
<table>
<thead>
<tr>
<th>Journal:</th>
<th>Canadian Journal of Civil Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>cjce-2018-0724.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Discussion</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>18-Jan-2019</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Beltaos, Spyros; Environment Canada</td>
</tr>
<tr>
<td>Keyword:</td>
<td></td>
</tr>
<tr>
<td>Is the invited manuscript for consideration in a Special Issue?:</td>
<td>Not applicable (regular submission)</td>
</tr>
</tbody>
</table>

Spyros Beltaos
Watershed Hydrology and Ecology Research Division, Environment and Climate Change Canada, Canada Centre for Inland Waters, 867 Lakeshore Rd, Burlington, ON, L7S 1A1
Spyros.Beltaos@Canada.ca

Word count: ~ 7000 (includes word-equivalents of 3 figures and 1 small table)

ABSTRACT

The writer systematically refutes all of the discussers’ criticisms and supplies more details on the rigour of his “slope-based” frequency analysis. He also identifies key flaws of alternative and conflicting statistical interpretations proposed by the two discussion teams; and provides detailed documentation of the large uncertainty associated with paleolimnological assessments of ice-jam flood frequency.

Keywords: Climate, flood frequency, ice jam, historical record, proxy-record uncertainty, regulation, slope-based method
1. INTRODUCTION

Though relentlessly critical, the discussers’ comments have engendered an opportunity to scrutinize and elucidate conflicting perspectives on the frequency of flooding caused by ice jams in the Lower Peace River. A key goal of both Discussion teams appears to be the disproof of the writer’s simple but robust flood frequency analysis, of which the results do not conform to the discussers’ long-held “no-regulation-effect” viewpoint. In the following sections, the two Discussions are shown to merely contain unfounded criticisms and, on a major issue, to contradict each other.

Before getting into detailed responses to Discussion comments, it is worth reviewing briefly the effects of regulation on the natural hydrograph of Peace River. Figure 1 shows monthly average flows at the Hudson Hope and Peace Point hydrometric stations that are operated by WSC (Water Survey of Canada). Such graphs have been presented, along with naturalized flows, before (e.g. Peters and Prowse 2001), while Fig. 1 presents updated, but observed-only, flow data. Regulation-induced changes to the hydrograph are considerable throughout the year, with the exception of brief transitional periods in spring and fall. Hydrologists and river ice scientists/engineers can easily tell that such changes are bound to have modified the flooding and ice regimes of the river.

Indeed, summer flooding is severely curtailed. Spring breakup flow at Peace Point, which is important in PS-PAD ice jamming, does not appear to have been greatly affected (on average) by regulation because breakup occurs near the time when the two lines (pre- and post-regulation) intersect. [Note: The abbreviation PS-PAD is used herein to denote the Peace Sector of the PAD]
and aims to eliminate inattention-driven confusion; Beltaos (2018) made it clear that his paper
pertained to the Peace Sector, but this was missed by Timoney et al].

The effects on the ice regime of the river are also considerable: (i) freezeup flows (November,
December) are now much greater than pristine ones, suggesting a similar increase in freezeup
water levels and previously unknown, or more frequent, formation of greatly thickened ice covers
via collapse and telescoping of initial surface juxtapositions of ice floes; and (ii) the length of
winter ice cover that is available to form breakup jams has been truncated. It is difficult to assess
the second effect (if any) with respect to the PS-PAD, because of the large distance from the dam.
The first effect inhibits the occurrence of ice-jam floods (IJFs for short) and contributes to the
statistically-demonstrated abrupt reduction in the frequency of ice-jam flooding. Table 1 illustrates
how increasing freezeup levels correspond to decreasing chances of ice-jam flooding in lower
Peace River. Of course, a correspondence does not, by itself, prove a causal association; but where
it is backed by physics, it strongly points in that direction. Interestingly, the positive effect of
moderate and low freezeup levels on ice-jam occurrence in the lower Peace River has been
acknowledged by the lead author of one of the Discussion teams (Timoney 2013, p. 188).

