

**Review of manuscript essd-2018-62 “Spatial and seasonal patterns of near-surface humidity in the foothills of the Canadian Rocky Mountains, 2005-2010” by Wendy H. Wood, Shawn J. Marshall, and Shannon E. Fargey.**

This paper presents a novel data set of humidity measurements from a network of near-surface sensors spanning the foothills of the Canadian Rocky Mountains near the city of Calgary. It is a new and significant data set that will be of use to researchers from a broad range of disciplines. Some interesting aspects of the data set are presented which provide some example of the utility of the data set.

While the data set is very much within the domain of ESSD and this special issue, I do have some major concerns that would need to be addressed before the resulting manuscript and accompanying data set is suitable for publication.

**Major comments**

It appears that an incorrect formula for calculating vapour pressure from RH may have been used. As most RH sensors report RH with respect to water, this equation should be used to calculate vapour pressure and specific humidity from the measurements. The alternative equation for the saturation vapour pressure of ice can be used to rescale RH to be with respect to ice, but this should only be performed after the calculation of vapour pressure. If the sensor used in this study reports RH with respect to water, then the specific humidity data presented here will be underestimated at low temperatures (i.e. at higher elevations and in the winter). This effect is significant - the saturation vapour pressure of ice is lower by 5% @-5C, 8.5% @-10C, 14% at -15C, 17.5% at -20C. This needs to be rectified and may affect the results presented (e.g. lapse rates of specific humidity).

There are many ambiguities in the methods used to quality control and derive the daily averages and gap fill the data. For example, it is unclear which daily averages are used in the gap filling and in the summary results presented. These prevent a confident use of the data set.

The text needs substantial work to remove ambiguities, avoid repetition or redundant text and provide more discussion of the potential uses of the data.

To clarify the method used to construct the data set, the manuscript would be better restructured

- 1 – Introduction
- 2 – Study area, sensor network and instrumentation (including sensor accuracy and quality control)
- 3 – Calculation of daily mean values (methods, results and discussion)
- 4 – Gap filling (method, results and discussion)
- 5 – Seasonal and spatial humidity variations and applications
- 6 – Data availability
- 6 – Summary

Also, the data set is currently not openly available on the Pangea webservice, even with a login, so I was unable to check the resulting data set.

**Line comments (page-line)**

Title: this should reflect the presentation of a data set, not the results interpreted from these. E.g “Near-surface humidity from a mesonet in the foothills of the Canadian Rocky Mountains, 2005-2010”

1-9: "2" should be in superscript here and 1-10.

1-20: Please add references to support these statements.

1-21: It would be worthwhile defining what you mean by 'humidity' here, as the term is used throughout and its current use is ambiguous.

1-21: Add "in the atmosphere" after "amount of water vapour"

1-22: Please explain the term vapour pressure as you do for other variables. e.g. "The partial pressure of water vapour in air"

1-22: Remove 'the' before mixing ratio and absolute humidity

2-1: RH is defined with respect to the saturation vapour pressure of pure water, so differs from the actual maximum water content - see comment on supersaturation below. Please revise. Also, please introduce the difference between RH with respect to ice and water here. See [http://glossary.ametsoc.org/wiki/Relative\\_humidity](http://glossary.ametsoc.org/wiki/Relative_humidity)

2-1: "Once the air is saturated (RH = 100%), any additional vapour will condense out." This is not strictly correct, as in certain circumstances air can become super-saturated. Please revise. See <http://glossary.ametsoc.org/wiki/Supersaturation>

2-7: "In operational..." this sentence would be better place alongside the discussion of the spatial variations in the dry line a few paragraphs down.

2-11: remove comma between "data" and "for"

2-11: change 'in' to 'of'

2-14: This paragraph should be part of the previous paragraph as it is not about human comfort, but how humidity variables are used.

2-24: "water vapour in the air". Water vapour in the atmosphere would be a more appropriate term here.

2-25: Please explain the term "rainout" or replace with a description

2-26: "humidity" do you mean RH or water vapour? please be consistent.

2-31: "while RH does the opposite" you don't explain why RH does the opposite. Please explain.

3-2: Please use "dry line" and not "dryline" throughout, as this is how you introduce the term.

3-6: "Other systematic spatial humidity structures can be expected as a function of elevation, aspect, surface type, and air mass/frontal interactions." please give more context for why we would expect these?

3-11: "humidity" do you mean RH or specific humidity?

3-11: Please reword to "...2010. The dataset of temperature is described..."

3-15: Please use a full stop between the description of each section.

3-21: Please change section title to "Study area, sensor network and instrumentation"

4-7: "Temperature and relative humidity..." This sentence is confusing - do you mean " Instantaneous measurements of T and RH were recorded every hour using SP-2000..." please revise.

4-15: "reliable data". please point readers to the quality control section below.

