# Solar X-ray variability in two distinct states and its real-time analysis based on a statistical autoregressive model

A. Stanislavsky[1], K. Burnecki[2], J. Janczura[2], K. Niczyj[2], A. Weron[2]

[1]Institute of Radio Astronomy, 4 Mystetstv St., Kharkiv, Ukraine

[2]Hugo Steinhaus Center, Wrocław University of Science and Technology,Wyb. Wyspianskiego 27, Wrocław, Poland

# Variability of radiation of space objects

- The nighttime sky seems immutable to the human eye, but this is not true.
- From ancient Egyptian, Greek, and Australian Aboriginal cultures it was known that a few stars (such as Algol, Mira, and Aldeberan) were recognized as variables.
- After starting telescopic studies from the seventeenth through twenty first centuries, more variable stars were found with a wide range of characteristics.
- NASA's Kepler mission has recently shown that most ordinary stars are variable when observed with ~0.001% accuracy.

# Causes of variability

- Due to pulsations, rotationally-modulated movements or eclipses of binary companions.
- Magnetic flares, eruptions, pulsations, gas accretion from companions and, most surprisingly, new and supernova explosions.
- The brightest sources in the X-ray and gamma-ray sky are highly variable, typically from accretion of gas onto neutron stars and black holes.

# Astronomy in the time domain

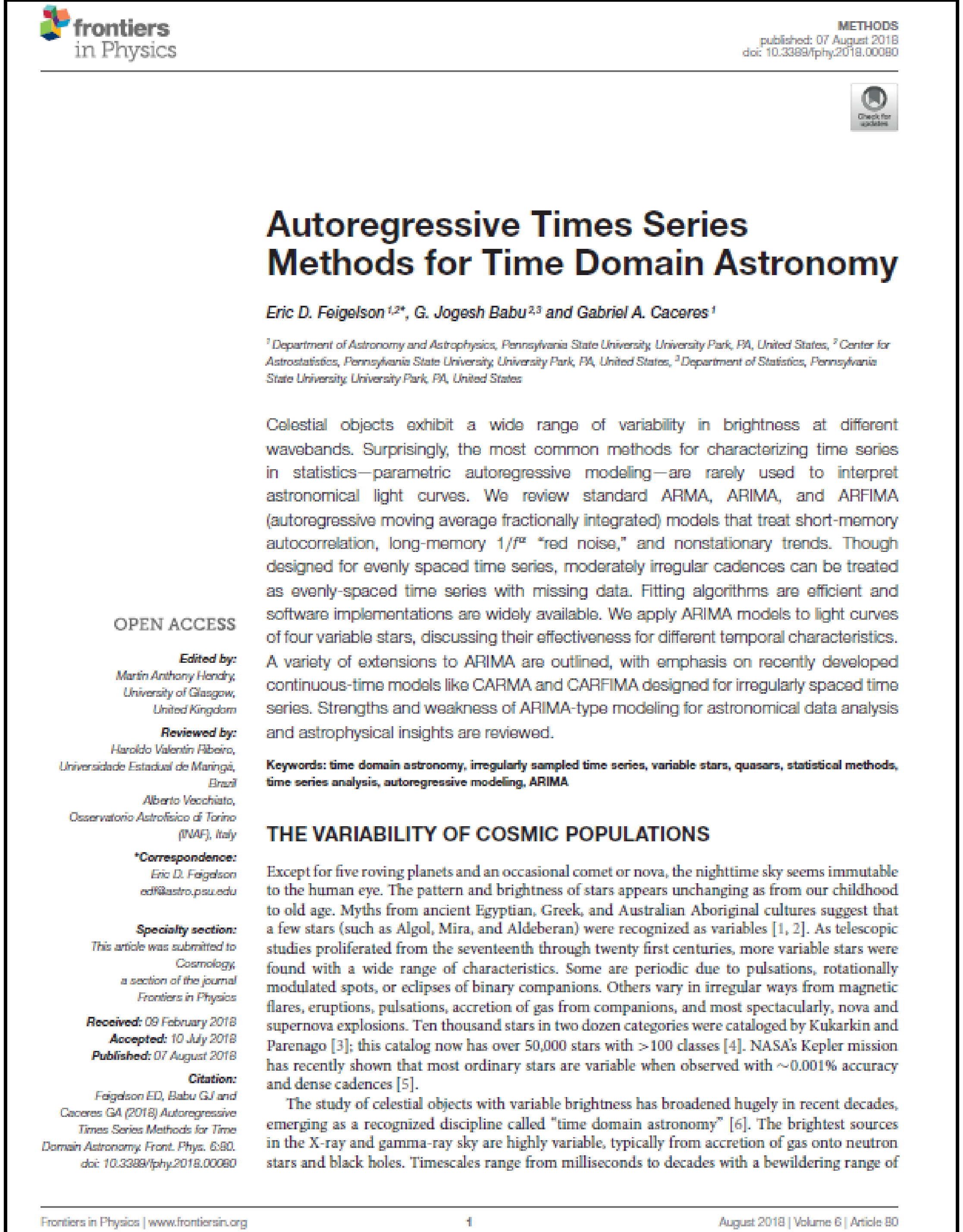


- Timescales range from milliseconds to decades with a bewildering range of periodic, quasi-periodic, stochastic and bursting characteristics.
- The Galactic black hole binary GRS 1915+105 alone has a dozen modes of variability.
- The radio sky has extragalactic quasars and blazars as well as Galactic pulsars and several varieties of fast radio bursts and transients.
- The non-photon gravitational wave observatories have recently emerged with rapid “chirps” from merging black hole and neutron star binaries.
- A huge industry searching for distant supernova explosions is propelled by their utility in tracing the accelerated expansion of the Universe.

# Methods of analysis

Accounting for this tremendous growth in the amount and complexity of astronomical time series data, we can ask what methods are common for their characterization and analysis.

