Manuscript prepared for Earth Syst. Dynam.
with version 5.0 of the LaTeX class copernicus.cls.
Date: 3 July 2015

Early warning indicators and their relation to Liouvillian dynamics.

M.H. Hebbink[1]

[1]Institute for Marine and Atmospheric research Utrecht, Utrecht University, Utrecht, the Netherlands

Abstract. The increasing global mean temperature will have a fargoing impact on the climate system. Especially systems that contain bifurcations are very vulnerable. The Meridional Overturning Circulation (MOC) in the Atlantic Ocean is such a system. The system is sensitive for the increasing amount of meltwater in the ocean due to meltdown of the big icesheets. At some point, only a small increase of the freshwater flux can result in a big transition of the circulation. There is a need for early warning indicators that can predict when a transition will happen, so that we can either prepare for this or, even better, prevent a transition from happening. The autocorrelation lag-1 indicator is the most well known indicator and should show an increase before the bifurcationpoint. However, this indicator shows only a smooth and linear increase. This makes it hard to set a proper level for an alarm. In this paper we show the link between the autocorrelation indicator and the transfer operator. The transfer operator is a tool that is used to study a system's ensemble behaviour, also called its Liouvillian dynamics. We will see that the autocorrelation reflects only a small piece of the rich information that this transfer operator can provide. Transfer operators haven't been used much in studies on climate physics before, but deserve to be studied more. Their spectra reveal a lot of the dynamics of a system and contain signatures of bifurcations. In future, indicators could be developed that capture more of the information that transfer operators can provide.

1 Introduction

Since the beginning of the industrial revolution the Earth's climate system has been exposed to a big increase in the amount of greenhousegases. A lot of research has been done on the implications that this increase may have on different components of the system. One of these subsystems that is of great interest is the Meriodional Overturning Circulation (MOC) in the Atlantic Ocean. In the current situation the MOC transports a lot of heat from the equator towards northern regions and is therefore very important for the heat distribution on Earth. Moverover, it is responsible for the relatively warm winters of western Europe compared to that of the east of North America. But the MOC can be strongly influenced by climate change. The current increase of the global mean atmospheric temperature will lead to more melt of ice sheets, causing an increased freshwater flux in northern (and southern) basins. The freshwater flux is a key parameter in the salt advection feedback, one of the most important positive feedback mechanisms of the MOC. Due to this positive feedback mechanism the system contains bifurcations. Increasing the amount of freshwater inflow can cause the system to undergo one of these bifurcations. If this happens, the circulation will change drastically and this will have major implications for the northward heat transport. Therefore, it is important to know how far the system is from such a transition.

A lot of research has been done on the development of early warning signals which should be able to predict from climate data when a bifurcation will happen. Most indicators are based on the idea of critical slowdown: just before the bifurcationpoint the system becomes less and less responsive to pertubations. This induces certain changes in some statistical quantities of the timeseries, which can therefore be used as indicators for the approach of a bifurcation. The most wel known indicator based on these ideas is the autocorrelation lag-1 indicator. The autocorrelation of a timeseries should increase before a bifurcation. However, it is questioned how good this indicator is. Ditlevsen and Johnsen (2010) show that conclusions based on solely this indicator are invalid and that it is hard to observe an increased autocorrelation with statistical significance. There is an ask for more trustful indicators that show a sharp change before a bifurcationpoint. Recently, there are also indicators developed based on complex networks (van der Mheen et al., 2013; Feng et al., 2014). The big difference between this approach and the classical one is that the network approach uses timeseries from different points in the basin. Therefore not only temporal correlations are taken into account, but also spatial correlations. This makes them potentially better indicators because they are based on physical changes of the circulation itself. In van der Mheen et al. (2013) the network indicators are compared with the classical indicators. A remarkable difference is that the network indicators show nonlinear changes before the approach of a saddle node bifurcation, while the classical indicators show a more or less linear increase. This makes the network indicators stronger indicators because it is more clear where to set the alarmlevel than in the case of a smooth and linear change. However, the reason for this different behaviour is not well understood.

Another method that is used to study bifurcations is the method of transfer operators. This method is less well known in climate physics, but it can provide a lot of information about the system and is therefore a strong tool. There are analyses done on the normal forms of the pitchfork bifurcation (Gaspard et al., 1995), the Hopf bifurcation (Gaspard and Tasaki, 2001) and the saddle node bifurcation (Tantet and Dijkstra). However, so far as known there are no studies on the transfer operator of the Atlantic MOC, although this could give us a lot of information about the dynamics of the system. Moreover, it could provide a lot of insight in the behaviour of the system before different types of bifurcations. This information will be important if one wants to find a more suitable bifurcation indicator, or even more powerful, a bifurcation specific indicator. We will show here that the autocorrelation indicator is directly linked to the spectrum of the transfer operator.

In this paper we will study the autocorrelation indicator and its relation to the transfer operator. Further, we will show how rich the information extracted from the transfer operator is. In the end we will show the strength of the network based indicators, although we cannot relate them yet to the transfer operators. The paper is organised as follows: in section 2 we explain the methods we used. In section 3.1 we present the bifurcationdiagram, in section 3.2 we present the results of the autocorrelation indicator and link it to the transfer operator based on one observable. In section 3.3 we discuss the richer spectrum of the transfer operator based on two observables. Section 3.4 shows network based indicators for the pitchfork bifurcation. The conclusions and discussions can be found in section 4.

2 Model and methodology

2.1 THCM

To simulate the MOC we use the 2 dimensional ThermoHaline Circulation Model (THCM) (den Toom et al., 2011), which is a fully implicit ocean model. Here we will summarize the most important model configurations. A more detailed model description can be found in appendix A1.

The model can be used to make timeseries, but it can also be used to follow an equilibrium while changing a certain parameter. In this way one can obtain a bifurcation diagram. All of our simulations were done on a meridional cross section in the Atlantic ocean bounded in the meridional direction by 60° S and 60° N and in the vertical direction by a flat bottom at a depth of 4000 m and a flat surface. The grid consists of 32 gridboxes in the meridional direction and 16 gridboxes in the vertical direction. In all the simulations windstress forcing will be absent and the Earth's rotation is set to zero.

At the bottom and lateral boundaries no slip and no flux conditions are applied. At the ocean's surface we impose mixed boundary conditions. The surface temperature T_s is restored according to the following formula:

$$T_s = T_0 + \frac{\Delta T}{2} \cos \frac{\pi \theta}{\theta_n} \tag{1}$$

Here, $T_0 = 15°$ C, $\Delta T = 20°$ C, $\theta_n = 60°$ N and θ is latitude.

The surface freshwater flux F_s is prescribed by:

$$F_s = \beta \frac{\cos \frac{\pi \theta}{\theta_n}}{\cos \theta} \tag{2}$$

In this formula β is the amplitude of the freshwater forcing and will be used as the bifurcation parameter. Notice that the freshwater forcing is symmetrical around the equator. We choose this because then we won't only obtain a saddle node bifurcation, but also a pitchfork bifurcation due to symmetry. Taking an asymmetrical flux with a bigger freshwater forcing in the northern hemisphere than in the southern hemisphere would be more realistic, but will only result in a saddle node bifurcation. We want to study both types of bifurcations because they will have very different implications for the change of the circulation and might show different signatures in the spectrum of the transfer operator.

2.2 Transfer operators

Transfer operators are used in many mathematical studies, but are not so common in climate physics yet. They are used to study the ensemble behaviour of a dynamical system, also called Liouvillian dynamics. They can provide a lot of information about the system and are therefore interesting to study. Loosely speaking, the transfer operator estimates for every i and j what the chance is that an observable, which is initially at position $\mathbf{x}_i$ in phase space, is at position $\mathbf{x}_j$ in phase space a time τ later. Especially the spectrum of the transfer operator is of great interest. To understand the relevance of this operator better, we provide some more background information. Consider the following dynamical system:

$$\dot{\mathbf{x}} = F(\mathbf{x}), \qquad \mathbf{x} \in X, \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{3}$$

Here, $X = \mathbb{R}^d$ is an Euclidean vector space and F: X $\rightarrow$ X a smooth vector field with associated flow $\{S_\tau\}_{\tau \geq 0}$. The evolution of the density of the trajectories $\rho(x)$ in phase space is described by the Liouville equation:

$$\dot{\rho}(t) = -\nabla \cdot (\rho F) = A\rho(x), \qquad \rho(0) = \rho_0 \tag{4}$$

The operator A is called the generator. It generates a one-parameter semigroup of transfer operators $\mathcal{L}_t$. The system of equation 4 has a unique solution given by: $\rho(t) = \mathcal{L}_t \rho_0$. The transfer operator $\mathcal{L}_t$,

which is also known as the Perron-Frobenius operator, yields the evolution of an initial distribution after a time t. More precisely, it is defined such that

$$< \mathcal{L}_\tau f, g >=< f, g \circ \mathcal{S}_\tau >, \quad \text{for all } g \in \mathcal{C}^\infty \tag{5}$$

where $\circ$ is the composition operator and $< f, g >$ is the action of the distribution f on the smooth test function g. The test function g can also be interpreted as an observable.

It is known that the point spectrum of the generator is related to the statistical properties of several mixing dissipative systems, such as the decay rate of correlations (Butterley and Liverani, 2007; Gouezel and Liverani, 2006). It can be shown that the rate at which correlations between two observables decay is given by $-Re(\alpha) > 0$ (Tantet et al., 2014), where α is an eigenvalue of the generator A. The closer an eigenvalue of the generator is to the imaginary axis, the slower the rate of decay of correlations will be. From this, one would expect the eigenvalues to come closer to the imaginary axis when a bifurcationpoint is approached, because then the system undergoes a critical slowdown and so the rate of decay of correlation decreases. In studies of normal forms of bifurcations it has been shown that the eigenvalues of the generator get indeed closer to the imaginary axis when a bifurcationpoint is approached. The way the eigenvalues approach the imaginary axis is bifurcation specific, at least, this is what happens in the case of normal forms of bifurcations. Before the pitchfork the eigenvalues of the generator are multiples of the Liapunov exponent (Gaspard et al., 1995), which is on the stable branch before the supercritical pitchfork bifurcation given by the bifurcation parameter itself. So, the eigenvalues are given by: $s_n = n \cdot \lambda$, where s_n are the eigenvalues, $n \in \mathbb{N}$ and λ is the bifurcation parameter. On the stable branch of a saddle node bifurcation the eigenvalues are also multiples of the Liapunov exponent (Tantet and Dijkstra), but here the Liapunov exponent is given by $\sqrt{\lambda}$. So the eigenvalues are given by: $s_n = n \cdot \sqrt{\lambda}$. Before a Hopf bifurcation the eigenvalues form complex pairs which lie in a triangle around the real axis (Gaspard and Tasaki, 2001). One has to realise that these results are based on studies of normal forms of bifurcations without noise. Here, we will study a more complex, stochastic system which contains multiple bifurcations, so the results of studies of normal forms might not hold in this case. However, we would expect to see some of the signatures of the different bifurcations back in the spectrum.

Only the transfer operator can be directly approximated from timeseries, the generator not. Therefore, the Specral Mapping Theory is needed which tells that the point spectrum $P\sigma(A)$ of the generator is related to the point spectrum $P\sigma(\mathcal{L}_\tau)$ of the transfer operator according to

$$P\sigma(\mathcal{L}_\tau) \backslash 0 = e^{\tau P\sigma(A)} \tag{6}$$

So, an eigenvalue λ of $\mathcal{L}_\tau$ is related to an eigenvalue α of the generator A as: $\lambda = e^{\tau\alpha}$. With this formula the eigenvalues of the generator can be calculated from the eigenvalues of the transfer operator.

