# Regression based polynomial chaos expansion for crop phenology estimation coupled with polsar imagery

## Conference or Workshop Item

# REGRESSION BASED POLYNOMIAL CHAOS EXPANSION FOR CROP PHENOLOGY ESTIMATION COUPLED WITH POLSAR IMAGERY

*M. F. Celik*[1], *O. Yuzugullu*[2], *E. Erten*[1,3]

[1] Faculty of Civil Engineering, Istanbul Technical University, TR-34469 Istanbul, Turkey
[2]Agricircle AG, Rapperswil, Switzerland
[3] Engineering and Innovation, MK7 6AA Milton Keynes, United Kingdom

## ABSTRACT

Crop phenology monitoring using Synthetic Aperture Radar (SAR) data is gaining popularity within the remote sensing community due to SAR's all weather and large coverage imaging capability. This paper introduces a polynomial chaos expansion (PCE) based regression algorithm to retrieve BBCH scale of crops, which identifies the phenology of crops in a standardized system. The impact and applicability of the proposed methodology is successfully illustrated using the TerraSAR-X dual-pol imagery that was acquired over the cultivation period of paddy-rice fields located in Turkey. To assess the applicability of the methodology, root mean square and correlation analysis were performed under different amount of training data and number of inputs.

***Index Terms*—** Polarimetry, SAR, precision agriculture, monitoring, crop phenology, optimization, metamodels

## 1. INTRODUCTION

Polarimetric Synthetic Aperture Radar (PolSAR) satellite measurements offer significant advantages for precision agriculture applications. With the capability of acquiring data independently of solar illumination and cloud coverage, the high sensibility of the radar signal to the geometrical and physical properties of crops, and the richness of information contained in the complex signal, PolSAR data provide a large potential for several agricultural applications. In remote sensing based precision agriculture application, the estimation of growth stage of crops, such as BBCH stage, is of special relevance for the continuous monitoring of fields. Recently, a number of crop growth stage estimation techniques using PolSAR measurements have been reported [1, 2, 3, 4].

In the recent studies, machine learning based regression algorithms have become popular for phenological studies compared to the Radiative Transfer Theory (RTT) based modeling approaches. The ease of their implementation and low computational cost make the machine learning regression algorithms appealing for phenological monitoring. *Learned* or data-driven statistical models do not deal with complex models (e.g. backscattering model for a crop), instead they are more flexible and they define the phenology estimation as a regression problem, but they rely heavily on the available training samples.

In this paper, we discuss the data-driven polynomial chaos based (non-linear) regression for crop phenology estimation. Polynomial Chaos Expansion (PCE), which relies on an assessment of impact of inputs' distribution on outputs' distribution, is well-known technique in uncertainty quantification (UQ). In this context, sparse PCE was used for simplified the complex backscattering function of paddy-rice [5, 6], and it has received much more attention because of its low computational cost and ability to deal with complex problems. Our work aims to expand the applicability of PCE as machine learning regression method, which ignores the physics behind the backscattering. Hence defining an approximate polynomial function between BBCH scale (phenology of crops) and the temporal polarimetric covariance matrix is the aim of the regression based phenological stage estimation. The efficiency of the data-driven PCE in terms of estimating phenology of crops with uncertain parameters from dual-pol covariance matrix will be discussed.

## 2. METHODS AND MATERIALS

### 2.1. Problem Statement

A sufficient static of dual-pol acquisition vector $k = [\mathrm{HH\ VV}]^T$ is estimated by maximum-likelihood method C=$\langle kk^{\dagger}\rangle$ and defined as multi-looked covariance matrix:

$$C = \begin{bmatrix} \langle |HH|^2\rangle & \langle HHVV^{\dagger}\rangle \\ & \langle |VV|^2\rangle \end{bmatrix} \quad (1)$$

where $\dagger$ and $\langle\rangle$ represent complex conjugate operator and spatial averaging, respectively. In (1), diagonal terms are the measurement of backscattering of crops in HH and VV polarization, where off-diagonal terms -which are symmetrical under lexicographic formulation- give information about their correlation and phase difference.

