European Commission

FAIRMODE
Forum for air quality modelling in Europe

WG3 session
Forecast indicators

FAIRMODE Technical Meeting

Dublin - Ireland, October 7-9 2024

agenda

Monday 7 Oct	Alexandra Monteiro	Feedback on last Hackathon (May 2024)
	Antonio Piersanti, Kees Cuvelier	Some insight on uncertainty parameters change and comparison with persistence model
Tuesday 8 Oct	Joanna Struszewska, Stijn Janssen	Feedback to the draft technical guidance document (ch. 6)
	Roberta Amorati	using DeltaTool to assess the performance of ARPAE probabilistic forecast model
	Antonio Piersanti, Alexandra Monteiro	Future challenges and proposal for the next steps

ITALIAN NATIONAL AGENCY FOR NEW TECHNOLOGIES, ENERGY AND SUSTAINABLE ECONOMIC DEVELOPMENT

universidade de aveiro
theoria poiesis praxis

European Commission

FAIRMODE
Forum for air quality modelling in Europe

WG3

STATUS OF WG3 ACTIVITIES

FAIRMODE Technical Meeting
Dublin - Ireland, October 7-9 2024

Antonio Piersanti, Alexandra Monteiro, Lina Vitali, Kees Cuvelier

OUTLINE

1. Feedback on last Hackathon (May 2024)
2. Some insight on uncertainty parameters change
3. Some insight on the comparison with the Persistence Model (by Kees)
4. Main discussion points

5. Future challenges and proposal for the next steps (Tue 8/10)

WG3 HACKATHON

Testing the effect of changing parameters for measurement uncertainty estimates

AN EXERCISE WAS LAUNCHED IN PARIS AT FAIRMODE PLENARY MEETING (FEBRUARY 2024)

Aim: Testing the effect on Forecast Evaluation Outcomes of adopting the new parameters proposed by AQUILA for measurement uncertainty estimates

AN HACKATHON WAS HELD ON 8th MAY 2024

Results and feedback of the exercise were shared and discussed

Contribution from

Paweł Durka, Aleksander Norowski	IEP-NRI (Poland)
Loris Colombo	ARPA Lombardia (Italy)
Alexandra Monteiro, Carla Gama	UniAveiro (Portugal)
Eivind G. Wærsted, Bruce R. Denby	MET Norway (Norway)

WG3 HACKATHON

FEEDBACK overview

Summary

- Changes in scatter and target plots – depending on pollutant
- Slight changes in MPI plots
- More stringent calculations for target plot – all pollutants
- In general - differences are rather small, but more tests are needed
- Is it a good idea to leave to options? Or should we leave one option after tests and discussions?
- Forecast treshold performance of model – need for „loosen" the stringency?

IEP-NRI (Poland)

BRIEF SUMMARY

- Test from 2019 to 2022 by using old criteria
- Test from 2019 to 2022 by using AQUILA criteria

- Main issues:

1) MQI becomes higher due to stringent parameters
2) MQI increasing (about 5-10%) only in one case more than 10% (O3)
3) Forecast plot is quite similar (except for NO2)

ARPA Lombardia (Italy)

UniAveiro (Portugal)

Questions & answers

Is forecast MQO too easy to be complied?
NO!

Should we plan to make MQIf criterion stricter?
We doubt...

Are new outcomes more reliable than old ones?
With the new parameters formulation, MQI is not fulfilled at some monitoring points where it was before (e.g. O3)

Is the new formulation better in pointing out critical contexts?
More time to conclude

MET Norway (Norway)

Comments

About the MQO

- New AQUILA parameters give stricter MQOf for our data, though for PM_{10} the difference is very small
- MQO for $PM_{2.5}$ (both assessment and forecast) still seem too easy to fulfill. Even in periods with large bias, we can fulfill MQOf.
  - May be a problem especially for low concentrations due to high obs. uncertainty
- Fulfilling MQOf might not mean it is a good forecast, just that persistence model would not do much better (particularly when day-to-day variability is large)
- What is the basis for the new values of β (for assessment)?
  (Philippe explained β was modified to give same stringency as before at the limit value)

Technical suggestions for Delta:

- Enable working with a period spanning more than 1 calendar year? (winter season)
- Scripted version of Delta? (it's inefficient to use GUI to create many plots)

WG3 HACKATHON

FEEDBACK overview

Summary

- Changes in scatter and target plots – depending on pollutant
- Slight changes in MPI plots
- More stringent calculations for target plot – all pollutants
- In general - differences are rather small, but more tests are needed
- Is it a good idea to leave to options? Or should we leave one option after tests and discussions?
- Forecast treshold performance of model – need for „loosen" the stringency?

IEP-NRI (Poland)

BRIEF SUMMARY

- Test from 2019 to 2022 by using old criteria
- Test from 2019 to 2022 by using AQUILA criteria

- Main issues:

1) MQI becomes higher due to stringent parameters
2) MQI increasing (about 5-10%) only in one case more than 10% (O3)
3) Forecast plot is quite similar (except for NO2)

ARPA Lombardia (Italy)

UniAveiro (Portugal)

Questions & answers

Is forecast MQO too easy to be complied?
NO!

Should we plan to make MQIf criterion stricter?
We doubt...

Are new outcomes more reliable than old ones?
With the new parameters formulation, MQI is not fulfilled at some monitoring points where it was before (e.g. O3)

Is the new formulation better in pointing out critical contexts?
More time to conclude

MET Norway (Norway)

Comments

About the MQO

- New AQUILA parameters give stricter MQOf for our data, though for PM_{10} the difference is very small
- MQO for $PM_{2.5}$ (both assessment and forecast) still seem too easy to fulfill. Even in periods with large bias, we can fulfill MQOf.
  - May be a problem especially for low concentrations due to high obs. uncertainty
- Fulfilling MQOf might not mean it is a good forecast, just that persistence model would not do much better (particularly when day-to-day variability is large)
- What is the basis for the new values of β (for assessment)?
  (Philippe explained β was modified to give same stringency as before at the limit value)

Technical suggestions for Delta:

- Enable working with a period spanning more than 1 calendar year? (winter season)
- Scripted version of Delta? (it's inefficient to use GUI to create many plots)

GENERAL OUTCOME

AQUILA parameters make the MQO_f criterion more stringent

WG3 HACKATHON

FEEDBACK overview

Summary

- Changes in scatter and target plots – depending on pollutant
- Slight changes in MPI plots
- More stringent calculations for target plot – all pollutants
- In general - differences are rather small, but more tests are needed
- Is it a good idea to leave to options? Or should we leave one option after tests and discussions?
- Forecast treshold performance of model – need for „loosen" the stringency?

IEP-NRI (Poland)

BRIEF SUMMARY

- Test from 2019 to 2022 by using old criteria
- Test from 2019 to 2022 by using AQUILA criteria

- Main issues:

1) MQI becomes higher due to stringent parameters
2) MQI increasing (about 5-10%) only in one case more than 10% (O3)
3) Forecast plot is quite similar (except for NO2)

ARPA Lombardia (Italy)

UniAveiro (Portugal)

Questions & answers

Is forecast MQO too easy to be complied?
NO!

Should we plan to make MQIf criterion stricter?
We doubt...

Are new outcomes more reliable than old ones?
With the new parameters formulation, MQI is not fulfilled at some monitoring points where it was before (e.g. O3)

Is the new formulation better in pointing out critical contexts?
More time to conclude

MET Norway (Norway)

Comments

About the MQO

- New AQUILA parameters give stricter MQOf for our data, though for PM_{10} the difference is very small
- MQO for $PM_{2.5}$ (both assessment and forecast) still seem too easy to fulfill. Even in periods with large bias, we can fulfill MQOf.
  - May be a problem especially for low concentrations due to high obs. uncertainty
- Fulfilling MQOf might not mean it is a good forecast, just that persistence model would not do much better (particularly when day-to-day variability is large)
- What is the basis for the new values of β (for assessment)?
  (Philippe explained β was modified to give same stringency as before at the limit value)

Technical suggestions for Delta:

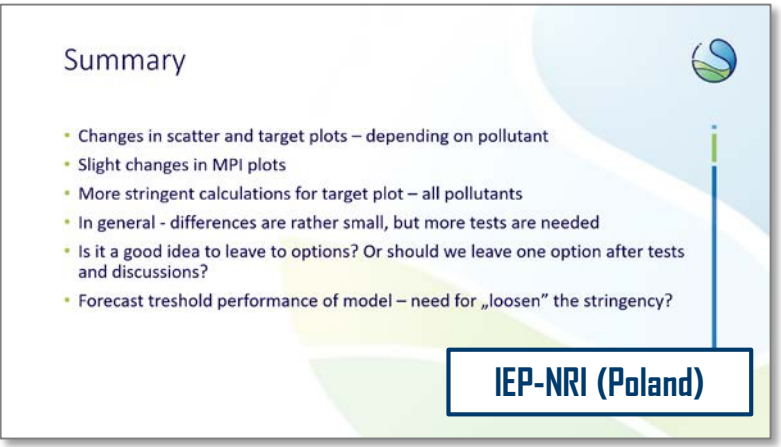
- Enable working with a period spanning more than 1 calendar year? (winter season)
- Scripted version of Delta? (it's inefficient to use GUI to create many plots)

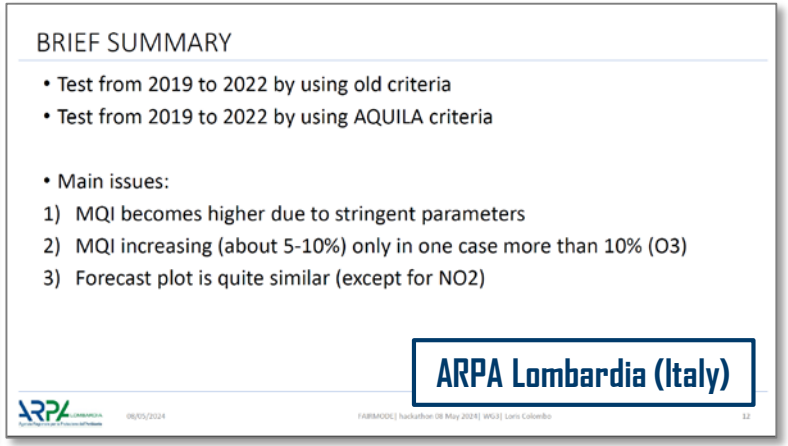
TO WHAT EXTENT?

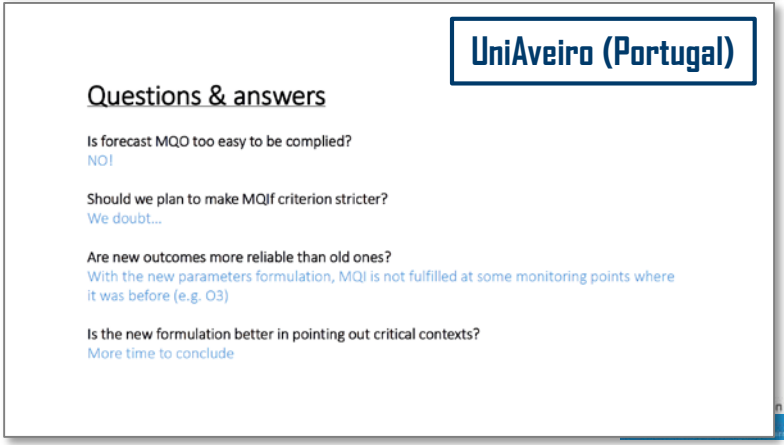
- *Poland: slight changes, small differences*
- *Lombardia (IT): MQI increasing by 5-10%*
- *Portugal: for O_3, MQO_f changes from being to not being fulfilled*
- *Norway: very small differences for PM10*

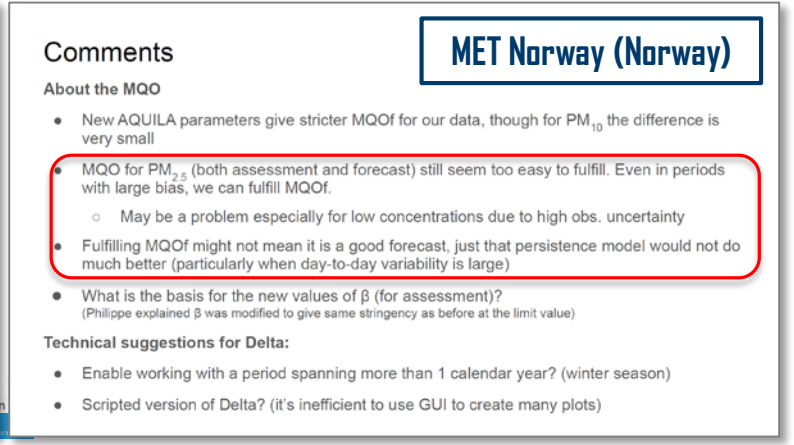
WG3 HACKATHON

FEEDBACK overview









IN ADDITION…

IN NORWAY

- *MQO for PM2.5 still too easy to be fulfilled*
- *Fulfilling MQO_f might not mean it is a good forecast*

[e.g. MQO turned out to be achieved in some cases where poor agreement is observed looking at the time series]

SOME INSIGHT ON PARAMETERS CHANGE

HOW DID PARAMETERS CHANGE?

