A FORMAL METHOD FOR IDENTIFYING DISTINCT STATES OF VARIABILITY IN TIME-VARYING SOURCES: SGR A* AS AN EXAMPLE

L. MEYER[1], G. WITZEL[1], F. A. LONGSTAFF[2], AND A. M. GHEZ[1]
[1] Department of Physics and Astronomy, University of California, Los Angeles, CA 90095-1547, USA
[2] UCLA Anderson School of Management, University of California, Los Angeles, CA 90095-1481, USA

ABSTRACT

Continuously time variable sources are often characterized by their power spectral density and flux distribution. These quantities can undergo dramatic changes over time if the underlying physical processes change. However, some changes can be subtle and not distinguishable using standard statistical approaches. Here, we report a methodology that aims to identify distinct but similar states of time variability. We apply this method to the Galactic supermassive black hole, where 2.2 μm flux is observed from a source associated with Sgr A* and where two distinct states have recently been suggested. Our approach is taken from mathematical finance and works with conditional flux density distributions that depend on the previous flux value. The discrete, unobserved (hidden) state variable is modeled as a stochastic process and the transition probabilities are inferred from the flux density time series. Using the most comprehensive data set to date, in which all Keck and a majority of the publicly available Very Large Telescope data have been merged, we show that Sgr A* is sufficiently described by a single intrinsic state. However, the observed flux densities exhibit two states: noise dominated and source dominated. Our methodology reported here will prove extremely useful to assess the effects of the putative gas cloud G2 that is on its way toward the black hole and might create a new state of variability.

Key words: accretion, accretion disks – black hole physics – Galaxy: center – methods: statistical

Online-only material: color figure

1. INTRODUCTION

Many different mechanisms can cause an astronomical source to be variable. Accreting black holes, for example, are variable electromagnetic sources for reasons such as oscillations in an accretion disk, changes in the accretion rate, or turbulent plasma processes like magnetic reconnection leading to an acceleration of electrons. Most often the observed variability in the flux can be well described by a single stochastic process such as a random walk (e.g., Kelly et al. 2009; MacLeod et al. 2010; Zu et al. 2013). If, however, the accretion rate suddenly jumps (e.g., due to the tidal disruption of a star or asteroid), the observed variability can change drastically and in a way that is not well described by a random walk. Such a state change in the variability could be obvious if the effects are large like an increase of the mean flux by orders of magnitude. However, there could be changes that are subtle and not trivial to detect. In this paper, we present a formal method to do just that. We apply the method to the massive black hole in the center of the Milky Way. This source is of particular interest since two distinct states have been claimed to be present in the past (Dodds-Eden et al. 2011), and the upcoming encounter with the gaseous object G2 (Gillessen et al. 2012, 2013a, 2013b; Phifer et al. 2013; Meyer et al. 2013) might or might not lead to a variability state change.

The emission associated with the accretion flow around the Galactic black hole, Sgr A*, has been detected in a few wavelength regimes (for recent reviews see Genzel et al. 2010; Morris et al. 2012; Falcke & Markoff 2013). While it was first detected at radio wavelengths a few decades ago (Balick & Brown 1974), its discovery in the X-rays (Baganoff et al. 2001) and near-infrared (Genzel et al. 2003; Ghez et al. 2004; Eckart et al. 2004) did not happen until the early 2000s when advanced imaging systems came online (*Chandra*/*XMM* in the X-rays and adaptive optics in the near-infrared). Sgr A* is in Eddington

terms the most under-luminous massive black hole accessible to observations. Its unexpected faintness inspired a class of radiative inefficient accretion flow models (e.g., Narayan et al. 1995; Blandford & Begelman 1999; Yuan et al. 2003, 2004).

While Sgr A* is variable across all observable wavelengths, its variability is most pronounced in the near-infrared (NIR) and X-rays with flux excursions that lie a factor of ∼10 (for the NIR) and ∼100 (for the X-ray), respectively, above the low flux levels. Early studies of Sgr A*'s NIR and X-ray variability reported the existence of a quasi-periodic oscillation (QPO) of ∼17 min in the flux (Genzel et al. 2003; Aschenbach et al. 2004). However, this finding was not confirmed by later statistical analyses (Do et al. 2009). The potential existence of a QPO was met with great interest since it potentially offers a way to measure the spin of the black hole and to test the curvature of the space-time close to it (e.g., Broderick & Loeb 2005; Meyer et al. 2006; Johannsen & Psaltis 2011).

One advantage of the NIR over the X-ray regime for studying the variability of Sgr A* is that it is visible at NIR wavelengths much more of the time. While Sgr A*'s X-ray emission peaks above the steady background created by the extended, larger-scale thermal accretion flow around 4% of the time (Neilsen et al. 2013), Sgr A*'s NIR emission is almost always (≳90%) detected at the highest angular resolution possible today with Keck Observatory (as we will show here and in Witzel et al., in preparation). Recent NIR studies have shown that Sgr A* is sufficiently modeled as a purely random process. The power spectral density (PSD) of Sgr A*'s time variability is a featureless power law for relatively high frequencies, a characteristic that can be modeled with a random walk and is more generally called red noise (Do et al. 2009; Meyer et al. 2008). At lower frequencies, the power law breaks to a shallower slope at a timescale of 150–600 minutes (Meyer et al. 2009; Witzel et al. 2012).

In addition to the PSD, another key quantity to describe Sgr A*'s variability is its flux density distribution. Recently, Dodds-Eden et al. (2011) and Witzel et al. (2012) looked at the total flux density distribution in the NIR by constructing a histogram of all observed flux density values. Interestingly, their interpretations are quite different: Dodds-Eden et al. (2011) use a lognormal distribution + a power-law tail—convolved with a Gaussian to account for measurement errors—to describe the distribution and argue that Sgr A* has two distinct states, one described by the lognormal part and the other by the power-law tail. In contrast, Witzel et al. (2012) find that only a power law (convolved with a Gaussian) is needed to accurately describe the flux density histogram of Sgr A*. It is important to note that both studies do not use timing information to argue for one or two states. The question whether or not multiple states can be inferred from the data was identified as a key question in understanding Sgr A* by Genzel et al. (2010).

In this work, we derive a formal method to answer the question of how many states can be inferred from Sgr A*'s light curve. In Section 2, we will present our methodology in detail. While it has been developed with the specific case of Sgr A* in mind, its application should be more general. It is useful whenever one wants to investigate if a time variable source exhibits distinct states of variability. This method is known in economics as "regime switching model" (see, e.g., Hamilton 1994). We then apply this approach to Sgr A* using the most extensive NIR data set constructed to date (Witzel et al., in preparation). We end by discussing how to assess the upcoming impact of the gaseous object G2 on the accretion flow around the black hole. Our methodology along with the unprecedented data set, which represents the best base line of Sgr A*'s behavior before G2 swings around in early 2014, is the ideal tool set to quantify Sgr A*'s response to any mass accreted from G2.