- Time series texts oriented toward engineers and meteorologists generally use spectral and wavelet analysis rather autoregressive modeling.
- Surprisingly, the most common methods for characterizing time series in statistics — parametric autoregressive time domain models — are seldom used to interpret astronomical brightness variations of stars.



frontiers in Physics

METHODS
published: 07 August 2018
doi: 10.3389/fphy.2018.00080

# Autoregressive Times Series Methods for Time Domain Astronomy

Eric D. Feigelson[1,2]*, G. Jogesh Babu[2,3] and Gabriel A. Caceres[1]

[1] Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, PA, United States, [2] Center for Astrostatistics, Pennsylvania State University, University Park, PA, United States, [3] Department of Statistics, Pennsylvania State University, University Park, PA, United States

Celestial objects exhibit a wide range of variability in brightness at different wavebands. Surprisingly, the most common methods for characterizing time series in statistics—parametric autoregressive modeling—are rarely used to interpret astronomical light curves. We review standard ARMA, ARIMA, and ARFIMA (autoregressive moving average fractionally integrated) models that treat short-memory autocorrelation, long-memory $1/f^{\alpha}$ "red noise," and nonstationary trends. Though designed for evenly spaced time series, moderately irregular cadences can be treated as evenly-spaced time series with missing data. Fitting algorithms are efficient and software implementations are widely available. We apply ARIMA models to light curves of four variable stars, discussing their effectiveness for different temporal characteristics. A variety of extensions to ARIMA are outlined, with emphasis on recently developed continuous-time models like CARMA and CARFIMA designed for irregularly spaced time series. Strengths and weakness of ARIMA-type modeling for astronomical data analysis and astrophysical insights are reviewed.

**Keywords: time domain astronomy, irregularly sampled time series, variable stars, quasars, statistical methods, time series analysis, autoregressive modeling, ARIMA**

OPEN ACCESS

**Edited by:**
*Martin Anthony Hendry, University of Glasgow, United Kingdom*

**Reviewed by:**
*Haroldo Valentin Ribeiro, Universidade Estadual de Maringá, Brazil*
*Alberto Vecchiato, Osservatorio Astrofisico di Torino (INAF), Italy*

