Review of the article "Mediterranean Tropical-Like Cyclones forecasts and analysis using the ECMWF Ensemble Forecasting System (IFS) with physical parameterizations perturbations" by Saraceni et al. submitted to *ACP*, Copernicus

Dear Dr. Stefano Galmarini,
As reported below, we tried to address all the requests of the two Reviewers. As a consequence, the revised version of the paper is very different from the first submission. All the noticeable changes we made can be deduced by comparing the two versions and by means of the included tracked-changed version (in **red** the parts that have been **deleted** and in **blue** those **added**).

REVIEWER #1

I read the manuscript with great interest. I appreciated the methods and richness of the modeling framework. However, my opinion is that the paper is not ready for publication. In terms of presentation, English and technical language need to be improved. I strongly suggest careful editing throughout the text. Especially awkward use of technical language leads to uncertainty about several concluding remarks. Please also refer to several specific comments up to line 385. In addition to the presentation, the introduction is not deepening into the subject and it seems that this extends to the interpretation and discussion of the results. In fact, important late advances in the state of the art are not taken into account and relationship between cyclones development and underlying processes is not adequately demonstrated. Finally, I believe that the scope of the study is not specific enough (see also specific comments). Actually the experiments might tell us more about the sensitivity of the cyclones to diabatic processes and less about the role of SPP in forecasts. To complete the latter, probably you should also use SPP+initial condition perturbations as it would be done in actual operational forecast conditions.
Given the important methods my suggestion to the authors is to withdraw and resubmit. Therefore, they could reorganise and be more specific on their objectives. For instance, they could break down cyclones into processes, discuss accordingly which ones are important for the different cases and make a more efficient linking with cyclone tracks and intensity.

Thanks to the extensive feedback from reviewer #1, the paper has significantly improved. First and foremost, we have restructured the entire manuscript from the Introduction Section to the Results Section. We have introduced the ensemble forecast with perturbed initial conditions and the SPP simultaneously, the TOT ensemble. As suggested by reviewer #1 this has been done primarily as a benchmark for the other experiments and for the sake of completeness in the initial part of the results to underline the role of SPP in forecasts. In the Introduction Section the authors have explicitly outlined the paper's objective, which is to assess the ensemble results concerning the ECMWF new ensemble predicts medicanes for extreme events such as Mediterranean cyclones and to investigate and justify these results by seeking similar physical processes and identifying the factors contributing to their successful or unsuccessful reproduction across different medicane cases. The goal of the study has been also addressed in the following responses to the reviewer #1 comments, where the changed paragraph in the Introduction Section has been quoted. As underlined in the Discussion and Conclusion Section of the manuscript the first outcome relates to the similarities between ensembles with SPP and perturbed initial conditions.

The second part of the Results Section has undergone a complete re-organization. Following the recommendations of reviewer #1, we have categorized the physical processes into formation and intensification processes. In each subsection, we initially discuss the operational analysis and subsequently delve into how the ensembles simulate each cyclone. We have paid particular attention to integrating the discussion of vertical profile analysis, which was previously presented separately, seamlessly within the process section. Much effort has been devoted to establishing connections between the process outcomes and the earlier Sections on tracking, intensity, and precipitation, both in the physical processes results and discussion.

The revised paper's organization is as follows:
Section 1: Introduction,
Section 2: Data and Methods,
    2.1 Ensemble Forecast Simulation
    2.2 Stochastically Perturbed Parameterizations Ensemble
    2.3 Validation Data
    2.4 Cyclone tracking
    2.5 Hart Parameters
Section 3: Overview of the storms,
Section 4: Results: Ensemble forecast evaluation,
    4.1 Tracking
    4.2 Intensity
    4.3 Precipitation
    4.4 Thermal structure and asymmetry
Section 5: Results: Physical processes analysis
    5.1 Cyclones Formation
5.2 Cyclones Intensification
5.3 The role of the Sea Surface Temperature
Section 6: Discussion and Conclusion.

In particular, the authors have added more concise explanations in the physical processes analysis Section and in the Discussion Section and have attempted to enrich the analysis with additional citations. According to what is presented in this paper, we tried to stress that Trixie's poor simulation is attributed to its weakness compared to the other two cyclones, the simulation of the cut-off low, and the absence of interaction with surface fluxes. Some of the new paragraphs are reported below:

"The surface level production of PV in Trixie is not enough to match with the upper tropospheric high PV field as reported in Fig. 12e and f, or they are not in phase. In the ensemble simulations starting on the 27th, the amount of diabatically generated PV comparable to the analysis is not able to align with the upper-level one, even though the alignment is better captured with starting dates closer to occurrence (Supplementary Fig. S6). This pertains to both the INI and SPP experiments, as mentioned before, meaning that the production of the right diabatic processes is equally sensitive to initial conditions and to physical and more specifically convective processes. The analysis points out that the cyclone dies before the 30th of October, due to the low-level vortex being too weak and not able to interact properly with the upper-level disturbances. Actually, the surface vortex disappears with the weakening of the upper-level cut-off low, and due to the absence of a reinforcement of the lower level PV production by the upper level one, meaning erosion of PV by the presence of intense diabatic processes, as for instance is happening in Zorbas (Portmann et al., 2020). Firstly, as mentioned in the previous Section, in the ensemble simulations of Trixie (both INI and SPP), the cut-off low that was crucial for the formation of the surface cyclone, disappears after the 29th and the formation of the second cut-off low weakens the first. With the weakening of the cut-off low, Trixie starts lowering its intensity. Only a few members can follow the weak cut-off low

(Supplementary Fig. S3). This has had an impact on the simulation of the already weak cyclone, Trixie. Indeed, it is hypothesized that Trixie is formed as a lower-level vortex mainly due to the thermodynamic disequilibrium generated by the upper-level cut-off low. If the latter is not well simulated, being weaker, it is not able to sustain the surface vortex. Secondly, when the surface level vortex is better reproduced, the misalignment with the upper-level PV anomaly does not permit reinforcement of the convection. Homar et al. (2003); Cioni et al. (2018) showed that the upper-level PV structure indirectly acts on the cyclone deepening through a modification of the surface circulation.

