CENTRE EUROPÉEN DE RECHERCHE ET DE FORMATION AVANCÉE EN CALCUL SCIENTIFIQUE

Accounting for model error in air quality forecasts: an application of 4DEnVar to the assimilation of atmospheric composition

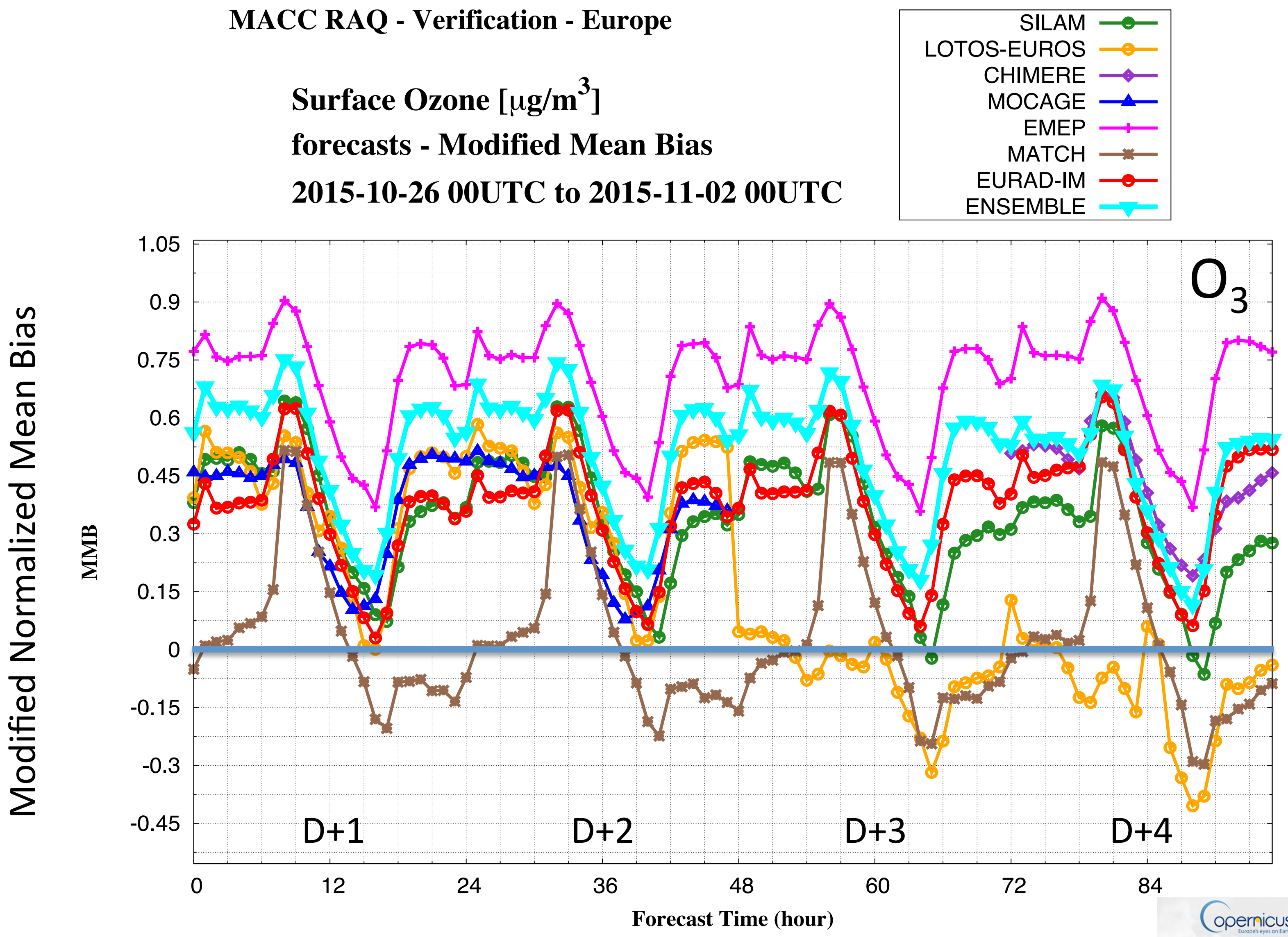
E. Emili, S. Gürol, D. Cariolle

Acknowledgments: G. Desroziers, P. Moinat, Y. Trémolet, M. Fisher

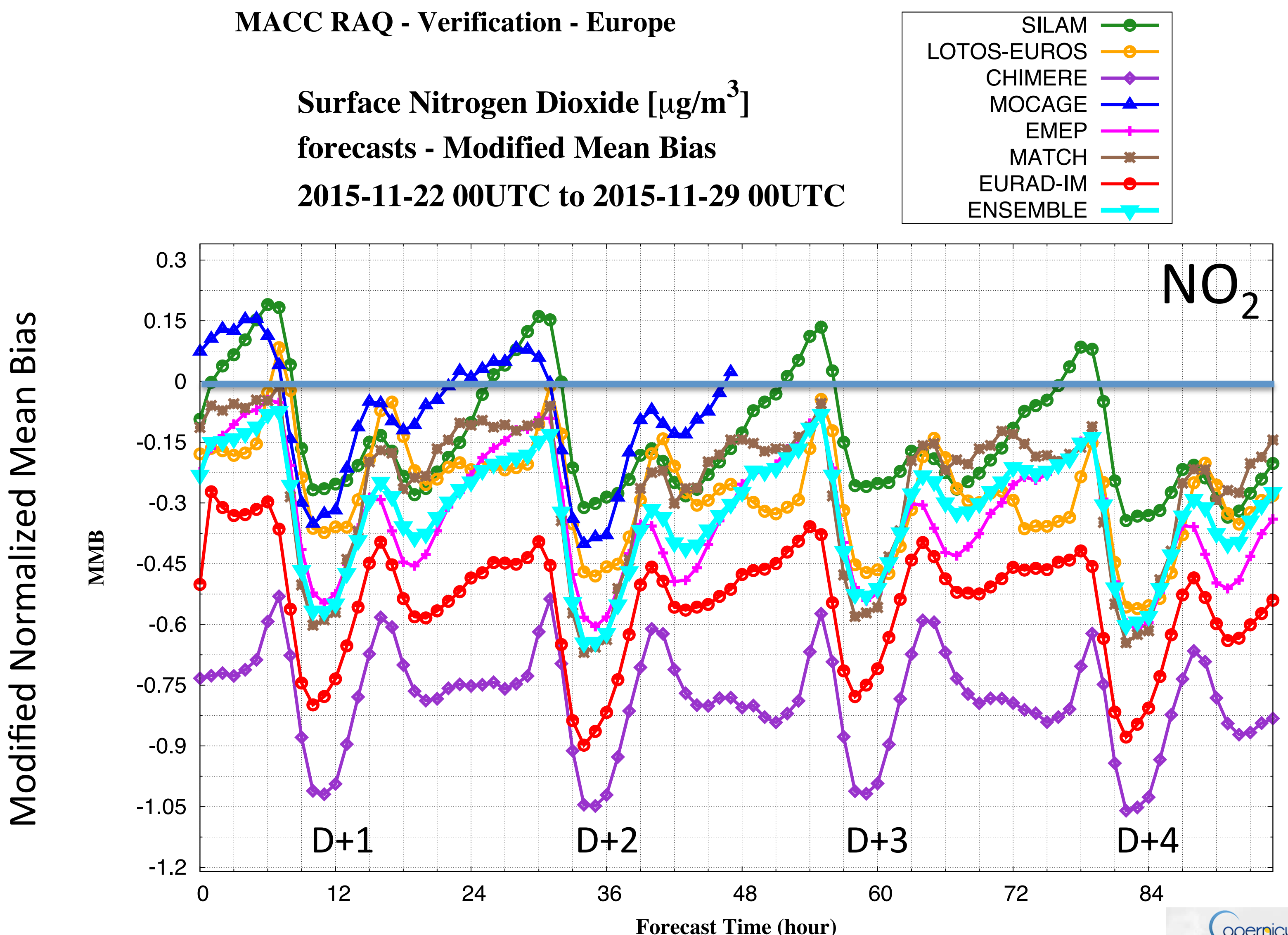
Outline

- Why model error is crucial for chemical forecasts
- QG-Chem: a new toy model for chemical DA
- DA experiments with a perfect model: 4DEnVar versus 3D-Var
- DA experiments with model errors (4DEnVar)
- Bias estimation and forecast bias correction (4DEnVar)
- Conclusions and perspectives

Forecasts biases in operational models



Forecasts biases in operational models



Which data assimilation algorithm?

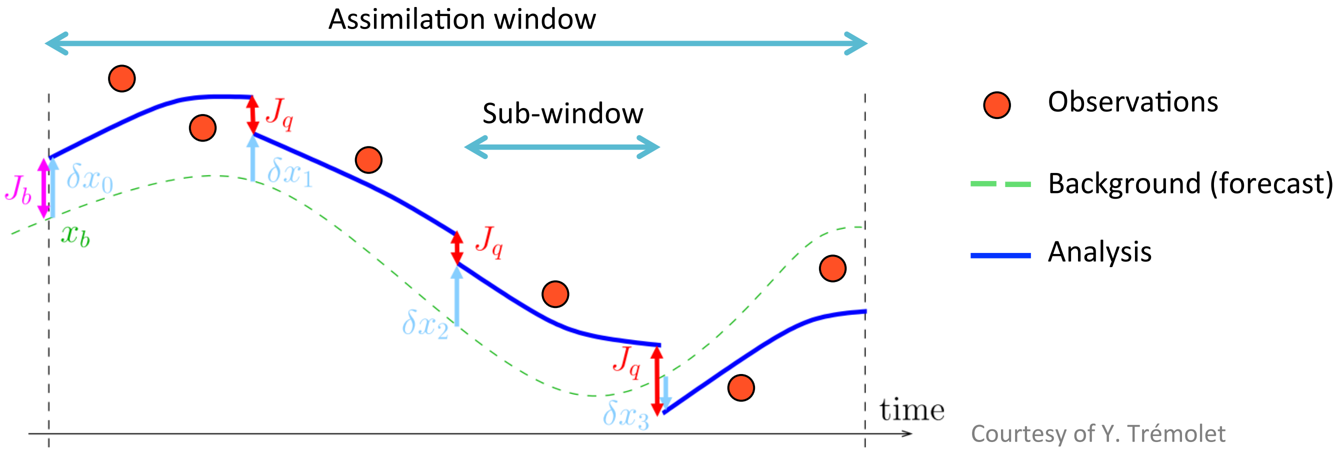
Chemistry facts:

- Heterogeneous, species- and time-dependent forecast errors
- Strong sensitivity to disparate and highly variable model parameters (chemical emissions, mixing layer height, clouds, meteorology etc. etc.)
- Highly variable (non-linear) chemical couplings

Basic needs for the DA algorithm:

- ❑ Initial condition **and** model errors accounted
- ❑ Long assimilation windows (≥ 12 hours)
- ❑ Multivariate

4DEnVar



- 4DEnVar exploits *ensembles* to mimic a ***weak constraint 4D-Var***
- … with no need of linearized and adjoint codes