***Correspondence:**
*Eric D. Feigelson*
*edf@astro.psu.edu*

**Specialty section:**
*This article was submitted to Cosmology, a section of the journal Frontiers in Physics*

**Received:** *09 February 2018*
**Accepted:** *10 July 2018*
**Published:** *07 August 2018*

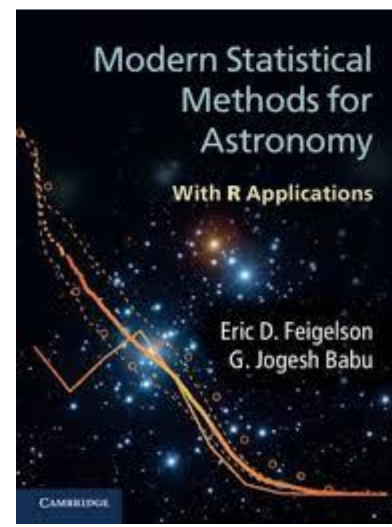
**Citation:**
*Feigelson ED, Babu GJ and Caceres GA (2018) Autoregressive Times Series Methods for Time Domain Astronomy. Front. Phys. 6:80. doi: 10.3389/fphy.2018.00080*

## THE VARIABILITY OF COSMIC POPULATIONS

Except for five roving planets and an occasional comet or nova, the nighttime sky seems immutable to the human eye. The pattern and brightness of stars appears unchanging as from our childhood to old age. Myths from ancient Egyptian, Greek, and Australian Aboriginal cultures suggest that a few stars (such as Algol, Mira, and Aldeberan) were recognized as variables [1, 2]. As telescopic studies proliferated from the seventeenth through twenty first centuries, more variable stars were found with a wide range of characteristics. Some are periodic due to pulsations, rotationally modulated spots, or eclipses of binary companions. Others vary in irregular ways from magnetic flares, eruptions, pulsations, accretion of gas from companions, and most spectacularly, nova and supernova explosions. Ten thousand stars in two dozen categories were cataloged by Kukarkin and Parenago [3]; this catalog now has over 50,000 stars with >100 classes [4]. NASA's Kepler mission has recently shown that most ordinary stars are variable when observed with ~0.001% accuracy and dense cadences [5].

The study of celestial objects with variable brightness has broadened hugely in recent decades, emerging as a recognized discipline called "time domain astronomy" [6]. The brightest sources in the X-ray and gamma-ray sky are highly variable, typically from accretion of gas onto neutron stars and black holes. Timescales range from milliseconds to decades with a bewildering range of





# Autoregressive model *AR*(*p*)

An autoregressive (AR) process has coefficients that quantify the dependence of current values on recent past values:

$$\boxed{y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + \varepsilon_t}$$

where $a_1 \ldots a_p$ are the corresponding coefficients for each lag up to order *p*, $a_0$ is constant (often for simplicity it is assumed to be zero), and $\varepsilon_t$ is a normally (Gaussian) distributed random error with zero mean and constant variance.

# Moving average model *MA(q)*

A moving average (MA) process has coefficients that quantify the dependence of current values on recent past random shocks to the system:

$$\boxed{y_t = \varepsilon_t - b_1\varepsilon_{t-1} - b_2\varepsilon_{t-2} - \ldots - b_q\varepsilon_{t-q}}$$

where $\varepsilon_t$ is the error term for the *t*-th time point, $b_1, \ldots, b_q$ are the coefficients for each lagged error term up to order *q*.

# ARMA(*p*, *q*)

Adding these two equations together gives a combined *ARMA*(*p*,*q*) process. Coefficients are estimated by standard regression procedures such as maximum likelihood estimation.

$$\boxed{y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + \varepsilon_t - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} - \ldots - b_q \varepsilon_{t-q}}$$

# Difference procedure

**Nonstationarity** and variable mean values can sometimes be removed by fitting a global regression model such as a polynomial, but often an adequate detrending regression model cannot be found. A flexible nonparametric procedure called **differencing** can remove nonstationarity in many such cases. Here one applies the backshift operator *B* that replaces the time series $y_t$ by another $y_t'$ consisting of the point-to-point difference in values:

$$y_t' = y_t - By_t = y_t - y_{t-1}$$

# *ARIMA(p,d,q)* and *ARFIMA(p,d,q)*

This combination of nonparametric differencing and integration with a parametric *ARMA* process is called the *ARIMA*(*p*,*d*,*q*) model where *d* represents the number of differencing operations applied and typically equals one.