On the other hand, both sets of discussers dispute the writer’s finding that a post-regulation
reduction in IJF frequency has occurred in a statistically significant way; and claim that the
experienced drying of portions of the delta is, instead, entirely due to an ongoing dry cycle that
started decades before regulation. Their view is not based on river ice science and engineering but
on paleolimnological and paleosedimentological proxy records, extending into the past for
hundreds and even thousands of years. It is not clear what can be learned with any confidence from

https://mc06.manuscriptcentral.com/cjce-pubs
proxy records, which no one can verify and no one can challenge for eras that have no observational information. It is not the writer but the authors, themselves, of a key proxy-evidence paper, who stated that:

“All proxy methods used to reconstruct ice-jam flooding rely on assumptions, many of which are difficult to resolve” (Hugenholtz et al 2009).

2. REPLY TO THE DISCUSSION BY TIMONEY ET AL. (abbreviated TAL)

2.1. Section on “Three Ecosystems, Not One”

It appears that TAL did not fully read Beltaos (2018). Had they done so, they would have noticed the following passages, which make it perfectly clear that the paper refers specifically to the Peace Sector and to the perched basins:

- “Major ice jams that occasionally form in lower Peace River generate extensive flooding, which helps replenish the perched basins…” (Beltaos 2018).

- “The Peace River sector of the PAD has experienced prolonged dry periods and…” (Beltaos 2018).

- “Previous work has indicated that occasional spring ice jams in lower Peace River, commonly initiated at the “toe” locations shown in Fig. 1, trigger much of the Delta inundation, particularly in the Peace River sector…” (Beltaos 2018).

- “Therefore, jams are particularly effective in replenishing the higher-elevation, or “perched”, basins of the PAD” (Beltaos 2018).

- “It is the “large” variety that generates widespread flooding (e.g., Straka and Gray 2014) and supplies significant water replenishment to the Peace sector of the Delta. Timoney’s compilation
does not include the years 2009 to 2017, during which the only noteworthy event in the Peace sector of the Delta was a large IJF in May of 2014.” (Beltoas 2018).

In view of the above quotations, this particular TAL section is superfluous.

2.2. Section on “Within-Delta Processes Affecting the Water Balance”

This is all irrelevant to the subject of Beltoas (2018), a limited-length “Technical Note”, which deals exclusively with the frequency of ice jams in the lower Peace River. Within-delta processes have zero, or next to zero, effect on the ice jamming regime of the lower Peace River.

2.3. Section on “A Concordant Environmental History”

In the first paragraph of this section, TAL stated that “The conclusion of Beltoas (2018) that paleolimnological findings “often differ from the observational record” is erroneous”.

The writer rejects this assessment and supplies detailed documentation in Appendix A. Very briefly, this appendix shows that the proxy records of two nearby lake sites are highly discordant between themselves, while the less unreliable of the two has much higher failure rate in detecting known flood events. It follows, moreover, that a subsequent negative TAL “appraisal” of the proven physics of ice jamming processes is facile.

In the fourth paragraph of the same TAL section, it is stated that: “Beltoas (2018) focused on the “large” floods reported by Timoney (2009). It is unclear why that analysis did not include the lower flood frequency 19th century data and instead focussed on flood-prone unregulated periods during the 20th century”.

The writer again finds it necessary to quote from Beltoas (2018):
Ideally, the effect of regulation on flooding should be assessed by comparing flood frequencies under regulated and “pristine-RP” conditions. The term “pristine-RP” is used herein to describe the unknown conditions that would have prevailed during the period of regulation (1968 to 2017), had Peace River not been regulated. An approximation to pristine-RP conditions can be furnished by the record of the unregulated period (up to, and including, 1967), but a question arises as to the duration of this period. If too short, it may simply describe a regime that is not representative of pristine-RP conditions, which refer to the years 1968 to 2017; if too long, it may be influenced by climatic regimes of long ago that have little resemblance to those of the latter period.” (Beltaos 2018).

