White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/132001/

Version: Accepted Version

Article:
Ayala-Solares, J.R., Wei, H.L. and Bigg, G.R. (2018) The variability of the Atlantic meridional circulation since 1980, as hindcast by a data-driven nonlinear systems model. Acta Geophysica, 66. pp. 683-695. ISSN: 1895-6572

https://doi.org/10.1007/s11600-018-0165-7

This is a post-peer-review, pre-copyedit version of an article published in Acta Geophysica. The final authenticated version is available online at: http://dx.doi.org/10.1007/s11600-018-0165-7.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

This paper was accepted for publication in Acta Geophysica on 24 May 2018
https://doi.org/10.1007/s11600-018-0165
https://link.springer.com/content/pdf/10.1007%2Fs11600-018-0165-7.pdf

The variability of the Atlantic Meridional Circulation Since 1980, as Hindcast by a Data-driven Nonlinear Systems Model

J. R. Ayala-Solares[1], Hua-Liang Wei[1], and G. R. Bigg[2]

[1]Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, U. K.,
[2]Department of Geography, University of Sheffield, Sheffield, U. K.

Email: { jrayalasolares1, w.hualiang, grant.bigg} @sheffield.ac.uk

Corresponding authors:

Hua-Liang Wei (w.hualiang@sheffield.ac.uk), Grant Bigg (grant.bigg@sheffield.ac.uk)

Abstract

The Atlantic Meridional Overturning Circulation (AMOC), an important component of the climate system, has only been directly measured since the RAPID array's installation across the Atlantic at 26°N in 2004. This has shown that the AMOC strength is highly variable on monthly timescales, however, after an abrupt, short-lived, halving of the strength of the AMOC early in 2010, its mean has remained ~ 15% below its pre-2010 level. To attempt to understand the reasons for this variability, we use a control systems identification approach to model the AMOC, with the RAPID data of 2004-2017 providing a trial and test data set. After testing to find the environmental variables, and systems model, that allow us to best match the RAPID observations, we reconstruct AMOC variation back to 1980. Our reconstruction suggests that there is inter-decadal variability in the strength of the AMOC, with periods of both weaker flow than recently, and flow strengths similar to the late 2000s, since 1980. Recent signs of weakening may therefore not reflect the beginning of a sustained decline. It is also shown that there may be predictive power for AMOC variability of around 6 months, as ocean density contrasts between the source and sink regions for the North Atlantic Drift, with lags up to 6 months, are found to be important components of the systems model.

1. Introduction

The Atlantic Meridional Circulation (AMOC) is a key feature of the Atlantic Ocean's circulation, as it provides a measure of the mean poleward flux of surface water and returning deep flow of colder water for the North Atlantic. Variation in the AMOC is believed to be strongly linked to variation in atmospheric climate, both now [*Buckley and Marshall*, 2016] and in the past [*Bigg et al.*, 2011]. However, despite the fundamental nature of this circulation parameter, continuous measurement of the AMOC only began with the deployment of the RAPID array of moored instruments along 26°N in spring 2004 [*Bryden et al.*, 2005]. This measurement program has continued ever since [*Frajka-Williams et al.*, 2016], with the most recent collection and deployment of instruments occurring in February 2017. The classical view of the AMOC was that, as it was a basin-wide measure, it could change only slowly over years. However, the first collection of the moored instruments of the RAPID array after just a year at sea showed that the AMOC was far more variable than had been suspected [*Cunningham et al.*, 2007], with the initial estimate of the AMOC having a large standard deviation: 18.7 ± 5.6 Sv. This strong, short-term, variability has continued ever since, with even a halving of the

AMOC strength being observed for a few months around early 2010 [*Smeed et al.*, 2014]. Since that abrupt drop, the strength of the AMOC has not recovered to its pre-2010 levels [*Frajka-Williams et al.*, 2016], with its recent mean being 2-3 Sv below the earlier level (Figure 1). Climate model simulations of the response of the AMOC to greenhouse warming suggest that it is very likely that the AMOC will decline in strength over the 21st Century [*Collins et al.*, 2013]. While models do not agree on the speed or size of this weakening, nor when it might be observed in the observations, the combination of this recent prolonged AMOC weakening with sustained global warming [*Kennedy et al*., 2016] raises the question of whether recent change falls within the range of natural variability.

There has been considerable interest in investigating past natural variability in the AMOC as simulated in climate or ocean models [*Danabasoglu et al*., 2015], and through ocean reanalysis reconstruction [*Tett et a*l., 2014; *Karspeck et al.*, 2015; *Jackson et al*., 2016]. While variability is a characteristic of the AMOC in all models, the climate models disagree on the change that has occurred at 26ºN over the last few decades, although most suggest an increase over the period from the mid-1970s to the mid-1990s [*Danabasoglu et al.*, 2015]. Ocean reanalyses might be expected to do better, as they assimilate a range of observations (although not the RAPID array) into reanalysis ocean model systems. They show AMOC strengths of the right order [*Tett et al.*, 2014], but many bear little resemblance to the RAPID observations [*Tett et al.*, 2014], or don't cover very much of the RAPID period [*Karspeck et al.*, 2015]. The latter do show an increase in the 1990s and early 2000s compared to the 1960s-80s, however. The one reanalysis product that does match the variation in the RAPID observations well is the GloSea5 analysis from the U. K. Met Office [*Jackson et al.*, 2016]. This only extends back to 1995, but suggests that the AMOC during the early 2000s was stronger than both earlier and latter in the time series, with late 1990s values being similar to those of recent years.
There are a small number of other observations of the AMOC near 26ºN, derived from geostrophic transport calculations for individual cruises stretching back to 1957 [*Bryden et al.*, 2005]. These have the disadvantage of being near-synoptic, and so may be unrepresentative of the mean flow at the time, as well as being from observations taken at different times of the year. The latter is particularly important as the annual cycle tends to have a late summer-autumn peak and a spring minimum (Figure 1), which may have led to bias in the calculations from individual cruise data (Table 1). Nevertheless, they provide another 5 observations to add to the RAPID set.

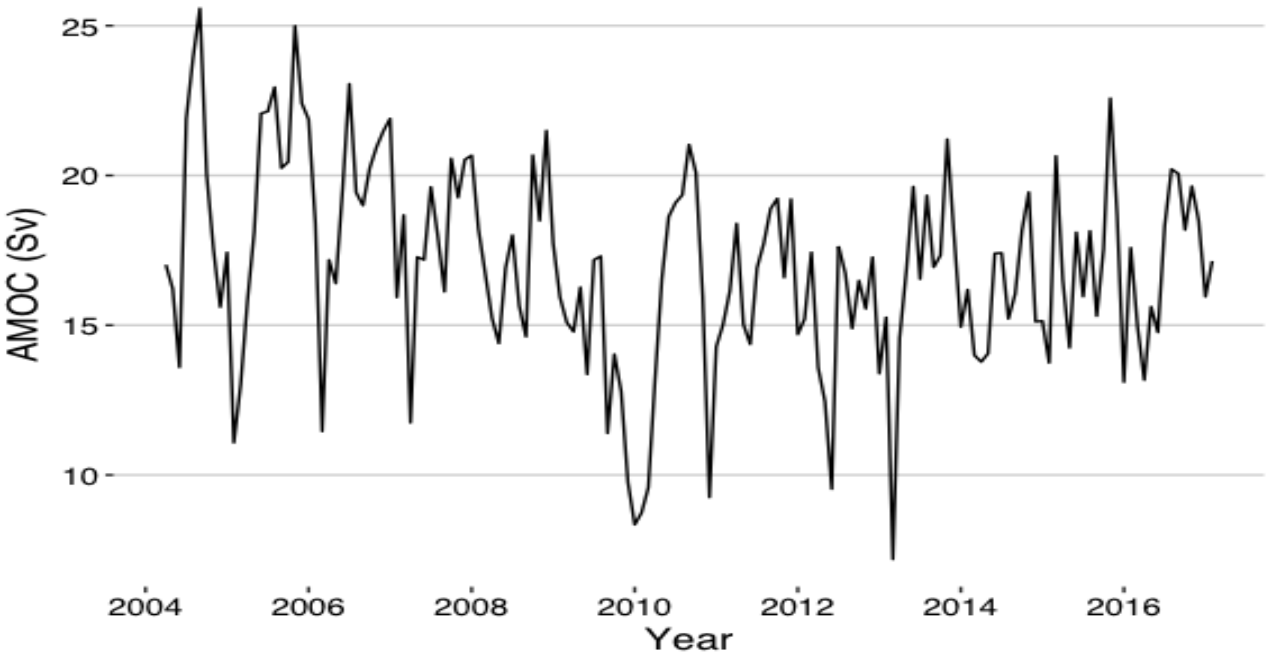

Figure 1. Monthly observed values of the AMOC over April 2004 – February 2017. The units are Sverdrups (Sv), or 10^6 m^3 s^{-1}.

