A two-stage approach for the spatio-temporal analysis of high-throughput phenotyping data

Diana M. Pérez-Valencia[1,2*]**, María Xosé Rodríguez-Álvarez**[1,3]**, Martin P. Boer**[4]**, Lukas Kronenberg**[5]**, Andreas Hund**[6]**, Llorenç Cabrera-Bosquet**[7]**, Emilie J. Millet**[4]**, and Fred A. van Eeuwijk**[4]

[1]BCAM - Basque Center for Applied Mathematics, Mazarredo 14, E48009 Bilbao, Basque Country - Spain
[2]Departamento de Matemáticas, Universidad del País Vasco UPV/EHU, Leioa, Spain
[3]IKERBASQUE, Basque Foundation for Science, Bilbao, 48009, Spain
[4]Biometris, Wageningen University & Research, Wageningen, 6708 PB, The Netherlands
[5]Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zürich, 8092 Zurich, Switzerland
[6]Crop Science, Institute of Agricultural Sciences, ETH Zürich, 8092 Zurich, Switzerland
[7]LEPSE, Univ Montpellier, INRAE, Institut Agro, Montpellier, France
*dperez@bcamath.org

ABSTRACT

High throuput phenotyping (HTP) platforms and devices are increasingly used for the characterization of growth and developmental processes for large sets of plant genotypes. Such HTP data require challenging statistical analyses in which longitudinal genetic signals need to be estimated against a background of spatio-temporal noise processes. We propose a two-stage approach for the analysis of such longitudinal HTP data. In a first stage, we correct for design features and spatial trends per time point. In a second stage, we focus on the longitudinal modelling of the spatially corrected data, thereby taking advantage of shared longitudinal features between genotypes and plants within genotypes. We propose a flexible hierarchical three-level P-spline growth curve model, with plants/plots nested in genotypes, and genotypes nested in populations. For selection of genotypes in a plant breeding context, we show how to extract new phenotypes, like growth rates, from the estimated genotypic growth curves and their first-order derivatives. We illustrate our approach on HTP data from the PhenoArch greenhouse platform at INRAE Montpellier and the outdoor Field Phenotyping platform at ETH Zürich.

Introduction

In recent years, the frequency of use of high-throughput phenotyping (HTP) techniques and platforms has strongly increased in plant genetics and physiology. HTP data provide quick, precise, non-destructive and cost-effective information on phenotypic traits with high spatial and temporal resolution[1]. Designed HTP experiments, either indoors or in a field, usually consist of experimental units (e.g., single plants in pots or plots) that are combined with a wide range of sensing equipment for the (almost) continuous monitoring of plant/plot phenotypic traits for large sets of genotypes. Treatments applied in HTP research designs may comprise not only different genotypes but also different management practices. High dimensional HTP data as derived from multiple sensors (e.g., images, point clouds, hyperspectral data) are typically filtered, condensed, integrated and summarised into features. Combinations of one or more features are used to approximate simple biological traits that are on the one hand still close to the data, and on the other hand are relatively far from the target traits of commercial interest (most often yield and quality parameters). Examples of such low-level traits are plant height, canopy cover, leaf area index, ear and tiller counts, canopy temperature or indices related to water or chlorophyll content. Thus, researchers and plant breeders have now access to large and detailed datasets, in the form of (long) time-series, enabling to track multiple low-level traits from, e.g., seed emergence to physiological maturity. The challenge in this setting is how to efficiently and adequately exploit the diversity and complexity of HTP data to (a) extract relevant information on several aspects of plant growth and development, (b) increase the biological understanding of plant systems; and (c) help and support the decision-making process in plant breeding programs[2]. On top of that, the vast amount of data typically produced by HTP platforms require not only appropriate but also rapid processing techniques.

Regardless of the fact that complex traits (like yield) are measured only once (as traditionally) or low-level phenotypic traits are measured several times (as in HTP), it is known that their expression is spatially affected by environmental factors such as soil heterogeneity in field experiments and temperature and light gradients in the greenhouse. This makes it necessary to correct for these nuisance factors when analysing agricultural experiments[2,3]. For "traditional" agricultural experiments, modelling

spatial variation through spatially correlated stochastic components or smoothing techniques has become ubiquitous[4–11]. Yet, data from HTP experiments present, on top of spatial variation, a time dimension which needs to be incorporated and adequately modelled in the analyses. Although interest may primarily focus on estimating (and further processing and analysing) the evolution over time of the genetic effect on a given phenotype, it will be essential to correct for environmental nuisance effects not only in a spatial but now also in a temporal dimension. Thus, from the statistical point of view, HTP data ask for spatio-temporal modelling approaches. To the best of our knowledge, there are few proposals in the HTP literature that tackle the problem from both perspectives, temporal and spatial. References for stage-wise analysis of HTP data include the work by van Eeuwijk *et al.* (2019)[2], where the authors propose to first estimate time-series of spatially adjusted genotypic means from low-level phenotypic traits, that are then subject to temporal dynamic modelling. A similar approach is followed by Kar *et al.* (2020)[12]. In the above-cited references, the information that is carried to the temporal analysis is not at the level of the experimental unit but at the genotype level. A different stage-wise approach is described in Roth *et al.* (2021), where the temporal analysis is performed first, followed by the spatial correction. Approaches that model simultaneously the spatial and temporal genetic and non-genetic variation in one-stage include the works by Verbyla *et al.* (2021)[13] and Pérez *et al.* (2020)[14]. While promising, both proposals report high computational complexity and scalability problems.

In this work we propose a two-stage approach that allows keeping the data resolution for the second stage at the experimental unit (plant/pot/plot) and is computationally efficient. The first stage focuses on correcting the low-level phenotypic data for experimental design factors and spatial variation, while the second stage aims at estimating the evolution over time of the genetic signal. In a similar fashion to van Eeuwijk *et al.* (2019)[2], in the first stage we fit a spatial model to the phenotypic data separately for each measurement time. The phenotypic data are subsequently corrected by only considering the (estimated) sources of variation which are of interest plus the residual component (i.e., the measurement error). The purpose of this stage and subsequent correction is two-fold: (1) to remove nuisance spatial variation from the phenotypic data; and (2) to keep the data resolution for the second stage at the level of the experimental unit (through the incorporation in the correction of the residual component). This is one of the main differences to the proposals described in van Eeuwijk *et al.* (2019)[2] and Kar *et al.* (2020)[12], and it is routinely applied for data derived from the field phenotyping platform of ETH Zurich[15–17]. Since analyses are performed separately for each measurement time, our modelling strategy implicitly permits the spatial variation to differ among measurement times, i.e., it allows correcting for both the spatial and temporal evolution of environmental variables and experimental design factors. For this first stage of spatial correction, different proposals are available in the literature[4–11]. The separable autoregressive model[5,6] has become the standard modelling strategy, but the recently proposed SpATS (Spatial Analysis of Field Trials with Splines) approach[11] has proved to be a powerful alternative. Indeed, comparisons of these two models have shown that SpATS is a very competitive approach, with the advantage that model selection is greatly simplified[18,19]. In addition, an R-package (also called `SpATS`) is freely available on CRAN (https://CRAN.R-project.org/package=SpATS), and the recently `statgenHTP`[20] R-package (available on https://CRAN.R-project.org/package=statgenHTP) allows for an easy fitting of SpATS models for different (and possibly large) measurement times. For all these reasons, we choose SpATS for this first stage.

The second stage of our proposal focuses on modelling the genetic signal as a function of time for the corrected phenotype obtained in the first stage. Data for this stage consist of time-series of spatially corrected phenotypic trait measurements per experimental unit. In other words, we have a sample of growth curves per plant/pot/plot that can be grouped by, at least, one treatment factor, the genotype[21]. Thus, generally speaking, this stage requires hierarchical or multilevel growth curve modelling approaches. A traditional data-driven approach to the analysis of growth curves is to use parametric (non-linear) models, being the logistic function one of the most commonly used[22]. However, while growth processes theoretically follow a clearly defined pattern that may be modelled using a parametric function, the observed dynamics may deviate considerably due to, for example, temporal changes in environmental conditions (e.g. cold spells) or the application of treatments (e.g. irrigation events). More flexible models that overcome the limitations of parametric specifications have been proposed in the literature. Examples include the use of smoothing or penalised splines (P-splines[23]) as well as functional principal components analysis (FPCA)[21,24,25]. For the particular case of data from HTP platforms, their temporal analysis by means of splines or FPCA has been discussed in Xu *et al.*[26–28], Wang *et al.*[29] and Brien *et al.*[30]. In this paper, we follow the same modelling philosophy as for the first stage and propose the use of P-spline hierarchical curve data models along the lines of the work by Durban *et al.* (2005)[24] and Greven and Scheipl (2017)[31]. Our approach presents several advantages. The connection between P-splines and linear mixed model offers a rich framework for estimation and inference[32,33]. Moreover, we can easily obtain the derivatives of the estimated curves at different levels of the hierarchy (e.g. genotype or experimental unit) which can provide important insights on genotype performance. Finally, from the estimated curves and their derivatives, different time-dependent and -independent characteristics (intermediate traits) can be easily extracted. Intermediate traits capture the dynamics of low-level traits and are closer to the target traits of plant breeders. Examples include the area under estimated growth curves, the timing of key stages (e.g., the timing at which final height is reached), or quantities at defined time points or periods (e.g., final height, final canopy cover or maximum speed rates)[34]. They can serve as inputs to subsequent statistical analyses that aim at modelling

genotype-by-environment interactions in biologically complex traits, like yield, in terms of underlying component traits.

Methods

We consider that, in the HTP experiment, genotypes have been allocated to the experimental units following an experimental design. For simplicity, in what follows we assume that only genotypes have been tested (i.e., we have a single factor treatment structure). However, we allow for population structure, modelled as different families, panels or populations of genotypes. Thus, the data present a three-level nested hierarchical structure, with plants/plots nested in genotypes, and genotypes nested in populations (see Supplementary Fig. S1 online). For clarity and simplicity, hereafter we refer to the experimental units of the experimental design as plants. The experimental unit is most commonly a plant for an indoor experiment and a plot containing several plants for a field experiment.

Data Structure

Let $y_{pgi}(t)$ denote the (observed) low-level phenotypic trait of interest for the i-th plant ($i = 1, \ldots, m_{pg}$) of the g-th genotype ($g = 1, ..., \ell_p$) in the p-th population ($p = 1, \ldots, k$) at time $t \in \{t_1, \ldots, t_n\}$. Let $L = \sum_{p=1}^{k} \ell_p$ denote the total number of genotypes and $M = \sum_{p=1}^{k} \sum_{g=1}^{\ell_p} m_{pg}$ the total number of plants. We consider that plants can be mapped to a coordinate system defined in terms of r rows and c columns, and denote as u_{pgi} and v_{pgi} the row and column position respectively ($u_{pgi} \in \{1, \ldots, r\}$ and $v_{pgi} \in \{1, \ldots, c\}$). In the first stage, the correction is made per time point for the whole experiment. As such, note that we assume that all plants in the experiment are measured at the same times, ($t_1, \ldots, t_n$). That is a simplification, as platform data is typically acquired within the order of minutes to hours. However, we presume that the factors that may affect the platform measurements within that period can be accounted for (and captured) by the experimental design (e.g., blocking structure). The assumption of the same measuring times, however, does not preclude the presence of incomplete data. The methods can handle missing values at both the plant and genotypic level (i.e., with plants or even genotypes not measured for some times).

