained in this way. This procedure corresponds to the calculation given in equation 1:

$$I_c(\lambda) = \frac{I_o(\lambda) - I_d(\lambda)}{I_w(\lambda) - I_d(\lambda)} \tag{1}$$

where $I_c(\lambda)$ is the corrected image intensity, $I_o(\lambda)$ the original image intensity, $I_d(\lambda)$ is the dark current recorded with the light source switched off and the lens covered, and $I_w(\lambda)$ is the intensity of the white reference for the wavelength λ. The dark current was already subtracted during the data collection.

It was found that with this type of referencing, a distinct periodic pattern can be seen in the referenced data (see Figure 3, left). This is caused by an irregularity on the surface of the roll. With 'blockwise' referencing, an attempt was therefore made to reduce this irregularity. For this purpose, the reference area of the measurement (uncoated PET film) was divided into equal sections, each covering a complete revolution of the roll. Ten of these sections were averaged to form a white reference of 74x1800x186 pixels, where the 74 pixels correspond to one revolution of the roll. The entire sample was then referenced with this white reference. The result is shown in Figure 3 (right) and shows a significantly improved result compared to the 'standard' referencing. The disadvantage of this referencing is that it cannot currently be used live during the measurement. In addition, it was found that the length of a roll revolution is not always exactly the same and the generated referencing is therefore not perfect.
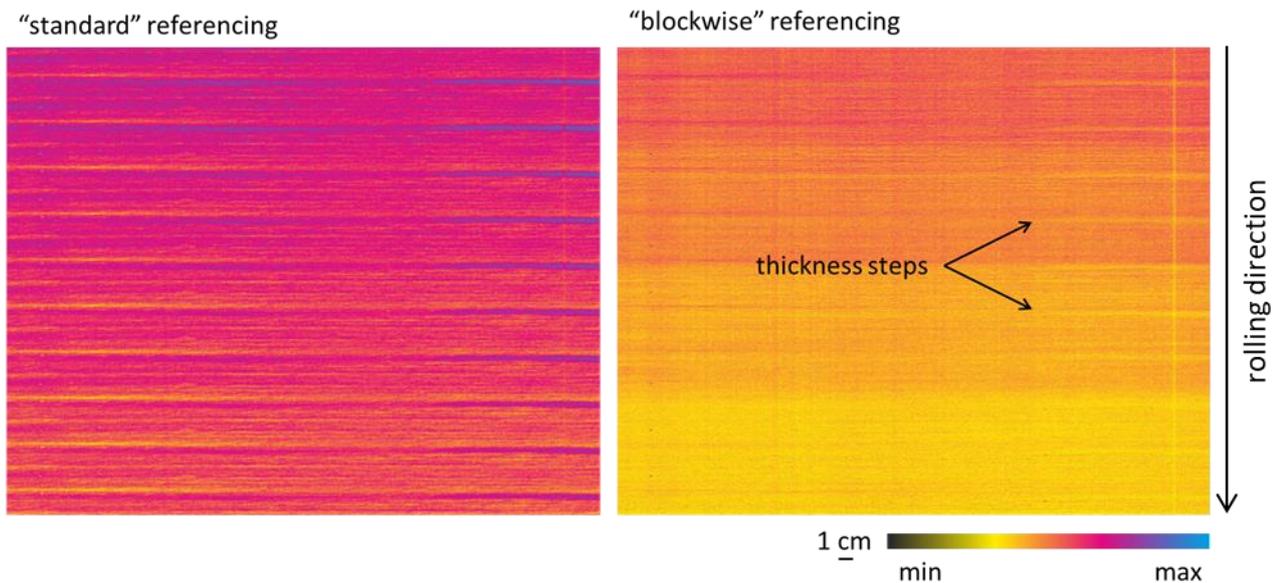
Figure 3: Comparison of 'standard' referencing and 'blockwise' referencing. The color-coded reflectivity at a wavelength of 650 nm for a ~30x40 cm area of the double layer sample is shown.

In the next step, the data for training the prediction models were extracted from the measurement. For this purpose, 800 lines (along the rolling direction) were selected from the center of each of the layer thickness steps. This was done automatically. A mean value spectrum was then formed from this area with the size 800x1800x186. The 110 mean value spectra obtained in this way, together with the corresponding ground truth layer thickness values, form the 'mean' data set. In addition, the spectrum most similar to the mean spectrum was

selected for each thickness step. The spectra obtained in this way from the 'similar' dataset. 1000 random spectra were also selected, which together with the corresponding ground truth values form the 'specs' dataset. These three data sets were used for further analysis.

Training, optimization and validation of regression models

The three obtained data sets were then used to train, optimize and validate various regression models. A separate model was calculated for each of the two target values (ZTO & ITO layer thickness). All calculations were performed with Python and the scikit-learn library. The necessary scripts were created by ourselves.

For the regression three different algorithms were compared: Partial Least Squares Regression (PLS)[1], Elastic Net Regression (EN)[2] and Neural Networks / Multilayer Perceptron (MLP)[3]. A description of the algorithms' functionality, their implementation in the scikit-learn library and the hyperparameters of the algorithms can be found in the respective literature.

The optimization of the hyperparameters of the algorithms was done by the random search method. This means that for n model calculations (here n = 50) random combinations of the hyperparameters of the algorithms are used. A detailed description of the random search method and a comparison with alternative methods of hyperparameter optimization can be found in Bengio et. al.[4] An overview of the optimized hyperparameters for the individual algorithms can be found in the Table 4 in the appendix. In addition to the hyperparameters of the algorithms, the pre-processing of the spectra by the random search procedure was also optimized. This includes a normalization of the spectra and a principal component analysis (PCA, not used for PLS) before the actual model training.

An important step in the creation of machine learning models is the validation and testing of the models. This involves investigating how well the model can predict the target values of unknown samples, i.e. samples that it did not see in the training step. In the present report, this validation was carried out by dividing the data into a training data set, a validation data set and a test data set. The division is carried out in such a way that all spectra of a layer thickness combination only appear in one of the data sets. The actual training and variation of the hyperparameters is carried out with the training data set. The quality of the models obtained is examined with the validation data set. The metrics root mean squared error (RMSE) and the coefficient of determination (R^2) are considered. Once the hyperparameter optimisation is complete, a new model is trained with the best hyperparameters found using the data from the training data set and the validation data set, and this model is finally used to predict the test data set and again calculate RMSE and R^2. This two-step procedure is necessary for hyperparameter optimisation, as otherwise an overfitting of the models can occur.

Results of the data evaluation for the measurements of the double layer sample

In the following, the results of the data evaluation for the double layer system are evaluated and compared. The data evaluation was performed with the three described algorithms (PLS, EN and MLP), the three described data sets ('mean', 'similar', and 'specs'), the two referencing variants ('standard' and 'blockwise') and with the complete and a reduced wavelength range (400 nm - 1000 nm and 430 nm - 730 nm). The tests with the reduced wavelength range were carried out because the LEDs originally used with a spectral range of 400 nm to 1000 nm were not suitable for use in the R2R CoFlex system. After only a short period of use, there was a drastic and only partially reversible drop in intensity. Therefore, a new light source with more robust white LEDs was developed and used for further measurements. However, these LEDs have a reduced wavelength range.

Table *1* shows an overview of the obtained RMSE and R^2 values of all calculated models:

Table 1: RMSE and R2 values for all the trained models.

---

[1] https://scikit-learn.org/stable/modules/generated/sklearn.cross_decomposition.PLSRegression.html
[2] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.ElasticNet.html
[3] https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html
[4] Bergstra, J. and Bengio, Y., 2012. Random search for hyper-parameter optimization. *Journal of machine learning research*, *13*(2).

NanoQI

RMSE	ITO				R2	ITO			
	reference wavelenght range	'standard' 400 - 1000 nm	'standard' 430 - 730 nm	'blockwise' 400 - 1000 nm		reference wavelenght range	'standard' 400 - 1000 nm	'standard' 430 - 730 nm	'blockwise' 400 - 1000 nm
'mean'	EN	11.08	18.73	11.55	'mean'	EN	0.89	0.62	0.87
	PLS	7.73	14.96	7.74		PLS	0.94	0.77	0.94
	MLP	6.22	10.51	7.20		MLP	0.96	0.88	0.94
'similar'	EN	14.82	19.65	26.02	'similar'	EN	0.80	0.66	0.42
	PLS	14.94	22.23	27.84		PLS	0.80	0.58	0.34
	MLP	12.23	16.86	24.62		MLP	0.87	0.75	0.53
'specs'	EN	21.99	26.00	19.57	'specs'	EN	0.65	0.51	0.72
	PLS	22.22	25.92	19.66		PLS	0.64	0.52	0.72
	MLP	17.57	25.96	21.51		MLP	0.78	0.54	0.77
RMSE	ZTO				R2	ZTO			
	reference wavelenght range	'standard' 400 - 1000 nm	'standard' 430 - 730 nm	'blockwise' 400 - 1000 nm		reference wavelenght range	'standard' 400 - 1000 nm	'standard' 430 - 730 nm	'blockwise' 400 - 1000 nm
'mean'	EN	6.33	5.31	6.22	'mean'	EN	0.95	0.97	0.95
	PLS	5.37	5.32	5.49		PLS	0.96	0.97	0.96
	MLP	10.21	8.27	9.34		MLP	0.84	0.91	0.86
'similar'	EN	20.44	26.08	22.05	'similar'	EN	0.49	0.21	0.41
	PLS	24.56	28.32	21.66		PLS	0.30	0.16	0.44
	MLP	14.21	16.89	21.85		MLP	0.70	0.66	0.47
'specs'	EN	27.68	30.30	27.52	'specs'	EN	0.27	0.07	0.27
	PLS	23.73	24.01	23.58		PLS	0.45	0.43	0.46
	MLP	20.24	20.84	19.55		MLP	0.57	0.56	0.60

The results show that the best models are obtained for the 'mean' dataset, followed by the 'similar' dataset, while the worst results are obtained for the 'specs' dataset. However, when looking at the prediction results, it was found that these results do not translate into a good prediction result and that much better prediction results are obtained for the 'specs' models. Further findings of the data evaluation are:

- Goodness of algorithms: EN ~ PLS < MLP for almost all models
- The smaller wavelength range reduces the model quality
- 'Blockwise' referencing has only small effect compared to 'standard' referencing
- Results for ZTO worse than for ITO

The following figures (Figure 4, Figure 5 & Figure 6) show the predicted layer thickness and layer thickness profiles for the complete measured sample (~120 m) with the best found 'specs' model with 'standard' referencing, a wavelength range from 400 nm to 1000 nm and the MLP algorithm:
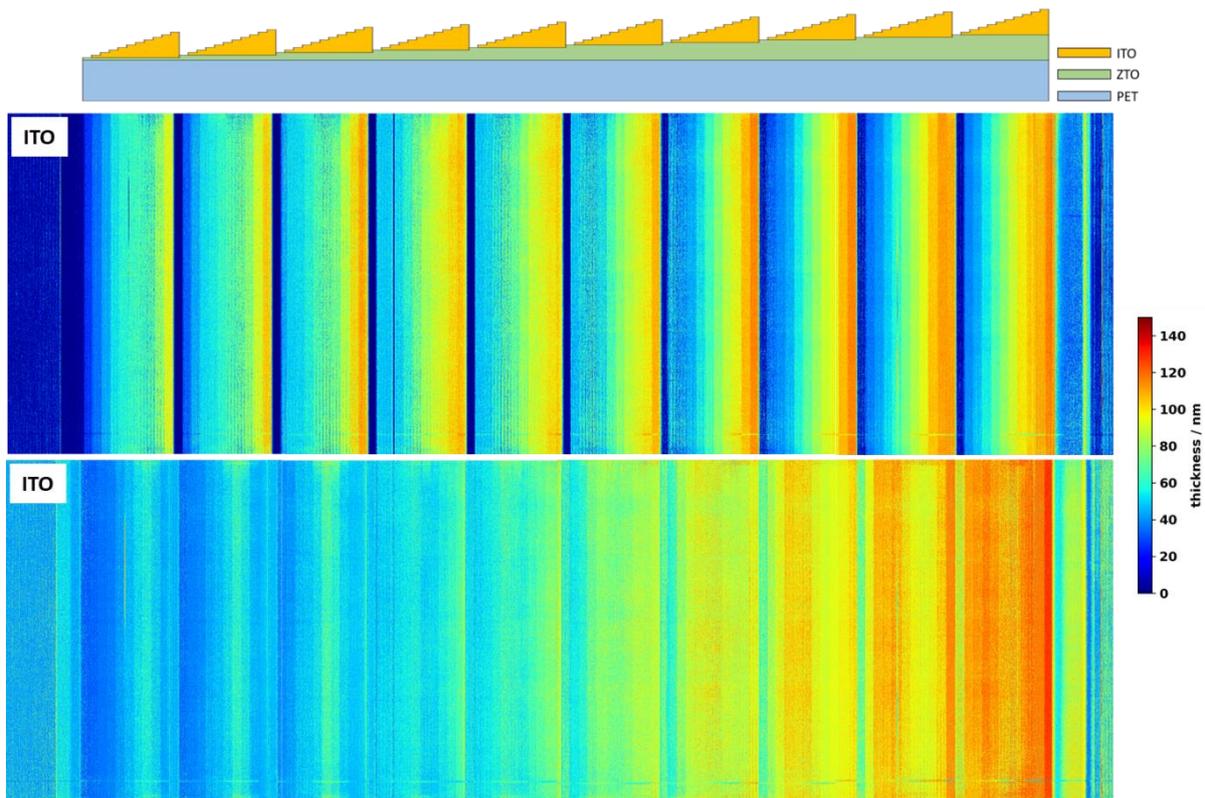
Figure 4: Predicted ITO and ZTO layer thickness for the whole measured sample, predicted with the MLP model trained on the 'specs' dataset with 'standard' referencing and a wavelength range from 400 nm to 1000 nm.
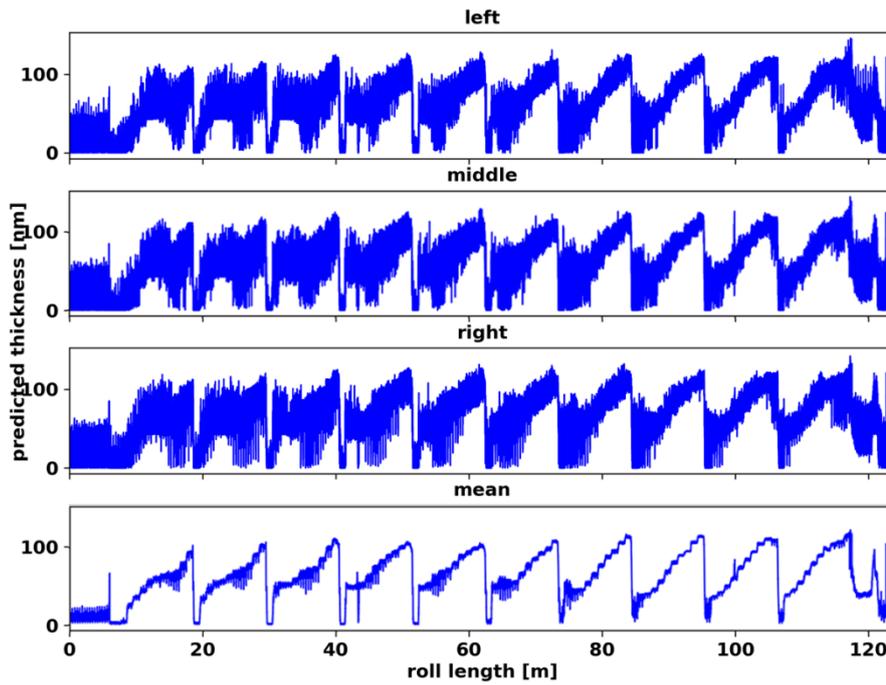
Figure 5: Predicted ITO layer thickness profiles for the whole measured sample, predicted with the MLP model trained on the 'specs' dataset with 'standard' referencing and a wavelength range from 400 nm to 1000 nm. Left & right: Profile at the left/right side of the roll. Middle: Profile at the middle of the roll. Mean: Mean profile over the complete roll, perpendicular to the rolling direction.
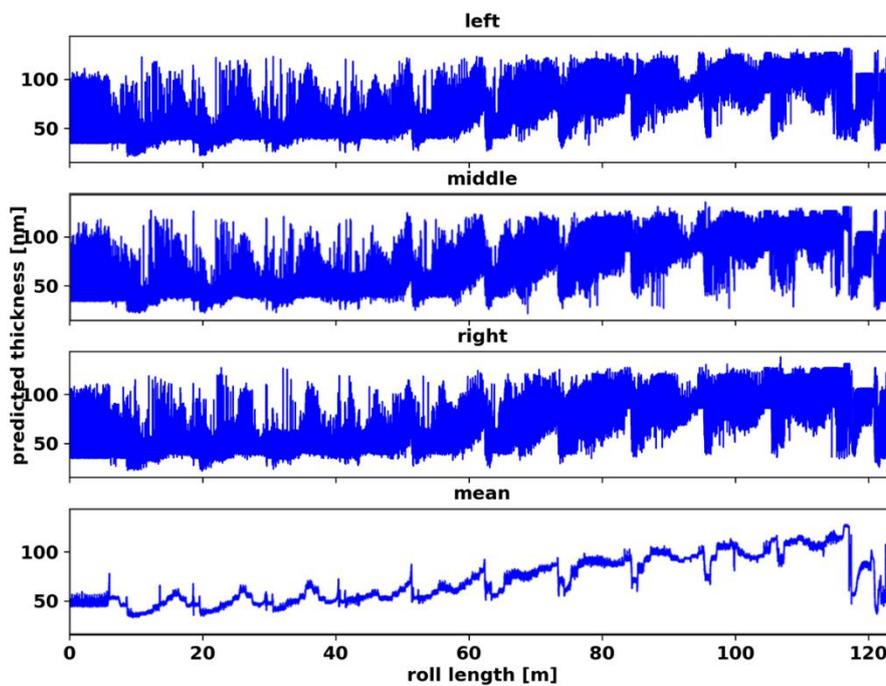
Figure 6: Predicted ZTO layer thickness profiles for the whole measured sample, predicted with the MLP model trained on the 'specs' dataset with 'standard' referencing and a wavelength range from 400 nm to 1000 nm. Left & right: Profile at the left/right side of the roll. Middle: Profile at the middle of the roll. Mean: Mean profile over the complete roll, perpendicular to the rolling direction.

The layer thickness prediction, especially for the ITO layer, corresponds very well to the theoretical layer thickness. Also for the ZTO layer, the stepwise increase of the layer thickness can be seen in principle. However, a closer look at the results (especially the layer thickness profiles) shows that the predicted layer thicknesses are subject to strong noise. This applies in particular to the edge areas of the roll. The following reasons can be given for this:

- Relatively low SNR of the measurements (caused, among other things, by the relatively low intensity of the light source)
- Irregularities of the roll surface which have a strong effect on the reflectivity
- Relatively thin layer thickness of the considered layers
- Problematic double layer system

Especially the last point is very relevant. Since the two layers under consideration are transparent materials with similar optical properties, a simultaneous prediction for both layer thicknesses is not always possible. This is not a limitation of the HSI measurement system or the data evaluation, but a physical limitation. However, the problem is further amplified by the relatively low quality of the reflectance spectra.

