Response to Referee 1 for GMD-2022-255

Philip G. Sansom and Jennifer L. Catto

December 2023

Referee 1

The authors present an updated version of an established automated front identification method geared toward reanalysis datasets that are available with ever higher spatial resolution, along with an objective calibration method for the algorithm to make the identified fronts comparable between datasets with different spatial resolutions. They make a convincing case that their adaptations improve the efficacy of the algorithm, especially for moderate-resolution datasets like ERA-Interim that are still widely used, as well as climate models run at comparable resolutions. Given the high degree of subjectivity involved in automated front identification methods (choices of variables, of the degree of smoothing, of parameter thresholds), any successful effort to introduce more objectivity is welcome, and the presented calibration method appears to work well for the datasets in question. The manuscript is well structured and well written, and in my opinion only requires minor adaptations. I therefore recommend it for publication pending minor revisions.

Following is a list of questions, requests and recommendations, roughly grouped into major, Figure-related and minor points.

The authors thank the reviewer for their detailed comments which are addressed individually below.

Major

1. Lines 1—11: Mention reanalysis datasets in the Abstract and quantify the improvements or changes over the old method (e.g., change in no. fronts overall and in storm track regions).

We have added mention of reanalysis datasets. The improvements are detailed in terms of the smoothness of the fronts and fewer breaks for example.

2. Lines 40—50: "This can impact the attribution of precipitation to fronts." How?

Clarified in text: "This was shown lead to large differences between datasets in the proportion of precipitation attributed to fronts."

3. Line 112: What does "joins between adjacent grid boxes" mean? Consider reformulating.

Clarified in text: "but only zero points located in adjacent grid boxes are considered for joining into lines"

4. Line 117: I understand "even moderately high resolution" as "moderately high or higher resolution" (i.e., as "low-to-high-resolution"), when what appears to be meant is the opposite, "moderately high or lower resolution" (i.e., "high-to-low-resolution"). Please reformulate to remove this unambiguity.

Clarified in text: "At or below the $0.75° \times 0.75°$ resolution of ERA-Interim"

5. Line 125: "More fronts" despite the fact that closing gaps between partial fronts decreases the overall number of front objects? Is this increase overcompensated by the fact that such joined fronts meet the minimum length criterion, which the partial fronts identified by the old approach might not? Or are there other factors at play?

A combination of increased lengths meeting the minimum length criteria, and the identification of front points and fronts missed completely by the old method, see for Figure 3 for examples. Clarified in text.

6. Line 149: Why do you choose the 25th percentile specifically, and not any

other percentile below the 50th ? Does it fit the data especially well? And how objective is this choice?

Comparison with charts and ability to produce continuous fronts. A range of combinations of thresholds K_1 and K_2 were explored and now shown in the supplementary material, but ultimately the choice is still subjective. Clarified in text.

7. Line 150: Why is the 50th percentile 0 K m-2; does this just happen to be the case by chance, or is there some theoretical reason for this symmetry around zero?

The TFP is essentially the second derivative of θ_W, so intuitively it makes sense that over a large area the average should be approximately zero, but over smaller areas or time scales, it need not be.

8. Line 153: What do you mean by "simple symmetry"?

Each is approximately the 50th percentile of the allowed range, since we require $K_1 < 0$. Clarified in text.

9. Line 164: Are the climatologies required to determine the quantiles? Please state this.

Clarified at the beginning of paragraph 3 of Section 3.2.

10. Lines 169—170: Which month did you end up using? Does it make a difference which month/season? How did you reach this conclusion, i.e., what did you test?

Various lengths of training data and spatial extents were tested from one month to 30 years. In practice we used January 2000 for the Norther Hemisphere extra-tropics. Clarified in text.

11. Line 174: Do I correctly understand that "object" refers to the front and for a zonally oriented front, you use the grid box to the left and that to the right of it? What happens for diagonally oriented fronts; do you use the top-left and the bottom-right neighboring boxes in that case? This could be formulated a bit more clearly.

Object refers to a grid point, Equations 1–4 are evaluated at each grid point, then interpolated to front points. The zonal and meridional components of the derivatives are computed separately using one grid box to the left and right, or above and below respectively. Clarified in text.

12. Lines 173/175—176: What is the difference between the "simple central finite difference" and "explicit central finite difference" approximations? A short explanation could be helpful (unless these terms are obvious to people more versed in numerical discretization than I am).

In the original code, approximations to the second and third derivatives required in Equations 1-4 were obtained by repeated applications of the "simple" finite difference approximation to the first derivative. We use an explicit finite difference approximation to the second derivative in order to evaluate ∇^2, avoiding the need for repeated applications of the "simple" approximation which degrade accuracy. Clarified in text.

