Referee comment, Alan Geer

Review of “Use of a GNSS-PRO forward operator for the evaluation of global and km-scale NWP models”

This manuscript compares PRO observations and simulations using two different forward models. Over France it also compares the PRO observations to simulations based on ground radar measurements. The work investigates the differences between the observations and the three simulations though it is not possible to come to any concrete explanations for those differences. However, the manuscript provides a good overview of the PRO technique and illustrates the possibility of using it with NWP models, both for model assessment and data assimilation. All these concepts are novel and likely of widespread interest.

The work is mostly in good shape and well presented. However, there are areas where the presentation needs to be more precise and thorough. Further, the hypotheses proposed to explain differences between the observations and various simulations are not always convincing. Therefore, there are some major issues to address, but the work would likely be suitable for publication after these are addressed.

The authors would like to thank the reviewer for its careful review of the manuscript and for providing valuable suggestions which significantly improve the quality of the work. In accordance with the suggestions, the authors have thoroughly checked the manuscript and shared the point by point responses to the comments. In the revised pdf file, the scientific changes are highlighted in orange. Other corrections, which are linked to some corrections made by the authors itself are displayed in blue in the revised pdf file. They are written and explained at the end of each answer.

Main issues

1) It is strange that the forward operator based on Bayesian estimation retains a bias against the observations that it was trained on (e.g. L298 - 305). Minimisation of equation 8 would naturally reduce the bias to zero over the whole profile, if the observation errors in R were constant. In practice the errors are not constant, so this effect may already explain the phenomenon. However, this needs to be explained and investigated in the manuscript. It is important to do this in order to demonstrate that the Bayesian inversion is working correctly. Further, it does not seem justified to attribute the differences to an underestimation of frozen hydrometeors in AROME: the Bayesian estimation should have already corrected this underestimation by adjusting the estimated mass scattering efficiency appropriately.

(L372-374) We would like to clarify this point, as we may not have explained it clearly in the submission. The reviewer is right that unequal observation errors in **R** affect the fit, but this is not the main reason the residual bias cannot be driven to zero. The retrieved scattering efficiencies are single scalars, one value per hydrometeor, with no height dependence. The minimization of Eq. 8 only guarantees that the weighted residual is orthogonal to the model’s regressors (water content), summed over the entire system of 27 occultations. It does not guarantee zero residual within any particular height bin.

This is a structural property of using a single scalar efficiency per species, not a sign that the inversion has failed to reach its optimum. We have revised the text accordingly to make this

explanation explicit, and have softened the attribution of the residual bias to an underestimation of frozen hydrometeors in AROME, since a height-invariant correction does not, by itself, establish that underestimation independently of this limitation.

2) The negative phase shifts in the simulations based on ground radar are very interesting. It is asserted on lines 272 - 274 that this comes from vertically oriented hydrometeors but it would be important to back this up with citations or further justification. And I am not convinced by the suggestion that the ground radar simulations are not representative of the PRO full ray path. In order to generate negative phase shifts averaged along the whole simulated ray path, the negative values of K_dp must be relatively widespread across multiple radar pixels along that path. To back up these assertions, there needs to be more investigation into how this signal looks in the ground radar. In particular, the issue might be much clearer after looking at maps of K_dp at each radar elevation.

Thanks a lot for this comment. First of all, after carefully checking our observations to better explain the differences between radar and GNSS-PRO observation, we have found needed corrections in the first submitted version:

- Conversion °/km to mm/km for the radar estimated ΔΦ (to have a consistent unit with the PRO observation).

- Correct nearest radar observation to PRO observation for case FM166.2023.190.10.03: In the submitted manuscript I have used 10:10 UTC radar observations, when in reality 10:05 UTC is the nearest version.

These two corrections slightly changed the results presented in the manuscript in Figure 3 for the case FM166.2023.190.10.03, and the Figure description has thus been updated as well. In particular, with the corrected time step (10:05 UTC instead of 10:10 UTC), the negative values of ΔΦ are not visible anymore in the radar observation (see below).

