

## Predicting potato plant vigor from the seed tuber properties

Atza, Elisa; Klooster, Rob; Hofstra, Falko; van der Werff, Frank; van Doorn, Hans; Budko, Neil

**DOI**

[10.1038/s41598-025-13024-0](https://doi.org/10.1038/s41598-025-13024-0)

**Publication date**

2025

**Document Version**

Final published version

**Published in**

Scientific Reports

**Citation (APA)**

Atza, E., Klooster, R., Hofstra, F., van der Werff, F., van Doorn, H., & Budko, N. (2025). Predicting potato plant vigor from the seed tuber properties. *Scientific Reports*, 15(1), Article 28650.  
<https://doi.org/10.1038/s41598-025-13024-0>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.



# OPEN Predicting potato plant vigor from the seed tuber properties

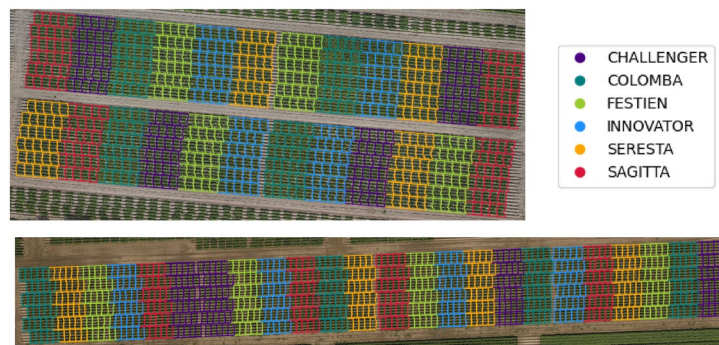
Elisa Atza<sup>1✉</sup>, Rob Klooster<sup>2</sup>, Falko Hofstra<sup>2</sup>, Frank van der Werff<sup>2</sup>, Hans van Doorn<sup>2</sup> & Neil Budko<sup>1✉</sup>

The vigor of potato plants is of crucial importance for potato seed producers, who are interested in predicting it at scale by exploiting the dependence of plant growth and development on the origin and physiological state of the seed tuber. In this article we present the results of a three-year long experiment in which we studied six potato varieties in three test fields. We identify a 73–90% overall correlation in the vigor of plants from the same seedlot grown in different test fields. Similarly, the biochemical tuber data produce plant vigor predictions that correlate up to 70–90% with the measurements. However, these relatively large data and prediction correlations are mostly due to the strong dependence of the seedlot vigor on the tuber genotype. For five out of six studied varieties, variety-specific cross-field and cross-year vigor predictions produce negligible or even negative correlations when the seed tubers and young plants experience environmental stress. At the same time, for the variety that appeared to be less sensitive to environmental stresses, we obtained cross-field and cross-year vigor predictions correlating up to 80% with the measurements. Analysis of individual predictor variables, such as the abundance of a particular metabolite, indicates that the vigor-enhancing properties of the seed tubers are also variety-specific and that the FTIR spectroscopy data is the most reliable predictor.

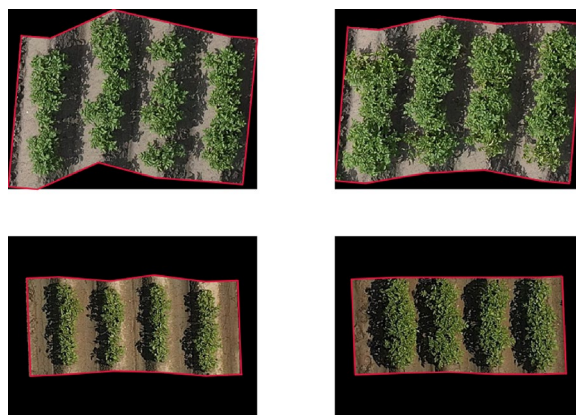
Potato (*Solanum tuberosum*) is a nutritious staple food, popular in many countries of the world, and an important source of starch and proteins. According to the Food and Agriculture Organization of the United Nations, in 2022, almost 375 million metric tonnes of potato were produced globally with the estimated value of 120 billion USD<sup>1</sup>. While potato production is a multi-billion dollar industry, the profit margins of the potato seed producers are relatively low compared to other popular crops, such as tomatoes, onions or lettuce. The reason is that potato is not grown from true seeds but from potato tubers, called seed tubers. These seed tubers are cultivated separately, in clean soil that is free from diseases, and stored in controlled conditions that prevent premature sprouting, loss of moisture, rotting and physiological disintegration of the tubers<sup>2</sup>. This relatively expensive and complicated production cycle is the main reason why some of the technologies common in the seed-based crops are still not available in the potato industry. In particular, there is no standard germination test for potatoes, such as the one, e.g., for barley, where the vigor of a seedlot is ascertained by letting some seeds germinate on a piece of wet paper. Furthermore, a reliable germination test for seed tubers, such as suggested by the National Institute of Agricultural Botany<sup>3</sup>, requires large climate controlled rooms, and lasts several weeks. This approach is costly and logistically hard to implement in the production of seed potato at scale (Fig. 1).

Over the years, producers of the seed potatoes have accumulated qualitative evidence about the lack of consistency in the performance of their produce. This lack of consistency manifests itself as a noticeable difference in the size of the emergent potato plants of the same variety, where some plants appeared to be less 'vital' or 'vigorous' than others. Since the seeds of the same variety are genetically identical, this difference was traced back to the difference in the production field or even the crop management techniques of some farmers. This paper reports on the results of the project 'Flight to Vitality' (FtV) which was the result of cooperation between an EU funding agency, two major seed producers, and two academic institutions in the Netherlands. The main idea of the FtV project is to study the relation between the properties of the seed tubers and the vigor of the resulting potato plants in order to obtain a faster and scalable alternative to germination tests. Exploiting the availability of cheap high-throughput imaging techniques, we investigate the association between the growth of plants in the field and a large set of biochemical data which can be measured at scale prior to planting. The ultimate goal was to build a data-driven model that would help the seed producers and the farmers to rank the seedlots of different production origin by their quality. The setup of this project and a detailed description of the data available are presented in "Methods" section, further technical detail as well as the data necessary to reproduce our results is available in a separate dataset<sup>4</sup>.

<sup>1</sup>Numerical Analysis, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, Netherlands. <sup>2</sup>HZPC Research B.V., Roptawei 4, 9123 JB Metslawier, Netherlands. ✉email: e.atza@tudelft.nl; n.v.budko@tudelft.nl



**Fig. 1.** Drone images of the V (top) and S (bottom) test fields in 2021 with the overlaid plot boundaries. Colors correspond to the six varieties planted in a randomized block design. Each variety block was repeated four times.



**Fig. 2.** Plots of weak (left) and strong (right) seedlots of the Sagitta variety showing consistent performance between the V (top) and S (bottom) test fields.

The size and health of the leaf canopy are the most readily observable features of potato plants. In fact, observations of the difference in canopy size have lead to the qualitative conclusions about the differences in the seed tuber vigor. These features are also relatively easy to monitor with a drone-mounted camera. Finally, the size of the leaf canopy at the time of tuber bulking is indicative of the eventual yield<sup>5</sup>. Therefore, the canopy area was chosen as the vigor parameter in this project (see “Methods” section). This choice also agrees with the accepted definition of the potato seed tuber’s vigor<sup>6,7</sup>. Figure 2 exemplifies the seedlot-dependence of plant vigor, one can see the plots of two Sagitta seedlots showing systematic differences in the canopy size on the chosen date in two test fields.

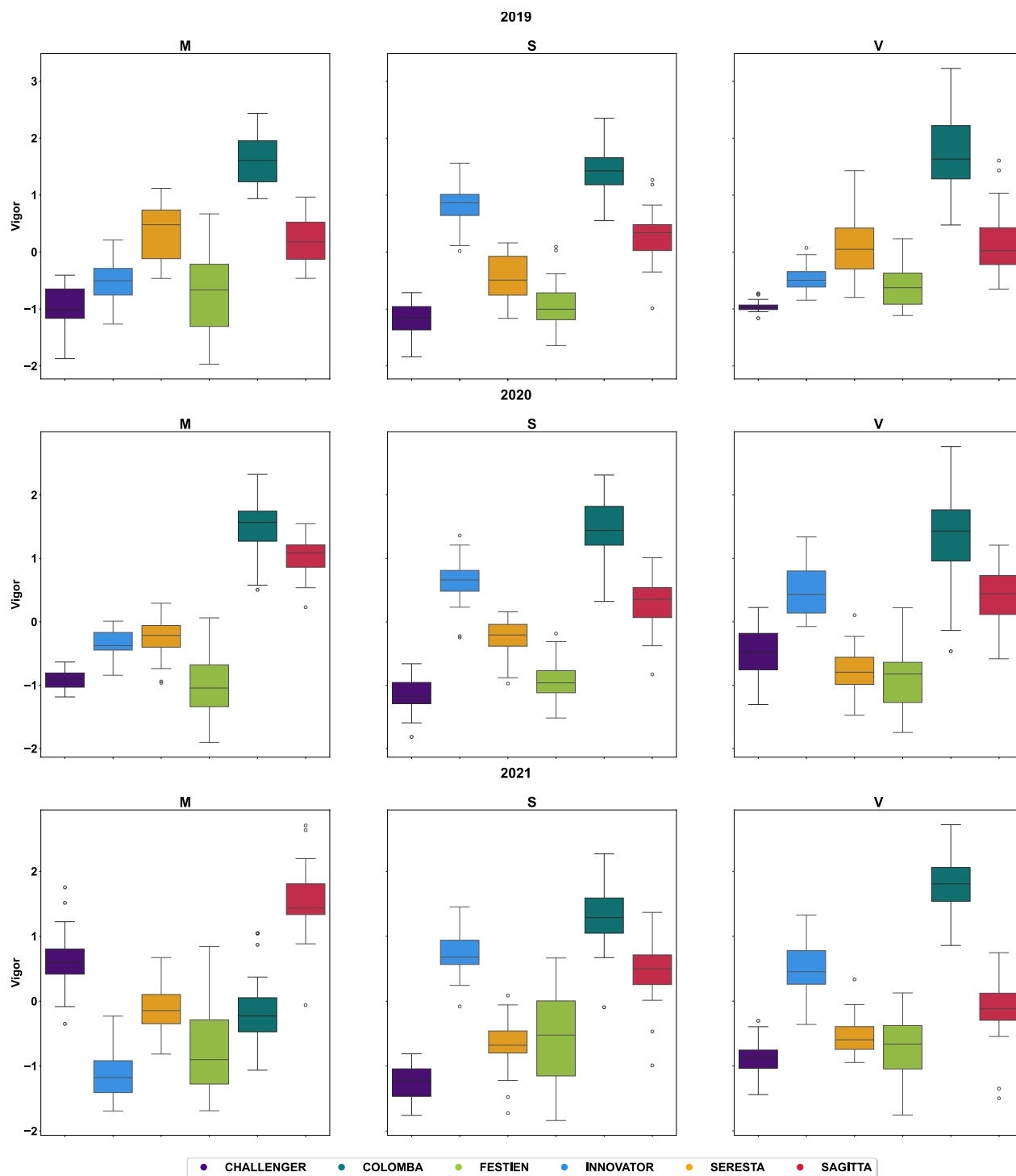
In our previous publication<sup>6</sup> which included partial results of the FtV project, the relation between the microbiome and the vigor of the plants has been considered, and the ability of microbiome to explain up to 20% of the vigor variation within some varieties in the tests ran over two consecutive years (2019 and 2020) has been demonstrated. There is certainly an intimate connection between the microbiome found on the skin and in the ‘eyes’ of the seed tuber and the tuber biochemistry. Some fungal and viral species may also directly affect the growth and development of the potato leaf canopy, e.g., by inducing diseases such as early or late blight, leaf roll, verticillium wilt, etc.<sup>8</sup> However, the presence and abundance of microbiome species on the surface of the tubers is to a large extent determined by the field of production (soil type, treatment, etc.) and represents only an indirect measure of the tuber biochemistry, which is known to affect the early growth stages<sup>9</sup>. In the present paper the focus is on the actual biochemical properties of the seed tubers as provided by the four standard laboratory techniques: X-ray Fluorescence (XRF), Fourier-Transform Infrared Spectroscopy (FTIR), Hyper-Spectral Imaging (HSI), and High-Resolution Mass Spectroscopy (HRMS, untargeted metabolome). These techniques quantify, respectively: the abundance of inorganic elements, the far-infrared spectra pertaining to organic compounds in the dry and wet samples, and the abundance of all detectable small molecules (metabolites).

Our goal is to identify the seed-tuber features associated with the plant vigor. We use the data from three years of field experiments to train and test an interpretable linear model and employ a variable-selection technique to isolate the most predictive features of seed tubers.

## Results

### Consistency of the seedlot vigor

Since each year seed tubers from the same seedlot were planted in three test fields at different locations and times, it was possible to investigate the consistency of the seedlot vigor (canopy cover area at a specific date after planting, see “Methods” section) in a quantitative way. We show the boxplots of the canopy coverage on the chosen dates in Fig. 3, separately for each variety, in order to facilitate the comparison among environments. To measure performance consistency across different environments we chose to look at the Pearson correlation



**Fig. 3.** Boxplots of the vigor parameter (normalized canopy area of a seedlot on a chosen date) in different test fields (M, S, V) over three years (2019, 2020, 2021) showing significant variations in vigor within and between the varieties. Inconsistency of the relative mean vigor of some varieties across fields and years is also visible.



coefficients between vigor data vectors, shown in Fig. 4, with the corresponding p-value in parentheses. There is a consistent positive significant correlation between the full vigor vectors (all varieties) in most of the fields, see the row marked as “All”. Results for the first two years (2019 and 2020) of the experiment were already presented in our previous publication about this project<sup>6</sup>. Here the third year of the experiment (2021) has been included as well and the canopy measurement algorithm has been improved for all three years (see “Methods” section).

The second and especially the third year of the project have demonstrated that, while the correlations for individual varieties are often strong, positive and significant, they can sometimes be affected by environmental conditions and external events. Specifically, in the test field V in 2020, the lack of rain around the emergence time has caused the remaining herbicide to hinder the growth of early-emerging varieties and seedlots, which otherwise would have exhibited a strong vigor. In the test field M in 2021, untypically low temperatures and strong winds around the planting time have delayed the emergence and affected the growth and development of early-emerging varieties and seedlots. This latter event has resulted in insignificant negative correlations between the field M and the other two fields in 2021 when all varieties were considered together Fig. 4 (bottom row), and in the significant negative correlation for the Festien variety (red entry).

The observed canopy size correlations confirm and quantify the qualitative evidence from potato farming practice that seed tubers of the same variety originating from different production fields indeed produce plants of consistently different vigor. It is important to note that the 30 seedlots of each variety are genetically identical. Hence, the systematic differences in vigor cannot be explained by genetic factors. To avoid the seed aging effect, the consistency of the difference between the seedlots was only tested within a single year following the year of production, i.e., it has been decided not to plant the seeds in the years after that.

It is also clear that not all varieties show the same level of correlation between test fields. One variety, namely, Festien, stands out as having the most consistent high correlations (70–90%) when the M field in the year 2021 is excluded (this test field also shows the most significant negative correlation with respect to other fields for this variety). While we also observe high correlations for other varieties, these are not as consistent. Note that only correlations above 50% result in useful predictions.

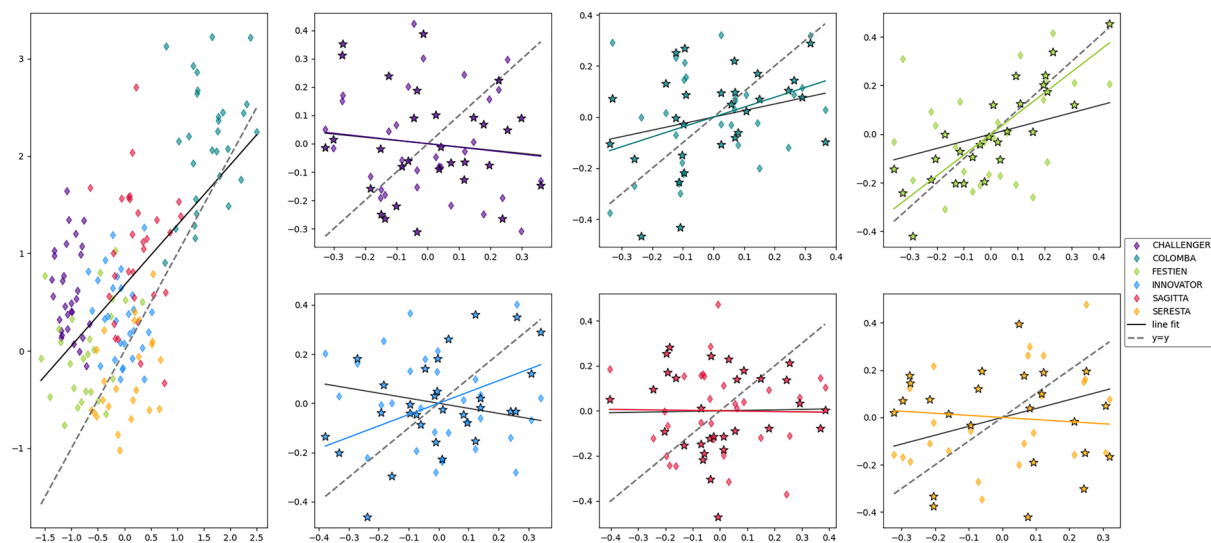
### Genotype dominates the seedlot vigor

The present experiments were not designed to recover the genotype-by-environment interaction, which was also not known *a priori* for the selected set of varieties. Therefore, the available environmental information (solar radiation, temperature, water etc.) could not be used to its full potential, i.e., in a way that would improve the prediction quality within the specific growth conditions. In these circumstances one can hope to predict the average vigor of a seedlot over several conditions, i.e., over several fields within the same test year. However, this strategy only works if the conditions are not too different between the fields. Since the V field in 2020 and the M field in 2021 did suffer from adverse growth conditions around the emergence times, these fields are considered to be statistical outliers and the corresponding data are treated with extra caution (Figs. 5, 6, 7, 8, 9, 10).

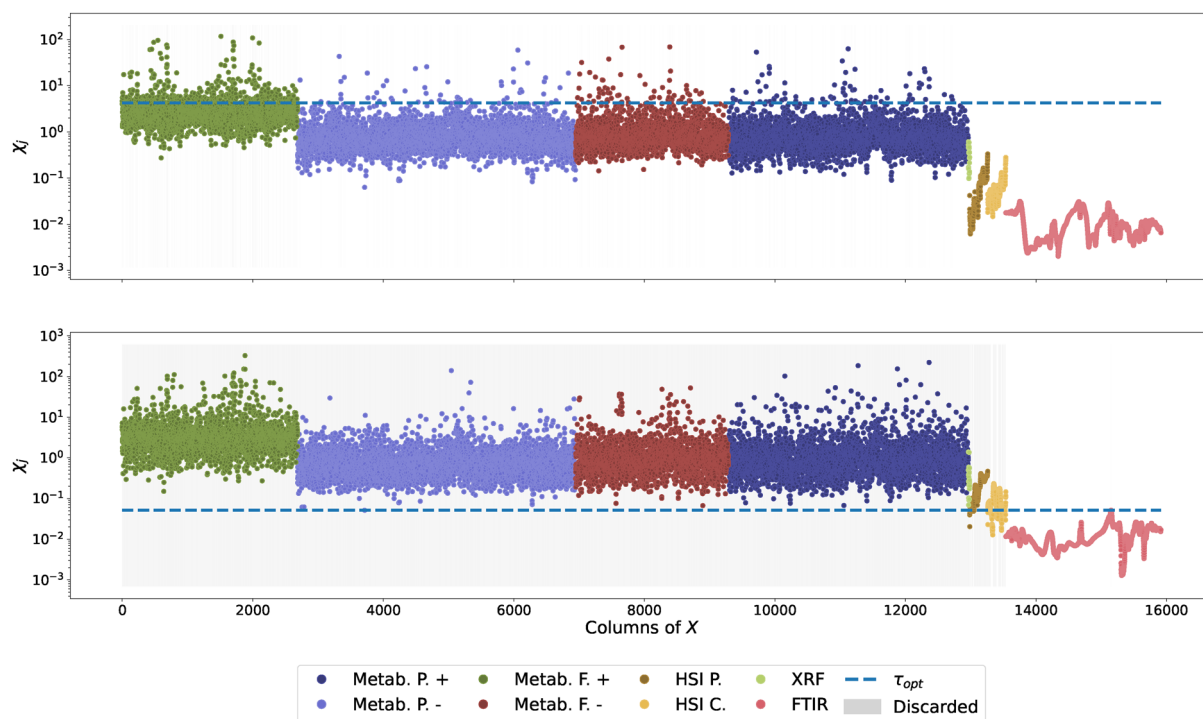
To predict the vigor of a plant from seed tuber properties, we follow the standard model validation and testing procedure which requires a strict separation between the data used for training and the data used for testing. Since the present experiment spanned three years and three fields in each year, there are many possibilities to train and to test the model. For completeness, all possible training-testing possibilities have been implemented

