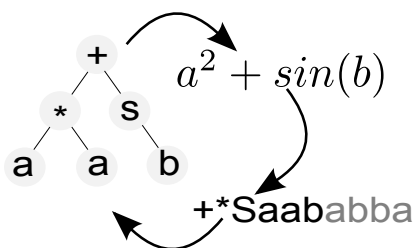# Supplemental Materials: Reverse engineering model structures for soil and ecosystem respiration: the potential of gene expression programming

Figure 1. The Karva language translation of a function containing a and b as variables and $\sin, +$, and $*$ as elementary functions. The mathematical structure can be translated into Karva coded genes and the genes expressed into expression trees that can be easily interpreted by machines. The dark coloured section of the gene (string) represents the active component of the gene that is translatable into mathematical expressions, the light coloured section is inactive at the moment.

GEP models for all log-transformed respirations types time series, before back-transformation.

$$\log\left(R_{eco}\right) = \frac{GPP_s}{T_{-10}} + \log\left(\log\left(T_{-10}\right)\right) \tag{1.1}$$

$$\log\left(R_{above}\right) = 0.1T_{-10} + 0.4\log\left(0.8\sqrt{SWC}\right) \tag{1.2}$$

$$\log\left(R_{soil}\right) = 1.2T_{-10}^{0.4} + 1.3SWC - 3.1 \tag{1.3}$$

$$\log\left(R_{root}\right) = 0.9\frac{1.2GPP_s - 8.1}{T_{-10}} \tag{1.4}$$

$$\log\left(R_{myc}\right) = 1.1\log\left(1.7T_{-10}\right) + 1.2T_{-10}^{SWC} - 7.4 \tag{1.5}$$

$$\log\left(R_{soil_a}\right) = 1.2T_{-10}^{0.5} + 2.5SWC - 4.9 \tag{1.6}$$

$$\log\left(R_{soil_h}\right) = -0.3 + 0.6\frac{1.1GPP_s - 3.6}{T_{-10}} \tag{1.7}$$

Figure 2 in supplemental material illustrates the change in the shape of the PDF estimated for each respiration type after log-transforming. For all time series, the skewness is visibly is reduced.

