# Forecast Benchmarking methodology within DELTA Tool

## ENEA's feedback on the Proposed New Forecast Plots

*26/05/2021 - FAIRMODE CT3 hackathon on forecast evaluation*

*Lina Vitali, Antonio Piersanti*

**ENEA, Laboratory of Atmospheric Pollution**

# BACKGROUND

At last FAIRMODE CT3 hackathon (09/02/2021) feedback was provided about how the comparison with the persistence model is performed within the Delta Tool

➢ *Is the persistence model a realistic benchmark?*

➢ *Which is the best way to compare model performances to persistence model ones?*

$$MQI_{forecast} = \sqrt{\frac{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2}}$$

$$P_i = O_{i-1-forecast\ horizon}$$

SOME **SHORTCOMINGS** OF

THE ORIGINAL $MQI_{forecast}$ FORMULATION

POINTED OUT AT LAST CT3 HACKATHON

✗ <u>different ranking of model performances</u> is observed, when comparing Forecast and Assessment Target Plots, despite they are based on the same statistical indicator (RMSE) [*ENEA presentation*]

✗ according to Forecast MQI outcomes <u>model performances seem to get better along with forecast days</u> [*Arpa Puglia presentation*]

ENEA

# FORECAST PLOTS – NEW PROPOSED FORMULATIONS

New Formulations and related Plots
proposed during and after the Hackathon
and implemented by Kees Cuvelier within a test version of the Delta Tool

A. **Forecast Target Plot (OU)**: similar to the current Forecast Target Plot, but based on a normalization term, taking into account not only the persistence model performances but also the observation uncertainty (OU).

B. **Forecast MQI&MFE Plot**: MFE is used instead of RMSE; in the final plot, not only the ratio (MQI) between forecast and persistence models skills is shown, but also the forecast performances score itself (MFE).

ENEA

$$MQI_{forecast} = \sqrt{\frac{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2}}$$

$$P_i = O_{i-1-forecast\ horizon} \pm OU(O_{i-1-forecast\ horizon})$$

➤ the main impact of adding uncertainty in the formulation is to improve both the bias and correlation

➤ including the observation uncertainty, we prevent the denominator to tend to zero (issue highlighted by ENEA) and therefore the overall indicator cannot tend to infinity

➤ since larger relative uncertainties are expected in the low concentration range, a good forecast at low concentration values is now less important than it was in the original formulation

# EXTRA VALUES

The **Forecast Target Plot (OU)** requires 3 extra values: Val1#Val2#Val3#, with

Val1 = threshold

Val2 = cut off value (*)

Val3 = forecast horizon (≥0)

(*) *Val2=0 in all the following plots*

ENEA

ENEA MAIN MESSAGES & SUGGESTIONS, DURING THE HACKATHON

1) We cannot look only at the ratio (MQI), as the role of the denominator may be too relevant.

→ *The plot should show not only MQI but also model performances score itself*

2) Using the normalized error could be useful to remove some artefacts due to possible differences in concentration values magnitude.

→ *from RMSE to MFE*

3) The criterion seems too demanding. It could be necessary introducing a β value greater than 1, to be lowered in perspective as the model performances increase.

→ *β > 1 in MQI formulation*
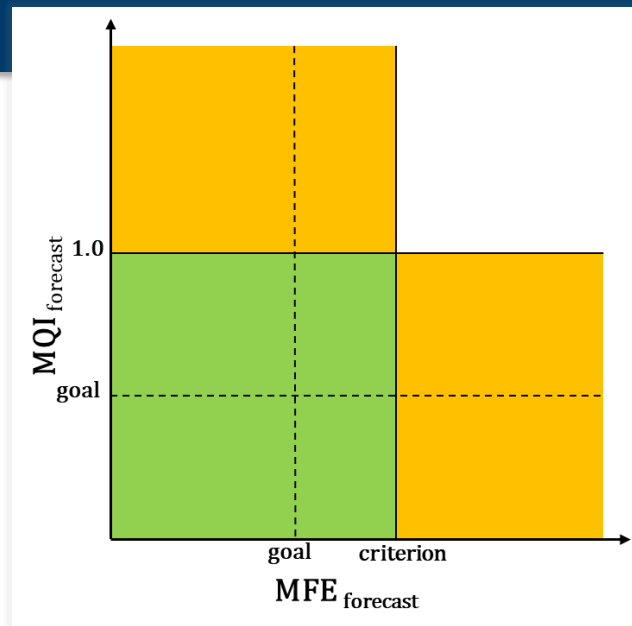
$$MQI_{forecast} = \frac{MFE_{forecast}}{\beta \, MFE_{persistence}}$$

$$MFE = \frac{2}{N} \sum_{i=1}^{N} \frac{|M_i - O_i|}{(M_i + O_i)}$$

# EXTRA VALUES

*"The diagram is requesting 5 input parameters. In a final version we can fix some of them or make them pollutant dependent. They are:*

*Forecast Horizon (0,1,2,3...), MFE goal (example 0.5), MFE criterion (example .75), MQI goal (example 0.5), and β (example 2.0)".*

(Kees Cuvelier, 11/02/2021)
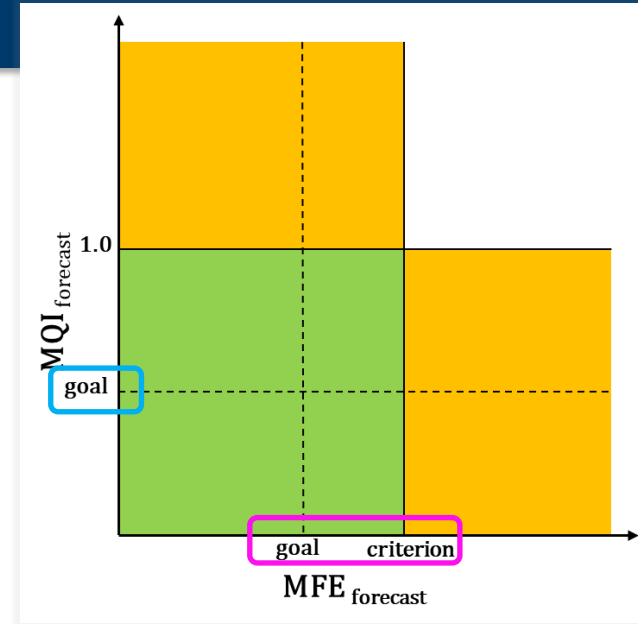


$$MQI_{forecast} = \frac{MFE_{forecast}}{\beta \cdot MFE_{persistence}}$$

$$MFE = \frac{2}{N} \sum_{i=1}^{N} \frac{|M_i - O_i|}{(M_i + O_i)}$$

## EXTRA VALUES – VALUES SET IN THE FOLLOWING PLOTS

### MFE_goal & MFE_criterion

I started my tests with PM10 and $O_3$, for whom MFE goals and criteria are suggested in literature.
So I fixed them as