First stage: Environmental Correction using SpATS

In the first stage, a SpATS model is fitted to the phenotypic data separately for each measurement time $t \in \{t_1, \ldots, t_n\}$. A SpATS model is a linear mixed model where large-scale and small-scale spatial dependence (or spatial trend) is explicitly modelled by a two-dimensional smooth surface defined over the row and column positions of the plants in the experiments, $f_t(u, v)$. This smooth function is constructed with tensor-product P-splines[23,35], i.e., through the combination of the tensor-product of marginal B-spline bases and an anisotropic discrete penalty on the coefficients. In its more general specification, and considering genotypes as random, a SpATS model has the following form (for more details, see Rodríguez-Álvarez *et al.*[11])

$$y(t) = \mathbf{1}_M \beta_{0t} + X\beta_t + Zc_t + Z_g c_{gt} + \underbrace{f_t(u, v)}_{\text{spatial trend}} + \varepsilon_t, \quad c_t \sim N(0, G_t), \quad c_{gt} \sim N(0, G_{gt}), \quad \varepsilon_t \sim N\left(0, \sigma_t^2 I_M\right), \tag{1}$$

where $y(t) = (y_{111}(t), \ldots, y_{k\ell_k m_{k\ell_k}}(t))^T$ is the low-level phenotypic trait at time t, X and Z represent column-partitioned matrices, associated respectively with fixed and random components, as for instance, row, column, replicate and/or (incomplete) block effects, and Z_g is the design matrix assigning plants to genotypes. Further, we decompose the fixed effects component in two terms

$$X\beta_t = X_h \beta_{ht} + X_q \beta_{qt}.$$

The first term, $X_h \beta_{ht}$, corresponds to the factors/covariates whose effects we are interested in modelling in the second stage, whereas $X_q \beta_{qt}$ corresponds to the factors/covariates whose effects we are interested in removing (e.g., those associated with experimental design factors). As said before, for simplicity, we consider that $X_h \beta_{ht}$ only contains information regarding the family/population to which the genotypes belong to (if any). Concerning $X_q \beta_{qt}$, we assume that it is associated with R experimental design factors (categorical covariates), and thus, $X_q = [X_{q1} | \ldots | X_{qR}]$ and $\beta_{qt} = (\beta_{qt1}^T, \ldots, \beta_{qtR}^T)^T$. The length of β_{qtr} (and therefore the number of columns in the associated design matrix X_{qr}) ($r = 1, \ldots, R$) corresponds to the number of different categories, say c_r, of the r-th experimental design factor minus one (as the intercept is included in the model).

Once the SpATS model in equation (1) is fitted, the phenotype of interest at time t, $y(t)$, is corrected by only considering the (estimated) sources of variation that are of interest, plus the residual component, i.e., the corrected phenotype, denoted as $\tilde{y}(t) = (\tilde{y}_{111}(t), \ldots, \tilde{y}_{k\ell_k m_{k\ell_k}}(t))^T$, is obtained as follows

$$\tilde{y}(t) = \underbrace{\mathbf{1}_M \hat{\beta}_{0t} + \sum_{r=1}^{R} \frac{1}{c_r} J_{qr} \hat{\beta}_{qtr} + X_h \hat{\beta}_{ht} + Z_g \hat{c}_{gt}}_{\hat{p}(t)} + \hat{\varepsilon}_t, \tag{2}$$

where $\boldsymbol{J}_{qr}$ are matrices of ones of appropriate dimensions (i.e., $M \times (c_r - 1)$). The correction is performed following the procedure for obtaining predictions (e.g., adjusted means) in linear mixed models described in Welham *et al.* (2004)[36]. In that paper, the authors propose a partition of the explanatory variables (see equation (1)) in three groups: (1) those for which predictions are required (i.e., population, $\boldsymbol{X}_h \hat{\boldsymbol{\beta}}_{ht}$, and genotypic, $\boldsymbol{Z}_g \hat{\boldsymbol{c}}_{gt}$, effects), (2) those to be averaged over (i.e., experimental design factors effects, $\sum_{r=1}^{R} \frac{1}{c_r} \boldsymbol{J}_{qr} \hat{\boldsymbol{\beta}}_{qtr}$); and, (3) those to be ignored (i.e., spatial trends, $\hat{f}_t(\boldsymbol{u}, \boldsymbol{v})$, and other random effects, $\boldsymbol{Z}\hat{\boldsymbol{c}}_t$).

As will be seen in the next section, in the second stage of our proposal we model $\tilde{\boldsymbol{y}}(t)$ ($t \in \{t_1, \ldots, t_n\}$). Thus, it is worth emphasising that, in the way it is constructed, $\tilde{\boldsymbol{y}}(t)$ (see equation (2)) only contains information about genetic populations and genotypes, as well as unexplained plant-to-plant variation (measurement error). In other words, for the second stage the predicted values for the genetic populations and the genotypes, $\hat{\boldsymbol{p}}(t)$, as well as the unexplained plant-to-plant variation, $\hat{\boldsymbol{\varepsilon}}_t$, are maintained as the "new" (spatially corrected) experimental unit values, while the spatial trends and other blocking factors to control for spatial variability are omitted. Also, note that the "observations" that enter the second stage, $\tilde{\boldsymbol{y}}(t)$, are not observed but estimated/predicted. Thus, we propose to propagate the uncertainty from the first stage to the second stage through the inclusion of weights, in a similar way to the weighted stage-wise analysis of multi-environment trials[37]. In particular, weights are obtained from the inverse of the variance-covariance (vcov) matrix for the predictions (plus the residual variance), i.e., $\boldsymbol{w}(t) = \text{diag}((\text{vcov}(\hat{\boldsymbol{p}}(t)) + \hat{\sigma}_t^2 \boldsymbol{I}_M)^{-1})$.

Second stage: P-spline Hierarchical Curve Data Model

The aim of the second stage is to model the spatially corrected phenotype obtained in the first stage. We can re-organise the data for this stage in such a way that they can be seen as a sample of plant growth curves, $\tilde{\boldsymbol{y}}_{pgi} = (\tilde{y}_{pgi}(t_1), \ldots, \tilde{y}_{pgi}(t_n))$ ($1 \le p \le k,\ 1 \le g \le \ell_p,\ 1 \le i \le m_{pg}$), with a nested hierarchical structure, where plants are nested in genotypes and genotypes are nested in populations. We propose to model this sample of curves by considering an additive decomposition of the phenotypic variation over time and use a three-level nested hierarchical growth model for this purpose[21]

$$\tilde{y}_{pgi}(t) = f_p(t) + f_{pg}(t) + f_{pgi}(t) + \varepsilon_{pgi}(t),\ \varepsilon_{pgi}(t) \sim N\left(0, \sigma^2 w_{pgi}(t)\right),\ 1 \le p \le k,\ 1 \le g \le \ell_p,\ 1 \le i \le m_{pg}, \quad (3)$$

where f_p is the growth/change over time of the (spatially corrected) phenotype for the p-th population (i.e., the p-th population mean function), f_{pg} is the genotype-specific growth deviation from f_p for the g-th genotype, and f_{pgi} is the plant-specific growth deviation from f_{pg} for the i-th plant. The additive modelling approach implies that $f_p + f_{pg}$ can be interpreted as the growth over time of the (spatially corrected) phenotype for the g-the genotype in the p-th population. Thus, on top of genotype-specific deviations from their overall population mean, we also obtain genotype-specific growth curves. Finally, w_{pgi} are the weights obtained in the first stage.

We use P-splines for hierarchical curve data[24,31] to estimate the model in equation (3). In this framework, each function in equation (3) is approximated by a linear combination of cubic B-spline basis functions, and its smoothness is controlled by a penalty on the differences of the B-spline coefficients (we use differences of order 2). The influence of the penalty is determined by a smoothing parameter. One of the attractive features of using B-splines to represent the functions in equation (3) is that their derivatives can be easily obtained[38]. Furthermore, we adopt the connection between P-splines and linear mixed models[32,33], where each function is treated as a sum of fixed (linear) and random (non-linear) components, and the smoothing parameter is replaced by a ratio of variances components. For clarity we omit here the more technical details, but it can be shown that, under this setting, each function in equation (3) is expressed as follows

$$f_p(t) = \beta_{p,0} + \beta_{p,1}t + \sum_{j=1}^{b_{\text{pop}}-2} u_{p,j} z_{\text{pop},j}(t),\ u_{p,j} \overset{iid}{\sim} N\left(0, \sigma_p^2\right),$$

$$f_{pg}(t) = \beta_{pg,0} + \beta_{pg,1}t + \sum_{j=1}^{b_{\text{gen}}-2} u_{pg,j} z_{\text{gen},j}(t),\ \beta_{pg,0} \overset{iid}{\sim} N\left(0, \sigma_{\text{gen},0}^2\right),\ \beta_{pg,1} \overset{iid}{\sim} N\left(0, \sigma_{\text{gen},1}^2\right),\ u_{pg,j} \overset{iid}{\sim} N\left(0, \sigma_{\text{gen}}^2\right), \quad (4)$$

$$f_{pgi}(t) = \beta_{pgi,0} + \beta_{pgi,1}t + \sum_{j=1}^{b_{\text{plant}}-2} u_{pgi,j} z_{\text{plant},j}(t),\ \beta_{pgi,0} \overset{iid}{\sim} N\left(0, \sigma_{\text{plant},0}^2\right),\ \beta_{pgi,1} \overset{iid}{\sim} N\left(0, \sigma_{\text{plant},1}^2\right),\ u_{pgi,j} \overset{iid}{\sim} N\left(0, \sigma_{\text{plant}}^2\right),$$

$1 \le p \le k,\ 1 \le g \le \ell_p,\ 1 \le i \le m_{pg}$, where $\{z_{\text{pop},j}(\cdot) : 1 \le j \le b_{\text{pop}} - 2\}$, $\{z_{\text{gen},j}(\cdot) : 1 \le j \le b_{\text{gen}} - 2\}$ and $\{z_{\text{plant},j}(\cdot) : 1 \le j \le b_{\text{plant}} - 2\}$ are sets of basis functions obtained from the connection between P-splines and linear mixed models[32]. In contrast to standard P-spline mixed models, here the linear component (intercept and slope) associated with f_{pg} (genotypic deviations) and f_{pgi} (plant deviations) is modelled with random rather than fixed effects. Also, for notational simplicity, we assume the same genetic variation across populations. However, this assumption can be easily relaxed by considering different values of $\sigma_{\text{gen},0}^2$, $\sigma_{\text{gen},1}^2$ and σ_{gen}^2 per population. A similar approach can be followed to allow for the plant-to-plant variation ($\sigma_{\text{plant},0}^2$, $\sigma_{\text{plant},1}^2$ and

σ^2_{plant}) to vary across genotypes. These generalisations might be worth exploring if there are sufficient number of genotypes per population and plants per genotype, respectively.

Before proceeding, we comment on the selection of the number of B-spline basis functions used to approximate f_p, f_{pg} and f_{pgi} (i.e., b_{pop}, b_{gen} and b_{plant}, respectively). In P-splines, it is recommended to choose a large number of bases to provide enough flexibility; the role of the penalty is to avoid over fitting[23]. In our setting the number of functions in the complete model equals $k + L + M$ (populations + genotypes + plants), and, hence, the number of regression coefficients (either fixed or random) to be estimated is $k \times b_{\text{pop}} + L \times b_{\text{gen}} + M \times b_{\text{plant}}$. This value can be very large, with the number of plants, M, and associated basis dimension, b_{plant}, playing the major role: the dataset may contain thousands of plants. Thus, to reduce the computational burden, one could be tempted to use different basis dimensions for f_p, f_{pg} and f_{pgi}, and to be less generous with f_{pgi}. However, this is not a good strategy. Simulation studies, as well as preparatory data analyses, have shown that results may be sensitive (and in some cases unreliable) to using different bases dimensions. We therefore recommend choosing the same value for b_{pop}, b_{gen} and b_{plant}, while keeping the number of coefficients at a reasonable level (i.e., a trade-off between flexibility and dimensionality).