Note that the eigenvalue $\lambda = 1$ of the transfer operator corresponds to the eigenvalue $\alpha = 0$ from the generator. This eigenvalue is the first eigenvalue of the spectrum and it will be always there because the corresponding eigenvector $\mathbf{v}$ is the invariant distribution, i.e., $\mathbf{v}\mathcal{L}_\tau = \mathbf{v}$. The invariant distribution respresents the stationary distribution to which all initial density distributions ρ_0 will converge after a long time.

We will approximate the transfer operator at different points before the pitchfork and saddle node bifurcation and from different timeseries as will be explained in section 2.4. Here we explain how the tranfer operator based on two observables is calculated from the timeseries of the full grid. The transfer operator can't be calculated directly from these observables because for this the observables should be independent of each other. Further, the phase space needs to have a lower dimension. Therefore, we project the data on a two dimensional phase space spanned by two independent observables. To do this we project the data on the first and second EOF of the covariance matrix of the different timeseries. This gives the first and second principal component as two orthogonal observables, so the phase space is then reduced to two dimensions. Then, we devide the two dimensional phase space into gridboxes and calculate for each box x_i and for each j, what the chance is that an observable in box x_i is in a box x_j a time τ later. This gives a transition matrix which is an approximation of the transfer operator. We will do the same analysis for another, one dimensional observable, namely a timeseries of the maximum of the streamfunction. The transfer operator can be calculated directly from this timeseries by discretising the one dimensional phase space into gridboxes. Although we expect this spectrum to be less rich than the two dimensional one because of the lower dimension it is based on, it makes sense to study. The autocorrelation indicator is usually calculated from a timeseries of a single observable like the maximum streamfunctionvalue and will therefore probably be more related to the spectrum of the transfer operator of the same one dimensional observable.

2.3 Networks

We will study the behaviour of network indicators based on the ideas of van der Mheen et al. (2013); Feng et al. (2014). The networks are constructed from timeseries of the different gridpoints. First we detrend the data in order to eliminate trends in the data. This is necessary because we are only interested in the short term variability in the timeseries and not in long term trends. We detrend the data using sliding windows and for each window we detrend the data linearly. After detrending the data we calculate the (0-lag) correlation between all the timeseries. If the correlation between two timeseries exceeds a certain threshold, we consider the corresponding gridpoints as linked. However, self connections are not allowed. So, the gridpoints form the nodes of the network and a link between two nodes means that the correlation between the timeseries of these nodes is sufficiently high. We will analyse the changes in the topological properties of the networks as the value of the

bifurcation parameter increases. We will focus our network analysis on the network degree. The degree is calculated for each node and is the total number of links that a node has. There are a lot more network measures like the clustering coefficient, the betweenness and the closeness. The network degree is, however, the most intuitive network measure and has proven its power many times before, which makes it the favorite choice.

It is necessary to choose a right threshold. The correlations between the timeseries must be statistically significant, so the threshold should be a least the smallest correlation value that is statistically significant for a certain confidencelevel. The statistical significance of correlations is determined by the t-distribution. The t-value can be calculated from

$$t = r\sqrt{\frac{N-2}{1-r^2}} \tag{7}$$

where r is the correlation value and N is the number of observations, i.e. the length of each timeseries. The number of degrees of freedom is given by $N-2$. The observed correlations should be high enough such that the corresponding t-value exceeds the critical t-value for which correlations are significant under a certain confidence level. This critical value of t can be looked up in tables from the one-tailed t-distribution or can be calculated from the cumulative distribution function. For a confidence level of 0.1% and 4998 degrees of freedom (we will take timeseries of 5000 years), the critical t-value is 3.0902. This gives a minumum correlation value of 0.044 and this is the minimum value of the threshold. We will take a much higher threshold because with such a low threshold the whole network will be connected. However, the results were tested for different threshold values.

We will use timeseries of the streamfunctionvalue at each gridpoint. The streamfunction is calculated on the borders of each gridbox, so the 16 gridboxes in the vertical give us 17 streamfunctionvalues in this direction. However, we apply no slip conditions on the boundaries of the domain so the streamfunctionvalues at the surface and bottom will always be zero. We will omit these boundarypoints because we are only interested in the variability in the timeseries. This leaves us with 15 nodes in the vertical. In the same way the 32 gridboxes in the meridional direction result in 31 nodes. So in total the network will contain $15 \cdot 31 = 465$ nodes. The maximum degree of a node is therefore 464. If all nodes have a degree of 464 we say that the network is fully connected.

2.4 Simulations

We will select 5 points in the bifurcationdiagram before both the pitchfork and saddle node bifurcation. The points are all equilibrium solutions of the system but for different values of β. These points will all be used as starting points for the timeseries. In these timeseries the bifurcation parameter is fixed. To represent the variability induced by other parameters which are not involved in the model, we put noise on the freshwater forcing. The formula we will use for the freshwater forcing with

noise is:

$$F_s = (1 + 0.1 \cdot \Delta w_r)\beta \frac{\cos \frac{\pi\theta}{\theta_n}}{cos\theta}$$

In this formula θ is the latitude and Δw_r represents white noise with zero mean and standard deviation one.

We will make two types of timeseries: timeseries of the maximum of the streamfunction and timeseries of the streamfunction at each gridpoint. The timeseries of the maximum of the streamfunction are used to calculate the autocorrelation indicator from and the one dimensional transfer operator. The timeseries of the streamfunction at each gridpoint are used to calculate the two dimensional transfer operator and to calculate the networks. All timeseries have a length of 50000 years and a resolution of 1 year, except for the timeseries used for the networks, they have the same resolution but have a length of 5000 years.

3 Results

3.1 Bifurcation diagram

The bifurcation diagram can be found in figure 1. On the x-axis in the figure is the bifurcation parameter β, which is the amplitude of the freshwater flux. On the y-axis the sum of the maximum and minimum of the streamfunction is given. Both before the supercritical pitchfork at $\beta = 1.54$ m year^{-1} and the saddle node on the upper branch at $\beta = 16.10$ m year^{-1} there are 5 points marked, which are the starting points for the timeseries as explained in section 2.4. The β-values of the marked points before the pitchfork can be found in table 1 and those before the saddle node in table 2. Initially, when β is very small, there is only one stable solution. This solution describes a symmetrical circulation with two circulation cells which are equal in strength. There is upwelling around the equator and sinking near the northern and southern boundaries of the basin. When β is increased to about 1.54 m year^{-1} the system undergoes a supercritical pitchfork bifurcation: the stable solution becomes unstable for bigger values of β and two new stable solutions appear. The new stable solution with positive Ψ_{sum} is a circulation with one overturning circulation cell. It has a sinking region in the north, and an upwelling region in the south. This situation is comparable to the present circulation of the Atlantic MOC. The other stable solution that has appeared is the reverse of this circulation: it is a one-cell circulation with a sinking region in the south and an upwelling region in the north. The two asymmetrical circulations have equal strength and are stronger than the symmetrical circulation. At a value of β equal to 12.52 m year^{-1}, there is second pitchfork bifurcation, but this one is subcritical: the unstable symmetrical solution becomes stable again and two new asymmetrical and unstable solutions appear. The two unstable branches of the second pitchfork are connected to the two stable branches of the first pitchfork at a value of β equal to 16.10 m year^{-1}. This forms

a saddle node bifurcation on each of the two stable branches of the first pitchfork. In a saddle node bifurcation a stable and an unstable solution come together in a critical point. After the critical point both solutions cease to exist.

From the paper by den Toom et al. (2011) we know that there is also a Hopf bifurcation at $\beta = 15.62$ m year^{-1}, just before the saddle node bifurcation. In the bifurcation diagram this should be between point 9 and 10. This bifurcation is not indicated in figure 1, but its dynamics will be present in the observations. In a Hopf bifurcation a stable solution becomes unstable after the critical point and a stable limit cycle appears. So, an observable will show oscillatory behaviour after the critical point of this bifurcation. We checked the presence of the Hopf bifurcation in our model by calculating the eigenvalues of the Jacobian. From this spectrum, a Hopf bifurcation can be recognised by a complex conjugate pair of eigenvalues that crosses the imaginary axis. The values of the complex pairs and their corresponding eigenvalue number at the different points before the pitchfork and saddle node can be found in table 3 and in table 4, respectively. We see that there is indeed a complex pair crossing the imaginary axis between point 9 and 10. The complex pair is already present before the pitchfork bifurcation. There, they form the fourth and fifth eigenvalues at most points. From point 7 on they are the two leading eigenvalues of the Jacobian spectrum.

3.2 Autocorrelation indicator and corresponding transfer operator

Figure 2 shows the timeseries of the grid mean streamfunction at each of the 5 points before the saddle node. The timeseries of point 9 and 10 contain a transition. In point 9 the timeseries makes a big oscillation before it transits. This oscillation is due to the Hopf bifurcation. One would expect that the system would stay on the stable limit cycle, but apparently the stable orbit is not stable enough for the amount of noise that is applied. In the bifurcationdiagram it was already seen that at point 9 there are multiple equilibria. Therefore the system can leave the stable orbit and can find another branch instead. In this case the system has turned into a symmetric circulation. Because of the transitions, we will not take point 9 and 10 into account here.

Usually, the autocorrelation at lag 1 is calculated. This value should increase before a bifurcation and can therefore directly be used as an indicator. To be able to show the relation between the autocorrelation indicator and the transfer operator, we will calculate an autocorrelation based indicator in a slightly different way, but the underlying thought is exactly the same. Furthermore, calculating the autocorrelation at lag 1 gives usually very high values and it is the question if changes in these values are then statistically significant. It is better to calculate the timelag for which the autocorrelation has decayed to $\frac{1}{e}$. This is the e-folding timelag τ_e. The e-folding timelag will increase before a bifurcationpoint because it takes more and more time to recover from pertubations when the bifurcationpoint is approached. From the e-folding timelag, we calculate the decay rate which is then

given by $\frac{1}{\tau_e}$ and this should decrease before the bifurcation.

We expect the autocorrelation rate to be linked to the rates of the generator, which are given by: $-Re(\alpha)$, where α is an eigenvalue of the generator. As explained in section 2.2, the eigenvalues of the generator can be calculated from the eigenvalues of the transfer operator. The transfer operator itself can be approximated from the timeseries. The transfer operator was calculated on a grid of 1600x1 and with a timelag of 51 years. Note that we use here the same timeseries as we used to calculate the autocorrelation rates, namely the timeseries of the maximum of the streamfunction. The spectrum of the transfer operator based on more than one observable will be more complete than the spectrum of only one observable. Therefore, the full spectrum of the generator will only be discussed for generators based on 2 observables (see section 3.3). Here, we only focus on the rates of the generator.

The two upper plots in figure 3 show the first 10 rates of the generators at the different points before the pitchfork and saddle node bifurcation. In both figures (almost) all rates decrease when the bifurcationpoint is approached. This is what was expected and what is observed in studies of normal forms as well. The rates are related to the decay rate of correlations and these decrease before a bifurcationpoint due to critical slowdown. The decreasing rates give a clear sign that the bifurcation is approached. The strong point of the rates of the generator is that not only one eigenvalue is changing, but that there are many eigenvalues affected by the bifurcation.