PM10      (Boylan and Russell, 2006):      **MFE_goal= 0.50**      **MFE_criterion=0.75**

$O_3$      (Chemel et al., 2010):      **MFE_goal=0.30**      **MFE_criterion= 0.45**

### β & MQI_goal

I tested two β values: β=2; β=1.5. MQI_goal was set according to β choice, since, concerning MQI outcomes, the dashed lines usually delimit the area where the «*ratio of the scores ≤ 1*».

i.e      *MQI_goal x β=1*   →   *MQI_goal=1/β*

For both PM10 and $O_3$, I tested 2 combinations (β=2 & MQI_goal=0.5) and (β=1.5 & MQI_goal=0.67).
In the following, outcomes are presented for the first combination (**β=2 & MQI_goal=0.5**).

### forecast horizon

I applied the approach to my data set of 3-day forecasts. **Forecast days=0,1,2** were available.

$MQI_{forecast}$(OU) AND FORECAST MQI&MFE
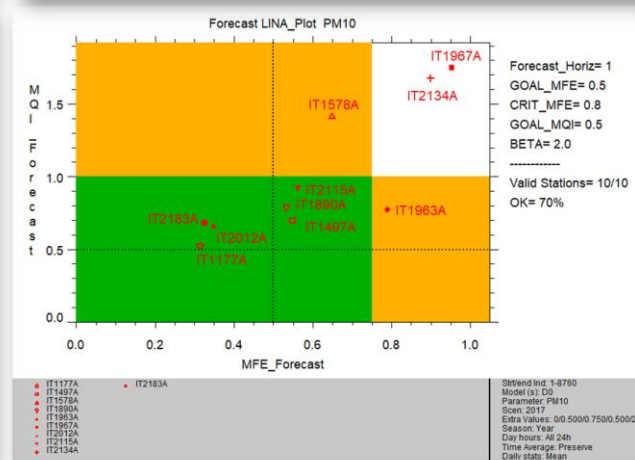
BEHAVIOR

CONCERNING THE MAIN SHORTCOMINGS OF THE ORIGINAL $MQI_{forecast}$ FORMULATION
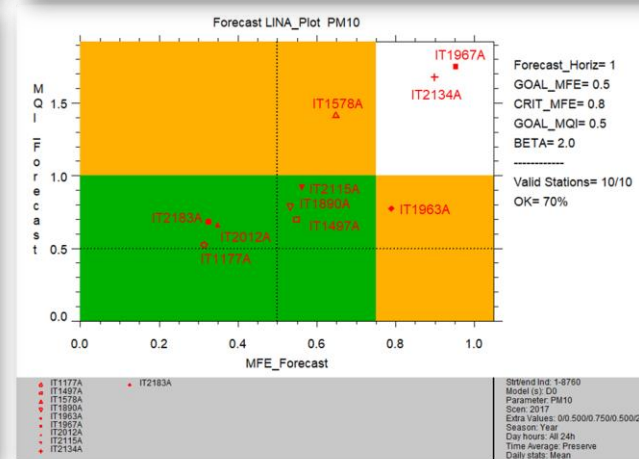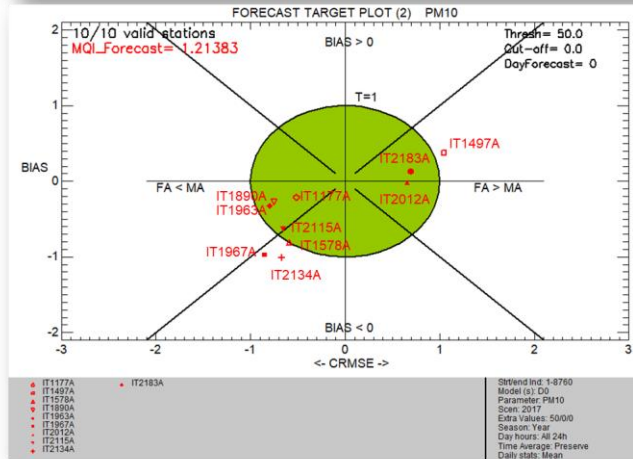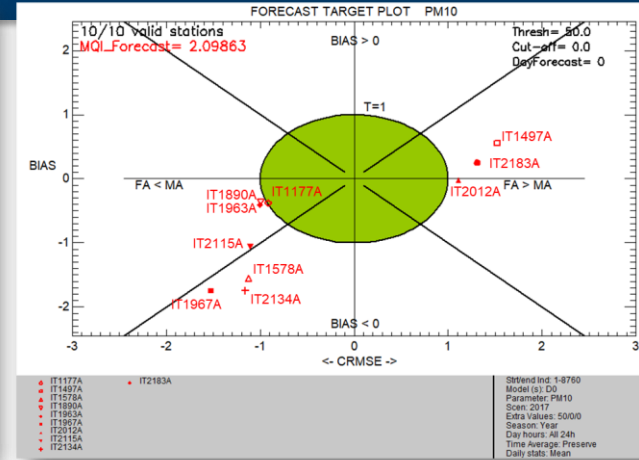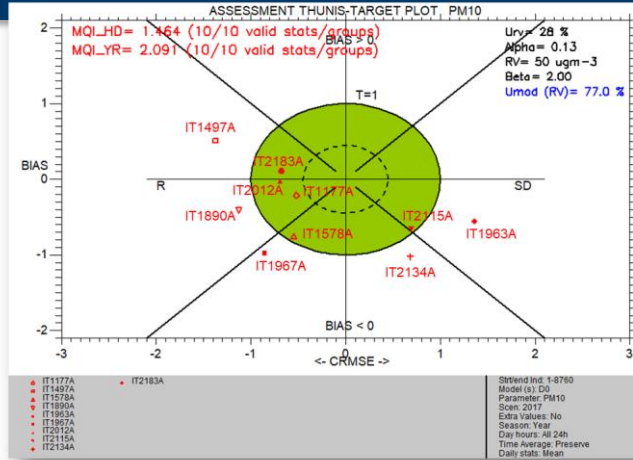
1. Is their behavior consistent concerning the <u>ranking of model performances</u>?

