SymReg

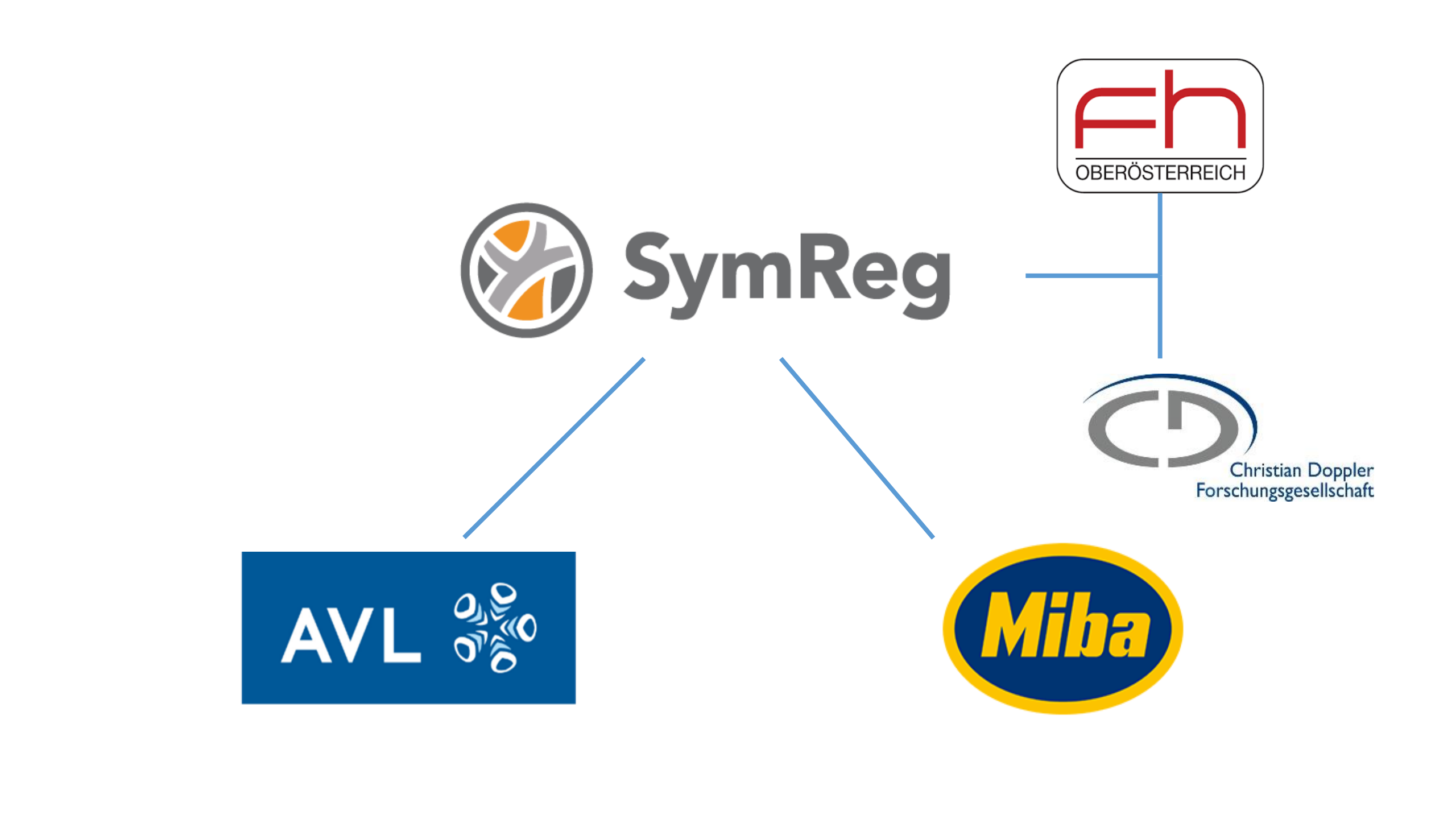
JOSEF RESSEL CENTER FOR
SYMBOLIC REGRESSION

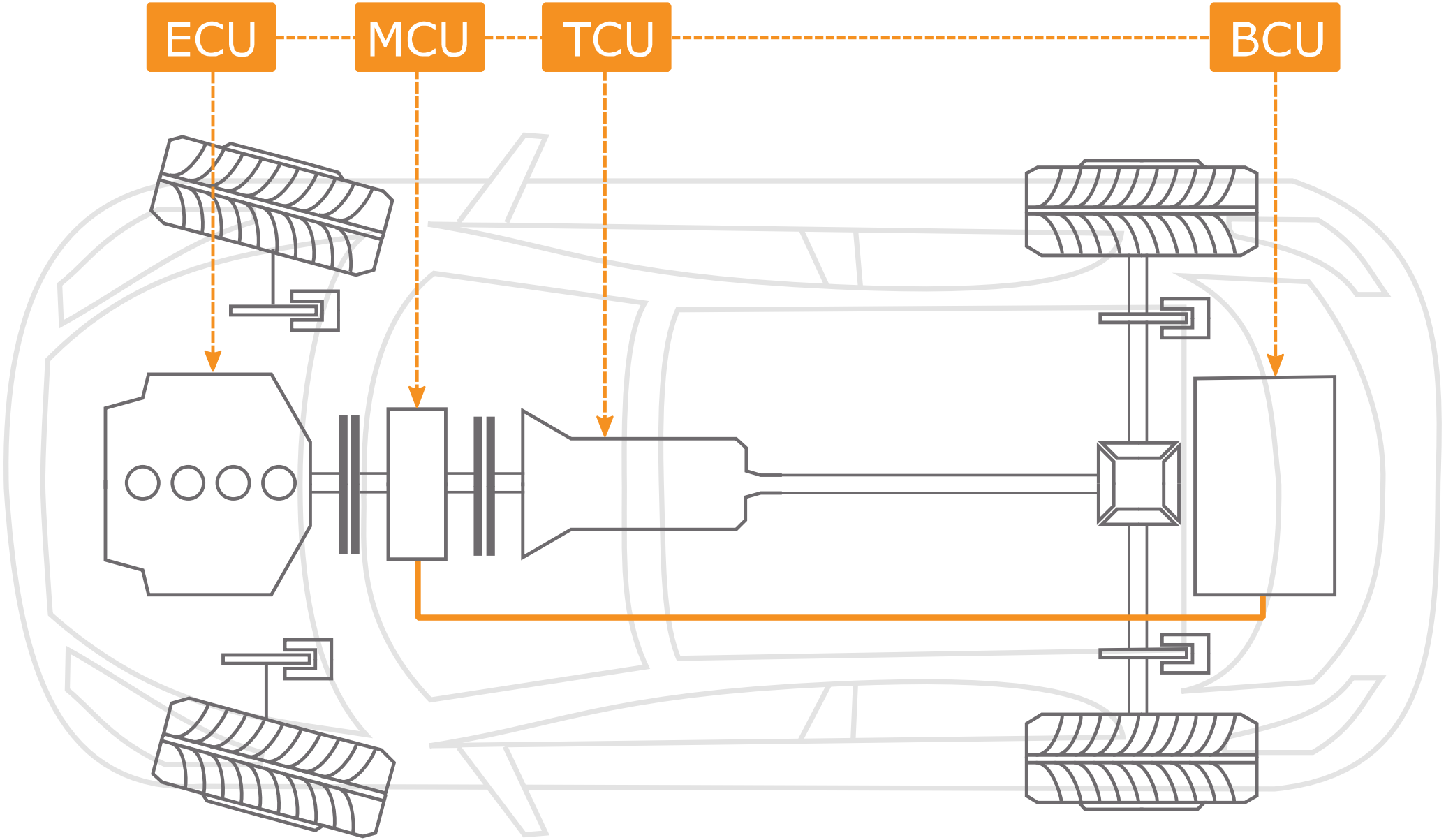
Real-world Applications of Symbolic Regression

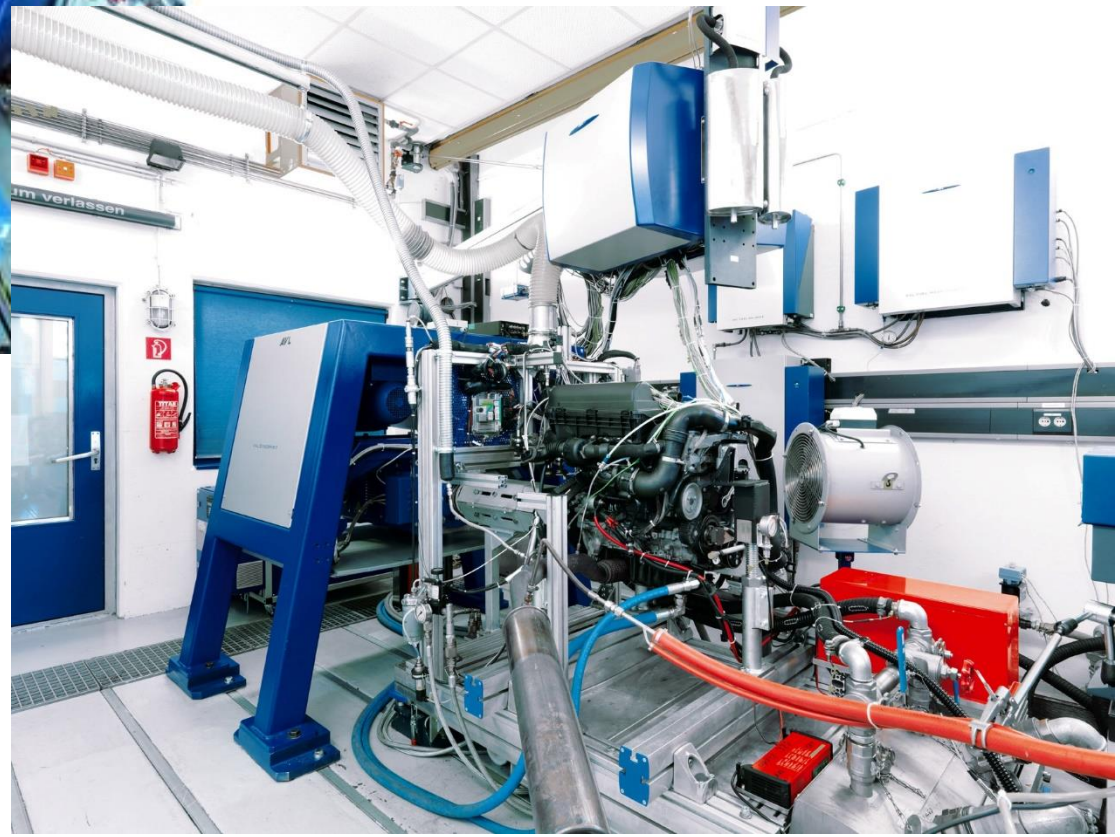
Gabriel Kronberger
University of Applied Sciences Upper Austria

