Emulating tropospheric chemistry with physics-informed machine learning

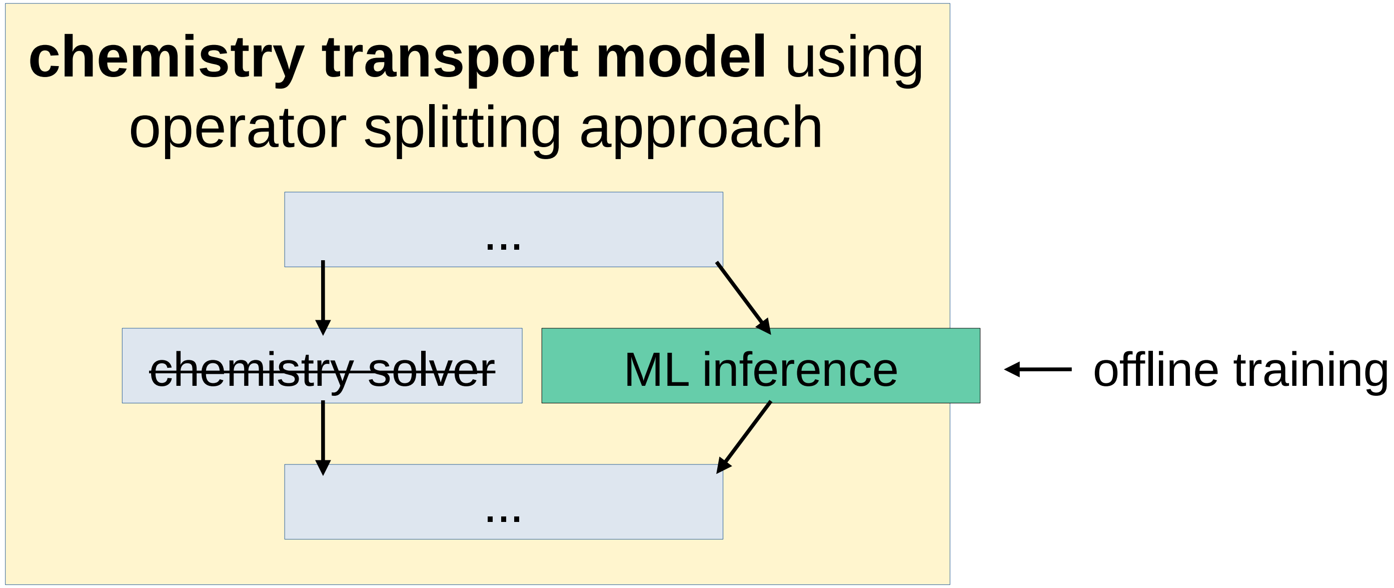
Alessio Melli*, Camille Mouchel-Vallon*, Isidre Mas Magre*, Hervé Petetin*, and Oriol Jorba Casellas*

** Barcelona Supercomputing Center, Barcelona (Spain)*

EGU25-5628

Vienna, 1st May 2025 – AS3.28 – Air Pollution Modeling

Goals of this study



- Implement baseline ML model in CTM: *fast inference*
- Gather insight from the use of smallest black box

aiming to retain physical consistency, and *possibly* interpretability

Case studies

toy mechanism

$$\mathrm{A} + \mathrm{B} \xrightarrow{k_1} \mathrm{C} \qquad k_1 = 2.0 \times 10^{-19}\,\mathrm{m^3 molec^{-1} s^{-1}}$$

$$\mathrm{C} \xrightarrow{k_2} 2\,\mathrm{B} \qquad k_2 = 1.0 \times 10^{-4}\,\mathrm{s^{-1}}$$

