# Introduction to Bayesian Statistics - 12

*Edoardo Milotti*

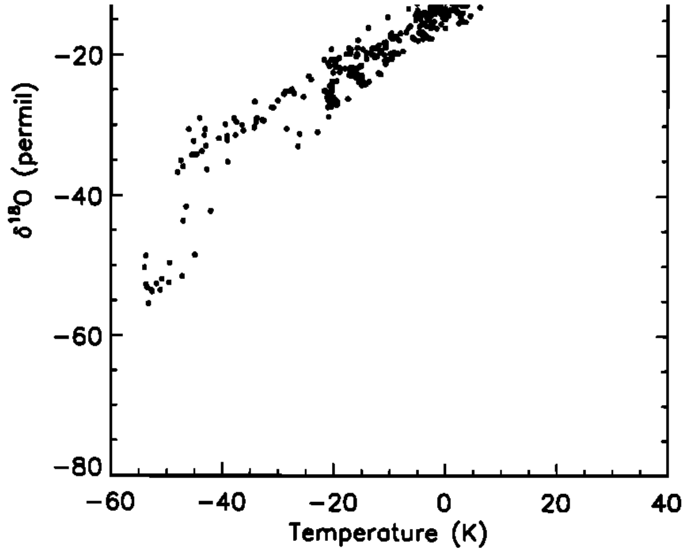
Università di Trieste and INFN-Sezione di Trieste

# The oxygen isotope ratio.

The oxygen isotope ratio is the primary method used to determine past temperatures from ice cores. Because isotopes have a different number of neutrons, they have different mass numbers. Oxygen's most common isotope has a mass number of 16 and is written as $^{16}O$. Most of the oxygen in water molecules is composed of 8 protons and 8 neutrons in its nucleus, giving it a mass number (the number of protons and neutrons in an element or isotope) of 16. About one out of every 1,000 oxygen atoms contains 2 additional neutrons and is written as $^{18}O$.

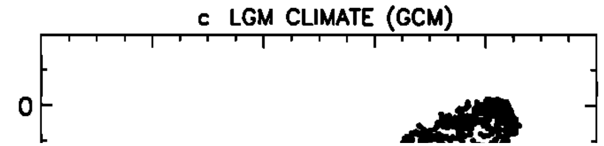
Depending on climate, the two types of oxygen ($^{16}O$ and $^{18}O$) vary in water. Scientists compare the ratio of the heavy ($^{18}O$) and light ($^{16}O$) isotopes in ice cores, sediments, or fossils to reconstruct past climates. They compare this ratio to a standard ratio of oxygen isotopes found in ocean water at a depth of 200 to 500 meters. The ratio of the heavy to light oxygen isotopes is influenced mainly by the processes involved in the water or hydrologic cycle.