13. Line 180: Add reference for NCL.

Done.

14. Line 196: Briefly mention Figure 4d to complete the discussion of the figure, e.g., "Figure 4d shows the fronts in ERA5 and will be discussed in Section X."

Figure 3d, Done.

15. Lines 256—257: Why is the performance improved over earlier versions? Does the new version maybe use optimized third-party libraries over handwritten code? Please add a short explanation.

Clarified in text: "Performance improvements come primarily from three areas: (i) the improved efficiency of the contouring algorithm compared to the line joining algorithm; (ii) vectorization of many calculations to avoid unnecessary loops; (iii) reduced memory usage by avoiding pre-allocating unnecessarily large arrays."

16. Line 258: I very much appreciate the performance numbers, but please replace "a single core of a 2 year old laptop" by a more meaningful hardware

description (processor type, number of cores etc.).

Clarified in text: "a single core of an Intel i7-8565U based laptop with a theoretical maximum speed of 4.6 GHz"

17. Line 268: Why is contouring cheaper than line joining? If possible, add a very brief explanation.

This was explained in Section 3.1, where it has been expanded.

18. Line 279: Consider replacing "performance" by a synonym when describing the efficacy or accuracy of the identification algorithm, as "performance" often refers to the compute performance (which, in this case, would likely rather be decreased than increased by moving to higher-order accuracy).

Clarified in text: "modest increases in both the number of fronts and front points identified"

19. Line 286: What kind of performance benefit do you refer to, compute performance or the performance (accuracy) of the algorithm?

This comment has been removed in response to comments from Reviewer 2. A more complete comparison would be enlightening, but is beyond the scope of the current study.

Figures

20. Figure 1, panels: Consider adding labels above the plots, e.g., "(a) Mask-then-join" and "(b) Contour-then-mask".

Done.

21. Figure 1, caption: Merge first two sentences; consider removing "Thick" before "black" (there are no thin contours other than country borders).

Done.

22. Figure 1, caption: Consider changing "black lines are contours of the"

to the, e.g., "black contours show" (shorter; likewise in other captions).

Done.

23. Figure 1, caption: Replace "Points" by "Circles".

Done.

24. Figure 2, plots: Add a legend for the line colors.

Done.

25. Figure 2, panels: Consider adding labels above the plots, e.g., "(a) Zonal-mean TFP" and "(b) Zonal-mean $|\nabla \theta_W|$".

Done.

26. Figure 2, plots: The yellowish lines are close to invisible on my (full-color) printout, and most likely hard to see on many projectors; consider adapting the colors.

Done.

27. Figure 2, caption: Add the threshold values.

Done.

28. Figure 3, panels: Add labels above the panels, e.g., "(a) Mask-then-join", "(b) Contour-then-mask (n=2)", "(c)Contour-then-mask (n=8)", "(d) ERA5".

Done.

29. Figure 3, caption: Mention that (a—c) are based on ERA-Interim data.

Done.

30. Figure 3, caption: Add the date and time of the snapshot.

Done.

31. Figures 4—10, plots: Use a different colormap for frequency plots; on the on hand, one comprised of multiple colors (not just shades of red) would make regional differences stand out more (especially in Figure 7 and 10, where all panels look the same at first glance) and increase the visual appeal of the maps; on the other hand, using the same shades of red for absolute frequencies and positive differences makes it hard to distinguish these two types of plots.

Done.

32. Figure 4, panels: Add labels, e.g., "(a) Dowdy and Catto (2017)", "(b) Mask-then-join", "(c) Contour-then-mask", "(c) Difference (b) – (a)", "(d) Difference (c) – (a)", "(f) Difference (c) – (b)".

Done.

33. Figure 5, plot: Fix overlapping colorbar labels (here and throughout the manuscript).

Done.

34. Figure 6, panels: Add labels, e.g., "(a) Cold fronts" etc.

Done.

35. Figure 6, panels: Reorder the panels such that the total front frequency comes first (top-left), which is more intuitive, especially given (d) is discussed first in the text before (a—c); the panels of a figure should ideally be discussed in order in the text, which reordering would achieve.

Done.

36. Figure 7, panels: Add labels above plots, e.g., "(a) Cold fronts in DJF" etc.

Done.

37. Figure 8, panels: Add labels above plots, e.g., "(a) ERA5", "(b) Difference ERA5 - ERA- Interim".

Done.

38. Figure 9, panels: Add labels and reorder panels as in Figure 6.

Done.

