Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions

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

Iulia Ilie[1], Peter Dittrich[2,3], Nuno Carvalhais[1,4], Martin Jung[1], Andreas Heinemeyer[5], Mirco Migliavacca[1], James I.L. Morison[8], Sebastian Sippel[1], Jens-Arne Subke[6], Matthew Wilkinson[8], and Miguel D. Mahecha[1,3,7]

[1]Max Planck Institute for Biogeochemistry, Department Biogeochemical Integration, Hans-Knoell-Str. 10, 07745 Jena, Germany
[2]Bio Systems Analysis Group, Institute of Computer Science, Jena Centre for Bioinformatics and Friedrich Schiller University, 07745 Jena, Germany
[3]Michael Stifel Center Jena for Data-Driven and Simulation Science, 07745 Jena, Germany
[4]CENSE, Departamento de Ciências e Engenharia do Ambiente, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, Caparica, Portugal.
[5]Department of Environment, Stockholm Environment Institute, University of York, York YO105NG, UK
[6]Biological and Environmental Sciences, School of Natural Sciences, University of Stirling, Stirling, UK
[7]German Centre for Integrative Biodiversity Research (iDiv), Deutscher Platz 5e, 04103 Leipzig, Germany
[8]Forest Research, Alice Holt Lodge, Farnham, Surrey, GU10 4LH, UK

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

Abstract.

Accurate modelling of land-atmosphere carbon fluxes is essential for future climate projections. However, the exact responses of carbon cycle processes to climatic drivers often remain uncertain. Presently, knowledge derived from experiments complemented with a steadily evolving body of mechanistic theory provides the main basis for developing the respective models. The strongly increasing availability of measurements may complicate the traditional hypothesis driven path to developing mechanistic models, but it may facilitate new ways of identifying suitable model structures using machine learning as well. Here we explore the potential to derive model formulations automatically from data based on gene expression programming (GEP). GEP automatically (re)combines various mathematical operators to model formulations that are further evolved, eventually identifying the most suitable structures. In contrast to most other machine learning regression techniques, the GEP approach generates models that allow for prediction and possibly for interpretation. Our study is based on two cases: artificially generated data and real observations. Simulations based on artificial data show that GEP is successful in identifying prescribed functions with the prediction capacity of the models comparable to four state-of-the-art machine learning methods (Random Forests, Support Vector Machines, Artificial Neural Networks, and Kernel Ridge Regressions). The case of real observations explores different components of terrestrial respiration at an oak forest in south-east England. We find that GEP retrieved models are often better in prediction than established respiration models. Furthermore, the structure of the GEP models offers new insights to driver selection and interactions. We find previously unconsidered exponential dependencies of respiration on seasonal ecosystem carbon assimilation and water dynamics. However, we also noticed that the GEP models are only partly portable across respiration components; equifinality issues possibly preventing the identification of a "general" terrestrial res-

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions

piration model. Overall, GEP is a promising tool to uncover new model structures for terrestrial ecology in the data rich era, complementing the traditional approach of model building.

Highlights

- We explore if the process of model building for describing ecosystem CO_2 fluxes can be automatized.
- We show that Gene Expression Programming combined with parameter optimization can be a useful algorithm to automatically derive models from ecological time series.
- We propose alternative models for the influence of key environmental variables on various respiratory fluxes CO_2 in an oak forest.
- Conventional ecosystem response functions can be revised by new models identified with GEP.

1 Introduction

One prerequisite to understand and anticipate the global consequences of anthropogenic climate change is an accurate quantitative description of the terrestrial carbon cycle (Bonan, 2008; Heimann and Reichstein, 2008; Luo et al., 2015). However, the description of the mechanisms underlying the total terrestrial efflux of CO_2 (Peng et al., 2014a), often referred to as “terrestrial ecosystem respiration" (R_{eco}), varies across the scientific literature and existing global models. This is partly because R_{eco} does not originate from a single process but is the sum of fluxes from different autotrophic and heterotrophic respiration processes that operate across different temporal and spatial scales and compartments (e.g. soil depths). Hence, it is experimentally very difficult to disentangle the main abiotic and biotic factors driving respiratory processes at the ecosystem level (Trumbore, 2006) and to derive suitable models for the individual respiration processes. In the remaining manuscript we use the term “model" as an equivalent of “response functions" i.e. some analytic description of how environmental drivers influence ecosystem fluxes.

Traditionally, respiration models have been based on some theoretical considerations but largely remain empirical in nature (e.g. Reichstein and Beer, 2008; Gilmanov et al., 2010; Hoffmann et al., 2015). Conventional model building (Fig. 1) is primarily hypothesis driven and capitalizes both on some understanding of the system and reported scaled experiments (Migliavacca et al., 2012; Richardson et al., 2008). Gupta et al. (2012) describe this common paradigm of model development as a four step approach involving *a*) observational, *b*) conceptual, *c*) mathematical and, *d*) computational phases (see also e.g. Bennett et al., 2010; Williams et al., 2009). During the observational phase, the system under scrutiny is monitored and observations are assembled, ideally representing process responses to hypothesized driving variables. Based on these observations, a conceptual model is proposed, which is subsequently guiding the formulations of mathematical representations of the system states and dependencies. The mathematical description then provides the basis for computational models that are used for simulations (Jakeman et al., 2006). Model-data integration may additionally lead to iterative structural revisions or parameter optimizations

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

(Williams et al., 2009). This conventional approach to model development is also characteristic to ecological model building of different kind, including the development of biogeochemical models (Williams et al., 2009).

The fundamental question addressed in this paper is whether models can be constructed more objectively, i.e. reducing the need for human intuition and expert knowledge. Specifically, we explore the possibility of reverse engineering offering an automated alternative to model development for predicting terrestrial carbon fluxes (Fig. 1). In reverse engineering, the work flow is fundamentally different (Bongard and Lipson, 2007): *a*) database set-up phase, *b*) computational phase, *c*) mathematical phase and a *d*) conceptual phase (Gupta et al., 2012). The rationale behind reordering the key phases is to minimize the human influence and perception biases that might shape the formulation of new hypotheses. Reverse engineering is aiming at identifying some mathematical representation of a system that is to a large degree independent from a priori conceptualizations; in the current case, the respiratory response of terrestrial ecosystems to environmental drivers. Reverse engineering leaves the model construction up to an algorithm and is therefore a way to empirically learn from observations with minimal user input. Therefore, reverse engineering is related to machine learning based regression techniques, where various candidate model formulations and specifications are explored in order to minimize the prediction error. The fundamental difference is that reverse engineering typically provides a symbolic regression, that is, the resulting structures are ideally directly readable as mathematical functions (i.e. response functions) and can be interpreted. Further, one can scientifically consider the applicability of the derived structures in other system domains (Ashworth et al., 2012). Here, we focus on the “Gene Expression Programming" (GEP, Ferreira, 2002) reverse engineering approach. GEP is an evolutionary algorithm that evolves mathematical response functions. The structural design of GEP allows for its use in a wide range of empirical modelling problems (Peng et al., 2014b; Khatibi et al., 2013; Traore and Guven, 2013), including (soil) hydrology (Fernando et al., 2009; Hashmi and Shamseldin, 2014). To the best of our knowledge the potential of GEP has not yet been explored for modelling biogeochmical fluxes in terrestrial ecosystems.

We seek to understand as well whether automating model development can provide new insights in understanding the dynamics of terrestrial respiration processes. We investigate if automatically derived model structures differ substantially from models conventionally used in the study of R_{eco} and its components or, if they are consistent with established theory. We base our study on data from a long-term monitoring experiment of R_{eco} components i.e. above ground respiration, root respiration, mycorrhiza respiration, soil autotrophic, and soil heterotrophic respiration. The monitoring was done separately but in a time-synchronized way over two years and is described in detail by Heinemeyer et al. (2012). The separation of R_{eco} into its components also allowed to test the portability of individual model structures across different respiration components. In this sense, we investigate whether a generic “respiration" response can be derived, or if specific formulations for a range of respiration components are required.

Our study is structured as follows: First we introduce the GEP methodology and explore its performance for symbolic regression type of problems using an artificial experiment under varying degrees of noise contamination designed to resemble R_{eco}. Second, we apply GEP to model the various respiration observations provided by Heinemeyer et al. (2012). This is an exceptional data record, as typically only integrated measurements of either soil (Heinemeyer et al., 2012) or ecosystem respiration are continuously and regularly measured, and the components measured offer a perfect test case for the GEP

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

methodology. For both the artificial experiment and real world observations, we systematically confront the prediction error of GEP with other state-of-the-art machine learning regression approaches. In addition, we adjust the modelling approach such that the objective function (or fitness function) accounts not only for absolute or relative error, but also reduces structure in the residuals. The discussion focuses on the comparison of the various GEP derived models, their equifinality, and performance compared to widely used literature models. Conclusions and outlook focus on the potential of the discussed GEP approach for the further applications in this branch of research.

2 Method

We rely on the GEP method (Ferreira, 2002) which automatically derives model structures from a set of given observations. As the models we want to obtain are mathematical structures, their extraction can be achieved by solving a symbolic regression (Kotanchek et al., 2013) type of problem. The general GEP approach in solving symbolic regressions is presented in the following section and is illustrated in Fig. 2.

2.1 Gene Expression Programming, GEP

In GEP the structure building process starts with a set of possible explanatory variables and a set of elementary functions that are given as input (e.g. $\sin, +, -, \times$). The variables and functions are subsequently mapped to a set of characters (e.g. $a, b, c, +, -$) according to an internal coding language called “Karva language"(Ferreira, 2006). The mapping process generates sets of strings that represent the basis for a manipulative evolution of operations. The mapped letters are randomly combined into fixed length strings called “genes" (Fig. 1 of suppl.). The use of fixed length linear strings for representing expression trees (ET) of varied shapes and sizes, simplifies the evolutionary process of GEP (Ferreira, 2002).

A set number of genes is aggregated into a chromosome with the help of a binary function (e.g. $+, -, \times$). The resulting chromosomes can be translated into expression trees encoding mathematical structures. The chromosomes are the objects involved in the evolution (extraction) of the final solution. The total sum of combinations of functions and variables, i.e. mathematical structures formed during an evolution time step make up a generation. The maximum number of generations needed to reach a solution is often used as a stopping criterion.

One deciding factors in constructing a model during an evolutionary algorithm is the selection process. Since the chromosomes can be translated into mathematical expressions that can be evaluated, “fitness values" (i.e. measures of model performance) are assigned to each chromosome during each generation. Depending on the fitness function, the fitness values can be crucial as they give a measure of how distant each of the current structure based predictions is from the observations. The fitness measures are assigned to the chromosomes by means of a fitness function that is optimized, usually by minimizing prediction error. Based on the fitness values, the chromosomes in a generation are sorted and a selection for a new time step generation is made.

The best chromosome, which is also replicated once for the subsequent generation, and the remaining selected chromosomes can go through a set of genetic manipulations that produce new individuals with new associated fitness measures. The manip-

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

ulation rate is an important hyper-parameter (Tab. 1) in GEP (as in other genetic programming approaches) since it is decisive in the amount of new individuals created from a generation to the other. For example, if the mutation rate (one of genetic variation operators) is too large, it can become disruptive and lead to loss of the information acquired along the previous evolutionary time steps and reduce the convergence of the algorithm. Conversely, if the rate is too low, one may not identify new relevant model structures in due time. The process of selection and genetic manipulation is repeated until a stopping criterion is reached (i.e. best fitness achieved, maximum number of unimproved generations is reached, etc.), and a solution in the form of a mathematical structure is returned.

The current implementation of the GEP approach does not contain an explicit population diversity management component. However in order to reduce stochastic bias and avoid getting stuck in local optima and produce over-fit results, we chose a practical approach of multi-start (multiple runs with the same settings) as proposed by Ferreira (2006).

The GEP method presented in this paper was implemented by the first author in the C++ language and is available on demand. All the experiments reported in this work were executed on a cluster containing 51 nodes, running SuSE SLES 11 SP1 and StorNEXT (global file system running on the IO nodes). In summary 868 CPU cores, 14.5 TB RAM, 1.2 PB file space. All the nodes are attached via GB LAN and OPENLAVA 3.1 is used as queueing system.

2.2 Fitness measure

In our study, the fitness measure is reported in terms of Nash–Sutcliffe modelling efficiency (MEF) coefficient (Nash and Sutcliffe, 1970; Bennett et al., 2010) which is often used in the context of quantifying the performance of terrestrial biosphere models (Mitchell et al., 2009; Migliavacca et al., 2015). The MEF is computed as

$$\mathrm{MEF} = 1 - \frac{\sum_{i=1}^{n}(o_i - p_i)^2}{\sum_{i=1}^{n}(o_i - \bar{o})^2} \tag{2.1}$$

where o_i is the observed value at step i and p_i is the predicted value at step i and $\bar{o}$ is the mean of observed values. MEF values range between $-\infty$ and 1, where an MEF value of 1 corresponds to the case where the predicted and observed values are identical. A negative MEF value means that the predictions are worse than the mean of the observations in recreating the observed signal. MEF=0 indicates that the models prediction are as good as a prediction by $\bar{o}$.

Although the MEF metric offers a straightforward interpretation, it does not take number of parameters of the models into account. In real-world applications, it might be desirable to derive models with lower number of parameters if those are not (much) worse in terms of prediction capacity than models with higher number of free terms. Thus, we include in our cost (fitness) function a normalized term related to number of parameters (ratio of current number of parameters to maxim number of possible parameters given the GEP run settings).