4-17: "pole extensions were added", please clarify if they were added and removed each season or if the sensors were just permanently mounted higher.

4-25: please change section title to "Calculation of vapour pressure and specific humidity and daily means" or similar

4-27: "actual humidity", please avoid using this term as you don't define it and is ambiguous. Do you mean the "amount of water vapour in the atmosphere"?

4-27: "We take different approaches to this calculation." please explain what you mean here - i.e. you take two different approaches compare them.

4-29: please explain why the pressure dependency of the saturation vapour pressure was omitted?

5-1: does the instrument report RH with respect to ice? Most instruments report RH with respect to water, so that vapour pressure should be calculated using 1a. 1b can be used to transform the sensor RH to RH with respect to ice after the calculation of vapour pressure. Please address this.

5-5: this is the equation for mixing ratio. Please explain that it is used here to approximate specific humidity.

5-17: please describe the approximate error you expect from using equation 3 vs actual pressure measurements.

5-25: please avoid using "actual humidity"- see earlier comment

5-26: "For applications..." this sentence is very confusing. Please reorder this section by, 1. defining the equations used to calculate vapour pressure and specific humidity 2. defining (and preferably naming) the two methods used to calculate daily means of each variable 3. describing which will be most appropriate for different applications.

6-3: please briefly explain how instruments were deployed and calibrated i.e. were different instruments used at each site, how often were instruments changed, were instruments pre and post calibrated? Were any adjustments to the raw data made?

6-12: "RH difference". do you mean the mean bias difference? or mean absolute difference? over what time period (i.e. hourly or daily). Please clarify.

6-12: "20%" do you mean 20% RH or 20% of the mean RH value (i.e. 50% RH +/- 10% RH)?

6-15: "read lower" by how much, and what do you think the cause of this is?

6-16: "biases of several %" your method would indicate that individual bias can be as large as 20%- please revise.

6-17: "although this may depend on the season of the sensor validation tests" please expand on what you mean by this?

6-17: "Uncertainties in RH are estimated" is this based on the standard deviation of the differences between WRS and Veriteq instruments?

6-20: "we estimate an accuracy of  $\pm 7\%$  in mean daily ev and qv." but this is assuming that the WRS RH is a true representation of the RH and the actual uncertainty is higher. Please revise.

6-24: "relative humidity" please be consistent in the use of either "RH" or "relative humidity" throughout the manuscript.

6-31: "exclude days" one presumes this is indicative of sensor malfunction, but please state the reason here.

7-10: change to "control steps result in"

7-15: "89% of site-days with valid data," do you mean that 89% of days have complete data (i.e. 24 valid measurements)? Please clarify.

8-1: change "is" to "was".

8-22: section "4.3.1 Kriging" is not needed as the results are not presented.

8-32: you do not present the kriging results so this sentence need revised.

8-33: please give the range of parameter values you tested for the IDW.

9-3: "<ean daily....saturates at 100%" This paragraph should be contained in section 3.

9-7: "all site-days" is this just days with good data or does it also include the gap filled data?

9-9: please move the definition of delta q to the start of the paragraph.

9-11: please change to "than that during"

9-13: "average daily specific humidity calculated from method (ii)" please use the terms you define and refer to qvh

9-13: "generally positive" the figure only indicates this is true for temperatures greater than -5C. please revise.

9-30: "jackknife" please the term you introduce earlier "leave-one-out"

9-31: "tested difference interpolation methods" please make a comment on how the kriging results compared or exclude these from your methods.

9-32: "qv" is this qvd or qvh? you discuss the differences and the need for thermodynamic constrain on one but not the other, so it is an important distinction.

10-7: "Relative (percentage) error" please clarify you mean in qv, not RH here

10-8: it would be worth adding another line to table 1 showing the statistics for qv as a percentage

10-9: "Interpolation errors in RH" what statistic is used here?

10-12: "7% of the mean annual values" why don't you list the average percentage error listed above?

10-13: ", respectively" do you mean in summer and winter or for qv and RH?

10-16: "actual humidity" please avoid this ambiguous term

10-17: please add "from the leave-one-out validation" after "RH values".

10-19: please replace both instances of "values" with "distributions"

10-19: "RH values differ" is this due to the 100% limit imposed during the gap filling or due to the differences in thermodynamic constraint in daily vs hourly data? Please discuss.

10-19: "These differences" please describe the specific differences in the figure you are referring to.

10-20: "This is the approach" it is still not clear how data were gap-filled. Are days with missing station data filled with daily mean qvh or qvd? Is qv this is transformed to ev and further transformed into RH using thermodynamic constraints. If qvh is used does this mean that the filled values in the qvh dataset will be biased with respect to the actual data values (keeping in mind figure 2)?

