



Supplement of
Benchmarking soil multifunctionality

E. R. Jasper Wubs

Correspondence to: E. R. Jasper Wubs (jasper.wubs@hvhl.nl)

The copyright of individual parts of the supplement might differ from the article licence.

Supplementary information

Table S1. The four plant species, and their characteristics, chosen by (Daou & Shipley 2019) as indicator plant in bio-assays to estimate generalized soil fertility.

Characteristics	Scale	<i>Festuca rubra</i> (L.) (creeping red fescue)	<i>Trifolium pratense</i> (L.) (red clover)	<i>Triticum aestivum</i> (L.) (hard red wheat)	<i>Arabidopsis thaliana</i> (L.) Heynh. (Columbia)
Plant family		Poaceae	Fabaceae	Poaceae	Brassicaceae
Growth rate		Slow	Fast	Fast	Fast
Associations		Mycorrhizal	Rhizobium	Mycorrhizal	None
Life span		Perennial	Perennial	Annual	Annual
Optimal day		Long-day	Short-day	Short-day	Long-day
Light	Shade (1) – light (9)	7	7	7-9	6
Air humidity	Dry (1) – wet (9)	5	5	2-4	5
Temperature	Cold (1) – Warm (9)	5	5	5-6	6
Continentality	Marine (1) – Continental (9)	5	5	4-8	5
Soil acidity (pH)	Acid (1) – Basic (9)	4	5	5-8	4
Soil moisture	Dry (1) – Wet (9)	4	4	3-4	3
Texture	Clay (1) – Rocks (9)	3	3	3-5	4
Nutrients	Poor (1) – Rich (9)	5	6	6-8	5
Salinity	Non-tolerant (1) – Very tolerant (9)	1	1	1-2	1
Soil organic matter	Poor (1) – Rich (9)	3	3	2-3	3

Supplementary results

Estimating generalized soil fertility on Dutch soils following Daou & Shipley (2019).

Soils were collected in 30 grassland sites in the Netherlands (Fig. S1) on clayey and sandy soils types. We took 12 soil cores (30 mm diameter, 25 cm deep) in a W-shape pattern on 2x2 m plots sampling three plots per site. Soil samples were homogenized by site and sieved (4 mm mesh) and placed in 7x7x8 cm (l x b x h) pots filled to the brim. For each site 8 pots were filled. In addition, 8 pots were filled with commercial potting soil (potgrond4 'Bloeiende PL. Klei'; Lensli, Bleiswijk, NL), and 8 pots with soil from an extremely sandy site (Zanderij Maarn), to serve as an internal benchmark. The 256 pots were incubated in a greenhouse (21/16 C day:night temperature, photoperiod 16:8h light:dark, relative air humidity >60%) for two weeks to dry and setting in randomized positions. Upon drying the pots were watered with 30 mL for three days. Pots were then seeded with the indicator species *Festuca rubra*, *Trifolium pratense*, *Triticum aestivum*, and *Arabidopsis thaliana*, that were commercially sourced (Cruydt-Hoeck, Nijeberkoop, NL). Based on prior germination trials 10 seeds of *F. rubra*, and 15 seeds for the other species were added to each pot, targeting to reach 4 individuals per pot. Three days after germination all excess plants were removed. For each soil two replicate pots per species were included. Subsequently, pots were watered with 30 mL three days per week (Monday, Wednesday, Friday). Twelve days after germination the water volume was adjusted to 20 mL to prevent waterlogging. Plant shoots were harvested at fixed days for the different species (day 47 and 63 for *F. rubra*, 28 and 44 for *T. pratense*, 20 and 34 for *T. aestivum*, 33 and 44 for *A. thaliana*). The harvest dates varied between species in order to insure sufficient growth had occurred between harvests to be able to measure relative growth rates. During the first harvest day two of the four individuals per pot were randomly selected to be clipped at ground level. We quantified aboveground relative growth rates (RGR, mg g⁻¹ d⁻¹) for each of the four species growing in each soil based on shoot dry mass (72 °C for 72 h).