		M-S	M-V	S-V
2019	CHA	0.64 (0.000)	0.17 (0.365)	0.14 (0.461)
	COL	0.67 (0.000)	0.44 (0.016)	0.69 (0.000)
	FES	0.80 (0.000)	0.90 (0.000)	0.79 (0.000)
	INN	0.84 (0.000)	0.78 (0.000)	0.71 (0.000)
	SAG	0.56 (0.001)	0.63 (0.000)	0.51 (0.004)
	SER	0.80 (0.000)	0.69 (0.000)	0.57 (0.001)
	All	0.73 (0.000)	0.90 (0.000)	0.73 (0.000)
2020	CHA	0.77 (0.000)	0.76 (0.000)	0.89 (0.000)
	COL	0.79 (0.000)	0.73 (0.000)	0.77 (0.000)
	FES	0.79 (0.000)	0.77 (0.000)	0.73 (0.000)
	INN	0.67 (0.000)	0.38 (0.037)	0.52 (0.003)
	SAG	0.72 (0.000)	0.55 (0.002)	0.81 (0.000)
	SER	0.70 (0.000)	0.47 (0.008)	0.45 (0.012)
	All	0.82 (0.000)	0.76 (0.000)	0.87 (0.000)
2021	CHA	-0.10 (0.610)	0.04 (0.813)	0.22 (0.243)
	COL	0.50 (0.005)	0.44 (0.015)	0.49 (0.006)
	FES	-0.54 (0.002)	-0.15 (0.439)	0.71 (0.000)
	INN	0.62 (0.000)	0.67 (0.000)	0.58 (0.001)
	SAG	0.60 (0.000)	0.60 (0.000)	0.70 (0.000)
	SER	0.06 (0.753)	0.32 (0.088)	0.59 (0.001)
	All	-0.10 (0.196)	-0.11 (0.160)	0.86 (0.000)

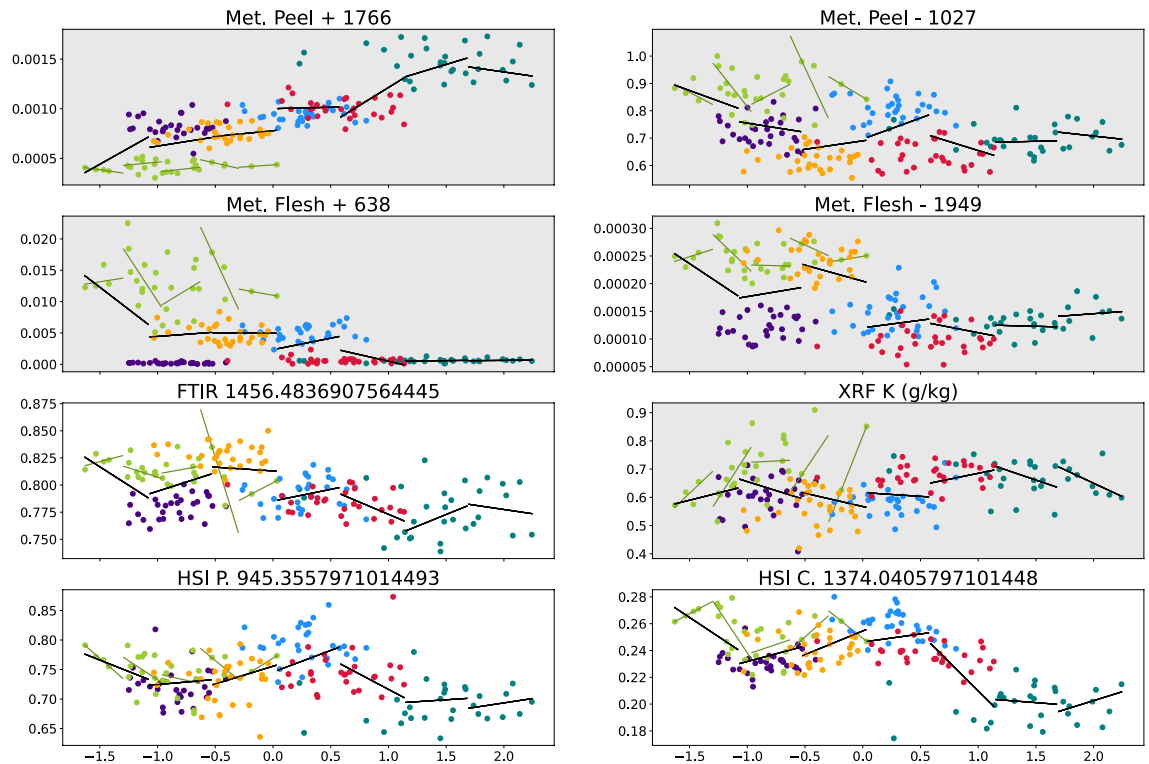
**Fig. 4.** Observed correlations (Pearson) in seedlot vigor between the three fields (M, S, and V) in each year. The rows show the correlations for each variety as well as for all varieties together (‘All’). Dark-green background highlights large significant positive correlations, red background indicates the significant negative correlation. The corresponding p-values are displayed in brackets, insignificant correlations, i.e., with p-value larger than 0.05, are displayed on a gray background.



**Fig. 5.** Performance of the variety-agnostic and variety-specific models trained on the dataset of the field S in 2021 and tested on the average vigor data of the fields M, S, and V in 2019. Horizontal axis—measured seedlot vigor  $y$ , vertical axis—predicted seedlot vigor  $\hat{y}$ . Leftmost vertical panel: variety-agnostic model with the seedlot results colored according to their variety. Dashed gray: the  $y = \hat{y}$  line; solid black: the best linear fit between the measured and predicted data. Six panels on the right: prediction per variety with the variety-agnostic model (diamonds) and the variety-specific model (stars) with the variety vigor data (both measured and predicted) shifted and scaled to have zero mean and unit norm. Dashed gray: the  $y = \hat{y}$  line; solid black: the best linear fit for the variety agnostic model; solid colored: the best linear fit for the variety-specific model.



**Fig. 6.** The  $X$ -data residuals for the variety-agnostic (top) and variety-specific (bottom) 2020 models tested on 2021 Festien tuber data. Colors indicate the data type (see legend). Dashed horizontal lines show the feature selection thresholds, with the rejected predictors marked by vertical gray lines.



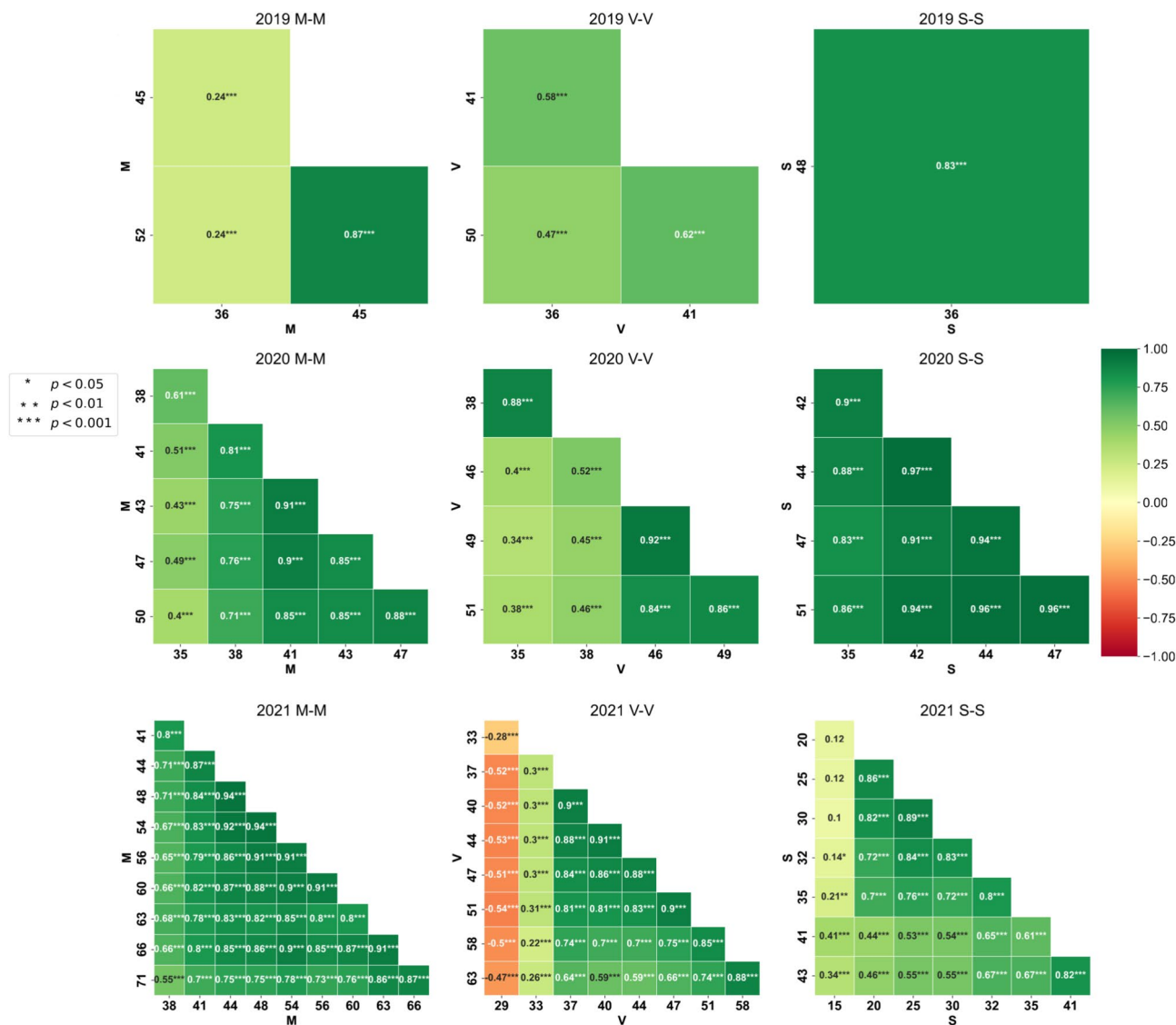
**Fig. 7.** Individual predictor data (vertical axes) fitted by piece-wise linear functions of the vigor parameter (horizontal axis). For each data type we show the predictor with the lowest residual in the Festien specific model. The training was performed on the average vigor of the three test fields and the resulting fits by the variety-agnostic model are shown as broken solid black lines. The fits produced by the variety-specific Festien model are shown as broken light-green lines. The data type is indicated in the titles of the plots together with the predictor label. Plots with the light-gray background show predictors that have been rejected by the variety-specific Festien model.

and the results are shown as tables in Figs. 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, where the rows correspond to the training dataset and the columns show the outcome of the testing. All the data used to obtain these results as well as model performance metrics for all trained models are available in a separate dataset<sup>4</sup>. Note that the block-diagonal portions of each table correspond to testing on a randomly selected portion of the dataset that was excluded from the training process. This means that the testing within the same year (diagonal blocks) was performed on a smaller dataset than across the years (off-diagonal blocks) and is therefore less reliable.

To obtain robust regression results, we opted for an innovative model<sup>10</sup> which allows choosing a mathematically optimal representation of the features so that the diverse tuber data could be handled simultaneously and on equal footing. As Fig. 7 illustrates, metabolomic measurements and spectral absorbancies are very different when viewed as functions of vigor (Inverse Regression approach<sup>10</sup>), with the latter showing a continuous dependence, whereas the former exhibiting jumps and extended zero-filled segments. To solve this problem, we employ a piecewise-linear Discontinuous-Galerkin representation of the features (see “Methods” section) which accommodates for both behaviors. In all other respects the employed regression method is mathematically equivalent to the classical Multiple Linear Regression, see<sup>10</sup>.

Results presented in Fig. 11 correspond to the variety-agnostic model which does not code explicitly for the known variety of the seedlot, and both the SSE residual (top table) and the correlation of the predicted and measured vigor (bottom table) are shown. It turns out that the data from 2020 are most suited for the training of the linear model as it results in the best predictions (smallest residuals and highest correlations) for years 2019 and 2020. Indeed, the variety-agnostic models trained on data collected in 2020 produce predictions that highly correlate with the measured vigor within the same year, with a minimum correlation of 65% ( $p$ -val 0.000) and a maximum correlation of 93% ( $p$ -val 0.000). Models trained on 2020 data also perform very well when tested on other years. Predictions for 2019 correlate to the measured vigor with a minimum correlation of 48% ( $p$ -val 0.000) and a maximum correlation of 80% ( $p$ -val 0.000), whereas predictions for 2021 correlate to the measured vigor with a minimum correlation of 61% ( $p$ -val 0.000) and a maximum of 86% ( $p$ -val 0.000). For each trained model, we provide results and respective  $p$ -values in the external dataset<sup>4</sup> in the “Regression results” folder. These correlations between the predicted and measured vitalities agree with the magnitude of the observed correlations (see Fig. 4, rows marked ‘All’) in vigor across the fields in each year.

However, these promising regression results mask the fact that the average vigor (canopy size) of the six considered varieties is markedly different, with the canopy of the ‘weakest’ variety Festien (around  $-1$  on the



**Fig. 8.** Kendall's  $\tau$  coefficient of the canopy sizes between the measurement dates (DAPs) in each test field and year. In this case, the Kendall  $\tau$  coefficient measures the stability of the seedlot order over three quantiles (low, middle, high) and would be equal one, if the seedlot canopies did not change their quantiles between the dates.

normalized field-centered scale) being almost three times smaller than that of the 'strongest' variety Colomba (around 2 on the normalized field-centered scale). Therefore, mathematically speaking, it is easy to achieve a relatively small residual by simply predicting the average vigor of each variety rather than the true vigor of the seedlot.

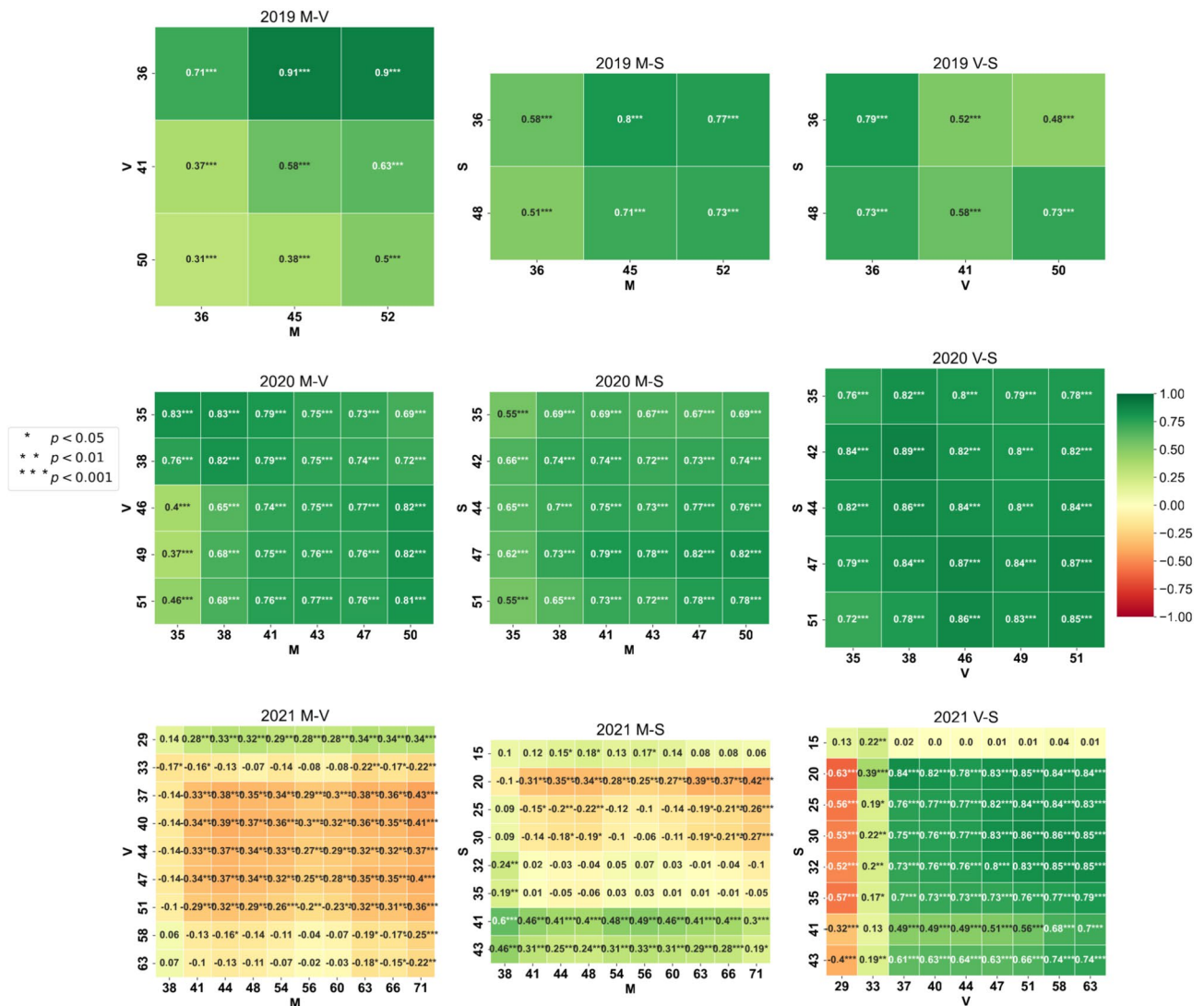
The fact that the genotype (variety) effect dominates the regression results can be deduced from the scattered plots of Fig. 5, where the vigor prediction of each seedlot is colored according to the corresponding genotype. It is obvious that, with some exceptions, the quality of predictions of the seedlot vigor within each individual variety is rather poor.

### Some varieties are more predictable than others

Since in practice farmers plant seed tubers of a single variety in the field, it is important to know how well one can predict the vigor of a seedlot within its specific variety since less vigorous seedlots in the delivered seed tubers will result in inhomogeneous growth. The assumption behind the variety-agnostic model is that the biochemical mechanisms resulting in a weak or a strong vigor are the same for all potato varieties. However, it is also possible that these mechanisms depend on the variety. Therefore, in addition to the variety-agnostic model we also trained six separate variety-specific IR models, the results are presented in Figs. 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 next to the corresponding segregated results of the variety-agnostic model.

From Figs. 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, it is clear that the quality of predictions within the varieties is much lower than across the whole set of varieties, Fig. 11, which confirms the strong genotype effect on the vigor of seedlots. Apart from the statistically less reliable within-the-year results (diagonal blocks in the





**Fig. 9.** Pearson's correlation coefficient for the canopy sizes on different dates (DAPs) between the three test fields in each year.

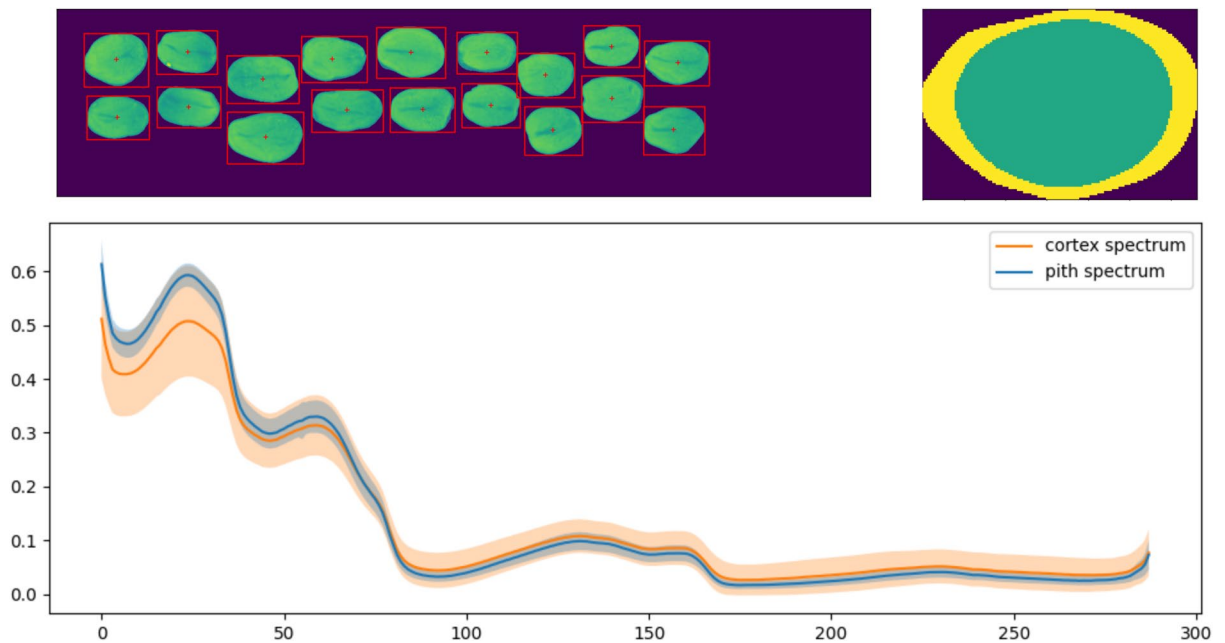
tables), for four of the six varieties (Challenger, Innovator, Seresta, Colomba) neither the variety-agnostic nor the variety-specific models are able to explain intra-variety variations in vigor when presented with the tuber data of a different year. Although, the correlations between predicted and measured vigor do fall just a little short of 50% for these varieties, this corresponds to residuals with magnitude above one, i.e., such predictions are practically useless.

The notable exception is the Festien variety, which exhibits a fairly predictable behavior, see Figs. 16, 17, with the variety-specific model showing somewhat better performance if compared to the variety-agnostic model. The variety-specific Festien model trained on tuber data of the year 2020 and the average vigor data of the M and S fields in 2020 produces predictions which highly correlate with the three-field average vigor in 2019 (Pearson correlation coefficient 65%,  $p$ -val 0.000) and with the three-field average vigor in 2021 (Pearson correlation coefficient 63%,  $p$ -val 0.000), when presented with the corresponding seed tuber data.

In the second place is the Sagitta variety, which shows a similar partially predictable behavior between the years 2020 and 2021, see Figs. 20, 21, 22, 23. However, this predictability does not extend to the year 2019.

