Shadowing the rotating annulus. Part I:
Measuring candidate trajectory shadowing times

Roland M. B. Young[a,b,*], **Roman Binter**[a], **Falk Niehörster**[a,c]

[a]Centre for the Analysis of Time Series, London School of Economics, London, UK
[b]Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK
[c]Bermuda Institute of Ocean Sciences, St. George's, Bermuda

Abstract

An intuitively necessary requirement of models used to provide forecasts of a system's future is the existence of *shadowing* trajectories that are consistent with past observations of the system: given a system-model pair, do model trajectories exist that stay reasonably close to a sequence of observations of the system? Techniques for finding such trajectories are well-understood in low-dimensional systems, but there is significant interest in their application to high-dimensional weather and climate models. We build on work by Smith et al. [2010, *Phys. Lett. A*, 374, 2618–2623] and develop a method for measuring the time that individual "candidate" trajectories of high-dimensional models shadow observations, using a model of the thermally-driven rotating annulus in the perfect model scenario. Models of the annulus are intermediate in complexity between low-dimensional systems and global atmospheric models. We demonstrate our method by measuring shadowing times against artificially-generated observations for candidate trajectories beginning a fixed distance from truth in one of the annulus' chaotic flow regimes. The distribution of candidate shadowing times we calculated using our method corresponds closely to (1) the range of times over which the trajectories visually diverge from the observations and (2) the divergence time using a simple metric based on the distance between model trajectory and observations. An empirical relationship between the expected candidate shadowing times and the initial distance from truth confirms that the method behaves reasonably as parameters are varied.

This paper was submitted to *Physica D* in 2010, but, after review, was not accepted. We no longer have the time or resources to work on this topic, but would like this record of our work to be available for others to read, cite, and follow up.

Keywords: Shadowing trajectories; Rotating annulus; Dimensional scaling; Weather and climate forecasting; Chaotic dynamics; Perfect Model Scenario.

1. Introduction

In forecasting one is often forced to use complex numerical models to predict the future behaviour of physical or mathematical systems. A good example are General Circulation Models (GCMs), which are used to predict the future behaviour of the atmosphere. GCMs have state spaces of over ten million dimensions and integrate differential equations describing complex dynamical and chaotic behaviour. Decisions on how to act over future timescales of days (in the case of weather) up to decades (in the case of climate) are made based on the output from these models, and so decision makers desire these forecasts to be informative.

An intuitive requirement of models used to provide forecasts of a system's future is the existence of model trajectories that are consistent with past observations of the system. Given a system-model pair, do model trajectories exist that stay reasonably close to a sequence of observations of the system? Such trajectories are said to *shadow* the observations. While the existence of a shadowing trajectory is not a *sufficient* condition for successful forecasts, it is intuitively a *necessary* one.

The concept of shadowing appears in several slightly different contexts, and it is important to differentiate between these. In its original context, so-called ϵ-shadowing, the shadowing problem is concerned with whether

---

*[*]Corresponding author. Current address: College of Science, UAE University, P.O. Box 15551, Al Ain, United Arab Emirates. *Email address:* roland.young@uaeu.ac.ae.*
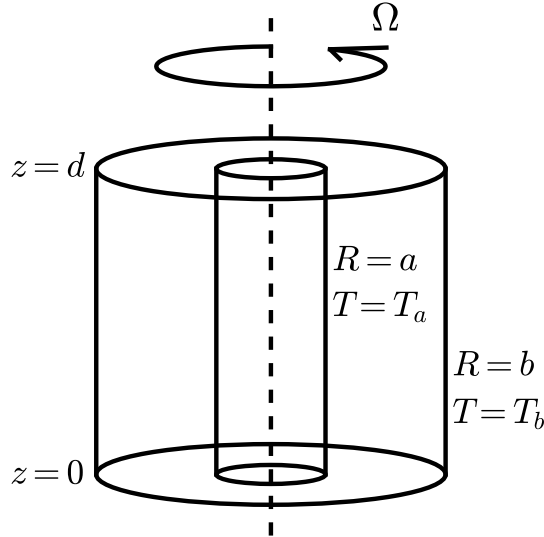
Figure 1: Schematic of the thermally-driven rotating annulus experiment. The inner cylinder is at radius $a = 2.5$ cm, the outer cylinder is at radius $b = 8.0$ cm, and the cylinders have a depth of $d = 14.0$ cm. The fluid between the cylinders is a mixture of glycerol and water, 17%/83% by volume. The cylinders are maintained at constant temperatures $T_a = 18\,°$C and $T_b = 22\,°$C respectively, giving a constant temperature difference across the fluid of $\Delta T = 4\,$degC. The whole system is rotated about a vertical axis with angular velocity $\Omega = 1.00\,$rad s^{-1}.

there exists a true solution to a differential equation that remains sufficiently close to a given numerical solution. If no true solution exists that shadows the numerical solution, can we learn anything at all about the true solution from the numerical solution? The ϵ-shadowing time is formalised in the Anosov-Bowen shadowing lemma (*Bowen*, 1975). Second, a superficially similar but fundamentally different context is ι-shadowing (*Gilmour*, 1998). This is concerned with whether a trajectory of a numerical model exists that shadows a given sequence of observations of a system. The ι-shadowing time is the maximum model trajectory length that, among all possible trajectories, remains consistent with the observational noise, and is defined with respect to a particular noise model (*Gilmour*, 1998, p. 47). In this context, if the model does not admit such a trajectory then there is no such initial condition that remains close to the observations, and the problem is then one of model error, not chaos. Third and finally, the ϕ-shadowing problem (*Smith*, 2000, p. 52) is concerned with whether there exists a model trajectory that remains close enough to a given set of observations for "useful" forecasts to be possible. As what is "useful" depends on context, this use of the term is necessarily subjective. For example, one could argue that a GCM solution that predicts the wrong amount of rain in a given place but does predict that rain will fall is "useful" enough to constitute a ϕ-shadowing trajectory.

In the atmospheric context we are interested in how well particular models shadow observations. Several methods exist that attempt to find shadowing trajectories directly from a sequence of observations (e.g. *Gilmour*, 1998; *Stemler and Judd*, 2009), and these methods are well-understood using low-dimensional (low-N) systems such as the Lorenz equations (*Lorenz*, 1963) ($N = 3$) and the Ikeda map (*Ikeda*, 1979) ($N = 2$). Whether these methods are feasible when applied to high-dimensional models of the atmosphere remains poorly understood, however, as is the time over which atmospheric models can shadow past observations.

Intermediate in complexity between low-dimensional systems and the atmosphere are laboratory experiments, such as the rotating annulus (*Hide*, 1953; *Hide and Mason*, 1975; *Hignett et al.*, 1985; *Read et al.*, 1992). This system consists of a fluid contained between two concentric cylinders, rotating about a vertical axis with a horizontal temperature gradient between the cylinders (Fig. 1). This setup mimics the major influences acting on a planet's atmosphere: the effects of rotation, gravity, and the temperature difference between low and high latitudes. The rotating annulus has been well-established over 50 years as a useful analogue for certain types of atmospheric dynamical behaviour, especially the general circulation of the atmosphere at mid-latitudes (*Winston and Krueger*, 1961; *Lorenz*, 1967). Models used to simulate the behaviour of fluid in the annulus are intermediate in complexity between low-dimensional systems and global atmospheric models; the dimension of these models are $O(10^4)$, compared with $O(10)$ for low-dimensional systems and $O(10^7)$ for GCMs. Experimental systems are also intermediate in their physical idealization of the system being studied, between GCMs with a low degree of idealization, and systems such as the Lorenz equations, which describe a highly-idealized representation of convective rolls at the onset of convection. This paper is the first in a larger body of work exploring shadowing

2

techniques using the rotating annulus. We aim to understand shadowing methods better by examining them in a real but idealised system, and also improve our understanding of the annulus and the models used to study it through their shadowing properties.

Later work will address the problem of finding the longest model trajectories that shadow a given sequence of observations. Here we focus on how to measure the shadowing time itself. Most methods that look for shadowing trajectories[1] generate a set of *candidate* trajectories, each of which may or may not shadow the observations. One then has to measure how long each candidate shadows, and the longest of these *candidate shadowing times* gives a lower bound on the true shadowing time for that model — observation pair. The main aim of this paper is to present a method for identifying shadowing times that can be used with high-dimensional systems, like the rotating annulus, when the observational error distribution is unbounded. We begin from a definition outlined in *Smith et al.* (2010) which has some problems when applied to high-dimensional systems, and present a modified form that solves these problems.

We use the annulus to illustrate the method by examining some of the shadowing properties of our annulus model using the perfect model scenario (PMS). The aim is to demonstrate that our method yields shadowing times that are reasonable. For example, we build a statistical distribution showing how long a candidate might expect to shadow observations if it starts a particular distance from the system's true state. When the initial conditions lie off the system's attractor then a reasonable method for identifying shadowing times in the PMS should, in general, yield longer times the closer the initial condition is to truth. This result will establish a benchmark for future work when we search for shadowing trajectories by generating candidates from observations using the gradient descent method (*Judd*, 2003; *Judd et al.*, 2004). As a sanity check we also use a simple distance-based metric to determine whether our shadowing times are reasonable; they should capture the time the model visually diverges from the observations. The PMS provides a robust framework for investigating shadowing in complex systems, as it allows many aspects of the setup to be controlled, and it also allows us to avoid dynamical problems due to the ϵ-shadowing properties of the model.

In Sect. 2 we present an analysis of the *Smith et al.* (2010) shadowing definition and describe our alternative for high-dimensional systems. Then in Sect. 3 we describe the rotating annulus model and the experimental setup, before investigating how the shadowing time depends on the distance of the candidate state from truth in Sect. 4. Finally we discuss our results and draw conclusions in Sect. 5.

2. A new method for measuring shadowing times in high-dimensional systems

2.1. The perfect model scenario

Consider a model f of a system $\tilde{f}$. If the model is a perfect representation of the system, $f \equiv \tilde{f}$, then there always exists a trajectory of the model that produces an infinitely long perfect shadow of the system. To investigate the limits of shadowing it is useful to work in the perfect model scenario (PMS), where the model and the system are equivalent ($f \equiv \tilde{f}$), but noise does obscure the system state. In the PMS, the only limit to shadowing the system is the uncertainty in a trajectory's initial state. When modelling physical systems this is never the case, because models are always imperfect representations of the system ($f \not\equiv \tilde{f}$). We call this setting the imperfect model scenario (IMS), and then model imperfections and dynamical aspects of shadowing a numerical solution (see below) limit the shadowing capabilities of the model. In practice, in real systems model imperfections will be the overwhelmingly larger of the two.

