Quarterly Journal of the Royal Meteorological Society A journal of the atmospheric sciences and physical oceanography



Complex Systems Modelling for Statistical Forecasting of Winter North Atlantic Atmospheric Variability: a New Approach

Journal:	QJRMS
Manuscript ID	QJ-18-0184.R3
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Hall, Richard J; University of Lincoln, UK, Geography Wei, Hua-Liang; University of Sheffield, Automatic Control and Systems Engineering Hanna, Edward; University of Lincoln, UK, Geography
Keywords:	NAO, seasonal forecast, NARMAX, predictability, jet stream, North Atlantic, winter
Country Keywords:	AAA - No country



2	
3	Complex Systems Modelling for Statistical Forecasting of Winter North Atlantic
4	Complex Systems foroutining for Statistical Forocusting of Whiter Foron Prantie
5	
6	Atmospheric variability: a New Approach
7	
8	Richard J. Hall ¹ , Hua-Liang Wei ² , Edward Hanna ¹
9	
10	
11	
12	¹ School of Geography and Lincoln Centre for Water and Planetary Health
13	School of Geography and Effeorit Centre for water and Flanciary freatur,
14	
15	University of Lincoln,
16	
17	Think Tank, Ruston Way,
18	
19	Lincoln LN6 7FL, UK
20	
21	
22	
23	² Department of Automatic Control and Systems Engineering
24	Department of Automatic Control and Systems Engineering,
25	
20	University of Sheffield,
28	
29	Mappin Street, Sheffield S1 3JD, UK
30	
31	Running Head: North Atlantic Seasonal Forecasting
32	
33	
34	
35	Corresponding outhors
36	Corresponding author:
37	
38	Richard Hall
39	
40	School of Geography and Lincoln Centre for Water and Planetary Health
41	
42	Think Tank
43	
44	Ruston Way
45	Ruston () uy
40	Lincoln I NG 7EI
47	
40	
50	1 el: 01522 835834
51	
52	Email: rihall@lincoln.ac.uk
53	
54	
55	
56	
57	
50	

Abstract

Seasonal forecasts of winter North Atlantic atmospheric variability have until recently shown little skill. Here we present a new technique for developing both linear and non-linear statistical forecasts of the winter North Atlantic Oscillation (NAO) based on complex systems modelling, which has been widely used in a range of fields, but generally not in climate research. Our polynomial NARMAX models demonstrate considerable skill in out-of-sample forecasts and their performance is superior to that of linear models, albeit with small sample sizes. Predictors can be readily identified and this has the potential to inform the next generation of dynamical models and models allow for the incorporation of non-linearities in interactions between predictors and atmospheric variability. In general there is more skill in forecasts developed over a shorter training period from 1980 compared with an equivalent forecast using training data from 1956. This latter point may relate to decreased inherent predictability in the period 1955-1980, a wider range of available predictors since 1980 and/or reduced data quality in the earlier period and is consistent with previously identified decadal variability of the NAO. A number of predictors such as sea-level pressure over the Barents Sea, and a clear tropical signal are commonly selected by both linear and polynomial NARMAX models. Tropical signals are modulated by higher latitude boundary conditions. Both approaches can be extended to developing probabilistic forecasts and to other seasons and indices of atmospheric variability such as the East Atlantic pattern and jet stream metrics.

Key Words: NAO, seasonal forecast, NARMAX, predictability, jet stream, North Atlantic, winter

30 1. Introduction

Winter North Atlantic (NA) atmospheric variability is dominated by the North Atlantic Oscillation (NAO; Hurrell and Deser, 2009; Hanna and Cropper, 2017). The NAO index gives a measure of the pressure difference between semi-permanent high pressure over the Azores and a semi-permanent low pressure over Iceland. There is a see-saw of atmospheric mass between these two nodes. The greater (smaller) the pressure difference, the more positive (negative) the NAO index. A positive NAO is associated with mild, wet and often stormy winters over northwestern Europe while a negative NAO is linked with cold, dry conditions in this region, but with wetter weather in the Mediterranean (Xoplaki et al., 2004).

The NAO can be regarded as arising from storm-track and jet-stream variability (Vallis and Gerber, 2008; Stendel et al., 2016), and is an indicator of the zonality of the atmospheric flow. A positive winter NAO tends to arise when the tropospheric jet and storm track are shifted further northwards, driving storms towards western Europe, with a more zonal jet stream, whereas a negative NAO indicates a southerly displacement with an increased meridonal jetstream component (Stendel et al., 2016) which can steer storms towards the Mediterranean and enable cold air outbreaks from the Arctic to lower latitudes. The NAO is more closely associated with shifts in jet latitude than it is with jet speed variability (Woollings et al., 2010a).

The NAO is a mode of internal atmospheric variability in idealised modelling experiments (e.g. James and James, 1989) and until recently it was considered that NAO variability on intraseasonal and interannual timescales was a result of internal atmospheric variability, or climate noise (e.g. Feldstein 2000) and largely unpredictable (e.g. Johansson, 2007; Kim et al., 2012). However, there is recent evidence that the storm track and jet stream are subject to forcing from slowly varying boundary conditions such as ocean temperatures and sea-ice

changes, together with solar variability and influences from the winter stratosphere (for a
review see Hall et al., 2015).

A number of recent studies indicates significant potential for winter seasonal forecasting in the NA region based on the influence of slowly-varying boundary conditions (Scaife et al., 2014; Riddle et al., 2013; Kang et al., 2014; Dunstone et al., 2016; Stockdale et al., 2015). There are also a number of older studies which identify significant skill in seasonal forecasting of the winter NAO using climate models (Palmer et al., 2004; Müller et al., 2005; Derome et al., 2005) and empirical approaches (Fletcher and Saunders, 2006). Scaife et al. (2014) report a correlation skill of 0.62 for hindcasts of the winter NAO over the period 1993-2012, based on the UK Met Office GloSea5 seasonal forecasting system (MacLachlan et al., 2015). Statistical forecasts are quick and cheap to implement (e.g. Cohen et al., 2018) and therefore complement the dynamical forecasts. They allow for identification of sources of potential predictability and may help to explain particular instances of poor forecasts in dynamical NWP models and inform their future development. Recent studies (Dunstone et al., 2016; Wang et al. 2017, Hall et al., 2017) have shown promising skill in predicting the winter NAO using a linear statistical framework and Folland et al. (2012) produced skillful forecasts of winter European temperatures based on a number of these factors. However, these studies only use linear combinations of predictors, not considering non-linear, cross-product and interaction terms. Often statistical models can be constructed for a training period with a very good fit, subsequently failing when making out-of-sample forecasting, due to non-stationary relationships, internal variability and overfitting in the training period.

 78 The slowly varying boundary conditions may act to reinforce or oppose one another and79 numerous studies examine remote causes of NA atmospheric variability, mostly in a linear

framework. However, the interaction of the atmosphere with boundary forcing can be non-linear (Petoukhov and Semenov, 2010), so a purely linear approach may only capture a limited portion of the variability. Here we further develop statistical seasonal forecasting by using a novel application of NARMAX (Non-linear Auto-Regressive Moving Average with eXogenous inputs) methodology (e.g. Billings, 2013), comparing linear and polynomial regression-based forecasting models. We aim to investigate whether the inclusion of non-linear interactions help to explain changes in the NAO. NARMAX modelling can reveal and characterize non-linear dynamic relationships among signals from recorded data and produces transparent models which demonstrate how a response variable (system output signal) is linked to a number of candidate explanatory variables (system input signals) and their combined interactions. The NARMAX approach will construct the simplest model to explain the system: therefore if a linear model provides a good representation of the system, the NARMAX method will go no further (Billings, 2013, p9). NARMAX modelling was first introduced to solve non-linear dynamical system identification and modelling problems in engineering, and it has been successful in revealing linear and non-linear relationships at a wide range of scales within the engineering, biological, ecological, medical, geophysical, and environmental sciences (e.g. Billings, 2013; Bigg et al., 2014; Ayala-Solares et al., 2018).

Here we review some of the drivers of North Atlantic climate variability identified in previous
research. A number of studies identify a sea-ice influence, particularly from the Barents-Kara
Sea (BK) region (e.g. Koenigk et al., 2016; Garcia-Serrano et al., 2017). A likely pathway of
influence is due to constructive interference of the atmospheric warming related to autumn seaice loss with climatological planetary wave patterns (Screen et al., 2018, Wu and Smith, 2016).
The location of the BK region is close to the climatological ridge of the zonal wave-1 and
wave-2 planetary waves, with localised warming acting to reinforce this pattern (Zhang et al.,

2018) which enhances vertical wave propagation. This then can weaken the stratospheric polar vortex with a subsequent downward propagation of this signal over a number of weeks (Baldwin and Dunkerton, 2001). A stronger (weaker) vortex is associated with a positive (negative) NAO, as a strengthened (weakened) circumpolar stratospheric jet induces a poleward (equatorward) shift in the tropospheric jet and storm tracks (Kidston et al., 2015). A cryospheric influence has also been detected from Siberian snow anomalies, which may enhance the Siberian high-pressure region, resulting in vertical wave propagation into the stratosphere and a weakening of the stratospheric vortex (Cohen et al., 2007). This effect has been observed in models, but can be weak, and may require modulation by the Quasi-biennial Oscillation (QBO) to be more effective (e.g. Tyrrell et al. 2018). The strength of the vortex can also be perturbed by factors such as the QBO phase (Boer and Hamilton 2008), the solar cycle (Ineson et al., 2011) and tropical volcanic eruptions (Robock and Mao, 1995; Driscoll et al., 2012).

Sea-surface temperatures (SST) are an important element of boundary layer forcing. The interaction between atmosphere and ocean is complex, with the atmospheric variability characterised by the NAO forcing SST variability in the NA, to produce the distinctive tripole pattern of SSTs (Deser et al., 2010). However, there is evidence for feedback of this SST pattern to the atmosphere at time lags of a few months, as the spring tripole anomalies are preserved beneath the summer mixed layer, re-emerging in winter as the mixed layer deepens (Rodwell et al., 1999, Deser et al., 2003; Czaja and Frankignoul, 1999). A complementary SST pattern is the North Atlantic Horseshoe (NAH; Czaja and Frankignoul, 2002), which may evolve from the tripole anomalies and where SST anomalies may lead the NAO by up to six months. These patterns fluctuate on a decadal scale, but Atlantic SSTs also experience

multidecadal variability known as the Atlantic Multidecadal Oscillation (AMO, e.g. Enfield etal., 2001) with warm and cool phases and a period of 65-80 years.

Other recently identified associations involve SSTs in the BK and Greenland-Norwegian (GIN) Seas (Kolstad and Årthun, 2018). In addition, sea-level pressure (SLP) in the BK region can precondition autumn sea-ice extent, and so itself may be a potential predictor of European winter weather variability (King and Garcia-Serrano, 2016). Furthermore, the temperature variability of the NA subpolar gyre (SPG) has been associated with changes in jet speed. A weakened gyre circulation can lead to increased poleward transport of warmer subtropical waters, and a decrease in meridional temperature gradient (Häkkinen et al., 2011; Woollings et al., 2018).

The role of solar fluctuations in North Atlantic climate variability is a subject of considerable
debate. A solar cycle signal has been detected in European winters (Lockwood et al., 2010;
Woollings et al., 2010b), with lower solar activity associated with colder European winters.
Recent studies suggest that there is a lag of 3-5 years between the solar signal and its impact
on the atmosphere, measured by the NAO (Scaife et al., 2013; Gray et al., 2013; 2016; Andrews
et al., 2015), possibly as a result of integration of the solar signal over time by SSTs.

The influence of tropical teleconnections is also evident. The El-Niño-Southern Oscillation (ENSO) signal can propagate via troposphere and stratosphere (Toniazzo and Scaife, 2006; Bell et al., 2009) and there is evidence of non-linearity, with stronger El Niño events not having the NAO-like impact of moderate events (Folland et al., 2012; Rao and Ren, 2016a; 2016b). This may be a consequence of a more eastward centre of action for stronger events (Takahashi and Dewitte, 2015) with a different pathway of propagation. In addition, links between Indian

Ocean SST anomalies and the NAO, (Hoerling et al., 2004; Li et al., 2010) with the NAO lagging SST anomalies by a month or more and between the NAO and the Madden-Julian Oscillation (Garfinkel et al., 2014 Tseng et al., 2018) have also been identified. Yu and Lin (2016) find an NAO response to tropical heating anomalies in the Indian and Atlantic regions, although not necessarily to SSTs.