### Predictive power of different tuber data types

The IR model provides a measure of the usefulness of each particular predictor variable<sup>10</sup>. When a new predictor-variable dataset (tuber properties) is considered for prediction purposes, the IR method can perform a 'prediction' of this known dataset by the model learned on the training dataset of tuber properties. For example, the dataset of tuber properties in the year 2021 can be predicted from the dataset of tuber properties in the year 2020. The quality of the tuber data prediction is measured by the normalized residuals for each individual predictor variable, i.e., abundance of a metabolite, spectral amplitude of the FTIR or HSI signature, abundance of an inorganic element. The smallness of such residuals is the necessary condition for the usefulness of the



**Fig. 10.** Extraction of the HSI spectral signatures. Top-left panel: spectrum-average image of the tray containing the seedlot tuber slices with the bounding boxes of individual slices. Top-right panel: the result of the segmentation of a single slice into its cortex (outer, yellow) and pith (inner, green) parts. Bottom panel: the spectra of the all cortex and pith pixels (shaded) and the corresponding average spectral signatures.

particular predictor variable. One can only achieve a good prediction of the dependent variable (vigor) if these residuals are small.

The predictor-variable residuals, or the *X*-data residuals, can be used in two ways: to rank the predictive power of the different data types, and to perform feature selection. Figure 6 shows the *X*-data residuals grouped and colored by the type of tuber data for the training dataset of 2020 and the testing dataset of 2021. In the top plot, the residuals of the 2020 variety-agnostic model predicting the 2021 Festien *X*-data are shown, and in the bottom plot, the residuals for the prediction of the same *X*-data by the 2020 variety-specific Festien model are displayed. In this and other training-testing configurations, the FTIR dataset consistently shows the smallest residuals, followed by the two HSI datasets, the XRF and the metabolome datasets.

The metabolome data demonstrate higher residuals than other datasets on average, and particularly, it has consistently higher residuals than FTIR. In mathematical terms, this means that the training metabolome dataset is less complete than the training FTIR dataset<sup>10</sup>. There can be several reasons why the IR algorithm indicates such incompleteness. First, the dependence of the majority of metabolites on the vigor parameter could be of ‘non-functional’ type, i.e., does not look like a sampled graph of a function. This is indeed what is observed in the trained variety-agnostic IR model. In Fig. 7 one can see the fits produced by the IR algorithm for the various predictor variables. The horizontal axis is the vigor parameter, and the vertical axis is the value of the corresponding predictor. Data points are colored in accordance with the seed tuber variety. The metabolome predictors shown in Fig. 7 (top four plots) exhibits a typical nonfunctional behavior<sup>10</sup>, since data points from different varieties represent vertically shifted clusters, i.e., it is impossible to draw a graph of a mathematical function of the vigor variable through these clusters.

The second reason for the apparent incompleteness of the metabolome dataset, which can be observed in the variety-specific Festien model, is the relatively high noise in the data, where data points appear to form a cloud rather than being aligned along a graph of a function, see Fig. 7 (left column, second plot from the top, green point cloud). The noise in these plots comes from two sources: the vigor data (error in the point location along the horizontal axis) and the metabolome data (error in the point location along the vertical axis). It is not clear which of the two noise sources is dominant, however, it is clear that the noise in the metabolome data is higher than in, e.g., the FTIR or HSI data.

## Discussion

In this project, we confirmed that potato plants grown from genetically identical seed tubers of different production origin and physiological state show systematic differences in vigor. These differences have been quantified here as significant correlations in the canopy area of seedlots across test fields. The answer to the question about the possibility to predict the seedlot vigor from the biochemical properties of the seed tubers is “sometimes yes”. The vigor of two varieties, Festien and to a lesser extent Sagitta, appears to be fairly predictable. Although, the model for Sagitta fails to predict the vigor in one of three test years. Notably, it was also the year which showed relatively small correlations in vigor between test fields for this variety. Festien and Sagitta were also the better predictable varieties in our previous microbiome study<sup>6</sup>, where it was shown that this level of



		2019							2020							2021						
Trained on	Tested on	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	0.56	0.88	0.73	0.62	0.61	0.47	0.38	1.04	0.83	0.98	0.88	0.89	0.95	0.86	1.77	1.22	1.12	1.09	1.25	1.17	1.14
	S	0.61	0.42	0.46	0.42	0.34	0.64	0.35	0.63	0.54	0.65	0.51	0.52	0.55	0.54	2.27	0.57	0.44	0.69	1.14	1.03	0.45
	V	0.51	0.98	0.88	0.81	0.83	0.84	0.64	0.92	0.79	0.99	0.83	0.80	0.89	0.85	1.83	1.20	1.08	1.09	1.28	1.19	1.11
	AVG	0.50	0.71	0.55	0.48	0.43	0.47	0.53	0.85	0.70	0.87	0.73	0.72	0.79	0.74	1.70	0.75	0.97	0.76	0.85	1.01	0.82
	MS	0.66	0.62	0.83	0.52	0.42	0.77	0.46	1.26	0.83	1.01	0.97	1.00	1.09	0.88	1.85	1.56	1.65	1.54	1.56	1.62	1.59
	MV	0.56	0.98	0.49	0.58	0.68	0.20	0.44	0.58	0.56	0.58	0.48	0.50	0.50	0.52	1.75	0.58	0.54	0.50	0.76	0.72	0.51
	SV	0.75	0.42	0.73	0.32	0.28	0.44	0.32	1.06	0.87	0.84	0.86	0.92	0.89	0.81	1.57	1.25	1.17	1.03	1.12	1.06	1.18
2020	M	0.64	0.69	0.64	0.55	0.56	0.60	0.56	0.24	0.35	0.45	0.31	0.27	0.31	0.35	1.53	0.50	0.47	0.31	0.53	0.50	0.43
	S	0.57	0.63	0.60	0.49	0.49	0.55	0.51	0.30	0.15	0.35	0.39	0.14	0.53	0.36	1.94	0.41	0.28	0.38	0.77	0.67	0.28
	V	1.03	0.81	1.03	0.87	0.84	1.01	0.84	0.38	0.44	0.61	0.51	0.52	0.56	0.51	1.80	0.51	0.51	0.47	0.74	0.74	0.45
	AVG	0.69	0.57	0.76	0.57	0.52	0.69	0.56	0.34	0.39	0.50	0.23	0.30	0.24	0.32	1.67	0.44	0.33	0.29	0.59	0.50	0.32
	MS	0.79	0.75	0.83	0.70	0.68	0.78	0.70	0.41	0.27	0.49	0.27	0.26	0.36	0.34	1.86	0.46	0.38	0.41	0.75	0.68	0.36
	MV	0.74	0.67	0.81	0.64	0.61	0.74	0.65	0.44	0.35	0.46	0.31	0.35	0.32	0.37	1.82	0.38	0.34	0.34	0.66	0.63	0.30
	SV	0.62	0.40	0.61	0.42	0.39	0.58	0.39	0.54	0.26	0.69	0.36	0.32	0.49	0.37	1.73	0.49	0.45	0.40	0.67	0.64	0.41
2021	M	1.65	1.90	1.78	1.76	1.76	1.71	1.83	1.46	1.99	1.71	1.70	1.71	1.56	1.85	0.39	2.48	2.20	1.60	1.21	1.06	2.28
	S	0.89	0.89	0.82	0.78	0.81	0.83	0.77	0.92	0.71	1.21	0.88	0.77	1.00	0.92	2.51	0.34	0.45	0.73	1.17	0.66	0.28
	V	0.63	0.66	0.67	0.54	0.54	0.61	0.56	1.11	0.77	0.88	0.85	0.89	0.94	0.78	2.01	0.40	0.27	0.37	0.81	0.70	0.24
	AVG	1.16	1.35	1.07	1.13	1.20	1.09	1.15	0.98	1.51	1.38	1.24	1.21	1.12	1.42	1.25	0.53	0.65	0.38	0.45	0.58	0.53
	MS	1.93	2.05	1.91	1.96	1.99	1.92	1.98	0.86	1.54	1.30	1.18	1.16	1.02	1.40	1.23	0.81	1.08	0.61	0.67	0.71	1.22
	MV	1.13	1.61	1.25	1.28	1.33	1.17	1.39	0.85	1.41	1.33	1.14	1.08	1.03	1.35	1.28	1.09	0.87	0.57	0.60	0.41	0.96
	SV	1.24	0.92	1.34	1.10	1.01	1.27	1.06	0.51	0.95	0.85	0.69	0.67	0.59	0.86	2.28	0.41	0.32	0.57	1.02	0.97	0.30
		2019							2020							2021						
Trained on	Tested on	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	0.72	0.56	0.63	0.69	0.70	0.77	0.81	0.48	0.59	0.51	0.56	0.56	0.52	0.57		0.39	0.44	0.45	0.37	0.42	0.43
	S	0.70	0.79	0.77	0.79	0.83	0.68	0.82	0.69	0.73	0.68	0.74	0.74	0.72	0.73		0.71	0.78	0.66	0.43	0.48	0.77
	V	0.75	0.51	0.56	0.60	0.59	0.58	0.68	0.54	0.61	0.50	0.59	0.60	0.55	0.58		0.40	0.46	0.45	0.36	0.40	0.44
	AVG	0.75	0.64	0.73	0.76	0.79	0.76	0.74	0.57	0.65	0.57	0.64	0.64	0.60	0.63	0.15	0.62	0.51	0.62	0.58	0.50	0.59
	MS	0.67	0.69	0.58	0.74	0.79	0.61	0.77	0.37	0.58	0.49	0.51	0.50	0.46	0.56		0.22	0.18	0.23	0.22	0.19	0.21
	MV	0.72	0.51	0.75	0.71	0.66	0.90	0.78	0.71	0.72	0.71	0.76	0.75	0.75	0.74		0.71	0.73	0.75	0.62	0.64	0.75
	SV	0.63	0.79	0.64	0.84	0.86	0.78	0.84	0.47	0.56	0.58	0.57	0.54	0.55	0.59	0.22	0.37	0.41	0.48	0.44	0.47	0.41
2020	M	0.68	0.66	0.68	0.73	0.72	0.70	0.72	0.88	0.83	0.78	0.85	0.87	0.85	0.82	0.24	0.75	0.76	0.84	0.73	0.75	0.79
	S	0.72	0.68	0.70	0.76	0.75	0.72	0.74	0.85	0.93	0.82	0.80	0.93	0.74	0.82		0.79	0.86	0.81	0.61	0.67	0.86
	V	0.48	0.59	0.48	0.56	0.58	0.50	0.58	0.81	0.78	0.70	0.75	0.74	0.72	0.75		0.75	0.75	0.77	0.63	0.63	0.77
	AVG	0.65	0.72	0.62	0.72	0.74	0.65	0.72	0.83	0.81	0.75	0.89	0.85	0.88	0.84	0.16	0.78	0.84	0.86	0.70	0.75	0.84
	MS	0.60	0.62	0.58	0.65	0.66	0.61	0.65	0.79	0.87	0.76	0.87	0.87	0.82	0.83		0.77	0.81	0.79	0.62	0.66	0.82
	MV	0.63	0.66	0.60	0.68	0.70	0.63	0.68	0.78	0.82	0.77	0.85	0.83	0.84	0.81		0.81	0.83	0.83	0.67	0.69	0.85
	SV	0.69	0.80	0.69	0.79	0.80	0.71	0.80	0.73	0.87	0.65	0.82	0.84	0.75	0.81		0.76	0.77	0.80	0.66	0.68	0.79
2021	M	0.17							0.27			0.15		0.22		0.80				0.40	0.47	
	S	0.56	0.55	0.59	0.61	0.60	0.59	0.61	0.54	0.64	0.40	0.56	0.62	0.50	0.54		0.83	0.77	0.64	0.42	0.67	0.86
	V	0.69	0.67	0.67	0.73	0.73	0.69	0.72	0.44	0.62	0.56	0.57	0.55	0.53	0.61		0.80	0.87	0.81	0.60	0.65	0.88
	AVG	0.42	0.32	0.46	0.43	0.40	0.45	0.42	0.51	0.24	0.31	0.38	0.40	0.44	0.29	0.37	0.73	0.67	0.81	0.77	0.71	0.74
	MS								0.57	0.23	0.35	0.41	0.42	0.49	0.30	0.39	0.60	0.46	0.70	0.67	0.65	0.39
	MV	0.43	0.19	0.38	0.36	0.34	0.41	0.31	0.58	0.30	0.33	0.43	0.46	0.49	0.33	0.36	0.45	0.56	0.72	0.70	0.80	0.52
	SV	0.38	0.54	0.33	0.45	0.50	0.36	0.47	0.75	0.52	0.58	0.66	0.67	0.70	0.57		0.80	0.84	0.72	0.49	0.51	0.85

**Fig. 11.** Performance of the variety-agnostic model tested on all six varieties simultaneously: the SSE residual  $r^2$  (top), the Pearson correlation coefficient  $c$  (bottom). Rows correspond to the training dataset, columns show the testing configuration. SSE residuals  $r^2 \geq 1$  are grayed out, correlations with  $p > 0.05$  are omitted.

predictability is enough to distinguish between three practically meaningful classes for the seedlots of these varieties: below average, average, and above average. Moreover, the feature selection algorithm applied to all biochemical tuber data, indicates that measuring the FTIR spectra of the dry samples of the seed tubers is sufficient to achieve this classification.

		2019							2020							2021						
Trained on	Tested on	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	1.08	1.19	0.87	1.94	0.97	0.21	0.53	1.69	1.70	1.30	1.43	1.67	1.37	1.40	1.41	2.16	2.22	1.72	1.55	1.60	2.25
	S	1.52	1.50	0.91	1.24	0.91	1.11	1.16	2.02	2.16	1.58	1.85	2.12	1.69	1.81	1.58	2.56	2.04	1.95	1.91	1.66	2.39
	V	1.20	0.58	1.58	0.99	1.54	2.76	2.69	1.45	1.91	1.50	1.57	1.72	1.44	1.63	1.92	2.47	1.69	2.02	2.18	1.79	2.11
	AVG	0.88	0.13	1.05	1.35	0.36	0.57	0.21	1.71	1.93	1.38	1.58	1.84	1.44	1.56	1.20	1.76	1.98	1.28	1.15	1.31	1.83
	MS	1.37	0.79	2.04	0.95	0.51	0.57	0.11	1.71	1.71	1.68	1.64	1.68	1.66	1.64	1.21	1.98	2.15	1.46	1.28	1.40	2.08
	MV	1.10	1.21	1.32	1.15	0.46	0.22	0.14	2.31	2.41	2.22	2.35	2.41	2.27	2.35	1.26	1.92	1.95	1.38	1.29	1.34	1.92
	SV	1.37	1.01	0.79	1.03	1.71	1.91	0.14	1.81	1.95	1.69	1.77	1.89	1.70	1.77	1.50	1.91	2.10	1.63	1.50	1.62	2.01
2020	M	2.03	2.21	1.38	2.02	2.12	1.88	2.01	2.41	1.97	1.34	0.74	0.19	1.19	0.46	2.50	2.06	2.08	2.44	2.48	2.46	2.09
	S	1.88	1.69	1.70	1.74	1.78	1.82	1.63	0.72	1.05	1.70	0.71	1.49	1.09	0.48	2.28	1.89	2.33	2.31	2.19	2.40	2.14
	V	1.26	1.33	1.80	1.22	1.22	1.26	1.32	1.55	0.89	0.26	0.97	1.52	0.74	0.57	1.59	2.59	2.03	2.30	1.94	1.66	2.41
	AVG	1.52	1.63	2.87	1.67	1.52	1.76	1.92	0.53	0.48	0.86	1.93	0.17	0.71	1.03	1.80	2.27	2.37	2.12	1.96	2.00	2.40
	MS	1.87	1.58	1.78	1.69	1.72	1.82	1.55	1.02	1.25	1.68	0.70	0.62	1.05	1.18	1.65	1.98	2.30	1.85	1.68	1.85	2.17
	MV	2.07	2.12	2.28	2.14	2.10	2.13	2.19	0.16	1.32	2.38	0.20	0.44	0.60	1.06	2.54	1.80	2.20	2.41	2.38	2.56	2.00
	SV	1.79	1.59	1.41	1.59	1.67	1.67	1.45	0.73	0.25	0.54	0.87	0.42	1.19	0.88	1.79	2.50	2.50	2.28	2.07	2.07	2.64
2021	M	1.00	1.30	2.36	1.14	1.05	1.16	1.46	2.16	2.10	2.18	2.17	2.13	2.18	2.17	2.81	1.78	1.55	1.12	1.52	0.95	2.59
	S	2.39	2.08	1.86	2.24	2.28	2.33	2.03	2.02	1.96	2.77	2.40	1.98	2.59	2.48	1.09	0.49	1.72	1.14	2.29	1.18	3.23
	V	1.99	1.75	2.18	1.90	1.87	2.03	1.82	2.28	2.28	2.04	2.20	2.31	2.12	2.17	2.66	2.24	2.98	1.26	2.39	1.45	2.93
	AVG	1.17	1.02	1.92	1.04	1.01	1.20	1.08	1.85	2.09	1.77	1.88	2.00	1.77	1.89	2.83	1.30	1.92	2.07	1.14	1.81	0.92
	MS	1.26	1.47	1.61	1.26	1.29	1.22	1.40	2.53	2.45	2.90	2.78	2.52	2.86	2.81	1.79	2.67	1.03	0.94	0.06	1.28	1.38
	MV	1.69	1.99	1.34	1.71	1.80	1.55	1.80	2.31	2.34	2.13	2.27	2.36	2.19	2.25	1.62	1.99	3.22	0.98	1.75	1.11	1.30
	SV	1.97	1.75	2.16	1.89	1.86	2.01	1.82	2.30	2.40	2.38	2.43	2.40	2.38	2.45	2.85	0.39	2.10	1.20	2.00	1.08	2.98