A fractional integrated procedure can be described by

$$(1-B)^d y_t = \varepsilon_t , \quad (1-B)^d = \sum_{k=1}^{\infty} \binom{n}{k} (-B)^k$$

where *d* can be a real (non-integer) order of differencing and *B* is the backshift operator defined above.

# Attractive models for astronomical time series analysis



*ARMA*, *ARIMA*, and *ARFIMA* models can be very useful for astronomical time series analysis for various reasons:

- they are very flexible, successfully modeling an astonishing variety of irregular or quasi-periodic, smooth or choppy, constant or variable mean light curves.
- the dimensionality of the models is relatively low with a moderate computational burden of the numerical optimization.
- error analysis on the parameters naturally emerges through the likelihood regression analysis.
- they are extensible to situations involving multivariate time series, combinations of stochastic and deterministic behaviors, change points, and (moderately) irregular observation spacing.

# Disadvantages of autoregressive modeling



- Autoregressive modeling is not well-adapted to situations with strictly periodic variations (where the signal is compactly concentrated in Fourier power coefficients)
- or with sudden eruptive events (where the nonstationary amplitude is not greatly reduced by differencing).

Despite their advantages, non-trivial *ARIMA* models have been used very rarely in time domain astronomy!!!

# Testing the Model

The first action that every analyst should take after *ARIMA*-type modeling is model validation based, in part, on residual analysis. There are two reasons why:

- even though the "best fit" has been obtained in a maximum likelihood sense, the entire model family may not apply to the dataset under study;
- the model may be correctly specified but its underlying mathematical assumptions may be violated.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS 485, 3970–3980 (2019)
Advance Access publication 2019 March 7
doi:10.1093/mnras/stz656

# Solar X-ray variability in terms of a fractional heteroskedastic time series model

Aleksander A. Stanislavsky,[1,2]★ Krzysztof Burnecki,[3]★ Joanna Janczura,[3] Karol Niczyj[3] and Aleksander Weron[3]★

[1] *Institute of Radio Astronomy, 4 Mystetstv St., Kharkiv UA-61002, Ukraine*
[2] *Department of Space Radio Physics, V.N. Karazin Kharkiv National University, 4 Svobody Sq., Kharkiv UA-61022, Ukraine*
[3] *Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wrocław University of Science and Technology, Wyb. Wyspiańskiego 27, Wrocław PL-50-370, Poland*

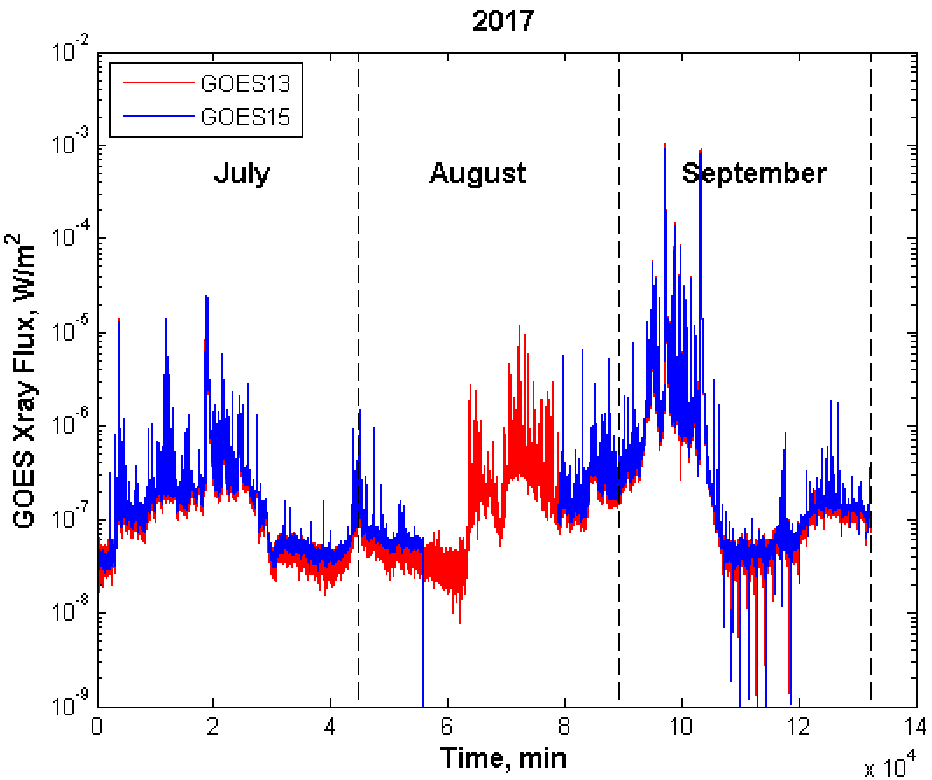
Accepted 2019 March 5. Received 2019 February 13; in original form 2018 December 11