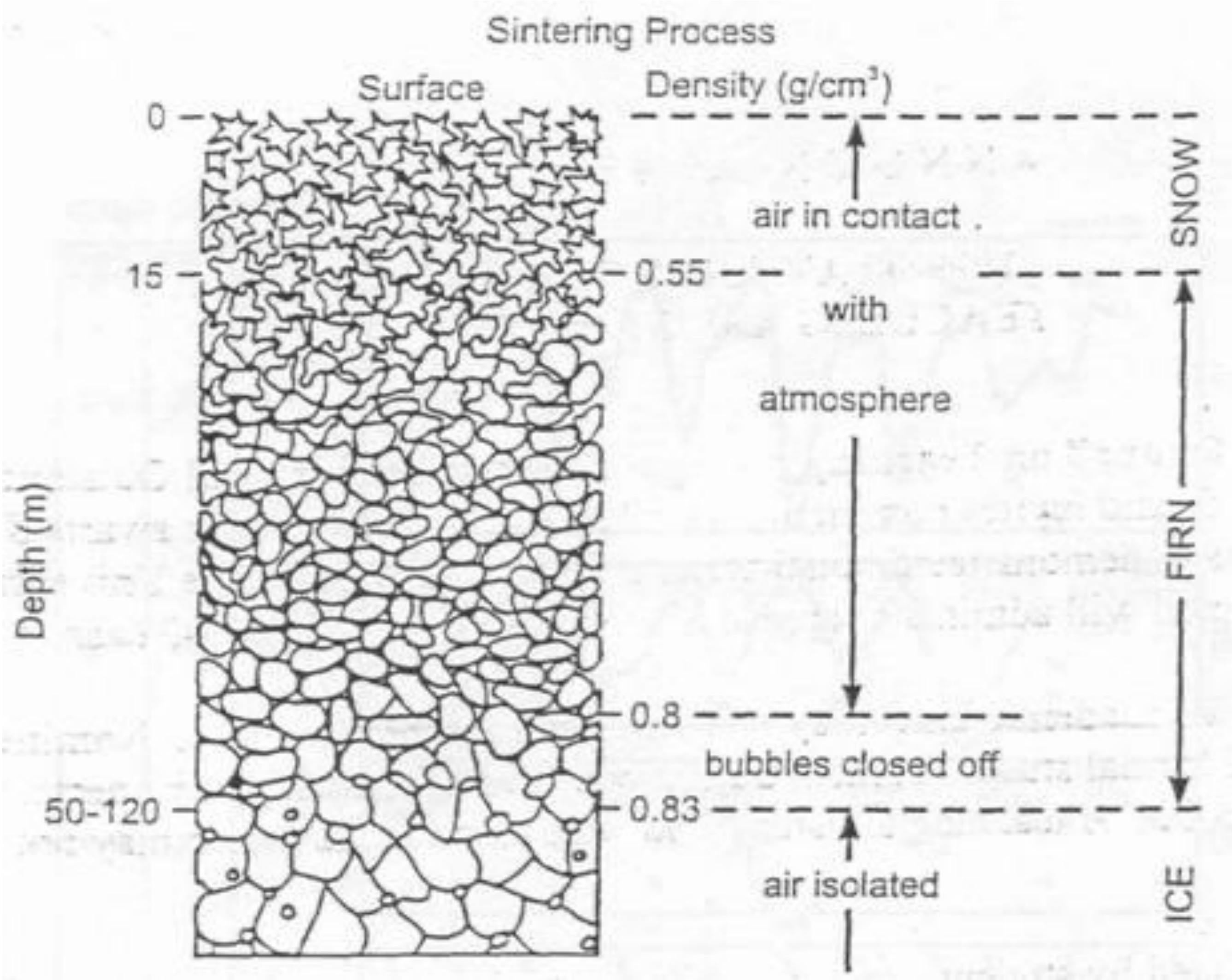
Evaporation and condensation are the two processes that most influence the ratio of $^{18}O$ to $^{16}O$ in the oceans. Water molecules containing $^{16}O$ evaporate slightly more readily than water molecules containing $^{18}O$. At the same time, water vapor molecules containing the $^{18}O$ condense more readily.