Table 1. Net northwards transport calculated by *Bryden et al.* [2005] for cruise-based AMOC-proxy calculations near the RAPID line.

Cruise date	Oct. 1957	Sep. 1981	Aug. 1992	Feb. 1998	Apr. 2004
AMOC (Sv)	25.3	20.8	20.6	18.3	17.3

There is general agreement that density anomalies along the western boundary current [*Buckley and Marshall*, 2016], and particularly in the Labrador Sea [*Jackson et al.,* 2016], lead to variation in the AMOC, through southward propagation of boundary waves rather than water masses [*Hodson and* Sutton, 2012; *Jackson et al.*, 2016]. This paper attempts to explore and quantify this suspected relationship between decadal-scale variability in the AMOC at 26°N and the density anomalies of the northern Atlantic, but also takes into account the density contrasts between the tropical source waters of the Gulf Stream and the northern convection zones as well, through use of a control systems identification approach to modelling the RAPID AMOC timeseries. Such control systems model formulations have been recently shown to be useful in understanding the causes underlying variability in a range of environmental processes, including iceberg discharge from the Greenland Ice Sheet [*Bigg et al.*, 2014; *Zhao et al.*, 2016] and fish population response to environmental and fishing pressures [*Marshall et al.*, 2016]. Having found the most important environmental variables underlying determination of the strength of the AMOC, the resulting model is used in hindcast mode to reconstruct the AMOC back to 1980, through use of ocean parameters produced in the GODAS reanalysis system. As an output from the model analysis, the possibility of some predictability of the AMOC will be discussed. The main contribution of the paper include the finding of a dominant timescale of ~ 6-7 months between changes in meridional density difference and AMOC strength, which means that the AMOC is predictable. Through a hindcasting study, it reveals that the recent slowing of the AMOC is not outside the range of model variability since 1980; this is important and helpful for understanding the recent behavior of the AMOC.

2. Data and Methods

The conceptual approach of the control systems identification model to be described in section 2.2 is to: 1) select key input variables that are a priori regarded as critical in determining the output variable, in this case the AMOC strength at 26°N; and 2) build a model containing terms involving linear or non-linear lagged combinations of these inputs, to be selected sequentially according to the magnitude of their contribution to the output variable's variance. The resulting equation is thus analogous in some respects to a multiple regression model, but using more complex terms of an initially unknown number, constructed in a statistically rigorous fashion. This approach has proven very successful in a range of engineering environments since its first development by *Chen and Billings* [1989], and has recently begun to show its versatility within environmental contexts, as noted above.

The current analysis uses a three-stage process, partly owing to its origin in the RAPID Challenge of late 2015 (www.rapid.ac.uk/challenge/; Smeed [2017]), to predict the AMOC over April 2014-September 2015 prior to the retrieval of the mooring data from the RAPID array [*Cunningham et al.*, 2007] from its (then) latest deployment. Input variables, described in section 2.1, are first used to construct a range of test models to fit the then existing series of RAPID AMOC data from April 2004 to March 2013. Trial predictions of the AMOC output variable are then made for April 2013 to March 2014, using input variables available over this time span. These trial predictions are then compared to the AMOC strength up to March 2014. The best test model, in terms of its reproduction of the AMOC signal from a range of statistical measures, is then selected. Finally, the model, having been verified as robust over a decade, is tested on the retrieved data up to February 2017 and used to hindcast the AMOC strength back to 1980.

2.1 Data used in the control systems model

The monthly mean AMOC values from the RAPID array over April 2004 to February 2017, used as the output in the trial and test phase of the control systems modelling, were supplied by *Smeed et al.* [2017]. For a full description of how the AMOC values are calculated see Bryden *et al.* [2005] and Cunningham *et al.* [2007]. The input variables used for the control systems model are composed of two major types, an atmospheric one and three ocean density variables. The AMOC strength has an upper ocean component directly related to the wind strength, through Ekman transport. To represent this at basin-scale, the North Atlantic Oscillation (N) index is chosen, because of its strong links to the relative strength and locations of the central North Atlantic atmospheric pressure systems, as seen through the Rotated Principal Component Analysis of *Barnston and Livezey* [1987]. Monthly, standardized and normalized, values of the NAO index were taken from http://www.cpc.ncp.noaa.gov/products/precip/CWlink/pna/nao_index.html. These are shown in Figure 2. Several density variables are considered for model development tests described in section 3. These are all surface densities averaged over regions, but involved a mix of northern regions, where winter convection, and so deep water formation, occurs and a southern region in the Gulf of Mexico, from which the upper ocean waters feeding the main northward flow of the Gulf Stream derive. This allows experimentation with the relative importance for producing a model with a good fit to the observed AMOC of purely northern source waters compared to a measure of the density difference between the sub-tropical and sub-polar Atlantic. The GODAS ocean reanalysis has ocean reanalysis data at 0.33^o x 1^o resolution from January 1980 up to approximately one month before present, providing the necessary near real-time ocean data this research required. A description of its production is available at http://www.cpc.ncep.noaa.gov/products/GODAS/pl/introduction_godas_web.pdf. Computation of the three basic density variables used each involved downloading surface potential temperature and salinity data over the respective areas, and then their input into the density formula given by *Gill* [1982] at zero pressure. The three input surface density variables (all with units of kgm^{-3}) are GM, averaged over the region in the Gulf of Mexico 23-30°N, 82-90°W, LS, averaged over the Labrador Sea region 51-65°N, 42-65°W, and NS, averaged over the southern Norwegian Sea area 60-65°N, 5°E-12°W (Figure 3). Note that the NS variable only represents a small part of the southern Norwegian Sea, as the GODAS reanalysis product does not extend polewards of 65°N. However, previous research suggests that it is the Labrador Sea that is most important for AMOC variability [*Jackson et al.*, 2016] so it is expected that this limitation to the study will not be a major one. Figure 2 shows the time series for these surface density variables, as well as the NAO.

2.2 Description of model construction

System identification aims to identify a model that captures the relationship between a system input and output from recorded data [*Billings*, 2013]. Several approaches have been proposed for this purpose, and among them, the Nonlinear AutoRegressive with eXogenous inputs (NARX) methodology has proved to be well-suited for nonlinear system identification problems [*Billings*, 2013; *Wei et al., 2004a,b*]. NARX model has several attractive properties, one of which is the easy interpretability, for example, it can easily be identified which predictors and cross-product terms are important and make significant contribution to explaining the system output [Billings, 2013]. In general, a NARX model can be written as

$$y(k) = f\left(y(k-1), \dots, y\left(k-n_y\right), u(k-1), \dots, u(k-n_u)\right) + e(k) \quad (1)$$

where $f(\cdot)$ is a function to be estimated from available data, $u(k)$ and $y(k)$ are the system input and output signals respectively, $e(k)$ is noise (with $k = 1, \dots, N$ where N is the total number of data points used for model estimation), and n_u and n_y are the maximum lags for the input and output

signals [*Wei and Billings*, 2008]. One of the most popular choices for $f(\cdot)$ is a polynomial representation. Therefore, equation (1) can be written as

