An assessment of biomarker-based multivariate classification methods versus the PIP_{25} index for paleo Arctic sea ice reconstruction

Denizcan Köseoğlu\textsuperscript{a}, Simon T Belt\textsuperscript{a,}\textsuperscript{*}, Katrine Husum\textsuperscript{b}, Jochen Knies\textsuperscript{c,d}

\textsuperscript{a}Biogeochemistry Research Centre, School of Geography, Earth and Environmental Sciences, Plymouth University, Plymouth, PL4 8AA, UK.

\textsuperscript{b}Norwegian Polar Institute, Fram Centre, NO-9296 Tromsø, Norway.

\textsuperscript{c}CAGE – Centre for Arctic Gas Hydrate, Environment and Climate, Department of Geosciences, UiT The Arctic University of Norway, 9037 Tromsø, Norway.

\textsuperscript{d}Geological Survey of Norway, N-7491 Trondheim, Norway.

*Author for correspondence

E-mail: sbelt@plymouth.ac.uk

Keywords: Arctic; Sea ice; HBI; Biomarker; Classification Tree; IP\textsubscript{25}; PIP\textsubscript{25}; Barents Sea
Abstract

The development of various combinative methods for Arctic sea ice reconstruction using the syngenic highly- branched isoprenoid (HBI) IP$_{25}$ in conjunction with pelagic biomarkers has often facilitated more detailed descriptions of sea ice conditions than using IP$_{25}$ alone. Here, we investigated the application of the Phytoplankton-IP$_{25}$ index (PIP$_{25}$) and a recently proposed Classification Tree (CT) model for describing temporal shifts in sea ice conditions to assess the consistency of both methods. Based on biomarker data from three downcore records from the Barents Sea spanning millennial timescales, we showcase apparent and potential limitations of both approaches, and provide recommendations for their identification or prevention. Both methods provided generally consistent outcomes and, within the studied cores, captured abrupt shifts in sea ice regimes, such as those evident during the Younger Dryas, as well as more gradual trends in sea ice conditions during the Holocene. The most significant discrepancies occurred during periods of highly unstable climate change, such as those characteristic of the Younger Dryas–Holocene transition. Such intervals of increased discrepancy were identifiable by significant changes of HBI distributions and correlations to values not observed in proximal surface sediments. We suggest that periods of highly- fluctuating climate that are not represented in modern settings may hinder the performance and complementary application of PIP$_{25}$ and CT-based methods, and that data visualisation techniques should be employed to identify such occurrences in downcore records. Additionally, due to the reliance of both methods on biomarker distributions, we emphasise the importance of accurate and consistent biomarker quantification.
1. Introduction

Arctic sea ice is a pivotal component of the global ecosystem. The receding sea ice edge is a site of primary productivity during the spring-summer melt season (e.g., Wassmann et al., 2006; Vancoppenolle et al., 2013, and references therein), while ice formation facilitates deep water formation and helps to maintain the global thermohaline circulation (e.g., Bitz et al., 2006). Additionally, sea ice is an effective reflector of incoming shortwave solar radiation, thus regulating the oceanic heat budget (e.g., Meier et al., 2014, and references therein) and the ocean-atmosphere heat exchange (Maykut, 1978). The sensitivity of the melting-freezing cycle and physical properties of sea ice to the global atmospheric and oceanic circulation of moisture and heat (e.g., Smedsrud et al., 2013) makes seasonal and interannual variability of sea ice cover a prime indicator of climate change (Vihma, 2014). The decline of Arctic sea ice extent and thickness observed via satellite passive microwave sensors since the 1970’s (Lindsay and Schweiger, 2015; Fetterer et al., 2017) is unprecedented within observational records covering recent centuries, at least (Divine and Dick, 2006; Walsh et al., 2017). Cumulative thinning and retreat of Arctic sea ice leads to pre-disposition of the ice cover for accelerated melting via various positive feedback mechanisms (Perovich and Polashenski, 2012), and augments the global temperature increase at high latitudes (Serreze and Barry, 2011). Thus, the Arctic is projected to become ice-free at its September minimum within the next few decades (Overland and Wang, 2013) due to potentially irreversible loss of ice cover (e.g., Lindsay and Zhang, 2005; Eisenman and Wettlaufer, 2009), with important implications for global oceanographic regimes, atmospheric heat circulation (Smedsrud et al., 2013), mid-latitude weather (Mori et al., 2014), Arctic food webs (Harada, 2016), and human activities (Meier et al., 2014). The assessment of such implications and accurate prediction of future trends requires the reconstruction of sea ice cover over geologically significant timescales (Stroeve et al., 2015).
IP$_{25}$, a source-specific highly-branched isoprenoid (HBI) biomarker synthesized by sympagic diatoms (Brown et al., 2014), has emerged as a relatively direct proxy of seasonal Arctic sea ice (Belt et al., 2007; Belt and Müller, 2013). Consistent with its source, IP$_{25}$ has been detected throughout the Arctic in surface sediments characterised by seasonal sea ice cover (Belt et al., 2007, 2015; Navarro-Rodriguez et al., 2013; Stoynova et al., 2013; Weckström et al., 2013; Xiao et al., 2013, 2015a; Ribeiro et al., 2017), and is mostly absent in ice-free settings (Müller et al., 2012; Méheust et al., 2013). Additionally, IP$_{25}$ appears to be relatively stable within sedimentary records for millions of years (Stein and Fahl, 2013; Knies et al., 2014; Stein et al., 2016). Combined, the seasonal sea ice selectivity, source-specificity and stability have facilitated the use of IP$_{25}$ for palaeo-sea ice reconstructions throughout the Arctic spanning a range of timescales (e.g., Andrews et al., 2009; Müller et al., 2009, 2012; Vare et al., 2010; Stein and Fahl, 2012, 2013; Berben et al., 2014, 2017; Knies et al., 2014, 2017; Müller and Stein, 2014; Belt et al., 2015; Xiao et al., 2015b; Cabedo-Sanz and Belt, 2016; Cabedo-Sanz et al., 2016; Hoff et al., 2016; Stein et al., 2016; Bartels et al., 2017).