POLLU mechanism **20 species, 25 reactions** [1]

$$\mathrm{NO} + \mathrm{O_3} \xrightarrow{k_{02}} \mathrm{NO_2} \qquad k_{02} = 1.8 \times 10^{-14}\,\mathrm{m^3 molec^{-1} s^{-1}}$$

$$\vdots$$

$$\mathrm{HCHO} + \mathrm{OH} \xrightarrow{k_{06}} \mathrm{HO_2} + \mathrm{CO} \qquad k_{06} = 1.0 \times 10^{-11}\,\mathrm{m^3 molec^{-1} s^{-1}}$$

stiff system
no *T* dep.
no photolysis

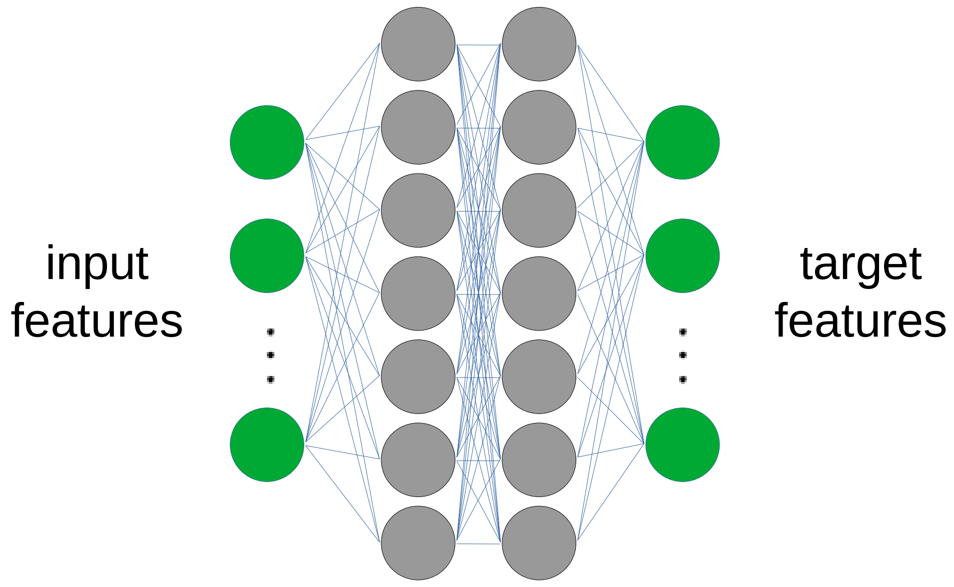
...complex chemistry will follow
e.g., Carbon bond 5 (59 species, 156 reactions) [2]

[1] Verwer, Jan G. *SIAM Journal on Scientific Computing* 15.5 (1994): 1243-1250
[2] Yarwood, Greg, *et al*. *Final report to the US EPA*, RT-0400675 8 (2005): 13

Baseline model

Simple neural network:

fully connected multilayer perceptron



small **hyperparameter tuning**
on limited search space*

↓

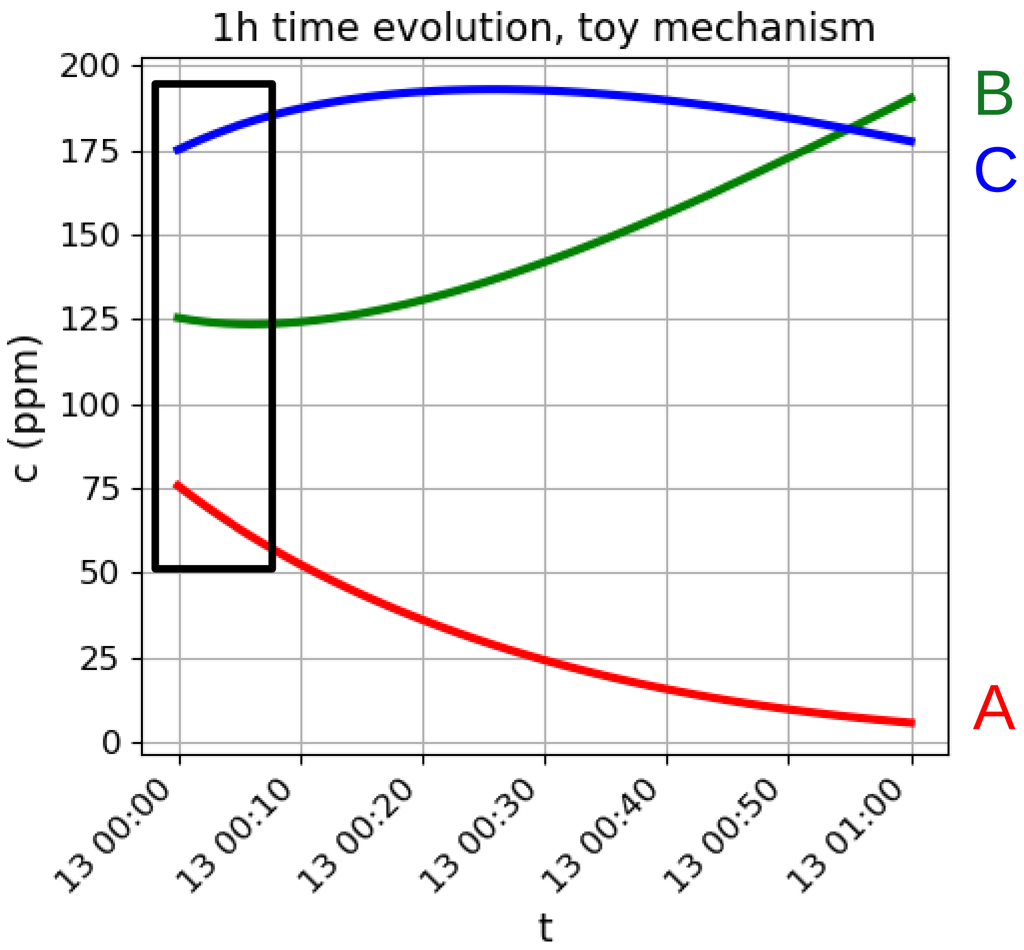
identify best configuration
using 80% of dataset

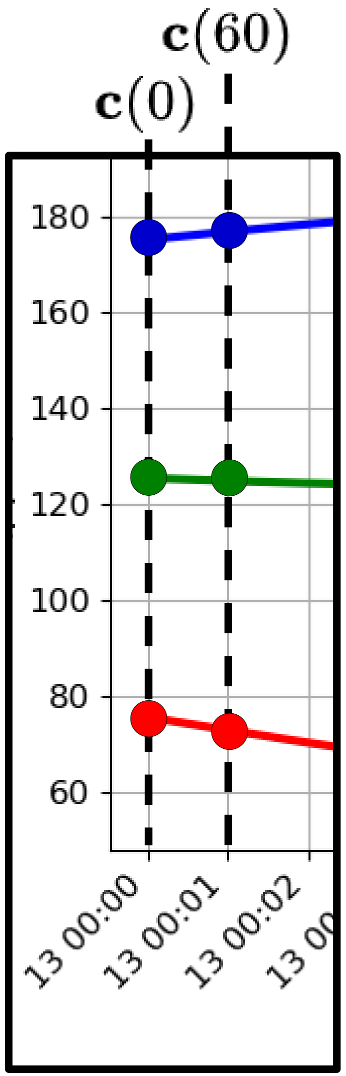
↓

test the tuned model on 20%

*: optimizer hyperparameters
neurons per layer
hidden layers

Data generation





$$\Delta c_s = c_s(60) - c_s(0)$$
$$\Delta t = 1\,\mathrm{min}$$

$$\boxed{N \sim 1 \times 10^6}$$

dataset design choice:

predict c

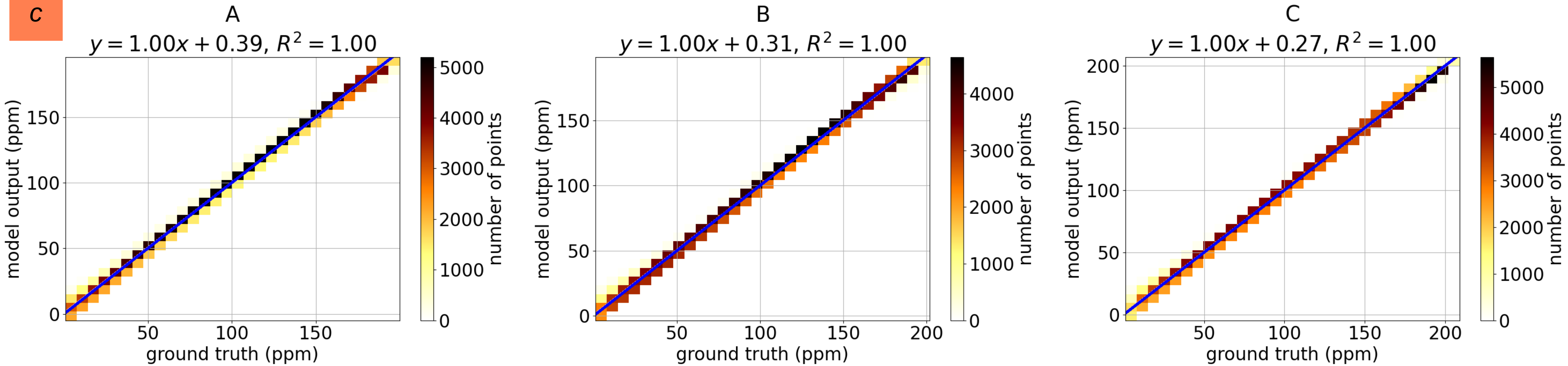
predict Δc

Baseline model: performance

x: ground truth
y: model output
1:1 line
linear regression

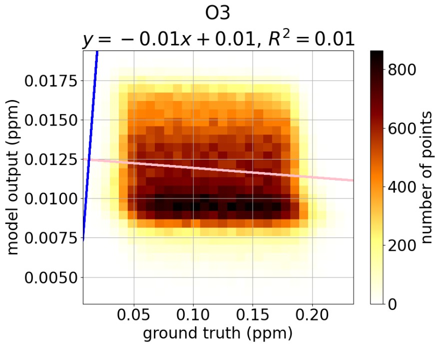
toy mechanism test set

c



~1 million parameters to capture the system

POLLU mechanism
no performance

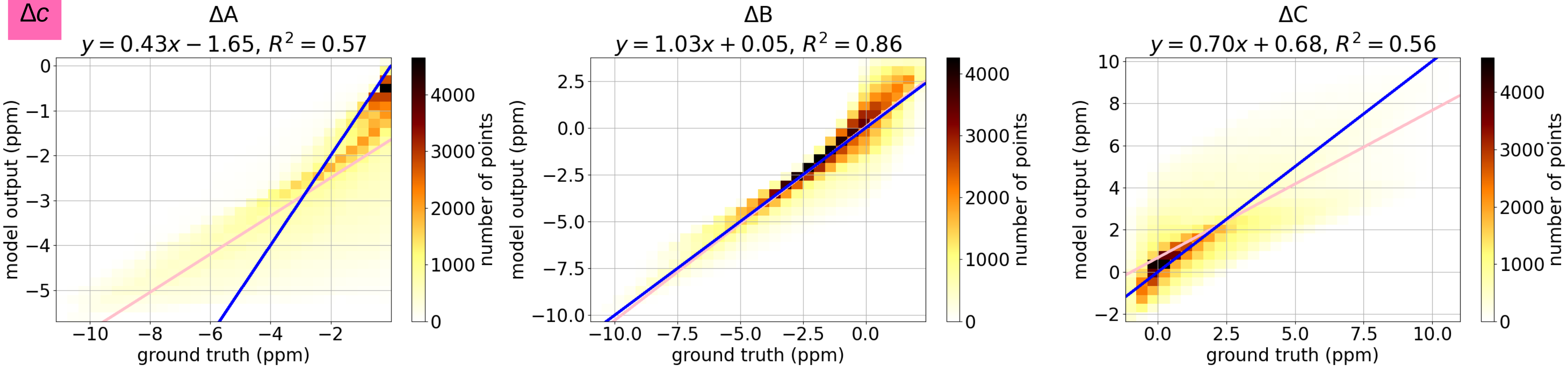


Baseline model: performance

x: ground truth
y: model output