In the plots at the bottom of figure 3, the rates of the autocorrelation are plotted as well, they are indicated by the red squares. The rates are almost exactly the same as the leading rates of the generator, they follow the same linear decrease. This shows that the autocorrelation is strongly linked with the spectrum of the transfer operator: the autocorrelation is mainly determined by the first eigenvalue of the corresponding transfer operator based on one observable. However, not only this first eigenvalue is of interest. As we saw, the rest of the spectrum is affected by the approach of the bifurcation as well. Because the autocorrelation indicator is mainly determined by the first eigenvalue of the transfer operator, and not by the other eigenvalues, it misses a lot of information which could potentially be used for the early detection of tipping points.

3.3 Transfer operator from two observables

Above the transfer operator was calculated from one observable. We expect the transfer operator based on two observables to capture more of the information of the system. Here we use the first and second principal component as our two observables. The transfer operator was calculated on a grid of 40x40 and the timelag we used is 51 years. The results for the pitchfork are robust for timelags between 31 and 71 years and those before the saddle node for timelags between 31 and 91 years.

For shorter lags the spectra change due to the limited length of the timeseries and the projection on a low dimensional observable. The robustness of the results is also tested for different gridsizes and shorter timeseries. The results are robust for different gridsizes and for the chosen grid size of 40x40 they are also robust for timeseries with a length equal to or bigger than 40000 years. The robustness of the spectra is shown in the appendix in section A2.

In figure 4 the spectra of the generator at the different points before the pitchfork are shown. The leading eigenvalues (those which are closest to the imaginary axis) are the most robust ones and they are the most important ones for the dynamics of the system. Furthermore, one has to be careful with drawing conclusion from eigenvalues which are far away from the imaginary axis since we calculated the transfer operator for a lag of 51 years so we might not be able to capture eigenvalues with a much smaller corresponding timescale. At point 1 the two leading eigenvalues form a complex pair. This complex pair is present at all 5 points and it doesn't seem to be affected by the pitchfork bifurcation because it has always a real part of 0.007 yr^{-1}. Behind this complex pair is a second complex pair with a larger imaginary part. This one seems to stay on the same place as well. There are more complex pairs in the back of the spectrum, but one has to be carefull with these since they are less robust. The complex pairs seem to lie in a triangle around the real axis. It is known that the generator of the normal form of a Hopf bifurcation has a spectrum that consists of complex pairs which lie in triangle (Gaspard and Tasaki, 2001). This, and the fact that the complex pairs are not affected by the pitchfork bifurcation make it very likely that the complex pairs in the spectrum are the signature of the Hopf bifurcation. It is very interesting that we can see the signature of the Hopf already far before the bifurcation actually takes place.

There are also multiple eigenvalues which are affected by the pitchfork bifurcation. These have a small imaginary part and move towards the imaginary axis. The approach of the eigenvalues to the imaginary axis can be better seen in figure 6, which shows the first 10 rates at each point before the pitchfork bifurcation. The shift of the rates towards zero can already be seen far before the bifurcation takes place. The first rate in point 1 to 4 is the complex pair, which has a constant rate. In point 5 the complex pair isn't the leading pair anymore. The smallest moving rate seems to depend linearly on the bifurcation parameter. The other rates that move to the imaginary axis don't seem to move in a particular way. However, as argued above, one has to be careful with drawing conclusions from them. Longer timeseries are needed to be able to say more about them. The linear approach of the first moving rate towards zero is in accordance with the results from studies on the normal form of the pitchfork. There, it is found that the rates depend linearly on the bifurcation parameter before a pitchfork bifurcation (Gaspard et al., 1995).

In the spectra of the generators at each of the three points before the saddle node (figure 5) we

observe a shift of eigenvalues towards the imaginary axis as well. This gives again a strong sign that the bifurcationpoint is approached. Now, also the complex pairs move, although they move very slowly. This is probably again the signature of the Hopf bifurcation. The Hopf lies between point 9 and 10, so we would also expect that the complex pairs move to the imaginary axis at points which are so close to the Hopf bifurcation. From figure 7 it can be clearly seen how the first 10 rates evolve when approaching the bifurcation. The rates seem to move linearly to the imaginary axis. From studies on the normal form of the saddle node bifurcation we know that the eigenvalues are multiples of the square root of the bifurcation parameter (Tantet and Dijkstra), but we don't observe that here. Of course, we are looking to a more complex system here so this relation doesn't need to hold here. The analysis is even made more complicated by the closeness of the Hopf bifurcation, so that it is hard to distinguish the signature of the saddle node and that of the Hopf bifurcation. Further, to have a better idea of what is happening one needs more than these three points.

3.4 Network indicators

In this last section of the results we present network based indicators based on the network degree. In van der Mheen et al. (2013); Feng et al. (2014) the strength of the kurtosis of the network degree distribution as an indicator for the saddle node bifurcation has already been shown. Here, we will do the same analysis for the pitchfork bifurcation. The runs we use for this purpose have a length of 5000 years. We use shorter timeseries here than we used before to show that these indicators also work with a more realistic timeseries length. We detrend the data linearly per sliding window and use a window size of 500 years and a sliding size of 10 years. The results are robust for different window- and sliding sizes. We calculate one network at each of the different points using a threshold of 0.65. So, a link between two nodes means that the correlation between the timeseries at these nodes is at least 0.65. From these networks we can calculate the degree of each node in the network.

Figure 8 shows the degree field of the networks at the 5 points before the pitchfork bifurcation. In the network at point 1 we see different bands of degree over the length of the basin. The highest values are found in the middle of the basin at a depth of about 1500 meters. Above and below this band are other bands which have a decreasing degree towards the top and the bottom of the basin. At each point closer to the pitchfork bifurcation the bands with the highest degree in the middle spread out to the top and bottom, so the degree in the whole basin increases. Remarkable is that the networks are not completely symmetric in the equator, as one would expect from a symmetrical circulation. This might be due to the limited length of the timeseries. However, the results were tested for shorter timeseries length and the results were qualitatively the same.

The increase of the degree in the basin can also be seen from the degree distributions in figure 9. As the bifurcationpoint is approached, there are less nodes with a low degree. Furthermore, the

peak of the number of nodes with a very high degree at the right end of the distribution increases. The approach of the bifurcation leads to a deformation of the shape of the degree distribution as the left tail becomes shorter. From this we expect the kurtosis to be a good indicator for the pitchfork bifurcation as well. We also expect the mean degree to increase so this might also be used as an indicator. In figure 10 the mean, variance, skewness and kurtosis of the degreedistributions at each point before the pitchfork bifurcation is given. All moments of the degreedistribution are affected by the pitchfork bifurcation. The mean and kurtosis increase and the variance and skewness decrease. In this sense they can all be used as bifurcation indicators. However, the way the moments change differs. The kurtosis and skewness show a clear nonlinear change. The mean and variance change more or less linearly. It seems that the higher the moment, the more nonlinear the behaviour of the indicator is. A good early warning indicator should show a sharp change before the bifurcationpoint because this makes it clear when the alarm signal should be given. This makes the kurtosis of degree a very good choice as an early warning indicator, especially when it is compared with the linear behaviour that we saw for the classical autocorrelation indicator.

The result above shows that the kurtosis of degree is not only a good indicator for the saddle node bifurcation, but that it also works before the pitchfork bifurcation. Unfortunately, we are not sure yet how to link the nonlinear behaviour of this indicator to the spectrum of the transfer operator.

4 Conclusions & Discussion

The autocorrelation indicator shows a very smooth and linear change before a bifurcationpoint. This makes it a quite poor indicator because it is hard to set a threshold at which an alarm signal should be given. In this paper we linked this linear behaviour to the spectrum of the transfer operator. The autocorrelation indicator follows mainly the first eigenvalue of this spectrum, which evolves linearly. In this way, it focusses only on a small piece of information that could be extracted from the spectrum. We saw that the spectrum of the transfer operator contains a lot of information about the dynamics of the system, especially when it is based on two observables. The signature of the Hopf bifurcation could already be seen long before the bifurcation actually happened. The shift of the multiple rates that are affected by the pitchfork and saddle node bifurcation could be seen well before the bifurcations took place as well. This makes the transfer operator a very powerful tool which can help us to understand the system better. Moreover, it could give more insight about what happens before different types of bifurcations in complex systems. This information will be useful for the development of better early warning indicators. One could think for example about indicators that capture more of the evolution of the other eigenvalues of the spectrum. The change of the spectra before the different bifurcations might also be bifurcation specific, such that bifurcation specific indicators could possibly be developed.

However, before we can do this, a lot of research must be done on the transfer operators in complex systems. Here we were only able to calculate the spectra at different points before the pitchfork reasonably. The analysis before the saddle node was made difficult due to the closeness of the Hopf bifurcation. We could only use three of our five selected points because of early transitions at the other two points. In future research one should study the spectra before the saddle node in more detail by taking more points. It will help to do simulations with less noise, or to study a different saddle node bifurcation without a Hopf in front. Further, the shown spectra before the pitchfork could be improved as well by taking longer timeseries. Although timeseries with a length of 50000 years might sound long, one actually needs even longer timeseries. It is therefore also not the idea that the spectrum of the transfer operator itself could serve as early warning indicator. Timeseries of that length are not available so this is not of practical use. The idea is more to find indicators that reflect more of the spectrum than only the first eigenvalue.

We have shown that the kurtosis of the network degree distribution might be a good indicator for the pitchfork bifurcation as well. The kurtosis showed a nonlinear increase before the bifurcation, which makes it more powerful indicator than the autocorrelation indicator. However, it is still the question why this indicator works so well. There is a theory which links the spectral gap to the roughness of parameters (Chekroun et al., 2014). The spectral gap is the difference between the zero eigenvalue and the first non-zero eigenvalue of the spectrum of the generator. The theory states that big changes in the statistics of an observable are only possible if the spectral gap is small. In our results we found that the spectral gap decreases before the bifurcation points, because the eigenvalues came closer to the imaginary axis. This makes it possible that some parameters show a more rough behaviour. It might be possible that the kurtosis is sensitive for this and that this is why it can increase in a nonlinear way. However, it is not clear how small the spectral gap has to be in order to give this rough behaviour. Further, we are not sure that the nonlinear increase in the kurtosis is indeed linked to this theory. More study is needed to have a better understanding of this indicator.

Bibliography

Butterley, O. and Liverani, C.: Smooth Anosov flows: correlation spectra and stability, Journal of Modern Dynamics, 1, 301–322, 2007.

Chekroun, M., Neelin, J., Kondrashov, D., McWilliams, J., and Ghil, M.: Rough parameter dependence in climate models and the role of Ruelle-Pollicott resonances, PNAS, 111, 1684–1690, 2014.

den Toom, M., Dijkstra, H., and Wubs, F.: Spurious multiple equilibria introduced by convective adjustment, Ocean Modelling, 38, 126–137, 2011.

Ditlevsen, P. D. and Johnsen, S. J.: Tipping points: Early warning and wishful thinking., Geophysical Research Letters, 37, 2010.

Feng, Q., Viebahn, J., and Dijkstra, H.: Deep ocean early warning signals of an Atlantic MOC collapse, Geophysical Research Letters, 41, 2014.

Gaspard, P. and Tasaki, S.: Liouvillian dynamics of the Hopf bifurcation, Physical Review E, 2001.