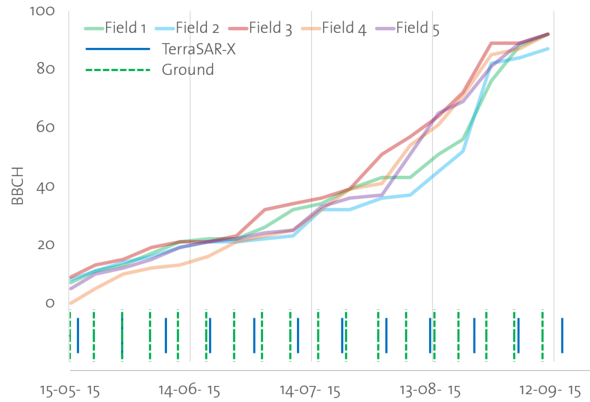


**Fig. 1:** Evolution of the BBCH stage for a complete growth cycle of paddy rice from 2015 ground campaign (May-October) given together with TerraSAR-X acquisition and ground measurement dates.

### 2.2. Least squares polynomial chaos regression

The polynomial chaos expansion metamodel learns the uncertainty of the output of nonlinear system ($Y$) from the uncertainty of the input parameters ($X$) using an approximation of a real model with a multidimensional polynomial expansion. In PCE, multidimensional orthogonal polynomials approximate the $Y$ as in (2):

$$Y = f(X) \cong \sum_{j=0}^{D} a_j \Psi_j(\xi) \tag{2}$$

where $X \in \mathfrak{R}^M$ is the random vector including the covariance matrix elements expressing BBCH scale $Y$, $a_j \in \mathfrak{R}$, $\Psi_j(\xi) \in \mathfrak{R}$ and $D$ are the coefficients, orthonormal basis and degree of the polynomial series with respect to the distribution of $X$ [7, 8]. Once all the coefficients are determined, the BBCH scale of the rice fields can be calculated with its variance from PolSAR image. Additionally, the coefficients can be interpretable compared to the other machine learning techniques.

There are different ways of calculating the coefficients in (2). In this paper, we consider the non-intrusive (regression) method to compute the coefficients of the generated PCE meta-model. If $X$ is a set of $N$ covariance matrix elements (inputs), the main interest is to find the coefficients minimize the sum of quadratic errors as:

$$\arg\min_{a} \sum_{i=1}^{N} \left( f\left(x^{(i)}\right) - a^T \Psi(x^{(i)}) \right)^2 \tag{3}$$

which aims to decrease the differences between the known and the predicted output values. The size and the variance of the training set determine the accuracy of the resulting polynomial model. Polynomial basis and coefficients are constructed using a least-square sense to directly evaluate the covariance matrix elements response to BBCH scale in MATLAB® within the UQLab framework by [9, 10].

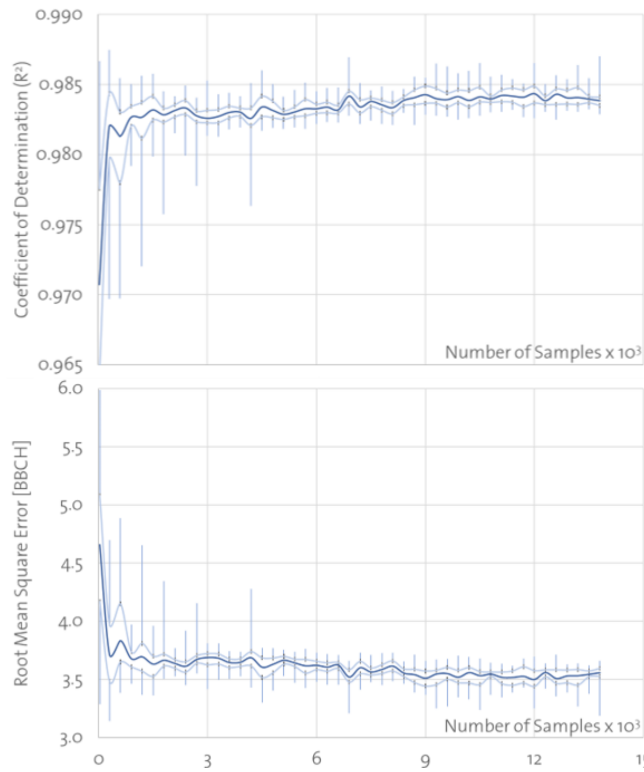


**Fig. 2:** The performance analysis of the PCE metamodel testings for the BBCH parameter for different sample sizes. The plots are prepared using the quantile information of 200 simulations for coefficient of determination and root mean square error.