In this paper we only use the PMS; the main advantage being that we know the true state of the system. This allows us to generate observations by purposely obscuring the true state by a known noise model of observational error Σ. Controlled experiments can then provide insights into the properties of shadowing trajectories under initial condition uncertainty alone. We measure how long a model trajectory started a given distance from the system state $f(\mathbf{x} + \delta)$ stays consistent with observations of the system $\tilde{f}(\mathbf{x}) + \mathbf{e}$, where $\mathbf{e} \sim \Sigma$ represents a realisation of the observational noise model. In the future, our aim is to explore the IMS as well by allowing the model to be imperfect; *Smith* (2002) discusses the subtleties of the PMS and IMS in more detail.

There is a rich literature on the shadowing of numerical solutions of differential equations by their true solutions, the so-called ϵ-shadowing problem (*Hammel et al.*, 1987; *Grebogi et al.*, 1990; *Sauer and Yorke*, 1991; *Sauer et al.*, 1997; *Lai et al.*, 1999, for example). Representing differential equations on a computer introduces round-off error into the numerical solution at each timestep, which tends to grow exponentially when the system is chaotic until the analytical and numerical solutions diverge (*Hammel et al.*, 1987). In the current context the PMS guarantees that this problem does not arise. By construction it is the numerical solution itself that is the system

---

[1]*Hammel et al.* (1988) is an exception, but would be very difficult in practice to extend to a high-dimensional system.

$\tilde{f}$, *not* the underlying equations. Hence the map from time t to $t + \delta t$ is the same for both model f and system $\tilde{f}$, *both* including the same numerics for a given $\mathbf{x}$, and so by construction we avoid the ϵ-shadowing problem entirely. Even when shadowing the laboratory experiment this phenomenon will not be relevant, because model imperfections will be the dominant source of error, leading to model drift (see, e.g., *Orrell et al.*, 2001, for an example in weather modelling).

2.2. Consistency in shadowing

The *candidate shadowing time* τ_S for a specific candidate trajectory is the time that trajectory remains consistent with observations, given a model of observational error. *Smith et al.* (2010) define the two as consistent if the distribution of the forecast residuals, $\mathbf{e}_t \sim \Phi$, and the distribution of noise realisations given by the noise model, $\tilde{\mathbf{e}} \sim \Sigma$, are equivalent. As long as $\Phi \equiv \Sigma$, the candidate trajectory is said to shadow the observations. For a particular model and observations, the *shadowing time* for that pair is the maximum among all possible candidates: $\tau_{S_t} = \max_{\mathbf{x}} \tau_S(\mathbf{x}, t)$. By measuring τ_S for a finite set of candidate trajectories one can only calculate a lower bound on τ_{S_t}.

In practice Φ is expected to differ from Σ and we aim to quantify how similar the two distributions are. The degree of similarity between Φ and Σ can be assessed by comparing equivalent percentiles of the two distributions. Suppose we have a single sample $x_1, \ldots, x_N$ from a distribution Φ and want to determine whether Φ is close to some known distribution Σ. We can proceed by computing the r-th order statistic $X_{(r)}$ of the sample from Φ and perform a statistical test to determine whether $X_{(r)}$ is close to the r-th order statistic $\tilde{X}_{(r)}$ from a sample $\tilde{x}_1, \ldots, \tilde{x}_N$ from Σ. This calculation requires the distribution $\tilde{F}_{(r)}$ of the order statistic $\tilde{X}_{(r)}$. The intuitive way to do this is to repeatedly sample N values $\tilde{x}_1, \ldots, \tilde{x}_N$, from Σ, calculate $\tilde{X}_{(r)}$ for each sample, and construct the empirical distribution $\tilde{F}_{(r)}$. We can then check whether $X_{(r)}$ falls within a particular confidence interval of $\tilde{X}_{(r)}$. If it does, we say that the order statistic $X_{(r)}$ is consistent with the distribution $\tilde{F}_{(r)}$ of the order statistics $\tilde{X}_{(r)}$ and hence the two distributions are sufficiently close.

For a more robust assessment more than one order statistic $X_{(r)}$ can be tested in this way, which we illustrate in Fig. 2. In that figure we draw 500 samples of size 1000 from a known noise distribution $\Sigma \sim N(0, 1)$, and calculate the 50th ($\tilde{X}_{(50\%)}$) and 90th ($\tilde{X}_{(90\%)}$) percentiles of each sample. The black bars in the figure show the distributions of these percentiles. We then produce a test sample of the same size but from a slightly different distribution $\Phi \sim N(0, 1.01)$, and calculate its order statistics $X_{(50\%)}$ and $X_{(90\%)}$, shown by circles in the lower figures. To retain the null hypothesis that the test sample is drawn from the same distribution as the noise model at significance level $p = 0.2$, the two circles must fall within the vertical lines enclosing the 80% confidence intervals of the percentile distributions. In this particular example they do not, so the two distributions are not sufficiently close.

Applying this procedure to a sample of residuals $\mathbf{e}_t$ and realizations of a noise model $\tilde{\mathbf{e}}_t$ determines whether the model trajectory is sufficiently close to shadow the target at a given significance level.

2.3. The Smith et al. (2010) method for measuring shadowing times

Smith et al. (2010) use the the Euclidean length of the residual vector $\|\mathbf{e}_t\|$ as the test statistic to measure the shadowing time in what we shall refer to as the "sequence" method. This is the distance in the N-dimensional model state space between the model trajectory and the target. They calculate this quantity at each time and form a distribution of distances over a trajectory time series, before calculating the median and 90th percentile of that distribution.

To obtain a sampling distribution of a given order statistic of the noise model, they draw a sample of noise realizations $\tilde{\mathbf{e}}$ of size N from Σ at each time. They calculate the Euclidean distance $\|\tilde{\mathbf{e}}\|$ for each time, giving a distribution of noise distances, and then locate the median and 90th percentile of *this* distribution. Repeating this sampling procedure M times yields a distribution of M medians and M 90th percentiles under the noise model. To determine whether the model trajectory shadows the target at this time, they test the following null hypothesis: for *both* order statistics (the 50th and 90th percentiles) the percentile of the residual distribution is drawn from the distribution of samples from the noise distribution. The comparison of medians tests the residual distribution for bias, and the comparison of the 90th percentiles tests it for dispersion. They use confidence intervals of 90% for the median and 99% for the 90th percentile. The procedure begins by using the trajectory segment with $t = 1$ only. If this segment shadows the target then they repeat with the trajectory segment increased in length by one, and so on, increasing the trajectory segment by one point at a time. The candidate shadowing time is determined once a trajectory segment rejects the null hypothesis.

2.3.1. The need for an alternative method

We see three problems with the *Smith et al.* (2010) definition that are particularly important when N becomes large.
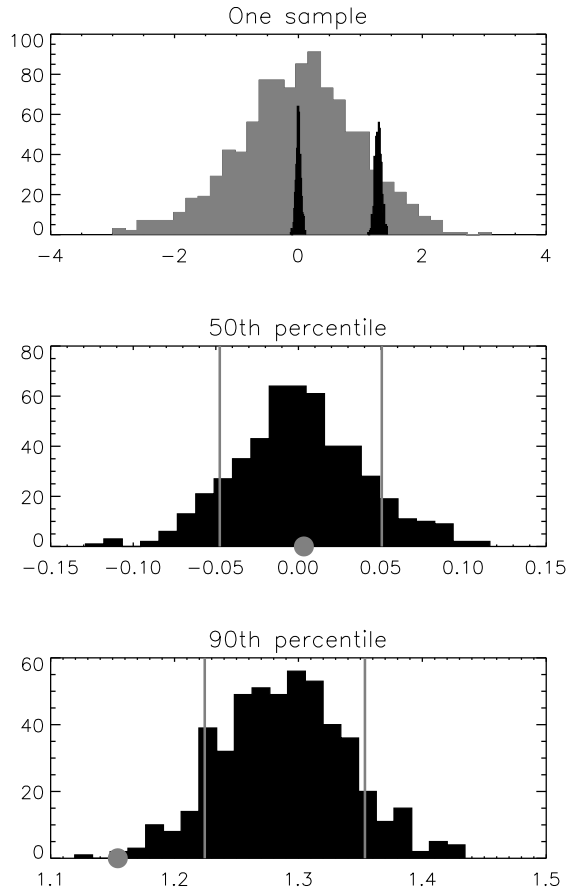
Figure 2: Schematic of the general order statistics approach to calculating a candidate's shadowing time. The top panel shows a histogram (grey) of the values generated from a sample of size 1000 from $N(0,1)$. The black histograms show, for 500 such samples, the distributions of the 50th and 90th percentiles of each sample. These percentile distributions are shown in the middle (50th) and lower (90th) panels. On the lower panels vertical lines are added at percentiles 10% and 90% in those distributions. The 50th and 90th percentiles of a single test distribution drawn from $N(0,1.01)$ are shown as grey circles.

First, *Smith et al.*'s (2010) sampling approach is computationally expensive. At each time one requires M samples of size N from the noise model to produce the distribution of percentiles. Of course M should be as large as possible to ensure the best correspondence to the true underlying distribution of each percentile. For systems of very low dimensionality this is feasible but in higher dimensions this sampling approach is no longer computationally practical.

Second, a single outlier at one point in the model state space in a very short trajectory may distort the distribution of the $\|\mathbf{e_t}\|$ and hence cause the model to fail to shadow for that trajectory length, while a longer trajectory including that point would reduce the effect of that outlier in the distribution and hence that longer time series might then shadow.

Finally, the length of the trajectory is important for another reason when using a distance method. If the trajectory is only length five, for example, then the distribution of distances is too small to obtain a sensible value of, say, the 90th percentile of that distribution. For a sensible shadowing time, therefore, the trajectory must be sufficiently long for statistical accuracy, but often the model is not capable of shadowing for the minimum of, say, $O(100)$ leadtimes required.

2.4. The "state" method for measuring shadowing times

To avoid some of these problems, we now present a slightly modified version of the method used to measure the shadowing time from the consistency between model trajectory and observations that is better suited to higher-dimensional systems. We also believe this method is more intuitive than using a distribution of distances from model to target along a trajectory. We call this the "state" method.

2.4.1. Avoiding sampling problems in high-dimensions

First we present a theoretical result that overcomes the sampling impracticality of the sequence method when applied to cases with $\gg O(10)$ dimensions, opening it up for use in the higher-dimensional context.