2. According to their outcomes <u>how do model performances change along with forecast horizon</u>?

**The new proposed FORECAST PLOTs outcomes are presented on top of current $MQI_{assessment}$ and $MQI_{forecast}$ outcomes.**

**ENEA**

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

**MODELLING SYSTEM DEVELOPMENT**
**SIMULATIONS OPERATIONAL MAINTENANCE**

*Mario Adani, Massimo D'Isidoro*



ENEA

THE MODELLING SYSTEM

**FORAIR - IT**
**3 Days Air Quality Forecasts**

Europe, 20 km
3 days

Italy, 4 km
3 days

Air quality forecasts over Italy
at 4km horizontal resolution

http://www.afs.enea.it/project/ha_forecast/

## PM10 available stations: ITA 2017 → 10 stations test case



Available background stations: 248

10 background monitoring sites were chosen, located in several geographical areas and emissions contexts, and with different performances

> When comparing Assessment and Forecast Target Plots, according to the original formulation, the ranking of model performances is noticeably different as already stated at February Hackathon

➤ When comparing Assessment and Forecast Target Plots, according to the original formulation, the ranking of model performances is noticeably different as already stated at February Hackathon

➤ When comparing Assessment and Forecast Plots, according to both the new formulations, still the ranking of model performances is someway different but generally consistent

*For example, according to both the new formulations, performances at IT2012A and IT2183A look better than at IT1890A and IT1963A (while at last Hackathon the behavior was not consistent).*
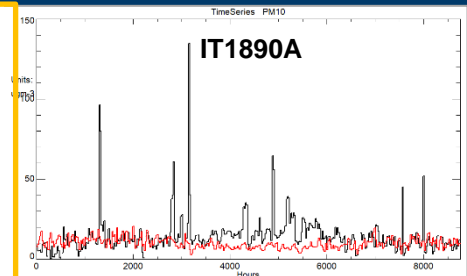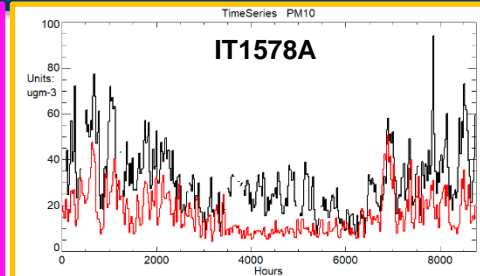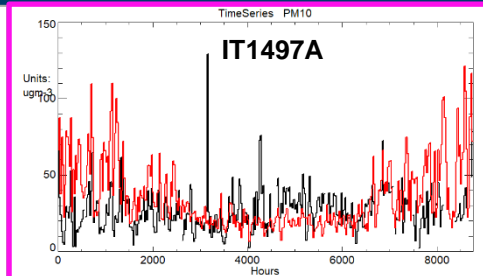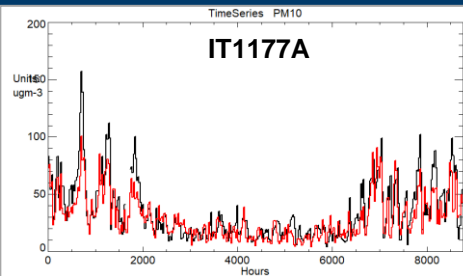
➢ When comparing the new Forecast Plots outcomes, it is worth noting that, even if they are based on different formulations, according to both of them, the criterion *MQI<1* is fulfilled at the same fraction of stations (seven out of ten), suggesting that a good choice of free parameters was probably made for Forecast MQI&MFE Plot.

*Two out of three stations where the criterion is not fulfilled are the same (IT1967A and IT2134A). The third is different, namely IT1497A according to Forecast Target Plot (OU) and IT1578A according to Forecast MQI&MFE Plot.*

# FORECAST PLOTS – NEW PROPOSED FORMULATIONS

## 1. IS THEIR BEHAVIOR CONSISTENT CONCERNING THE RANKING OF MODEL PERFORMANCES?
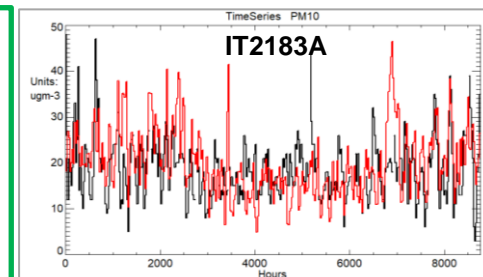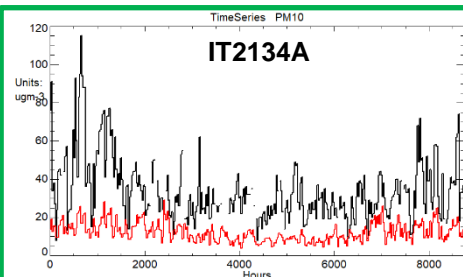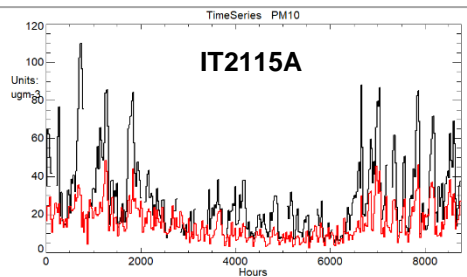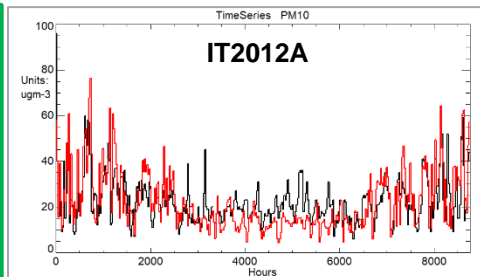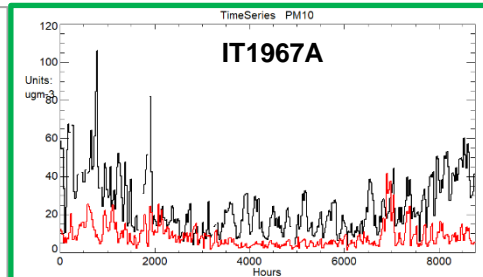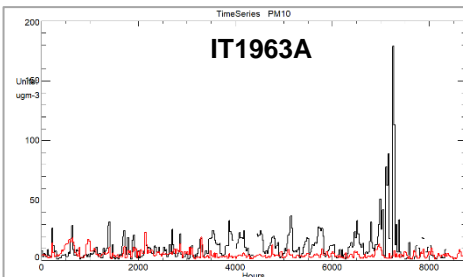
**When comparing the new Forecast Plots outcomes...**

Two out of three stations where the criterion is not fulfilled are the same: IT1967A and IT2134A

The third is different, namely IT1497A according to Forecast Target Plot (OU) and IT1578A according to Forecast MQI&MFE Plot.

