Reviewer #1:

General Comments:

1. *Absence of NODA in comparisons/figures*

Although announced in section 6.3, the NODA is not part of any figure as far as I can see. Therefore, the reader can't be sure that the assimilation improves forecasts. It could be that both methods are not much different from NODA.

We thank the reviewer for this important observation. We agree that including the NODA results in the comparisons is essential to properly assess the impact of the assimilation process.

In response to this comment, we have now incorporated NODA results into the relevant figures and discussions. The updated figures are included in the Section "New Figures" below in the present document, ensuring that the impact of data assimilation is clearly demonstrated by comparing EnKF, 3DVar and NODA runs. Additionally, we have revised the corresponding discussion to explicitly analyze how each DA method performs relative to the NODA run.

We appreciate the reviewer's input, as this addition strengthens the conclusions regarding the effectiveness of data assimilation in improving forecasts.

2) Inconsistency of verification measures (shown in results) and indicated aims (section 2)

The paper intends to show ... (indicated aims, L161ff)

1. *the improved prediction of small-scale extreme weather events*
2. *enhanced accuracy of atmospheric conditions in the pre-convective environment*
3. *impact of assimilating in-situ conventional and remote sensing observations*

Ad 1: There is no comparison to NODA, thus no improvement measurable. Moreover, the paper evaluates precipitation FSS(>1mm/h), RMSE(1h/6h), but does show to which extent the observed extreme precipitation could be forecasted.

Ad 2: Where did you evaluate the pre-convective environment?

We thank the reviewer for this valuable comment. We acknowledge the initial concerns regarding the inconsistency between the verification measures and the stated aims of the study. However, these issues have already been addressed in response to the other reviewer's comments, and we have made significant revisions to improve the clarity of the study's objectives. Consequently, the aims outlined in the original manuscript have been updated in the revised version. Regarding Ad1, we have now incorporated NODA results into the relevant figures and discussions, ensuring that the impact of data assimilation is clearly demonstrated. These additions allow for a proper evaluation of the extent to which both 3DVar and EnKF improve forecasts relative to a no-data-assimilation scenario. We have also revised the precipitation verification by incorporating higher thresholds for FSS and ensuring that FSS and RMSE calculations are based on the same time window. This provides a clearer assessment of the model's ability to capture intense precipitation events, making the results more relevant to the study's stated objectives. In terms of Ad 2, we agree that our aim (2) in L165 was not clearly stated and may have been misleading. Most DA studies focus on assimilating observations once the weather phenomena has already initiated, rather than incorporating observations several hours before the event begins (i.e., in the pre-convective phase). To the best of our knowledge, very few studies focus explicitly on DA strategies in pre-convective conditions, particularly in the Mediterranean region. This distinction underscores the novelty of the present study and its contribution to the existing body of research. This study aims not to analyze how the pre-convective environment is modified after assimilation, but rather to evaluate the forecast impact of assimilating observations during the pre-convective phase, as opposed to assimilating observations after convection has initiated.Our objective is to assess whether assimilating pre-convective observations ultimately leads to an improved forecast of extreme weather events. The improvements observed in the forecasted convective evolution and precipitation fields suggest that assimilating data in the pre-convective phase contributes to a better representation of the event.

To ensure clarity, we have revised the objectives of the study, which now reads as follows:

"*On overall, this study aims at:*

(a) Assessing the impact of 3DVar in comparison with the EnKF system to predict small-scale extreme weather events initiated over maritime regions with lack of in-situ observations.

(b) Investigate the potential of using 3DVar and EnKF in the pre-convective environment, hours before the mature stage of convective systems are reached, to improve forecast lead time and warning capabilities for extreme weather events.

(c) Compare the forecast impact from assimilating in-situ conventional observations in comparison to assimilating high spatial and temporal resolution data from remote sensing instruments.

(d) Provide a quantitative assessment between two different DA schemes using several statistical verification methods."

We appreciate the reviewer's feedback, as it has helped us refine the presentation of our study's objectives.

3) Inconsistency of conclusions and results

L974-975: "Similar skill" of EnKF and 3DVar in FSS (Fig8) and Taylor diagram (Fig10)

That contradicts what I see in Fig 8 and 10.

We thank the reviewer for this observation. We agree that this statement does not accurately reflect the results shown in these figures.

In response to this comment, we have modified this sentence to ensure consistency between the results and the discussion.

This sentence now reads as follows:
"*For IOP13, both the Filtering method and Taylor diagram verification show that EnKF slightly outperforms 3DVar, though the differences are not substantial.*

We appreciate the reviewer's feedback, as it has helped improve the clarity and accuracy of our conclusions.

L976: "significantly improved the forecast". There was no comparison to NODA. Thus no improvement visible.

We thank the reviewer for this comment. We agree that the statement in L976 regarding "significantly improved the forecast" was initially made without a direct comparison to NODA, making it difficult to quantify the improvement.

This issue has already been addressed in response to previous comments, as we have now included the NODA experiment in our analysis. The updated figures and discussions clearly demonstrate the impact of data assimilation relative to a no-data-assimilation experiment, providing a more robust assessment of forecast improvements.

We appreciate the reviewer's input, as it has helped ensure that our conclusions are better supported by the presented results.

L979: "EnKF provides worst results." This is not a disadvantage of EnKF. The ensemble mean of a cyclone pressure field is as useful as the ensemble mean of a precipitation field. As it is not Gaussian, it should not be expected to perform well.

We thank the reviewer for this insightful comment. We agree that this is not a disadvantage of EnKF, but rather an inherent characteristic of ensemble-based methods. While the ensemble mean provides valuable information, it is not always the most appropriate metric for variables with non-Gaussian distributions, such as cyclone pressure fields or precipitation fields. In such cases, ensemble forecasts are better interpreted in a probabilistic framework, rather than relying solely on the ensemble mean.

In response to this comment, we have revised the discussion to more accurately reflect these considerations and removed the misleading statement. Now this sentence reads as follows:

"*For the Qendresa event, while the **ensemble mean of EnKF underestimates the intensity of the medicane compared to 3DVar**, some individual **EnKF ensemble members produce more accurate results than 3DVar**.*"

We appreciate the reviewer's feedback, as it has helped us improve the interpretation of our results.

<u>Minor comments:</u>

Figure 8: The authors employ FSS of precipitation >1 mm/h for three regions (Fig 8a-c). This score evaluates correct positioning of precipitation in forecasts, but doesn't show improved prediction of extreme events (1 mm/h is hardly extreme).