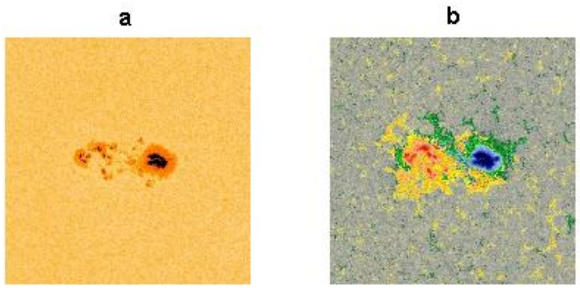
## ABSTRACT

The Sun is variable in activity with changes on time-scales as short as minutes to as long as a solar cycle. Although the most accurate measurements are limited to the satellite era, the past four decades, looking at the solar variability over this period provides a possible link between complex dynamics of the Sun and the accompanying radiation. Measurements of the latter and their analysis by sophisticated time series methods encourage forecasting future values of the time series. Our data analysis work focuses on the soft X-ray emission observed at the current solar minimum, in 2017 September July. We have found two different (active and inactive) states of the solar activity using a Hidden Markov Model, and we show that in the periods of high-solar activity the energy distribution of soft X-ray solar flares is well described by an ARFIMA-GARCH model, whereas in the case of low-solar activity an ARFIMA model is best fitted. Switching from the inactive state to the active one is caused by explosive phenomena in the Sun. The model describes three effects detected in our empirical studies. One of them is a long-term dependence, the second is variance changing in time, and the third corresponds to heavy-tailed distributions of the X-ray data. Moreover, the model takes into account memory effects in soft X-ray emission due to the Sun's magnetic field evolution. All this together allows us to suggest a statistically justified model for explaining the solar activity variability at the current solar minimum.

**Key words:** methods: statistical – space vehicles: instruments – Sun: activity.

# X-ray observations of our star

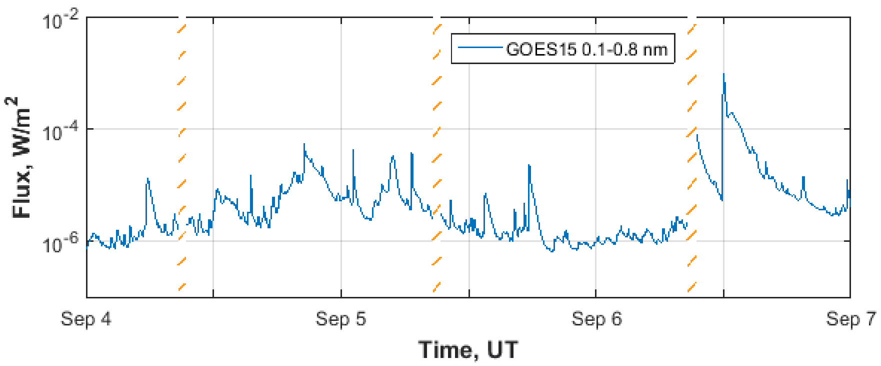


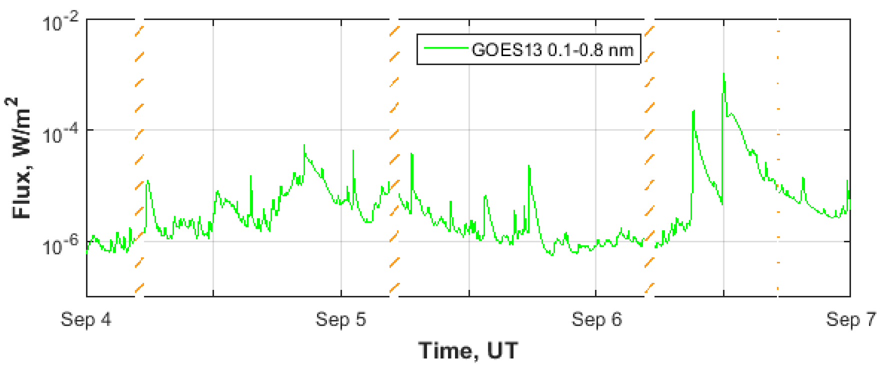




Solar imaginary of Active Region AR12665, provided by the space observatory SDO (Solar Dynamics Observatory):