39. Figure 10, panels: Add labels above plots, e.g., "(a) Cold fronts in DJF" etc.

Done.

Minor

40. Line 2: Consider adding "weather prediction" after "operational".

Done.

41. Lines 4—5: Sentence is a bit hard to read as the two approaches ("applying a mask then joining frontal points" and "contouring the thermal field then applying the mask") don't stand out very well; could be improved by, e.g., adding "(i)" and "(ii)" or so before "applying" and "contouring", respectively.

This sentence has been reformulated in response to comments from Reviewer 2.

42. Line 6: Add "for" after "allows".

Done.

43. Line 7: Consider replacing "have made" by "present" or similar.

Done.

44. Line 23: Remove "on" after "impact".

Done.

45. Line 25: Remove "prior to then" (he would have been hard pressed to summarized future methods).

Done.

46. Lines 26—29: Revise "Tomas [...] thresholds": Several incomplete sentences that could be merged into one (e.g., "[...] parameter: first, [...] 850 hPa; second, [...] derivative; and [...]").

Done.

47. Line 31: Remove "a thermal variable" (the whole paragraph is about approaches based on thermal variables) and remove parentheses around "equivalent potential temperature".

Done.

48. Line 31: Replace "the second derivative of that variable" by "its second derivative".

Done.

49. Line 33: Replace "which" by "who".

Done.

50. Line 34: Remove "and" before "placing" or replace the latter by "placed".

Done.

51. Line 40: Is "benefiting" really the right word here?

Replaced by "tailored to suit".

52. Line 54: Move "Hewson" out of parentheses (probably used "citep" instead of "citet" in LaTeX source code).

Done.

53. Line 55: Consider streamlining "We demonstrate a method that can be used [...]".

Done.

54. Lines 61—62: Add "the" before "European and consider moving "reanalysis" after "ERA- Interim"; define "ERA-Interim"; consider putting the definitions of ECMWF and ERA-Interim in parentheses instead of the acronyms (easier to read than vice versa).

Done, but elsewhere since first references to these have been made earlier.

55. Line 64: Remove "The wet bulb potential temperature θ_W is computed" and join the sentence with the previous one ("[...] identify fronts, using the [...]").

Done.

56. Line 67: Define ERA5 and move reference here from line 71.

Done.

57. Line 68: Add "of," "among" or similar before "standard" (something is missing there).

Added "among".

58. Line 71: Reference ERA5 at first occurrence on line 67.

Done.

59. Line 77: Are both equations necessary, i.e., is it not obvious that $\nabla \times \nabla$ is the same as ∇^2?

∇^2 is sometimes also used to indicate the Hessian matrix, so we have retained both equations for clarity, and compatibility with Hewson [1998].

60. Line 79: As far as I can tell, "Thermal Front Locator" should not be written in uppercase.

Changed to lower case.

61. Line 83: Add comma before "where".

Done.

10

62. Lines 91—93: Something appears to be missing from this sentence (around "K_2 a fraction of a grid length").

Clarified - "at a point a fraction of grid length"

63. Line 100: Consider replacing "1, 2, 3 and 4" by "1—4".

Done.

64. Line 101: Consider adding after "improvement" something along the lines of "described in the next section".

Done - "described in Section 3.1"

65. Line 110: What does "simultaneously" refer to? If it means that zero points are identified in parallel, then reformulate accordingly. (Grammatically, it could also refer to identifying points and joining them into lines in parallel, but that doesn't make sense as far as I can tell.)

Removed. In the algorithm used by Berry et al. [2011] locating zero points and joining them into lines were two separate steps, which in our implementation are done by a single function call. The more important performance distinction has been clarified in the next sentence.

66. Line 135: Replace "smooth [features]" by "continuous" (provided I understand the sentence correctly).

Done.

67. Line 136: Consider joining the sentences, e.g., "[. . .] Equation 1, which will appear [. . .]".

Done.

68. Lines 137—138: Consider replacing "the contours of TFL = 0 K m-3" by "these contours".

Done.

69. Line 139: Consider adding "filter" after "average" and "data" or "fields"

after "ERA-Interim".

Added "data" after ERA-Interim.

70. Line 146: What does "gradient" refer to?

Equation 3, clarified in text.

71. Line 146: Consider adding "the" before "gradient".

Done.

72. Line 147: Replace "very [different]" by, e.g., "more variable and different" ("variable" in contrast to "constant" in the extratropics, and "different" because the values differ).

This paragraph has been reformulated in response to comments from Reviewer 2.

