Optimal time scaling for plant growth analysis

Sabine K. Schnabel[1,3], Paul H.C. Eilers[1,2] and Fred A. van Eeuwijk[1,3]

[1] Biometris, Wageningen UR, Postbus 100, 6700 AC Wageningen, The Netherlands; sabine.schnabel@wur.nl (communicating author)
[2] Erasmus MC, Department of Biostatistics, Postbus 2040, 3000 CA Rotterdam, The Netherlands
[3] Center for Biosystems Genomics, Postbus 98, 6700 AB Wageningen, The Netherlands

Abstract: In field trials the development of plants is regularly scored on a visual scale. Plots of the data show strongly curved relationships with time. We investigate optimal scaling of the time axis in order to get linear curves and apply it to decay data of potato plants.

Keywords: Optimal scaling; time axis; linear model

1   Introduction

In plant research field experiments are a common instrument to study the behavior of a plant population under different environmental conditions. However, there are many aspects that contribute to the uncertainty of the data. On the one hand some phenotypic traits for plant development are only registered on an ordinal scale by qualitatively judging the level of development. It is not clear how this translates into a numeric scale used in data analysis. The distances between two subsequent levels of the observation are often not known. On the other hand another data problem might result from the different environmental conditions during a field experiment. Weather cannot be kept constant. Temperature and exposure to daylight are the main factors driving plant development. For better comparison between field experiments in different environments it is necessary to have an uniform and adapted time scale that can capture these differences. In this analysis we will focus on transforming the time axis.

2   Data and objective

During a field experiment in Finland in 2004, haulm senescence of 200 potato varieties was recorded at 11 days (Zaban et al., 2006) on a discrete

scale from "green plant" (1), "upper leaves with first signs of yellowing" (2) etc. to "dead plant" (7). Figure (1.2) shows examples of three varieties. The horizontal axis is not calendar time, but it is beta-thermal time (PBTT) (Yin et al., 1995). This is a scale, developed by plant physiologists, in which the history of daylength and temperature is integrated over the growing season.

To summarize the senescence data for each variety, we would like to fit a simple curve to them, so that only a few clearly interpretable curve characteristics can be carried on to a genetic analysis. The simplest curve is a straight line. We assume that PBTT is a first step in the direction of a linear relationship between ϕ, i.e. transformed PBTT (indicated by τ), and y, the observed scores. The same transformation of PBTT is to be used for all varieties of the population.

We do not consider transformation of the response scale — nor of both scales simultaneously — but we will return to this issue in the Discussion.

3   Theory

Let there be m time points and n varieties. The senescence scores are collected in a matrix $Y = [y_{ij}], i = 1, \ldots, m, j = 1, \ldots, n$. For simplicity we assume Y to be complete. If that is not the case, an appropriate indicator vector can be introduced easily.

Our goal is to find a vector ϕ with m elements, such that ϕ is the optimal transformed PBTT τ. Optimal means that we get the best possible linear correlation between ϕ and each column of Y. This leads to the following objective function to be minimized:

$$S = \sum_j \sum_i (y_{ij} - \alpha_j - \beta_j \phi_i)^2$$

Given ϕ, we are looking for the least squares regression line for each variety. This is an ill-posed problem, because any arbitrary shifting and scaling of ϕ can be compensated by inverse scaling of β_j and shifting of α_j. In order to find the unique solution we want ϕ to be standardized, i.e. $\sum_i \phi_i = 0$ and $\sum_i \phi_i^2 = m$. This is an arbitrary constraint, and it is only used for fitting the model. Afterwards a linear transformation to a more meaningful scale can be applied (see Figure 1 for an example, where we have made minimum and maximum of ϕ equal to those of τ).

An intuitive algorithm repeats the following steps which start with an approximate solution $\tilde{\phi}$:

1. Estimate (new) α_j and β_j by linear regression of column j of Y on $\tilde{\phi}$ for each variety j.

2. Improve $\tilde{\phi}$ by linear regression of $y_{ij} - \alpha_j$ on β_j for each time point i.

In our experience this works well and convergence is obtained in few iterations. As starting values for $\tilde{\phi}$ we take the integers from 1 to m and standardize them.