a) HMI Intensitygram, b) HMI Magnetogram. These images come from the Helioseismic and Magnetic Imager (HMI), an instrument on SDO. The image b) shows the magnetic field directions near the surface of the Sun. Red and blue areas indicate opposite magnetic polarities, with red showing outward polarity and blue showing inward polarity.

# Major X-class solar flare X9.3




An interesting X-class X9.3 solar flare was at 12:02 UTC on September 6, 2017. The event began at 11:53, peak at 12:02 and ended at 12:10 UTC. It was the second X-class solar flare that day. It happened a few hours after the X2.2 flash at 09:33 UTC. The previous record for the strongest cycle flare was X6.9 on August 9, 2011. The hatching in the figure indicates a lack of data.

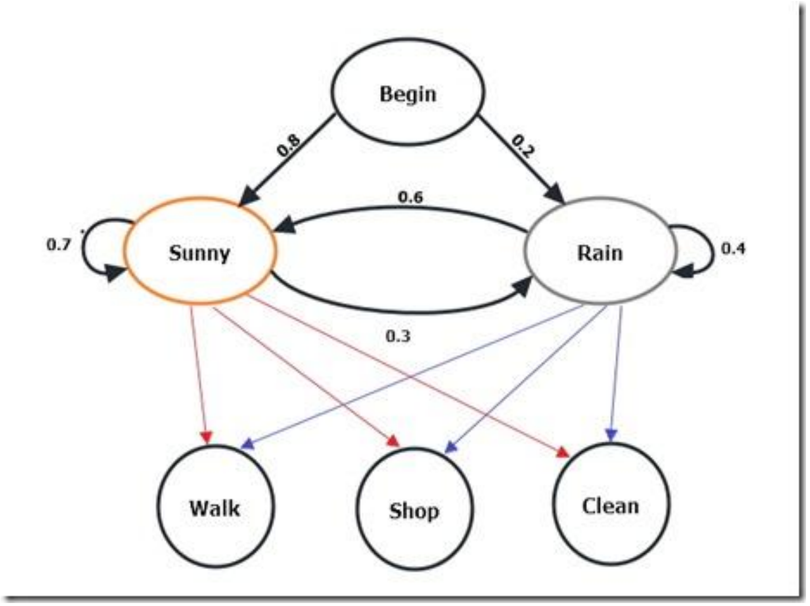
# *ARFIMA* for solar X-ray time series

- Using an *ARFIMA* model estimation directly to the entire time series is problematic, as the parameters *d*, *p*, and *q* are changing.
- But if we take a rolling window of 512 data points and scan the entire segment from beginning to the end, observing how the parameters change, then the parameter estimate becomes a successful procedure.
- From this point of view the best model is (2,*d*,0).
- The third parameter *d* characterizes memory effects caused by the Sun's magnetic field. The memory of the solar cycle plays an important role in predictions because it determines how much of the past history of solar activity determines its future output.

# Hidden Markov Model



**Hidden Markov Model (HMM)** is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (i.e. hidden) states. In the hidden Markov model, the state is not directly visible, but the output (in the form of data), dependent on the state, is visible.




Imagine two friends talking on the phone every night, what they did today during the day. One of them can do only three things: walk in the park, go shopping or clean the room. His choice is based only on the weather, which was at the time of the decision. Another friend know nothing about the weather in the region, where his friend lives, but he can, based on decisions of his friend, try to guess what the weather was. The weather behavior can be represented as a Markov chain, it has two states: sunny or rainy, but the second friend cannot see it himself, therefore, it is hidden from him.

Every day, the first friend makes one of three possible decisions: a walk, shopping or cleaning. The second can know about his decision, so this is an observable value. In general, we get HMM.