Influences from these boundary conditions at a range of lead times means there could be a significant component of predictability within the winter NA atmospheric circulation (Smith et al., 2016). As the NAO is such a significant factor in determining the winter weather around the Atlantic basin, skillful seasonal forecasts of the NAO will have considerable economic and societal benefits. The NAO is related to hydrological outlooks and flood risk (e.g. Svensson et al., 2015; Bell et al., 2017) and is significantly related to energy demand (Thornton et al., 2017; Clark et al., 2017).

Section 2 presents the data used, and section 3 explains the methods, including how the
NARMAX methodology is applied. Results are presented in section 4 and interpreted in section
5. This is followed by some concluding remarks in section 6.

² 171

5 172 **2. Data**

Data used are summarised in Table I with additional information in Table S1. HadISST1 (Rayner et al., 2003) is used for SST-based and sea-ice variables. We select monthly predictor variables taken at lead times of one month up to a maximum of seven months (the preceding May) preceding the winter in question. For this study it is assumed that there is no prediction skill derived from the previous year's NAO, although November values are available for selection. The sea-ice regions are taken from Screen (2017), where nine distinct sectors are

identified with limited covariability suggesting a large degree of regional independence. The NA SST tripole index is constructed following Fan and Schneider (2012). Similarly, the NAH and SST gradient indices are derived by calculating mean SSTs over two regions and subtracting them. See Table S1 for details. Blanca Ayarzaguena provided the T100 index which is a measure of the strength of the stratospheric polar vortex (SPV). This is an index of daily temperature anomalies at 100hPa, averaged over 65-90N, derived from JRA-55 reanalysis data (Kobayashi et al., 2015). Monthly mean values are constructed from an average of daily means. Tropical rainfall anomalies provide an indication of convective activity and divergence aloft, which can generate Rossby waves which propagate away from the source and are capable of influencing extratropical atmospheric circulation (Hoskins and Karoly, 1981). Data are obtained from the Global Precipitation Climatology Project (GPCPv2.3, Adler et al., 2003). However, such data are not available prior to 1979 so tropical SSTs for similar regions are used in the 1956 models, although there is not always a causal link between SSTs and tropical rainfall, depending on the region under consideration. The MJO is another index which can provide a tropical signal and is obtained from the Climate Prediction Center (CPC). Here ten phases are used rather than the more common eight, and a negative value indicates the active convective state (Baxter et al., 2014). QBO data (Naujokat, 1986, updated) use the 30hPa level following Hamilton (1984). All data are normalised to the period 1981-2010.

To capture the non-linear relationship between the NAO and the ENSO signal, we use the discontinuous N3.4 index of Folland et al. (2012) alongside the conventional index, where values are set to zero for normalised values between ± 1 . More negative values are set to -1, values in the range 1-1.75 are set to one, while values greater than 1.75 are set to zero. The volcanic index used by Folland et al. (2012) is also applied. Here, for two years following a volcanic eruption, values are set to one, and are zero for all other years, as Robock and Mao

(1995) identify the effects of eruptions in the first and second winters after the eruption. This
simulates the persistence of volcanic aerosols in the stratosphere after a major eruption. Major
tropical eruptions are as identified by Stenchikov et al. (2006).

Two versions of the NAO are used: the PC-based NAO (HPC; Hurrell et al., 2003) and a station-based index derived from SLP data from Reykjavik and Ponta Delgada, Azores (station NAO). This is taken as the difference between SLP at the two stations, which is then normalised. This is the approach followed by Scaife et al. (2014), although other indices are typically constructed by normalising the pressure at each station before subtracting (e.g. Hurrell, 1995; Cropper et al., 2015). The correlation between these two indices for winter is 0.90 (1956-2017); however, this does vary slightly over time (13-year running correlations between the two reach a low of 0.82 for a period in the early 1990s). The PC-based NAO better captures the spatial patterns of the NAO, although is dependent on the area selected for analysis, whereas due to shifts in the NAO centres of action a fixed-point index will not always represent this optimally. A disadvantage of the PC-based approach is that the whole index time series needs recalculating every time a new value is incorporated, and being a mathematical construct, does not necessarily represent climate physics (Dommenget and Latif, 2002)

Data are not detrended as we aim to forecast as closely as possible to the real world. In order
to address any trends present in the data, indices of atmospheric carbon dioxide and global
temperature are available for incorporation in the models. The winter season is defined as
December-January-February (DJF), where the winters refer to the year of the January.

3. Methods

Models are constructed for training periods from 1956-2010 and 1980-2010. The year 2010 is used as the cutoff, as in that year there are extreme values of the NAO and these are included in the training dataset, as simulation experiments show that models trained without extreme values perform poorly when used for prediction. Such extreme values are particularly important in view of the relatively small sample sizes available. This leaves 2011-2018 for use as a validation period (2011-2017 for the station-based NAO due to data availability). We predict values for this period without adapting the model. An alternative, retroactive verification approach (Mason and Baddour, 2008) constructs a model over the training period, then forecasts the next year only based on that model. Forecasts for subsequent years are based on models which incorporate the previous year's observation, and the models are allowed to evolve, both in terms of coefficients and predictors selected. This approach was tested for linear models and produced almost identical forecasts with only slight changes in coefficients, but the same predictors, with no appreciable improvement in forecast quality.

242 3.1 The NARMAX Method

One of the most attractive features of the NARMAX model, distinguishing it from other non-linear data-driven modelling techniques, is its power to build transparent and interpretable models where the mathematical significance of each model term is meaningful (e.g. Billings, 2013, p10) and in most real applications the selected model terms are physically interpretable (e.g. Billings, 2013, Chapter 14 Case Studies). Essentially, this approach treats each of the candidate predictors as a possible underlying cause of change in the response variable of interest and uses a set of model detection algorithms to automatically identify and pick out the most important predictors, based on which it then establishes a quantitative relationship that best relates the possible forcing variables to the response variable. Note that the dependence

relation of response on potential predictors can be either linear or non-linear. While traditional
linear modelling approaches such as ARMA and ARMAX might be able to capture main linear
relationships, they fail to reveal or capture any non-linear dynamics that are inherent in weather
and climate (e.g. Easterling et al., 2000; Hoerling et al., 2001; Dell et al., 2013; Burke et al.,
2015).

15 257

Taking the case of a one input (*u*) and one output (*y*) problem as an example, the NARMAX
model for *y* is written as (Billings, 2013):

$$y(k) = F(y(k-1), y(k-2), L, y(k-n_y),$$

$$u(k-d), u(k-d-1), L, u(k-d-n_u),$$

$$e(k-1), e(k-2), L, e(k-n_e)) + e(k)$$
(1)

where v(k) and u(k) are the measured system output (response) and input (explanatory), respectively, at time k; e(k) is a noise sequence which is not measurable but can be estimated once a model is built; n_y , n_u , and n_e are the maximum lags for the system output, input, and noise; $F(\bullet)$ is some non-linear function to be determined; and d is a time delay (typically d=0 or d=1). For an identified model, the noise e(k) can be estimated as the prediction errors: $e(k) = y(k) - \hat{y}(k)$, where $\hat{y}(k)$ is the predicted value at time instant k generated by an estimated model. The noise terms are included to accommodate the effects of measurement noise, modelling errors, and/or unmeasured disturbances. For the purposes of this study, we consider the predictability that can be obtained from the input (predictor) variables of preceding months, and so do not consider past outputs (previous winter NAO values) further.

₅₃ 270

There are many model subset selection methods such as the traditional forward selection
 (Faraway, 2002; Wilks, 2011). NARMAX uses an orthogonal forward selection algorithm,
 called Forward Regression Orthogonal Least Squares (FROLS; Billings, 2013), to select the

important terms. The efficiency of FROLS can be attributed to its use of mutual information (in addition to the correlation function), to measure not only linear but also potential non-linear dependent correlation of the target signal and the candidate explanatory signals. Furthermore, unlike traditional stepwise selection which uses hypothesis-tests and p-values to measure the significance of variables (terms), FROLS uses an effective index called the error reduction ratio (ERR) to measure the contribution made by each of the individual terms to explaining the change in the response variable, based on which the most significant term will be selected in each search step. The number of model terms can be determined using either the APRESS (Billings and Wei, 2008) statistics or the penalized error-to-signal ratio (PESR, a variant of APRESS) proposed in Wei et al. (2010). This is similar in principle to the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) but is developed specifically for non-linear systems. A leave-K out cross-validation is normally used with NARMAX, where K is around 10% of the training sample. A common model is identified from subsets of the training period, and common model parameters are then estimated using all the training data. In this study, a leave-one-out approach is employed due to the small sample size in the training data.

291 3.1.1. Linear models

NARMAX includes several linear model structures, e.g. autoregressive with exogenous inputs
(ARX) and autoregressive moving average with exogenous inputs (ARMAX) as a special case.
A general linear model structure of NARMAX is as follows:

$$y(k) = a_1 y(k-1) + \dots + a_{n_y} y(k-n_y) + b_0 u(k-d) + b_1 u(k-d-1)\dots + b_{n_u} u(k-n_u),$$

+ $e(k) + c_1 e(k-1) + \dots + c_n e(k-n_e)$ (2)

where $\underline{a_1, a_2, ..., a_{n_y}}$, $\underline{b_0, b_1, ..., b_{n_u}}$, and $\underline{c_1, c_2, ..., c_{n_e}}$ are model parameters. The above single input single output (SISO) case model (2) can be extended to multiple input single output (MISO) or multiple input multiple output (MIMO) cases. The commonly used multiple linear regression model is a special case of the MISO linear model. For example, for a case where there are *r* inputs, $u_1, u_2, ..., u_r$, by setting $d=0, n_y=0, n_u=0$ and $n_e=0$ yields:

$$y(k) = a_1 u_1(k) + a_2 u_2(k) + \dots + a_r u_r(k) + e(k)$$
(3)

302 3.1.2. Polynomial NARMAX models

In practice, many types of model structures are available to approximate the unknown function $F(\bullet)$ in (1), including power-form polynomial models and rational models (Chen and Billings, 1989), radial basis function networks (Wei et al., 2007), and wavelet neural networks (Billings and Wei, 2005; Wei et al., 2010). Power-form polynomial models are the most commonly used representation because such models have a number of unique, attractive properties (Billings 2013, pp35-37): for example for most applications the resulting models are transparent, physically interpretable and simple (parsimonious). This is the model form used in this study.

311 In practical applications, it is usual to consider many input signals (or explanatory variables) 312 and investigate how the explanatory variables (e.g. $u_1, u_2, ..., u_r$) influence the response variable 313 of interest. A case of many inputs can be represented by a special form of the NARMAX model 314 as follows:

$$y(k) = f(u_1(k), u_1(k-1), L, u_1(k-n_u), L, u_2(k), u_2(k-1), L, u_2(k-n_u), L, u_r(k), u_r(k-1), L, u_r(k-n_u)) + e(k)$$
(4)

where $n_u \ge 0$. In this study, $n_u = 0$, for which the model reduces to a polynomial model which belong to the family of multiple linear regression models:

317
$$y(k) = f(u_1(k), u_2(k), ..., u_r(k) + e(k))$$
 (5)

2	
3	
4	
5	
6	
7	
/	
8	
9	
10	
11	
11	
12	
13	
14	
15	
16	
17	
10	
10	
19	
20	
21	
22	
23	
24	
25	
25	
20	
27	
28	
29	
30	
31	
27	
22	
33	
34	
35	
36	
37	
38	
20	
29	
40	
41	
42	
43	
44	
45	
16	
40	
47	
48	
49	
50	
51	
52	
52	
52	
54	
55	
56	
57	
58	
59	

318 For example, with two inputs u_1 and u_2 , the initial full model of degree 2 is:

319
$$y(k) = a_0 + a_1 u_1(k) + a_2 u_2(k) + a_3 u_1^2(k) + a_4 u_1(k) u_2(k) + a_5 u_2^2(k) + e(k)$$
 (6)

320 Note that in practice it may not be necessary for all the six model terms in (6) to be included in a final predictive model, and only those that are important for explaining the variation of the 321 response should be included in the final model. 322

323

3.2. Small Sample Size Problem and Model Averaging 324

The NARMAX method has been successfully applied to solve a wide range of real-world 325 326 problems, and in most cases the method produces a robust reliable model that can be used for system analysis and prediction (simulation). For small sample-size problems where the number 327 of observations is small and much smaller than the number of regressors (explanatory variables 328 and their cross-product interactions), the identified model can be sensitive to adding or 329 removing a variable or cross-product term. In order to reduce the risk of using a single model 330 331 and mitigate the uncertainty in the prediction of a single model, this study proposes a simple 332 model-averaging approach to deal with the small sample-size problem. Models used in model 333 averaging only differ from each other in the number of terms used, thus all contain a common 334 core of predictors, with only minor differences in the coefficients.

