



Supplement of

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

Iulia Ilie et al.

Correspondence to: Iulia Ilie (ilie@bgc-jena.mpg.de) and Miguel D. Mahecha (mmahecha@bgc-jena.mpg.de)

The copyright of individual parts of the supplement might differ from the CC BY 3.0 License.

Supplemental Materials:

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

Table S1. Standard error of the MEF at validation values for all MLM for different SNR values when the MEF values are computed against the noisy data.

SNR	GEP	KRR	RF	SVM	ANN
9.82	0.00	0.00	0.02	0.00	0.00
8.18	0.00	0.00	0.02	0.02	0.00
7.01	0.00	0.00	0.02	0.01	0.00
6.14	0.00	0.00	0.02	0.01	0.00
5.45	0.00	0.00	0.02	0.02	0.01
4.46	0.00	0.00	0.02	0.01	0.00
3.27	0.01	0.01	0.02	0.01	0.01
2.73	0.01	0.01	0.02	0.01	0.01
2.34	0.02	0.01	0.02	0.01	0.01
1.96	0.02	0.02	0.02	0.02	0.01
1.75	0.02	0.02	0.02	0.03	0.02
1.40	0.05	0.03	0.02	0.02	0.02
1.23	0.03	0.03	0.02	0.03	0.03
1.09	0.04	0.03	0.03	0.04	0.03
1.00	0.04	0.03	0.02	0.03	0.03

Table S2. Standard error of the MEF at validation values for all MLM for different SNR values when the MEF values are computed against the clear data.

SNR	GEP	KBB	RE	SVM	ANN
SINK				5 1 10	
9.82	3e-07	4e-05	2e-02	4e-03	4e-03
8.18	3e-07	6e-05	2e-02	2e-02	2e-03
7.01	3e-07	4e-05	2e-02	1e-02	2e-03
6.14	2e-06	7e-05	2e-02	2e-02	2e-03
5.45	2e-06	1e-04	2e-02	2e-02	4e-03
4.46	6e-06	1e-04	2e-02	2e-02	2e-03
3.27	9e-06	2e-03	2e-02	1e-02	3e-03
2.73	4e-05	4e-04	2e-02	1e-02	6e-03
2.34	4e-05	6e-04	2e-02	9e-03	3e-03
1.96	8e-05	1e-03	2e-02	1e-02	3e-03
1.75	2e-04	8e-04	1e-02	1e-02	5e-03
1.40	8e-04	1e-03	1e-02	2e-02	5e-03
1.23	1e-04	2e-03	1e-02	2e-02	4e-03
1.09	4e-03	3e-03	1e-02	2e-02	5e-03
1.00	7e-04	3e-03	1e-02	5e-02	6e-03



Figure S1. Change in estimated density function of observations before and after log-transforming for all studied respiration types.

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

5

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

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

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

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

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

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

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



Figure S2. Residuals computed for the GEP models against the log-transformed targets before back-transformation.

Figure 1 in supplemental material illustrates the change in the shape of the PDF estimated for each respiration type after 10 log-transforming. For all time series, the skewness is visibly is reduced.

From Fig. 5 and 6 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 S3. Distributions of the residuals after smear bias correction computed for the GEP models after training on log-transformed data.



Figure S4. Monthly averaged error values for some literature models for and the GEP generated model for daily soil CO_2 efflux in the two studied years. The centre of the plots is -1. The scale of the fluxes is given in $gC/m^2/day$.



Figure S5. Candidate driver linear correlations with residuals computed after bias corrected transformation of the GEP models from runs with settings given in Tab ?? for R_{eco} , R_{above} and R_{soil} . The drivers are on the X axis and the residuals on the Y axis. The candidate driver is given as title of each row and the type of respiration is given as title of the column.



Figure S6. Candidate driver linear correlations with residuals computed after bias corrected transformation of the GEP models from runs with settings given in Tab ?? for R_{root} , R_{myc} , R_{a} and R_{soil_h} . The drivers are on the X axis and the residuals on the Y axis. The candidate driver is given as title of each row and the type of respiration is given as title of the column.