However, while relative changes in IP$_{25}$ concentration are generally consistent with corresponding shifts in sea ice conditions (Massé et al., 2008; Andrews et al., 2009; Vare et al., 2010; Axford et al., 2011), the considerable differences between absolute sedimentary IP$_{25}$ concentration ranges for locations experiencing similar sea ice conditions (Stoynova et al., 2013; Xiao et al., 2015b) limits comparison of sea ice variability for different Arctic regions.

To help provide more detailed and regionally comparable descriptions of sea ice conditions, Müller et al. (2011) first introduced the Phytoplankton-IP$_{25}$ index (PIP$_{25}$; Eq. 1), based on IP$_{25}$ abundance relative to that of open-water biomarkers produced by marine phytoplankton, such as brassicasterol and dinosterol (Volkman, 1986, 2006). Thus, a normalised uniform scale (0–1) removed the influence of absolute concentrations, allowing
more consistent comparisons of sea ice conditions between Arctic regions, while a
centrations balance factor (i.e. \( c \)-factor) compensated for different concentration ranges
commonly observed for IP\(_{25} \) and sterols (Müller et al., 2011; Cabedo-Sanz and Belt, 2016). In
practice, correlation of sterol-based PIP\(_{25} \) indices and overlying sea ice concentrations yielded
variable results for different Arctic regions (Müller et al., 2011; Navarro-Rodriguez et al.,
2013; Xiao et al., 2015b), partially attributed to the lower source specificity of sterols, which
are produced ubiquitously by a variety of marine, terrigenous (Yunker et al., 2005; Rampen
et al., 2010), and even ice-obligate sources (Belt et al., 2013, 2018). Moreover, the value of
the \( c \)-factor greatly affected PIP\(_{25} \) estimates in downcore records due to large and variable
differences in IP\(_{25} \) and sterol concentration ranges in different core sections representing
periods of abrupt shifts between contrasting climate conditions (Belt and Müller, 2013;
Berben et al., 2014; Cabedo-Sanz and Belt, 2016). More recently, substitution of sterols with
a source-specific tri-unsaturated HBI biomarker (hereafter referred to as HBI III (Z); Fig. 1)
derived from some pelagic *Pleurosigma* and *Rhizosolenia* spp. (Belt et al., 2000, 2017;
Rowland et al., 2001) has significantly reduced the influence of the \( c \)-factor, in some cases
(e.g., Belt et al., 2015; Cabedo-Sanz and Belt, 2016), as a result of comparable concentration
ranges of IP\(_{25} \) and HBI III (Z). A regional calibration of P\(_{III}IP_{25} \) (i.e. PIP\(_{25} \) using HBI III (Z)
as the pelagic biomarker) versus satellite-derived spring sea ice concentration (%SpSIC)
allowed calculation of semi-quantitative SpSIC estimates in the Barents Sea and the western
Svalbard margin (Eq. 2; Cabedo-Sanz and Belt, 2016; Smik et al., 2016; Berben et al., 2017).
Nonetheless, the magnitude of the \( c \)-factor, which relies on average biomarker concentrations
throughout a dataset when a regional P\(_{III}IP_{25} \)–SpSIC calibration is not available (Müller et al.,
2011), could still potentially introduce bias to P\(_{III}IP_{25} \)-derived SpSIC estimates. Further, the
objective choice of a suitable pelagic counterpart to the sympagic IP\(_{25} \) remains a challenge,
and is made more difficult by the inherent limitation of the PIP\(_{25} \) index to one such
biomarker. To address these challenges, Köseoğlu et al. (2018) recently constructed a multivariate classification tree (CT) model (Breiman et al., 1984) using percentage abundances of a group of HBI biomarkers (Fig. 1) in surface sediments encompassing the full range of sea ice variability in the Barents Sea. Each surface sediment was classified into marginal (<10 % satellite SpSIC), intermediate (10–50 % SpSIC) or extensive (>50 % SpSIC) class of sea ice conditions (Fig. 2). The CT model was subsequently used to reconstruct classes of sea ice conditions throughout recent centuries in four short sediment cores collected from Barents Sea sites experiencing variable modern sea ice cover (Vare et al., 2010). For these cores, the CT class predictions were consistent with P_{III}P_{25}-derived semi-quantitative SpSIC estimates and, perhaps more importantly, the observational record of sea ice cover (Divine and Dick, 2006; Walsh et al., 2017). The CT method offered automatic selection of HBI biomarkers that achieved the highest classification rate, quantitative model evaluation via performance metrics, and independence from the $c$-factor. Thus, the CT model was ca. 92 ± 6 % accurate using $IP_{25}$ and a further tri-unsaturated HBI (hereafter HBI III (E); Fig. 1) as primary descriptive variables representing ice algal and pelagic productivity, respectively. HBIs II and III (Z) were utilised as surrogate variables, probably due to their high correlation to $IP_{25}$ and HBI III (E), respectively.

Despite this development, CT models require further evaluation before their wider applicability as a statistical tool for classification of sea ice conditions can be established. Specifically, the agreement of model-derived categorical assessment of sea ice conditions and $PIP_{25}$-derived semi-quantitative SpSIC estimates in sediment cores needs to be determined across longer timescales encompassing major changes in sea ice dynamics, such as those occurring during the Younger Dryas–Holocene transition (e.g., Ślubowska et al., 2005; Rasmussen et al., 2007; Cabedo-Sanz et al., 2013; Kristensen et al., 2013) and the Holocene cooling (e.g., Duplessy et al., 2001, 2005; Risebrobakken et al., 2010, 2011). Potential error
sources that could impact the accuracy and applicability of both methods also require
discussion, such as the possibility of differential degradation of HBIs based on their variable
degree of unsaturation (Rontani et al., 2014a, 2018a). To achieve this, in the current study we
assessed the agreement of P$_{III}$IP$_{25}$-based SpSIC estimates (Smik et al., 2016) and CT model
predictions (Köseoğlu et al., 2018) in three marine sediment cores located at sites of
contrasting contemporary sea ice conditions in the Barents Sea (Fig. 2) and spanning periods
of both abrupt and gradual climate shifts throughout the last ca. 16 cal kyr BP. The core sites
were chosen based on the availability of previous climatological reconstructions (Cabedo-
Sanz et al., 2013; Berben et al., 2014, 2017), a comprehensive suite of HBI concentrations
(Belt et al., 2015), and surface sediment-based regional calibrations for both P$_{III}$IP$_{25}$ and CT
approaches (Smik et al., 2016; Köseoğlu et al., 2018). This allowed us to readily identify the
causes of discrepancies between the two methods and provide recommendations regarding
their complementary application to downcore records.