An examination of a model trained on the 'mean' data set (Figure 7, 'mean', 400 - 1000 nm, 'standard' referencing, MLP) shows that the prediction for the complete roll is worse than for the 'specs' data set, although better RMSE values are obtained for this model (6.22 vs. 17.57). The reason for this is that in the 'mean' dataset the variation of the spectra (due to noise, roll irregularities, etc.) is not included in the training and the model obtained is therefore less robust. This is not considered when creating the model, and thus when calculating the RMSE and R^2 values.
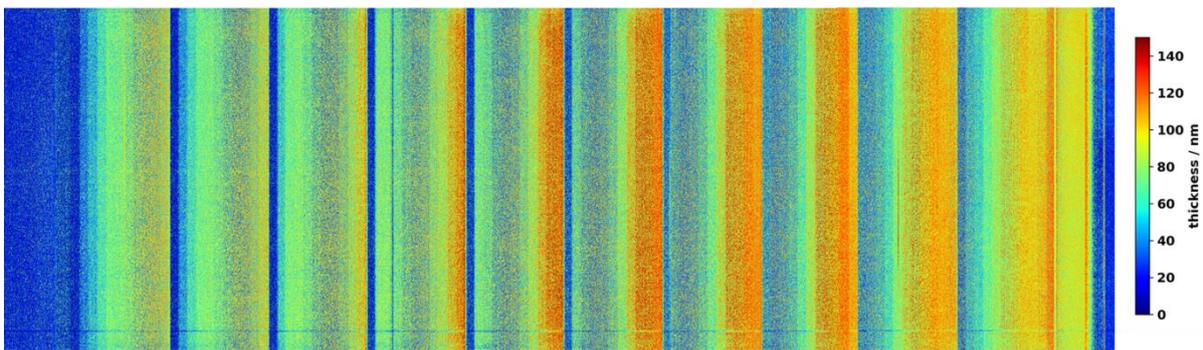
Figure 7: Predicted ITO layer thickness for the whole measured sample, predicted with the MLP model trained on the 'mean' dataset with 'standard' referencing and a wavelength range from 400 nm to 1000 nm.

If we look at the layer thickness profiles for the prediction with one of the models with 'blockwise' referencing (Figure 8, 'mean', 400 - 1000 nm, 'blockwise' referencing, MLP), we also see that this hardly improves the prediction. The strong noise of the predicted layer thickness remains. The reason for this is that although

'blockwise' referencing reduces the visible variation due to roll inhomogeneity, it does not completely eliminate it. Thus, the effect on the model quality is small.
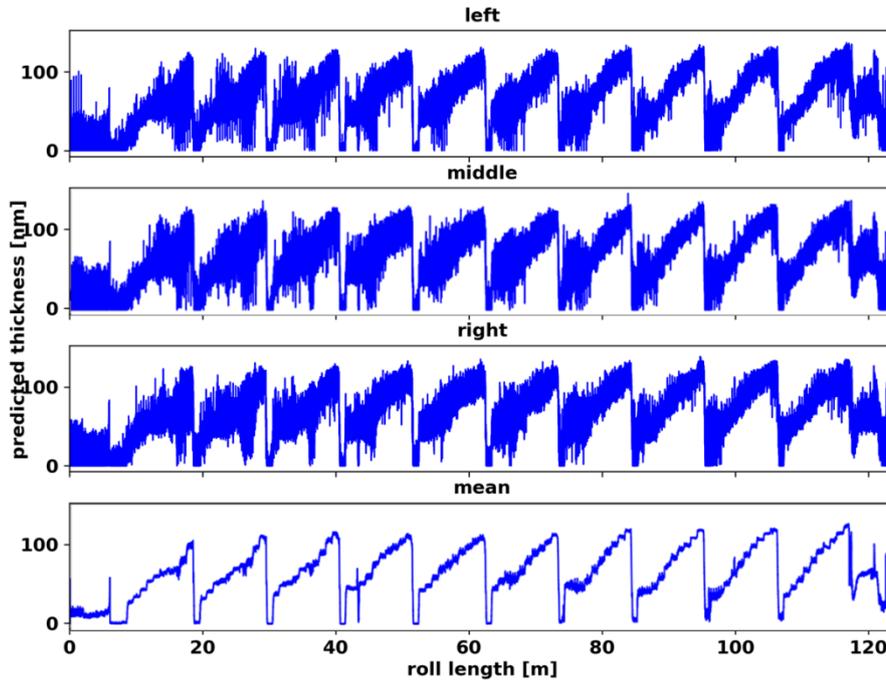
Figure 8: Predicted ITO layer thickness profiles for the whole measured sample, predicted with the MLP model trained on the 'specs' dataset with 'blockwise' referencing and a wavelength range from 400 nm to 1000 nm. Left & right: Profile at the left/right side of the roll. Middle: Profile at the middle of the roll. Mean: Mean profile over the complete roll, perpendicular to the rolling direction.

2.3. Single layer measurement

Description of the samples

In addition to the double layer sample four additional single layer samples were coated and measured:

- **ITO@Al@PET**; 0 nm & 20 nm to 110 nm in 10 nm steps, 3 m per step → CoFlex
- **ITO@PET**; 0 nm & 20 nm to 110 nm in 10 nm steps, 3 m per step → CoFlex
- **AlN@PET**; 0 nm & 20 nm to 110 nm in 10 nm steps, 3 m per step → CoFlex
- **AlOx@PET**; 0, 10, 20, 30, 50 & 80 nm steps, 50 m per step → from NORD

After fabrication, the samples were measured using a NEO VNIR HSI camera (HySpex VNIR-1800) in the spectral range from 400 to 1000 nm. Because the new illumination system was used for the measurements, only the wavelength range from 430 m to 780 nm was usable. The measurement was done only with dark current correction. Therefore 200 frames were measured with the shutter of the camera closed. For the white referencing, a not coated part of the PET foil was also measured, except for the ITO@Ag@PET foil, for which a part of the sample coated only with silver was used.

The FOV of the camera was 48 cm and the samples was were measured with the parameters in Table 2.

Table 2: Frame rate and roll speed for the single layer measurements.

Sample	Framerate [fps]	Roll speed [m/min]
ITO@Al@PET	31.3	4.05
ITO@PET	31.3	4.05
AlN@PET	31.3	4.05
AlOx@PET	100	6.47

Generation of ground truth data

As ground truth data the nominal thickness of the coated layers were used, because the XRR measurements were not yet completed at the time of the evaluation.

Data selection and preparation

The data sets were divided in a first step into partial data sets with a size of 1000x1800x186 pixels. A Python tool developed by NEO was used for this purpose. In the same step the size of the hypercubes was cut to 1000x1575x110 to remove dark edges on the side of the measured roll (due to shadowing) and to reduce the wavelength range to 430 nm to 780 nm.

The next step was the referencing of the measurements. Only the 'standard' referencing approach was used here. For the samples ITO@PET, AlN@PET and AlOx@PET, the measurement of the uncoated PET film was used, and for the sample ITO@Ag@PET, the measurement of the uncoated Ag was used. The average of the uncoated part was taken in the direction of roll movement (around 600 frames), resulting in a dataset of 1x1575x110 pixel. Subsequently, each line of the sample data set was divided by the white reference obtained in this way. This procedure corresponds to the calculation given in equation 1:

$$I_c(\lambda) = \frac{I_o(\lambda) - I_d(\lambda)}{I_w(\lambda) - I_d(\lambda)} \tag{1}$$

where $I_c(\lambda)$ is the corrected image intensity, $I_o(\lambda)$ the original image intensity, $I_d(\lambda)$ is the dark current recorded with the light source switched off and the lens covered, and $I_w(\lambda)$ is the intensity of the white reference for the wavelength λ. The dark current was already subtracted during the data collection.

The distinct periodic pattern of the roll can be seen in the referenced data for the samples ITO@PET, AlN@PET and AlOx@PET, while for the sample ITO@Ag@PET no pattern is visible, because the Ag layer acts as a mirror.

In the next step, the data for training the prediction models were extracted from the measurement. For this purpose, 500 lines (along the rolling direction) were selected from the centre of each of the layer thickness steps. This was done automatically. A mean value spectrum was then formed from this area with the size 500x1575x110. 1000 random spectra were selected, which together with the corresponding ground truth values form the 'specs' dataset. This data sets were used for further analysis.

Training, optimization and validation of regression models

The obtained 'specs' data sets for the four single layer samples were then used to train, optimize and validate various regression models. All calculations were performed with Python and the scikit-learn library. The necessary scripts were created by ourselves.

For the regression two different algorithms were compared: Partial Least Squares Regression (PLS)[5], and Neural Networks / Multilayer Perceptron (MLP)[6]. A description of the algorithms' functionality, their implementation in the scikit-learn library and the hyperparameters of the algorithms can be found in the respective literature.

The optimization of the hyperparameters of the algorithms was done by the random search method. This means that for n model calculations (here n = 50) random combinations of the hyperparameters of the algorithms are used. A detailed description of the random search method and a comparison with alternative methods of hyperparameter optimization can be found in Bengio et. al.[7] An overview of the optimized hyperparameters for the individual algorithms can be found in the Table 5 in the appendix. In addition to the hyperparameters of the algorithms, the pre-processing of the spectra by the random search procedure was also optimized. This includes a normalization of the spectra and a principal component analysis (PCA, not used for PLS) before the actual model training.

---

[5] https://scikit-learn.org/stable/modules/generated/sklearn.cross_decomposition.PLSRegression.html
[6] https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html
[7] Bergstra, J. and Bengio, Y., 2012. Random search for hyper-parameter optimization. *Journal of machine learning research*, *13*(2).

An important step in the creation of machine learning models is the validation and testing of the models. This involves investigating how well the model can predict the target values of unknown samples, i.e. samples that it did not see in the training step. In the present report, this validation was carried out by dividing the data into a training data set, a validation data set and a test data set. The actual training and variation of the hyperparameters is carried out with the training data set. The quality of the models obtained is examined with the validation data set. The metrics root mean squared error (RMSE) and the coefficient of determination (R^2) are considered. Once the hyperparameter optimisation is complete, a new model is trained with the best hyperparameters found using the data from the training data set and the validation data set, and this model is finally used to predict the test data set and again calculate RMSE and R^2. This two-step procedure is necessary for hyperparameter optimisation, as otherwise an overfitting of the models can occur.