### 2.3. Experimental Setup

The selected test area, Ipsala, is one of the biggest rice cultivation sites in Turkey with an approximate acreage of ~190 $km^2$. As shown in Fig.1, field campaigns were conducted almost simultaneously with the SAR acquisitions to have a representative rice phenology information. For the analysis, the noise in the TerraSAR-X data was reduced using $13 \times 13$ boxcar averaging windows. Each pixel within the fields is considered as a separate sample for regression analysis and is labeled with a single BBCH value. Thus, the measured BBCH value of the field is defined by a distribution of covariance matrix elements. A total of 5 paddy-rice fields provides a total number of 30000 samples. The results are reported with their coefficient of determination ($R^2$) and the root-mean-square error (RMSE) values.

## 3. RESULTS AND DISCUSSION

Here, we present the PCE based regression results for the BBCH-scale estimation from TerraSAR-X images. The dual-pol covariance matrix elements in (1) and BBCH-scale measurements are used as input ($X$) and output dataset ($Y$) in (2), respectively. Two groups of experiments are designed to evaluate the effectiveness of the proposed method. Firstly, the effect of number of samples is studied, then the PCE is implemented to dataset with setting training sample size fixed. Since the performance of the regression is strongly dependent on the training dataset, 200 experiments (for each plot) are conducted with different training and test dataset.

As the first study, PCE metamodel was implemented several times by taking different sized training sets ranging from 100 to 15000 with 100 sample increments. The outcomes of the final PCE metemodels in their estimation capability of BBCH scale are given in Fig.2. The dark blue line shows the mean value of the simulation results. Besides, light blue continuous lines represents the 25% quartile from the training dataset with different samples. The analysis shows that the PCE metamodel training converges to a stable state around 5000 samples, which results in an $R^2$ of 0.985 and RMSE of 3.6 BBCH unit. Therefore, we have fixed the training sample size to 5000 for the next steps of the analysis.

After performing the regression analysis with different training size, 16.6% of the data (5000 samples) is randomly chosen as a training dataset, and the rest (25000 samples) as a test dataset. The maximum polynomial degree was set to 20 for the PCE training. As presented in Fig.3(b), the BBCH scale is estimated successfully ($R^2 \geq 0.97$) with the polynomial function constructed from the training dataset. Uncertainty of BBCH scale varies in time due to the heterogeneity of the input dataset. The performance of the PCE metamodel is better for rice canopies having BBCH scale less than 30 and more than 80. The higher uncertainty in the reproductive stage (30 < BBCH scale < 70) can be related to the variance of the input variables related to the presence of plants with varying morphological structure within the fields. The RMSE values for training and testing were calculated to be less than 5 BBCH unit, which are acceptable for phenology estimation. Fig.4 shows the BBCH scale map for a selected paddy-rice field, which was estimated with PCE metamodel constructed with the training dataset having 5000 samples. Inspecting these temporal maps, one could check the growing trend on a field-by-field basis in the pixel-level and observe the different stages within fields.

To put in a nutshell, this paper proposes an least square polynomial chaos based regression for phenology estimation of crops. The proposed approach takes the advantage of low computational cost and reliability. However, several issues that need to be further addressed in the future studies. First, Sobol indices can be calculated analytically from the polynomial coefficients. Regarding this, the proposed approach can be used for dimensionality reduction, which can improve the regression performance in terms of computational cost and accuracy [11]. Another aspect to be investigated is related to the application of the PCE. Previously, [12] has also implemented the PCE approach for phenology monitoring, but with the training dataset including the combination of ground measurements and simulated data generated by a electromagnetic scattering model. The proposed regression based (statistical)