The Astrophysical Journal, 791:24 (9pp), 2014 August 10 doi:10.1088/0004-637X/791/1/24


# 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
*Received 2014 March 20; accepted 2014 June 17; published 2014 July 22*

## 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 $\mu$m 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.

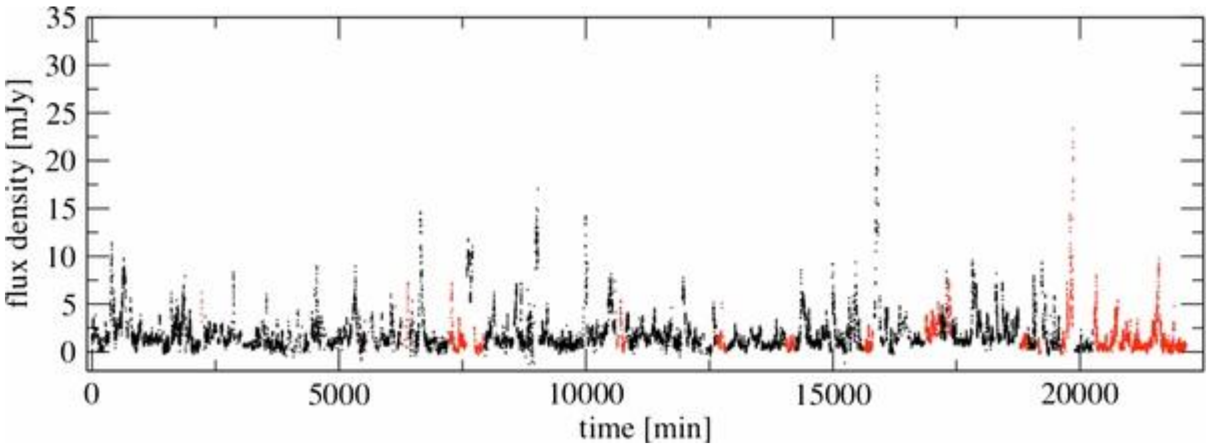
*Key words:* accretion, accretion disks – black hole physics – Galaxy: center – methods: statistical

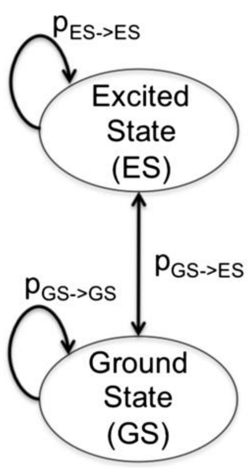
*Online-only material:* color figure

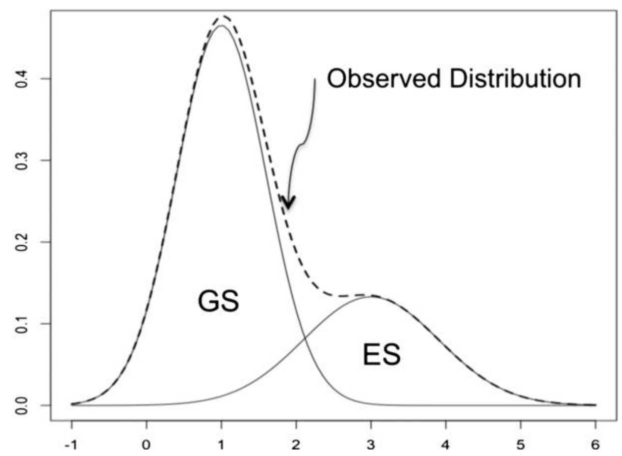
**Sagittarius A*** is a bright and very compact astronomical radio source at the center of the Milky Way, near the border of the constellations Sagittarius and Scorpius.