Results of the data evaluation for the measurements of the single layer samples

ITO@Al@PET sample

Figure 9 shows the raster image (PC 1 to 3) of the measurement of the entire sample. The gradations of the ITO layer thickness are well visible and because of the Al layer the inhomogeneity of the roll is not visible.
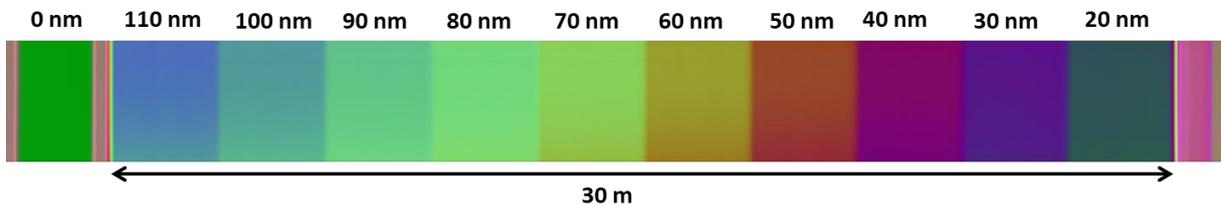
Figure 9: Score image (PCs 1 to 3) of the ITO@Al@PET sample with the nominal thicknesses of the ITO layer

Figure 10 shows the mean spectra (left) and the spectra most similar to the mean spectrum of the respective layer thickness steps. A clear change as a function of the layer thickness can be seen and the noise of the individual spectra is relatively low.
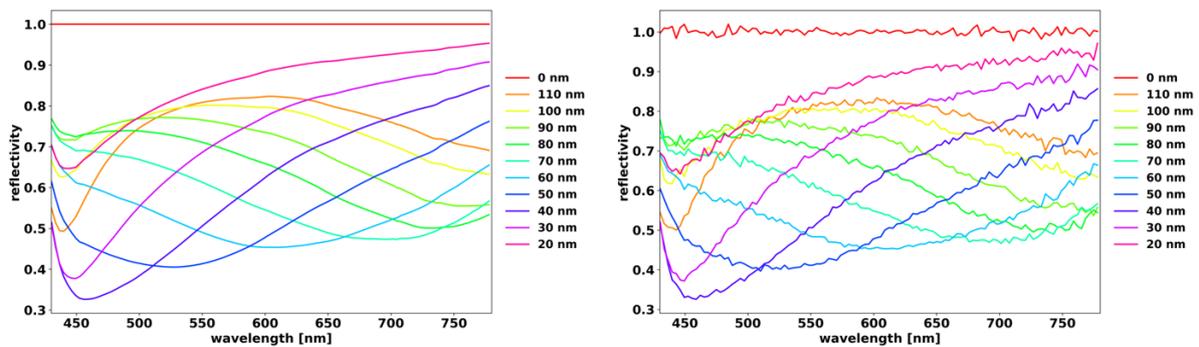
Figure 10: Mean spectra (left) and spectra most similar to the mean spectra (right) for each thickness step of the ITO@Al@PET sample.

Figure 11 shows the model plots of the optimized PLS and MLP regression model. For both models, a good correlation between measured and predicted layer thicknesses can be seen. In particular, an almost perfect fit is obtained for the MLP model. It is possible that an overfitting has occurred in this model.
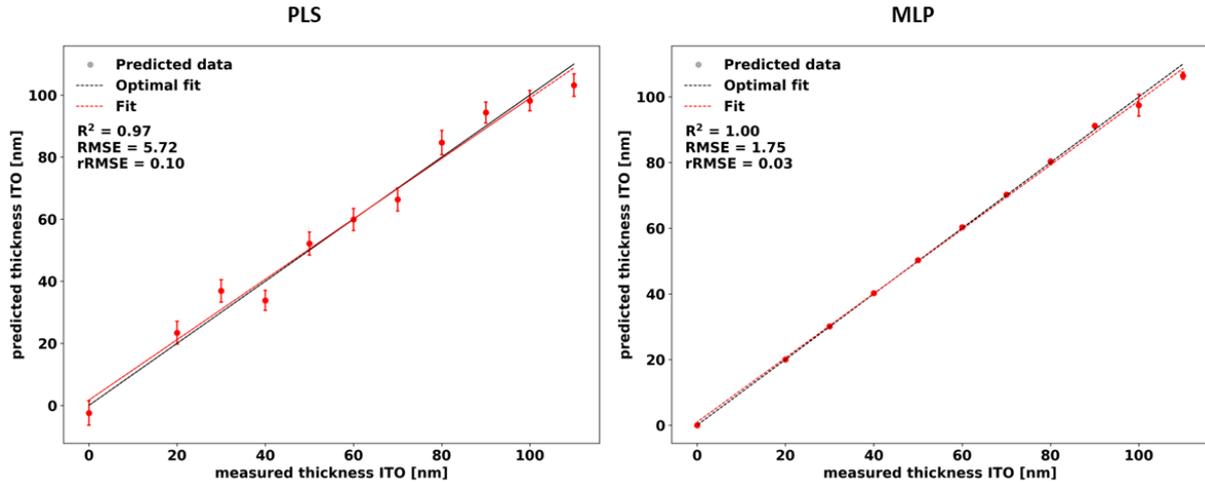
Figure 11: Modell plots of the best found PLS (left) and MLP (right) regression models for the ITO@Al@PET sample calculated with the nominal layer thickness.

Figure 12 shows the predicted ITO thickness for the whole measured sample with the optimised PLS and MLP models. The thickness steps are very well visible for both models. For the MLP model, there is almost no noise in the predicted thickness, which again indicates an overfitting of the model.
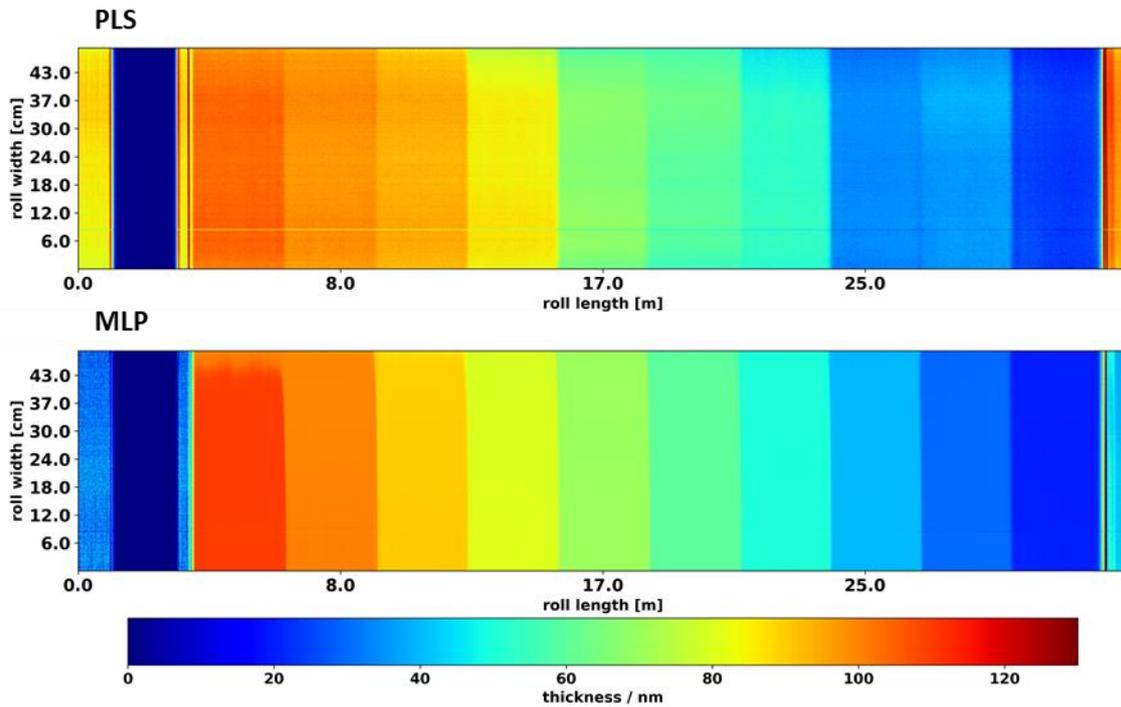
Figure 12: Predicted ITO thickness for the complete ITO@Al@PET sample with the best found PLS (top) and MLP model (bottom).

Figure 13 shows the ITO thickness profiles for the predicted sample using the optimized PLS and MLP sample respectively. The 'left' and 'right' profiles are 5 cm from the edge of the measurement range. It can be seen that the thickness steps for the MLP model are very well resolved and that there is hardly any noise in the calculated thickness (indicating overfitting). For the PLS model, more noise is visible and the calculated layer thickness steps do not exactly correspond to the nominal layer thicknesses in all cases.
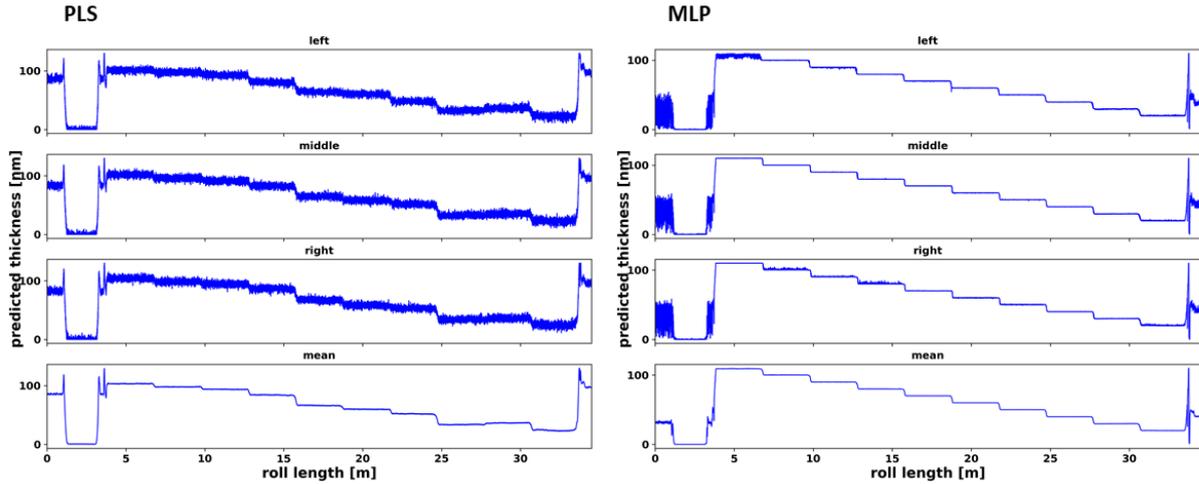
Figure 13: ITO thickness profiles fort he predicted ITO@Al@PET sample with the best found PLS (left) and MLP (right) regression model.