The reason, therefore, for limiting oneself to the 20\textsuperscript{th} century is to ensure that the pristine period that is chosen for the comparison with the regulated period, matches, as much as possible, the latter in climatic conditions. If one wishes to find a time interval that has similar climate to that of the regulation period, it is advisable to look at the recent, not the distant, past. In addition to climate, present-era channel morphology (which also influences ice jam occurrence and magnitude), could be very different from what it was in, say, AD 1150, 1750, etc. A further concern pertains to the completeness of the historical records at different time periods. Nowadays, IJFs in the lower Peace do not go unnoticed or unrecorded. But is this true of the 1700s, 1800s, or even the pre-regulation years of the 1900s? Even as late as 1967, a flood was reported without mention of ice jamming (Timoney 2009, Table 1). Yet, recent numerical modelling using observed spring and summer flow peaks has indicated that a flood could have only occurred in the presence of an ice jam (Beltaos, unpublished data). This inference is consistent with the high spring flow of that year and
the large volume of water that was diverted towards the PAD, which is comparable to volumes delivered by large IJFs (Peters and Buttle 2010).

In the fourth paragraph of the same TAL section, it is also stated that: “Nor did Beltaos (2018) analyze “small” or “moderate” floods reported by Timoney (2009) even though the assignment of flood magnitude is tentative in many historic flood accounts”.

Had TAL fully read the paper, they would have noticed that Beltaos (2018) did analyze moderate floods and stated: “Similar results are obtained if one considers cumulative numbers of large plus-moderate floods (frequency decreases from ~0.3 before, to ~0.15 events per year, after 1967)” (Beltaos 2018); here, the word “similar” refers to the statistical results obtained with the large events. Moreover, the writer prefers to focus on large events because the “moderate” category is of doubtful effectiveness for perched-lake replenishment. For instance, Timoney (2009) classified the 1994 ice-jam event as moderate; however, Peters and Buttle (2010) indicated that this event diverted a negligible volume of water towards the PAD, relative to what large floods deliver. The writer is even more skeptical about the potential contribution of “small” floods to the perched lakes.

Regarding the questioned wording “fading local memory”, the writer used it figuratively: he had meant, and maintains, that what is stored in current “memory”, e.g. in a contemporary book or on a computer hard disc, may not include all that has actually occurred because records of events are likely to be increasingly incomplete as one goes farther into the past. The word “local” obviously refers to the lower Peace River and the Peace Sector of the PAD, an area of very low population
density, where pre-regulation IJFs could go unnoticed or unrecorded more readily than in a densely populated area.

### 2.4. Section on “Statistical Analysis of Flood History”

A major, and perhaps the most important, point of the discussers’ statistical criticism pertains to their perceived lack of normality in the writer’s regression residuals: “Normality of the regression residuals is questionable (Fig. 2A)....”. Apparently, the discussers’ assertion is based on the fact that the line drawn through the data points in their Fig. 2A is not completely straight (as it should be in the case of a normal distribution). For a direct visual check, the writer at first plotted the empirical probability distribution of the ranked standardized residuals, using the Excel plotting position formula \[(\text{rank}-0.5)/N\]; and then superimposed the standard normal distribution to the data points, as shown in Fig. 2. No perfect coincidence can be claimed, but the residuals could well belong to a normal distribution since they constitute a “sample” of finite size; it is well known that finite samples, drawn randomly from a normal population, often exhibit departures from the normal line when plotted in the form of Fig. 2 or that of Fig. 2A of TAL. Similar plots are exhibited by all pre-regulation series listed in Table A1 of Beltaos (2018). The writer is, moreover, skeptical about the discussers’ claim that his residuals are not homoscedastic. The inclined linear segments in Fig. 2B of TAL simply result from the step-like form of the time series \(Y(t)\) (\(Y = \text{cumulative number of floods}; t = \text{year}\)). During periods of no flooding, \(Y\) is constant; subtraction from the regression line yields a set of residuals that are necessarily arranged along an inclined linear segment. However, there is no consistent change of the width of the scatter band. Consistent change of this width, e.g. continuously increasing or...
decreasing, implies a similarly changing variance and is a defining feature of heteroscedasticity. Figure 3 shows a longer record, with repeated occurrences of inclined linear segments, and again without consistent change in the width of the scatter band. Similarly, the dependency exhibited by the residuals is largely apparent: the variable Y and the resulting residuals arise from the summing of binary outcomes; both are bound to exhibit dependence, even though the binary flood/no-flood occurrences are independent (see also later discussion of IJF independence in this section).