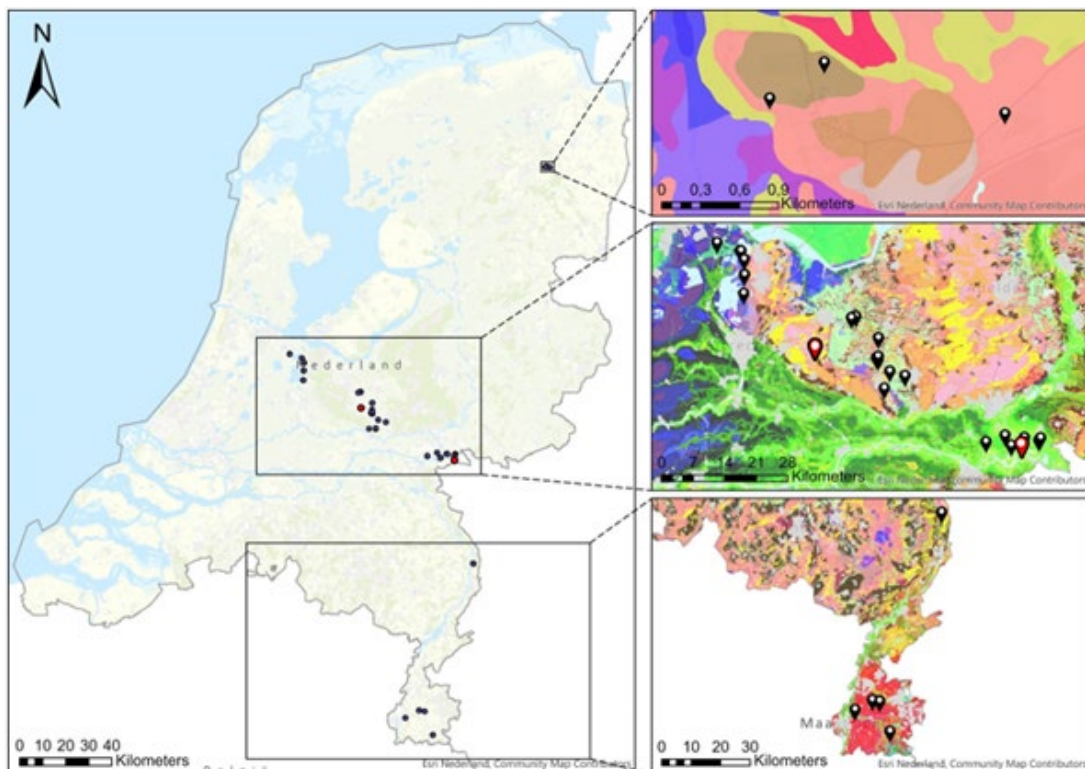


Fig. S1. Sampling locations in the Netherlands. Esri Nederland, Community Map Contributors.

In a companion experiment conducted at the same time we fertilized three soils using half strength Hoagland solution at control, low and high rates. We sampled a sandy and a clayey soil from the 30 sites above (Zanderij Maarn and Millingerwaard, respectively) in the same way as above and made a homogenized mixture of these two extremes as a third soil type. Every week the soils were fertilized. The control received 100% water, the low treatment $\frac{1}{3}$ 0.5 Hoagland and high $\frac{2}{3}$ 0.5 Hoagland solution, mixed with water to make a total of 30 or 20 mL as part of the watering regime.

We estimated RGR values of both experiments jointly using linear mixed-effects models applied separately for each species respectively as $\ln(\text{biomass, mg}) \sim \text{age (fixed effect, days)} + \text{soil (random effect)}$. The slopes (i.e., RGR) were random effects whose values per soil were estimated as Best Linear Unbiased Predictors (BLUPs). The intercept was a fixed parameter because this measures the plant mass of that species at the day of germination ($t = 0$) and this mass will not vary between soils. Finally, we estimated the pure measurement error variances of the RGR values of each species as the sum of the squares of the standard errors associated with the fixed and random components of the slopes. We used the average value per species in the measurement model described next.