Following *David* (1981), assume independently and identically distributed (IID) samples of size N from a cumulative distribution function (CDF) $P(x)$, each sample providing N random variables $X_1, \ldots, X_N$. When the sample elements are sorted in ascending order, i.e. $X_{(1)} \leq X_{(2)} \ldots \leq X_{(N)}$, we call $X_{(r)}$ the rth order statistic. Denote by $F_{(r)}$ the CDF of the rth order statistic $X_{(r)}$. *David* (1981, pp. 8–9) showed that $F_{(r)}$ is

$$F_{(r)}(x) \quad = \quad Pr\{X_{(r)} \leq x\} \tag{1}$$

$$= \quad \sum_{i=r}^{N} \binom{N}{i} P^i(x) \left[1 - P(x)\right]^{N-i} \tag{2}$$

The relation between the binomial sum and the incomplete β-function (*NIST*, 2010, Eq. 8.17.5) means we can replace Eq. (2) with

$$F_{(r)}(x) = I_{P(x)}(r, N - r + 1) \tag{3}$$

Suppose we are interested in the 90th percentile of the distribution of the rth order statistic $X_{(r)}$ (i.e. $F_{(r)} = 0.90$), with X being sampled from $P(x)$. We require the x that corresponds to the 90th percentile of the distribution of $X_{(r)}$. Since $F_{(r)} = I_{P(x)}(r, N - r + 1)$, we may find the probability that $P(x)$ assigns the required value x by taking the inverse of the incomplete beta function, i.e.

$$P(x) = I_{F_{(r)}}^{-1}(r, n - r + 1) \tag{4}$$

Finally we obtain the value x which gives the required percentile of the distribution of $X_{(r)}$ using the inverse of $P(x)$: $x = P^{-1}(x)$.

This theoretical result provides an exact distribution of any order statistic $X_{(r)}$ at negligible computational cost. We have overcome the intractability of sampling in high dimensions, and so under the assumption of IID samples we can now perform formal hypothesis testing on a sample of any size.

### 2.4.2. Definition and properties of the "state" method

With this result in mind, we present a modified form of the *Smith et al.* (2010) sequence method used to calculate shadowing times. Our method overcomes the problems listed in Sect. 2.3.1 above while remaining computationally practical.

At each time t in a model trajectory we have a model state $\mathbf{x}_t$, a target state $\tilde{\mathbf{x}}_t$ (whose observational error is well-defined by the distribution Σ in the PMS), and a residual error vector $\mathbf{e}_t = \mathbf{x}_t - \tilde{\mathbf{x}}_t$, each with dimension N. Starting at $t = 0$, we test the null hypothesis at significance level p that the residual error vector $\mathbf{e}_t$ is a sample drawn from the observational error distribution Σ. To perform the test we measure the median and the 90th percentile of the residual error vector, and find the distributions of the equivalent percentiles of the noise distribution analytically using the order statistics method discussed above. To shadow the observations at time t we require both the median and 90th percentiles of the residual distribution to fall within the respective confidence intervals of the noise distribution at significance level p. If the model shadows at $t = 0$ then we repeat the procedure for $t = 1, 2, \ldots$, and the shadowing time for that model trajectory is the longest time for which the null hypothesis is retained.

This method differs from the "sequence" method in a number of ways. First, we avoid sampling the noise distribution by using the inverse β-function as described in the previous section[2]. This avoids problem one in Sect. 2.3.1. Second, our method applies to each point in the model trajectory in turn, rather than using the whole trajectory at once. Our method uses the distribution of values over the components of the state vector, while *Smith et al.* (2010) map the vector of components to a single number. This property means we avoid problem two in Sect. 2.3.1. Finally, we avoid problem three in Sect. 2.3.1 by requiring the state vector dimension N to be large. Therefore our method is not appropriate for use with low-dimensional systems, where N is not large enough to compute sensible percentile statistics for the residual error distribution. Take, for instance, the *Lorenz* (1963) system, with $N = 3$. Determining the 90th percentile in a sample of three values does not produce a sensible result. Our method is suitable for medium- and high-dimensional systems, with a dimensionality of at least $N = O(100)$ and preferably $N = O(\gg 100)$.

We should note that a caveat with our method is that by examining each trajectory point in turn we may not detect small but systematic outliers. If an outlier is small but present at each trajectory point the model should shadow at that point. If the whole trajectory is considered, however, as in *Smith et al.* (2010), the outliers at all times will accumulate and the failure of the model trajectory will be detected.

---

[2]Note that now this result has been obtained, sampling can be avoided for the "sequence" method as well.
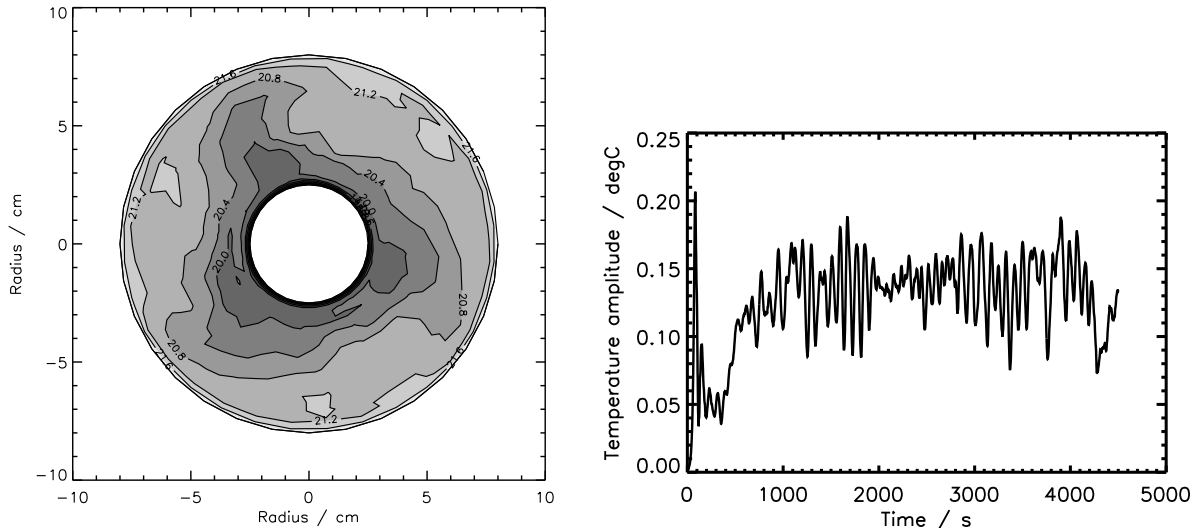
(a) Snapshot of an equilibrated wavenumber-3 temperature structure in a horizontal section at $z = 5.61$ cm after 3000 s.

(b) Time series of the wavenumber-3 temperature mode amplitude in the same simulation. The amplitude is found at each timestep from a Fourier transform of the temperature field at mid-radius / mid-height.

Figure 3: Temperature output from a rotating annulus simulation using the model described in the text with the setup in Fig. 1.

3. Experimental methods

To demonstrate our method for measuring shadowing times from the previous section and investigate some shadowing properties of our rotating annulus model we run some perturbation experiments; this section describes how they are set up.

3.1. The rotating annulus simulation

Figure 1 shows a schematic of the rotating annulus with the particular configuration used in this work, which is the "small annulus" setup used by *Hignett et al.* (1985). We work in the PMS, so ignore details of the laboratory setup that introduce complications due to imperfections in the apparatus, for example, and we take the numerical representation of Fig. 1 as a perfect representation of the system.

The mathematical model used to simulate the rotating annulus experiment is the Met Office / Oxford Rotating Annulus Laboratory Simulation (MORALS) (*Farnell and Plumb*, 1976; *Hignett et al.*, 1985; *Read et al.*, 2000). This model solves the Navier-Stokes, heat transfer, and continuity equations subject to the Boussinesq approximation, in cylindrical polar coordinates (see Appendix). There are four prognostic variables: three velocity directions **u** (radial), **v** (azimuthal), **w** (vertical), and temperature **T**. The fifth field is a kinetic pressure field $\Pi \equiv \mathbf{p}/\rho_0$, which is diagnostic (calculated using a Poisson equation from the other four fields); ρ_0 is the fluid density at a reference temperature. The fields are discretized on a staggered Arakawa C grid (*Arakawa and Lamb*, 1977), and are non-uniform in the radial and vertical directions to resolve the boundary layers. All velocities are set to zero at the boundaries, and the temperature gradient is zero across the top and bottom boundaries. The parametrisation of fluid properties used to set up a 17% glycerol / 83% water fluid (by volume) is described in *Hignett et al.* (1985, Table 1). The model runs with a timestep of $\delta t = 0.02$ s and a spatial resolution of $(N_R, N_\theta, N_z) = (16, 32, 16)$ nodes in the radial, azimuthal, and vertical directions respectively. The number of independent variables[3] is $N = 24192$.

Fig. 3 shows temperature output from an annulus simulation using this configuration, with $\Omega = 1$ rad s^{-1} and $\Delta T = 4$ degC. Figure 3a is a typical horizontal temperature field near mid-height, showing a wavenumber-3 structure to the flow, and Fig. 3b shows the amplitude of the wavenumber-3 azimuthal temperature mode as a function of time. The flow starts from rest in the rotating frame at $t = 0$, so there is some transient behaviour in the first ~ 1000 s.

To determine whether the flow regime is chaotic we estimated the largest Lyapunov exponent for this run. Chaos requires aperiodic behaviour in a deterministic system with at least one positive Lyapunov exponent (*Strogatz*, 1994, p. 323). To estimate the largest Lyapunov exponent we followed the procedure that *Young*

---

[3]From $16 \times 32 \times 16 \times 4$ minus the number of points on and outside the fluid boundary (some outside points are required by the boundary conditions when using a staggered grid).

7

and Read (2008) used on annulus simulations in various flow regimes (*Wolf et al.*, 1985; *Hegger et al.*, 1999), in this case working with the wavenumber-3 temperature time series. The procedure was the same as documented in *Young and Read* (2008), with 62 s the optimal delay time in this case, and we extended the time series in Fig. 3b to 100000 s to minimise noise in the calculation. We found the largest Lyapunov exponent to be 0.0061 ± 0.0025 bits s^{-1}, which corresponds to an error-doubling time of 1.5–3.0 oscillation periods (125–250 s). This exponent is comparable to values in the chaotic structural vacillation flow regime (0.0013–0.0055 bits s^{-1}), and is an order of magnitude larger than values in non-chaotic flow regimes (-0.00012–0.00063 bits s^{-1}) (*Young and Read*, 2008), so we might expect perturbations to grow at a similar rate to structurally vacillating flow. The model is deterministic, and the flow is aperiodic for the length of the time series, so the positive Lyapunov exponent demonstrates (for all practical purposes) that this flow is chaotic[4].