Moreover, any systematic signature in the model residuals (the differences of model predictions and observations) needs to be reduced as the latter should ideally only represent uncorrelated noise. To meet this criterion, we complement the fitness function with a term related to the information content (entropy) in the residual time series, i.e. derived from information-theoretic considerations. Entropy values would thus be maximized for data without structure (i.e. white noise), and lower

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions

entropy values would be obtained for structured data, e.g. correlated stochastic or deterministic processes (Rosso et al., 2007) . The information content in a time series is typically quantified by the Shannon Entropy (SE, C. E. Shannon (1948)) , i.e. a term of the form

$$\mathrm{S}[P] \;=\; -\sum_{i=1}^{N} p_i \ln [p_i] \;. \tag{2.2}$$

Here, $P = \{p_i; i = 1, \ldots, N\}$ denotes a probability distribution with $\sum_{i=1}^{N} p_i = 1$ and N possible states. To calculate Shannon's entropy measure from a time series, the series thus has to be adequately partitioned into a suitable probability distribution. As our aim is to minimize structure in the residuals, the temporal order becomes important . Here, we extract ordinal patterns from the time series and derive a (discrete) probability distribution through counting the occurrence probabilities of each pattern, following Bandt and Pompe (2002). This approach is fully based on the temporal dynamics in the residuals (i.e. the order within the time series) and largely non-parametric, as only the window length has to be specified. This parameter is set to $n_{demb} = 4$ throughout the paper, following previous work on ecosystem gross primary productivity dynamics (Sippel et al., 2016).

The final normalized form the fitness function further used in our work is:

$$\mathrm{CEM} = \sqrt{(1-\mathrm{MEF})^2 + (\frac{P}{P_{max}})^2 + (1-\mathrm{SE})^2} \tag{2.3}$$

$$P_{max} = gN \times h \tag{2.4}$$

where, CEM will stand from here on for complexity corrected efficiency in modelling, *P* is the number of parameters present in a model structure, P_{max} is the maximum numbers of parameters possible for each individual from a GEP run set-up, *gN* is the number of genes in a chromosome and *h* is the length of a gene (Fig. 1 of suppl.).

2.3 Parameter optimization

The GEP algorithm does not have a specific treatment of constants in the building of model formulations but mutations can change both the model structure and constants. However, the scaling of constant values (model parameters) might be a decisive factor in adequately determining the fitness of a formulation. Without this, a model structure might be discarded regardless of potentially being a very powerful candidate. Furthermore, model parameters are often very informative regarding a system's sensitivity to some modifications of the drivers. These aspects have led to the addition of a final parameter optimization step at the end of each GEP run.

In order to obtain an optimal set of parameters for the GEP extracted model structures, an approach that would be applicable in a large set of generated search spaces was necessary. Here we use the "Covariance Matrix Adaptation Evolution Strategy"(CMA-ES) (Hansen et al., 2003) for optimization. The CMA-ES is a stochastic optimization algorithm that seeks to minimize a fitness function by estimating and adapting a covariance matrix according to a sampling from a multivariate normal

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions

distribution (Beyer and Schwefel, 2002; Auger and Hansen, 2005). According to Hansen (2006), one of the main arguments in favour of the CMA-ES approach is that it has shown good results even in the case of ill-posed problems (Kabanikhin, 2008), which may very well be the case for some of the GEP structures that are automatically generated.

The CMA-ES version used for the final step of optimization is the Hansen Python implementation available at "https://pypi.python.org/py

3 Experimental design

For exploring the possibility of using GEP in developing relevant model structures for describing the terrestrial carbon fluxes, two case studies were designed: Firstly, an experiment based on artificially generated data to better understand and present the general properties and capacities of GEP. Secondly, we explored the use of GEP on real measurements of various respiratory flux components monitored continuously over two years in an oak forest (Heinemeyer et al., 2011).

3.1 Artificial experiments

These experiments were designed to explore whether our implementation of the GEP method is suitable for symbolic regression type of problems, and how robust/vulnerable it is across various signal to noise ratios. We explored a set of functions with increasing levels of non-linearity to generate data points.

$$f(x_1) = 2x_1 + 1 \tag{3.1}$$

$$f(x_1) = x_1^2 + 3x_1 + 5 \tag{3.2}$$

$$f(x_1) = e^{x_1} + 1 \tag{3.3}$$

$$f(x_1) = e^{-x_1} - x_1 \tag{3.4}$$

$$f(x_1) = x_1^2 - 4\sin(x_1) \tag{3.5}$$

$$f(x_1) = x_1^3 + 6x_1^2 + 11x_1 - 6 \tag{3.6}$$

$$f(x_1, x_2) = x_2 x_1 \tag{3.7}$$

$$f(x_1, x_2) = x_2 x_1 - 3\cos(x_1) \tag{3.8}$$

$$f(x_1, x_2) = 2x_1^2 + 3x_2^2 \tag{3.9}$$

$$f(x_1, x_2, x_3) = 2x_1^2 + 3x_2^2 + 2\sin(x_3) \tag{3.10}$$

2000 data points were randomly generated with $x_1 \in [1, 20]; x_2 \in [1, 5]; x_3 \in [1, 100]$ and each functional values were computed based on these. Out of the 2000 data points, 1000 data points were used for training, while 1000 data points were reserved for validation. The GEP settings used for each of the 20 runs are given in Table 1.

In order to investigate the response of the GEP approach to noise contaminated data, we simulated Gaussian noise that scales with signal amplitude as often observed in the case of terrestrial ecosystem fluxes (Lasslop et al., 2012). The signal-to-noise ratio (SNR, measured as ratios of standard deviations) was varied between 10 and 1 in six steps.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions

For each of these functions and SNR levels, 20 GEP runs were performed and the retrieved model structure with the highest MEF values at the validation points was chosen. If a returned structure was identical to the originally prescribed function or if $(1 - \mathrm{MEF}) \leq 1e05$ at validation, the retrieval of the original structure was considered to be a success.

As the choice of fitness function was crucial for the construction of structures in a GEP type of approach, we also investigated in one experiment the effects of minimizing the CEM values (eq. 2.3) as opposed to using MEF (eq. 2.1) as acceptance criteria.

3.1.1 Alternative Machine Learning Methods

The prediction performance of the best GEP derived models was compared with the prediction performance of four commonly used state-of-the-art machine learning methods (MLM), i.e Artificial Neural Networks, ANN, (Yegnanarayana, 2009), Support vector Machines, SVM (Hearst, 1998), Random Forests, RF (Breiman, 2001) and Kernel Ridge Regressions, KRR (Hoerl and Kennard, 1970). The toolboxes used for generating the predictions of the ANN and KRR methods are described by Tramontana et al. (2016), the predictions of the SVM were obtained by using the LIBSVM package (Chang and Lin, 2011) and the RF predictions were given by the Matlab statistics toolbox implementation. All the runs were performed with default settings.

3.2 Measured ecosystem CO_2 fluxes

In the second experiment we tried to reverse engineer model structures R_{eco} and its components based on measured data. Specifically, we explored GEP derived model structures for various components of terrestrial ecosystem respiration fluxes collected in an 80 year old deciduous oak plantation in the Alice Holt forest in SE England as described in (Heinemeyer et al., 2012; Wilkinson et al., 2012).

3.2.1 Alice Holt in-situ data

The particular strength of the Alice Holt data set is that component fluxes of R_{eco} were measured separately. R_{eco} and the total influx of CO_2 to the ecosystem as mediated via photosynthesis (gross primary production, GPP) were estimated from eddy covariance measurements of the forest net CO_2 exchange (NEE, Eq. 3.11), and the various soil respiration components were measured separately for two years with hourly time resolution for total soil respiration (R_{soil}), root respiration (R_{root}), mycorrhiza respiration (R_{myc}), soil autotrophic respiration (R_{soil_a}) and soil heterotrophic respiration (R_{soil_h}) and above ground respiration (R_{above}) estimated by difference (Eq. 3.12). Additionally, we have access to derived measurements of GPP, as well as direct measurements of soil moisture (SWC), air temperature, surface temperature, and soil temperature taken at 2, 10 and 20 cm depth.

R_{eco} and GPP were obtained from a micro-meteorological measurement tower at the same site that reports half hourly integrals of net ecosystem exchange (NEE) with the eddy covariance (EC) methodology (Moncrieff et al., 1997). The Reichstein et al. (2005) procedure was used for gap-filling and separation of NEE into GPP and R_{eco}. Given that R_{soil} is a fraction of R_{eco}, above ground respiration can calculated as the difference between R_{eco} and R_{soil}. For an in-depth description of other site conditions and measurements see Heinemeyer et al. (2012).

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

A multiplexed chamber system was used for measuring R_{soil} , using a continuous sampling method at fixed locations. In order to partition the R_{soil} flux into its components, mesh-bags that are not penetrable by roots, but allow for mycorrhizal hyphae development were installed. Deep steel collars were applied to stop both root and mycorrhizae in-growth. As a result, R_{root} is given by the difference of R_{soil} and the respiration recorded in the mesh bag chambers, R_{myc} is given by subtracting the steel collar flux from the mesh bag chamber flux, and the R_{soil_h} is given by the CO_2 efflux at the steel collar chambers and (R_{soil_a} is the sum of R_{myc} and R_{root} (Eq. 3.13 and 3.14) .

$$R_{eco} = \mathrm{NEE} + GPP \tag{3.11}$$

$$R_{above} = R_{eco} - R_{soil} \tag{3.12}$$

$$R_{soil_a} = R_{root} + R_{myc} \tag{3.13}$$

$$R_{soil} = R_{soil_a} + R_{soil_h} \tag{3.14}$$

The computation of R_{above} as difference between R_{eco} and R_{soil} might be highly uncertain because of the different techniques used to compute the two respiration components, the completely different footprints, and the typical high flux underestimation and low flux overestimation of R_{eco} from EC (Wehr et al., 2016). The limitations of the separation of R_{eco} into its components and the uncertainty of the estimates are further discussed by Heinemeyer et al. (2011), Heinemeyer et al. (2012) and Wilkinson et al. (2012).

3.2.2 Data processing

We used the following candidate driver variables: soil volumetric moisture measurements, air temperature (from micrometeorological station), and temperatures at different soil depths, and GPP. A number of recent studies have shown a tight linkage between GPP and R_{soil}, reflecting dynamics of respiratory substrate supply to roots and mycorrhizal fungi from recently assimilated C in plants. (Moyano et al., 2008; Mahecha et al., 2010; Migliavacca et al., 2011, amongst others). We use GPP obtained from EC measurements at the site, but acknowledge the conceptual problem that R_{eco} and GPP were derived from the same observations of NEE. In order to minimize the potential spurious correlation between R_{eco} and GPP as well as redundancy of possible GPP influence with the meteorological drivers, we considered low-frequency variability of GPP only (i.e. low-pass filtered modes of GPP which corresponds to variability beyond a 60 days periodicity only, see Mahecha et al., 2010). “Singular Spectrum Analysis" as described and implemented by Buttlar et al. (2014) was used to obtain a smooth GPP signal. The seasonal cycle was extracted with the SSA method as the assumption is that GPP affects mainly the seasonality of the respiration while the variability at the high frequency is assumed to be more related to meteorological drivers (e.g. temperature, Mahecha et al., 2010).

To reduce the skewness and the search space that the GEP evolution would have to cover in order to construct valuable solutions (Keene, 1995), we log-transformed the seven target respiration data sets (see Figure 2 in supplemental material) and applied a back-transformation when reporting the respective model structures. The time series used for the candidate drivers remain unchanged.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

3.2.3 GEP set-up

For each combination of respiration target and possible drivers, 50 subsets of 500 target time steps each were randomly selected and used for the training of GEP models using the settings found in Table 1. The 50 subsets of the remaining 113 time steps are used for cross-validation and the model with the lowest average validation CEM value is finally selected for each respiration type.

We were particularly interested in determining the general character of each extracted model with respect to the different respiration fractions. We therefore re-optimized the parameters of all extracted model structures when applying one extracted model as the candidate function for a different respiration term. For example, the model formulation extracted for R_{eco} is re-calibrated for all the other types of respiration, creating six parameter sets (one for each respiratory flux) per equation. To cross-validate parameter sets, we computed performances for each train–validation data set pair and report averaged MEF values.

As in the artificial example, we compared the returned GEP solutions predictions performance with that of other common MLM such as SVN, KRR, ANNs, and RF. All methods were used for generating 50 subsets of 113 prediction values, after training on the 50 subsets of 500 time steps of observations presented in the start of section 3.2.3. Then, a mean MEF value was computed for all methods for all respiration components and the best mean MEF values were reported and compared with those of the GEP extracted models. The comparison is done in terms of MEF as number of model parameters were not available and CEM could not be computed.

3.2.4 GEP in the context of other known ecological models: Real observational data

A comparison was done between the GEP built models and some common literature respiration models with different structures and driving variables that were also optimized using CMA-ES. The optimization was performed for each respiration dataset and its candidate drivers and parameters (Table 2). The structures and prediction performances of the GEP models were then compared with those of the optimized literature models.

4 Results

4.1 Artificial experiments

In the first artificial experiment the GEP approach is used to verify if it can reconstruct prescribed functions. Following the training of the 20 independent GEP runs, the initial functions were successfully reconstructed for all 10 equations defined in section 3.1.