73. Line 157: If "previously" refers to Berry et al., replace it by "their study", "Berry et al." or the full reference.

Added full reference.

74. Line 160: Consider replacing "the datasets of interest" by "them".

Done.

75. Lines 163—164: Replace "at [any reasonable]" by "for" ("at" doesn't fit to "number of smoothing passes") and "or [number]" by "and".

Done.

76. Line 165: Replace "so [that]" by "such".

Done.

77. Line 166: I think "magnitude" should be plural.

Don't believe so, rephrased.

78. Line 168: By "can be compared [...] only for", do you mean that it is not possible to compare the percentile for a larger subset of the data (let alone the whole datasets), or do you mean that it is sufficient to only use a small subset of each dataset to meaningfully compare them? "Can be [...] only" reads like the former (i.e., a limitation), but judging from "avoiding the need [...]", I think you mean the latter (i.e., an opportunity). Please reformulate to clarify this.

The latter, rephrased for clarity.

79. Line 168: Add "that of" before "the gradient field".

Done.

80. Line 172: Consider swapping "single" and "biggest".

This sentence has been removed.

81. Line 174: Add "on" before "either side".

Done.

82. Line 184: Replace "criteria" by "criterion".

Done.

83. Line 190: Consider replacing "smoothing passes" by "n" for consistency with "K_1" and "K_2" (and add a comma after "i.e."); add "UTC" after "00:00" (or after the date).

Done.

84. Lines 198—199: Join sentences with comma after "i.e., [add comma] the [...] K_2)" or make the second sentence ("Specifically, [...]") complete by adding a verb.

Joined sentences.

85. Lines 215—216: Consider shortening the title by removing "identified" and maybe even "data"; alternatively, replace "Fronts identified" by

13

"front climatology" to distinguish the section from those before, which already discussed fronts in ERA-Interim but focused on different identification approaches rather than on the climatologies.

Shortened to "Front climatology from ERA-Interim"

86. Line 217: Add "for" after "allows".

Done.

87. Line 225: Remove "in [in]".

Done.

88. Lines 225—229: Put all figure references in parentheses instead of writing "in Figure X".

Done.

89. Line 230: Same as line 215.

Shortened to "Front climatology from ERA5"

90. Line 233: Add "UTC" after "00:00" (or after the date).

Done.

91. Lines 234—235: Replace "[calibrated] in Figure 3(c)" by "(Figure 3c)".

Done.

92. Line 235: Add "that in" or similar before "ERA-Interim" (the frequency of fronts in ERA5 is not compared with ERA-Interim itself, but with the frequency of fronts in ERA-Interim).

Done.

93. Line 236: What are the thresholds identical to?

Clarified - "to those used for ERA-Interim"

94. Line 240: Replace "[fronts] seen where there fronts [were]" by "where they".

Done.

95. Line 242: Replace "[at] it's [native]" by "its".

Done.

96. Line 243: Add "Figure" before "[in] 8".

Done.

97. Line 249: Consider replacing "associated" with a more suitable word (it feels a bit off to me).

Replaced "associated with" with "due to"

98. Lines 262—263: Reformulate to avoid using "contour-then-mask approach" twice in close succession.

Reformulated.

99. Line 269: Reformulate to avoid using "involve" and "involving" in close succession.

Replaced "involve" with "require"

100. Line 270: I understand the usage of "human" to highlight the subjectivity introduced by a meteorologist's expert judgement, but since I am not aware of the existence of nonhuman meteorologists (weather-sensitive animals aside), please consider reformulating this or simply removing "human".

Removed "human".

101. Line 277: "Alternative choices" of what?

Meteorological field, clarified in text.

102. Line 277: Replace "As well as" by, e.g., "In addition to".

Done.

103. Line 280: "Implied" is probably a typo and should be "increased" or similar.

Replaced "implied" with "increased".

104. Line 285: Consider replacing "6, 7 & 8" by "6—8" (or at least "6, 7, 8") and "Equations 6—8 of that paper" by "their Equations 6—8".

Removed as part of changes in response to comments by Reviewer 2.

105. Line 288: Add some separator before "while", e.g., "Furthermore, [while]", to gain some separation from the previous, unrelated sentence.

Done.

References

Gareth Berry, Michael J. Reeder, and Christian Jakob. A global climatology of atmospheric fronts. *Geophysical Research Letters*, 38(4):1–5, 2011. doi: 10.1029/2010GL046451.

T. D. Hewson. Objective fronts. *Meteorological Applications*, 5(1):37–65, 1998. ISSN 13504827. doi: 10.1017/S1350482798000553.