As mentioned above, the simulation of strong convective activity seems to be important in all three cyclones, which is associated with latent heat release, developing 600 in the central region of the cyclones (Fig. 8, 9 and 10). As recognized in the literature, either the surface fluxes or convective activity can predominate in the intensification of the surface level vortex, as identified for different medicanes (Davolio et al., 2009; Miglietta et al., 2017; Chaboureau et al., 2012; Fita and Flaounas, 2018). While surface fluxes seem to have also played a role in the intensification of both Ianos and Zorbas (Supplementary Fig. S7), in the case of Trixie specifically deep moist convection seems to be the main mechanism leading to the maintenance and deepening of the system, as due to the interaction of convection with the upward forcing induced by the PV streamer (Chaboureau et al., 2012). Indeed, for Trixie is the long-lasting deep convective activity that may have played an important role in their intensification later than their genesis time, as also recognized by Dafis et al. (2020), and its weakening or even absence compromises the forecast, as happens in the presented ensemble experiments. While the tracking is mainly influenced by the position of the cutoff low and the PV steamer position, the simulation intensification phase and timing are dependent on how the convection is reproduced and interacts with the upper-level PV structure."

And also :

"Specifically, regarding the reproduction of the minimum central pressure by the ensemble forecasts (Fig. 4) there is generally a time shift in the reproduction of the minimum intensity, where both the Ianos and Zorbas ensemble means reach the maximum intensity with a delay. Since the delay decreases as the forecast starting date gets closer to the event, this is due to the improved initial conditions. Indeed, it is found that the maximum intensity is reached when the upper-level PV streamer and the surface-level vortex are aligned, and the latter is better captured by starting the simulation closer to the cyclone's onset. This is consistent with what was found in Flaounas et al. (2015) where they underlined that the intensity of the surface cyclone increases while the streamer is on the western side of the cyclone and begins to decrease as the streamer is wrapping over its centre. It is noteworthy that for Zorbas the intense phase (i.e. low values of minimum core pressure) is maintained for a longer period, compared to the operational analysis, as underlined by the minimum core pressure trend in Fig. 4, and by the PV tower reported in Fig. 12. This is probably due to the fact that most members (especially regarding the SPP ensembles) spend most of their lives at sea and do not cross land over Greece, thus managing to sustain themselves for a longer period through condensational heating and air-sea interaction."

And also :

“Then, it has been shown that after the formation of the three cyclones, the operational analysis represents a clear interaction between the surface level vortex and the PV anomaly in the upper troposphere in all three cases, the PV tower, in order for them to intensify and eventually reach “tropical-like” features (Carrió et al., 2017; Cioni et al., 2018; Flaounas et al., 2022). This is present in the ensembles (both for SPP and INI) in the case of Ianos and Zorbas, where, the upper-level PV anomaly brought by the PV streamer is able to match with the lower-level PV production by diabatic processes. As also pointed out, this surface cyclone is sustained by convective heating release by condensation, with surface fluxes supporting the cyclone in the most intense phase along with convective activity (Supplementary Figure S7). In the case of Ianos, the convective activity peaks during the moist intense phase, as underlined by the PV anomaly in the PV cross-section in Figure 11, while in the case of Zorbas, convective activity is present also before the cyclone tropical like phase (as also evidenced in Figure 11 by the PV anomaly).

The ensemble simulated cyclones are sensitive to the positioning of the PV streamer, which, as observed in the literature, has a clear impact on the simulation of the medicane. In the case of Zorbas, as noted in Portmann et al. (2020), a westward shift of the PV streamer on average leads to weaker cyclones, which can be observed in the ensemble mean of Zorbas in Figure 4. The displacement of the PV streamer is linked to the alignment with the surface level and the consequent creation of a PV tower. Furthermore, the intensity of the PV anomalies, as well as the height reached by the 2 PVU isosurface, may have played a role in the cyclone’s greater or lesser intensity. In fact, in the case of Ianos, it can be argued that the cyclone is more intense in the ensembles compared to the analysis due to the higher simulated PV anomaly and deeper PV streamer. In other studies, also using the ECMWF ensemble forecasting system (Chaboureau et al., 2012), it was found that later forecasts are able to capture better the thermal structure of the medicanes, due to the lower uncertainty of the positioning of PV streamer connected to the generation of the medicane. In general, in our study is assessed that the earlier the starting date the greater the misalignment (or the worse positioning of the upper-level disturbance) between the lower and the upper PV production. This is true particularly in the case of Trixie, where on the simulation starting too early, on the 25th, the lower-level PV production is absent, and for the simulation starting on the 26th and the 27th it is very weak and not able to be reinforced by the upper level one.

Trixie is also a weaker medicane compared to Ianos and Zorbas (the central minimum pressure of Trixie is lower than the other two cyclones, as reported in Table 2), influencing the simulations. In a recent study by (Panegrossi et al., 2023), while Ianos and Zorbas have been recognized as two of the most intense medicanes, Trixie has been characterized as one of the weakest, not even reaching the status of deep warm core, as established by looking at passive microwave measurements and products. Here, it has been shown firstly that the surface fluxes in the phase of tropicalization and intensification are lower than the other medicanes. Secondly it is underlined that the SST anomaly in the case of Trixie was low both in the analysis and in the ensembles, yielding to a weaker cyclone, and lastly that Trixie tends to shut off when the cut-off low vanishes in the simulations and generally tends to follow the PV streamer position. This makes it subject to being more dependent on the simulation of large-scale processes. More specifically the lower-level PV production present in the analysis in Figure 11 is absent in the simulations in Figure 12. Since the surface fluxes and the anomaly of SST do not play a role, this makes the convective instability brought by the cutoff low necessary to sustain the cyclone convection. The deep cut-off low absence in the simulations can be linked to the medicane missing development in the case of Trixie. This is supported by the findings of Fischer et al. (2017), who hypothesized that tropical cyclones intensification rate after tropical cyclogenesis, in environments of upper-tropospheric troughs, is closely linked to the structure and temporal evolution of the upper-level trough."

However, interpreting sensitivity to diabatic processes was challenging, as we couldn't investigate specific sources of uncertainty beyond analysing the difference and similarities within the SPP and the SPP-Conv experiments. Furthermore, given the random nature of the applied perturbations, which change at each time step, it is not feasible to isolate the most influential source of uncertainty throughout the simulation. This aspect extends beyond the scope of this article and will be explored in subsequent studies, encompassing a broader range of cyclones.

