

The authors apply BCSD (actually SD first then BC) to ensemble seasonal forecasts from ECMWF, for five basins in four arid regions. They use ERA5-Land (hourly) as reference data but that in itself represents a down-scaled (time and space replay of ERA5 atmosphere to global land) product. They make numerous references to the importance of topography but then give it almost no attention in results.

I do not think this fits in ESSD. Nothing about ESSD handling or not handling model products. Instead a fundamentally different approach to error terms and uncertainties. A forecast has some skill realized against actual outcomes: forecast 20 mm of rain in a given future period, validated or not against measured rainfall (with spatial and measurement errors!) during that forecast period. Some weather services extract probabilities from their ensemble forecasts, e.g. 50% chance of rain or snow, combined with some publicly acknowledged uncertainty of amounts, e.g. up to 3 cm of rain or snow expected, for shorter-term forecasts. I accept that seasonal forecasts present different challenges. Here, however, authors treat the forecasts as perfect (= certain) and likewise the reanalyses as certain and then, despite having introduced substantial but unspecified additional uncertainty by downscaling to 10 km and hourly, spend their efforts trying close gaps between forecasts and higher-resolution reanalyses. Nothing wrong with their approach, but ESSD focuses explicitly and extensively on real-world uncertainties (e.g. read 'uncertainty' paragraphs in ESSD guidelines at <https://www.earth-syst-sci-data.net/10/2275/2018/>). A typical ESSD paper describes uncertainties of a measurement (e.g. PM2.5 in Christchurch) in terms of instrument errors, measurement errors, operational errors, etc. Then and only then would one attempt to calculate uncertainty of an air quality forecast. 'Uncertainty' is a different problem for ECMWF and for these authors than in most ESSD papers. That difference causes the mismatch? In review that follows I express the view that authors tend to over-sell their product but I do not doubt their motivation or their skill. I doubt that their description belongs in ESSD.

Page 1 line 19 and following: Domain numbers e.g. DO4 come from ECMWF forecasts, from DKRZ labelling, or for author convenience? Used extensively in some sections of results and figures but in other places authors seem to rely more on geographic acronyms e.g. CC-basin. Use / need both?

Page 3 line 29: "huge" another press opinion or outcome of a peer-reviewed study?

Page 3 line 30: "urgent need" expressed by who? The authors?

Page 4 lines 8 to 11: previous limitations mostly applied to 'short-term' not 'seasonal' forecasts. The authors make very high claims for this product without any evidence.

Page 4 line 14: what does 'reference' mean in this sentence?

Page 4 line 20: 5 days before the present?

Page 4 line 25 to 29: this text comes almost verbatim from the landing page for ERA5-Land. Authors should cite that?

Page 5 Table 1: Nothing about elevation or topographic complexity of basins. Earlier, authors listed elevation corrections as a necessary or desirable feature?

Page 5 line 11: bias correcting to what?

Page 5 line 15: readers will likely know forecast skill score but the term "highest" conveys nothing about skill level?

Page 6 line 6: “crucial” to understand orography but authors give only generalities (“up to 4000 m” Fig 1 not much help, only color-coded 2-D. Give us an elevation profile for stream level 1?

Page 6 line 9: no doubt, but by who’s definition? Or what reference?

Page 6 line 10: “dangerous”?

Page 6 line 12: “assumed to experience an increase in the frequency and severity” assume by who, what references. Likely true but on what basis? References that follow in this paragraph document past extreme events but largely avoid prediction?

Page 7 line 8: “these anomalies” - the remaining differences between forecast and reference data once the climatological mean reference has been subtracted?

Page 8 line 9: “fairly large number of samples for both the reference” but these represent data sparse regions?

Page 9 Model Biases: extensive discussion of how the uncorrected forecasts fail but why do we care? Useful discussion starts at line 24?

Page 9 line 29: “RMSE of SEAS5 BCSO is much lower compared to the raw forecasts.” Strong statement not supported by Figure 3. This statement from line 33 “other cases where the bias-correction shows almost no improvement” seems more accurate. For this reader, Fig 3 shows that when RMSE differences occur, they generally favor the BCSO product, while in other cases one can not distinguish RMSE terms between raw and corrected. We also need, as the authors hint but do not show, some uncertainty limits here? All precip RMSE, except for one station, lie below 2 mm/day, often below 1 mm/day. Do the authors claim such accuracy in their base numbers? One doubts. For tas, again except for 1 station, essentially all RMSE lie below 1k. The authors expect us to believe with their tools they can distinguish products at 2 mm/day and 1k? Remarkable if true but they give us no evidence. A low correlation error (RMSE) between two products of assumed ‘perfection’ but almost certainly with high inherent fundamental uncertainties seems of little relevance?

Page 10: Reader needs to jump from Fig 3 in 4.1 to Fig 6 in 4.2 then back to Fig 4 in 4.3. Reason for this hopping around? Hopping will disappear once Figures take their appropriate place in final document but then sequence will look wrong?

Page 10 Section 4.2 resolution: no uncertainties here? These are average sums of 4-month periods from 25 to 51 ensemble runs over 35 years. They must have SD, 95CI, etc? Almost every number and result across the manuscript has substantial uncertainty ranges but authors treat everything as exact?

Page 10 section 4.3 lead-time: without ranges or uncertainties, reader has no basis to accept any of these supposed differences or patterns.

Page 10 line 20: weather patterns may shift but locations do not shift, southward or any other direction

Page 11 line 3: reader needs to go from Fig 4 in previous paragraph now to Fig 7. Consider a more helpful and logical sequence??

Page 11 line 5: reader now moves from geographic codes KA or CC back to domain codes D03. Why? Confusing!

Page 11 section 4.5 overall skill: many readers will know these skill scores but will usually have seen them expressed as a range. This reader has no confidence in an absolute CRPSS of 0.4 but might accept a range from 0.3 to 0.5? Again, authors treat their results as absolute when in fact they contain substantial uncertainty!

Page 11 section 5 Discussion: helpful discussion of regional factors follows, intended apparently as justification for why corrected products seem occasionally but not consistently to outperform original forecasts. Very real regional challenges, no doubt. But if the original forecast products lacked sufficient skill when confronted by meteorological and topographic details of each basin, bias correction to higher resolution will not remove that fundamental detail-driven uncertainty? It may raise skill scores but still miss key local details. E.g. it will continue to show high fundamental uncertainty! Vis “spatial and temporal inconsistencies in the forecasted spatial extent and intensity” (Page 12 line 7) of precip, of temperature, of clouds, etc. represent the real-world uncertainty not included and certainly not overcome! The authors themselves make this point (Page 12 line 13) that for basins with skill score improvements of 0 and no differences in RMSE, fundamental uncertainty has defeated their good efforts!

Page 27 Figure 2: These are composite biases (areal sum of daily data) for source forecast vs ERA5-Land reanalysis? The colours - almost impossible to distinguish even in the label) represent different lead times from 0 to 11 months? Or are these monthly averages? Not clear. After working extensively similar Fig 3, I still find these graphics difficult to read and interpret.

At this point this reader largely ‘gave up’. The data description for DKRZ seems easy to use and very helpful. Authors have provided useful guidance to static products and how to find updates. Generally ESSD does not allow: ‘contact the author’ (Page 15 line 23). Appendices provide useful documentation on BC, on error calculations, and on skill scores. Overall the authors have provided useful information. Their approach however still seems orthogonal to the intent of ESSD.