- … no need to construct and estimate complex error covariances operators **B**, **Q** (localization operator **C** needed instead)
- Physical-based specification of model error covariance (**Q**) through stochastic perturbation of model parameters
- No need of ensemble inflation

Desroziers et al. 2014. 4D-En-Var: link with weak-constraint 4D-Var and different possible implementations. QJRMS

A new toy-model for atmospheric chemistry DA: **QG-Chem**

Quasi-Geostrophic model (QG):

- Two-layers 3-D model, mesoscale dynamics at mid-latitudes
- **Model state**: stream function $\psi(N_x,N_y,2)$
- Conservation of potential vorticity q
- 12000*6000 km^2 domain, cyclic in E-W, fixed boundary conditions N-S
- q advected with a semi-lagrangian scheme

Tropospheric chemistry model:

- 0-D chemical solver $C_t=F_{ASIS}(C_{t-1})$
- **Model state**: C(96) chemical concentrations
- RACM chemical scheme (Stockwell et al. 1997)
- Emissions of primary pollutants
- Diurnal effect (photo-chemistry)

QG-Chem = QG meteorology + tropospheric chemistry

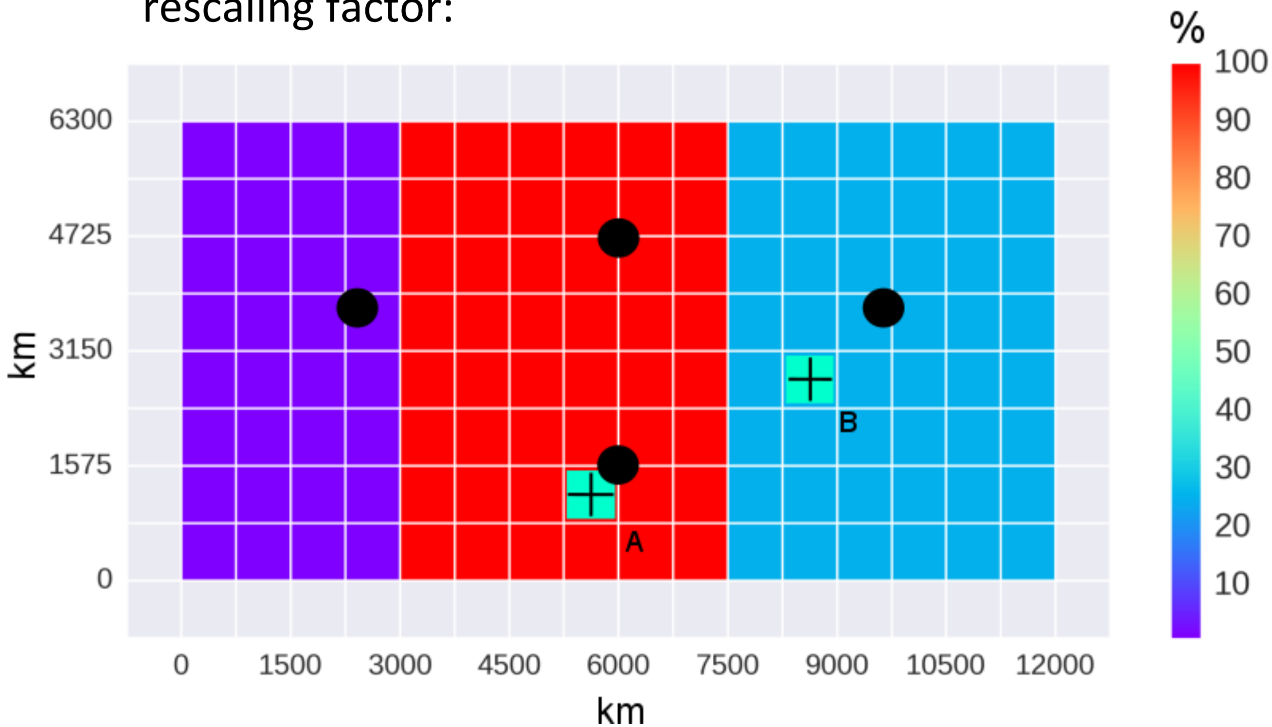
- **Model state**: 3-D stream function $\psi(N_x,N_y,2)$ + 3-D chemical concentrations $C(N_x,N_y,2,96)$
- Chemicals advected with QG semi-lagrangian scheme
- 2-D field of pollutants emissions $E(N_x,N_y,N_{emi})$
- N-S chemical boundary conditions (set to climatology)