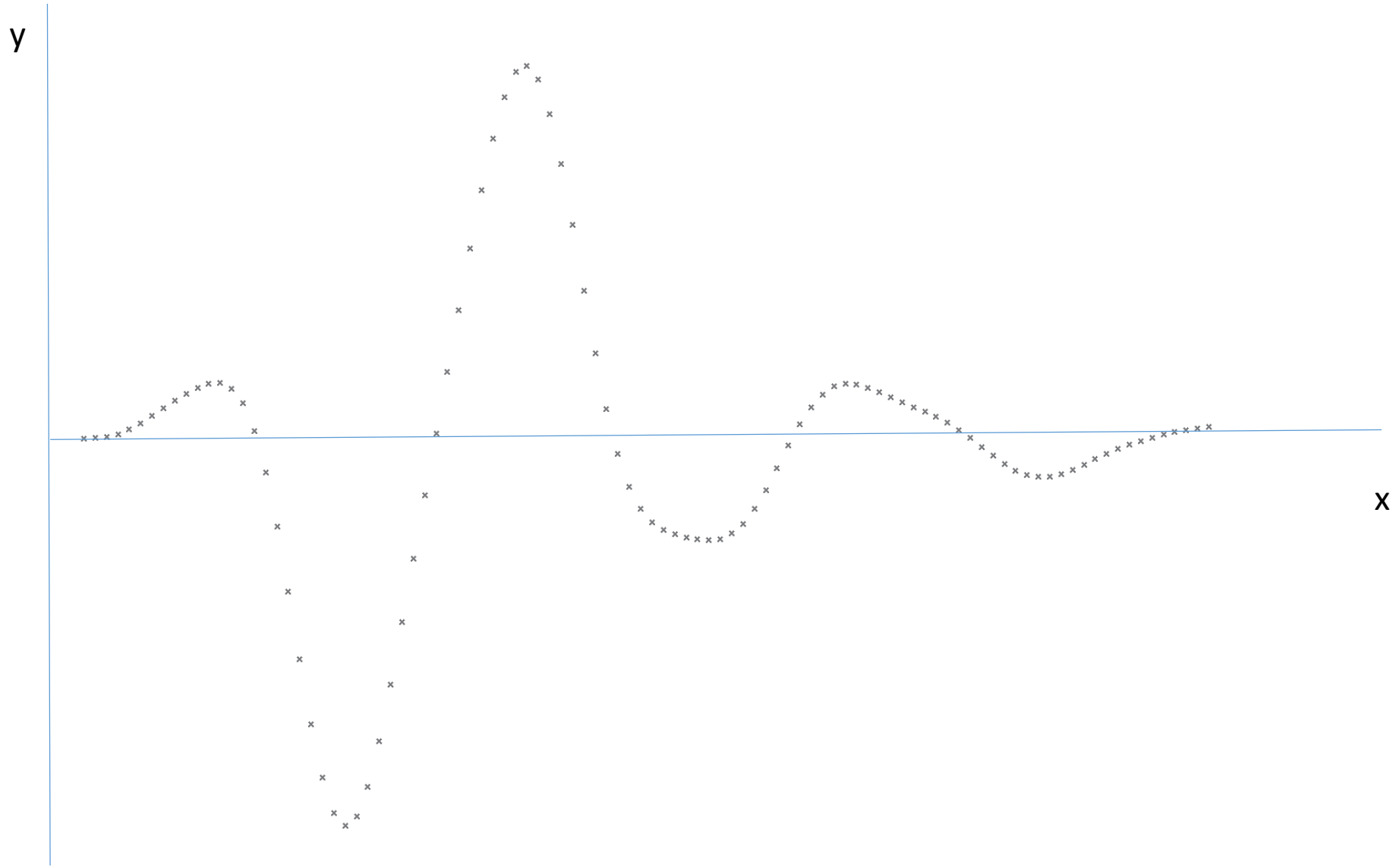

Ireland
Liverpool
London
Hamburg
Amsterdam
Netherlands
Berlin
Poland
Belarus
Воронеж
Sign in
Kharkiv
Харків
Волгоград
Kyiv
Київ
Brussels
Cologne
Germany
Belgium
Frankfurt
Prague
Rostov-on-Don
Ростов-на-Дону
Ukraine
Luxembourg
Paris
University of Applied Sciences Upper Austria...
Krasnodar
Краснодар
Moldova
Chisinau
Odesa
Одеса
Munich
Budapest
Austria
Hungary
France
Switzerland
Romania
Slovenia
Zagreb
Belgrade
Београд
Bucharest
Bay of Biscay
Black Sea
Milan
Croatia
Bosnia and Herzegovina
Serbia
Sarajevo
Monaco
Porto
Andorra
Italy
Podgorica
Подгорица
Kosovo
Bulgaria
Istanbul
Ankara
Barcelona
Macedonia (FYROM)
Portugal
Madrid
Rome
Tirana
Bursa
Turkey
Albania
Spain
Valencia
Tyrrhenian Sea
Adana
Izmir
Greece
Athens
Αθήνα
Antalya
Seville
Málaga
Gibraltar
Algiers
Cyprus
Lebanon
Google