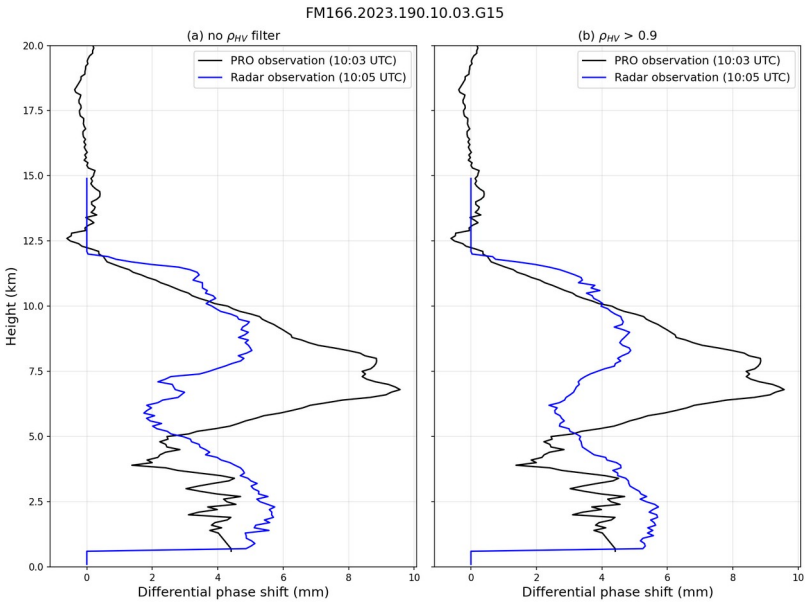


Figure 1: (Figure 3 from the revised version) with radar observation at 10:05 UTC, instead of the 10:10 UTC: Comparison between the PRO ΔΦ observation (black line) and the ΔΦ from ARAMIS for case FM166.2023.190.10.03.G15.

In this updated Figure, the ΔΦ observed by the radar is slightly lower because of the conversion to mm. However, the values slightly increase when applying a more restrictive ρ_{HV} filtering, that has been added to remove the negative Kdp echoes along the ray suspected to be due to non uniform beam filling contamination (see explanation below).

Negative peak explanation encountered at 10:10 UTC

We have plotted the radar observables into the 2D ray geometry of PRO, in order to identify the negative values that generate this peak in FM166.2023.190.10.03.G15.