We thank the reviewer for this comment. This concern was also raised by Reviewer 1, and we have already addressed it by incorporating higher precipitation thresholds in the FSS analysis to better assess the prediction of extreme precipitation events. We have updated Fig. 8 to include the new FSS results using progressively higher thresholds, ensuring a more meaningful evaluation of extreme precipitation forecasts. The updated figures are included in the Section below "**New Figures**", and the corresponding discussion has been modified accordingly.

We appreciate the reviewer's input, as this improvement enhances the robustness of our precipitation verification.

Figure 8: It is unclear what RMSE shows. Is it RMSE of ensemble mean prediction of precipitation? [mm/h]?

The RMSE shown in Figure 8 corresponds to the root mean square error of the predicted precipitation field, evaluated against observations not assimilated. Specifically, for the 3DVar, RMSE is computed from the deterministic forecast, and for EnKF, it is computed from the ensemble mean precipitation field. The RMSE values are expressed in mm/h.

To improve clarity, we have updated the figure caption and revised the text in the manuscript to explicitly state this information. We appreciate the reviewer's suggestion, as it helps ensure that the methodology is clearly communicated.

Figure 8: Is the FSS of EnKF computed from the ensemble mean forecast or from the whole ensemble?

We thank the reviewer for this comment. The FSS for EnKF is Figure 8 has been computed from the ensemble mean forecast, rather than from individual ensemble members.

To ensure clarity, we have updated the figure caption and revised the corresponding section in the manuscript to explicitly state this.

Figure 10: Which observations have been used for this figure?

The observations used in Figure 10 refers to the precipitation observations, as those used in Figure 8.

To improve clarity, we have updated the figure caption and revised the manuscript text to explicitly state this.

The paper is 31 pages. I suggest to shorten the text in the interest of the reader and journal guidelines. For example, the introduction is very long and does not always on point (why are particle filters discussed?). The methods contain a revision of DA equations. I don't see how that serves the rest of the paper.

We thank the reviewer for this suggestion. We agree that the manuscript can be shortened to improve readability and better align with journal guidelines. In response to this comment, **we have reduced the section introducing EnKF and 3DVar, keeping only the essential details relevant to this study**. While we have removed the general revision of DA equations, we have retained key information on the specific parameters used to configure these DA schemes, as this was explicitly requested by reviewers in previous studies.

L162-172: These are general points and considering the limited set of observations for verification, such general questions cannot reasonably be answered, as you state in L174-177. Maybe you can edit L161 to include something like "we address these questions for two high-impact cases"

We thank the reviewer for this comment. The section L161-172 has been revised following suggestions from the other reviewer, and the concerns raised here are no longer applicable. Specifically, we have clarified the study's objectives, ensuring that they align with the limited set of observations and case study approach.

We appreciate the reviewer's feedback, as it has contributed to improving the focus and clarity of our manuscript.

L162, and others: "high-resolution data assimilation": I don't see what the difference between "high-resolution 3DVar" and 3DVar is. Clearly, if applied at high resolution, any method becomes a high-resolution method. Is there any more to it?

We thank the reviewer for this comment and appreciate the opportunity to clarify our use of the term "high-resolution data assimilation".

In this study, "high-resolution" refers not only to the spatial resolution of the model but also to the high-temporal frequency of data assimilation cycles. This approach significantly increases computational demands, as it requires assimilating observations at short intervals while maintaining fine-grid resolutions.

While 3DVar and EnKF can be applied at various resolutions, the combination of high spatial and temporal resolution DA is computationally expensive and less commonly explored in the literature. Our study aims to evaluate these methods under this computationally demanding configuration, which differs from many conventional DA studies that use lower temporal update frequencies or coarser spatial resolutions.

To clarify this point, we have revised the manuscript to better define what we mean by high-resolution data assimilation in this context. We appreciate the reviewer's input, as it helps improve the precision of our terminology.

L171-172: Isn't this the same as point (a) before?

We thank the reviewer for this comment. This section has been revised following suggestions from the other reviewer, and the concerns raised here are no longer applicable because this section is now different. The structure and wording have been adjusted to improve clarity and avoid redundancy.

Missing table: It would be useful to collect the assimilated observation types per case in a table.

We thank the reviewer for this helpful suggestion. We agree that including a table summarizing the assimilated observation types for each case will improve the clarity of the manuscript.

In response to this comment, we have added a new table that lists the types of observations assimilated in each case study. This table ensures a clearer presentation of the observational data used in the assimilation process.

Event	Observation Type	Data Sources	Assimilation Frequency	Coverage	Additional Processing
IOP13	Conventional *in-situ* data	MADIS (NOAA)	Hourly	Entire Domain	Quality-controlled
IOP13	Radar Reflectivity	Météo-France Doppler Weather Radars (Aleria & Nimes)	Every 15 minutes	Ligurian Sea & Gulf of Genoa	Quality controlled and Interpolated using Cressman Objective Analysis (6 km grid)
Qendresa	Conventional *in-situ* data	MADIS (NOAA)	Hourly	Mediterranean Region	Quality-controlled
Qendresa	Satellite-Derived Winds (RSAMVs)	EUMETSAT (SEVIRI instrument onboard MSG)	Every 20 minutes	Entire atmosphere over the Mediterranean Region	Quality-controlled, superobbing (128x128 km, 25 hPa vertical)

L940: high-resolution DA techniques

Well, plain-vanilla 3D-Var is not a high-resolution DA technique, especially without hydrometeor control variables.

We thank the reviewer for this comment and appreciate the opportunity to clarify our terminology. We acknowledge that standard 3DVar, particularly without hydrometeor control variables, is not inherently a high-resolution DA technique.

In this study, we refer to "high-resolution DA" in the context of both high spatial and high temporal resolution, rather than implying modifications to the 3DVar formulation itself. Our experimental setup involves frequent assimilation cycles and fine-grid numerical simulations, which significantly increase the computational demands and are less commonly explored in previous studies.

To prevent any misunderstanding, we have revised the wording in the manuscript to ensure that our use of "high-resolution DA techniques" is accurately conveyed. We appreciate the reviewer's feedback, as it has helped improve the precision of our terminology.

Figure 6: The figure is split over two pages. It would be good if it would not be separated from the caption. Labels a-f could be replaced by SYN, CNTRL, NODA, if possible.