335

336 Assume the training dataset S contains a total of N data points. Rather than using only a single 337 model to calculate predictions, we use a weighted average of multiple models (in this case three) to carry out predictions. The model averaging scheme is described below. 338

1 2		
3 4	340	Let $M_1, M_2,, M_s$ be the s best models identified by the PESR from the training dataset S, and
5 6 7	341	assume the values of the mean squared errors (MSE) of the s models over the training dataset
8 9 10	342	are $mse_1, mse_2,, mse_s$, respectively. Define
10 11 12	343	$c_1 = 1/mse_1, c_2 = 1/mse_2,, c_s = 1/mse_s$ (7a)
13 14 15	344	$c = c_1 + c_2 + \dots + c_s$ (7b)
16 17 18	345	$w_1 = c_1 / c, w_2 = c_2 / c,, w_s = c_s / c$ (7c)
19 20 21	346	Let $\hat{y}_1, \hat{y}_2,, \hat{y}_s$ be the predicted output values by the <i>s</i> models, the model averaging
22 23	347	prediction is defined as:
24 25 26	348	$\hat{y} = w_1 \hat{y}_1 + w_2 \hat{y}_2 + \dots + w_s \hat{y}_s \tag{8}$
20 27 28	349	
29 30	350	3.3. Forecast Verification
31 32 33	351	Correlation, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used to
34 35	352	give an indication of forecast skill, for both training and testing periods. A further measure is
36 37	353	used, based on the mean squared error (MSE), the Mean Square Error Skill Score (MSESS,
38 39 40	354	e.g. Wilks, 2011, p328). This compares the skill of the forecast with a reference forecast, in
41 42	355	this case a forecast based on climatology. There is generally lower skill in seasonal forecasting
43 44	356	compared with shorter term forecasts and an out-of-sample forecast of 8 years is not large
45 46 47	357	enough to allow robust statistical conclusions to be drawn about the quality of the forecast. A
47 48 49	358	generalized discrimination score, D, (Mason, 2012; Mason and Weigel, 2009) indicates
50 51	359	whether the forecasts are potentially useful, despite bias and poor calibration in a low-skill
52 53	360	situation (Mason, 2012), based on whether forecast values increase (decrease) with an increase
54 55 56	361	(decrease) in observation values, regardless of the magnitude of the change. For a pair of
57 58	362	observations, D gives the probability that a forecaster can discriminate the observations based
59 60	363	on the corresponding forecasts. D is related to Kendall's correlation coefficient τ by:

1 2		
3 4	364	$\tau = 2D-1 \tag{9}$
5 6 7	365	τ is effectively scaled from 0 to 1, a value of 0.5 means there is no skill in the forecast as the
7 8 9	366	probability of correctly discriminating between the size of two observations is 50%. Values
10 11	367	greater than 0.5 suggest that the forecast is potentially useful.
12 13 14	368	
14 15 16	369	Comparing the forecast methods objectively and determining whether there is a significant
17 18	370	difference is difficult with a small out-of-sample testing set (N=8; seven for station NAO). To
19 20	371	address this, the longer time series is split into even and odd years for the station NAO data.
21 22 22	372	Models are trained on the even years (N=32), the odd years forming the testing period (N=31).
23 24 25 26 27 28 29 30 31 32 33 34	373	The differences between linear and polynomial NARMAX model correlations for the testing
	374	period are assessed for significance with the "cocor" package for R (Diedenhofen and Musch,
	375	2015), using tests for overlapping dependent samples (the correlations both concern the NAO
	376	and the compared coefficients have a shared variable, the observed NAO). The package runs
	377	six different tests and their variations, for assessing the significance of the difference between
35 36	378	the correlations, together with a confidence interval test. For details see Diedenhofen and
37 38 39	379	Musch (2015).
40 41	380	
42 43	381	A further test is performed to examine whether the results obtained in the testing period for
44 45 46	382	NARMAX models are likely to have occurred by chance. For the HPC80LIN model, 100
47 48	383	surrogate datasets (Kugiumtzis, 2000; Schrieber and Schmitz, 2000) are created for each of the
49 50	384	predictors selected, allowing 100 surrogate NAO timeseries to be constructed for each of the
51 52	385	models with different numbers of terms (models with five to nine terms were examined).
55 55	386	Further details of this approach are given in the online supplementary information.
56 57	387	
58 59	388	4. Results

> In this section, NAO models are designated by type (station-based NAO: stat; Hurrell principal-component NAO: HPC), by year, and whether linear (LIN) or polynomial (POLY) NARMAX models. Thus stat80POLY is the polynomial NARMAX model for the station-based NAO, using 1980-2010 as the training dataset.

4.1 Linear Models

The predictors used in the linear models, and their coefficients are summarised in Table II and verification statistics for the averaged models are shown in Table III. It is the averaged models that are subsequently discussed, unless otherwise stated. An example of the model fit is shown for the station-based NAO in Figure 1, with other models shown in Figure S2. There is some consistency among the predictors selected for the different NAO indices, both within and between the different training periods (Table II). October Barents Sea SLP and October Barents-Kara Sea ice are selected for all models. Bering Sea and East Siberian-Laptev Sea ice in November are selected for three of the four models; both 1980 linear models, and HPC56LIN (Bering Sea ice only) and stat56LIN (East Siberian-Laptev Sea ice only). There is some discrepancy in the cryospheric terms, with Greenland Sea ice being selected only for the 1956 models. While Greenland Sea ice is chosen in the 1956 NAO models, October Bering Sea ice is selected for HPC80LIN as an additional term. Tropical influences are identified in both 1980 models, and for stat56LIN, however these show considerable variation, as variables selected for the 1980 models such as the MJO and tropical rainfall, are not available for the 1956 models. Consequently the greater availability of tropical predictors post-1980 is reflected in these models: November MJO (phase 8, HPC80LIN and phase 9, stat80LIN) and tropical Atlantic rainfall (July; stat80LIN) and August (HPC80LIN) are present in both 1980 linear models. Despite being previously identified as an important predictor of winter extratropical atmospheric variability, N3.4 is not particularly prominent, only being selected for both station-

based NAO models (July N3.4I). Extratropical oceanic SST-related terms are selected for all
except stat80LIN; October GIN SST for HPC80LIN and NAH (September-stat56LIN,
October-HPC56LIN). The only model to detect a stratospheric influence is HPC56LIN, where
the stratospheric polar vortex indices for October and November are selected, although
interestingly these are of opposite signs and are only weakly correlated (r=0.21).

Correlations of the 1980 models for the 2011-2018 testing period (Table IIIa) are 0.90 (HPC80LIN) and 0.82 (stat80LIN). However, correlations can mask a systematic negative bias in the forecasts, particularly in the linear model forecasts (with the exception of HPC80LIN), which becomes evident when MAE and RMSE are examined. For example, although stat80LIN correlates very well in the testing period (r=0.82), MAE and RMSE are 1.11 and 1.26 respectively, considerably higher than the training period values. MSESS scores are negative, the forecasts having less skill than a climatological forecast. The correlations indicate that the two NAO models are able to capture aspects of the local maxima and minima of the observed NAO during the testing period, but with a large negative bias (Figure 1), which reduces skill according to MSESS and inflates the MAE and RMSE. However, D-scores suggest that the models for the NAO have considerable potential usefulness (0.80, stat80LIN; 0.75, HPC80LIN, Table IIIa).

For the 1956 linear models, over the testing period the correlation is weaker for both (HPC56LIN, 0.46; stat56LIN, 0.69; Table III) compared with the 1980 models although no significant difference can be determined due to small sample size. MAE and RMSE values for test data are similar to the 1980 linear model for the station NAO, but greater for the HPC NAO, but are always larger than those for the training period, reflecting the negative bias of

forecasts in the testing period. MSESS are negative, but D-scores indicate potential usefulness (0.80, stat56LIN; 0.68, HPC56LIN).

> Many of the terms used in construction of the linear models, for example sea-ice, SST and tropical rainfall terms contain significant trends, which are likely to contribute to the negative bias of the forecasts. It is notable that the station NAO forecasts from 1980 training data appear to have overall negative trends (Figure 1), probably due to the influence of the sea-ice inputs to the models, as found in Hall et al. (2017).

4.2 Polynomial Models

The polynomial models are presented in detail in Table IV and compared with linear models in Figures 1 and S2. The comparisons use the averaged linear and polynomial models and an example of model averaging for the stat80POLY model is shown in Figure S3. Verification statistics are in Table III. For the HPC80 forecast the NARMAX algorithm produces a linear model only, as the inclusion of interaction terms does not significantly improve prediction performance.

There is less consistency amongst the predictor variables selected (Table IV) compared with the linear models, as terms are more complex and often based on interactions between variables, and with short datasets the models are very sensitive to slight differences in predictors. Equally, different predictors selected may be trying to capture the same variability or acting as proxies for some hidden variable. Some predictors are selected for both linear and polynomial models but may differ by a month or two.

Page 21 of 59

Forecasting verification statistics for the polynomial model averages are shown in Table III. In almost all cases, MAE, RMSE, MSESS and correlation are superior to the linear models in both training and testing periods, for both time periods. The D-score is consistently greater than 0.7, indicating considerable potential usefulness of the forecasts and MSESS are always positive. Verification statistics are mostly poorer for the 1956 models. Initially this may appear to be surprising as the models are constructed over a longer training set, although data from the earlier part of the time series prior to the satellite era are of poorer quality, and some variables are not available for the earlier period. See also section 5.1 on variable inherent predictability over the time period. The 1956 polynomial models show considerable improvements in all verification metrics compared with linear models on the longer time-scale, although results from linear models are broadly comparable to results obtained using the shorter training set. Due to the small sample size, the statistical significance of any improvements achieved by polynomial over linear models cannot always be satisfactorily determined. As would be expected for such a small sample size, when correlations in the testing period are compared using the suite of tests in the "cocor" package, differences are not significant.

In order to create a longer testing period and provide greater confidence in the performance of the models in out-of-sample prediction, the 1956-2017 period was split into even years for the training dataset and odd years for the testing dataset, using the station-based NAO index. The predictive skill of NARMAX polynomial and linear models when even years only are used as the training period is shown in Figure S4 and verification statistics are presented in Table S2. Correlations for the testing period (odd years) are 0.45 (linear) and 0.60 (polynomial). While both are significant (p<0.05), the polynomial correlation coefficient is appreciably higher, consistent with results from the shorter testing datasets, yet still not deemed significantly so by the range of comparison tests in "cocor". However, this does not tell the whole story. For the

earlier part of the testing period (1957-1979), correlations for the linear and polynomial models
are very similar (0.6, linear; 0.66, polynomial), however in the later period (1981-2017), the
correlations are significantly different using "cocor" (0.04, linear; 0.44 polynomial, p<0.1, one-
tailed test). In addition both MAE and RMSE for 1981-2017 are around two thirds of the values
for 1957-1979. This is suggestive of some change in the NAO in the later period, which is
better captured by the polynomial models.

Verification statistics taken over the 100 surrogate models for HPC80LIN show that model performance is much poorer for the surrogates in the testing period (Table S3), with significant differences between model performance and the surrogate data for all verification metrics. This indicates that the good prediction performance obtained with NARMAX is not a result of chance but is due to the efficacy of the algorithm.

Figure 2 allows closer inspection of model performance in the testing period. Both linear and polynomial 1980 models capture the local maxima and minima of both versions of the NAO for the first three years (Figure 2a,b), although the amplitude of the response is much reduced in the stat80LIN model (Figure 2b), which also display a negative bias, as discussed above, while any bias in stat80POLY is minimal and non-systematic. Similarly1956 linear models for both NAO indices reproduce the interannual change from 2011-2014, after which the response is damped. In general, polynomial models better match the local maxima and minima, with the exception of HPC56POLY which forecasts a negative rather than positive NAO in 2014 (Figure 2c).

4 509

Figures 3 and 4 give a visual representation of the selected predictors for the different NAO
 models, for both training periods, in order to highlight common predictor variables, although

some of these are used in interaction terms in the NARMAX models. For 1980 (Figure 3), Barents Sea SLP and Bering Sea ice are used in all three models, Barents-Kara Sea ice, East Siberian-Laptev Sea ice and tropical Atlantic rainfall are selected by both linear models while MJO phase 8 occurs for both linear and polynomial HPC models. For the 1956 models (Figure 4), some predictors (Greenland Sea ice, NAH, Barents Sea SLP) are included in all four models, while October Barents-Kara Sea ice and the N3.4I index are selected in three. A number of predictors occur in only one model, due to them occurring as interaction terms, and the sensitivity of the model to slight changes in variables given the short time series. For the 1956 training data, linear and polynomial models will select predictors that were suboptimal in the 1980 models, as the best predictors are not always available for the longer time series. Thus NARMAX selects N3.4I for the 1956 models, whereas for the 1980 models the optimal tropical influences seems to come from the MJO and tropical rainfall. While models are capturing essentially the same signals, there is high sensitivity to small variations in input data due to small sample size: hence different predictors are selected, which represent slightly different aspects of a common signal.