Gaspard, P., Nicolas, G., and Provata, A.: Spectral signature of the pitchfork bifurcation: Liouville equation approach., Physical Review E, 51, 1995.

Gouezel, S. and Liverani, C.: Banach spaces adapted to Anosov systems., Ergodic Theory and Dynamical Systems, 26, 189–217, 2006.

Tantet, A. and Dijkstra, H.: Liouvillian dynamics of the saddle-node bifurcation.

Tantet, A., van der Burgt, F., and Dijkstra, H.: An early warning indicator for atmospheric blocking events using transfer operators., Chaos, 25, 2014.

van der Mheen, M., Dijkstra, H., Gozolchiani, A., den Toom, M., Feng, Q., Kurths, J., and Hernandez-Garcia, E.: Interaction network based early warning indicators for the Atlantic MOC collapse, Geophysical Research Letters, 40, 2013.

5 Tables

Point number:	β (m year^{-1})
1	0.248
2	0.434
3	0.6448
4	0.992
5	1.426

Table 1: β-values of the marked points before the pitchfork bifurcation

Point number:	β (m year^{-1})
6	13.64
7	14.88
8	15.31
9	15.52
10	15.77

Table 2: β-values of the marked points before the saddle node bifurcation

Point number:	Value of complex pair:	Eigenvalue number:
1	$-0.11 \cdot 10^{-1} \pm 0.34 \cdot 10^{-2} i$	4th and 5th
2	$-0.11 \cdot 10^{-1} \pm 0.34 \cdot 10^{-2} i$	4th and 5th
3	$-0.11 \cdot 10^{-1} \pm 0.34 \cdot 10^{-2} i$	4th and 5th
4	$-0.11 \cdot 10^{-1} \pm 0.34 \cdot 10^{-2} i$	4th and 5th
5	$-0.11 \cdot 10^{-1} \pm 0.35 \cdot 10^{-2} i$	5th and 6th

Table 3: Values of the complex pair of eigenvalues of the Jacobian at the different points before the pitchfork bifurcation and their corresponding eigenvalue number.

Point number:	Value of complex pair:	Eigenvalue number:
6	$-0.58 \cdot 10^{-2} \pm 0.11 \cdot 10^{-1} i$	3rd and 4th
7	$-0.41 \cdot 10^{-2} \pm 0.43 \cdot 10^{-2} i$	1st and 2nd
8	$-0.16 \cdot 10^{-2} \pm 0.45 \cdot 10^{-2} i$	1st and 2nd
9	$-0.58 \cdot 10^{-3} \pm 0.41 \cdot 10^{-2} i$	1st and 2nd
10	$+0.68 \cdot 10^{-3} \pm 0.32 \cdot 10^{-2} i$	1st and 2nd

Table 4: Values of the complex pair of eigenvalues of the Jacobian at the different points before the saddle node bifurcation and their corresponding eigenvalue number.

6 Figures

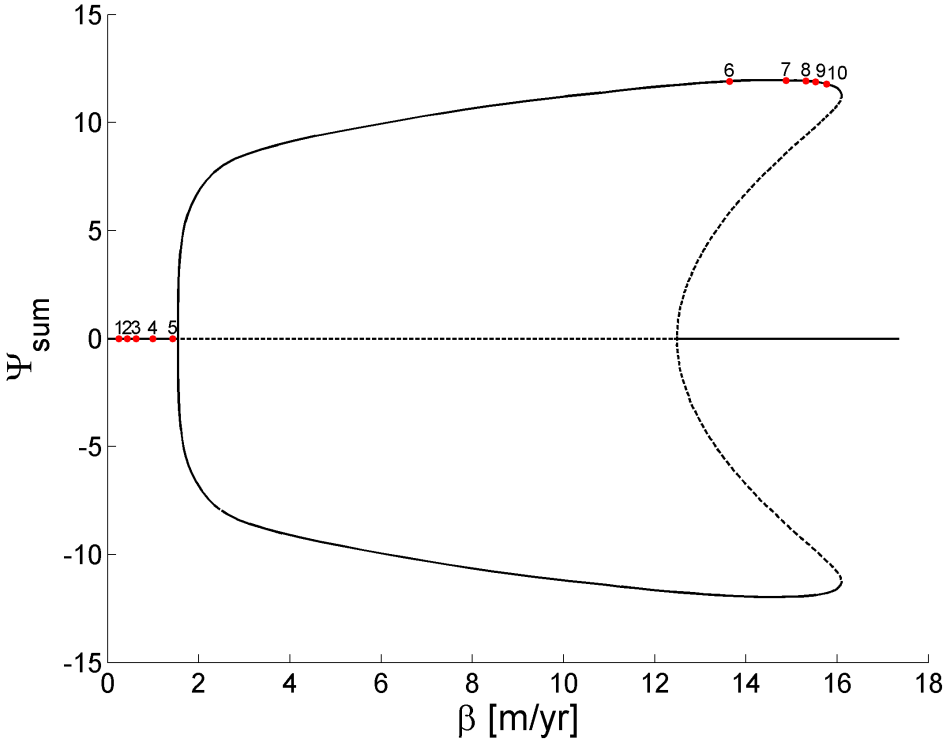


Fig. 1: The deterministic bifurcationdiagram for the symmetrically forced Atlantic MOC. On the y-axis is the sum of the minimum and maximum of the streamfunction, on the x-axis is the bifurcationparameter β. Stable equilibria are indicated with solid lines, unstable equilibria are indicated with dashed lines. The marked points before the pitchfork and saddle node bifurcation are the starting points for the timeseries.

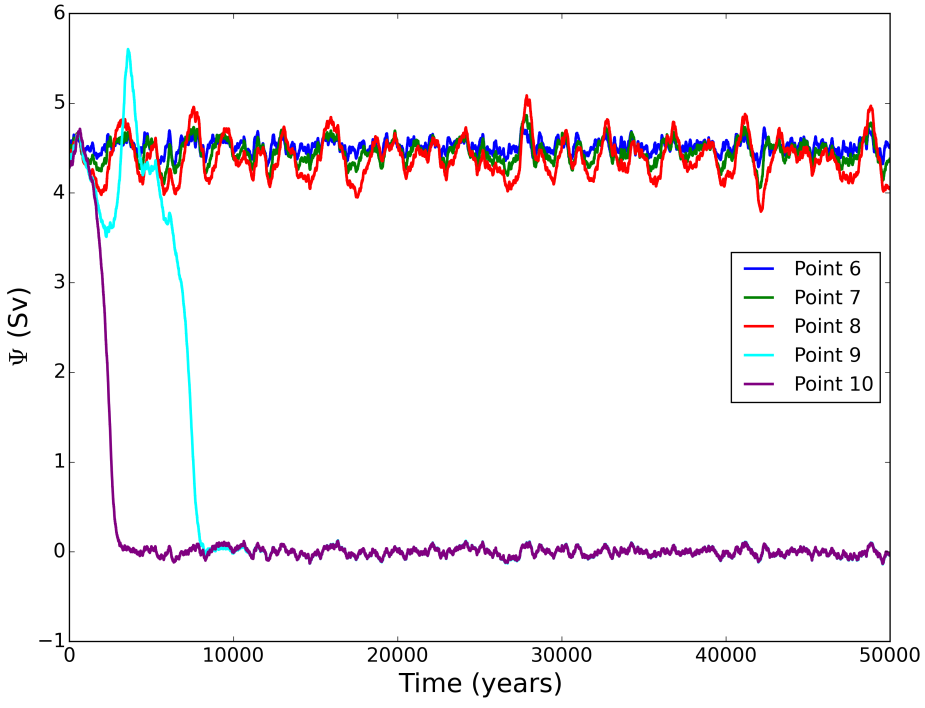
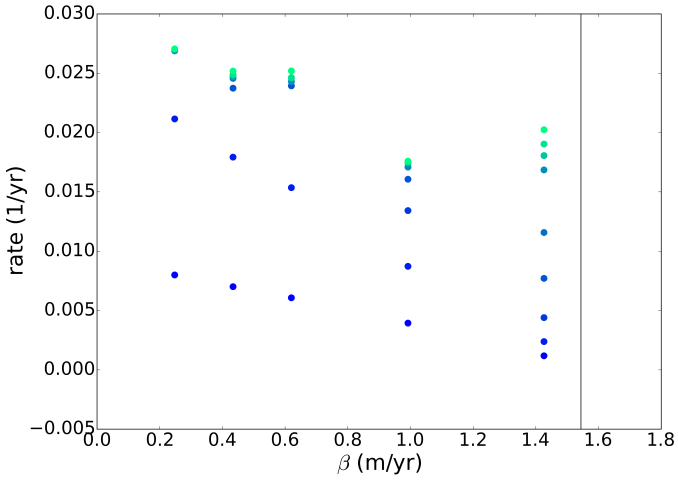
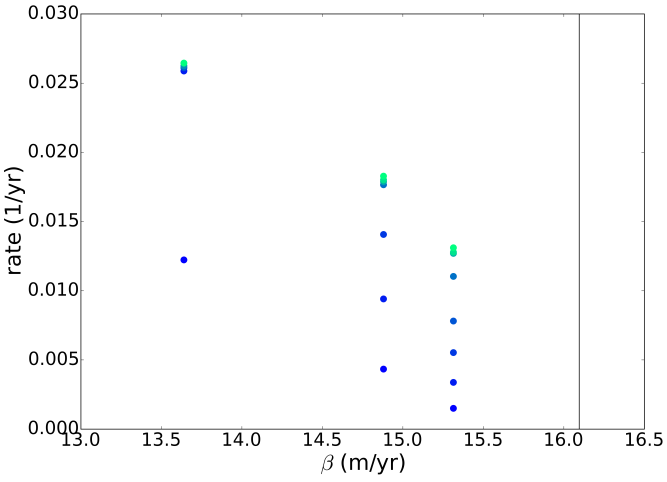


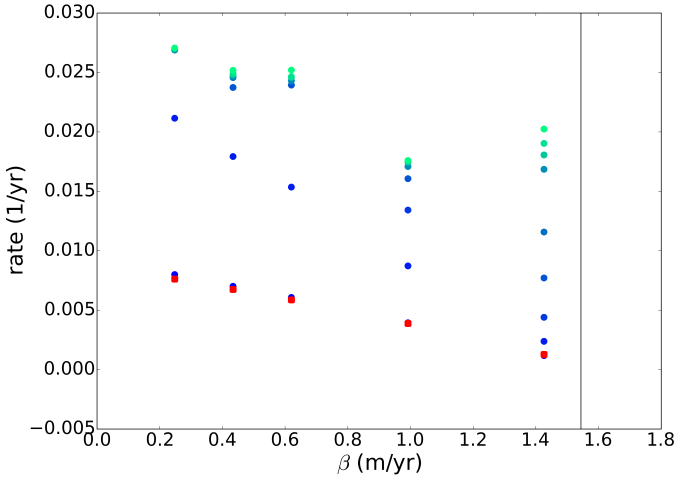
Fig. 2: Timeseries of the 5 different points before the saddle node bifurcation.