3.2. MORALS perfect model setting and observations

Let $\mathbf{x}[t]$ and $\tilde{\mathbf{x}}[t]$ be model and true system states at time t, and $f^{\Delta t}$ and $\tilde{f}^{\Delta t}$ be integration of states by the model and system for time Δt. So $\mathbf{x}[t + \Delta t] = f^{\Delta t}(\mathbf{x}[t])$ and $\tilde{\mathbf{x}}[t + \Delta t] = \tilde{f}^{\Delta t}(\tilde{\mathbf{x}}[t])$ describe the model and system exactly. Let $\mathbf{s}[t] = \tilde{\mathbf{x}}[t] + \tilde{\mathbf{e}}$ be observations of the system at time t, and $\Delta \mathbf{x}$ be perturbations to initial states. In the PMS we simply choose the system being modelled, $\tilde{f}$, to be MORALS with the same parameters as the model: $f \equiv \tilde{f}$.

We generate a true system trajectory and observations of that trajectory as follows. First we run MORALS for 1500 s to spin up the model from rest until a coherent flow structure is present (e.g. Fig. 3) and any transient behaviour has decayed. From this we run MORALS for an additional 1500 s to obtain an estimate, at each grid point, of the range of values admitted by the model when it is in dynamical equilibrium. This gives the natural variability of the system at each point in space. Let the vector $\mathbf{r}$ be this range, where each element contains the range of values that encloses 99% of the values over these 1500 s at that grid point:

$$\mathbf{r} = \tilde{\mathbf{x}}[t_0 - 1500s, ..., t_0]_{99.5\%} - \tilde{\mathbf{x}}[t_0 - 1500s, ..., t_0]_{0.5\%} \tag{5}$$

The final state in this run (i.e. after a total of 3000 s) will be the first state in the true system trajectory. To make things simpler later, we redefine this point as $t = 0$. We then obtain a long system trajectory by integrating this state using MORALS for time T to obtain $\tilde{\mathbf{x}}[t]$, $t \in [0, \ldots, T]$.

We then generate observations $\mathbf{s}[t]$ every 5 s along the system trajectory by adding a vector of random numbers $\tilde{\mathbf{e}}$ (noise) to the system trajectory. The random numbers are independently and identically distributed (IID) and are originally drawn from the Gaussian distribution $N(0, 1)$. These random numbers are converted to observational error by multiplying by a fixed fraction σ (= 0.1 throughout this paper) of the natural variability $\mathbf{r}$ at each grid point, i.e.

$$\tilde{\mathbf{e}} \sim \sigma \mathbf{r} N(0, 1) \tag{6}$$

The effect of the scaling by $\mathbf{r}$ ensures that, statistically, the amount of noise added at each point and to each field is the same fraction of the natural variability. For this reason, much of the analysis to follow will be done in terms of quantities expressed as a fraction of the natural variability, as this non-dimensional metric provides a convenient way to compare and combine the variables. We shall call this state space the "scaled space", as opposed to the (dimensional) "physical space". For two points in the physical state space, $\mathbf{x}_1$ and $\mathbf{x}_2$, their separation $\Delta \mathbf{x}'$ in the scaled space is

$$\Delta \mathbf{x}' \equiv (\mathbf{x}_2 - \mathbf{x}_1)' = (\mathbf{x}_2 - \mathbf{x}_1) \circ \mathbf{r}^{-1} \equiv \Delta \mathbf{x} \circ \mathbf{r}^{-1} \tag{7}$$

where $\circ$ denotes the Hadamard (pointwise) product.

3.3. Setup of the perturbation experiments

To investigate the shadowing properties of the rotating annulus model MORALS in the PMS, we calculate trajectories of the model starting from initial *candidate* states perturbed from the true state at $t = 0$. We use these *candidate trajectories* to demonstrate the "state" method developed in the previous section, by calculating how long each candidate shadows the observations in Sect. 3.2. By varying the size of the perturbation we shall determine how the distribution of candidate shadowing times depends on the original distance of the candidate from the true state.

---

[4]The reader should note that in the laboratory the flow under this external forcing is usually regular wavenumber-3 or 2, which is also the case when running MORALS at higher resolution (*Young and Read*, 2008, Fig. 3b). Because we consider only the PMS in this work, however, direct correspondence with the laboratory experiment is not so important here.

To generate a candidate state perturbed from the true state $\tilde{\mathbf{x}}[0]$ we add a random perturbation $\Delta\mathbf{x}$ as follows. First, we draw a random vector $\mathbf{n}$ of length N from the normal distribution $N(0, 1)$. Second, we scale the Euclidean length of the vector to be δ, i.e.

$$\mathbf{n}' = \delta\frac{\mathbf{n}}{\|\mathbf{n}\|} \tag{8}$$

Finally, to obtain a perturbation in physical space (i.e. the perturbation that is actually added to generate the candidate state), we multiply $\mathbf{n}'$ by the natural variability at each grid point, $\mathbf{r}$. Hence the perturbation $\Delta\mathbf{x}$ is

$$\Delta\mathbf{x} = \delta\frac{\mathbf{n}}{\|\mathbf{n}\|} \circ \mathbf{r} \tag{9}$$

By construction,

$$\|\Delta\mathbf{x} \circ \mathbf{r}^{-1}\| = \delta \tag{10}$$

so δ is the length of the perturbation in the scaled space. All the perturbations of length δ lie on a N-dimensional hypersphere of radius δ centred on the truth, and the Gaussian distribution guarantees that the perturbations are distributed uniformly in direction. By repeating this procedure any number of candidates may be generated.

This perturbation method is rarely used to efficiently sample initial condition uncertainty; usually methods such as breeding vectors (*Toth and Kalnay*, 1993), singular vectors (*Buizza and Palmer*, 1995), or even so-called "nightmare" vectors (*Smith et al.*, 1999) are used, which all aim to give shorter candidate shadowing times by sampling the fastest-growing directions. Here we just use a simple method to demonstrate our method for calculating shadowing times. We note that the distribution of candidate shadowing times would probably be different were we to use these other methods.

3.4. Measuring candidate shadowing times

For each candidate state used to begin a model trajectory we calculate the candidate shadowing time τ_S with respect to the observations using the "state" method described in the previous section. At each time t we use the model $f^t(\mathbf{x}[0])$ and observations $\mathbf{s}[t]$ to construct the residual error vector $\mathbf{e}[t] = f^t(\mathbf{x}[0]) - \mathbf{s}[t]$. The observational error $\tilde{\mathbf{e}}$ is well-defined by Eq. (6). We determine whether the distribution of scaled forecast residuals $\mathbf{e}[t] \circ \mathbf{r}^{-1}$ is consistent with the distribution of scaled observational error $\tilde{\mathbf{e}} \circ \mathbf{r}^{-1}$, testing the null hypothesis at significance level p for the 50th and 90th percentiles. If either test statistic lies outside the confidence interval of the noise distribution then the candidate trajectory does not shadow the observations at time t. The candidate shadowing time is the final t at which the test is passed, counting upwards from zero.

The shadowing time should capture the time the model diverges from the observations. Hence as a sanity check on our shadowing times we also use a simple measure of how long the model remains close to observations. The Euclidean norm $\|\mathbf{x}\| = (\mathbf{x}'\mathbf{x})^{1/2}$ of a vector of length N whose components are drawn from the Gaussian distribution $N(0, \sigma^2)$ follows a χ-distribution with N degrees of freedom and scale parameter σ (*Cramér*, 1946, p. 236). For $N \to \infty$ the asymptotic form of the χ-distribution is a normal distribution with mean $\sigma\sqrt{N - 1/2}$ and variance $\sigma^2/2$ (*Stuart and Ord*, 1994, p. 544–5). So if a vector of N observations $s_i, i = 1, \ldots, N$ is constructed by adding IID random numbers with distribution $N(0, \sigma^2)$ to a true state $x_i, i = 1, \ldots, N$, then the expected Euclidean distance from the truth to the observations in our scaled space is very close to $\sigma\sqrt{N}$ for the value of N we use.

The Euclidean distance in the scaled space between a model candidate trajectory $f^t(\mathbf{x}[0]))$ and the observations $\mathbf{s}(t)$ is

$$D(t) = \|\{f^t(\mathbf{x}[0]) - \mathbf{s}[t]\} \circ \mathbf{r}^{-1}\| \tag{11}$$

so we say that the candidate trajectory remains close to the observations as long as

$$D(t) \leq m\sigma\sqrt{N} \tag{12}$$

is satisfied, i.e. once the distance from model to observations becomes greater than $m\sigma\sqrt{N}$ then the candidate trajectory is no longer close. We call this time τ_D. Based on test data we use $m = 2$, as this represents a considerable deviation from the expected distance between observations and truth, but it turns out that τ_D is not very sensitive to m. Now, this is basically a simple RMS error metric, and while that is not an optimal error measure for nonlinear models (see, e.g., *McSharry and Smith*, 1999), the purpose of using this metric is to catch when the trajectory is *not* close to the observations, and the RMS error will be large when that is the case, so it is useful as a secondary measure here.
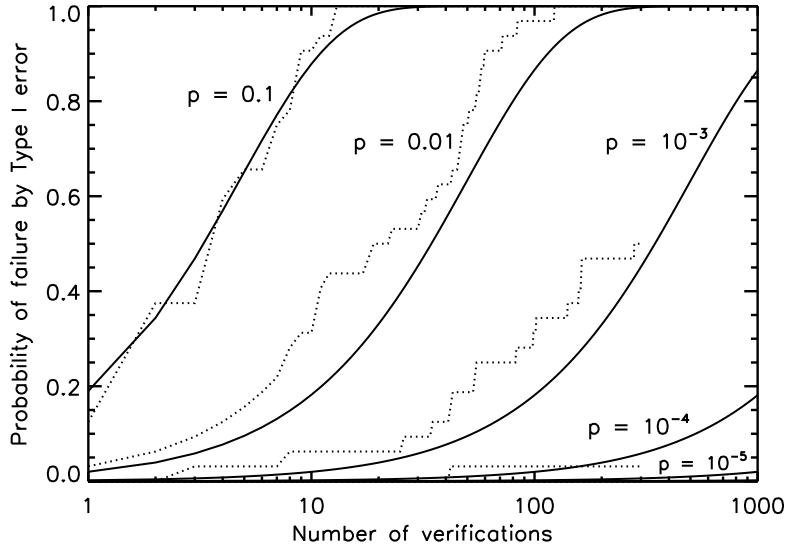
Figure 4: Consequences of the Type I error property of the "state" method for low significance levels. For each significance level (indicated by its p-value) the solid line shows the expected fraction of candidates (from Eq. 13) that fail to shadow because of a Type I error as a function of the number of verifications, and the dotted line shows the actual fraction that failed out of the 32 noise realisations. The dotted lines stop at 300 verifications, corresponding to maximum trajectory lengths of 1500 s.