From the results, it is remarkable that while linear models show more limited skill, particularly over the longer training dataset, polynomial model forecasts, whether linear or non-linear, particularly those based on the 1980 training set, are able to replicate local maxima and minima, the amplitude of the observation and have minimal bias for the testing dataset.

5. Discussion

5.1 Differences when using longer and shorter training periods

Based on the verification statistics (Table III), polynomial models generally outperform the equivalent linear model (lower MAE and RMSE; higher correlation, MSESS and D-score),

although the significance of this is hard to assess on a small sample size. 1956 models generally perform less well than the equivalent from 1980. There is evidence from other studies to suggest that from the 1950s to 1960s there is less skill in dynamical forecasting hindcasts, with the possibility that during this period the atmospheric circulation was inherently less predictable (e.g. Weisheimer et al., 2016), associated with a more prevalent negative NAO. However, dynamical models appear to be skillful in predicting strongly negative NAO events, but less skillful regarding weaker negative NAO events. O'Reilly et al. (2017) concur that in the mid 20th century, forecast skill for the NAO is reduced, and attribute this to weaker forcing from tropical Pacific SSTs during this period. An alternative perspective is presented by Woollings et al. (2015). Decadal variability of the NAO is identified, which is associated with changes in strength of the jet and different dynamical behaviour (eddy-mean flow interaction, Rossby wavebreaking and blocking), the period from 1980 being associated with relatively higher jet speeds. Therefore when statistical models are trained on a longer period, with the assumption of stationarity, these decadal variations are averaged out and predictors that are selected are likely to be sub-optimal. Furthermore, if there is a return to relatively weak forcing from Pacific SSTs, or a period of reduced jet speeds, forecasts trained on the recent period may perform less well (O'Reilly et al., 2017). Weisheimer et al. (2016) report that forecast skill actually increases further back in the 20th century, in the 1930s and 1940s, suggesting the issue may not be attributed to reduced data quality. This corresponds to the decadal fluctuations identified by Woollings et al. (2015), suggesting that periods of lower jet speed may contribute to the reduced predictability. In this study, a longer training dataset does not equate to a better forecasting model, in agreement with the studies above. It is also notable that when tested on all the odd years, a polynomial model significantly outperforms a linear model for the period 1981-2017, better capturing the NAO for both early and late periods (Figure S4, Table S2).

Page 25 of 59

Greenland Sea ice is a predictor in all 1956 models (Figure 4) but is not selected for models based on the shorter training period. The months selected (May-July) represent the time of maximum decrease in the regional annual sea-ice cycle (October is selected as the 8th and final term in stat56LIN, Table IIc, and makes a minimal contribution). Interannual variability and mean values of Greenland Sea ice are notably larger prior to 1980; consequently contributions of the predictor terms to the models are greatest prior to 1980. Given the relatively small coefficients used in the models, contributions between 1980 and 2010 are negligible, also indicating why this variable is not selected for the 1980 models. Figure S5 shows that the four Greenland Sea ice terms used essentially make similar contributions to each model, and can be regarded as different, somewhat imperfect representations of some unknown predictor variable. It is possible that the association in the 1960s and 1970s is related to the Great Salinity Anomaly (e.g. Dickson et al., 1988), which circulated in the North Atlantic at this time and had its origins in increased ice export and freshwater release through the Greenland Sea region. The association between the Greenland Sea ice and NAO is physically plausible at this time, particularly as the predictors capture the transition from a more negative to a positive NAO over the period.

² 570

579 5.2. Tropical forcing and interaction terms

580 Evidence for forcing of the winter NAO from the tropics is evident in all models; the 1980 581 training data favours the selection of the MJO and tropical rainfall, whereas these are not 582 available for the 1956 training data, so the N3.4 Index becomes more prominent, along with 583 West Indian Ocean SST in stat56LIN and HPC56POLY. It is also notable that HPC56LIN 584 includes no tropical forcing, although the use of stratospheric polar vortex terms here could 585 incorporate aspects of tropical forcing, such as the ENSO-stratospheric-mid-latitude 586 teleconnection (e.g. Bell et al., 2009). Tropical variables selected are somewhat inconsistent,

in terms of both month and predictor, which is likely to be a result of model sensitivity to slight
variations due to the short time series used. However, the evidence for tropical signals is clear.

There are multiple interaction terms identified in the polynomial models, which can be very difficult to explain effectively without subsequent further analysis. N3.4I terms are selected in both 1956 polynomial models, possibly due to the unavailability of MJO indices as discussed above. However, in each case it occurs as an interaction term with a high latitude predictor (terms 9 and 10, stat56POLY; terms 4 and 8, HPC56POLY, Table IVb,c). A closer inspection of the interaction terms' time series reveals that the main input parameter is N3.4I, with the high latitude term modulating the signal in terms of amplitude and sometimes polarity (not shown). This, along with term 1 in stat56POLY (Table IVb), is suggestive of a tropical signal, propagated by Rossby waves, which is then modulated by slowly-varying boundary conditions at higher latitudes (e.g. Ding et al., 2014). Term 5 in HPC56POLY (June Hudson Bay ice x volcanic index; Table IVc) operates in a similar way, the magnitude of the tropical volcanic signal being the dominant input, modulated by the sea-ice term.

A composite plot of low minus high years for MJO phase 8 in October shows a wave train emanating from the central Pacific, over North America and to the Atlantic (Figure S6). Poleward propagating Rossby waves emanate from only a few source regions, whose activity varies with tropical rainfall variability from year to year (Scaife et al., 2017). Of note in this wavetrain is a node boundary near the Labrador Sea. An interaction term between September Labrador Sea ice and the October MJO8 is identified for stat80POLY (Figure 3, Table IVa). It is possible that in years with high sea-ice in this region, there is an interaction with the MJO signal, and the effect of this interaction is sustained into winter. As the interaction term is multiplicative, then it is the sign of each index which is particularly important. Both high (low)

sea-ice and high (low) MJO8 will combine to make the NAO prediction more positive. If,
however, one input is positive and the other negative, that will combine towards a more
negative NAO forecast.

5.3. Higher latitude forcing and interaction terms.

Both cryospheric and extratropical forcings are included in all models, often as interaction terms with tropical forcing as discussed above. Key sea-ice predictors are October Barents-Kara Sea (five out of seven models), November East Siberian -Laptev Sea (four out of seven), Bering Sea (five out of seven). The Greenland Sea has been discussed in Section 5.1 above. Sea-ice in the Barents-Kara Seas has frequently been identified as a key source of potential winter NAO predictability (Scaife et al., 2014; Garcia-Serrano et al., 2015; Wang et al., 2017; Hall et al., 2017). The importance of Bering Sea ice is more surprising, but could well be a proxy for atmospheric variability in the region, such as the Pacific-North American pattern (PNA), with associations with North Atlantic atmospheric variability (e.g. O'Reilly et al., 2017), which in turn is linked to tropical Pacific SST variability.

3 627

Another input variable worthy of more detailed discussion is Barents SLP. This is selected by all NAO models (Figures 3, 4). The most commonly selected month is October (four times). In polynomial models Barents Sea SLP occurs as an interaction term with tropical signals (stat80POLY, Table Iva, HPC56POLY, Table IVc), as a single term (stat56POLY, Table IVb), an interaction term with SSTs (stat56POLY, Table IVb) and the QBO (HPC56POLY, Table IVd). A cyclonic (anticyclonic) anomaly in the BK region in October can lead to positive (negative) sea-ice anomalies there in November (King and Garcia-Serrano, 2016). In other words, the preceding pressure/geopotential height anomaly is a precursor of the Barents-Kara Sea ice ice which has been frequently identified as a predictor in other studies (e.g. Garcia-

Serrano et al., 2015; Hall et al., 2017). However, it also has the advantage that it is not subject to the same long-term dramatic trends as autumn sea-ice. The interaction terms support the concept of geopotential height anomalies over the BK Seas in autumn modulating signals from the tropics, stratosphere and cryosphere in a non-linear way (e.g. Vihma et al., in review, International Journal of Climatology). There are a number of input terms from these high latitude regions that are combined in polynomial models: [November east Siberian Sea ice x November GIN seas SST], [October Barents SLP x November GIN seas SST] (stat56POLY); [September GIN seas SST x October Atlantic SST gradient] (HPC56POLY). It seems likely that these terms are capturing different aspects of autumn variability in the northern seas, that would merit further investigation as important predictors of the winter NAO. Labrador Sea ice is only selected as a predictor in polynomial models, as an interaction term

with tropical (see section 5.2 above) or extratropical forcings. An interaction term selected in both HPC56POLY and stat80POLY is [October Labrador Sea ice x October NAH SST pattern]. The NAH pattern of SSTs is associated with forcing of the winter NAO by persistent SST anomalies at up to six months lead time (Czaja and Frankignoul, 2002). The separate input terms and the resulting multiplicative term are shown in Figure S7a. An examination of the 13-year running correlation between the October NAH and winter NAO reveals a consistent negative correlation, except for the period from 1995 to 2005 when the correlation coefficient increases sharply, becoming positive, before a rapid return to negative values (Figure S7b). This positive excursion coincides with large positive Labrador Sea ice anomalies in October, these being negative for the rest of the time series. The regions for both these variables have a partial overlap. Therefore when there is a negative ice anomaly in the Labrador Sea, the inverted NAH index provides a predictor of the winter NAO interannual variability; however for the brief period with positive ice anomalies, there is a positive relationship between the

NAH and NAO interannual variability (Figure S7a). The sea-ice in the Labrador Sea modulates
the NAH/NAO interaction, or there is a hidden North Atlantic variable for which Labrador Sea
ice is a proxy.

¹²₁₃ 666 **6. Summary**

The NARMAX approach shows appreciable potential skill in out-of-sample forecasting, albeit with small testing datasets, for both linear and polynomial models which both outperform a more conventional ordinary least squares approach to multiple regression (e.g. Hall et al., 2017). There are strong correlations with observations, reproducing local maxima and minima of the observations, and the amplitude of the observed signal. Model fits are improved when based on a shorter training dataset from 1980-2010. This may partly relate to a wider range of potential predictors being available for this period, but is also because of reduced inherent predictability of circulation indices during the 1950s and 1960s. The skill of polynomial models is consistently greater than that of equivalent linear models, and error statistics are reduced, but small sample size means that further work is needed to confirm the significance of this result. However, an analysis based on using odd years for the testing data is strongly suggestive of a better performance by a polynomial model over that of a linear model, particularly in that it better represents the transition from the early more negative NAO period to the end of the 1970s to the more positive phase post-1980. NARMAX can identify important predictors of winter North Atlantic atmospheric variability. Discrepancies between predictor selection in models is likely to arise through increased sensitivity to small fluctuations in input, due to the small sample sizes available. This means that the models capture the same signals, but select them in slightly different ways.

- ⁵⁵ ⁵⁶ 685

An important result of the study is that polynomial NARMAX models are capable of revealing the potential modulation of tropical forcing by higher latitude boundary conditions. Barents Sea SLP can also play a crucial role in modulating cryospheric and extratropical signals in addition to those from the tropics. This could be significant in developing the next generation of NWP models.

Our NARMAX approach can be extended to other circulation indices such as the East Atlantic and Scandinavian Patterns and jet latitude and speed, and to other seasons. This may be especially beneficial for summer seasonal forecasts, where there is currently relatively little predictability from dynamical models (e.g. Ossó et al., 2018; Dunstone et al., 2018). It is also possible to extend the approach to probabilistic forecasting and - by utilising links between North Atlantic circulation patterns and, for example, UK regional temperature and precipitation patterns (Hall and Hanna, 2018) - provide enhanced seasonal forecasts that should be useful for a wide range of stakeholders. NARMAX can also be used to assess how contributions from different atmospheric circulation predictors vary over time using a moving window approach. Future work will use Coupled Model Intercomparison Project (CMIP) 5/6 output to construct models using a longer timeseries, enabling the use of a longer testing dataset, to confirm whether polynomial forecasts are significantly improved compared to linear versions. The models can be further extended to include previous years' predictor values, and by increasing the lead-time at which forecasts are issued for a given season.

Acknowledgements

We thank Blanca Avarzaguena for providing the stratospheric polar vortex data, and the providers of the numerous datasets used in the study. We thank the three anonymous reviewers for their insightful comments which greatly improved the manuscript.