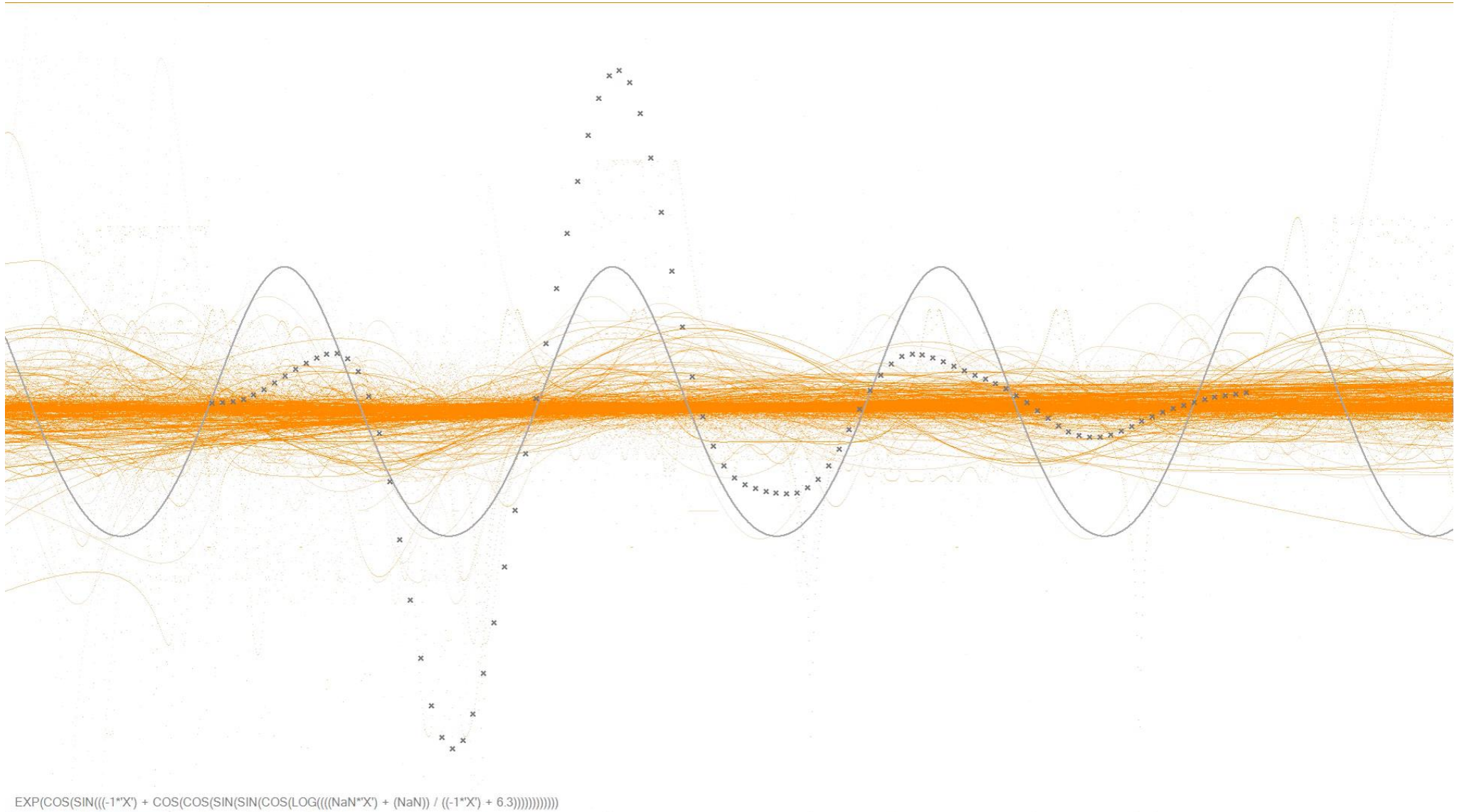

University of Applied Sciences Upper Austria...
Linz
Prague
Czechia
Austria
Slovakia
Nuremberg
Munich
Vienna
Bratislava
Brno
Ostrava
Katowice
Kraków
Pilsen
Karlovy Vary
Hradec Kralove
Pardubice
Ceske Budejovice
Český Krumlov
Regensburg
Straubing
Ingolstadt
Passau
Stuttgart
Ulm
Augsburg
Würzburg
Bamberg
Bayreuth
Wels
Steyr
St Pölten
Salzburg
Rosenheim
Kempten
Konstanz
Garmisch-Partenkirchen
Innsbruck
Liechtenstein
Gallen
Wiener Neustadt
Sopron
Győr
Budapest
Trnava
Nitra
Zlin
Olomouc
Bielsko-Biala
Wieliczka
Zakopane
Národný park Nízke Tatry
Eger
Vác
Szolnok
Google


FH
OBERÖSTERREICH
SymReg
Christian Doppler
Forschungsgesellschaft
AVL
Miba


ECU
MCU
TCU
BCU



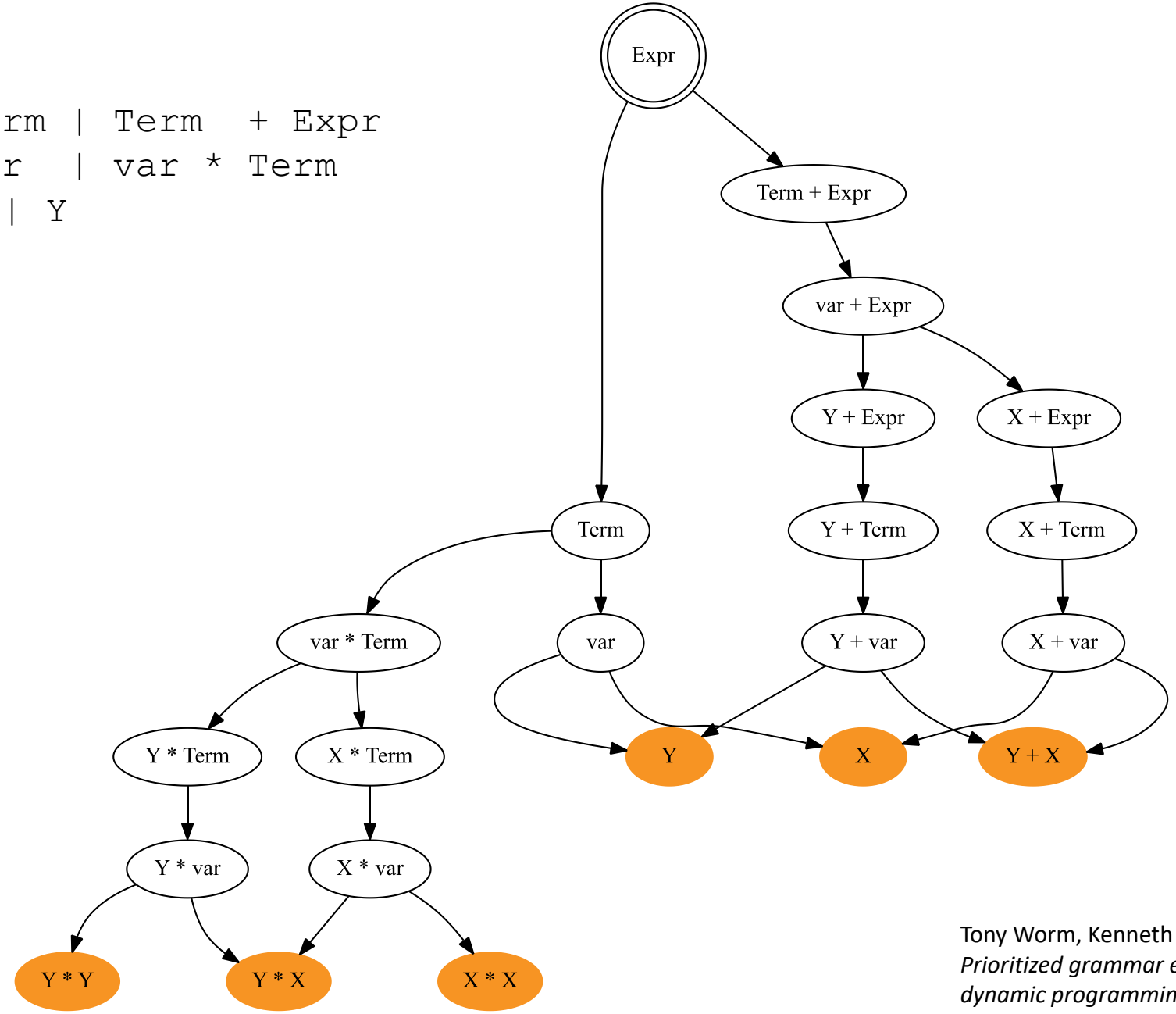

y
x


EXP(COS(SIN(((-1*'X') + COS(COS(SIN(SIN(COS(LOG((((NaN*'X') + (NaN)) / ((-1*'X') + 6.3)))))))))))

$$y = \left(\frac{\log\left(\left(\log\left(\left(c_0 \cdot \mathrm{X} + \cos\left(c_1 \cdot \mathrm{X}\right) + c_2\right)\right) + \sin\left(\cos\left(\left(c_3 \cdot \mathrm{X} + \sin\left(c_4 \cdot \mathrm{X}\right)\right)\right) \cdot c_5\right)\right) \cdot \left(c_6 \cdot \mathrm{X} + c_7\right) \cdot \sin\left(c_8 \cdot \mathrm{X}\right) \cdot c_9}{\left(\exp\left(\mathrm{X}\right) + \frac{1}{c_{11} \cdot \mathrm{X}} + c_{12}\right)} + c_{13} \right) + \epsilon$$

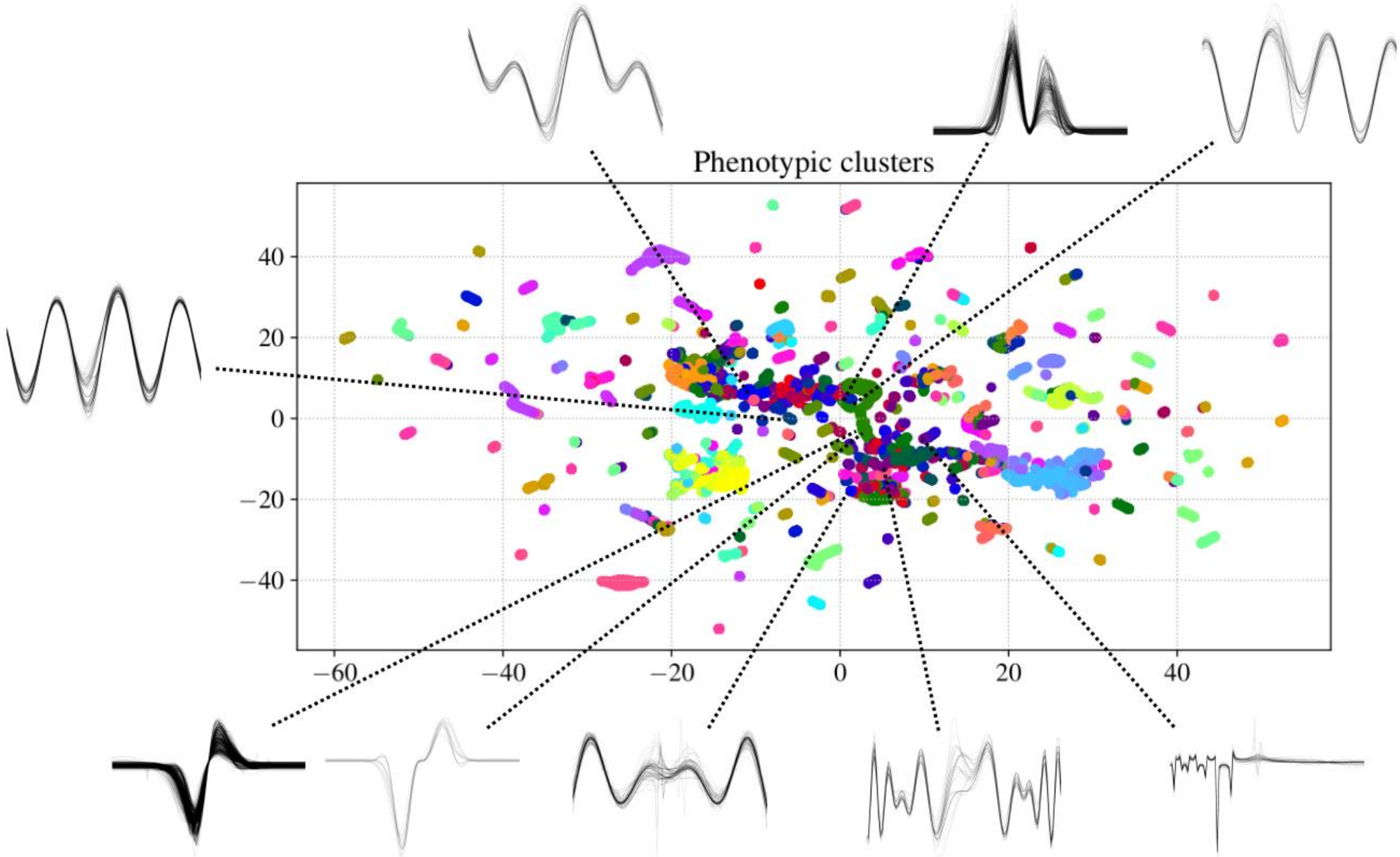
$$\begin{aligned}
c_0 &= -1.0 \\
c_1 &= 2.0 \\
c_2 &= -8.8304 \\
c_3 &= -1.2639 \\
c_4 &= 1.7887 \\
c_5 &= -1.0 \\
c_6 &= -1.0 \\
c_7 &= -8.8304 \\
c_8 &= 2.0 \\
c_9 &= -2.7642 \\
c_{11} &= 0.0058959 \\
c_{12} &= 7.3565 \\
c_{13} &= 8.8304
\end{aligned}$$