MEF values for the GEP extracted models and for the predictions generated by ANN, RF, KRR and SVM are illustrated in Fig. 3a. These MEF values were obtained through cross validation against independent, yet equally noise contaminated data points (the SNR values are given on the x axis in reverse order for visualizing the increase in noise levels). There is a clear

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

pattern of decreasing MEFs with increasing noise contamination. This was expected, as none of the methods should fit the noise added to the signal.

Figure 3b shows MEF values equivalent to fig. 3a, but applied to noise-free data points of the validation set, in order to compare GEP outputs to the “true" structure underlying the artificial data set. In this set-up, the MEF values remained relatively constant across SNR values above 2. When SNR level was set to 1, predictions for all investigated machine learning methods, except for GEP predictions, show decreased fitness, with MEF values decreasing to a minimum of 0.8.

Figure 3c compares the two validation approaches described above - SNR is now represented by a colour code. This figure suggests that we may not expect MEF values of $\gtrapprox$0.9 under real scenarios (where no noise-free validation is possible).

In order to verify the effects of changing the fitness function from MEF to CEM, we compare the distributions of MEF values for all runs for all studied SNR. Figure 4 exemplifies outputs for equation 3.10; panel a shows a drop of prediction capacity of the GEP models with noise increase for all types of fitness functions when compared with noise-infused data. This contrasts the reduced MEF assessed against original data, where a slight drop in MEF with noise increase for the MEF optimization structures was seen, and where the CEM optimized structures show stability in MEF with noise. The new CEM leads to a reduced number of returned parameters compared to MEF (Fig.4c), as well.

4.2 Measured ecosystem CO_2 fluxes

Applying GEP on the Alice Holt data set yielded a series of model structures for each respiration type. The returned model structures are illustrated in equations 4.1-4.7.

$$R_{eco} = \log(T_{-10}) \times e^{(\frac{GPP_s}{T_{-10}})} \tag{4.1}$$

$$R_{above} = 0.9SWC^{0.2} \times e^{(0.1GPP_s)} \tag{4.2}$$

$$R_{soil} = e^{(1.2T_{-10}^{0.4}+1.3SWC-3.1)} \tag{4.3}$$

$$R_{root} = e^{(0.9\frac{1.2GPP_s-8.1}{T_{-10}})} \tag{4.4}$$

$$R_{myc} = 1.8T_{-10} \times e^{(1.2T_{-10}^{SWC}-7.4)} \tag{4.5}$$

$$R_{soil_a} = e^{(1.2T_{-10}^{0.5}+2.5SWC-4.9)} \tag{4.6}$$

$$R_{soil_h} = e^{(-0.3+0.6\frac{1.1GPP_s-3.6}{T_{-10}})} \tag{4.7}$$

where, GPP_s is gross primary production that has been smoothed using the SSA method with a 60 day window ; T_{-10} is soil temperature measured at 10 cm depth; and SWC is volumetric soil water content. The corresponding cross-validation MEF values are given in Table 3, indicating a range of capacities for GEP models to represent different respiration types.

Whilst GEP-derived models may differ between respiration types, there are a number of equivalent models for different respiration components. R_{soil} and R_{soil_a} were described by identical model structures (but distinctive parameter values), and R_{root} and R_{soil_h} were described by similar (but not identical) models. Overall, the most common selected drivers were T_{-10}, SWC and GPP.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

The highest performance in terms of MEF value was recorded for R_{soil_a} and for R_{soil}, that is 0.82 and 0.81 respectively. The lowest capacity of process representation, with an MEF value of 0.28, was recorded for R_{above} (Table 3), possibly because this specific component would need to include active versus inactive periods determined by dormancy and leaf fall (i.e. seasonality in this deciduous forest). A comparison of the predicted values and observed fluxes for all types of respiration can be seen in Figures 5 and 6. In order to explore the capacity of the GEP models generated for the R_{eco} components to recreate the larger, across compartmental sum fluxes, we summed the predictions of the models and compare them with the original fluxes. We find that the global modelling performance of the derived models remained in a very small range of the initial trained for the sum fluxes GEP models (Fig. 7), indicating that the larger fluxes actually exhibit sensitivity to some of the non-selected drivers, except that the sensitivity is present only in a certain compartment.

The residuals depict some remaining patterns (Fig 8 and Fig 3 of suppl.) and the null hypothesis of normal distribution was rejected for all seven respiration component residuals at 5% significance level with the one-sample Kolmogorov-Smirnov test. Hence, we might expect additional information that could be extracted from the residuals. In order to check whether the remaining structure was missed in the first training routine because of imposing a multiplicative form in the models by log-transforming the target data, we performed GEP runs on the residuals and combined the models. The improvement in overall modelling performance is minimal, yet model structures become overly complex. The capacity of the GEP approach to retrieve new information from the residuals is illustrated in Fig. 10 in comparison with that of the other MLM presented in section 3.1.1. When correlation values were computed between the candidate drivers and the residuals, no significant linear correlations were found (Fig. 5 of suppl.).

4.2.1 Model transferability

We investigated the capacity of each extracted model structure (equations 4.1-4.7) to represent a component of R_{eco} not seen in the training procedure. This was done by means of new CMA-ES optimization steps. The new prediction performances are illustrated in Tab. 4.

After optimization, none of the structures show an overall best MEF for all the R_{eco} components (i.e. we clearly cannot identify an optimal general model). However, we identify certain model structures that tend to perform overall better than others. This is the case for the R_{myc} model (eq. 4.5). It can also be seen that after the individual model optimizations, the structures for R_{eco} and that for R_{soil_a} have similar prediction capacities.

The prediction capacity of the GEP generated models in the context of other commonly utilized MLMs was assessed as well. KRR, ANN, SVM and, RF were used for generating 113 predicted data points as described in section 3.2 (Fig. 9). The prediction performance of GEP, KRR, ANN, SVM and, RF are shown in Fig. 10. Panel a contains the average MEF values computed for all MLM methods predicted values when compared to the original observations for $R_{eco}, R_{above}, R_{soil}, R_{root}, R_{myc}, R_{soil_a}, R_{soil_h}$. For all other cases, the performance is in the same range for all methods, but the GEP derived models having the lowest mean MEF values. Panel b shows that when all MLM were trained on the residuals obtained from comparing the GEP outputs with the observations, the GEP approach has the lowest capacity of capturing new relevant signals and is strongly outperformed by

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

the rest of the MLM, indicating that amount of information retrievable by GEP with the current fitness and settings is limited and captured already in the first run.

4.2.2 Comparing with literature models

Lastly, the GEP generated models were compared with some of the most commonly used literature models for describing respiration. The resulting MEF values obtained after individual parameter optimization using the CMA-ES procedure for each literature model are given in Tab. 5. The literature model structure that performed best overall in terms of prediction capacity measured as MEF is the $WaterQ_{10}$ model (Fig. 11). Figure 11 shows as well that certain types of respiration are easier to represent by all models, including the models GEP generated, whilst other types of respiration are poorly predicted by all models. Nevertheless, for all respiration types, the highest MEF values are generally recorded by the GEP models.

As the studied literature models performed best in modelling R_{soil}, we focus on contrasting GEP model results to literature model outcomes for this ecosystem respiration component. Of all models included, the GEP model and Q_{10} model including SWC dependency captured seasonal variability best, but no model satisfactorily represented short-term CO_2 flux variations (Fig. 12, panel a). All models show the largest range of residuals for the months May to July in 2008, and June/July in 2009 (Fig. 12, panel b), with the two best-performing models (GEP and $WaterQ_{10}$) having the narrowest range of absolute residuals. Monthly mean average errors (MAE) indicate a systematic underestimation of soil CO_2 efflux in the first year (Fig. 4 of suppl.).

5 Discussion

5.1 On the GEP method

In this work, the primary reason for the artificial experiments was obtaining a better understanding of the capacity of GEP to solve symbolic regression types of problems. We put an emphasis on GEP performance in the presence of noise. This aspect was important, given that monitoring data from terrestrial ecosystem CO_2 effluxes are typically contaminated by sometimes substantially large random uncertainties and measurement noise. In the case of NEE flux measurements, Lasslop et al. (2008) and Richardson et al. (2008) show that the measurement error typically scales with the magnitude of the flux, leading us to simulate that type of situation by adding noise that scales with signal to an already known function, equation 3.10. The results show that all the studied methods are stable to presence of noise in the training set. These results increase our confidence in the predictions generated by studied machine learning methods; in particular GEP derived modes can tolerate SNRs of 1. Considering that the SNR in the R_{eco} observations (if noise is only considered as random error) is probably larger than 4 which is where the curve starts decreasing in Fig 3, the noise presence in the data should not influence the automated model construction process and the real signals should be accurately captured when data uncertainties follow the pattern described here. On the other hand, for R_{soil} and other CO_2 fluxes measured with other techniques the magnitude and the distribution of the uncertainty can be different (Ryan and Law, 2005; Pérez-Priego et al., 2015), and we cannot state what the response of the present MLM is in the presence of different types of uncertainties and measurement noise.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Our findings illustrate that selection of CEM over MEF as a fitness function for optimization has a minor effect on the global mean MEF (Fig. 4a). We also notice that due to the constraints on presence of structure in the residuals and the length of the parameter vector, the final mean number of parameters is lower when CEM is chosen.

5.1.1 Limitations

One of the critical aspects in our work is that GEP, as implemented here, can only represent and derive "$n \rightarrow 1$" type of response functions. We are not able to generate model structures that encode e.g. system-intrinsic dynamics like feedback loops, which are expected from our current understanding of biogeochemical cycles in terrestrial ecosystems (Ehrenfeld et al., 2005; Friedlingstein et al., 2006). Hence, we believe GEP is suitable to e.g. understand and describe the sensitivities and non-linear responses to changes in hydro-meteorological drivers, but fails to represent more complex carbon or water reservoir dynamics. Pools and pool transfers cannot be introduced currently in the input, unless the depletion/repletion equations are known and can be included in the set of functions that can participate in the evolution.

Lagged responses can only be detected if the number of lags from a driver is correctly included in the input, which already implies sufficient knowledge of their existence and behaviour. Whilst in the current implementation of the GEP algorithm, shifts in conditions and responses cannot be encoded or detected; these could be addressed with the inclusion of a conditional operator in the set of functions encoded in the GEP evolution individuals.

Nevertheless, it would be fair to mention that the same limitations can affect the results of the other MLM and empirical models presented in this paper.

5.2 The value of GEP for modelling ecosystem respiration fluxes

We generated a series of model structures to describe terrestrial CO_2 respiration fluxes (equations 4.1-4.7). Most of these structures (5 out of 7) were of rather low complexity i.e. requiring 4 free parameters (which is certainly an effect of the chosen cost function CEM). The most complex structure is found for the R_{myc} model, which is in line with previous findings (Shi et al., 2012). Nevertheless, there is need for more in-depth analysis for determining whether the described processes make actual biological sense and the selected drivers and their interactions represent real processes and responses.

Interestingly, the models derived for R_{eco} and R_{soil} are structurally very similar. That is also the case of R_{root} and heterotrophic respiration, where the difference lies in the set of parameters and the added presence of an intercept in the formulation of the R_{soil_h} model. This finding suggests a consistency in the response of the R_{soil} components to their drivers, considering that the separation of the R_{soil} into its components might still lack accuracy (e.g. P. J. Hanson, N. T. Edwards and Andrews, 2000; Kuzyakov, 2006; Subke et al., 2006; Heinemeyer et al., 2011). However, all selected GEP generated models led to an underestimation of the high respiration fluxes (Fig. 6) and typically do not capture the peaks (fast responses). This phenomenon is in some cases a systematic pattern, and sometimes affects only certain times of the year. We suspect that is partly due to surface moisture affecting litter decomposition and fungal activity, as soil moisture was only monitored over the average 8 cm surface but the top few centimetres were most likely the most active and partly due to some potential pro-

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

cesses/drivers like lags between GPP and respiration (Hölttä et al., 2011) or phenology (Migliavacca et al., 2015) that were not specifically included in the learning process.

However, semi-empirical models similarly struggle to adequately simulate CO_2 flux peaks and in some cases monthly flux averages (Fig. 12). Structurally, the GEP-derived models share some key features with the semi-empirical approaches for temperature dependencies of CO_2 fluxes, which are typically captured by exponential relationships. A major difference is that when SWC has been chosen as driver, GEP often also identifies some exponential dependency, i.e. there are only certain parts of the signal that are strongly sensitive to varying SWC. This is a very intuitive pattern, which has not yet been reported in the literature and requires further exploration.

5.3 Data quality

During our study, it was apparent that the highest MEF values were obtained for all the studied methods in the case of the respiration types that had direct measured observations and were not derived. It might be the case that when fluxes are obtained from derivations, the measurement error will also increase, and the partition of clear signal existing in the observations is not sufficient for constructing a good model with GEP.

5.4 High frequency variability

For some of the modelled respiration components (e.g. R_{eco}) a large amount of high frequency variability present in the observations was lost (Fig. 5). The question is whether the GEP method lacks the ability to build models that correctly represent the processes and their dynamic responses, or whether the candidate drivers and the observations used for their representation are simply not sufficient for generating representative models. In the end, the response of R_{soil} and R_{eco} to external drivers might be too complex to describe solely with the currently available measurements and with the selected drivers.

Another explanation for missing some of the (high flux) variability could be in our choice of fitness function. As we decided to penalize during the learning process for structures with many parameters, it is likely that some structures were eliminated early-on during this process, even though they may be well-suited for describing a given process from a modelling efficiency point of view. However, this is a case of trade-off between a good fit and structural simplicity, and in our approach, we decided that simplicity of structure, i.e. the possibility of interpretation is a very important asset.