2. Regional setting

The Barents Sea is a seasonally ice-covered shelf area of the Arctic Continental Shelf,
which experiences extensive seasonal sea ice cover in winter (October–March) and remains
almost entirely ice-free in September following the insolation-induced summer melt (May–
August; Sakshaug et al., 2009). It is also characterised by phytoplankton blooms occurring in
the highly-productive Marginal Ice Zone (MIZ) along the receding sea ice edge (e.g.,
Wassmann et al., 1999). The steep salinity and temperature gradients created through mixing
of Atlantic Water (AW) and Arctic Water (ArW) mark the position of the Polar Front (PF)
and define the winter maximum sea ice extent. The PF position is relatively stable in the
western and central Barents Sea (Loeng and Drinkwater, 2007), but becomes more variable to
the east where it is characterised by separate temperature and salinity gradients (Oziel et al.,
Most of the inter-annual sea ice variability occurs in winter and is largely dictated by the volume and temperature of inflowing AW (e.g., Loeng et al., 1997; Smedsrud et al., 2010). While most of the Atlantic-derived heat energy is lost to the atmosphere (Smedsrud et al., 2010, 2013), the volume and temperature of AW inflow have been increasing due to forcing from atmospheric circulation modes, such as the North Atlantic Oscillation (e.g., Loeng et al., 1997; Vinje, 2001; Ingvaldsen et al., 2004; Sorteberg and Kvingedal, 2006), and further amplification via positive feedback mechanisms, including the ice-albedo feedback (Smedsrud et al., 2013). Strengthening of AW inflow has been linked to the diminishing sea ice extent and increasing temperature of the Barents Sea, and continued ‘Atlantification’ of the region could influence its capacity to limit heat flux to the central Arctic Ocean and hinder sequestration of atmospheric CO$_2$ by halting deep water formation (Screen and Simmonds, 2010; Årthun et al., 2012).

3. **Materials and methods**

3.1 **Sediment material**

Three sediment cores from sites characterised by different modern sea ice conditions in the Barents Sea were selected for this study. Specific descriptions of all core locations and chronologies are available elsewhere (Ebbesen and Hald, 2004; Rüther et al., 2012; Berben et al., 2014, 2017; Belt et al., 2015). Core NP05-11-70GC (78.67°N, 32.70°E; 293 m water depth), hereafter referred to as core 70, was collected from the Olga Basin to the South of Kong Karls Land (East Svalbard) aboard the RV Lance in August of 2005. Core chronology is based on three calibrated $^{14}$C Accelerated Mass Spectrometry (AMS) dates from mixed foraminifera (Berben et al., 2017). Concentrations of IP$_{25}$ and HBI III (Z) (Fig. 1), P$_{III}$IP$_{25}$ indices, and associated semi-quantitative SpSIC estimates of core 70 for the last ca. 9.4 cal kyr BP were presented previously (Belt et al., 2015; Berben et al., 2017). Core JM09-KA11-
GC (74.87°N, 16.48°E; 345 m water depth), hereafter referred to as core 11, was obtained from the Kveithola Trough (South off Svalbard) aboard RV Jan Mayen in 2009. We use the age model spanning ca. 16 cal kyr BP presented in Belt et al. (2015) and based on merged $^{14}$C AMS dates from previous studies (Rüther et al., 2012; Berben et al., 2014).

Micropaleontological distributions, stable isotope analyses (Dylmer et al., 2013; Groot et al., 2014), IP$_{25}$ and HBI III (Z) concentrations (and P$_{III}$IP$_{25}$ values) were presented previously for core 11 (Belt et al., 2015), but not SpSIC estimates. Piston core JM99-1200 (69.27°N, 16.42°E; 475 m water depth), hereafter referred to as core 1200, was retrieved from Andfjorden (northern Norway) aboard the RV Jan Mayen in November 1999. Herein, we used the age model of Cabedo-Sanz et al. (2013) corresponding to ca. 14.0–7.0 cal kyr BP (Bølling-Allerød to middle Holocene). Concentrations of IP$_{25}$ and HBI III (Z) and P$_{III}$IP$_{25}$ values (but not P$_{III}$IP$_{25}$-derived SpSICs) of core 1200 were reported previously (Cabedo-Sanz et al., 2013; Belt et al., 2015), in addition to sedimentological, isotopic and micropaleontological analyses (Knies et al., 2003; Ebbesen and Hald, 2004). The CT model outputs for cores 70, 11 and 1200 are presented here for the first time.

3.2 Proxy and statistical methods

Modern SpSIC (April-June average for the 1988–2007 period) for each core site was inferred from the Nimbus-7 SMMR and DMSP SSM/I-SSMIS satellite dataset (Cavalieri et al., 1996) used for PIP$_{25}$ and CT model calibrations (Xiao et al., 2015a; Smik et al., 2016; Köseoğlu et al., 2018). Previously published concentrations of IP$_{25}$ and HBI III (Z) for cores 1200 (Cabedo-Sanz et al., 2013) and 11 (Belt et al., 2015) were re-examined to calculate P$_{III}$IP$_{25}$ indices (Eq. 1; Müller et al., 2011) and derive SpSIC estimates (Eq. 2) using the regional $c$-factor ($c = 0.63$) and P$_{III}$IP$_{25}$–SpSIC calibration of Smik et al. (2016), respectively. Additionally, a threshold P$_{III}$IP$_{25}$ value of 0.8, corresponding to a SpSIC of 68% (Eq. 2), was
used to indicate the occurrence of at least some (>5%) sea ice cover during the summer
months (July–September; Smik et al., 2016). Absolute concentrations (ng/g dry sed.) are
denoted by square brackets in all equations.

\[
P_{IIIIP_{25}} = \frac{[IP_{25}]}{([IP_{25}]+[HBI III (Z)] \times 0.63)} \quad \text{(Eqn. 1)}
\]

\[
SpSIC (%) = \frac{(P_{IIIIP_{25}} - 0.0692)}{0.0107} \quad \text{(Eqn. 2)}
\]

Previously obtained chromatographic and mass spectrometric (MS) data were re-
examined to quantify additional HBI lipids (viz. HBIs II and III (E)) required to obtain CT
model predictions of sea ice conditions following the method of Köseoğlu et al. (2018).