1:1 line
linear regression

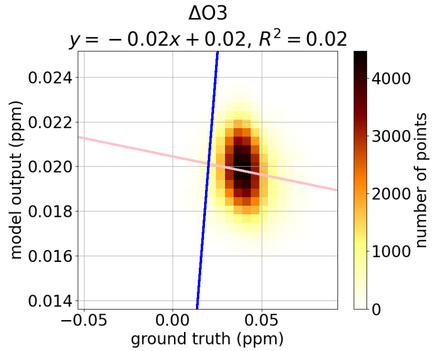
toy mechanism test set

Δc



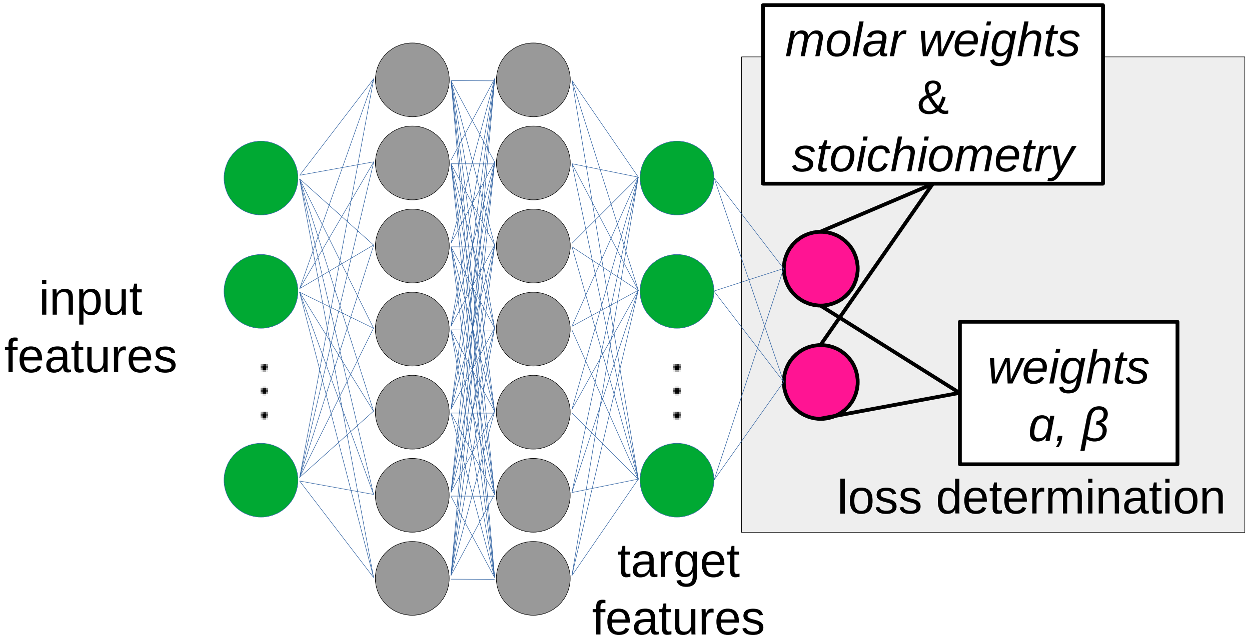
model struggles for low concentration changes: customizing the loss function?

POLLU mechanism
no performance



Baseline model + soft constraints

1. penalty for **weighted sum of concentration changes ≠ 0**
2. penalty for **final concentration values < 0**