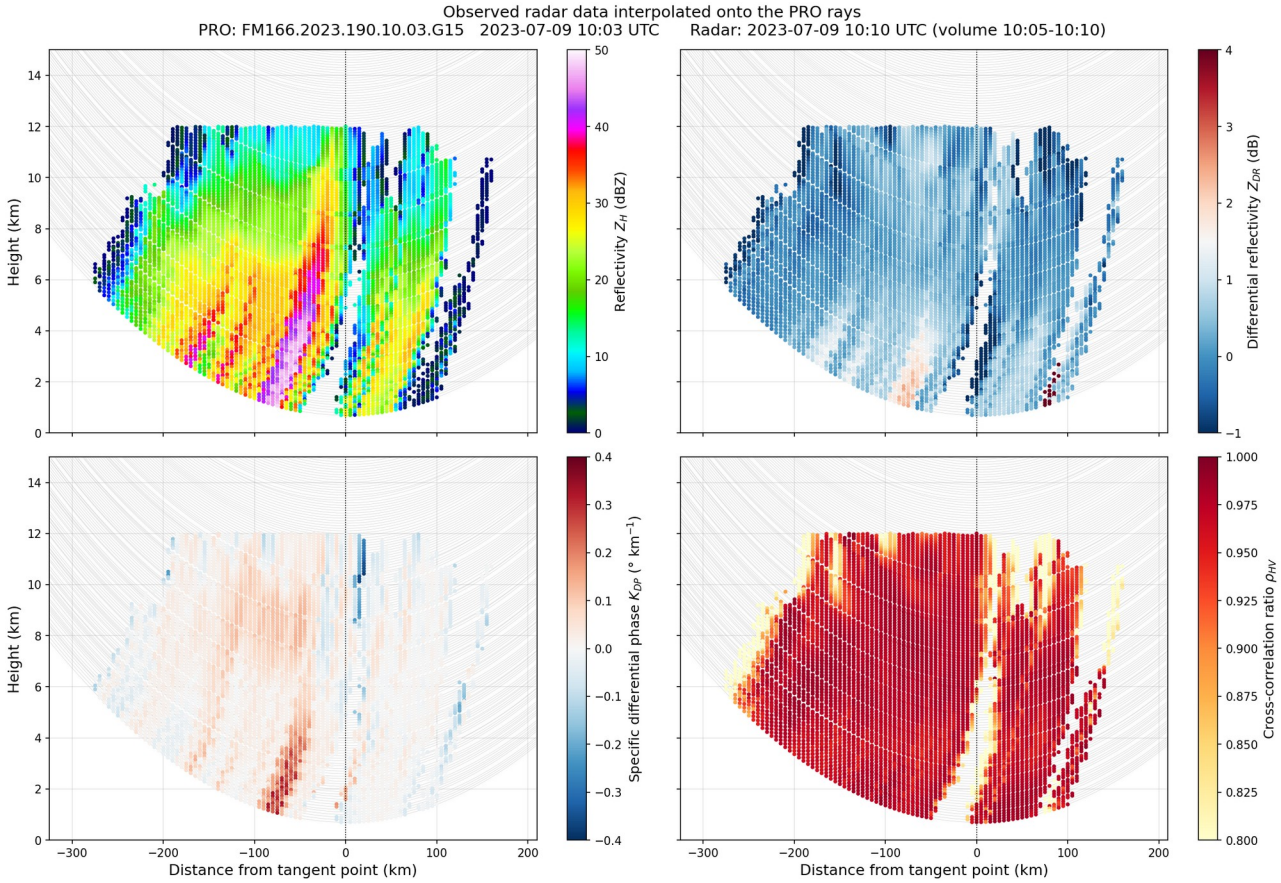


Figure 2: Radar variables interpolated to the 2D ray PRO geometry at 10:10 UTC.

Negative Kdp values are visible around the tangent height of 10-12km, as in the vertical profile of ΔΦ, with values around -0.2 to -0.4 °/km. This region with negative Kdp values is associated with low cross-correlation coefficient values (ρ_{HV} < 0.9) and also low reflectivities (< 5 dBZ). Contrary to our previous interpretation where we attributed these values to the presence of vertically oriented particles, we now think that this is more probably related to quality issue effects and probably non uniform beam filling (NBF). Indeed, examples of effects of vertically oriented particle due to electrical field alignement shown in the literature usually correspond to higher ρ_{HV} and relfectivity values (eg: Fig 5 c, d from Hubbert et al., 2014, or Fig 6.35 from yzhkov and Zrnic, 2019). In our case, those negative Kdp values are at high altitudes, thus corresponding to distances far away from the radar where the beam is large and where the polarimetric variables can be contaminated by NBF in case of strong precipitation gradients within the beam (as explained for example in Ryzhkov et al., 2007).