QG-Chem: experiments setup

Surface emissions typical of Paris (Crassier et al. 2000):

Species	Value (10^9 molec m^{-2} s^{-1})
NO	121.29
CO	2500
CH_4	802
ETH	6.25
HC3	37.67
HC5	44.43
HC8	19.14
ETE	22.33
OLT	39.67
OLI	6.37
TOL	9.02
HCHO	5.77
ALD	14.45
KET	5.7
XYL	14.55
CSL	3.68

QG-Chem domain (**N**$_x$=16, **N**$_y$=8) and emissions rescaling factor:

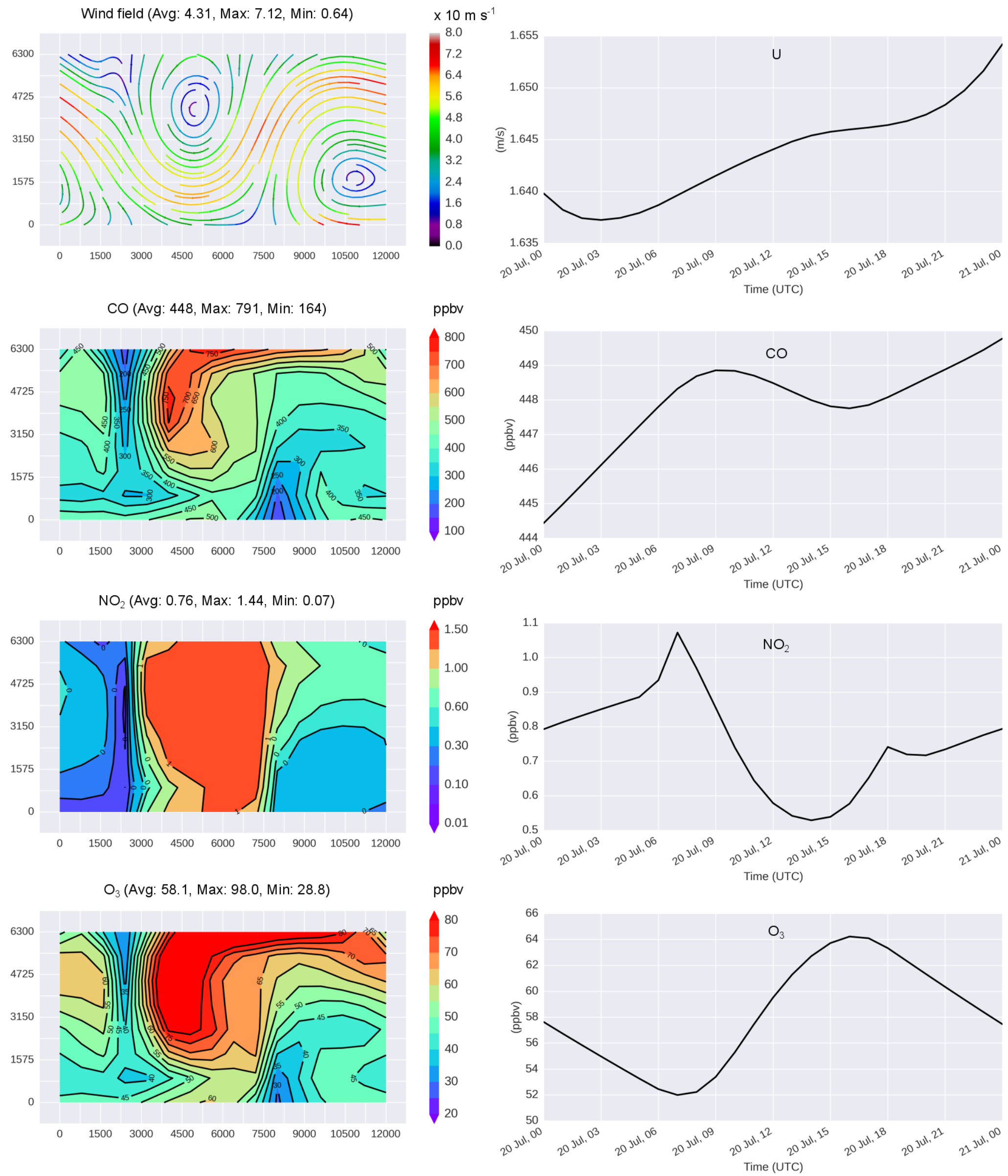


- Location of assimilated measurements

A, B : Observed locations

QG-Chem fields

QG-Chem on day twenty, Rossby circulation, polluted atmosphere



Winds

CO

NO_2

O_3

DA experiments setup

- 24 hours long experiments
- Control variable (ψ,C): size = 97·16·8·2 = **24832 variables**
- 4 synthetic observations of four species (O_3, CO, CO_2, NO_2) assimilated hourly: 4·24 = **96** daily **observations** for each species
- Observations perturbed with gaussian noise (function of the species)
- **Reference unperturbed simulation to validate the results (truth)**

Perfect model experiment

Initial Condition (IC) randomized with $\mathbf{B}^{1/2}$

3D-Var

- Analysis/forecast cycles of **1h**
- Same **B** as above in assimilation step (*climatological* covariance)

4DEnVar

- 1 window of **24h**, 24 sub-windows of 1h
- **16** ensemble members
- $\mathbf{B}^{1/2}$ to create ensemble perturbations

