Growth curves of sorghum roots via quantile regression with P-splines

Curve di crescita per radici di sorgo attraverso regressione quantilica con P-splines

Alessio Pollice, Vito M.R. Muggeo, Federico Torretta, Rocco Bochicchio and Mariana Amato

Abstract Plant roots are a major pool of total carbon in the planet and their dynamics are directly related to greenhouse gas balance. Composted wastes are increasingly used in agriculture for environmental and economic reasons and their role as a substitute for traditional fertilizers needs to be assessed on all plant components. Here we propose a quantile regression approach based on P-splines to quantify and compare the root growth patterns in two treatment groups respectively undergoing compost and traditional fertilization.

Abstract *Gli apparati radicali delle piante rappresentano una delle principali scorte di carbonio organico totale del pianeta e le loro dinamiche hanno un'influenza diretta sul bilancio dei gas a effetto serra. I rifiuti compostati sono sempre pi utilizzati in agricoltura per ragioni ambientali ed economiche ed il loro comportamento in rapporto ai fertilizzanti tradizionali deve essere analizzato per tutte le componenti della pianta. In questo lavoro viene proposto un modello di regressione quantilica con spline penalizzate per valutare, quantificare e confrontare la crescita delle radici in due gruppi di trattamento riferiti rispettivamente alla fertilizzazione mediante compost e tradizionale.*

Key words: growth curves, quantile regression, penalized splines, noncrossing curves

---

Alessio Pollice
Dipartimento di Scienze Economiche e Metodi Matematici, Università degli Studi di Bari Aldo Moro, Largo Abbazia Santa Scolastica 53, 70124 Bari, Italy, e-mail: alessio.pollice@uniba.it

Vito M.R. Muggeo, Federico Torretta
Dipartimento di Scienze Statistiche e Matematiche "S. Vianelli", Università degli Studi di Palermo, Viale delle Scienze, Parco d'Orleans, 90128 Palermo, Italy e-mail: {vito.muggeo, federico.torretta}@unipa.it

Rocco Bochicchio, Mariana Amato
Dipartimento di Scienze dei Sistemi Colturali, Forestali e dell'Ambiente, Universit degli Studi della Basilicata, Potenza, Italy, e-mail: mariana.amato@unibas.it, bochicchiorocco@gmail.com

1 Introduction

Plant roots are a major pool of total carbon in the planet, and their dynamics are directly relevant to greenhouse gas balance. Composted wastes are increasingly used in agriculture for environmental and economic reasons, but their role as a substitute for traditional fertilizers needs to be evaluated and tested on all plant components. At this aim a three-year experiment (2007-2009) was carried out by Dipartimento di Scienze dei Sistemi Colturali, Forestali e dell'Ambiente, Università degli Studi della Basilicata, Potenza, Italy. Compost application was compared to traditional fertilization with regard to growth of roots of Sorghum bicolor Moench x S. sudanense (Piper) Stapf. in Battipaglia (Sa), Italy. After sowing and treatment of compost or traditional fertilization, plant roots were monitored through sequential images taken with a digital microscope from 4 transparent acrylic access tubes per treatment, buried in the soil at 45° angle up to the soil depth of 60 cm from the surface. A total of 18 images representing depths from 0 to 60 cm were analyzed from all 8 tubes at each date for a total of 3024 images (3 years x 7 dates x 18 depths x 8 tubes). Each image represents an investigated area of 207 mm^2. Image analysis was carried out through a dedicated software (WinRhizoTron MF 2009a, Regent Instruments Canada Inc.) and three root growth measurements were obtained for each image and four tracked root types: *total length*, *total surface area* and *average diameter* for *total*, *alive*, *white* and *dark* roots. The aim of the experiment was to assess root growth across *days after sowing*, by emphasizing differences due to two treatment 'arms': compost vs. traditional fertilization.

The usual modeling framework for growth curves is via mean regression, namely by means of specification of a regression equation for the expected value of the response conditional distributions (Pollice et al., 2013). However there are at least two issues that should be emphasized when modelling throughout mean regression. Firstly, the non-negligible portion of zeroes cannot be ignored and it needs to be modelled properly, for instance via mixture models; secondly, and more importantly, mean regression does not provide a complete picture of data when interest lies in studying growth patters, particularly with strongly heteroscedastic data. In order to analyze root growth we propose an approach based on quantile regression (QR); more specifically we aim at modelling the growth patterns, i.e. the growth curves for different quantiles, with respect to days after sowing by emphasizing possible differences due to the two aforementioned treatment groups. There are several additional advantages in using QR, including robustness to outliers and no need to specify the response distribution, see Koenker (2005) for details.