We now present the model in equation (3) based on the specification in equation (4) in matrix notation. Let's first order the data by population, genotype, plant, and time (in that order), i.e., $\tilde{\boldsymbol{y}} = (\tilde{y}_{111}(t_1), \ldots, \tilde{y}_{111}(t_n), \ldots, \tilde{y}_{k\ell_k m_{k\ell_k}}(t_1), \ldots, \tilde{y}_{k\ell_k m_{k\ell_k}}(t_n))^T$. Thus, in a compact way, the three-level nested hierarchical growth model can be expressed as

$$\tilde{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon}, \ \boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{G}), \ \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{W}), \tag{5}$$

where

$$\boldsymbol{X} = [\boldsymbol{Q}_{\text{pop}} \otimes \boldsymbol{X}_{\text{pop}}],$$
$$\boldsymbol{Z} = [\boldsymbol{Q}_{\text{pop}} \otimes \boldsymbol{Z}_{\text{pop}} \mid \boldsymbol{Q}_{\text{gen}} \otimes \boldsymbol{X}_{\text{gen}} \mid \boldsymbol{Q}_{\text{gen}} \otimes \boldsymbol{Z}_{\text{gen}} \mid \boldsymbol{I}_M \otimes \boldsymbol{X}_{\text{plant}} \mid \boldsymbol{I}_M \otimes \boldsymbol{Z}_{\text{plant}}] \text{ with } \boldsymbol{u} = (\boldsymbol{u}^T_{\text{pop}}, \boldsymbol{\beta}^T_{\text{gen}}, \boldsymbol{u}^T_{\text{gen}}, \boldsymbol{\beta}^T_{\text{plant}}, \boldsymbol{u}^T_{\text{plant}})^T,$$

and

$$\boldsymbol{G} = \begin{pmatrix} \text{blockdiag}(\sigma^2_1 \boldsymbol{I}_{b_{\text{pop}}-2}, \ldots, \sigma^2_k \boldsymbol{I}_{b_{\text{pop}}-2}) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_L \otimes \begin{pmatrix} \boldsymbol{\Sigma}_{\text{gen}} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma^2_{\text{gen}} \boldsymbol{I}_{b_{\text{gen}}-2} \end{pmatrix} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_M \otimes \begin{pmatrix} \boldsymbol{\Sigma}_{\text{plant}} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma^2_{\text{plant}} \boldsymbol{I}_{b_{\text{plant}}-2} \end{pmatrix} \end{pmatrix}.$$

The specific form of the matrices and vectors involved in equation (5) is given in Section "Three-level Nested Hierarchical Growth Model: Mixed Model Formulation". Note that model in equation (5) is a standard linear mixed model, and, thus, estimation can be carried out with any mixed-model software, such as the R-packages `ASReml-R`[39], `nlme`[40] and `lme4`[41], or the `PROC MIXED`[42] procedure in SAS®. Also, we can use the R-package `mgcv`[43] for that purpose. However, high-throughput phenotypic data are usually characterised by a large number of observations, which, together with the number of regression coefficients in equation (5), might make estimation with the above-mentioned software computationally expensive. Thus, we have implemented in the R language[44] our own code (provided along with the paper), which resorts to the recently proposed SOP (Separation of Overlapping Penalties) method[45]. Empirical best linear unbiased estimates (BLUE) and predictors (BLUP) are obtained by the solution of Henderson's mixed model equations[46], and variance components by means of restricted maximum likelihood (REML)[47]. Construction of (approximate) confidence intervals for the estimated curves and their derivatives is based on the prediction error variance[36]. To speed up computation, we take advantage of the array structure of the data through Generalised Linear Array Models (GLAM)[48] and of the sparse structure of the matrices involved in the model. It is to note that the model in equation (5) presents a standard variance-covariance matrix for the random effects (i.e., it is linear in the variance parameters) and thus, the SOP method reduces to the estimating algorithm described in Harville (1977)[49].

Applications

We illustrate the potential of our approach with data from two experiments from two different HTP platforms: 1) the PhenoArch platform (INRAE Montpellier)[50] (greenhouse, Figure 1(a)) and, 2) the FIP (FIeld Phenotyping) platform (ETH Zürich)[51] (field, Figure 1(b)). The PhenoArch platform data are used to describe our approach and illustrate the outcomes at each step. The FIP platform data are further analysed to discuss the extraction of important (genotype-specific) attributes from the estimated genotype-specific growth curves and their first-order derivative, as well as their possible use in the decision-making process[2,52,53].
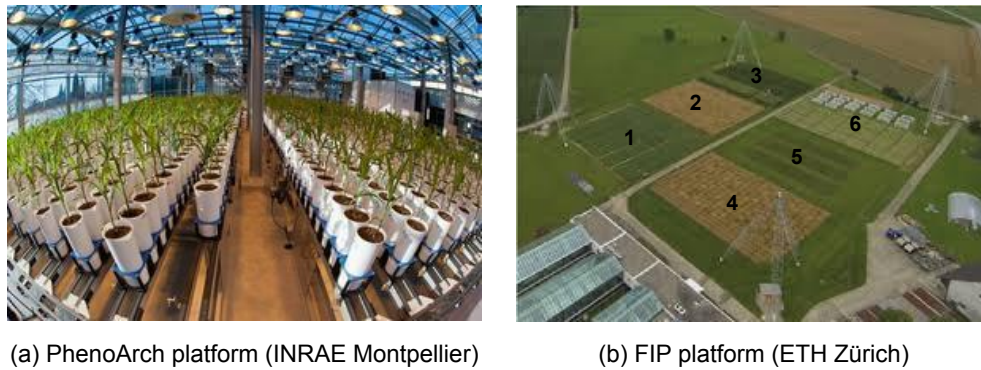
| (a) PhenoArch platform (INRAE Montpellier) | (b) FIP platform (ETH Zürich) |

Figure 1. Overview of the two high-throughput phenotyping platforms.

PhenoArch platform (INRAE Montpellier)

The PhenoArch platform hosted at M3P, Montpellier Plant Phenotyping Platforms (https://www6.montpellier.inra.fr/lepse/M3P) is composed of a conveyor belt structure of 28 lanes carrying 60 carts with one pot each (i.e. 1680 pots on a rectangular grid of $r = 60$ rows and $c = 28$ columns (see Figure 1(a)), plus a conveyor belt system that feeds the imaging or the watering units. Pots are daily moved to be imaged and/or watered. They are then moved back to the same positions and orientation, so that the plant position with respect to its neighbours is conserved throughout the experiment. The data analysed here correspond to an experiment including two different panels of commercial maize hybrids representative of breeding history in Europe and US during the last 60 years. This material covers a wide range of plant architecture, growth and development. A total of 60 genotypes were tested in Panel 1 and 30 genotypes were tested in Panel 2, and each genotype was replicated between 4 (Panel 2) and 14 (Panel 1) times. All genotypes were tested under two levels of soil water content: (1) mild water deficit (WD, soil water potential of -0.5 MPa) and (2) retention capacity (WW, soil water potential of -0.05 MPa). The experiment was carried out in 2017 between April 13th and May 15th, which correspond to 103 and 135 days since January 1st (hereafter referred as DOY, Day of the Year), respectively. Red-green-blue (2056×2454) images taken from 13 views (12 side views from 30° rotational difference and one top view) were captured daily for each plant. Plant pixels from each image were segmented from those of the background and used for estimating the whole plant leaf area (among other features) as described in Brichet *et al.* (2017)[54]. The dataset consists of $n = 32$ leaf area measurements on $M = 1656$ plants ($1656 \times 32 = 52992$ observations, including missing data). For a proper analysis of this dataset, panels and treatments should be modelled as crossed effects. However, for the sake of simplicity, for our analyses we combined panels and water regimes, such that we considered $k = 4$ "populations" (Panel 1 - WD, Panel 1 - WW, Panel 2 - WD and Panel 3 - WW) and $L = 180$ "genotypes" (60 genotypes in Panel 1 and WW treatment, 60 genotypes in Panel 1 and WD treatment, 30 genotypes in Panel 2 and WW treatment and 30 genotypes in Panel 1 and WD treatment).

First stage results: PhenoArch platform

A SpATS model was fitted for each individual measurement time point of the leaf area data. In addition to the spatial trend and the random genotypic effects (see equation (1)), the model included the population (panel by water regime combination) as fixed effect, and the row and column positions as random effects. A different genetic variance for each of the four populations (panel by water regime combination) was considered. Regarding the spatial trend (i.e., the tensor-product P-spline), B-spline bases of dimension 60 and 28 were chosen for the row and column positions, respectively, and nested bases, with half the dimension, were used[55]. The spatially corrected leaf area included the estimated population and genotypic effects, as well as the residual (see equation (2)). The comparison of the spatially corrected leaf area when modelling genotypes as fixed (as usually in stage-wise analyses) or random (as done here) effects shows essentially identical results (Supplementary Fig. S3 online): shrinkage of the genotypic BLUPs is counteracted by the inclusion of the residual component into the correction in equation (2). Results show that, as expected, the spatial pattern observed in the raw data disappears in the spatially corrected phenotype, as illustrated with the leaf area at four different measurement times (Figure 2). Moreover, the correction reduced the variability among plants (i.e., replicates) of the same genotype and water treatment combination (Figure 3).