**PM10 available background stations**

**O$_3$ available background stations**



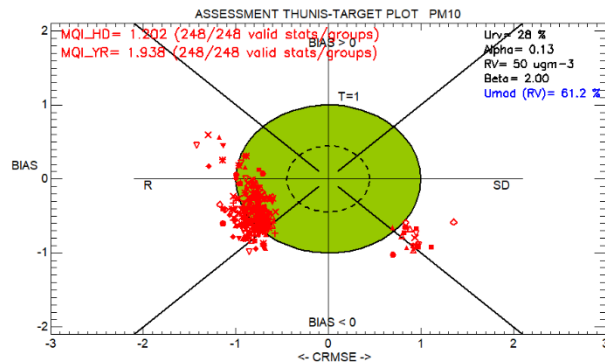*All the available background monitoring sites were used in the following tests*

ENEA

# FORECAST PLOTS – NEW PROPOSED FORMULATIONS
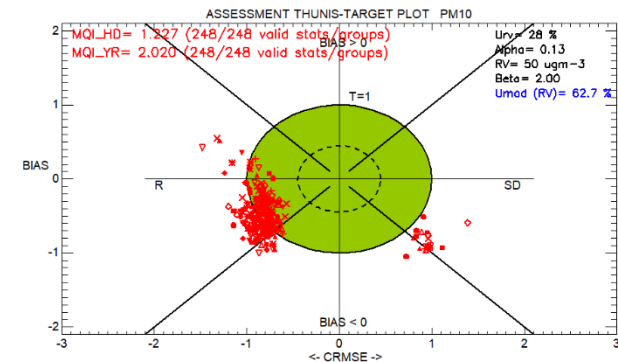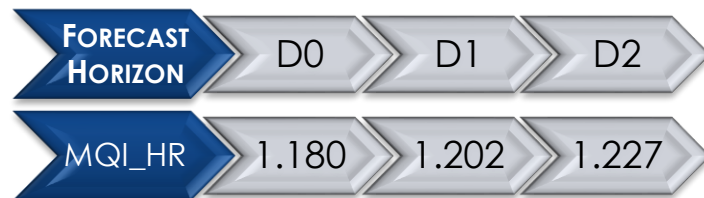## 2. HOW DO MODEL PERFORMANCES CHANGE ALONG WITH FORECAST HORIZON?
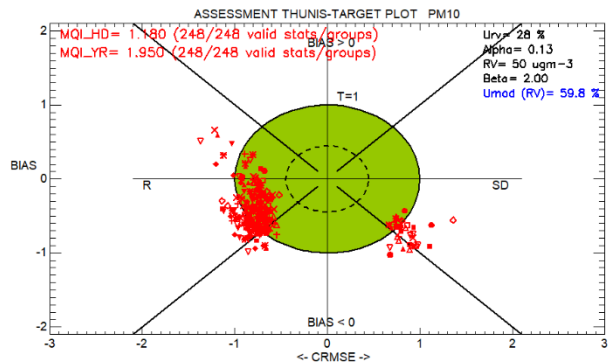
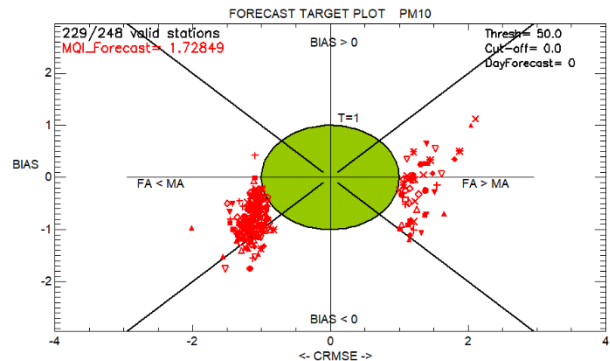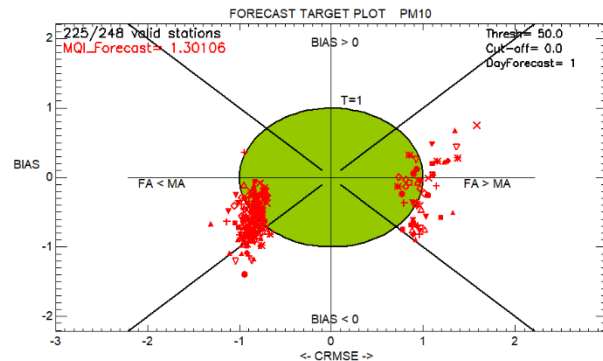**forecast day = 0**  **forecast day = 1**  **forecast day = 2**

ASSESSMENT THUNIS TARGET PLOT - PM10

| FORECAST HORIZON | D0 | D1 | D2 |
|---|---|---|---|
| MQI_HR | 1.180 | 1.202 | 1.227 |

Model performances <u>deteriorate</u> along with forecast horizon

forecast day = 0     forecast day = 1     forecast day = 2
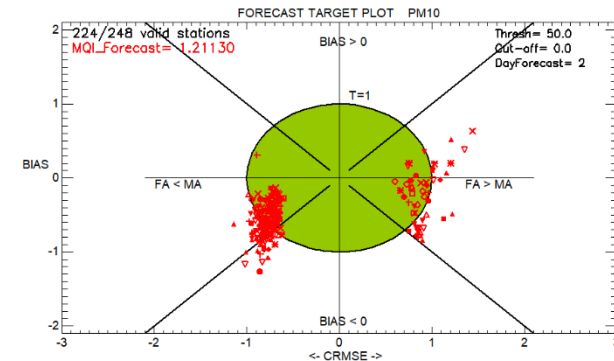
## FORECAST TARGET PLOT - PM10