4. Results

We are interested in how long a candidate trajectory might expect to shadow observations if the setup of an experiment is perfect and it starts at a given distance from the truth. This is an important benchmark to measure, because in practice with physical systems the setup is never perfect. A shadowing time measured in the imperfect model scenario or using laboratory data is of limited use if we have nothing to compare it against. By using the PMS we can obtain an upper limit to the shadowing properties of a particular system under this type of initial condition uncertainty. We expect a priori that, for perturbations in random directions, the time a trajectory remains close to observations should decrease as its initial distance from truth increases.

4.1. Candidate trajectories started from perfect initial conditions

First, however, we must calibrate the "state" method by selecting an appropriate significance level p. To do so we ran a "null" candidate trajectory started from perfect initial conditions, i.e. with $\delta = 0$. We measured the shadowing time against 32 different realisations of the observational noise model. By construction, the real shadowing time for this case is infinity, but an intrinsic property of our "state" method (and any other test based on repeated significance tests, including the *Smith et al.* (2010) "sequence" method, is that some fraction of the candidates will fail to shadow by a given time, even if the forecast is perfect, because the method is based on significance testing. This fraction depends critically on the significance level at which the null hypothesis is tested. So by calculating the parameters required to avoid this problem we can determine *a priori* the minimum significance level at which our "state" method should be used in practice.

The problem occurs because at each verification time along the trajectory we compute two test statistics, both of which must fall within prescribed confidence intervals in order to retain the null hypothesis at significance level p and hence shadow at that time. The probability of falsely rejecting the null hypothesis when it is true (Type I error) is therefore $1 - (1 - p)^2$. Within the first n verifications (in this case we check the shadow every 5 s along the trajectory), the probability of at least one Type I error is

$$r = 1 - (1 - p)^{2n} \tag{13}$$

which is the expected fraction of candidates that will fail to shadow within the first n verifications purely from Type I errors. For significance levels commonly used in statistical testing ($p = 0.1$, 0.01, say) this probability becomes large after $O(10)$ verifications. We must avoid these Type I errors because they erroneously assign a candidate shadowing time shorter than the true value.

10

In Fig. 4 we compare the fraction of candidates that failed to shadow because of a Type I error at each t against theoretical predictions from Eq. (13); the two correspond well. By $t = 50\,$s ($n = 10$), at significance level $p = 0.1$ almost all the candidates have rejected the null hypothesis and hence recorded a (false) shadowing time, as have almost all of the candidates by $t = 500\,$s ($n = 100$) for $p = 0.01$. The 25th percentile of the distribution for $p = 0.001$ also falls around this time, there is a sole outlier for $p = 0.0001$, and for $p \leq 10^{-5}$ none of the candidates suffer a Type I error.

This property is a fundamental limitation of any method that relies on repeated statistical significance tests, but we can largely avoid it in practice through our choice of p. We can find the largest value of p such that in the event that the true shadowing time is the equivalent of n verifications, fewer than R out of a total of E candidates will suffer a Type I error. This condition is $r < R/E$, and hence

$$p < 1 - \left(1 - \frac{R}{E}\right)^{1/2n} \tag{14}$$

is required, by rearranging Eq. (13). This problem with Type I errors is well-known in the statistical and medical literature (*Abdi*, 2007); the expression in Eq. (14) is an example of a *Šidák equation*.

4.2. Candidate trajectories started a fixed distance from truth

Now we demonstrate our method for identifying shadowing times by examining how long candidate trajectories started a fixed distance from truth shadow the observations.

We generated eight sets of initial candidate states using the method in Sect. 3.3, seven sets containing 64 candidates and one containing 256 (in order to have one with better statistics). While each set may be small compared with N, because the perturbation directions are random and initial conditions almost surely lie off the attractor, we would simply expect a smoothing of the distribution of candidate shadowing times if the number of candidates were increased further. Much longer or shorter shadowing times might occur in the special case of a perturbed initial condition falling on the attractor, but this occurs with probability zero. Within each set the distance from truth of each candidate was the same in the scaled space. We used $\delta = 10^{-4}$, 10^{-3}, 10^{-2}, $e.10^{-2}$, $e.10^{-1}$, 1, and e as the initial distances from truth in the scaled space for the 64-candidate sets, and $\delta = 0.1$ for the set with 256 candidates[5]. Using $\sigma = 0.1$ the expected distance between observations and truth is $0.1\sqrt{24192} = 15.6$ (Sect. 3.4). Hence the largest distance between candidate state and truth ($\delta = e$) corresponds to about 15–20% of the observational error, $\delta = 0.1$ corresponds to about 0.6%, and $\delta = 10^{-4}$ corresponds to about 0.0005% of the observational error. The candidate trajectories were integrated by MORALS for 1500 s of simulated time, recording the state every 5 s.

To calculate the appropriate p-values to use when measuring those candidates' shadowing times we used Eq. (14) with either $E = 64$ or $E = 256$ and $n = 301$ (maximum trajectory length of 1500 s, verifying every 5 s, plus a verification at $t = 0$). So for $R < 1$ (fewer than one candidate suffering a Type I error) we found that $p < 2.6 \times 10^{-5}$ when $E = 64$ and $p < 6.5 \times 10^{-6}$ when $E = 256$, so we used $p = 10^{-5}$ for the 64-candidate sets and $p = 10^{-6}$ for the 256-candidate set.

Figure 5a shows one grid point in a typical MORALS candidate trajectory and corresponding (temperature) observations in the PMS using the setup described in Sect. 3.1. The residual error between the trajectory and observations is shown in Fig. 5b, and the trajectory in Fig. 5a diverges from the observations at this particular grid point around 800 s, corresponding to the time the residual in Fig. 5b grows, on average, to be larger than the $\pm 2\sigma$ error bars.

We measured the distance between the trajectories and observations as a function of time:

$$D(t) = \|\{f^t(\mathbf{x}[0]) - \mathbf{s}[t]\} \circ \mathbf{r}^{-1}\| \tag{15}$$

Figure 6 shows this distance for each trajectory in the 256-candidate set (an initial distance from truth of size $\delta = 0.1$ in the scaled space). By construction, in the scaled space the expected initial distance between each candidate and the observations is $\sigma\sqrt{N} \sim 15.6$.

In this figure we see a rapid growth of the distance between candidate trajectories and observations from the expected initial observational error to the size of the attractor over a period of $\sim 300\,$s between $t = 500\,$s and $800\,$s. Intuitively, we require our method to show that a candidate shadows observations until the distance between candidate and observations diverges from from the expected initial distance $\sigma\sqrt{N}$.

Using the "state" method described in Sect. 3.4, we measured the candidate shadowing times for each candidate in the eight sets, using a range of eleven significance levels in each case from $p = 0.1$ up to the maximum

---

[5]Multiples of e were used to ensure approximately regular spacing of the perturbation sizes on a logarithmic scale.
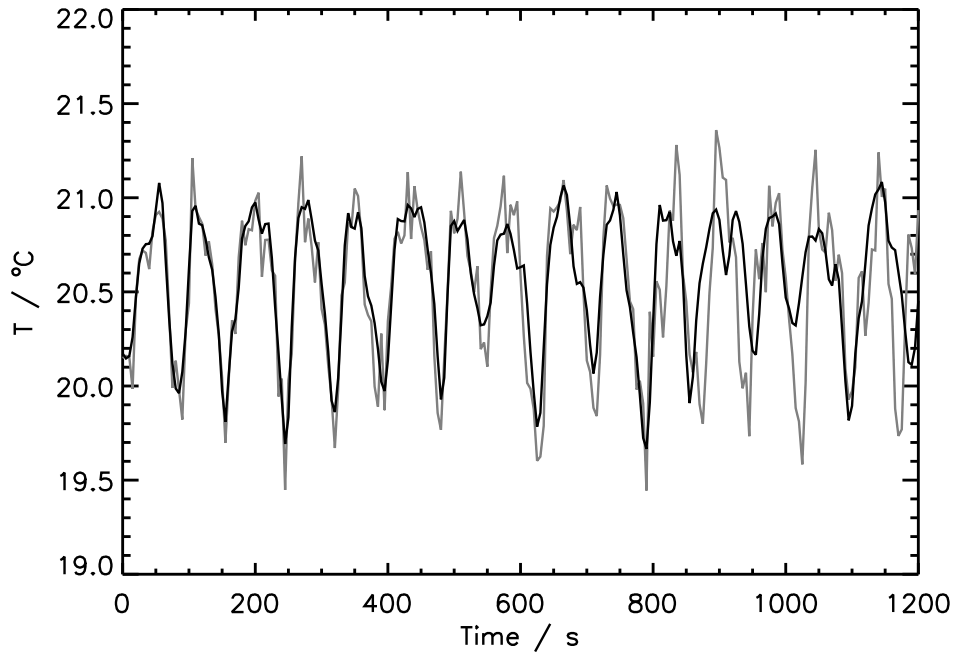
(a) Temperature model trajectory (black line) and corresponding observations (grey line).
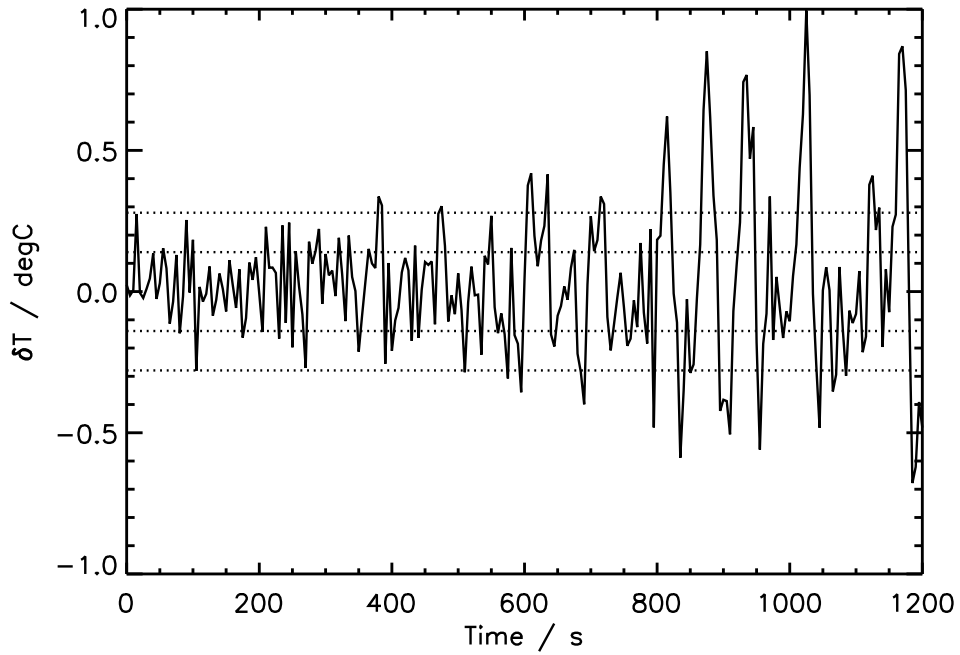
(b) Temperature residual error (black line) and corresponding $\pm\sigma$ and $\pm2\sigma$ observational error (dotted lines).