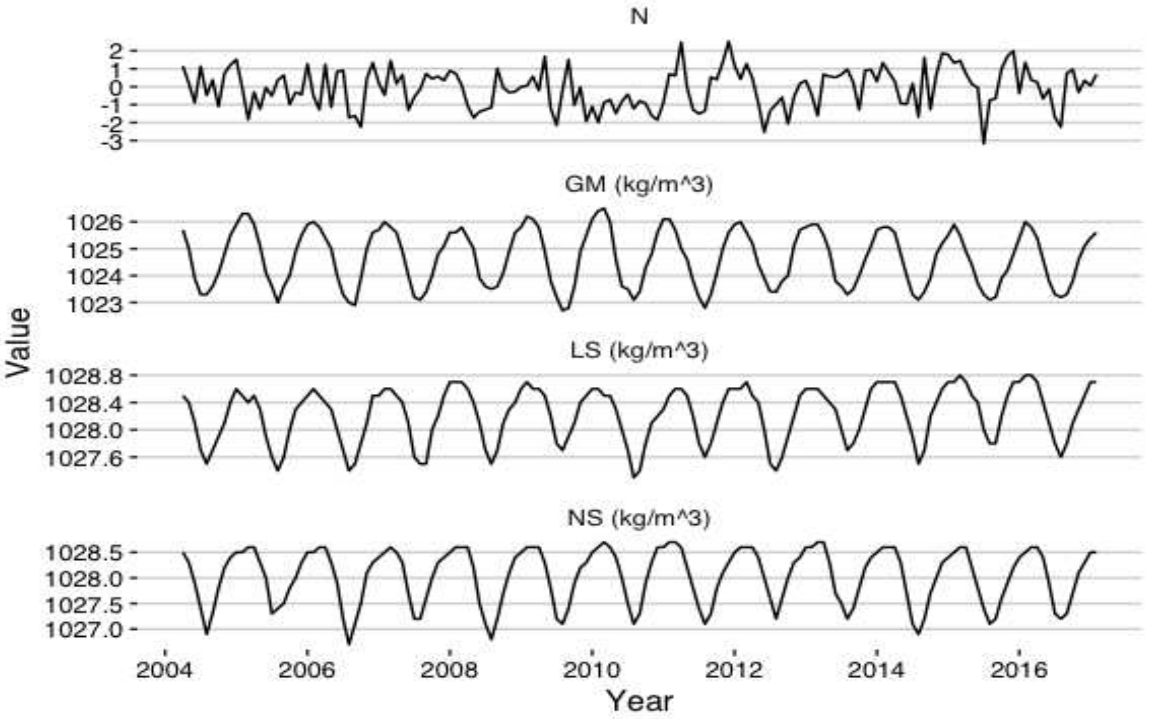

Figure 2. Monthly observed values of N (standardized NAO index), density variables GM (Gulf of Mexico), LS (Labrador Sea) and NS (Norwegian Sea) over April 2004 – February 2017. Units for density variables are kg m^{-3}, while the units for N are with respect to the standard deviation of the NAO pressure difference index.

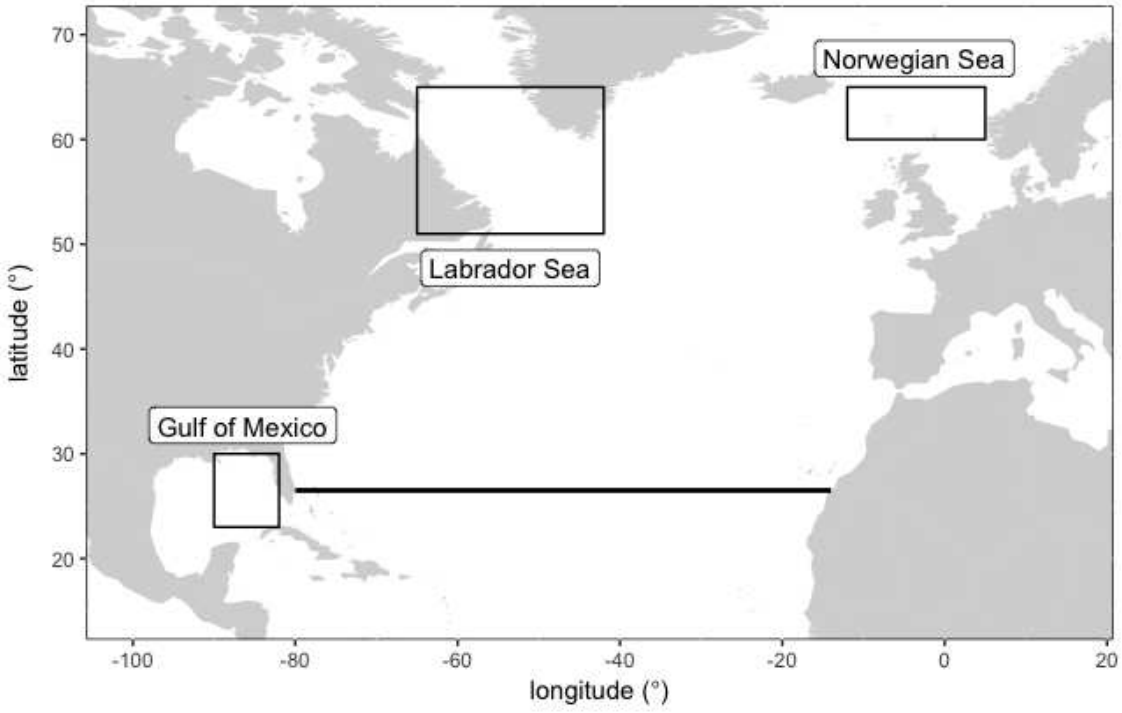

Figure 3. Map showing the position of the RAPID line (in black), and the regions used to calculate regional surface density variables (in squares). See text for precise variable definitions.

$$y(k) = \theta_0 + \sum_{i_1=1}^{n} \theta_{i_1} x_{i_1}(k) + \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \theta_{i_1 i_2} x_{i_1}(k) x_{i_2}(k) + \cdots + \sum_{i_1=1}^{n} \cdots \sum_{i_\ell=i_{\ell-1}}^{n} \theta_{i_1 i_2 \ldots i_\ell} x_{i_1}(k) x_{i_2}(k) \ldots x_{i_\ell}(k) + e(k) \quad (2)$$

where

$$x_m(k) = \begin{cases} y(k-m) & 1 \le m \le n_y \\ u(k-m+n_y) & n_y + 1 \le m \le n = n_y + n_u \end{cases} \quad (3)$$

The θ's are the model parameters, and ℓ is the nonlinear degree of the polynomial model. A NARX model of degree ℓ implies that the degree of each term in equation (2) is not higher than ℓ. For example, $x_1 x_2$ is a term of nonlinear degree 2 while $x_1^2 x_2$ is a nonlinear term of degree 3.

The most popular algorithm for building NARX models is the Orthogonal Forward Regression (OFR) algorithm [*Billings*, 2013; *Guo et al.*, 2015]. This is a stepwise algorithm that identifies the most significant predictors and regressors that explain the output variable's variance using an Error Reduction Ratio (ERR) index [*Chen et al.*, 1989]. A comprehensive explanation of the meaning of ERR may be found in *Chen et al.* [1989], *Wei et al.* [2004] or *Billings* [2013]. In recent years, several improvements to the original OFR algorithm have been developed. One such improvement is the use of new metrics, such as mutual information (MI) [*Billings and Wei*, 2007] and distance correlation [*Ayala Solares and Wei*, 2015], given that the ERR index, defined as a squared correlation function [*Wei and Billings*, 2008], is only able to capture linear dependencies.

Conventionally, classical correlation or squared correlation (also called energy correlation) is used to measure the correlation relationship between two signals, but recently it has been shown that distance correlation works better [*Ayala Solares and Wei*, 2015; *Ayala Solares*, 2017] for characterising both linear and nonlinear dependency. The distance correlation is thus used in lagged variable dual cross-product term selection in the NARX modelling procedure. Another concern is related to the stop criterion when building a model. Originally, when the sum of the ERR values of the predictors selected is above a given threshold, the model training process is stopped. This of course requires a careful selection of the threshold. If it is too small, the identified model cannot capture the dynamics of the system completely. However, if the threshold is too large, it can lead to an overfitted model that does not generalize well to new observations. In this work, a different model selection approach is used to select the most appropriate number of model terms. This makes use of performance metrics where the data is separated into training and validation sets. As the names suggest, the former is used to train the model, and the latter is used to evaluate the performance of the model using an evaluation metric (EM) [*James et al.*, 2013]. Furthermore, as it is desired that the modelling fitting performance over the training set and the prediction performance over the validation set are similar they are weighted equally. This allows the computation of a weighted evaluation metric (WEM), i.e.

$$\text{WEM} = 0.5 * \text{EM}(\text{training set}) + 0.5 * \text{EM}(\text{validation set}) \quad (4)$$

In (5), taking the commonly used root mean squared errors (RMSE) as an example and assuming the length of the training and validation datasets are *m* and *n*, respectively, then,

$$\text{EM(training set)} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} e_{i,\text{training}}^2} \quad (5a)$$

$$\text{EM(validation set)} = \sqrt{\frac{1}{n}\sum_{j=1}^{n} e_{j,validation}^2} \quad (5b)$$

where $e_{i,\text{training}}$ and $e_{j,\text{validation}}$ are the errors on the ith and jth data point in the training and test data sets, respectively. The WEM approach helps to build a model that captures efficiently the system dynamics without under- or overfitting the data.

Performance metrics have advantages and disadvantages. In general, performance metrics provide a better estimate of the test error, and make fewer assumptions about the true underlying model [*James et al.*, 2013]. However, by splitting the data into training and testing sets, the sample size is reduced for both model training and testing. It is also computationally expensive since the process may need to be repeated several times to achieve good estimates of accuracy.

