The authors present a well-written and insightful paper that proposes to use a clastic varve records covering over 1,500 yrs to track past hydroclimates in northern America. By correlating the record of varve thickness to regional precipitation patterns as well as other regional records, the variations observed are attributed to changes in winter snowfall. These interpretations are then placed in a wider context to determine the role of large-scale synotic climate, such as the North-Atlantic Oscillation and Greendand Blockage on winter precipitation. The 1,500 yr long varved record is therefore suggested to provide insights into past NAO/GB variability.
While this is undoubtly a valuable record and a publication that falls within the scope of Climate of the Past, I would like to raise a few issues regarding the predicting power of the record to track winter precipitation and NAO. I do not refute the conclusions *per se*, but I question the robustness of the statistical methods on which the argumentation relies and try to propose additionnal ways to explore the data.

Thank you very much for this general comment.

Varves:

- I know this is standard in varve research, but there is no mention that the sediment blocks were embedded in resin. Since this paper is adressing a larger audience, I would recommend to add a line on that. Also are the SEM pictures taken on the thin sections or sediment blocks?

  Yes, the thin sections were made using epoxy resin. Thanks for your point, we will add clarification about it.

- I don't doubt that these laminations are real varves (indeed a textbook example of clastic varves!) but I was wondering whether you had any hint of years with more than two sublayers, i.e., with more than one discharge event?

Yes, additional layers are very clear in the cores proximal to the two main tributaries. This was reported in our previous paper (Gagnon-Poiré *et al*. 2021). In this distal site, there are a few (3-4) intercalated coarser layers, but their origin cannot be attributed, for sure, to additional discharge events.

Data analysis:

- Determination of drivers of sediment input (VT). Correlation of precipitation at Goose A and sediment accumulation (VT) (fig 5): Please provide r^2 instead of (or in addition to) r. The r value shows that the variables are positively and significantly correlated (aka. yes, total precipitation/snowfall influences the amount of sediment deposited at the site). The r^2 value shows how much of the variance in the dataset can be explained by the driver considered – since you are using this relationship to make predictions about winter precipitation and the NAO, it is not trivial. In your case, an $r^2$of ca. 0.1 (which is more or less what you must have here) means that most of the variance in the dataset (90%) is explained by something else than total precipitation or snow precipitation. Did you investigate other drivers? I am wondering whether running a multiple linear regression with adjusted r^2 would help to constrain the role of e.g., rain+snow+temperature on sediment input. In terms of processes, I can well imagine that a combination between how fast the temperature rises in the spring and how much snow has been deposited in the winter might exert a strong control on discharge and flow strength.

That's an excellent point. We will include R^2 values in the figures alongside the correlation coefficients (R values). Our multiple linear regression analysis indicates stronger correlations and more significant p-values for the period 1972–2017. The reason for this is unclear, but it may be related to a change in meteorological instrumentation following the 1970s. We attempted to obtain more information about the meteorological station at Goose Bay but have not yet been successful. In any case, the strongest correlation observed is with snow deposition.

From 2017-1943:

```
Call:
lm(formula = VT ~ SNOW_Nov_May + temp + raingoosebay, data = df_scaled_subset)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.15641 -0.66563 -0.03165  0.42732  2.65285 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)   
(Intercept)  -3.057e-17  1.093e-01   0.000  1.00000   
SNOW_Nov_May  3.258e-01  1.129e-01   2.887  0.00515 **
temp          1.636e-01  1.157e-01   1.414  0.16164   
raingoosebay  8.733e-02  1.154e-01   0.757  0.45171   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.947 on 71 degrees of freedom
Multiple R-squared:  0.1396,	Adjusted R-squared:  0.1032 
F-statistic: 3.839 on 3 and 71 DF,  p-value: 0.01318
```

From 2017-1972:

```
Call:
lm(formula = VT ~ SNOW_Nov_May + temp + raingoosebay, data = df_scaled_subset)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.4210 -0.7043 -0.1026  0.4959  2.2126 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)   
(Intercept)  4.393e-17  1.308e-01   0.000  1.00000   
SNOW_Nov_May 4.017e-01  1.337e-01   3.004  0.00443 **
temp         1.199e-01  1.439e-01   0.833  0.40931   
raingoosebay 2.288e-01  1.433e-01   1.596  0.11780   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8965 on 43 degrees of freedom
Multiple R-squared:  0.2486,	Adjusted R-squared:  0.1962 
F-statistic: 4.743 on 3 and 43 DF,  p-value: 0.006044
```

- Same comment for Figure 9c: use r2 instead of r if you want to show that GB is a good predictor for winter precipitation. By the way: how is the GBI calculated?