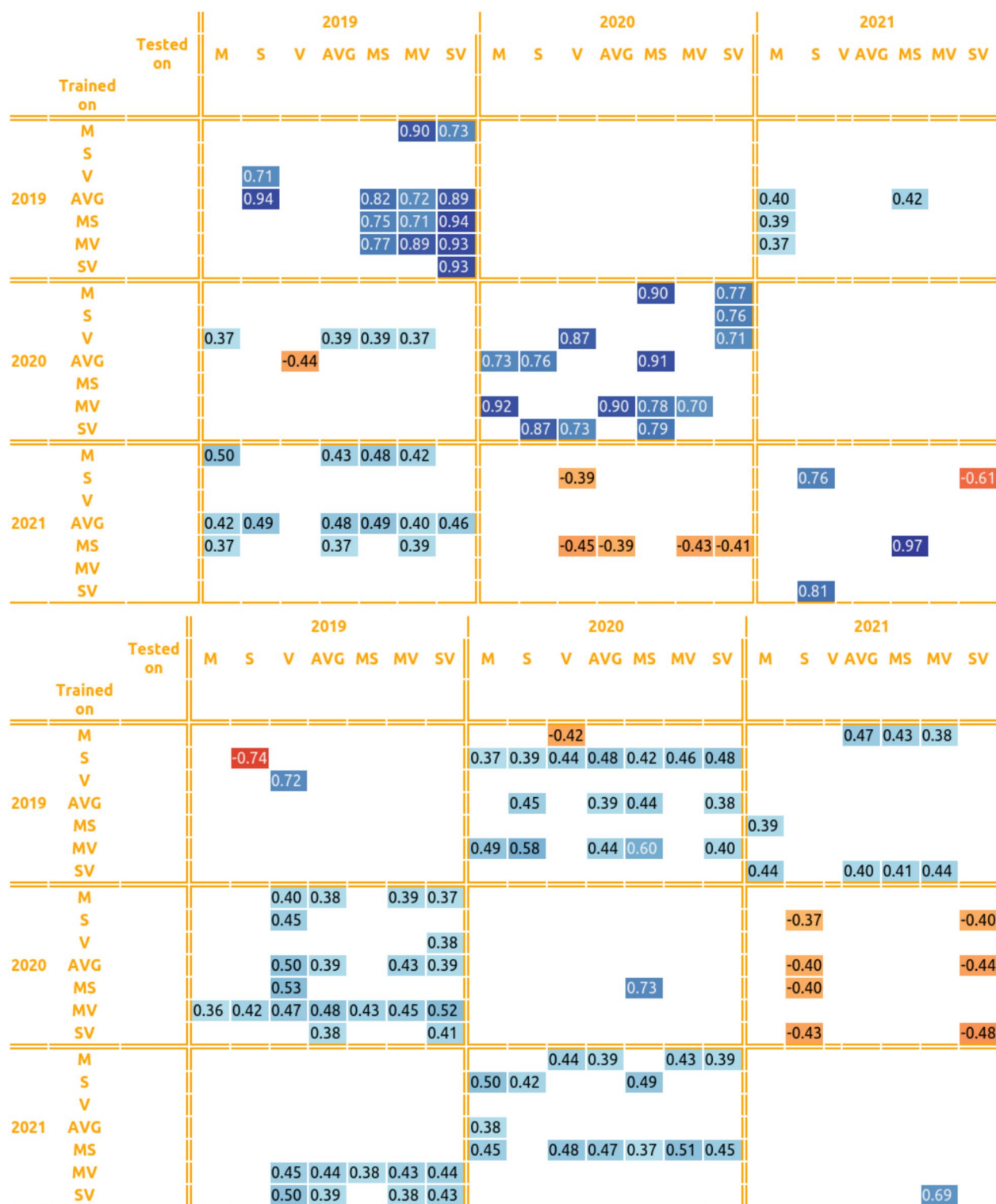
  

		2019							2020							2021						
Trained on	Tested on	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	3.11	3.19	2.64	3.28	3.28	1.67	1.65	2.15	2.04	2.84	2.50	2.09	2.69	2.57	1.38	1.43	1.51	1.07	1.14	1.23	1.32
	S	2.45	3.47	1.70	1.09	1.19	0.86	1.54	1.26	1.23	1.11	1.04	1.16	1.09	1.04	1.94	2.39	1.90	2.08	2.15	1.90	2.19
	V	2.49	1.85	0.55	1.75	1.94	1.71	3.33	1.78	1.79	1.85	1.78	1.77	1.82	1.80	1.58	1.93	2.08	1.69	1.59	1.68	2.01
	AVG	2.48	2.38	1.19	2.29	0.85	1.13	1.04	1.34	1.11	1.51	1.22	1.11	1.41	1.24	1.54	2.56	2.39	2.06	1.88	1.80	2.61
	MS	1.59	1.10	2.32	1.28	1.15	1.48	1.36	2.42	2.25	2.35	2.38	2.34	2.40	2.35	1.22	2.46	2.15	1.67	1.53	1.40	2.39
	MV	1.74	1.16	1.53	1.50	1.56	1.23	1.94	1.02	0.85	1.64	1.12	0.81	1.41	1.19	1.92	1.87	2.39	2.05	1.86	2.12	2.16
	SV	1.90	1.16	2.10	1.93	1.81	2.18	1.88	2.05	2.11	2.51	2.33	2.10	2.40	2.38	1.11	1.95	1.76	1.20	1.17	1.13	1.82
2020	M	1.36	1.44	1.20	1.24	1.34	1.22	1.25	2.64	0.76	1.07	1.69	0.92	1.20	0.77	1.51	2.39	2.26	1.91	1.76	1.70	2.41
	S	1.87	1.98	1.10	1.77	1.91	1.66	1.71	1.86	1.75	1.15	1.23	1.77	1.20	1.24	1.83	2.73	2.52	2.41	2.23	2.10	2.80
	V	1.64	1.40	1.29	1.40	1.49	1.50	1.24	1.01	1.75	1.80	2.02	2.35	1.82	2.05	2.18	2.61	1.97	2.40	2.49	2.14	2.38
	AVG	1.34	1.47	0.99	1.21	1.34	1.14	1.22	1.27	1.32	0.88	1.27	1.27	1.38	1.28	1.54	2.81	2.57	2.25	2.00	1.88	2.88
	MS	1.63	1.74	0.94	1.50	1.65	1.41	1.45	0.67	0.73	0.92	1.25	0.54	0.82	1.35	1.73	2.80	2.22	2.24	2.17	1.87	2.66
	MV	1.28	1.15	1.07	1.03	1.14	1.11	0.95	0.87	1.62	1.64	1.34	1.40	1.37	1.36	1.73	2.58	2.16	2.12	2.06	1.84	2.48
	SV	1.41	1.31	1.37	1.23	1.30	1.30	1.18	1.88	0.78	1.74	0.68	1.05	1.64	1.38	1.82	2.87	2.64	2.52	2.30	2.16	2.97
2021	M	1.63	1.87	2.35	1.77	1.71	1.74	1.98	1.37	1.56	1.13	1.21	1.44	1.13	1.21	1.96	1.32	1.36	1.33	1.41	1.09	1.76
	S	2.20	2.26	1.92	2.23	2.25	2.17	2.21	1.00	1.16	1.79	1.32	1.02	1.52	1.45	1.42	1.66	1.80	1.05	0.96	1.11	1.41
	V	1.85	2.07	1.40	1.85	1.94	1.71	1.89	1.96	2.24	2.41	2.31	2.15	2.29	2.39	1.89	1.92	1.94	2.00	2.17	2.22	1.44
	AVG	1.88	2.22	2.19	2.06	2.03	1.93	2.26	1.24	1.57	1.43	1.35	1.40	1.33	1.42	2.38	2.88	2.46	1.70	1.49	1.27	1.83
	MS	1.88	1.84	2.05	1.86	1.85	1.90	1.87	1.11	1.43	1.04	1.05	1.25	0.98	1.09	1.71	1.55	1.39	1.38	2.12	1.99	2.74
	MV	1.30	1.32	1.10	1.13	1.24	1.13	1.11	1.31	1.70	1.46	1.43	1.52	1.36	1.50	2.32	2.43	2.49	2.10	1.62	2.24	2.21
	SV	1.43	1.39	1.01	1.23	1.35	1.23	1.15	2.08	1.81	2.31	2.11	1.90	2.26	2.11	2.04	2.45	3.01	2.64	2.23	0.61	1.04

**Fig. 12.** Performance (SSE residual  $r^2$ ) of the variety-specific Challenger model (top) and the variety-agnostic model tested on the Challenger variety (bottom). Rows correspond to the training dataset, columns show the testing configuration. SSE residuals  $r^2 \geq 1$  are grayed out.

For the other four varieties (Challenger, Colomba, Innovator, and Seresta) the vigor of the seedlots could not be predicted in any reliable way. It is worth noting that the vigor correlations across the test fields were also not as consistent for these varieties, which may be a sign of a purely stochastic nature of vigor variations within these varieties. Nevertheless, in our opinion, the reasons behind this lack of predictability are not entirely clear. On the one hand, it is possible that the biochemical composition of seed tubers simply does not define the vigor of these





**Fig. 13.** Performance (Pearson correlation  $c$ ) of the variety-specific Challenger model (top) and the variety-agnostic model tested on the Challenger variety (bottom). Rows correspond to the training dataset, columns show the testing configuration. Correlations with  $p \geq 0.05$  are omitted.

varieties to the same extent as with Festien, and there are other processes at play that were not measured in the present experiment. A seed tuber is a slowly changing dynamical system. Therefore, measuring its biochemical constitution at a few points in time may deliver better predictors of the potato plant vigor. This could mean that Festien tubers are somehow more 'stationary' than tubers of other varieties.

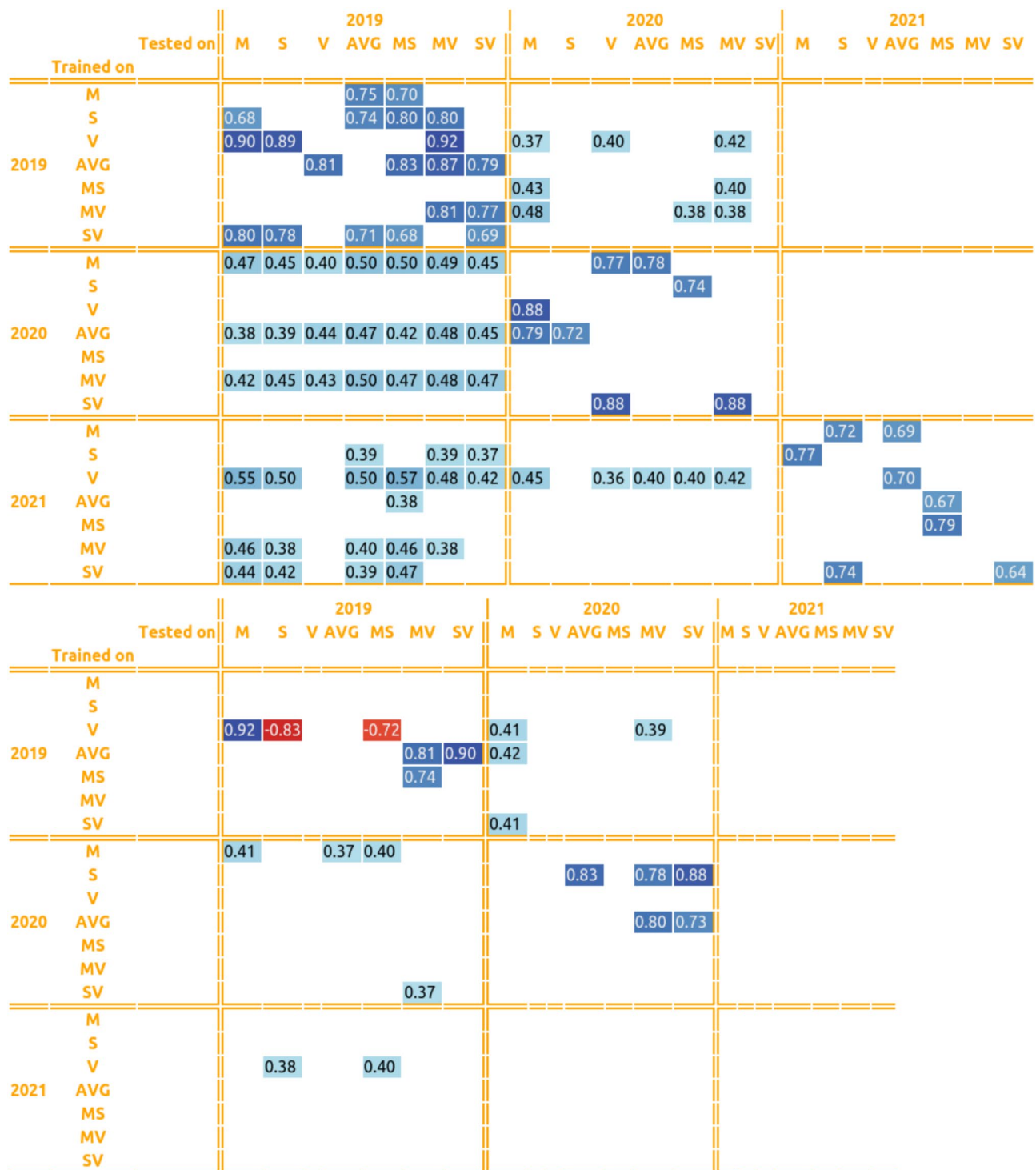
		2019							2020							2021						
Trained on	Tested on	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	1.52	1.01	0.72	0.51	0.59	1.82	1.88	1.52	1.98	1.78	1.75	1.74	1.66	1.85	1.92	1.40	1.56	1.54	1.61	1.70	1.40
	S	0.64	1.90	1.51	0.52	0.41	0.41	0.80	1.93	2.03	1.65	1.82	1.98	1.74	1.79	1.93	2.04	1.84	1.92	1.98	1.87	1.93
	V	0.21	0.22	2.03	0.99	2.22	0.16	1.00	1.26	1.71	1.19	1.30	1.45	1.16	1.35	1.90	1.87	1.64	1.77	1.87	1.74	1.72
	AVG	0.96	0.94	0.38	1.23	0.34	0.27	0.43	1.37	1.76	1.88	1.68	1.54	1.66	1.82	1.91	1.35	1.71	1.58	1.58	1.78	1.45
	MS	0.80	1.40	0.96	0.90	1.39	0.60	0.98	1.14	1.71	1.32	1.33	1.39	1.20	1.44	1.85	1.64	1.51	1.59	1.71	1.63	1.51
	MV	1.08	1.48	0.68	0.80	2.38	0.38	0.46	1.03	1.53	1.44	1.29	1.24	1.23	1.44	2.16	1.90	1.84	1.96	2.03	2.01	1.85
	SV	0.40	0.45	1.07	0.58	0.63	1.25	0.62	1.40	1.73	1.38	1.43	1.54	1.34	1.48	1.68	1.84	1.52	1.61	1.72	1.53	1.63
2020	M	1.05	1.10	1.21	1.00	0.99	1.02	1.10	1.31	1.04	0.47	0.44	3.26	0.62	1.07	2.10	1.82	1.91	1.94	1.96	2.01	1.85
	S	2.66	2.49	2.10	2.42	2.63	2.37	2.26	1.58	3.12	1.26	1.82	0.51	0.72	1.46	2.08	2.59	2.00	2.27	2.37	2.05	2.34
	V	2.42	2.20	1.85	2.12	2.34	2.08	1.97	0.23	1.79	2.30	1.42	2.42	2.05	2.47	2.00	2.43	1.81	2.10	2.24	1.89	2.14
	AVG	1.25	1.22	1.12	1.06	1.17	1.05	1.10	0.43	0.56	1.19	3.04	1.31	1.00	2.32	1.99	2.61	1.68	2.12	2.34	1.81	2.17
	MS	2.56	2.35	2.14	2.37	2.50	2.35	2.24	1.82	1.36	0.70	0.88	1.68	0.92	1.90	1.88	2.16	2.01	2.02	2.02	1.93	2.10
	MV	1.17	1.10	1.15	1.01	1.05	1.03	1.06	0.72	0.93	1.11	1.66	0.87	0.84	1.04	1.84	2.18	1.66	1.87	2.00	1.71	1.91
	SV	1.70	1.70	1.60	1.60	1.67	1.58	1.61	0.67	2.22	0.25	2.74	1.53	0.25	2.52	2.02	2.64	1.87	2.22	2.37	1.94	2.30
2021	M	2.27	2.19	1.90	2.36	2.25	2.40	2.35	2.20	1.81	2.13	2.07	2.01	2.17	2.01	0.52	0.55	1.43	0.63	2.17	1.40	1.82
	S	1.37	1.37	1.28	1.23	1.31	1.21	1.26	1.72	1.91	1.85	1.82	1.80	1.79	1.87	0.46	2.11	0.87	1.11	2.24	2.80	1.64
	V	0.89	1.00	1.35	0.99	0.85	1.05	1.16	1.10	1.40	1.28	1.20	1.21	1.15	1.28	2.49	1.69	1.48	0.59	1.16	1.14	1.55
	AVG	1.29	1.32	1.66	1.38	1.24	1.45	1.50	1.76	1.94	1.73	1.78	1.84	1.72	1.80	1.81	1.81	1.08	1.75	0.67	1.33	0.85
	MS	2.07	2.00	2.50	2.29	2.04	2.39	2.34	1.79	1.47	1.37	1.47	1.61	1.50	1.37	1.29	2.94	1.00	1.12	0.43	2.29	0.97
	MV	1.09	1.23	1.49	1.20	1.08	1.24	1.35	1.56	1.72	1.35	1.47	1.62	1.39	1.46	1.01	1.54	1.39	1.10	1.46	1.26	0.91
	SV	1.12	1.16	1.54	1.22	1.06	1.29	1.35	1.60	1.67	1.59	1.58	1.62	1.57	1.60	1.19	0.52	1.94	1.03	1.19	1.50	0.71

		2019							2020							2021						
Trained on	Tested on	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	0.74	1.16	1.69	1.03	0.92	2.43	1.91	1.73	2.20	1.85	1.90	1.96	1.79	1.98	1.69	1.42	1.45	1.41	1.49	1.50	1.34
	S	3.14	3.01	2.50	1.33	1.37	1.23	1.42	2.11	1.77	1.95	1.94	1.93	2.01	1.87	2.34	1.70	2.18	2.10	2.03	2.31	1.93
	V	0.16	3.66	3.00	3.38	3.44	3.13	3.19	1.19	1.61	1.33	1.31	1.37	1.22	1.40	1.95	1.63	1.58	1.66	1.77	1.73	1.55
	AVG	1.37	1.34	1.10	1.28	0.61	0.37	0.19	1.16	1.54	1.54	1.39	1.31	1.36	1.52	2.13	1.68	1.72	1.81	1.90	1.92	1.65
	MS	1.69	1.67	0.86	0.64	0.73	0.53	0.76	1.53	1.91	1.66	1.67	1.71	1.58	1.74	2.00	1.90	1.56	1.78	1.94	1.75	1.69
	MV	1.37	2.13	1.52	1.71	1.81	2.00	1.96	1.67	1.80	1.80	1.74	1.72	1.73	1.79	2.03	1.46	1.98	1.78	1.72	2.01	1.67
	SV	1.92	1.81	1.79	1.73	1.80	1.74	1.65	1.19	1.79	1.46	1.43	1.46	1.31	1.56	1.77	1.38	1.44	1.42	1.51	1.54	1.32
2020	M	1.19	1.36	1.47	1.26	1.20	1.27	1.38	1.69	1.14	0.72	1.32	1.75	0.99	0.67	1.97	1.86	1.91	1.89	1.90	1.93	1.87
	S	1.81	1.88	1.79	1.79	1.83	1.77	1.81	0.82	1.33	1.59	0.34	1.05	0.44	0.24	2.32	2.23	2.19	2.31	2.32	2.30	2.25
	V	2.32	2.45	2.13	2.31	2.42	2.24	2.27	1.56	1.17	1.67	1.99	2.24	1.84	1.92	2.47	1.77	2.29	2.23	2.15	2.45	2.03
	AVG	1.68	1.67	1.57	1.57	1.65	1.55	1.57	2.50	2.01	1.48	0.63	2.29	0.40	0.55	2.09	2.31	1.66	2.03	2.23	1.86	1.99
	MS	2.56	2.53	2.08	2.39	2.59	2.30	2.26	1.02	0.68	1.50	1.08	0.71	1.30	0.81	2.56	2.18	2.36	2.46	2.43	2.55	2.32
	MV	1.49	1.85	2.04	1.81	1.63	1.81	1.97	3.15	1.69	2.63	2.46	3.06	2.26	2.09	2.05	1.70	1.78	1.81	1.86	1.90	1.70
	SV	1.37	1.47	1.34	1.29	1.37	1.26	1.34	1.24	0.64	1.12	2.46	0.95	1.28	0.77	2.19	1.83	2.03	2.02	2.01	2.13	1.91
2021	M	1.74	2.08	2.48	2.19	1.90	2.23	2.36	1.51	1.49	1.61	1.51	1.47	1.54	1.54	0.99	0.82	2.29	1.62	1.33	1.69	1.89
	S	1.31	1.42	1.79	1.49	1.31	1.55	1.63	1.70	1.78	1.74	1.72	1.73	1.70	1.74	2.29	1.53	0.84	1.41	1.82	2.45	2.69
	V	1.30	1.25	1.48	1.27	1.21	1.32	1.35	1.49	1.78	1.53	1.55	1.61	1.48	1.60	1.48	1.21	2.16	1.51	1.09	1.81	1.66
	AVG	1.76	1.66	1.77	1.69	1.68	1.73	1.71	1.63	1.50	1.54	1.51	1.54	1.54	1.49	0.87	1.94	2.21	2.17	2.13	2.22	0.76
	MS	1.91	2.25	2.27	2.19	2.08	2.15	2.28	1.32	1.56	1.78	1.55	1.41	1.58	1.68	1.32	1.51	1.46	1.34	1.91	1.79	2.29
	MV	1.68	1.66	1.90	1.74	1.64	1.79	1.80	1.57	1.50	1.51	1.48	1.51	1.50	1.48	0.68	1.71	1.44	1.27	1.22	0.93	1.70
	SV	1.35	1.46	1.60	1.41	1.35	1.43	1.51	1.89	1.98	2.03	1.98	1.93	1.98	2.01	1.29	2.16	1.22	1.49	1.68	1.83	1.45

**Fig. 14.** Performance (SSE residual  $r^2$ ) of the variety-specific Colomba model (top) and the variety-agnostic model tested on the Colomba variety (bottom). Rows correspond to the training dataset, columns show the testing configuration. SSE residuals  $r^2 \geq 1$  are grayed out.

On the other hand, these conclusions are also subject to the choice of the vigor parameter—leaf canopy area on a selected date. Since the selected date was at the end of the exponential canopy growth period, even small variations in the emergence date or the rate of growth could lead to significant changes in the ordering of the seedlots at the measurement date, thus, reducing the correlation in the vigor parameter between the fields



**Fig. 15.** Performance (Pearson correlation  $c$ ) of the variety-specific Colomba model (top) and the variety-agnostic model tested on the Colomba variety (bottom). Rows correspond to the training dataset, columns show the testing configuration. Correlations with  $p \geq 0.05$  are omitted.

and the eventual predictability of vigor. It is possible that a different, possibly dynamic, vigor measure, while requiring more phenotyping effort, could be both more stable and more predictable.

