Seasonal predictions: Confidence and multi-decadal variability of skill estimates of the winter NAO

Antje Weisheimer

ECMWF
University of Oxford
In a nonlinear system the growth of initial uncertainty is flow dependent.

The set of initial conditions (black circle) is located in different regions of the attractor in a), b) and c) and leads to different error growth and predictability in each case.
The climate is a chaotic system where the future state of the system can be very sensitive to small differences in the current (initial) state of the system.

In practice, the initial state of the system is always uncertain.

Our forecast models are not perfect in all aspects (e.g. small-scale features such as clouds).

Ensemble forecasting takes into account these inherent uncertainties by running a large number of similar but not identical versions of the model in parallel. The resulting forecasts are expressed in probabilities.
El Niño Southern Oscillation – a coupled atmosphere-ocean mode of variability
El Niño Southern Oscillation – a source of predictability on seasonal timescales

[Map of El Niño Weather Patterns December - February]

[Map of El Niño Weather Patterns June - August]
Observations Forecast

Current state of the atmosphere

Atmospheric model

Coupler

Current state of the ocean

Ocean model

Data Assimilation

Forecast models for seasonal predictions

Forecast Products

Coupled model

Products

Current state of the ocean

Current state of the atmosphere
Seasonal forecasts aim to predict an anomaly from the default climatological probability.

Probability density distributions of a hypothetical climatology and forecast given an observation.

“Ideal” situation

“Real” situation
Climate forecasts are not crucially sensitive to the initial conditions. They are a mixed initial-boundary condition (forcing) problem in a chaotic system.
The North Atlantic Oscillation (NAO)

- Dominant mode of variability on a range of time scales over the North Atlantic-European region
- Typically defined as the 1st EOF of MSLP or Z500
- NAO index: 1st Principal Component or sometimes (mostly for historical reasons) as normalised MSLP difference between Iceland and the Azores
Seasonal forecasts of the winter NAO

Ensemble hindcasts of the NAO index 1993-2012 with the Met Office model (GloSea GA3)

Scaife et al. (GRL 2014)

Siegert et al. (JClim 2016)

Forecasters crack formula to predict long-range weather

Extreme winters will be predicted with greater reliability than before after the world’s best long-term weather forecast model was developed by British scientists, the Met Office said.

The breakthrough may have a substantial impact on the economy, allowing power companies and wind farms to anticipate energy demands while airports and councils can estimate how much grit and anti-freeze is likely to be required.
Sources of predictability

ENSO

Atlantic sub-polar gyre
Oceanic heat content

Kara Sea ice

QBO

GloSea5 MSLP obs

Scaife et al. (GRL 2014)
Seasonal forecasts of the weather and climate over Euro-Atlantic region are difficult due to

- low signal-to-noise ratios in predictability of extratropical atmosphere
- teleconnections from tropical forcings are less direct, and perhaps more manifold, than for other areas in the world
- sample sizes are intrinsically small (mainly limited by number of observed seasons, usually $O(30)$)

Estimates of seasonal predictability, skill and reliability suffer from rather large uncertainties.
Illustrative example of correlation drawbacks after Anscombe (1973):

- Four pairs of x-y variables
- The four y variables have the same mean (=7.5), variance (=4.1) and correlation (=0.82)
- However, distributions of variables are very different

Anscombe’s quartet

Normally distributed, “well behaved”
Not normally distributed, non-linear relationship
Perfect linear relationship, Except for one outlier
No relationship with one outlier

Anscombe (Amer. Statist. 1973)
Atmospheric Seasonal Forecasts of the 20th Century (ASF-20C)

- A new very long data set of seasonal hindcasts to study changes in predictability
- Use of ECMWF’s re-analysis of the 20th Century (ERA-20C) that spans the 110-year period 1900 to 2010 to initialise atmospheric seasonal forecasts with ECMWF’s forecast model
- SSTs and sea-ice are prescribed using HadISSTs
- Seasonal re-forecast experiments over the period 1900-2010
- Large ensemble of 51 perturbed members
- Focus here: 4-month forecast initialised on 1st of Nov each year to cover boreal winter (DJF) season
- More details in Weisheimer et al. (QJRMS 2017) and O’Reilly et al. (GRL 2017)
Global mean 2m temperature forecast anomalies in DJF

DJF global mean 2m temperature in ERA-20C (red) and the re-forecast ensemble of ASF-20C (blue). Uncertainty estimates from the reanalysis and the re-forecast ensemble are shown in orange (full range of the 10-member ensemble) and with blue shades (light blue: full range; darker blue: interquartile 25%-75% range; blue dots: ensemble median), respectively.