ITO@PET sample

Figure 14 shows the raster image (PC 1 to 3) of the measurement of the entire sample. The gradations of the ITO layer thickness are visible but get less visible with lower layer thickness. The inhomogeneity caused by the role is clearly visible.
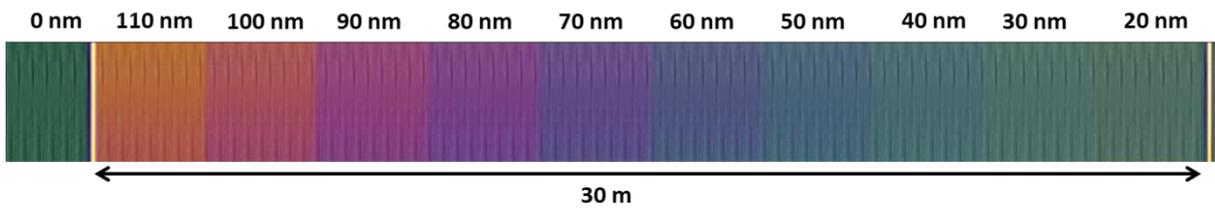
Figure 14: Score image (PCs 1 to 3) of the ITO@PET sample with the nominal thicknesses of the ITO layer

Figure 10 shows the mean spectra (left) and the spectra most similar to the mean spectrum of the respective layer thickness steps. A clear change as a function of the layer thickness can be seen for layer thicknesses above around 60 nm. For the lower ITO thicknesses there is no clear correlation between thickness and spectra visible. Also the noise of the spectra increases this problem.
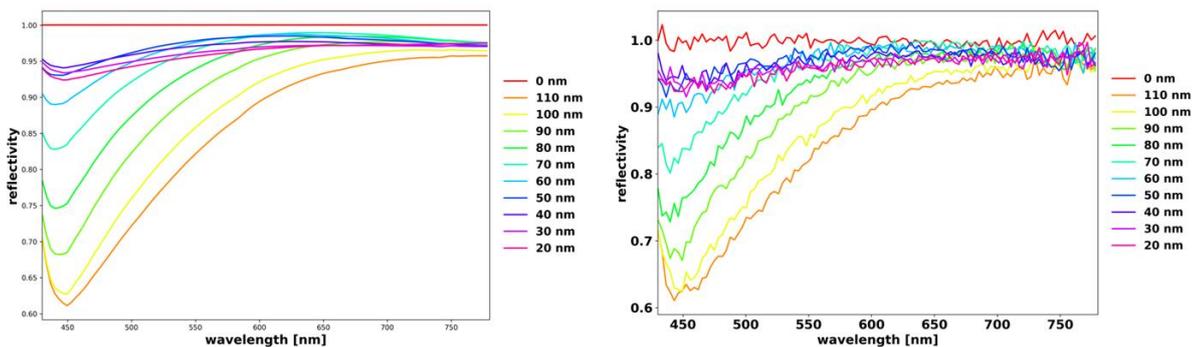
Figure 15: Mean spectra (left) and spectra most similar to the mean spectra (right) for each thickness step of the ITO@PET sample.

Figure 16 shows the model plots of the optimized PLS and MLP regression model. For both models, a correlation between ground truth and predicted values is visible. But especially for thicknesses below around 60 nm the prediction gets worse and also the prediction error increases. The results of the MLP model are better compared to the PLS model. In the PLS model, the largest layer thickness of 110 nm is also not predicted correctly.
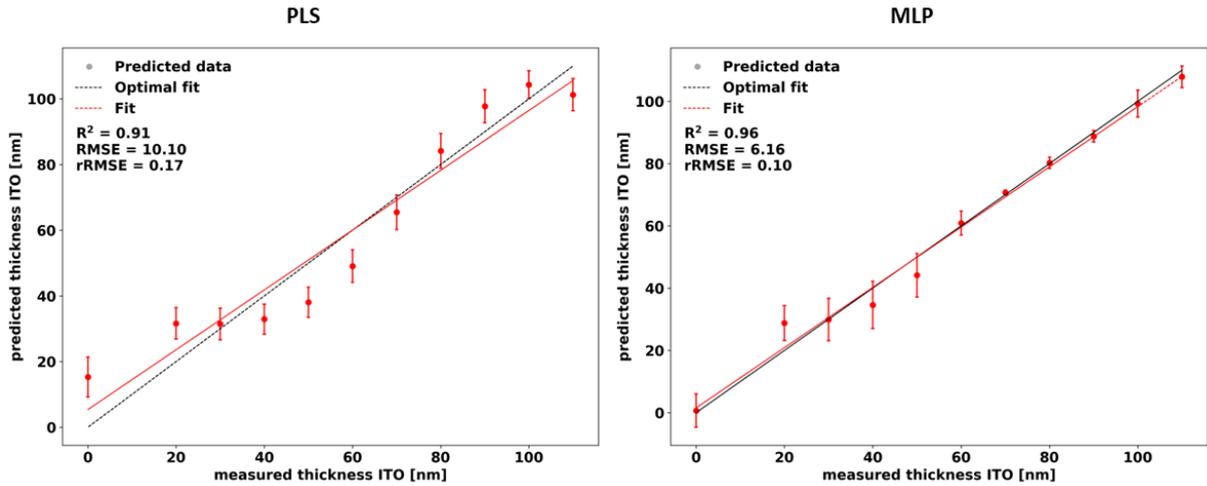
Figure 16: Modell plots of the best found PLS (left) and MLP (right) regression models for the ITO@PET sample calculated with the nominal layer thickness.

Figure 17 shows the predicted ITO thickness for the whole measured sample with the optimised PLS and MLP models. The thickness steps are very well visible for both models for layer thicknesses above around 60 nm and less visible for lower thicknesses. The predicted thickness fluctuations caused by the roll inhomogeneity are clearly visible for both models.
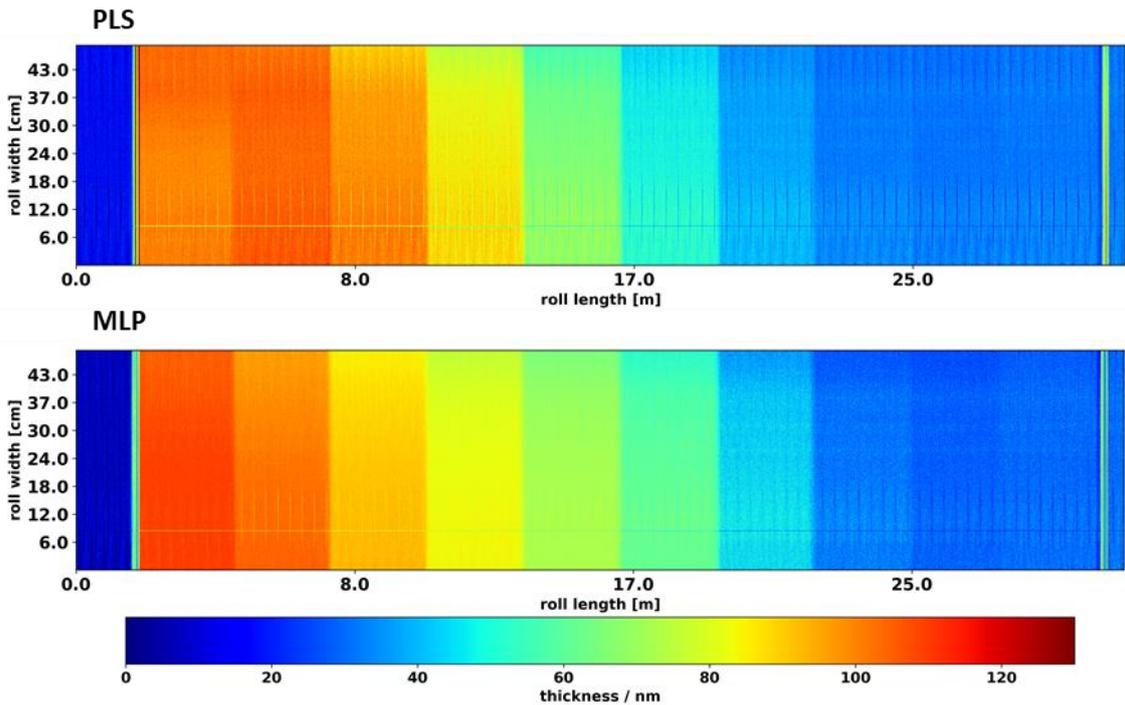
Figure 17: Predicted ITO thickness for the complete ITO@PET sample with the best found PLS (top) and MLP model (bottom).