Finally, observations in the year 2021 indicate a strong influence of the environment on some seedlots and varieties, which can be taken as the evidence of the genotype-by-environment interaction. The definition of the potato plant vigor and the quality of its prediction can be significantly improved if the influence of the environment is taken into account by incorporating the readily available environmental data into a physiology-informed growth model. This, however, requires a separate set of dedicated experiments or suitable historical



		2019							2020							2021						
Trained on	Tested on	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	0.50	0.70	0.66	0.41	0.14	0.24	0.12	0.98	1.40	1.00	1.01	1.10	0.93	1.10	2.55	1.15	1.16	1.39	1.66	1.96	1.09
	S	0.48	1.86	0.87	0.62	0.37	1.02	0.56	0.78	1.06	0.84	0.77	0.83	0.73	0.84	2.80	0.75	1.26	1.33	1.50	2.24	0.88
	V	0.50	1.65	0.47	0.33	0.32	0.74	0.24	0.87	1.11	0.82	0.81	0.91	0.77	0.86	3.02	0.73	1.23	1.45	1.71	2.42	0.85
	AVG	0.25	0.41	0.13	1.37	0.14	0.67	0.10	1.06	1.22	1.13	1.05	1.07	1.04	1.11	2.98	0.73	1.03	1.33	1.66	2.26	0.76
	MS	0.46	0.35	0.15	0.20	0.34	0.10	0.06	1.02	1.12	1.15	1.01	1.01	1.02	1.07	3.31	0.91	1.62	1.94	2.19	2.91	1.14
	MV	0.41	0.53	0.36	1.87	0.30	0.01	0.55	0.74	1.10	0.78	0.74	0.82	0.68	0.83	2.95	0.70	1.15	1.34	1.60	2.31	0.80
	SV	0.10	0.62	2.07	0.75	0.38	0.45	0.10	2.18	2.17	2.21	2.21	2.18	2.21	2.21	2.81	1.28	1.80	1.95	2.07	2.59	1.45
2020	M	0.94	0.98	0.95	0.89	0.90	0.92	0.91	2.04	0.48	0.58	1.66	0.45	0.46	0.64	2.39	1.38	2.02	1.85	1.75	2.36	1.61
	S	3.27	2.91	3.14	3.20	3.19	3.25	3.07	0.99	2.10	1.04	1.40	0.44	1.59	1.32	2.15	1.34	1.81	1.56	1.46	2.01	1.50
	V	0.68	0.61	0.72	0.59	0.58	0.67	0.59	0.23	0.13	1.42	0.18	0.75	0.54	0.96	2.43	0.70	0.74	0.81	1.06	1.59	0.62
	AVG	2.57	2.12	2.33	2.41	2.42	2.50	2.23	0.82	1.28	2.17	0.40	0.19	1.09	0.69	2.47	1.08	1.34	1.37	1.51	2.00	1.12
	MS	0.76	0.69	0.88	0.70	0.67	0.78	0.72	1.60	0.30	0.46	0.36	0.32	1.21	0.73	2.46	0.56	0.75	0.73	0.94	1.62	0.53
	MV	1.01	0.86	1.00	0.90	0.89	0.98	0.87	0.44	0.80	0.75	0.58	0.49	1.28	0.39	2.45	0.97	0.92	1.09	1.37	1.72	0.87
	SV	1.22	0.91	1.00	1.02	1.05	1.13	0.90	0.18	0.24	0.43	1.35	0.89	0.89	0.52	2.36	0.94	1.15	1.12	1.25	1.79	0.95
2021	M	3.32	3.46	3.48	3.48	3.45	3.41	3.54	2.99	3.05	2.98	3.09	3.07	3.05	3.08	0.56	1.72	2.64	2.18	1.99	0.42	3.08
	S	0.39	0.37	0.40	0.29	0.29	0.35	0.30	1.06	0.80	1.14	0.94	0.90	1.04	0.94	0.09	0.64	1.38	2.19	1.34	1.83	0.50
	V	2.23	1.98	2.36	2.20	2.14	2.28	2.16	1.75	1.94	1.22	1.56	1.81	1.45	1.46	1.09	1.13	2.06	0.99	0.47	1.96	1.51
	AVG	2.71	2.85	2.90	2.84	2.80	2.79	2.92	1.67	1.66	1.48	1.56	1.65	1.55	1.51	2.70	0.88	1.06	1.41	0.74	1.53	0.96
	MS	2.54	2.64	2.77	2.66	2.61	2.63	2.74	1.23	1.50	1.08	1.17	1.30	1.10	1.19	1.34	0.25	1.94	0.16	0.50	1.46	0.87
	MV	2.77	2.83	2.98	2.89	2.84	2.86	2.94	2.01	2.21	1.64	1.91	2.10	1.81	1.85	0.80	3.52	1.47	0.41	0.96	0.51	2.04
	SV	1.34	2.10	1.78	1.65	1.61	1.48	1.95	1.47	1.73	1.74	1.61	1.55	1.58	1.72	1.89	0.43	0.50	0.32	0.73	2.82	0.34

		2019							2020							2021						
Trained on	Tested on	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	1.66	2.06	1.05	1.36	1.78	2.31	2.66	0.80	1.04	0.59	0.67	0.84	0.61	0.67	2.41	0.99	0.82	1.03	1.35	1.63	0.84
	S	2.39	2.81	1.33	1.43	0.92	1.75	1.91	2.14	2.28	1.76	2.03	2.21	1.94	1.96	1.95	1.22	1.06	0.99	1.12	1.37	1.08
	V	0.70	3.15	1.65	2.22	2.70	1.88	1.93	1.62	1.60	1.46	1.51	1.59	1.51	1.48	1.61	1.76	1.51	1.35	1.35	1.35	1.63
	AVG	2.36	2.38	2.46	2.48	1.68	1.89	1.96	1.10	1.03	1.62	1.22	1.02	1.32	1.35	2.38	0.58	1.30	0.95	0.89	1.89	0.79
	MS	1.47	1.46	1.91	1.64	1.50	1.51	1.95	1.34	1.56	1.54	1.42	1.40	1.41	1.51	2.42	0.99	1.12	1.17	1.36	1.82	0.97
	MV	1.79	1.77	1.01	1.40	1.76	0.98	1.01	0.92	0.74	0.99	0.81	0.79	0.89	0.82	2.31	0.99	1.53	1.30	1.25	1.98	1.15
	SV	1.89	0.53	1.41	0.68	0.40	0.74	1.83	0.97	1.23	0.74	0.85	1.02	0.78	0.85	1.76	1.32	1.01	0.91	1.03	1.17	1.13
2020	M	0.54	0.94	0.76	0.63	0.62	0.59	0.80	1.34	1.64	1.83	0.48	1.76	1.84	1.74	2.61	0.73	1.26	1.19	1.28	2.07	0.86
	S	0.87	1.22	1.08	0.97	0.95	0.92	1.11	0.58	0.79	0.77	0.87	0.65	0.81	0.91	1.75	1.42	1.57	1.24	1.13	1.51	1.44
	V	1.93	2.49	2.25	2.18	2.15	2.04	2.40	1.12	1.21	2.09	1.66	1.18	1.79	1.84	2.25	1.09	1.04	1.10	1.30	1.62	1.00
	AVG	0.83	1.30	1.18	0.99	0.95	0.93	1.20	0.46	0.67	0.81	1.67	0.56	1.82	1.60	1.89	1.05	1.16	0.89	0.88	1.38	1.03
	MS	0.84	1.51	1.20	1.07	1.05	0.94	1.33	0.63	1.18	0.99	0.33	0.94	0.87	0.37	2.03	1.15	1.47	1.20	1.14	1.70	1.23
	MV	1.43	1.85	1.81	1.63	1.57	1.56	1.82	1.10	0.73	0.98	0.79	0.64	0.62	1.12	2.27	0.77	1.02	0.87	0.97	1.62	0.78
	SV	1.96	2.20	2.17	2.08	2.05	2.03	2.20	0.65	0.72	1.07	1.18	0.96	1.19	0.85	2.32	1.09	1.90	1.55	1.37	2.23	1.38
2021	M	0.72	1.19	0.90	0.84	0.84	0.76	1.01	1.44	1.55	1.53	1.46	1.45	1.45	1.51	1.74	0.36	0.41	0.34	0.49	0.27	0.42
	S	1.67	1.07	1.45	1.41	1.41	1.58	1.21	1.34	1.01	1.35	1.20	1.17	1.30	1.16	2.13	1.82	3.13	2.49	1.96	1.09	0.48
	V	0.64	0.54	0.66	0.54	0.53	0.62	0.52	1.26	1.26	0.96	1.07	1.22	1.05	1.01	1.36	1.08	0.74	0.29	0.30	0.33	0.30
	AVG	2.01	2.02	2.14	2.05	2.02	2.06	2.08	1.63	1.95	1.33	1.56	1.75	1.45	1.54	0.23	1.56	1.78	1.62	1.54	1.66	0.74
	MS	2.42	2.47	2.57	2.50	2.46	2.48	2.54	1.32	1.76	1.13	1.30	1.47	1.17	1.33	2.29	2.14	1.87	2.13	0.68	0.68	1.47
	MV	1.14	1.42	1.39	1.24	1.21	1.21	1.37	1.70	2.21	1.52	1.73	1.90	1.58	1.77	0.87	2.67	0.20	0.32	0.47	0.24	0.31
	SV	1.37	1.37	1.38	1.34	1.34	1.36	1.34	1.65	2.02	2.04	1.88	1.79	1.84	2.03	0.88	1.03	0.78	0.81	0.91	0.91	0.98

**Fig. 16.** Performance (SSE residual  $r^2$ ) of the variety-specific Festien model (top) and the variety-agnostic model tested on the Festien variety (bottom). Rows correspond to the training dataset, columns show the testing configuration. SSE residuals  $r^2 \geq 1$  are grayed out.

data, in order to learn the genotype-specific environmental component and also calibrate the model for each particular genotype. One way forward is to incorporate the biochemical tuber data into the available physiology-informed models<sup>11</sup> calibrated for specific potato varieties.



		2019							2020							2021						
		Tested on							Tested on							Tested on						
Trained on		M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	0.75		0.67	0.80	0.93	0.88	0.94	0.51		0.50	0.49	0.45	0.54	0.45		0.43	0.42				0.46
	S	0.76			0.69	0.82		0.72	0.61	0.47	0.58	0.62	0.59	0.63	0.58	-0.40	0.62	0.37				0.56
	V	0.75		0.76	0.83	0.84		0.88	0.56	0.44	0.59	0.59	0.54	0.61	0.57	-0.51	0.63	0.39				0.57
	AVG	0.88	0.79	0.94		0.93		0.95	0.47	0.39	0.43	0.48	0.46	0.48	0.45	-0.49	0.64	0.48				0.62
	MS	0.77	0.83	0.92	0.90	0.83	0.95	0.97	0.49	0.44	0.43	0.49	0.50	0.49	0.46	-0.65	0.54				-0.45	0.43
	MV	0.79	0.73	0.82		0.85	0.99	0.73	0.63	0.45	0.61	0.63	0.59	0.66	0.59	-0.48	0.65	0.43				0.60
	SV	0.95	0.69			0.81	0.78	0.95								-0.41	0.36					
2020	M	0.53	0.51	0.53	0.55	0.55	0.54	0.54		0.76	0.71		0.78	0.77								
	S	-0.63	-0.45	-0.57	-0.60	-0.60	-0.62	-0.53					0.78									
	V	0.66	0.69	0.64	0.70	0.71	0.67	0.70	0.88	0.93		0.91		0.73			0.65	0.63	0.60	0.47		0.69
	AVG											0.80	0.90				0.46					0.44
	MS	0.62	0.65	0.56	0.65	0.67	0.61	0.64		0.85	0.77	0.82	0.84				0.72	0.63	0.63	0.53		0.73
	MV	0.50	0.57	0.50	0.55	0.55	0.51	0.57	0.78			0.71	0.76		0.80		0.51	0.54	0.46			0.57
	SV	0.39	0.55	0.50	0.49	0.47	0.44	0.55	0.91	0.88	0.79				0.74		0.53	0.43	0.44	0.37		0.53
2021	M	-0.66	-0.73	-0.74	-0.74	-0.73	-0.70	-0.77	-0.49	-0.52	-0.49	-0.54	-0.53	-0.52	-0.54	0.72					0.79	
	S	0.81	0.82	0.80	0.86	0.86	0.82	0.85	0.47	0.60	0.43	0.53	0.55	0.48	0.53	0.68						0.75
	V										0.39								0.76			
	AVG		-0.42	-0.45	-0.42	-0.40	-0.40	-0.46														
	MS								0.39		0.46	0.41		0.45	0.40	0.88		0.92	0.75			
	MV		-0.41	-0.49	-0.44	-0.42	-0.43	-0.47								-0.76		0.80		0.74		
	SV															0.78	0.75	0.84				0.83

		2019							2020							2021						
		Tested on							Tested on							Tested on						
Trained on		M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M								0.60	0.48	0.71	0.67	0.58	0.70	0.66	0.51	0.59	0.48				0.58
	S															0.39	0.47	0.50	0.44			0.46
	V																					
	AVG								0.45	0.49		0.39	0.49			0.71		0.52	0.56			0.61
	MS															0.51	0.44	0.41				0.52
	MV								0.54	0.63	0.50	0.60	0.61	0.55	0.59	0.51			0.37			0.43
	SV		0.73			0.80			0.52	0.38	0.63	0.57	0.49	0.61	0.57		0.49	0.54	0.48	0.42	0.43	
2020	M	0.73	0.53	0.62	0.68	0.69	0.71	0.60				0.76				0.64	0.37	0.41				0.57
	S	0.57	0.39	0.46	0.52	0.53	0.54	0.44	0.71									0.38	0.44			
	V															0.45	0.48	0.45				0.50
	AVG	0.59		0.41	0.50	0.52	0.54	0.40	0.77			0.72				0.47	0.42	0.56	0.56			0.49
	MS	0.58		0.40	0.47	0.48	0.53				0.84			0.82		0.42		0.40	0.43			0.39
	MV															0.62	0.49	0.56	0.52			0.61
	SV															0.45						
2021	M	0.64	0.40	0.55	0.58	0.58	0.62	0.50								0.82	0.80	0.83	0.76	0.87	0.79	
	S		0.46					0.40	0.49		0.40	0.42		0.42								0.76
	V	0.68	0.73	0.67	0.73	0.74	0.69	0.74	0.37	0.37	0.52	0.46	0.39	0.47	0.49			0.85	0.85	0.84	0.85	
	AVG															0.88						
	MS										0.43			0.41								
	MV	0.43				0.40	0.39											0.90	0.84	0.77	0.88	0.84
	SV																					

**Fig. 17.** Performance (Pearson correlation  $c$ ) of the variety-specific Festien model (top) and the variety-agnostic model tested on the Festien variety (bottom). Rows correspond to the training dataset, columns show the testing configuration. Correlations with  $p \geq 0.05$  are omitted.

## Conclusion

The hypothesis of the project that the biochemical analysis of seed tubers provides information about the vigor of the resulting potato plants has been confirmed. The prediction of seedlot vigor by a data-driven multiple linear regression model holds across different environments as long as the variety performance is consistent in these environments. Such a consistent performance is observed for one of the studied varieties (Festien).

Trained on	Tested on	2019							2020							2021						
		M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	0.42	0.34	1.54	0.20	0.85	0.51	1.42	2.08	2.11	2.30	2.22	2.11	2.27	2.24	1.70	1.16	1.54	1.39	1.36	1.57	1.29
	S	2.15	2.86	1.92	0.52	0.85	1.49	1.16	1.63	1.03	1.75	1.35	1.18	1.66	1.33	1.58	0.75	1.30	1.09	1.07	1.37	0.94
	V	1.16	0.80	1.29	0.84	0.71	1.94	2.44	1.65	1.72	2.09	1.83	1.67	1.93	1.91	2.37	1.59	1.80	1.89	1.97	2.06	1.67
	AVG	0.02	0.44	0.66	0.58	0.62	0.05	1.10	1.89	1.39	0.91	1.14	1.53	1.10	1.00	1.27	0.66	1.18	0.90	0.85	1.15	0.82
	MS	2.28	3.27	1.91	2.69	0.85	0.19	1.14	2.38	1.85	1.40	1.72	2.04	1.68	1.55	2.56	1.62	2.04	2.08	2.09	2.29	1.83
	MV	0.86	0.38	0.20	1.40	1.65	0.82	1.89	1.73	2.06	2.74	2.34	1.94	2.48	2.49	1.72	2.12	1.70	1.81	1.91	1.68	1.88
	SV	0.31	2.84	0.80	1.41	1.84	0.47	2.00	1.58	1.35	1.75	1.48	1.38	1.64	1.50	2.10	1.14	1.52	1.51	1.57	1.76	1.27
2020	M	2.59	2.41	2.79	2.61	2.52	2.71	2.60	0.98	1.50	3.52	0.92	1.90	1.33	0.42	1.70	2.09	1.95	1.91	1.88	1.83	2.02
	S	1.79	1.77	1.73	1.75	1.77	1.76	1.74	1.75	2.50	2.31	0.53	0.74	0.56	1.28	0.96	1.14	1.23	0.99	0.95	1.03	1.09
	V	2.69	2.48	2.75	2.67	2.61	2.75	2.63	2.59	1.75	1.92	1.26	0.04	1.26	0.37	1.87	1.42	1.75	1.63	1.60	1.79	1.55
	AVG	2.18	1.80	2.35	2.08	1.99	2.26	2.01	2.59	1.04	0.92	1.02	1.00	0.18	1.01	1.33	1.16	1.68	1.32	1.16	1.48	1.38
	MS	2.51	2.12	2.68	2.43	2.33	2.61	2.36	2.27	0.50	1.27	1.16	0.29	0.31	0.53	0.65	1.45	1.48	1.10	0.95	1.03	1.40
	MV	2.44	2.08	2.61	2.37	2.27	2.53	2.31	2.18	0.74	0.30	1.16	1.69	2.08	0.33	1.14	1.39	1.58	1.29	1.18	1.33	1.43
	SV	2.24	1.91	2.37	2.16	2.08	2.30	2.09	1.79	0.81	0.75	1.28	0.99	0.59	0.54	1.51	1.87	1.66	1.64	1.66	1.56	1.73
2021	M	2.38	2.25	2.42	2.37	2.33	2.41	2.34	1.32	1.11	1.51	1.19	1.12	1.35	1.23	1.17	1.48	1.61	1.60	1.75	0.95	2.31
	S	1.27	0.85	1.42	1.08	1.02	1.29	1.00	1.77	1.02	1.69	1.35	1.23	1.67	1.29	2.12	3.07	1.97	1.44	0.55	2.48	0.85
	V	1.71	1.71	1.52	1.64	1.70	1.62	1.61	2.01	2.14	2.88	2.52	2.10	2.70	2.62	0.34	1.96	2.85	0.55	1.94	0.55	0.62
	AVG	2.07	1.97	2.23	2.08	2.02	2.14	2.08	2.36	0.98	1.98	2.01	2.03	2.12	1.90	1.73	1.82	0.57	1.28	1.01	1.53	1.42
	MS	2.69	2.27	2.66	2.56	2.50	2.72	2.45	2.33	1.75	1.60	1.77	1.96	1.82	1.62	0.49	2.24	1.29	0.29	1.64	0.56	1.52
	MV	1.31	1.21	1.24	1.20	1.23	1.25	1.16	1.34	0.97	2.25	1.51	1.03	1.95	1.61	1.46	1.12	1.39	0.65	0.42	1.24	1.65
	SV	1.26	1.15	1.36	1.18	1.17	1.26	1.17	1.68	0.96	2.17	1.55	1.16	2.01	1.56	1.22	0.75	2.17	0.64	1.96	1.52	0.89