The measurement model is specified as a confirmatory factor model (Fig. 1). This model forces the patterns of covariation between the RGR values of the four species across soils to be entirely caused by a single latent variable representing the productive capacity (“generalized” fertility, F_G) of a given soil to which these different species are responding in the same way but with different quantitative values. The true values in the statistical population (RGR'_{ij}) are only estimated with a known measurement error by the actual measures (RGR_{ij}). Thus, the path coefficients between the actual measured values of RGR for each species i growing in soil j in our experiment and their true values in the statistical population (RGR'_{ij}) are fixed to 1, with variances of the pure statistical measurement errors (e_{ij}) fixed to those values estimated in the mixed-model regressions: $RGR_{ij} = 1RGR'_{ij} + e_{ij}$. The RGR'_{ij} values are then linked to the latent F_G as $RGR'_{ij} = \alpha_i F_{G,j} + \epsilon_i$. Units of F_G were defined by fixing the path from it to the true RGR of wheat to 1; thus, a one-unit increase in F_G causes an increase of $1 \text{ mg g}^{-1} \text{ d}^{-1}$ in RGR of wheat. The variance of ϵ_i is variation in the RGR of species i in soil j that is not caused by either generalized soil fertility (F_G) or by statistical measurement error (e_{ij}) and is uncorrelated with that of any other species if the measurement model is correct. We have labelled this “specific fertility” (F_S) because it represents a growth response to a soil that is unique to that species and cannot be generalized to other species. The data were fit to this measurement model using maximum likelihood structural equations modelling (Shipley 2016).

This measurement model has two degrees of freedom. The fit of the observed and predicted patterns of covariation was measured by the maximum likelihood chi-square statistic. If the data fit the model, this χ^2 value will not show significant lack of fit. The maximum likelihood estimates of the latent generalized fertility are obtained via the predict() function.

Results & Discussion

Overall the measurement model fitted the data well ($\chi^2 = 4.257$, d.f. = 2, $p = 0.119$ Fig. S2), and there was a strong relationship between the estimated F_G and the RGRs of the species (Fig. S3), only for *F. rubra* the relationship was weak ($R^2 = 0.29$). Overall there was a strong correlation between both replicates of each soil ($r = 0.87$; Fig. S4), with two outliers, and F_G values were significantly higher under fertilization (Fig S5). The potting soil and extreme sand worked well as high and low benchmarks, giving the lowest and highest F_G values respectively (Fig. 4).

These results indicate that the generalized soil fertility approach work well in the Netherlands as it did in Canada and France. The method shows the expected patterns, is consistent across replicates and sensitive to changes in nutrient availability. I therefore conclude that the method is valid, it measures soil fertility, shows good differentiation among soils, produces an ordered response scale that fits with prior expectations, results are repeatable and objective. With the internal benchmarks I have also calibrated the method for compatible use across laboratories.