OK, we'll do. We will add a sentence that describes the GBI along with the reference from Hanna et al. 2016. Thanks.

- Minor point: in Fig 5, did I understand correctly that you compare log(VT) to log(precip)? That would make sense since both datasets are constrained (R+, only positive).

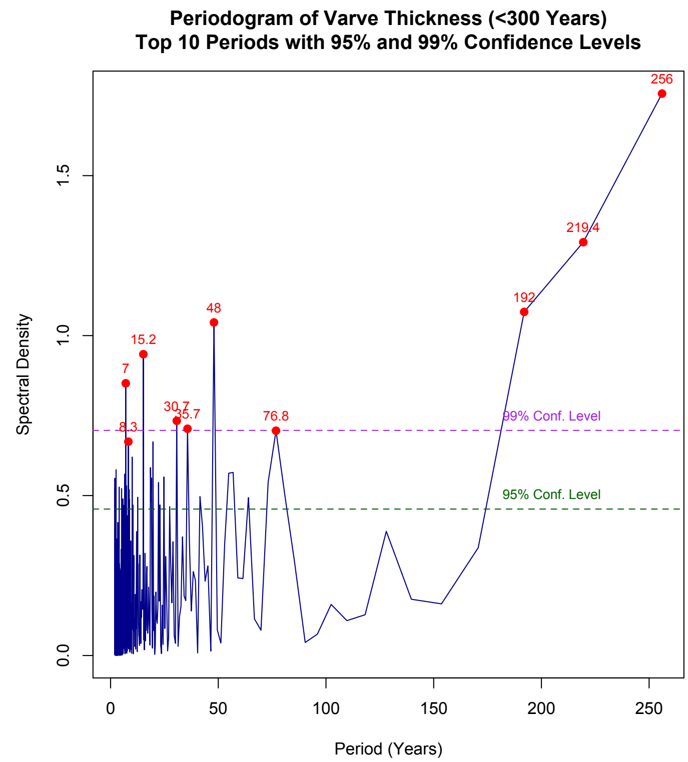
The data are only normalized (relative to the mean and the standard deviation).

- p values need to be reported and discussed (see for instance the recent paper of *Benjamin et al., Nature Human Behaviour 2018*)

OK, we will discuss p-values further.

- time-series analyses: the data have been analysed using a wavelet analysis (Fig. S2) (using which software?). I am a bit surprised that the analysis (which is unique as being annually-resolved) shows little high-frequency variability, esp. in the range of those originating from the NAO (7-yr if I am right?), if this is supposed to be an important driver. Do the authors have an explanation for this? Did they also perform a periodogram (e.g., multi-tapper)? Did you run time-series analyses on the xrf records?

There is actually a period of significant variability centered around 7 years (>99% conf. level: see periodogram below), which is also the case when the data are log transformed (lower panel). This is also observed in a new wavelet analysis that is better detailed (below).
No, we haven't run time series analyses on the XRF data, which is not the scope of this paper.