		FAIRMODE (FM)					
		$Ur(RV)$	RV	α	Np	Nnp	beta
Short term	NO2	0,24	200	0,2	5,2	5,5	2
	O3	0,18	120	0,79	11	3	2
	PM10	0,28	50	0,25	20	1,5	2
	PM2.5	0,36	25	0,5	20	1,5	2
Long term	NO2	0,24	200	0,2	5,2	5,5	2
	O3	0,18	120	0,79	11	3	2
	PM10	0,28	50	0,25	20	1,5	2
	PM2.5	0,36	25	0,5	20	1,5	2

	AQUILA (CURR.)					
	$Ur(RV)$	RV	α	Np	Nnp	beta
NO2	0,15	200	0,2	1	1	4,23
O3	0,15	100	0,4	1	1	2,4
PM10	0,25	45	0,35	1	1	2,26
PM2.5	0,25	25	0,6	1	1	2,88
NO2	0,3	20	0,97	1	1	1,53
O3	0,15	70	0,4	1	1	1,93
PM10	0,2	20	0,6	1	1	1,55
PM2.5	0,3	10	0,8	1	1	2,49

SOME NOTES

- Not only uncertainty parameters, but also β values were modified (within WG2 activities). Note that this change impacts only on MQO for Assessment, since β is not included within MQO formulation for Forecast
- On the "on-the-fly MQI" portal, AQUILA-based parameters are used as the "CURRENT" version for MQI calculation (for Assessment)

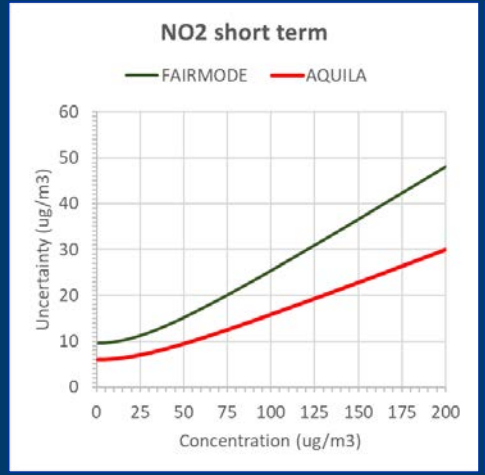
SOME INSIGHT ON PARAMETERS CHANGE

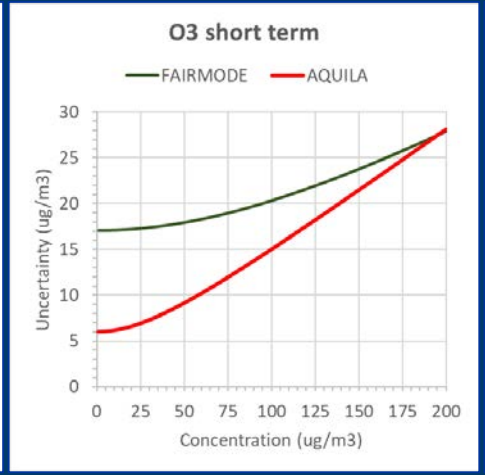
PARAMETERS IN THE FORMULAS

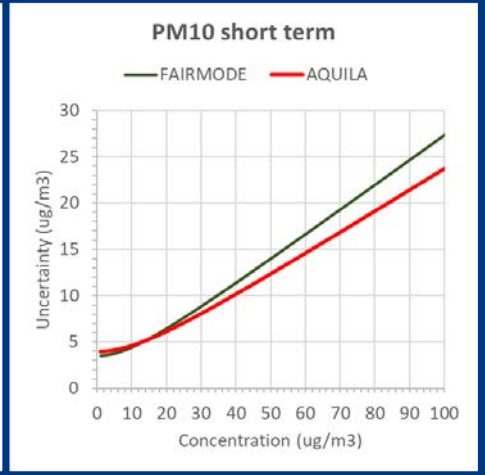
Measurement Uncertainty		$U(O_i) = U_r(RV)\sqrt{(1-\alpha^2)O_i^2 + \alpha^2 RV^2}$
MQI Assessment	Short term	$MQI = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - M_i)^2}}{\beta\sqrt{\frac{1}{N}\sum_{i=1}^{N}U(O_i)^2}} = \frac{\mathrm{RMSE}}{\beta RMS_U}$ with $RMS_U = \sqrt{\frac{\sum_{i=1}^{N}(U(O_i))^2}{N}} = U_r(RV)\sqrt{(1-\alpha^2)(\bar{O}^2 + \sigma_o^2) + \alpha^2 RV^2}$
	Long term	$MQI = \frac{\lvert\bar{O} - \bar{M}\rvert}{\beta U(\bar{O})}$ with $U(\bar{O}) = U_r(RV)\sqrt{\frac{(1-\alpha^2)}{N_p^*}(\bar{O}^2 + \sigma_o^2) + \frac{\alpha^2 . RV^2}{N_{np}}} \cong U_r(RV)\sqrt{\frac{(1-\alpha^2)}{N_p}\bar{O}^2 + \frac{\alpha^2 . RV^2}{N_{np}}}$
MQI Forecast		$MQI_{forecast} = \sqrt{\frac{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2}}$ with $P_i = O_{i-1-forecast\ horizon} \pm U(O_{i-1-forecast\ horizon})$

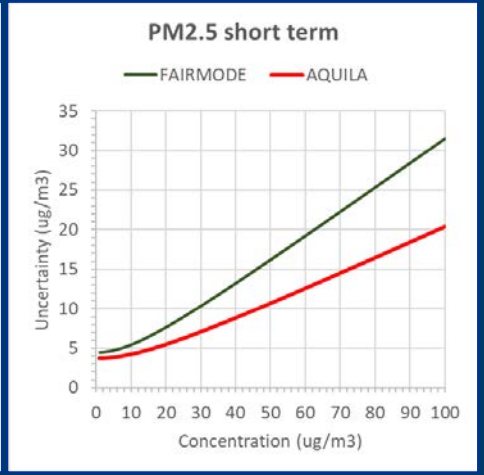
How do new parameters impact on measurement uncertainty estimates?

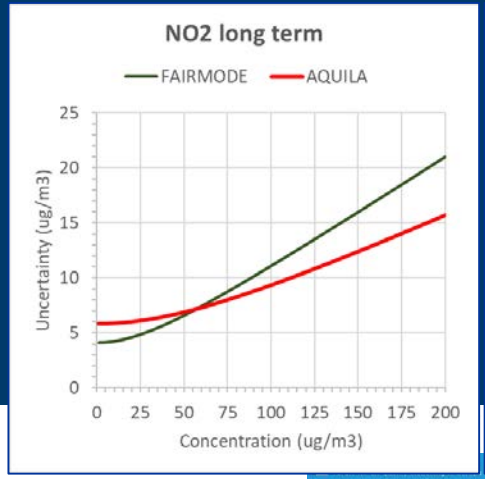
$$U(O_i) = U_r(RV)\sqrt{(1-\alpha^2)O_i^2 + \alpha^2 RV^2}$$

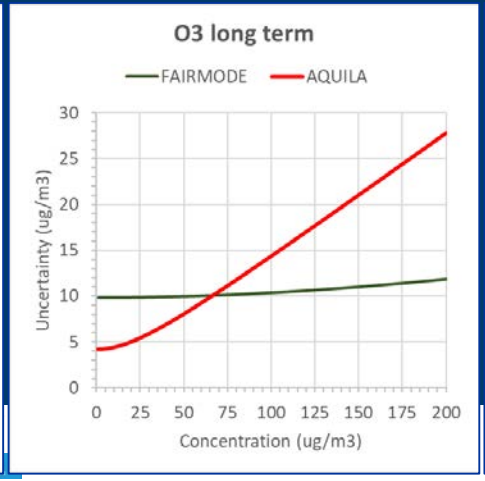


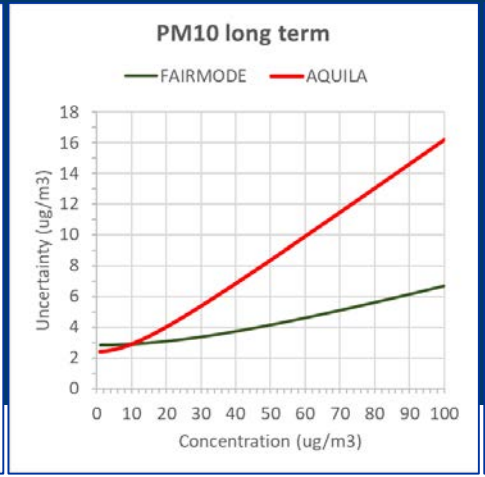


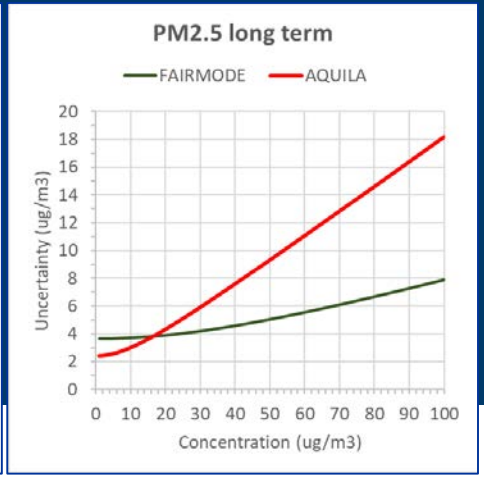












SOME FACTS

1) Sometimes new AQUILA-based parameters make the measurement uncertainty estimate decrease (i.e. MQO criterion gets more stringent); sometimes the opposite happens.
   It depends on the pollutant, on the concentration value, on the type of analysis (short-term or long-term).
2) For short-term analysis, the measurement uncertainty estimates get lower in most of the cases (i.e. criteria get more stringent)
3) Within specific Forecast evaluations, only short-term analyses are considered (i.e. MQI_f based on daily values)

Moreover

4) β values are modified too, but this change impacts only on MQO for Assessment, and not on MQO for Forecast

✓ Concerning MQO for Assessment:

- Due to 1), criteria get more stringent (outcomes get worse) in some cases, less stringent (outcomes get better) in other
- Due to 4), outcomes do not change so much

✓ Concerning MQO for Forecast:

- Due to 2) and 3), criteria get more stringent (outcomes get worse) in most of the cases
- Due to 4), there is no “control knob” to compensate the effect of changing uncertainty parameters

In summary
adopting AQUILA-based parameters
impacts on Assessment MQI and Forecast MQI outcomes
very differently

ENEA
universidade de aveiro
theoria poiesis praxis
European Commission
FAIRMODE

Different impacts on Assessment MQI and Forecast MQI outcomes

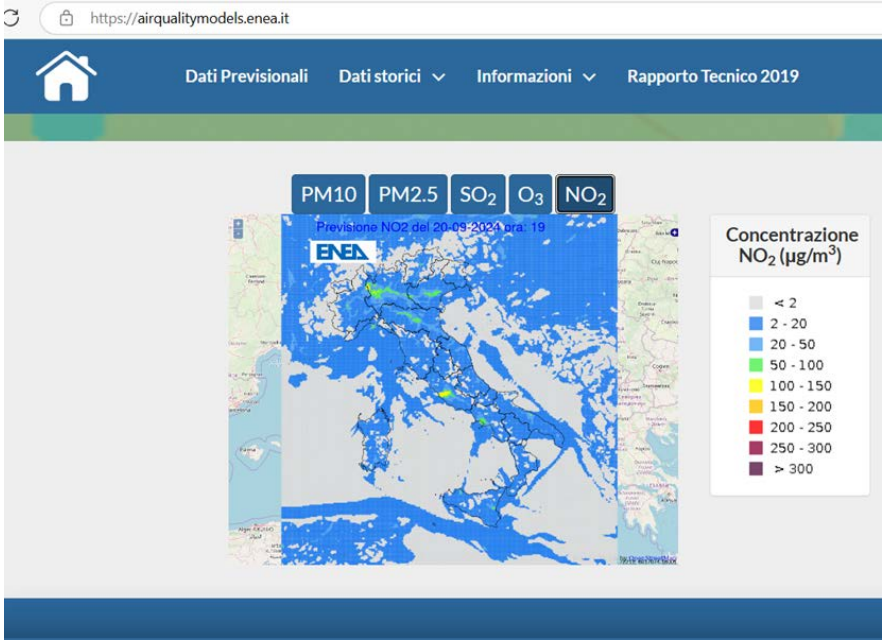
An Example: Evaluation of National Italian Forecast System FORAIR- IT

FORAIR-IT Modeling System

- Europe at 20 km, Italy at 4 km hor. res.
- Meteo: NCEP + WRF
- BC: CAMS
- Emissions: TNO on Europe, NEI on Italy + MEGAN BVOCs
- CTM: FARM (SAPRC-99 + aero3 + ISOROPIA + SORGAM)
- No assimilation of observations

Validation Features

- Year: 2022
- Pollutants: NO_2, O_3, PM10, PM2.5
- Validation on all available Background stations:
  370 (NO_2); 300 (O_3); 340 (PM10); 199 (PM2.5)



Modelling System Development, Simulations Operational Maintenance
Massimo D'Isidoro (ENEA)