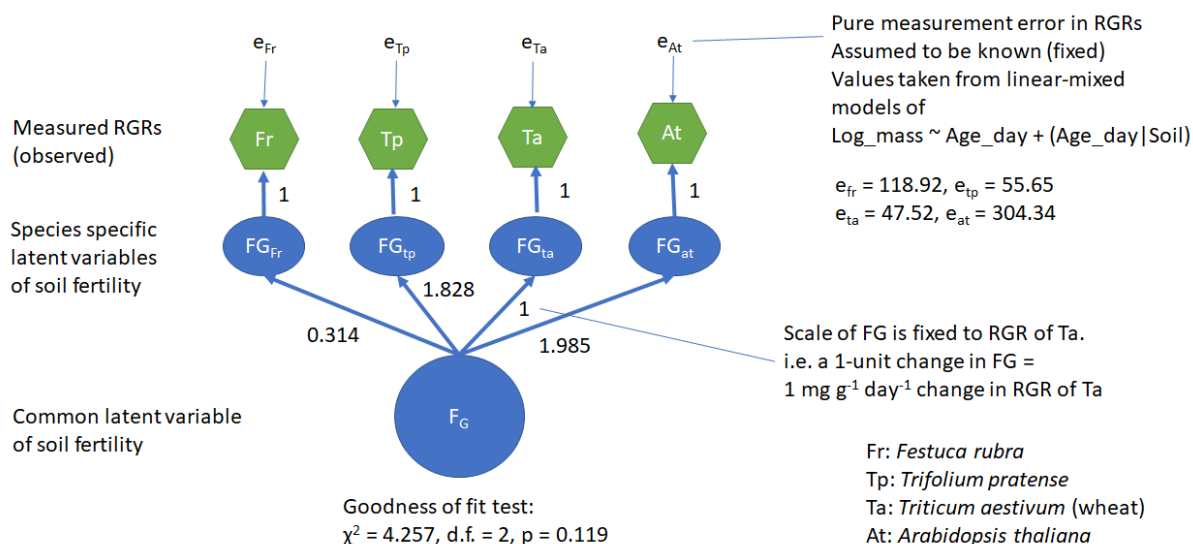


Fig. S2. Summary of the fitted measurement model for generalized soil fertility (F_G).

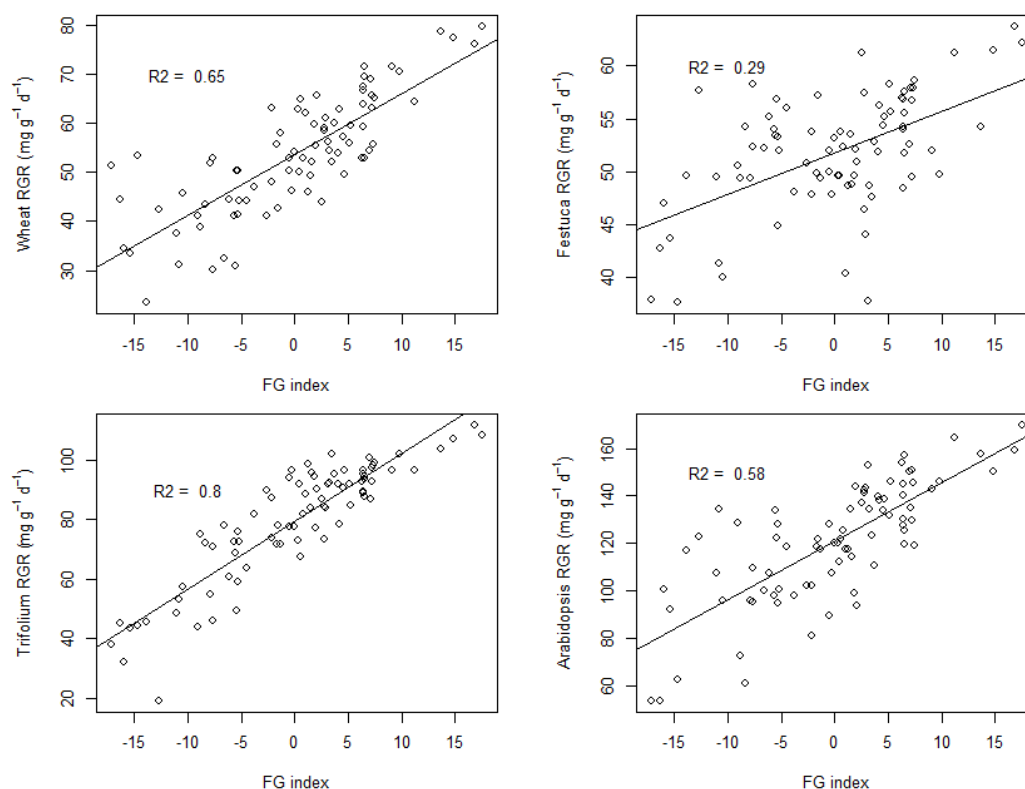


Fig. S3. Relationship between the estimated generalized soil fertility and the relative growth rates of each species.