We thank the reviewer for this suggestion. We agree that keeping Figure 6 and its caption on the same page will improve readability and presentation. We have adjusted the formatting to ensure that the figure is not split across multiple pages, and we have also updated the labels (a-f) to more descriptive names such as SYN, CNTRL and NODA, making it easier for the reader to interpret the figure.
We appreciate the reviewer's input, as these improvements enhance the clarity and accessibility of the figure.

L566: The year should be 2012 not 2021.

We thank the reviewer for catching up this typo. We have corrected the year from 2021 to 2012 in L566 to ensure accuracy. We appreciate the reviewer's attention to detail.

Finally, I would suggest to mention the opportunities from satellite data assimilation for convective-scale forecasting. Future studies could benefit greatly from that.

We thank the reviewer for this valuable suggestion. We agree that satellite data assimilation presents significant opportunities for improving convective-scale forecasting, particularly in data-sparse maritime and remote regions.

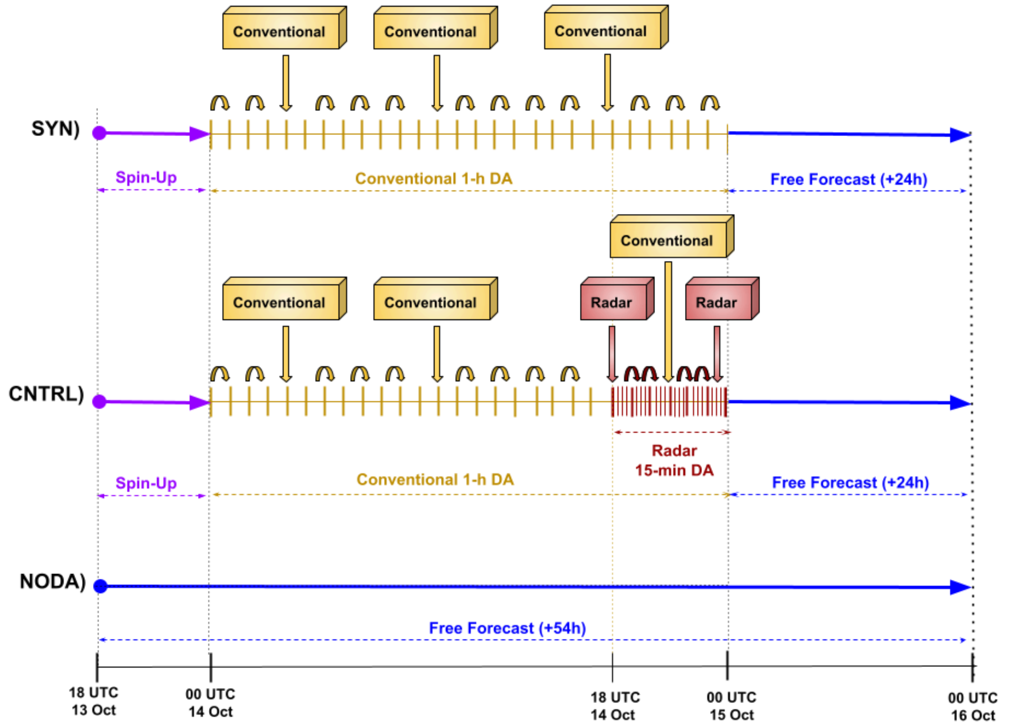
In response to this comment, we have included the following in the conclusions section, highlighting the potential benefits of future studies incorporating satellite-based observations:

"*In addition, it is important to highlight that satellite-based data assimilation presents a significant opportunity for advancing convective-scale forecasting, particularly in data-sparse maritime regions such as the Mediterranean, where the formation of extreme weather events like tropical-like cyclones is increasingly impacting densely populated areas. Future studies integrating high-resolution satellite observations, such as cloud top highs, thermodynamic profiles or cloud properties, could further enhance the accuracy of convective-scale predictions, improving early warning capabilities and disaster preparedness.*"

We appreciate the reviewer's input, as this addition strengthens the discussion on potential future advancements in data assimilation techniques.

New Figures:

Fig. 6: I have modified the figure according to reviewer's comments.