From Fig. 5 it is worth mentioning the apparent correlation, although weak in terms of R^2 value, of the R_{myc} residuals with GPP_s, even when this was not chosen as a driver, indicating that the relation was not strong enough for an explicit model inclusion but it could show a dependency to a driver for which GPP_s acts as a proxy such as phenology, or substrate availability. Such weak correlations are present as well between R_{soil} and R_{soil_h} residuals and T_{air}.
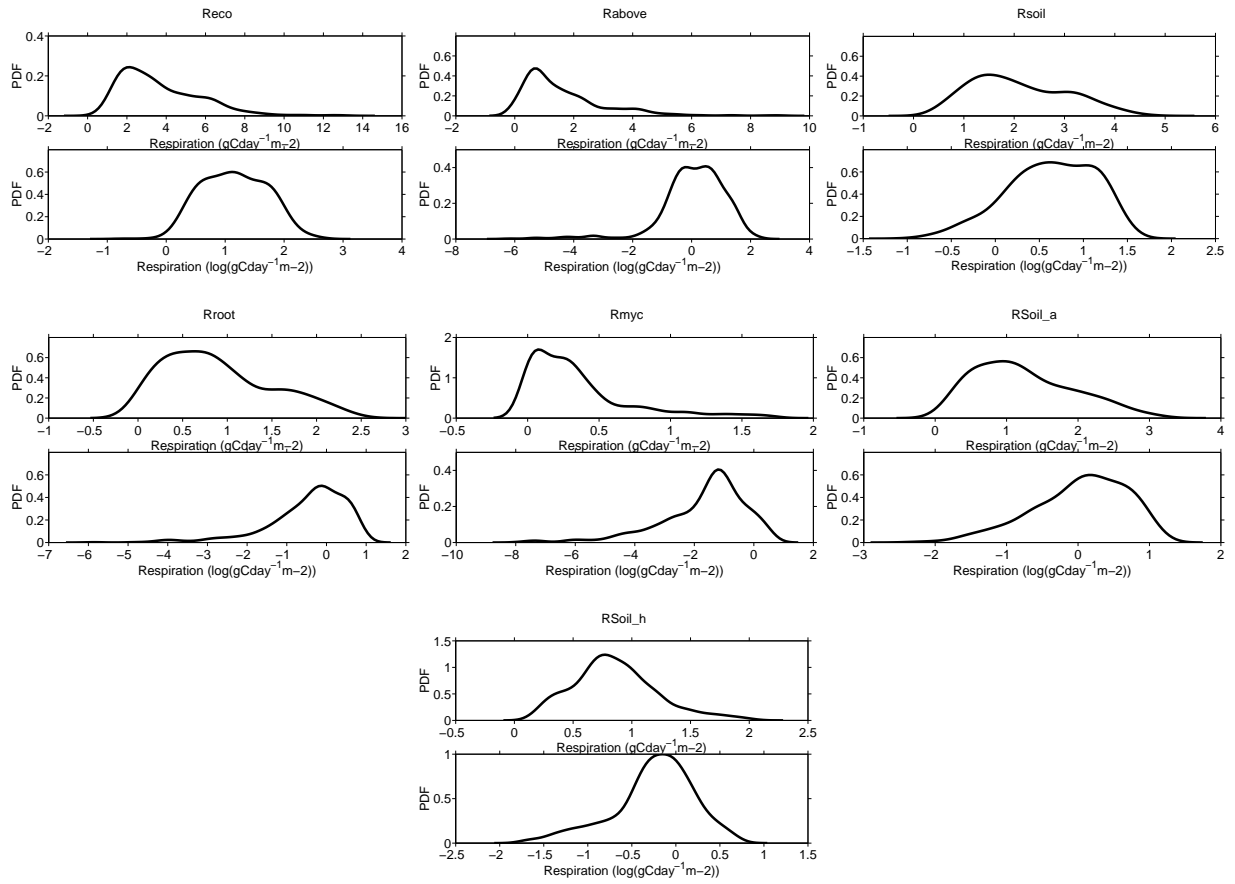
Figure 2. Change in estimated density function of observations before and after log-transforming for all studied respiration types.
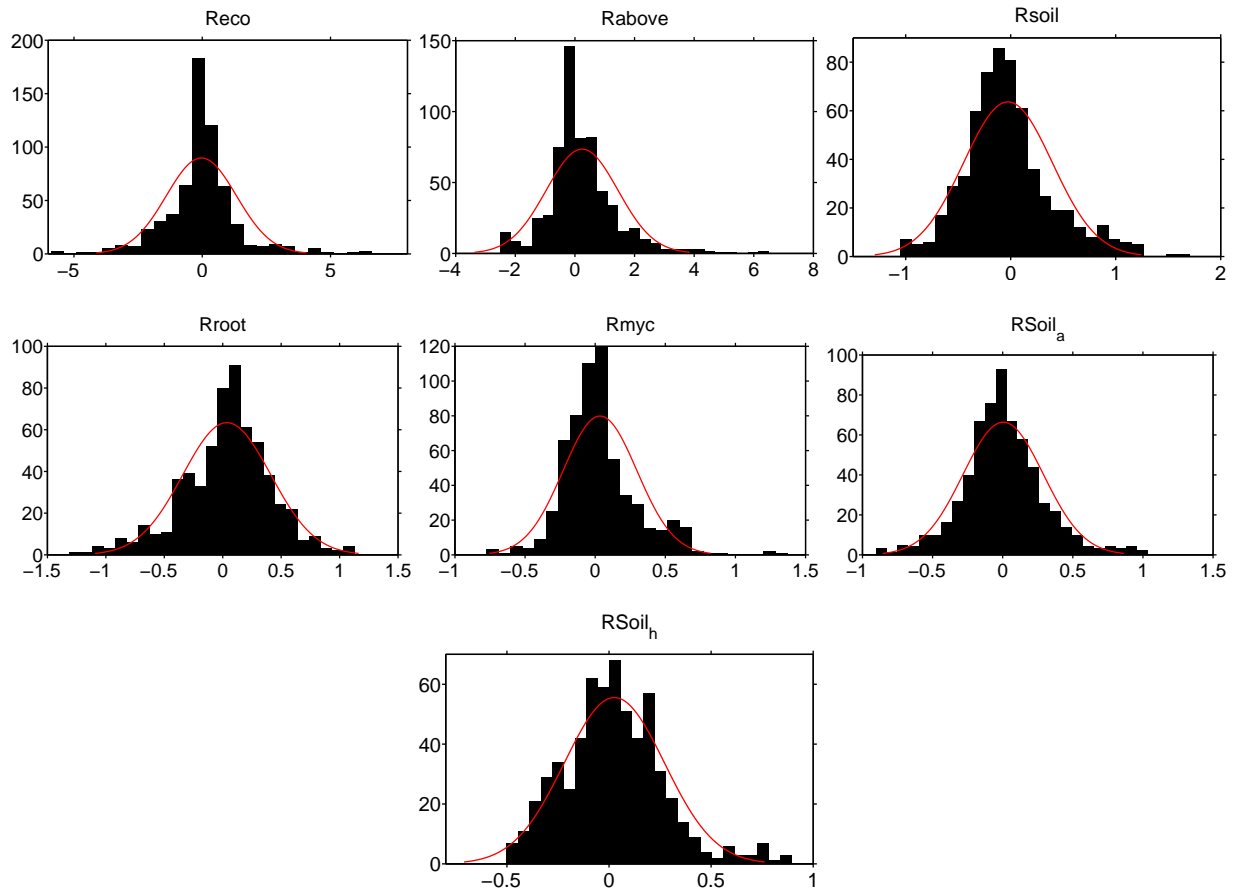