# Illustration of a two-state hidden Markov model





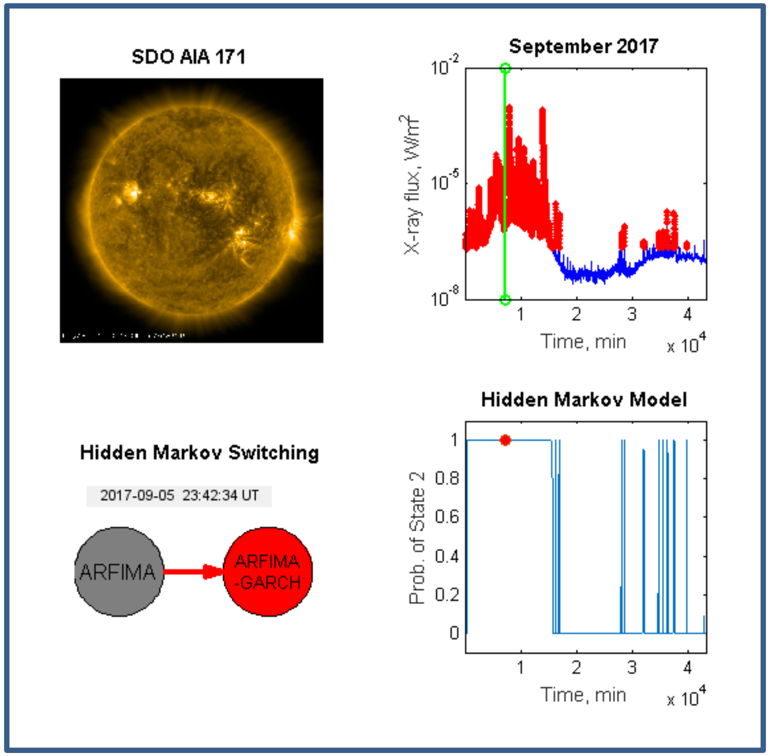



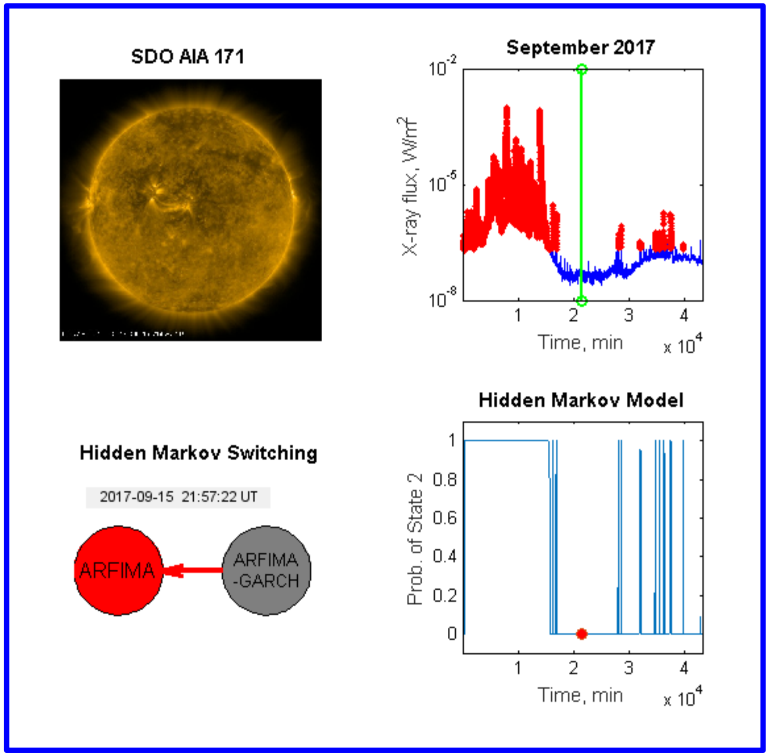




The observed value is presented as the total dotted distribution. It can be decomposed into the ground state (GS) and the excited state (ES), which are not directly observable.

# Two regimes in switching of solar activity in September 2017





# Conclusions

- Soft X-ray emission demonstrates a long-term dependence, variance changing in time, and heavy-tailed distributions.
- Evidence for the existence of regime (state) switching behaviour was found.
- The observed flux densities exhibit two states: background dominated and flares dominated.
- The piecewise ARFIMA and ARFIMA–GARCH models were built separately for both states. In the models the solar X-ray data are considered as a non-stationary time series, whereas the short intervals may be stationary, but their duration is random.
- Our findings were confirmed by rigorous ARFIMA and ARFIMA–GARCH residual diagnostics which shown that for most of the case a plain ARFIMA model is not enough.
- In the framework of ARFIMA and ARFIMA–GARCH models the evolution of their parameters is connected with changes of solar activity (in X-ray emission). The correlations are very noticeable.