We suspected as well that the underestimation of the carbon flux variability was caused by the log-transformations we did on the observations. That could have introduced a bias that excluded interesting components of the model structures by forcing the method to build multiplicative models. However, when the GEP was run again on the residuals, without log-transforming, no new meaningful information was retrieved, indicating that multiplicative models were sufficient for reconstructing the studied R_{eco} components.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

5.5 Equifinality

Table 4 shows that when optimizing the parameters for all structures, the prediction performance becomes similar, which leads to the question of equifinality of dynamical systems, where different models that try to capture their structure, might have different formulations, but represent the same response.

A critical question for the applicability of any ecosystem model is whether the model structure is more important than the parametrisation of a given “best" model. For this question to be addressed, however, we need a larger sample of ecosystem types representative for different types of responses where we can explore the importance of the obtained structures and their parameter sets.

5.6 GEP models in the context of other machine learning methods

The comparison of GEP generated models machine-learning methods showed a narrow range of predicted fluxes (Fig. 10). The analysis of training all the MLM on the GEP residual output showed that the GEP approach is not able to retrieve any new meaningful structural components, but that the remaining MLM are much better at reconstructing the signal left in the residuals. This indicates that although the GEP is actually a reliable MLM when it comes reconstructing the underlying R_{eco} fluxes and is not prone to over-fitting, it could be that the current set-up of the GEP is not sufficient for an exhaustive description of those fluxes, or that might be overly strict on complexity of models compared to other MLM. The GEP approach has, nevertheless, the benefit of producing mathematical model structures that can be basis for future interpretation.

6 Conclusions and Outlook

Overall, our results suggest that the GEP approach is a potentially powerful tool of reverse engineering, particularly helpful for building ecological models when there is a minimum of a priori system understanding. We exemplified this conceptually using artificial data, but also show that GEP always yields as good or better results compared to conventionally used models in the case of ecosystem respiration. Based on data from a long-term monitoring site of different respiratory fluxes, and using GEP as a reverse engineering tool, we found new structures for modelling R_{eco} components. The GEP derived models outperform conventionally used models and generally differ by the way temperature and GPP, but also SWC are interpreted, indicating that conventional respiration models might have to be revised. At the same time, we found that when the GEP derived models are mutually compared, there are sufficient structural particularities for each terrestrial respiration type as to not allow for the formulation of a general R_{eco} law. More research is needed on a larger set of sites to identify widely usable models and for their interpretation. A particular matter of concern is the apparent equifinality of selected model structures, indicating that many response functions are yielding predictions of almost similar quality. A study of multiple sites would enable an investigation of whether specific ecosystem types result in similar model structures, or if whether response functions apply across contrasting ecosystem types.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

The current study has also revealed methodological aspects that could be improved. In particular, we found the inclusion of a parameter optimization step very helpful to further test the transferability of model structures. But this approach could be potentially integrated into the GEP evolution. More specifically, we think that the next development of GEP could include the parameter optimization as an intermediate step before selection during each evolution generation (Ilie et al.). In this way, a model structure could be chosen according to not only the current state of parameters but also on its potential and convergence to a global solution might be achieved faster.

Code and data availability

All code and data used to produce the results of this paper can be provided upon request by contacting Iulia Ilie or Miguel D. Mahecha.

Acknowledgements

We thank Markus Reichstein for useful comments and suggestions.

This work was supported by the International Max-Planck Research School for global Biogeochemical Cycles (IMPRS-gBGC), Jena, by the European Union's H2020 research and innovation programme project BACI; grant agreement 640176 and by NOVA grant UID/AMB/04085/2013. The Alice Holt Forest GHG Flux site is funded by the UK Forestry Commission.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

References

Ashworth, J., Wurtmann, E. J., and Baliga, N. S.: Reverse engineering systems models of regulation: Discovery, prediction and mechanisms, Current Opinion in Biotechnology, 23, 598–603, doi:10.1016/j.copbio.2011.12.005, http://www.ncbi.nlm.nih.gov/pubmed/22209016, 2012.

Auger, a. and Hansen, N.: A restart CMA evolution strategy with increasing population size, 2005 IEEE Congress on Evolutionary Computation, 2, 1769–1776, doi:10.1109/CEC.2005.1554902, http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1554902, 2005.

Bandt, C. and Pompe, B.: Permutation entropy: a natural complexity measure for time series., Physical review letters, 88, 174 102, doi:10.1103/PhysRevLett.88.174102, http://www.ncbi.nlm.nih.gov/pubmed/12005759, 2002.

Bennett, N. D., Croke, B. F., Jakeman, A. J., Newham, L. T. H., and Norton, J. P.: Performance evaluation of environmental models, 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, pp. 1–9, http://www.iemss.org/iemss2010/papers/S20/S.20.01.Performanceassessmentofenvironmentalmodels-ANTHONYJAKEMAN.pdf, 2010.

Beyer, H.-g. and Schwefel, H.-p.: Evolution Strategies, Natrual Computing, pp. 3–52, 2002.

Bonan, G. B.: Ecological Climatology: Concepts and Applications, Cambridge University Press, http://www.amazon.co.uk/Ecological-Climatology-Applications-Gordon-Bonan/dp/0521693195, 2008.

Bongard, J. and Lipson, H.: Automated reverse engineering of nonlinear dynamical systems., Proceedings of the National Academy of Sciences of the United States of America, 104, 9943–9948, doi:10.1073/pnas.0609476104, 2007.

Breiman, L.: Random forests, Machine Learning, 45, 5–32, doi:10.1023/A:1010933404324, http://link.springer.com/article/10.1023/A:1010933404324, 2001.

Buttlar, J. V., Zscheischler, J., and Mahecha, M. D.: An extended approach for spatiotemporal gapfilling: Dealing with large and systematic gaps in geoscientific datasets, Nonlinear Processes in Geophysics, 21, 203–215, doi:10.5194/npg-21-203-2014, 2014.

C. E. Shannon: A Mathematical Theory of Communication, The Bell System Technical Journal, Vol. 27, 379–423, 1948.

Chang, C.-C. and Lin, C.-J.: Libsvm, ACM Transactions on Intelligent Systems and Technology, 2, 1–27, doi:10.1145/1961189.1961199, http://dl.acm.org/citation.cfm?doid=1961189.1961199, 2011.

Coello, C. a. and Montes, E. M.: Constraint-handling in genetic algorithms through the use of dominance-based tournament selection, Advanced Engineering Informatics, 16, 193–203, doi:10.1016/S1474-0346(02)00011-3, http://www.sciencedirect.com/science/article/pii/S1474034602000113, 2002.

Ehrenfeld, J. G., Ravit, B., and Elgersma, K.: Feedback in the plant-soil system, Annual Review of Environment and Resources, 30, 75–115, doi:10.1146/annurev.energy.30.050504.144212, http://www.annualreviews.org/doi/10.1146/annurev.energy.30.050504.144212, 2005.

Fernando, D., Shamseldin, a. Y., and Abrahart, R. J.: Using gene expression programming to develop a combined runoff estimate model from conventional rainfall-runoff model outputs, in: IMACS / MODSIM Congress, July, pp. 748–754, 2009.

Ferreira, C.: Gene Expression Programming in Problem Solving, in: WSC6, 1992, pp. 1–9, 2002.

Ferreira, C.: Gene Expression Programming - Mathematical Modeling by an Artificial Intelligence, Springer-Verlag Berlin Heidelberg, 2 edn., doi:10.1007/3-540-32849-1, http://www.springer.com/us/book/9783540327967, 2006.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K.-G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., Zeng, N., Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K.,

Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K.-G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., and Zeng, N.: Climate–Carbon Cycle Feedback Analysis: Results from the C ⁴ MIP Model Intercomparison, Journal of Climate, 19, 3337–3353, doi:10.1175/JCLI3800.1, http://journals.ametsoc.org/doi/abs/10.1175/JCLI3800.1, 2006.

Gilmanov, T. G., Aires, L., Barcza, Z., Baron, V. S., Belelli, L., Beringer, J., Billesbach, D., Bonal, D., Bradford, J., Ceschia, E., Cook, D., Corradi, C., Frank, a., Gianelle, D., Gimeno, C., Gruenwald, T., Guo, H., Hanan, N., Haszpra, L., Heilman, J., Jacobs, a., Jones, M. B., Johnson, D. a., Kiely, G., Li, S., Magliulo, V., Moors, E., Nagy, Z., Nasyrov, M., Owensby, C., Pinter, K., Pio, C., Reichstein, M., Sanz, M. J., Scott, R., Soussana, J. F., Stoy, P. C., Svejcar, T., Tuba, Z., and Zhou, G.: Productivity, Respiration, and Light-Response Parameters of World Grassland and Agroecosystems Derived From Flux-Tower Measurements, Rangeland Ecology & Management, 63, 16–39, doi:10.2111/REM-D-09-00072.1, http://www.bioone.org/doi/abs/10.2111/REM-D-09-00072.1, 2010.

Gupta, H. V., Clark, M. P., Vrugt, J. a., Abramowitz, G., and Ye, M.: Towards a comprehensive assessment of model structural adequacy, Water Resources Research, 48, n/a–n/a, doi:10.1029/2011WR011044, http://doi.wiley.com/10.1029/2011WR011044, 2012.

Hansen, N.: Towards a New Evolutionary Computation, Studies in Fuzziness and Soft Computing, 192, 75–102, doi:10.1007/3-540-32494-1, http://www.scopus.com/inward/record.url?eid=2-s2.0-33845271655{&}partnerID=tZOtx3y1, 2006.

Hansen, N., Müller, S. D., and Koumoutsakos, P.: Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)., Evolutionary computation, 11, 1–18, doi:10.1162/106365603321828970, http://www.ncbi.nlm.nih.gov/pubmed/12804094, 2003.

Hashmi, M. Z. and Shamseldin, A. Y.: Use of Gene Expression Programming in regionalization of flow duration curve, Advances in Water Resources, 68, 1–12, doi:10.1016/j.advwatres.2014.02.009, http://linkinghub.elsevier.com/retrieve/pii/S0309170814000323, 2014.

Hearst, M. A.: Support vector machines, IEEE Intelligent Systems and Their Applications, 13, 18–28, doi:10.1109/5254.708428, http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=708428, 1998.

Heimann, M. and Reichstein, M.: Terrestrial ecosystem carbon dynamics and climate feedbacks., Nature, 451, 289–92, doi:10.1038/nature06591, http://dx.doi.org/10.1038/nature06591, 2008.

Heinemeyer, a., Di Bene, C., Lloyd, a. R., Tortorella, D., Baxter, R., Huntley, B., Gelsomino, a., and Ineson, P.: Soil respiration: Implications of the plant-soil continuum and respiration chamber collar-insertion depth on measurement and modelling of soil CO 2 efflux rates in three ecosystems, European Journal of Soil Science, 62, 82–94, doi:10.1111/j.1365-2389.2010.01331.x, 2011.

Heinemeyer, a., Wilkinson, M., Vargas, R., Subke, J. a., Casella, E., Morison, J. I. L., and Ineson, P.: Exploring the overflow tap theory: Linking forest soil CO 2 fluxes and individual mycorrhizosphere components to photosynthesis, Biogeosciences, 9, 79–95, doi:10.5194/bg-9-79-2012, 2012.

Hoerl, A. E. and Kennard, R. W.: Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics, 12, 55–67, doi:10.1080/00401706.1970.10488634, http://amstat.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634, 1970.

Hoffmann, M., Jurisch, N., Albiac Borraz, E., Hagemann, U., Drösler, M., Sommer, M., and Augustin, J.: Automated modeling of ecosystem CO2 fluxes based on periodic closed chamber measurements: A standardized conceptual and practical approach, Agricultural and Forest Meteorology, 200, 30–45, doi:10.1016/j.agrformet.2014.09.005, http://linkinghub.elsevier.com/retrieve/pii/S0168192314002160, 2015.

Hölttä, T., Mencuccini, M., and Nikinmaa, E.: A carbon cost-gain model explains the observed patterns of xylem safety and efficiency, Plant, Cell & Environment, 34, 1819–1834, doi:10.1111/j.1365-3040.2011.02377.x, http://doi.wiley.com/10.1111/j.1365-3040.2011.02377.x, 2011.

Ilie, I., Mahecha, M. D., Jung, M., Carvalhais, N., and Dittrich, P.: Evolving compact symbolic expressions by a GEP CMA-ES hybrid approach, Genetic Programming and Evolvable Machines.

Jakeman, a. J., Letcher, R. a., and Norton, J. P.: Ten iterative steps in development and evaluation of environmental models, Environmental Modelling and Software, 21, 602–614, doi:10.1016/j.envsoft.2006.01.004, 2006.

Kabanikhin, S. I.: Definitions and examples of inverse and ill-posed problems, Journal of Inverse and Ill-Posed Problems, 16, 317–357, doi:10.1515/JIIP.2008.019, 2008.

Keene, O. N.: The log transformation is special, Statistics in Medicine, 14, 811–819, doi:10.1002/sim.4780140810, http://doi.wiley.com/10.1002/sim.4780140810, 1995.

Khatibi, R., Naghipour, L., Ghorbani, M. a., Smith, M. S., Karimi, V., Farhoudi, R., Delafrouz, H., and Arvanaghi, H.: Developing a predictive tropospheric ozone model for Tabriz, Atmospheric Environment, 68, 286–294, doi:10.1016/j.atmosenv.2012.11.020, http://linkinghub.elsevier.com/retrieve/pii/S1352231012010722, 2013.