2. METHODOLOGY

The notion of hidden states in a time series has been applied to many diverse fields for a few decades (e.g., Rabiner 1989). Simply put, the idea is that observables entail information about states that are not directly accessible to the observer. Figure 1 sketches a simple system that can be in a ground state and an excited state, and the distribution of the observable is different for each state. Overall, a mixture of distributions would be observed. A two-state hidden Markov model that looks at the time sequence of observations and assumes a general form for the distributions is able to determine their parameters, the probability to transition from the ground to the excited state and vice versa, as well as the probability to be in a certain state for every observation. Modeling the stochastic process of the state variable as Markovian means that the process satisfies the Markov property: one can make predictions for the future state based solely on its present state just as well as one could knowing the process's full history.

The key to this approach is that we move beyond the simple unconditional flux distribution—measured by the histogram of all flux densities—and use the information in the time series to identify conditional flux distributions. To illustrate the difference between unconditional and conditional distributions, consider the following analogy. Imagine that one wanted to forecast the high temperature on the Santa Monica beach tomorrow. One way to do this would be to compute the distribution of all high temperatures over the entire year and use the mean of this unconditional distribution as the forecast. A much better way to proceed, however, would be to estimate the distribution of day-ahead temperatures conditional on today's high temperature being, say, 50 degrees. By conditioning on today's temperature, the distribution for tomorrow's temperature would be much tighter and more informative that would be the unconditional distribution based on the entire year. This analogy illustrates the intuition behind our approach. We use the current flux measurement to specify the conditional distribution of the subsequent flux measurement and estimate a range of conditional distributions, one for each state, rather than just using the single unconditional distribution.

In the following, we will first describe our approach and the construction of the likelihood function in more detail with a simplified example (adopted from Hamilton 2008) and then generalize this to the more complex case of Sgr A*.

2.1. Simple Example

Let us consider a hypothetical flux time series where y_t is the flux measured at time t, and the data are well described by a stochastic model of the form

$$y_t = c_1 + \phi y_{t-1} + \epsilon_t, \qquad (1)$$

where c_1 and $|\phi| < 1$ are constants and $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Let us further assume that we would like to test whether a model that allows a change in mean flux gives a better description of the data. For a permanent change of the mean flux, we could just write down a different parameter c_2 for some $t > t_1$,

$$y_t = c_2 + \phi y_{t-1} + \epsilon_t. \qquad (2)$$

We are, however, interested in a situation where we can jump back and forth between both models. We will refer to the different models as different states, and at a given time, the source emitting the flux is in state $s_t = j$, where $j = 1$ or 2 in our example. We will assume that this state is not directly observable (e.g., with spectroscopic observations) but has to be inferred from the light curve itself.

The unobserved state is modeled using a Markov switching approach. Concretely, let p_{ij} be a matrix that reflects the transition probabilities of switching to another state or remaining in the current state,

$$p_{ij} = \text{Prob}(s_t = j | s_{t-1} = i), \qquad (3)$$

and let Π_{jt} be the probability of being in state j at time t,

$$\Pi_{jt} = \text{Prob}(s_t = j | y_t). \qquad (4)$$

The probability density of y_t for our model is then

$$f(y_t | s_t = j, y_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_t - c_j - \phi y_{t-1})^2}{2\sigma^2}\right)$$
$$= f_{jt}. \qquad (5)$$

With these notations, the conditional probability density of y_t can be written as

$$h(y_t | y_{t-1}; \theta) = \sum_{i,j} p_{ij} \Pi_{it-1} f_{jt}, \qquad (6)$$

where θ means the set of parameters in the model (here, c_j, ϕ, and p_{ij}). Since Π_{jt} is not directly observable, it has to be recursively calculated by observing that

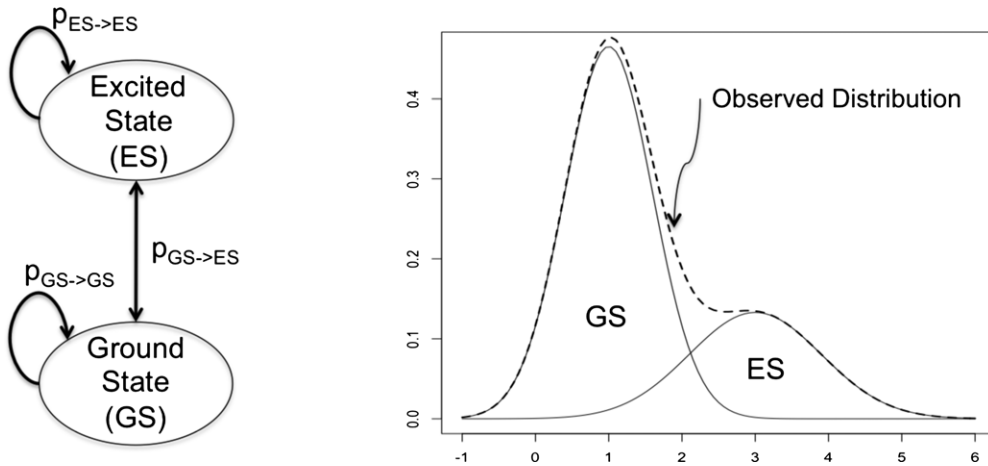$$\Pi_{jt} = \frac{\sum_i p_{ij} \Pi_{it-1} f_{jt}}{h(y_t | y_{t-1}; \theta)}. \qquad (7)$$

Figure 1. Illustration of a two-state hidden Markov model. The observable quantity shows the dashed overall distribution. It can be decomposed into a ground state (GS) and an excited state (ES), both of which are not directly observable. Given a sequence of observations, the parameters of the individual distributions as well as the transition probabilities can be solved for. Note that this illustration is overly simplified, and in our application to time series, the overall, unconditional distribution cannot simply be calculated as the weighted sum of the conditional ones.

Starting at t_0, we can execute the iteration to solve for all Π_{jt}. This means that the final log-likelihood function of the whole light curve is given by

$$\log h(y_1, y_2, y_3, ..., y_T | y_0; \theta) = \sum_{t=1}^{T} \log h(y_t | y_{t-1}; \theta). \quad (8)$$

The maximum of this log-likelihood function gives the preferred values for c_j, ϕ, and p_{ij}.

2.2. Extension to Sgr A*

While the simple model elucidates the construction of the likelihood function, it is not well suited to be applied to light curves from Sgr A*. The main reason is that the real data set shows unevenly sampled measurements with big gaps representing limited telescope time, the night/day cycle, the observability of the Galactic center from the ground throughout a year, measurements of sky background, and instrument failures. This requires us to use a more appropriate model that is time continuous and not discrete as in the above example. Furthermore, the flux density distribution is not Gaussian, and the data contain measurement noise. This is where the astronomical application of the regime switching approach departs from mathematical finance. We will deal with these points step by step in the following.