2.3 Model Evaluation

The data set is divided in three parts. The first part contains data from April 2004 to March 2013, which is used for training several models using the ERR and MI indices, together with performance metrics. The second part uses data from April 2013 to March 2014 for model validation/comparison and model evaluation. The last part contains data from April 2014 to February 2017, which is used to test models' predictive performance on data that were not used in the model identification and selection phase.

If we define e_i as the error between the ith prediction $\hat{y}_i$ and the ith output value y_i, i.e. $e_i = \hat{y}_i - y_i$, then the three evaluation metrics considered are:

- Mean Error: $ME = \frac{1}{N}\sum_{n=1}^{N} e_i$ (6a)
- Mean Absolute Error: $MAE = \frac{1}{N}\sum_{n=1}^{N} |e_i|$ (6b)
- Root Mean Squared Error: $RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (e_i)^2}$ (6c)

The above metrics are widely used in traditional modelling practices. The ME is used to check if the mean of the model error is close to zero. The $RMSE$ is usually used to measure the overall performance of the model, while the MAE can be used to measure the model predictive power to detect extreme or peak values of the system response.

3 Results

3.1 The models

Based on the data description given in Section 2.1, three sets of variables are used to build three different NARX models of the AMOC. As described in section 2.3, data from April 2004 to March 2014 are used for training and validation, while that from April 2014 to February 2017 are used for testing. The maximum order of the polynomials used in model construction (equation 2) was 2, in accord with the findings of section 3.2. In all cases, the ERR and MI are used together with the performance metrics of (equation 6) to determine the most appropriate models.

The three cases involve selections of varying groups of input variables as summarised below, where the variables are defined in section 2.1. All cases assume that both atmospheric and oceanic

quantities contribute towards the observed AMOC variation, in line with the mix of Ekman transport and density-driven elements contributing to the flow. In all cases the atmospheric component is represented by the large-scale atmospheric circulation measure of the NAO. However, we examine three different ways in which density may contribute to the AMOC: Case 1 uses the mean surface density over the origin and sinking regions for the North Atlantic Drift; Case 2 considers the density gradient between the sinking and origin regions; and Case 3 allows both of these density measures to play a role in the model.

Note that the main objectives of the study are twofold: a) to investigate which input variables are most important, and how the change of AMOC depends on the interactions of these important input variables; and b) to investigate the predictive power of these important variables for forecasting the AMOC. We therefore do not consider autoregressive model terms, that is, a lagged AMOC is not included in the models.

- Case 1 – forced by the relative contributions of the atmosphere and ocean mean states:
  - AMOC strength (output variable)
  - NAO index - N (input variable)
  - Mean of the density variables (input variable) defined as

$$U = \frac{GM + LS + NS}{3} \tag{7}$$

- Case 2 – forced by the relative contributions of the atmosphere and the meridional density difference between surface and deep water source waters:
  - AMOC strength (output variable)
  - NAO index - N (input variable)

Difference of the density variables (input variable) defined as

$$V = \frac{LS + NS}{2} - GM \tag{8}$$

- Case 3 – forced by the relative contributions of the atmosphere and the contrasting mean and meridional differences in ocean density:
  - AMOC strength (output variable)
  - NAO index - N (input variable)
  - Mean of the density variables - U (input variable)
  - Difference of the density variables - V (input variable)

The time series for the new variables U and V are shown in Figure 4.

It is noteworthy that the variables U and V, while retaining the annual cycle visible in Figure 2, have opposite extremes, namely, the highest value of the mean density, U, is during the winter, while the largest difference in density, V, occurs during the summer. It is also important to mention that for each of the three variables, AMOC, U, and V, the corresponding mean value is removed prior to the model building procedure. The mean values of the three variables, estimated based on the training data (i.e. data from April 2004 to March 2013) are 16.97 Sv for AMOC, 1026.98 kg m^{-3} for U, and 3.4 kg m^{-3} for V, respectively. This is done partly because the magnitudes of the density variables are much larger than the N index and the AMOC strength. Removing the mean value ensures that the

density variables do not dominate the training and validation phases, and that the resulting models are more robust.

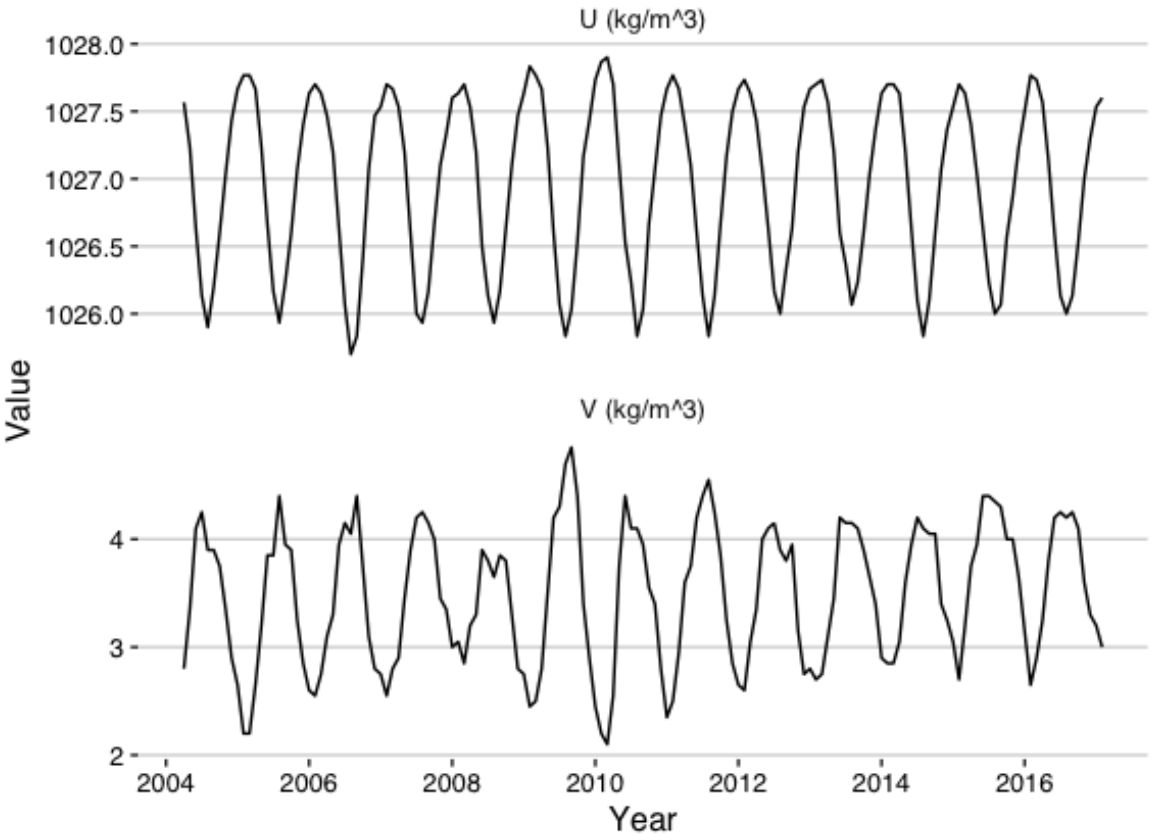

Figure 4. Monthly observed values of U (mean of surface density variables) and V (difference of surface density variables) over April 2004 – February 2017. Units are kg m^{-3}.

For all three cases, the model with the best performance is that selected by means of the ERR metric, as the models obtained with the MI metric had a poorer performance. The model terms selected are shown in Table 2, where the terms show the lag element through $(k-m)$, where m is the number of months by which the variable is lagged relative to the current month, k.

Note that the variables in the models reported in Table 2 are standardized, so the model for Case 3 should be read as below and models for Cases 1 and 2 should be used in the same manner:

$$y(k) = -2.5[V(k-7) - 3.41] + 1.207N(k) - 1.24N(k)[U(k-6) - 1026.98] \quad (9)$$

Accordingly, the model predicted AMOC strength is:

$$AMOC(k) = y(k) + 16.97 \quad (10)$$

Each trained model is evaluated using the training and validation data sets up to March 2014. The weighted performance metrics for the three models are shown in the bottom panel of Table 2. From these, it is argued that the Case 2 model performs best overall as the lowest value for each of the metrics occurs for this Case. This suggests that the difference in density between the deep-water formation areas and the upstream Gulf Stream source region seven months ago provides the best

indication of variation in the AMOC strength, this being the leading term for Case 2. Furthermore, an important observation is that all three cases agree that the current NAO index plays a discernible role in the AMOC strength, as all models have a second term linearly dependent on N(t), with a lagged element of N being involved in all higher order terms (Table 2).