Weisheimer et al. (QJRMS 2017)
Multi-decadal variability of NAO forecast skill
- estimates from 30-year moving windows -

Anomaly correlation coefficient of the DJF NAO index between the ensemble mean and ERA-20C computed for moving 30-year windows by one year. Values are plotted at the 15th year of each window. The horizontal line indicates the $t$-test 95% significance level of the correlations and the red vertical bars show 90% confidence intervals estimated from bootstrap re-sampling (1000 times) with replacement for three representative periods.
ROC skill scores and NAO distribution

(a) NAO percentiles (1900-2009)

(b) ROC skill score (1900-2009)

30-year running NAO thresholds (from ERA-20C)
Multi-decadal variability of NAO forecast skill

- persistence and intraseasonal variability -
Seasonal forecasts aim to predict an anomaly from the default climatological probability.

Probability density distributions of a hypothetical climatology and forecast given an observation.
Signal and noise

$x_{m,n}$: variable $x$ with member $m$ and year $n$

Mean: $\bar{x} = \frac{1}{NM} \sum_{n} \sum_{m} x_{m,n}$  

Ensemble mean: $\langle x_n \rangle = \frac{1}{M} \sum_{m} x_{m,n}$

Variance: $VAR_{total} = \frac{1}{NM} \sum_{n} \sum_{m} (x_{m,n} - \bar{x})^2$

$VAR_{total} = VAR_{signal} + VAR_{noise}$  

$\Rightarrow S/N = \frac{VAR_{signal}}{VAR_{noise}}$

$VAR_{signal} = \frac{1}{N} \sum_{n} (\langle x_n \rangle - \bar{x})^2$

$VAR_{noise} = \frac{1}{NM} \sum_{n} \sum_{m} (x_{m,n} - \langle x_n \rangle)^2$

ensemble mean variance  

$\Rightarrow$ “signal”

variance of ensemble members about  

ensemble mean (=spread)  

$\Rightarrow$ “noise”
Correlation skill and signal-to-noise (S/N) ratio

The *expected value* for various measures of skill for seasonal climate predictions is determined by the S/N ratio.

“The expected value, however, is only realized for long verification time series. In practice, the verifications for specific seasons seldom exceed a sample size of 30. The estimates of skill measure based on small verification time series, because of sampling errors, can have large departures from their expected value.”

Probability of expected correlation for a given realised value of skill

*Kumar (MWR 2009)*
The Ratio of Predictable Components (RPC)

Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?

Rosie Eade¹, Doug Smith¹, Adam Scaife¹, Emily Wallace¹, Nick Dunstone¹, Leon Hermanson¹, and Niall Robinson¹

¹Met Office Hadley Centre, Exeter, UK

Abstract Seasonal-to-decadal predictions are inevitably uncertain, depending on the size of the predictable

\[
RPC = \frac{PC_{obs}}{PC_{model}} \geq \frac{r(\text{obs, ens mean})}{\sqrt{VAR_{signal}/VAR_{total}}}
\]

Predictable Components (PCs) … predictable part of the total variance

observed \( PC_{obs} \) … estimated from explained variance = \( r^2(\text{obs, ensmean}) \)

model \( PC_{model} \) … estimated from ratio of signal variance to total variance

Eade et al. (GRL 2014)
Perfect model ensembles and potential skill

What is a perfect model ensemble?