Undoubtedly, neither Fig. 2 nor Fig. 3 conforms to “textbook-like” requirements on residuals. However, it seems probable that a larger sample, such as one comprising all the results of numerous n-trial binary strings (n = total years of applicable record), would plot even closer to the normal line in Fig. 2 and “fill up” the gaps in the scatter band of Fig. 3. Statistical imperfections of the residuals can indeed impact t-statistics and P-values, as noted by the discussers; Wilks (2006) indicates that the main concern pertains to underestimation of the sample variance and thence of P-values. To examine the potential magnitude of such impacts, the writer doubled the calculated variances of the regulation and pre-regulation series examined in Beltaos (2018). Though this is an extreme change, the P-values shown in Table A1 of Beltaos (2018) remained infinitesimally small. The null hypothesis (no change in slope between the compared time periods) is still rejected with near-certainty under the OLS regression approach (Method 1 of Beltaos 2018); and this continues to be consistent with the results of the completely different, non-parametric, Method 2 (P-value = $4 \times 10^{-9}$).

With reference to the Mann-Kendall test, which the writer used in his Method 2, the discussers stated: “Beltaos (2018) also applied a Mann-Kendall test as a non-parametric alternative to the
parametric OLS-based t-test. A Mann-Kendall test is used to evaluate if a trend exists within a single time series (Hirsch et al. 1982). It is not intended to test for a difference between segments of a single time series as was done in Beltaos (2018)."

It is not factual that the Mann-Kendall test was applied as a non-parametric alternative to the OLS-based test within a single time series. The test was applied to a new, single, series equal to “the difference between cumulative numbers of floods for equal-duration periods, 1922 to 1967 (pre-regulation) and 1972 to 2017 (post-regulation). The starting year in each series is assigned a value of 1 and subsequent years are assigned values of 2, 3, ..., 46.” (Beltaos 2018). The obvious physical background is that the series 1922-1967 is, in Method 2, considered an approximation to what would have occurred during 1972-2017, had the river not been regulated (pristine-RP condition), as had been reasoned earlier in the paper. The difference between pristine-RP flood numbers and observed flood numbers should remain constant if there were no regulation effect, meaning that the Mann-Kendall test would result in a large P-value. A very small P-value, on the other hand, points to a statistically significant trend and thence to a regulation effect. Method 2 is another approach for detecting a statistical difference between the flood frequencies of two time series, and completely unrelated to Method 1. A non-parametric test was sensibly chosen because quantification of the slope of the new series was not needed. It is worth recalling here that Method 2 indicated a highly significant flood “deficit” for the regulation period (P-value = 4 × 10⁻⁹), which is in full accord with the results of Method 1.

Moreover, the use of a non-parametric test in Method 2 nullifies the discussers’ next objection (see below), which claims that there is correlation in the flood series, rendering parametric OLS
approaches “inappropriate”. Even if the OLS results were inappropriate (which they are not, as shown earlier), the non-parametric P-value of Method 2 results in a non-OLS-derived, near-certain rejection of the null hypothesis.

The preceding paragraphs fully refute the criticisms regarding the validity of the writer’s analysis. But the discussers argued next that: “Because flood events tend to occur in correlated runs (flood years tend to follow flood years and non-flood years tend to follow non-flood years (Timoney et al. 1997), parametric approaches based on OLS (e.g. Beltaos 2018) or Bernoulli trials (Timoney 2013), which assume independent events, are inappropriate. We instead use a block bootstrapping approach...”.

This statement is questionable, for the following reasons:

(i) With reference to large IJFs, Timoney’s (2009) Table 1 indicates that since the year 1900, there have been 4 instances in which flood years followed flood years (1932-33, 33-34, 42-43, and 96-97); and 12 instances in which flood years were followed by non-flood years (1900, 04, 20, 34, 43, 58, 63, 65, 72, 74, 97; plus 2014, which occurred after publication of Timoney’s 2009 paper). Consequently, the assertion “flood years tend to follow flood years” is not factual.