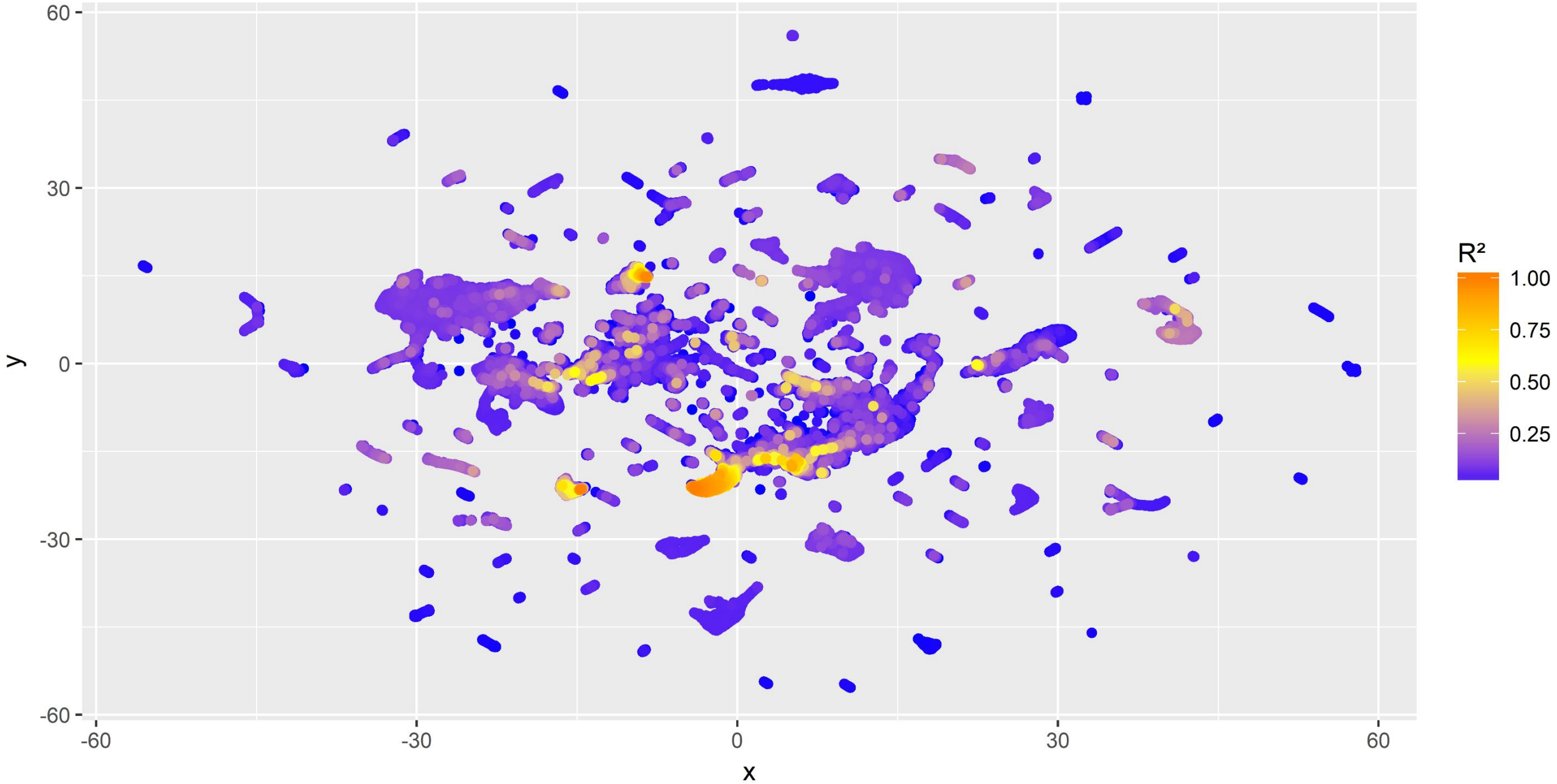
```
G(Expr):
Expr → Term | Term + Expr
Term → var | var * Term
var → X | Y
```



Tony Worm, Kenneth Chiu:
Prioritized grammar enumeration: symbolic regression by dynamic programming. GECCO 2013: 1021-1028

```
G(Expr):
Expr       -> "const" Term "*" Expr "+" | "const" Term "*" "const" "+"
Term       -> RecurringFactors Term "*" | RecurringFactors | OneTimeFactors

RecurringFactors -> VarFactor | LogFactor | ExpFactor | SinFactor
VarFactor  -> <variable>
LogFactor  -> SimpleExpr "log"
ExpFactor  -> "const" SimpleTerm "*" "exp"
SinFactor  -> SimpleExpr "sin"

OneTimeFactors -> InvFactor SqrtFactor "*" CbrtFactor "*" |
                  InvFactor SqrtFactor "*"                |
                  InvFactor               CbrtFactor "*" |
                            SqrtFactor    CbrtFactor "*" |
                  InvFactor  |
                  SqrtFactor |
                  CbrtFactor
InvFactor  -> InvExpr "inv"
SqrtFactor -> SimpleExpr "sqrt"
CbrtFactor -> SimpleExpr "cbrt"

SimpleExpr -> "const" SimpleTerm "*" SimpleExpr "+" | "const" SimpleTerm "*" "const" "+"
SimpleTerm -> VarFactor SimpleTerm "*" | VarFactor

InvExpr -> "const" InvTerm "*" InvExpr "+" | "const" InvTerm "*" "const" "+"
InvTerm -> RecurringFactors InvTerm "*" |
           RecurringFactors SqrtFactor "*" CbrtFactor "*" |
           RecurringFactors SqrtFactor "*"                |
           RecurringFactors               CbrtFactor "*" |
                            SqrtFactor    CbrtFactor "*" |
           RecurringFactors |
           SqrtFactor       |
           CbrtFactor
```