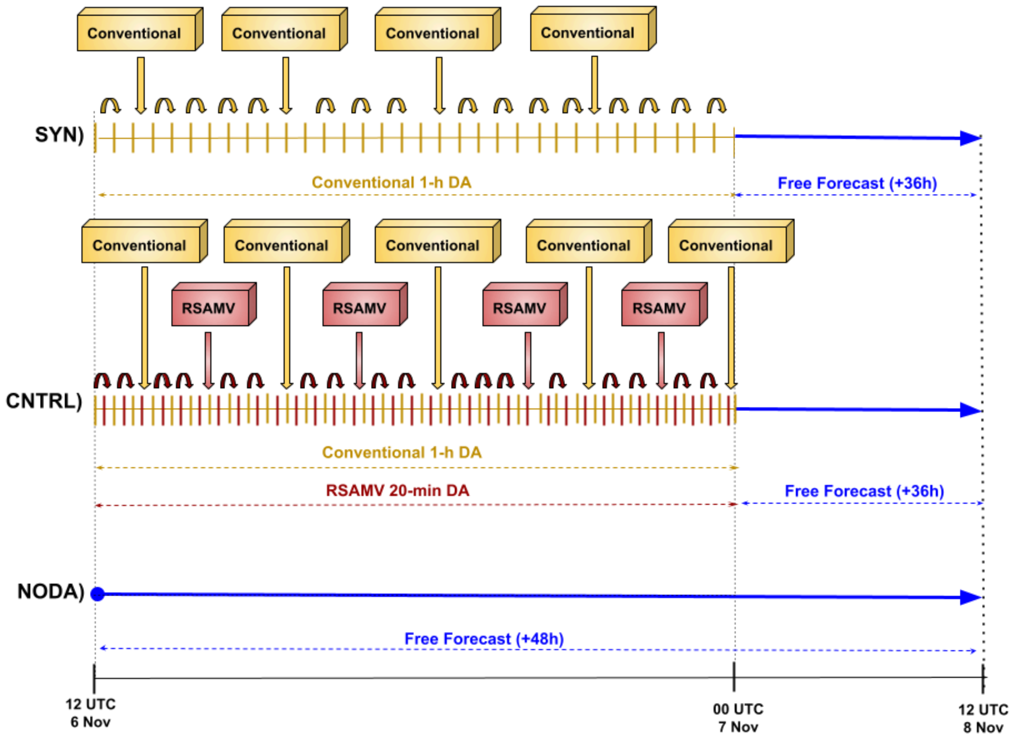
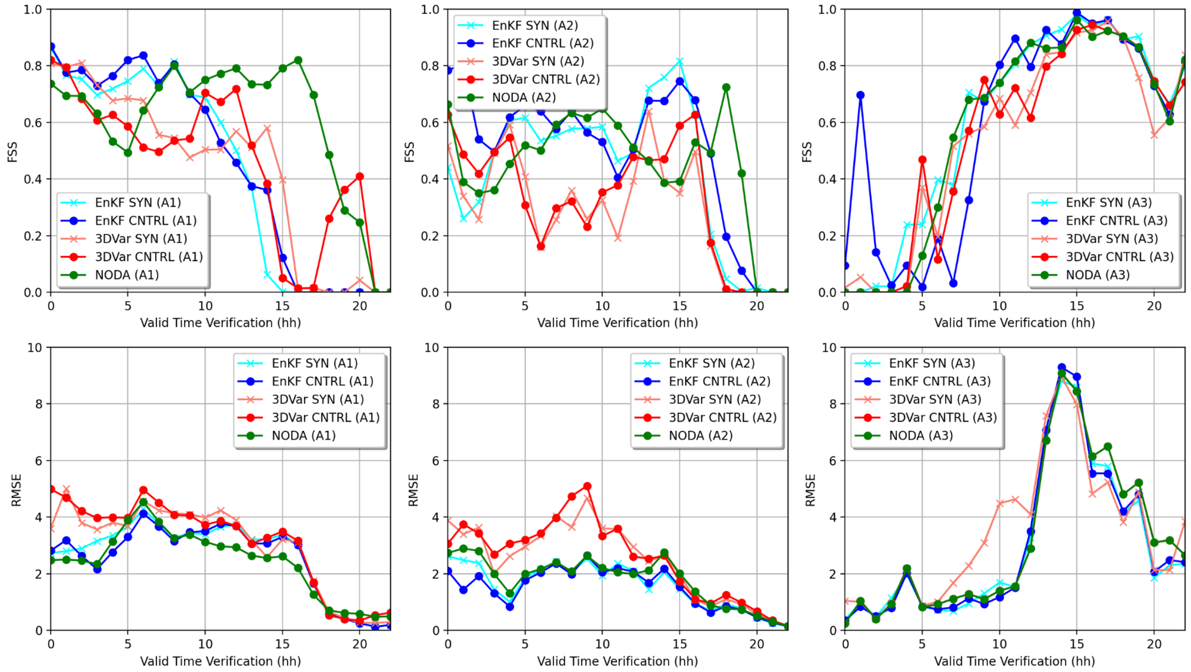
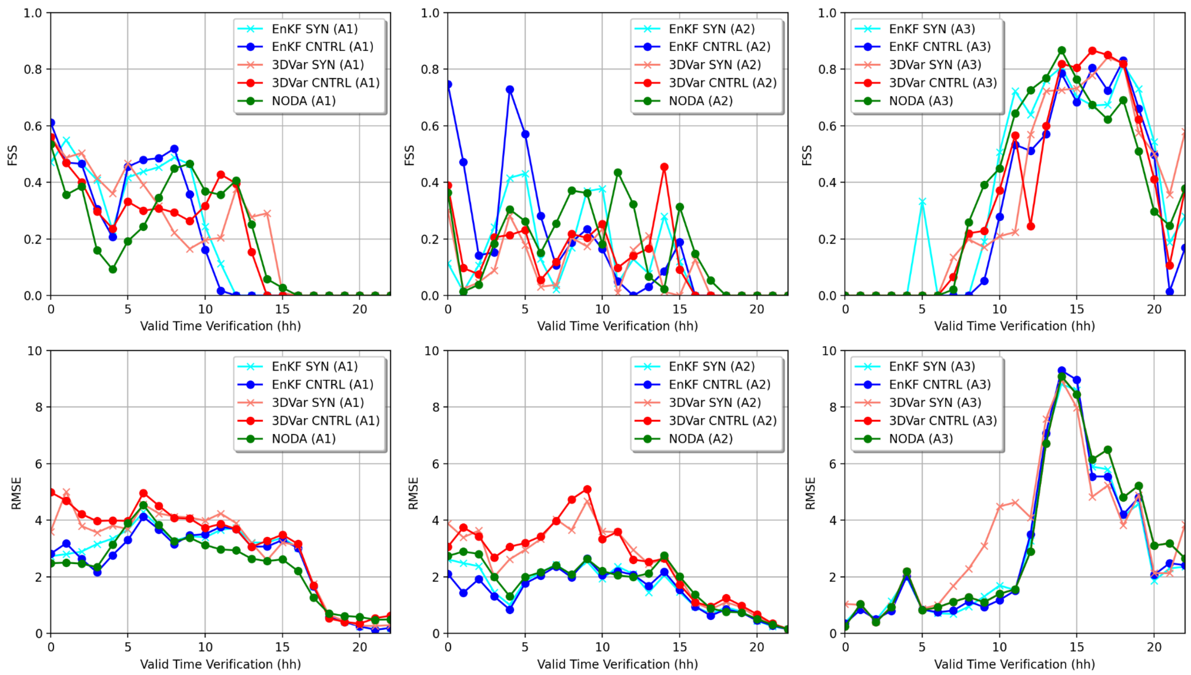


Fig. 8: I have added results from NODA simulation.