Kotanchek, M. E., Vladislavleva, E., and Smits, G.: Symbolic Regression Is Not Enough : It Takes a Village to Raise a Model, in: Genetic Programming Theory and Practice X, pp. 187–203, Springer Science+Business Media New York, doi:10.1007/978-1-4614-6846-2, 2013.

Kuzyakov, Y.: Sources of CO2 efflux from soil and review of partitioning methods, Soil Biology and Biochemistry, 38, 425–448, doi:10.1016/j.soilbio.2005.08.020, 2006.

Lasslop, G., Reichstein, M., Kattge, J., and Papale, D.: Influences of observation errors in eddy flux data on inverse model parameter estimation, Biogeosciences Discussions, 5, 751–785, doi:10.5194/bgd-5-751-2008, 2008.

Lasslop, G., Migliavacca, M., Bohrer, G., Reichstein, M., Bahn, M., Ibrom, a., Jacobs, C., Kolari, P., Papale, D., Vesala, T., Wohlfahrt, G., and Cescatti, a.: On the choice of the driving temperature for eddy-covariance carbon dioxide flux partitioning, Biogeosciences, 9, 5243–5259, doi:10.5194/bg-9-5243-2012, 2012.

Lloyd, J. and Taylor, J. a.: On the temperature dependence of soil respiration, Functional Ecology, 8, 315–323, 1994.

Luo, Y., Keenan, T. F., and Smith, M. J.: Predictability of the terrestrial carbon cycle, Global Change Biology, 21, 1737–1751, doi:10.1111/gcb.12766, http://www.ncbi.nlm.nih.gov/pubmed/25327167, 2015.

Mahecha, M. D., Reichstein, M., Carvalhais, N., Lasslop, G., Lange, H., Seneviratne, S. I., Vargas, R., Ammann, C., Arain, M. A., Cescatti, A., Janssens, I. a., Migliavacca, M., Montagnani, L., and Richardson, A. D.: Global convergence in the temperature sensitivity of respiration at ecosystem level., Science (New York, N.Y.), 329, 838–40, doi:10.1126/science.1189587, http://www.ncbi.nlm.nih.gov/pubmed/20603495, 2010.

Migliavacca, M., Reichstein, M., Richardson, A. D., Colombo, R., Sutton, M. a., Lasslop, G., Tomelleri, E., Wohlfahrt, G., Carvalhais, N., Cescatti, A., Mahecha, M. D., Montagnani, L., Papale, D., Zaehle, S., Arain, A., Arneth, A., Black, T. A., Carrara, A., Dore, S., Gianelle, D., Helfter, C., Hollinger, D., Kutsch, W. L., Lafleur, P. M., Nouvellon, Y., Rebmann, C., Humberto, R., Rodeghiero, M., Roupsard, O., Sebastià, M. T., Seufert, G., Soussana, J. F., and Michiel, K.: Semiempirical modeling of abiotic and biotic factors controlling ecosystem respiration across eddy covariance sites, Global Change Biology, 17, 390–409, doi:10.1111/j.1365-2486.2010.02243.x, http://doi.wiley.com/10.1111/j.1365-2486.2010.02243.x, 2011.

Migliavacca, M., Sonnentag, O., Keenan, T. F., Cescatti, a., O'Keefe, J., and Richardson, a. D.: On the uncertainty of phenological responses to climate change, and implications for a terrestrial biosphere model, Biogeosciences, 9, 2063–2083, doi:10.5194/bg-9-2063-2012, 2012.

Migliavacca, M., Reichstein, M., Richardson, A. D., Mahecha, M. D., Cremonese, E., Delpierre, N., Galvagno, M., Law, B. E., Wohlfahrt, G., Andrew Black, T., Carvalhais, N., Ceccherini, G., Chen, J., Gobron, N., Koffi, E., William Munger, J., Perez-Priego, O., Robustelli, M., Tomelleri, E., and Cescatti, A.: Influence of physiological phenology on the seasonal pattern of ecosystem respiration in deciduous forests, Global Change Biology, pp. 363–376, doi:10.1111/gcb.12671, 2015.

Mitchell, S., Beven, K., and Freer, J.: Multiple sources of predictive uncertainty in modeled estimates of net ecosystem CO2 exchange, Ecological Modelling, 220, 3259–3270, doi:10.1016/j.ecolmodel.2009.08.021, http://www.sciencedirect.com/science/article/pii/S0304380009006000, 2009.

Moncrieff, J., Massheder, J., de Bruin, H., Elbers, J., Friborg, T., Heusinkveld, B., Kabat, P., Scott, S., Soegaard, H., and Verhoef, A.: A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide, Journal of Hydrology, 188-189, 589–611, doi:10.1016/S0022-1694(96)03194-0, http://www.sciencedirect.com/science/article/pii/S0022169496031940, 1997.

Moyano, F. E., Kutsch, W. L., and Rebmann, C.: Soil respiration fluxes in relation to photosynthetic activity in broad-leaf and needle-leaf forest stands, Agricultural and Forest Meteorology, 148, 135–143, doi:10.1016/j.agrformet.2007.09.006, 2008.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282–290, doi:10.1016/0022-1694(70)90255-6, http://www.sciencedirect.com/science/article/pii/0022169470902556, 1970.

P. J. Hanson, N. T. Edwards, C. T. G. and Andrews, J. A.: Separating Root and Soil Microbial Contributions to Soil Respiration: A Review of Methods and Observations on JSTOR, http://www.jstor.org/stable/1469555?seq=1{#}page{_}scan{_}tab{_}contents, 2000.

Peng, S., Ciais, P., Chevallier, F., Peylin, P., Cadule, P., Sitch, S., Piao, S., Ahlström, A., Huntingford, C., Levy, P., Li, X., Liu, Y., Lomas, M., Poulter, B., Viovy, N., Wang, T., Wang, X., Zaehle, S., Zeng, N., Zhao, F., and Zhao, H.: Benchmarking the seasonal cycle of CO2 fluxes simulated by terrestrial ecosystem models, Global Biogeochemical Cycles, pp. 46–64, doi:10.1002/2014GB004931.Received, 2014a.

Peng, Y., Yuan, C., Qin, X., Huang, J., and Shi, Y.: An improved Gene Expression Programming approach for symbolic regression problems, Neurocomputing, 137, 293–301, doi:10.1016/j.neucom.2013.05.062, http://linkinghub.elsevier.com/retrieve/pii/S0925231214002598, 2014b.

Pérez-Priego, O., López-Ballesteros, A., Sánchez-Cañete, E. P., Serrano-Ortiz, P., Kutzbach, L., Domingo, F., Eugster, W., and Kowalski, A. S.: Analysing uncertainties in the calculation of fluxes using whole-plant chambers: random and systematic errors, Plant and Soil, 393, 229–244, doi:10.1007/s11104-015-2481-x, http://link.springer.com/10.1007/s11104-015-2481-x, 2015.

Reichstein, M. and Beer, C.: Soil respiration across scales: The importance of a model-data integration framework for data interpretation, Journal of Plant Nutrition and Soil Science, 171, 344–354, doi:10.1002/jpln.200700075, http://doi.wiley.com/10.1002/jpln.200700075, 2008.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J. M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm, Global Change Biology, 11, 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x, 2005.

Richardson, A. D., Mahecha, M. D., Falge, E., Kattge, J., Moffat, A. M., Papale, D., Reichstein, M., Stauch, V. J., Braswell, B. H., Churkina, G., Kruijt, B., and Hollinger, D. Y.: Statistical properties of random CO2 flux measurement uncertainty inferred from model residuals, Agricultural and Forest Meteorology, 148, 38–50, doi:10.1016/j.agrformet.2007.09.001, http://linkinghub.elsevier.com/retrieve/pii/S0168192307002365, 2008.

Rosso, O. A., Larrondo, H. A., Martin, M. T., Plastino, A., and Fuentes, M. A.: Distinguishing Noise from Chaos, Physical Review Letters, 99, 154 102, doi:10.1103/PhysRevLett.99.154102, http://link.aps.org/doi/10.1103/PhysRevLett.99.154102, 2007.

Ryan, M. G. and Law, B. E.: Interpreting, measuring, and modeling soil respiration, Biogeochemistry, 73, 3–27, doi:10.1007/s10533-004-5167-7, http://www.springerlink.com/index/10.1007/s10533-004-5167-7, 2005.

Shi, Z., Wang, F., and Liu, Y.: Response of soil respiration under different mycorrhizal strategies to precipitation and temperature, Journal of Soil Science and Plant Nutrition, 12, 411–420, doi:Doi 10.4067/S0718-95162013005000053, 2012.

Sippel, S., Lange, H., Mahecha, M., Hauhs, M., Gans, F., Bodesheim, P., and Rosso, O.: Diagnosing the dynamics of observed and simulated ecosystem gross primary productivity with time causal information theory quantifiers, PLoS ONE, 4/2016, in, 2016.

Subke, J.-A., Inglima, I., and Francesca Cotrufo, M.: Trends and methodological impacts in soil CO2 efflux partitioning: A metaanalytical review, Global Change Biology, 12, 921–943, doi:10.1111/j.1365-2486.2006.01117.x, http://doi.wiley.com/10.1111/j.1365-2486.2006.01117.x, 2006.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences, 13, 4291–4313, doi:10.5194/bg-13-4291-2016, http://www.biogeosciences.net/13/4291/2016/, 2016.

Traore, S. and Guven, A.: New algebraic formulations of evapotranspiration extracted from gene-expression programming in the tropical seasonally dry regions of West Africa, Irrigation Science, 31, 1–10, doi:10.1007/s00271-011-0288-y, http://link.springer.com/10.1007/s00271-011-0288-y, 2013.

Trumbore, S.: Carbon respired by terrestrial ecosystems–recent progress and challenges, Global Change Biology, 2, 141–153, doi:10.1111/j.1365-2486.2005.01067.x, http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2486.2006.01067.x/full, 2006.

Wehr, R., Munger, J. W., McManus, J. B., Nelson, D. D., Zahniser, M. S., Davidson, E. A., Wofsy, S. C., and Saleska, S. R.: Seasonality of temperate forest photosynthesis and daytime respiration, Nature, 534, 680–683, doi:10.1038/nature17966, http://www.nature.com/doifinder/10.1038/nature17966, 2016.

Wilkinson, M., Eaton, E. L., Broadmeadow, M. S. J., and Morison, J. I. L.: Inter-annual variation of carbon uptake by a plantation oak woodland in south-eastern England, Biogeosciences, 9, 5373–5389, doi:10.5194/bg-9-5373-2012, 2012.

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., and Jung, M.: Improving land surface models with FLUXNET data, Biogeosciences, pp. 1341–1359, http://www.biogeosciences.net/6/1341/2009/bg-6-1341-2009.pdf, 2009.

Yegnanarayana, B.: Artificial neural networks, PHI Learning Pvt. Ltd., 2009.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

Glossary

chromosome individual used in automatically evolving an optimal solution comprised of a set of genes that are connected with a binary operation (e.g. $+ \times -$). 4

CMA-ES covariance matrix adaptation evolutionary strategy. 6

expression tree binary tree used to represent algebraic expressions. 4

gene set of characters of fixed length that encodes an expression tree. 4

generation time step of an evolution. 4

genetic manipulation change produced in the structure of a chromosome and the expression tree it encodes by altering the strings composing the genes (e.g. mutation, inversion, recombination, etc.) . 4

GEP **g**ene **e**xpression **p**rogramming, machine learning method that evolves chromosome structures with the purpose of minimizing a cost function. 3

ill-posed problem a problem for which the solutions might not be unique or unstable, also known as an inverse problem. 7

manipulation rate probability of a genetic manipulation to occur during a generation. 4

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

Table 1. GEP settings

Parameter	Artificial data	Real observations
Number of chromosomes	2000	2000
Number of genes	3	2
Head length	5	6
Functions	$+,-,/,*,x^y,\sqrt{},\ln,\exp,\sin,\cos$	$+,-,/,*,x^y,\sqrt{},\ln,\exp$
Terminals	x_1,x_2,x_3	$GPP_s,T_{Air},T_{-10},SWC$
Link function	+	+
Max run time	1200 seconds	1800 seconds
Fitness function	CEM	CEM
Selection method for replication	tournament(Coello and Montes, 2002)	tournament
Mutation probability	0.2	0.2
IS and RIS transpositions probabilities	0.05	0.05
Two-point recombination probability	0.3	0.3
Inversion probability	0.05	0.05
One point recombination probability	0.4	0.4

Table 2. Respiration model formulations commonly used in the environmental science community

Model	Formulation[1]	Reference
Arrhenius	$a \times e^{-E_0/RT}$	(Lloyd and Taylor, 1994)
Q_{10}	$\phi_1 \times \phi_2^{(\frac{T-T_{ref}}{10})}$	(Reichstein and Beer, 2008)
Water Q_{10}	$\phi_1 \times \phi_2^{(\frac{T-T_{ref}}{10})} \times \frac{SWC}{SWC+\phi_3} \times \frac{\phi_4}{SWC+\phi_4}$	(Richardson et al., 2008)
$LinGPP$	$(R_0 + k_2 GPP) \times e^{E_0(\frac{1}{T_{ref}-T_0} - \frac{1}{T_A-T_0})} \times \frac{\alpha k + SWC(1-\alpha)}{k+SWC(1-\alpha)}$	(Migliavacca et al., 2011)
$ExpGPP$	$[R_0 + R_2(1-e^{k_2 GPP})] \times e^{E_0(\frac{1}{T_{ref}-T_0} - \frac{1}{T_A-T_0})} \times \frac{\alpha k + SWC(1-\alpha)}{k+SWC(1-\alpha)}$	(Migliavacca et al., 2011)
$addLinGPP$	$R_0 \times e^{E_0(\frac{1}{T_{ref}-T_0} - \frac{1}{T_A-T_0})} \times \frac{\alpha k + SWC(1-\alpha)}{k+SWC(1-\alpha)} + k_2 GPP$	(Migliavacca et al., 2011)
$addExpGPP$	$R_0 \times e^{E_0(\frac{1}{T_{ref}-T_0} - \frac{1}{T_A-T_0})} \times \frac{\alpha k + SWC(1-\alpha)}{k+SWC(1-\alpha)} + R_2(1-e^{k_2 GPP})$	(Migliavacca et al., 2011)

Where $a, E_0, \phi_1, \phi_2, \phi_3, \phi_4, R_0, R_2, k, k_2$ and α are model parameters that can be optimized

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

Table 3. Modelling performance for all extracted model structures after cross validation over 90 cases.