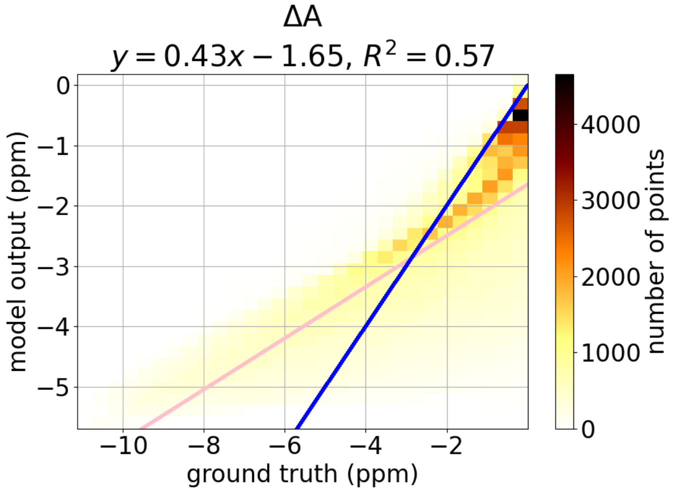
search space:
optimizer hyperparameters
constraints' weights

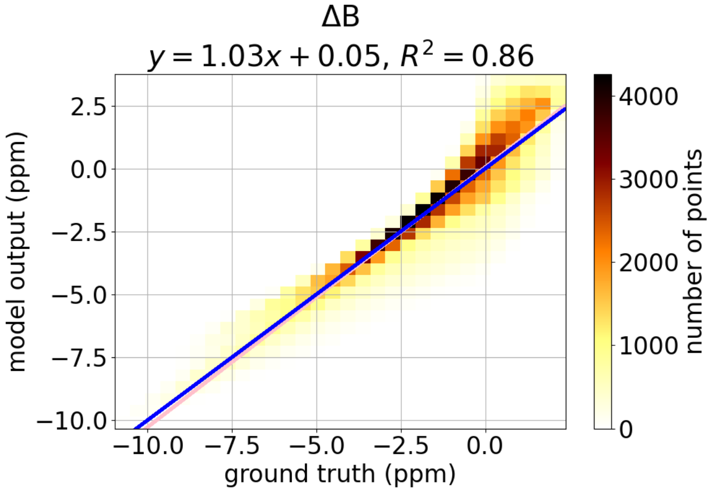
Soft constraints: performance

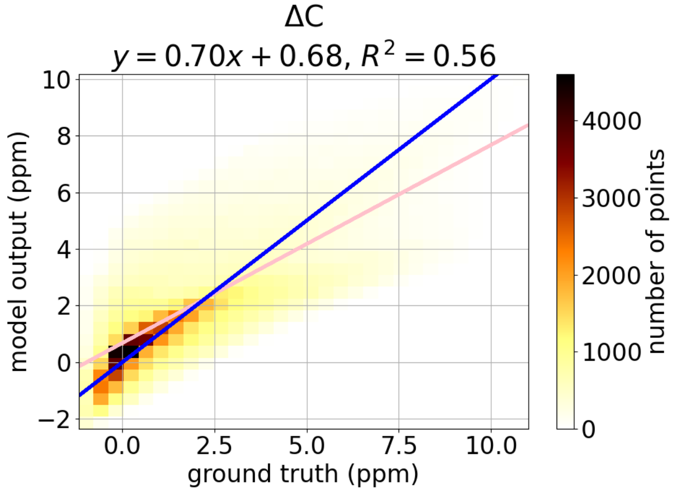
x: ground truth
y: model output
1:1 line
linear regression