A popular model to describe the random variability of quasars is a so-called damped random walk, which is the only process that is stationary, Gaussian, and Markov. This process is also known as an Ornstein–Uhlenbeck (OU) model, which is the nomenclature we will use here. It is similar to red-noise models with a broken power law as the power spectral density, which have been used to model Sgr A* in the past (Do et al. 2009; Meyer et al. 2009; Witzel et al. 2012). Most recently, Dexter et al. (2014) used an OU process to successfully describe the submillimeter variability of Sgr A*. Additionally, Kelly et al. (2009), MacLeod et al. (2010), and Zu et al. (2013) have shown that an OU process is an excellent description of active galactic nucleus (AGN) flux variations. Its key advantage for our purpose is that it is a time continuous model that makes the handling of sampling gaps easier.

The OU model is determined by three parameters, the mean μ, the speed of mean reversion k, and the volatility σ, and

the dynamics can be described by the following stochastic differential equation:

$$dy = k(\mu - y)dt + \sigma dZ_t, \quad (9)$$

where dZ_t is the increment of a Brownian motion with $Z_t \sim \mathcal{N}(0, t)$. A solution to this equation is the conditional distribution

$$f(y_{t+\Delta t} | y_t) = \frac{1}{\sqrt{2\pi \tilde{\sigma}^2}} \exp\left(\frac{-(y_{t+\Delta t} - \tilde{\mu})^2}{2\tilde{\sigma}^2}\right), \quad (10)$$

$$\tilde{\mu} = y_t e^{-k\Delta t} + \mu(1 - e^{-k\Delta t}), \quad (11)$$

$$\tilde{\sigma}^2 = \frac{\sigma^2(1 - e^{-2k\Delta t})}{2k}. \quad (12)$$

Our goal is to identify whether more than one state is present in the light curves from Sgr A*. Past studies of this source have already revealed a great deal of information, and in particular imply that any additional state will be subtle: as reported by Witzel et al. (2012) and Dodds-Eden et al. (2011), a unimodal, heavy-tailed, power-law-like distribution convolved with a Gaussian describes the flux density distribution accurately. In our case, it is therefore important that the baseline model is already a good description of this overall flux distribution and power spectral density; otherwise, a model with two or more states might be wrongly preferred over the baseline model of one state if the one state is modeled incorrectly. An OU process matches the observed characteristics of Sgr A*'s power spectrum: a broken power law that is otherwise featureless. The observed flux distribution, however, is not Gaussian. In order to match the power-law-like distribution convolved with a Gaussian, we will also use the exponential and double-exponential of an OU process to model Sgr A*. Since this translates into the equivalent of a lognormal (and loglognormal) distribution, we will use the notation of a log- and log log-OU process in the following. Please note that taking the logarithm twice is purely empirically motivated and not rooted in physical considerations. Ideally, we would like to use a power-law distribution, but a stochastic model similar to Equation (9) which results in a power-law density is not known.

The conditional distributions for the log- and log log-OU processes are given as follows[3]: (1) for the log of a OU process,

$$f(y_{t+\Delta t}|y_t) = \frac{1}{\sqrt{2\pi\tilde{\sigma}^2}y_{t+\Delta t}} \exp\left(\frac{-(\ln(y_{t+\Delta t}) - \tilde{\mu})^2}{2\tilde{\sigma}^2}\right), \quad (13)$$

$$\tilde{\mu} = \ln(y_t)e^{-k\Delta t} + \mu(1 - e^{-k\Delta t}), \quad (14)$$

$$\tilde{\sigma}^2 = \frac{\sigma^2(1 - e^{-2k\Delta t})}{2k}, \quad (15)$$

and (2) for the log log of a OU process,

$$f(y_{t+\Delta t}|y_t) = \frac{1}{\sqrt{2\pi\tilde{\sigma}^2}y_{t+\Delta t}\ln(y_{t+\Delta t})}$$
$$\times \exp\left(\frac{-(\ln(\ln(y_{t+\Delta t})) - \tilde{\mu})^2}{2\tilde{\sigma}^2}\right), \quad (16)$$

$$\tilde{\mu} = \ln(\ln(y_t))e^{-k\Delta t} + \mu(1 - e^{-k\Delta t}), \quad (17)$$

$$\tilde{\sigma}^2 = \frac{\sigma^2(1 - e^{-2k\Delta t})}{2k}. \quad (18)$$

The choice of a time continuous model leads to the parameter Δt in the equations above and therefore offers a direct way to deal with the gaps in Sgr A*'s light curve for the conditional distribution. However, the transition matrix $P = p_{ij}$ is calculated for a specific time difference τ and must also be modified when gaps are present.[4] A straightforward way to do this is to multiply the transition matrix with itself N times for a gap that is $\tau \cdot N$:

$$A = P^N, \quad \text{with} \quad N = \max(1, \text{round}(\Delta t/\tau)). \quad (19)$$

For Sgr A*'s data set, which has an average sampling of one measurement per 1.2 minutes (not counting the large nightly/yearly gaps), we chose a final value of $\tau = 1$ minute. We have explored much shorter values but found it to make no significant difference. For values of $\Delta t > 1000$ minutes, we set $\Delta t = 1000$ minutes since these gaps are safely greater than the mean-reversion timescale of Sgr A* (Meyer et al. 2009; Witzel et al. 2012), and the flux density points after these gaps are therefore a new, independent realization.

The stochastic model in Equation (9) aims to describe the intrinsic properties of the source under consideration. In a realistic setting, however, an additional noise component is present that reflects the measurement process. This measurement noise is typically white noise, i.e., it does not depend on the previous observation, and it is well described by a Gaussian and therefore fully specified by one parameter σ_{meas}. Often, there exists an estimate of σ_{meas}. For the case of Sgr A*, for example, nearby stars of similar magnitude visible in the same image offer a straightforward way to estimate the measurement noise since these stars have constant flux intrinsically. In case an independent estimate of σ_{meas} is present, it can be advantageous to include it in the

stochastic model of the source. In order to do this, the conditional distribution from the OU process has to be convolved with a Normal distribution $\mathcal{N}(0, \sigma_{\text{meas}})$,

$$\tilde{f}(y_{t+\Delta t}|y_t; \sigma_{\text{meas}}) = \int f(x|y_t) \cdot \frac{1}{\sqrt{2\pi\sigma_{\text{meas}}^2}}$$
$$\times \exp\left(\frac{-(y_{t+\Delta t} - x)^2}{2\sigma_{\text{meas}}^2}\right) dx. \quad (20)$$

For the important cases of a log- or log log-OU process (Equations (13) and (16)), this integral has to be calculated numerically in every term of the sum in the likelihood function of Equation (8). Note that this dramatically increases the computing time. For us, the duration of computing the posterior with the numerical integration increased to the order of days from just a few hours without it. Please also note that this approach of incorporating the measurement noise assumes that it is constant. While photon noise leads to an increase of noise with flux, the data properties are well modeled with the assumption of constancy and may reflect the dominance of point-spread function measurement noise (Witzel et al. 2012; G. Witzel et al., in preparation).