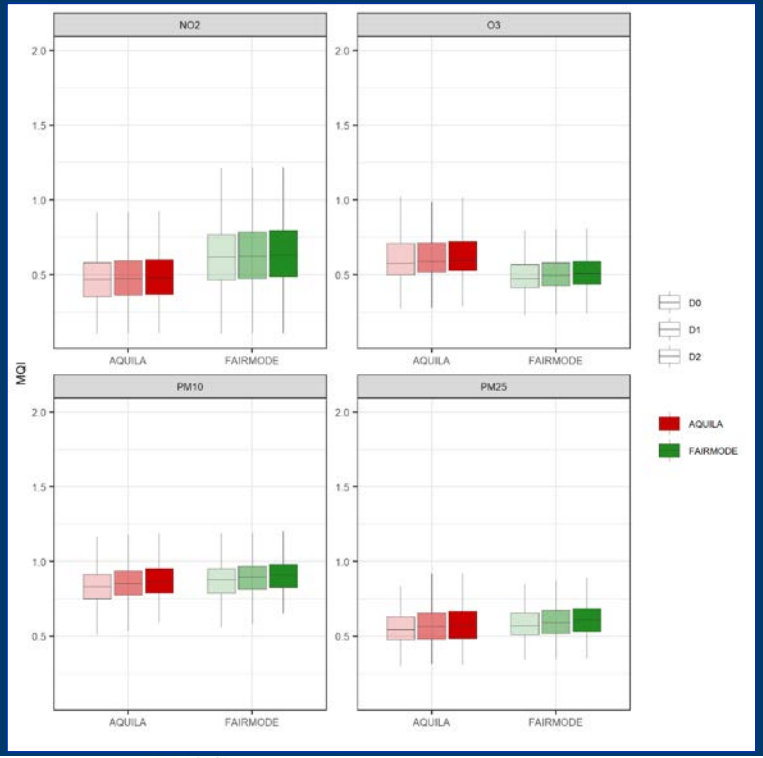
Validation Data Base Setting up
Maria Gabriella Villani (ENEA)

https://airqualitymodels.enea.it

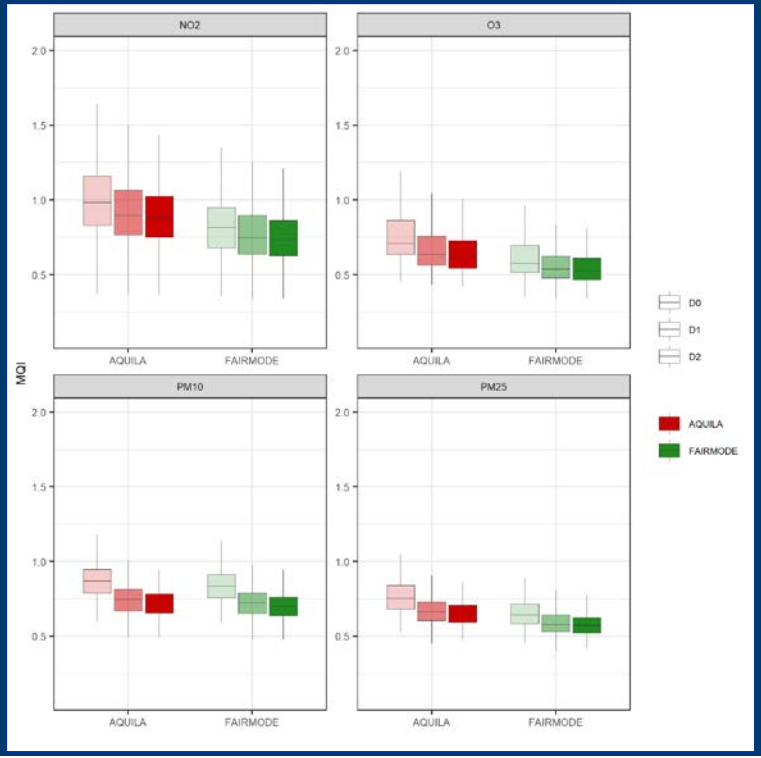
DIFFERENT IMPACTS ON ASSESSMENT MQI AND FORECAST MQI OUTCOMES

AN EXAMPLE: EVALUATION OF NATIONAL ITALIAN FORECAST SYSTEM FORAIR- IT

MQI ASSESSMENT



MQI FORECAST

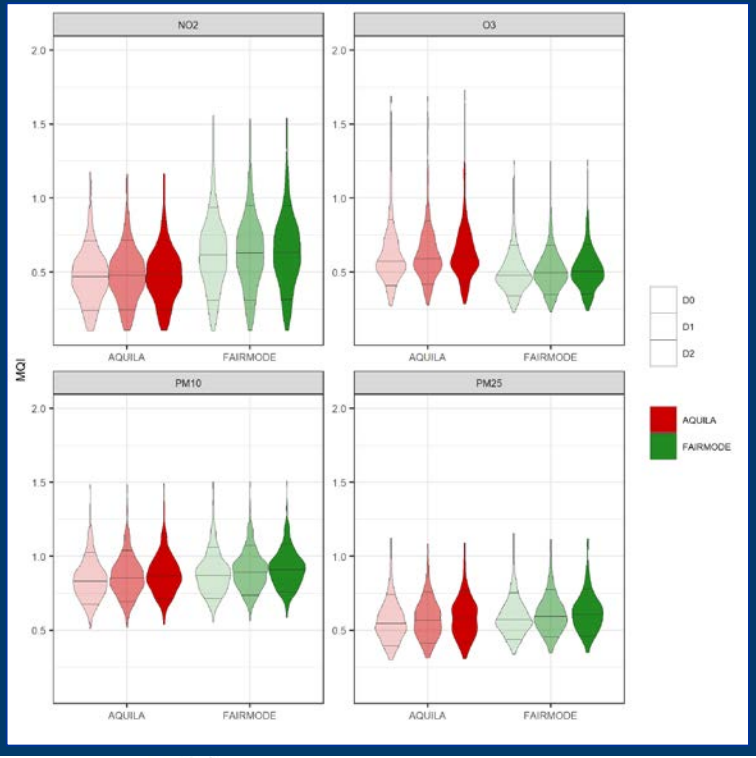


Box plots 25°-75° perc.: 1 line inside, for 50°perc.

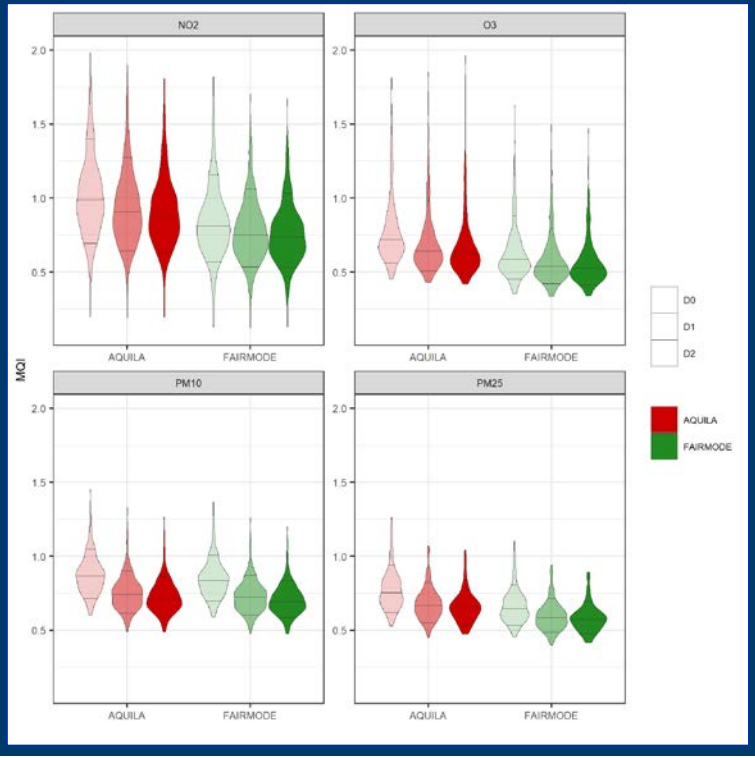
Different impacts on Assessment MQI and Forecast MQI outcomes

An Example: Evaluation of National Italian Forecast System FORAIR- IT

MQI Assessment



MQI Forecast



Violin plots: 3 lines inside, for 10°-50°-90° perc.

First Points for Discussion

A1. Should we introduce β within Forecast MQI formulation?

A2. Ideas for setting β values?

- preserve the consistency with current AQUILA-based Assessment MQO
- preserve the consistency with previous Forecast MQO
- statistical analysis on a large amount of data to set the level of stringency that makes the X-percentile of the simulations to turn out fit-for-purpose
- …….

THE COMPARISON WITH THE PERSISTENCE MODEL

SOME INSIGHT

Not only RMSE….

An analysis by Kees Cuvelier
about the comparison with the Persistence Model
from different points of views

Validation of CAMS AQ ForeCast 2021

9 Models:

CHIMERE, DEHM, EMEP, EURADIM, GEMAQ, LOTOS, MATCH, MOCAGE, SILAM

$\Rightarrow$ ENS (Median Ensemble model)

Each day at 0h00 there is a 96-hourly forecasts (i.e. 4 days)

Download: ALL models: Jan & Jul 2021 ENS: Full year 2021

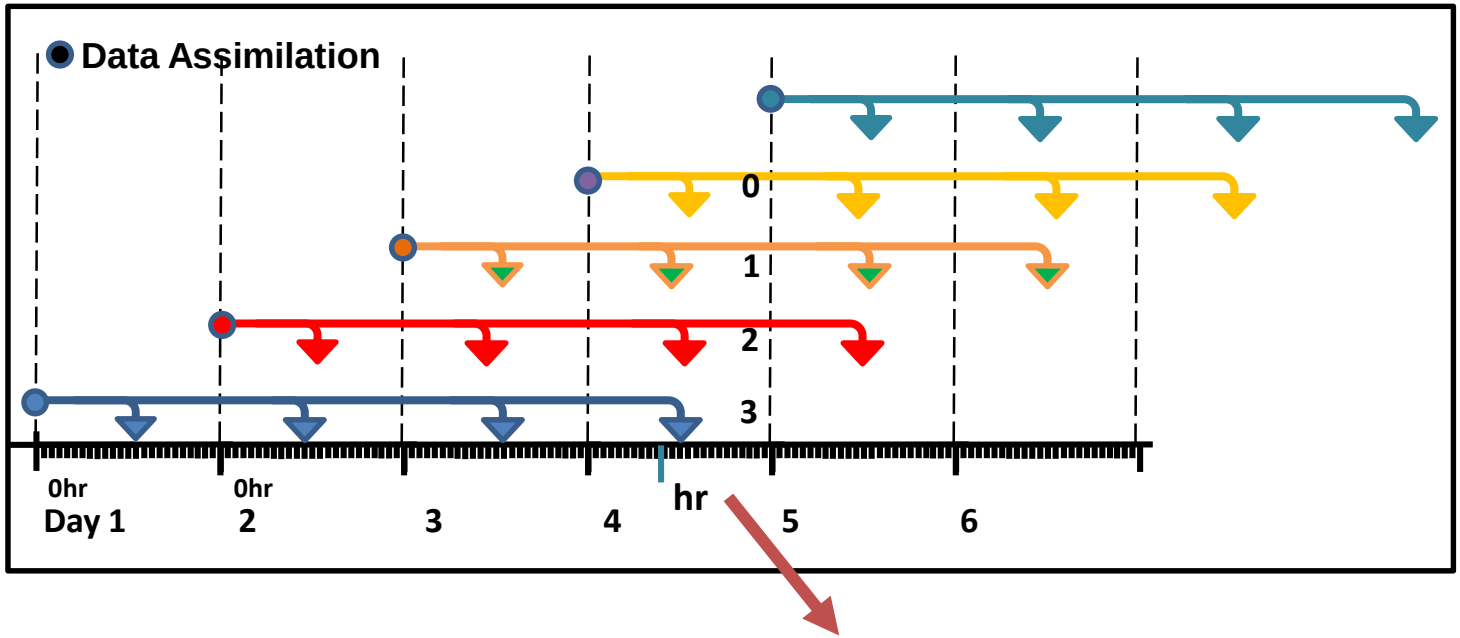
Converted into model files of size (700, 420, 8760) with id FC 0, 1, 2, 3

NO2, O3, PM2.5, PM10

ES, FR, DE, IT/POV, PL, EUR

EEA stations for Validation (> 3000)

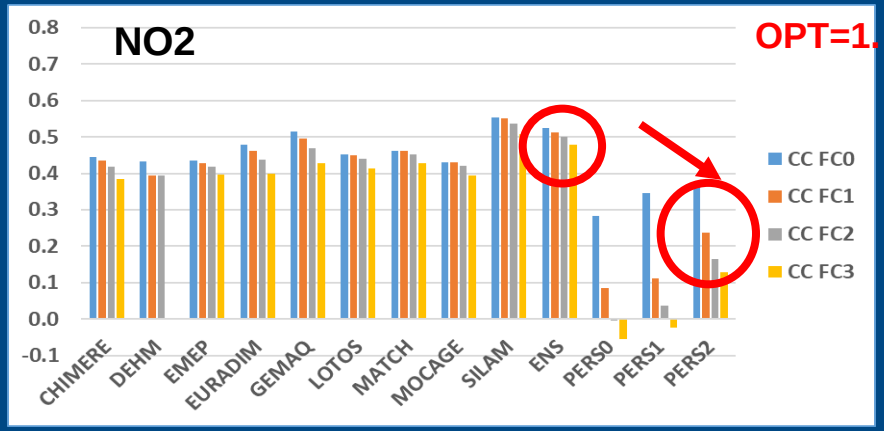
Model Forecast schematically

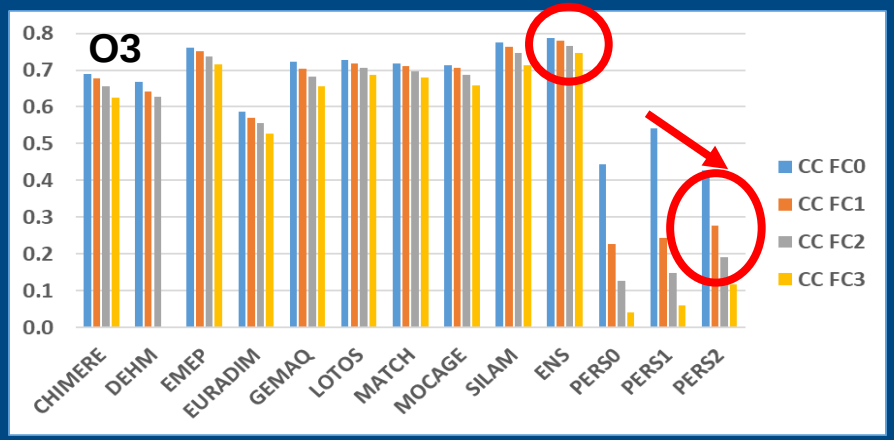


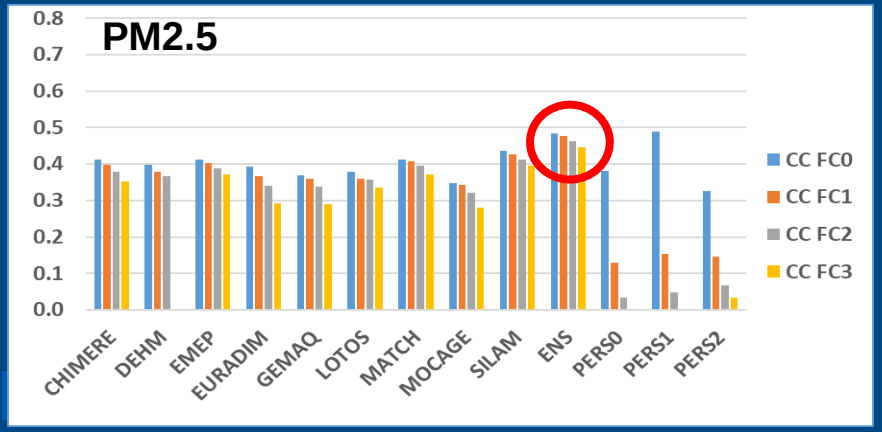
In general, each hour/day contains 4 types of hourly FC data (D0, D-1, D-2, D-3)