However, we have attempted to emphasize in which cyclones the surface fluxes were crucial for cyclone intensification and in which they were not, as for instance the subsection "the role of the sea surface temperature" reports.

Specific comments

Line 16: "dynamically driven" is meant advected?

Thanks to Reviewer #1 we revised the phrase accordingly.

Lines 16-17: "that is driven by surface heating and stratiform and convective condensational heating" I suggest you replace with "diabatically produced by latent heat".

Thanks to Reviewer #1 we revised the phrase accordingly.

Line 21: please modify as "visual similarity" or "phenomenological". The way it is formulated, the phrase suggests that the two kinds of systems are governed by similar dynamics.

Thanks to Reviewer #1 we revised the phrase accordingly.

Lines 39-40: The two phrases do not articulate. The first one compares Medicanes to othe Mediterranean cyclones. The second starts with "indeed" but compares tropical cyclones to Medicanes.

Line 39: Actually, this issue is still debatable. In fact there are more evidence that cyclones already identified as Medicanes do not differ much in terms of processes from any other extratropical cyclone. Please revise or remove this phrase.

Lines 41-43: This part is not very accurate. Almost no intense cyclone may develop in the Mediterranean without baroclinic forcing. In fact, one of the main community questions about

the formation of medicanes is whether an already formed cyclone may be sustained uniquely by diabatic forcing, i.e. hether the so called tropical transition applies to medicanes (please refer to the review paper by Flaounas et al. 2022)

From lines 39-43 the whole paragraph was revised according to reviewer #1 suggestions and it read as follows:

“Medicanes differ from other Mediterranean cyclones in the complexity of their formation and evolution. However, unlike hurricanes, which develop in regions with near-zero baroclinicity and draw their energy from warm tropical oceans, medicanes form from pressure lows under moderate to strong baroclinicity, which is a typical condition of midlatitudes and Mediterranean cyclogenesis. Indeed, medicanes are regarded as baroclinic cyclones evolving into vortices with structural characteristics similar to tropical cyclones (Flaounas et al., 2022). The debate is still open on which processes sustain the cyclone development, baroclinic instability or pure diabatic forcing which also marks the tropical transition phase (Flaounas et al., 2022; Miglietta and Rotunno, 2019; Flaounas et al., 2021).”

Line 43: I am not sure I understand this phrase. How can warm sea interact with a trough? Is it meant that stability is reduced if colder air in a trough is advected right above warmer air masses at the sea surface? If so, I would argue that convection in this case might be favoured due to forced large scale ascent. Furthermore, in the cited papers, it is discussed the development of medicanes in conjuction with the WISHE mechanism but WISHE supposes no "upper-level troughs" to develop a warm core.

Lines 45-47: Please be more thorough to explain what is referred as "mechanics". In connection with my previous comment, please note that McTaggart-Cowan et al. (2015) also refer to the concept of tropical transition.

Lines 49-51: Large scale and diabatic forcing might indeed have an important role in medicanes development however this is true for any extratropical cyclone. What makes medicanes rather different is the relative contribution of these two forcings in different stages of their lifecycle. You might want to refer to the three paradigms in the conclusions of Miglietta and Rotunno (2019).

Line 43: What the authors meant was indeed that stability is reduced if colder air in a trough is advected right above warmer air masses at the sea surface and agree that convection, in this case, is favoured to forced large-scale ascent. The paper by Emanuel et al. (2005) was cited in this context as it refers to this as the type of environment that helps the development of Mediterranean tropical-like cyclone : “Mediterranean hurricanes usually, and perhaps always, develop under deep upper tropospheric troughs, in regions of small baroclinicity but large air-sea thermodynamic disequilibrium owing to the unusually deep, cold air associated with the trough. (This is also true of “polar lows”; Rasmussen et al., 1992.)”. However, it might be more appropriate to cite the paper of Pytharoulis et al., (2000) to refer to this and only use the paper by Emanuel et al. (2005) only to refer to the role of WISHE on medicanes. This part of the Introduction was changed according to the reviewer #1 comment.

Lines 45-47: Mechanism refers to the increase in efficiency of the conversion of thermal energy into mechanical energy given the cold air intrusion when passing over warm air in the extratropics. The authors agree with reviewer #1 as this reference was used here maybe too lightly. Thus this part was repetitive and unnecessary and was deleted.

Lines 49-51: The authors agree with reviewer #1 comments and therefore have changed this part of the paragraph to account for Mediterranean tropical-like cyclones classification based on Miglietta and Rotunno, (2019) and the review paper by Flaounas et al., (2022).

Given all these changes the paragraph in the Introduction has been changed and now reads:

“Typically, the initial phase of a medicane life cycle is similar to that of an extratropical cyclone, where the medicane intensifies through the interaction of an upper tropospheric disturbance (Potential Vorticity streamer (Flaounas et al., 2015)) with a low-level baroclinic area. However, their development is what makes them different given the relative contribution of large-scale forcing, air-sea interactions, and convection at different stages of their lifetime. Recently a classification has been produced for this type of phenomenon by Miglietta and Rotunno (2019) and Dafis et al. (2020). Medicanes have been grouped into three categories: those where baroclinic instability plays an essential role throughout the cyclones’ lifetime and most of their intensification can be attributed to convection; those where baroclinicity is relevant only in the initial stage, and, the theory of wind-induced surface heat exchange (WISHE (Emanuel, 1986)) can explain their intensification through positive feedback between latent heat release and air–sea interactions, although WISHE may only take place after the occurrence of tropical transition, i.e. after organized convection near the cyclone centre is capable of sustaining the vortex; and finally those, including smaller-scale vortices, that develop within the circulation associated with a synoptic-scale cyclone.”

Line 51: what is meant by "limited data availability"? Also, why the low frequency of occurrence is related to poor predictability?

With limited data availability, the authors meant that the cyclones usually spend most of their life at sea, limiting the amount of data available to study them. Moreover, the data availability for the authors was connected to the low frequency of occurrence, thus the low number of cases that happen every year. Nonetheless, we agree with reviewer #1 that both the limited data availability and the low frequency of occurrence are related to the ability to forecast medicanes and have little to do with their predictability, thus we delete the reference to predictability from that sentence.

Line 55: Not sure I understand this argument. In this paragraph it is argued that large scale forcing and convection are the two main processes that develop cyclones into medicanes. Both might be explicitely or well reproduced even in relatively coarse resolutions (e.g. 10-25 km). Could you please elaborate more on the added value of convection-permitting simulations.
Line 57: It is not clear to me what is meant by "rather than observational aspects"?