2 Methods

Our QR modelling framework is based on Muggeo et al. (2013): Let Y be the growth variable, here the total length of roots, $Q_{\tau_k}(Y|t,x_i)$ the τ_kth quantile of Y conditional to covariates x_i and time t . We consider the following quantile regression model

$$Q_{\tau_k}(Y|t,x_i) = x_i^T \beta_{\tau_k} + s_{\tau_k}(t) \tag{1}$$

where β_{τ_k} quantifies the linear effect of p covariates and $s_{\tau_k}(z_i)$ accounts for the growth pattern with respect to days after sowing. Since growth patters are typically nonlinear, $s_{\tau_k}(\cdot)$ is a smooth but unspecified function, and we use B-splines at this goal, namely $s_{\tau_k}(\cdot) = \sum_j^J b_{jk} B_j(\cdot)$.

By setting $\theta_k = (\beta_k^T, b_k^T)^T$ and $w_i = (x_i^T, B_i^T)^T$, the objective function to be minimized can be written as

$$\sum_i \rho_k(y_i - w_i^T \theta_k) + \lambda \sum_{j=1}^{J-d} (\Delta^d b_k)_j^2, \tag{2}$$

where $\rho_k(u) = u(\tau_k - I(u<0))$ is the so-called check function and the penalty term $\lambda \sum_{j=1}^{J-d} (\Delta^d b_k)_j^2$ controls the wigglyness of fitted curve. Δ^d is the order d difference operator whereby d affects the curve behaviour as $\lambda \to \infty$. Notice that the objective function (2) is somewhat unusual as it combines L_1-norm fidelity and L_2-norm penalty; some simulations carried out in Muggeo et al. (2013) have shown that such objective criterium does not perform worse that the more usual 'L_1 fidelity plus L_1 penalty' pair. Finally objective (2) is extended to allow multiple estimation of several quantile curves with noncrossing constraints.

3 Results

Figure 1 displays the distributions of dark roots length across days after sowing by the two treatment groups: the plots emphasize strong asymmetry and the marked zero inflation, making somewhat awkward the usual application of mean regression. We propose a more refined analysis of the distribution of dark roots total lengths, rather summarized by six quantiles than by one expectation.

Notice that the zeroes excess in the continuous response variable corresponds to images containing no roots, and can be understood as roots with no growth. When modeling the expectation of zero-inflated responses common alternatives include mixture modeling (Zuur et al., 2012) and the use of Tweedie distribution models (within the exponential dispersion family framework, see Pollice et al., 2013 and references therein). However QR is robust to the presence of zeroes excess and it does not need to specify any probability distribution for the response. In order to constrain the fitted quantiles to take only nonnegative values, we model the log values and then come back to the original growth scale by exponentiating the fitted values; this is legitimate as quantiles are invariant to monotone transformations. We employ the aforementioned methodology, as implemented in the R package `quantregGrowth`, to fit the dark root length data.

Figure 2 displays the fitted quantiles at percentiles $(0.10, 0.25, 0.50, 0.75, 0.90, 0.95)$. The quantile curves at low percentiles are indistinguishable due to the presence of zero values in both treatment groups; however at higher percentiles the two treat-

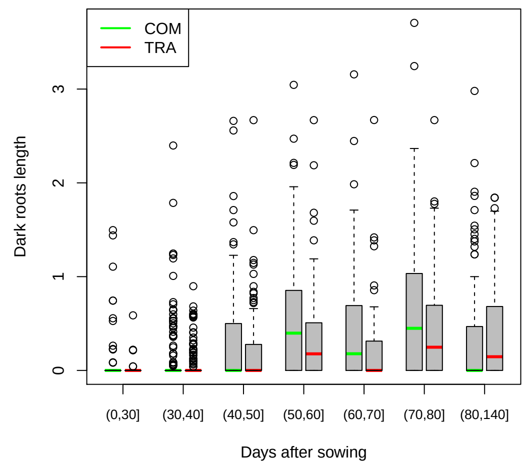

Fig. 1 Boxplots of dark roots total lengths with respect to days after sowing classes by treatment groups.

ments lead to quite different profiles: in the COM group quantile curves are higher and steeper, suggesting better performance, particularly within about 100 days after sowing.

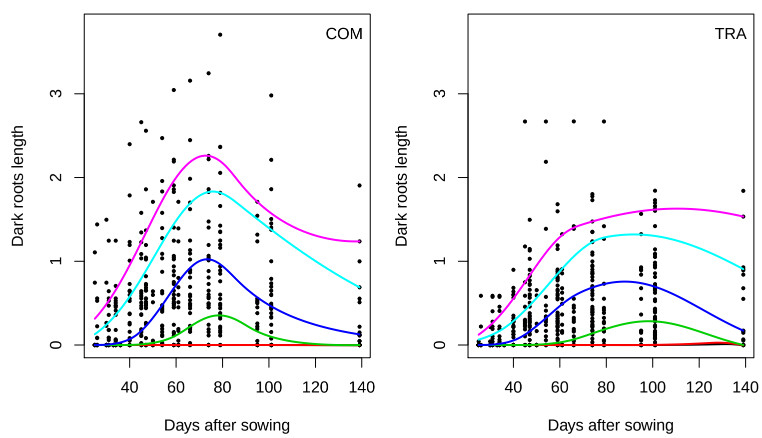

Fig. 2 Fitted regression quantiles at percentiles $(0.10, 0.25, 0.50, 0.75, 0.90, 0.95)$ for dark roots total length in the two treatment groups. Due to the presence of zeroes in both treatments, quantile curves at low percentiles are indistinguishable.