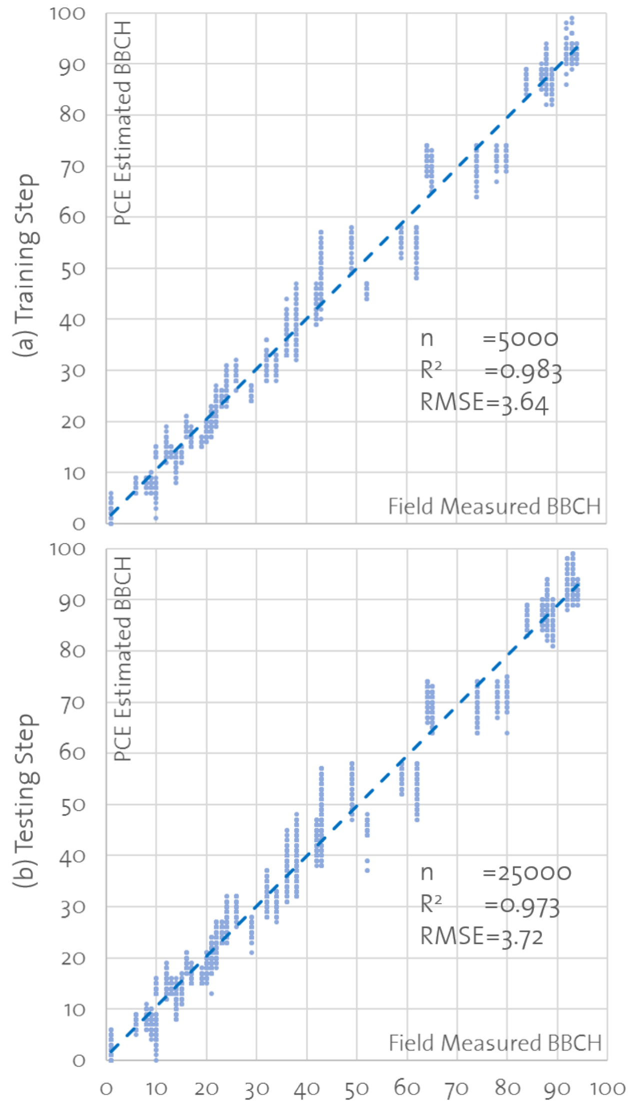


**Fig. 3:** PCE accuracy analysis for BBCH estimation from rice fields with a total number of n=30000 samples.

method should be compared with the this hybrid (statistical and physical) method in the context of computational cost, accuracy and stability.

## 4. ACKNOWLEDGMENT

This work has been supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Project ID 113Y446, by the German Aerospace Center (DLR) under Project ID XTILAND1476.

## 5. REFERENCES

[1] F. Vicente-Guijalba, T. Martinez-Marin, and J. M. Lopez-Sanchez, "Dynamical approach for real-time monitoring of agricultural crops," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3278–3293, 2015.