Briefly, percentage abundances of IP\textsubscript{25} and HBIs II, III (Z) and III (E) (Fig. 1) to their totals
were calculated for each core horizon from absolute concentrations (ng/g dry sed.) using Eq.

\[
HBI (%) = \frac{[HBI]}{\sum([IP_{25}],[HBI II],[HBI III (Z)],[HBI III (E)])} \times 100 \quad \text{(Eqn. 3)}
\]

Subsequently, the CT model constructed from a Barents Sea surface sediment dataset via
the R Statistical Package (R Core Team, 2017) was used to classify each core horizon into
one of three classes representing marginal (<10% satellite SpSIC), intermediate (10–50%
SpSIC), and extensive (>50% SpSIC) spring sea ice conditions. The performance metrics of
the CT model used for classification of core horizons are shown in Supplementary Table S1
(Köseoğlu et al., 2018). Biomarker concentrations, P\textsubscript{IIIIP_{25}}-derived SpSIC and CT model
outcomes for all cores are available at https://doi.pangaea.de/10.1594/PANGAEA.891102.

Additional statistical analyses were carried out to supplement the comparison of CT and
P\textsubscript{IIIIP_{25}}-based sea ice assessments. Thus, Pearson’s correlations for IP\textsubscript{25} versus HBI II and
HBI III (Z) versus HBI III (E) were calculated from surface sediment data (Fig. 2; Köseoğlu
et al., 2018). The biomarker pairings for correlation were chosen due to previous evidence of
co-production of sympagic IP\textsubscript{25} and HBI II (Navarro-Rodriguez et al., 2013; Brown et al.,
2014, Belt et al., 2016), pelagic HBIs III (Z) and III (E) (Belt et al., 2000; Rowland et al., 2001), as well as significant correlation of these biomarker pairs in the Barents Sea and other Arctic regions (Navarro-Rodriguez et al., 2013; Navarro-Rodriguez, 2014). Thus, the product-moment correlation coefficient was used as a measure of correlation to distinguish between negative and positive linear relationships and identify periods of anomalously deteriorated correlations in downcore records compared to those characteristic of relatively modern settings represented by surface sediments. Further, rolling Pearson’s correlations were calculated for all downcore records using a sampling window of nine, corresponding to a time window of between ca. 0.5–2.0 cal kyr BP. All correlation and CT model analyses were incorporated into functions within the R statistical package (R Core Team, 2017). The source code and supporting materials are available at https://doi.org/10.5281/zenodo.1346305.

4. Results and Discussion

The applicability of multivariate CT models and $P_{IIIIP_{25}}$-based semi-quantitative SpSIC estimates (Eq. 1 and 2) as complementary methods for sea ice reconstruction spanning both abrupt and gradual climate shifts of the Younger Dryas and Holocene (ca. last 13.0 cal kyr BP) was assessed by comparing the results of both approaches in cores 11, 70 and 1200 (Figs. 3–6). The assessment of consistency between the two methods was contextualised further by considering findings of previous studies (e.g., Belt et al., 2015).

4.1 Core 70 (northern Barents Sea)

The core 70 site is characterised by extensive modern sea ice conditions ($\approx$80% SpSIC) and the downcore record represents a gradual evolution of sea ice cover in the...
northern Barents Sea from ice-free conditions during the early Holocene to prolonged seasonal sea ice presence prevalent in the region today. The primarily insolation-controlled southward expansion of sea ice cover previously inferred for the core site throughout the Holocene (Belt et al., 2015; Berben et al., 2017) is reflected in the CT model assessment (Fig. 3). Consistent with the onset of the Holocene Thermal Maximum and the resulting proximity of the annual maximum sea ice edge to the core site between ca. 9.5–8.5 cal kyr BP evident from low P_{III}IP_{25}-derived SpSIC (ca. 5–15%), the CT model predicts mostly marginal sea ice conditions during this interval. Similarly, the southward migration of sea ice beginning ca. 8.5 cal kyr BP as a response to decreasing summer insolation (Berben et al., 2017) is also reflected by a switch of CT model assessment from marginal to intermediate sea ice conditions. Finally, following a period of consistent intermediate ice conditions (ca. 30 ± 4 % SpSIC), a further southward migration of the ice edge between ca. 6.5–5.9 cal kyr BP, previously attributed to further decreasing solar insolation and reduced AW influence (Berben et al., 2017), is reflected by an associated shift of CT model predictions from intermediate to extensive sea ice conditions at ca. 6.0 cal kyr BP. This trend agrees with previous reports of Neoglaciation in the Barents Sea, a period characterised by glacier advances and increased sea ice export via the Fram Strait when modern-type oceanic circulation was re-established (Werner et al., 2013, 2016; Rasmussen and Thomsen, 2015).

Overall, the timing of CT prediction shifts was consistent with previously inferred climate evolution at the core 70 site (Berben et al., 2017), and no discrepancies from semi-quantitative SpSIC estimates were observed (Fig. 3), with all horizons consistently classified within the satellite SpSIC boundaries defined in the CT model training set (Fig. 2). Our data supports the complementary application of both approaches when describing gradual changes in sea ice conditions at millennial timescales, where CT predictions have the potential to identify the timing of switches between distinct sea ice conditions, as hypothesized.
previously (Köseoğlu et al., 2018). However, we note that the CT model is limited by the class boundaries assigned to the training set (Fig. 2), which may result in insufficiently detailed assessment of changing sea ice conditions. In this case, a distinct ice expansion ca. 2.7 cal kyr BP (SpSIC values of <80%; Fig. 3b) previously linked to insolation decreases (Berben et al., 2017) was not identified by the CT model since the 50–100% SpSIC range is only represented by a single class within the training set (Fig. 2). Such limitations of the training set, in this case driven by reduced sample density around Svalbard, should be considered when interpreting model output. The otherwise high agreement with the $P_{\text{III}}$ $P_{25}$ approach is potentially attributable, at least in part, to significantly similar HBI distributions and data structure between the surface sediment training set and core 70. Like most supervised classification methods, CTs rely on distributional similarity of predictive variables (e.g., HBI percentages) between the training set and new samples to be classified. Thus, in our study, the model is only likely to function correctly when the overall relative abundance ranges and relationships between HBIs observed in the surface sediment training set (Fig. 2) are reproduced in downcore records. This is the case for core 70, where consistently significant positive correlations are observed for $P_{25}$ versus HBI II and HBI III (Z) versus HBI III (E) biomarker pairs (Fig. 3c), also evident in surface sediments with different overlying SpSIC (Fig. 7) used to build the CT model. Relative HBI abundances (Eq. 3) are also comparable and generally dominated by sympagic biomarkers in both datasets (Fig. 8); although we stress that inherent HBI variability in surface sediments and core 70 prevent any detailed interpretation of spatially and temporally averaged values. Nonetheless, such visualisations of data ranges and structure help inform the expected CT performance for a given dataset and suggest that the model is likely to perform well for core 70.

