This is a clear and well motivated description of experiments contributing to the DCPP. I have only minor comments, which I think may improve the text (descriptions and motivation) a little.

1. Pg 3, L25, fig. 1, It is difficult to argue that initialisation has enhanced prediction skill in years 2 and beyond, as there seems to be equal or greater areas of negative skill. I suggest a more careful formulation, supported by reference to process understanding. In particular, the skill in the NA sub polar gyre has been attributed to initialisation.
2. Pg 5, L15, I undertand that by "analysis" is used here to refer to data assimilation. This terminology may not be obvious to many readers. Also the list of contributors misses "enhancement of the observing system" which could be argued to be the most important. Furthermore, you could include statistical (flux correction, and anomaly coupling, and anomaly initialisation) methods that reduce forecast drift.
3. Pg 6, I believe that developing an understanding of the impact of initial shock on forecast skill should be mentioned under scientific aspects (perhaps under point 2).
4. Pg 17, Appendix A and respective place in main text, I can partly understand the reasoning for limiting the tier-1 hindcasts experiments to years 1-5, however, I would call them mulitannual and not near-term or decadal. Personally, I feel the greatest benefit comes from the longer 1-10 year period that focus on capturing predictability associated with the low-frequency component of climate variability rather than the interannual that is dominated by ENSO (which is not predictable beyond a year). It could be useful to make clear why shorter hindcasts are being encouraged and also called "near-term", which I understand refers to 10-30 years periods.
5. Appendix C. Is there any shock expected from applying a temperature anomaly essentially instantaneously in experiments C1.1-C1.8? If there is one, it could introduce an artefact into the results. How will it be assessed? For experiments C1.9 and C1.10, is there a reason for suggesting to start the extended pacemaker experiments exactly in 1920. The early century warming started in 1920, and it wouldn't be prudent to start the runs a little earlier if this is of interest.
6. Pg 29, I think it is important to also include salinity data (surface, upper 300m, 700m and 2000m) in the 2D Ocean data. These quantities are important for verifying the mechanisms for multi-decadal variability in the North Atlantic.

Typos

1. Pg8, L24, It should be: "to what extent can"
2. Pg 17, Table 17. A1, It should be "and start ….are recommended"
3. C1.1, I believe you mean a 50m deep mixed layer.