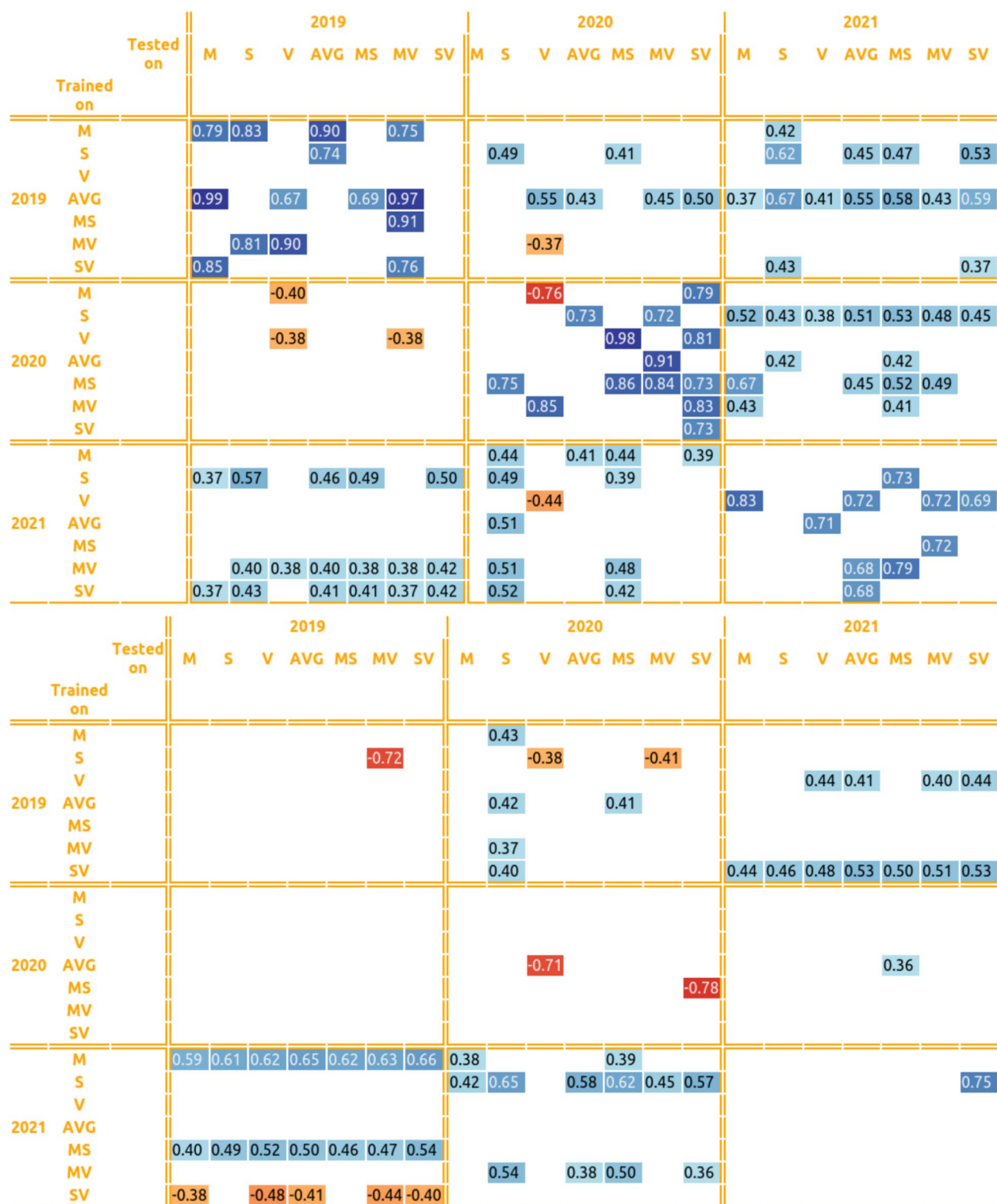
  

Trained on	Tested on	2019							2020							2021						
		M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	2.06	2.06	2.89	2.50	2.06	1.64	1.26	1.96	1.14	1.98	1.60	1.39	1.97	1.53	1.49	1.74	1.69	1.59	1.58	1.57	1.68
	S	1.52	1.48	1.16	3.25	2.71	3.43	3.36	2.56	2.26	2.76	2.64	2.40	2.81	2.60	2.61	1.90	2.10	2.22	2.27	2.36	2.01
	V	1.91	1.64	1.34	1.19	1.23	1.11	2.87	2.23	1.56	2.60	2.17	1.78	2.56	2.14	1.46	1.32	1.13	1.18	1.32	1.20	1.12
	AVG	1.99	1.81	3.05	2.45	0.87	0.73	1.06	1.39	1.17	2.48	1.73	1.19	2.14	1.85	2.24	1.50	1.45	1.66	1.85	1.78	1.41
	MS	0.65	1.66	1.96	1.75	1.66	1.61	2.06	1.91	1.50	2.42	1.97	1.62	2.29	2.00	1.77	1.67	1.38	1.52	1.68	1.51	1.45
	MV	2.14	2.14	1.50	1.83	2.14	0.75	1.19	1.78	1.26	1.88	1.56	1.40	1.82	1.54	2.32	1.51	1.83	1.86	1.90	2.06	1.65
	SV	2.80	2.26	2.32	1.90	1.79	1.75	2.17	2.09	1.19	2.13	1.74	1.48	2.14	1.65	1.13	1.07	1.04	0.93	1.00	0.99	0.94
2020	M	1.95	1.55	1.99	1.79	1.74	1.96	1.70	2.73	2.04	2.11	1.29	1.96	2.07	2.11	2.09	1.31	1.71	1.65	1.65	1.87	1.47
	S	2.47	2.03	2.58	2.36	2.27	2.54	2.26	0.74	2.22	2.37	3.08	1.35	3.13	2.80	2.26	1.46	1.91	1.85	1.83	2.07	1.67
	V	2.11	2.26	2.17	2.19	2.19	2.14	2.24	2.28	2.38	2.33	1.08	0.79	1.21	1.49	2.58	2.29	2.37	2.47	2.48	2.51	2.37
	AVG	2.46	1.98	2.57	2.33	2.23	2.53	2.22	2.55	2.45	3.42	2.75	2.52	2.68	2.96	1.31	1.39	1.58	1.35	1.28	1.41	1.43
	MS	1.79	1.41	1.85	1.63	1.59	1.80	1.55	1.39	1.56	1.23	2.90	1.44	1.19	3.56	2.30	1.63	1.94	1.94	1.95	2.11	1.77
	MV	1.84	1.43	1.89	1.68	1.62	1.85	1.58	0.97	1.58	3.39	3.04	2.49	3.23	2.83	1.74	1.62	1.72	1.65	1.64	1.71	1.63
	SV	1.92	1.87	2.00	1.92	1.89	1.95	1.92	2.64	0.91	2.45	1.87	1.13	1.98	1.88	2.46	1.64	1.93	2.00	2.05	2.18	1.78
2021	M	0.83	0.77	0.76	0.70	0.75	0.74	0.67	1.25	1.31	2.34	1.68	1.22	1.98	1.84	1.41	2.45	2.58	2.39	2.28	2.86	1.89
	S	2.48	2.22	2.48	2.41	2.37	2.51	2.34	1.16	0.69	1.28	0.84	0.76	1.11	0.86	1.44	1.59	0.69	0.96	1.18	1.22	0.49
	V	2.56	2.13	2.67	2.46	2.36	2.64	2.36	1.86	1.29	1.79	1.55	1.45	1.78	1.49	2.58	1.53	1.76	0.84	0.87	1.24	0.77
	AVG	2.28	2.00	2.05	2.13	2.15	2.20	2.02	1.64	1.30	1.97	1.58	1.37	1.83	1.61	1.73	0.72	2.06	1.63	1.41	2.54	0.79
	MS	1.19	1.02	0.97	1.00	1.07	1.06	0.92	2.31	1.56	2.29	2.04	1.82	2.35	1.95	1.50	0.85	1.58	1.15	1.69	1.51	1.11
	MV	2.38	1.96	2.53	2.28	2.18	2.46	2.20	1.37	0.93	1.74	1.24	1.01	1.55	1.27	1.03	1.75	1.44	1.37	1.36	1.20	1.63
	SV	2.76	2.61	2.96	2.81	2.72	2.88	2.80	1.56	1.45	1.93	1.61	1.45	1.78	1.67	1.72	1.22	1.43	1.28	1.27	2.37	1.82

**Fig. 18.** Performance (SSE residual  $r^2$ ) of the variety-specific Innovator model (top) and the variety-agnostic model tested on the Innovator variety (bottom). Rows correspond to the training dataset, columns show the testing configuration. SSE residuals  $r^2 \geq 1$  are grayed out.

Hence, our goal of creating a high-throughput proxy for the germination test has been achieved for this variety. Seedlots of other varieties (Challenger, Colomba, Innovator, Sagitta, and Seresta) did not perform consistently across environments, and their vigor could not be predicted by the seed-tuber properties to the same extent as for Festien. Nevertheless, our previous study<sup>6</sup> suggests that the obtained prediction accuracy for these varieties might still be sufficient to achieve a coarse categorization of seedlots into low, medium, and high vigor. Through





**Fig. 19.** Performance (Pearson correlation  $c$ ) of the variety-specific Innovator model (top) and the variety-agnostic model tested on the Innovator variety (bottom). Rows correspond to the training dataset, columns show the testing configuration. Correlations with  $p \geq 0.05$  are omitted.

variable selection we identified the data obtained with FTIR spectroscopy as the most useful dataset for vigor prediction. This is a fast and high-throughput methodology which could enable practitioners to make vigor predictions at scale. In future research, the variations of seedlot vigor across different environments can be accounted for by a physiology-informed plant-growth model incorporating environmental data. Also, a more robust dynamic measure of the plant vigor should be considered.

Trained on	Tested on	2019							2020							2021						
		M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	1.02	1.89	0.29	0.95	1.01	1.04	1.00	1.42	1.30	1.47	1.34	1.30	1.39	1.37	2.03	1.93	1.79	1.86	1.91	1.90	1.77
	S	0.53	1.19	0.51	1.01	1.23	0.34	1.78	1.15	1.08	1.25	1.08	1.05	1.11	1.13	1.95	1.44	1.31	1.52	1.68	1.61	1.32
	V	1.72	1.29	0.50	0.23	1.84	1.64	0.26	1.92	1.75	1.46	1.63	1.81	1.58	1.57	2.76	2.44	2.07	2.50	2.69	2.49	2.27
	AVG	0.14	1.16	1.27	1.43	1.49	0.50	1.57	2.50	2.16	1.86	2.13	2.33	2.11	1.99	3.00	2.56	2.28	2.72	2.89	2.73	2.45
	MS	0.96	0.43	0.94	1.55	1.14	1.01	0.56	2.03	1.52	1.36	1.53	1.72	1.56	1.40	2.55	1.82	1.47	1.97	2.24	2.04	1.61
	MV	0.80	0.42	0.60	1.45	1.42	0.26	0.65	1.92	1.51	1.38	1.51	1.66	1.53	1.41	2.05	1.58	1.55	1.70	1.81	1.80	1.53
	SV	1.04	1.09	0.35	0.66	1.64	0.41	1.37	2.86	2.63	2.34	2.63	2.78	2.60	2.49	3.09	2.85	2.53	2.96	3.09	2.92	2.74
2020	M	1.77	2.16	2.62	2.28	1.97	2.29	2.48	1.92	1.32	0.96	0.77	0.40	1.54	0.45	1.74	1.28	1.48	1.44	1.48	1.57	1.33
	S	2.03	2.14	2.67	2.38	2.09	2.45	2.50	1.02	2.05	0.42	1.12	0.71	0.76	0.31	0.98	1.00	0.98	0.83	0.78	0.73	0.63
	V	1.72	1.83	1.87	1.78	1.75	1.79	1.83	1.89	1.92	1.27	0.87	1.11	1.10	0.51	2.00	1.48	1.63	1.67	1.73	1.80	1.52
	AVG	2.55	2.16	2.73	2.59	2.39	2.72	2.56	1.42	1.65	1.12	0.94	0.48	0.46	2.06	1.36	1.38	0.99	1.13	1.29	1.09	1.11
	MS	2.25	2.07	2.51	2.36	2.18	2.45	2.37	0.84	1.43	1.62	0.24	1.95	0.67	0.46	1.60	1.29	0.94	1.18	1.39	1.21	1.03
	MV	1.85	2.07	1.88	1.91	1.96	1.85	1.95	0.91	1.45	0.46	0.33	0.74	0.76	2.49	1.76	1.45	1.07	1.36	1.57	1.37	1.20
	SV	2.31	2.17	2.75	2.53	2.27	2.63	2.58	0.53	0.14	0.43	0.40	0.69	1.51	0.07	1.32	1.22	0.90	1.02	1.18	1.02	0.97
2021	M	2.27	2.00	2.78	2.47	2.15	2.63	2.51	1.21	1.34	1.43	1.27	1.23	1.27	1.36	1.92	0.50	0.76	2.86	1.01	1.61	2.32
	S	1.88	1.89	2.33	2.03	1.87	2.16	2.16	1.96	1.93	1.77	1.86	1.94	1.82	1.84	1.03	0.34	0.76	0.14	1.18	0.90	1.39
	V	2.02	1.96	1.97	1.98	1.99	1.99	1.96	1.06	0.82	0.76	0.72	0.84	0.73	0.72	2.80	2.78	0.24	1.92	0.46	0.88	0.25
	AVG	1.76	1.74	1.98	1.82	1.72	1.88	1.86	1.15	0.90	0.92	0.85	0.93	0.88	0.85	1.66	0.42	1.71	1.92	0.96	2.31	1.39
	MS	1.96	2.00	2.52	2.24	1.98	2.32	2.34	1.13	1.41	1.67	1.39	1.24	1.41	1.54	0.42	0.54	1.07	1.18	1.57	0.86	1.73
	MV	1.81	1.89	2.19	1.99	1.84	2.04	2.07	1.01	0.92	0.99	0.86	0.88	0.87	0.91	0.30	1.34	0.99	1.74	1.47	2.50	0.34
	SV	2.03	1.94	2.53	2.24	1.98	2.36	2.32	0.91	1.03	1.12	0.93	0.90	0.92	1.04	1.11	2.50	0.79	0.73	1.22	0.81	0.46

Trained on	Tested on	2019							2020							2021						
		M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	0.53	0.46	0.38	0.29	0.29	0.82	0.89	1.39	1.06	1.25	1.13	1.14	1.22	1.13	1.49	1.28	1.17	1.22	1.32	1.26	1.16
	S	2.26	2.71	2.61	0.37	0.37	0.33	0.49	1.30	1.06	1.19	1.08	1.10	1.13	1.08	2.44	2.01	1.75	2.09	2.27	2.12	1.87
	V	1.64	2.31	2.63	2.62	2.61	2.74	1.37	1.14	1.29	1.43	1.23	1.17	1.24	1.33	1.53	1.33	1.63	1.43	1.38	1.53	1.44
	AVG	1.60	2.99	1.69	1.88	0.45	0.79	0.58	1.09	1.66	1.82	1.54	1.38	1.50	1.74	2.11	1.76	2.24	2.05	1.94	2.19	2.01
	MS	0.88	0.22	0.89	0.91	0.97	0.70	1.22	1.40	1.85	2.09	1.82	1.63	1.81	1.98	1.53	1.66	1.93	1.66	1.54	1.69	1.78
	MV	2.76	2.62	2.87	2.77	2.72	1.04	1.05	1.54	1.33	1.34	1.32	1.38	1.34	1.30	2.20	1.71	1.97	1.97	1.97	2.11	1.83
	SV	0.67	0.92	1.30	0.67	0.65	0.79	0.61	0.99	0.85	0.84	0.75	0.83	0.76	0.79	1.40	1.07	0.79	0.96	1.16	1.00	0.83
2020	M	1.66	2.21	2.29	2.10	1.94	2.03	2.29	3.38	3.03	3.10	1.28	2.64	2.77	3.15	1.90	1.71	1.84	1.80	1.79	1.86	1.76
	S	1.80	2.32	2.38	2.23	2.08	2.15	2.41	1.27	1.15	1.08	0.80	1.18	0.85	0.71	2.12	1.77	1.81	1.90	1.95	1.97	1.77
	V	1.86	2.21	2.31	2.18	2.05	2.14	2.30	1.89	2.74	1.38	1.51	1.62	1.46	1.47	2.28	2.18	1.94	2.16	2.26	2.14	2.06
	AVG	1.93	2.36	2.69	2.44	2.17	2.41	2.63	1.95	2.53	2.26	2.02	2.19	2.01	2.33	1.12	1.15	1.04	0.97	1.04	0.98	1.02
	MS	2.21	2.12	2.76	2.48	2.18	2.59	2.55	1.51	1.55	1.11	0.92	1.52	1.28	1.05	1.83	1.65	1.28	1.53	1.72	1.52	1.41
	MV	2.42	2.46	2.70	2.65	2.50	2.64	2.68	0.51	0.68	0.45	1.07	1.28	0.82	1.35	1.43	1.50	1.19	1.28	1.40	1.24	1.28
	SV	2.06	2.43	2.89	2.61	2.29	2.61	2.80	0.35	1.56	0.97	0.76	1.81	1.33	1.11	1.93	2.14	2.43	2.18	2.03	2.19	2.32
2021	M	2.44	2.14	2.54	2.46	2.32	2.55	2.42	1.27	1.11	1.07	1.03	1.12	1.04	1.04	2.94	0.96	2.62	2.53	2.41	1.75	2.73
	S	2.03	2.23	1.72	1.96	2.15	1.83	1.93	1.52	1.21	1.19	1.20	1.29	1.22	1.16	2.28	1.89	3.02	2.66	2.12	1.65	2.41
	V	1.45	1.90	1.59	1.58	1.64	1.49	1.68	1.21	0.79	0.78	0.77	0.89	0.81	0.73	2.37	0.65	0.56	1.07	1.16	1.39	1.04
	AVG	1.96	1.84	1.47	1.67	1.88	1.64	1.57	1.34	0.95	0.86	0.90	1.05	0.92	0.84	1.82	0.59	0.69	0.91	1.12	1.93	0.98
	MS	2.81	2.00	2.22	2.38	2.44	2.51	2.14	1.21	1.24	0.87	0.97	1.17	0.87	0.98	2.32	1.19	1.62	1.22	2.29	2.06	2.08
	MV	2.17	2.21	2.21	2.24	2.21	2.21	2.24	1.26	1.28	1.35	1.23	1.22	1.23	1.28	1.29	1.91	1.66	1.57	1.58	1.35	1.90
	SV	1.63	2.26	2.23	2.08	1.95	1.98	2.28	1.88	2.01	2.40	2.16	1.95	2.24	2.24	2.70	0.53	0.86	0.71	0.76	1.02	2.93

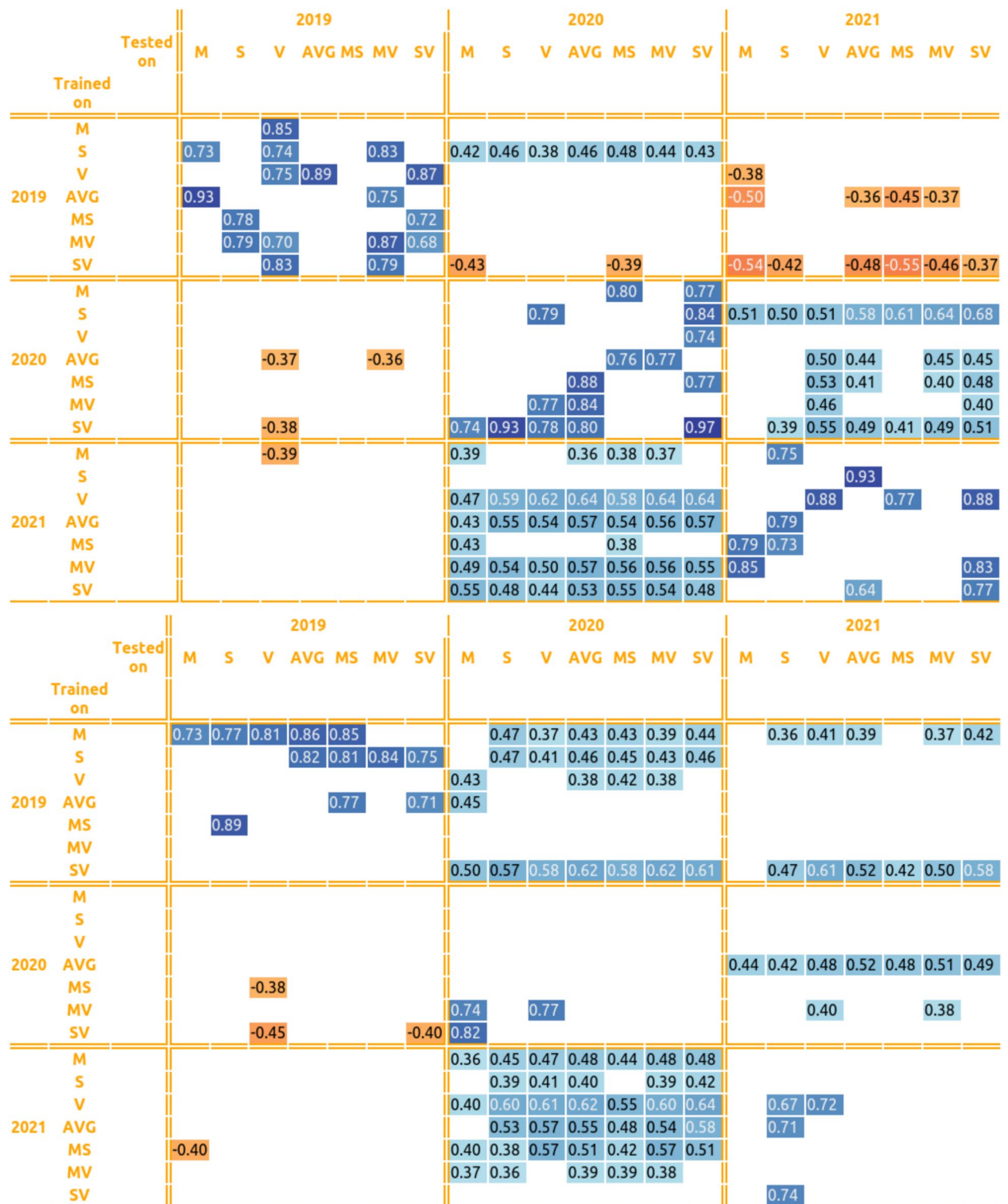
**Fig. 20.** Performance (SSE residual  $r^2$ ) of the variety-specific Sagitta model (top) and the variety-agnostic model tested on the Sagitta variety (bottom). Rows correspond to the training dataset, columns show the testing configuration. SSE residuals  $r^2 \geq 1$  are grayed out.

## Methods

### Measuring the plant vigor

#### Design of the field experiments

In this project, six potato varieties (Challenger, Colomba, Festien, Innovator, Sagitta, and Seresta) were studied, each represented by 30 seedlots of different production origin. The tubers from these seedlots were planted in



**Fig. 21.** Performance (Pearson correlation  $c$ ) of the variety-specific Sagitta model (top) and the variety-agnostic model tested on the Sagitta variety (bottom). Rows correspond to the training dataset, columns show the testing configuration. Correlations with  $p \geq 0.05$  are omitted.

three test fields, according to a split-plot in randomized complete block design, see Fig. 1. The test fields were located in Montfrin (France), Veenklooster (Netherlands), and Kolummerwaard (SPNA, Netherlands), and will be denoted as the fields M, V, and S, respectively. The experiment, including the selection of the new 30 seedlots of each variety, biochemical and microbial analysis of the tubers, and the subsequent planting and observation in the three test fields, was repeated over three consecutive years (2019, 2020, and 2021). Each of the experimental



		2019							2020							2021						
Trained on	Tested on	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	0.90	0.61	1.06	1.48	0.63	1.90	0.27	1.98	1.83	1.32	1.62	1.89	1.56	1.47	2.33	2.07	1.87	2.14	2.27	2.15	1.98
	S	0.60	3.06	1.50	2.23	0.44	0.57	1.04	1.67	1.95	1.88	1.80	1.79	1.75	1.90	2.04	1.50	1.81	1.69	1.67	1.92	1.59
	V	1.03	1.60	1.40	1.56	2.60	0.35	0.93	1.34	1.36	1.87	1.46	1.30	1.57	1.58	2.33	2.37	2.17	2.40	2.48	2.31	2.32
	AVG	2.04	0.90	1.57	2.69	1.15	1.30	0.50	1.76	1.44	1.67	1.55	1.57	1.66	1.49	2.61	2.14	2.29	2.46	2.50	2.56	2.23
	MS	0.42	0.34	1.19	1.59	1.27	0.48	1.03	1.55	1.49	1.39	1.36	1.48	1.37	1.34	2.59	2.60	2.18	2.65	2.82	2.50	2.47
	MV	1.20	1.54	0.97	0.40	0.56	0.76	1.41	1.97	1.65	1.73	1.74	1.80	1.81	1.64	2.66	2.64	2.34	2.76	2.89	2.63	2.57
	SV	1.19	0.22	2.13	1.68	1.23	1.45	0.42	2.13	1.72	1.83	1.87	1.93	1.96	1.74	2.33	2.33	1.93	2.29	2.46	2.18	2.18
2020	M	1.50	2.01	1.43	1.54	1.71	1.41	1.60	3.44	0.71	2.84	1.34	0.90	1.34	0.62	2.44	2.65	2.20	2.61	2.76	2.41	2.51
	S	1.87	1.88	2.63	2.22	1.86	2.33	2.39	0.82	0.99	2.20	0.68	0.41	0.37	1.29	2.54	1.92	2.34	2.34	2.29	2.55	2.11
	V	2.07	2.18	1.35	1.76	2.12	1.63	1.61	1.19	1.08	2.52	1.00	1.61	1.15	1.35	2.53	2.66	2.55	2.80	2.82	2.67	2.69
	AVG	1.30	1.61	1.44	1.36	1.40	1.32	1.44	0.64	0.90	1.10	2.31	2.20	2.84	1.12	2.60	2.49	2.13	2.58	2.74	2.47	2.38
	MS	1.67	1.87	1.58	1.64	1.75	1.58	1.65	0.52	1.40	1.51	1.85	2.12	0.84	0.60	2.24	2.45	2.10	2.38	2.48	2.21	2.33
	MV	2.44	2.20	1.53	1.99	2.36	1.92	1.75	0.67	0.28	1.49	0.03	1.89	1.27	1.08	2.02	2.21	1.56	1.95	2.17	1.76	1.92
	SV	2.27	2.07	2.01	2.12	2.19	2.13	2.04	1.81	0.73	2.07	0.60	1.51	0.58	1.94	2.02	1.88	2.13	2.00	1.92	2.09	1.99
2021	M	2.74	2.48	1.93	2.37	2.66	2.31	2.15	2.01	1.93	1.81	1.89	1.97	1.88	1.84	2.25	2.80	0.49	1.41	1.17	1.29	0.44
	S	2.38	2.17	1.98	2.18	2.31	2.17	2.05	2.41	2.31	1.49	2.04	2.39	1.90	1.83	1.99	1.17	1.04	0.86	1.89	1.10	0.89
	V	2.59	2.35	1.62	2.13	2.51	2.04	1.87	2.11	1.92	1.84	1.94	2.02	1.96	1.85	0.90	1.65	0.86	0.97	1.28	0.48	1.51
	AVG	2.62	2.25	2.41	2.50	2.49	2.55	2.39	2.02	2.00	1.73	1.89	2.01	1.84	1.83	0.67	0.84	1.42	1.63	2.23	0.33	1.78
	MS	2.89	2.57	1.89	2.44	2.80	2.36	2.16	2.22	2.11	2.22	2.22	2.18	2.25	2.20	1.44	1.95	0.74	0.91	1.63	0.96	1.59
	MV	3.10	2.65	2.19	2.68	2.96	2.64	2.40	2.04	1.82	1.92	1.91	1.92	1.97	1.85	2.01	1.04	1.03	0.73	2.45	1.76	1.06
	SV	2.83	2.29	1.75	2.27	2.63	2.24	1.94	2.20	2.04	1.90	2.05	2.13	2.04	1.96	1.50	1.84	0.22	0.40	1.15	0.70	1.37