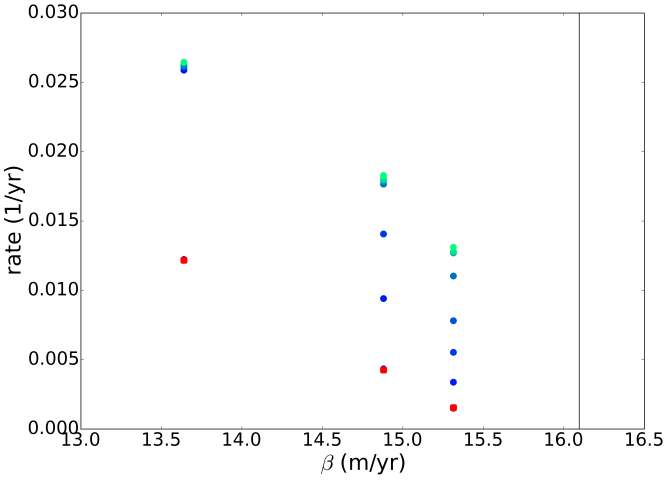
(a) Rates of the generator before pitchfork



(b) Rates of the generator before saddle node

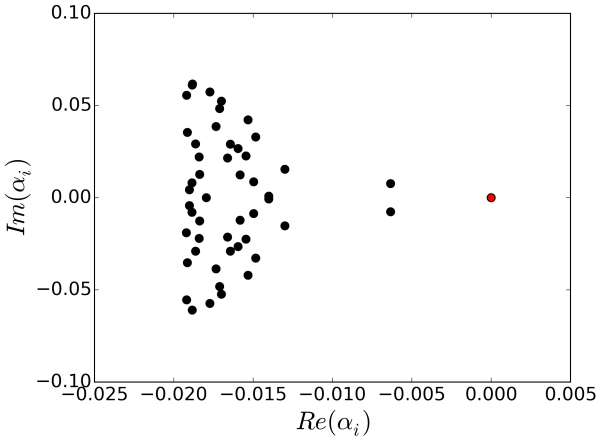


(c) Rates of the generator with rates of autocorrelation before pitchfork

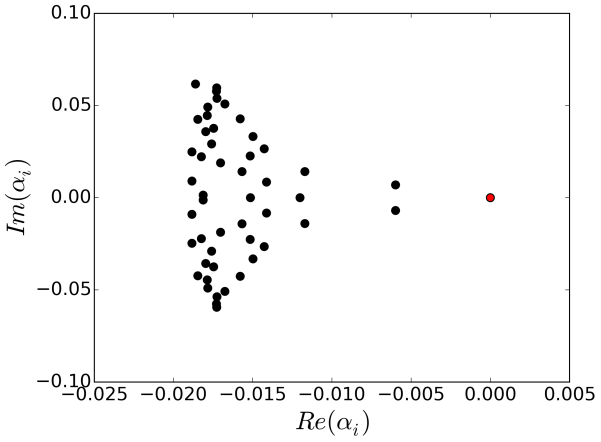


(d) Rates of the generator with rates of autocorrelation before saddle node

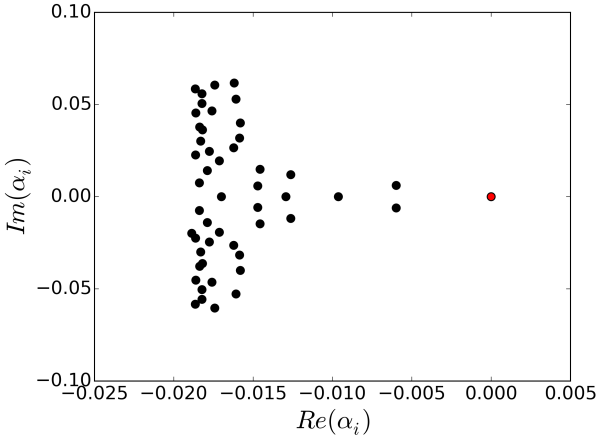
Fig. 3: The plots at the top show the rates of the generator, calculated from one observable, versus the bifurcation parameter. In the plots at the bottom the rates of the e-folding timelag of the autocorrelation are indicated as well by the red squares. The vertical lines at $\beta = 1.54$ m year^{-1} before the pitchfork bifurcation and at $\beta = 16.10$ m year^{-1} before the saddle node bifurcation indicate the position of the bifurcationpoint.



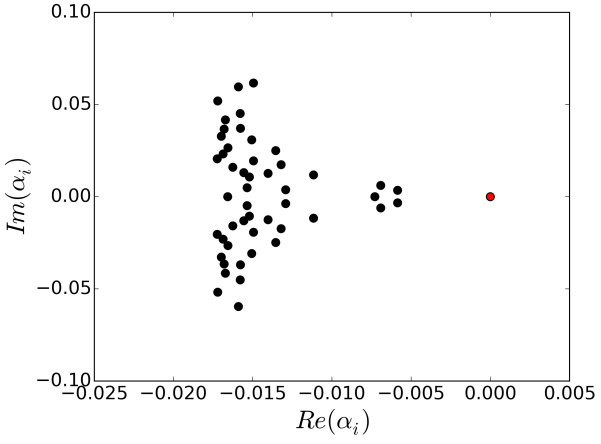
(a) Point 1



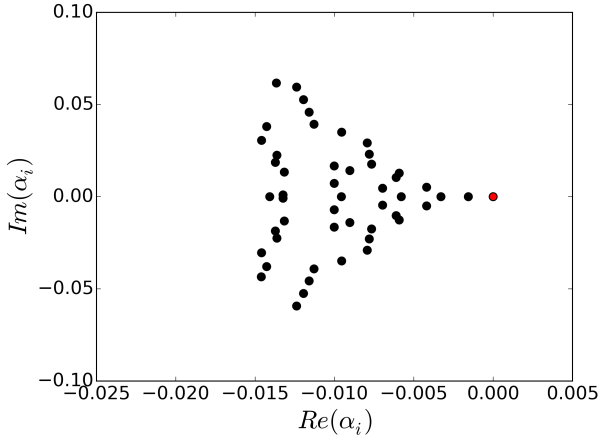
(b) Point 2



(c) Point 3

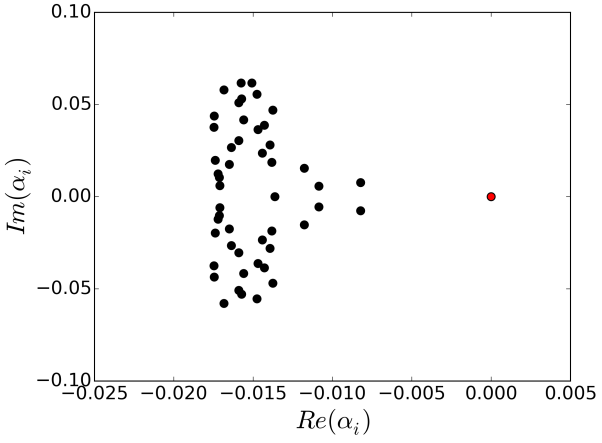


(d) Point 4

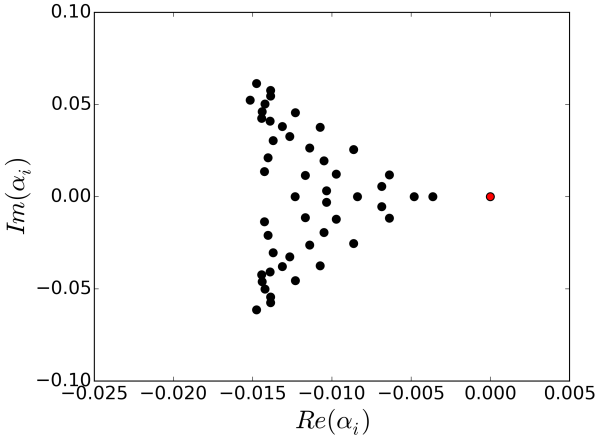


(e) Point 5

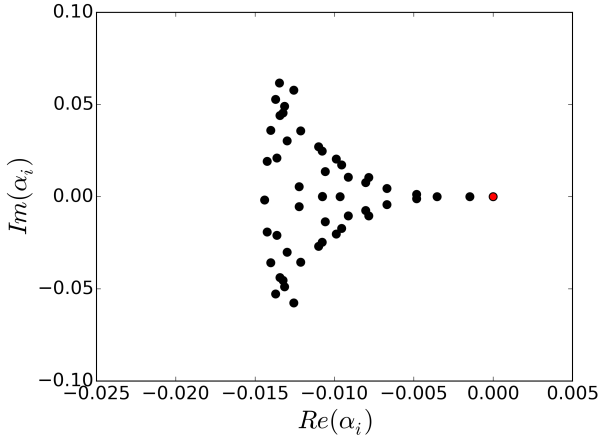
Fig. 4: Eigenvalues α_i of the generators at the different points before the pitchfork bifurcation in complex space. The transfer operator is calculated on a 40x40 grid and for a lag of 51 years.



(a) Point 6



(b) Point 7



(c) Point 8

Fig. 5: Eigenvalues α_i of the generators at the different points before the saddle node bifurcation in complex space. The transfer operator is calculated on a 40x40 grid and for a lag of 51 years.

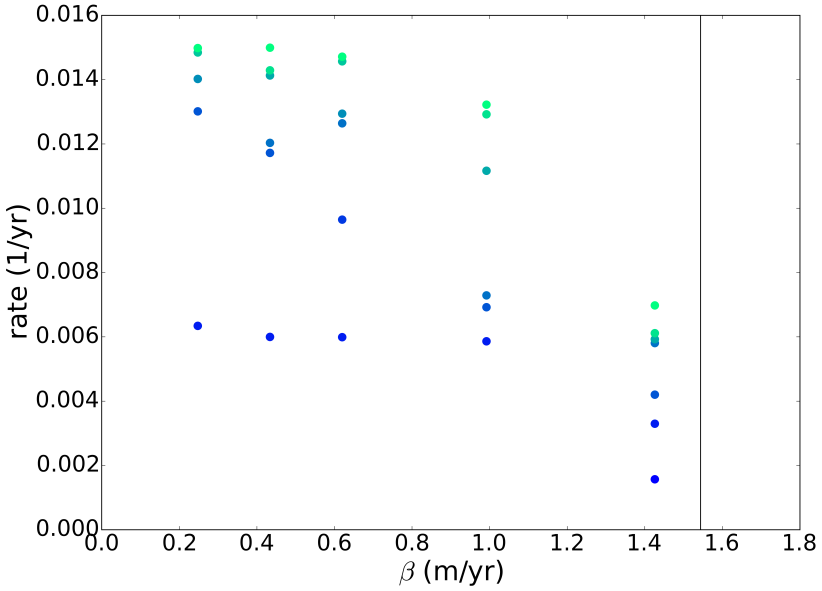


Fig. 6: The first 10 rates of the generator at each point before the pitchfork bifurcation, calculated on a 40x40 grid and for a lag of 51 years. The vertical line at $\beta = 1.54$ m year^{-1} indicates the position of the bifurcation point.