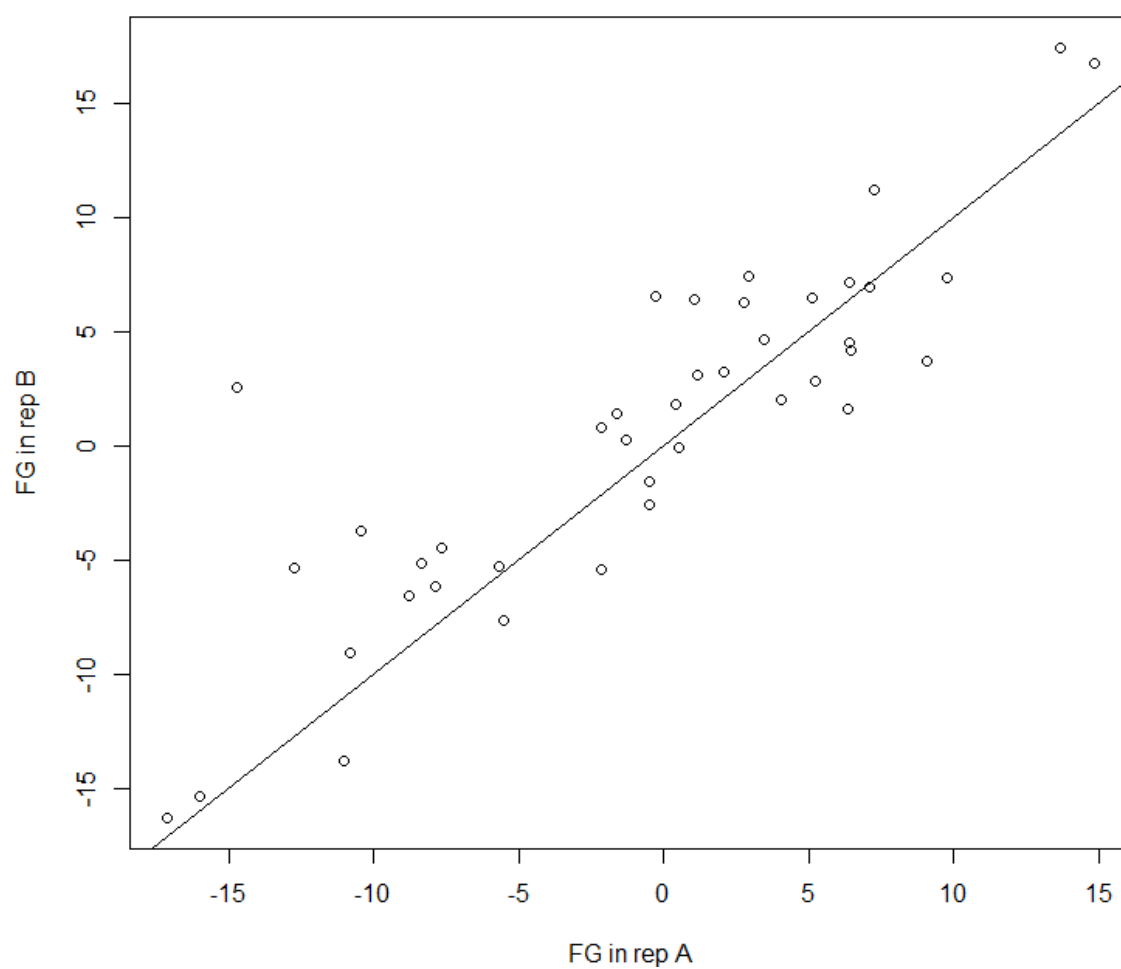


Fig. S4. Correspondence in F_G values among the two replicates per soil. Black line is 1:1 relationship.