Second stage results: PhenoArch platform

The three-level nested hierarchical growth model in equation (3) was fitted to the spatially corrected leaf area. We used cubic B-spline basis of dimension 13 to represent f_p, f_{pg} and f_{pgi} (i.e., $b_{pop} = b_{gen} = b_{plant} = 13$). Under this configuration, the mixed model in equation (5) has 52992 observations (there are missing values), 23920 regression coefficients (both fixed and random) and 11 variance components. Estimation took approximately 9 minutes. Computations were performed in (64-bit) R
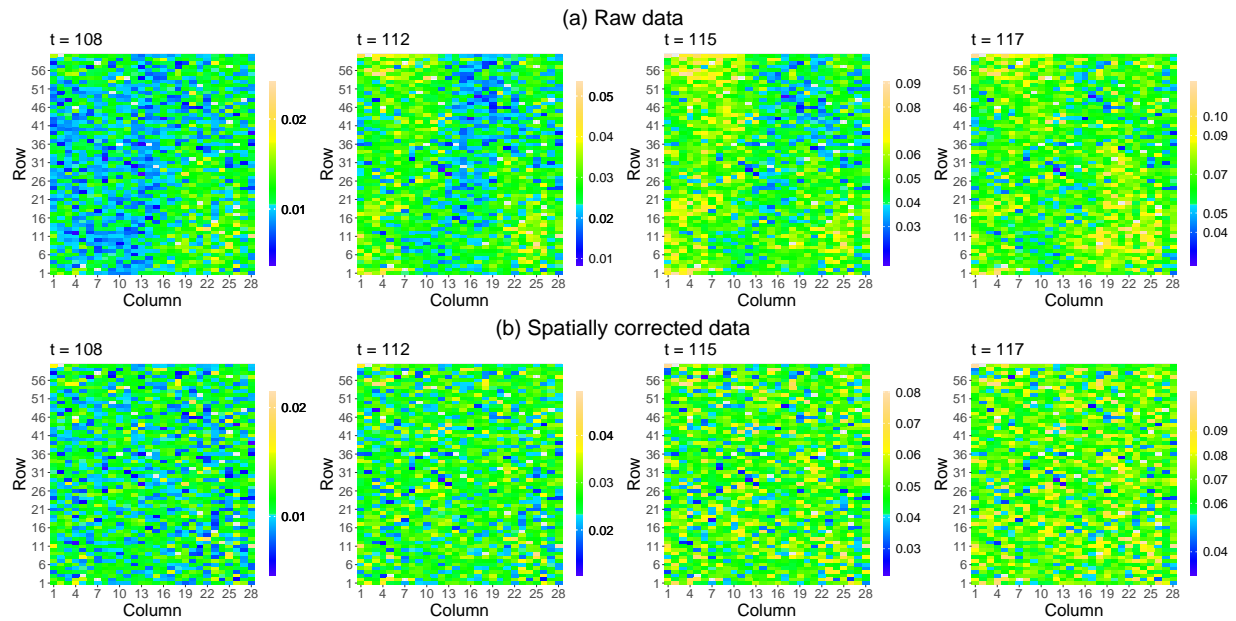
Figure 2. Results of the first stage of analysis for the PhenoArch platform: Spatial distribution of the (a) raw and (b) spatially corrected leaf area at four different measurements times ($t = 109, 113, 116, 118$ DOY). The white areas denote missing data. The colour scale is different for each time point.
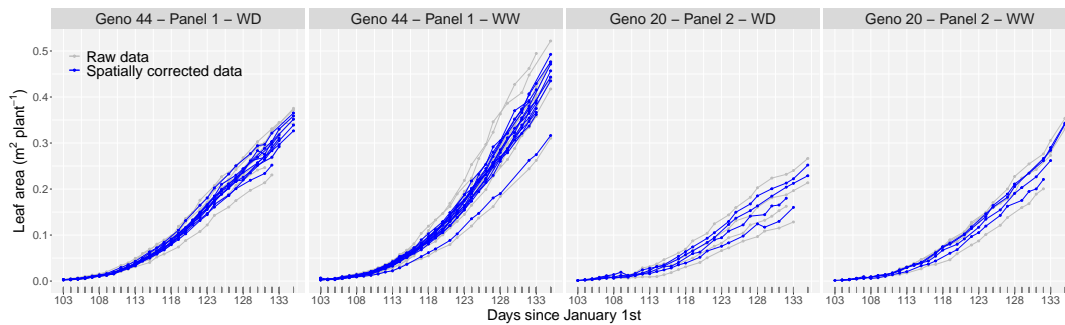

Figure 3. Results of the first stage of analysis for the PhenoArch platform: Evolution over time of the raw (grey lines) and spatially corrected (blue lines) leaf area for the plants (replicates) of two genotypes, one per panel, under the two water regimes (as illustration). WD stands for water deficit and WW for well watered.

4.1.0, and a 2.40GHz $\times$ 4 Intel® Core™ i7 processor computer with 15.6GB of RAM and Ubuntu 16.04 LTS OS.

Estimates at the three levels of the hierarchy were obtained: (1) population growth curves ($\hat{f}_p$), (2) genotype-specific deviations ($\hat{f}_{pg}$) and respective growth curves ($\hat{f}_p + \hat{f}_{pg}$), and (3) plant-specific deviations ($\hat{f}_{pgi}$) and respective growth curves ($\hat{f}_p + \hat{f}_{pg} + \hat{f}_{pgi}$). Note that the four populations (panel by water regime combination) show a different growth pattern. Plants in water deficit show a lower growth compared to well watered plants for both panels (Figure 4(a)). Genotype-specific deviations from their estimated population mean are shown in Figure 4(b) where positive and negative deviations refer, respectively, to better and worse genotypic performance compared to the mean population. As expected, the magnitude of the deviations (and, thus, the differences among genotypes performance) increases with time. Also, genotypes from Panel 1 show the largest genetic variation under both water regimes. This is in concordance with the spatially corrected data (grey lines in Figure 4(a)). The genotype-by-water regime interaction is illustrated for two genotypes per Panel in Figure 4(c). When compared to the average performance of the genotypes in the Panel, genotype 44 in Panel 1 and genotype 20 in Panel 2 behave similarly under both water regimes (they depict similar deviations under both treatments). In contrast, genotype 43 in Panel 1 presents a better performance under WD regime, while genotype 03 in Panel 2 displays a better performance under WW regime. Finally, the model is able to successfully recover the evolution over time of the spatially corrected leaf area (Figure 4(d), grey lines), while appropriately handling the missing data, as illustrated with the plant-specific growth curves (Figure 4(d), dotted blue lines) of

two genotypes (see also results for the populations means in Figure 4(a)). Additionally, genotype-specific growth curves (Figure 4(d), orange lines) seem to properly summarise/describe the behaviour of the plant curves. To asses the relative accuracy of our model in predicting the plant-specific growth curves, i.e. $\hat{f}_p + \hat{f}_{p_g} + \hat{f}_{pgi}$, we computed the mean absolute percentage error (MAPE). For this data, the MAPE value was 2.9%, indicating a good performance.
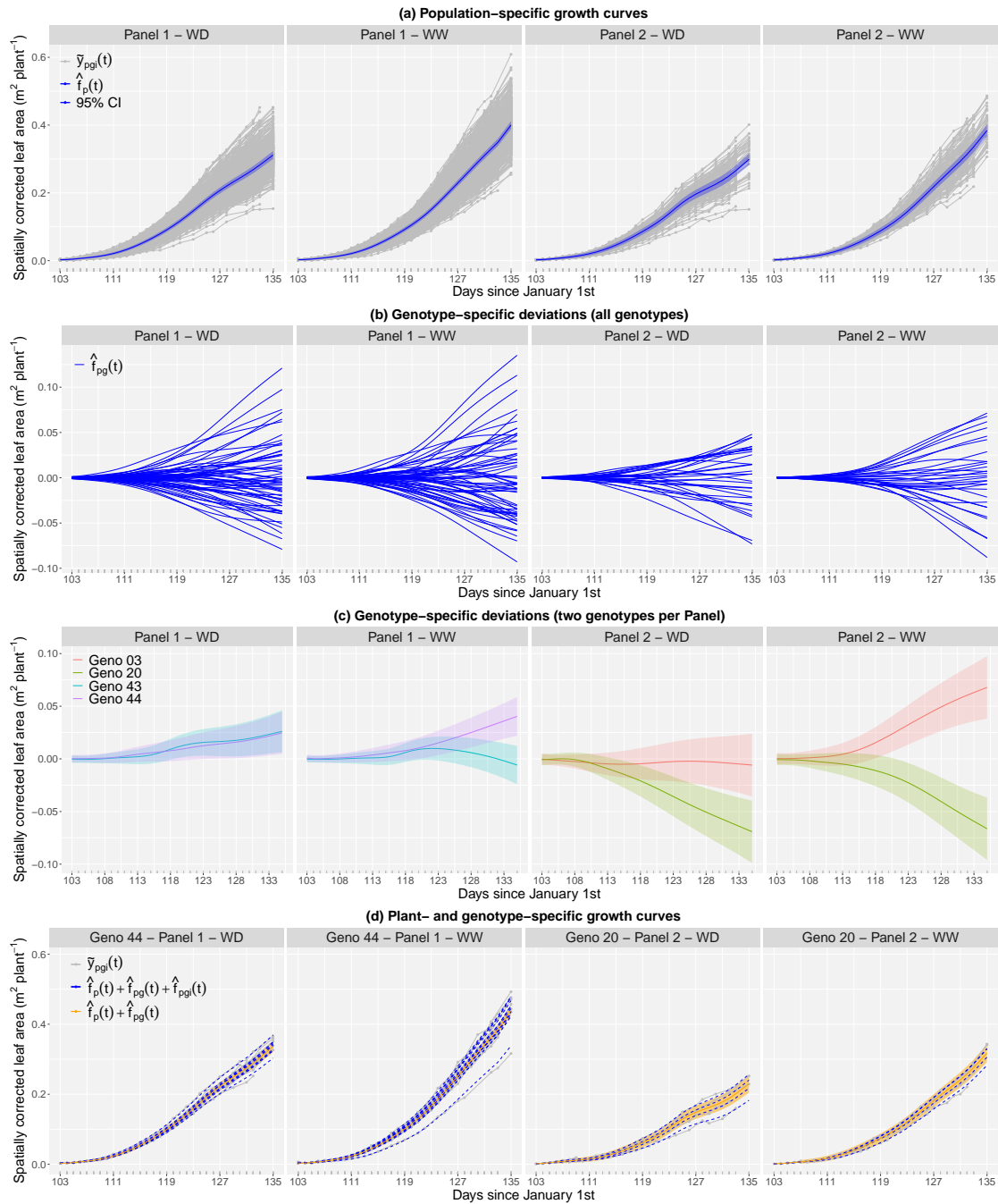

Figure 4. Results of the second stage of analysis for the PhenoArch platform: (a) Estimated population growth curves (continuous blue lines) with 95% pointwise confidence intervals (blue shaded areas), (b) estimated genotype-specific deviations for all genotypes, (c) estimated genotype-specific deviations for three genotypes per population (as illustration) with 95% pointwise confidence intervals (shaded areas); and (d) estimated plant- (dotted blue lines) and genotype- (orange continuous lines) specific growth curves with 95% pointwise confidence intervals (orange shaded areas) for two genotypes, one per panel, under the two water regimes (as illustration). WD stands for water deficit and WW for well watered. In (a) and (d) the grey lines represent the spatially corrected leaf area at the plant level (first stage).

FIP platform (ETH Zürich)

The FIP platform, located at the ETH research station in Lindau-Eschikon (Switzerland), is a cable-suspended multi-sensor platform designed for automated, accurate and supervised high throughput data acquisition on an area of 1 hectare[56]. From 2015 to 2017, the FIP platform was used to measure the development of canopy height on a diverse panel of European wheat genotypes (GABI wheat[57]), including a panel of Swiss varieties[15, 16, 51]. Here we focus on the 2017 data where height measurements started well before jointing. Figure 1(b) shows the FIP platform with its crop rotation allocated to six different lots; in 2017 the wheat experiment was planted in lots 2 and 6. Details on the experiment and experimental design can be found in Kronenberg *et al.* (2017, 2021)[16, 51]. In short, the experimental unit was a plot to which the genotypes were allocated as only treatment factor in an augmented 2D design. Three checks were placed in nine complete replications per lot, and test genotypes were allocated in a row-column design assuming that each row within a replicate (lot) received different environmental conditions (due to variability of crop husbandry measures, such as crop protection and fertilisation) while the upper, central and lower range of each lot received similar conditions (due to the similar slope direction within both lots). Canopy height measurements were carried out in irregular intervals of 2 to 13 days between February 27th and June 30th 2017 (which correspond to 58 and 181 DOY, respectively), using a terrestrial laser scanner mounted on the FIP sensor head[51]. To analyse the experiment, we arranged the two replicates (lots) diagonally in a virtual grid of $r = 42$ rows by $c = 36$ columns. Additionally, the country in which a genotype was first inscribed into the European variety catalogue was also considered in the analyses. We used this information to allocate the genotypes to different wheat populations targeted for specific regions within Europe. We will refer to these wheat populations (groups of genotypes) as regions. Accordingly, plots are nested in genotypes, and genotypes are nested in regions (wheat populations). The number of genotypes per region was as follows: Austria/Czechia (AT/CZ): 17; Switzerland (CH): 39; Germany (DE): 114; Sweden/Denmark (SE/DK): 11; France (FR): 90; Great Britain (GB): 38; Poland (PL): 25. The final dataset consists of $n = 23$ dates at which canopy height was measured on $M = 720$ plots ($720 \times 23 = 16560$ observations, including missing data), $L = 334$ genotypes, and $k = 7$ populations (regions).