Respiration type	MEF	σMEF	Equation
R_{eco}	0.56	0.14	4.1
R_{above}	0.28	0.13	4.2
R_{soil}	0.81	0.13	4.3
R_{root}	0.59	0.10	4.4
R_{myc}	0.42	0.13	4.5
R_{soil_a}	0.82	0.13	4.6
R_{soil_h}	0.51	0.11	4.7

Table 4. Average validation MEF performance for all extracted model structures when re-optimized against all other respiration CO_2 flux observations.

trained for/ opt. for	R_{eco}	R_{above}	R_{soil}	R_{root}	R_{myc}	R_{soil_a}	R_{soil_h}
R_{eco} (Eq. 4.1)	0.56	0.25	0.77	0.51	-0.06	0.67	0.42
R_{above} (Eq. 4.2)	0.56	0.27	0.69	0.52	0.01	0.62	0.47
R_{soil} (Eq. 4.3)	0.50	0.13	0.81	0.35	0.29	0.82	0.40
R_{root} (Eq. 4.4)	0.34	0.22	0.61	0.57	0.03	0.65	0.51
R_{myc} (Eq. 4.5)	0.54	0.16	0.81	0.50	0.43	0.84	0.51
R_{soil_a} (Eq. 4.6)	0.50	0.13	0.81	0.35	0.29	0.82	0.40
R_{soil_h} (Eq. 4.7)	0.55	0.23	0.76	0.53	-0.03	0.67	0.51

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

Table 5. Average validation MEF performance for CMA-ES optimized selected literature model formulations when compared with respiration CO_2 flux observations.

Model formulation	R_{eco}	R_{above}	R_{soil}	R_{root}	R_{myc}	R_{soil_a}	R_{soil_h}
Arrhenius	0.41	0.15	0.65	0.50	0.07	0.61	0.38
Q_{10}	0.47	0.19	0.69	0.52	0.09	0.62	0.46
Water Q_{10}	0.50	0.20	0.79	0.55	0.40	0.81	0.43
$LinGPP$	0.55	0.25	0.74	0.57	0.17	0.70	0.49
$ExpGPP$	0.58	0.30	0.76	0.57	0.20	0.72	0.54
$addLinGPP$	0.55	0.27	0.73	0.56	0.12	0.67	0.48
$addExpGPP$	0.56	0.27	0.73	0.54	0.20	0.69	0.49

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

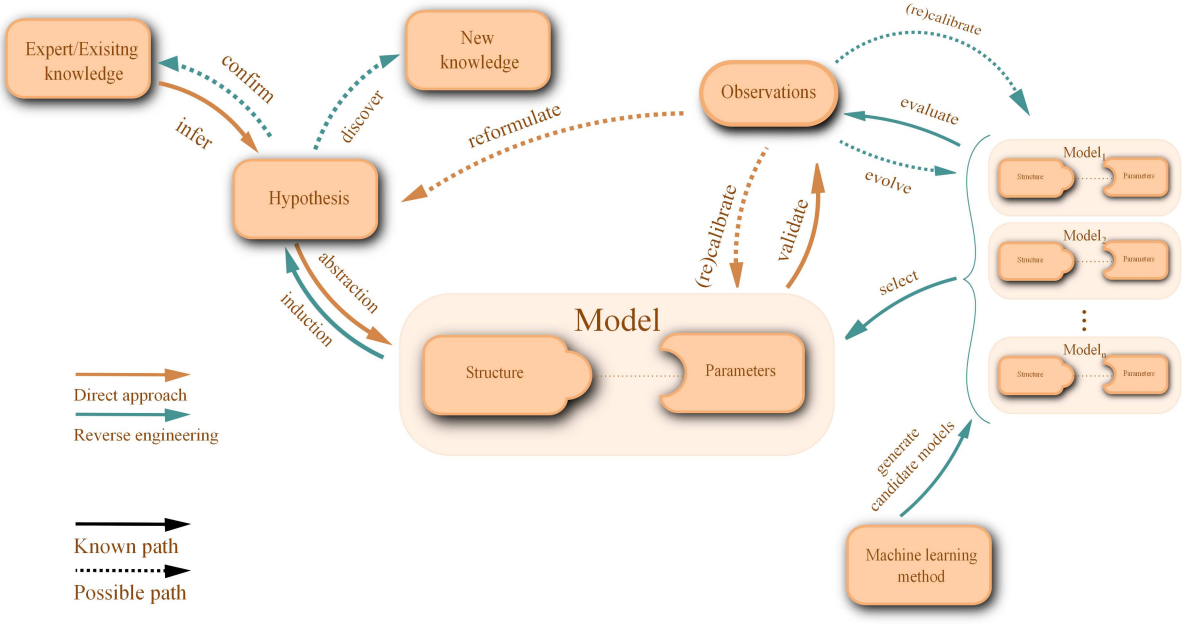


Figure 1. Direct approach and reverse engineering in model development for describing dynamical systems. Existing and possible steps needed in the process of building a model. For the direct approach, the process starts with the building of hypothesis from existing knowledge, the hypothesis is then subject of abstraction and summarized in a mathematical model that has two components: the structure and the parameters. The mathematical model can be translated into a computational form that will generate predictions. Depending on how well the predicted values manage to recreate the available observations, the model's parameters are calibrated or if the general trends are missed, there might be need for structural reformulation. On the other hand, in the reverse engineering approach, a machine learning method is used to generate a set of candidate models that are then compared with the available observations and which according to the prediction capacity may have to go through structural changes by automatic evolution or through a final parameter adaptation. From the set of evolved models, the best model in terms of prediction capacity is chosen and its structure will be the basis for hypothesis building, as an expert would try to explain why a specific structure was automatically evolved and whether the structure of the model can be explained from the studied system intrinsic processes. If that will be the case, and the structure has not emerged randomly, the conclusions can be compared with the existing knowledge which can be either reconfirmed or new aspects of the studied system might be brought into light.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

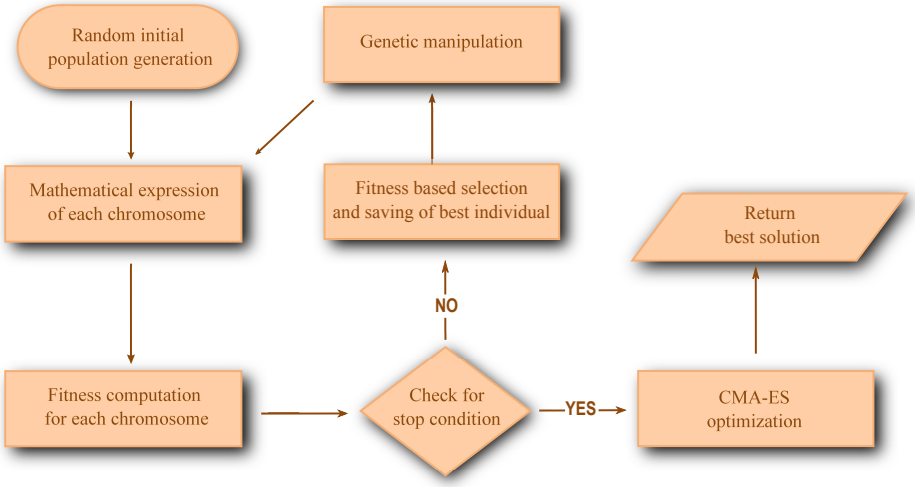
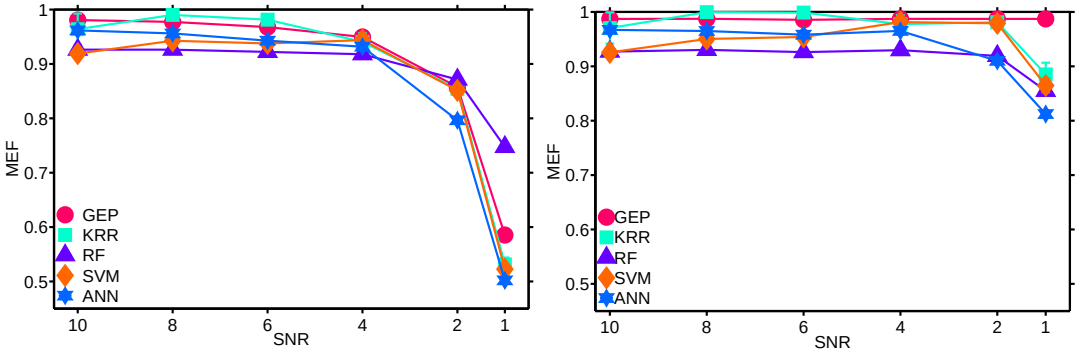


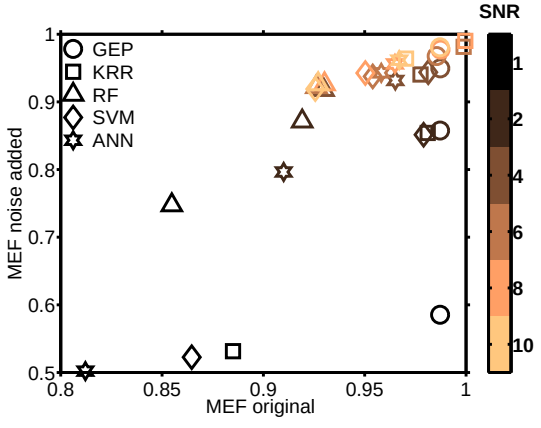
Figure 2. The work flow used in solving symbolic regression problems with GEP.. The process of evolving an optimal solution from observations starts with randomly generating a set number of evolution individuals called chromosomes. The chromosomes are composed of genes that are sets of strings encoding expression trees that can be translated into mathematical expressions in the subsequent step. Following the mathematical expression comes the evaluation of each emerging individual (model) against the target variable values and for each one a fitness values is assigned. If the stopping criterion has not been reached (e.g.. best fitness possible, highest number of generations allowed, convergence etc.) the best individual in terms of fitness is saved and the remaining set of chromosomes are selected for genetic manipulation. When the stop criterion is reached, the parameters of the best chromosome is calibrated against the training data with an optimization approach, the CMA-ES, and the best solution is returned.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016



(a) Validating against noise containing datasets.

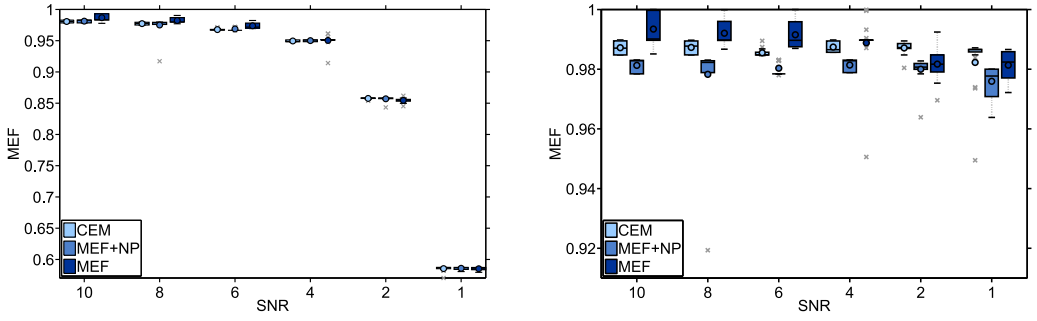
(b) Validating against noise-free datasets.



(c)

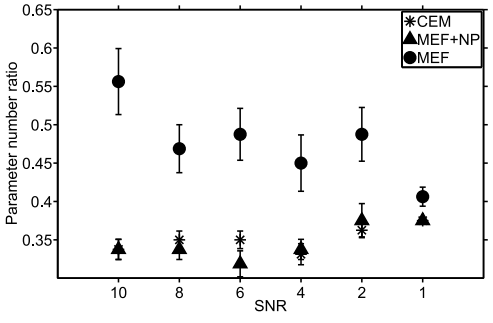
Figure 3. Effect of adding noise to original signal on prediction capacity for GEP, KRR, RF, SVM and ANN. The first panel contains the evolution of mean MEF values from 20 independent runs for each increasing level of noise. MEF is computed after learning from a data set of 200 data points and validating against 1000 data points containing noise. The second panel shows the evolution of mean MEF values from 20 independent runs for each increasing level of noise where MEF is computed after after learning from a data set of 200 data points and validating against 1000 data points generated from equation 3.10. Panel c shows the compared individual MEF evolutions of the studied machine learning methods with noise.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016



(a) Mean MEF when validation against noisy data after 20 GEP runs with different fitness functions.