The best model (i.e. the Case 2 model) is applied to the test data of April 2014 to February 2017, and its performance is shown in Table 3. Figure 5 shows a comparison between the model simulation output (also known as model predicted output) and the actual measurements. These show that the model captures the main dynamics of the AMOC process, although it is worth noting that the reduced annual cycle component of the AMOC in 2014/15 worsened the metric scores for the test period (Table 3) compared to the training period. Many of the predictions in the RAPID Challenge also experienced difficulty in predicting this unusual feature [Smeed, 2017].

It is noteworthy that over the whole period of the RAPID dataset, the correlation between the model simulation output (from the Case 2 model) and the observations is 0.62, statistically significant well beyond the 1% level. It is also worth noting that Figure 5 shows that the model successfully captures the transition from a semi-regular annual cycle prior to 2012 to the more chaotic variability since then. Nevertheless, as well as the poor performance of the model in 2014/15, high peak levels tend to be under-estimated throughout (Figure 5). It is not clear what has caused this, however, only large-scale measures of atmospheric and oceanic conditions have been used as inputs to the model so any more locally related variability will not be captured by the model.

Table 2. TOP: Model terms selected when modeling the AMOC strength using the NAO index and means of density variables for each of the three Model Cases. BOTTOM: Weighted performance metrics on the training and validation data sets for each of the three Model Cases.

Case 1			**Case 2**			**Case 3**		
Variable	Parameter	ERR [%]	Variable	Parameter	ERR [%]	Variable	Parameter	ERR [%]
$\boldsymbol{U(t-7)}$	2.221	17.95	$V(t-7)$	-2.449	20.60	$V(t-7)$	-2.500	20.60
$\boldsymbol{N(t)}$	1.307	13.88	$N(t)$	1.316	14.46	$N(t)$	1.207	14.46
$\boldsymbol{N(t)U(t-6)}$	-1.363	6.63	$N(t)V(t-6)$	1.237	5.27	$N(t)U(t-6)$	-1.240	5.90
$\boldsymbol{N(t-8)U(t-3)}$	1.096	4.42	$N(t-8)V(t-3)$	-1.065	5.10			
			$N(t-3)V(t-3)$	1.018	4.61			
				Performance Metrics				
ME **(Sv)**	-0.3188		*ME* (Sv)	**-0.2123**		*ME* (Sv)	-0.3479	
MAE **(Sv)**	1.8603		*MAE* (Sv)	**1.6908**		*MAE* (Sv)	1.8940	
RMSE **(Sv)**	2.3282		*RMSE* (Sv)	**2.0761**		*RMSE* (Sv)	2.2852	

Table 3. Performance metrics on the training dataset (April 2004 – March 2013), validation dataset (Aril 2013 - March 2014), test dataset (April 2014 – February 2017), and validation + test dataset (April 2013 – February 2017), using the best model found (model from Case 2).

Performance Metrics				
	Training	Validation	Test	Validation+Test
ME	-0.2123 Sv	0.7304 Sv	-0.3626 Sv	-0.0836 Sc
MAE	1.6908 Sv	1.3569 Sv	2.1810 Sv	1.9706 Sv
RMSE	2.0761 Sv	1.6092 Sv	2.6476 Sv	2.4251 Sv

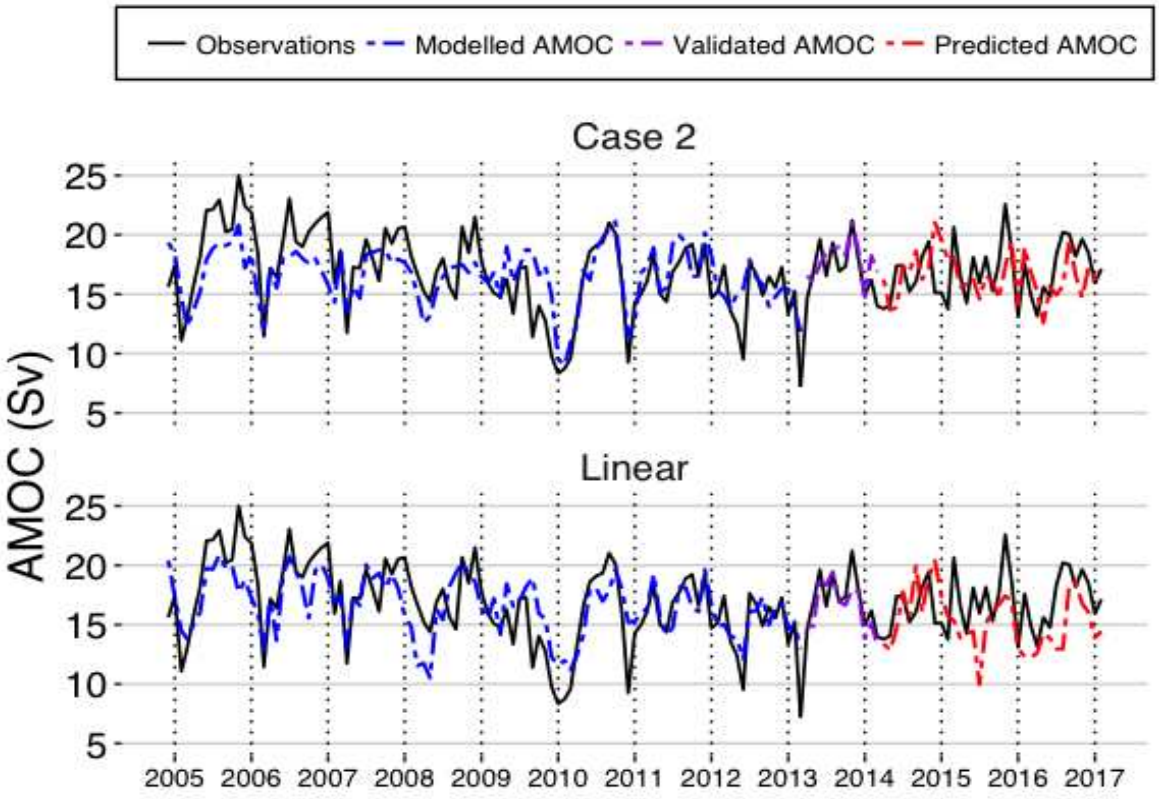

Figure 5. TOP: Modelled and predicted AMOC anomaly obtained using the best model found (model from Case 2). BOTTOM: Modelled and predicted AMOC anomaly obtained using the linear model (10).

3.2 Non-linear versus linear models

It is interesting to notice that the two leading terms of all 3 Case models shown in Table 2 are linear. To test whether use of a non-linear model was a statistically significant improvement over use of a purely linear model, a NARX model of maximum degree 1 of the AMOC was developed for the training period. This is given by:

$$\begin{aligned} y(k) = &-3.133[V(k-7) - 3.4] + 1.268N(k) \\ &- 1.960[V(k-1) - 3.4] \\ &- 0.104[U(k-7) - 1026.98] \\ &- 4.145[V(k-3) - 3.4] \\ &- 5.322[U(k-4) - 1026.98] \\ &- 4.571[U(k) - 1026.98] \end{aligned} \tag{11}$$

with the AMOC strength prediction computed with (10).

The above linear model is applied to predict the AMOC strength. The weighted performance metrics of the linear model on the training and validation data sets are: $ME = 0.5393$ Sv, $MAE = 1.9481$ Sv, and $RMSE = 2.5678$ Sv, all of which are significantly larger than the metrics for the Case 2 model in Table 2 and so clearly suggesting that the purely linear model is inferior to the Case 2 model. This is consistent with the poorer fit of the linear model shown in the bottom panel of Figure 5.

The above statement can be confirmed by means of the Ramsey Regression Equation Specification Error Test (RESET) [*Ramsey*, 1969]. This test was designed to examine the null hypothesis that a linear model is enough to explain the output signal, whereas the alternative hypothesis suggests that a nonlinear model would perform better. The results from this test are shown in Table 4. These suggest that there is enough evidence to use a nonlinear model whose nonlinearity degree is 2 (i.e. the polynomial power is 2), to represent the preprocessed data, while not enough evidence is available to choose a model of power 3 (i.e. nonlinearity degree of 3) at the 5% significance level. This result drove our choice of the maximum polynomial in the model constructions.

Table 4. P-values obtained from the Ramsey Regression Equation Specification Error Test (RESET) to determine the appropriate order of the model.

Polynomial Order	P-value
2	**4.591e-05**
3	0.9573

3.3 Hindcasting

The Case 2 model is used to hindcast the AMOC strength back to January 1980. The hindcast and predicted AMOC values are shown in Figure 6. It is clear that the mean of the recovered AMOC from the model has changed little since 1980. The mean before the establishment of the RAPID array is 16.8±1.9 Sv, while since April 2004 the modeled AMOC has become 16.6±2.2 Sv, showing no statistical difference in the mean or variance. The tendency for an irregular annual cycle, with a winter minimum, a spring maximum, and a typical annual range of 2-3 Sv, also extends throughout the dataset (Figure 7), although this has occasionally broken down in the past (e.g. around 1989) as during the RAPID program (e.g. around 2009).