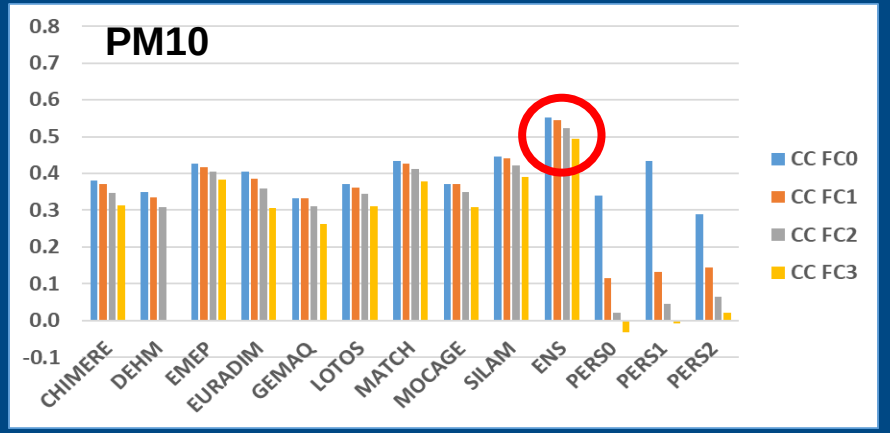
CorrCoef for FC0, FC1, FC2, FC3

JAN (JUL) (all ES, FR, DE, POV, PL stations)



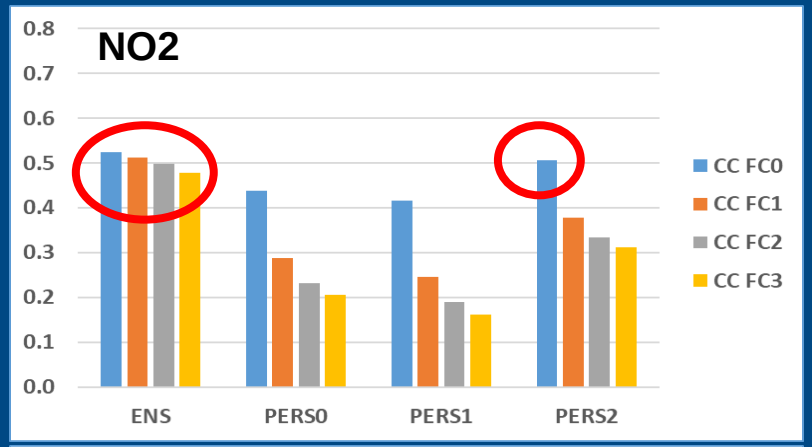


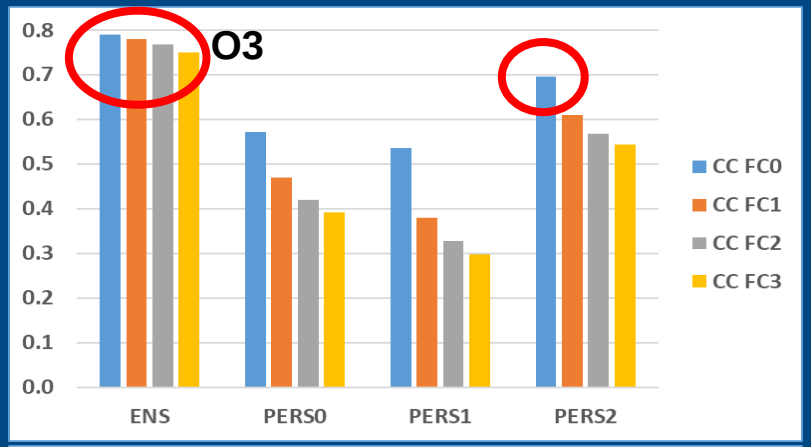


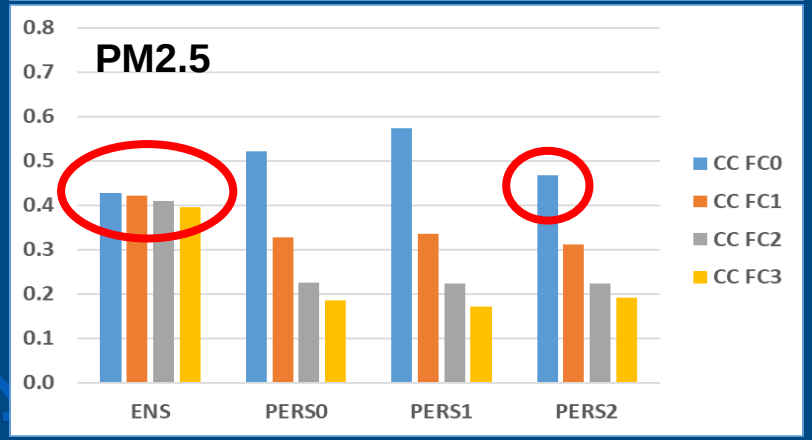


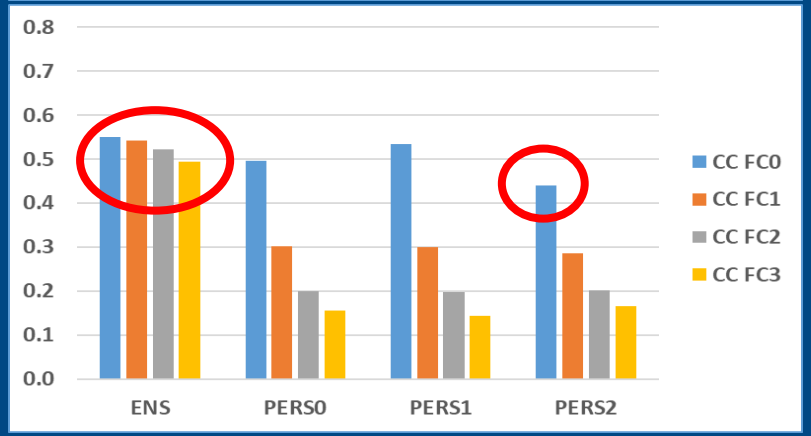
CorrCoef ENS, PERSx for FC0, FC1, FC2, FC3

YEAR (all ES, FR, DE, POV, PL stations)



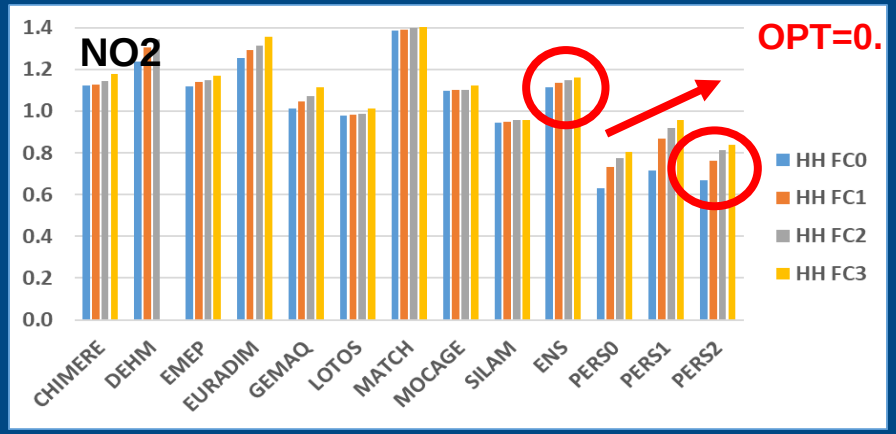


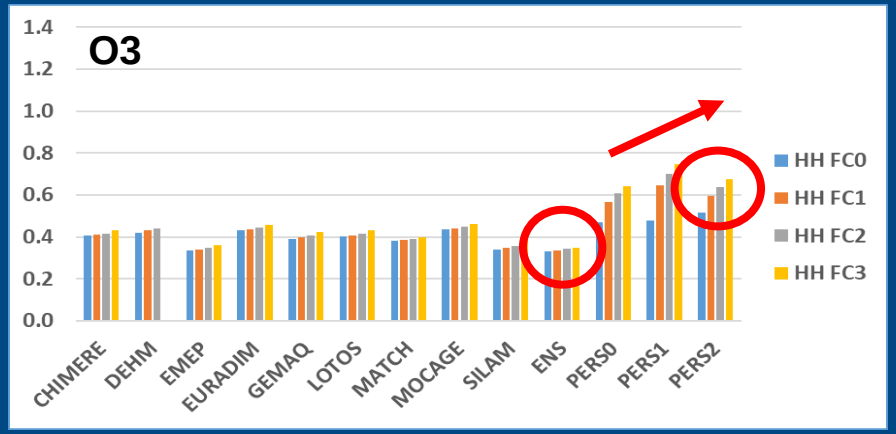


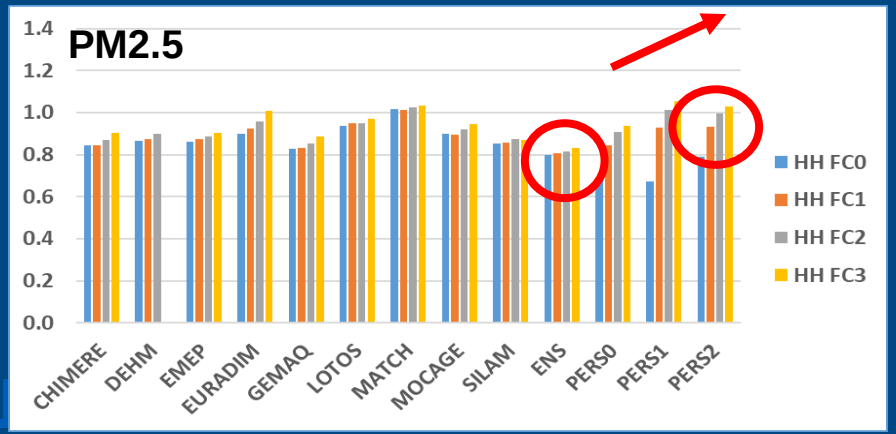


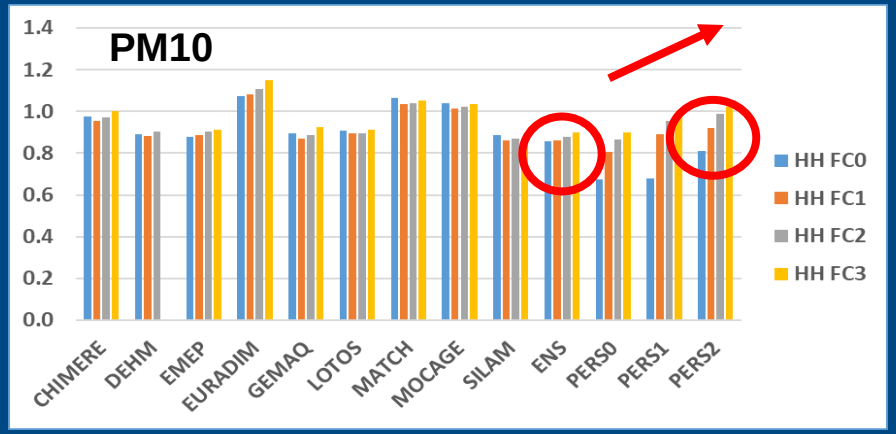
NRMSE (HH) for FC0, FC1, FC2, FC3

JAN (JUL) (all ES, FR, DE, POV, PL stations)