2.3. Is an Additional State Justified?

An important question is how to decide whether more than one state is needed at all, and if so, how many different states can be inferred from the data. Here lies an important distinction between astrophysical and economic analyses: while the latter are mainly interested in a precise model of a time series to make accurate forecasts, the former are insight driven. The necessity of different states could point to different physical mechanisms and elucidate the astrophysics of the source under consideration. Whether or not an additional state is necessary in the model of an astronomical time series should be assessed by several metrics. Bayesian methods such as comparing the Bayesian evidences for different models belong to the standard methods well suited to comparing models but have the (dis)advantage of forcing one to write down specific priors for the parameters, which are often ambiguous. Note that Hamilton (2008) warns that methods relying on likelihood ratio tests fail to satisfy necessary regularity conditions.

A quite robust way of assessing the necessity of an additional state is to look at the global likelihood/posterior distribution in another way: too many assumed distinct states will lead to a highly multi-modal, very irregular looking posterior distribution. If only the maximum of the distribution is determined, the optimizer might often fail to converge at all in that case. Most importantly, the individual states should show persistence, if they are real. In any solution with a superfluous additional state, the probability of remaining in that state p_{22} will be $\ll 0.1$, while a significant additional state should show persistence with $p_{22} \gtrsim 0.8$. If the additional state has a very low probability of remaining in that state, two things might occur. (1) The probability of remaining in the first state is also very low, meaning that the states fluctuate from point to point. It seems extremely unnatural that state changes in the observed source occur exactly at the sampling rate of the measurements. (2) The probability of remaining in the first state is very high. In this case, the source will be in the additional, second state hardly at all. Only very few flux points will be assigned to that state, and a natural explanation is that these are outliers for the assumed first state conditional distribution. This can easily happen if any assumption of the stochastic model is not quite accurate, e.g., if the flux

---

[3] Please note that for ease of notation we will use the same symbols for the mean and variance as above. They are related through a change of variables.

[4] We use a discrete time transition matrix since the images are recorded at essentially the same cadence or at integral multiples thereof. We note that this approach could be extended to a continuous time formulation by allowing transitions to be triggered by the realizations of continuous time Poisson processes with constant intensities.
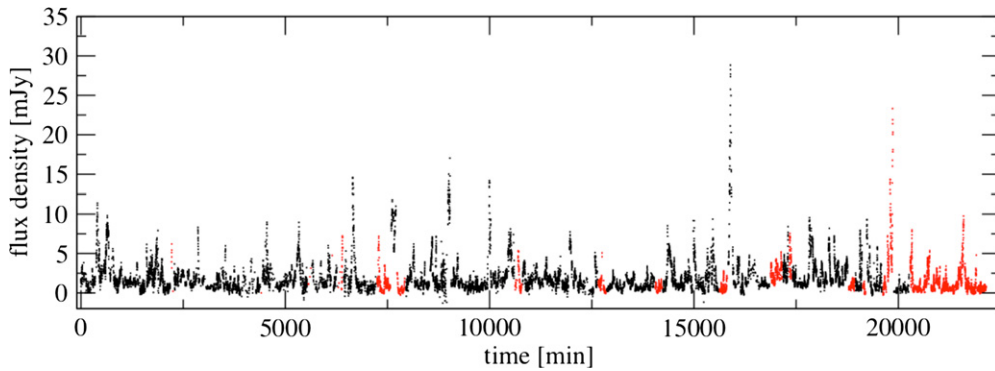
Figure 2. Longest, most comprehensive NIR light curve of Sgr A* that is available today (13,800 data points, displayed here without gaps longer than 30 minutes); see G. Witzel et al. (in preparation). These data have been taken in the *K* band from 2003 to 2013 with both the VLT (black) and the Keck observatory (red) and show a typical cadence of about 1 minute for the individual night.

(A color version of this figure is available in the online journal.)

density distribution is not quite lognormal or the measurement noise is not strictly constant.

3. RESULTS FOR SGR A* DATA

In this section, we will show the results of the regime-switching approach applied to Sgr A*. We will use two data sets, both taken with adaptive optics in the near-IR K band: all publicly available (up to 2010) Very Large Telescope (VLT) data as published in Witzel et al. (2012) and all AO Keck data taken from Sgr A* (up to and including 2013; see Witzel et al., in preparation). The photometry has been extracted in the same way in both data sets. We refer the reader to Witzel et al. (2012) and G. Witzel et al. (in preparation) for details. Key features of each data set are as follows.

1. The VLT data were taken with NaCo in Ks band (2.18 μm; 68 mas resolution), the Keck data with NIRC2 in K' band (2.12 μm; 53 mas resolution).
2. The VLT data contain 10,639 quality-selected data points, taken between 2003 June 13 and 2010 June 16, and the Keck data contain 3157 quality-selected points between 2004 July 16 and 2013 July 19.
3. The average sampling of the covered time periods is 1.2 minutes (VLT) and 1.1 minutes (Keck).
4. The integration time is 28 s for Keck and ranges between 30 and 40 s for VLT.
5. Both data sets use consistent flux density calibration using 13 non-variable stars throughout all epochs.
6. Both data sets are corrected for extinction with $m_{ext} = 2.46$ and for confusion levels (epoch by epoch).
7. The (Gaussian) measurement noise is determined to be 0.32 mJy (VLT) and 0.16 mJy (Keck).
8. Typical background fluxes are 0.6 mJy (VLT) and 0.3 mJy (Keck).
9. The data cover a de-reddened flux density range of 0–29 (VLT)/23 (Keck) mJy, which is consistent with a power-law distribution of intrinsic fluxes in both cases.

Figure 2 shows the complete data set. Since the two data sets come from different telescopes and instruments and show substantial differences in noise characteristics, we mainly analyze them separately.

In the following, we will first model SgrA* using a single state only. This will serve as our baseline model. We will then go on modeling the data with two states and see whether a substantial improvement has been achieved. We will first present

the analysis without accounting for the measurement (white) noise component. We think that any first pass of a new data set should be done without the very time-consuming convolution in Equation (20). The treatment including the measurement noise follows after that.

3.1. Baseline Model

Up to now, the timing behavior of Sgr A* has always been modeled using only a single state in the literature. An advantage of this approach is that it allows for any choice of the overall flux distribution, e.g., a power law convolved with a Gaussian as in Witzel et al. (2012). In our methodology, however, where we explicitly model conditional distributions, we do not have the freedom to choose a power-law behavior since a dynamic model similar to Equation (9) that results in power-law distributions is not known. Therefore, we have to show first that a log- or log log-OU process can accurately model the overall, unconditional flux distribution as well. If this was not the case, any additional state would likely be preferred just because a single state does not appropriately reflect the distribution of fluxes.[5] The key feature we are looking for is a different timing behavior since this is expected from a significantly distinct state that reflects a distinct physical process in the accretion flow.