1		
2 3 4	711	
5 6	712	
7 8 0	713	
10 11	714	References
12 13 14	715	Adler, R. F., Huffmann, G. J., Chang, A., Ferrar, R., Xie, P., Janowiak, J., Arkin, P.
15 16 17	716	(2003): The version-2 Global Precipitation Climatology Project (GPCP) monthly
17 18 19	717	precipitation analysis (1979-present). Journal of Hydrometeorology, 4, 1147-1167,
20 21 22	718	doi:10.1175/1525-7541(2003)004, 1147:TVGPCP.2.0.CO;2
22 23		
24 25	719	Andrews, M. B., Knight, J.R., & Gray, L.J. (2015). A simulated lagged response of the North
26 27	720	Atlantic Oscillation to the solar cycle over the period 1960–2009. Environmental
28 29 30	721	Research Letters, 10, 054022, doi:10.1088/1748-9326/10/5/054022
31 32 33	722	Ayala-Solares, J.R., Wei, HL., & Bigg, G.R. (2018). The variability of the Atlantic
34 35	723	meridional circulation since 1980, as hindcast by a data-driven nonlinear systems
36 37	724	model. Acta Geophysica 66, 683-695, doi: 10.1007/s11600-018-0165-7
38 39		
40 41	725	Baxter, S., Weaver, S., Gottschalck, J., & Xue, Y. (2014). Pentad evolution of wintertime
42 43	726	impacts of the Madden-Julian Oscillation over the contiguous United States. Journal
44 45 46	727	of Climate, 27: 7356-7367, doi:10.1175/JCLI-D-14-00105.1
47 48 49	728	Baldwin, M. P., & Dunkerton, T.J. (2001). Stratospheric harbingers of anomalous weather
50 51 52	729	regimes. Science, 294, 581-584, doi:10.1126/science.1063315
53 54 55	730	Bell, C. J., Gray, L.J., Charlton-Perez, A.J., & Joshi, M.M. (2009). Stratospheric
56 57	731	communication of El Niño teleconnections to European winter. Journal of Climate, 22,
58 59 60	732	4083–4096, doi:10.1175/ 2009JCLI2717.1

1 ว		
2 3 4	733	Bell, V.A., Davies, H.N., Kay, A.L., Brookshaw, A., & Scaife, A.A. (2017). A national-scale
5 6	734	seasonal hydrological forecast system: development and evaluation over Britain.
7 8 9	735	Hydrological and Earth System Sciences Discussions, doi: 10.5194/hess-2017-154
10 11 12	736	Bigg, G.R., Wei, HL., Wilton, D.J., Zhao, Y., Billings S.A., Hanna, E., & Kadirkamanathan,
13 14 15	737	V. (2014). A century of variation in the dependence of Greenland iceberg calving on ice
15 16 17	738	sheet surface mass balance and regional climate change. Proceedings of the Royal Society
19 18 19	739	Series A, 470. doi:10.1098/rspa.2013.0662.
20 21 22	740	Billings, S. A. (2013). Non-Linear System Identification: NARMAX Methods in the Time,
22 23 24	741	Frequency, and Spatio-Temporal Domains. London: Wiley.
25 26	742	Billings, S.A. & Wei, HL. (2005). The wavelet-NARMAX representation: A hybrid model
27 28	743	structure combining polynomial models with multiresolution wavelet decompositions.
29 30 31	744	International Journal of Systems Science, 36: 137–52.
32 33 34	745	Billings, S.A. & Wei, HL. (2008). An adaptive orthogonal search algorithm for model
35 36	746	subset selection and non-linear system identification. International Journal of Control,
37 38 39 40	747	81(5), 714-724, doi: 10.1080/00207170701216311.
41 42	748	Boer, G.J., & Hamilton, K. (2008). QBO influence on extratropical predictive skill. Climate
43 44 45	749	Dynamics, 31: 987-1000, doi: 10.1007/s00382-008-0379-5
46 47 48	750	Burke, M., Hsiang, S.M., & Miguel, E. (2015). Global non-linear effect of temperature
49 50	751	on economic production. Nature, 527: 235–23.
51 52	752	Chen, S., & Billings, S.A. (1989). Representation of non-linear systems: The NARMAX
53 54 55 56 57 58 59 60	753	model. International Journal of Control, 49: 1013–32.

1		
2 3 4	754	Clark, R.T., Bett, P.E., Thornton, H.E., & Scaife, A.A. (2017). Skilful seasonal predictions
5 6 7	755	for the European energy industry. Environmental Research Letters 12: 024002. doi:
/ 8 9	756	10.1088/1748-9326/aa57ab
10 11	757	Cohen, J., Barlow, M., Kushner, P.J., & Saito, K. (2007). Stratosphere-troposphere coupling
12 13	758	and links with Eurasian land surface variability. Journal of Climate, 20: 5335-5343, doi:
14 15 16	759	10.1175/2007JCLI1725.1
17 18	760	Cohen, J., Coumou, D., Hwang, J., Mackey, L., Orenstein, P., Totz, S., & Tziperman, E.
19 20	761	(2018). S2S reboot: an argument for greater inclusion of machine learning in
21 22 23	762	subseasonal to seasonal forecasts. WIREs Climate Change 2018e00567, doi:
24 25	763	10.1002/wcc.567.
26 27	764	Cropper, T., Hanna, E., Valente, M.A., & Jónsson, T. 2015. A daily Azores-Iceland North
28 29 30	765	Atlantic Oscillation index back to 1850. Geoscience Data Journal, 2(1): 12-24, doi:
31 32	766	10.1002/gdj3.23
33 34 35	767	Czaja, A., & Frankignoul, C. (1999). Influence of the North Atlantic SST on the atmospheric
36 37 38	768	circulation. Geophysical Research Letters, 26: 2969-2972.
39 40 41	769	Czaja, A., & Frankignoul, C. (2002). Observed impact of Atlantic SST anomalies on the
42 43 44	770	North Atlantic Oscillation. Journal of Climate, 15: 606-623.
45 46	771	Dell, M., Jones, B.F., & Olken, B.A. (2013). What do we learn from the weather? The new
47 48 49	772	climate-economy literature. Journal of Economic Literature, 52(3): 740:798.
50 51	773	Derome, J., Lin, H., & Brunet, G. (2005). Seasonal Forecasting with a simple general
52 53	774	circulation model: predictive skill in the AO and PNA. Journal of Climate 18:597-609.
54 55 56	775	Deser, C., Alexander M.A., & Timlin, M.S. (2003): Understanding the persistence of sea
57 58 59 60	776	surface temperature anomalies in midlatitudes. Journal of Climate, 16, 57-72.

3 4	777	Deser C., Alexander, M.A., Xie, SP., & Phillips, A.S. (2010). Sea surface temperature
5 6 7	778	variability: patterns and mechanisms. Annual Reviews in Marine Science, 2: 115-143,
/ 8 9	779	doi: 10.1146/annurev-marine-120408-151453
10 11	780	Diedenhofen, B., & Musch, J. (2015). cocor: a comprehensive solution for the statistical
12 13	781	comparison of correlations. PLoS ONE 10(4): e0121945, doi:
14 15 16	782	10.10371/journal.pone.0121945
10 17 18	783	Dickson, R.R., Meincke, J., Malmberg, SA., & Lee, A.J. (1988). The "Great Salinity
19 20	784	Anomaly" in the Northern Atlantic, 1968-1982. Progress in Oceanography 20, 103-151.
21 22 23	785	Ding, Q., Wallace, J.M., Battisti, D.S., Steig, E.J., Gallant, A.J.E., Kim, H-J, & Geng, L.
23 24 25	786	(2014). Tropical forcing of the recent rapid Arctic warming in northeastern Canada and
26 27	787	Greenland. Nature 509, 209-212, doi: 10.1038/nature13260
28 29 30	788	Dommenget, D., & Latif, M. (2002), A cautionary note on the interpretation of EOFs. Journal
30 31 32	789	of Climate 15, 216-225.
33 34	790	Driscoll, S., Bozzo, A., Gray, L.J., Robock, A., & Stenchikov, G. (2012). Coupled Model
35 36	791	Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions.
37 38 39	792	Journal of Geophysical Research, 117: D17105, doi: 10.1029/2012D017607
40 41	793	Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N.,Knight, J.
42 43	794	(2016). Skilful predictions of the winter North Atlantic Oscillation one year ahead.
44 45 46 47	795	Nature Geoscience, 9: 809-814. doi: 10.1038/NGEO2824
48 49	796	Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Fereday, D., O'Reilly, C., Belcher, S.
50 51	797	2018. Skilful seasonal predictions of summer European rainfall. Geophysical Research
52 53 54 55 56 57 58 59 60	798	Letters, 45: 3246:3254. doi: 10.1002/2017GLO76337

2		
3 4	799	Easterling, D.R., Meehl, G.A., Parmesan, C., Changnon, S.A., Karl, T.R., & Mearns, L.O.
5 6	800	(2000). Climate extremes: observations, modeling, and impacts. Science, 289: 2068-
/ 8 9	801	2074.
10 11	802	Enfield, D.B., Mestas-Nuñez, A.M., & Trimble, P.J. (2001). The Atlantic multidecadal
12 13	803	oscillation and its relation to rainfall and river flows in the continental U.S. Geophysical
14 15	804	Research Letters, 28(10): 2077-2080.
16 17 18	805	Fan, M., & Schneider, E.K. (2012). Observed decadal North Atlantic tripole SST variability.
19 20	806	Part i: weather noise forcing and coupled response. Journal of the Atmospheric
21 22	807	Sciences, 69: 35-50, doi: 10.1175/JAS-D-11-018.1
23 24 25	808	Faraway, J (2002) Practical Regression and ANOVA Using R. Available at: http://cran.r-
25 26 27	809	project.org/doc/contrib/Faraway-PRA.pdf. Accessed November 2018.
28		
29 30	810	Feldstein, S. (2000). The timescale, power spectra, and climate noise properties of
31 32 33	811	teleconnection patterns. Journal of Climate, 13: 4430-4440
34 35 36	812	Fletcher, C.G., & Saunders, M.A. (2006). Winter North Atlantic Oscillation Hindcast Skill:
37 38	813	1900-2001. Journal of Climate, 19: 5762-5776.
39 40		
41 42	814	Folland, C.K., Scaife, A.A., Lindesay, J., & Stephenson, D.B. (2012). How predictable is
43 44	815	northern European winter climate a season ahead? International Journal of Climatology,
45 46	816	32: 801-818, doi: 10.1002/joc.2314
47 48 40	817	García-Serrano, J., Frankignoul, C., Gastineau, G., & De La Cámara, A. (2015). On the
49 50		
51 52	818	predictability of the winter Euro-Atlantic climate: lagged influence of autumn sea ice.
53 54	819	Journal of Climate 28, 5195-5216, doi: 10.1175/JCLI-D-14-00472.1.
55 56 57	820	García-Serrano, J., Frankignoul, C., King, M.P., Arribas, A., Gao, Y., Guemas, V.,
58 59 60	821	Sanchez-Gomez, E. (2017). Multi-model assessment of linkages between eastern

1 2		
3 4	822	Arctic sea-ice variability and the Euro-Atlantic atmospheric circulation in current
5 6 7	823	climate. Climate Dynamics, 49: 2407-2429, doi: 10.1007/s00382-016-3454-3.
8 9 10	824	Garfinkel, C.I., Benedict, J.J., & Maloney, E.D. (2014). Impact of the MJO on the boreal
11 12	825	winter extratropical circulation. Geophysical Research Letters, 41: 6055-6062, doi:
13 14 15	826	10.1002/2014GL061094
16 17 18	827	Gray, L.J., Woollings, T.J., Andrews, M., & Knight, J. (2016). Eleven-year solar cycle signal
19 20	828	in the NAO and Atlantic/European blocking. Quarterly Journal of the Royal
21 22 23	829	Meteorological Society, 142: 1890-1903, doi: 10.1002/qj.2782
24 25	830	Gray, L.J., Scaife, A.A., Mitchell, D.M., Osprey, S., Ineson, S., Hardiman, S.,Kodera K.
26 27 28	831	(2013). A lagged response to the 11 year solar cycle in observed winter
29 30	832	Atlantic/European weather patterns. Journal of Geophysical Research: Atmospheres,
31 32 33	833	118: 13 405-13-420, doi: 10.1002/2013JD020062
34 35 36	834	Häkkinen, S., Rhines, P.B., & Worthen, D.L. 2011. Atmospheric blocking and Atlantic
37 38 39	835	multidecadal ocean variability. Science, 334: 655-659, doi: 10.1126/science.1205683
40 41	836	Hall, R.J., & Hanna, E. (2018). North Atlantic circulation indices: links with summer and
42 43	837	winter UK temperature and precipitation and implications for seasonal forecasting.
44 45 46 47	838	International Journal of Climatology, 38(S1): e660-e677, doi: 10.1002/joc.5398
48 49	839	Hall, R., Erdélyi, R., Hanna, E., Jones, J.M., & Scaife, A.A. (2015). Drivers of North Atlantic
50 51 52	840	polar front jet stream variability. International Journal of Climatology, 35: 1697-1720.
52 53 54	841	doi: 10.1002/joc.4121
55 56	842	Hall, R.J., Scaife, A.A., Hanna, E., Jones, J.M., & Erdélyi, R. (2017). Simple statistical
57 58	843	probabilistic forecasts of the winter NAO. Weather and Forecasting 32: 1585-1601. doi:
59 60	844	10.1175/WAF-D-16-0124.S1