First stage results: FIP platform

In this case, the SpATS model included, besides the spatial trend and the random genotypic effects, fixed effects for the two lots (experimental design factor) and the seven wheat populations (region of origin), as well as random effects for the row and column positions. We also considered different genetic variances for each of the seven populations (regions). For the spatial trend, basis dimensions of 42 and 36 were assumed for the row and column positions of the virtual grid, respectively. The correction included the estimated population and genotypic effects and the residuals, and we averaged over the lot fixed effect to eliminate its impact. The spatially corrected canopy heights obtained when modelling genotypes either as fixed or random were, also here, very similar (see Supplementary Fig. S5 online). As for the PhenoArch experiment, the correction performed in the first stage reduced the variability among replicates of the same genotype (see Supplementary Fig. S4 and S6 online). This reduction is due to the lot effect and the spatial variation. The spatial distribution of the raw and spatially corrected canopy height, at four different measurement days (DOY), is shown in Figure 5.
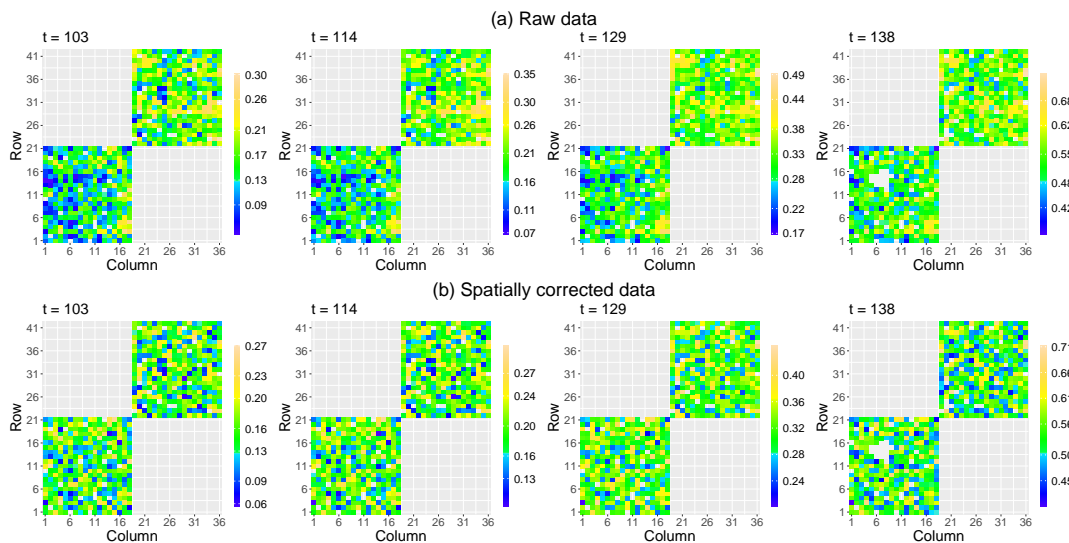

Figure 5. Results of the first stage of analysis for the ETH field phenotyping platform: Spatial distribution of the (a) raw and (b) spatially corrected canopy height at four different measurement days ($t = 103, 114, 129, 138$ DOY). The white areas denote missing data. The colour scale is different for each day.

Second stage results: FIP platform

For this analysis, and to allow the genetic variation to vary across regions (as suggested by the spatially corrected data; see Supplementary Fig. S4(b) online), we assumed, in equation (4), different values of $\sigma^2_{gen,0}$, $\sigma^2_{gen,1}$ and σ^2_{gen} per region. The growth curves for the spatially corrected phenotype show here a more complex pattern than for the PhenoArch platform (see Supplementary Fig. S2(b) and S4(b) online), so we considered cubic B-spline bases of dimension 20 for the three levels of the hierarchy (f_p, f_{pg} and f_{pgi}). Under this configuration, mixed model in equation (5) has a total of 21220 coefficients, 16560 observations (including missing data) and 32 variance components. The fitting processes needed approximately 10 minutes. The MAPE was 1.94%.

In addition to the estimated growth curves and deviations at the three levels of the hierarchy, for this dataset we also obtained the first-order derivatives of the region and genotype-specific growth curves (Figure 6); they have shown to be good descriptors of genotype specific growth habit[52] (recall that the first-order derivative is an indicator of the growth rate, i.e. the speed of canopy height development). Results show differences between regions (and between genotypes within a region) in, e.g., growth patterns (Figure 6(a)), growth rates (Figure 6(b)), and genotypes performance (Figure 6(c)). To better illustrate differences among regions, Figure 7(a) shows the estimated region growth curves together with the daily average temperature along the experiment. In Figure 7 (b) important maxima (i.e. largest growth rates) are indicated in the first-order derivatives of the respective region growth curves. Results for the replicate-specific growth curves are included in Supplementary Fig. S7 online (for one genotype per region, as illustration).
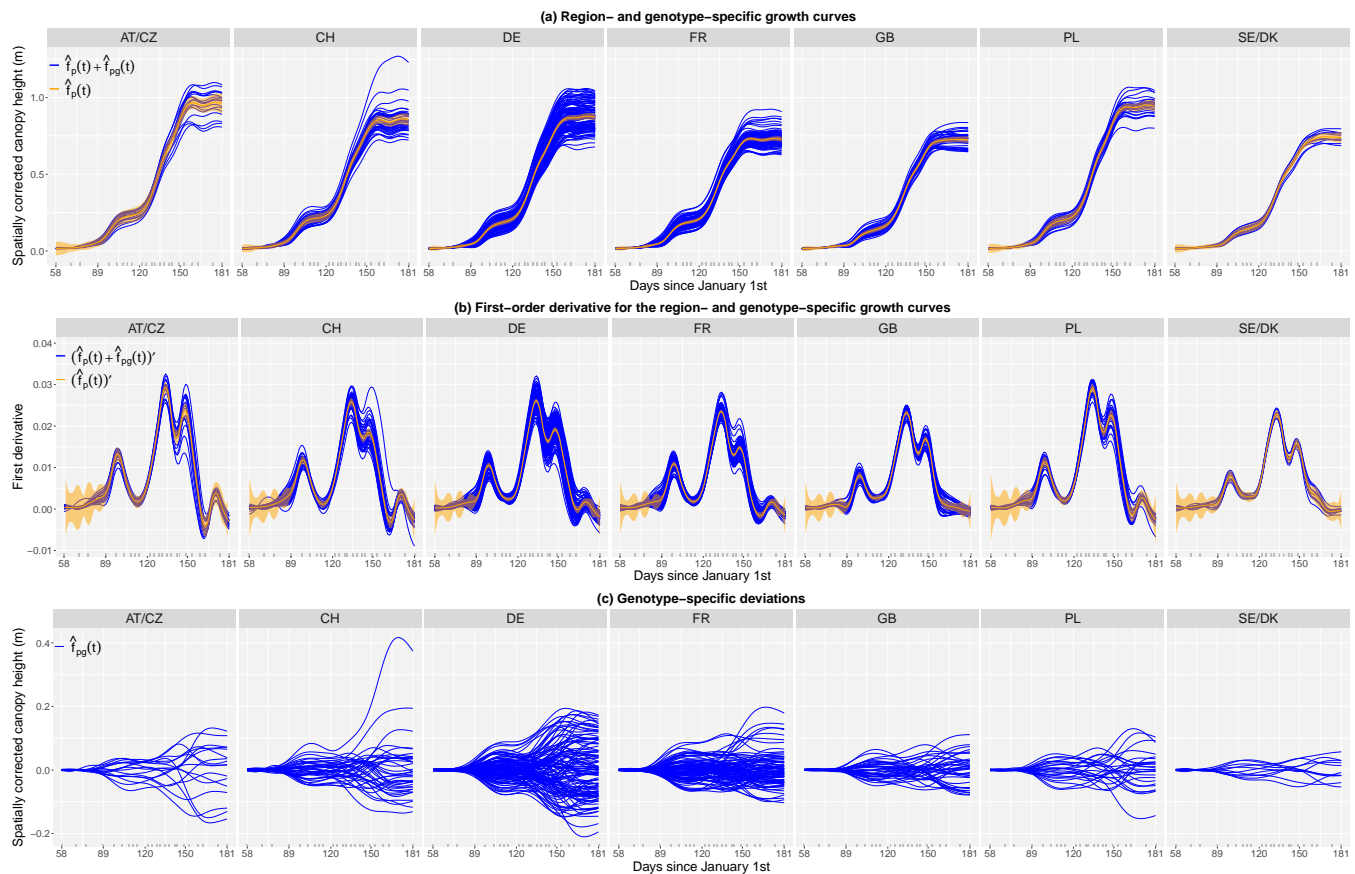

Figure 6. Results of the second stage of analysis for the ETH field phenotyping platform: (a) estimated region (orange) and genotype-specific (blue) growth curves, (b) estimated region (orange) and genotype-specific (blue) first-order derivatives, and (c) estimated genotype-specific deviations. In (a) and (b) the orange shaded areas denote 95% pointwise confidence intervals at the region level. AT/CZ: Austria/Czechia; CH: Switzerland; DE: Germany; FR: France; GB: Great Britain; PL: Poland; SE/DK: Sweden/Denmark.

Extracting time-independent attributes to characterise genotypes

To characterise the genotypes, we extracted the maximum corrected canopy height (maxTrait) from the estimated genotype-specific growth curves (Figure 6(a)) as well as three maximum speed rates (maxSpeed1, maxSpeed2 and maxSpeed3) from
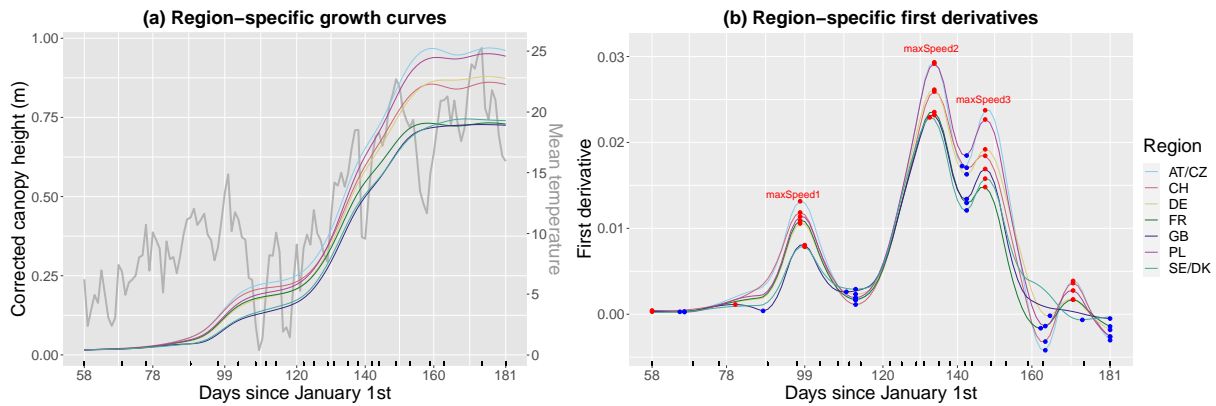
Figure 7. Results of the second stage for the ETH field phenotyping platform: (a) region-specific growth curves (coloured lines) vs. mean temperature (grey line), and (b) region-specific first-order derivatives; blue and red points indicate (local) minima and maxima, respectively. AT/CZ: Austria/Czechia; CH: Switzerland; DE: Germany; FR: France; GB: Great Britain; PL: Poland; SE/DK: Sweden/Denmark.

their first-order derivatives (Figures 6(b) and 7(b)). These speed rates correspond to maxima around DOYs 97, 133 and 147, respectively. Maximum speed rates around DOY 133 (maxSpeed2) and 189 (maxSpeed3) can be interpreted, respectively, as recoveries after a severe cold period in April (DOY 110-120) and a milder one in May (DOY 140) (Figure 7(a), grey line). However, it appears that already for maxSpeed3 the growth rates declined as plants approached their final height. We also estimated the area under the genotype-specific deviations (Figure 6(c)) as a global measure of a genotype performance over time when compared to the genotypes of the same region. A positive (negative) area indicates a genotype performance better (worse) than the regional average. Here, the area (AUC) was estimated for the complete time interval where the genotypes were measured. Nothing, however, precludes to focus attention on a restricted time interval of interest.