Figure 5: Example annulus candidate trajectory showing the temperature time series and residual error at a single grid point: $R = 4.81\,\mathrm{cm}$, $\theta = 6.19\,\mathrm{rad}$, $z = 5.61\,\mathrm{cm}$. This particular candidate trajectory begins a distance $\delta = 0.1$ from the truth in the scaled space (Eq. 10).
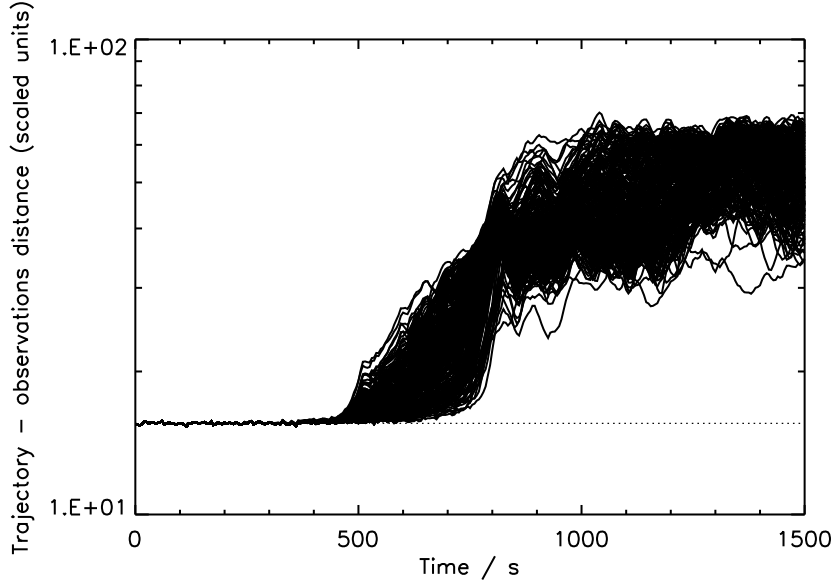
Figure 6: Distance between 256 candidate trajectories and one realisation of the observational noise (using Eq. 15) for candidates begun a distance $\delta = 0.1$ from truth in the scaled state space introduced in Eq. (7). The vertical scale shows the Euclidean distance between candidate trajectory and observations in the scaled state space, and is logarithmic. The dotted line is at $D = \sigma \sqrt{N}$. This distribution of error growth over the candidates is robust with respect to different realisations of the observational noise (four were tried in total, with similar results).

significance permitted by floating point precision ($p = 10\epsilon \approx 4.4 \times 10^{-15}$). Figure 7 shows the distribution of candidate shadowing times for the trajectories in the 256-candidate set ($\delta = 0.1$) for each of these significance levels.

Although above we calculated the "appropriate" significance level to use in this context, it is important to determine how sensitively the results depend on this quantity. The figure shows that our candidate shadowing times are generally not very sensitive to the significance level. Once the candidate trajectory diverges from the observations, it does so rapidly, reflecting the trajectory's rapid divergence from the observations as shown in Fig. 6. As the residual error distribution diverges from the observational error distribution, the test rejects the null hypothesis even at the highest significance levels, and the robustness of our method in this respect is very encouraging. Below $p = 0.001$ the candidate shadowing times are significantly smaller, which is expected from the analysis of the "null" trajectories above.

At $p = 10^{-6}$, the "appropriate" significance level for this candidate set, the candidate shadowing times are in the range 450–700 s. This concurs with a visual analysis of the trajectories in Fig. 5 and the visual estimate of about 500 s from Fig. 6.

To determine whether the shadowing times are sensitive to position on the model attractor, we compared the candidate shadowing time distributions against two candidate trajectory sets started from other points in the state space. To start candidate trajectories from other points in the state space, the spinup time was changed from 1500 s to 1525 s and 1550 s. At this point in $(\Omega, \Delta T)$ parameter space the timescale of the fluid's major oscillation (which determines the extent of the model attractor) is about 75 s (Fig. 5a), so trajectories started from these other two spinup times are well-spaced about the model attractor. At each position we ran two 64-candidate sets with $\delta = 10^{-3}$ and $e.10^{-1}$. The distributions of candidate shadowing times for $p = 10^{-6}$ were very similar, with extrema and medians varying by only 5-10%. Hence the distribution of candidate shadowing times is not generally sensitive to position in state space, at least for the points tested.

Now we have established that our method measures reasonable candidate shadowing times, we consider what the expected candidate shadowing time should be given its initial distance from truth. Conversely, how close to the truth must a candidate trajectory begin at to be expected to shadow for a particular time? With our perturbation experiments we can approach this question in a controlled fashion.

Using $p = 10^{-5}$ for the 64-candidate sets, and $p = 10^{-6}$ for the 256-candidate set, we plot in Fig. 8 the distribution of candidate shadowing times for the eight different initial distances from truth δ. There is clearly a logarithmic relationship between these candidate shadowing times and the initial distance from truth. An approx-
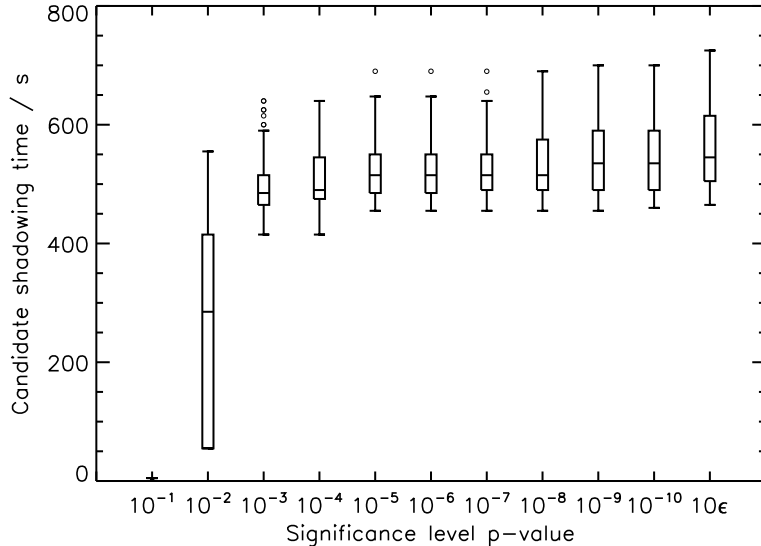
13

Figure 7: Distributions of candidate shadowing times measured using the "state" method for the 256 candidate trajectories in Fig. 6 with $\delta = 0.1$. Distributions for eleven different significance levels (p-values) are shown, with the p-value on the x-axis. Each box-whisker plot shows the distribution of candidate shadowing times at one significance level. The box represents the interquartile range, the middle line is the median, and the two whiskers represent the maximum/minimum or the 75%/25% values +/- 1.5 times the interquartile range, whichever is closest to the 75%/25% value. Outliers are shown with circles.

imate empirical best fit line to the data is

$$\tau_S \approx 250 \left(1 - \log_{10} \delta\right) \tag{16}$$

where τ_S is the candidate shadowing time, measured in seconds, and δ is the initial distance from truth (Eq. 10). Conversely, we predict that a candidate will have an expected shadowing time of τ_S when the initial distance from truth in the scaled space is approximately

$$\delta \approx 10^{1 - \tau_S / 250} \tag{17}$$

To test this prediction, we ran a final set of candidate trajectories which we expect to shadow for around $\tau_S = 2000 \, \text{s}$. From Eq. (17) this corresponds to $\delta \approx 10^{-7}$. We started the final 64-candidate set from the same point in state space as the previous eight sets, with an initial distance from truth of $\delta = 10^{-7}$, and found the distribution of candidate shadowing times to be close to the predicted 2000 s:

Percentile (%)	0	25	50	75	100
τ_S (s)	1925	2000	2090	2130	2405

Note that the maximum shadowing time τ_{S_t} over all possible candidates is not what is being measured here. Candidates beginning on the attractor are special cases that the dynamics allows for but will almost never occur with our perturbation method. We loosely use the term "on the attractor" here to include points on the attractor and on the relevant stable manifold. In these cases a candidate further away from the truth may have a longer shadowing time, but our perturbation method generates an initial condition on the attractor with probability zero. Hence we can only conclude that the expected candidate shadowing time for a trajectory beginning a given distance from truth varies logarithmically with that distance; the maximum shadowing time may be much longer.

Our aim was to demonstrate that the method behaves sensibly, and it does so: the statistical distribution of shadowing times decreases as initial distance from the truth increases. The functional form of this relationship is close to logarithmic, however, which is perhaps a little surprising. *Smith et al.* (1999) showed that, even for low-dimensional systems like the Lorenz equations, the error q-pling time τ_q behaves such that the error-quadrupling time τ_4 is generally much less than twice the error-doubling time τ_2, indicating explosive and episodic uncertainty growth (their Fig. 8). By contrast, exponential divergence of two trajectories implies $\tau_{q^2} \approx 2\tau_q$. So one would not necessarily expect an logarithmic dependence of the shadowing time on initial distance from truth, as that would be equivalent to $\tau_{q^2} \approx 2\tau_q$. This is what we have observed, however, in Fig. 8. Our initial perturbations are possibly too large for this phenomenon to be relevant; *Smith et al.* (1999) use infinitesimal perturbations in their analysis. Furthermore, they state that $\tau_{q^2} \approx 2\tau_q$ may hold once q is large enough. $q = 32$ is too small for the systems they use, but our experiments represent q as high as ≈ 400, which may be sufficient.
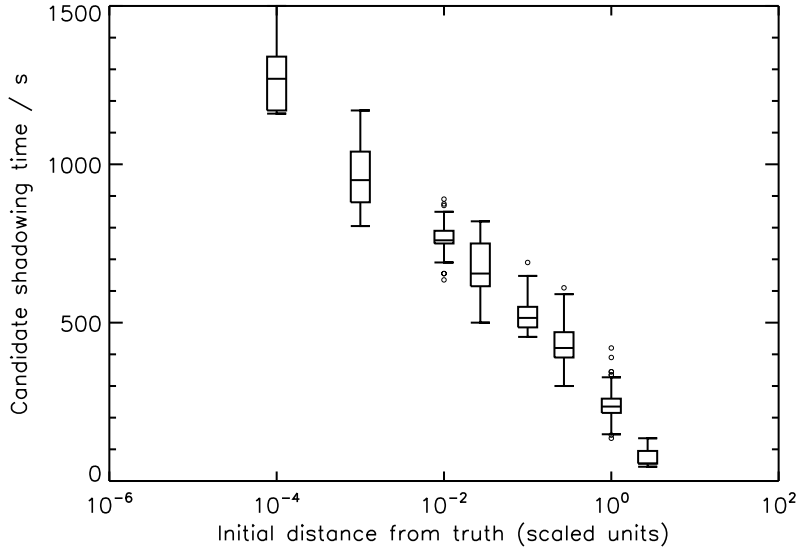
14

Figure 8: Distributions of candidate shadowing times for eight initial distances from truth δ, using significance level $p = 10^{-5}$ for all but $\delta = 0.1$, which uses $p = 10^{-6}$. The boxes are constructed in the same way as in Fig. 7.