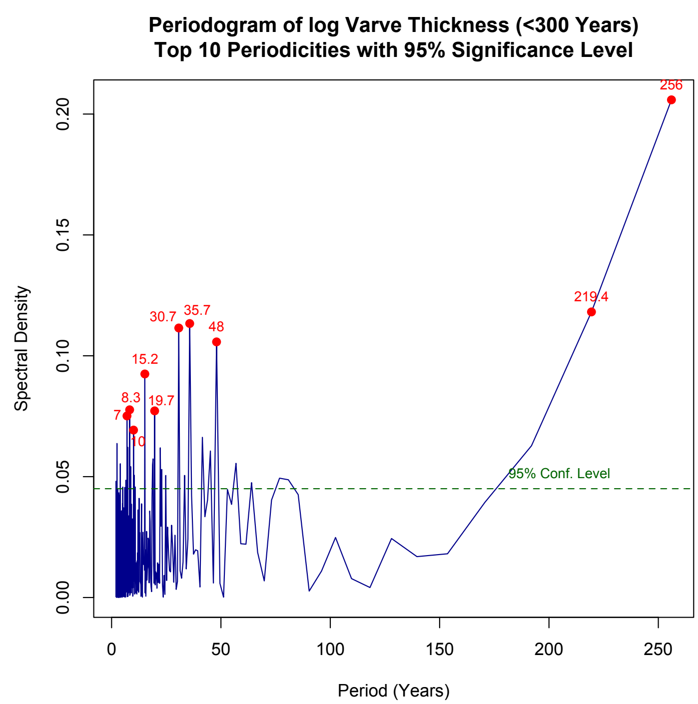
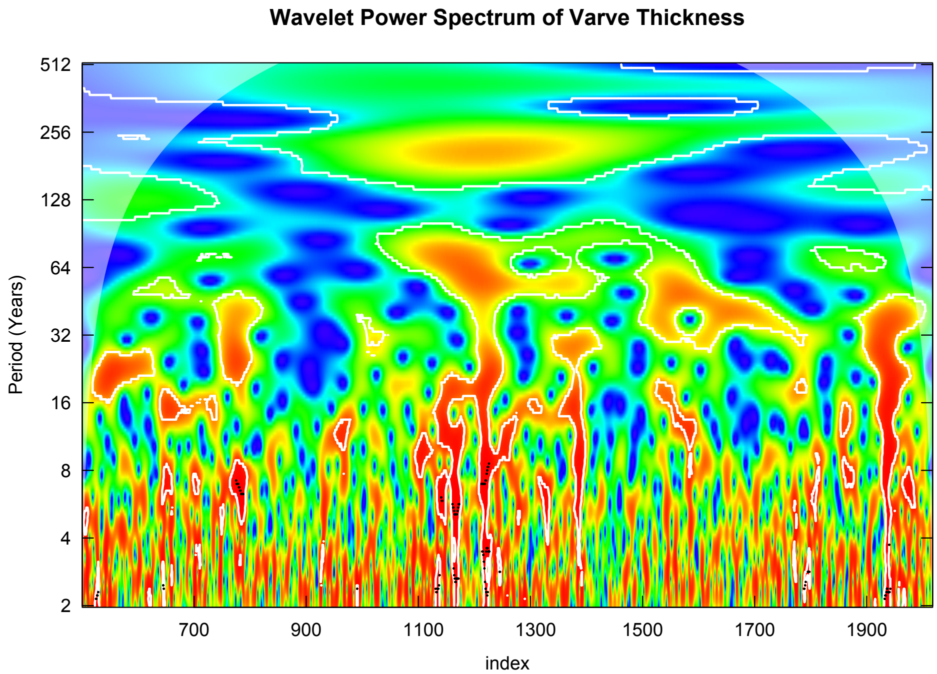


Fig. Upper panel: Periodogram of the varve thickness (VT) record from Grand Lake. The periodogram displays the distribution of spectral power across different frequencies, revealing dominant periodic components in the time series. To assess the statistical significance of spectral peaks, the results are compared against a null hypothesis that the time series follows a red-noise (AR(1)) process, which represents a baseline of natural variability with no true periodicity. Peaks that exceed the 95% confidence threshold—determined from the theoretical red-noise spectrum—are considered statistically significant, indicating the presence of persistent periodic signals unlikely to result from random fluctuations alone. Lower panel : same as the upper panel, but analysis of the log varve thickness. Many spectral peaks exceed the 95% confidence level (including at a 7-year cycle), suggesting that they reflect periodic signals.