Figure 18 shows the ITO thickness profiles for the predicted sample using the optimized PLS and MLP sample respectively. The 'left' and 'right' profiles are 5 cm from the edge of the measurement range. It can be seen that the thickness steps for the MLP model are relatively well predicted and show only low noise for thicknesses above around 60 nm. For the lower thicknesses the noise increases significantly and the steps are not well resolved. The PLS model shows a worse and partially incorrect prediction of the thicknesses and significantly more noise for all thickness levels.
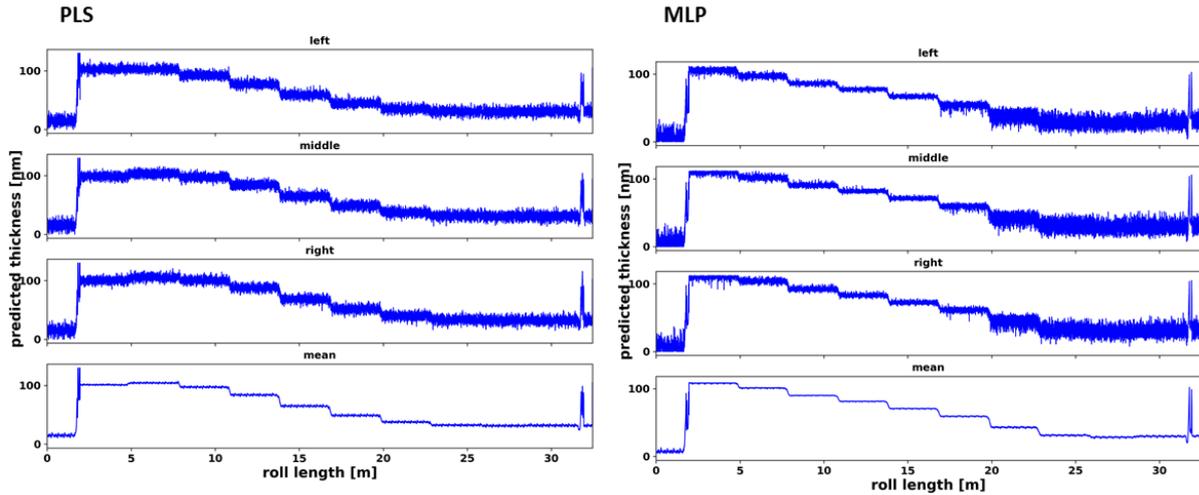
Figure 18: ITO thickness profiles fort he predicted ITO@PET sample with the best found PLS (left) and MLP (right) regression model.

AlN@PET sample

Figure 19 shows the raster image (PC 1 to 3) of the measurement of the entire sample. The gradations of the AlN layer thickness are visible but get less visible with lower layer thickness. The inhomogeneity caused by the role is clearly visible.
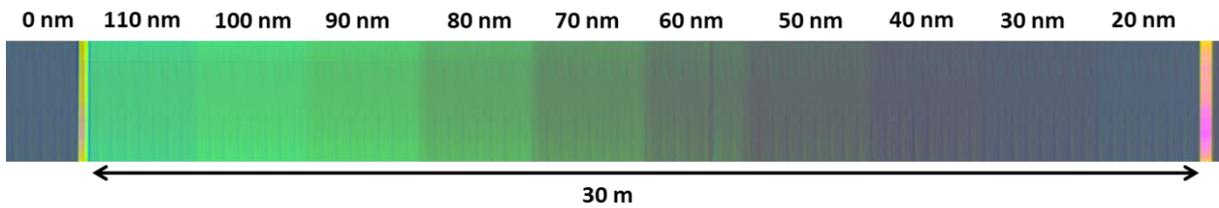
Figure 19: Score image (PCs 1 to 3) of the AlN@PET sample with the nominal thicknesses of the ITO layer

Figure 20 shows the mean spectra (left) and the spectra most similar to the mean spectrum of the respective layer thickness steps. A clear change as a function of the layer thickness can be seen, but the noise of the spectras is relatively high, especially for the lower layer thicknesses. This could make the regression difficult.
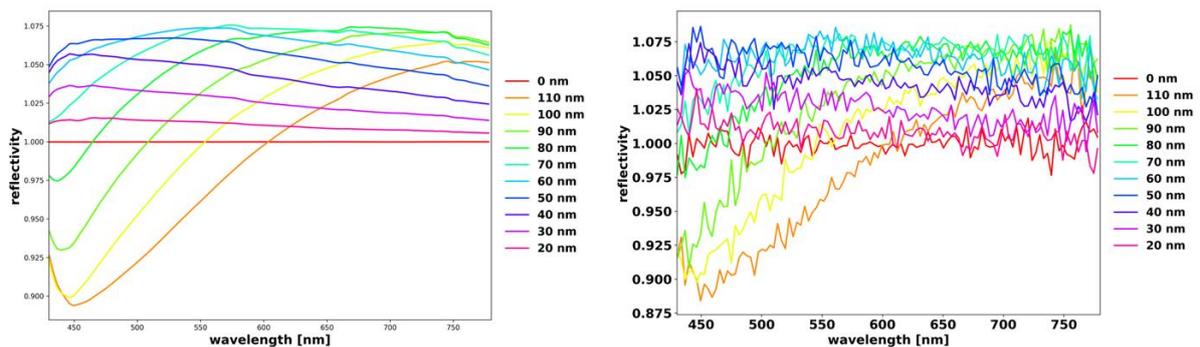
Figure 20: Mean spectra (left) and spectra most similar to the mean spectra (right) for each thickness step of the AlN@PET sample.

Figure 21 shows the model plots of the optimized PLS and MLP regression model. For both models, a correlation between ground truth and predicted values is visible. But especially for thicknesses below around 70 nm the prediction gets worse and also the prediction error increases. The results of the MLP model are better compared to the PLS model. In particular, the very high and very low layer thicknesses are poorly predicted in the PLS model.
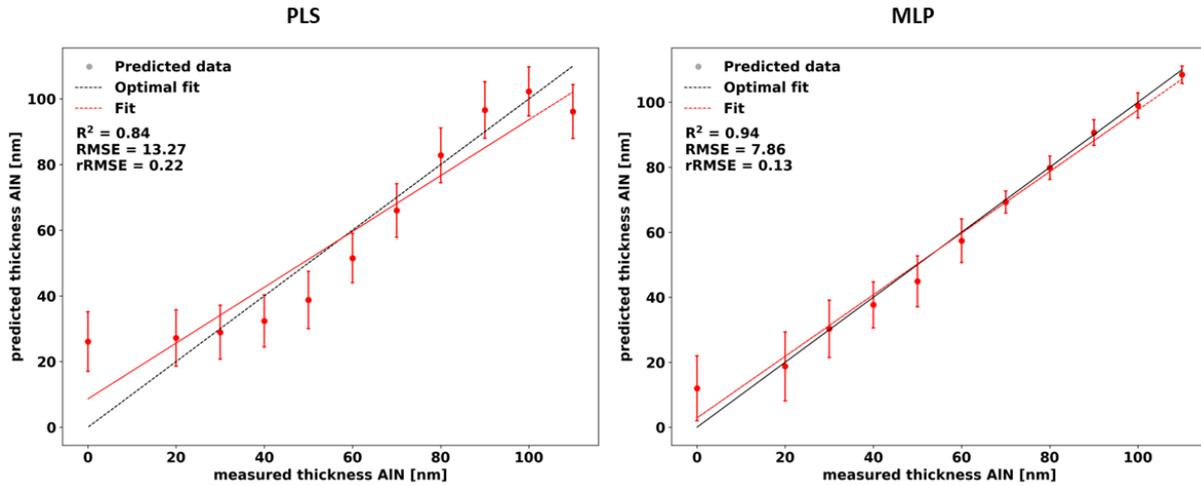
Figure 21: Modell plots of the best found PLS (left) and MLP (right) regression models for the AlN@PET sample calculated with the nominal layer thickness.

Figure 12 shows the predicted AlN thickness for the whole measured sample with the optimised PLS and MLP models. The thickness steps are very well visible for the MLP model. For the PLS model the poor prediction results for the high and the low thicknesses are observable. The predicted thickness fluctuations caused by the roll inhomogeneity are clearly visible for both models.
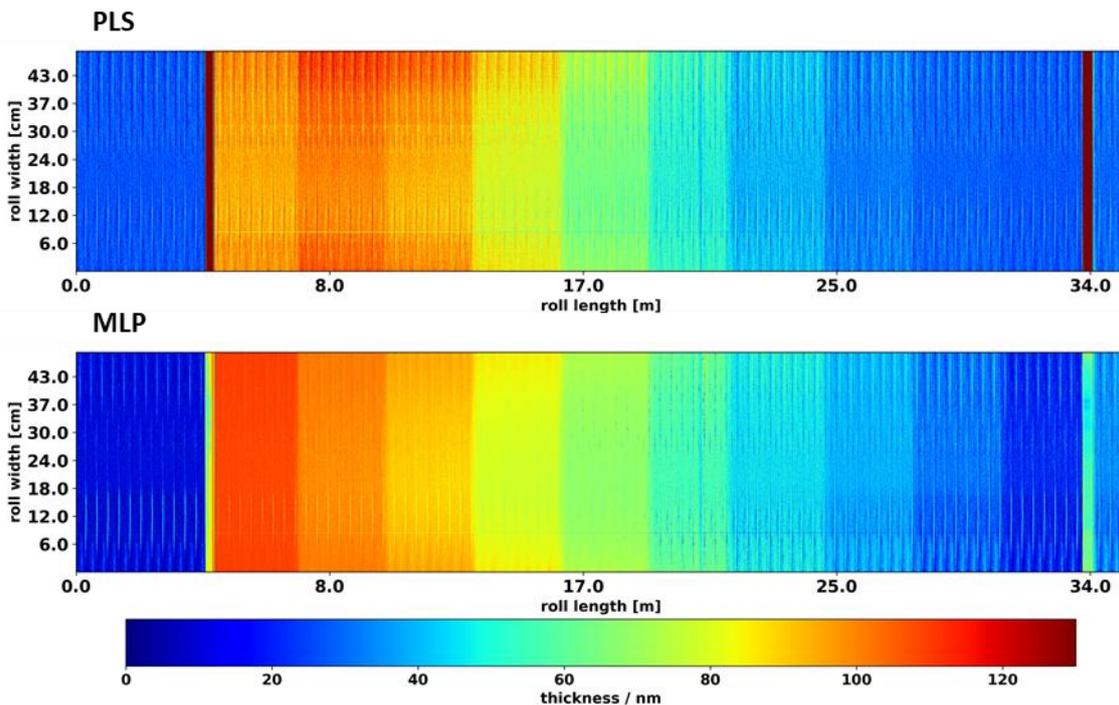
Figure 22: Predicted ITO thickness for the complete AlN@PET sample with the best found PLS (top) and MLP model (bottom).