| FORECAST HORIZON | D0 | D1 | D2 |
|---|---|---|---|
| MQI | 1.72849 | 1.30106 | 1.21130 |

Model performances get better along with forecast horizon
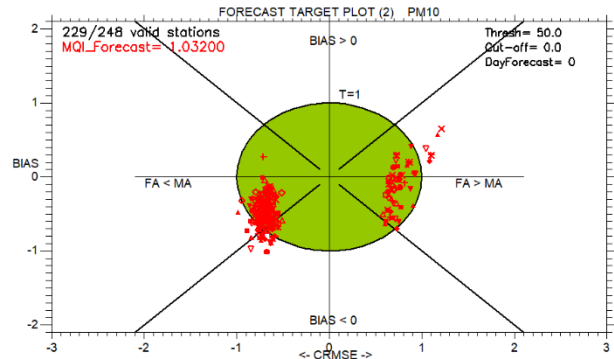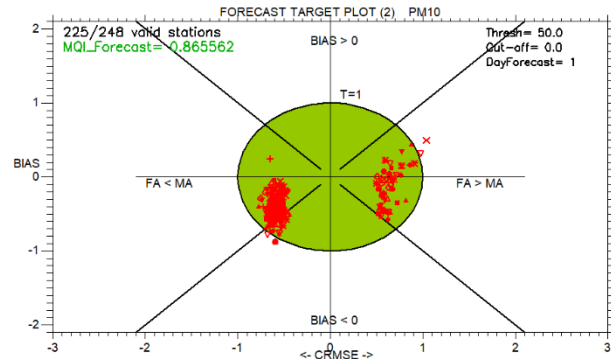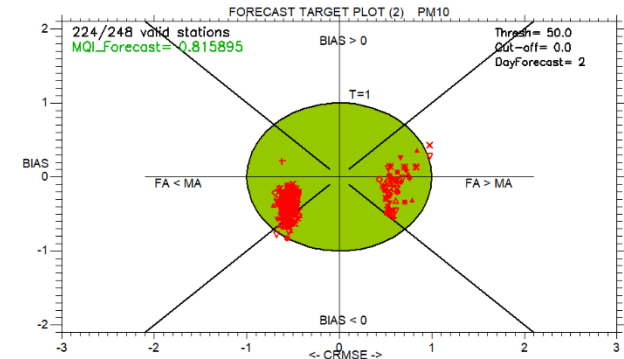
# FORECAST PLOTS – NEW PROPOSED FORMULATIONS
## 2. HOW DO MODEL PERFORMANCES CHANGE ALONG WITH FORECAST HORIZON?

**forecast day = 0**

**forecast day = 1**
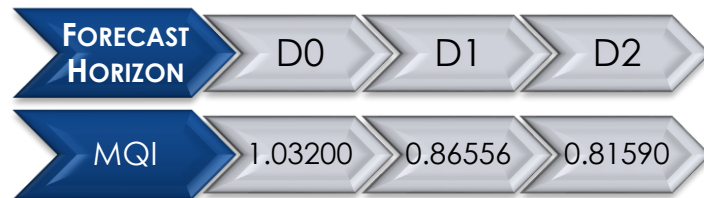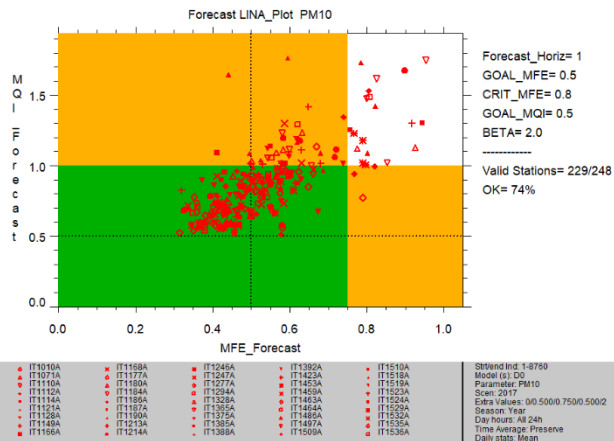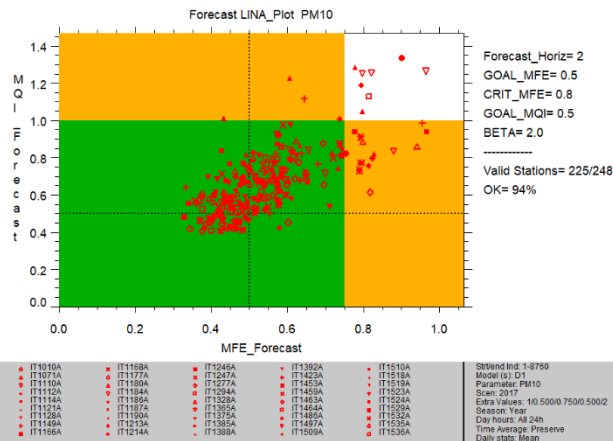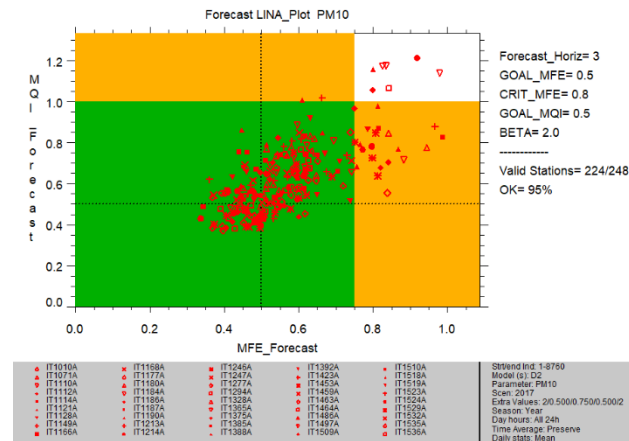
**forecast day = 2**

## FORECAST TARGET PLOT (OU) - PM10

| FORECAST HORIZON | D0 | D1 | D2 |
|---|---|---|---|
| MQI | 1.03200 | 0.86556 | 0.81590 |

Model performances <u>get better</u> along with forecast horizon

forecast day = 0 | forecast day = 1 | forecast day = 2

**FORECAST MQI&MFE PLOT - PM10**

| FORECAST HORIZON | D0 | D1 | D2 |
|---|---|---|---|
| OK | 74% | 94% | 95% |

Model performances _get better_ along with forecast horizon

**forecast day = 0**

**forecast day = 1**

**forecast day = 2**



## FORECAST MQI&MFE PLOT - PM10

| FORECAST HORIZON | D0 | D1 | D2 |
|---|---|---|---|
| OK | 74% | 94% | 95% |

More in detail, according to MQI outcomes (y axis) model performances get better along with forecast days.
Anyway, the fact that this improvement is due to persistence model performances degradation is clear from the plot, since according to MFE outcomes (x axis), model performances slightly deteriorate along with forecast days.

**ENEA**

Model performances get better along with forecast horizon

# FORECAST PLOTS – NEW PROPOSED FORMULATIONS

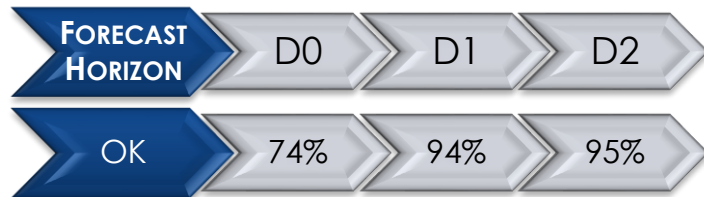## 2. HOW DO MODEL PERFORMANCES CHANGE ALONG WITH FORECAST HORIZON?