The bivariate scatterplots of the extracted genotype-specific attributes show that the genotypes cluster according to their region of origin (Figure 8, lower off diagonal, one color per region). Some atypical genotypes can also be identified, for example an old, tall (>1.2m) Swiss variety. In the upper off diagonal, we report the Pearson's correlation coefficient (marginal and by region). The strongest marginal correlations are between maxSpeed3 and maxTrait, and maxSpeed2 and maxTrait; maxSpeed3 and maxSpeed2 also present a high correlation. It is noteworthy that the marginal correlation between maxTrait and AUC is lower than the conditional correlation by region. Univariate analysis of each attribute using boxplots also shows regional clustering, and that clusters might change along time (e.g., compare clusters in maxSpeed1, maxSpeed2 and maxSpeed3).

Use of time-independent attributes to characterise regional adaptation

While a deeper physiological analysis is beyond the scope of this publication, we will use the extracted attributes to highlight the potential benefit of an in-depth analysis of spline-based growth patterns. Here, we use the regional groups but a similar analysis could be done using individual genotypes. The observed height development follows a principally logistic growth curve: stem elongation started after the plants were vernalised over winter (by means of cold exposure) and ended around flowering. However, the height development plateaued between day 103 and 118 in all genotypes, most likely due to a cold period in April. When looking at the region specific growth curves, this short and extreme phase caused even rank changes in growth (first-order derivative in Figure 7): the regional groups showing most vigorous growth before the stress (maxSpeed1) stopped growth completely while those which grew slowest could maintain some growth during the cold (first local minima following maxSpeed1). Such pattern may point to physiological adaptations to the different climatic regions of Europe as the slow-growing northern types from Great Britain (GB) and Denmark and Sweden (SE/DK) showed least response to cold while the fast-growing continual types from Poland (PL) and Austria and Czechia (AT/CZ) stopped growing. Moreover, the genotypes from the south-west – France (FR) and Switzerland (CH) – did not recover growth up to the same level as the more northern and eastern varieties did (compare maxSpeed1 with maxSpeed2). We acknowledge that a solid interpretation of this pattern and its significance will require a multi-year analysis to sample the typical genotype or region-specific average development.

Discussion

In agricultural and breeding research, non-destructive data acquisition of phenotypic traits by HTP platforms has emerged in recent years as a rich source of new information on plant growth and development as well as on genotypic performance. This has been accompanied by the need for novel and appropriate statistical methods of analysis. In this paper, we focus on
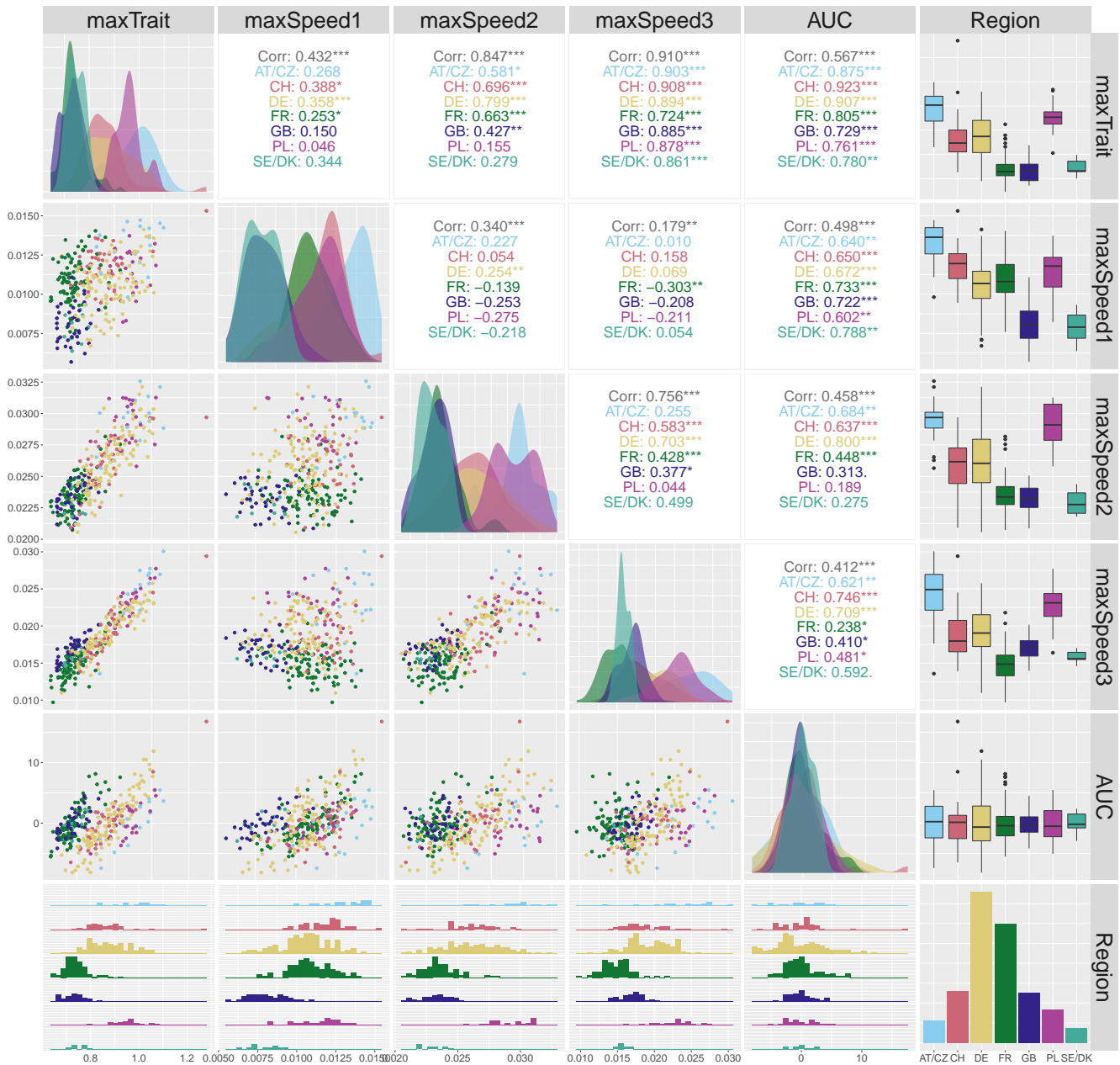
Figure 8. Scatterplots matrix with the extracted attributes at the genotype level for the ETH field phenotyping platform. The lower off diagonal depicts bivariate scatterplots, the diagonal shows the conditional densities of each attribute per region, the upper off diagonal indicates the bivariate Pearson correlation (marginal and by region), the last column displays the boxplots of each attribute per region, the last row depicts the conditional histograms of each attribute per region, and the bottom right barplot shows the number of genotypes per region. AT/CZ: Austria/Czechia; CH: Switzerland; DE: Germany; FR: France; GB: Great Britain; PL: Poland; SE/DK: Sweden/Denmark.

the analysis of a common type of HTP data in the form of time-series observed on phenotypic traits that are still close to the platform measurements (low-level traits), with these time-series occurring within a hierarchical structure, of plants or plots nested in genotypes, and genotypes nested in populations (of genotypes). We propose a statistically flexible and computationally feasible method that decomposes the required spatio-temporal analysis into two stages. In the first stage, we correct the "raw" HTP data for (nuisance) spatial variation and obtain spatially corrected time-series at the resolution of plants or plots with reduced between replicates/plots variability. The second stage consists of a temporal analysis with a hierarchical curve data model to jointly estimate growth curves at each level of the hierarchy (plant or plot, genotype, and population) as well as

their first-order derivatives. From the estimated curves and their derivatives, we calculate new phenotypic traits, attributes, that we call intermediate traits. Although we have not explicitly covered it in the paper, these new phenotypic traits, as well as the estimated curves, can be used for selection purposes in plant breeding, i.e., to differentiate between genotypes. For instance, estimates for intermediate-level traits can be used as genotypic covariates in models for higher-order phenotypic traits as described by van Eeuwijk *et al.* (2019)[2]. Higher order phenotypic traits can be understood as functions of low and intermediate traits, for either or both of biological and statistical reasons. For example, yield can be interpreted as a high-order phenotypic trait that can be modelled as a function of yield components, where the yield components may represent low and intermediate level phenotypic traits. In Roth *et al.* (2021)[34], intermediate traits obtained from modelling HTP data are included into genotype-by-environment interaction analyses, and Moreira *et al.* (2020)[58] discuss the use of information obtained from HTP time-series traits for genomic selection and the detection of QTL and causal variants. Finally, while we have focused in the paper on data with a nested structure, the proposed modelling framework can easily be extended to accommodate more complex structures, such as data with crossed levels of grouping[21] (e.g., when modelling genotype-by-treatment interactions is of interest).

Regarding the statistical methods used at each stage of our modelling approach, the first stage relies on the SpATS model, and (hierarchical) P-splines are used for the second stage. However, our approach is flexible with respect to the choice for both the first and second stages. For instance, the separable autoregressive model[6] represents a clear alternative for the first stage, while, for the second stage, hierarchical functional principal component analysis can be used[28]. We believe that our double P-spline approach is attractive both computationally and for interpretation, and the HTP data that we analysed in this paper and in other projects show that it works well (see Millet *et al.*, 2021[20] and https://eppn2020.plant-phenotyping.eu/ for more examples). Nevertheless, fairness requires us to mention some limitations as well. The P-spline hierarchical data model used in the second stage relies on specifying each function in the model using B-spline basis expansions. The model contains, consequently, a large number of parameters. Typically, computational times are within an acceptable range (for the experiments analysed in this paper, around 10 minutes). Yet, the approach may not scale well to experiments where the number of plants (and associated basis dimension) is very large (due to the size of the system of equations to be solved). Regarding the number of B-spline basis functions, we recommend using the same value for the three levels of the hierarchy, even if this increases computation. The final number does not seem to impact results (estimated curves) significantly, provided it is large enough to capture the underlying patterns. However, the estimated first-order derivatives have shown to be more sensitive to the number of basis functions.

In the first stage of our two-stage approach, analyses are done separately per time point. As such, information on spatial heterogeneity is not shared across different measurement times, and there is the need that all plants or plots are measured at (approximately) the same times. From a modelling perspective, the simultaneous modelling of spatial and temporal genetic and non-genetic variation in a one-stage model will serve to share information on common spatial variability across measurement times and it may solve the problem of variable measurement times. Yet, one-stage approaches have the limitation of being very computationally demanding, especially when the number of observations and/or the parameters to be estimated is very large. Preliminary and proof-of-concept results can be found in Pérez *et al.* (2020)[14] and in Verbyla *et al.* (2021)[13]. In a similar vein to the work presented here, Pérez *et al.* (2020)'s approach is based on using three-dimensional P-splines for modelling the non-genetic spatio-temporal (noise) process and hierarchical P-splines for the longitudinal genetic signal. In Verbyla *et al.* (2021), multi-dimensional splines are also used for the spatio-temporal noise, whereas a factor analytic structure is considered for the genetic effects. Although promising, both approaches suffer from severe computational limitations (e.g., Verbyla *et al.* 2021, report convergence requiring several days). More work is thus required to render them feasible for routine use, and to date, stage-wise approaches represent not only computationally feasible but also statistically valid alternatives.