4.2 Core 1200 (southwestern Barents Sea)
In contrast to contemporary ice-free conditions that characterise the core 1200 site, harsh glacial conditions with short ice-free summers during the majority of the Younger Dryas cold stadial spanning ca.13.0–11.9 cal kyr BP were previously inferred (Cabedo-Sanz et al., 2013; Belt et al., 2015). Accordingly, our records show elevated $P_{III25}$-derived SpSIC values (>75%) accompanied by CT model predictions of extensive sea ice conditions during this period (Fig. 4b–c). Subsequently, ameliorated conditions are evident during the Younger Dryas–Holocene transition, with precipitous decrease of SpSIC estimates (to ca. 10–40%) and a switch of CT model assessment from extensive to marginal (<10% SpSIC) sea ice conditions. In core 1200, the ice retreat is characterised by rapid fluctuations of SpSIC estimates, consistent with switching of CT model assessment between intermediate and marginal classification of sea ice cover during the 11.9–11.5 cal kyr BP period, which is followed by ice-free conditions for the remainder of the record (Fig. 4b). Similar unstable conditions, likely attributable to the return of thermohaline circulation during this period (e.g., Bakke et al., 2009), were previously inferred for core 1200 from PIP$_{25}$ records (Cabedo-Sanz et al., 2013; Belt et al., 2015), as well as sea surface temperature (SST) and sea surface salinity (SSS) reconstructions based on stable isotope measurements of planktic foraminifera (Ebbesen and Hald, 2004).

The apparent consistency of CT predictions and $P_{III25}$-derived SpSIC in core 1200 suggests that both methods respond similarly to extremes of sea ice conditions observed both during (ca. 13.0–12.0 cal kyr BP) and after (11.5 cal kyr BP onwards) the Younger Dryas stadial. However, while the CT model also detected rapid fluctuations of sea ice cover during the climatically unstable YD–Holocene transition (11.9–11.5 cal kyr BP), several horizons ($n = 7$) were classified outside of the categorical sea ice boundaries based on satellite SpSIC thresholds (Fig. 2 and 3b). Nonetheless, such discrepancies with the $P_{III25}$-based SpSIC record were always within the RMSE of the regional $P_{III25}$–SpSIC calibration (ca. ±11%.
SpSIC; Smik et al., 2016), and are potentially attributable to various error sources associated
with $\text{P}_{\text{III}}$IP$_{25}$-based SpSIC estimates and CT models. For example, the dependence of $\text{P}_{\text{III}}$IP$_{25}$
values on the $c$-factor may significantly influence the regional comparability of the resulting
SpSIC estimates. While the general trends of $\text{P}_{\text{III}}$IP$_{25}$-derived SpSIC were previously shown
to be unaffected by the magnitude of the $c$-factor (Belt et al., 2015; Smik et al., 2016), $\text{P}_{\text{III}}$IP$_{25}$
values (and associated SpSIC estimates; Eq. 1 and 2) may vary by ca. 10% when omitting the
$c$-factor ($c = 1$; Smik et al., 2016). Further, the $c$-factor used in the current study ($c = 0.63$;
Eq. 1) was calculated based on average IP$_{25}$ and HBI III (Z) concentrations from a wide range
of Barents Sea locations characterised by different sea ice conditions (Smik et al., 2016).
Thus, the regional applicability of a spatially averaged $c$-factor remains a challenge, and the
same value may not provide fully comparable SpSIC estimates for all downcore locations in
our study. In contrast, the CT approach is based on a multivariate set of HBI biomarkers, is
therefore independent of the $c$-factor, and probably provides results that are more comparable
between locations within the geographical coverage of the surface sediment dataset used for
model training (Köseoğlu et al., 2018). Together, these caveats imply that the interpretation
of $\text{P}_{\text{III}}$IP$_{25}$-derived SpSIC variability and any discrepancies with CT model predictions within
the associated RMSE (±11 % SpSIC) should be avoided, and broader changes beyond this
time period of high misclassification spanning 11.9–11.5 cal kyr BP in core 1200 was
characterised by the return of enhanced biogenic production (Knies, 2005), fluctuating AW
inflow, and similarly variable sea ice conditions resulting from meltwater and nutrient input
from waning ice sheets (Cabedo-Sanz et al., 2013). It is possible that CT performance
suffered during this interval of significantly unstable sea ice cover and primary productivity
regimes, of which the latter could potentially have contributed to degraded correlations
observed in our record (Fig. 4c) due to inconsistent, variable HBI production. Overall, our
results suggest that CT predictions potentially become more prone to misclassification, and
therefore less consistent with semi-quantitative SpSIC estimates, in rapidly shifting climate
conditions observed during stadial-interstadial transitions. Finally, misclassification errors
may also be associated with information loss due to insufficient representation of the SpSIC
range (0–100%) in the CT model. Specifically, the surface sediment dataset used for model
construction contains no samples with modern overlying SpSIC of 16–22% and 56–67%
(Köseoğlu et al., 2018), potentially resulting in an incomplete model definition near the
marginal-intermediate (10% satellite SpSIC) and intermediate-extensive (50% satellite
SpSIC) sea ice class boundaries, respectively. Such potential error sources may be mitigated
by expansion of the model training set to increase sample density and include such under-
represented SpSIC ranges. In the meantime, we suggest that shifts in CT model class
predictions should be interpreted as broader changes between sea ice regimes, rather than
between definitive SpSIC threshold values. Thus, the marginal, intermediate, and extensive
sea ice classes included in the model (Fig. 2) likely represent ice-free or proximal maximum
ice edge conditions, the highly-productive MIZ during the spring melt season, and more
northern regions where ice cover persists until ca. August–September (e.g., North-East of
Svalbard), respectively.