(ii) The notion that independence is negated by the fact that “non-flood years tend to follow non-flood years” is illogical. In a series of independent Bernoulli trials (e.g. flood/no flood, heads/tails, 1/0), with one outcome being much more likely than the other, one would expect the more likely outcome to “tend to follow” any one trial. In the case of rare events, such as floods, non-flood
years occur much more frequently than flood years. Therefore a non-flood year is much more likely (than a flood year) to follow a non-flood year.

Independence of IJFs is further reinforced by the near-coincidence of P-values obtained by the TAL bootstrapping exercise and by the simple binomial approach applied by Timoney (2013), which assumes full independence.

At the same time, the results of the bootstrapping application are not persuasive because the formulation of the null hypothesis is incomplete: as has been shown by the writer (Beltaos 2018), it is not only the number of floods but also their temporal order of occurrence that determine flood frequency. Therefore, a more accurate null hypothesis would have been “no more than the observed number of floods in the regulated period and slope of their cumulative graph not exceeding the observed value”. It is very likely that the correct hypothesis would have resulted in lower P-values, possibly even below 0.05. It is, moreover, unclear whether and how the changing frequency during the examined time periods has been quantified in the bootstrapping applications. Such deficiencies highlight the tenuous nature of the discussers’ statistical interpretations, while underscoring the robustness of the writer’s slope-based method, which takes into account both the number and the temporal order of flood occurrences.

It is moreover noteworthy that, in conducting statistical tests, the discussers prefer to use very long pre-regulation periods, extending as far back as 1826, because: “Using a short period of record (1900-1967) as the null distribution….risks biasing the results to only wet futures”.

https://mc06.manuscriptcentral.com/cjce-pubs
The reference to “futures” is puzzling. The writer merely wished to compare the regulated flood frequency, as it is revealed by observed events during the period 1972-2017 with the closest approximation to the “pristine-RP” series (that would have occurred in the same period had regulation not been implemented). The future is irrelevant to this very simple, but apparently overlooked, concept. As for the writer “biasing the results”, the opposite is actually true. As stated earlier, as well as in Beltaos (2018), the best possible choice of a pristine period is one that is subject to the same, or approximately the same, climatic and geomorphic conditions as the regulation period. Ice-jam location, severity, and frequency depend on climate and channel morphology, both of which change over time. For instance, one of the anonymous reviewers of this Reply indicated that “channel conditions are likely not representative of the conditions over 100 years later”. Moreover, instances of unrecorded flood events are likely to increase as one goes farther into the past. Therefore, flood frequencies of very long ago are irrelevant, while their use leads to conclusions that happen to favour the discussers’ no-regulation-effect viewpoint. [Due to differing climatic and geomorphic conditions and/or incomplete and unreliable past-era records, apparent IJF frequency of centuries ago is very low relative to the present (~3% in 1600; 6% in 1800; Timoney 2009, Fig. 3c)].

Summarizing, it has been shown in the reply to TAL that:

(a) The discussers’ notion of a “Concordant Environmental History” has weak factual basis, while the paleolimnological flood record is discordant and unreliable.

(b) Both parametric and non-parametric frequency comparisons of pristine-to-regulation periods, which have been performed by the writer, correctly reject the null hypothesis with near-certainty.
(c) The discussers’ bootstrapping analysis possibly ignores the variability of flood frequency during the examined pristine periods, and adopts a deficient null hypothesis that very likely results in overestimation of the P-value.

(d) The discussers’ use of very long pristine periods introduces bias, which happens to favour their “no-regulation-effect” viewpoint.

3. REPLY TO DISCUSSION BY HALL ET AL. (abbreviated HAL)

HAL objected to the writer’s analysis “…because residuals are not independent (as shown in Beltaos (2018) Figure 3, which demonstrates that non-flood years tend to follow non-flood years)”. The fact that non-flood years tend to follow non-flood years is not evidence of dependence, but a necessary property of a series of independent binary trials in which the two possible outcomes have greatly unequal probabilities (e.g. floods/non-floods). Because a flood is a low-frequency event, what is more likely to follow a non-flood year?

Immediately following the above-quoted text, HAL stated that: “But even if we excuse this, the use of quadratic regression as shown in Figure 3 of Beltaos (2018) is flawed because it implicitly predicts that flood frequency increases over time, for which there is no physical basis nor does it align with independent, paleolimnological reconstructions of flood frequency and magnitude from oxbow lake sediment records in the delta that are proximal to the Peace River (Wolfe et al. 2006). Thus, it is misleading to construe an expectation of accelerating flood frequency during the post-regulation era.”