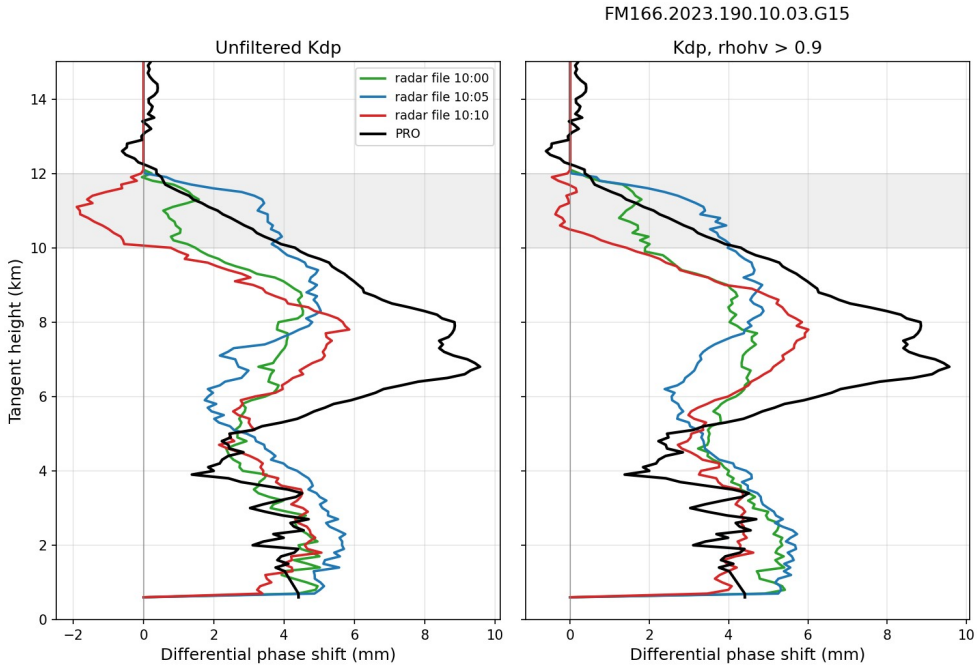


Figure 3: ΔΦ profiles from radar observations at times: 10:00, 10:05 and 10:10. Without and with filtering.

In Figure 3, we can appreciate how the negative peak is only seen in the 10:10 UTC observation, and how it disappears after applying a more restricitve filter on ρ_{HV}. No negative peak is observed at 10:05UTC, the closest observation to PRO observation.

3) L295-296 assert that “both forward operators capture the leading-order vertical structure”. The evidence presented in the manuscript is not sufficient to justify this, and errors of a few mm could be quite large in profiles such as those shown in Figure 3. To provide more evidence, it would be important to show some examples of the phase shift profiles - observations versus the two simulations - and not just averaged differences.

(L352-365) We agree that the height-resolved median bias in Fig. 4, aggregated across 27 occultations, does not on its own demonstrate that individual profiles are well captured, and that a bias of a few milimeters needs to be read relative to the amplitude of each profile rather than in absolute terms. To better illustrate how the two operators compare with observations, we have added a new figure (now Figure 5). It shows observed and simulated ΔΦ profiles for a representative subset of individual occultations, spanning weak, moderate, and strong cases in terms of ΔΦ, across the five geographical domains. We have also revised the text to qualify the “leading order vertical structure” statement accordingly, and to note explicitly that the reduced chi-squared and RMSE values in Table 1 should be interpreted relative to each region’s typical ΔΦ amplitude rather than in asbolute mm.

4) The mathematical presentation in section 2.3 needs to be tightened up. It is current incomplete and unclear:

We thank the reviewer for this comments. We addressed each point in turn and have rewritten Section 2.3.

a) Representing water content by “WC” means the notation needs to use “·” to clarify multiplication. It would be clearer to use a single character to denote water content (W or C would be fine) and then it would be possible to drop the dot notation.

We have replaced the two-letter “WC” by a single symbol “W”, for water content, which removes the need for the “·”.

b) The A in equation 3 and A in equation 7 are different, so ideally distinguish these more clearly.

We have renamed the liquid power-law coefficients from (A,B) to (a_{liq}, b_{liq}), reserving the bold, capital **A** exclusively for the observation matrix introduced in Eq. 7.

c) The notation for y, the observations or their simulated equivalent, follows two different conventions. In equation 7 the simulated observation is not distinguished (it is just Ax) but in equation 13 there are y and y_sim. I think the latter is clearest.

We now use *y* consistently for the observed $\Delta\Phi$ and y^{sim} for the corresponding simulated value, including the inversion section, rather than leaving the simulated quantity as the unlabelled product **A**·x in on place and y^{sim} in another.

d) Equation 7 does not define epsilon

We have added an explicit definition of epsilon as a zero-mean noise vector representing observational and representativeness error, with covariance **R**.

e) Equation 7 subtracts the liquid contribution from y, whereas Equation 13 presumably includes it. There should be some way in the notation to distinguish this.

We now distinguish explicitly between the full observation *y* (used in Eqs. 13-15) and the liquid-substracted quantity used as the inversion target, which we denote $\tilde{y} = y - \Delta\varphi_{liq}^{sim}$, and we state explicitly that $y^{sim} = \Delta\varphi_{liq}^{sim} + \mathbf{A}\cdot x$.

f) Vector, matrix and dot notation in equation 7 would ideally be defined. Especially the link between x_i and bold x should be made. These things may be implicitly clear, we hope, but being precise saves the reader effort and confusion. For example, vector x could run over both vertical levels and hydrometeors, or it could be fixed across all vertical levels.