4.3 Core 11 (western Barents Sea)
The core 11 site in our dataset is presently characterised by marginal SpSIC (<5%) with a proximal spring sea ice edge (Fig. 2). As seen for core 1200, the site experienced SpSIC values of ca. 80% during the Younger Dryas (ca. 13.0–12.0 cal kyr BP), but exhibited a more gradual ice retreat and a step-wise switch of CT model predictions from extensive towards marginal sea ice conditions during the 12.0–11.5 cal kyr BP period (Fig. 5b). A period of highly unstable sea ice cover is instead observed between ca. 11.0–10.0 cal kyr BP, with fluctuating SpSIC estimates (ca. 0–65%) and CT model predictions. This is consistent with the return of enhanced sub-surface AW inflow to the core site after ca. 11.5 cal kyr BP inferred from benthic foraminiferal census data (Groot et al., 2014), with a contrastingly colder surface water layer dominated by ArW inferred from reduced SSTs (Berben et al., 2014). From ca. 10.0–1.5 cal kyr BP, ice-free conditions characterised the core 11 site, as evidenced by consistently low SpSIC (ca. <10%) and marginal sea ice conditions predicted by the CT model, and further supported by an enhancement of AW inflow to the core site from ca. 9.8 cal kyr BP (Groot et al., 2014). Finally, re-emergence of highly fluctuating sea ice cover during the last ca. 0.9–0.6 cal kyr BP (Berben et al., 2014; Belt et al., 2015), despite increasing AW inflow (Dylmer et al., 2013), was also captured by the CT model, which switches from marginal to intermediate sea ice conditions at this time (Fig. 5b).

Consistent with outcomes from core 1200, several horizons \(n=20\) from core 11 were classified outside of the sea ice class boundaries based on satellite SpSIC thresholds. Notably, these differences also exceeded the standard RMSE (±11% SpSIC) of the regional P_{IIIIP25}-SpSIC calibration (Eq. 2; Smik et al., 2016) for four horizons, where marginal sea ice cover (<10% satellite SpSIC) was inferred for P_{IIIIP25}-derived SpSIC values in excess of ca. 40% (Fig. 5b). Most significant misclassification was observed during periods of high climatic variability (11.5–10.0 and 0.9–0.6 cal kyr BP), consistent with the return of variable sub-surface AW inflow. As with core 1200, we suggest that this stems from considerably
different distributions and data structure of HBIs in the misclassified horizons compared to those of surface sediments used for model construction, potentially caused by climate fluctuations during intervals of rapid climate change. Indeed, running correlations between IP$_{25}$ and HBI II severely degrade towards negative $r$ values during the 11.5–10.0 cal kyr BP interval and the last 0.7 cal kyr BP (Fig. 5c), thus deviating from the consistently high positive associations observed in the surface training set (Fig. 7). Indeed, these distributional changes coincide with CT misclassification beyond the $P_{III}$IP$_{25}$-based SpSIC error of 11% and are better illustrated when considering individual concentration profiles of IP$_{25}$ and HBI II, as well as rolling correlations of corresponding relative differences (Fig. 6). It is evident that disproportional increases of HBI II relative to IP$_{25}$ contribute to the correlation reduction. Similar increases in the HBI II/IP$_{25}$ ratio were previously observed across the Arctic during periods of increased warm water inflow from the North Atlantic and North Pacific (e.g., Fahl and Stein, 2012; Hörner et al., 2016; Ruan et al., 2017), conditions that also characterised the core 11 site during intervals of elevated HBI II concentration (Berben et al., 2014; Belt et al., 2015). Overall, our data support the results from core 1200 and suggest that CT performance and consistency with the $P_{III}$IP$_{25}$ approach suffer when HBI distributions included in the model training set are not represented in downcore records characterised by different or unstable climate. Visualisation of the variables used in the CT assessment (Fig. 5 and 6) is essential when identifying such cases. It is important to acknowledge that this limitation is potentially amplified as variations in relative abundances of any biomarker included in the model inherently affect the overall HBI composition due to data normalisation used for the CT (Eq. 3). The dependence of CT performance and viability on the consistency of data distribution in the training set with that of new samples highlights the necessity of constructing separate training sets for different Arctic regions, which often exhibit significantly different HBI distributions despite similar seasonal ice conditions (Stoynova et
al., 2013; Xiao et al., 2015a). Ideally, downcore records to be classified should be within the geographical coverage of the surface sediment training set, as is the case in our study. Further, the choice of an optimal time interval for the satellite SpSIC data is potentially problematic due to the often unavailability of accumulation rates for surface sediments. Thus, the integrated biomarker signal at each surface location potentially corresponds to a variable temporal window (Köseoğlu et al., 2018) and some surface sediments may not even represent recent accumulation. This is likely to influence the accuracy of both the CT model and P₃IP₂₅-based SpSIC estimates.

Another important consideration is the consistent quantification of all biomarkers between the training set and new samples. Accurate quantification of HBIs via mass spectrometric techniques involves the use of an instrumental Response Factor (RF), usually obtained from calibration with authentic standards, to account for mass spectral fragmentation efficiency differences between individual biomarkers and the internal standard (Belt et al., 2012; Belt et al., 2014). Moreover, HBIs usually exhibit vastly different RF values (Belt et al., 2014), necessitating instrument calibration via separate standard series for each biomarker and subsequent quality monitoring using a reference sediment material of known HBI concentration. Any RF changes thus affect the HBI distribution and resulting CT model rules, such that the use of different quantification methods (RF values) for the model training set and new samples will cause the model to fail when classifying the latter. For our data, assigning the same value to all RFs shifts the HBI composition towards higher relative abundances of HBI III (Z) and HBI III (E) (Fig. 8a). When these modified downcore distributions are classified using the CT model trained with correctly quantified surface sediments (Köseoğlu et al., 2018), the model fails to identify the extensive sea ice class completely and exhibits a high discrepancy with P₃IP₂₅-based SpSIC. On the other hand, when consistent methods are used to obtain the RFs, model performance is largely unaffected.
(Fig. 8b). Directly comparable quantification is therefore necessary for the training and new sample sets, and the CT must be re-built with a new training set should a change in quantification methods occur.