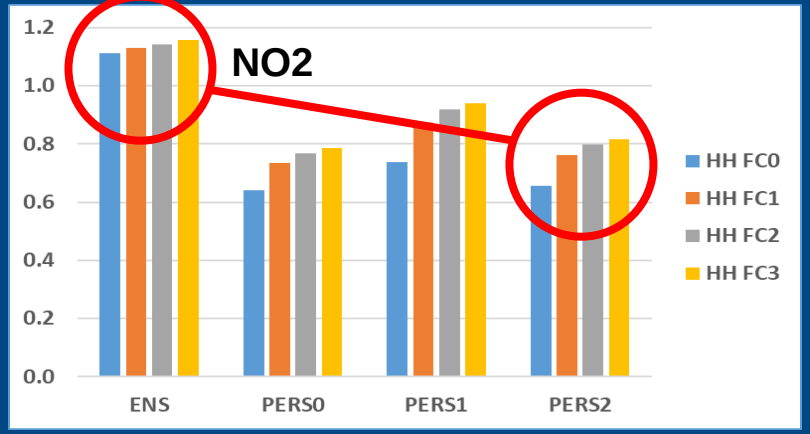


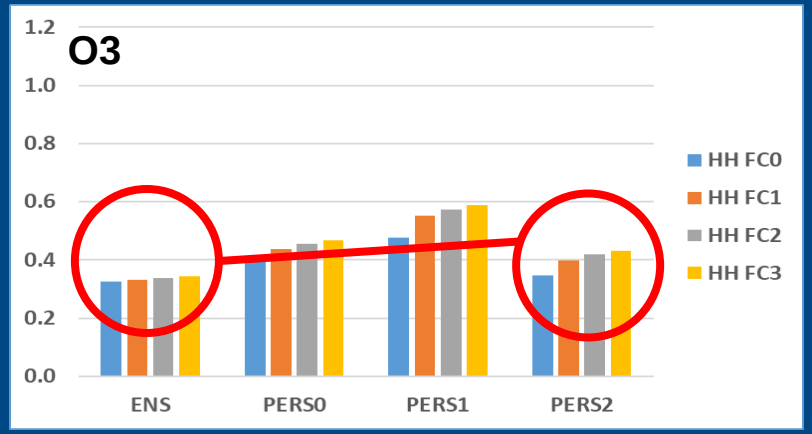


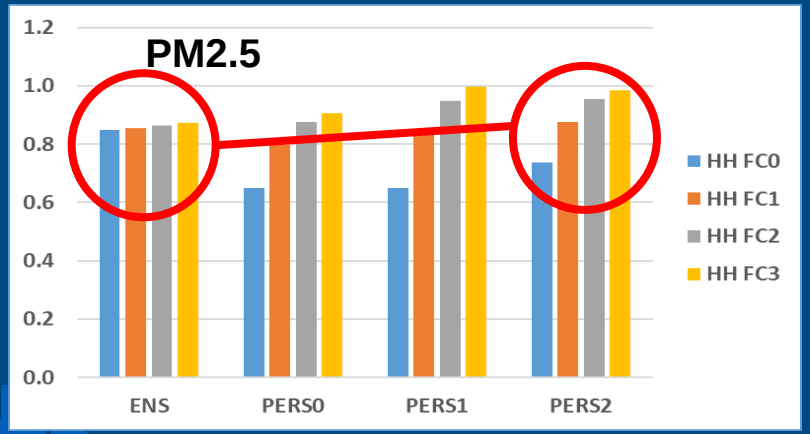


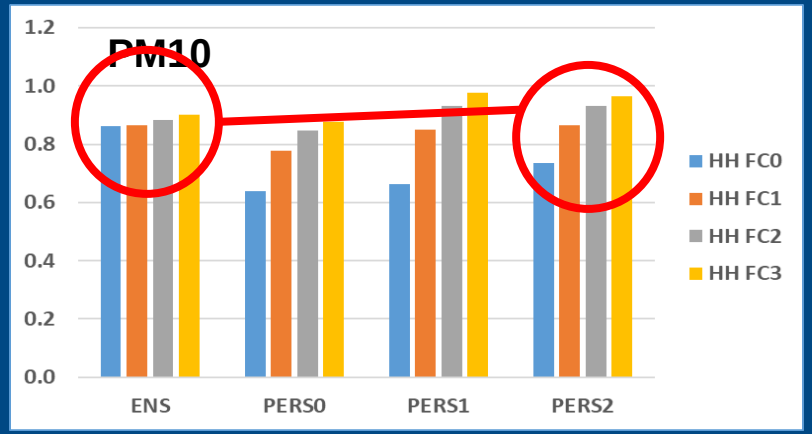
NRMSE ENS, PERSx for FC0, FC1, FC2, FC3

YEAR (all ES, FR, DE, POV, PL stations)

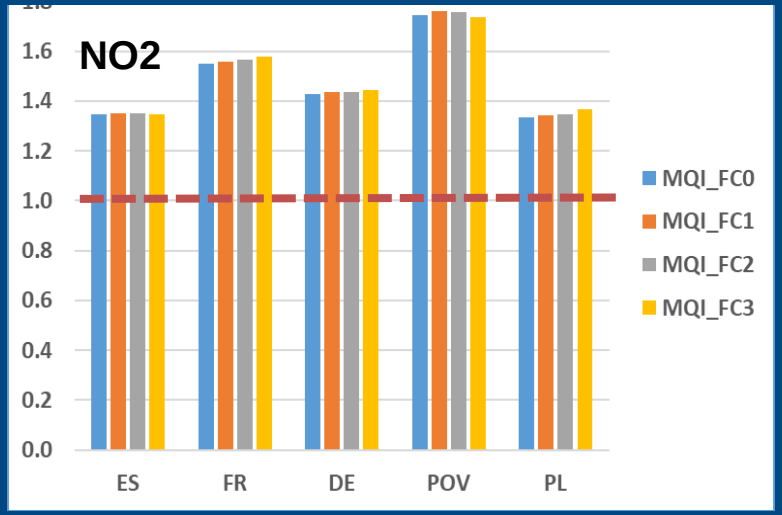


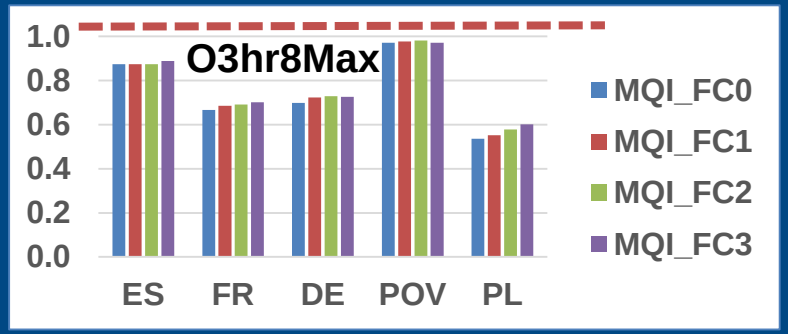


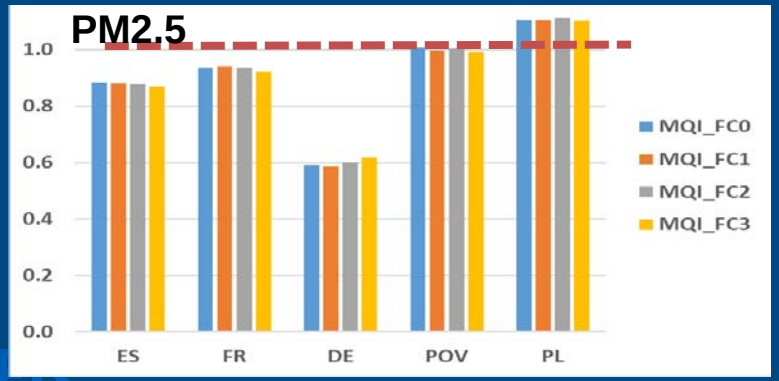


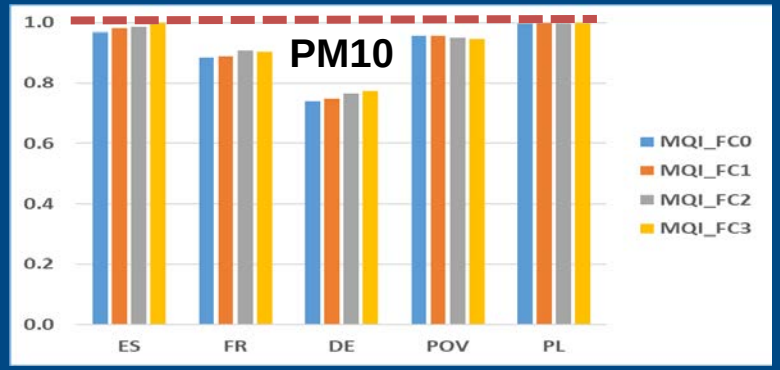


ENS (Year) MQI_FC for FC0, FC1, FC2, FC3









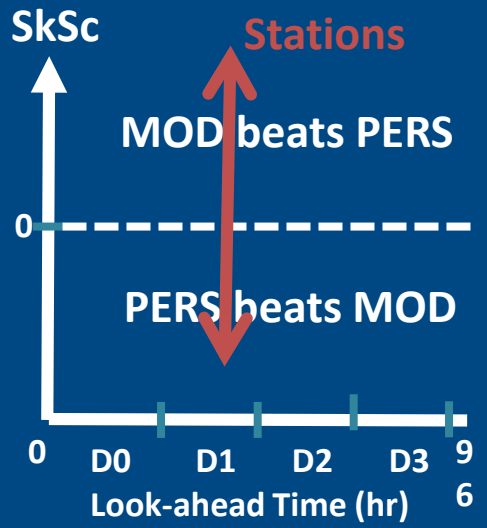
Look-ahead time

For each Forecast hour (00-23) we have 365 values of FC0 + OBS => Indic / Statistics
For each Forecast hour (24-47) we have 365 values of FC1 + OBS => Indic / Statistics
etc

Indicator/Statistics (IND) examples: Correlation, RMSE, BIAS, …

$$\text{SkillScore} = \begin{cases} \dfrac{\text{IND(Mod)} - \text{IND(Pers)}}{1 - \text{IND(Pers)}} & \text{for Correlation} \\ \dfrac{\text{IND(Pers)} - \text{IND(Mod)}}{0 - \text{IND(Pers)}} & \text{for RMSE} \end{cases}$$

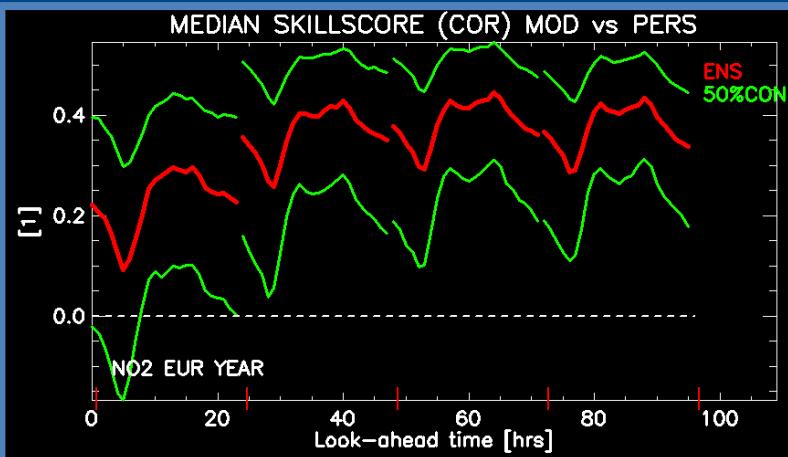
Pers(ih) is value at ih at Day -1, -2, -3, -4