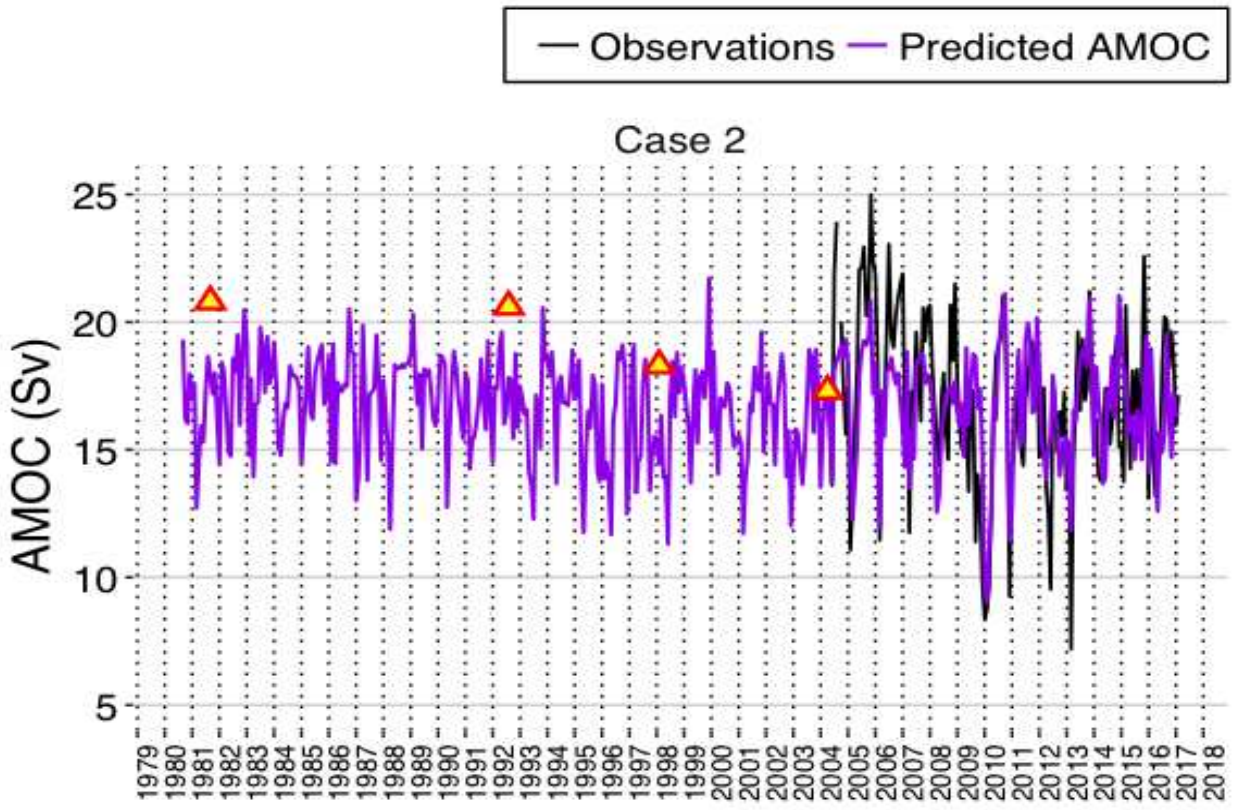

Figure 6. Hindcast and predicted AMOC obtained using the Case 2 model. The section calculations from Table 1 that fall within the interval studied are shown by triangles.

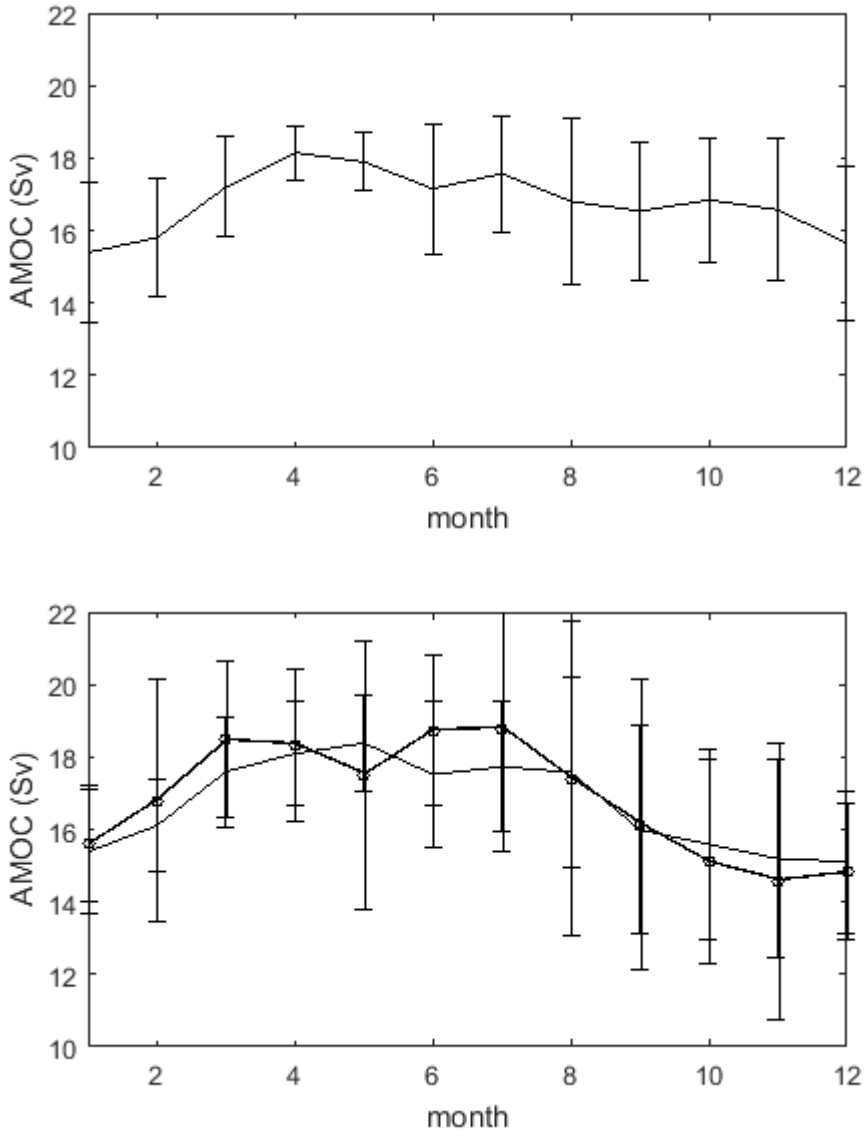

Figure 7. Model AMOC annual cycle, with standard deviation errors: upper panel) September 1980-August 2004; lower) September 2004-August 2016. In the lower panel the observed AMOC annual cycle over the same period is shown in bold.

4. Discussion

The NARX model of the AMOC strength at 26^{o}N has been shown in Figure 5, Figure 7 and Table 2 to match reasonably well the RAPID training data, while producing the right magnitude of the recent, test, dataset and its irregular nature compared to a "normal" annual cycle. This gives confidence in the broad structure of the hindcast back to 1980. Note also that the more recent cruise calculations from Table 1 agree reasonably well with the model estimates of the AMOC (Figure 6). Nevertheless, details of the variation in the AMOC are not always well captured. The extrema during the training and test period are often under-estimated, although there are periods when these are captured well. This under-estimation of the extrema seems particularly true of the maxima, while the occurrence of major negative excursions is found within the model. In this context it is notable that the extended

reduction in observed AMOC strength around the beginning of 2010 is well predicted by the model (Figure 5). This is linked to an extreme variation in the mean density difference, V, between a peak maximum in 2009 and a peak minimum in 2010, associated with the prolonged negative excursion in NAO around this period (Figure 2), which led to the coldest winter in the UK since 1978/9 [*Prior and Kendon*, 2011]. On the other hand, all the high AMOC peaks in the mid-2000s are under-estimated by the model. In some measures an averaging of the three case models gives a good performance (see Appendix), but this inability to reproduce high extremes remains (Figure A1).