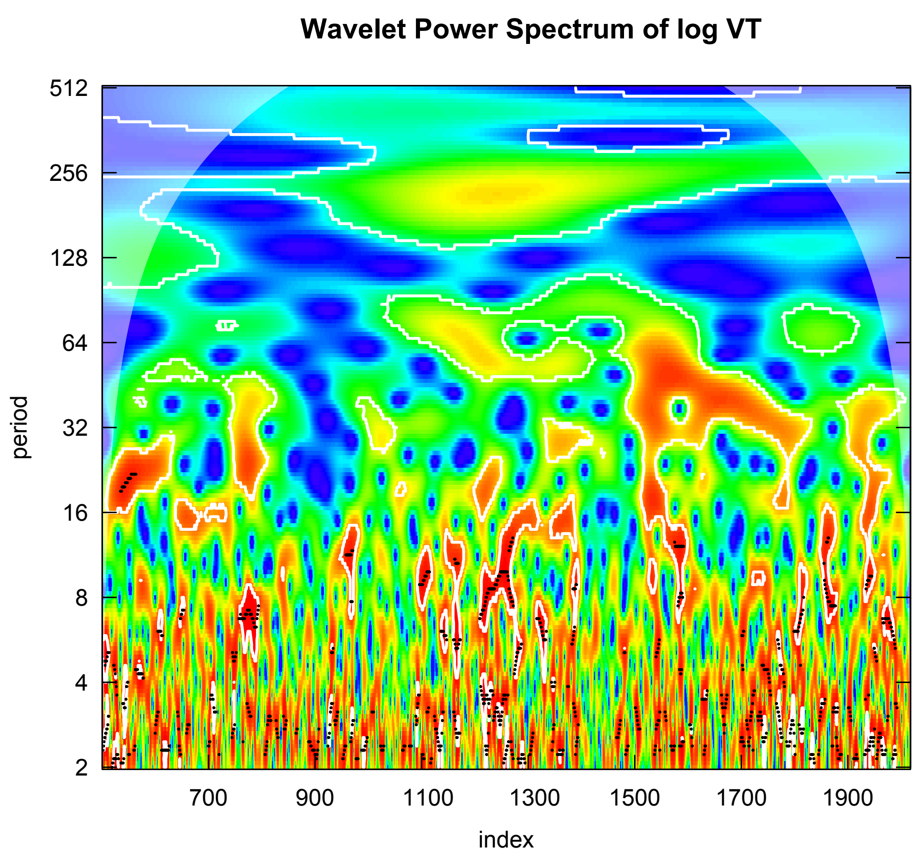


Fig. Wavelet analysis of the 1,500-varve record from Grand Lake, Labrador. White contours delimit areas of >95% confidence levels. Statistical significance was assessed using 100 Monte Carlo simulations were performed to estimate the distribution of wavelet power under a red-noise (AR(1)) null model. The black dot represents a grid cell where the observed power exceeds the significance threshold; they are helpful for visually spotting localized significant features within the broader white contour region. Short periods of around 7-year cycles were particularly active during 600-800AD, the Medieval climate anomaly (1100-1380) and the latter part of the Little Ice Age.

In terms of structure, the authors might consider moving Fig. 5 and the associated paragraphs to the discussion, after section 5.1. The attemps to determine the drivers will come more logically after having explained the depositional process of the varves.

This is a great comment. We will move this part to the discussion. Thanks

Interpretations:

- Line 235: "One of the most notable features in the varve thickness series at GL is the major warm peak between the 1150s-1170s CE." I don't understand this sentence. In Fig. 5, the only predictor tested is precipitation, not temperature. The correlations with temperature proxies (not temperature records) only tell us that general trends in hydroclimates might be consistent regionally. Please rephrase accordingly.

That is correct; however, paleo evidence suggests this period was likely characterized by elevated temperatures and higher sea surface temperatures (SSTs). This is supported by comparisons between our record and reconstructed Northern Hemisphere temperature datasets, including those reflecting AMV phases. These findings suggest that increased precipitation during this time was likely driven by warmer regional and/or hemispheric temperatures. Nevertheless, we will rephrase the sentence accordingly.

- As in most fluvial systems, the VT record might be affected by memory/autocorrelation effects (and nonlinear relationships between precipitation/flood strength/sediment transport). Did you try to use differentiation ($y_{diff}=y-y_{-1}$) to reduce these effects? Or investigate the autocorrelation patterns?

The VT time series shows strong autocorrelation at lags extending up to 40 years (left panel). When applying the same analysis to the differenced series (right panel: $Y_t = X_t - X_{t-1}$), the data appear more random, with values falling within the dashed confidence bounds. This indicates that the original autocorrelation likely stems from a long-term trend rather than true cyclicity. There is strong evidence from paleoclimate records suggesting this trend may be linked to long-term oceanic variability, such as the Atlantic Multidecadal Variability (AMV). Additionally, spectral analysis reveals significant multidecadal variability, with notable peaks around 48 and 76 years (which could be linked to AMV).