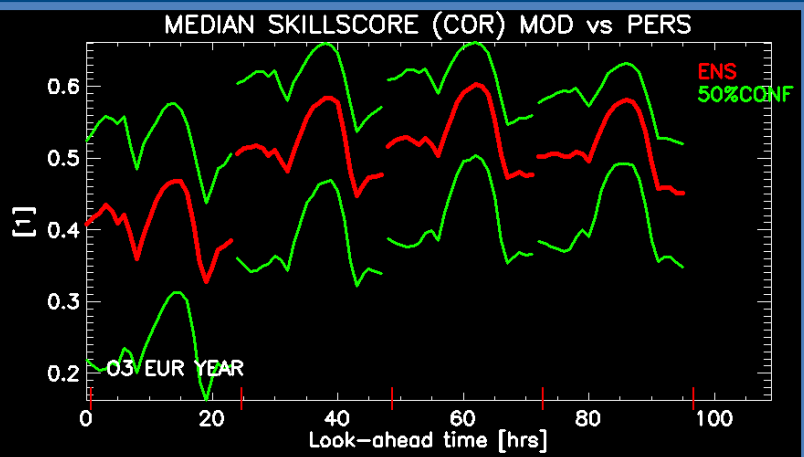


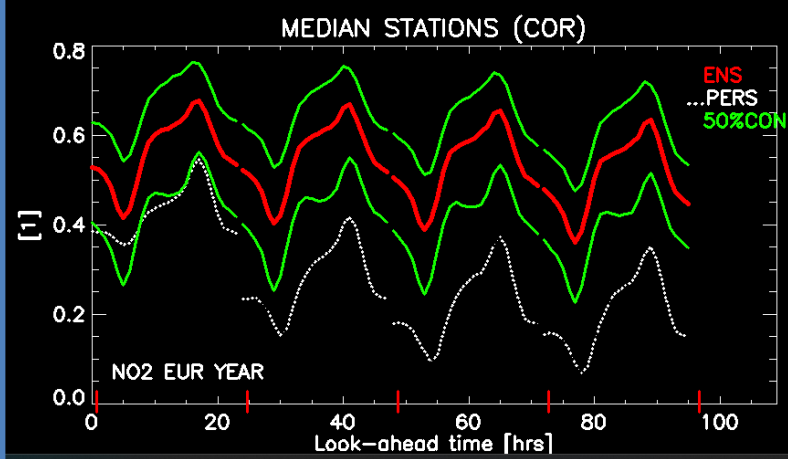
NO2 EUR YEAR COR O3

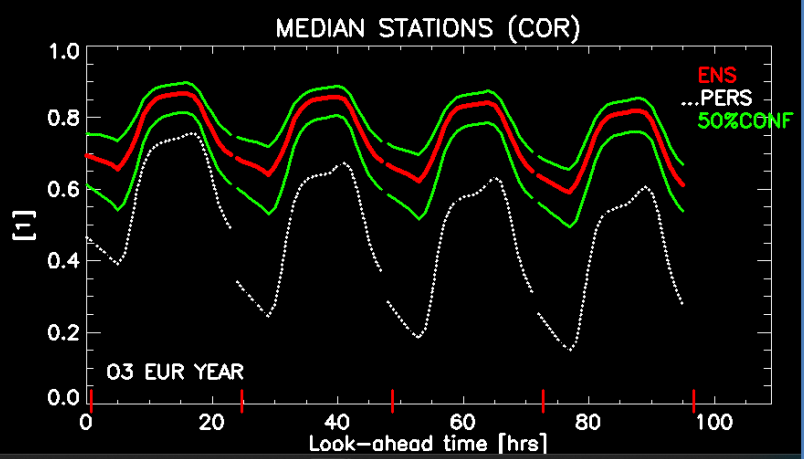
SkillScore

Median







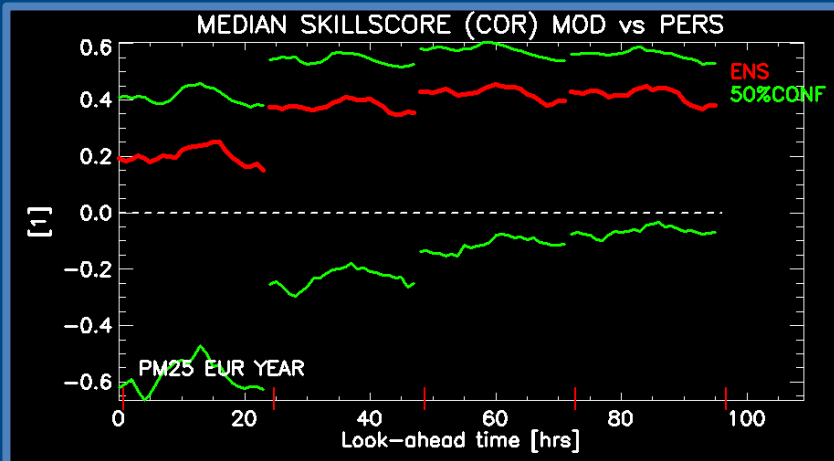


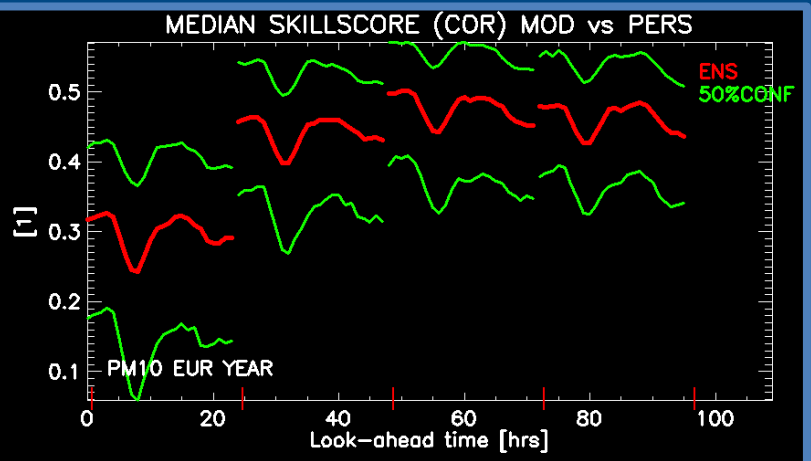
EUR YEAR COR

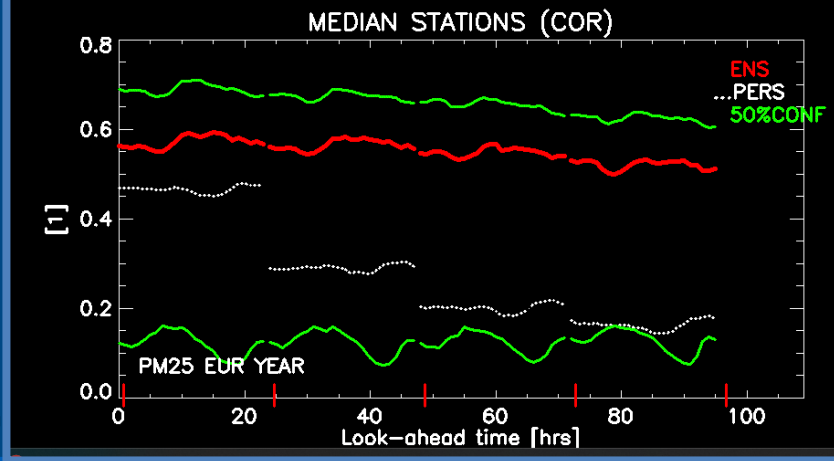
PM2.5 PM10

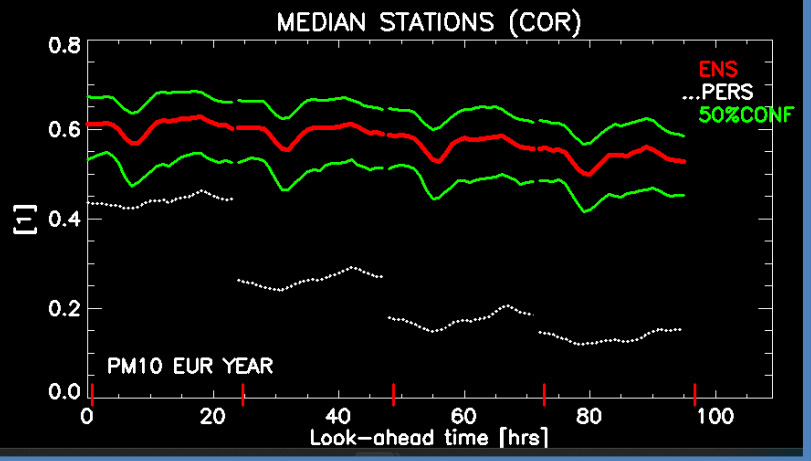
SkillScore

Median







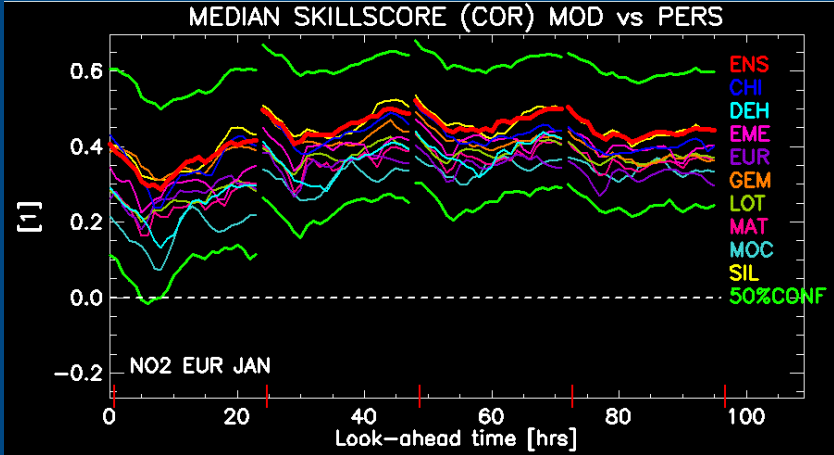


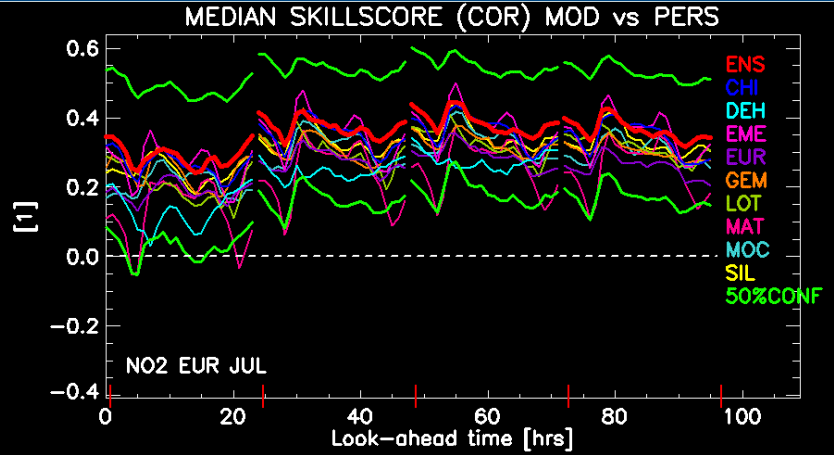
EUR SKILLSCORE COR

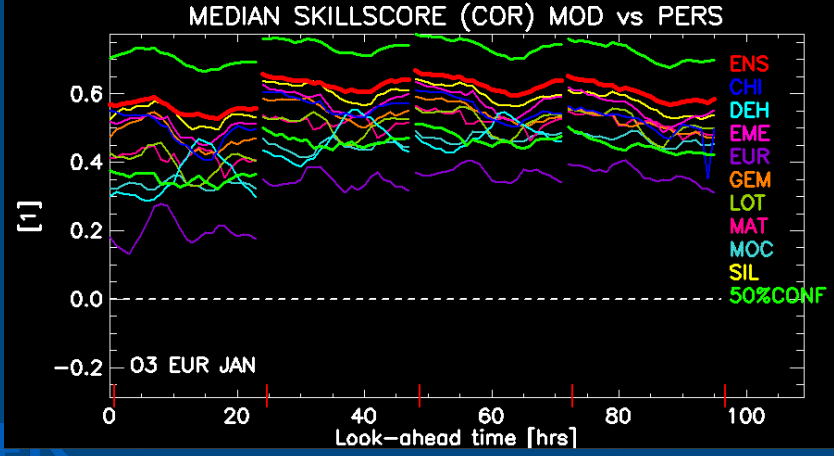
JAN JUL

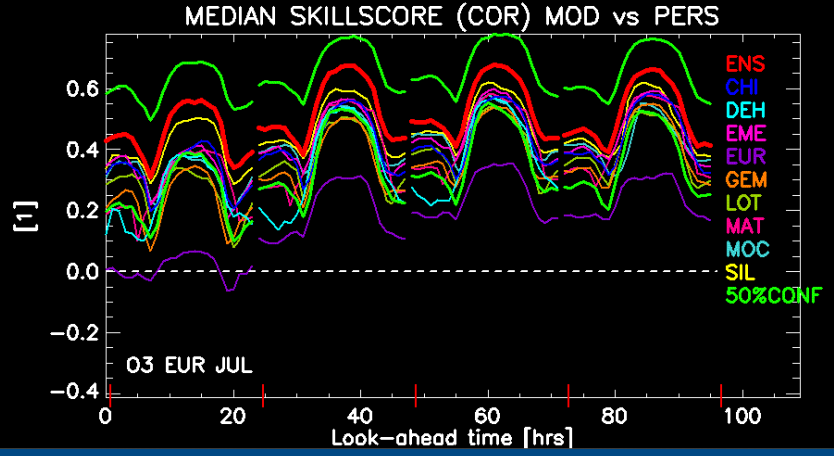
NO2

O3





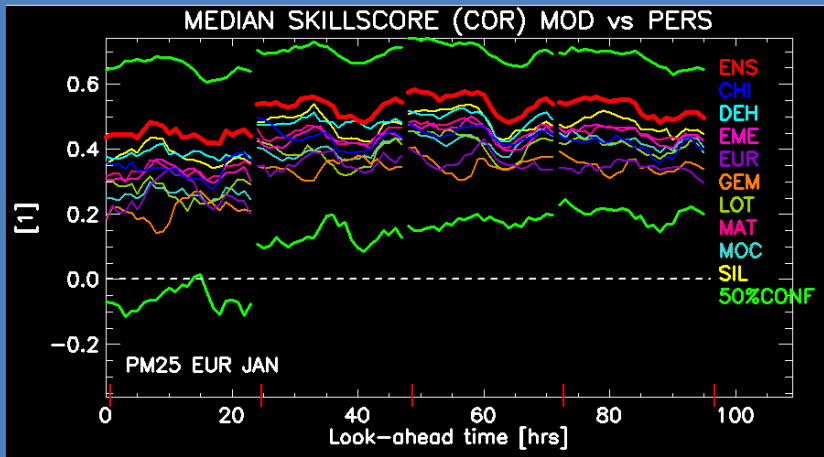


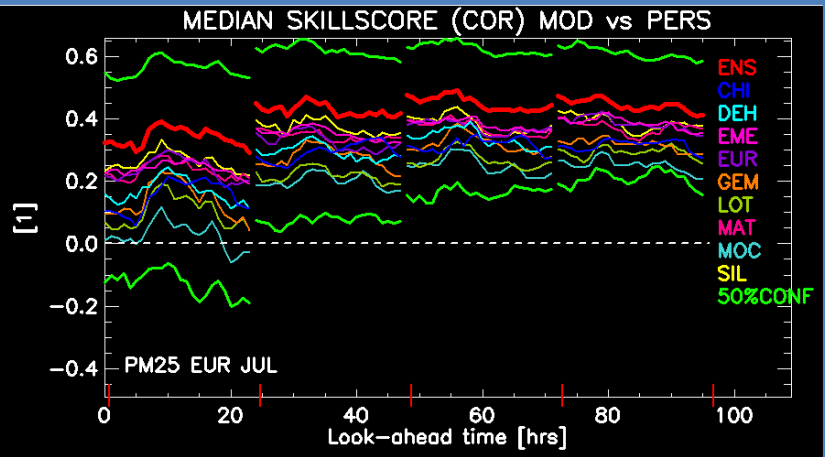


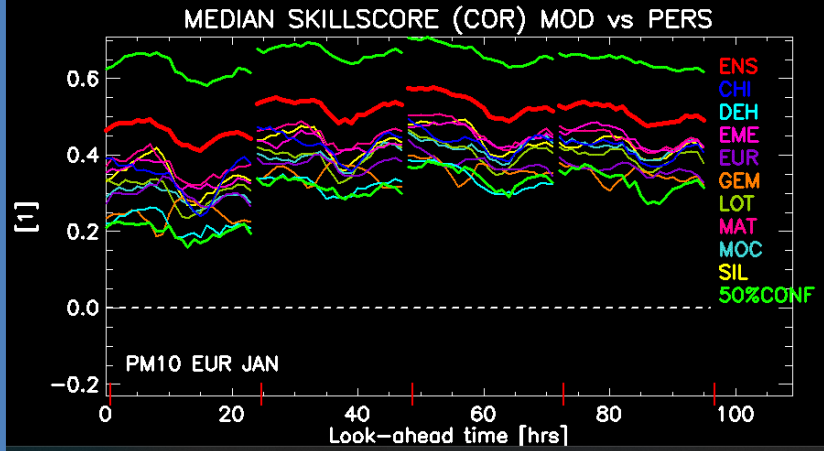
EUR SKILLSCORE COR

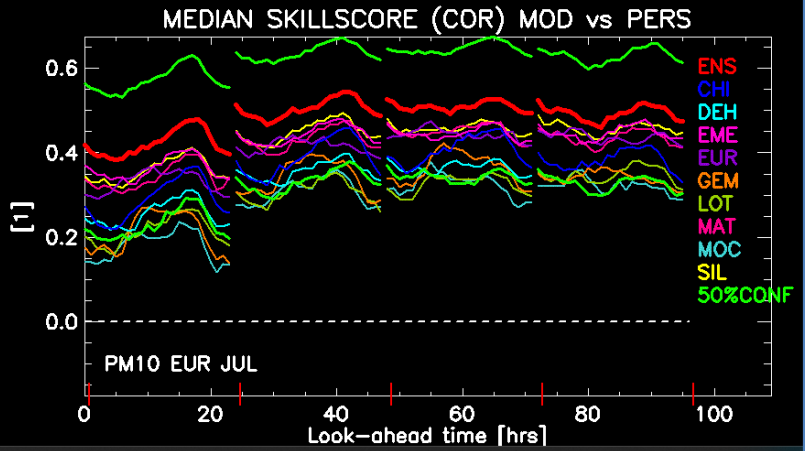
JAN JUL

PM2.5 PM10





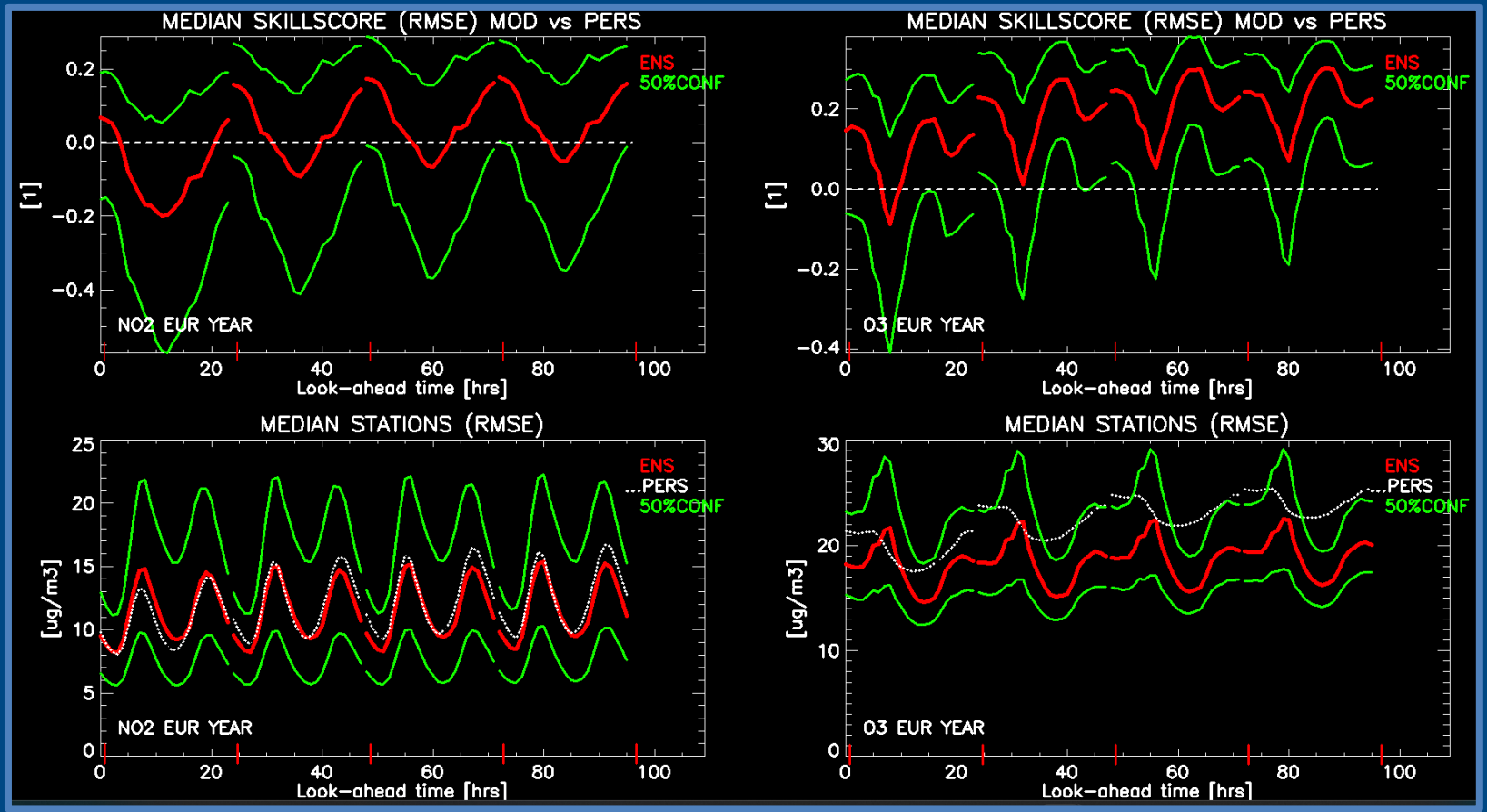




NO2 EUR YEAR RMSE O3

SkillScore

Median

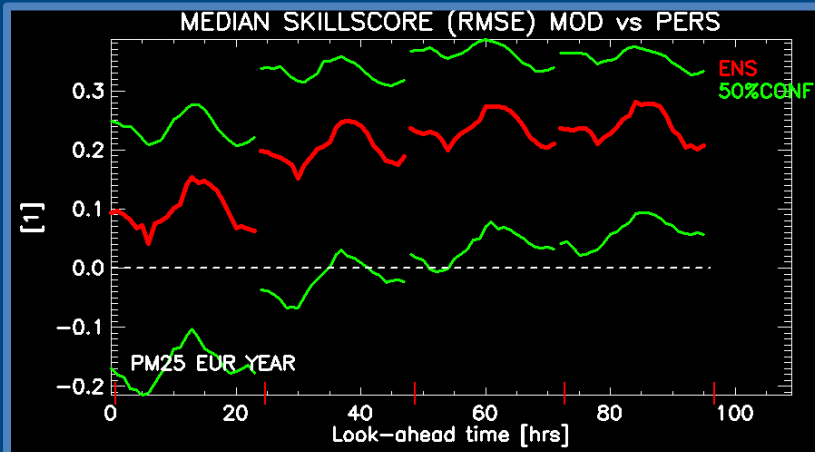


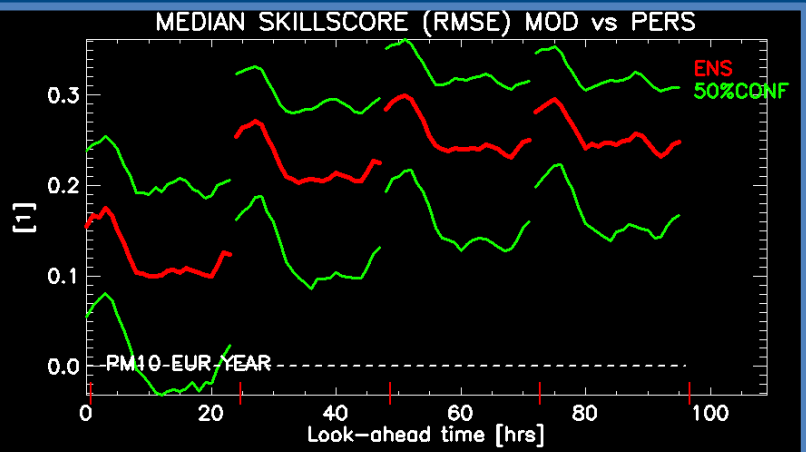
EUR YEAR RMSE

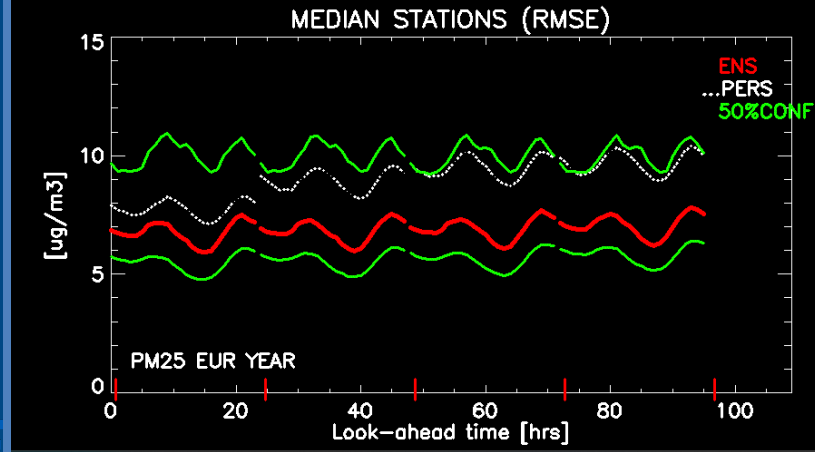
PM2.5 PM10

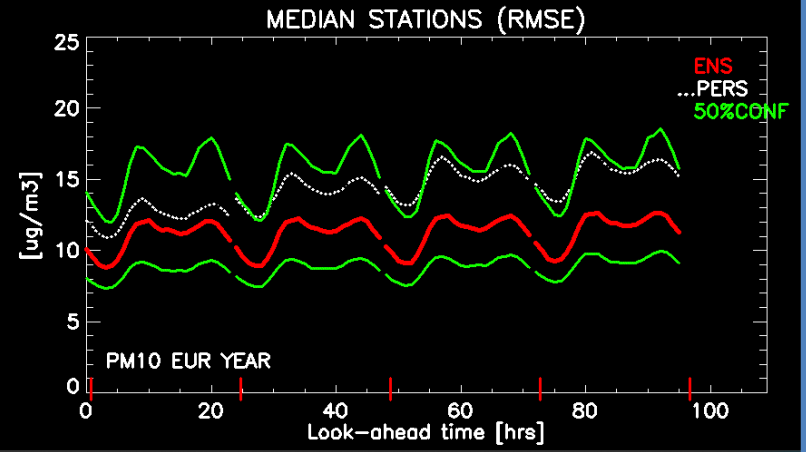
SkillScore

Median







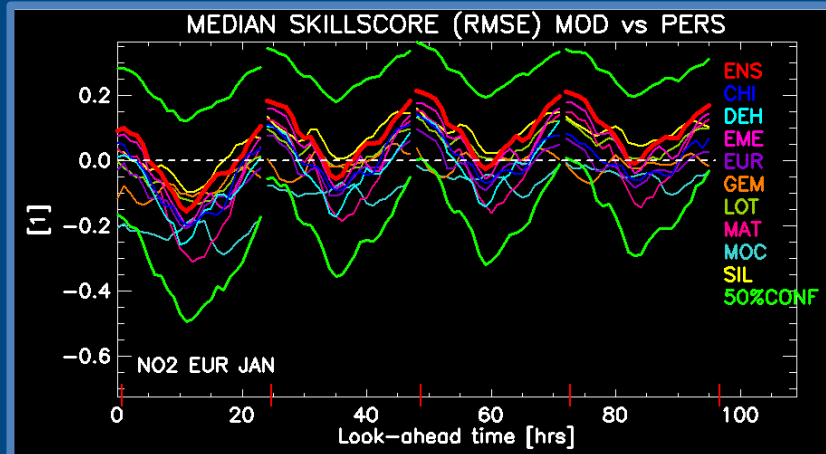


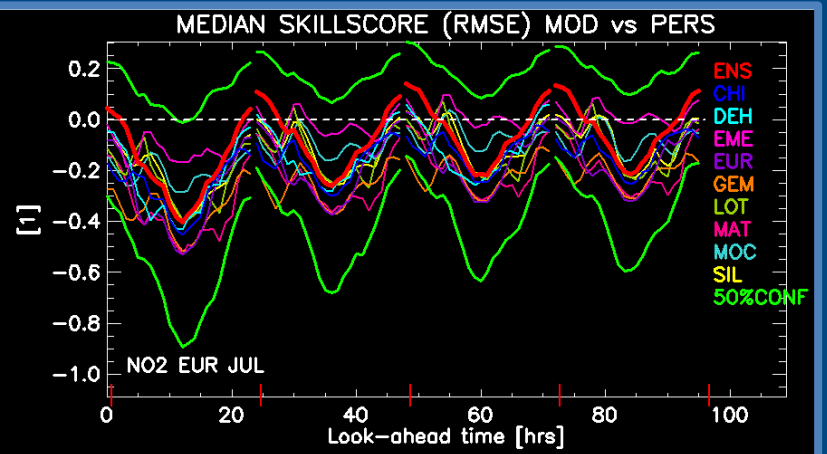
EUR SKILLSCORE RMSE

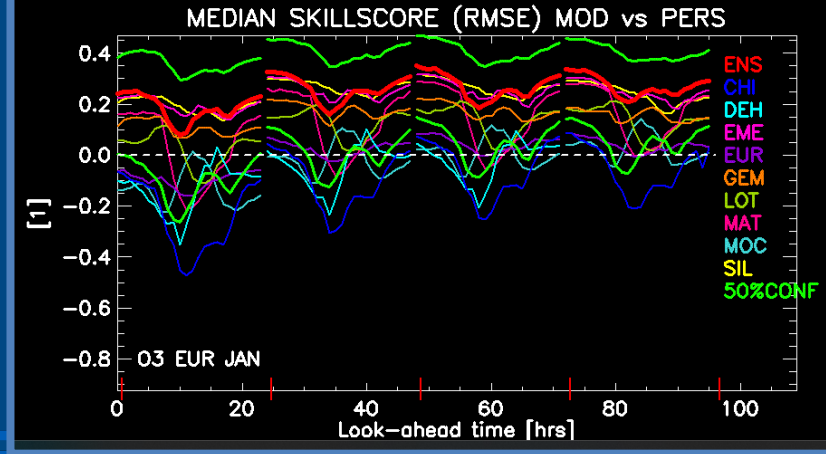
JAN JUL

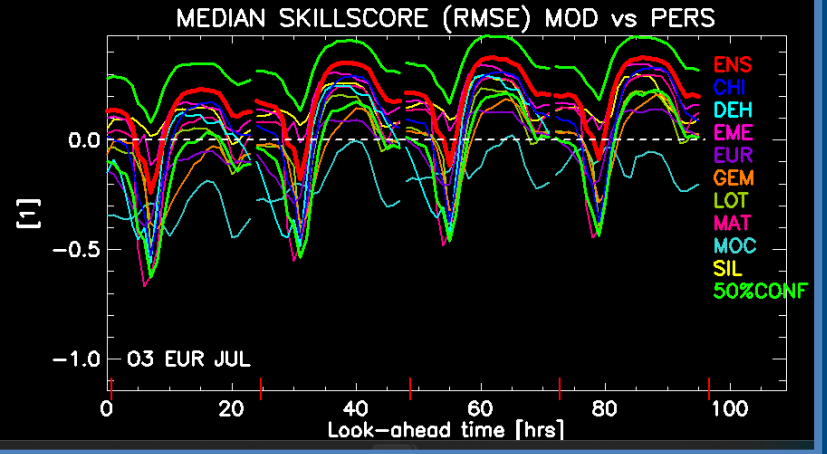
NO2

O3







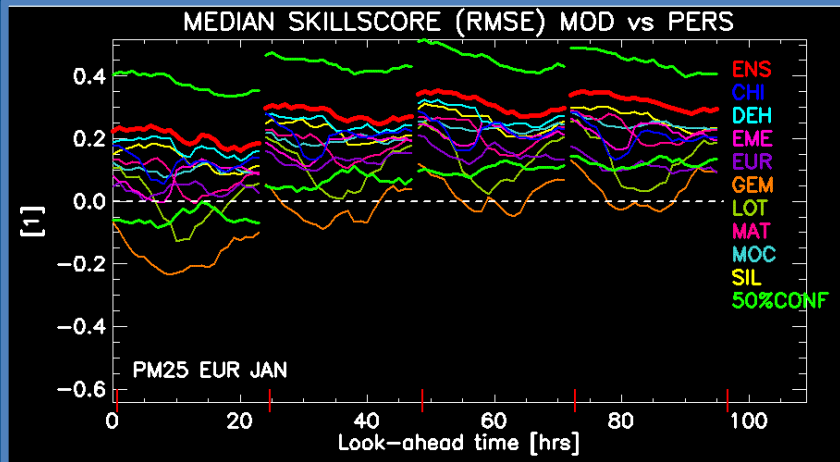


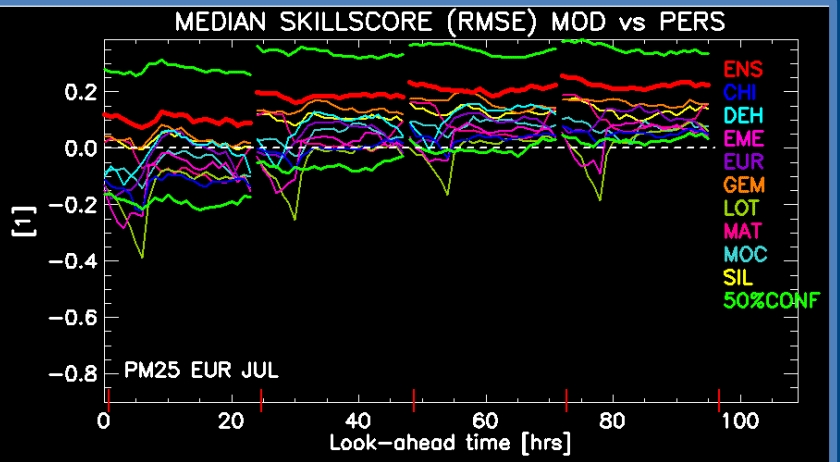
EUR SKILLSCORE RMSE

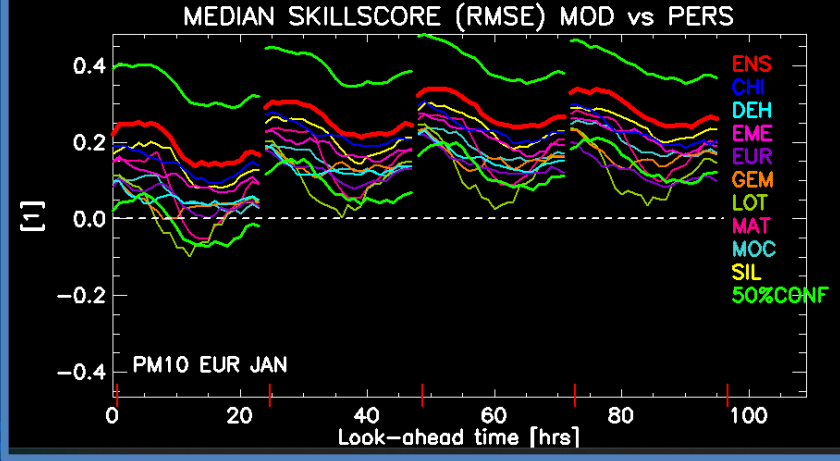
JAN JUL

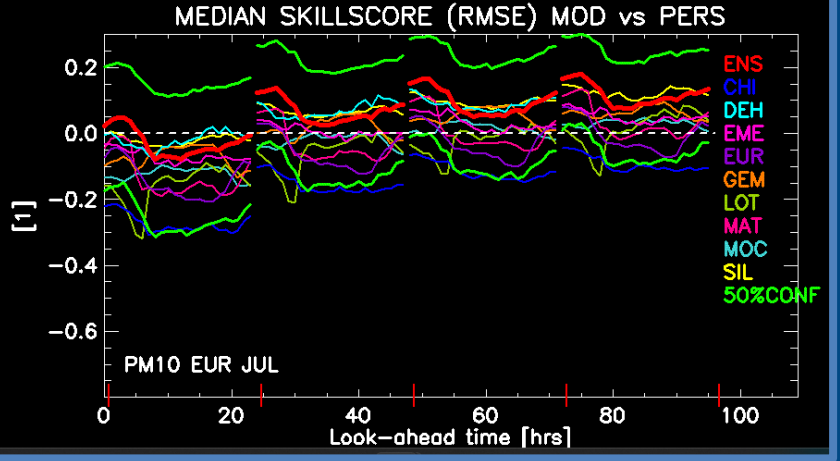
PM2.5

PM10

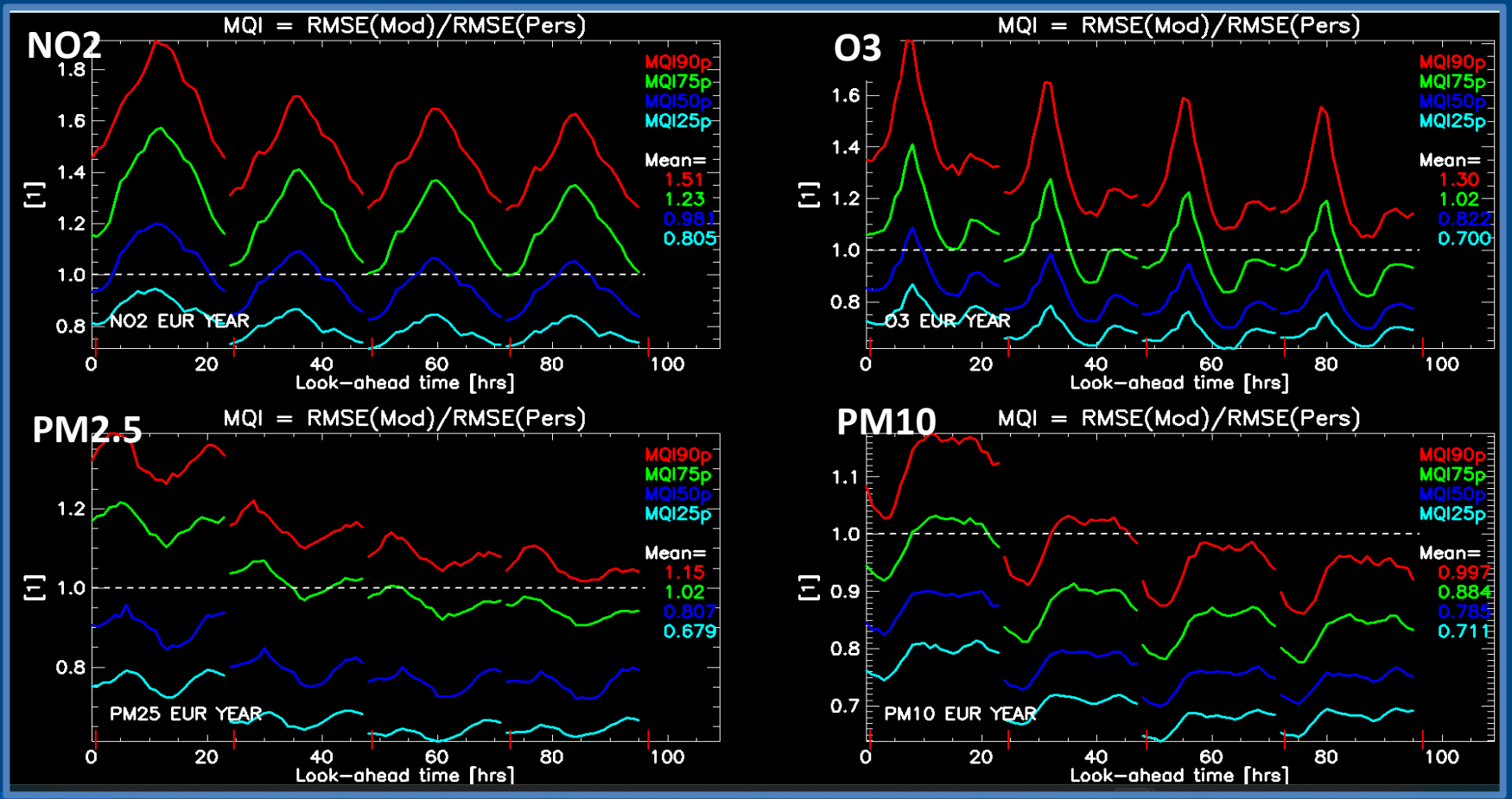








MQI EUR YEAR


NO2
MQI = RMSE(Mod)/RMSE(Pers)
MQI90p
MQI75p
MQI50p
MQI25p
Mean=
1.51
1.23
0.981
0.805
[1]