Figure 23 shows the ITO thickness profiles for the predicted sample using the optimized PLS and MLP sample respectively. The 'left' and 'right' profiles are 5 cm from the edge of the measurement range. For both models, a strong noise of the predicted layer thicknesses can be seen, which is caused in particular by the inhomogeneity of the roll. For the PLS model, the poor prediction for the high and the low layer thicknesses is recognizable. The MLP model gives very good results for film thicknesses above about 70 nm, while the noise of the prediction increases strongly for smaller film thicknesses.
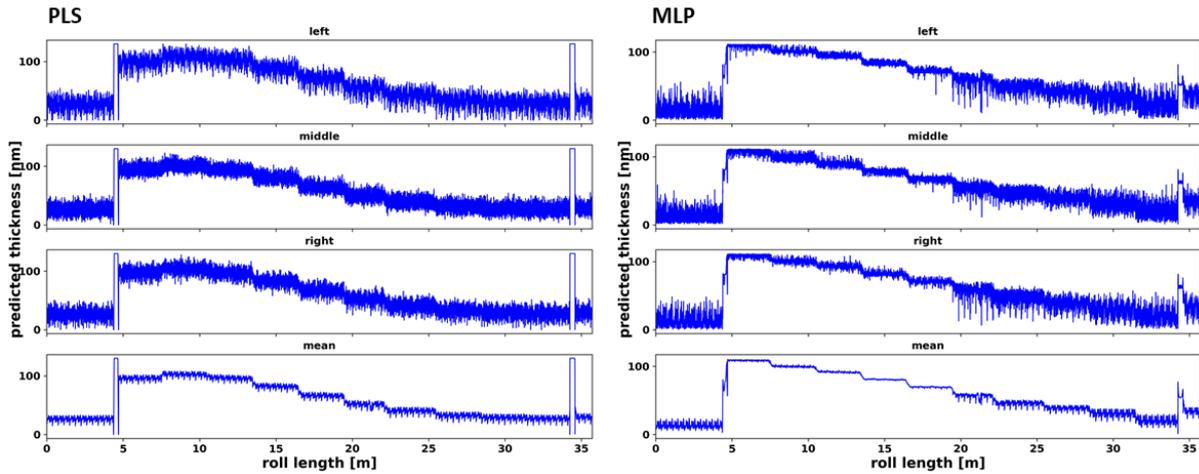
Figure 23: ITO thickness profiles fort he predicted AlN@PET sample with the best found PLS (left) and MLP (right) regression model.

AlOx@PET sample

Figure 24 shows the raster image (PC 1 to 3) of the measurement of the entire sample. The gradations of the AlOx layer thickness are well visible but there is also some inhomogeneity in the thickness steps visible.
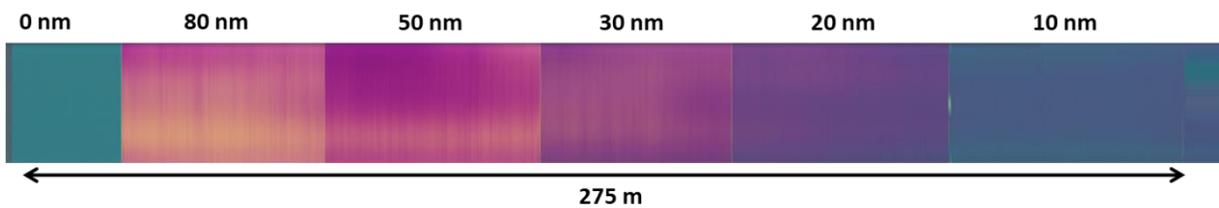
Figure 24: Score image (PCs 1 to 3) of the ITO@Ag@PET sample with the nominal thicknesses of the ITO layer

Figure 25 shows the mean spectra (left) and the spectra most similar to the mean spectrum of the respective layer thickness steps. A clear change as a function of the layer thickness can be seen and the noise of the individual spectra is relatively low.
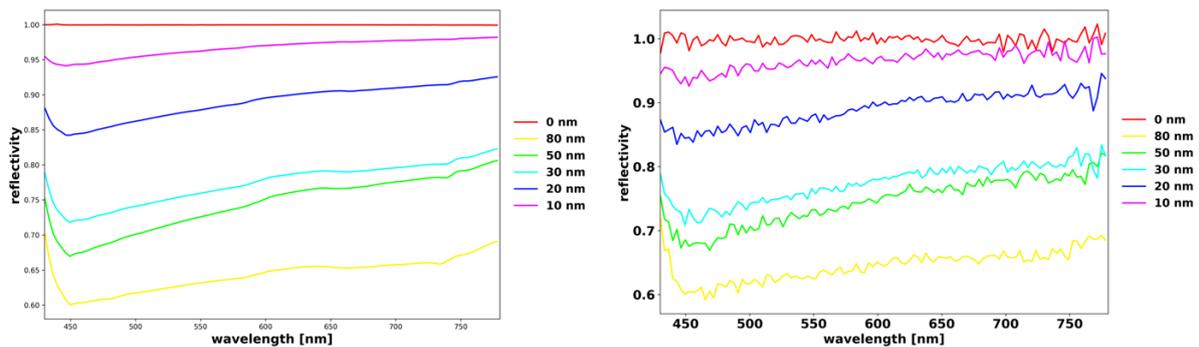
Figure 25: Mean spectra (left) and spectra most similar to the mean spectra (right) for each thickness step of the ITO@Ag@PET sample.

Figure 26 shows the model plots of the optimized PLS and MLP regression model. For both models, a good correlation between measured and predicted layer thicknesses can be seen. For both models, a relatively large variation of the predicted values can be seen for some layer thicknesses, which is possibly caused by the inhomogeneous coating.
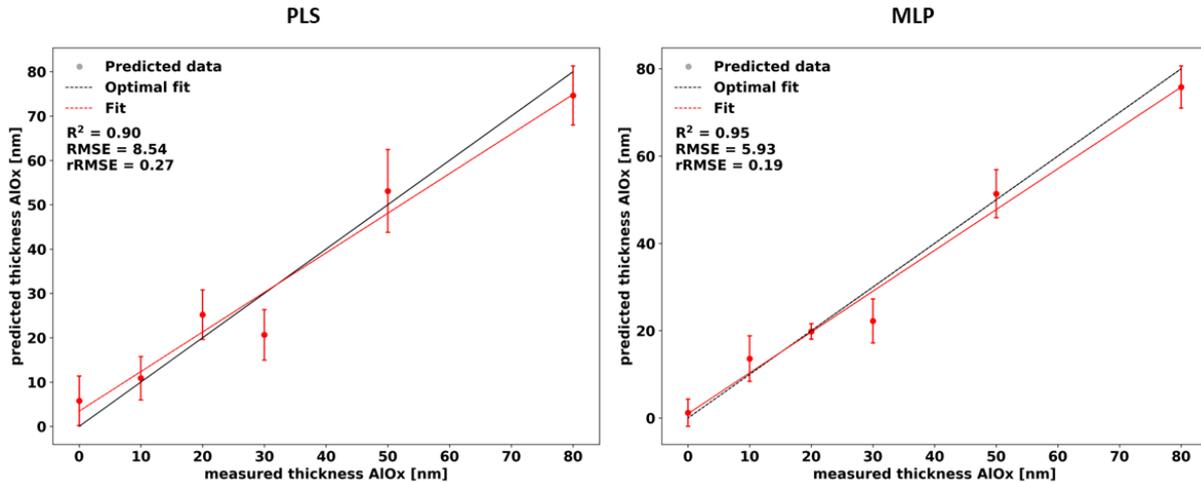
Figure 26: Modell plots of the best found PLS (left) and MLP (right) regression models for the ITO@Ag@PET sample calculated with the nominal layer thickness.

Figure 27 shows the predicted AlOx thickness for the whole measured sample with the optimised PLS and MLP models. The thickness steps are very well visible for both models. In addition to the variations in coating thickness caused by the inhomogeneity of the roll, the areal inhomogeneities of the individual coating thickness steps can also be seen, which may originate from the coating process.
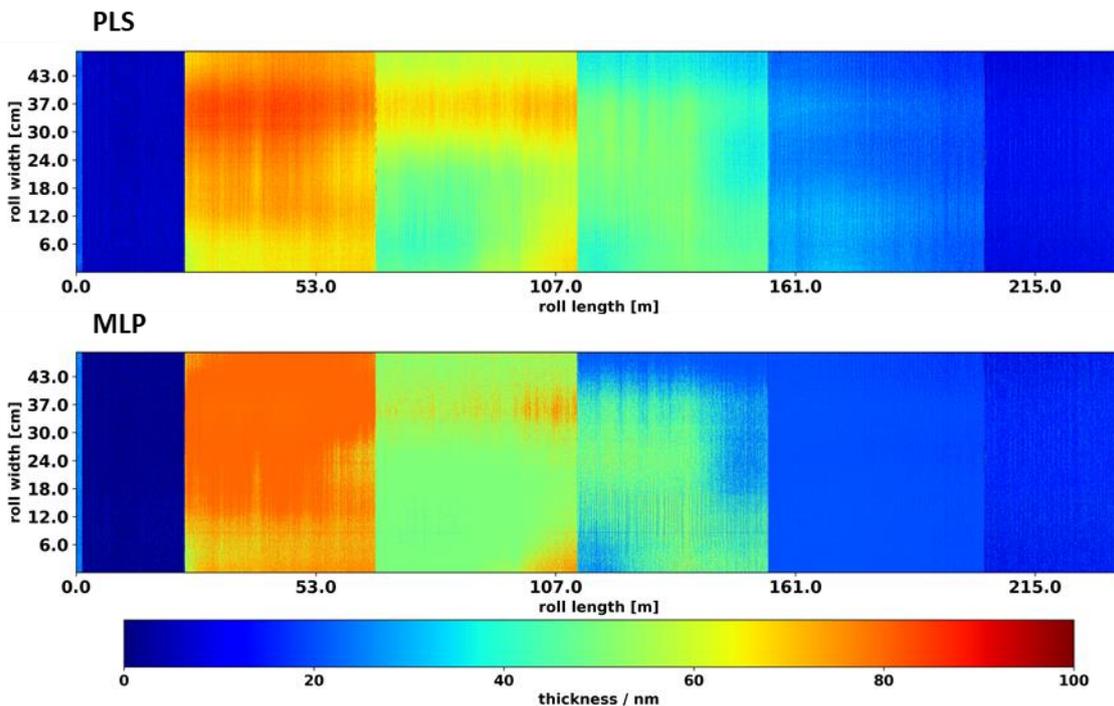
Figure 27: Predicted ITO thickness for the complete ITO@Ag@PET sample with the best found PLS (top) and MLP model (bottom).