No less than four points are wrong in this short statement:
(i) The writer’s Fig. 3 does not “predict”, but merely detects, an increase in frequency over time.

(ii) There is a physical basis for gradual changes in flood frequency: it is called climatic variability.

(iii) It is not the writer, but the historical record itself (on which the writer’s analysis was performed) that does not align with paleolimnological reconstructions (see also reply to TAL and Appendix A).

(iv) No expectation of accelerating flood frequency for the post-regulation era should be ascribed to the writer: in Fig. 3 of Beltaos (2018) the pre-regulation polynomial trendline ends in 1967. Extrapolations of statistical trends beyond the record for which they have been developed are neither rigorous nor advisable. It is for this reason that the writer assessed pristine frequency for different scenarios of pre-1968 time intervals (Table A1 of Beltaos 2018), rather than adopt the frequency that would be indicated by extrapolation of the polynomial line to 2017.

Next, HAL stated that: “Furthermore, inflation of the effective sample size by counting every year as an observation for the purposes of regression analysis, as performed by Beltaos (2018; i.e., cumulative floods for 1920 = 4, 1921 = 4, 1922 = 4… 1931 = 4 are part of his regression so n = 88), results in excessive confidence in the estimated frequency of floods (and slope of the relation) between 12 major pre-regulation flood events over 88 yrs. We have attempted to account for some of these shortcomings by applying a more conservative linear regression model, which only includes observed floods as data points to characterize the baseline (i.e., 1880-1967) interval (Figure 1A).”

The writer could not agree less. Inclusion of non-flood years is not “inflation”, but acknowledgment of reality: each non-flood year carries important information (a flood did not occur in that year), which a statistician cannot ignore. Consider, for example, a hypothetical time
interval of, say 20 years, which contains, say 4 floods, all occurring in consecutive years, say years 12-15. The discussers’ “pruned-record” scheme would then indicate a flood frequency equal to 1(!), while the prediction-interval lines (such as those shown in Fig. 1 of HAL) would collapse onto the regression line (!). The latter feature would persist even if the floods did not occur in consecutive years, so long as they were temporally spaced at equal intervals. There are more issues with the HAL scheme and the attendant regression extrapolations, but its striking lack of statistical rigour suggests that no further attention need be paid to it.

At this point, the writer cannot help reflecting that both teams of discussers (TAL and HAL) dispute his method of analysis, but each for a different reason: HAL actually adopt the slope-based approach, but question the “legitimacy” of non-flood years. TAL, on the other hand, do not dispute the status of non-flood years, but question the slope-based approach. It follows that they contradict each other on both of these points, while each of HAL and TAL agree with the writer on one point.

The discussers also expressed reservations about the writer’s flood data set: “For example, the dataset assumes each flood possesses the same magnitude and each flood is inferred to have the same substantial effect on the water balance of perched lakes in the delta.” This statement is not factual. Consistent with common hydrological practice, no such assumptions and inferences are made when counting floods. A flood is simply understood to be an event of sufficient magnitude to cause flooding. Mathematically speaking, it is defined as an event of which the magnitude equals or exceeds an agreed-upon flood threshold. This leads to the concept of exceedance probability, which is quantified by assigning a value of 1 to each flood year, and 0 to each non-flood year. The writer’s slope-based method adheres to this concept and enables
statistically more robust frequency estimates than the conventional division of the number of floods by the number of years of record; it also captures changes in frequency that may have occurred during the time interval of interest. Hydrologists understand the binary 1/0 (flood/no-flood) concept and do not infer that the floods are equal in magnitude. The issue of the hydro-ecological effectiveness of various floods is both interesting and complex, but beyond the scope of Beltaos’ (2018) short paper.

HAL further stated that: “Beltaos (2018) also counters conclusions that we have drawn from our studies that lake drying began several decades before the construction of the WAC Bennett Dam. Instead, he argues in favour of the apparent quadratic-inferred pre-1968 (i.e., pre-dam) increase in the observational ice-jam flood frequency record for the lower Peace River”.