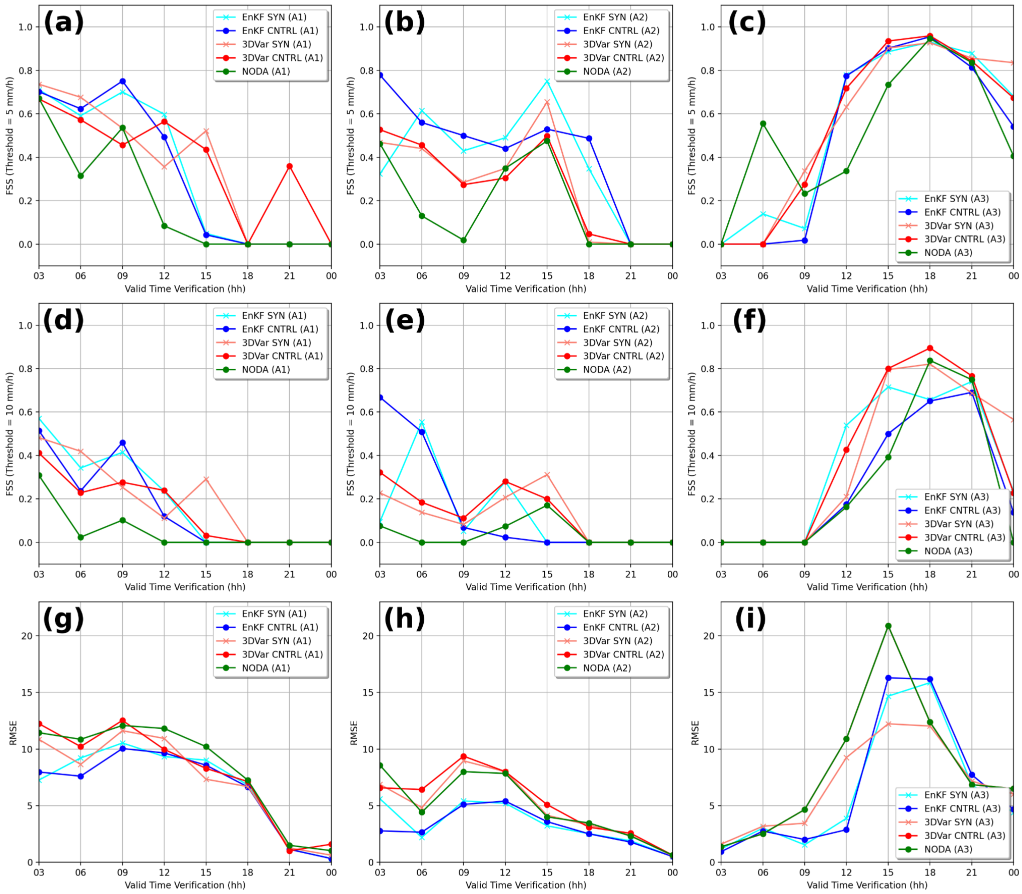
1h Accumulated precipitation (threshold: 1mm/h):



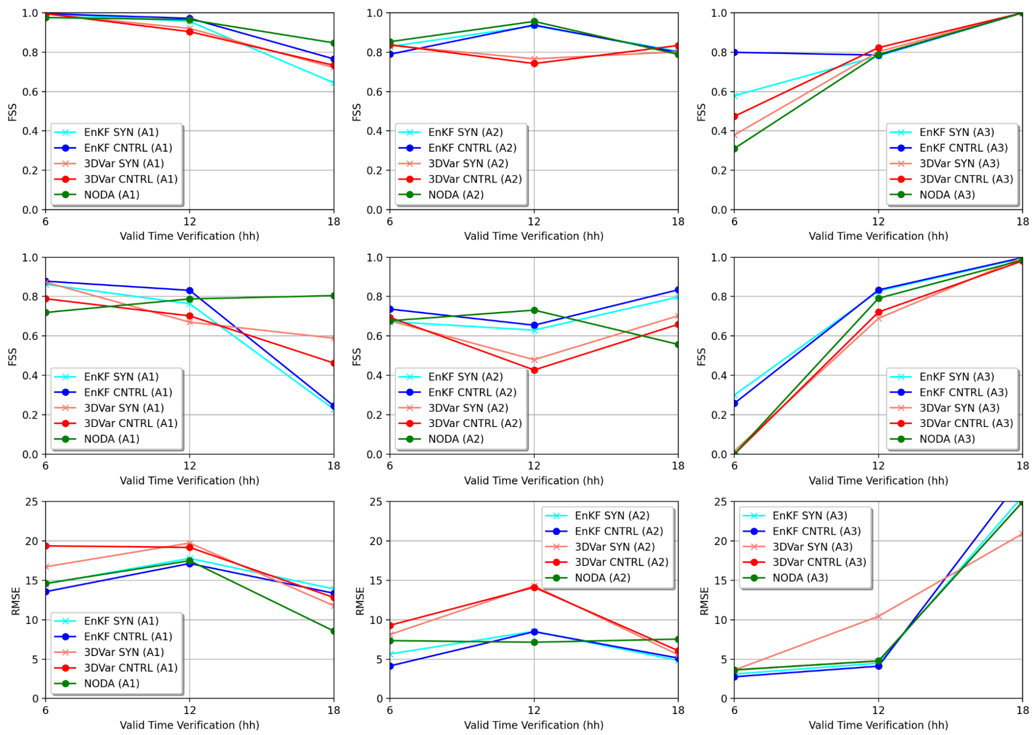
1h Accumulated precipitation (threshold: 3mm/h):



3h Accumulated precipitation (threshold: 5 mm/h and 10 mm/h):



6h Accumulated precipitation (threshold: 3mm/h):



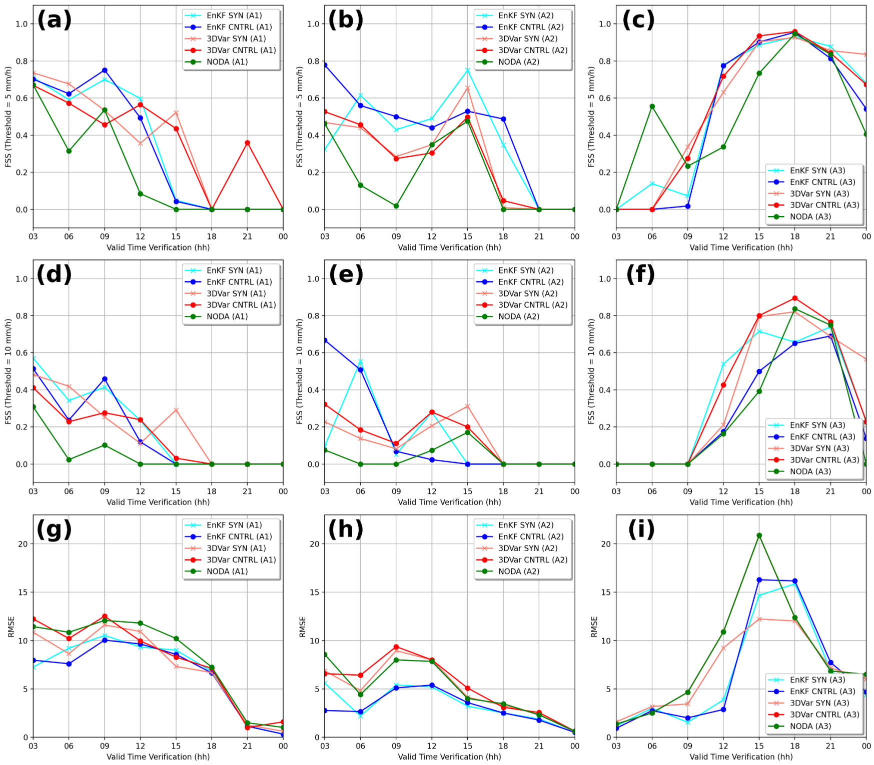


Figure 8. *Upper panels: Evolution of the FSS during the first 24-h hours of free forecasts for 3-h accumulated precipitation in the Italian sub-regions: R1 (first column), R2 (second column) and R3 (third column). Two thresholds are used: > 5 mm·h*$^{-1}$ *(first row) and > 10 mm·h-1 (second row). Lower panels: Evolution of the RMSE associated with each experiment during the first 24 hours of free forecast in the different sub-regions. Simulations assimilating both conventional and radar observations (CNTRL) and simulations assimilating only conventional observations (SYN) associated with the 3DVar and the EnKF are shown here. As a reference, NODA results are also included.*