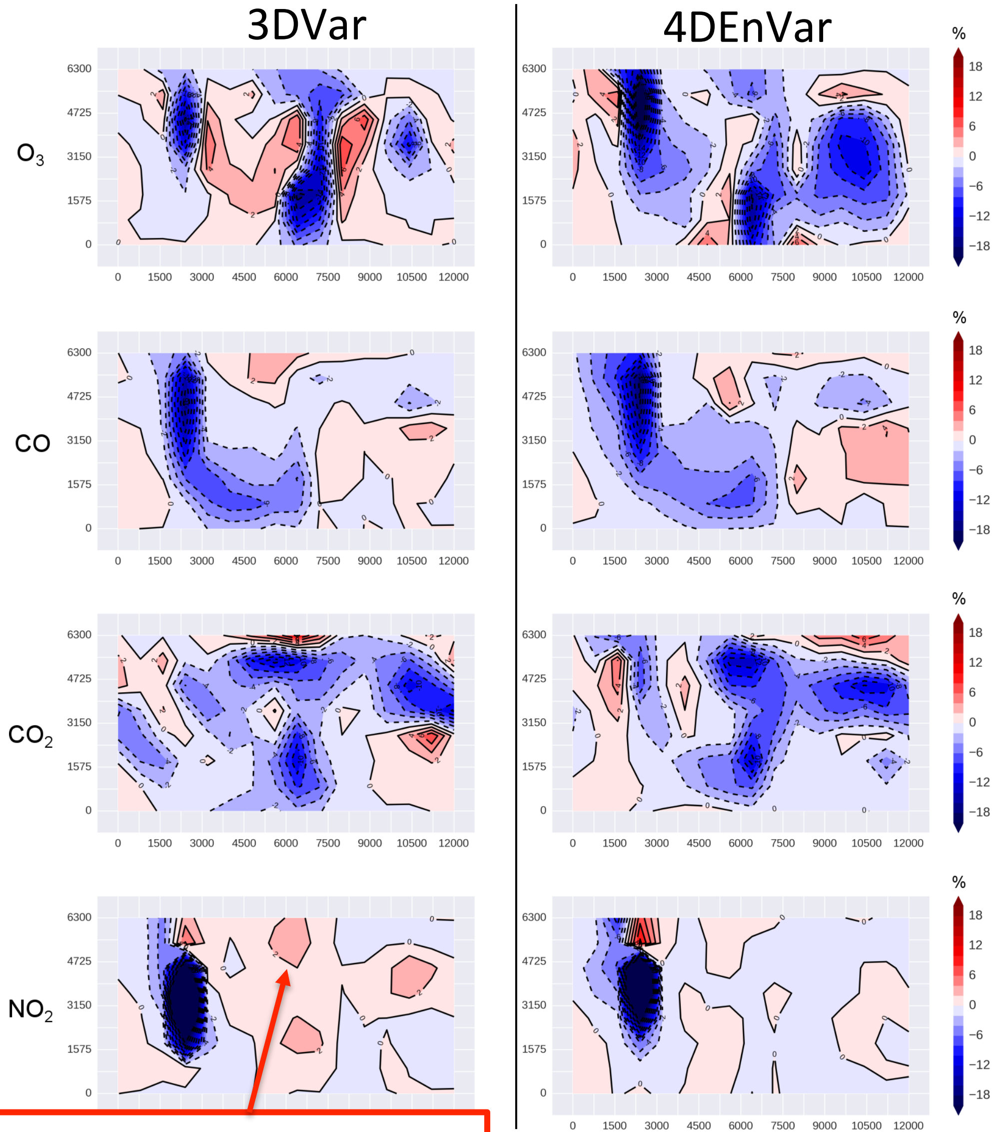
Model error experiment

IC randomized with $\mathbf{B}^{1/2}$ +
Log-normal perturbation of NO emissions (model parameter)

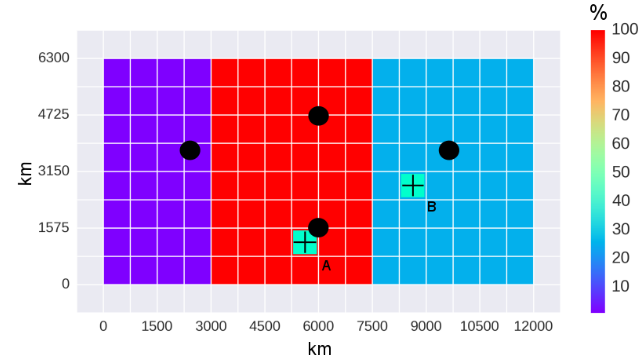
4DEnVar

As in the left side, but model emissions perturbed as well to generate the ensemble

Perfect model experiment: RMSE gain (%)



Degradation at observed locations!



Reanalyses skills: 4DEnVar on par or better than 3DVar when the model is perfect. Impact of flow dependent **B**?

gain (%) = (RMSE$_{analysis}$ – RMSE$_{forecast}$) / MEAN(x$_{truth}$)

< 0 (**blue**) when DA improves the 4D state

> 0 (**red**) otherwise

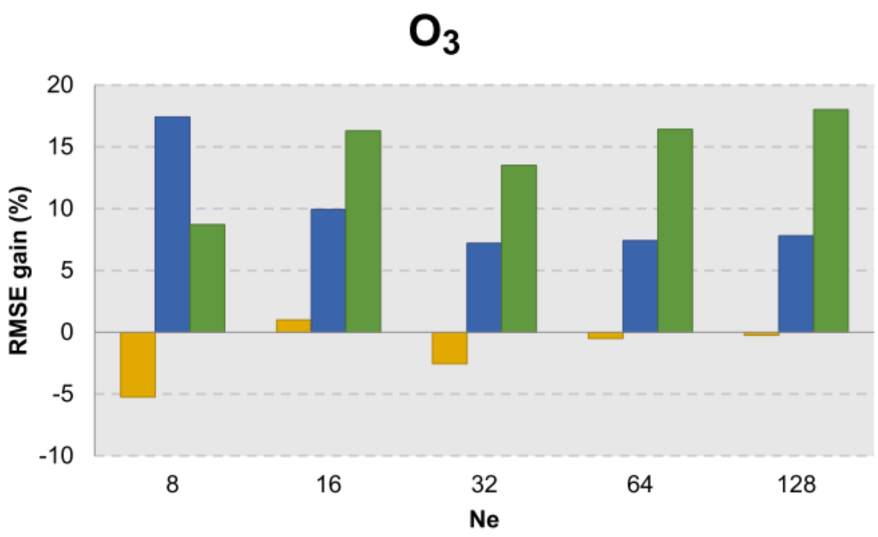
Perfect model experiment: 3DVar vs 4DEnVar

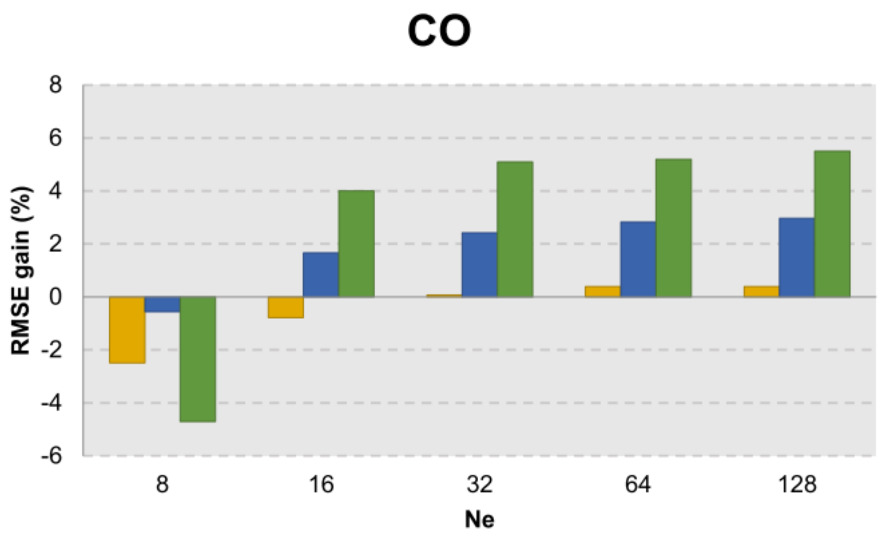
	3D-Var			4DEnVar		
Species	Max. gain (%)	Min. gain (%)	Avg. gain (%)	Max. gain (%)	Min. gain (%)	Avg. gain (%)
O_3	17.11	−7.67	0.92	$\boxed{27.03}$	$\boxed{-6.68}$	$\boxed{2.55}$
CO	17.15	$\boxed{-3.33}$	1.22	$\boxed{18.82}$	−4.11	$\boxed{1.62}$
CO_2	13.08	−10.60	1.76	$\boxed{13.62}$	$\boxed{-7.51}$	$\boxed{1.93}$
NO_2	$\boxed{57.39}$	$\boxed{-6.15}$	0.95	48.07	−10.52	$\boxed{1.21}$