4.3. Sanity check using the simple distance-based metric

In Sect. 3.4 we introduced a simple distance-based metric to determine the time τ_D that a candidate trajectory remains within $m\sigma\sqrt{N}$ of the observations in the scaled state space (Eqs 11 and 12). This estimates the time at which the candidate trajectory diverges from the observations.

The trajectory and residual in Fig. 5 show an example at a single model grid point, where the model trajectory diverges from the observations around $t = 800$ s, but the definition of τ_D should identify the time at which this occurs for the *whole* state vector. In Fig. 9 we show a visualisation of the same candidate trajectory (of which Fig. 5 was a single grid point) that extends the visualisation of residual error at one grid point to the whole state vector. There is an abrupt transition from small distances (dark regions) to large distances (light regions) over the whole state vector, quite similar over most of the grid points, which is the time at which the candidate trajectory diverges from the observations. Below we list several percentiles of the probability distribution function for τ_D measured for each trajectory in the 256-candidate set.

Percentile (%)	0	25	50	75	100
Shadowing time (s)	675	735	760	775	980

In Fig. 9, τ_D is identified by a solid horizontal line. There is general agreement between τ_D and the time at which the absolute distance between candidate trajectory and observations increases rapidly, i.e. from dark to light in the figure as we move upwards. τ_D corresponds well to this transition for all candidates for which such diagrams were produced, being within ± 50 s in each case. Hence τ_D estimates well the time at which a candidate trajectory diverges from observations. Furthermore, the time between diverging from the expected distance between observations and truth and reaching the saturation distance (distance between two points randomly chosen on the attractor) is relatively short compared with the length of a trajectory (Fig. 6). So as long as m is not so large as to make $m\sigma\sqrt{N}$ larger than the saturation distance, then τ_D will be quite insensitive to m.

The dotted lines show τ_S measured in the previous section for this particular candidate for each of the eleven p-values. In this case, as in the others, $\tau_S < \tau_D$, which demonstrates that the distribution of residual error loses consistency with the observational noise distribution some time before the trajectory and observations diverge visually. Overall, however, our simple distance metric confirms that our method for measuring shadowing times is a reasonable measurement of the time model trajectories diverge from observations.

5. Discussion and conclusions

We have presented a study of shadowing in the rotating annulus using the perfect model scenario. First we developed a version of the *Smith et al.* (2010) "sequence" method of calculating the shadowing time that is more suitable for use with high-dimensional systems. In the course of this analysis we found a way to avoid the
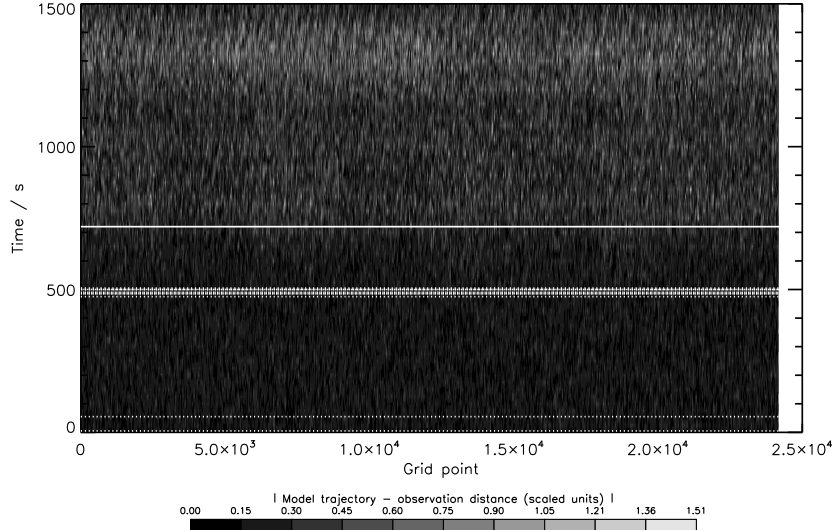
Figure 9: Contour plot of the unsigned distance $|\mathbf{e}[t] \circ \mathbf{r}^{-1}|$ between candidate trajectory and observations for each grid point in a candidate with $\delta = 0.1$ (same run as Fig. 5), for each point in the trajectory. A time series for a single grid point is a single vertical line in this diagram, and for a particular t all the model grid points are in a horizontal line (the order of these grid points along the x-axis is unimportant). The horizontal solid line is at $t = \tau_D$ for this candidate (using $m = 2$), and the dotted lines show τ_S for the 11 different p-values in Fig. 7. Note that to enhance the contrast, the contour intervals are irregularly spaced.

practical sampling problems of applying the sequence method in higher dimensions. Our "state" method measures the consistency of the residual error distribution and the observational error distribution, and requires consistency at each time up to t in order to shadow for a time t.

We demonstrated our state method using rotating annulus simulations in a series of perturbation experiments. Generating sets of candidate states a fixed distance from the true state, we measured the candidate shadowing times with respect to a set of artificial observations. Because our perturbation method ensured that, with probability one, initial conditions would not lie on the model attractor or stable manifold, we expected (almost surely) the trajectories to diverge from the observations, and they did so. The shadowing times decreased as the trajectories' initial distances from truth increased, with a clear logarithmic relationship between the initial distance from truth and the distribution of candidate shadowing times. We used the relationship to predict how close to the truth the candidate would have to be for its expected shadowing time to be 2000 s, and the results agreed well with the prediction. The candidate shadowing times (Fig. 7) also corresponded well to the time trajectories remained close to observations measured using a simple distance metric (Fig. 9), concurred with a visual analysis of the forecasts (Fig. 5), and appeared to be similarly distributed at other points in the model state space, all of which strengthen our conclusion that the method we have presented to measure shadowing times is reasonable.

By using the PMS we were able to measure shadowing times against the true state of the system using a simple setup. In particular, deliberately simplifying the problem in this way avoided complications introduced by the ϵ-shadowing properties of the model, which are currently unknown. We expect the model's equations to be non-hyperbolic and so long ϵ-shadowing trajectories probably don't exist, for various reasons (*Lai et al.*, 1999). The larger the number of dimensions the more complications this is expected to introduce: for example, see *Hayes* (2003) for an analysis in $O(100)$ dimensions, which is still a factor of $O(100)$ smaller than MORALS. To investigate these properties further, one could determine whether phenomena such as glitching (*Sauer and Yorke*, 1991; *Dawson et al.*, 1994) or unstable-dimension variability (*Kostelich et al.*, 1997) might be important. *Dawson et al.* (1994) showed that if a finite-time local Lyapunov exponent fluctuates about zero then ϵ-shadowing will fail quickly, and so quantifying this using the hyperbolicity exponent (*Sauer*, 2002), which is possible in principle although difficult in practice, would provide some insight. Another simple experiment would be to compare a "clean" model trajectory with one where noise on the order of the roundoff error is introduced at each timestep, which would give a lower bound on the ϵ-shadowing times. This approach has precedent in *Sauer et al.* (1997); see also *Smith and Spiegel* (1987) for an example using numerical representations of 1–1 chaotic maps without round-off error. In any case, we expect the model's intrinsic structural error to be larger than these other sources of error, so in practice shadowability of the real experiment will almost certainly be limited by model error rather than anything else.

16

These results define a benchmark for further study of shadowing in this system. From Eq. (17) we have a quantitative measure of whether a candidate shadowing time of, say, 400 s is "good" or "bad", because it tells us how close to the true state we should expect a randomly orientated perturbation to be in order to expect that candidate shadowing time. In addition, in the PMS the candidate shadowing time tells us approximately how close to the truth our candidate was initially, assuming random perturbations, and hence how much we might expect to be able to improve it. Because of our perturbation method our candidate shadowing times will be lower bounds in the PMS — a method that generates candidates close to or on the attractor (such as gradient descent) should yield longer candidate shadowing times. With an imperfect model or observational data, however, we would expect these shadowing times to be an upper bound given that then the model and system attractors will then be disjoint or almost disjoint.

It should be stressed that these experiments were performed with one particular set of model parameters. The particular candidate shadowing times obtained would therefore be expected to change for different points in the parameter space. In a non-chaotic flow regime, for example, it is conceivable that a candidate trajectory might shadow observations for all t. Nevertheless, in other chaotic regions of parameter space we predict that the candidate shadowing times should vary in a similar way as a function of initial distance from the truth.

The work we have presented in this paper falls into a larger body of work examining the shadowing properties of the rotating annulus system by applying the method of gradient descent (*Judd*, 2003). This method searches for shadowing trajectories by relaxing a sequence of observational states onto the model attractor. Gradient descent has been used in low- (*Judd*, 2003, and elsewhere), intermediate- (*Judd et al.*, 2004) and high- (*Judd et al.*, 2008) dimensional models for the purposes of state estimation. We intend to combine our "state" method for measuring shadowing times with an algorithm for gradient descent and use this to determine whether gradient descent can produce candidates with long shadowing times in the rotating annulus system. In the perfect model scenario we will know how far away a state obtained via gradient descent is from the truth, and hence from Eq. (16) we will have an *a priori* estimate of how long we might expect a candidate to shadow observations. This work will be a useful step towards determining whether the gradient descent method can produce model trajectories with long shadowing times in high-dimensional models of weather and climate.