Fig. 9: I have added results from NODA (Deterministic) simulation.

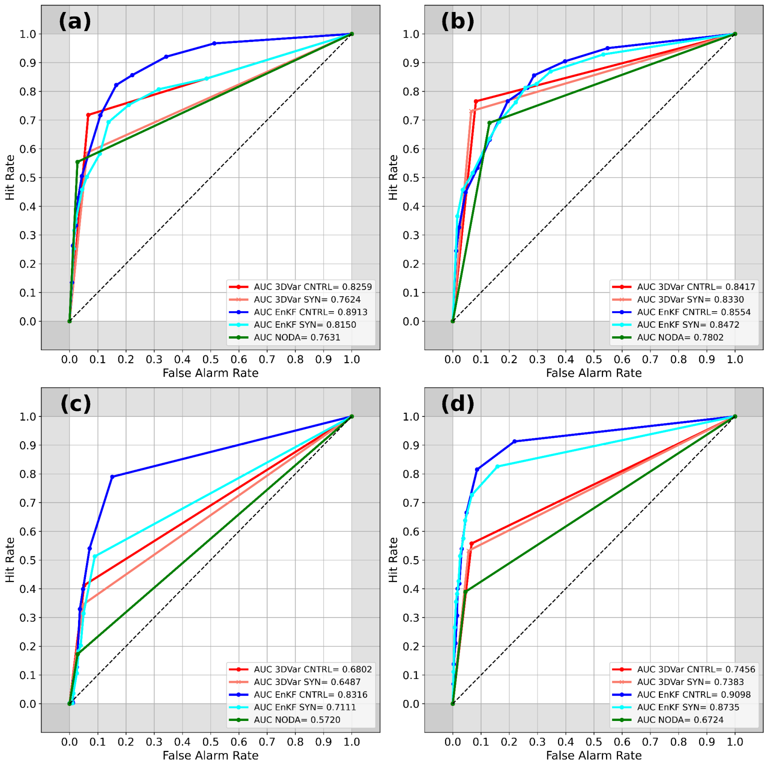


Figure 9. *ROC curves and AUC associated with the 3DVar (red and pink colors), EnKF (blue and cyan colors) and NODA (green color) for the 3-hour accumulated precipitation using (a) 1 mm and (b) 10 mm threshold and 6-hour accumulated precipitation using (c) 1 mm and (d) 10 mm threshold, computed over the entire inner domain.*

Fig. 10: I have added results from NODA simulation.

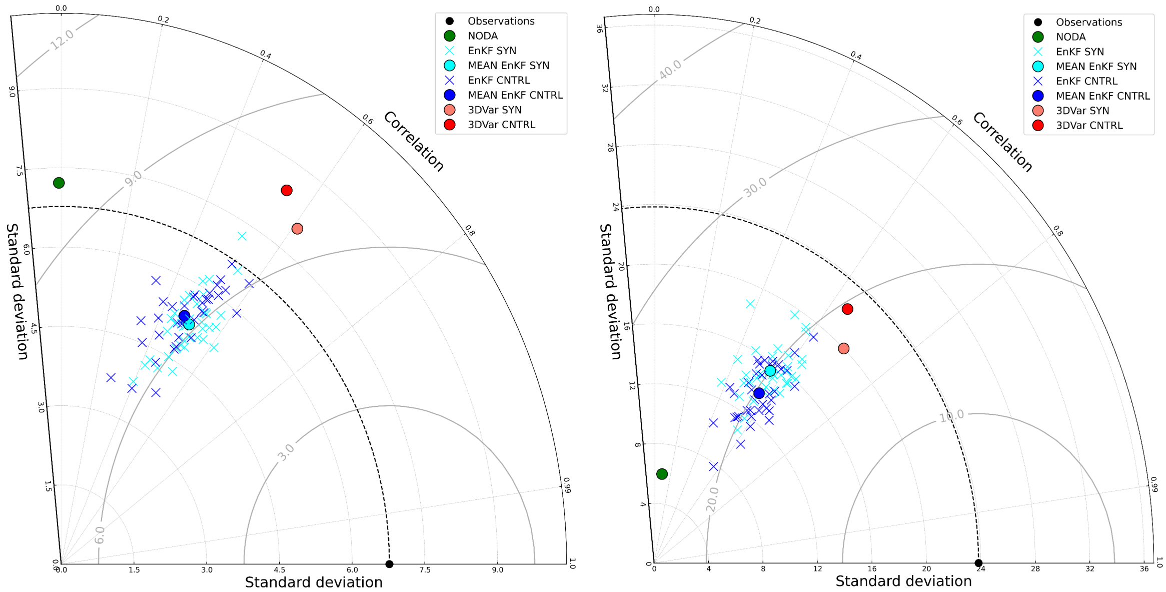


Figure 10. *Taylor diagram comparing the performance of 3DVar (red), EnKF (blue) and NODA (green) for the 6-hour (left panel) and 24-hour (right panel) accumulated precipitation, valid at 06 UTC 15 October 2012.*

Fig. 11: I have added results from NODA simulation.