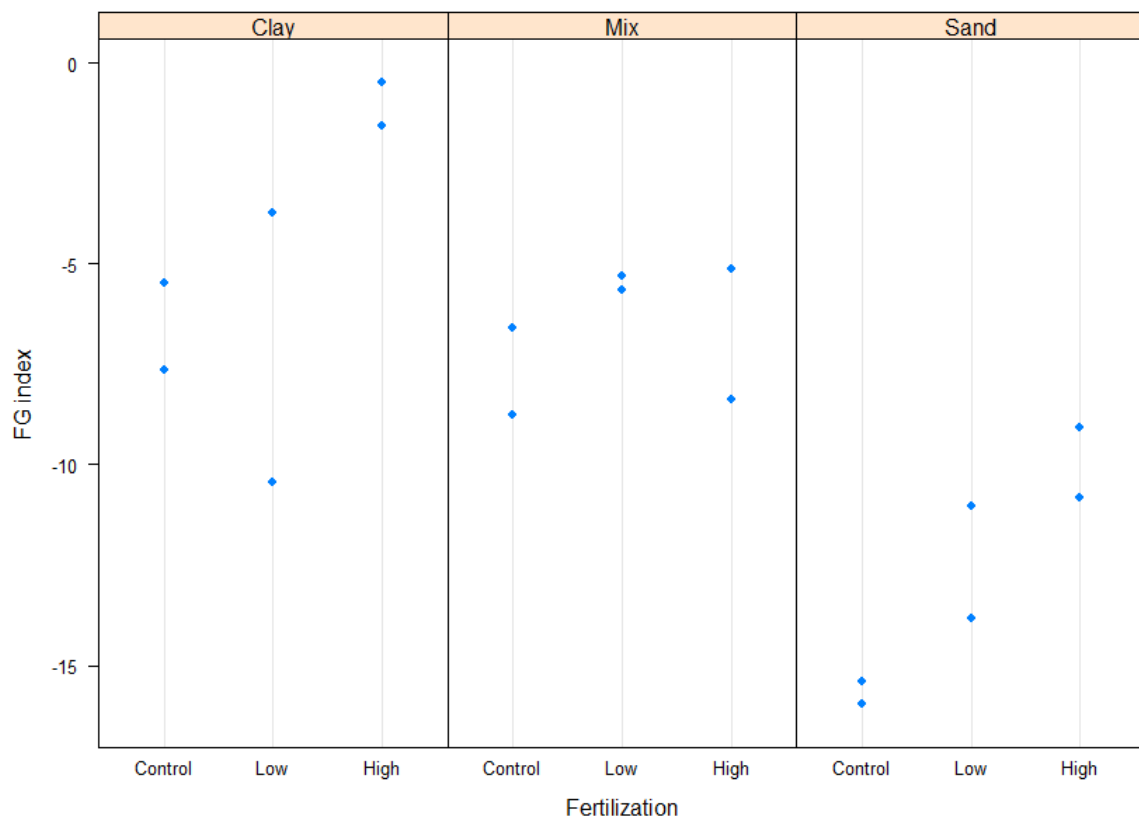


Fig. S5. Effect of fertilization in sand, clay and mixed soils on the F_G index. Estimated F_G was influenced by soil type (Two-way ANOVA, $F_{2,9} = 23.40$; $p < 0.0005$) and fertilization ($F_{2,9} = 5.95$; $p = 0.024$), but there was no interaction ($F_{4,9} = 1.92$; $p = 0.19$). $N = 2$ per treatment.

Supplementary References

- Daou, L. & Shipley, B. (2019). The measurement and quantification of generalized gradients of soil fertility relevant to plant community ecology. *Ecology*, 100, e02549.
- Shipley, B. (2016). *Cause and correlation in biology: a user's guide to path analysis, structural equations, and causal inference with R*. Cambridge University Press, Cambridge, UK.