Most importantly, the total flux distribution of Sgr A* is peaked and highly skewed. Exploring the distribution visually, it is noticeable that the histogram of log(flux) and log(log(flux)) resembles a Gaussian, although it is still skewed in the latter case. This skew is minimized when a constant is added to the entire light curve, i.e., $y_t = y_t + c$ for all t, where y_t denotes the flux density at time t. For the loglog of the flux, this constant is evaluated at $c = 1.25$ mJy for the VLT data and $c = 1.35$ mJy for the Keck data, and in the case of log(flux), it is 0 mJy for both data sets. Note that a constant of $c > 1$ would be required anyway in order for the loglog to be defined since the light curves are normalized such that $\min(y_t) = 0$. Table 1 shows the Bayesian evidences and the best-fit parameter values for both models and data sets. The preferred model in both cases is the loglog one, although the difference from the log model is marginal for the VLT data. We adopt the loglog-OU process

---

[5] We would like to point out that, in general, a single-state model must not accurately describe the flux distribution, e.g., in sources where the flux distribution is multi-modal. However, since for Sgr A* it has been shown that a single-state model can accurately model the flux distribution, we require this for our single-state baseline model as well. Note that this does not imply that a second state is not present.

Table 1
Multi-state Modeling of Sgr A* (without Measurement Noise Component)

Model	log(Evidence)	Parameter Values[a]
VLT data		
log log-OU	−9639	$(k, \mu, \sigma) = (0.13, 0.16, 0.17)$
log-OU	−9642	$(k, \mu, \sigma) = (0.10, 0.69, 0.30)$
log log-OU/OU	−5364	$(k_1, \mu_1, \sigma_1, p_{11}, k_2, \mu_2, \sigma_2, p_{21}) = (0.016, 0.24, 0.061, 0.96, 0.52, 2.90, 0.84, 0.11)$
loglog-OU/log-OU	**−4813**	$(k_1, \mu_1, \sigma_1, p_{11}, k_2, \mu_2, \sigma_2, p_{21}) = (0.52, 0.016, 0.32, 0.91, 0.017, 1.29, 0.079, 0.02)$
log log-OU/log log-OU	−5177	$(k_1, \mu_1, \sigma_1, p_{11}, k_2, \mu_2, \sigma_2, p_{21}) = (0.019, 0.25, 0.059, 0.97, 0.42, 0.002, 0.28, 0.08)$
log log-OU/log log-OU[b]	−5334	$(k_1, \mu, \sigma, p_{11}, k_2, p_{21}) = (0.031, 0.16, 0.060, 0.97, 0.14, 0.08)$
Keck data		
log log-OU	−733	$(k, \mu, \sigma) = (0.04, 0.023, 0.12)$
log-OU	−939	$(k, \mu, \sigma) = (0.04, 0.29, 0.26)$
log log-OU/OU	−158	$(k_1, \mu_1, \sigma_1, p_{11}, k_2, \mu_2, \sigma_2, p_{21}) = (0.02, 0.03, 0.08, 0.92, 0.016, 3.86, 0.33, 0.13)$
loglog-OU/log-OU	**−32**	$(k_1, \mu_1, \sigma_1, p_{11}, k_2, \mu_2, \sigma_2, p_{21}) = (0.17, 0.04, 0.29, 0.74, 0.015, 1.06, 0.073, 0.02)$
log log-OU/log log-OU	−180	$(k_1, \mu_1, \sigma_1, p_{11}, k_2, \mu_2, \sigma_2, p_{21}) = (0.017, 0.10, 0.067, 0.96, 0.10, 0.014, 0.21, 0.16)$
log log-OU/log log-OU[b]	−175	$(k_1, \mu, \sigma, p_{11}, k_2, p_{21}) = (0.021, 0.036, 0.067, 0.96, 0.067, 0.15)$

Notes. The bold entries have the largest evidence values and are therefore the preferred models.
[a] The unit for the k parameter is always minute^{-1}. The units for μ and $\sigma * \sqrt{t}$ depend on the model and are either mJy, log(mJy), or log(log(mJy)). The prior for all parameters is a uniform distribution $\mathcal{U}(0, 10)$.
[b] This model only allows for changes in the timing behavior. Here, we set $\mu_1 = \mu_2$ and $\sigma_2 = k_2/k_1 \, \sigma_1$. This tests whether the additional state is mainly driven by a different timing behavior (manifested in the mean reversion time k), or by the unconditional flux distribution (described by μ and σ).
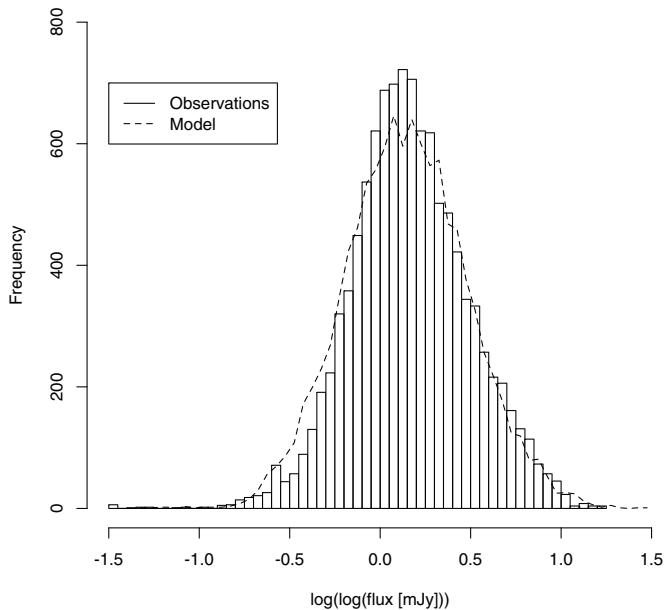

Figure 3. Overall flux distribution of Sgr A* from the VLT data set presented in Witzel et al. (2012; solid line) and our single-state baseline model (dashed line). Concretely, the log log of the flux after a constant value of 1.25 mJy has been added is shown. The observed distribution closely resembles a Gaussian, which shows that our assumptions about the baseline model are accurate.

as the baseline when measurement noise is not included via Equation (20), and the resulting model as well as the observed flux density histograms for the VLT data are shown in Figure 3. The comparison of baseline model and data shows that while the agreement is very good, it is not perfect. However, since we cannot use power-law distributions and do not explicitly model the measurement noise here, this is not surprising.

We have used the Bayesian nested sampler *MultiNest* (Feroz & Hobson 2008; Feroz et al. 2009) to explore the posterior distribution and assumed uniform priors for the parameters $\mathcal{U}(0, 10)$.

3.2. Modeling without the Measurement Noise Component

We can now turn to the question we set out to answer: does Sgr A* have more than one state? While the choice of the conditional distribution for the baseline model was motivated to most closely match the observed overall flux distribution (for a single state, the conditional flux distribution obviously equals the overall one), the two-state case is more complex. Fundamentally, the selection of the form of the conditional distributions (e.g., OU, *log*-OU, etc.) is an assumption to make and cannot be derived from the data. In a practical approach, we have chosen many different combinations and looked at them individually.