Δc

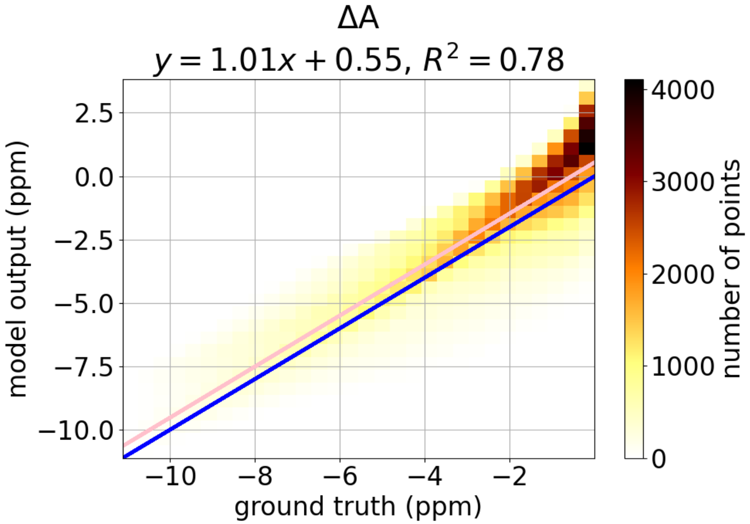
toy mechanism test set

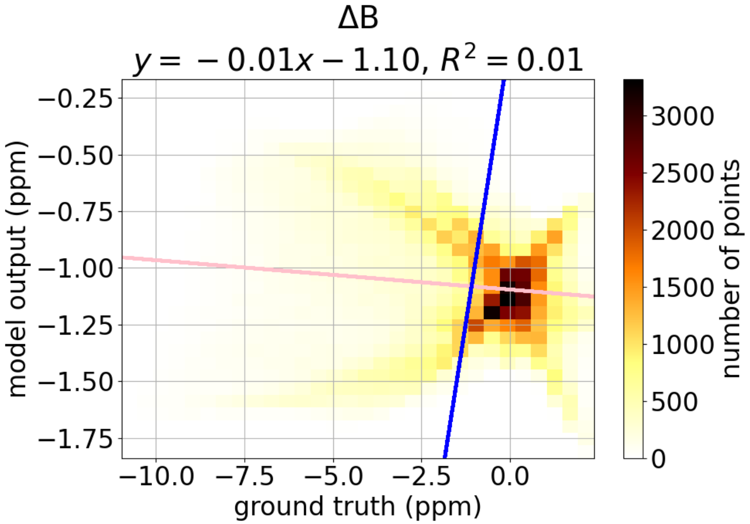


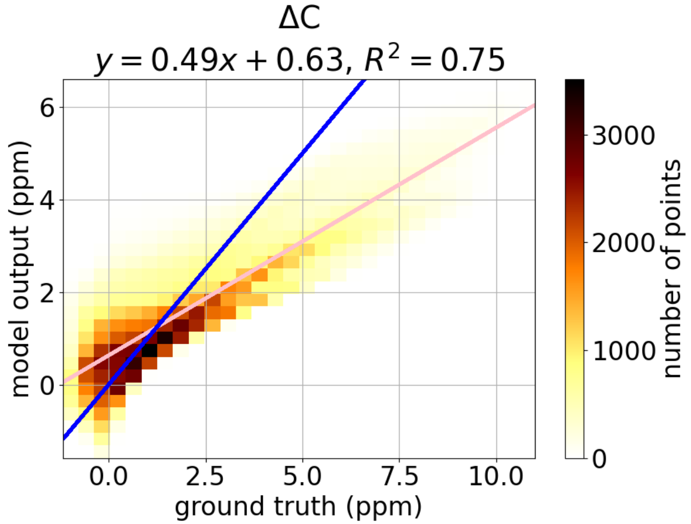




Δc *with soft constraints*







Thanks for your kind attention

Next steps

Significant data reduction: optimize dataset generation

Data pre-processing: how to handle properly different concentration ranges

Generalizability: assess out-of-distribution performance

EACH: This research has been supported by the Grant PCI2024-155075-2 funded by MICIU/AEI /10.13039/501100011033 and Cofunded by the European Union (EACH Project).

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Cofinanciado por la Unión Europea

AGENCIA ESTATAL DE INVESTIGACIÓN

Funded by the European Union NextGenerationEU

GOBIERNO DE ESPAÑA VICEPRESIDENCIA TERCERA DEL GOBIERNO MINISTERIO PARA LA TRANSICIÓN ECOLÓGICA Y EL RETO DEMOGRÁFICO

Plan de Recuperación, Transformación y Resiliencia

AEMet Agencia Estatal de Meteorología

BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación

RESPIRE: We acknowledge the RESPIRE project, which is part of the Recovery, Transformation and Resilience Plan (Plan de Recuperación, Transformación y Resiliencia, PRTyR) funded by the European Commission to repair the harm caused by the Covid-19 crisis.