2 3	845	Hamilton, K. (1984). Mean wind evolution through the quasi-biennial cycle in the tropical
4 5	846	lower stratosphere Journal of the Atmospheric Sciences 41: 2113-2125
6 7	047	Home E. & Crommer T.E. (2017). North Atlantic Oscillation. Oxford Descerab
8 9	847	Hanna, E, & Cropper, T.E. (2017). North Atlantic Oscillation. Oxford Research
10 11	848	Encyclopedia of Climate Science. doi: 10.1093/acrefore/9780190228620.013.22
12 13	849	Hoerling, M.P., Kumar, A., & Xu, T. (2001). Robustness of the non-linear climate response
14 15	850	to ENSO's extreme phases. Journal of Climate 14:1277-1293.
16 17 18	851	Hoerling, M.P., Hurrell, J.W., Xu, T., Bates, G.T., & Phillips, A.S. (2004). Twentieth century
19 20	852	North Atlantic climate change. Part II. Understanding the effect of Indian Ocean
21 22	853	warming. Climate Dynamics, 23: 391-405.
23 24	854	Hoskins, B.J., & Karoly, D.J. (1981). The steady linear response of a spherical atmosphere to
25 26 27	855	thermal and orographic forcing. Journal of the Atmospheric Sciences, 38:1179-1196.
28 29	856	Hoskins, B.J., & Valdes, P.J. (1990). On the existence of storm-tracks. Journal of the
30 31 22	857	Atmospheric Sciences, 47: 1854-1864.
32 33 34	858	Hurrell, J.W. (195) Decadal trends in the North Atlantic Oscillation: regional temperature and
35 36	859	precipitation. Science, 269: 676-679.
37 38	860	Hurrell, J.W., Kushnir, Y., Visbeck, M., & Ottersen, G. (2003). An overview of the North
39 40 41	861	Atlantic Oscillation. In Hurrell, J.W., Kushnir, Y., Ottersen, G., & Visbeck, M., (eds).
42 43	862	The North Atlantic Oscillation, Climatic Significance and Environmental Impact. AGU
44 45 46	863	Geophysical Monograph 134, pp1-35.
47 48 49	864	Hurrell, J.W., & Deser, C. (2009). North Atlantic climate variability: the role of the North
50 51	865	Atlantic Oscillation. Journal of Marine Systems, 78: 228-41. doi:
52 53	866	10.1016/j.jmarsys.2008.11.026
54 55 56	867	Ineson, S., Scaife, A.A., Knight, J.R. Manners, J.C., Dunstone, N.J., Gray, L.J., & Haigh, J.D.
57 58	868	(2011): Solar forcing of winter climate variability in the Northern Hemisphere. Nature
59 60	869	Geoscience, 4, 753-757, doi: 10.1038/ngeo1282

3 4	870	James, I.N., & James, P.M. (1989). Ultra-low-frequency variability in a simple atmospheric
5 6 7	871	circulation mode. Nature, 342: 53-55.
, 8 9	872	Johansson, Å. (2007). Prediction skill of the NAO and PNA from daily to seasonal time
10 11 12	873	scales. Journal of Climate, 20: 1957-1975, doi: 10.1175/JCLI4072.1
13 14 15	874	Kalnay, E., Kanamitsu. M., Kistler, R., Collins, W., Deavon, D., Gandin, L., Joseph, D.
16 17	875	(1996). The NCEP/NCAR 40-year reanalysis project. Bulletin of the American
18 19 20	876	Meteorological Society, 77: 437-471.
21 22 23	877	Kang, D., Lee, MI., Im, J., Kim, D., Kim, HM., Kang, HS., MacLachlan, C. (2014).
24 25	878	Prediction of the Arctic Oscillation in boreal winter by dynamical seasonal forecasting
26 27	879	systems. Geophysical Research Letters, 41: 3577-3585, doi: 10.1002/2014GL060011
28 29 20	880	Kidston, J., Scaife, A.A., Hardiman, S.C., Mitchell, D.M., Butchart, N., Baldwin, M.P., &
30 31 32	881	Gray, L.J. (2015). Stratospheric influence on tropospheric jet streams, storm tracks and
33 34	882	surface weather. Nature Geoscience, 8: 433-440, doi: 10.1038/NGEO2424
35 36	883	Kim, HM., Webster, P., & Curry, J. (2012). Seasonal prediction skill of ECMWF system 4
37 38 39	884	and NCEP CFSv2 retrospective forecast for the Northern Hemisphere winter. Climate
40 41	885	Dynamics, 39: 2957-2973, doi: 10.1007/s00382-012-1364-6
42 43	886	King, M.P., & García-Serrano, J. (2016). Potential ocean-atmosphere preconditioning of late
44 45 46	887	autumn Barents-Kara sea ice concentration anomaly. Tellus A, 68: 28580,
40 47 48	888	doi:10.3402/tellusa.v68.28580
49 50	889	Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Takahashi, K.
51 52	890	(2015). The JRA-55 reanalysis: general specifications and basic characteristics. Journal
53 54 55 56 57 58 59 60	891	of the Meteorological Society of Japan, 93: 5-48, doi: 10.2151/jmsj.2015-001

1 ว		
2 3 4	892	Koenigk, T., Caian, M., Nikulin, G., & Schimanke, S. (2016). Regional Arctic sea ice
5 6	893	variations as predictor for winter climate conditions. Climate Dynamics, 46: 317-337,
7 8 9	894	doi:10.1007/s00382-015-2586-1
10 11	895	Kolstad, E.W., & Årthun, M. (2018). Seasonal prediction from Arctic sea surface
12 13	896	temperatures: opportunities and pitfalls. Journal of Climate, 31:8197-8210, doi:
14 15	897	10.1175/JCLI-D-18-0016.1
16 17 18	898	Kugiumtzis, D. (2000). Surrogate data test for nonlinearity including nonmonotonic
19 20	899	transforms. Physical Review E 62, R25-R28, doi: 10.1103/PhysRevE.62.R25
21 22	900	Li, S., Perlwitz, J., Hoerling, M.P., & Chen, X. (2010). Opposite annular responses of the
23 24 25	901	northern and southern hemispheres to Indian Ocean warming. Journal of Climate, 23:
26 27	902	3720-3738, doi: 10.1175/2010JCLI3410.1
28 29	903	Lockwood, M., Harrison, R.G., Woollings, T., & Solanki, S.K. (2010). Are cold winters in
30 31 32	904	Europe associated with low solar activity? Environmental Research Letters, 5: 024001,
32 33 34	905	doi: 10.1088/1748-9326/5/2/024001
35 36	906	MacLachlan, C., Arribas, A., Peterson, K.A., Maidens, A., Fereday, D., Scaife, A.A.,
37 38	907	Madec, G. (2015) Global seasonal forecast system version 5 (GloSea5): a high
39 40 41	908	resolution seasonal forecast system. Quarterly Journal of the Royal Meteorological
42 43	909	Society, 141:1072-1084, doi: 10. 1002/qj.2396
44 45	910	Mason, S.J. (2012). Seasonal and longer-range forecasts. In Forecast Verification: a
46 47 48	911	Practitioner's Guide in Atmospheric Science. eds: IT Jolliffe, I.T., & Stephenson, D.B.
49 50	912	Chichester, Wiley Blackwell, pp203-220.
51 52	913	Mason, S.J., & Baddour, O. (2008). Statistical Modelling. In Seasonal Climate Variability:
53 54 55	914	Forecasting and Managing Risk. eds: Troccoli, A. Harrison, M.S.J., Anderson, D.L.T.,
56 57	915	& Mason, S.J. Dordrecht, Springer pp163-201.
58 59		
60		

1 2		
- 3 4	916	Mason, S.J., & Weigel, A.P. (2009). A generic forecast verification framework for
5 6	917	administrative purposes. Monthly Weather Review, 137: 331-349, doi:
7 8 0	918	10.1175/2008MWR2553.1
) 10 11	919	Morice, C.P., Kennedy, J.J., Rayner, N.A., & Jones, P.D. (2012). Quantifying uncertainties in
12 13	920	global and regional temperature change using an ensemble of observational estimates:
14 15 16	921	the HadCRUT4 dataset. Journal of Geophysical Research, 117: D08101, doi:
17 18	922	10.1029/2011JD017187
19 20	923	Müller, W.A., Appenzeller, C., & Schär, C. (2005). Probabilistic seasonal prediction of the
21 22 22	924	winter North Atlantic Oscillation and its impact on near surface temperature. Climate
23 24 25	925	Dynamics 24: 213-226, doi: 10.1007/s00382-004-0492-z
26 27	926	Naujokat, B. (1986). An update of the observed quasi-biennial oscillation of the stratospheric
28 29 30 31	927	winds over the tropics. Journal of the Atmospheric Sciences, 43: 1873-1877.
32 33	928	van Oldenborgh, G.J., te Raa, L.A., Dijkstra, H.A., & Philip, S.Y. (2009). Frequency- or
34 35	929	amplitude-dependent effects of the Atlantic meridional overturning on theller tropical
36 37 29	930	Pacific Ocean. Ocean Science, 5: 293-301, doi: 10.5194/0s-5-293-2009
30 39		
40 41	931	O'Reilly, C.H., Heatley, J., MacLeod, D., Weisheimer, A., Palmer, T.N., Schaller, N., &
42 43	932	Woollings, T. (2017). Variability in seasonal forecast skill of Northern Hemisphere
44 45 46	933	winters over the twentieth century. Geophysical Research Letters, 44: 5729-5738, doi:
47 48 49	934	10.1002/2017GL073736
50 51	935	Ossó, A., Sutton, R., Shaffrey, L., & Dong, B. 2018. Observational evidence of European
52 53	936	summer weather patterns predictable from spring. Proceedings of the National Academy
54 55 56	937	of Sciences, 115: 59-63, doi: 10.1073/pnas.1713146114
57 58 59 60	938	Palmer, T.N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Délécluse, P.,

2 3 4	939	Thomson, M.C. (2004). Development of a European multimodel ensemble system for
5 6	940	seasonal-to-interannual prediction (DEMETER). Bulletin of the American
7 8 9	941	Meteorological Society 85: 853-872, doi: 10.1175/BAMS-85-6-853.
10 11 12	942	Petoukhov, V., & Semenov, V.A. (2010). A link between reduced Barents-Kara sea ice and
13 14	943	cold winter extremes over northern continents. Journal of Geophysical Research, 115:
15 16 17	944	D21111, doi: 10.1029/2009JD013568
18 19 20	945	Rao, J., & Ren, R. (2016a). Asymmetry and non-linearity of the influence of ENSO on the
21 22	946	northern winter stratosphere: 1 observations. Journal of Geophysical Research:
23 24	947	Atmospheres, 121: 9000-9016, doi: 10.1002/2015JD024520
25 26		
27 28	948	Rao, J., & Ren, R. (2016b). Asymmetry and non-linearity of the influence of ENSO on the
29 30	949	northern winter stratosphere: 2. Model study with WACCM. Journal of Geophysical
31 32 33	950	Research, 121: 9017-9032, doi: 10.1002/2015JD024521
34 35 26	951	Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P.,
30 37 38	952	Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night
39 40	953	marine air temperature since the late nineteenth century. Journal of Geophysical
41 42 43	954	Research: Atmospheres, 108(D14): 4407. doi: 10.1029/2002JD002670
44 45 46	955	Riddle, E.E., Butler, A.H., Furtado, J.C., Cohen, J.L., & Kumar, A. (2013). CFSv2 ensemble
47 48	956	prediction of the wintertime Arctic Oscillation. Climate Dynamics, 41: 1099-1116, doi:
49 50	957	10.1007/s00382-013-1850-5
51 52 53	958	Robinson, D.A., & Estilow, T.W. NOAA CDR program 2012. Accessed 5th March 2017.
54 55	959	NOAA Climate Data Record (CDR) of Northern Hemisphere (NH) Snowcover extent
56 57 58 59 60	960	(SCE), v01r01. NOAA National Climatic Data Center. doi: 10.7289/V5N014G9

∠ 3 4	ç
5 6	ç
7 8 9	ç
) 10 11	ç
12 13	ç
14 15 16	ç
17 18	ç
19 20	ç
21 22 23	ç
23 24 25	ç
26 27	ç
28 29	ç
30 31 32	ç
33 34	ç
35 36 37 38	C
39 40 41	ç
41 42 43	ç
44 45	ç
46 47 48	ç
40 49 50	ç
51 52	ç
53 54	ç
55 56 57	ç
58 59	

961 Robock, A., & Mao, J. (1995): The volcanic signal in surface temperature observations.
962 Journal of Climate, 8: 1086-1103.