**forecast day = 0**
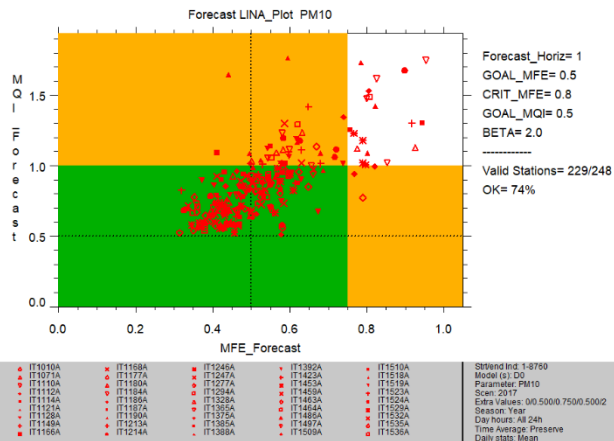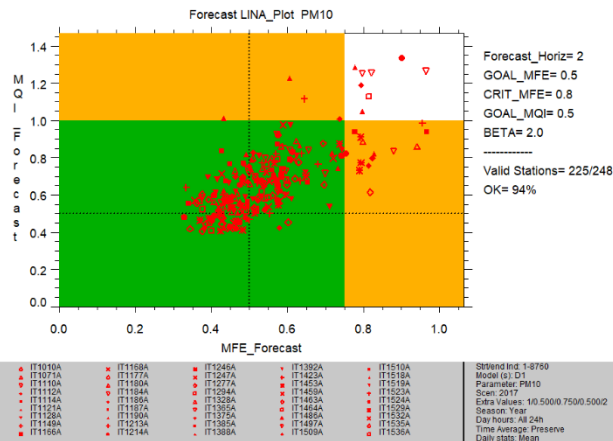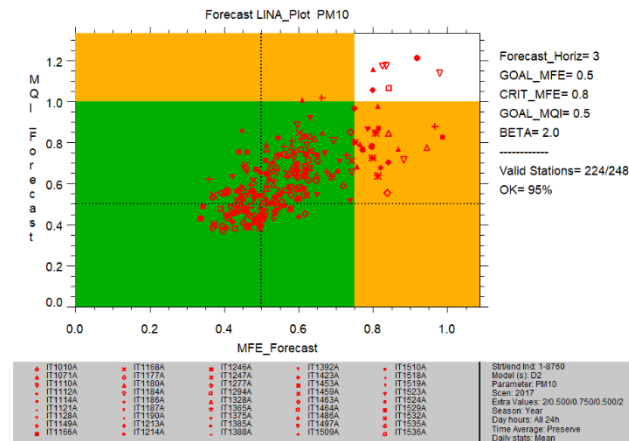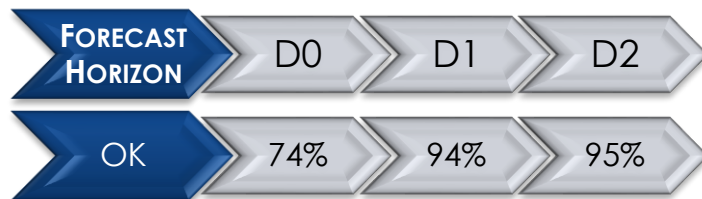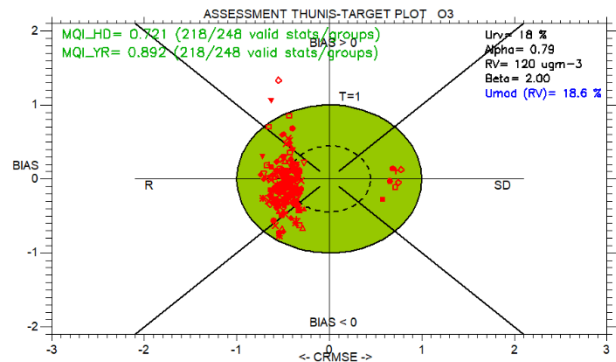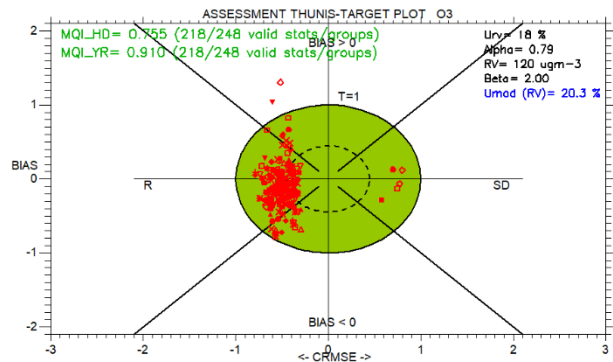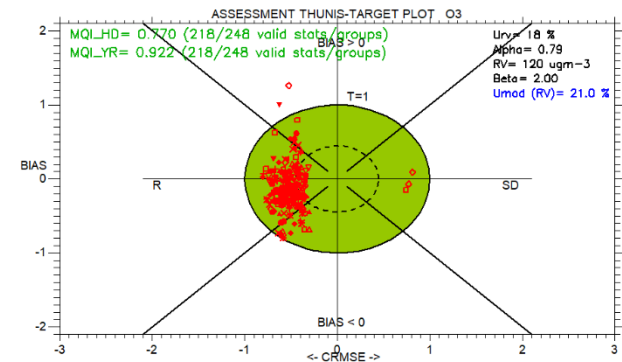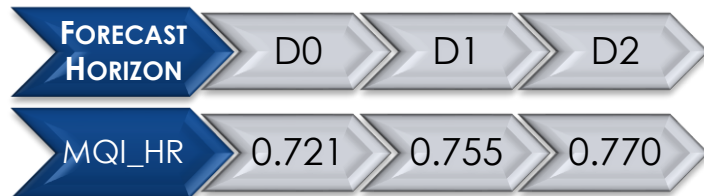
**forecast day = 1**

**forecast day = 2**

## ASSESSMENT THUNIS TARGET PLOT – $O_3$

| FORECAST HORIZON | D0 | D1 | D2 |
|---|---|---|---|
| MQI_HR | 0.721 | 0.755 | 0.770 |

Model performances **deteriorate** along with forecast horizon
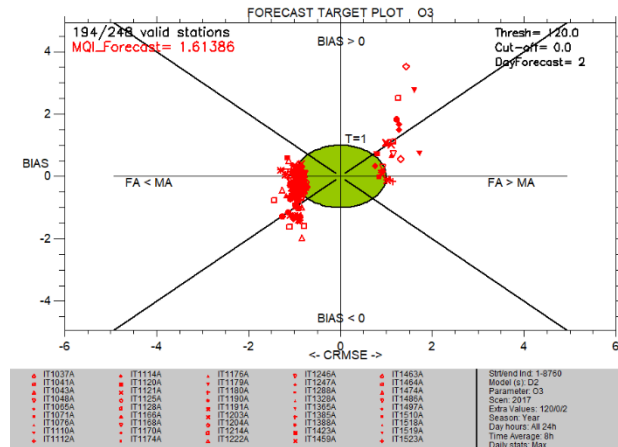
**forecast day = 0**      **forecast day = 1**      **forecast day = 2**