As is thoroughly documented in Appendix A, in which the reliability of paleolimnological evidence is scrutinized, the notion that lake drying began several decades before construction of the Bennett Dam is refuted by the historical IJF record and by the floodbed-sequence proxy record (Fig. 3c of Timoney 2009). Guided by common sense, the writer chose to base his analysis on the historical record. If one were to accept the discussers’ claim that lake drying began decades before construction of the Bennett Dam, one would also need to assume that replenishment of the perched basins is not linked to IJFs, which at that time were becoming more, not less, frequent. Such assumption seems highly unrealistic to this writer.

Summarizing, the HAL criticisms have been fully refuted in the preceding paragraphs. The proposed regression scheme, which “expels” non-flood years, was shown to be particularly flawed, while the claimed commencement of a drying trend decades before regulation was challenged.
4. CONCLUDING REMARKS

Despite their highly critical, albeit tenuous, content, the two Discussions have prompted helpful elucidation of various aspects of IJF frequency, which will hopefully contribute towards eventual convergence of conflicting viewpoints. The writer has refuted the discussers’ criticisms and disproved their arguments in favour of using centuries-long unregulated periods to compare with the regulation period. He also pointed out that their proposed alternative frequency analyses are not only unsound but also contradict each other: the HAL exclusion of non-flood years from the historical record spawned a fanciful regression scheme; while the TAL bootstrapping method very likely overestimated P-values and possibly did not account for frequency variability during the examined pre-regulation intervals. Moreover, the discussers’ comments elicited scrutiny of paleolimnological proxy records of IJFs and exposure of their serious limitations.

APPENDIX A. SCRUTINY OF PALEOLIMNOLOGICAL FINDINGS

In his comprehensive review of the paleolimnological evidence pertaining to the PAD, which was submitted by BC Hydro to the Joint Review Panel for the (then-proposed) Site C dam, Smol (2013) stated:

“According to the ~180-year flood record generated from PAD 54, maximum flood frequency occurred during the first third of the 20th century (see Figure PAD 54/15-9). Between ~1900 and ~1935, 11 major floods are recorded at this site, and only 8 major floods occurred during the remainder of the 20th century (including 1996 and 1997) with half of these occurring during a 10-year time period from about the early 1940s to early 1950s. Notably, this record shows an
estimated 14-year hiatus of major floods between ~1951 and ~1965, immediately prior to the
construction of the WAC Bennett Dam in 1968. At the apparently less flood-prone PAD 15 site,
maximum flood frequency is inferred to have occurred earlier between ~1875 and ~1900 when
four major floods can be identified. Only 5 major floods can be distinguished (including 1996 and
1997) during the 20th century. While differences in the timing of maximum flood frequency at these
two sites are currently difficult to resolve, these data suggest that flood frequency had been in
decline for several decades preceding Peace River regulation, beginning in the early to mid-20th
century (at PAD 54) or perhaps as early as the late 19th century (at PAD 15).[Smol 2013].

The “major” floods referred to by Smol (2013) are events equal in magnitude, or greater than, the
flood of 1974 (a large IJF), as inferred by Wolfe et al (2006) using magnetic susceptibility data at
two lake sites in the PS-PAD, labelled PAD 54 and PAD 15. One may note the large discrepancy
between the proxy flood records of the two sampling sites, which, according to Wolfe et al (2006),
“may reflect that the sill elevation is higher at PAD 15 than at PAD 54 or the location of PAD 54
is more prone to flooding”. Earlier in that paper, however, it was noted that the sill elevation at
PAD 54 may only be slightly lower than that at PAD 15. Considering also the proximity of the
two sites to each other and to the southern bank of Peace River, it is hard to imagine that only one
of them would receive flood water during a large IJF. Consequently, it seems very likely that the
discordant records of PAD 54 and PAD 15 simply indicate that the paleolimnological approach is
influenced by uncontrolled factors that can lead to large errors.