- 963 Rodwell, M.J., Rowell, D.P., & Folland, C.K. 1999. Oceanic forcing of the wintertime North
 964 Atlantic Oscillation and European Climate. Nature, 398: 320:323.
- ² 965 Scaife, A.A., Ineson, S., Knight. J.R., Gray, L., Kodera, K., & Smith, D.M. (2013). A
- 5 966 mechanism for lagged North Atlantic climate response to solar variability. Geophysical
- 967 Research Letters 40(2): 434-439. doi: 10.1002/grl.50099
- 968 Scaife, A.A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R.T., Dunstone, N., ...
- 969 Williams, A. (2014). Skillful long-range prediction of European and North American
- 970 winters. Geophysical Research Letters 41: 2514-2519. doi: 10.1002/2014GL059637
- ⁶ 971 Scaife, A.A., Comer, R.E., Dunstone, N.J., Knight, J.R., Smith, D.M., MacLachlan, C., ...
 - 972 Slingo, J. (2017). Tropical rainfall, Rossby waves and regional winter climate
- 973 predictions. Quarterly Journal of the Royal Meteorological Society, 143: 1-11, doi:
- ³ 974 10.1002/qj.2910
 - 975 Schrieber, T. & Schmitz, A. (2000). Surrogate time series. Physica D 142: 346-382
- 976 Screen, J.A. (2017). Simulated atmospheric response to regional and pan-Arctic sea ice loss.
 977 30: 3945-3962, doi: 10.1175/JCLI-D-16-0197.1
- 978 Screen, J.A., Deser, C., Smith, D.M., Zhang, X., Blackport, R., Kushner, P.J., ...Sun, L.
- 979 (2018). Consistency and discrepancy in the atmospheric response to Arctic sea ice loss
 980 across climate models. Nature Geoscience. 11: 155-163, doi: 10.1038/s41561-018-
- ⁵¹ 981 0059-y

60

982 Smith, D.M., Scaife, A.A., Eade, R., & Knight, J.R. (2016). Seasonal to decadal prediction of
983 the winter North Atlantic Oscillation: emerging capability and future prospects.

2		
3 4	984	Quarterly Journal of the Royal Meteorological Society 142: 611-617. doi:
5 6 7	985	10.1002/qj.247
/ 8 0	986	Stenchikov, G., Hamilton, L., Stouffer, R.J., Robock, R., Ramaswamy, V., Santer, B., &
9 10 11	987	Graf, HF. (2006). Arctic Oscillation response to volcanic eruptions in the IPCC AR4
12 13	988	climate models. Journal of Geophysical Research, 111: D07107, doi:
14 15	989	10.1029/2005JD006286
16 17 19	990	Stendel, M., van den Besselaar, E., Hannachi, A., Kent, E.C., Lefebvre, C., Schenk, F.,
19 20	991	Woollings, T. (2016). Recent change-atmosphere in North Sea Region Climate
21 22	992	Change Assessment (eds: Quante, M., Coliijn, F.). Regional Climate studies, doi:
23 24	993	10.1007/978-3-319-39745-0_2
25 26 27	994	Stockdale, T.N., Molteni, F., & Ferranti, L. (2015). Atmospheric initial conditions and the
28 29	995	predictability of the Arctic Oscillation. Geophysical Research Letters, 42: 1173-1179.
30 31 22	996	doi: 10.1002/2014GL062681
32 33 34	997	Svensson, C., Brookshaw, A., Scaife, A.A., Bell, V.A., Mackay, .JD., Jackson, C.R.,
35 36	998	Stanley, S. (2015). Long-range forecasts of UK winter hydrology. Environmental
37 38	999	Research Letters 10: 064006. doi: 10.1088/1748-9326/10/6/064006
39 40 41	1000	Takahashi, K., & Dewitte, B. (2016). Strong and moderate non-linear El Niño regimes.
41 42 43	1001	Climate Dynamics, 46: 1627-1645, doi: 10.1007/s00382-015-2665-3
44 45	1002	Tseng, KC., Barnes, E.A., & Maloney, E.D. (2018). Prediction of the midlatitude response
46 47 48	1003	to strong Madden-Julian Oscillation events on S2S time scales. Geophysical Research
40 49 50	1004	Letters, 45: 463-470, doi: 10.1002/2017GL075734
51 52	1005	Tyrell, N.L., Karpechko, A.Y., Räisänen, P. (2018). The influence of Eurasian snow extent
53 54	1006	on the northern extratropical stratosphere in a QBO resolving model. Journal of
55 56 57 58 59 60	1007	Geophysical Research : Atmospheres, 123: 315-328, doi: 10.1002/2017JD27378

3 4	1008	Thornton, H.E., Scaife, A.A., Hoskins, B.J., & Brayshaw, D.J. (2017). The relationship
5 6	1009	between wind power, electricity demand and winter weather patterns in Great Britain.
/ 8 0	1010	Environmental Research Letters 12: 064017. doi: 10.1088/1748-9326/aa69c6.
9 10 11	1011	Toniazzo, T., & Scaife, A.A. (2006). The influence of ENSO on winter North Atlantic
12 13	1012	climate. Geophysical Research Letters, 33: L24704, doi: 10.1029/2006GL027881
14 15	1013	Vallis, G.K., & Gerber, E.P. (2008). Local and hemispheric dynamics of the North Atlantic
16 17 18	1014	Oscillation, annular patterns and the zonal index. Dynamics of Oceans and
19 20 21	1015	Atmospheres, 44: 184-212. doi: 10.1016/j.dynatmoce.2007.04.003
22 23	1016	Vihma, T., Graversen, R., Chen, L., Handorf, D., Skific, N., Francis, J.A., Overland, J.E.
24 25 26	1017	Effects of the tropospheric large-scale circulation on European winter temperature
27 28	1018	during the period of amplified Arctic warming. International Journal of Climatology.
29 30 31	1019	submitted
32 33 34	1020	Wang, L., Ting, M., & Kushner, P.J. (2017). A robust empirical prediction of winter NAO
35 36 37	1021	and surface climate. Science Reports, 7: 279, doi: 10.1038/s41598-017-00353-y
38 39	1022	Wei, HL., Zhu, D., Billings, S.A., & Balikhin, M.A. (2007). Forecasting the geomagnetic
40 41	1023	activity of the Dst index using multiscale radial basis function networks. Advances in
42 43 44	1024	Space Research 40: 1863–70.
45 46	1025	Wei, HL., Billings, S. A., Zhao, Y., & Guo, L. (2010). An adaptive wavelet neural network
47 48	1026	for spatio-temporal system identification. Neural Networks, 23: 1286–99.
49 50 51	1027	Weisheimer, A., Schaller, N., O'Reilly C., MacLeod, D.A., & Palmer, T. (2016).
52 53	1028	Atmospheric seasonal forecasts of the twentieth century: multidecadal variability in
54 55	1029	predictive skill of the winter North Atlantic Oscillation (NAO) and their potential for
56 57 58	1030	extreme event attribution. Quarterly Journal of the Royal Meteorological Society, 143:
59 60	1031	917-926, doi: 10.1002/qj.2976

1 2		
3 4	1032	Wilks, D.S. (2011). Statistical Methods in the Atmospheric Sciences. Third edition. Academic
5 6 7	1033	Press, Oxford.
8 9 10 11	1034	Woollings, T., Hannachi, A., & Hoskins, B. (2010a). Variability of the North Atlantic eddy-
11 12	1035	driven jet stream. Quarterly Journal of the Royal Meteorological Society, 136: 856-868,
13 14 15 16 17	1036	doi: 10.1002/qj.625
	1037	Woollings, T., Lockwood, M., Masato, G., Bell, C., & Gray, L. (2010b). Enhanced signature
18 19	1038	of solar variability in Eurasian winter climate. Geophysical Research Letters, 37:
20 21	1039	L20805, doi: 10.1029/2010GL044601
22 23 24	1040	Woollings, T., Barnes, E., Hoskins, B., Kwon, YO., Lee, R.W., Li, C., Williams, K.
24 25 26	1041	(2018). Daily to decadal modulation of jet variability. Journal of Climate, 31: 1297-
27 28 29 30 31 32 33	1042	1314, doi: 10.1175/JCLI-D-17-0286.1
	1043	Woollings, T., Franzke, C., Hodson, D.L.R., Dong, B., Barnes, EA, Raible, C.C, & Pinto,
	1044	J.G. (2015). Contrasting interannual and multidecadal NAO variability. Climate
34 35	1045	Dynamics 45: 539-556, doi: 10.1007/s00382-104-2237-y
36 37	1046	Wu, Y., & Smith, K. (2016). Response of northern hemisphere midlatitude circulation to
38 39 40	1047	Arctic Amplification in a simple Atmospheric General Circulation model. Journal of
41 42	1048	Climate, 29: 2041-2058, doi: 10.1175/JCLI-D-15-0602.1
43 44	1049	Xoplaki, E., González-Rouco, J.F., Luterbacher, J., & Wanner, H. (2004). Wet season
45 46 47	1050	Mediterranean precipitation variability: influence of large-scle dynamics and trends.
47 48 49	1051	Climate Dynamics, 23: 63-78
50 51	1052	Yua V. Higging W. & Kousky V. (2002). Influences of the Madden Julian Oscillations on
52 53	1032	Aue, T., Higgins, W., & Kousky, V. (2002). Influences of the Madden-Junan Oscillations on
54 55	1053	temperature and precipitation in North America during ENSO-neutral and weak ENSO
56 57	1054	winters. A Workshop on Prospects for Improved Forecasts of Weather and Short-term
58 59 60		

1 ว		
2 3 4	1055	Climate Variability on Subseasonal (2 week to 2 month) Time Scales, NASA/Goddard
5 6 7	1056	Space Flight Center, April 16-18, 2002.
8 9 10	1057	Yu, B., Lin, H. (2016). Tropical atmospheric forcing of the wintertime North Atlantic
11 12	1058	Oscillation. Journal of Climate, 29: 1755-1772, doi: 10.1175/JCLI-D-15-0583.1
13 14 15	1059	
15 16 17	1060	Zhang, P., Wu, Y., & Smith, K.L. (2018). Prolonged effect of the stratospheric pathway in
18 19	1061	linking Barents-Kara Sea sea ice variability to the midlatitude circulation in a simplified
20 21 22	1062	model. Climate Dynamics, 50: 527-539, doi: 10.1007/s00382-017-3624-y
22 23 24 25 26 27 28 29 31 23 34 35 37 38 30 41 42 34 45 46 47 48 951 52 53 45 56 57 89 60	1063 1064 1065 1066 1067 1068 1069	

3	1070
4	10/0
5	
6	
/	
ð	
9 10	
10	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24 25	
25	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
3/ 20	
30	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49 50	
50 E1	
כו בי	
52 52	
55	

for per peries



Figure 1 Example of hindcasts (solid) and forecasts (dashed) using linear (LIN) and polynomial (POLY) NARMAX models, derived from the 1980 training period for the station-based NAO.

186x124mm (300 x 300 DPI)



Figure 2. Out-of-sample forecasts compared with observations for linear (LIN) and polynomial (POLY) NARMAX models. Note there is no polynomial model for HPC80 and the observed station NAO is missing for 2018 in (b) and (d).

207x141mm (300 x 300 DPI)





Figure 3. Predictors selected for the NAO models based on the 1980 training period. N3.4I =El Niño 3.4 discontinuous index; EIR=tropical East Indian Ocean rainfall, NAH=North Atlantic Horseshoe SST pattern; snow=Eurasian snowcover; BK ice=Barents-Kara Sea ice; GIN SST= Greenland-Iceland Norwegian SST; ES/L= East Siberian/Laptev Sea ice; TAR= tropical Atlantic rainfall; Bering ice=Bering Sea ice; MJO8/9=Phase 8 Madden-Julian Oscillation; Barents SLP=Barents Sea regional SLP; WPR=tropical West Pacific Ocean rainfall; SST gradient=North Atlantic SST gradient; LAB ice=Labrador Sea ice; lead 4 SS =sunspot cycle leading by 4 years. Where predictor month is not specified, it is indicated by a white hexagon linked to the variable. This is used where different models select different months of a common predictor.