(b) Mean MEF when validation against noise-free data after 20 GEP runs with different fitness functions.



(c) Ratio of predicted number of parameters to true number of parameters after 20 GEP runs with different fitness functions.

Figure 4. Effects on modelling performance and parameter number caused by choice of fitness function during GEP training for artificial noisy data generated by equation 3.10, where MEF is defined in equation 2.1 and CEM is defined in equation 2.3.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

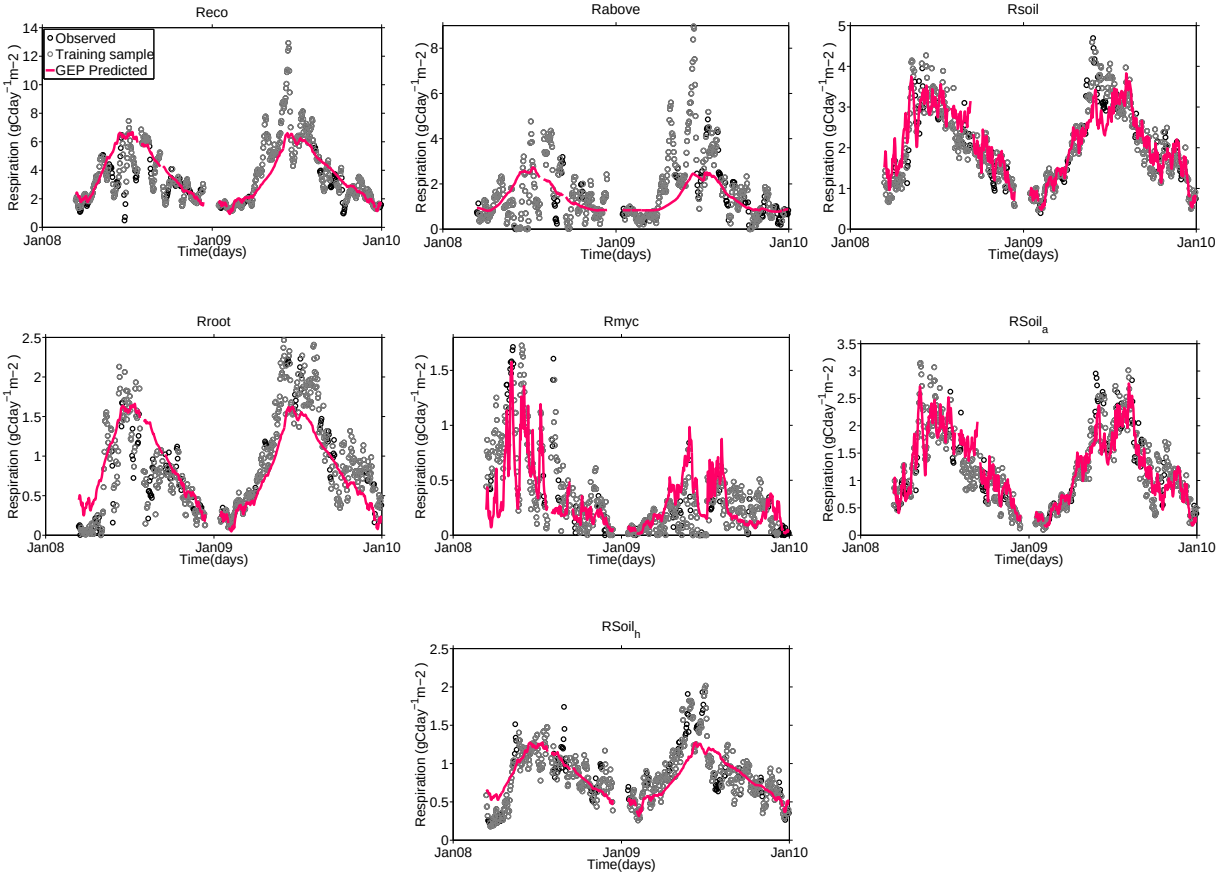


Figure 5. Observed and predicted outgoing CO_2 fluxes. 613 time steps of daily averaged CO_2 effluxesvfor two years at the Alice Holt oak forest site. The predicted values are generated with the models extracted by the GEP approach with the settings given in table 1 for the following types of respiration: $R_{eco}, R_{above}, R_{soil}, R_{root}, R_{myc}, R_{soil_a}, R_{soil_h}$. The models are given in equations: 4.1-4.7

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

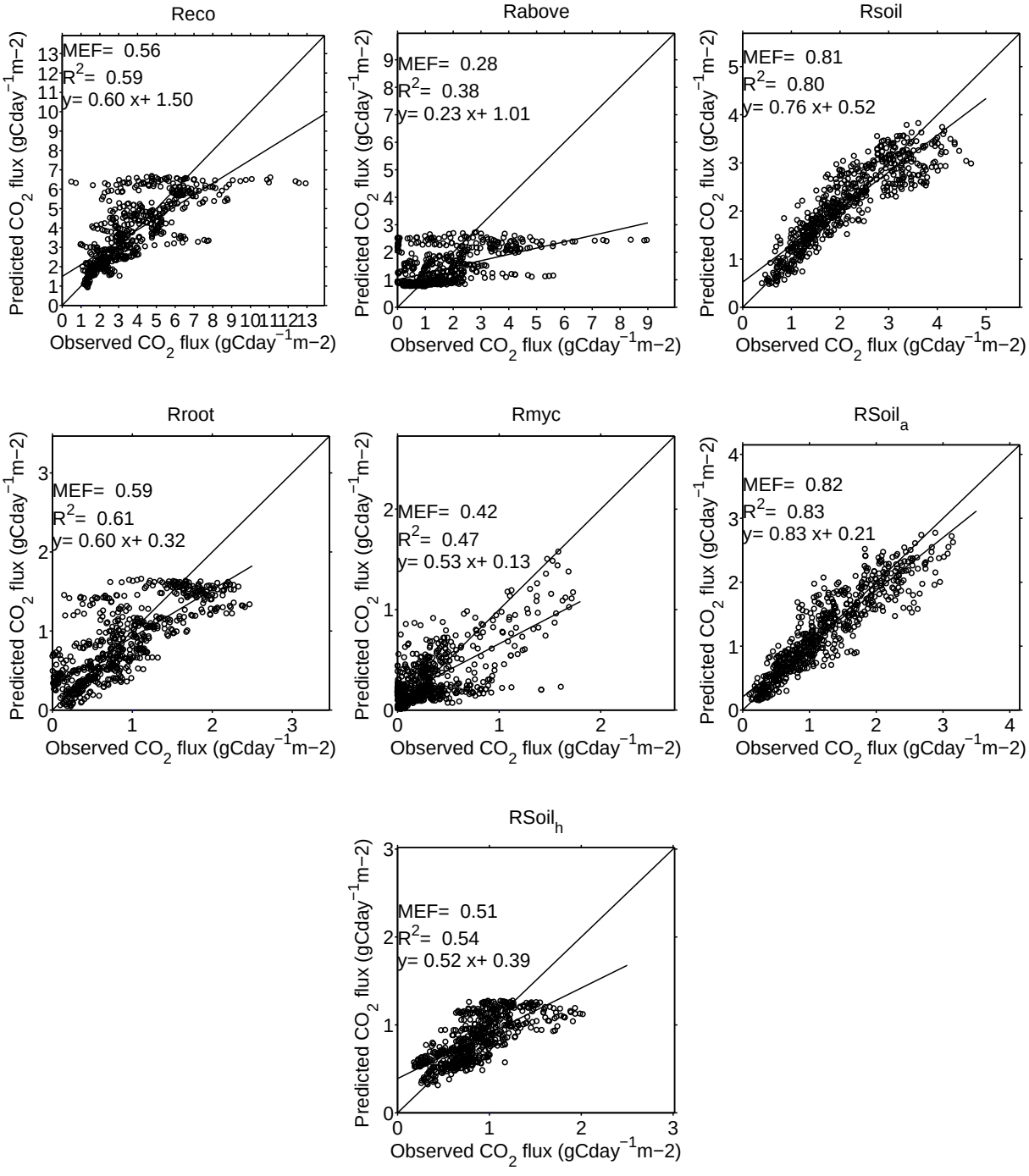


Figure 6. Observed and predicted outgoing CO_2 fluxes.613 time steps of daily averaged CO_2 effluxes for two years at the Alice Holt oak forest site. The predicted values are generated with the models extracted by the GEP approach with the settings given in table 1 for the following types of respiration: $R_{eco}, R_{above}, R_{soil}, R_{root}, R_{myc}, R_{soil_a}, R_{soil_h}$. The models are given in equations: 4.1-4.7

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

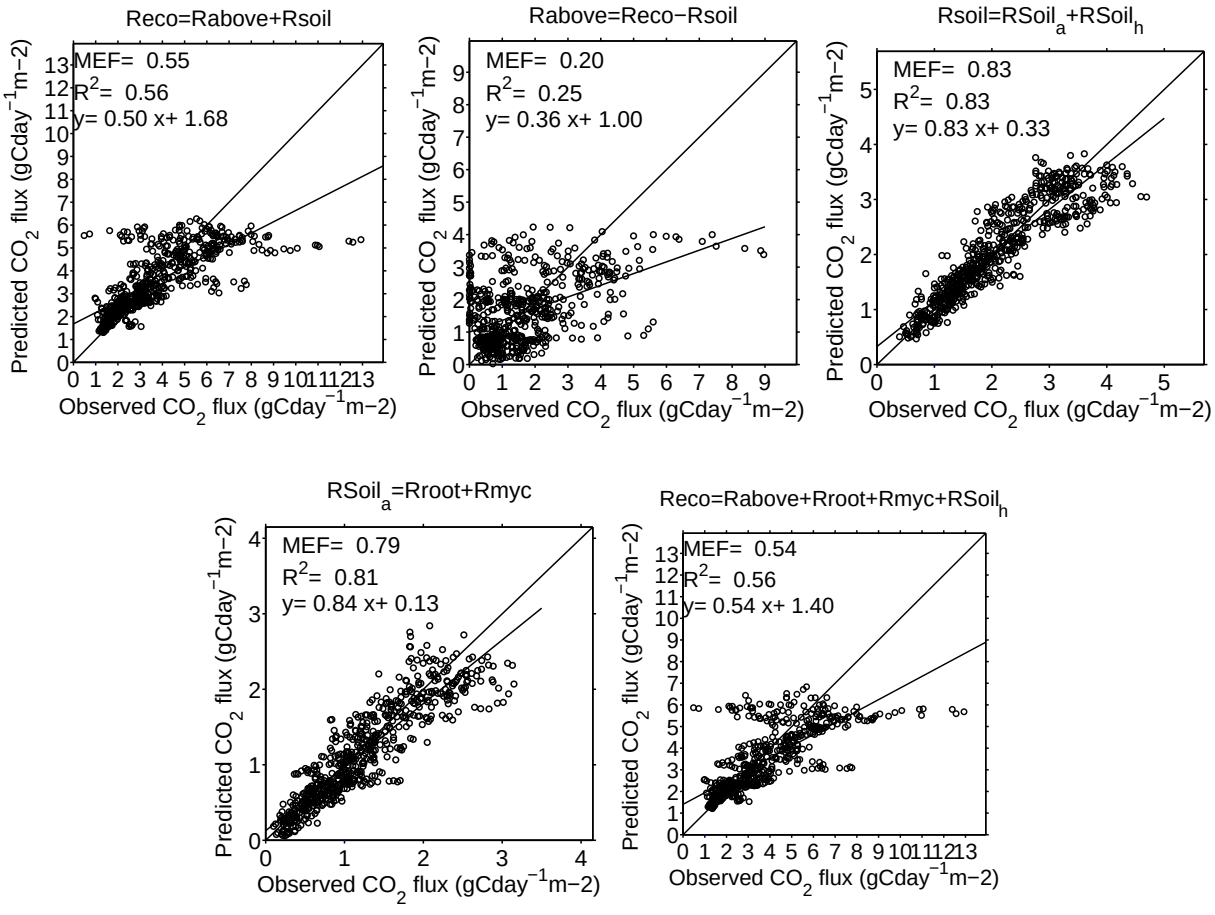


Figure 7. Observed versus predicted R_{eco} components fluxes, where predicted values are computed as derived fluxes.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