The models found that a dominant lag time in many terms tends to be around 6-8 months, particularly in V, the density difference between the convection regions and the Gulf Stream source. This timescale agrees well with those found in previous studies that suggest AMOC variation is linked to the transit time for boundary waves generated by density fluctuations in the Labrador Sea and then travelling south along the American shelf [*Hodson and Sutton*, 2012; *Jackson et al.*, 2016]. The AMOC's variation is thus driven by wave signals and not direct change in water mass properties, which would have a much longer timescale if important. However, some modulation of this signal is found in more short term signals from the atmosphere, through the NAO, where Table 2 shows a strong, but secondary, signal of instantaneous linear terms in N. These terms arise from direct responses of the upper ocean due to the wind-induced Ekman transport. Our analysis also showed that while the leading terms of each model are linear, the best model has distinct nonlinear components, involving a modulation of the wind and density difference variables. This nonlinearity was important in providing the best reproduction of the observed AMOC variation, and its inclusion was statistically robust. This necessity for including nonlinearity to provide the best model is consistent with the nonlinear nature of many density-driven wave processes [*Gill*, 1982].

Looking at the longer model reconstruction, back to 1980, an element of decadal-scale change is visible (Figure 6). While there is essentially no trend over the whole record (-0.02 Sv/yr), the 1980s tended to have a higher modeled AMOC (17.2±1.7 Sv) than the late 1990s (16.2±2.1 Sv over 1995-1999). Furthermore, it is also notable that the hindcasted AMOC varies in a range approximately between 13 and 20 Sv. Rapid and significant change in the strength of the AMOC within this range is a characteristic of the longer term pattern, and recent changes since 2010 are not unprecedented.
Case 2's nature (Table 2) suggests the possibility of some predictive ability for the AMOC, because the dominant term contains a time lag of 7 months, through the density difference driving the variability. For comparison purposes, a simpler model was examined using just the first model term from Case 2. This is shown in Figure 8. Much less of the variability of the AMOC signal is captured when just using this term, and in particular the nonlinear terms in the case 2 model are required to reproduce the extreme minimum in early 2010 (cf. Figures 5 and 8). Nevertheless, the correlation with the RAPID series during 2004-15 is still a statistically significant 0.44 largely because the model annual cycle is controlled by the strong annual cycle in the salinity component of V. While such a simple model is clearly not of significant predictive usefulness in itself, it suggests that the AMOC may be predictable up to 7 months in advance.

5. Conclusions

Using the control system identification model, NARX, it has been shown that the variation in the AMOC during the RAPID observational program is consistent with variability over the preceding 25 years. This includes the ability to experience periods of distinctly reduced flows, and decadal-scale variation in the long term flow strength. Thus, recent slower flows (AMOC observed mean over 2009-2013 is 15.6 Sv) are not dissimilar to model hindcasts for the late 1990s (AMOC modeled mean over 1995-1999 is 16.2 Sv), given that the model is not normally able to capture short term negative excursions. In addition, the difference in transect measurements of the AMOC from before the RAPID era shown in Table 1 agree well with the model predictions since the mid-1990s. Those from 1981 and 1992 are ~2 Sv above the model estimates for cruise months (18.7 and 17.9 Sv respectively), which was within the error estimate for these observations [*Bryden et al.,* 2005].

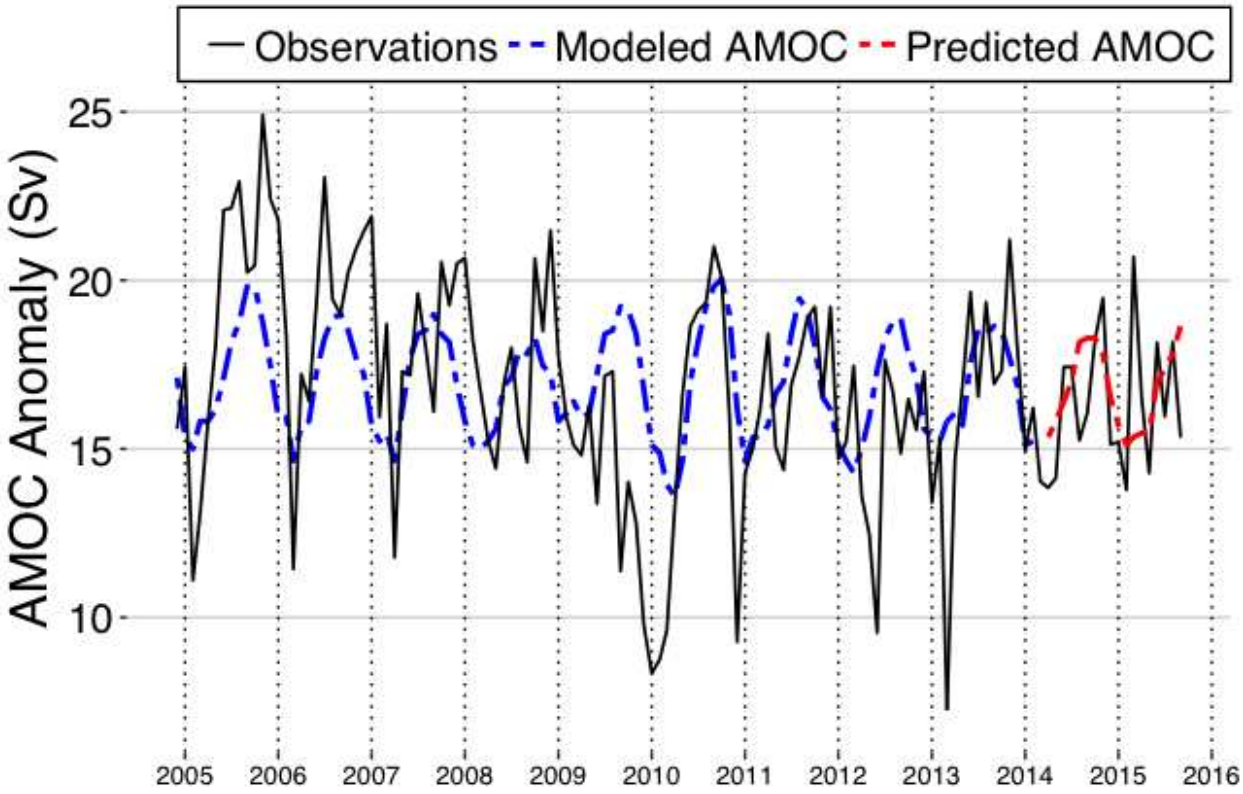

Figure 8. Modelled and predicted AMOC anomaly obtained using just the first model term from Case 2, i.e. $V(k-7)$.

It has also been shown that the variation of the AMOC is linked strongly to the variation in the density difference between the northern sinking waters and the Gulf of Mexico source waters of the main overturning current, with a time lag of ~ 7 months, commensurate with the physical driving force being boundary density waves. This offers the future opportunity for some predictive power of the strength of the AMOC in the subtropical North Atlantic.

Acknowledgments, Samples, and Data

We thank the UK RAPID programme for providing the AMOC data at http://www.rapid.ac.uk/rapidmoc/. GODAS data was provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/.

The authors have no financial conflicts of interest in carrying out this research.

Appendix – Model Averaging

One of the model performance metrics, RMSE (root mean squared error), given in Table 2 is used to implement a model averaging scheme. Note that the RMSE values for the three models are MSE1 = 2.2382, MSE2 = 2.0761, and MSE3 = 2.2852, respectively. Using these values, we define the following parameters:

$$c_1 = \frac{1}{\text{MSE}_1}, \quad c_2 = \frac{1}{\text{MSE}_2}, \quad c_3 = \frac{1}{\text{MSE}_3} \tag{A1}$$

$$w_1 = \frac{c_1}{c_1 + c_2 + c_3} = 0.3276, \quad w_2 = \frac{c_2}{c_1 + c_2 + c_3} = 0.3563, \quad w_3 = \frac{c_3}{c_1 + c_2 + c_3} = 0.3201, \tag{A2}$$

Let $\hat{y}_1$, $\hat{y}_2$, and $\hat{y}_3$ be the predicted output values by the three models of the three cases in Table 2, the model averaging prediction is defined as:

$$\hat{y} = w_1\hat{y}_1 + w_2\hat{y}_2 + w_3\hat{y}_3 \tag{A3}$$

The three model performance metrics, ME, MAE, and RMSE produced by the model averaging prediction are listed in Table A1.

Table A1 A comparison of the prediction performance of the best linear model, the model for Case 2 (in Table 2), and the model averaging of the three models in Table 2. Performance metrics are calculated over the test dataset of the period from April 2014 to February 2017

Model Case	Performance Metrics		
	ME	MAE	MSE
Best linear model	1.6318	2.4935	3.1891
Best model for Case 2	-0.3626	2.1810	2.6476
Model averaging	0.1077	1.8448	2.3355

The graphical illustration of the model averaging prediction performance is shown in Fig A1.