## FORECAST TARGET PLOT - O₃

| FORECAST HORIZON | D0 | D1 | D2 |
|---|---|---|---|
| MQI | 2.23010 | 1.75344 | 1.61386 |

Model performances _get better_ along with forecast horizon
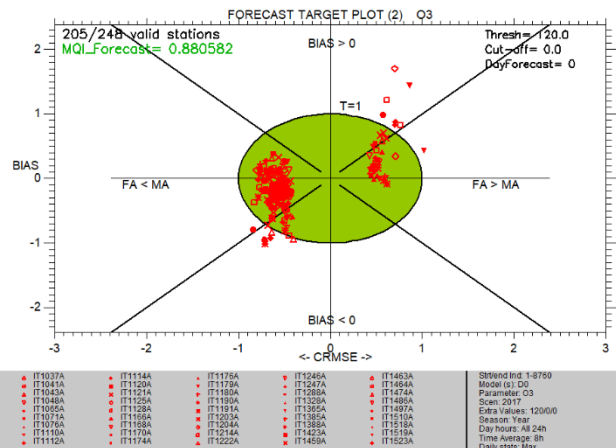
**forecast day = 0** **forecast day = 1** **forecast day = 2**
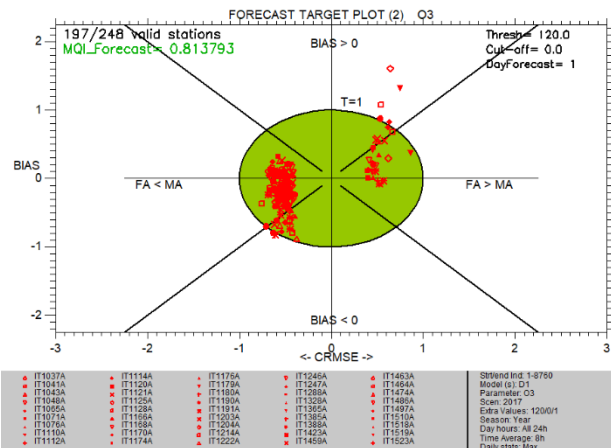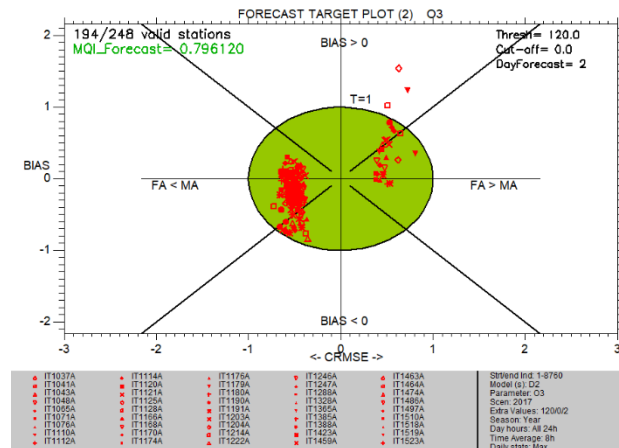


## FORECAST TARGET PLOT (OU) - O₃

| FORECAST HORIZON | D0 | D1 | D2 |
|---|---|---|---|
| MQI | 0.88058 | 0.81379 | 0.79612 |

Model performances <u>get better</u> along with forecast horizon

forecast day = 0

forecast day = 1
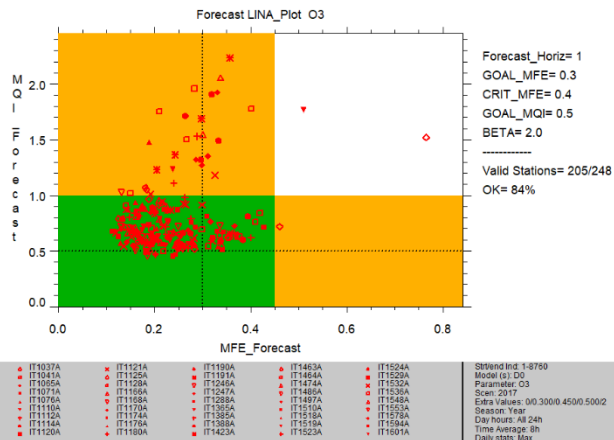
forecast day = 2

## FORECAST MQI&MFE PLOT - $O_3$
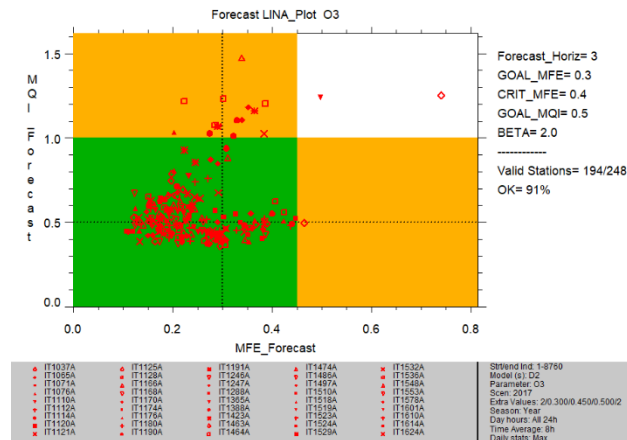
| FORECAST HORIZON | D0 | D1 | D2 |
|---|---|---|---|
| OK | 84% | 89% | 91% |
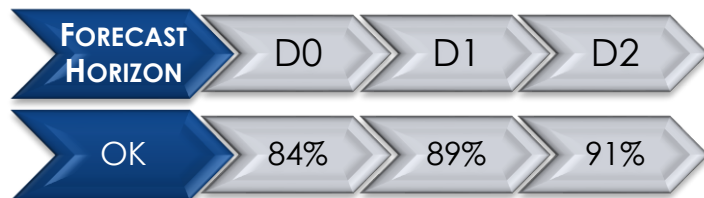
Model performances <u>get better</u> along with forecast horizon