Finally, the relatively high susceptibility of more unsaturated HBIs, particularly those with trisubstituted double bonds (HBIs III (Z) and III (E) in this case), towards degradative processes could also alter biomarker distributions in downcore sedimentary sequences relative to those in surface sediments. The lower stability of HBI trienes towards photodegradation and autoxidation in sea ice and the water column (Rontani et al., 2014a,b) possibly implies their increased potential for aerobic degradation in upper oxic sediments, which was shown recently to affect even the more diagenetically stable IP$_{25}$ (Rontani et al., 2018a,b). Thus, selective removal of HBIs III (Z) and III (E) from the HBI distribution may adversely affect P$_{III}$IP$_{25}$-based SpSIC estimates and CT model performance, especially under conditions of high light penetration, long residence times of algal cells in the photic zone, and low sedimentation rates, where diagenetic processes are more likely to have an effect. While it is not feasible to analytically diagnose the relative impacts of climate change and selective HBI degradation due to the extremely high reactivity of associated photo- and oxidation products (Rontani et al., 2014a,b), examination of HBI triene concentration profiles suggests a prevailing influence of climate on our data. Specifically, the concentrations of pelagic HBI III (Z) reach and surpass those of IP$_{25}$ (Fig. 3a, 4a, and 5a) during periods of reduced sea ice cover and generally ameliorated climate conditions inferred in previous studies, while reduced concentrations only coincide with harsh glacial conditions of the Younger Dryas and the late Holocene ice expansion (Cabedo-Sanz et al., 2013; Berben et al., 2014, 2017). Moreover, downcore concentrations of HBIs III (Z) and III (E) are often higher than maximum values observed in surface sediments from the highly-productive MIZ, which are ca. 40 ng/g and 20 ng/g for HBIs III (Z) and III (E), respectively (Köseoğlu et al., 2018). This
suggests that, in this case, sedimentary aerobic degradation or other processes prior to deposition are unlikely to significantly alter downcore HBI content relative to that of proximal surface sediments. Finally, as noted previously, the alteration of HBI distributions and data structure due to disproportional and even opposing concentration increases of II relative to those of IP$_{25}$ is the likely cause of discrepancies between P$_{III}$IP$_{25}$- and CT-based methods for cores 1200 and 11, in particular (Fig. 4–6). Nonetheless, a diagenetic influence on downcore HBI concentrations cannot be discounted, particularly in older core sections or when overlying climate conditions are more likely to promote accelerated or prolonged oxidation and photodegradation. Thus, we suggest that HBI distributions should be combined with degradation proxies, such as the recently utilized ratio of brassicasterol to 24-methylenecolesterol (Rontani et al., 2018a), and that uncharacteristically low concentrations of HBIs III (Z) and III (E) relative to otherwise inferred climate conditions (e.g., using other proxies) should be interpreted with caution.

5. Conclusions

Downcore records encompassing different modern sea ice conditions in the Barents Sea were used to assess the spatio-temporal consistency between CT model predictions and P$_{III}$IP$_{25}$-based SpSIC estimates. A good overall agreement between both approaches was observed for all cores, and the CT model was able to capture both abrupt and fluctuating shifts in sea ice regimes, such as those evident during the Younger Dryas stadial, as well as more gradual trends in sea ice conditions during the Holocene. However, shifts of CT model predictions occurred at variable threshold values of P$_{III}$IP$_{25}$-based SpSIC estimates in different downcore records (ca. ±11% for 16 samples, >11% for 4 samples). This variability was attributed partially to the occurrence of downcore HBI distributions, which are not
represented in the model training dataset, most notably during intervals of unstable and rapid
climate change characterising stadial-interstadial transitions. While it is not feasible to avoid
this limitation, examination of data structure and distribution may pinpoint intervals where
CT performance is likely to decrease. A consequence of CT dependency on HBI
distributional changes is the necessity to use consistent quantification methods for model
training and new (downcore) samples, and to use separate training sets for different Arctic
regions. Selective removal of more unsaturated HBIs via degradation processes represents
another potential error source, although this was likely not the case for our data. Further
potential error sources of both methods were also identified, including the uncertain regional
applicability of a uniform \( c \)-factor and insufficient sample density for representation of the
entire SpSIC range (0–100%) within the CT model. We suggest, therefore, that only
variations of \( \text{P}_{10} \text{I}_{25} \)-based SpSIC exceeding the associated RMSE of 11% be considered
significant, and that CT model predictions should be interpreted in terms of broader changes
in sea ice regimes (i.e. open water or proximal ice edge, MIZ conditions, and stable sea ice
cover) rather than inflexible satellite-based numeric SpSIC thresholds (i.e. 10% and 50%
SpSIC).

Acknowledgments

This work was jointly supported by the University of Plymouth and the Research Council of
Norway through its Centre of Excellence funding scheme for CAGE, (project 223259). We
thank Yunping Xu and an anonymous reviewer for providing valuable comments that
improved the overall quality of the manuscript.
6. References

Andrews, J.T., Belt, S.T., Olafsdottir, S., Massé, G., Vare, L.L., 2009. Sea ice and marine climate variability for NW Iceland/Denmark Strait over the last 2000 cal. yr BP. The Holocene 19, 775-784.


Hörner, T., Stein, R., Fahl, K., Birgel, D., 2016. Post-glacial variability of sea ice cover, river run-off and biological production in the western Laptev Sea (Arctic Ocean) – A high-resolution biomarker study. Quaternary Science Reviews 143, 133-149.


Lindsay, R., Schweiger, A., 2015. Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations. The Cryosphere 9, 269-283.


Weckström, K., Massé, G., Collins, L.G., Hanhijärvi, S., Bouloubassi, I., Sicre, M.-A., Seidenkrantz, M.-S., Schmidt, S., Andersen, T.J., Andersen, M.L., Hill, B., Kuijpers,


Figure Legends
Figure 1. Structures of $C_{25}$ Highly-Branched Isoprenoid (HBI) biomarkers used for sea ice reconstruction in the current study.

Figure 2. Map of the Barents Sea showing locations of downcore records 1200, 11, and 70 (labelled black squares). The maximum April–June sea ice edge for the 1988–2007 period corresponding to an SpSIC threshold of 0% is shown by a solid black line. Circles correspond to surface sediments used for CT model training, with overlying marginal (<10% satellite SpSIC; red markers), intermediate (10–50% SpSIC; yellow markers), and extensive (>50% SpSIC; green markers) sea ice cover. A simplified representation of the major AW surface currents is illustrated by red arrows, with abbreviations for: WSC – West Spitsbergen Current; NCaC – North Cape Current.