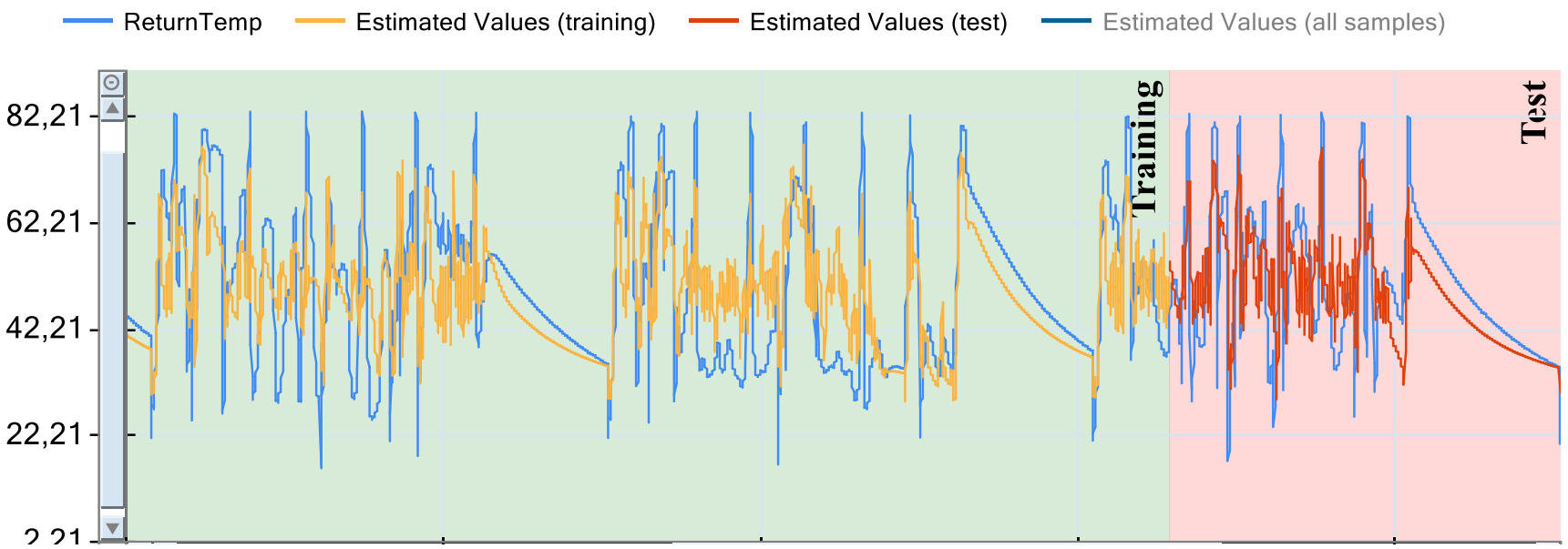

Phenotypic clusters
40
20
0
−20
−40
−60
−40
−20
0
20
40


60
30
0
-30
-60
y
-60
-30
0
30
60
x
R²
1.00
0.75
0.50
0.25

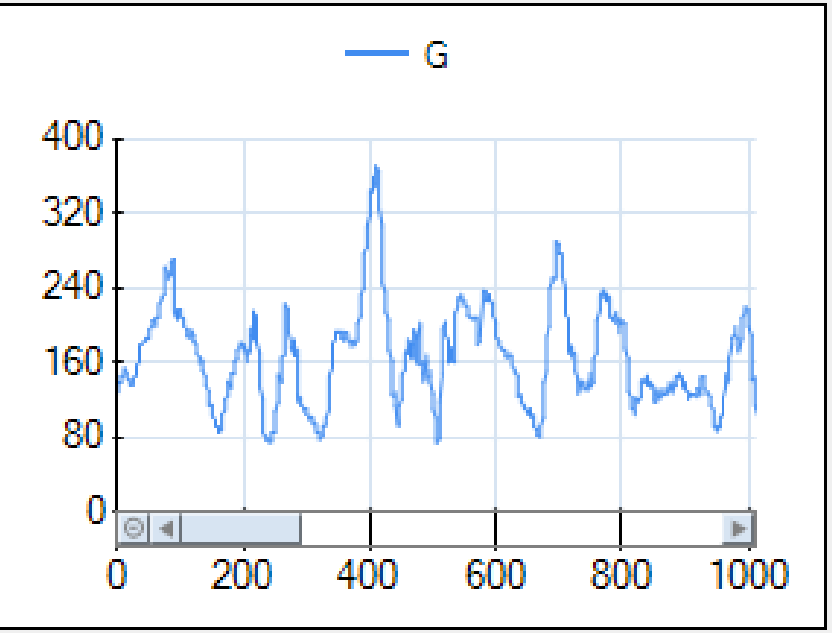
Modelling Dynamical Systems

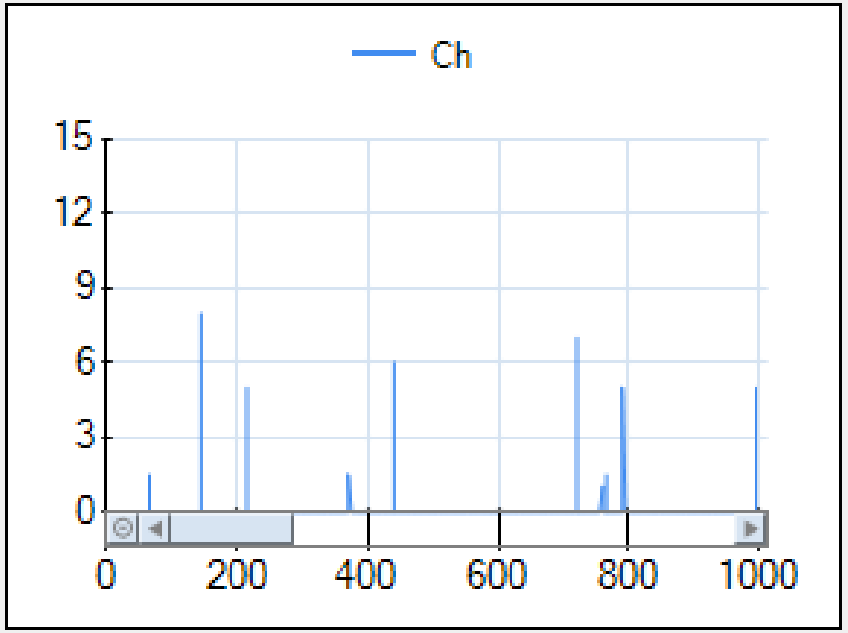
SymReg

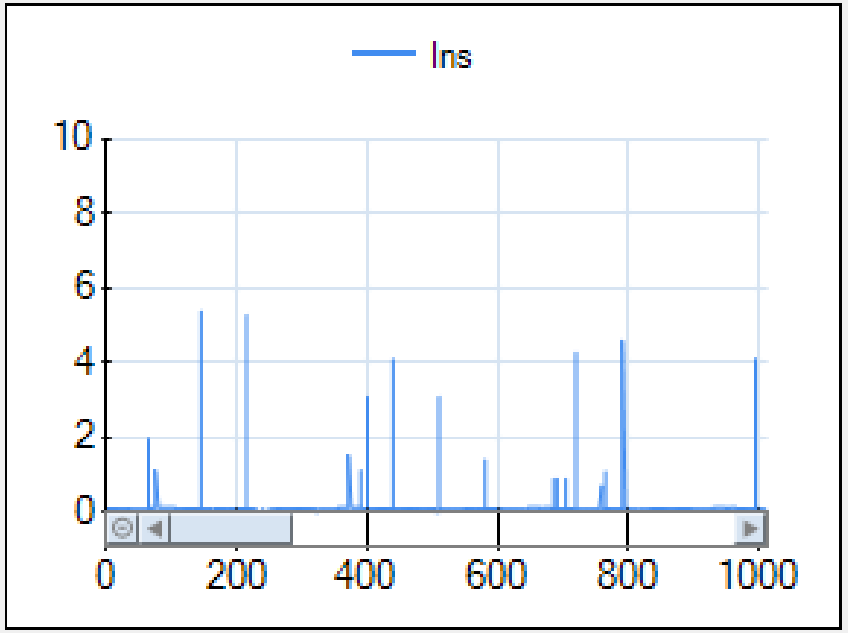
$$\frac{d\ \text{Temp}}{dt} = f_1(\text{Power}, \text{Temp}, \text{Env})$$



Blood Glucose Concentration Prediction for Artificial Pancreas







J Med Syst (2017) 41:142
DOI 10.1007/s10916-017-0788-2

PATIENT FACING SYSTEMS

Data Based Prediction of Blood Glucose Concentrations Using Evolutionary Methods

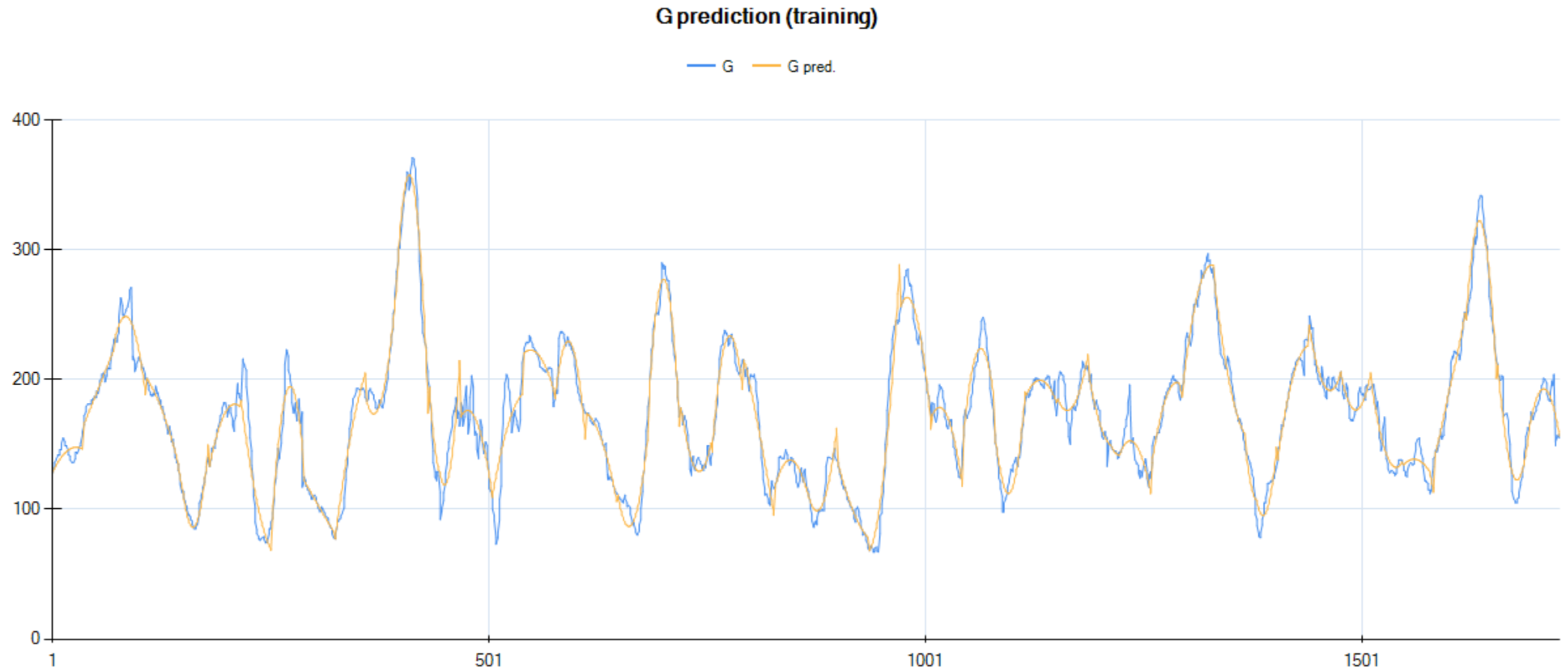
J. Ignacio Hidalgo[1] · J. Manuel Colmenar[2] · Gabriel Kronberger[3] · Stephan M. Winkler[3] · Oscar Garnica[1] · Juan Lanchares[1]