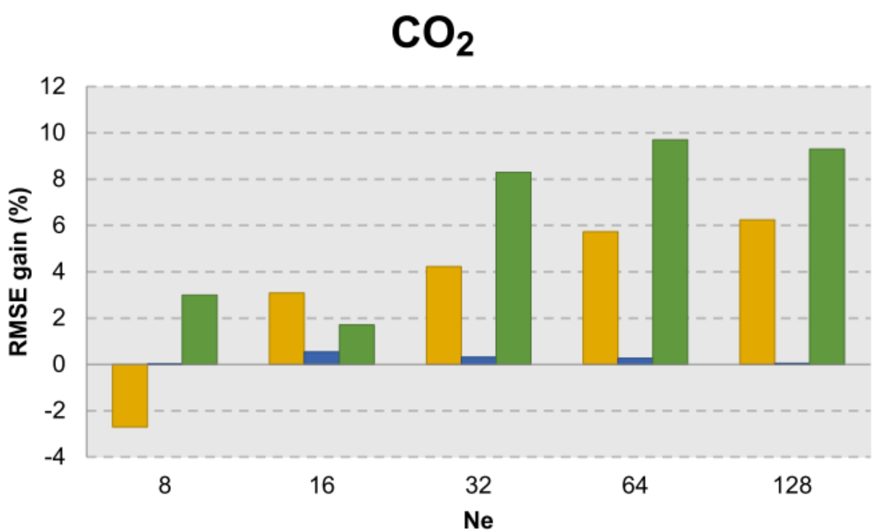
gain (%) = ($RMSE_{analysis}$ – $RMSE_{forecast}$) / MEAN(x_{truth})

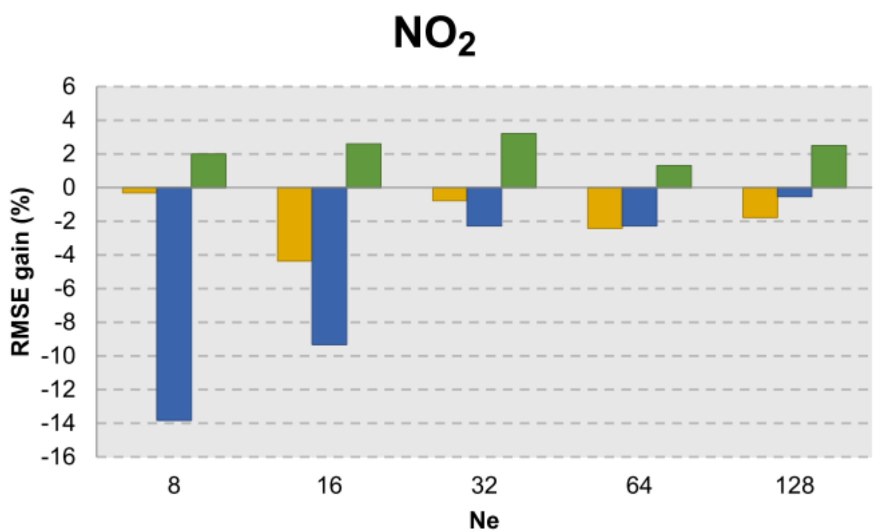
- Comparison done for one cycle of 24 hours (24 cycles of 3DVar)
- 4DEnVar gains slightly better than 3DVar
- Better average gains with 4DEnVar for all chemical species
- Locally some larger errors with 4DEnVar for some species (CO, NO_2). Ensemble size? Localization?

Perfect model experiment: ensemble size

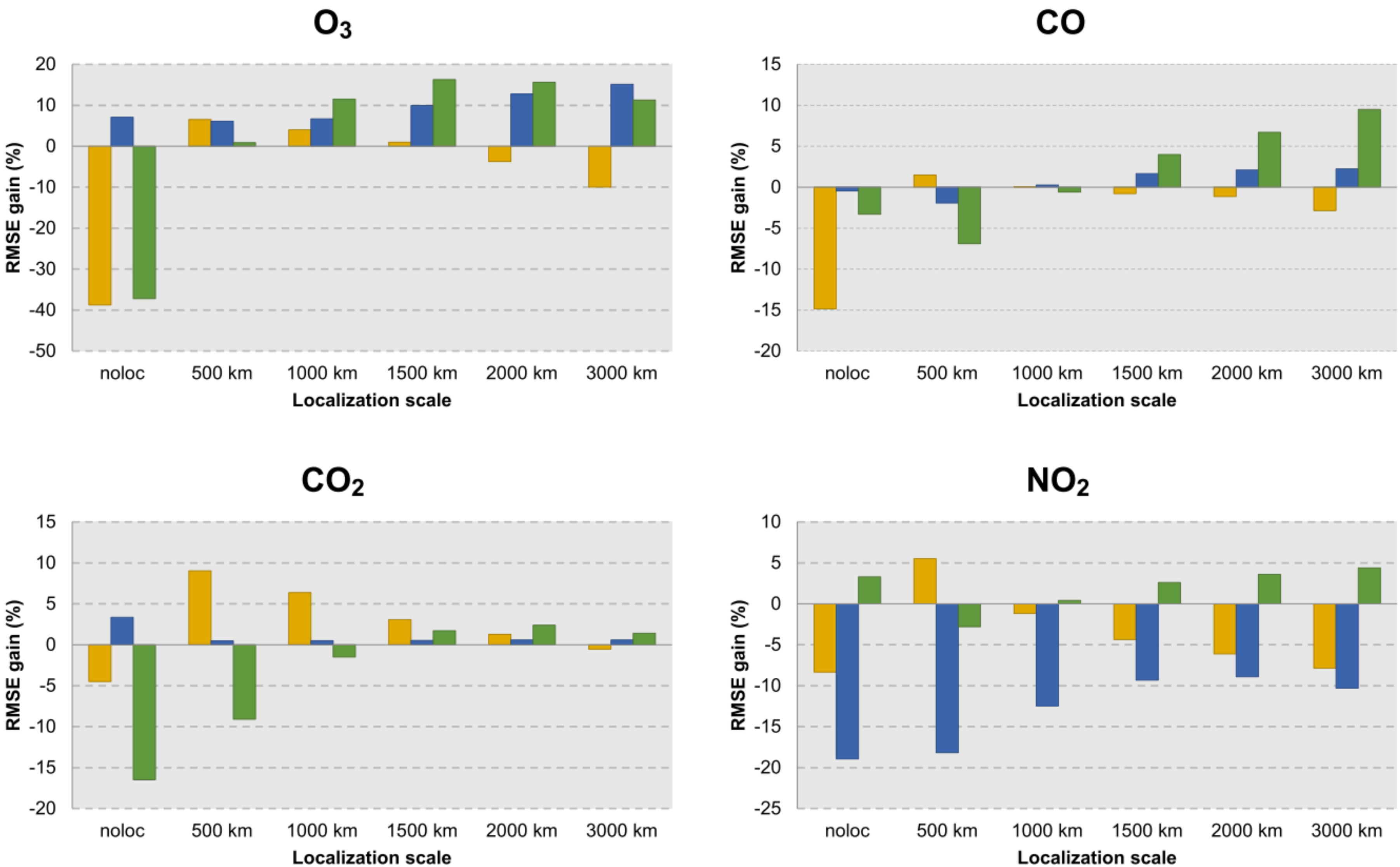








Perfect model experiment: localization



Effective model error (forecast bias)