The results of the modeling with different assumptions for the conditional distributions are summarized in Table 1. Strikingly, all two-state scenarios are preferred over the single-state model for both data sets as indicated by the Bayesian evidence. All posteriors are well behaved and not more than bimodal (for a two-state model, the posterior will often be bimodal[6]), and the probability of remaining in a given state is high, indicating persistence of these states. Taken together, this is strong evidence for the presence of more than one variability state in the observed flux from Sgr A*. The overall preferred two-state model is one consisting of a log log-OU and a log-OU state. This is true for both the VLT and Keck data sets.

In Figure 4, we show the observed VLT flux density time series in a decomposed way: the upper panel contains all points that are with probability >0.5 in the log-OU state, while the lower panel shows all flux points that are with probability >0.5 in the log log-OU state. The differences are clearly visible. In fact, in all explored two-state models, one state is always reverting to its mean fast and has a lower mean, while the other state is slower mean reverting with a higher mean. A fast mean-reversion here means that this timescale is at the order of the sampling. We have also tested that the second state is not just preferred because of a better description of the overall flux distribution: forcing the two states to have the

---

[6] This is because the labels for states 1 and 2 could be switched, while the value of the likelihood function would remain the same.
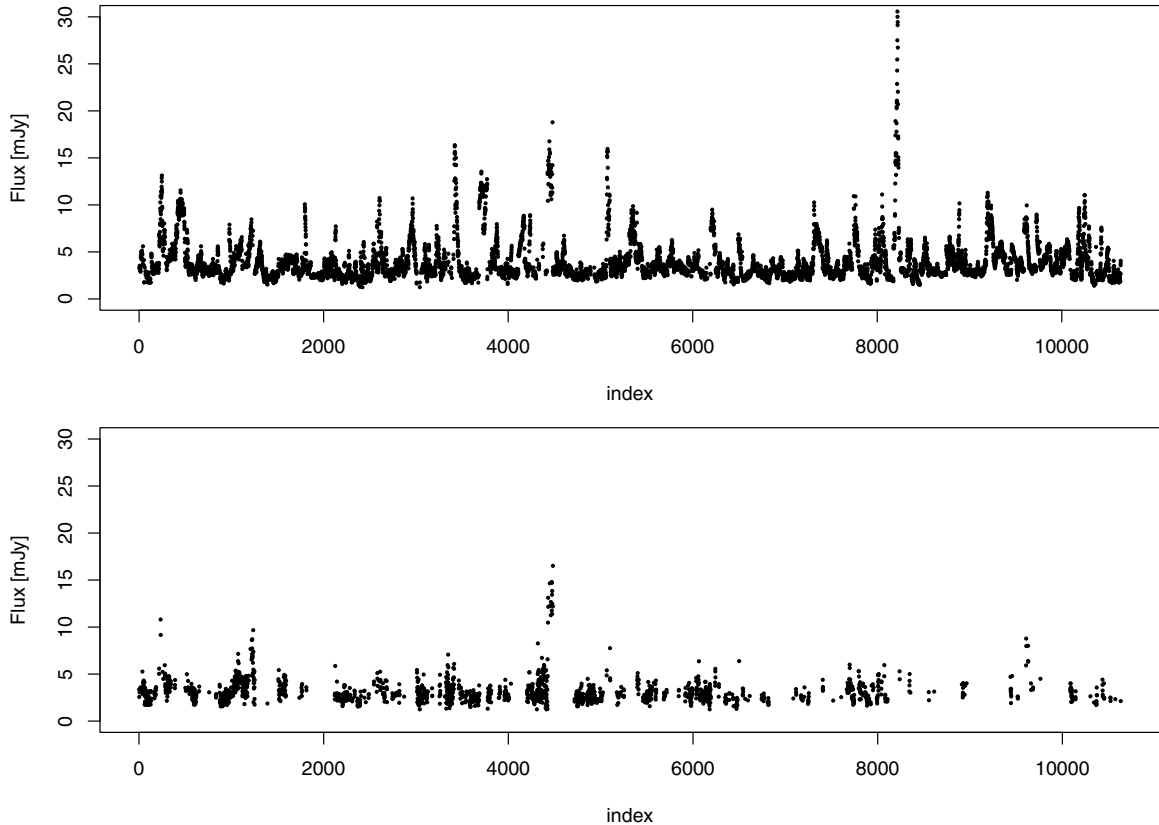
Figure 4. Decomposition of the VLT data into the best-fit two-state model (see Table 1, row 4). Upper panel: all flux points of the time series that have a probability of >0.5 in the log-OU state (83% of all points). Lower panel: all flux points that have a probability of >0.5 in the log log-OU state (17% of all points). This visualizes the differences in the two states: the log log-OU state is quickly reverting to a rather low mean, while the log-OU state is more slowly mean reverting and the mean is higher. The log log-OU state represents the state where the changes in flux are dominated by measurement noise, and the log-OU state represents the changes in flux that are source dominated.

same form of the unconditional distribution by setting $\mu_1 = \mu_2$ and $\sigma_2 = k_2/k_1 \; \sigma_1$, we still see strong evidence for a second state (see last lines in Table 1). It is the timing that drives the significance of a second state.

The fact that one of the two states is quickly mean reverting to a fairly low mean suggests that the interpretation of this state predominantly describes the parts of the light curve that are dominated by measurement noise, which is approximately white noise. Note that even if this is the case, the above analysis is still valuable since it offers a recipe to label each point-to-point change of the flux series as "noise dominated" or "source dominated." Consistent with this interpretation is the observation that the less noisy Keck data are only "noise dominated" 6% of the time, while the VLT data are "noise dominated" 17% of the time.

3.3. Modeling with Measurement Noise

Our hypothesis is that the quickly mean-reverting state represents the measurement, white noise process. We will test this in two ways. Since we have an estimate of σ_{meas}, we can make use of Equation (20). Furthermore, Witzel et al. (2012) developed a Monte Carlo tool to simulate Sgr A* light curves (assuming a single state only), and we can analyze these mock data separately.

We summarize our results in Table 2 for the real data and in Table 3 for the mock data. Regarding the real data, the evidence for a second state vanishes when the measurement noise is incorporated into the model. For both data sets, the second state becomes highly non-persistent with a probability of remaining

in that state of $p_{22} \lesssim 1\%$. This is in stark contrast to the case of modeling without the extra noise component, where $p_{22} \gtrsim 90\%$ generally. For the VLT data, there is also no improvement in the Bayesian evidence. While there is some improvement in the evidence for the Keck data, the structure of the solution suggests that a few outliers drive this behavior (see also the discussion in Section 2.3).