NO2 EUR YEAR
Look-ahead time [hrs]
O3
MQI = RMSE(Mod)/RMSE(Pers)
Mean=
1.30
1.02
0.822
0.700
O3 EUR YEAR
Look-ahead time [hrs]
PM2.5
MQI = RMSE(Mod)/RMSE(Pers)
Mean=
1.15
1.02
0.807
0.679
PM25 EUR YEAR
Look-ahead time [hrs]
PM10
MQI = RMSE(Mod)/RMSE(Pers)
Mean=
0.997
0.884
0.785
0.711
PM10 EUR YEAR
Look-ahead time [hrs]

Similar pictures for

Seasons: JAN, JUL, YEAR, DJF, MAM, JJA, SON

Statistics: COR, RMSE, NRMSE, NMAE, MBIAS, NBI

Regions: EUR, ES, FR, DE, IT/POV, PL

Conclusions:

- ENS not always the 'best' model, oké for CORR, not for RMSE
- Persisting issue with the Persistence model !
- Is this what we want in the context of the AQ directives ?
- Focus on Forecast of Exceedance days, AQ index (categories) ?

OTHER POINTS FOR DISCUSSION

B1. Should we persist with the Persistence Model comparison?

B2. Should we look at the comparison with the Persistence Model in a more comprehensive way?

C1. Should we focus on setting criteria for the evaluation of the capability in forecasting Exceedance days?

i.e. defining objective criteria for threshold exceedance's indicators (at least POD, SR, ACC), also considering the new daily limit for PM2.5

C2. Ideas for point C1.?

- statistical analysis on a large amount of data, covering different contexts and air quality regimes
- …..

POINTS FOR DISCUSSION

A1. Shall we introduce β within Forecast MQI formulation?

A2. Ideas for setting β values?

- preserve the consistency with current AQUILA-based Assessment MQO
- preserve the consistency with previous Forecast MQO
- statistical analysis on a large amount of data to set the level of stringency that makes the X-percentile of the simulations to turn out fit-for-purpose