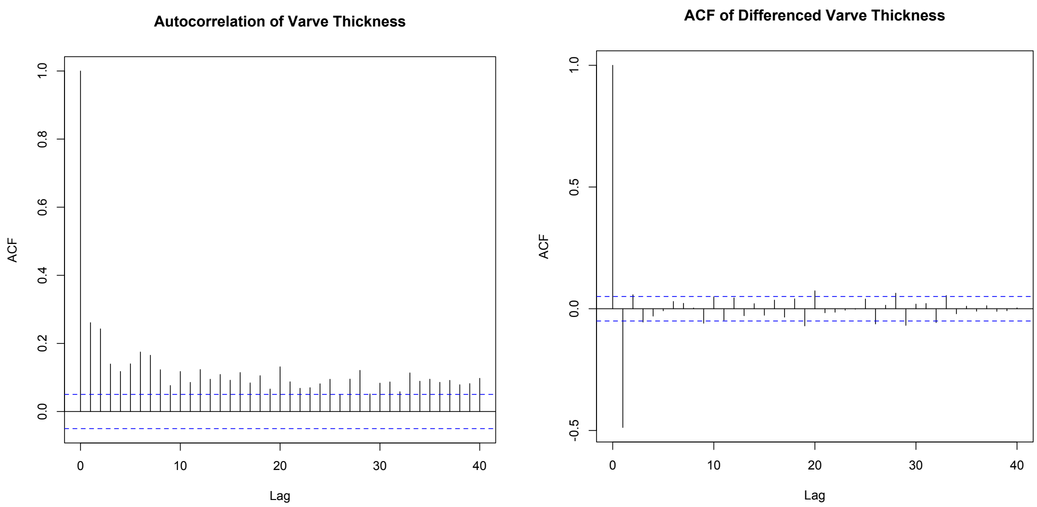


Fig. Autocorrelation functions (ACF) of the original and differenced varve thickness (VT) record from Grand Lake over the past 1,500 years. The ACF of the original VT series (left panel) shows positive autocorrelation at short lags, with values remaining above the 95% confidence bounds up to 40 years. This indicates the presence of low-frequency persistence, likely related to long-term climate variability. In contrast, the ACF of the first-differenced VT series (right panel) displays values that mostly fall within the 95% confidence bounds, consistent with white noise. This suggests that the autocorrelation observed in the original series is primarily driven by long-term trends rather than cyclic behaviour.

- I found the discussion on weather patterns interesting and have no comments on that part.

Thank you.

- A purely curiosity-driven question: do the authors notice changes in interannual variability of VT (potentially precipitation) throughout the record?

Detecting changes in year-to-year variability is challenging due to the high level of noise in the varve thickness record. We think that the NAO plays a significant role, as evidenced by the marked reduction in precipitation during the 1990s, which coincided with a pronounced intensification of the NAO+. We are currently preparing a second paper that applies change point analysis to identify and characterize abrupt shifts in the time series. This work also aims to examine how these shifts align in timing and phase with other high-resolution proxy records from the subpolar gyre.

Figures:
Figure 2: pannel b is hard to see in an otherwise well-constructed and clear figure. Pannel c: what do the codes mean (abp, abo, ab…?)

We will better explain it in the figure caption. Thanks

In Fig. 5,6,7: please add underlines or boxes to show where the MCA and the LIA are.
Figure 6: please plot the original data with the resampled ones so that one can keep track of the data transformation.

Ok, we'll do.

Technical comment: There is ample litterature to emphasize the need to use a log-ratio transformation of xrf data (which are by essence compositional and therefore constrained, see for instance *Kucera and Malmgren, Marine Micropaleontology 1998*). Data presented as count rates can provide spurious results and should therefore be transformed (see recent publication by *Bertrand et al., Earth Science Reviews 2024*, with many clear examples).
I would therefore recommend that the authors revise figure 2 to show the iron relative concentration as log ratios (or centered log-ratios), with a clear x-axis and unit given.

We are perfectly aware that central log-ratios (clr) are the best way to use the XRF compositional datasets. However, clr are needed to help interpret variations of a group of elements over an entire sediment sequence and to deal with potential biases due to, for instance, compaction. Here, we only use Fe peaks to help identify the clay caps of the varves over a very short interval. In this very case, we don't think that log-ratios will provide spurious results. We are therefore respectfully declining the invitation to revise figure 2.

Minor mistakes:
Line 156: Three years *were* added
Thanks