Following reviewer #1 comments the authors changed the whole paragraph. Indeed, the authors meant to list previous studies of different kinds, starting with climatological studies, research based on observations, which has been rephased instead of using “observational aspects”, convective-permitting simulations and ensemble forecasting studies. We summed up and rephrased the paragraph differently and while we inserted some sentences to explain why so many studies are concerned with high-resolution modeling, we agree with reviewer #1 claim that large-scale forcing and convection might be well reproduced even in relatively coarse resolutions. Indeed, in the review paper by Flaounas et al., (2022) is mentioned that: “… Although this result appears promising, a systematic gain from kilometre-scale resolution has not been generally demonstrated for cyclones yet. Besides, improvements of physical parameterizations are needed, in particular for microphysics and turbulence, since they remain two major sources of uncertainty."

Thus, the whole paragraph from the Introduction now reads as follows:

"Because of their small size, low frequency of occurrence (Cavicchia et al., 2014), and the complex geography of the Mediterranean region, predicting medicanes is a challenge for numerical weather forecasting. There are some climatological studies on medicanes, using synthetic production of tracks and 3D numerical simulation (Romero and Emanuel, 2013; Cavicchia et al., 2014; Zhang et al., 2019) which have assessed the climatological medicanes number per year, seasonal pattern, areas of occurrence and intensity. There are fewer studies based on observations (Pytharoulis et al., 2000; Moscatello et al., 2008; Miglietta et al., 2013) given the above-mentioned low frequency of occurrence and scarcity of data, which were focused on the analysis of medicanes convective activity. Numerous modeling studies of medicanes using convective permitting models and general circulation models include Davolio et al. (2009); Miglietta et al. (2011, 2013); Mazza et al. (2017); Cioni et al. (2016); Ricchi et al. (2019) with the importance of model resolution discussed in the review paper by Flaounas et al. (2022). Among the others, Carrió et al. (2020) was able to capture, with high-resolution modeling (2.5 km) the development of a small-scale cyclone and its relationship to convection, especially highlighting the role of diabatic heating in its intensification. Cioni et al. (2018) found out that explicit convection is necessary to capture the track, intensity, and thermal structure of a specific medicane in 2014. However, in the above-mentioned review paper, it is concluded that: "...a systematic gain from kilometer-scale resolution has not been generally demonstrated for cyclones yet"."

Line 65: If I am not mistaken, Di Muzio et al. (2019) suggest that the prediction skill is highly variable and most of the times predictability is poor in lead times of more than 4 days. Indeed, this comes in contrast with the following phrases.

This was corrected according to reviewer #1 comment.

Line 73-75: Is it meant that in the studies of Di Muzio et al (2019) and Portmann et al (2020) the ensemble members did not use SPPT? If yes, please revise accordingly.

Since the authors meant exactly that this sentence has been fixed, following the reviewer #1 comment.

Line 82: This is awkward phrasing. Please specify how are forecasts improved, and what is meant by "specifically with ECMWF ensembles and tropical cyclones". Also please comment on the added value of SPP. Especially concerning the improvement of the forecast, could you please be more specific on the added value of SPP. I think that forecast reliability with SPP is comparable to the one when using SPPT.

Lines 82-84: Awkward phrasing. Please be more specific that SPP is perturbing physical parameters within the parametrisations, instead of parametrisation outputs (tendencies) as in SPPT.

The Stochastically Perturbed Parametrization Tendencies (SPPT) and the Stochastically Perturbed Parametrization (SPP) present similar reliability as reviewer #1 points out. However, the authors probably failed to underline properly the added value of the SPP. Indeed, the latter is that the SPP perturbs the amplitude and the shape of the tendencies from the individual physical processes, thereby also allowing for the generation of clouds and convection, thus it does not only perturb the amplitude of the total physics tendency as the SPPT does. One could ask why the need for a new scheme to represent model uncertainty. As mentioned above, the SPPT scheme randomly perturbs the net of the physical tendencies at each time step, and the perturbation is produced with multiplicative noise. This leads to greater uncertainty attributed to larger tendencies while preserving the relative balances between the tendencies of different physical processes. However, SPPT is not able to trigger new states (for a single time-step), but it is only able to enhance or diminish the effect of the represented physical processes, and it introduces inconsistency between the perturbed physics tendencies and fluxes that are computed from the unperturbed tendencies. Since there is no correction for the surface fluxes, after perturbing the atmospheric tendencies, an energy imbalance is introduced into the system. Furthermore, the SPP can also represent uncertainty beyond a simple amplitude error, unlike the SPPT; e.g. the uncertainty in the shape of a heating profile (Leutbecher et al., 2017). A in depth discussion on the comparison of different Stochastic representations of model uncertainties within the ECMWF model was made in Leutbecher et al., (2017). Regarding the forecast improvements and the connection to tropical cyclones, the authors can refer also to the paper by Ollinaho et al., (2017) and Lang et al., (2021) as the most recent papers on SPP and SPPT comparison papers. Thus, following the reviewer #1 comment and suggestions we revised that sentence and added a small paragraph in the Introduction, which now reads:

"Indeed, this is a novel stochastic representation of model uncertainties which is still under development at ECMWF in order to replace the Stochastically Perturbed Parameterization Tendency scheme (SPPT) (Palmer et al., 2009). SPP consists of a set of physical parameters in the model being perturbed (Ollinaho et al., 2017). The added value of SPP is that it perturbs the amplitude and the shape of the tendencies from the individual physical processes, thereby also allowing for the generation of clouds and convection, thus it does not only perturb the amplitude of the total physics tendency as with SPPT. Leutbecher et al. (2017) and Lang et al. (2021) report on skillful forecast with SPP in the ECMWF ensemble system including for tropical cyclones, where SPP increases the spread of the tropical cyclone core pressure while presenting similar statistics to SPPT for the cyclone tracks. As discussed in Frogner et al. (2022) ensemble applications using SPP are clearly on the rise as it allows the representation of uncertainty close to the actual source of error and maintains physical consistency, particularly with local conservation of energy and humidity (Lang et al.,2021)."

Line 93: should physical be changed to diabatic?

Line 117: "are run". Familiar language

Line 117: "forecasted period" please change to "duration of simulations"?