- .....

B1. Shall we persist with the Persistence Model comparison?

B2. Should we look at the comparison with the Persistence Model in a more comprehensive way?

C1. Should we focus on setting criteria for the evaluation of the capability in forecasting Exceedance days?

i.e. defining objective criteria for threshold exceedance's indicators (at least POD, SR, ACC), also considering the new daily limit for PM2.5

C2. Ideas for point C1.?

- statistical analysis on a large amount of data, covering different contexts and air quality regimes
- .....

Open discussion, we really need FAIRMODErs contributions to elaborate a roadmap!

→ Tomorrow we will report the results of the discussion

ENEA universidade de aveiro theoria poiesis praxis European Commission FAIRMODE

POINTS FOR DISCUSSION #2

- Two different needs: communication to stakeholders (agencies, policy. Etc.) and “expert” evaluation
- Stakeholders need a clear indication: performance on exceedances (Norway, Germany, Portugal…).
  This is reinforced by the new AAQD!

→ Working on ranges of acceptability categorical indicators for exceedances (POD, SR, ACC…), collecting data + best practices and analyzing CAMS 2021 data

POINTS FOR DISCUSSION #2

- Now we have 2 different MQIf, based on FAIRMODE-old and AQUILA-CEN-new uncertainty parameters: bear this in mind when presenting the MQI to stakeholders!!!
- If we kill the MQIf today, would you be happy? Or upset?

→ Survey on actual use of MQI in Fairmode:

→ no?

→ yes, for stakeholders?

→ yes, for expert evaluation?

For deciding about investing or not time on new analysis of CAMS 2021 (and/or other data)

FUTURE CHALLENGES AND PROPOSAL FOR THE NEXT STEPS

- .....