Apart from this work, van Eeuwijk *et al.* (2019)[2] also proposed a two-stage approach for the analysis of HTP data, where they first correct for spatial variation, and then focus on estimating, and further processing, the temporal dynamic of the genetic effects. In that paper, spatially adjusted genotypic means are carried to the temporal analysis. In contrast, our proposal allows keeping the data resolution for the second stage at the experimental unit. Also, in the second stage, we jointly model the whole sample of spatially corrected growth curves, while in van Eeuwijk *et al.* (2019)[2] analyses are done separately per genotype. Our hierarchical approach thus allows borrowing strength across plant/plot curves for more efficient estimation of genotype and population growth trajectories. This is particularly important in the presence of incomplete data. We are aware that when choosing a two-stage approach, one might also choose, as done in Roth *et al.* (2021)[34], to first model the longitudinal variation at a plant or plot level and subsequently apply a spatial correction to extracted features. In that paper, a P-spline model is first fitted separately for each plot time-series, from which the timing of key stages (among other features) are extracted. These intermediate traits are then processed to obtain spatially adjusted genotypic means for further analyses. We feel that both options - with the spatial or the temporal analysis first - represent valid alternatives. The choice for one or another method will depend on the relative magnitudes of the various spatial, temporal and spatio-temporal genetic and non-genetic processes and may be difficult to assess beforehand. A study of the proposal that best suits particular situations represents an interesting area

of study. It should be realised that stage-wise analyses make it necessary to propagate uncertainty from stage to stage. Here, we did it by weighting the second stage with the (inverse) of the estimated variance associated with the spatially corrected trait. In the HTP context, weighting has shown to improve results[34].

All in all, the two-stage approach described in this paper represents a good compromise between accuracy, adequacy, computational efficiency and interpretability. The results show that our proposal is feasible on standard computers and delivers good descriptions of the genetic and non-genetic variation in the temporal dimension together with useful summary statistics for breeding purposes. We believe that it represents a powerful tool for routine application in phenotyping experiments with dense time series. To allow practitioners to use our proposal, the R-functions implementing it are publicly available at `https://gitlab.bcamath.org/dperez/htp_two_stage_approach`, where we also provide the code to reproduce the analyses and results shown in the paper.

Three-level Nested Hierarchical Growth Model: Mixed Model Formulation

On the basis of model in equation (3) and specification in equation (4), we use the notation $\boldsymbol{\beta}_p = (\beta_{p,0}, \beta_{p,1})^T$, $\boldsymbol{\beta}_{pg} = (\beta_{pg,0}, \beta_{pg,1})^T$, $\boldsymbol{\beta}_{pgi} = (\beta_{pgi,0}, \beta_{pgi,1})^T$, $\boldsymbol{u}_p = (u_{p,1}, \dots, u_{p,b_{gen}-2})^T$, $\boldsymbol{u}_{pg} = (u_{pg,1}, \dots, u_{pg,b_{pop}-2})^T$, $\boldsymbol{u}_{pgi} = (u_{pgi,1}, \dots, u_{pgi,b_{plant}-2})^T$, $\boldsymbol{z}_{pop,j} = (z_{pop,j}(t_1), \dots, z_{gen,j}(t_n))^T$, $\boldsymbol{z}_{gen,j} = (z_{gen,j}(t_1), \dots, z_{gen,j}(t_n))^T$ and $\boldsymbol{z}_{plant,j} = (z_{plant,j}(t_1), \dots, z_{plant,j}(t_n))^T$. Then, in matrix notation, the three-level nested hierarchical growth model can be expressed as

$$\tilde{\boldsymbol{y}} = \underbrace{[\boldsymbol{Q}_{pop} \otimes \boldsymbol{X}_{pop}]\boldsymbol{\beta} + [\boldsymbol{Q}_{pop} \otimes \boldsymbol{Z}_{pop}]\boldsymbol{u}_{pop}}_{f_p} + \underbrace{[\boldsymbol{Q}_{gen} \otimes \boldsymbol{X}_{gen}]\boldsymbol{\beta}_{gen} + [\boldsymbol{Q}_{gen} \otimes \boldsymbol{Z}_{gen}]\boldsymbol{u}_{gen}}_{f_{pg}}$$
$$+ \underbrace{[\boldsymbol{I}_M \otimes \boldsymbol{X}_{plant}]\boldsymbol{\beta}_{plant} + [\boldsymbol{I}_M \otimes \boldsymbol{Z}_{plant}]\boldsymbol{u}_{plant}}_{f_{pgi}} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{W}),$$

where $\otimes$ is the Kronecker product, $\boldsymbol{Q}_{pop}$ and $\boldsymbol{Q}_{gen}$ are matrices assigning, respectively, plants to populations and plants to genotypes, and

$$\boldsymbol{X}_{pop} = [\boldsymbol{1} \mid \boldsymbol{t}] \text{ and } \boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_k^T)^T,$$
$$\boldsymbol{Z}_{pop} = [\boldsymbol{z}_{pop,1} \mid \dots \mid \boldsymbol{z}_{pop,b_{pop}-2}] \text{ and } \boldsymbol{u}_{pop} = (\boldsymbol{u}_1^T, \dots, \boldsymbol{u}_k^T)^T \sim N(\boldsymbol{0}, \text{blockdiag}(\sigma_1^2\boldsymbol{I}_{b_{pop}-2}, \dots, \sigma_k^2\boldsymbol{I}_{b_{pop}-2})),$$
$$\boldsymbol{X}_{gen} = [\boldsymbol{1} \mid \boldsymbol{t}] \text{ and } \boldsymbol{\beta}_{gen} = (\boldsymbol{\beta}_{11}^T, \dots, \boldsymbol{\beta}_{k\ell_k}^T)^T \sim N(\boldsymbol{0}, \boldsymbol{I}_L \otimes \boldsymbol{\Sigma}_{gen}) \text{ with } \boldsymbol{\Sigma}_{gen} = \text{diag}(\sigma_{gen,0}^2, \sigma_{gen,1}^2),$$
$$\boldsymbol{Z}_{gen} = [\boldsymbol{z}_{gen,1} \mid \dots \mid \boldsymbol{z}_{gen,b_{gen}-2}] \text{ and } \boldsymbol{u}_{gen} = (\boldsymbol{u}_{11}^T, \dots, \boldsymbol{u}_{k\ell_k}^T)^T \sim N(\boldsymbol{0}, \sigma_{gen}^2\boldsymbol{I}_L \otimes \boldsymbol{I}_{b_{gen}-2}),$$
$$\boldsymbol{X}_{plant} = [\boldsymbol{1} \mid \boldsymbol{t}] \text{ and } \boldsymbol{\beta}_{plant} = (\boldsymbol{\beta}_{111}^T, \dots, \boldsymbol{\beta}_{k\ell_k m_{k\ell_k}}^T)^T \sim N(\boldsymbol{0}, \boldsymbol{I}_M \otimes \boldsymbol{\Sigma}_{plant}) \text{ with } \boldsymbol{\Sigma}_{plant} = \text{diag}(\sigma_{plant,0}^2, \sigma_{plant,1}^2),$$
$$\boldsymbol{Z}_{plant} = [\boldsymbol{z}_{plant,1} \mid \dots \mid \boldsymbol{z}_{plant,b_{plant}-2}] \text{ and } \boldsymbol{u}_{plant} = (\boldsymbol{u}_{111}^T, \dots, \boldsymbol{u}_{k\ell_k m_{k\ell_k}}^T)^T \sim N(\boldsymbol{0}, \sigma_{plant}^2\boldsymbol{I}_M \otimes \boldsymbol{I}_{b_{plant}-2}).$$

Finally, $\boldsymbol{W}$ is diagonal matrix whose diagonal entries are the weights from the first stage. We finish by noting that, under the previous model specification, there is the assumption that observations arising from the same plant, genotype or population are serially correlated, and the correlation increases as a function of the shared grouping levels[21]. In particular, curves $\tilde{\boldsymbol{y}}_{pgi}$ and $\tilde{\boldsymbol{y}}_{p'g'j'}$ have covariance

$$cov(\tilde{\boldsymbol{y}}_{pgi}, \tilde{\boldsymbol{y}}_{p'g'j'}) = \begin{cases} \boldsymbol{0} & p \neq p' \\ \sigma_p^2\boldsymbol{Z}_{pop}\boldsymbol{Z}_{pop}^T & p = p', g \neq g' \\ \sigma_p^2\boldsymbol{Z}_{pop}\boldsymbol{Z}_{pop}^T + \boldsymbol{X}_{gen}\boldsymbol{\Sigma}_{gen}\boldsymbol{X}_{gen}^T + \sigma_{gen}^2\boldsymbol{Z}_{gen}\boldsymbol{Z}_{gen}^T & g = g', i \neq i' \\ \sigma_p^2\boldsymbol{Z}_{pop}\boldsymbol{Z}_{pop}^T + \boldsymbol{X}_{gen}\boldsymbol{\Sigma}_{gen}\boldsymbol{X}_{gen}^T + \sigma_{gen}^2\boldsymbol{Z}_{gen}\boldsymbol{Z}_{gen}^T + \boldsymbol{X}_{plant}\boldsymbol{\Sigma}_{plant}\boldsymbol{X}_{plant}^T + \sigma_{plant}^2\boldsymbol{Z}_{plant}\boldsymbol{Z}_{plant}^T + \sigma^2\text{diag}(\boldsymbol{w}_{pgi}) & i = i'. \end{cases}$$

Data availability

Both the PhenoArch and FIP datasets analysed during the current study are available within the article and its "Supplementary information".

Code availability

The R-functions implementing the two-stage approach, as well as the R-codes needed to reproduce the analyses for the PhenoArch and FIP platforms are available within the article and its "Supplementary information".

References

1. Tardieu, F., Cabrera-Bosquet, L., Pridmore, T. & Bennett, M. Plant phenomics, from sensors to knowledge. *Curr. Biol.* **27**, R770–R783, DOI: https://doi.org/10.1016/j.cub.2017.05.055 (2017).

2. van Eeuwijk, F. A. *et al.* Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Sci.* **282**, 23–39, DOI: https://doi.org/10.1016/j.plantsci.2018.06.018 (2019).

3. Araus, J. L. & Cairns, J. E. Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* **19**, 52–61, DOI: https://doi.org/10.1016/j.tplants.2013.09.008 (2014).

4. Green, P., Jennison, C. & Seheult, A. Analysis of field experiments by least squares smoothing. *J. Royal Stat. Soc. Ser. B* **47**, 299–315, DOI: https://doi.org/10.1111/j.2517-6161.1985.tb01358.x (1985).

5. Cullis, B. & Gleeson, A. Spatial analysis of field experiments-an extension to two dimensions. *Biometrics* 1449–1460, DOI: https://doi.org/10.2307/2532398 (1991).

6. Gilmour, A. R., Cullis, B. R. & Verbyla, A. P. Accounting for natural and extraneous variation in the analysis of field experiments. *J. Agric. Biol. Environ. Stat.* 269–293, DOI: https://doi.org/10.2307/1400446 (1997).

7. Verbyla, A. P., Cullis, B. R., Kenward, M. G. & Welham, S. J. The analysis of designed experiments and longitudinal data by using smoothing splines. *J. Royal Stat. Soc. Ser. C* **48**, 269–311, DOI: https://doi.org/10.1111/1467-9876.00154 (1999).

8. Besag, J. & Higdon, D. Bayesian analysis of agricultural field experiments. *J. Royal Stat. Soc. Ser. B* **61**, 691–746, DOI: https://doi.org/10.1111/1467-9868.00201 (1999).