- Perfect sampling of the underlying probability distribution of the true state
- Over a large number of forecasts, the statistical properties of the truth are identical to the statistical properties of a member of the ensemble
- I.e., the truth is indistinguishable from the ensemble

→ Replace observation with ensemble member
Perfect model ensembles and potential skill

Properties of a perfect model ensemble

- Time-mean ensemble spread == RMSE of ensemble mean forecast
- \( r \) (perfect model) = corr(ens mean, ens members) \( \rightarrow \) “potential skill”
- RPC of a perfect ensemble == 1
- Observed correlation \( \leq \) perfect model correlation ??
Perfect model ensembles and potential skill

Implications for non-perfect ensembles

- Time-mean ensemble spread ≠ RMSE of ensemble mean forecast
  - ensemble spread < RMSE  → ensemble is underdispersive
  - ensemble spread > RMSE  → ensemble is overdispersive

- RPC ≠ 1
  - RPC > 1  → underconfidence; $VAR_{signal}$ too small, model underestimates predictability of real world, observed correlation > perfect model correlation
  - RPC < 1  → overconfidence; observed correlation < perfect model correlation model predictability is larger than in real world
The signal-to-noise “conundrum” or “paradox”

\[
\text{RPC of DJF MSLP in GloSea5 } \quad (\text{RPC} = \frac{\text{PC}_{\text{obs}}}{\text{PC}_{\text{model}}})
\]

Ensemble hindcasts of the AO index 1981-2010 with the ECMWF System 4

\[r=0.61 \quad S/N=0.1\]

The real world seems to have higher predictability than the model.
The signal-to-noise “conundrum” or “paradox”

- 95% uncertainty intervals on $r=0.62$ are [0.19;0.68]
- S/N_{obs} is larger than S/N_{model}
  - raw forecasts should not be taken as representative scenarios of the observations (not exchangeable)
  - predictable signal in model too weak
- The particular 20-yr period is unusual and produces higher-than-normal correlation skill
What is the empirical evidence on shorter forecast ranges that
i) models are overdispersive
    and/or
ii) model estimates of predictability are too low (underconfidence)?
Spread-skill relationship in medium-range forecasts

ECMWF model for NHem Extratropics Z500 DJF 2016/17

courtesy David Richardson (ECMWF)
ECMWF model for NHem Extratropics Z500

RMSE

spread
courtesy David Richardson (ECMWF)
Spread-skill relationship in monthly forecasts

Z500 projections on EOF 1 (NAO) during winter 1999-2010 from S2S re-forecasts

courtesy Laura Ferranti (ECMWF)
Z500 projections on EOF 1 (NAO) during winter 1999-2010 from S2S re-forecasts

Spread-skill relationship in monthly forecasts
courtesy Laura Ferranti (ECMWF)
1st Nov start date 1981-2010      Z500 seasonal forecasts S4     51 ens members

rmse spread spread/RMSE

no stochastic physics
Week 6

RMSE

spread

spread/RMSE
Correlations in medium-range forecasts (TIGGE models)

real world
model world
(potential or perfect model skill)

courtesy Mio Matsueda (Uni Oxford)
Correlations in monthly forecasts

1995/96 – 2016/17 hindcasts with 11 ensemble members
CY41R1 T255L60 atmosphere only experiments with observed SSTs

courtesy Dan Rowlands (Cumulus)
Correlations in monthly forecasts

1995/96 – 2016/17 hindcasts with 11 ensemble members
CY41R1 T255L60  atmosphere only experiments with observed SSTs

courtesy Dan Rowlands (Cumulus)
Correlations in week 4

courtesy Dan Rowlands (Cumulus)
Seasonal NAO predictions in the EUROSIP models

\[ RPC = \frac{r}{\sqrt{\frac{VAR_{signal}}{VAR_{total}}}} \]

(d) ACC

(e) Variance ratio

(f) RPC

- Met Office GloSea5-GA3
- Met Office GloSea5-GA6
- Météo France S3
- Météo France S4
- ECMWF S4
- NCEP S2

courtesy Laura Baker (Uni Reading)
Spread-RMSE relationship in ASF-20C

courtesy Dave MacLeod (Uni Oxford)
NAO skill and RPC in ASF-20C
using 30-year moving windows across the 110-year period

Weisheimer et al. (QJRMS 2017)
Contributions to covariance in ASF-20C

\[ r = \frac{1}{N} \sum_{n} \frac{x'_\text{obs} x'_\text{ensmean}}{\sigma_{\text{obs}} \sigma_{\text{ensmean}}} \]

courtesy Damien Decremer (ECMWF)
Z500 anomalies for largest contributions to covariance

courtesy Damien Decremer (ECMWF)
1st Nov start date 1981-2010  Z500 seasonal forecasts S4  51 ens members

correlation skill  DJF mean  perfect model correlation skill

cor correlation skill minus perfect model correlation skill
1st Nov start date 1981-2010  MSLP seasonal forecasts S4  51 ens members

correlation skill  DJF mean  perfect model correlation skill

correlation skill minus perfect model correlation skill
1st Nov start date 1981-2010  Z50  seasonal forecasts S4  51 ens members

correlation skill

DJF mean

perfect model correlation skill
Underconfidence in seasonal forecasts?