172x225mm (300 x 300 DPI)



Table I. Summary of datasets used. For more detailed information, see Table S1.

dataset	Obtained from	Variable used
Hurrell PC_NAO	https://climatedataguide.ucar.edu/climate-data/hurrell- north-atlantic-oscillation-nao-index-pc-based	DJF NAO index
Station-based NAO	https://rda.ucar.edu/datasets/ds570.0	MSLP, Azores and Iceland
АМО	www.climexp.knmi.nl	HadSST3.1.1 SST
HadISST1	www.climexp.knmi.nl	SST, SIC
GPCPv2.3	www.climex.knmi.nl	Tropical precipitation
Carbon dioxide	www.esrl.noaa.gov/gmd/ccgg/trends/data.html	Annual CO ₂ level
QBO	www.geo.fu- berlin.de/en/met/ag/strat/produkte/qbo/index.html	Mean zonal wind, 30hPa
sunspots JRA-55 SPV_T100	http://sidc.oma.be Blanca Ayarzaguena, University of Exeter	Sunspot number Temperature 100hPa
NCEP/NCAR SLP	www.climexp.knmi.nl	Sea level pressure
Rutgers Global Snow Lab snow cover extent	https://climate.rutgers.edu/snowcover/docs.php?target =datareq	Snow cover extent
HadCRUT4.6	https://www.metoffice.gov.uk/hadobs/hadcrut4/	2m temperature anomaly
MJO Indices	www.cpc.ncep.noaa.gov/products/precip/CWlink/dail y_mjo_index/pentad.html	200hPa velocity potential anomalies

a) stat80LIN		Model Parameter		
	Model Term	7-term	8-term	9-term
1 July N3.4I		-0.77	-0.39	-0.23
2 July tropica	l Atlantic rainfall	-0.10	-0.16	-0.22
3 November	Bering Sea ice	0.40	0.46	0.55
4 October Ba	rents-Kara Sea ice	0.53	0.67	0.73
5 October Ba	rents Sea SLP	-0.30	-0.38	-0.37
6 November	East Siberian-Laptev Sea ice	-0.32	-0.29	-0.34
7 November MJO phase 9		-0.38	-0.53	-0.63
8 July tropical East Indian Ocean rainfall			0.29	0.30
9 lead 4 year s	sunspot cycle			-0.18
Individual M	odel Performance			
Training	MAE	0.29	0.27	0.22
data	RMSE	0.37	0.31	0.27
	correlation	0.92	0.94	0.95
Testing data	MAE	1.02	1.25	1.07
0	RMSE	1.14	1.38	1.26
	correlation	0.77	0.83	0.81

b)HPC80LIN		Mo	del Param	eter
	Model Term	7-term	8-term	9-term
1 October Bar	ents Kara Sea ice 💦 💦	0.91	0.94	0.85
2 October Bar	ents Sea SLP	-0.42	-0.47	-0.48
3 November E	Bering Sea ice	0.57	0.35	0.32
4 November E	East Siberian-Laptev Sea ice	-0.44	-0.47	-0.44
5 constant		0.28	0.27	0.28
6 October GIN	N SST	0.33	0.42	0.39
7 November N	AJO phase 8	-0.46	-0.62	-0.57
8 October Beri	ng Sea ice		0.35	0.37
9 August tropi	cal Atlantic rainfall			0.23
Individual Mo	odel Performance			
Training	MAE	0.33	0.29	0.23
data	RMSE	0.41	0.35	0.28
	correlation	0.93	0.95	0.97
Testing data	MAE	0.33	0.30	0.36
	RMSE	0.49	0.47	0.48
	correlation	0.92	0.88	0.88

2	
2	
2	
4	
5	
6	
7	
, ,	
8	
9	
1	0
1	1
1	2
1	2
I	3
1	4
1	5
1	6
1	7
1	, 0
1	ð
1	9
2	0
2	1
2	2
2	2
2	3
2	4
2	5
2	6
2	7
2	, ^
2	ð
2	9
-	-
3	0
3	0 1
333	0 1 2
333	0 1 2
333	0 1 2 3
- 3 3 3 3 3	9 1 2 3 4
- 3 3 3 3 3 3 3	0 1 2 3 4 5
- 3 3 3 3 3 3 3 3 3	0 1 2 3 4 5 6
- 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	0 1 2 3 4 5 6 7
- 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	0 1 2 3 4 5 6 7 0
	0 1 2 3 4 5 6 7 8
- 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	0123456789
- 3 3 3 3 3 3 3 3 3 3 3 3 4	01234567890
- 3 3 3 3 3 3 3 3 3 4 4	012345678901
- 3 3 3 3 3 3 3 3 3 4 4 4	0123456789012
- 3 3 3 3 3 3 3 3 3 4 4 4 4	01234567890123
- 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4	01234567890123
- 3 3 3 3 3 3 3 3 3 4 4 4 4 4	012345678901234
- 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4	0123456789012345
- 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4	01234567890123456
- 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4	012345678901234567
- 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4	0123456789012345678
_ 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4	01234567890123456780
- 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4	01234567890123456789
- 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 5	012345678901234567890
- 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 5 5	0123456789012345678901
- 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 5 5 5	01234567890123456789012
- 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 5 5 5 5	012345678901234567890123
- 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 5 5 5 5	0123456789012345678901234
- 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 5 5 5 5	0123456789012345678901234
- 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 5 5 5 5	01234567890123456789012345
_ 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 5 5 5 5	012345678901234567890123456

1

-) -4-45(1 IN		N/-	J.1 D	4
c) stat56LIN		NIO(del Parame	ter
	Model Term	6-term	7 -term	8-term
1 May Green	land Sea ice	-0.15	-0.16	-0.14
2 September	NAH	-0.27	-0.27	-0.27
3 July N3.4I		-0.61	-0.65	-0.61
4 October Ba	rents-Kara Sea ice	0.37	0.41	0.44
5 October Ba	rents Sea SLP	-0.23	-0.22	-0.25
6 November	tropical West Indian Ocean SST	0.14	0.16	0.14
7 November E	East Siberian-Laptev Sea ice		-0.16	-0.15
8 October Gre	enland Sea ice			-0.08
Individual M	odel Performance			
Training	MAE	0.52	0.51	0.51
data	RMSE	0.66	0.65	0.64
	correlation	0.65	0.67	0.69
Testing data	MAE	1.27	0.95	1.03
	RMSE	1.38	1.07	1.14
	correlation	0.56	0.74	0.73
d) HPC56LIN	N N	Mo	odel Parame	eter
	Model Term	5-term	6-term	7-term
1 October NA	АН	-0.36	-0.33	-0.28
2 November	SPV	-0.38	-0.29	-0.28
3 October SP	V	0.41	0.34	0.36
4 July Greenl	and Sea ice	-0.12	-0.22	-0.27
5 October Ba	rents Sea SLP	-0.31	-0.41	-0.41
6 October Ba	rents-Kara Sea ice		0.44	0.44
7 November	Bering Sea ice			0.18
Individual M	odel Performance			
Training	MAE	0.68	0.61	0.58
data	RMSE	0.79	0.72	0.69
	correlation	0.75	0.80	0.81
Testing data	MAE	0.83	1.22	1.29
	DMSE	0.00	1 4 4	156

Table II. Selected predictors, model coefficients and verification statistics for linear NARMAX models.

correlation

0.41

0.49

0.42

a) models	MA	NE	RM	SE	correl	ation	MSESS	D
1980 linear	training	testing	training	testing	training	testing		
stat80LIN	0.25	1.11	0.29	1.26	0.95	0.82	-0.96	0.80
HPC80LIN	0.27	0.29	0.31	0.48	0.96	0.90	0.82	0.75
1980 polynomial								
stat80POLY	0.35	0.43	0.41	0.52	0.89	0.92	0.68	0.89
b) models	MA	E	RM	SE	corre	lation	MSESS	D
1956 linear	training	testing	training	testing	training	testing		
stat56LIN	0.51	1.08	0.64	1.19	0.76	0.69	-0.44	0.80
HPC56LIN	0.61	1.10	0.71	1.31	0.81	0.46	-0.28	0.68
1956 NARMAX								
stat56POLY	0.40	0.68	0.50	0.81	0.87	0.77	0.34	0.79
HPC56POLY	0.40	0.59	0.52	0.70	0.90	0.73	0.68	0.79

Table III. Verification statistics for averaged linear and polynomial NARMAX models, for a) periods 1980-2018 and b) 1956-2018. Bold figures show if polynomial outperforms the linear model.

tatistics ... d b) 1956-2018. ..

a)		Model Parameter			
stat80POLY	Model Term	4-term 5-term		n 6-term	
1 Jul tropical	West Pacific Rainfall* Nov Barents Sea SLP	-0.34	-0.31	-0.33	
2 Oct Labrad	lor Sea ice LAB * Oct NAH	0.63	0.64	0.58	
3 Sep snow *	* Nov_SST gradient	0.48	0.55	0.56	
4 Oct Bering	Sea ice * Nov NAH	0.22	0.22	0.27	
5 Sep Labrad	lor Sea ice * Oct MJO phase 8		0.51	0.51	
6 Jul tropical	West Pacific Rainfall * Sep snow			0.33	
Individual M	Iodel Performance				
Training	MAE	0.40	0.36	0.33	
data	RMSE	0.48	0.43	0.38	
	correlation	0.84	0.88	0.91	
Testing data	MAE	0.43	0.53	0.48	
-	RMSE	0.50	0.58	0.62	
	correlation	0.90	0.88	0.95	

b)			Paramet	er
stat56POLY	Model Term	8-term	9-term	10-term
1 Jul tropica	l West Indian Ocean SST * Sep GIN SST	-0.06	-0.05	-0.07
2 Nov Berin	g Sea ice	-0.33	-0.67	-0.49
3 Sep NAH	Ciborian Lontov Socioo * Nov CIN SST	-0.28	-0.28	-0.27
5 Jul Greenl	and Sea ice * Jul Greenland Sea ice	-0.25	-0.19	-0.18
6 Oct Barent	ts-Kara Sea ice	-0.03	-0.03	-0.03
7 Oct Barent	ts Sea SLP * Nov GIN SST	0.37	0.33	0.32
9 Jul N3 4I	is Sea SLP * Sen Canadian Archinelago-Baffin, sea ice	-0.21	-0.22	-0.24
10 Jul N3.4I *	* Sep NAH	-0.22	-0.20	-0.21
			0.19	0.18
				0.46
Individual M	odel Performance			
Training	RMSE	0.53	0.51	0.49
data	MAE	0.43	0.40	0.40
	Correlation	0.85	0.86	0.87
Testing	RMSE	0.74	0.85	0.85
data	MAE	0.67	0.68	0.69
	Correlation	0.70	0.63	0.82

1
2
3
4
5
5
0
/
8
9
10
11
12
13
14
15
16
17
17 10
10 10
19
20
21
22
23
24
25
26
27
27
20
29
30
31
32
33
34
35
36
37
38
39
40
40 //1
40
42
43
44
45
46
47
48
49
50
51
52
52
5/
55
22
20
5/

c)			Parameter	
HPC56POL	Y Model Term	6-term	7-term	8-term
1 Oct Labra	dor Sea ice * Oct NAH	0.45	0.41	0.43
2 Aug N3.4	* lead 3 year sunspot cycle	0.63	0.69	0.73
 3 atmospheric CO₂ * Jun Greenland Sea ice 4 Nov N3.4I * Sep Barents Sea SLP 5 Jun Hudson Bay sea ice * volcanic index 6 Sep GIN SST * Oct SST gradient 7 Sep QBO * Nov Barents Sea SLP 8 Sep N3 4I * Nov GIN SST 		0.04	0.04	0.03
		0.82	0.97	1.19
		1.00	1.00	0.99
		0.25	0.29	0.32
			0.26	0.28
· F				-0.56
Individual N	Model Performance			
Training	RMSE	0.59	0.55	0.49
data	MAE	0.45	0.42	0.37
	Correlation	0.86	0.89	0.91
Testing	RMSE	0.60	0.68	0.75
data	MAE	0.52	0.53	0.57
	Correlation	0.79	0.73	0.68

Table IV. As for Table II, but for polynomial models.

e perez

Graphical Abstract

Complex Systems Modelling for Statistical Forecasting of Winter North Atlantic Atmospheric Variability: a New Approach

Richard J. Hall*, Hua-Liang Wei, Edward Hanna

A new approach to seasonal forecasting based on complex systems modelling is presented.

The focus is on North Atlantic winter atmospheric circulation, specifically the NAO.

Polynomial models show greater skill than linear versions and out-of-sample forecasts show

promising skill, closely matching the observed time series. Potential non-linear interactions

ee perie

between predictors are identified.