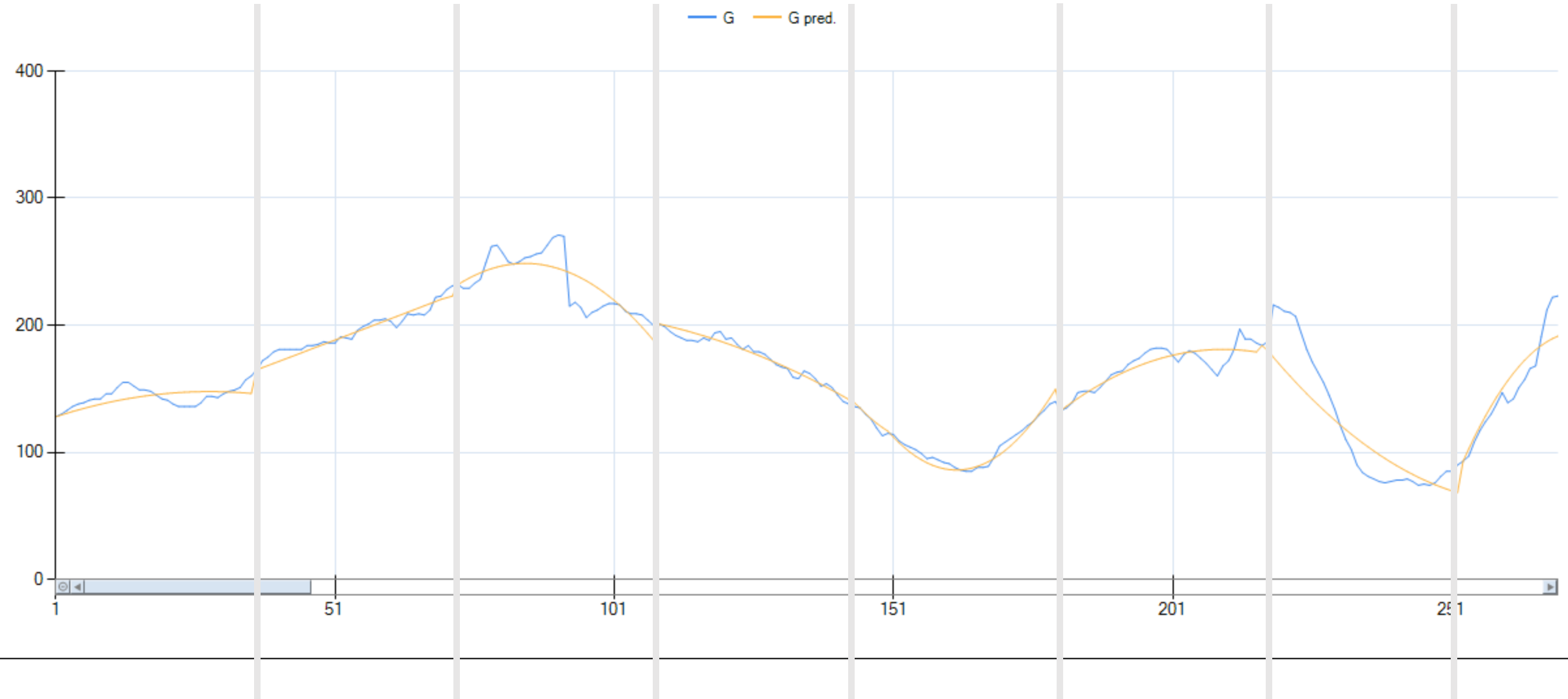
Received: 8 May 2017 / Accepted: 20 July 2017

Abstract Predicting glucose values on the basis of insulin and food intakes is a difficult task that people with diabetes need to do daily. This is necessary as it is important to maintain glucose levels at appropriate values to avoid not only short-term, but also long-term complications of the illness. Artificial intelligence in general and machine learning techniques in particular have already lead to promising results in modeling and predicting glucose concentrations. In this modeling and prediction of glucose concentrations using as inputs the values measured by a continuous monitoring glucose system as well as also previous and estimated future carbohydrate intakes and insulin injections. In particular, we use the following four techniques: genetic programming, random forests, k-nearest neighbors, and grammatical evolution. We propose two new enhanced modeling algorithms for glucose prediction, namely (i) a variant of grammati-

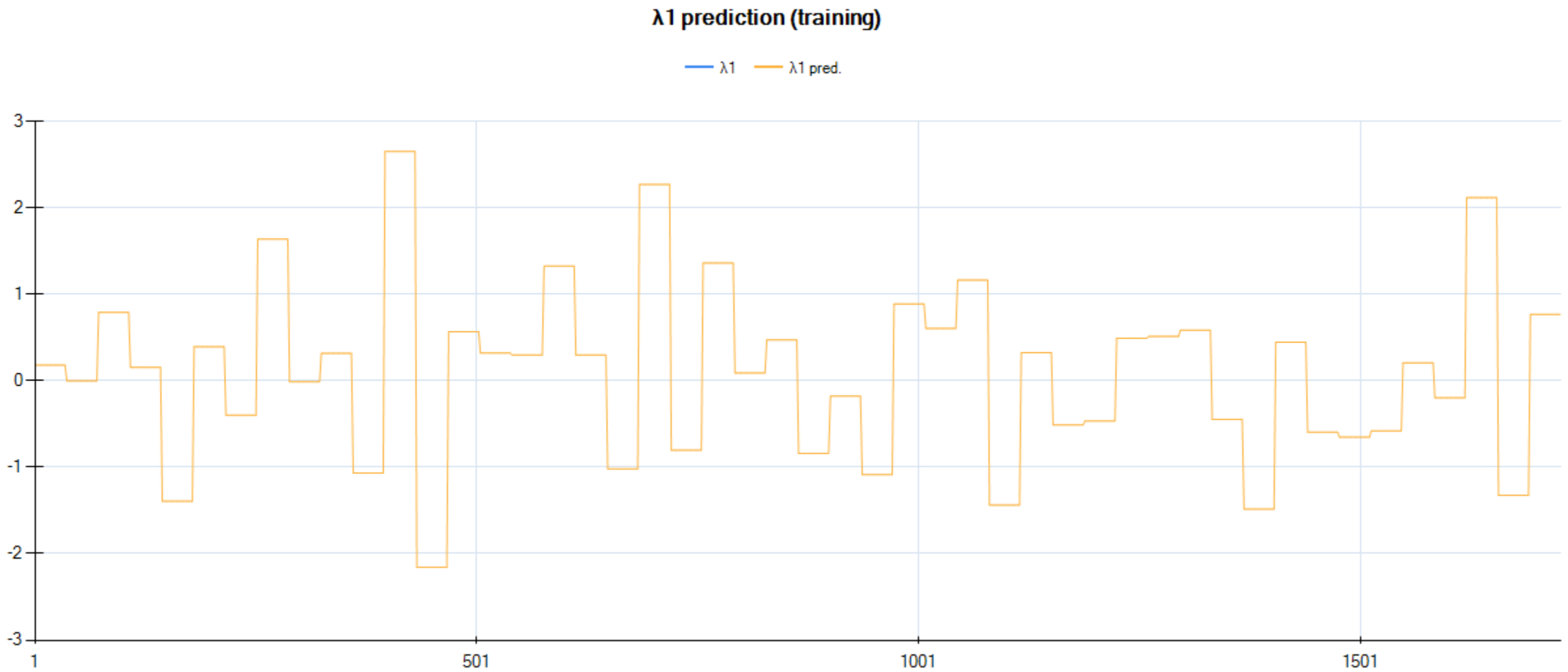
Our goal is to develop the following models:

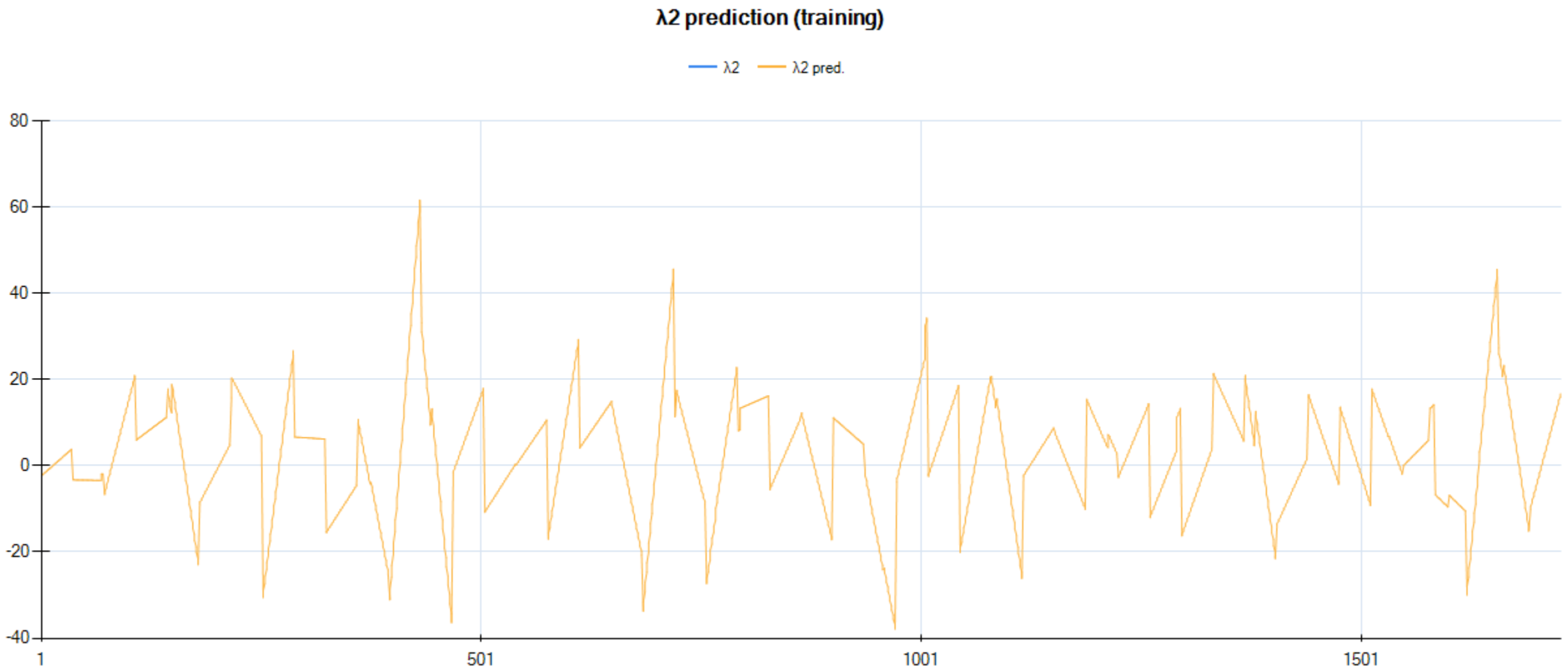
$$
\begin{aligned}
\hat{G}_{t+30} &= f_{t+30}(G_{t+i}, i \in (-240 \ldots 0)), \\
&\quad (I_{t+i}, C_{t+i}), i \in (-240 \ldots +30)) \\
\hat{G}_{t+60} &= f_{t+60}((G_{t+i}, i \in (-240 \ldots 0)), \\
&\quad (I_{t+i}, C_{t+i}, i \in (-240 \ldots +60)) \\
\hat{G}_{t+90} &= f_{t+90}((G_{t+i}, i \in (-240 \ldots 0)), \\
&\quad (I_{t+i}, C_{t+i}, i \in (-240 \ldots +90)) \\
\hat{G}_{t+120} &= f_{t+120}((G_{t+i}, i \in (-240 \ldots 0)), \\
&\quad (I_{t+i}, C_{t+i}, i \in (-240 \ldots +120))
\end{aligned}
$$


G prediction (training)
G
G pred.
400
300
200
100
0
1
501
1001
1501

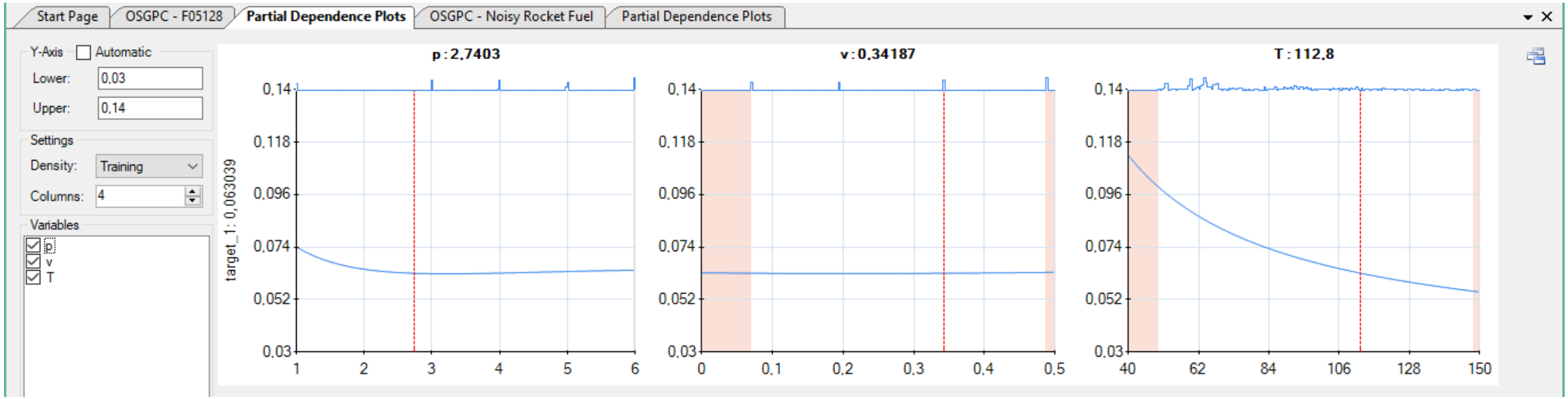

G prediction (training)
G
G pred.
400
300
200
100
0
1
51
101
151
201

$$\dot{G} = -0.29\lambda_2 + 0.7$$
$$\dot{\lambda_1} = 0$$
$$\dot{\lambda_2} = \lambda_1 + ch$$