Regarding the mock data, Table 3 shows that the mock data lead to a very similar solution as the real data, which is encouraging. Since the mock data algorithm was calibrated to the VLT data in terms of the sampling function and applied noise, the agreement between the mock data and the real VLT data is stronger than with the Keck data. When the measurement noise is not incorporated into the modeling, a second state is clearly preferred. Since the mock data follow a single-state process by construction, this shows that the white measurement noise leads to a discernible state in addition to the intrinsic red noise process of Sgr A*. Again, when the noise is explicitly convolved into the conditional distributions, the evidence for a second state vanishes. This solution puts the probability of remaining in the second state at $p_{22} \lesssim 1\%$.

In summary, the evidence for two distinct states in the flux from Sgr A* disappears when the measurement noise is taken into account. Therefore, we see our hypothesis verified: the quickly mean-reverting, lower-mean state represents the state when the observed flux changes are dominated by instrumental noise, and the slowly mean-reverting, higher-mean state represents the state when the flux changes are dominated by Sgr A* itself.

Table 2
Multi-state Modeling of Sgr A* with Measurement Noise Component[a]

Model	log(Evidence)	Parameter values[b]
VLT data		
log log-OU	−7899	$(k, \mu, \sigma) = (0.06, 0.14, 0.10)$
log-OU	−7436	$(k, \mu, \sigma) = (0.04, 0.75, 0.17)$
loglog-OU/log-OU	−7420	$(k_1, \mu_1, \sigma_1, p_{11}, k_2, \mu_2, \sigma_2, p_{21}) = (0.07, 0.17, 0.10, 0.99, 0.72, 1.39, 0.56, 0.98)$
		2nd state irrelevant and non-persistent ($p_{22} = 2\%$)
Keck data		
log log-OU	−432	$(k, \mu, \sigma) = (0.02, 0.04, 0.07)$
log-OU	−417	$(k, \mu, \sigma) = (0.02, 0.37, 0.16)$
loglog-OU/log-OU	−194	$(k_1, \mu_1, \sigma_1, p_{11}, k_2, \mu_2, \sigma_2, p_{21}) = (0.01, 0.04, 0.05, 0.97, 0.20, 1.33, 0.28, 0.99)$
		2nd state irrelevant and non-persistent ($p_{22} = 1\%$)

Notes.
[a] The value for σ_{meas} is $\sigma_{\mathrm{meas}} = 0.32$ mJy for the VLT data (Witzel et al. 2012), and $\sigma_{\mathrm{meas}} = 0.16$ mJy for the Keck data (G. Witzel et al., in preparation).
[b] The unit for the k parameter is always minute^{-1}. The units for μ and $\sigma * \sqrt{t}$ depend on the model and are either log(mJy) or log(log(mJy)). The prior for the parameters is a uniform distribution $\mathcal{U}(0, 1)$, only for μ_2 it is $\mathcal{U}(0, 3)$.

Table 3
Multi-state Modeling of Mock Data[a]

Model	log(Evidence)	Parameter values[b]
Modeling without measurement noise component		
log log-OU	−9157	$(k, \mu, \sigma) = (0.10, 0.15, 0.15)$
log-OU	−10477	$(k, \mu, \sigma) = (0.11, 0.87, 0.29)$
loglog-OU/log-OU	−7975	$(k_1, \mu_1, \sigma_1, p_{11}, k_2, \mu_2, \sigma_2, p_{21}) = (1.25, 0.001, 0.28, 0.97, 0.04, 1.48, 0.11, 0.01)$
Modeling with measurement noise component[c]		
log log-OU	−8486	$(k, \mu, \sigma) = (0.04, 0.25, 0.07)$
log-OU	−8286	$(k, \mu, \sigma) = (0.04, 0.87, 0.14)$
loglog-OU/log-OU	−8248	$(k_1, \mu_1, \sigma_1, p_{11}, k_2, \mu_2, \sigma_2, p_{21}) = (0.03, 0.04, 0.06, 0.50, 0.07, 1.75, 0.09, 0.999)$
		2nd state non-persistent ($p_{22} = 0.1\%$)

Notes.
[a] The model for the mock data is taken from Witzel et al. (2012), where it has been calibrated with the VLT data and uses a single state only.
[b] The unit for the k parameter is always minute^{-1}. The units for μ and $\sigma * \sqrt{t}$ depend on the model and are either log(mJy) or log(log(mJy)). When no measurement noise is modeled, the prior for all parameters is a uniform distribution $\mathcal{U}(0, 10)$. When measurement noise is modeled, the prior for the parameters is a uniform distribution $\mathcal{U}(0, 1)$, only for μ_2 it is $\mathcal{U}(0, 3)$.
[c] The value for σ_{meas} is $\sigma_{\mathrm{meas}} = 0.32$ mJy.

It is interesting to note that the same result holds true when the analysis is done on a combined Keck and VLT data set (see Figure 2). While two states are again being picked up when the measurement noise is unaccounted for, one state tends to describe white noise and the other Sgr A* intrinsically. With the measurement noise convolved into the conditional distributions, the evidence for a second state vanishes.

3.4. The Parameters of the Single-state (Best-fit) Model

We have shown that Sgr A* is sufficiently described by a single-state process once the measurement noise is accounted for in the modeling. The accurate model for Sgr A*'s light curve is therefore the log-OU process convolved with Gaussian noise (see Table 2). While it is tempting to interpret these parameters and compare them to different modeling approaches of the past, one has to be cautious when the single-state parameter inference is done as the limiting case of a multi-state model, as we have done here.

The reason for caution is as follows. Our approach models the conditional distribution from one measurement to the next, $f(y_{t+\Delta t}|y_t)$, which is necessary in the context of a multi-state Markov model. If only a single state is assumed, this means, however, that only the typical sampling horizon Δt is used to

estimate the mean-reversion timescale k. If the OU process is completely accurate and describes the true nature of the source, this is unproblematic. If, however, there are deviations from the OU process in the data, a different sampling horizon would lead to a different estimate of k. Ideally, a global approach using all time lags in the data would be used to estimate k for a single-state process. Since we focus here on multi-state modeling, this is beyond the scope of this work. We would like to note, however, that the analysis of our mock data lead to consistent results with the analysis of the real data, which means we explicitly confirm the results of Witzel et al. (2012).

4. DISCUSSION AND CONCLUSION

Using the most extensive light curve of Sgr A* to date with combined data from the VLT and Keck Observatory (G. Witzel et al., in preparation), we arrive at two main results: (1) the observed flux from Sgr A* shows two distinct states, one being noise dominated and the other source dominated, and (2) the intrinsic variability of Sgr A* is sufficiently described by a single-state stochastic process. Both findings have interesting implications, and we will discuss them in turn.

The presence of a noise-dominated and a source-dominated state in Sgr A* is a manifestation of the colloquial language

used in the Galactic center community, where a quiescent and a flaring state of Sgr A* in the NIR is often talked about. We have shown here that this intuitive distinction is correct when describing observed light curves; however, Sgr A* is intrinsically not in quiescent and flaring states but rather in one single state only. The language is therefore somewhat imprecise, and the word "flare" should not be used to describe Sgr A* in the NIR. This also means that the lack of statistical evidence for a QPO of ∼17 minutes can likely not be explained by distinct states of which only one is accompanied by a QPO, a possibility raised by Genzel et al. (2010). The fact that the intrinsic behavior of Sgr A* shows no evidence for a second state explores and falsifies the idea first suggested by Dodds-Eden et al. (2011). Physically, this means that one stochastic process without the addition of time-resolved, discrete events—such as the disruption of asteroids—fully explains our data set.