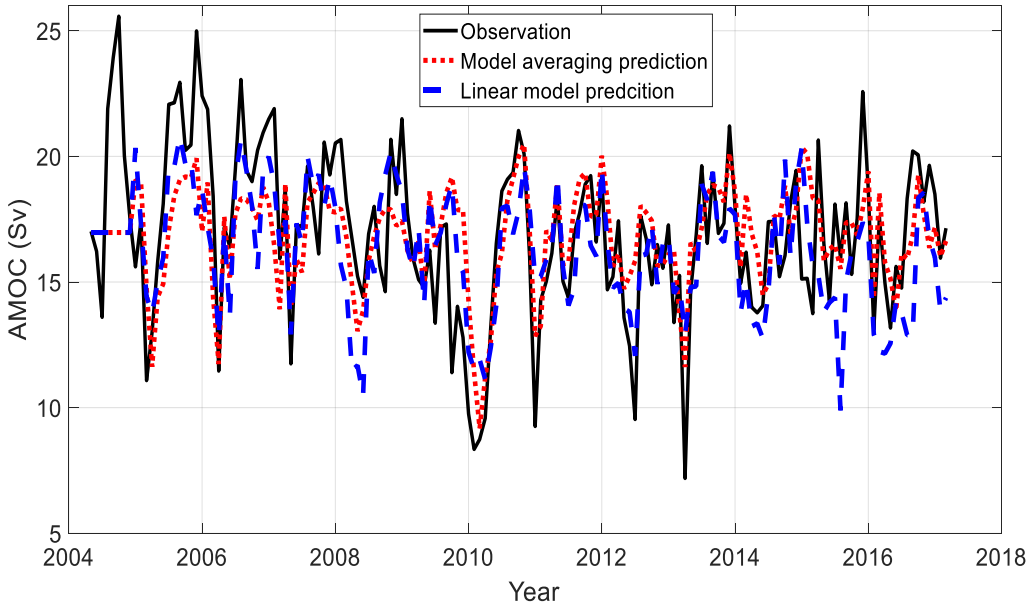

Figure A1. An illustration of the model averaging prediction performance. Training data: April 2004 – March 2014; Test data: April 2014 – February 2017.

References

Ayala Solares, J. R., and H.-L. Wei (2015), Nonlinear model structure detection and parameter estimation using a novel bagging method based on distance correlation metric, Nonlinear Dynamics 82, 201–215.

Ayala Solares, J. R. (2017), Machine Learning and Data Mining for Environmental Systems Modelling and Analysis, PhD Thesis, University of Sheffield.

Barnston, A. G., and R. E. Livezey (1987), Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, Mon. Wea. Rev., 115, 1083-1126.

Bigg, G. R., R. C. Levine and C. J. Green (2011), Modelling abrupt glacial North Atlantic freshening: rates of change and their implications for Heinrich events, Glob. Planet. Change, 79, 176-192.

Bigg, G. R., H. L. Wei, D. J. Wilton, Y. Zhao, S. A. Billings, E. Hanna, V. Kadirkamanathan (2014), A century of variation in the dependence of Greenland iceberg calving on ice sheet surface mass balance and regional climate change, Proc. Roy. Soc Ser. A, 470, 20130662.

Billings, S. A. (2013), Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains, Wiley.

Billings, S. A., and H.-L. Wei (2007), Sparse Model Identification Using a Forward Orthogonal Regression Algorithm Aided by Mutual Information, IEEE Transactions on Neural Networks, 18 (1), 306–310.

Bryden, H. L., H. R. Longworth and S. A. Cunningham (2005), Slowing of the Atlantic meridional overturning circulation at 25°N, Nature, 438, 655-657.

Buckley, M.W., and J. Marshall (2016), Observations, inferences, and mechanisms of the Atlantic Meridional Overturning Circulation: A review, Rev. Geophys., 54, 5-63.

Chen, S., and S. A. Billings (1989), Representations of non-linear systems: the NARMAX model, Int. J. Control, 49, 1013-1032.

Chen, S., S. A. Billings, and W. Luo (1989), Orthogonal least squares methods and their application to non-linear system identification. Int. J. Control, 50, 1873–1896.

Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, et al. (2013), Long-term Climate Change: Projections, Commitments and Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex & P.M. Midgley (eds.)], 1029-1136. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Cunningham, S. A., T. Kanzow, D. Rayner, M. O. Baringer, W. E. Johns, J. Marotzke et al. (2007), Temporal variability of the Atlantic Meridional Overturning Circulation at 26.5°N, Science, 317, 935-938.

Danabasoglu, G., S. G. Yeagera, W. M. Kimb, E. Behrensc, M. Bentsend, D. Bie, et al. (2015), North Atlantic simulations in coordinated ocean-ice reference experiments phase II (CORE-II). Part II: inter-annual to decadal variability, Ocean Model., 97, 65-90.

Frajka-Williams, E., C. S. Meinen, W. E. Johns, D. A. Smeed, A. Duchez, A. J. Lawrence, D. A. Cuthbertson, G. D. McCarthy, H. L. Bryden, B. I. Moat, and D. Rayner (2016), Compensation between meridional flow components of the Atlantic MOC at 26°N, Ocean Sci., 12, 481-493.

Gill, A. E. (1982), Atmosphere-ocean dynamics, Academic Press, London, 662 pp.

Guo, Y., L. Guo, S. Billings, and H.-L. Wei (2015a), An iterative orthogonal forward regression algorithm, International Journal of Systems Science , 46 (5), 776–789.

Hyndman, R. J., and G. Athanasopoulos (2014), Forecasting: principles and practice. OTexts.

Hodson, D., and R. Sutton (2012). The impact of resolution on the adjustment and decadal variability of the Atlantic meridional overturning circulation in a coupled climate model. Clim. Dyn., 39, 3057-3073.

Jackson, L. C., K. A. Peterson, C. D. Roberts and R. A. Wood (2016), Recent slowing of Atlantic overturning circulation as a recovery from earlier strengthening, Nature Geosci., 9, 518-522.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). An introduction to statistical learning (Vol. 6). New York: Springer.
Karspeck, A. R., et al. (2015), Comparison of the Atlantic meridional overturning circulation between 1960 and 2007 in six ocean reanalysis products, Clim. Dyn., doi:10.1007/s00382-015-2787-7.
Kennedy, J., C. Morice, D. Parker and M. Kendon (2016). Global and regional climate in 2015, Weather, 71, 185-192.
Marshall, A., G. R. Bigg, S. M. van Leeuwen, J. K. Pinnegar, H.-L. Wei, T. J. Webb, and J. L. Blanchard (2016), Quantifying heterogeneous responses of fish community size structure using novel combined statistical methods, Glob. Change Biol., 22, 1755-1768.
Prior, J., and M. Kendon (2011) The UK winter of 2009/2010 compared with severe winters of the last 100 years, Weather, 66, 4-10.
Ramsey, J. B. (1969), Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis, Royal Statistical Society, 31 (2), 350-371.
Smeed, D. A. (2017), The RAPID Challenge: observational oceanographers challenge their modelling colleagues, Ocean Challenge, 22, 16-18.
Smeed, D. A., D. A. Smeed, G. McCarthy, S. A. Cunningham, E. Frajka-Williams, D. Rayner, W. E. Johns, C. S. Meinen, M. O. Baringer, B. I. Moat, A. Duchez, and H. L. Bryden (2014), Observed decline of the Atlantic meridional overturning circulation 2004–2012, Ocean Sci., 10, 29-38.
Smeed, D., G. McCarthy, G., D. Rayner, B.I. Moat, W.E. Johns, M.O. Baringer, and C.S. Meinen (2017), Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2017. British oceanographic Data Centre – Natural Environment Research Council, UK. doi: 10.5285/5acfd143-1104-7b58-e053-6c86abc0d94bTett, Simon F. B., et al. (2014), How much has the North Atlantic Ocean overturning circulation changed in the last 50 years?, Journal of Climate, 27 (16), 6325-6342.
Wei, H.-L., S. A. Billings, and M. A. Balikhin (2004a), Prediction of the Dst index using multiresolution wavelet models, Journal of Geophysical Research: Space Physics , 109 (A7), A07212.
Wei, H.-L., S. A. Billings, and J. Liu (2004b), Term and variable selection for non-linear system identification, International Journal of Control, 77 (1), 86–110.
Wei, H.-L., and S. A. Billings (2008), Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information, International Journal of Modelling, Identification and Control , 3 (4), 341–356.
Zhao, Y., G. R. Bigg, S. A. Billings, E. Hanna, A. J. Sole, H.-L. Wei, V. Kadirkamanathan, and D. J. Wilton (2016), Inferring the variation of climatic and glaciological contributions to West Greenland iceberg discharge in the Twentieth Century, Cold Reg. Sci. Technol., 121, 167-178.