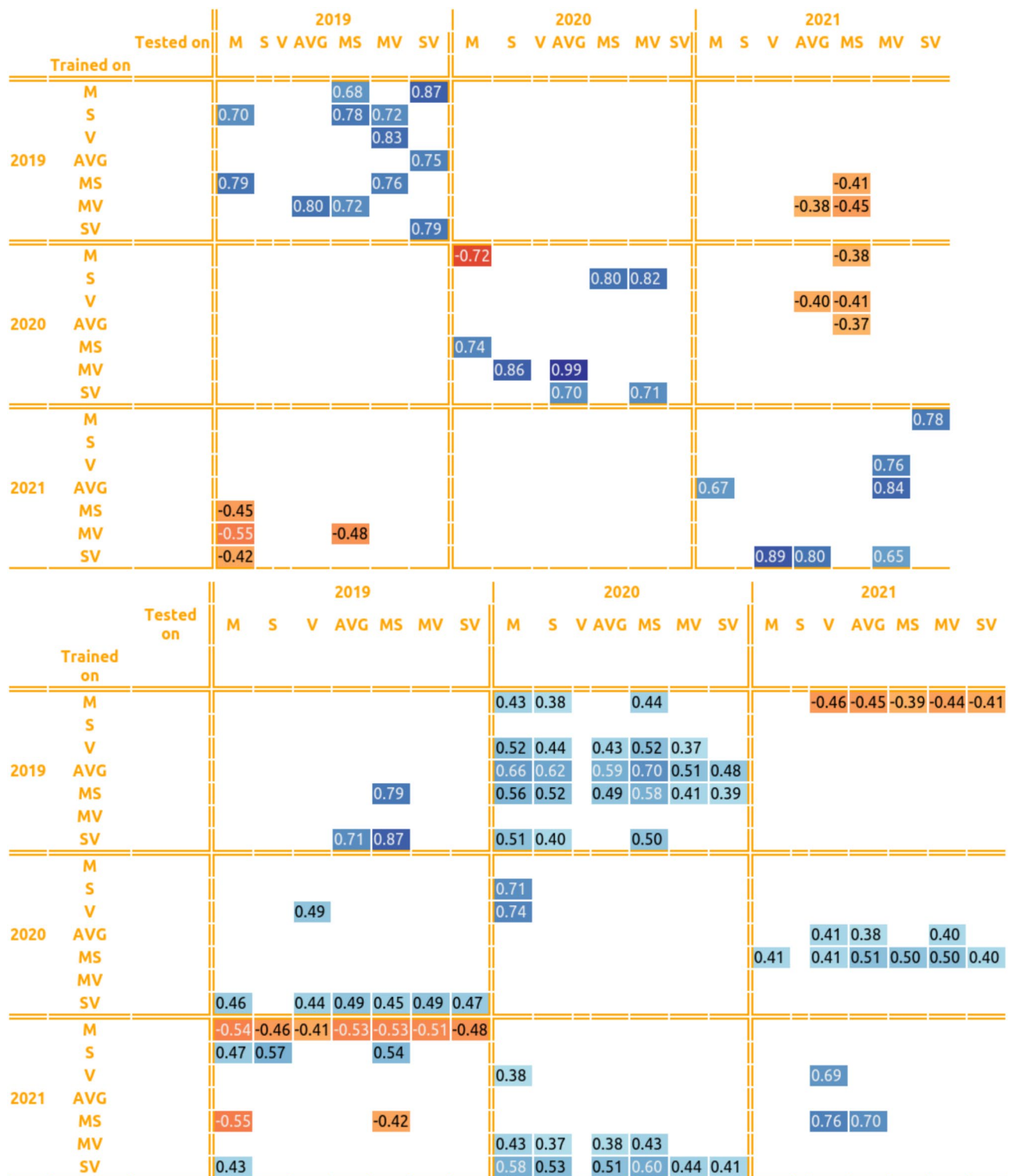
  

		2019							2020							2021						
Trained on	Tested on	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV	M	S	V	AVG	MS	MV	SV
2019	M	0.63	0.87	1.92	1.22	0.60	1.66	1.79	1.14	1.25	2.30	1.53	1.13	1.73	1.80	2.55	2.59	2.92	2.91	2.78	2.89	2.82
	S	2.05	2.64	2.27	1.15	1.10	1.05	1.29	1.73	1.89	1.93	1.82	1.79	1.81	1.89	1.95	1.82	2.24	1.98	1.84	2.10	2.00
	V	1.03	0.94	2.18	1.30	0.82	1.53	2.95	0.95	1.13	1.71	1.15	0.96	1.26	1.35	2.36	2.67	2.31	2.62	2.71	2.41	2.57
	AVG	1.57	1.16	2.48	1.97	1.35	1.18	1.23	0.67	0.76	1.52	0.82	0.61	0.99	1.04	1.98	2.00	1.65	1.86	1.99	1.79	1.83
	MS	1.99	2.23	2.09	1.08	0.43	1.27	1.53	0.89	0.96	1.64	1.03	0.83	1.18	1.22	2.22	2.69	2.23	2.55	2.65	2.28	2.55
	MV	1.86	1.79	1.90	1.85	1.81	1.02	1.02	1.40	1.66	2.37	1.81	1.49	1.91	2.06	2.11	1.75	2.04	1.93	1.89	2.09	1.86
	SV	0.66	1.24	1.56	0.58	0.26	0.60	0.92	0.99	1.19	1.95	1.29	1.01	1.42	1.54	1.78	1.96	2.10	1.92	1.83	1.92	2.03
2020	M	1.23	1.28	1.77	1.40	1.21	1.49	1.55	1.79	1.65	1.18	1.57	1.58	1.30	1.36	1.87	1.73	1.70	1.69	1.72	1.74	1.68
	S	1.26	1.44	2.28	1.71	1.31	1.82	1.97	0.58	1.11	1.10	1.33	0.69	1.30	1.46	1.54	1.61	1.50	1.40	1.42	1.41	1.51
	V	1.74	1.81	1.02	1.38	1.76	1.26	1.23	0.52	0.84	2.87	2.67	2.45	2.93	2.61	2.41	2.00	2.40	2.34	2.26	2.50	2.19
	AVG	1.44	1.57	2.29	1.81	1.47	1.91	2.03	1.15	1.69	1.86	1.09	1.41	1.19	1.08	1.49	1.59	1.18	1.23	1.37	1.20	1.34
	MS	2.70	2.55	2.45	2.63	2.67	2.61	2.54	0.92	0.93	2.09	2.64	0.84	1.57	2.40	1.19	1.36	1.18	0.98	1.01	1.00	1.20
	MV	2.44	2.09	2.12	2.24	2.30	2.28	2.12	2.66	2.73	3.05	1.48	1.24	1.58	1.60	2.07	2.24	2.01	2.16	2.22	2.05	2.16
	SV	1.09	1.22	1.13	1.02	1.09	1.02	1.07	0.77	1.88	0.98	1.11	1.43	0.94	1.30	1.99	1.94	1.60	1.81	1.94	1.77	1.77
2021	M	3.07	2.91	2.83	3.05	3.06	3.02	2.96	1.97	1.93	1.90	1.92	1.95	1.92	1.90	2.32	1.45	1.32	1.65	1.95	1.46	1.61
	S	1.06	0.85	1.82	1.25	0.92	1.45	1.41	1.82	1.86	2.28	2.01	1.83	2.08	2.11	1.77	2.24	2.47	2.14	1.97	1.44	1.11
	V	2.56	1.66	2.80	2.13	1.58	2.29	2.42	1.24	1.53	2.00	1.53	1.33	1.59	1.75	2.38	0.78	0.63	1.59	1.62	1.67	1.55
	AVG	1.08	1.59	2.62	2.22	1.86	2.43	2.27	1.55	1.79	1.72	1.62	1.64	1.59	1.71	1.69	1.92	1.86	1.74	1.70	1.73	0.87
	MS	3.09	2.41	2.35	2.69	2.84	2.74	2.41	1.98	1.94	2.05	1.99	1.95	2.02	2.00	0.99	1.07	0.49	0.59	2.33	2.49	2.09
	MV	1.67	1.52	2.60	2.03	1.58	2.21	2.23	1.15	1.26	1.65	1.24	1.13	1.33	1.39	1.03	1.75	1.16	1.21	1.39	1.13	1.23
	SV	1.15	1.37	1.53	1.28	1.20	1.30	1.41	0.83	0.94	1.60	0.98	0.79	1.12	1.18	3.11	2.63	1.30	2.29	2.99	1.89	1.86

**Fig. 22.** Performance (SSE residual  $r^2$ ) of the variety-specific Seresta model (top) and the variety-agnostic model tested on the Seresta variety (bottom). Rows correspond to the training dataset, columns show the testing configuration. SSE residuals  $r^2 \geq 1$  are grayed out.

fields was planted according to a split-plot in randomized complete block design. The planting was done with a semi-automated planting machine, tubers were sown with a distance of 25 cm in the vertical direction, along the ridge, and of 75 cm in the horizontal direction, inter-ridge. The resulting plots for a single seedlot are 3 meters wide and 1.5 meters long. The distance between consecutive plots in the ridge direction is of 50 cm. The fields were fertilized with 80 kg of Nitrogen fertilizer per hectare. Only the field of Montfrin was irrigated twice at the





**Fig. 23.** Performance (Pearson correlation  $c$ ) of the variety-specific Seresta model (top) and the variety-agnostic model tested on the Seresta variety (bottom). Rows correspond to the training dataset, columns show the testing configuration. Correlations with  $p \geq 0.05$  are omitted.

beginning of the season. Herbicide was applied once (Boxer + Proman). During the growing season fungicides were applied weekly. All seed tubers planted in the experiments have been planted whole and measured between 30 and 55 mm. The tubers were harvested in August of the year previous to planting, were sampled uniformly in October, and planted starting April of the experiment's year. For precise planting dates per field we refer to online data<sup>4</sup>.

	M	S	V
2019	52	48	36
2020	47	47	51
2021	54	30	58

**Table 1.** Measurement dates (in DAPs) per test field and year, when the canopy area is considered to be the seedlot vigor parameter.

*Drone-image segmentation and spatial correction*

The growth of the potato plants was monitored with the drone-mounted camera and the drone operator company delivered the stitched images (orthophotos) of each field taken approximately once per week. The leaf canopy of the potato plants was extracted from the images and the average canopy area per plant for each seedlot was calculated. The general image processing protocol and the example code demonstrating the extraction of the vigor data in the years 2019 and 2020 has been published in connection with the previous paper<sup>6,12</sup>. The results presented here include the canopy data from the year 2021, the extraction of which required additional processing steps. For uniformity, the new processing pipeline was then retrospectively applied to the years 2019 and 2020 as well. The updated detailed protocol for the extraction of the canopy area and the corresponding vigor and seed tuber data are available online<sup>4</sup>.

One of the image processing steps is the semi-automatic procedure that detects the boundaries of the seedlot plots. This was done in the image taken on the date when it was easy to detect the plots and distinguish between them. Due to the large scale of the experiment, the planting was performed from a moving tractor, which resulted in the plots having slightly irregular boundaries, that had to be represented by polygons rather than simple rectangles, see Fig. 2. Moreover, the irregularity of the planting scheme required some manual input while detecting the plot boundaries. To avoid repeating this laborious procedure in every snapshot of the same field, the field images were transformed to the same spatial reference frame with the help of permanent markers that were installed on the ground at the time of planting and later automatically detected in all images.

Once the plots were identified and assigned to the corresponding seedlots, the plot sub-images, see Fig. 2, were segmented by color to identify and count the plant pixels. This is achieved on the HSV representation of the image by selecting pixels for which the weighted average of some vegetation indices exceeds a threshold. The weights and thresholds are provided per measurement date in the protocol<sup>4</sup>. The raw canopy area for each plot is found by counting the identified canopy pixels and converting the result to cm<sup>2</sup> with the help of the known inter-ridge distances in each field. In the final step, the spatial trend induced by the test-field heterogeneity was removed from the raw canopy area data with the specialized R-package SpATS<sup>13</sup>, resulting in 180 canopy surface estimates per test field snapshot.

*Choice of the vigor parameter*

From the observational point of view, the relative difference in the vigor of potato plants can manifest itself in several ways. First, some seed tubers can sprout sooner than others. Second, the sprouts can grow at different rates, emerging above the ground at different times. Finally, the rates of growth of the leaf canopies can be different, so that canopies achieve different sizes at the moment when the plant blooms and produces the tubers. While the first two processes are certainly indicative of the seed tuber vigor, the size of the canopy at the time of tuber bulking is more important for the final tuber yield of the potato plant. That is why, in this project, we focused on the canopy size at a fixed date, later in the season.

By analyzing the canopy images taken on consecutive dates, it was found that the seedlots achieve a stable order in terms of the size of their canopy size towards the end of the exponential growth period, see Fig. 8. The date which showed a large (Kendall) correlation with the previous dates and, simultaneously, a significant variance between the canopies of the variety, was chosen as the date when the canopy area was taken to be the vigor parameter of the seedlot. The second condition is necessary because eventually the canopies of the neighboring plants merge, become indistinguishable in the drone images, and start competing for the incoming light.

It appears that in most cases the 47-th Day After Planting (DAP) is the best choice, since the plants of all six varieties have completely emerged, the canopy-based ordering of the seedlots has stabilized, and the variations between the seedlots are still significant. In the year 2019, with only a few dates available, the choice was mostly dictated by the seedlot order stability. In 2020 and 2021 the 47-th DAP was chosen where available and in the test field V in 2020, the choice was again based on the date of maximal order stability. The full list of selected dates is provided in the Table 1.

In Fig. 9 one can see how the choice of the DAP affects the (Pearson) correlation in the vigor parameters between the test fields—the quantity which is eventually displayed in Fig. 4. It is clear that, due to the adverse weather conditions in France in 2021, no choice DAP could significantly improve the correlations with the test field M in 2021.

**Tuber data**

The seed lots were collected from the various seed growers and stored at 8 °C. From each seed lot 4 batches of 5 tubers were washed, dried and peeled. Both the flesh and peel fraction was cut into small pieces using ceramic knives and frozen in liquid nitrogen. Two 50 ml tubes were filled with frozen potato flesh and peel respectively, weighed and stored at – 80 °C before freeze drying. A separate 150 ml container was filled with 100 g frozen

potato flesh, weighed and stored at  $-20^{\circ}\text{C}$  before freeze drying. All samples were weighed after freeze drying and dry matter percentages were calculated.

#### *XRF data*

From the 150 ml container, 8.3 grams of freeze dried potato flesh was milled in a zirconium oxide grinder and sieved. A mixture of 8.0 g powder and 2.0 g binder was well homogenized and pressed into a pellet. The dry weight concentrations of Mg, K, Ca, Fe, Cu, Zn, P, S, Cl and Mn were determined using energy dispersive X-ray fluorescence spectroscopy (Bruker S2 PUMA) and converted to concentrations in fresh weight using the dry matter percentage.

#### *FTIR data*

The freeze dried potato flesh of the 50 ml tube was ground to a fine powder and homogenized. The powder of each batch was put into 3 wells of a 96-well sample plate and measured in a high-throughput FTIR spectrometer (Bruker Tensor II + HTS-XT). Spectral outliers among the replicates were identified and the corresponding samples were remeasured.

#### *HSI data*

From each seed lot, 50 tubers were selected and a 1 cm wide longitudinal slice from the center of each tuber was cut. The slices were placed on a moving platform and scanned using a push broom SWIR hyperspectral camera (900–2500 nm, Specim). Before each scan, a dark and white reference was scanned and used to convert the data to absorbance units.

For each seedlot, a hyper-spectral image containing a number of tuber slices was acquired, see Fig. 10 (upper-left image). The sub-image of each individual slice was automatically detected and segmented into two biologically meaningful partitions: pith and cortex. The average spectral signature per partition was calculated as the mean of spectra for all pixels of that part. Then, the average of the spectral signatures over all slices in the image was taken. Thus, the HSI seedlot data represent the average spectral signatures of the cortex and pith of a seedlot, see Fig. 10 (bottom plot).

Technically, the most involved stage of this process is the automated segmentation of the slices into the pith and cortex parts. The outer boundary of each slice was easy to detect on the gray-scale image obtained by averaging over the spectral dimension, see Fig. 10 (upper-left image). Geometrically, the cortex part was defined as the domain of the slice bounded on the outside by the outer boundary of the slice and on the inside by the level curve of the potential function that solved the two-dimensional Poisson equation with the point source at the geometrical center (centroid) of the slice and zero Dirichlet boundary conditions at the outer boundary of the slice. The main property of this level curve is that it conforms to the outer boundary for the zeroth level of the potential and gradually changes its shape towards a circle around the point source located at the centroid for higher levels of the potential. For each slice, the level of the curve was chosen in such a way that the area of the cortex part would constitute 30% of the total slice area, see Fig. 10 (upper-right image).

#### *Metabolome*

From both sample types (flesh and peel) untargeted metabolic profiles were measured in an incomplete block design following guidelines from the Metabolomics Quality Assurance and Quality Control Consortium (mQACC)<sup>14</sup>. Pooled samples were created by mixing 60 random lots, 10 per variety, for flesh and peel separately. These samples were used in separate dilution series and as precision references every 10 sample injections in the experiment runs.

The freeze dried sample (100 mg) was weighed in a 2 ml Eppendorf tube, 1.3 ml of methanol was added, vortexed and shaken for 30 minutes. After centrifugation (14,000 rpm, 5 min.), 1 ml supernatant was transferred to a new tube and dried using a vacuum concentrator (1 mbar,  $35^{\circ}\text{C}$ , 150 min.). The residue was redissolved in cyclohexane (200  $\mu\text{l}$ ), milliQ water (300  $\mu\text{l}$ ) was added, shaken (10 min.) and centrifuged (14000 rpm, 10 min.). Using extended length tips, 180  $\mu\text{l}$  of the lower aqueous phase was transferred to a 0.2  $\mu\text{l}$  PVDF filter plate and centrifuged (1200 rpm, 4 min.).

The filtered samples were analyzed using a Waters Acquity I-class UPLC coupled with a Waters Xevo G2-XS QTOF MS. Chromatographic separation was achieved on a reverse-phase Acquity UPLC HSS T1.8  $\mu\text{m}$  (2.1 x 100 mm) column (Waters) at  $40^{\circ}\text{C}$ . The mobile phases employed were A: water (MilliQ) containing 0.1% formic acid (UPLC grade, BioSolve) and B: acetonitrile (UPLC grade, BioSolve). The flow rate was maintained at 0.5 mL/min, and the injection volume was 5  $\mu\text{L}$ . A gradient elution was employed, starting with 99% A for the first minute, followed by a linear gradient to 30% A over the next 10 minutes. A cleanup step was performed at 99% B for 1 minute, followed by 3 minutes of re-equilibration at the initial conditions. The total runtime was 15 minutes. The ESI source was operated in both positive and negative ionization modes with the following settings: capillary voltage, 3.00 kV; cone voltage, 40 V; source temperature,  $120^{\circ}\text{C}$ ; desolvation temperature,  $350^{\circ}\text{C}$ ; gas flow rate, 800 L/h (N<sub>2</sub>); cone gas flow rate, 50 L/h (N<sub>2</sub>). Leucine enkephalin was used as a Lock Spray reference.

Mass data were collected over the  $m/z$  range of 50–1200. MSe (Multiplexed Selected Ion Monitoring) was employed to acquire both parent ions and fragmentation data in a single run. The collision energy ramp was set from 10 to 40 V, allowing for the fragmentation of precursor ions across a range of energies.

Alignment of the chromatograms and peak picking was performed in Progenesis Q1. The precision reference results were used to perform batch and drift correction for each individual peak in the nPYc-Toolbox<sup>15</sup>. Feature selection was done based on the linearity in the dilution series and relative standard deviation in the precision references.