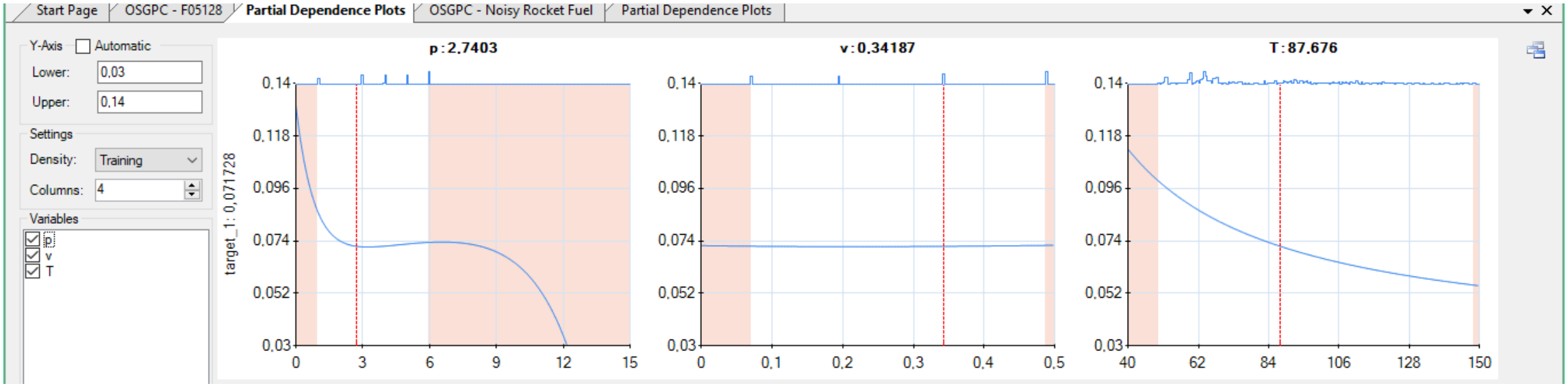

λ1 prediction (training)
λ1
λ1 pred.
3
2
1
0
-1
-2
-3
1
501
1001
1501


λ2 prediction (training)
λ2
λ2 pred.
80
60
40
20
0
-20
-40
1
501
1001
1501

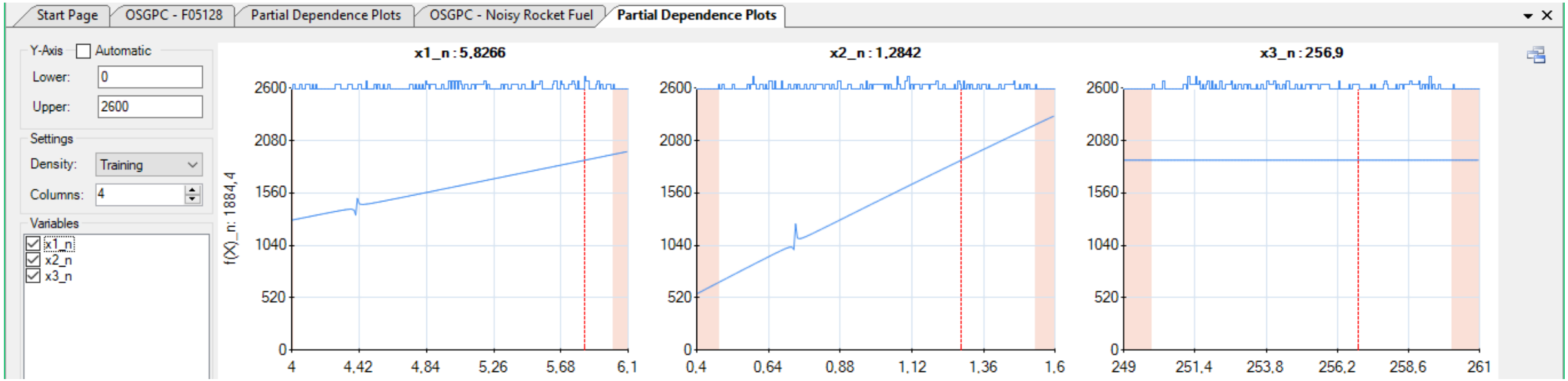
Interpolation and Extrapolation



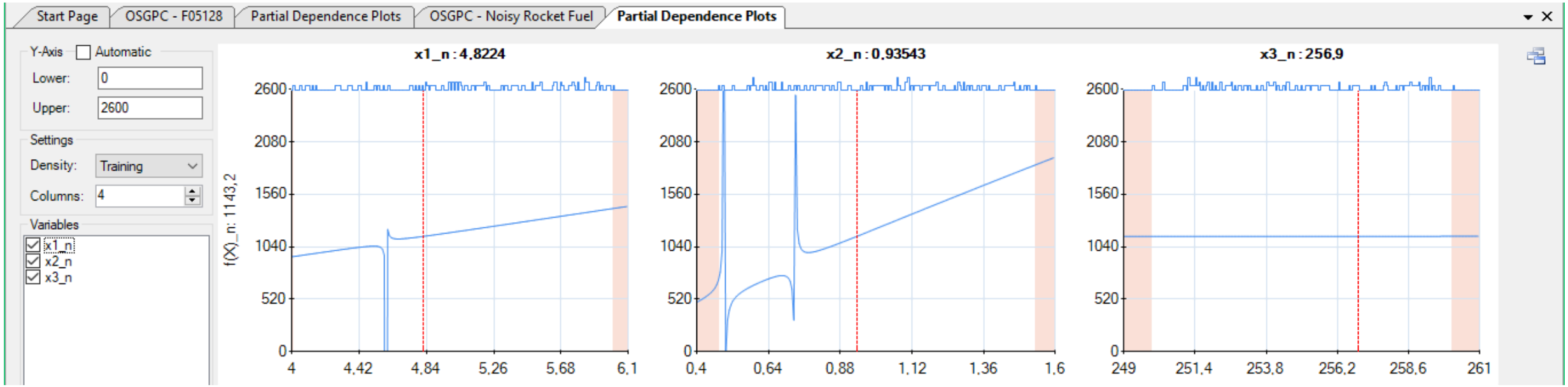
Interpolation and Extrapolation



Interpolation and Extrapolation



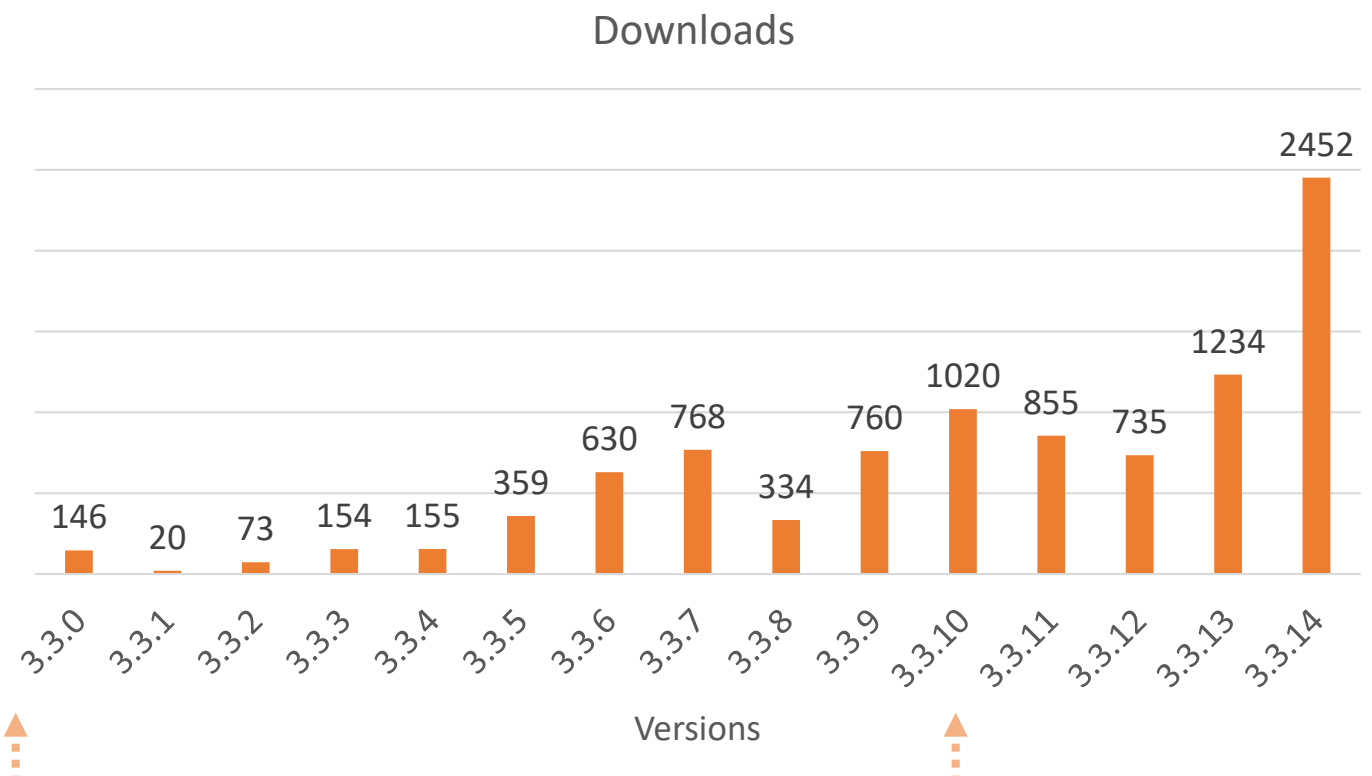
Interpolation and Extrapolation



SymReg

JOSEF RESSEL CENTER FOR SYMBOLIC REGRESSION

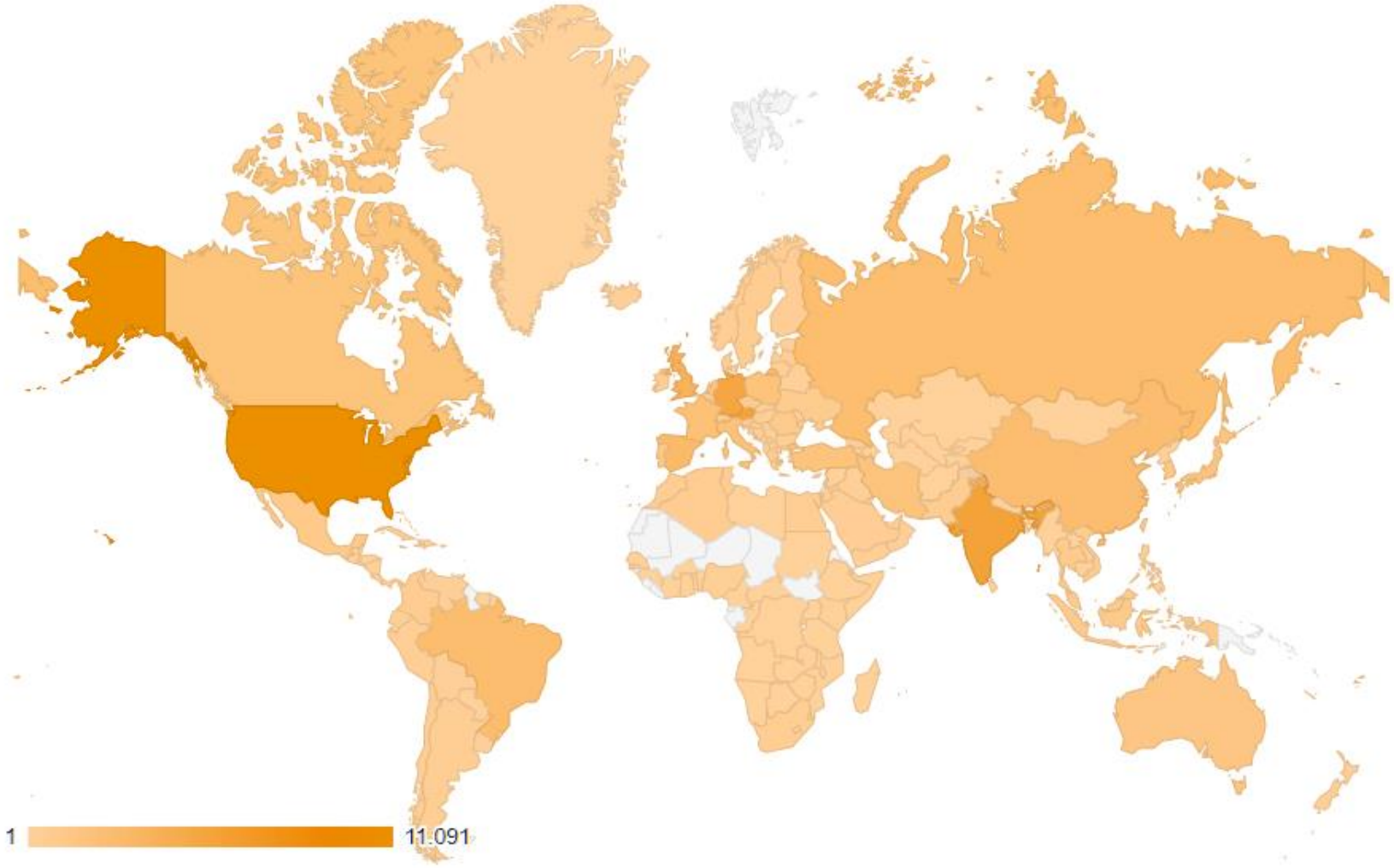
HeuristicLab



Josef Ressel-Zentrum HEUREKA!

Heuristic Optimization in Production and Logistics

SymReg
JOSEF RESSEL CENTER FOR SYMBOLIC REGRESSION



HeuristicLab

County	Website Users (dev.heuristiclab.com)
USA	11091
Austria	5766
India	5678
Germany	5096
UK	3743
Spain	2809
Russia	2645
Brazil	2607
China	2575
France	2099

Feb. 2018