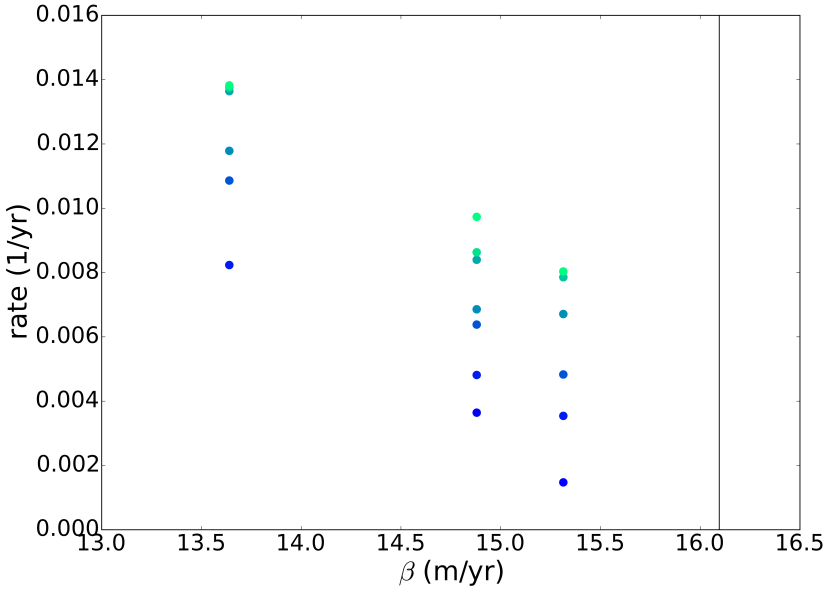


Fig. 7: The first 10 rates of the generator at each point before the saddle node bifurcation, calculated on a 40x40 grid and for a lag of 51 years. The vertical line at $\beta = 16.10$ m year^{-1} indicates the position of the bifurcation point.

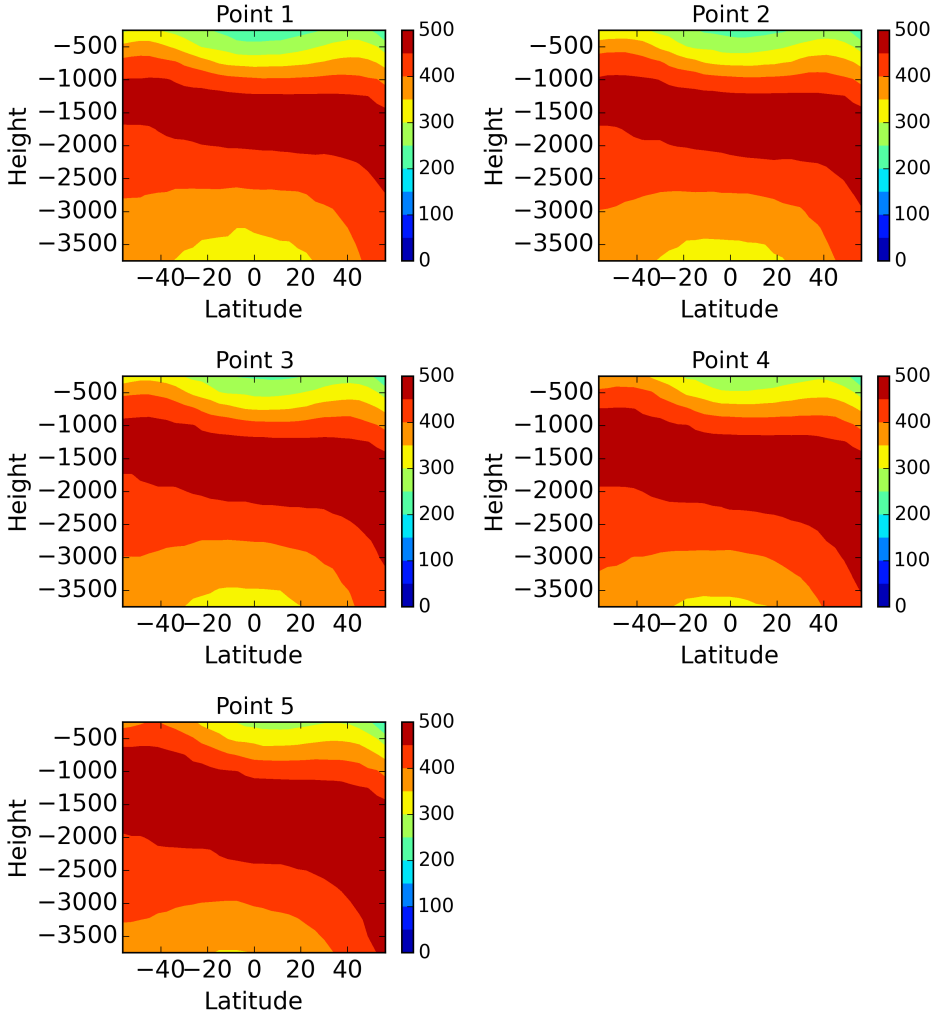


Fig. 8: The degreefield of the networks at the 5 points before the pitchfork bifurcation. On the x-axis is the latitude in degrees, on the y-axis the height in meters.

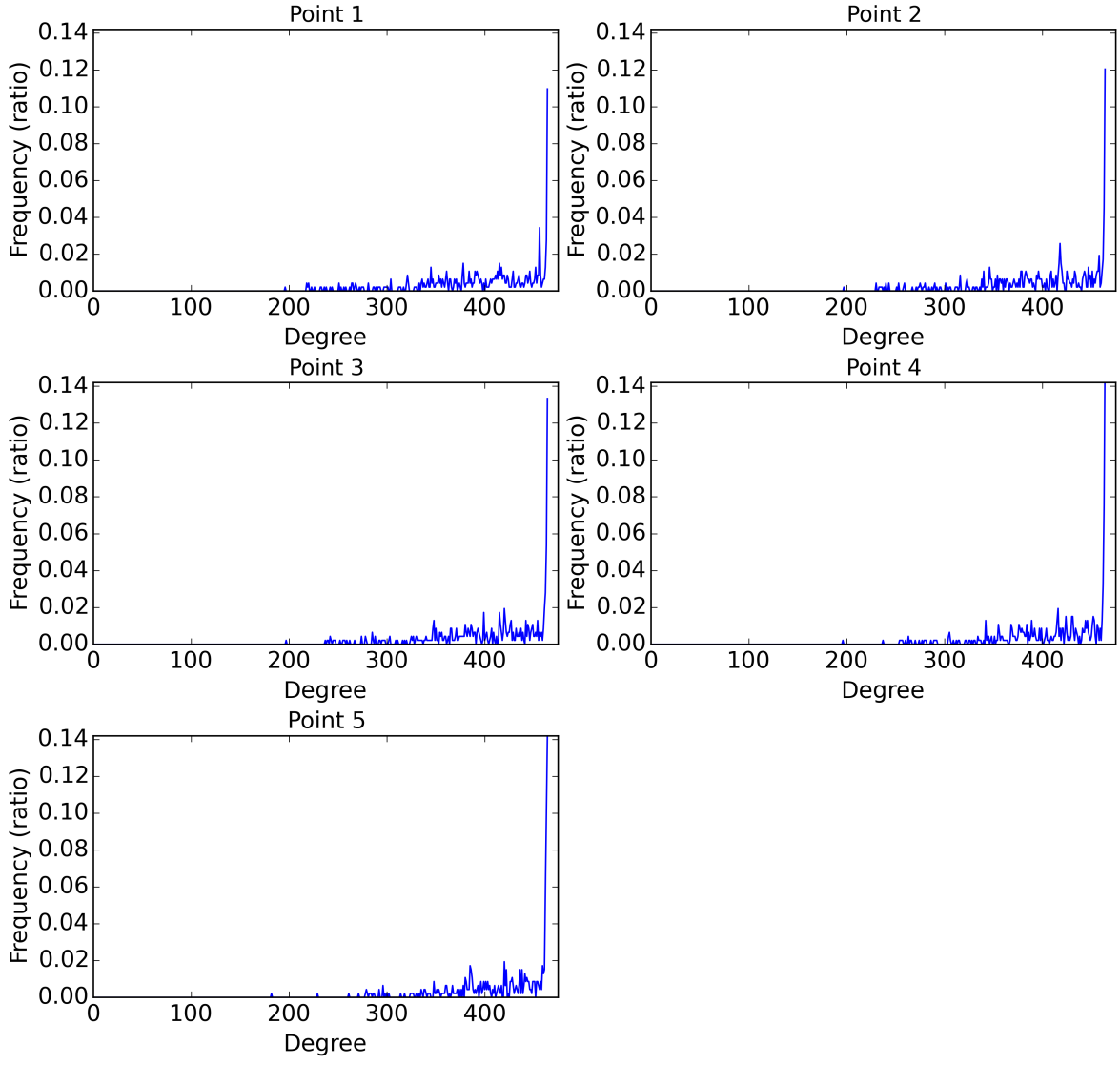


Fig. 9: The degree distribution of the networks at the different points before the pitchfork bifurcation. On the x-axis is the degree, on the y-axis is the number of nodes given as a part of the total number of nodes.

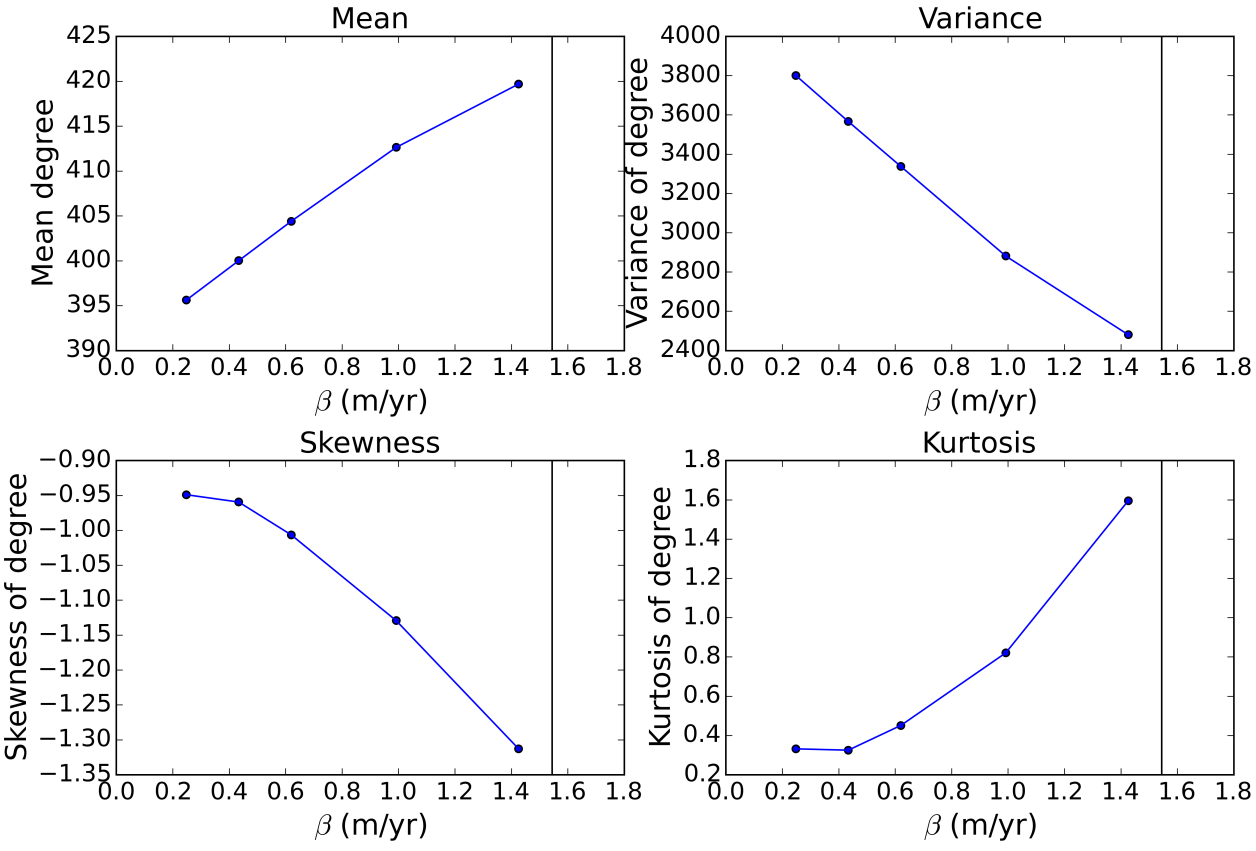


Fig. 10: Mean, variance, skewness and kurtosis of the degree distribution of the networks at the different points before the pitchfork bifurcation. On the x-axis is the bifurcation parameter β, on the y-axis is one of the moments of the degree distribution. The vertical lines at 1.54 m year^{-1} indicate the position of the deterministic bifurcationpoint.