Figure 3. Proxy data and correlations for core 70 (northern Barents Sea): (a) Absolute concentrations of $IP_{25}$ (black line with circle markers) and HBI III (Z; green line); (b) $P_{III}IP_{25}$-based % SpSIC profile (black line) with an RMSE error of ca. 11% (thin black lines; Smik et al., 2016; Köseoğlu et al., 2018) and superimposed CT predictions of marginal (red circles), intermediate (yellow squares), and extensive (green diamonds) sea ice cover. Dashed horizontal lines represent the satellite SpSIC boundaries used to separate the sea ice classes in the CT model (Köseoğlu et al., 2018), where 10% and 50% satellite SpSIC thresholds correspond to the marginal-intermediate and intermediate-extensive boundaries, respectively; (c) Running Pearson’s correlations for $IP_{25}$ versus HBI II (black line with circle markers) and HBI III (Z) versus HBI III (E; red line) with a sampling window of 9 core horizons. Grey diamonds represent correlations significant at $p = 0.05$, and the dashed line separates positive ($r > 0$) and negative ($r < 0$) correlations.
Figure 4. Proxy data and correlations for core 1200 (southwestern Barents Sea). The illustrated profiles are analogous to those shown in Fig. 3. The light blue vertical bar highlights the Younger Dryas stadial (12.95–11.70 cal kyr BP).

Figure 5. Proxy data and correlations for core 11 (western Barents Sea). The illustrated profiles are analogous to those shown in Figs. 3 and 4. In (b), crossed red squares represent horizons where the discrepancy between CT predictions and P_{III}IP_{25}-based SpSIC estimates exceeded the P_{III}IP_{25}–SpSIC calibration RMSE of ca. 11% (Smik et al., 2016). The light blue vertical bar highlights the Younger Dryas stadial (12.95–11.70 cal kyr BP).

Figure 6. Biomarker profiles and correlations for core 11: (a) Concentrations of IP_{25} and HBI II represented by black (with circle markers) and green lines, respectively; (b) Running correlation of first relative (%) differences, with a sampling window of 9 horizons. Positive and negative correlations are separated by a dashed horizontal line, while grey diamonds show correlations significant at a 95% confidence level. In both (a) and (b), crossed red squares represent samples for which CT model predictions significantly differed from P_{III}IP_{25}-derived % SpSIC values, with a discrepancy exceeding the RMSE of the P_{III}IP_{25}–SpSIC calibration (ca. ±11% SpSIC). The light blue vertical bar highlights the Younger Dryas stadial (12.95–11.70 cal kyr BP).

Figure 7. Pearson’s correlations of IP_{25} versus HBI II (upper panel series A) and HBI III (Z) versus HBI III (E) (lower panel series B) in Barents Sea surface sediments with marginal (<10% SpSIC; red circles), intermediate (10–50% SpSIC; yellow triangles), and extensive (>50% SpSIC; green squares) overlying sea ice conditions. Surface sediment biomarker data was taken from Köseoğlu et al. (2018).

Figure 8. Biomarker composition and CT model output for cores 70, 1200, and 11 using different RF combinations: (a) Averaged relative abundances of IP_{25} and HBIs II, III (Z) and
III (E) in surface sediments and cores using calibrated and arbitrarily equated (uncalibrated) RFs; (b) Percentage distribution of core horizons classified into three categories of sea ice conditions by the CT model. “Correct RFs” and “incorrect RFs” indicate the respective use of calibrated RFs and arbitrarily equated RFs for both the CT training set (surface sediments) and cores. “Mismatched RFs” represents the use of calibrated and uncalibrated RFs for the training set and downcore records, respectively. Values within white circles correspond to the number of horizons classified into a given category of sea ice conditions where a discrepancy with $P_{III IP_{25}}$-based SpSIC estimates of at least 1% was observed. No samples were misclassified into the extensive (>50% SpSIC) sea ice category.
Figure 1

Figure: "Disclaimer: This is a pre-publication version. Readers are recommended to consult the full published version for accuracy and citation."
Figure 6

(a) HBI II (ng g⁻¹ sed.) and IP 25 (ng g⁻¹ sed.) over time (cal ka BP).

(b) Pearson's r (relative differences) with age (cal ka BP).

Disclaimer: This is a pre-publication version. Readers are recommended to consult the full published version for accuracy and citation.
Figure 7

A

\[ y = 0.1273x + 1.3412 \]

\[ r = 0.950 \]

HBI II (ngg \(^{-1}\) sed.)

B

\[ y = 0.1822x + 0.355 \]

\[ r = 0.960 \]

HBI II (ngg \(^{-1}\) sed.)

\[ y = 0.244x + 0.9323 \]

\[ r = 0.965 \]

HBI II (ngg \(^{-1}\) sed.)

\[ y = 1.9736x - 1.4367 \]

\[ r = 0.929 \]

HBI III (Z) (ngg \(^{-1}\) sed.)

\[ y = 1.1333x + 0.0229 \]

\[ r = 0.996 \]

HBI III (E) (ngg \(^{-1}\) sed.)

\[ y = 1.8227x - 0.2983 \]

\[ r = 0.980 \]

HBI III (E) (ngg \(^{-1}\) sed.)

\[ y = 1.333x + 0.0229 \]

\[ r = 0.996 \]

HBI III (E) (ngg \(^{-1}\) sed.)

*Disclaimer: This is a pre-publication version. Readers are recommended to consult the full published version for accuracy and citation.*
Figure 8

(a) HBI relative abundance (%)

Correct RFs (calibrated) vs. Incorrect RFs (all equal to 1)

(b) CT-classified samples (%)

Cores (correct RFs) vs. Cores (incorrect RFs) vs. Cores (mismatched RFs)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>5</th>
<th>11</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>1</td>
<td>6</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>1200</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>26</td>
<td>18</td>
<td>15</td>
</tr>
</tbody>
</table>

Legend:
- Red: Marginal (<10% SpSIC)
- Yellow: Intermediate (10–50% SpSIC)
- Green: Extensive (>50% SpSIC)

Figure disclaimer:
"Disclaimer: This is a pre-publication version. Readers are recommended to consult the full published version for accuracy and citation."