## Multiple linear regression using discontinuous-Galerkin inverse regression model

It is well known that the quality of predictions may depend on the choice of the mathematical model that seeks to connect the predictor variables with the dependent variable. In the present case, the number of predictors (biochemical information about the tubers) significantly exceeds the number of experiments (measurements of the vigor parameter)—this case is known as overparameterization or  $p \gg n$  in the Machine Learning context. Nevertheless it has been shown<sup>16,17</sup> that the use of complex models does not result in a performance improvement for datasets of limited size such as ours. In other words, there is no reason to expect a multilayered Artificial Neural Network (ANN), typically applied to overparameterized problems, to perform any better than a simple Multiple Linear Regression (MLR) model. Hence, while in our previous paper<sup>6</sup> a nonlinear Random Forest (RF) model was considered, in this paper we turn to an MLR model, common in the FTIR and HSI applications<sup>18–20</sup>. Specifically, a version of the fixed-effect MLR model called the Inverse Regression (IR) model<sup>10</sup> is applied. The IR model is especially suited for severely overparameterized problems with diverse predictor data and has been further adapted here to accommodate for the possible discontinuities in the dependence of the tuber properties on the plant vigor. The IR method also allows to compare the quality of the predictor data and to remove the features (predictors) that are either too noisy or do not comply with the linearity assumptions of the MLR model.

The explanatory and predictive model of the potato plant vigor in terms of seed tuber properties combines diverse datasets with some of them, like the metabolome data, potentially causing the so called zero inflation. This happens when a certain metabolite is detected only in a few samples and is absent in the remaining samples. Other datasets, like FTIR and HIS, where each feature is the absorbance at a given frequency, are generally continuous not only over the frequency range but also across the samples. The Inverse Regression (IR) model<sup>10</sup> accommodates for this potential diversity in the predictor data via the freedom to choose a suitable functional basis to represent the predictor variables  $x_j$ ,  $j = 1, \dots, p$  as functions of the dependent variable  $y$ . Due to the aforementioned possibility of discontinuities in the predictor data across the samples, we apply the Discontinuous Galerkin Finite-Element Method (DG-FEM) with the first-order (linear) Lagrange basis functions<sup>21</sup>. Since this is the first time the FEM is applied in the multiple linear regression and possibly also in the general Machine-Learning context, this section elaborates the technical details of the approach.

Let  $X \in \mathbb{R}^{n \times p}$  be the predictor data matrix and  $\mathbf{y} \in \mathbb{R}^n$  – the vector of vigor data sorted by the magnitude of its entries. The idea of the IR method is to represent the predictor data as  $x_{i,j} = x_j(y_i)$ , i.e., as the (sampled) functions of the dependent variable  $y$ . These functions are assumed to belong to a finite-dimensional function space defined by the choice of a suitable basis, i.e., a finite-dimensional set of basis functions. With the FTIR data<sup>10</sup> we have applied the polynomial basis and a very simple projection procedure leading to the decomposition  $X = VA$  of the data matrix in terms of the basis matrix  $V$  and the coefficient matrix  $A$ . A more rigorous and general projection procedure, suitable for a wider class of functions, is the Galerkin projection scheme widely used in the numerical solution of partial differential equations<sup>22</sup>.

The derivation of the DG-FEM formulation of the IR model starts from the discretization of the  $y$ -range, i.e., the interval  $I = [\min(\mathbf{y}), \max(\mathbf{y})]$ , into  $m$  finite elements  $I_k$ . The number of elements  $m$  plays the role of a regularization parameter that will eventually be chosen through a cross validation (CV) procedure, similarly to the maximal polynomial degree in the polynomial-basis IR method<sup>10</sup> or the maximal dimension of the Krylov subspace in the PLS method<sup>18</sup>. This  $y$ -range discretization yields  $m + 1$  delimiters  $\min(\mathbf{y}) = y_0, y_1, \dots, y_{m-1}, y_m = \max(\mathbf{y})$ . Here, for simplicity, we consider the uniform mesh with the step size  $h = y_{k+1} - y_k$ .

Next, on each element  $I_k = [y_k, y_{k+1}]$ , we introduce two local linear Lagrange basis functions  $v_{k,1}(y)$  and  $v_{k,2}(y)$ , defined as:

$$v_{k,1}(y) = \begin{cases} \frac{y_{k+1}-y}{h}, & \text{if } y \in I_k \\ 0, & \text{otherwise} \end{cases}, \quad v_{k,2}(y) = \begin{cases} \frac{y-y_k}{h}, & \text{if } y \in I_k \\ 0, & \text{otherwise} \end{cases}. \quad (1)$$

The main assumption of the IR method is that the predictor data are ‘generated’ by the functions  $x_j(y)$ ,  $j = 1, \dots, p$ ; all of which can be represented in terms of the chosen basis as:

$$x_j(y) = \sum_{k=1}^m \sum_{q=1}^2 \alpha_{k,q}^{(j)} v_{k,q}(y), \quad (2)$$

and the observed predictor data  $x_{i,j}$  are simply the sampled values  $x_j(y_i)$ ,  $j = 1, \dots, p$ ;  $i = 1, \dots, n$ .

The goal of the model training is to determine (learn) the expansion coefficients  $\alpha_{k,q}$ ,  $k = 1, \dots, m$ ,  $q = 1, 2$  in the representation (2). For this purpose, we apply the Galerkin scheme, where one chooses the *test* functions to be equal to the basis functions. Multiplying both sides of the Eq. (2) with a test function and integrating over the  $y$ -range, we arrive at the  $2m$  equations for the  $2m$  unknown coefficients  $\alpha_{k,q}$ :

$$\int_{y_0}^{y_m} x_j(y) v_{\hat{k},\hat{q}}(y) dy = \sum_{k=1}^m \sum_{q=1}^2 \alpha_{k,q}^{(j)} \int_{y_0}^{y_m} v_{k,q}(y) v_{\hat{k},\hat{q}}(y) dy, \quad \hat{k} = 1, \dots, m; \quad \hat{q} = 1, 2. \quad (3)$$

To further specify the system matrix and the right-hand side vector of this linear algebraic system of equations, we first note that, by the nature of the DG finite-element basis, the following partial orthogonality condition holds:

$$\int_{y_0}^{y_m} v_{k,q}(y) v_{\hat{k},\hat{q}}(y) dy = \delta_{k,\hat{k}} \int_{y_0}^{y_m} v_{k,q}(y) v_{\hat{k},\hat{q}}(y) dy. \quad (4)$$

Therefore, the equations for the coefficients  $\alpha_{k,q}$  corresponding to the different elements  $I_k$  decouple, and it is possible to find all coefficients by solving  $m$  separate  $2 \times 2$  linear systems:

$$\int_{y_{\hat{k}-1}}^{y_{\hat{k}}} x_j(y) v_{\hat{k},\hat{q}}(y) dy = \sum_{q=1}^2 \alpha_{\hat{k},q}^{(j)} \int_{y_{\hat{k}-1}}^{y_{\hat{k}}} v_{\hat{k},q}(y) v_{\hat{k},\hat{q}}(y) dy, \quad \hat{k} = 1, \dots, m; \quad \hat{q} = 1, 2. \quad (5)$$

The integrals in (5) are further approximated by a numerical quadrature rule<sup>23</sup>. Since the values of  $x_j(y)$  are only given at the available data points  $y_1, \dots, y_n$ , the trapezoidal rule seems to be appropriate:

$$\begin{aligned} \int_{y_{\hat{k}-1}}^{y_{\hat{k}}} v_{\hat{k},q}(y) v_{\hat{k},\hat{q}}(y) dy &= \sum_{m=1}^{n_{\hat{k}}} w_{\hat{k},m} v_{\hat{k},q}(y_{\hat{k},m}) v_{\hat{k},\hat{q}}(y_{\hat{k},m}) + \mathcal{O}(\max_m |y_{\hat{k},m+1} - y_{\hat{k},m}|^2), \\ \int_{y_{\hat{k}-1}}^{y_{\hat{k}}} x_j(y) v_{\hat{k},\hat{q}}(y) dy &= \sum_{m=1}^{n_{\hat{k}}} w_{\hat{k},m} x_j(y_{\hat{k},m}) v_{\hat{k},\hat{q}}(y_{\hat{k},m}) + \mathcal{O}(\max_m |y_{\hat{k},m+1} - y_{\hat{k},m}|^2), \end{aligned} \quad (6)$$

where the data points  $y_{\hat{k},m}$ ,  $m = 1, \dots, n_{\hat{k}}$  are situated within the element  $I_{\hat{k}}$  bounded by the delimiters  $y_{\hat{k}}$  and  $y_{\hat{k}+1}$ . The weights  $w_{\hat{k},m}$  of the trapezoidal quadrature can be explicitly written out as:

$$\begin{aligned} w_{\hat{k},1} &= |y_{\hat{k},1} - y_{\hat{k}}| + \frac{1}{2} |y_{\hat{k},2} - y_{\hat{k},1}|, \\ w_{\hat{k},m} &= \frac{1}{2} |y_{\hat{k},m} - y_{\hat{k},m-1}| + \frac{1}{2} |y_{\hat{k},m+1} - y_{\hat{k},m}|, \quad m = 2, \dots, n_{\hat{k}-1}, \\ w_{\hat{k},n_{\hat{k}}} &= |y_{\hat{k},n_{\hat{k}}} - y_{\hat{k}+1}| + \frac{1}{2} |y_{\hat{k},n_{\hat{k}}} - y_{\hat{k},n_{\hat{k}}-1}|. \end{aligned} \quad (7)$$

Therefore we can write a linear approximation of Eq. (5) for a given interval  $I_{\hat{k}}$  with the following notation:

$$V_{\hat{k}} = \begin{bmatrix} v_{\hat{k},1}(y_{\hat{k},1}) & v_{\hat{k},2}(y_{\hat{k},1}) \\ v_{\hat{k},1}(y_{\hat{k},2}) & v_{\hat{k},2}(y_{\hat{k},2}) \\ \vdots & \vdots \\ v_{\hat{k},1}(y_{\hat{k},n_{\hat{k}}}) & v_{\hat{k},2}(y_{\hat{k},n_{\hat{k}}}) \end{bmatrix}, \quad W_{\hat{k}} = \text{diag} \left( \begin{bmatrix} w_{\hat{k},1} \\ \vdots \\ w_{\hat{k},n_{\hat{k}}} \end{bmatrix} \right), \quad \mathbf{x}_j = \begin{bmatrix} x_j(y_{\hat{k},1}) \\ \vdots \\ x_j(y_{\hat{k},n_{\hat{k}}}) \end{bmatrix} \quad (8)$$

as

$$V_{\hat{k}}^T W_{\hat{k}} \mathbf{x}_j = V_{\hat{k}}^T W_{\hat{k}} V_{\hat{k}} \begin{bmatrix} \alpha_{\hat{k},1}^{(j)} \\ \alpha_{\hat{k},2}^{(j)} \end{bmatrix} = V_{\hat{k}}^T W_{\hat{k}} V_{\hat{k}} \boldsymbol{\alpha}_j \quad (9)$$

where the column vector  $\boldsymbol{\alpha}_j$  is a subvector of  $\mathbf{a}_j$ , the  $j$ -th column of  $A$ , relative to the  $\hat{k}$ -th interval.

Since the linear systems concerning each interval are decoupled, the full linear system is:

$$V^T W X = V^T W V A, \quad V^T W \mathbf{y} = V^T W V \mathbf{a}_0 \quad (10)$$

where

$$V = \begin{bmatrix} V_1 & & & \\ & \ddots & & \\ & & V_k & \\ & & & \ddots \\ & & & & V_m \end{bmatrix}, \quad W = \begin{bmatrix} W_1 & & & \\ & \ddots & & \\ & & W_k & \\ & & & \ddots \\ & & & & W_m \end{bmatrix}, \quad \mathbf{a}_j = \begin{bmatrix} \alpha_{1,1}^{(j)} \\ \alpha_{1,2}^{(j)} \\ \vdots \\ \alpha_{m,1}^{(j)} \\ \alpha_{m,2}^{(j)} \end{bmatrix}. \quad (11)$$

Therefore the sought after matrix  $A$  and vector  $\mathbf{a}_0$  solving Eq. (10) are:

$$A = (V^T W V)^{-1} V^T W X \quad (12)$$

and  $\hat{\mathbf{y}}_t$ , the prediction of the vector  $\mathbf{y}_t$ :

$$\hat{\mathbf{y}}_t = \hat{V} \mathbf{a}_0 \quad (13)$$

$$= X_t A^T (A A^T)^{-1} (V^T W V)^{-1} V^T W \mathbf{y}_t \quad (14)$$



where  $\hat{V} = X_t A^T (A A^T)^{-T}$ .

By denoting the full dataset as  $S = [y|X]$ , and the test set as  $S_t$ , the predictions for both the dependent and independent variables can be written as:

$$\hat{S} = S_t A^T (A A^T)^{-T} A. \quad (15)$$

### Measures of prediction quality

The standard measures of the MLR model quality, such as the  $R^2$  coefficient, are not applicable to the severely overparameterized case, where the number of predictors is significantly larger than the number of experiments<sup>24</sup>. In this paper, both the measured,  $y$ , and the predicted,  $\hat{y}$ , vigor data vectors, each of length  $n$ , are normalized as:

$$\tilde{y} = \frac{1}{\|y - n^{-1} \mathbf{1} \mathbf{1}^T y\|_2} (y - n^{-1} \mathbf{1} \mathbf{1}^T y), \quad \tilde{\hat{y}} = \frac{1}{\|\hat{y} - n^{-1} \mathbf{1} \mathbf{1}^T \hat{y}\|_2} (\hat{y} - n^{-1} \mathbf{1} \mathbf{1}^T \hat{y}). \quad (16)$$

Here,  $\mathbf{1} \in \mathbb{R}^n$  is the vector of all ones. After this transformation, we have  $\|\tilde{y}\|_2 = \|\tilde{\hat{y}}\|_2 = 1$ , where  $\|\cdot\|_2$  denotes the Euclidean norm of the vector, and  $n^{-1} \mathbf{1}^T \tilde{y} = n^{-1} \mathbf{1}^T \tilde{\hat{y}} = 0$ , i.e., both vectors have the unit norm and the zero mean.

For the normalized vectors we compute the squared norm of the residual vector, also known as Sum of Squared Errors (SSE):

$$r^2 = \|\tilde{y} - \tilde{\hat{y}}\|_2^2, \quad (17)$$

which is a measure of the unexplained variance. In this case, the  $r^2$  is related to the Pearson correlation coefficient  $c(\tilde{y}, \tilde{\hat{y}})$  between the measured and predicted data vectors as follows:

$$r^2 = 2(1 - c). \quad (18)$$

Since  $-1 \leq c \leq 1$ , we have the bounds  $0 \leq r^2 \leq 4$ . We consider the model to be predictive if for the testing dataset it gives the values  $r^2 < 1$  and  $c > 0.5$ , with the latter parameter also subject to the  $p$ -value analysis. The upper bound  $r^2 = 1$  for the SSE and the corresponding lower bound  $c = 0.5$  for the correlation, stem from the fact that for  $1 \leq r^2 \leq 4$  the prediction  $\hat{y}$  is usually of very poor quality, making it impossible to correctly classify the vigor of the seedlot as belonging to the low, middle or high quantile.

### Training, validation and testing

Both the variety-agnostic and the six variety-specific DG-IR models contain two hyper-parameters that need to be tuned. The first is the number  $m$  of segments, see the Eq. (2), that divide the range of the vigor parameter. The more segments are used, the better is the fit for each predictor variable. The 10-fold Cross Validation is used to find the optimal number of segments and avoid overfitting. In the DG-IR method with linear basis functions, at least two data points should be present within each segment. This puts a practical upper bound on the number of uniform segments, so that one only has to check for a smaller number of segments in search of the optimal parameter  $m$ . The results presented in Fig. 7 illustrate the fits with the optimal number of segments determined by the CV procedure.

The second hyper-parameter is the threshold  $\tau$  which allows discarding the predictors with the  $X$ -data residual above  $\tau$ , see the dashed horizontal lines in Fig. 6. Tuning of this parameter is performed with the number of segments  $m$  fixed at the optimal value.

To tune the above-mentioned hyper-parameters of the DG-IR model, the dataset is split into the training-validation and testing subsets. When the testing is performed on a different year, then the complete dataset of that year is used as the testing set. This means that the complete dataset of the training year can be used for training and validation (parameter tuning). When, however, the testing is performed on the same year as the training, then 33% of the variety data is reserved for testing and the remaining 67% is used for training and validation.

### Data availability

Both the seed tuber and plant canopy datasets are available at <http://doi.org/10.4121/3a97fa0c-8c7d-451a-b8fe-d521f1cec55e>. The data also includes the Python code necessary to reproduce the results of regression.

Received: 23 October 2024; Accepted: 21 July 2025

Published online: 05 August 2025

### References

- Food and Agriculture Organization of the United Nations, *FAOSTAT Database*. <https://www.fao.org/faostat/en/#data/QCL> (2022).
- Struik, P. C. & Wiersema, S. G. *Seed Potato Technology* (BRILL, 2023).
- National institute of agricultural botany, <https://www.niab.com/services/niab-labtest-analytical-services/potato-disease-testing/germination>.
- Atza, E. & Budko, N. Data underlying the publication: Predicting potato plant vigor from the seed tuber properties. *4TU ResearchData* <https://doi.org/10.4121/3a97fa0c-8c7d-451a-b8fe-d521f1cec55e> (2024).
- Oliveira, J., Brown, H., Gash, A. & Moot, D. An explanation of yield differences in three potato cultivars. *Agron. J.* **108**, 1434–1446 (2016).



6. Song, Y. et al. Seed tuber microbiome can predict growth potential of potato varieties. *Nat. Microbiol.* **10**, 28–40. <https://doi.org/10.1038/s41564-024-01872-x> (2025).
7. Zarzyńska, K., Boguszewska-Mańkowska, D., Feledyn-Szewczyk, B. & Jończyk, K. The vigor of seed potatoes from organic and conventional systems. *Agriculture* **12**, 1764 (2022).
8. Fiers, M. A. A. Potato soil-borne diseases. A review. *Agron. Sustain. Dev.* **32**, 93–132 (2012).
9. Davies, H. Mother tuber reserves as factors limiting potato sprout growth. *Potato Res.* **27**, 209–218 (1984).
10. Atza, E. & Budko, N. Overparameterized multiple linear regression as hyper-curve fitting. *Under review in SIAM J. Math. Data Sci.*, preprint (2024). [arxiv:2404.07849](https://arxiv.org/abs/2404.07849).
11. ten Den, T. et al. Modelling potential potato yields: Accounting for experimental differences in modern cultivars. *Eur. J. Agron.* **137**, 126510 (2022).
12. Atza, E. & Budko, N. Data underlying the publication: Seed tuber microbiome is a predictor of next-season potato vigor. *4TU ResearchData* <https://doi.org/10.4121/21892a06-078a-4600-8386-1abe46f42271> (2024).
13. Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A. & Eilers, P. H. C. Correcting for spatial heterogeneity in plant breeding experiments with p-splines. *Spat. Stat.* **23**, 52–71. <https://doi.org/10.1016/j.spasta.2017.10.003> (2018).
14. Broadhurst, D. et al. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **14**, 72. <https://doi.org/10.1007/s11306-018-1367-3> (2018).
15. Sands, C. J. et al. The nPYc-Toolbox, a Python module for the pre-processing, quality-control and analysis of metabolic profiling datasets. *Bioinformatics* **35**, 5359–5360. <https://doi.org/10.1093/bioinformatics/btz566> (2019).
16. Dong, Y. et al. Deep learning methods may not outperform other machine learning methods on analyzing genomic studies. *Front. Genet.* **13**, 992070. <https://doi.org/10.3389/fgene.2022.992070> (2022).
17. He, T. et al. Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. *Neuroimage* **206**, 116276. <https://doi.org/10.1016/j.neuroimage.2019.116276> (2020).
18. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst. Syst.* **58**, 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1) (2001).
19. Linker, R., Shmulevich, I., Kenny, A. & Shaviv, A. Soil identification and chemometrics for direct determination of nitrate in soils using FTIR-ATR mid-infrared spectroscopy. *Chemosphere* **61**, 652–658. <https://doi.org/10.1016/j.chemosphere.2005.03.034> (2005).
20. Rozali, N. L. et al. Fourier transform infrared (FTIR) spectroscopy approach combined with discriminant analysis and prediction model for crude palm oil authentication of different geographical and temporal origins. *Food Control* **146**, 109509. <https://doi.org/10.1016/j.foodcont.2022.109509> (2023).
21. Di Pietro, D. & Ern, A. *Mathematical Aspects of Discontinuous Galerkin Methods*. Mathématiques et Applications, vol. 69 (Springer-Verlag, 2011).
22. Brenner, S. & Scott, R. *The Mathematical Theory of Finite Element Methods* (Springer, 2005).
23. Vuik, K., Vermolen, E., van Gijzen, M. & Vuik, T. *Numerical Methods for Ordinary Differential Equations* (VSSD, TU Delft Open, 2023).
24. Izenman, A. J. *Modern Multivariate Statistical Techniques* (Springer, 2008).

## Acknowledgements

We are grateful to our industrial partners HZPC and Averis Seeds B.V. for collecting the seed tuber data and supporting the field trials. In particular, we would like to thank Johan Hopman from Averis Seeds B.V. (currently at McCain Foods) for the many fruitful discussions and valuable insights.

## Author contributions

EA and NVB performed quantitative analysis of drone images, modeling and data analysis. EA, NVB, and RK wrote the manuscript. FvdW, HvD, FH and RK coordinated sampling collection and experimental field trials. HZPC Research was responsible for all instrumental measurements. FH and RK were responsible for data collection. All authors reviewed the manuscript.

## Funding

This work was funded by HZPC Research B.V., Averis Seeds B.V., BO-Akkerbouw, and European Agricultural Fund for Rural Development.

## Declarations

## Competing interests

The authors declare that this study received funding from HZPC Research B.V. and Averis Seeds B.V. The funder had the following involvement in the study: study design, sample collection, and the decision to submit it for publication. RK, FH, FvdW, HvD are currently employed by HZPC research B.V.

## Additional information

**Correspondence** and requests for materials should be addressed to E.A. or N.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025