The actual PBTT times, τ, do not occur in the estimation only their index i does. This is a consequence of the fact that all varieties have been scored on the same days. If this would not be the case, a second matrix $T = [\tau_{ij}]$ would give the actual observation times. Instead of a vector ϕ we would have to estimate a continuous function $f(\tau)$ that gives the transformation at every point in time.

A possible approach is to use B-splines for this purpose: $f(\tau) = \sum_k B_k(\tau)\gamma_k$. The second step of the algorithm above would then involve fitting the B-splines, scaled by b_j, to $y_{ij} - \alpha_j$.

4   Application

Optimal time scaling to linearity seems to be suitable for the present data. The results for the transformation of the scale as well as the linear fit before and after transformation are presented in Figure 1. The model is parsimonious and the estimated coefficients can be directly related to the development process. The slope b_j describes the speed of senescence for variety j. An important other characteristic deduced from the results is the halfway point of the senescence process, i.e. the transformed time at which the score halfway between 1 and 7 is reached.

5   Discussion and outlook

Optimal scaling is a standard tool in psychometric research and practice. To our knowledge it has not been used in plant research yet. As shown we got interesting — but also somewhat worrying — results: linearity is much better on the real time scale (just counting the days after planting) than on PBTT. The transformation we found was almost the inverse of the one from real time to PBTT.

Similar ideas can be used to transform only the senescence scores in relation to time. Technically it also possible to transform both time and scores simultaneously. However, interpretation of the results is unclear. Any monotone transformation of time combines with the corresponding inverse transform of the scores. There is a fundamental identification problem, for which we have no solution yet.

We did not show it in the examples, but a number of varieties show early saturation at the highest possible score. This leads to an S-shaped curve. No time transformation can accommodate that. Our next step will be to introduce a standard S-shaped curve –like the logistic function– with $a_j + b_j\phi_i$ as its argument. This is similar to the link function in generalized linear models.

A still more ambitious effort will be to investigate time transformations including cumulative solar radiation and temperature as explanatory variables.
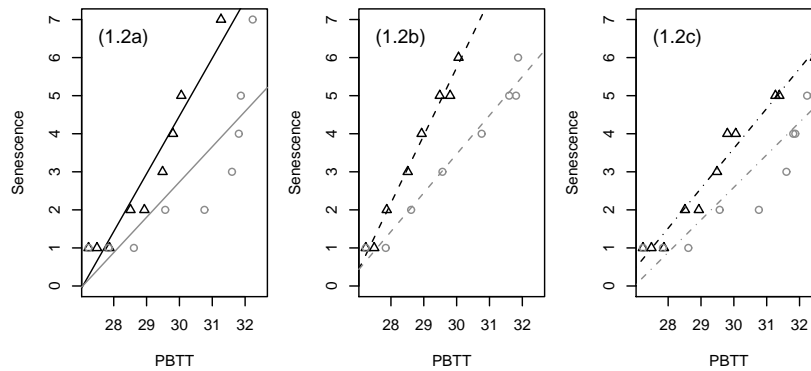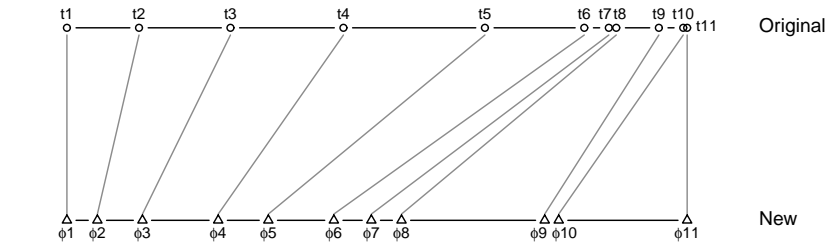

FIGURE 1. (1.1) Original scale versus new scale retransformed to original range. (1.2a-c) Original data (grey) and data on transformed time (black). Linear fit before (grey) and after transformation (black) for three selected potato varieties.

References

Yin X., Kropff M.J., McLaren G., and Visperas R.M. (1995) A nonlinear model for crop development as a function of temperature. *Agricultural and Forest Meteorology*, **77**, 1- 16.

Zaban A., Veteläinen M., Celis-Gamboa C., van Berloo R., Häggman H., and Visser R.G.F. (2006) Physiological and genetic aspects of a diploid potato population in the Netherlands and Northern Finland. *Maataloustieteen Päivät*, 1-7.