Thanks to Reviewer #1 we revised lines 93 and 117 accordingly.

Lines 119: Please provide motivation for the chosen initalisation dates.

Following reviewer #1 a sentence was added to explain how the initialization dates were chosen. We did a previous analysis of each cyclone by using ERA5 reanalysis, thus we picked the dates by looking at the timing of cyclone intensification, meaning the deepening phase (maximum intensity) as reproduced by the reanalysis. For this reason, we added: "The three dates were chosen as three days before the intensification phase of each cyclone, based on the reference data of ERA5 reanalysis"

We acknowledge that the selected starting dates vary among cyclones, especially in comparison to the operational analysis, for instance, are too early for some of the cyclones compared to the start of the track in the operational analysis. Nevertheless, the results of the presented study would remain consistent even if we had extended the period of initialization dates.

Lines 122-123: Please provide more information about the coupling, e.g. resolution of the models, frequency of exchanges etc.

As requested by reviewer #1, this information have been provided with the following paragraph added to the data and method Section:

"The ensemble forecast is coupled to the ECMWF Wave Model (ecWAM:(IFS Documentation CY47R3, 2021c)) and to the Nucleus for European Modelling of the Ocean (NEMO) ocean model (Mogensen et al., 2012). The ecWAM model provides the atmospheric model with the Charnock parameter, thus controlling the sea surface fluxes via the surface roughness. The atmospheric oceanic surface heat and moisture fluxes are controlled by the SSTs computed from NEMO every 20 minutes using a 0.25° horizontal grid with 75 vertical levels. Due to the limited horizontal resolution of NEMO, the IFS is forced in the middle latitudes only, with fixed SSTs from the OSTIA product (Donlon et al., 2012) up to day 4, while beyond day 4 the SSTs from NEMO are used."

Line 130: "the whole parametrisation...". This phrase is unclear please improve technical language. What is "sub-grid orography"? What kind of parametrisation is meant by "large-scale precipitation"?

The sub-grid orography scheme is a part of the turbulence and diffusion scheme that is included to represent the turbulent orographic form drag induced by small-scale (< 5 km) orography. Indeed, the effects of unresolved orography on the atmospheric flow are parametrized as a sink of momentum (drag) (Beljaars et al., 2004). This orographic drag parametrization (Chapter 4 of the IFS documentation of physical processes) represents the effects of low-level blocking due to unresolved orography (blocked flow drag) and the absorption and/or reflection of vertically propagating gravity waves (gravity wave drag) on the momentum budget (for more details see Lott and Miller, 1997).

Instead, the clouds and large-scale precipitation scheme (Chapter 7 of the IFS documentation of physical processes) stands for a parameterization of clouds and precipitation with prognostic equations for cloud liquid, cloud ice, rain and snow water contents and a sub-grid fractional cloud cover. The cloud scheme represents the sources and sinks of cloud and precipitation due to the major generation and destruction processes, including cloud formation by detrainment from cumulus convection, condensation, deposition, evaporation, collection, melting and freezing. The scheme is based on (Tiedtke, 1993) but with an enhanced representation of mixed-phase clouds and prognostic precipitation (Forbes and Tompkins, 2011; Forbes et al., 2011).

Line 130, according to reviewer #1 comment, has been changed to "In the latter (SPP), the convective, radiative, clouds and large-scale precipitation, turbulence, diffusion and sub-grid orography parameterization parameters of the IFS model are perturbed, as briefly discussed below." For the sake of brevity, the authors didn't think of including a thorough analysis of each parameterization of the IFS model. However, following reviewer #1 comment, they just added some information on the parameterization schemes to the already present description of the parameters chosen to be perturbed by ECMWF in the Supplementary information.

Line 138: what is meant by "poorly constrained"?

Lines 139-142: Without being an expert on SPP, I have the impression that here you describe SPPT. Does SPP include "in-space varying noise derived from in-time evolving 2D random number fields"?

In general reviewer #1 is right, to generate perturbations, SPP makes use of the same generic 2D random field generator as SPPT. However, SPPT applies this to the physical tendencies of wind components, temperature and specific humidity and SPP applies it to "poorly constrained" parameters and variables within the parametrization schemes. The current implementation of SPP allows simultaneous perturbations of up to 27 parameters and variables in the deterministic IFS parametrizations; as already mentioned in the paper; and SPP samples the distributions independently for each parameter and variable. Thus, the perturbations are uncorrelated (Lang et al., 2021).

What does poorly constrained mean?

Sub-grid parametrization schemes simulate processes that are either too fast or too small-scale to be resolved directly by the model. The schemes in the IFS model use assumptions and some simplifications, as well as many uncertain parameters (e.g. due to lack of direct observations). These latter are then used in the parametrizations to reproduce bulk descriptions of processes that are otherwise either absent from or inadequately represented by the model (e.g. turbulent and microphysical processes). Indeed, during the initial development phase of SPP, experts working on the IFS parametrization of individual processes identified 20 parameters and variables that are considered uncertain and when changed introduce significant changes in the forecast (Ollinaho et al. 2017) and then extended to 27 (Lang et al 2021). They are also chosen to describe the uncertainty across a range of meteorological regimes/phenomena and the entire atmosphere from the boundary layer and free troposphere to the stratosphere. The choices behind these parameters have been summed up, following Ollinaho et al. (2017), in the Supplementary material.

What is the main difference between the SPP and the SPPT perturbations?

SPPT generates perturbed parametrization tendencies p stochastically by multiplying the net physics tendency ***pt*** provided by the physics package with a 2D random field *r* which samples a Gaussian distribution:

$$\boldsymbol{p} = (1 + \mu r)\boldsymbol{p}t$$

where p represents the vector of perturbed tendency of wind components, specific humidity and temperature of a model column and μ is a scaling function that depends on the model level only and has values in the range [0, 1] (Buizza et al., 2009). On the other hand, in the SPP scheme, if $\hat{\xi}_j$ denote the unperturbed value of the jth parameter (with j ≤ 27), which is the value of the respective parameter used in the deterministic forecasts; the perturbed parameters are referred to as ξj:

$$\xi_j = \exp(\psi_j)\ \hat{\xi}_j \qquad \text{with } \psi_j \sim N(\mu_j, \sigma_j^2)$$

Here, the perturbations ψ_j sample a Gaussian distribution with a mean μj and a standard deviation σj, both determined individually for each perturbed parameter j. At present, two options are considered for determining the mean μj: (a) μj = −σ2 j /2 and (b) μj = 0. Option (a) implies that the mean of the distribution of the perturbed parameter ξj is equal to the unperturbed value ξj, while option (b) implies that the median of the distribution is equal to the unperturbed value. The standard deviation values are based on expert estimates about the uncertainties of each parameter and variable (Lang et al., 2021).