9. Durban, M. *et al.* The practical use of semiparametric models in field trials. *J. Agric. Biol. Environ. Stat.* **8**, 48, DOI: https://doi.org/10.1198/1085711031265 (2003).

10. Piepho, H. & Williams, E. Linear variance models for plant breeding trials. *Plant Breed.* **129**, 1–8, DOI: https://doi.org/10.1111/j.1439-0523.2009.01654.x (2010).

11. Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A. & Eilers, P. H. Correcting for spatial heterogeneity in plant breeding experiments with p-splines. *Spatial Stat.* **23**, 52–71, DOI: https://doi.org/10.1016/j.spasta.2017.10.003 (2018).

12. Kar, S. *et al.* Spatemhtp: A data analysis pipeline for efficient processing and utilization of temporal high-throughput phenotyping data. *Front. Plant Sci.* **11**, 1746, DOI: https://doi.org/10.3389/fpls.2020.552509 (2020).

13. Verbyla, A., De Faveri, J., Deery, D. & Rebetzke, G. Modelling temporal genetic and spatio-temporal residual effects for high-throughput phenotyping data. *Aust. & New Zealand J. Stat.* DOI: https://doi.org/10.1111/anzs.12336 (2021).

14. Pérez, D. M., Rodríguez-Álvarez, M. X., Boer, M. P., Millet, E. J. & van Eeuwijk, F. A. Spatio-temporal and hierarchical modelling of high-throughput phenotypic data. In *Proceedings of the 35th International Workshop on Statistical*, 394–397 (2020). http://hdl.handle.net/10810/45863.

15. Perich, G. *et al.* Assessment of multi-image uav based high-throughput field phenotyping of canopy temperature. *Front. Plant Sci.* **11**, 150, DOI: https://doi.org/10.3389/fpls.2020.00150 (2020).

16. Kronenberg, L. *et al.* Temperature response of wheat affects final height and the timing of stem elongation under field conditions. *J. Exp. Bot.* **72**, 700–717, DOI: https://doi.org/10.1093/jxb/eraa471 (2021).

17. Anderegg, J. *et al.* Spectral vegetation indices to track senescence dynamics in diverse wheat germplasm. *Front. Plant Sci.* **10**, 1749, DOI: https://doi.org/10.3389/fpls.2019.01749 (2020).

18. Velazco, J. G. *et al.* Modelling spatial trends in sorghum breeding field trials using a two-dimensional p-spline mixed model. *Theor. Appl. Genet.* **130**, 1375–1392, DOI: https://doi.org/10.1007/s00122-017-2894-4 (2017).

19. Andrade, M. H. M. L. *et al.* Accounting for spatial trends to increase the selection efficiency in potato breeding. *Crop. Sci.* DOI: https://doi.org/10.1002/csc2.20226 (2020).

20. Millet, E. J. *et al. statgenHTP: High Throughput Phenotyping (HTP) Data Analysis* (2021). R package version 1.0.3.

21. Brumback, B. A. & Rice, J. A. Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Am. Stat. Assoc.* **93**, 961–976, DOI: https://doi.org/10.1080/01621459.1998.10473755 (1998).

22. Paine, C. T. *et al.* How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. *Methods Ecol. Evol.* **3**, 245–256, DOI: https://doi.org/10.1111/j.2041-210X.2011.00155.x (2012).

23. Eilers, P. H. & Marx, B. D. Flexible smoothing with b-splines and penalties. *Stat. Sci.* 89–102, DOI: https://doi.org/10.1214/ss/1038425655 (1996).

24. Durban, M., Harezlak, J., Wand, M. P. & Carroll, R. J. Simple fitting of subject-specific curves for longitudinal data. *Stat. Medicine* **24**, 1153–1167, DOI: https://doi.org/10.1002/sim.1991 (2005).

25. Di, C.-Z., Crainiceanu, C. M., Caffo, B. S. & Punjabi, N. M. Multilevel functional principal component analysis. *The Annals Appl. Stat.* **3**, 458, DOI: https://doi.org/10.1214/08-AOAS206SUPP (2009).

26. Xu, Y., Li, Y. & Nettleton, D. Nested hierarchical functional data modeling and inference for the analysis of functional plant phenotypes. *J. Am. Stat. Assoc.* **113**, 593–606, DOI: https://doi.org/10.1080/01621459.2017.1366907 (2018).

27. Xu, Y., Qiu, Y. & Schnable, J. C. Functional modeling of plant growth dynamics. *The Plant Phenome J.* **1**, 1–10, DOI: https://doi.org/10.2135/tppj2017.09.0007 (2018).

28. Xu, Y., Li, Y. & Qiu, Y. Growth dynamics and heritability for plant high-throughput phenotyping studies using hierarchical functional data analysis. *Biom. J.* **(in press)**, DOI: https://doi.org/10.1002/bimj.202000315 (2021).

29. Wang, R., Qiu, Y., Zhou, Y., Liang, Z. & Schnable, J. C. A high-throughput phenotyping pipeline for image processing and functional growth curve analysis. *Plant Phenomics* **2020**, DOI: https://doi.org/10.34133/2020/7481687 (2020).

30. Brien, C., Jewell, N., Watts-Williams, S. J., Garnett, T. & Berger, B. Smoothing and extraction of traits in the growth analysis of noninvasive phenotypic data. *Plant Methods* **16**, 1–21, DOI: https://doi.org/10.1186/s13007-020-00577-6 (2020).

31. Greven, S. & Scheipl, F. A general framework for functional regression modelling. *Stat. Model.* **17**, 1–35, DOI: https://doi.org/10.1177/1471082X16681317 (2017).

32. Currie, I. D. & Durban, M. Flexible smoothing with P-splines: a unified approach. *Stat. Model.* **2**, 333–349, DOI: https://doi.org/10.1191/1471082x02st039ob (2002).

33. Wand, M. P. Smoothing and mixed models. *Comput. Stat.* **18**, 223–249, DOI: https://doi.org/10.1007/s001800300142 (2003).

34. Roth, L., Rodríguez-Álvarez, M. X., van Eeuwijk, F., Piepho, H.-P. & Hund, A. Phenomics data processing: A plot-level model for repeated measurements to extract the timing of key stages and quantities at defined time points. *bioRxiv* DOI: https://doi.org/10.1101/2021.05.02.442243 (2021).

35. Eilers, P. H. C. & Marx, B. D. Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemom. Intell. Lab. Syst.* **66**, 159–174, DOI: https://doi.org/10.1016/S0169-7439(03)00029-7 (2003).

36. Welham, S., Cullis, B., Gogel, B., Gilmour, A. & Thompson, R. Prediction in linear mixed models. *Aust. & New Zealand J. Stat.* **46**, 325–347, DOI: https://doi.org/10.1111/j.1467-842X.2004.00334.x (2004).

37. Buntaran, H. *et al.* Cross-validation of stage-wise mixed-model analysis of swedish variety trials with winter wheat and spring barley. *Crop. Sci.* DOI: https://doi.org/10.1002/csc2.20177 (2020).

38. de Boor, C. *A practical guide to splines*, vol. 27 (springer-verlag New York, 1978).

39. Butler, D., Cullis, B., Gilmour, A., Gogel, B. & Thompson, R. *ASReml-R reference manual Version 4*. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK (2017).

40. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models* (2021). R package version 3.1-152.

41. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48, DOI: https://doi.org/10.18637/jss.v067.i01 (2015).

42. SAS Institute Inc. 2015. SAS/STAT®. *14.1 User's Guide.* Cary, NC: SAS Institute Inc.

43. Wood, S. *Generalized Additive Models: An Introduction with R* (Chapman and Hall/CRC, 2017), 2 edn.

44. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2021).

45. Rodríguez-Álvarez, M. X., Durban, M., Lee, D.-J. & Eilers, P. H. On the estimation of variance parameters in non-standard generalised linear mixed models: Application to penalised smoothing. *Stat. Comput.* 483–500, DOI: https://doi.org/10.1007/s11222-018-9818-2 (2019).

46. Henderson, C. R. Selection index and expected genetic advance. *Stat. Genet. Plant Breed.* **982**, 141–163 (1963).

47. Patterson, H. D. & Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554, DOI: https://doi.org/10.1093/biomet/58.3.545 (1971).

48. Currie, I. D., Durban, M. & Eilers, P. H. C. Generalized linear array models with applications to multidimensional smoothing. *J. Royal Stat. Soc. Ser. B* **68**, 259–280, DOI: https://doi.org/10.1111/j.1467-9868.2006.00543.x (2006).

49. Harville, D. A. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* **72**, 320–338, DOI: https://doi.org/10.2307/2286796 (1977).

50. Cabrera-Bosquet, L. *et al.* High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. *New Phytol.* **212**, 269–281, DOI: https://doi.org/10.1111/nph.14027 (2016).

51. Kronenberg, L., Yu, K., Walter, A. & Hund, A. Monitoring the dynamics of wheat stem elongation: genotypes differ at critical stages. *Euphytica* **213**, 157, DOI: https://doi.org/10.1007/s10681-017-1940-2 (2017).

52. Hurtado, P. X. *et al.* Dynamics of senescence-related QTLs in potato. *Euphytica* **183**, 289–302, DOI: https://doi.org/10.1007/s10681-011-0464-4 (2012).

53. Bustos-Korts, D. *et al.* Combining crop growth modeling and statistical genetic modeling to evaluate phenotyping strategies. *Front. Plant Sci.* **10**, DOI: https://doi.org/10.3389/fpls.2019.01491 (2019).

54. Brichet, N. *et al.* A robot-assisted imaging pipeline for tracking the growths of maize ear and silks in a high-throughput phenotyping platform. *Plant Methods* **13**, 1–12, DOI: https://doi.org/10.1186/s13007-017-0246-7 (2017).

55. Lee, D.-J., Durbán, M. & Eilers, P. Efficient two-dimensional smoothing with p-spline anova mixed models and nested bases. *Comput. Stat. & Data Analysis* **61**, 22–37, DOI: https://doi.org/10.1016/j.csda.2012.11.013 (2013).

56. Kirchgessner, N. *et al.* The eth field phenotyping platform fip: a cable-suspended multi-sensor system. *Funct. Plant Biol.* **44**, 154–168, DOI: https://doi.org/10.1071/FP16165 (2016).

57. Kollers, S. *et al.* Whole genome association mapping of fusarium head blight resistance in european winter wheat (triticum aestivum l.). *PLoS One* **8**, e57500, DOI: https://doi.org/10.1371/journal.pone.0057500 (2013).

58. Moreira, F. F., Oliveira, H. R., Volenec, J. J., Rainey, K. M. & Brito, L. F. Integrating high-throughput phenotyping and statistical genomic methods to genetically improve longitudinal traits in crops. *Front. Plant Sci.* **11**, 681, DOI: https://doi.org/10.3389/fpls.2020.00681 (2020).

Funding

Author contributions statement

D.M.P.: methodology, software, data curation, data analysis, visualization, original draft, review and editing. M.X.R.: conceptualization, methodology, software, review and editing, supervision. M.B.: methodology, software, review and editing. A.H: data collection, review and editing. L.K.: data collection, data curation, review and editing. Ll.C.B.: data collection, data curation, review and editing. E.M.: methodology, software, data curation, review and editing. F.E: conceptualization, methodology, review and editing, supervision. All authors reviewed and approved the manuscript.

Ethics declaration

Competing interests. The authors declare no competing interests.

Supplementary information

Supplementary information (code and data) is available for this paper in the Basque Center for Applied Mathematics repository, https://gitlab.bcamath.org/dperez/htp_two_stage_approach.