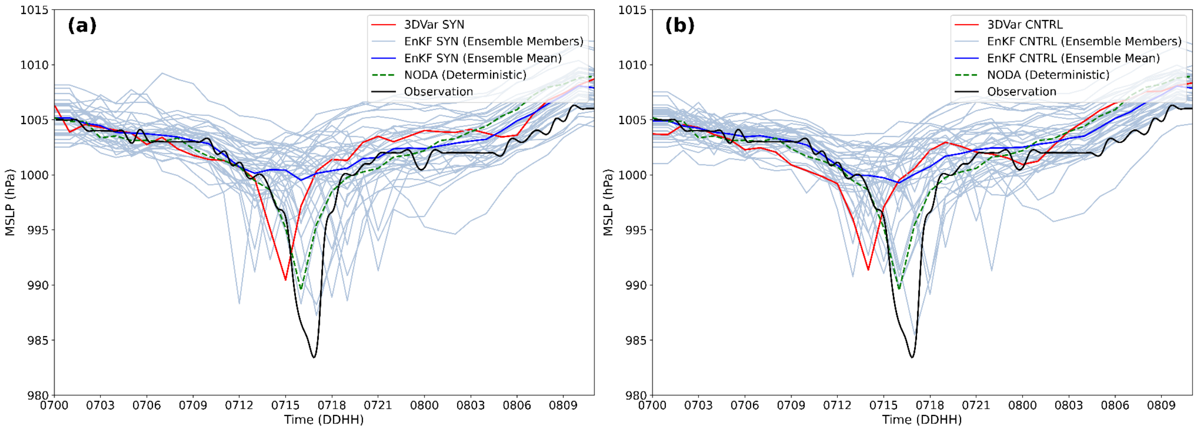


Figure 11. *Temporal surface pressure evolution at the closes grid point to Malta for the (a) SYN and (b) CNTRL experiments associated with the EnKF (blue), 3DVar (red) and NODA (green), compared to the observed surface pressure registered by METARs in Malta's airport (black line).*

Fig. 12: I have added results from NODA simulation.

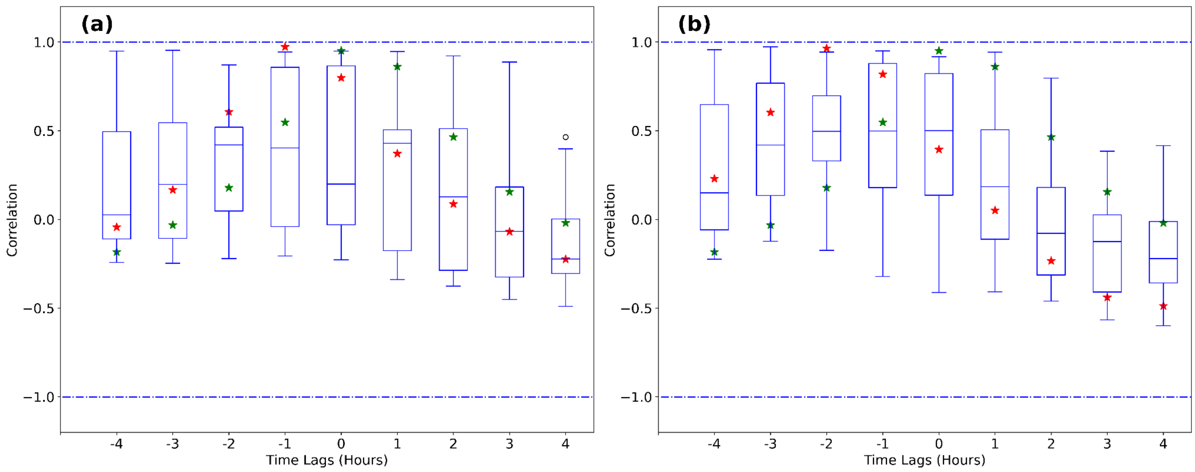


Figure 12. *Whisker plots depict the lagged correlation values between the observations and the EnKF (blue boxes), the 3DVar (red stars) and NODA (green stars) for the (a) SYN and (b) CNTRL experiments. The correlation is computed considering that the observed V-shape pressure signature associated with the observations is shifted 4 hours to the left and 4 hours to the right.*

Fig. 13: I have added results from NODA simulation.