10-21: "qv" again, you need to clarify if this is qvh or qvd

10-24: what ev and qv are used in Table 2 - qvh or qvd? Please clarify in the text and the table caption

11-2: please add "(Table 2)" after "in the summer"

11-7: "other three-seasons" there is a lot of spatial structure in the mountains during most seasons that it would be worth mentioning.

11-11: "running eastwards along FCA lines 3 and 4, where line 1 is the northernmost line (see Figure 1)." This description is rather awkward - it would be much clearer if the figures could be annotated with the bow valley and the approximate transition from the mountains to prairie sites.

11-17: "is a significant relationship from November through January" the R2 for this relationship must be very low - please list the R2 for each month and the annual value in the table and consider revising this statement.

11-29: "8.5% km<sup>-1</sup>" is this the average of the two months? why not just list each month? It gets confusing when different time periods are introduced in this section.

11-31: "both specific and relative humidity" from what I can tell, the relationship with elevation for RH must only explain a small amount of the relationship (less than 40%), so this result is tenuous. Please revise.

11-33: "Figure 5b also support this" please define what you mean by 'this'. The increase in qv with longitude can be explained by increasing temperature, so doesn't necessarily support decreased RH at higher elevations. The mechanisms behind the observed differences between April-May and July-August would be worth investigating further but need further analysis that is beyond the scope of ESSD. Please revise this section.

12-4: This paragraph (and figure 6) would be better placed at the start of the section, to give context for spatial and elevation patterns presented.

12-11: it would be more intuitive for the reader if the order of these examples aligned with the order in figure 7.

12-24: “high humidity levels” but this is just for RH, indicating that temperature is warmer but that the air mass has a similar water vapour content. Please revise.

12-25: “Daily mean ...” this sentence seems very out of context is not needed.

12-30: “inversion” you don't explain this term or present data to show elevation gradients, so this sentence is confusing. Please remove the term and revise.

13-4: “common structure” common with what? Please revise.

13-26: “estimated (interpolated)” it is still unclear if the interpolated data are based on qvh or qvd?

14-2: “complicated” “complex” is a more common term to describe terrain – consider changing.

15-2: “15 sites” this differs from the 18 sites mentioned in the methods. Please revise.

15-6: “High summer humidity ... particularly in the spring and autumn.” These sentences are more speculative and should be more removed from the summary.

15-10: “also creates a general increase in RH with altitude, with an average gradient of +3% km<sup>-1</sup>.” but this result explains only a small fraction of the variation, so cannot be considered robust result. please revise.

15-13: This last paragraph belongs in the discussion section.

Table 1: please explain how the standard error was derived in section 4.

Table 2: please place the standard deviations in parentheses as the use of +/- is not appropriate.

Table 2: R<sup>2</sup> > 0.4 seems like a very low benchmark. It would be better to list the R<sup>2</sup> values in parentheses after the lapse rates to indicate the strength of the relationship in each month.

Figure 1: please explain what the different symbols (dots or crosses, grey and black) mean?

Figure 2: what is the red line in these plots? if it is a best fit line, please describe in the text and provide some statistics.

Figure 2: what is the justification for the single vertical line at 10C? either add regularly spaced vertical grids or remove

Figure 2: this figure would be better suited to a kernel density plot or to having smaller circles to allow the reader to see how frequently the points overlap. it is currently hard to tell.

Figure 2: The caption needs reworded to “Relation between  $\Delta q_v$  and (a, c) temperature and (b, d) daily temperature range.  $\Delta q_v$  is the difference in specific humidity calculated from daily mean T and RH (qvd) and from hourly data (qvh):  $\Delta q_v = q_{vd} - q_{vh}$ . The top panels are actual differences, in g kg<sup>-1</sup>, and the lower panels are normalized by mean daily  $q_{vx}$ , expressed as the % difference.” Also the caption needs to clarify if they are normalised by qvh or qvd

Figure 3: please clarify the “interpolated” data interpolated using the leave-one-out method. Also, the last sentence - “Specific humidity distributions are statistically equivalent, but the RH distributions differ” - should be in the main body text, not the caption.

Figure 4: the order of these panels is confusing. Please place in consecutive order, perhaps starting with winter as this is the first season discussed in the text.

Also, the colour scale is confusing and would ideally be consistent between each panel. Consider using a perceptually uniform colour palette such as 'parula' in MATLAB or 'viridis' in Python

Figure 6: It would be much more useful to plot the seasonal mean (as in figure 5) as well as plot the saturation vapour pressure and specific humidity curves on panel a to give context for the results in figure 5

Figure 7: it would be much better for the labels in each panel to describe the conditions rather than the date. e.g. 'summer cyclonic', 'summer convective', 'winter polar airmass', 'winter chinook',

Also, the colour scale is confusing and would ideally be consistent between each panel. Consider using a perceptually uniform colour palette such as 'parula' in MATLAB or 'viridis' in Python.