Thus, following the reviewer #1 comments, we revised the paragraph slightly and now reads as:

"The new scheme, developed by Ollinaho et al. (2017), following the work of Baker et al. (2014) and Christensen et al. (2015), is based on applying perturbations directly to a selected number of parameters and/or equations within the parameterization schemes, usually those known to be specific sources of uncertainty for the model. The perturbations follow horizontal patterns that evolve stochastically in space and time. Each perturbed parameter is assigned an individual random field and different random fields are statistically independent. The lognormal distribution has been chosen for practical reasons as it ensures that the perturbed parameter values retain their original sign."

Furthermore, for the sake of clarity, we added a part in the Supplementary material explaining the way the perturbation is applied in the SPP scheme.

Lines 143-145: Please improve technical language in the description of parametrisations.

See previous comment. This has been added to the Supplementary Material, following reviewer #1 comment.

Lines 173-177: What was the input dataset for the calculation of cyclone phase space (CPS) diagrams? Please be more specific. It is unclear to which life stage of the cyclone refer to the values in Table 2.

The input dataset for the cyclone phase space diagrams is the geopotential height at certain levels (the ones required by the parameters calculation, thus, 400, 700 and 925 hPa) and the values presented in Table 2 refer to the deep warm core phase as determined by the calculation of the CPS with the analysis. However, the authors revised completely this part since it was creating confusion regarding Table 2. In general, as suggested by reviewer #1, also in the following comments the part with the CPS and its computation explanation has been moved to the Data and Methods Section. Furthermore, a sentence has been added in the Overview of the Storm section to better explain the parameters of Table 2. This part now reads as follows:

"A summary of the main features as retrieved by the analysis data: the storm duration, the period, the region of occurrence, the asymmetry B, and upper-level thermal wind −V UT is provided in Table 2. The latter two parameters have been used for the computation of the cyclone phase space diagrams. The intensity (central pressure) and trajectory of each storm are shown in Figure 1 along with the ensemble track."

Line 185: What is meant by "a clear mid-level warm core"? Presumably CPS diagrams provide lower or upper level diagnostics.

Yes, this sentence is based on the diagnostic and this whole paragraph is based on pre-existing literature.

Lines 203-208: In the following, I copy text lines and provide comments to highlight uncertainty in writing style.

"Medicane Trixie formed on the 28th of October, as the consequence of a deep cut-off low which emerged on 26–27 October and moved from northern to southern Italy in the following days, triggering deep convective storms along the Italian west coast."

Here I guess that it is meant that Trixie was formed due to baroclinic forcing from a cut-off low. It is still not clear what is meant by "emerged". Usually a cut-off low is the remnants of an intruding trough. With "storms" is it meant other cyclonic systems or just convective events? It is also unclear how these storms are related to Trixie.

"This PV anomaly crossed the Adriatic Sea on 28 October at O4:00 UTC (EUMETSAT analysis by Scott Bachmeier, Jochen Kerkmann, and Djordje Gencic) and then quickly moved to Sicily and approached Tunisia and Algeria."

Here probably PV anomaly refers to the cut-off low. It could also however refer to the convective storms as an area of concise diabatiocally-produced PV. Still it is not clear what the described displacement of this PV anomaly tells us about Trixie and how/if it contributes to its development. The citation here seems to be provided in plain language.

"Then between the 26 and the 28 of October, the medicane started to develop in the area of the old PV anomaly. On the 29th of October, it deepened and moved to the east of Malta, then on the 30th of October, it moved eastward towards Greece (Figure 1g)."

What "old anomaly" refers to is unclear. In total, I am not sure what we learn about Trixie in terms of processes and how this is useful for the subsequent analysis.

This whole part was deleted, as mentioned above, following reviewer #1 comments. The authors agree that this part is unnecessary and is not useful for the subsequent analysis. Therefore, only a summary of this part is left in the Section:

" The three storms formed and developed in the same area, the Southern Mediterranean, in the Ionian and Aegean Seas. Thisregion has one of the highest medicanes occurrences, as recognized in the literature (Cavicchia et al., 2014; Zhang et al., 2021). They occurred in the same period of the year, between September and November, the most frequent period for medicane occurrence (Romero and Emanuel, 2013). From Fig. 1 it can be gathered that there are some differences in duration and intensity, with Trixie being the longest-lasting of the three medicanes (in terms of the deepening phase) and Zorbas and Ianos being deeper than Trixie, as mentioned in the Introduction. As the track suggests (Fig. 1a) Ianos originated in the Gulf of Sidra. Then between the 14th and 15th of September 2020, it emerged in the Gulf of Sidra and spent most of its life over the Mediterranean Sea, eventually reaching Greece on the 17th and turning southeastward, dissipating around the 21st. The analysis is capable of reproducing a value similar to the observed pressure minimum of 995 hPa (Comellas Prat et al., 2021). Zorbas formed on the 27th of September 2018 close to North Africa and then moved into the central Mediterranean, turning eastward and moving over Greece into the Aegean Sea, where it finally decayed four days after its formation (Fig. 1d). Zorbas reached its maximum intensity (observed 992 hPa), which is well captured by the analysis. Medicane Trixie formed on the 28th of October 2016. On the 29th, it moved to the east of Malta, then on the 30th of October, it moved eastward towards Greece (Fig. 1g) while dissipating. In the analysis, there was only a short intensification period evident, and the minimum pressure was fluctuating between 1010 and 1014 hPa during the period from the 29th and the 30th of October, which might have been highly underestimated.”

Line 225: If it is "largely influenced by large-scale processes" then how it is a "reflection of the model’s capability to reproduce multi-scale processes"? .

Line 230: Hart (2003) developed a diagnostic not a "theory". Please revise.

Lines 228-231: These are motivational phrases that mostly belong to the introduction. Also they do not really align with the objective in line 90. Please be more precise in the objectives.