The historical review by Timoney (2009) begins in 1710 and ends in 2008, but it is best to focus
on more recent times because it is likely that instances in which IJFs have gone unrecorded become
more frequent as one goes farther back into the past. In the 20\textsuperscript{th} century, Timoney’s Table 1 lists large IJFs in 1900, 04, 20, 32, 33, 34, 42, 43, 58, 63, 65, 72, 74, 96, and 97. Clearly, only PAD 54 need be compared to the historical record, since doing so for PAD 15 would be hopeless. Between 1900 and 1935, PAD 54 detected 11 major floods (according to the above quotation by Smol 2013), which can be compared to the 6 events of the historical record. For the remainder of the 20\textsuperscript{th} century, PAD 54 detected 8 major floods, as compared to the 9 events in the historical record. However, the 14-year hiatus of the proxy record between 1951 and 1965 does not match the historical record, which contains major floods in 1958 and 1963 as well as in 1965. The failure to detect the 1963 flood is particularly disappointing, because that flood generated the largest, by far, spring breakup reverse flow volume (~5 billion m\textsuperscript{3}) during the period of the Peace Point hydrometric record (Peters and Buttle 2010), which began in 1959. By comparison, the proxy-record’s “threshold” flood of 1974 delivered 1.8 billion m\textsuperscript{3}.

A more detailed comparison, based on historical major events and proxy-based flood years indicated in Fig. 8 of Wolfe et al. 2006, resulted in the following (for PAD 54): Matches and likely matches: 1920, 33, 42, 43, 65. False positives (PAD 54 detected a flood that did not actually occur): 1905-07, 09-10, 13-14, 19, 22-23, 25, 27, 28, 44, 49-50, 51. False negatives (PAD 54 failed to detect a flood that actually did occur): 1900, 04, 32, 34, 58, 63, 72. False negatives also occurred for the flood years 1996 and 1997, but were excused because the failure was apparently due to “very high water content in these uppermost sediments”. Clearly, even the proxy record of the PAD 54 site contains many more failures than successes. As for PAD 15, a detailed comparison would have been futile.
At this point, it is worth recalling two very pertinent sentences from the Smol (2013) quotation:

“According to the ~180-year flood record generated from PAD 54, maximum flood frequency occurred during the first third of the 20th century...”. And: “…these data suggest that flood frequency had been in decline for several decades preceding Peace River regulation, beginning in the early to mid-20th century (at PAD 54) or perhaps as early as the late 19th century (at PAD 15”).

Together, these statements summarize a key aspect of the notion that the drying of the perched basins began well before construction of the Bennett Dam (e.g. see HAL Discussion). Yet, the historical record, as is graphically illustrated in Fig. 3 of Beltaos (2018), indicates a steadily increasing frequency during the period 1900-1967; there is not the slightest hint of a “maximum” occurring in the first third of the 20th century. Increasing 20th century frequency is also indicated in Fig. 3c of Timoney (2009) up to ~1960 (though this is based on proxy sedimentological data and is thence less reliable than the historical record). These considerations suggest that the claimed occurrence of an early-twentieth century flood frequency “maximum” and subsequent decline are merely products of the very large uncertainty associated with the PAD 54 and PAD 15 records.

REFERENCES


FIGURE CAPTIONS

Fig. 1. Pre- and post-regulation hydrographs of monthly flows at the Hudson Hope (HH) and Peace Point gauges (PP), respectively located ~30 and ~1135 km downstream of the Bennett Dam. Based entirely on WSC data.

Fig. 2. Probability distribution of Beltaos’ (2018) residuals, compared to the normal distribution (1972-2017 series).

Fig. 3. Plot of regression residuals versus fitted Y values (Y = cumulative number of floods). 1900-1967 time series.
Table 1. Large IJF occurrences in different ranges of freezeup levels at Peace Point (period of hydrometric record: 1959-present); reservoir filling years 1968-71 are excluded

<table>
<thead>
<tr>
<th>Freezeup elevation (m)</th>
<th>Number of years</th>
<th>Number of IJFs</th>
<th>IJFs per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;213</td>
<td>11</td>
<td>3</td>
<td>0.27</td>
</tr>
<tr>
<td>213-214</td>
<td>17</td>
<td>3</td>
<td>0.18</td>
</tr>
<tr>
<td>214-215</td>
<td>18</td>
<td>1</td>
<td>0.06</td>
</tr>
<tr>
<td>&gt;215</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>