**forecast day = 0**

**forecast day = 1**

**forecast day = 2**

## FORECAST MQI&MFE PLOT - O₃



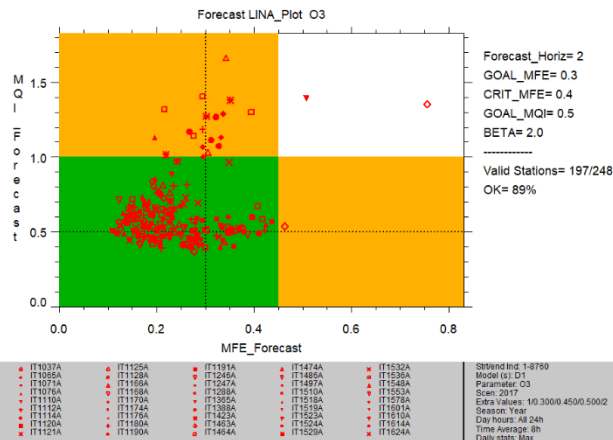| FORECAST HORIZON | D0 | D1 | D2 |
|---|---|---|---|
| OK | 84% | 89% | 91% |
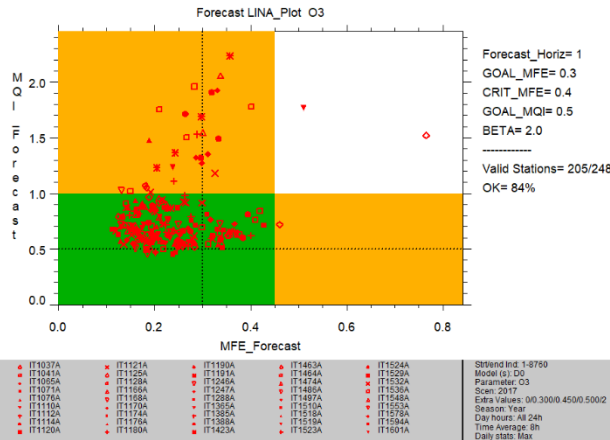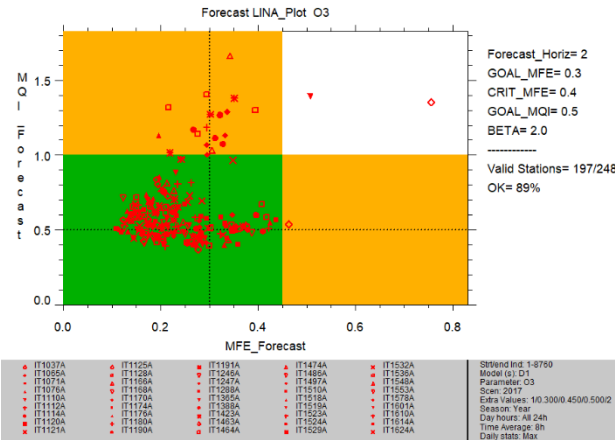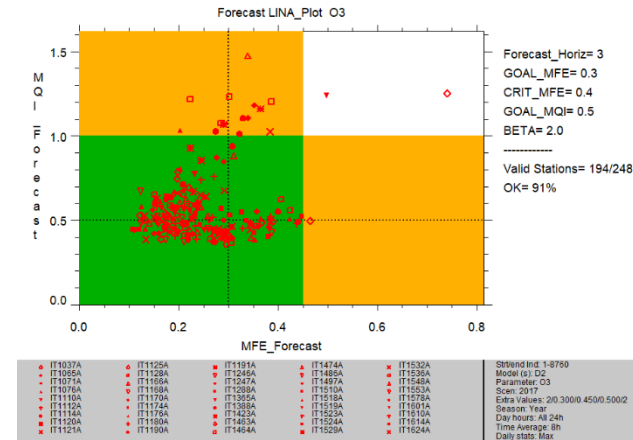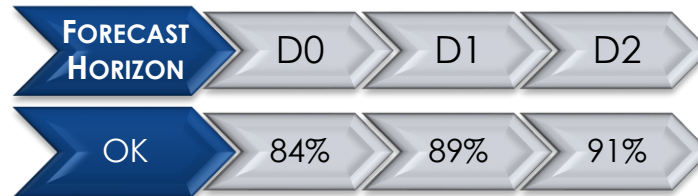
Again..
according to <u>MQI outcomes</u> (y axis) model performances <u>get better</u> along with forecast days
but
according to <u>MFE outcomes</u> (x axis) model performances <u>slightly deteriorate</u> along with forecast days.

Model performances <u>get better</u> along with forecast horizon

**In summary, the Forecast MQI&MFE Plot helps to clarify**.

This outcome can be easily explained to a policy maker:

*My model performances turn out to be good (according to MFE criteria).*

*They slightly deteriorate along with forecast days but persistence model does it worse. Indeed, performances ratios (i.e. MQI values) get better.*

*Another good reason, besides spatial coverage, for using the forecast model instead of persistence!*

$MQI_{forecast}$(OU) AND *FORECAST MQI&MFE* BEHAVIOR
CONCERNING THE MAIN SHORTCOMINGS OF THE ORIGINAL $MQI_{forecast}$ FORMULATION

1.  Is their behavior consistent concerning the <u>ranking of model performances</u>?

2.  According to their outcomes <u>how do model performances change along with forecast horizon</u>?

$MQI_{forecast}$(OU) AND $FORECAST\_MQI\&MFE$ BEHAVIOR
CONCERNING THE MAIN SHORTCOMINGS OF THE ORIGINAL $MQI_{forecast}$ FORMULATION

1. Is their behavior consistent concerning the <u>ranking of model performances</u>?

   ✓ When comparing Assessment and Forecast Plots outcomes, the ranking of model performances is more consistent according to both the new formulations, than it was in the original formulation

   ✓ When comparing the new Forecast Plots outcomes, they turn out to be quite consistent, even if they are based on different formulations

2. According to their outcomes <u>how do model performances change along with forecast horizon</u>?

**ENEA**

## $MQI_{forecast}$ (OU) AND $FORECAST\_MQI\&MFE$ BEHAVIOR CONCERNING THE MAIN SHORTCOMINGS OF THE ORIGINAL $MQI_{forecast}$ FORMULATION

1. Is their behavior consistent concerning the <u>ranking of model performances</u>?

   ✓ When comparing Assessment and Forecast Plots outcomes, the ranking of model performances is more consistent according to both the new formulations, than it was in the original formulation

   ✓ When comparing the new Forecast Plots outcomes, they turn out to be quite consistent, even if they are based on different formulations

2. According to their outcomes <u>how do model performances change along with forecast horizon</u>?

   ✗ according to both the new Forecast formulations, model performances seem to get better along with forecast horizon, as it was in the original formulation

   ✓ Forecast MQI&MFE Plot helps to clarify the reason

# CONCLUSIONS

The new formulation of the **Forecast Target Plot (OU)** sounds promising: most of the artefacts and shortcomings of the original formulation seem to be overcome

The **Forecast MQI&MFE Plot** can be used to support the interpretation of results, since it helps to clarify

ENEA

Thank you

lina.vitali@enea.it