This whole part from line 224 to line 231 was moved to the Introduction and revised according to the reviewer #1 comments. Indeed, the authors agree with reviewer #1 about the fact that the sentence in line 225 was contradictory indeed the track is a reflection of mostly large-scale processes, so it is not possible to gain insight into multi-scale processes only by looking at the tracking. Thus, we clear that out. We also changed the word theory into diagnostic. Now in the Introduction, there is a part that reads as follows:

“Thus, a comparison between three ensemble forecast experiments is set up. One ensemble is run with only initial condition perturbations, through the Ensemble Data Assimilation (EDA), one is run with the entire physical parameterizations perturbed and one is run with only the convective parameterization perturbed. The last experiment comprises both the initial conditions and the model parameterizations perturbations. Each of these experiments is applied to the three above-mentioned chosen medicanes and the goal of this study is to determine whether these forecasts can accurately predict them, if there are possible biases presented by the ensemble forecasts and if the ensembles compare in terms of spread and error. Furthermore, the assessment of which of the perturbation experiments can capture the medicane more accurately is carried out trying also to understand what diabatic processes, among the ones already studied in the literature, influence the forecast and if different ensembles predict these.

The outline of the paper is the following. In Section 2 the data and methods used are described, with an in-depth description of the ensemble forecast experiments carried out, a description of the SPP, of the tracking method, and of the Hart (2003) diagnostic. In Section 3 the assessment of the ensemble predictions of medicanes in terms of tracking, intensity, precipitation, and thermal structure is presented. Then, in Section 5 the relevant process involved in the evolution and prediction of the medicanes are investigated in relation to the results of the previous sections and in Section 6, the results are discussed and the concluding remarks are given."

Line 239: I do not understand what "sensibly" means. Please quantify your errors or just make reference that Trixie is following an opposite direction.

Revised accordingly to reviewer # 1 comment.

Line 265: it is stated that INI experiments present largest spread. On the other hand, line 286 suggests that the spreads between the SPP and INI experiments are comparable. Actually, the conclusion of 4.1 is not clear to me. I guess your results suggest that the cyclone track (not the development as mentioned) is sensitive to both initial conditions and parametrized processes.

The authors agree with the reviewer in stating that the cyclone is sensitive to both initial conditions and parametrized processes. When writing development, the authors had track development in mind. Furthermore, the point of these sentences was to underline that convection specifically, has the most important role here, since the spread of the SPP-Conv experiment is similar to the SPP.
Thus, we revised this paragraph according to the reviewer #1 comments. This now reads as follows:

"Interestingly, both SPP experiments, and specifically the SPP-Conv, produce cyclone spread that is comparable at the later stages of the simulation to the experiment with initial condition perturbations only. Similar results have been found in Lang et al., (2012), underlining the initial condition and physical heating-related sources of uncertainty in the tracking of these cyclones."

Lines 357-358: Here there is an analysis on the usefulness of precipitation forecasts. But in an actual forecast there would be a use of both initial condition perturbations and SPP(T). Please revise your concluding remarks.

All this Section (Section 4) was revised with the inclusion of the new set of ensemble members for the experiment including the initial condition perturbation with EDA and the physical parameterization perturbation with SPP, the TOT ensemble.

Lines 360-385: Please move this part to the methods section.

As mentioned above this part has been removed from this section to the Data and Methods one, according to reviewer #1 suggestion.

REVIEWER #2

Review of "Mediterranean Tropical-Like Cyclones forecasts and analysis using the ECMWF Ensemble Forecasting System (IFS) with physical parameterizations perturbations" by Miriam Saraceni et al.

General comments

In this new study, Saraceni et al. discuss ensemble simulations of tropical-like cyclones over the Mediterranean using ECMWF's IFS ensemble system. The ensemble simulations are conducted using perturbed initial conditions from the ensemble data assimilation (INI), the recently developed Stochastically Perturbed Parameterizations representing uncertainty due to model physics with perturbed parameters (SPP), and a reduced set of the SPP focusing on convection (SPP-Conv). The study is comparing simulations for three medicanes, Ianos, Zorbas, and Trixie in 2020, 2018, and 2016, respectively. The three cases considered here represent a reasonable selection of cases with different mechanisms responsible for generating and intensifying the medicanes and showing varying prediction skills of the ECMWF IFS ensemble system. The study evaluates the ensemble simulations of the medicanes based on different observations and metrics, namely track and intensity, precipitation, thermal structure and asymmetry, and tropical-like phase characteristics. Benefits and deficits of the modeling system and the different ensemble sets (INI, SPP, SPP-Conv) are assessed.

Overall, I think the study was carefully and designed and conducted, with plausible results. Except for being a bit long, the manuscripts is mostly well-written and clear. I would like to recommend considering the paper for publication in ACP subject to a few minor comments and suggestions as listed below.

Specific comments

l5: All simulations were conducted with 9 km horizontal resolution. While this seems to work fine here, tropical cyclone simulations are often conducted with higher horizontal resolution, in order to better cover their characteristics and dynamics like the sharp gradients in their structure near the center. It would be interesting if the authors could elaborate on whether the results of their study are significantly affected by the choice of resolution of the forecast model or not?

We thank reviewer #2 for their comment. Following what was found in the paper by Majumdar et al. (2023), it is demonstrated that the IFS 4 km simulation closely aligns with observations in terms of structure and intensity of tropical cyclones (TCs). They tested the experimental 4-km global ECMWF model (EC4) with upgraded moist physics is compared with a 9-km version (EC9) to evaluate the influence of resolution on the Structure and intensity forecasts of 19 TCs during the 2020 Atlantic hurricane season. EC4 is then benchmarked against the 4-km regional COAMPS–Tropical Cyclones (COAMPS-TC) system (CO4) to compare systems with similar resolutions. They found out that EC4 produced stronger TCs than EC9, with a >30% reduction of the maximum wind speed bias in EC4, resulting in lower forecast errors. However, both ECMWF predictions struggled to intensify initially weak TCs, and the radius of maximum winds was often too large. In contrast, CO4 had lower biases in central pressure, maximum wind speed, and radius of maximum winds. Regardless, minimal statistical differences between CO4 and EC4 intensity errors were found for ≥36-h forecasts.