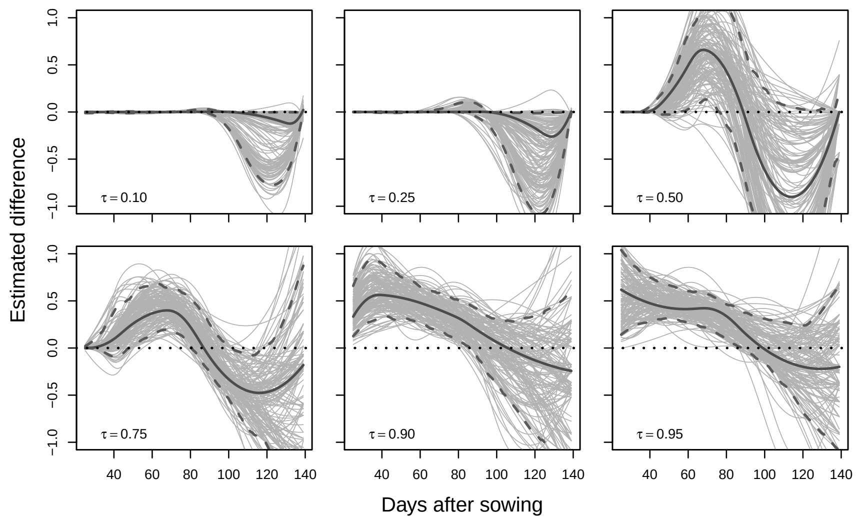

Fig. 3 Displaying difference profiles (3) at different quantile curves. In each panel the light grey lines represent the estimated difference profiles for different bootstrap samples and the dark solid lines indicate the estimated difference in the observed sample. The dashed lines portray the 90% point-wise confidence intervals.

In order to quantify the treatment effect on root growth we consider the difference of estimated quantile curves at each percentile τ_k

$$\hat{s}_k^{\mathrm{COM}}(t) - \hat{s}_k^{\mathrm{TRA}}(t) = \sum_j (\hat{b}_{jk}^{\mathrm{COM}} - \hat{b}_{jk}^{\mathrm{TRA}}) B_j(t). \tag{3}$$

The rationale is plain: if the two treatments do not make any difference the difference profile should settle around zero. Asymptotic theory for penalized quantile regression is far from being well established and it is instead a hot and challenging topic (Koenker, 2005); thus, in order to obtain a sample distribution for difference quantiles, we rely on bootstrap according to the following steps:

1. Resample data independently from the two treatment groups;
2. Fit two noncrossing quantile regressions with *P*-splines using the same basis;
3. Compute the difference quantiles (3) for each percentile τ_k.

By repeating these steps a large number of times we obtain a bootstrap distribution of the difference quantiles.

Figure 3 reports the results for each of six analyzed quantile regressions. At lower percentiles differences are negligible, but for the others ($\tau \geq 0.5$) there are noteworthy differences between the two treatments. At early stage, namely within about 80 days, treatment COM appears to provide better performance growth with respect to the TRA treatment; on the other hand, at late stage the difference between the two treatments is reversed, although uncertainty is much higher making quite hazardous any speculation. However the differential evolution of dark root lengths along time for the two treatments highlights agronomic instances that are worth considering in deeper detail.

References

1. Bochicchio, R.: The contribution of roots to the sustainable management of cerals in a changing climate: experimental evidence from Sorghum (*Sorghum bicolor* (L.) Moench x S. sudanense (Piper) Stapf) under biosolid amendment, and Wheat (Triticum aestivum L.) under simulated climate change. Doctoral dissertation, Università degli Studi della Basilicata (2013)
2. Koenker, R.: Quantile Regression. Cambridge University Press (2005).
3. Min, Y., Agresti, A.: Modeling Nonnegative Data with Clumping at Zero: A Survey. J. Iranian Stat. Soc. **1**, 7–33 (2002)
4. Muggeo, V., Sciandra, M., Tomasello, A., Calvo, S.: Estimating growth charts via nonparametric quantile regression: a practical framework with application in Ecology. Environmental and Ecological Statistics. **20**, 519–531 (2013)
5. Pollice, A., Bochicchio, R., Amato, M.: Nonlinear growth curves of sorghum roots under compost/traditional fertilization. TIES 2013 the 23th Annual Conference of The International Environmetrics Society. Anchorage, Alaska, USA, 10-14 June 2013
6. Zuur, A.F., Saveliev, A.A., Ieno, E.R.: Zero Inflated Models and Generalized Linear Mixed Models with R. Highland Statistics (2012).