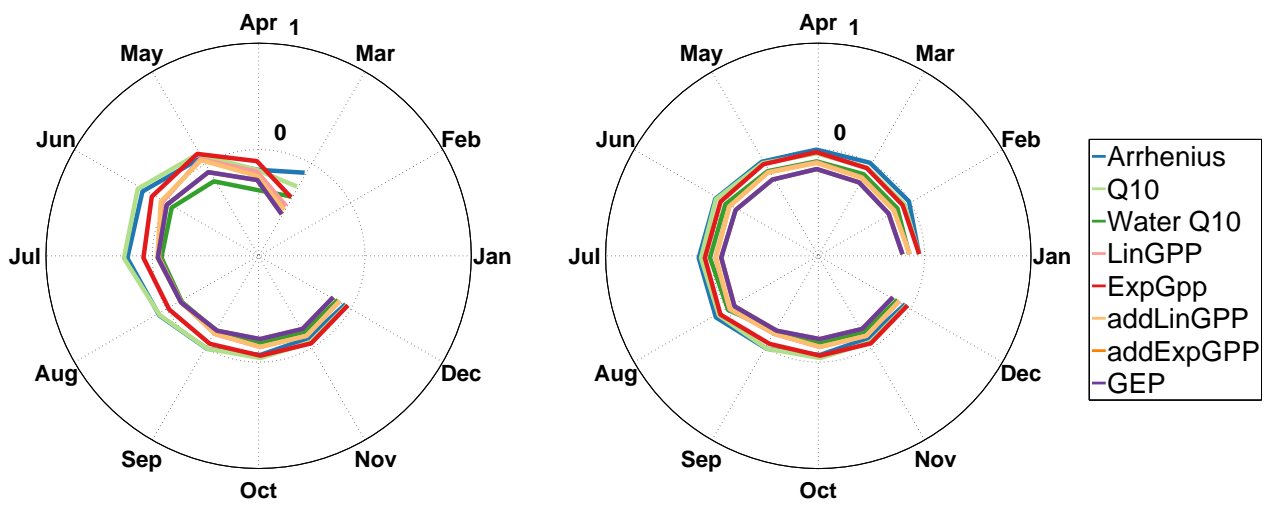**Figure 3. Residuals computed for the GEP models after training on log-transformed data.**
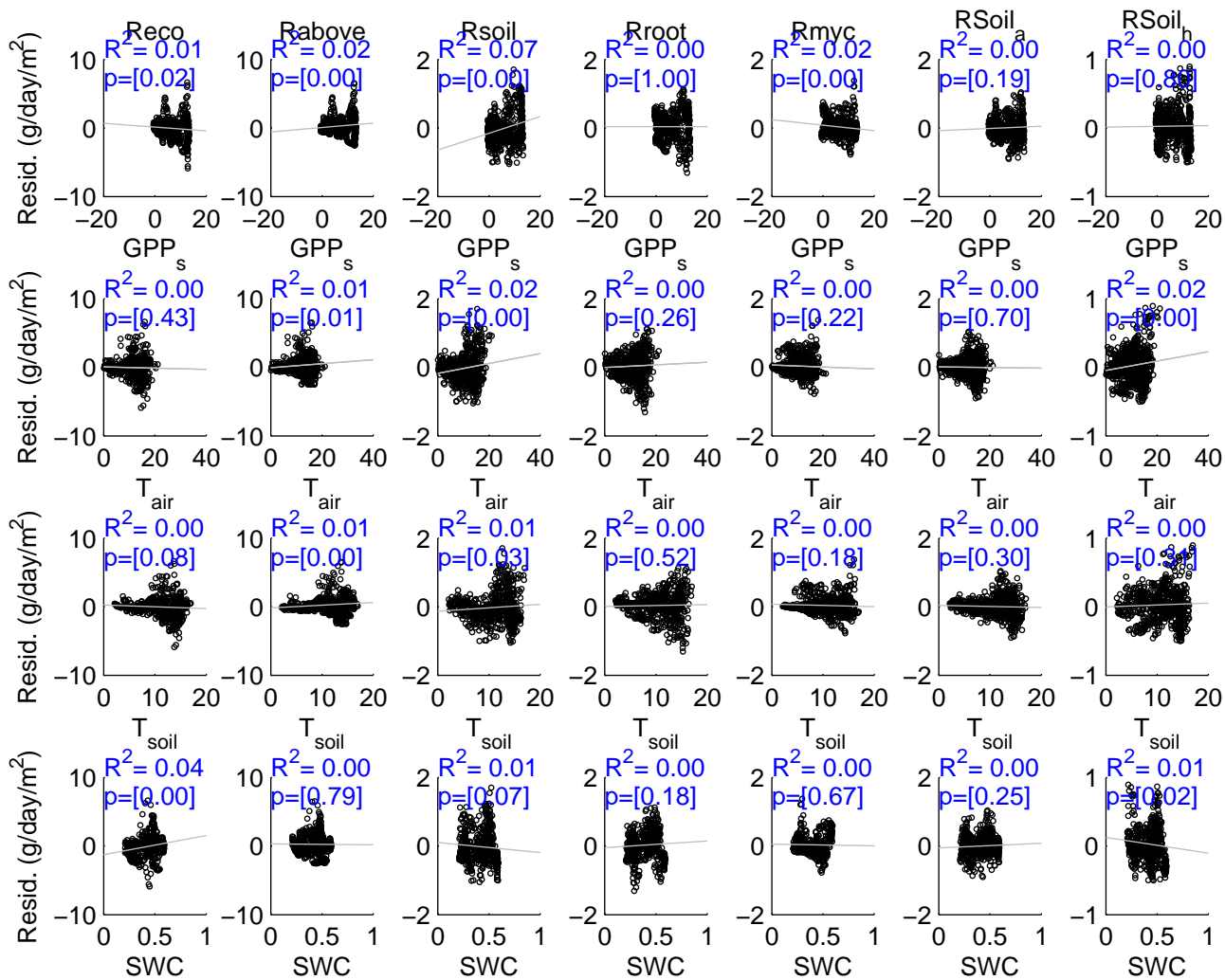
Figure 4. Monthly averaged error values for some literature models for and the GEP generated model for daily soil CO_2 efflux.

Figure 5. Candidate driver linear correlations with GEP model residuals.