Effective model error (bias)

4DEnVar analysis

Non-linear forecast initialized with $\mathbf{x}_a(t_0)$

$$\Delta\underline{\mathbf{x}}^{(t>0)} = \boxed{\underline{\mathbf{x}}_a^{(t>0)}} - \boxed{\underline{\mathbf{x}}_f^{(t>0)}}$$

Truth

Non-linear model

Sub-window index

Model error

$$\boldsymbol{\eta}^{(t)} = \mathbf{x}_*^{(t)} - \mathcal{M}_{(t-1)\to(t)}(\mathbf{x}_*^{(t-1)})$$

Analysis error

$$\mathbf{e}_a^{(t)} = \mathbf{x}_a^{(t)} - \mathbf{x}_*^{(t)}$$

Linearization

$$\mathcal{M}_{(t-1)\to(t)} \sim \mathbf{M}_t.$$

Analysis error contribution

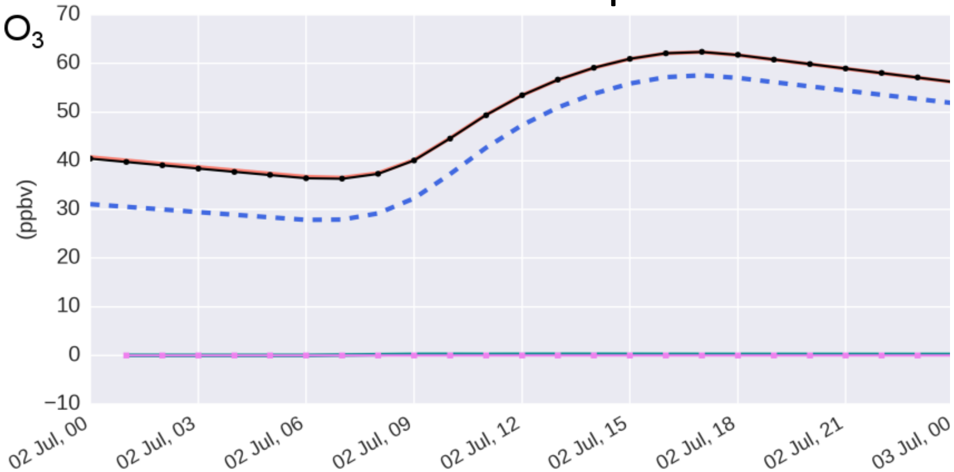
<<

Model error contribution

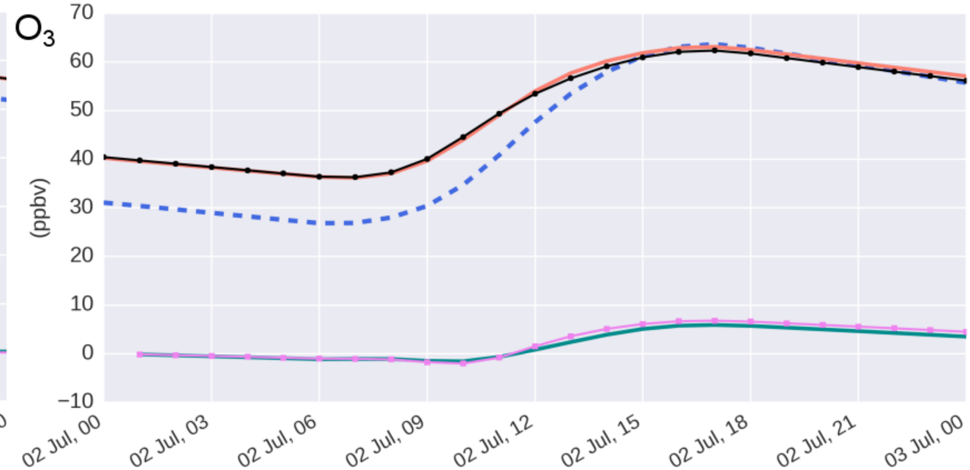
$$\Delta\boldsymbol{x}^{(t)} \simeq \boxed{\boldsymbol{e}_{\mathrm{a}}^{(t)} - \prod_{l=0}^{t}\mathbf{M}_l\boldsymbol{e}_{\mathrm{a}}^{(0)}} + \boxed{\boldsymbol{\eta}^{(t)} + \sum_{j=1}^{t-1}\prod_{l=j+1}^{t}\mathbf{M}_l\boldsymbol{\eta}^{(j)}}$$