Appendix A

Model description

A1 THCM

The ThermoHaline Circulation Model (THCM) is a fully implicit ocean model. We use a two dimensional version of this model with no wind-stress forcing and zero rotation. The governing model equations are the hydrostatic primitive equations which are then given by

$$0 = -\frac{1}{\rho_0 r_0}\frac{\partial p_*}{\partial \theta_*} + A_V\frac{\partial^2 v_*}{\partial z_*^2} + \frac{A_H}{r_0^2}\Big(\frac{1}{\cos\theta_*}\frac{\partial}{\partial\theta_*}\Big(\cos\theta_*\frac{\partial v_*}{\partial\theta_*}\Big) + (1-\tan^2\theta_*)v_*\Big) \tag{A1}$$

$$\frac{\partial p_*}{\partial z_*} = \rho_* g, \tag{A2}$$

$$0 = \frac{\partial w_*}{\partial z_*} + \frac{1}{r_0}\frac{\partial v_*}{\partial\theta_*} - \frac{v_*\tan\theta_*}{r_0}, \tag{A3}$$

$$\frac{dT_*}{dt_*} = \frac{K_H}{r_0^2\cos\theta_*}\frac{\partial}{\partial\theta_*}\Big(\frac{\partial T_*}{\partial\theta_*}\cos\theta_*\Big) + K_V\frac{\partial^2 T_*}{\partial z_*^2}, \tag{A4}$$

$$\frac{dS_*}{dt_*} = \frac{K_H}{r_0^2\cos\theta_*}\frac{\partial}{\partial\theta_*}\Big(\frac{\partial S_*}{\partial\theta_*}\cos\theta_*\Big) + K_V\frac{\partial^2 S_*}{\partial z_*^2}. \tag{A5}$$

Here, $\frac{d}{dt_*} = \frac{\partial}{\partial t_*} + \frac{v_*}{r_0}\frac{\partial}{\partial\theta_*} + w_*\frac{\partial}{\partial z_*}$ is the material derivative, θ_* the latitude and z_* depth. The radius of the Earth is represented by r_0, v_* and w_* are the meridional and vertical velocity components, respectively, pressure is represented by p_*, temperature by T_* and salinity by S_*. The density ρ_* is related to the temperature and salinity by the linear equation of state

$$\rho_* = \rho_0(1 - \alpha_T(T_* - T_0) + \alpha_S(S_* - S_0))$$

with expansion coefficients α_T and α_S and reference temperature T_0, salinity S_0 and density ρ_0.

Mixing is represented by eddy diffusivities, with horizontal and vertical diffusivities K_H and K_V for both heat and salt, and friction coefficients A_H and A_V for momentum. At the lateral and bottom boundaries, no-slip and no-flux conditions are imposed. At the ocean-atmosphere interface mixed boundary conditions are imposed. The surface temperature is restored to a temperature profile T_S,

$$T_S = T_0 + \frac{\Delta T}{2}\cos\frac{\pi\theta}{\theta_N},$$

where $\Delta T = 20°C$ and $\theta_N = 60°N$. The freshwater forcing is symmetrical around the equator and is prescribed by

$$F_S = \beta\frac{\cos\frac{\pi\theta}{\theta_N}}{\cos\theta}.$$

β is the amplitude of the freshwater forcing and is used as bifurcation parameter.

The equations are discretised in space using an Arakawa B-grid that places p, T and S in the center of a grid cell and the v and w on its boundaries. This is decribed in more detail in den Toom et al. (2011). To calculate branches of steady states directly as a function of the control parameter, pseudo-arclength continuation is used. With this technique, unstable solutions can also be determined. To converge to individual solutions, the Newton-Raphson method is used. The model also implements the Jacobi-Davidson QZ method to solve linear stability problems.

We consider a two dimensional, meridional cross section in the Atlantic Ocean with a width of $64°$, which is relevant for the value of the strength of the MOC. The cross section is bounded in the vertical by a flat bottom at a depth of 4000 meters and a flat surface. In the meridional direction it is bounded by the latitudes $60°$S and $60°$ N. The grid contains 32 points in the meridional direction and 16 points in the vertical direction. This gives then a meridional resolution of $3.75°$ and a vertical resolution of 250 m.

All model parameter values can be found in the table below.

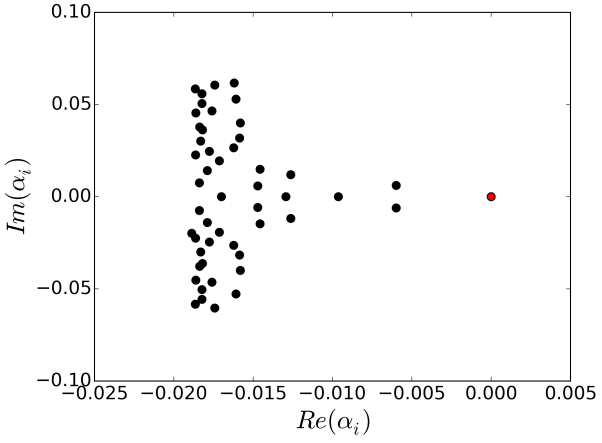
$r_0 = 6.37 \cdot 10^6$ m	$\rho_0 = 1.0 \cdot 10^3$ kg m^{-3}
$g = 9.8$ m s^{-2}	$\alpha_T = 1.0 \cdot 10^{-4}$ K^{-1}
$A_H = 2.2 \cdot 10^{12}$ m^2 s^{-1}	$\alpha_S = 7.6 \cdot 10^{-4}$ psu^{-1}
$A_V = 1.0 \cdot 10^{-3}$ m^2 s^{-1}	$T_0 = 15°C$
$K_H = 1.0 \cdot 10^3$ m^2 s^{-1}	$S_0 = 35.0$ psu
$K_V = 1.0 \cdot 10^{-4}$ K^{-1}	$H = 4000$ m
$\theta_N = 60°$	

Table 5: Parameter values of the two-dimensional model.

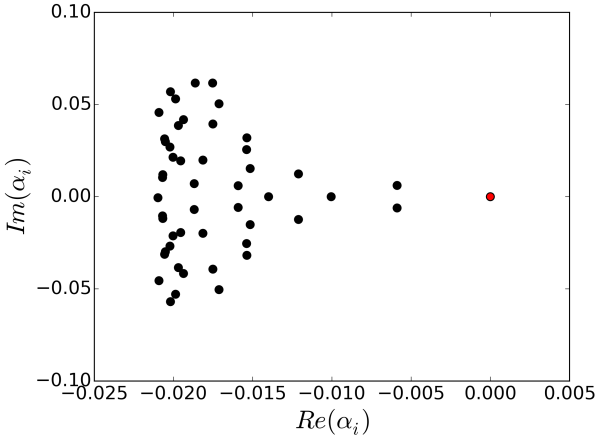
A2 Robustness of the 2D transfer operator

The robustness of the spectra of the generator was checked for different gridsizes, different timelags and shorter timeseries length. This was done for point 3 and point 7.

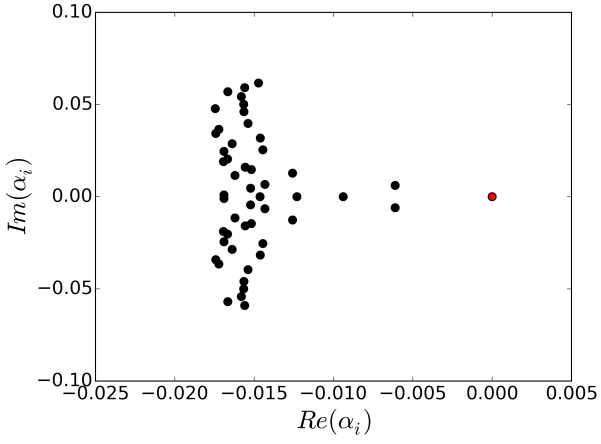
A2.1 Robustness for the chosen gridsize



(a) Spectrum for a 40x40 grid

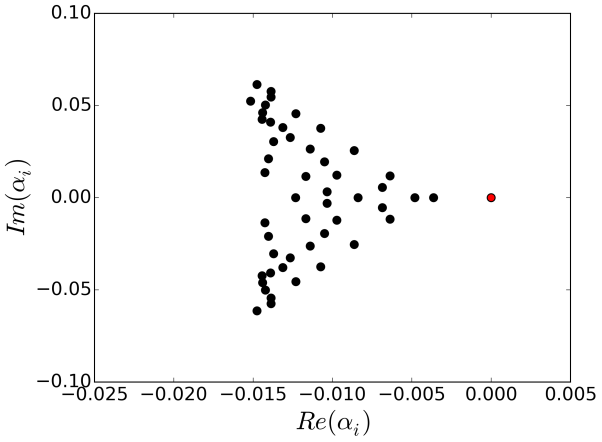


(b) Spectrum for a 30x30 grid

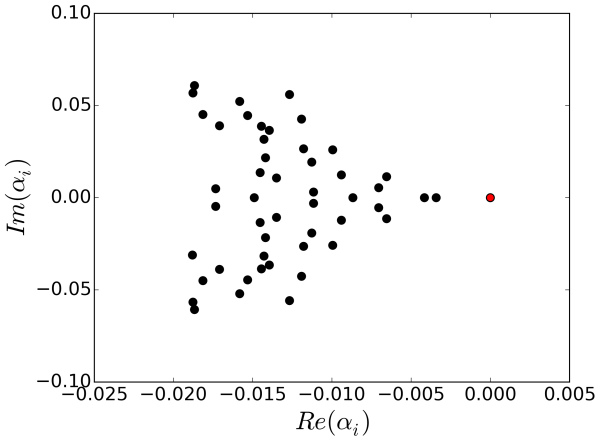


(c) Spectrum for a 45x45 grid

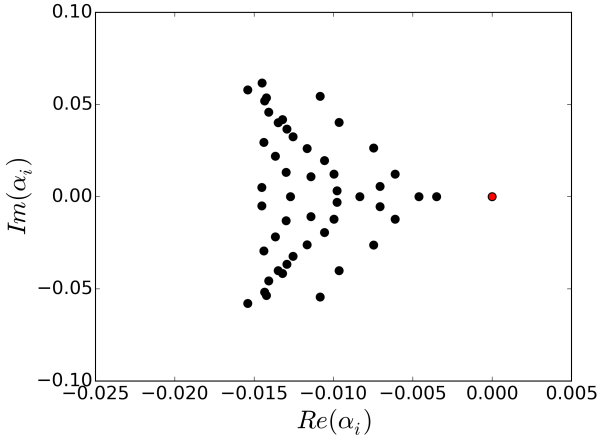
Fig. 11: Spectra at point 3 for different gridsizes. The spectrum with a grid of 40x40 was shown in the results. All spectra are calculated for a timelag of 51 years and for a timeserieslength of 50000 years.