Figure 28 shows the ITO thickness profiles for the predicted sample using the optimized PLS and MLP sample respectively. The 'left' and 'right' profiles are 5 cm from the edge of the measurement range. For the PLS model the layer thickness steps are well resolved, but there is a significant noise in the predicted layer thicknesses. It can also be seen that the layer thickness of the individual steps does not appear to be homogeneous along the rolling direction of the sample.

The profile of the MLP prediction shows even more noise and the prediction seems to be much worse. A possible reason is that the thickness steps are relatively inhomogeneous. In combination with the overfitting through the MLP model this leads to a worse model performance compared to the less overfitted PLS model.
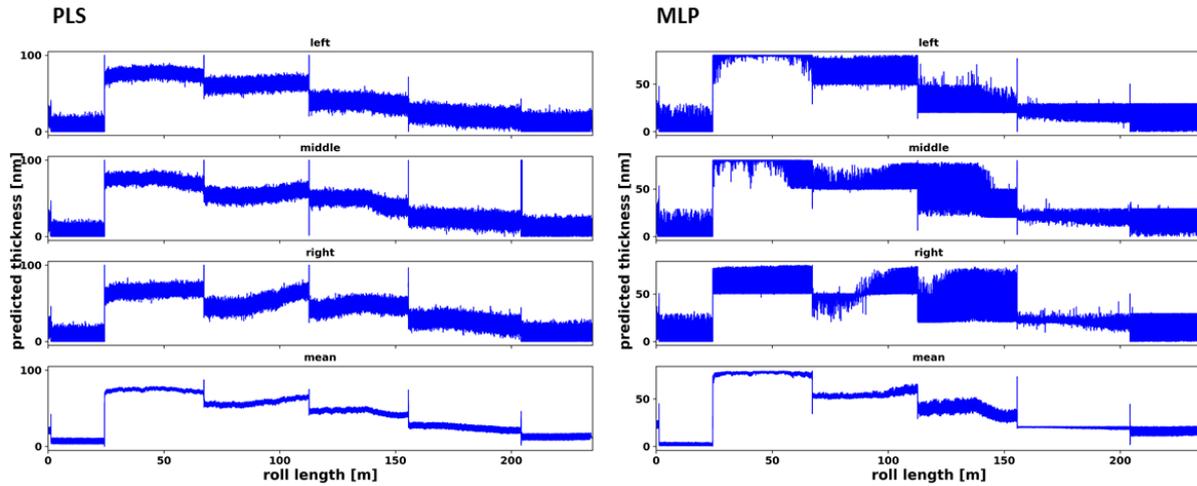
Figure 28: ITO thickness profiles fort he predicted ITO@Ag@PET sample with the best found PLS (left) and MLP (right) regression model.

Comparison of all samples

Table 3 shows the comparison of the results of all the model trainings performed. The table shows the RMSE, the normalized RMSE (nRMSE) and the coefficient of determination (R^2) for all tested samples. The results were determined for the 'specs' dataset and the 'standard' referencing. For the bilayer sample (ITO@ZTO@PET), the wavelength range from 400 nm to 1000 nm was used for the calculation, while the wavelength range from 430 nm to 780 nm was used for the other samples.

The results show that the prediction of the layer thicknesses for the double layer model leads to the worst results (see reasons described above). For the single layer samples, the best results are obtained for the sample with the Al interlayer, since the problem of inhomogeneity of the roll does not occur for this sample and makes the prediction more easy. For the other samples, the results are similar. Differences in the prediction quality result here in particular from the strength of the change in the spectra with the layer thickness. For the AlN@PET sample, for example, only small differences between the spectra of the individual thickness steps are visible, which leads to a comparatively high regression error. It can also be seen that better results are obtained for the MLP algorithm in comparison to the PLS algorithm. The reason therefore is probably the nonlinear relationship between the reflection spectra and the layer thickness. However, since MLP models tend to overfit, these results should be confirmed by further measurements before industrial application.

In summary, complete and high-resolution prediction of the thickness of individual layers on films in an R2R using hyperspectral imaging and machine learning is possible with relatively high accuracy. Further improvements, for example, in terms of signal noise, model stability and homogeneity of the rolls are desirable. In addition, the results should be validated by further, independent measurements. The prediction of two transparent layers on one transparent film (double-layer sample) yielded less good results, but is also difficult from a physical point of view.

Table 3: Comparison of the results of all trained regression models. Shown are the RMSE, the normalized RMSE (nRMSE) and the coefficient of determination (R2) for all tested samples. The results were determined for the 'specs' dataset and the 'standard' referencing. For the bilayer sample (ITO@ZTO@PET), the wavelength range from 400 nm to 1000 nm was used for the calculation, while the wavelength range from 430 nm to 780 nm was used for the other samples.

Model	Algorithm	RMSE	nRMSE	R2
ITO@ZTO@PET ITO	PLS	22.224	0.322	0.64
	MLP	17.572	0.255	0.78
ITO@ZTO@PET ZTO	PLS	23.727	0.340	0.45
	MLP	20.238	0.290	0.57
ITO@Ag@PET	PLS	5.718	0.096	0.97
	MLP	1.753	0.030	1.00
ITO@PET	PLS	10.100	0.170	0.91
	MLP	6.164	0.104	0.96
AlN@PET	PLS	13.271	0.224	0.84
	MLP	7.865	0.133	0.94
AlOx@PET	PLS	8.539	0.271	0.90
	MLP	5.996	0.190	0.95

2.4. Four layer antireflection stack

The task description enclosed the analysis of a four-layer antireflection system at the coFlex 600. After the experience of the single layer and the double layer stacks, this topic was considered of being too complex. The non-specific software approach based on general regression models and training of neural networks is not sensitive enough for a reliable analysis of a four-layer system Moreover, the preparation of reproducible training data is too time consuming for practical purposes.

The general reasons for these challenges are:

- The stability of the signal is influenced by a moving substrate.
- The antireflective coating (corresponding to a high transmittance of the sample) is increasing the influence of the backside reflection of the rotating roller. This critical influence was already commented for the other transparent layers and layer stacks in this report. It is more serious for an antireflection layer since the signal to noise ratio is reduced compared to the other examples.

Based on the results of NanoQI is seems necessary to develop special models for complex layer systems. It seems inevitable to generate training data based on the computer. Additionally, the reliability of the measurements is increased if the perturbation of a target spectrum by deviations of the single layers is investigated, rather than the spectra related to a complete free variation of the four single layers.

This different approach would be beyond the focus of the project. However, the results achieved give a clear indication of the necessary tasks in future industrial installations.

3. Conclusions

The report provided an overview about the validation work done with the HSI installation at the coFlex 600 pilot coater.

The system demonstrated its capability with several single layers and double layer systems on polymer film. The samples were either directly coated on the sputter coater or they were delivered by the project partner Nord from a production line. All layers could be interpreted by the Imanto software. Several fitting procedures were tried. The neural network approach proved to be the most powerful. However, overfitting is still a problem and could not be excluded in all cases.

The models were validated with the help of ground truth data from different sources. Among them were also XRR data. So, the interplay of the analysis software in NanoQI could be demonstrated.

Challenges had been identified for the application of HSI in roll-to roll systems.

- The interaction of the frame rate and the line speed results into a deformation of the field of view. This needs to be taken into account if the analytical results are related to exact positions on the web.

- The moving substrates limits the accuracy of the results, especially if the background reflection is changing as well.

These points pose a limitation for the analysis of complex structures. Therefore, the 4-layer stack was too demanding for the software algorithm as it was tested in NanoQI. A computer-aided generation of training data and an adapted evaluation algorithm seem to be inevitable for such complex solutions.

4. Degree of progress

Degree of fulfilment is 100%.

5. Dissemination level

This Deliverable is Public.

6. Appendix 1

Table 4: Overview of the optimized hyperparameters of the used regression models for the double layer sample.

algorithm	hyperparameter[8]	Value range
All	Standardization	None, L1, L2, SNV[9]
All	n components of PCA	5 - 20
PLS	n latent variables of PLS	5 - 20
EN	Alpha L1 ratio	$10^{-5} - 10^{3}$ 0 - 1
MLP	hidden layer size activation alpha learning rate learning rate limit	100 / 50-50 / 100-1 / 50-50-50 / 20 / 20-20 / 20-20-20 relu, tanh, logistic $10^{-1} - 10^{-7}$ constant, adaptive 0.01, 0.1

Table 5: Overview of the optimized hyperparameters of the used regression models for the single layer samples.

algorithm	hyperparameter[10]	Value range
All	Standardization	None, L1, L2, SNV[11]
All	n components of PCA	5 - 10
PLS	n latent variables of PLS	5 - 10
EN	Alpha L1 ratio	$10^{-5} - 10^{3}$ 0 - 1
MLP	hidden layer size activation alpha learning rate limit	100 / 50-50 / 100-1 / 50-50-50 / 20 / 20-20 relu, tanh, logistic $10^{-1} - 10^{-4}$ 0.01, 0.1

---

[8] For a detailed description of the hyperparameters see the scikit-learn documentation
[9] SNV = standard normal variate correction
[10] For a detailed description oft he hyperparameters see the scikit-learn documentation
[11] SNV = standard normal variate correction