Thus, the resolution actually has an impact at least for tropical cyclones. Nonetheless, we acknowledge that the work and preliminary results carried out with a IFS 4 km resolution by the Destination Earth (DestinE) project (Gascón et al. 2023), even though utilizing a resolution that is still experimental for the IFS model, did not yield significant changes in Medicanes tracking and intensity, except for slightly more intense winds. This is true for both Trixie cyclones (https://confluence.ecmwf.int/x/soQvEQ) and Ianos (https://confluence.ecmwf.int/x/w53mE) where the 9 km resolution simulation and the 4.5 km simulation are mostly similar for the above-cited quantities.

Following these considerations, the authors added some lines in the Discussion section of the paper:

"It could be argued that a finer resolution than a 9 km resolution would improve the reproduction of the intensity and track of the tropical-like Mediterranean cyclones. A recent single forecast experimental tropical cyclone simulations at 4 km resolution by (Majumdar et al.,2023) reveal that the 4 km simulations produce deeper and more realistic tropical cyclones in terms of radial wind structure compared to the observations than the 9 km forecasts. However, in the aforementioned work done within the DestinE project (Gascón et al., 2023), preliminary simulations carried out with IFS at 4 km resolution did not show any significant changes in the simulated tracks and intensity of Medicanes, except for slightly more intense winds. This applies to both Trixie (https://confluence.ecmwf.int/x/soQvEQ) and Ianos (https://confluence.ecmwf.int/x/soQvEQ) where the 9 and 4 km resolution simulations are roughly equivalent."

Gascón, E., Sandu, I., Vannière, B., Magnusson, L., Forbes, R., Polichtchouk, I., ... & Balsamo, G. (2023). *Advances towards a better prediction of weather extremes in the Destination Earth initiative* (No. EMS2023-659). Copernicus Meetings.

Majumdar, S. J., Magnusson, L., Bechtold, P., Bidlot, J. R., & Doyle, J. D. (2023). Advanced tropical cyclone prediction using the experimental global ECMWF and operational regional COAMPS-TC systems. *Monthly weather review*.

l210: Is there a reference/DOI for the EUMETSAT report?

As pointed out by the reviewer #2 the DOI/reference wasn't found. The authors can only provide the link to the website (https://www.eumetsat.int/mediterranean-cyclones). However, this whole part has been deleted from the paper for the sake of brevity.

Fig. 1: Perhaps showing the reference data (operational analysis) in front of the ensemble tracks would be more clear? In the caption, please fix "operation analysis".

Fixed according to the reviewer #2 suggestion.

l243-245: Is this something you would actually like to demonstrate, that the spread from INI, SPP, and SPP-Conv is comparable? It might be helpful for the reader to clarify whether this would be a goal of the study?

Following the reviewer #2 comment, the authors have slightly changed the Introduction in order to account for this. Indeed, we would like to retain this as one of the objectives of the paper. In the following the sentences added to the Introduction:

"… and the goal of this study is to determine whether these forecasts can accurately predict them, if there are possible biases presented by the ensemble forecasts and if the ensembles compare in terms of spread and error."

Furthermore, some sentences have been added to the Results Sections, when dealing with the tracking and intensity results.

l301: At this point, I was wondering what is actually defining the size of the initial perturbations. I realized only later, the size of the initial perturbations would be guided by their related uncertainties/perturbations found during the data assimilation? It might be helpful to add 1-2 sentences clarifying this.

Following the reviewer #2 comment the authors added the following sentence in the Ensemble explanation in Section Data and Methods: " ..the size of the initial perturbations stems from the analysis uncertainty due to observation errors and model uncertainties including SPPT also in the trajectory of the variational data assimilation. ". Indeed, the flow dependent analysis errors are sampled by the EDA through random perturbations of observations and physics.

l315: Maybe rephrase "... thus a point-by-point verification presents several problems and is guaranteed by using satellite data."? I found it a bit unclear.

Changed according to the reviewer #2 suggestion to: “Matching the forecast with the verifying data is hampered by the irregularly spaced ground network of observations and by the spatial variability of precipitation. Therefore, satellite products, specifically the above-discussed Integrated Multi-satellite Retrievals for GPM (GPM-IMERG), are used.”

l333-334: It would be good to add 1-2 sentences explaining why the secondary maxima are not well captured by the simulations. Later in the text it is suggested that this might be related to the resolution of the simulations, it seems?

See above response to comment l5. In this case the following sentence was added:

“This is possibly related to the simulation resolution not being able to capture completely the precipitation structure. Preliminary results from the recent work of the Destination Earth (DestinE) project (Gascón et al. 2023) shows that the accumulated precipitation pattern is slightly better captured by the 4 km simulation compared to the 9 km one with the IFS model https://confluence.ecmwf.int/x/w53m for Ianos, reported as an example).”

Figs. 11 and 12: It might be easier to disentangle the contour lines of sea level pressure, if the coast lines would be plotted in (light) gray color.

Fixed according to the reviewer #2 suggestion.

l537: Maybe rephrase to "reduced convective activity _near the center_" to stress this, as the Trixie maps show much larger CAPE south of the medicane than the Zorbas maps, which might be confusing.

Changed according to the reviewer #2 suggestion to: “In the ensembles, the reduced convective activity near the centre reported in Figure 13 for Trixie underlines the lower energy conversion from diabatic heating.”

Technical corrections

l103: please check throughout the paper the proper use of abbreviations (Sect., Fig., ...) according to the Copernicus manuscript preparation guidelines

l105: In Sects. 5 and 6, the results... (?)

l129: SSP -> SPP

l205: O4:00 -> 04:00

l206: please check proper use of time/date formats throughout the paper (e.g., 26_th_ of October, ...)

l270: ...)_,_ a great uncertainty (?)

Fig. 9: fix october -> October in the caption

Fig. 13: fix Potential -> potential in the caption; _at_ altitudes from 25° to 50°_N_; fix date/time format (16_th_, 27_th_,...)

l521 and 524: Emanuel (2005) -> (Emanuel, 2005)

l550: fix 2°_C_

l573: do not use contractions, "isn't" -> "is not"

l580: Indeed, _it_ is ... (?)

l612: do you mean "underlining" or "stressing" instead of "underlying"?

l624: remove comma, "... dataset tends..."

l661 and 666: fix references to the figures

l697-698: remove some of the "mores" in this sentence

l707: _of_ medicanes

All the technical corrections suggested by reviewer #2 have been taken into account and the text has been changed accordingly.