\[ \text{corr(}\text{obs, ensmean}\text{)} \text{ minus corr(}\text{ens, ensmean}\text{)} \]

S4

ASF-20C

SEAS 5

courtesy Damien Decremer (ECMWF)
Underconfidence in seasonal forecasts (ASF-20C)?

corr(obs, ensmean) minus corr(ens, ensmean)
Arctic amplification?

Z500 trend 1981-2009

courtesy Damien Decremer (ECMWF)
Role of the Tropics

as a major source of predictability on longer timescale?
Anomaly Correlation of the NAO in monthly forecast experiments
1995/96 – 2016/17 hindcasts with 11 ensemble members
CY41R1 T255L60 atmosphere only experiments with observed SSTs

courtesy Dan Rowlands (Cumulus)
Anomaly Correlation of the NAO in monthly forecast experiments
1995/96 – 2016/17 hindcasts with 11 ensemble members
CY41R1 T255L60  atmosphere only experiments with observed SSTs

actual skill (tropical relaxation)
potential/perfect model skill

courtesy Dan Rowlands (Cumulus)
Role of the Tropics in ASF-20C

- corr(obs, ensmean)
- corr(ens, ensmean) = perfect model

control

tropical relaxation
Why is RPC > 1 during decadal periods when NAO > 0?

→ Tim Palmer’s regime hypothesis
Circulation regimes over the Euro-Atlantic area

Dawson et al. (GRL 2012)
Effect of non-linear regime error

\[ \rho_{obs}(X) \quad \rho_{model}(X) \]

\[ +\text{NAO} \]

\[ \mu_{model} = \mu_{obs} = 0 \]
\[ \sigma_{model} = \sigma_{obs} = 1 \]

Model’s first two moments agree perfectly with observations

\[ \rho_{obs}(X|X > 0) \quad \rho_{model}(X|X > 0) \]

\[ +\text{NAO} \]

\[ \mu_{obs} = 1.0 \quad \mu_{model} = 0.8 \]
\[ \sigma_{obs} = 0.3 \quad \sigma_{model} = 0.6 \]

\[ \sigma_{model} > \sigma_{obs} \]

courtesy Tim Palmer (Uni Oxford)
Summary and Conclusions (I)

Dynamical predictions of the winter NAO remain a challenge.

Distinct multi-decadal variability of the winter NAO forecast skill:

- No general evidence that model cannot predict negative NAO winters but asymmetry in predictive skill of NAO phase
- Lack of skill in mid-Century: Flow-dependent non-linear model error or lower intrinsic predictability of the atmosphere?

Mid-Century period stands out as an important period on which to test the performance of future seasonal forecast systems.

Achieving good forecast skill for recent decades, with predominantly positive NAO winters, is no guarantee for a similar good performance in the future during possible periods with more negative NAO winters.
Summary and Conclusions (II)

It has recently been suggested that predictability estimates of seasonal forecast models of the winter NAO underestimate the real world predictability. These findings are based on multi-decadal simulations when the NAO was predominantly in its positive phase.

Spread-RMSE diagnostics across forecast time scales give no indication of over-dispersive behaviour. Correlation skill does indicate situations with perfect model skill > actual skill on time scales of ~14d onwards.

However, correlation measures suffer from large uncertainties due to small samples taken over specific long-term (decadal-centennial) climate regimes.

Long seasonal hindcasts covering the full 20th Century allow to put the predictability situation of the recent decades into a longer climate context. Over the entire period RPC~1.

Recent decades see high levels of NAO skill and a tendency to underestimate the real skill. Previous climate periods do not show indications for such a behaviour.

Preferred flow pattern of most skillful years point towards strong Z500 anomalies over Greenland and parts of the Arctic. Observed predictability is higher throughout the atmospheric column in these regions but only during recent decades.

“Conundrum” (or paradox) is a plausible manifestation of model deficiencies in representing non-linear circulation regimes.
References