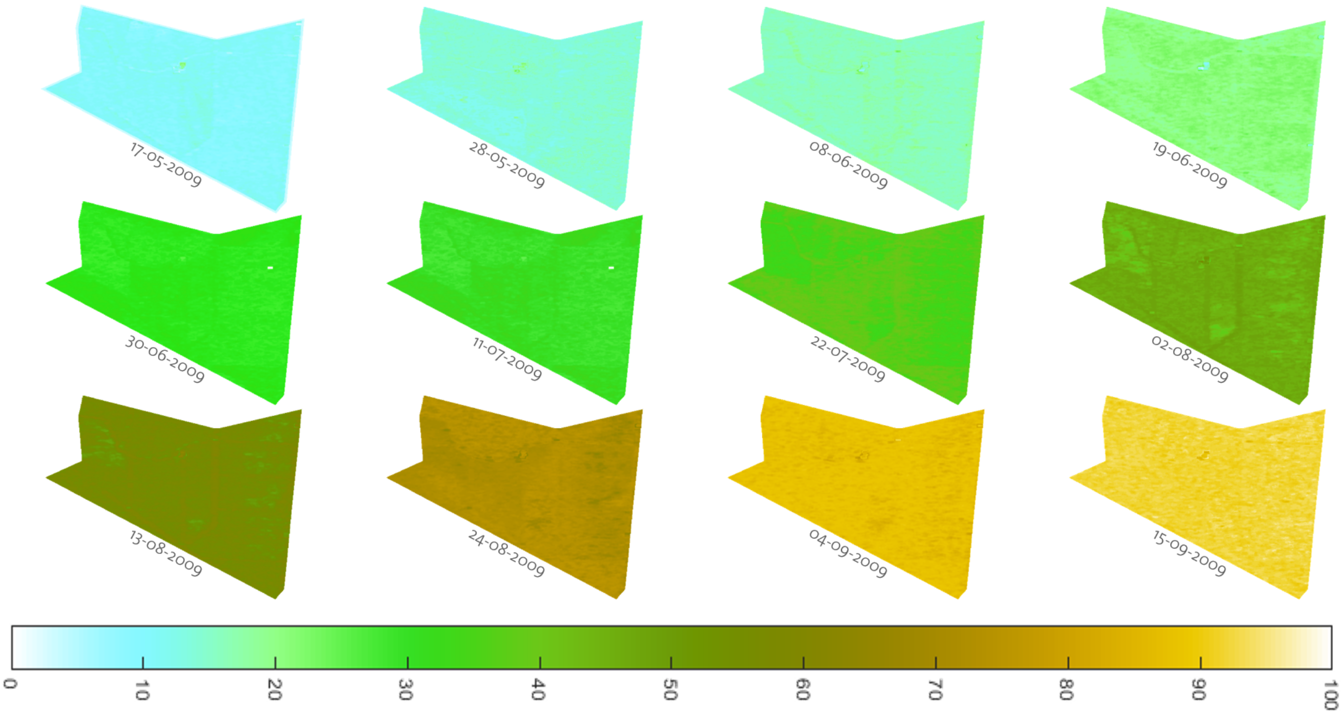


**Fig. 4:** BBCH-scale maps of rice fields with exploiting the temporal behavior.


[2] Zhi Yang, Yun Shao, Kun Li, Qingbo Liu, Long Liu, and Brian Brisco, "An improved scheme for rice phenology estimation based on time-series multispectral HJ-1A/B and polarimetric RADARSAT-2 data," *Remote Sensing of Environment*, vol. 195, pp. 184 – 201, 2017.

[3] Damian Bargiel, "A new method for crop classification combining time series of radar images and crop phenology information," *Remote Sensing of Environment*, vol. 198, no. Supplement C, pp. 369 – 383, 2017.

[4] Caglar Kucuk, Gulsen Taskin, and Esra Erten, "Paddy-Rice Phenology Classification Based on Machine-Learning Methods Using Multitemporal Co-Polar X-Band SAR Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 6, pp. 2509–2519, June 2016.

[5] E. Erten, J. M. Lopez-Sanchez, O. Yuzugullu, and I. Hajnsek, "Retrieval of agricultural crop height from space: A comparison of SAR techniques," *Remote Sensing of Environment*, vol. 187, pp. 130–144, Dec. 2016.

[6] O. Yuzugullu, S. Marelli, E. Erten, B. Sudret, and I. Hajnsek, "Determining rice growth stage with X-band SAR: A metamodel based inversion," *Remote Sensing*, vol. 9, no. 5, pp. 460, 2017.

[7] N. Fajraoui, S. Marelli, and B. Sudret, "Sequential design of experiment for sparse polynomial chaos expansion," *SIAM/ASA J. Uncertainity Quantification*, vol. 5, pp. 1061 – 1085, 2017.

[8] A. Camacho, A. Talavera, A. E. Alexandre, M. A. C. Pacheco, and J. Zanni, "Uncertainty quantification in reservoir simulation models with polynomial chaos expansions: Smolyak quadrature and regression method approach," *Journal of Petroleum Science and Engineering*, vol. 153, no. Supplement C, pp. 203 – 211, 2017.

[9] S. Marelli and B. Sudret, *UQLab: A framework for uncertainty quantification in Matlab*, chapter 257, pp. 2554–2563, 2014.

[10] S. Marelli and B. Sudret, "UQLab user manual - Polynomial Chaos Expansion," Tech. Rep., Chair of Risk, Safety & Uncertainty Quantification, ETH Zurich, 2015, Report - UQLab-V0.9-104.

[11] O. Yuzugullu, S. Marelli, E. Erten, B. Sudret, and I. Hajnsek, "Global sensitivity analysis of a morphology based electromagnetic scattering model," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 2743–2746.

[12] O. Yuzugullu, E. Erten, and I. Hajnsek, "A multi-year study on rice morphological parameter estimation with X-band PolSAR data," *Applied Sciences*, vol. 7, no. 6, pp. 602, 2017.