(see below)

g) In particular, the dimensions of matrix A and vector x should be stated

We now state explicitly that x collects one scattering efficiency per frozen species, and is held fixed across all vertical levels of a given profile, it is the along-ray-integrated frozen water content profiles; that italic x_i denotes the i-th scalar component and bold x the length-n_s vector; and also the shape of **A**.

h) Equation 10 is not enough to specify the full linear system. It needs to be stated mathematically how the “global R” is constructed and whether the global x is one vector for all locations or whether again it is a stacked vector of the kind shown in equation 10, so that a different x can be estimated for each location. Indeed, both approaches are used in this work.

We have added the missing specifications: **R** is diagonal (occultations are assumed to have independent errors), and we now state explicitly that both approaches referred to by the reveiewer are used in this work, a single shared x retrieved jointly from all 27 occultations stacked as in the

original Eq. 10, and an independent per-case vector x_m retrieved from each occultation *m* alone by minimizing the single-case cost function.

5) The potential issue of non-uniform beam filling, and the treatment of sub-grid hydrometeor inhomogeneity, needs to be better discussed throughout the paper. It would be good to state already in the methods section whether and/or how these effects are represented in the various simulations. Note that on L433, it is true that (for example in RTTOV) the hydrometeor water content is normalised by the precipitation fraction before being used to generate reflectivities, but it is also then important to multiply the resulting reflectivities by the hydrometeor fraction that actually generates them. Coupled with (presumably) the lack of representation of the area being sampled normal to the ray path, the treatment of hydrometeor inhomogeneity may be one of the largest remaining errors in the forward simulations.

(L511-519) We agree our original treatment was too brief, and incomplete in the way the reviewer identifies. We distinguish two effects: non-uniform beam filling, and sub-grid hydrometeor inhomogeneity (the NWP grid-mean W already diluting the true in-cloud water content by the precipitation fraction, *f*). Since both operators use grid-mean model fields, is the latter that matters here.

Whether this dilution biases the forward operator depends on linearity. For the frozen term, $K_{dp}^{frozen} = x_i W_i$ is linear in W_i, so applying it to the diluted grid-mean W_i gives exactly the same result as the correct un-normalize/re-scale procedure. For the liquid term, $K_{dp}^{liquid}=aW_{liq}^{b}$ is a power-law, so this equivalence fails: applying it directly to grid-mean W_{liq} as we currently do, understates the true contribution relative to the two-step procedure the reviewer describes (normalize by *f*, apply the power law, re-scale by *f*).

We have revised the manuscript to state this clearly.

Minor issues

1) L64-65 (and also L509) discuss PRO being able to “constrain” snow and thermodynamic fields within NWP systems. The word “constrain” is ill defined and the wording could be read as suggesting that PRO observations could single-handedly provide all this information, without need for other observations. The remedy could be as simple as saying that the PRO observations could “contribute to constraining” the relevant parameters.

(L64-65) We agree and have adopted the reviewer's suggested phrasing. We have revised "constrain" to "contribute to constraining" (or equivalent), and checked the manuscript for other instances of the same overstatement, revising them for consistency.

2) Figure 1 shows apparently solid coloured shapes representing the occultation rays. The caption needs to explain how a series of rays can end up being shown as a solid shape rather than a set of distinct lines.

At the scale of the full domain map, the densely spaced points along each ray’s horizontal projection render as a filled wedge/cone shape rather than a visibly discrete set of points. We have revised the caption to state this explicitly.

3) Line 175: “lhy and fhy define the model’s liquid and frozen hydrometeor species.” Here or somewhere else in the introduction, there needs to be a clearer statement of what these are, and how they map onto the hydrometeor representations of AROME and ARPEGE (which are different).

Note that there seem to be configurations with both 3 and 4 frozen hydrometeors, when “wet graupel” is included. It is better to introduce all these configurations in the methods section, rather than in the results section.

We agree. Lhy and fhy are introduced but never defined. This has been corrected.

We also have added Table 1 in Section 2.3 to state explicitly the different types of hydrometeors treated in the study.