It will be interesting to see whether a change of Sgr A*'s intrinsic variability state changes when the gaseous, red, emission-line object G2 has passed the black hole in early 2014 (Gillessen et al. 2012, 2013a, 2013b; Phifer et al. 2013; Meyer et al. 2013). Since there is gas associated with this object, it could potentially offer a unique probe to observe the response of the accretion flow to a sudden increase of mass. However, the amount of gas is uncertain, and it is unclear if there will be any visible effect to the flux emission from Sgr A*. Two scenarios are likely: either something obvious will happen to Sgr A*'s flux distribution like a shift toward a much higher mean or—if there is any change at all—it could be very subtle, requiring a detailed statistical method like the one we have developed here. This statistical methodology, in addition to the best possible data baseline by merging the VLT and Keck data, might prove crucial in understanding G2's impact on Sgr A*.

Compared to simpler methods, our Hidden Markov Model approach presented here has the distinct advantage that it not only answers the question whether two states are needed, but it also solves for the times when the state changes happen. For G2, this is important because although the time and distance of the closest approach is fairly well known for G2's orbit, it is unclear when the gas has come down from $\sim 2400 R_S$ all the way to a few R_S where the near-infrared flux gets emitted. Given enough observations, solving for the time of state change (if there will be one) can directly test accretion flow dynamics around Sgr A*.

Another benefit of our approach that assigns a probability to each flux change to be noise or source dominated is that it offers a recipe to get an astrometric position of Sgr A* in the infrared. The astrometry of stars orbiting around the black hole aims to detect deviations from a purely Keplerian orbit as the next milestone (e.g., Ghez et al. 2008; Gillessen et al. 2009; Meyer et al. 2012). The astrometry is currently limited by the construction of an absolute reference frame that is used to transform all star positions into a common frame. This frame of reference is defined by seven maser sources visible in the radio as well as infrared (Yelda et al. 2010). Since Sgr A* is also visible at both of these wavelengths and sits at a position that is most important to anchor the reference frame, it could be of tremendous help in overcoming the current limitations. However, reliable astrometry of Sgr A* has been elusive. Its position changes as a function of brightness, probably because unresolved stellar sources bias its position, and the resulting astrometric shift is dependent on the relative contribution of Sgr A*'s intrinsic flux. Our two-state modeling approach effectively filters out the flux point variations in the time series that are dominated by Sgr A* itself. Getting a position

based on the images only where the apparent flux from Sgr A* is source dominated should improve the astrometry and lead to a consistent determination of its position.

As a final remark, we would like to note that the multi-state methodology presented here is directly applicable to other sources such as AGNs. The only key assumption is the validity of the OU process (also called a damped random walk). This also means that the method is not wavelength specific. The analysis of optical AGN light curves, e.g., should be possible in exactly the same way as is presented here.

REFERENCES

Aschenbach, B., Grosso, N., Porquet, D., & Predehl, P. 2004, A&A, 417, 71
Baganoff, F. K., Bautz, M. W., Brandt, W. N., et al. 2001, Natur, 413, 45
Balick, B., & Brown, R. L. 1974, ApJ, 194, 265
Blandford, R. D., & Begelman, M. C. 1999, MNRAS, 303, L1
Broderick, A. E., & Loeb, A. 2005, MNRAS, 363, 353
Dexter, J., Kelly, B. C., Bower, G. C., et al. 2014, MNRAS, 442, 2797
Do, T., Ghez, A. M., Morris, M. R., et al. 2009, ApJ, 691, 1021
Dodds-Eden, K., et al. 2011, ApJ, 728, 37
Eckart, A., Baganoff, F. K., Morris, M., et al. 2004, A&A, 427, 1
Falcke, H., & Markoff, S. B. 2013, CQGra, 30, 4003
Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Genzel, R., Eisenhauer, F., & Gillessen, S. 2010, RvMP, 82, 3121
Genzel, R., Schödel, R., Ott, T., et al. 2003, Natur, 425, 934
Ghez, A. M., Salim, S., Weinberg, N. N., et al. 2008, ApJ, 689, 1044
Ghez, A. M., Wright, S. A., Matthews, K., et al. 2004, ApJ, 601, L159
Gillessen, S., Eisenhauer, F., Trippe, S., et al. 2009, ApJ, 692, 1075
Gillessen, S., Genzel, R., Fritz, T. K., et al. 2012, Natur, 481, 51
Gillessen, S., Genzel, R., Fritz, T. K., et al. 2013a, ApJ, 763, 78
Gillessen, S., Genzel, R., Fritz, T. K., et al. 2013b, ApJ, 774, 7
Hamilton, J. D. 1994, Time Series Analysis (Princeton, NJ: Princeton Univ. Press)
Hamilton, J. D. 2008, in New Palgrave Dictionary of Economics, ed. S. Durlauf & L. Blume (2nd ed.; London: Palgrave MacMillan)
Johannsen, T., & Psaltis, D. 2011, ApJ, 726, 11
Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009, ApJ, 698, 895
MacLeod, C. L., Ivezic, Z., Kochanek, C. S., et al. 2010, ApJ, 721, 1014
Meyer, L., Do, T., Ghez, A., et al. 2008, ApJ, 688, L17
Meyer, L., Do, T., Ghez, A., et al. 2009, ApJ, 694, L87
Meyer, L., Ghez, A. M., Schödel, R., et al. 2012, Sci, 338, 84
Meyer, L., Ghez, A. M, Witzel, G., et al. 2013, IAUS, 303, 264
Meyer, L., Eckart, A., Schödel, R., et al. 2006b, A&A, 460, 15
Morris, M. R., Meyer, L., & Ghez, A. 2012, RAA, 12, 995
Narayan, R., Yi, I., & Mahadevan, R. 1995, Natur, 374, 623
Neilsen, J., Nowak, M. A., Gammie, C., et al. 2013, ApJ, 774, 42
Phifer, K., Do, T., Meyer, L., et al. 2013, ApJL, 773, L13
Rabiner, L. 1989, Proc. IEEE, 77, 257
Witzel, G., Eckart, A., Bremer, M., et al. 2012, ApJS, 203, 18
Yelda, S., Lu, J. R., Ghez, A., et al. 2010, ApJ, 725, 331
Yuan, F., Quataert, E., & Narayan, R. 2003, ApJ, 598, 301
Yuan, F., Quataert, E., & Narayan, R. 2004, ApJ, 606, 894
Zu, Y., Kochanek, C. S., Kozlowski, S., & Udalski, A. 2013, ApJ, 765, 106