(a) Spectrum for a 40x40 grid



(b) Spectrum for a 30x30 grid

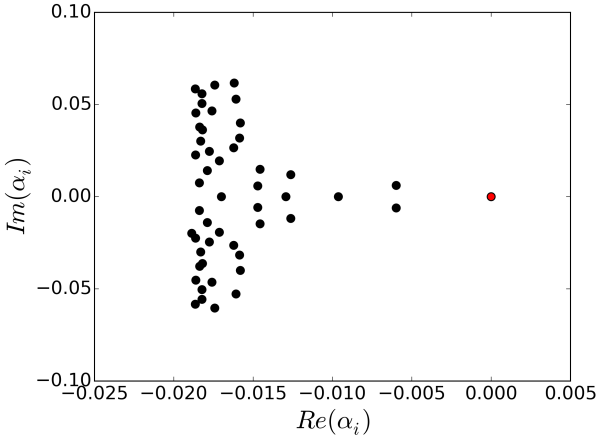


(c) Spectrum for a 45x45 grid

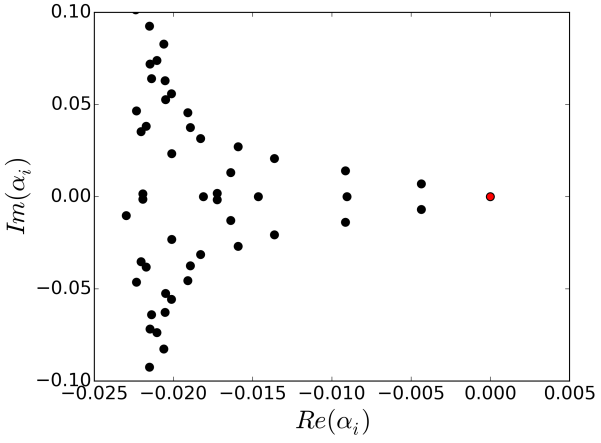
Fig. 12: Spectra at point 7 for different gridsizes. The spectrum with a grid of 40x40 was shown in the results. All spectra are calculated for a timelag of 51 years and for a timeserieslength of 50000 years.

The spectra are quite robust for different gridsizes. At least the first 6 rates stay approximately the same. Important is as well that the complex pairs are also observed for different gridsizes.

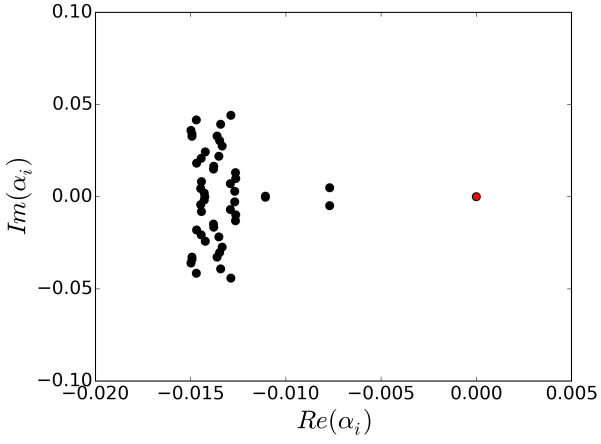
A2.2 Robustness for the chosen timelag



(a) Spectrum for a 51 year timelag

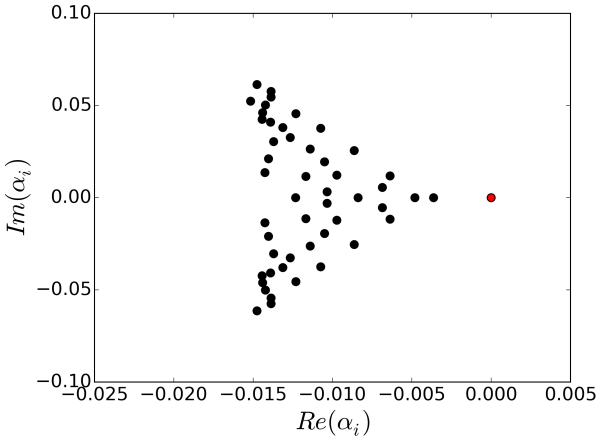


(b) Spectrum for a 31 year timelag

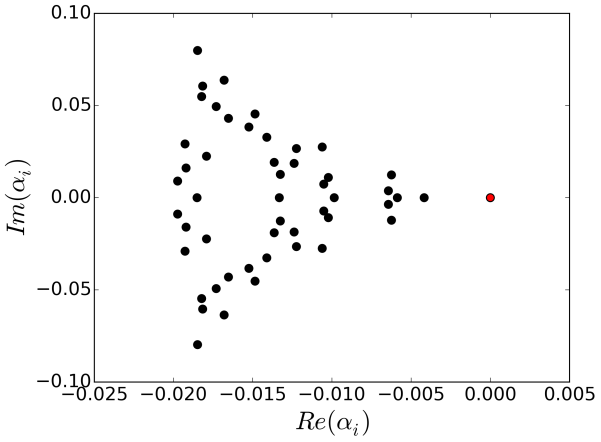


(c) Spectrum for a 71 year timelag

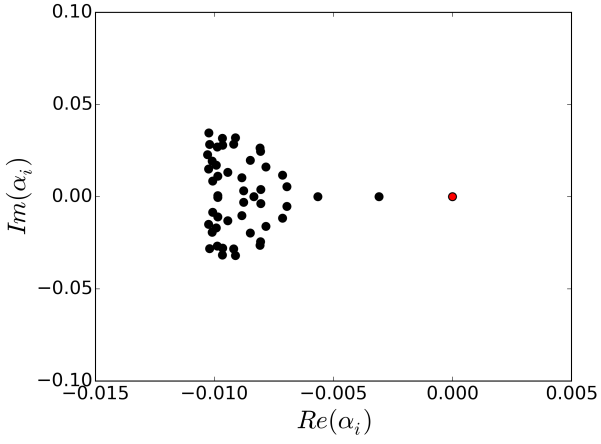
Fig. 13: Spectra at point 3 for different timelags. The spectrum for a lag of 51 years was shown in the results before. All spectra are calculated on a 40x40 grid and for a timeserieslength of 50000 years.



(a) Spectrum for a 51 year timelag



(b) Spectrum for a 31 year timelag



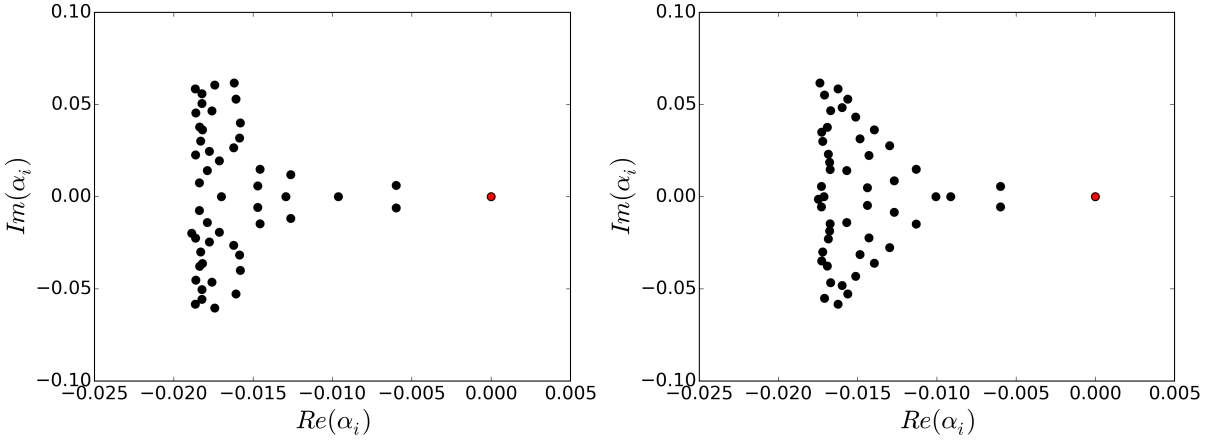
(c) Spectrum for a 91 year timelag

Fig. 14: Spectra at point 7 for different timelags. The spectrum for a lag of 51 years was shown in the results. All spectra are calculated on a 40x40 grid and for a timeserieslength of 50000 years.

The spectrum of the pitchfork is quite robust for timelags in a range of 31 to 71 years. For this range, the first few rates stay qualitatively the same. The spectrum of the saddle node is more robust for the chosen timelag, it is robust in a range of 31 to 91 years. Theoretically, the spectra shouldn't be dependent on the chosen timelag. However, because of the limited length of the timeseries, the projection on a lower dimensional observable and the discretisation of the phase space they become a bit more lag dependent.

A2.3 Robustness for shorter timeseries

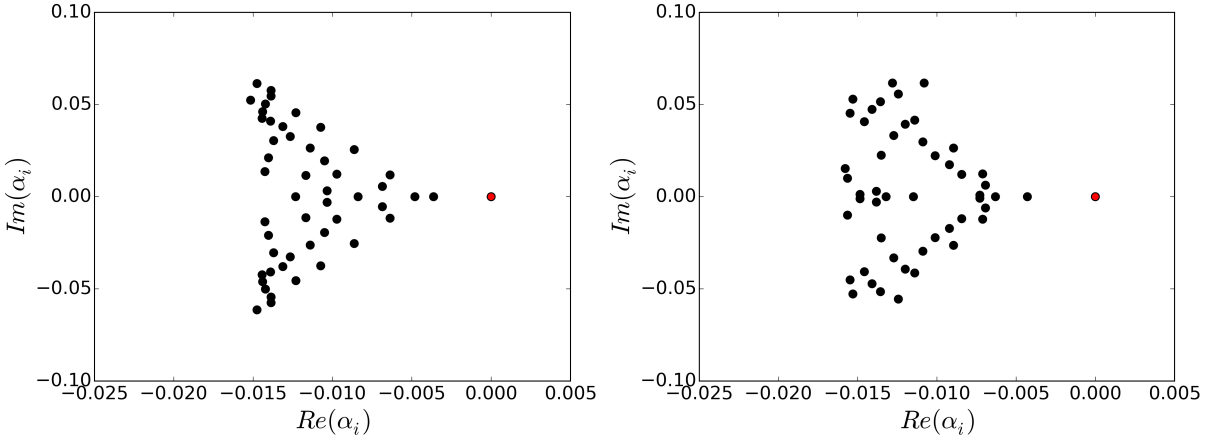
Last, we check the robustness for shorter timeseries.



(a) Spectrum for a timeseries of 50000 years.

(b) Spectrum for a timeseries of 40000 years.

Fig. 15: Spectra at point 3 for different timeseries length. The spectrum with a timeseries length of 50000 years was shown in the results. Both spectra were calculated on a 40x40 grid and for a timelag of 51 years.



(a) Spectrum for a timeseries of 50000 years.

(b) Spectrum for a timeseries of 40000 years.

Fig. 16: Spectra at point 7 for different timeseries length. The spectrum with a timeseries length of 50000 years was shown in the results. Both spectra were calculated on a 40x40 grid and for a timelag of 51 years.

The spectra are both quite robust for shorter timeseries length. Again, the most important eigenvalues, namely the leading ones, are most robust.