4) Equation 9 and L189. It needs to be explained why the standard deviation over 1 second intervals represents a sensible observation error. Further, the significance of the number 50 in equation 9 needs to be made clear.

We agree this needs justification. N=50 in Eq 9 is the number of raw 50 Hz samples within each 1-second averaging interval, and the equation is the standard error of the mean of those samples. We have added this explanation and noted that this uncertainty characterizes random measurement noise specifically.

5) Introduction of Operator in 2.4 needs to be expanded a little, specifically how the particle shape and orientation is represented in the T-matrix calculations. This is needed to support the description of canting angles on line 217.

The description of operadar has been slightly expanded as suggested. A new table detailing the axis ratios and canting standard deviation of each hydrometeor specie was added.

6) L264 “the interpolation process cannot reflect exactly the real ray geometry”. I am not convinced by the hypothesis that interpolation is to blame here. It ought to be possible to represent the exact ray geometry, even if a more sophisticated approach might be needed. To explain noisier results from ground radar, this could happen if the volume sampled by ground radar were smaller than the volume sampled by PRO: there must be a finite horizontal and vertical width of the atmosphere sampled by the PRO measurement. The text needs to investigate these issues in a bit more detail.

We agree that attributing this to imperfect numerical interpolation was not well justified, and thank the reviewer for pointing toward a more physically grounded explanation. We now attribute the noisier character of $\Delta\Phi_{aramis}$ primarily to a mismatch in effective sampling volume between the two measurements. The ARAMIS network observes along discrete conical beams from a finite number of elevation angles, so the volume contributing to K_{dp} at a given point in the 3D composite has a finite horizontal and vertical extent and is built from sparse set of the beam intersections, whereas PRO integrates continously along a ray path with a comparatively narrow Fresnel-zone footprint.

7) L270: rho_HV is not defined.

Agreed. We have added a brief definition at first use.

8) The source of the reflectivity in Fig. 2 (right) needs to be stated - there are multiple simulations available.

The caption currently states the right panel shows “column-maximum reflectivity dBZ simulated by AROME, which is ambiguous since AROME itself does not output polarimetric reflectivity directly, this is a forward simulated quantity. We have revised the caption to state explicitly that the reflectivity shown is produced by operadar driven by the AROME hydrometeor fields.

9) L288 - 2980: The merged graupel configuration (and other configurations) need to be explained properly in the methods section (see previous comments). This is needed to better support discussions such as on L452-459. More generally, throughout the manuscript there needs to be more clarity over which configuration is being used in each set of results.

As noted under comment 3), we will move the justification of the merged wet graupel configuration into Section 2.3, and also stating which results use one or another configuration.

10) L292: Since section 2.5 defined the metrics RMSE, bias and chi^2, it is surprising to find the first metric presented is a median, which has not been mentioned to this point.

We have added the median explicitly to Section 2.5 alongside the other three metrics, defining it as the per-height median of (y_i-y_i^{sim}) across occultations.

Typos and editorial

On page 2 lines 43 and 45 the delta Phi symbol is erroneously repeated

Corrected.

L110 K_dp does not seen to have been introduced at this point, either scientifically or as an acronym.

Corrected.

L201 the “sigma” needs explaining.

Corrected.

References

Hubbert, J. C., Ellis, S. M., Chang, W.-Y., Rutledge, S., & Dixon, M. (2014). Modeling and Interpretation of S-Band Ice Crystal Depolarization Signatures from Data Obtained by Simultaneously Transmitting Horizontally and Vertically Polarized Fields. *Journal of Applied Meteorology and Climatology*, *53*(6), 1659–1677. https://doi.org/10.1175/JAMC-D-13-0158.1

Ryzhkov, A. V. (2007). The Impact of Beam Broadening on the Quality of Radar Polarimetric Data. *Journal of Atmospheric and Oceanic Technology*, *24*(5), 729–744. https://doi.org/10.1175/JTECH2003.1

Ryzhkov, A. V., & Zrnic, D. S. (2019). *Radar Polarimetry for Weather Observations*. Springer International Publishing. https://doi.org/10.1007/978-3-030-05093-1