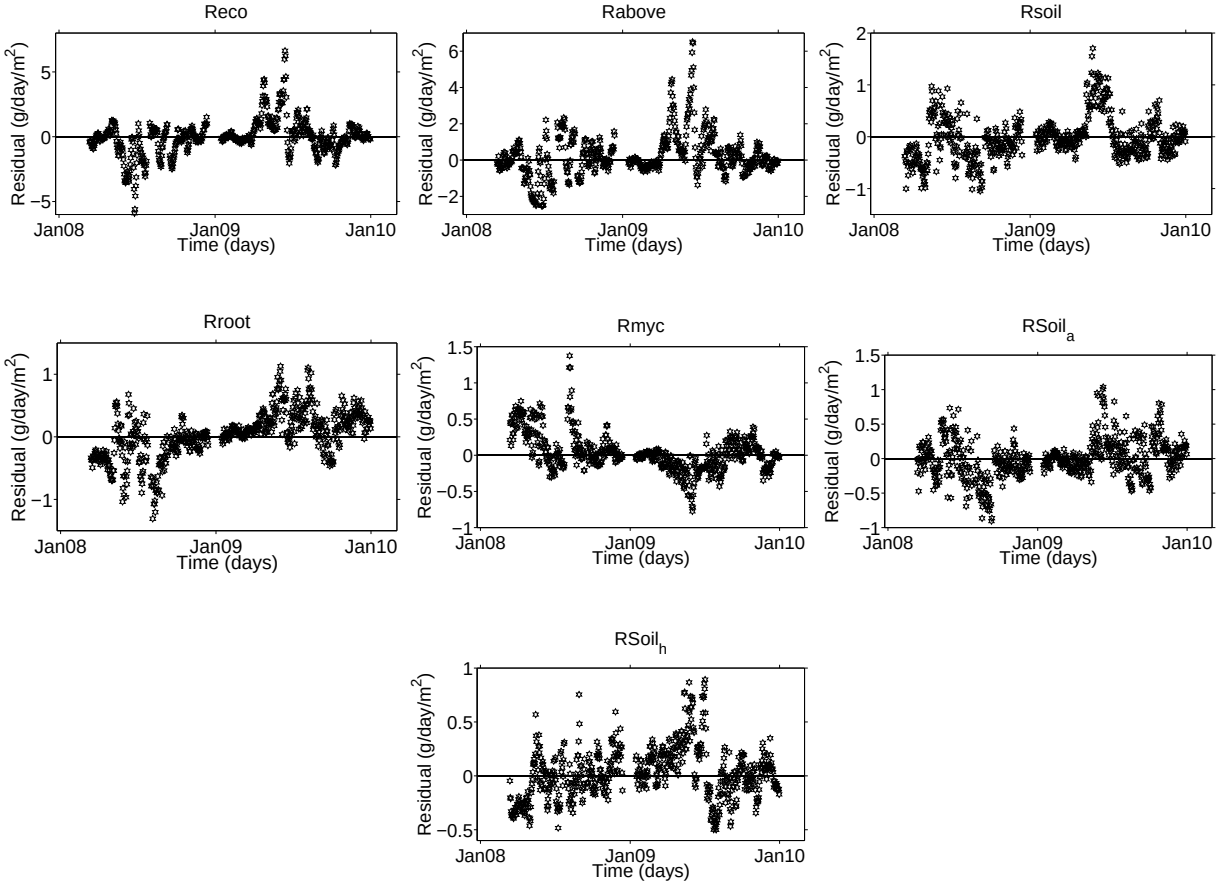


Figure 8. Residuals computed for the GEP models after training on log-transformed data.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

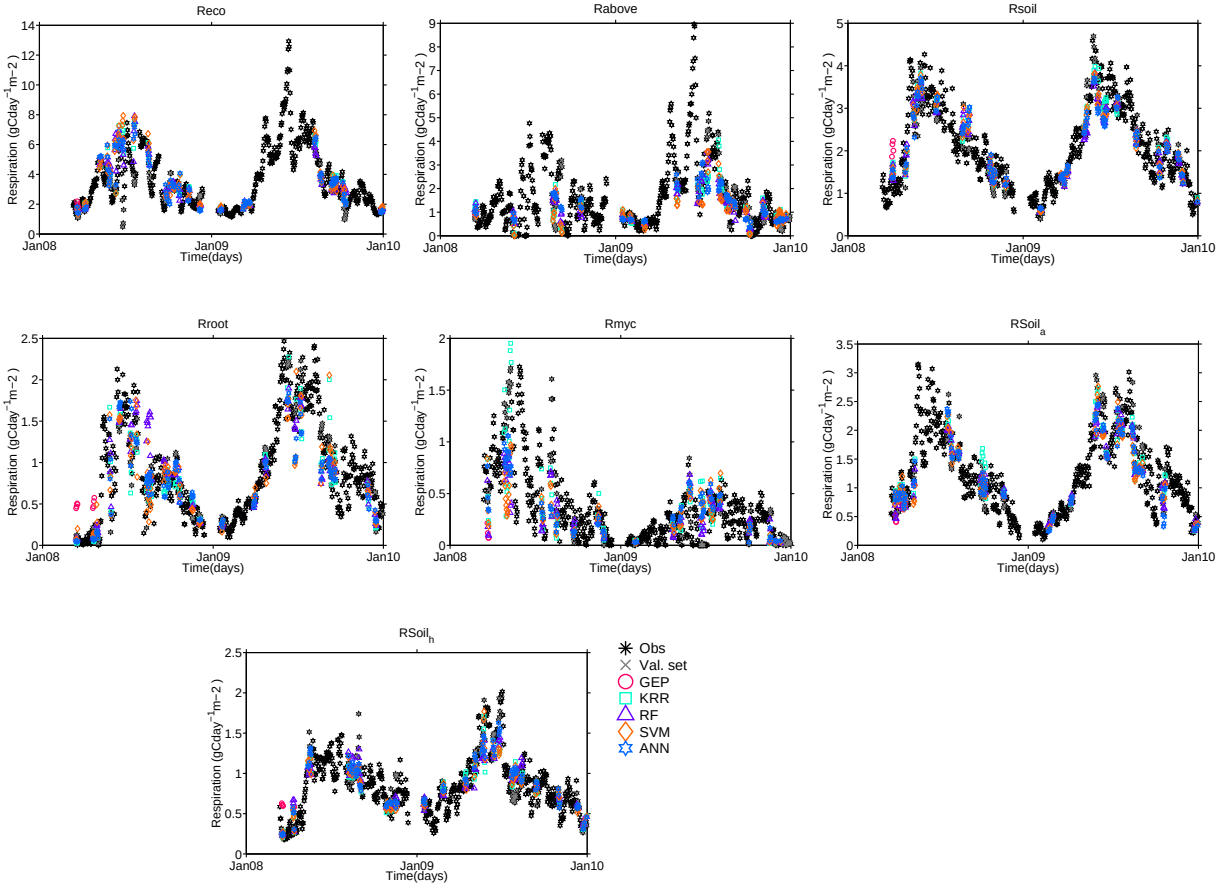


Figure 9. Observed CO_2 fluxes and one set of 113 predicted values given by the some common machine learning methods (MLM) after training on 500 data points.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

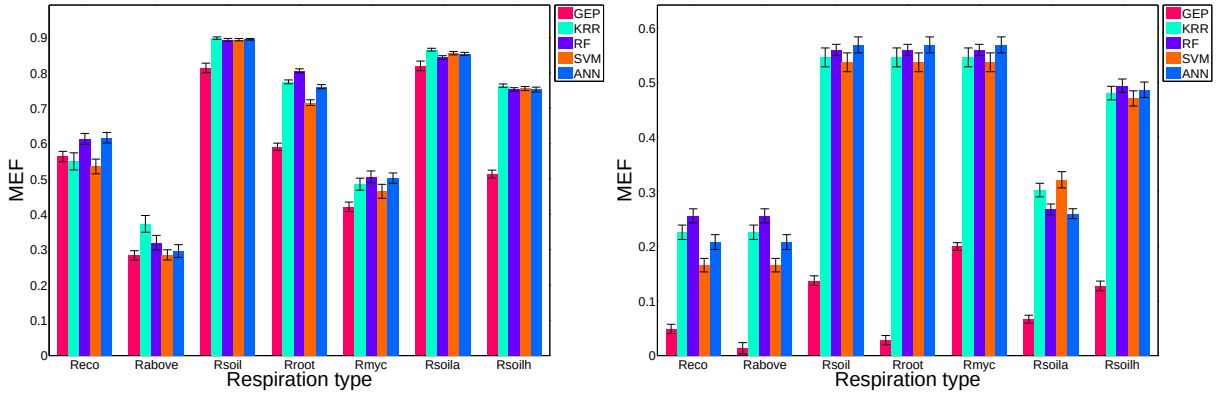


Figure 10. Machine learning methods (MLM) prediction performance for all respirations components (left) and for the residuals (right) resulting from the GEP trained models. The MEF values obtained for validation by all the MLM methods for $R_{eco}, R_{above}, R_{soil}, R_{root}, R_{myc}, R_{soil_a}, R_{soil_h}$

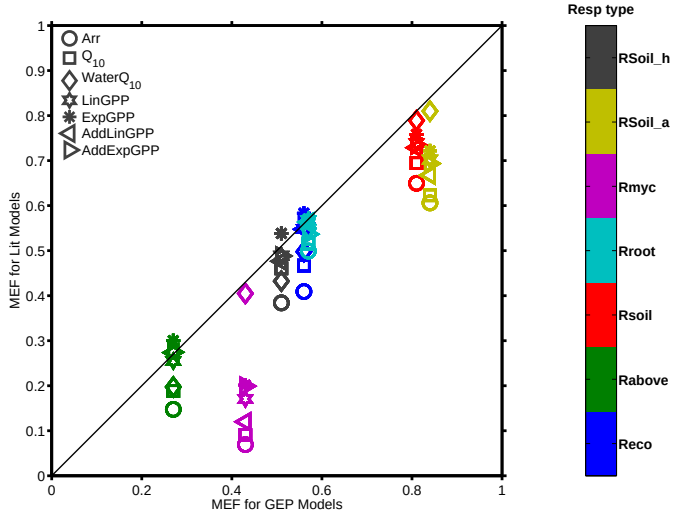


Figure 11. MEF validation values for literature models and for the best GEP model in terms of MEF at each respiration level. Each R_{eco} flux component is shown in a separate colour.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-242, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 7 November 2016

Geoscientific Model Development Discussions
Open Access
EGU

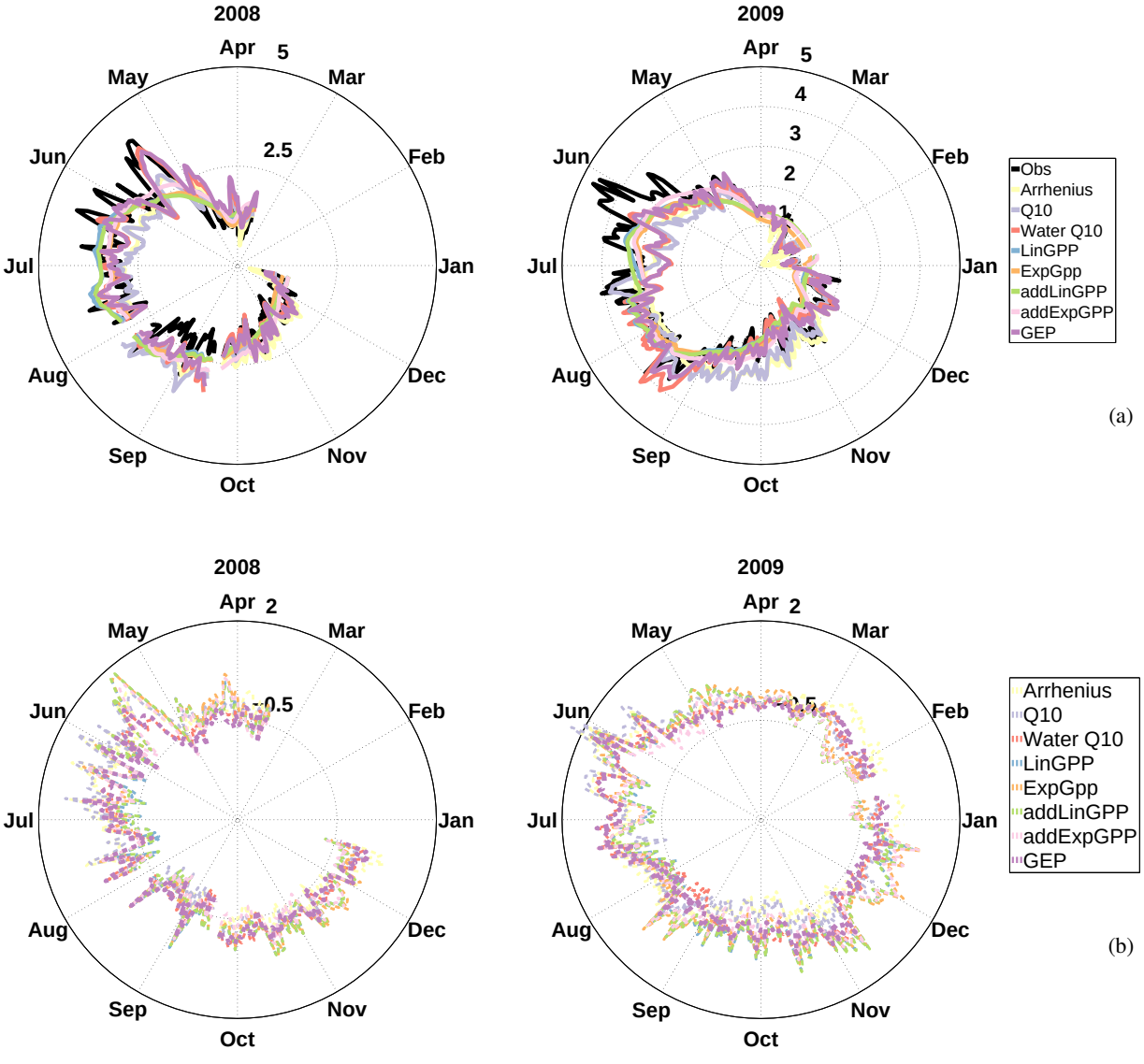


Figure 12. Daily R_{soil} fluxes illustrated in the context of the two studied years and residual values of the total soil daily CO_2 outgoing fluxes as simulated by the investigated literature models and the GEP emerged model. The fluxes shown here are the real flux measured at the site and the predicted fluxes generated according to the GEP model and some of the models used in the environmental science community. The scale of the fluxes is given in gC/day/m^2.