Model error experiment: ozone

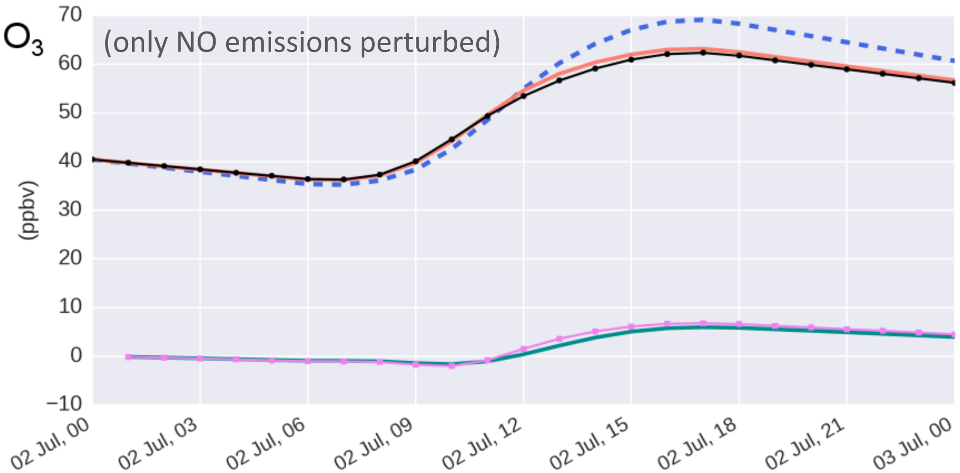
Initial Cond. error and perfect model



Both IC and model error



Perfect IC and model error



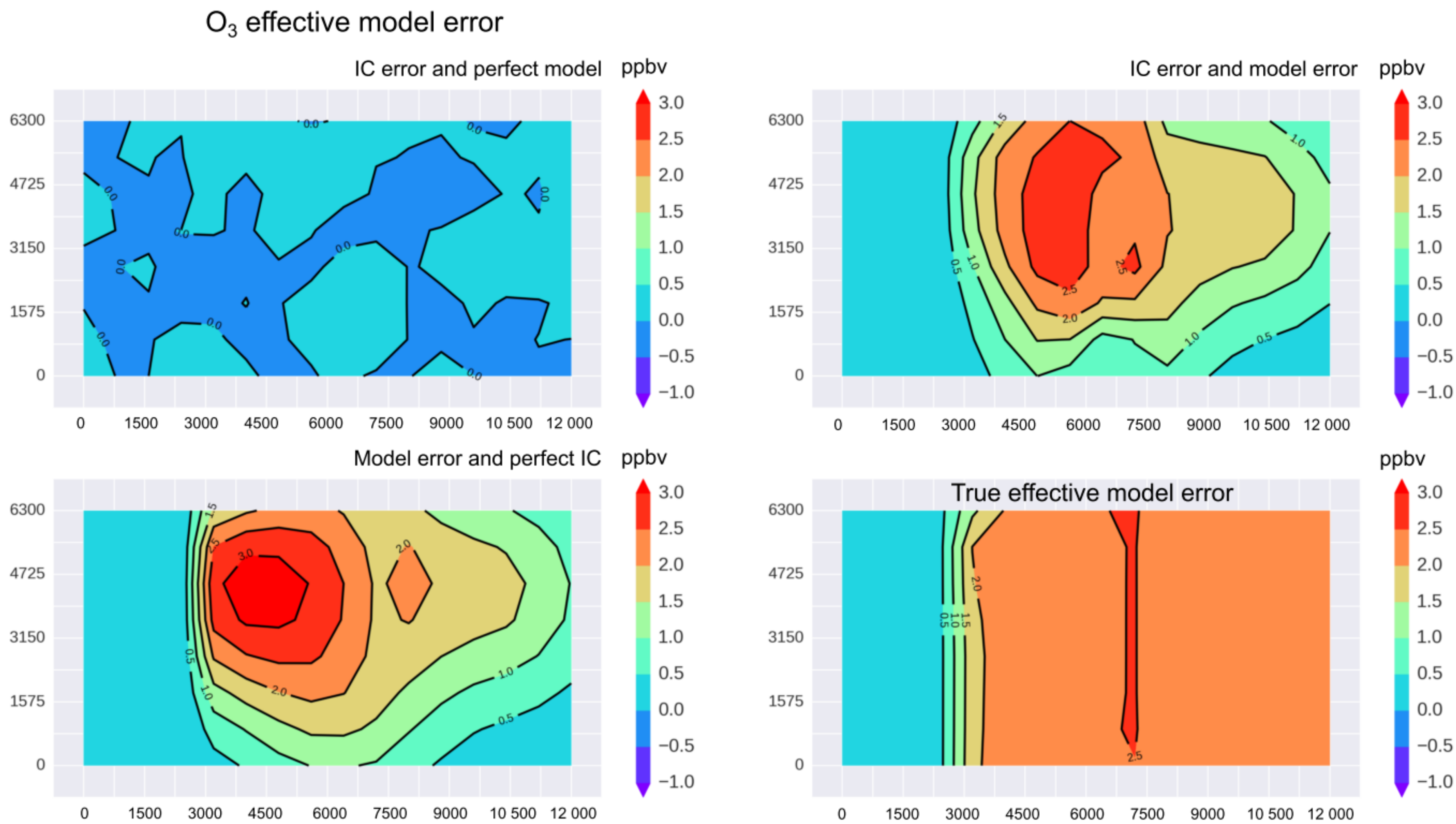
Truth
Free forecast
4D-EnVar analysis (only O_3 assimilated)
True model bias
Model bias estimation

4DEnVar model bias estimation (4D):

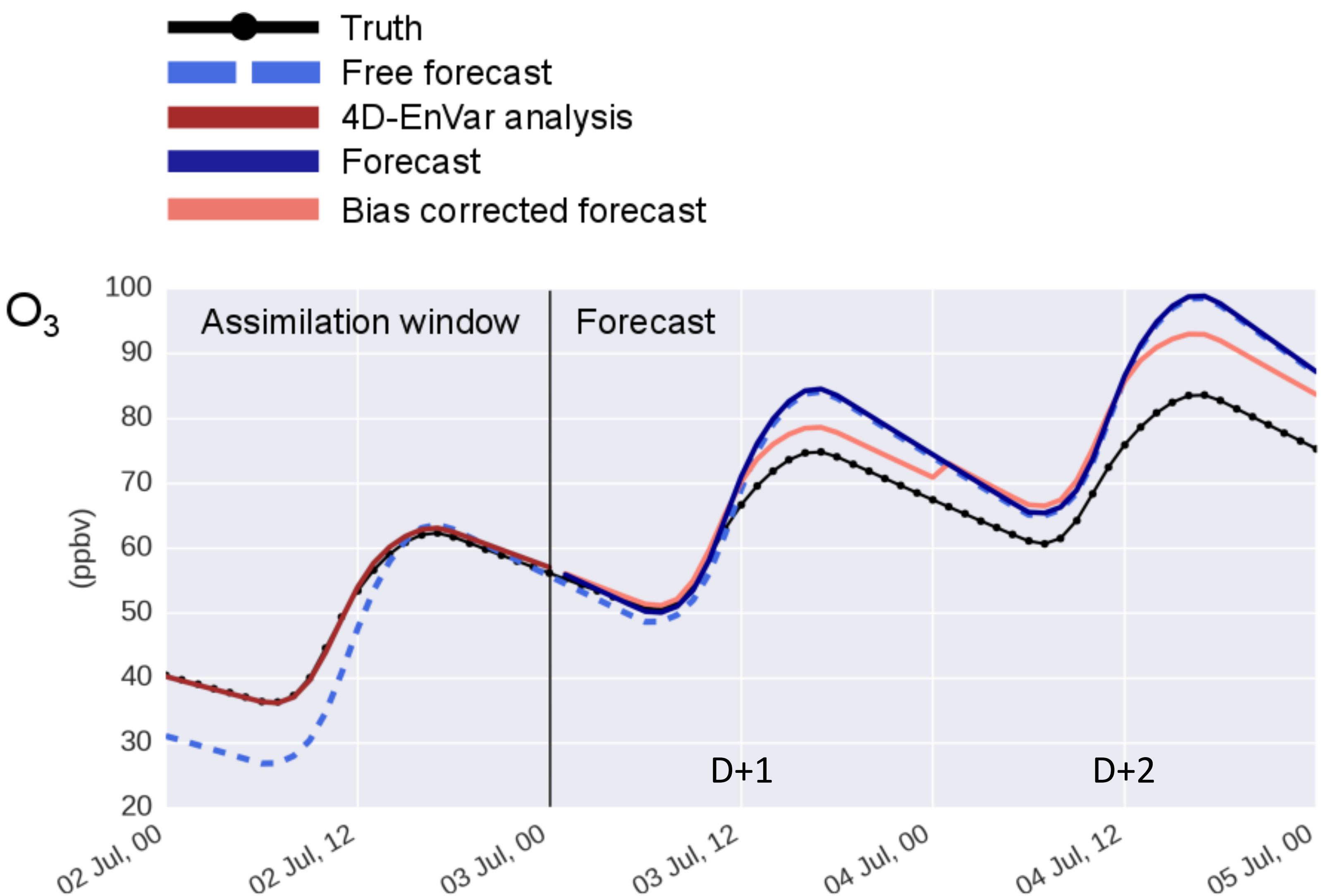
$\Delta \mathbf{x} = M_{t_0 \text{->} t_f}(\mathbf{x}_a(t_0)) - \mathbf{x}_a$