Acknowledgements

Appendix A. MORALS equations

MORALS solves the fluid equations in cylindrical polars (R, ϕ, z) for a 3D velocity field $\mathbf{u} = (u, v, w)$ and temperature T in a reference frame rotating at angular velocity Ω using a finite-difference integration scheme with leapfrog time-stepping. The pressure Π is found diagnostically using a Poisson equation (*Farnell and Plumb*, 1976, Sect. 5.1). T is relative to a reference temperature $T_R = 22\,°C$, and Π is relative to a reference pressure $\Pi_0(R, z) = \frac{1}{2}\Omega^2 R^2 + g(d - z)$. The equations solved are the continuity, heat, and Navier-Stokes equations:

$$\nabla \cdot \mathbf{u} = 0 \tag{A.1}$$

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \nabla \cdot (\kappa \nabla T) \tag{A.2}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + 2\mathbf{\Omega} \times \mathbf{u} - (\Omega^2 R \hat{\mathbf{R}} - g\hat{\mathbf{z}})\frac{\Delta\rho}{\rho_0} + \nabla\Pi = \mathbf{F} \tag{A.3}$$

Here $\mathbf{F}$ is the viscous term, but is more complicated than the standard $\nu\nabla^2\mathbf{u}$ because it allows spatial variations of the viscosity as well; see *Farnell and Plumb* (1976, Eqs 2.1–2.3, 2.6) for the full expression. These equations are closed by an equation of state for density $\rho(T) = \rho_0(1 + \rho_1 T + \rho_2 T^2)$ with $\Delta\rho = \rho - \rho_0$ and two "constitutive" relations for viscosity $\nu(T) = \nu_0(1 + \nu_1 T + \nu_2 T^2)$ and thermal diffusivity $\kappa(T) = \kappa_0(1 + \kappa_1 T)$. On all boundaries $\mathbf{u} = \mathbf{0}$ and $\hat{\mathbf{n}} \cdot \nabla\Pi = 0$, on the top and bottom boundaries $\hat{\mathbf{n}} \cdot \nabla T = 0$, and on the inner and outer cylinders $T = T_a - T_R$ and $T = T_b - T_R$ respectively.

ABDI, H. (2007), The Bonferonni and Šidák corrections for multiple comparisons, in *Encyclopedia of Measurement and Statistics*, edited by N. J. Salkind, pp. 103–107, Sage.

ARAKAWA, A., AND V. R. LAMB (1977), Computational Design of the Basic Dynamical Processes of the UCLA General Circulation Model, *Meth. Comput. Phys.*, 17:173–265, doi:10.1016/B978-0-12-460817-7.50009-4.

BOWEN, R. (1975), omega-Limit Sets for Axiom A Diffeomorphisms, *J. Differ. Equations*, 18:333–339, doi:10.1016/0022-0396(75)90065-0.

BUIZZA, R., AND T. N. PALMER (1995), The singular-vector structure of the atmospheric global circulation, *J. Atmos. Sci.*, 52:1434–1456, doi:10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2.

CRAMÉR, H. (1946), *Mathematical Methods of Statistics*, 11th print ed., Princeton University Press.

DAVID, H. (1981), *Order statistics*, second ed., Wiley.

DAWSON, S., C. GREBOGI, T. SAUER, AND J. A. YORKE (1994), Obstructions to shadowing when a Lyapunov exponent fluctuates about zero, *Phys. Rev. Lett.*, 73:1927–1930, doi:10.1103/PhysRevLett.73.1927.

FARNELL, L., AND R. PLUMB (1976), Numerical integration of flow in a rotating annulus II: three dimensional model, *Tech. rep.*, Occasional Note Met O 21 76/1, Geophysical Fluid Dynamics Laboratory, Meteorological Office, Bracknell, Berkshire.

GILMOUR, I. (1998), 'ɩ-shadowing, probabilistic prediction and weather forecasting', Ph.D. thesis, University of Oxford.

GREBOGI, C., S. M. HAMMEL, J. A. YORKE, AND T. SAUER (1990), Shadowing of physical trajectories in chaotic dynamics: Containment and refinement, *Phys. Rev. Lett.*, 65:1527–1530, doi:10.1103/PhysRevLett.65.1527.

HAMMEL, S. M., J. A. YORKE, AND C. GREBOGI (1987), Do numerical orbits of chaotic dynamical processes represent true orbits?, *J. Complexity*, 3:136–145, doi:10.1016/0885-064X(87)90024-0.

HAMMEL, S. M., J. A. YORKE, AND C. GREBOGI (1988), Numerical orbits of chaotic processes represent true orbits, *B. Am. Math. Soc.*, 19:465–469, doi:10.1090/S0273-0979-1988-15701-1.

HAYES, W. B. (2003), Shadowing-based reliability decay in softened n-body simulations, *Astrophys. J.*, 587:L59–L62, doi:10.1086/375263.

HEGGER, R., H. KANTZ, AND T. SCHREIBER (1999), Practical implementation of nonlinear time series methods: The TISEAN package., *Chaos*, 9:413–435, doi:10.1063/1.166424.

HIDE, R. (1953), Some experiments on thermal convection in a rotating liquid, *Q. J. Roy. Meteor. Soc.*, 79:161, doi:10.1002/qj.49707933916.

HIDE, R., AND P. MASON (1975), Sloping convection in a rotating fluid, *Adv. Phys.*, 24:47–100, doi:10.1080/00018737500101371.

HIGNETT, P., A. A. WHITE, R. D. CARTER, W. D. N. JACKSON, AND R. M. SMALL (1985), A comparison of laboratory measurements and numerical simulations of baroclinic wave flows in a rotating cylindrical annulus, *Q. J. Roy. Meteor. Soc.*, 111:131–154, doi:10.1002/qj.49711146705.

IKEDA, K. (1979), Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system, *Opt. Commun.*, 30:257–261, doi:10.1016/0030-4018(79)90090-7.

JUDD, K. (2003), Nonlinear state estimation, indistinguishable states, and the extended Kalman filter, *Physica D*, 183:273–281, doi:10.1016/S0167-2789(03)00180-5.

JUDD, K., L. SMITH, AND A. WEISHEIMER (2004), Gradient free descent: Shadowing, and state estimation using limited derivative information, *Physica D*, 190:153–166, doi:10.1016/j.physd.2003.10.011.

JUDD, K., C. A. REYNOLDS, T. E. ROSMOND, AND L. A. SMITH (2008), The Geometry of Model Error, *J. Atmos. Sci.*, 65:1749–1772, doi:10.1175/2007JAS2327.1.

KOSTELICH, E. J., I. KAN, C. GREBOGI, E. OTT, AND J. A. YORKE (1997), Unstable dimension variability: A source of nonhyperbolicity in chaotic systems, *Physica D*, 109:81–90, doi:10.1016/S0167-2789(97)00161-9.

LAI, Y.-C., C. GREBOGI, AND J. KURTHS (1999), Modeling of deterministic chaotic systems, *Phys. Rev. E*, 59:2907–2910, doi:10.1103/PhysRevE.59.2907.

LORENZ, E. N. (1963), Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20:130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

LORENZ, E. N. (1967), Laboratory Models of the Atmosphere, in *The Nature and Theory of the General Circulation of the Atmosphere*, pp. 114–126, World Meteorological Organization.

MCSHARRY, P. E., AND L. A. SMITH (1999), Better nonlinear models from noisy data: attractors with maximum likelihood, *Phys. Rev. Lett.*, 83:4285–4288, doi:10.1103/PhysRevLett.83.4285.

NIST (2010), NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/.

ORRELL, D., L. SMITH, J. BARKMEIJER, AND T. N. PALMER (2001), Model error in weather forecasting, *Nonlinear Proc. Geoph.*, 8:357–371, doi:10.5194/npg-8-357-2001.

READ, P. L., M. J. BELL, D. W. JOHNSON, AND R. M. SMALL (1992), Quasi-periodic and chaotic flow regimes in a thermally driven, rotating fluid annulus, *J. Fluid Mech.*, 238:599–632, doi:10.1017/S0022112092001836.

READ, P. L., N. P. J. THOMAS, AND S. H. RISCH (2000), An evaluation of Eulerian and semi-Lagrangian advection schemes in simulations of rotating, stratified flows in the laboratory. Part I: Axisymmetric flow., *Mon. Weather Rev.*, 128:2835–2852, doi:10.1175/1520-0493(2000)128<2835:AEOEAS>2.0.CO;2.

SAUER, T., AND J. A. YORKE (1991), Rigorous verification of trajectories for the computer simulation of dynamical systems, *Nonlinearity*, 4:961–979.

SAUER, T., C. GREBOGI, AND J. A. YORKE (1997), How Long Do Numerical Chaotic Solutions Remain Valid?, *Phys. Rev. Lett.*, 79:59–62, doi:10.1103/PhysRevLett.79.59.

SAUER, T. D. (2002), Shadowing breakdown and large errors in dynamical simulations of physical systems, *Phys. Rev. E*, 65:036,220, doi:10.1103/PhysRevE.65.036220.

SMITH, L. A. (2000), Disentangling uncertainty and error: On the predictability of nonlinear systems, in *Nonlinear Dyanmics and Statistics*, edited by A. Mees, pp. 31–64, Birkhäuser Boston.

SMITH, L. A. (2002), What might we learn from climate forecasts?, *P. Natl Acad. Sci. USA*, 99:2487–2492, doi:10.1073/pnas.012580599.

SMITH, L. A., AND E. A. SPIEGEL (1987), Strange Accumulators, *Ann. NY Acad. Sci.*, 497:61–65, doi:10.1111/j.1749-6632.1987.tb48711.x.

SMITH, L. A., C. ZIEHMANN, AND K. FRAEDRICH (1999), Uncertainty dynamics and predictability in chaotic systems, *Q. J. Roy. Meteor. Soc.*, 125:2855–2886, doi:10.1002/qj.49712556005.

SMITH, L. A., M. C. CUÉLLAR, H. DU, AND K. JUDD (2010), Exploiting Dynamical Coherence: A geometric approach to parameter estimation in nonlinear models, *Phys. Lett. A*, 374:2618–2623, doi:10.1016/j.physleta.2010.04.032.

STEMLER, T., AND K. JUDD (2009), A guide to using shadowing filters for forecasting and state estimation, *Physica D*, 238:1260–1273, doi:10.1016/j.physd.2009.04.008.

STROGATZ, S. H. (1994), *Nonlinear Dynamics and Chaos*, Studies in Nonlinearity, 1st ed., Perseus Books Publishing LLC.

STUART, A., AND J. K. ORD (1994), *Kendall's advanced theory of statistics. Volume 1: Distribution Theory*, 6th ed., Edward Arnold.

TOTH, Z., AND E. KALNAY (1993), Ensemble Forecasting at NMC: The Generation of Perturbations, *B. Am. Meteorol. Soc.*, 74:2317–2330, doi:10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.

WINSTON, J. S., AND A. F. KRUEGER (1961), Some Aspects of the Cycle of Available Potential Energy, *Mon. Weather Rev.*, 89:307–318, doi: 10.1175/1520-0493-89.9.307.

WOLF, A., J. B. SWIFT, H. L. SWINNEY, AND J. A. VASTANO (1985), Determining Lyapunov exponents from a time series, *Physica D*, 16:285–317, doi:10.1016/0167-2789(85)90011-9.

YOUNG, R. M. B., AND P. L. READ (2008), Flow transitions resembling bifurcations of the logistic map in simulations of the baroclinic rotating annulus, *Physica D*, 237:2251–2262, doi:10.1016/j.physd.2008.02.014.