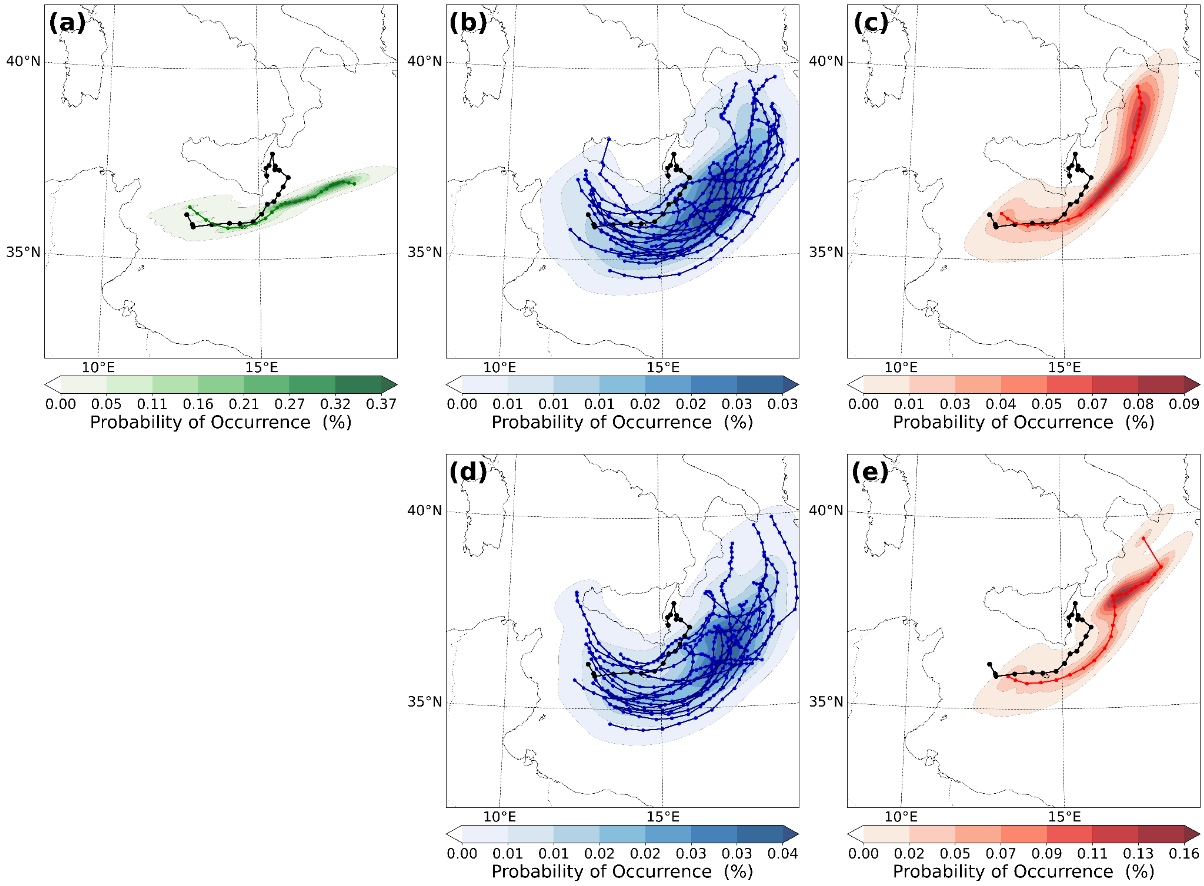


Figure 13. *Probability of cyclone center occurrence computed using Gaussian KDE for (a) NODA, (b) EnKF (SYN), (c) 3DVar (SYN), (d) EnKF (CNTRL) and (e) 3DVar (CNTRL), from 11 UTC 7 November to 12 UTC 8 November 2014. Qendresa's trajectory observed via satellite imagery is depicted in black.*

Fig. A1

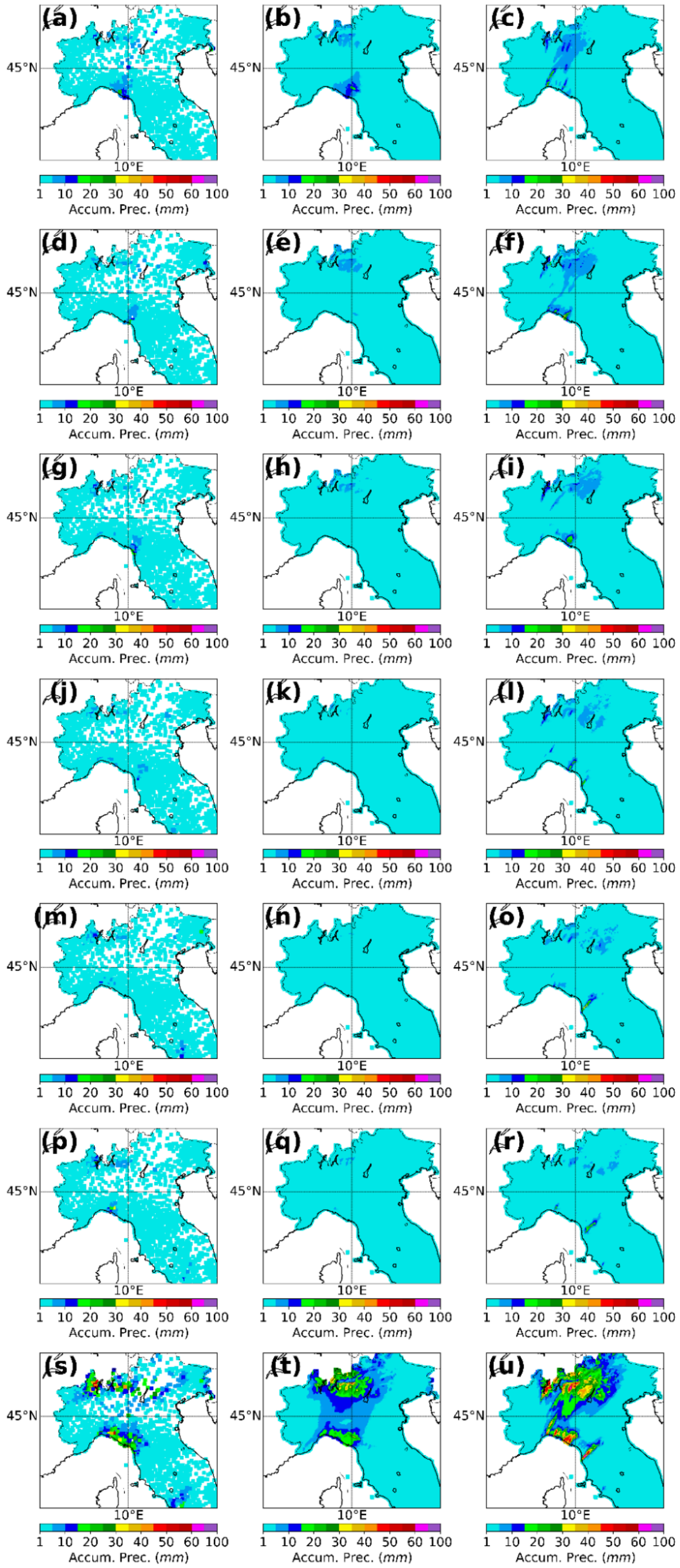

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
(m)
(n)
(o)
(p)
(q)
(r)
(s)
(t)
(u)
45°N
10°E
1 10 20 30 40 50 60 100
Accum. Prec. (mm)

Fig. A1. *1-h accumulated precipitation computed from 00-06 UTC 15 October 2012 associated with Observations (first column), NODA (second column), EnKF (CNTRL) (third column) and 3DVar (CNTRL) (fourth column).*

Fig. A2

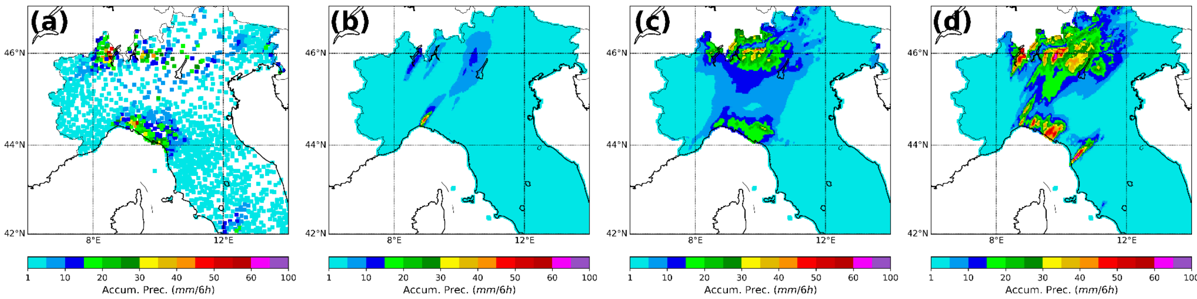


Fig. A2. *6-h accumulated precipitation computed from 00-06 UTC 15 October 2012 associated with (a) Observations, (b) NODA, (c)EnKF (CNTRL), (d) 3DVar (CNTRL).*