True model bias: $\Delta \mathbf{x} = M_{t_0 \text{->} t_f}(\mathbf{x}_*(t_0)) - \mathbf{x}_*$

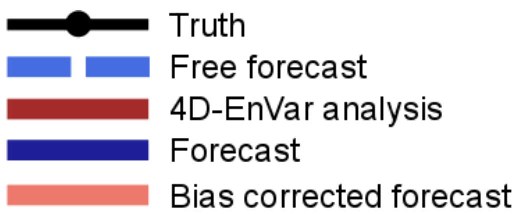
Model error experiment: ozone

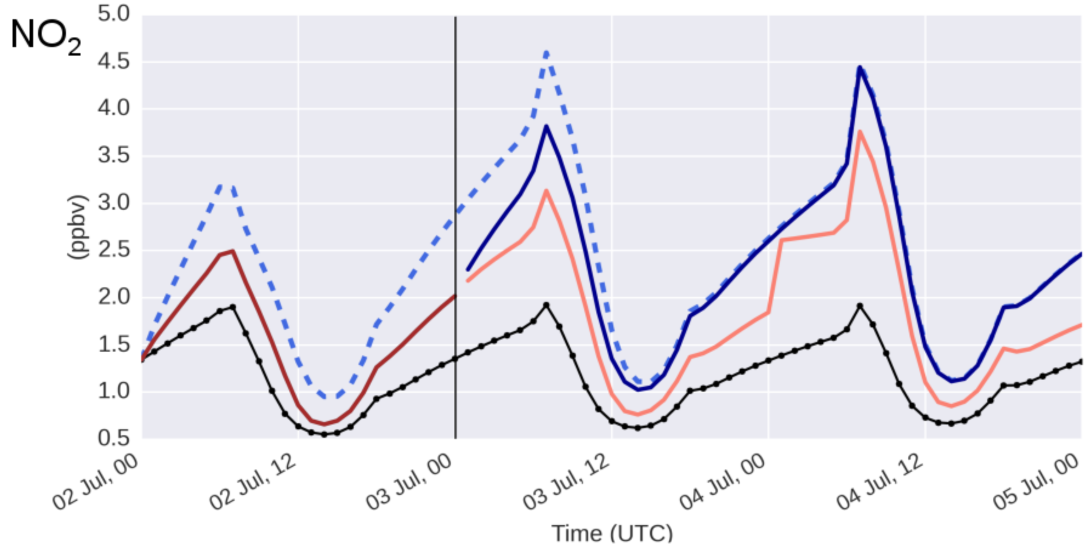


Model error experiment: forecast correction

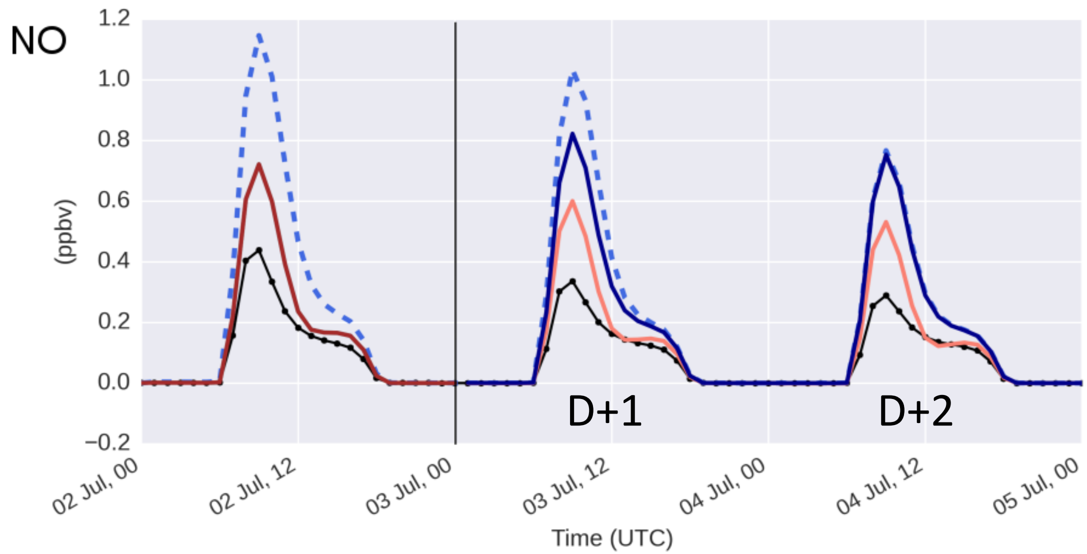


Model error experiment: multivariate forecast correction





- Unobserved species can benefit from O_3 assimilation through multivariate control



Model error experiments: 3DVar vs 4DEnVar

Species	Reanalysis						Forecast					
	3D-Var			4DEnVar			3D-Var			4DEnVar		
	ϵ_{min}	ϵ_{max}	ϵ_{avg}	ϵ_{min}	ϵ_{max}	ϵ_{avg}	ϵ_{min}	ϵ_{max}	ϵ_{avg}	ϵ_{min}	ϵ_{max}	ϵ_{avg}
O_3	2.8	25.7	14.3	2.0	26.4	13.3	5.1	23.9	14.4	4.9	21.7	11.3
CO	1.3	42.8	17.1	1.1	27.7	13.1	6.7	90.8	37.3	7.2	50.1	22.1
NO_2	7.7	88.9	41.2	4.1	69.4	23.3	23.6	117.8	74.0	11.3	90.3	37.2
CO_2	3.1	15.3	8.6	1.4	17.5	8.6	1.7	13.7	5.7	1.9	13.6	6.9

ε (%) = $RMSE_{analysis/forecast}$ / $MEAN(x_{truth})$

- Comparison done for 7 windows of 24 hours (Reanalysis) and the follow-up 24 hours forecast
- All chemical emissions perturbed (model error) plus the initial condition for the four species of interest
- Initial condition set to the truth at beginning of each window (no cycling)
- Forecast bias correction applied within 4DEnVar
- 4DEnVar significantly better for chemical species that depend strongly from model error (CO, NO_2, O_3), both in reanalysis than in forecast mode

Summary

- ✓ QG-Chem developed to test state-of-the-art chemical DA
- ✓ 4DEnVar slightly better than 3DVar for chemical reanalyses when the model is perfect …
- ✓ … but much more promising for multivariate chemical assimilation in presence of generic model errors
- ✓ A simple 4DEnVar diagnostic can help detecting and reducing forecast biases due to strong model errors
- ➢ Future work:
  - ❑ Non-linear chemical regimes (outer loops?)
  - ❑ Better localization (advection of localization, species dependent length scales)
  - ❑ Cycling
  - ❑ Prototype implementation on a real CTM

More details on Emili et al., GMD 2016