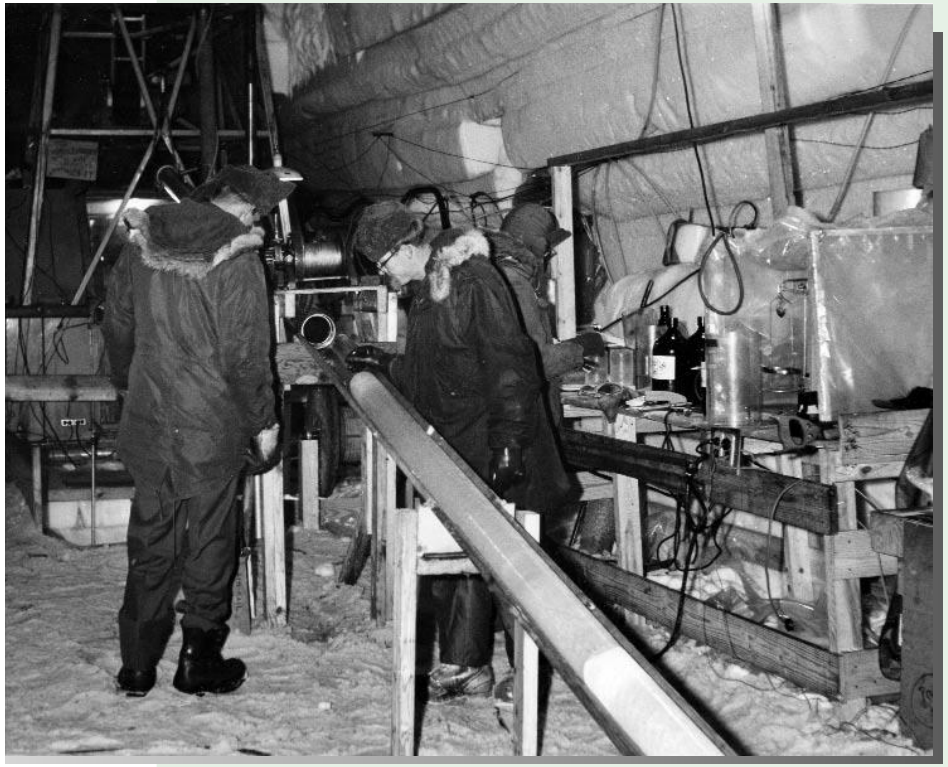



(adapted from https://www.ces.fau.edu/nasa/module-3/how-is-temperature-measured/isotopes.php and https://earthobservatory.nasa.gov/features/Paleoclimatology_OxygenBalance)

The concentration of $^{18}O$ in precipitation decreases with temperature. This graph shows the difference in $^{18}O$ concentration in annual precipitation compared to the average annual temperature at each site. The coldest sites, in locations such as Antartica and Greenland, have about 5 percent less $^{18}O$ than ocean water. (from Jouzel et al., 1994)




Sintering Process
Surface
Density (g/cm³)
0
15
50-120
Depth (m)
0.55
0.8
0.83
air in contact
with
atmosphere
bubbles closed off
air isolated
SNOW
FIRN
ICE



from Weart, Phys. Today 56, 30 (2003)

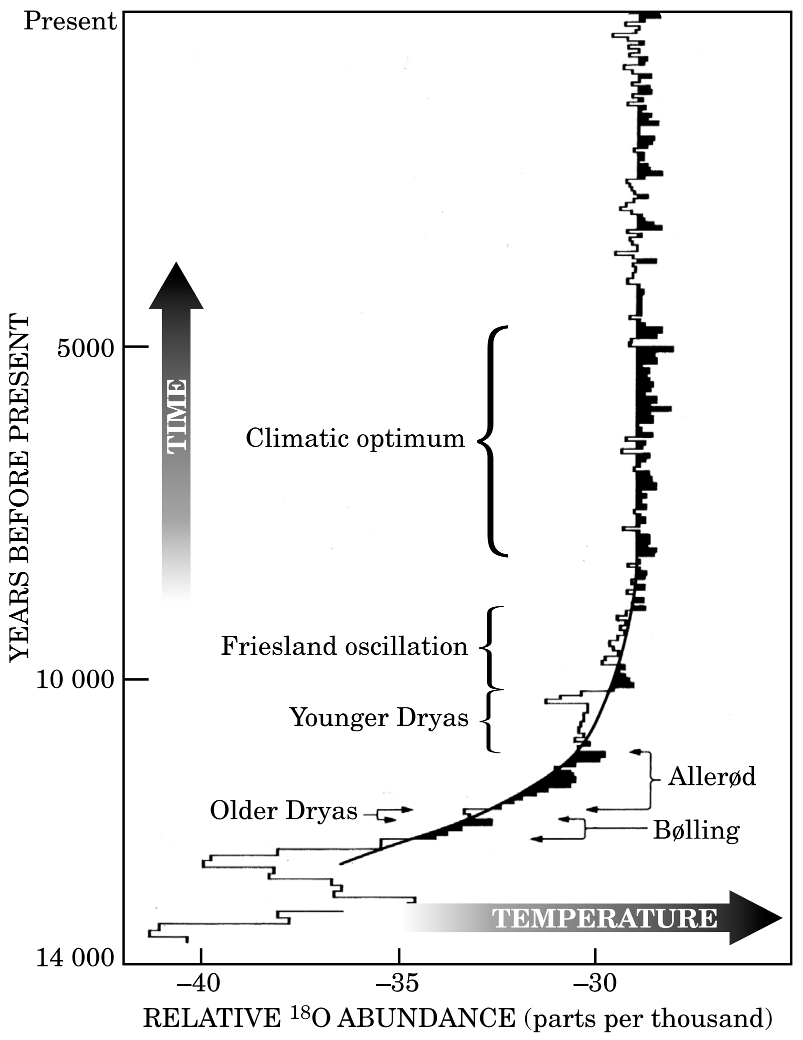
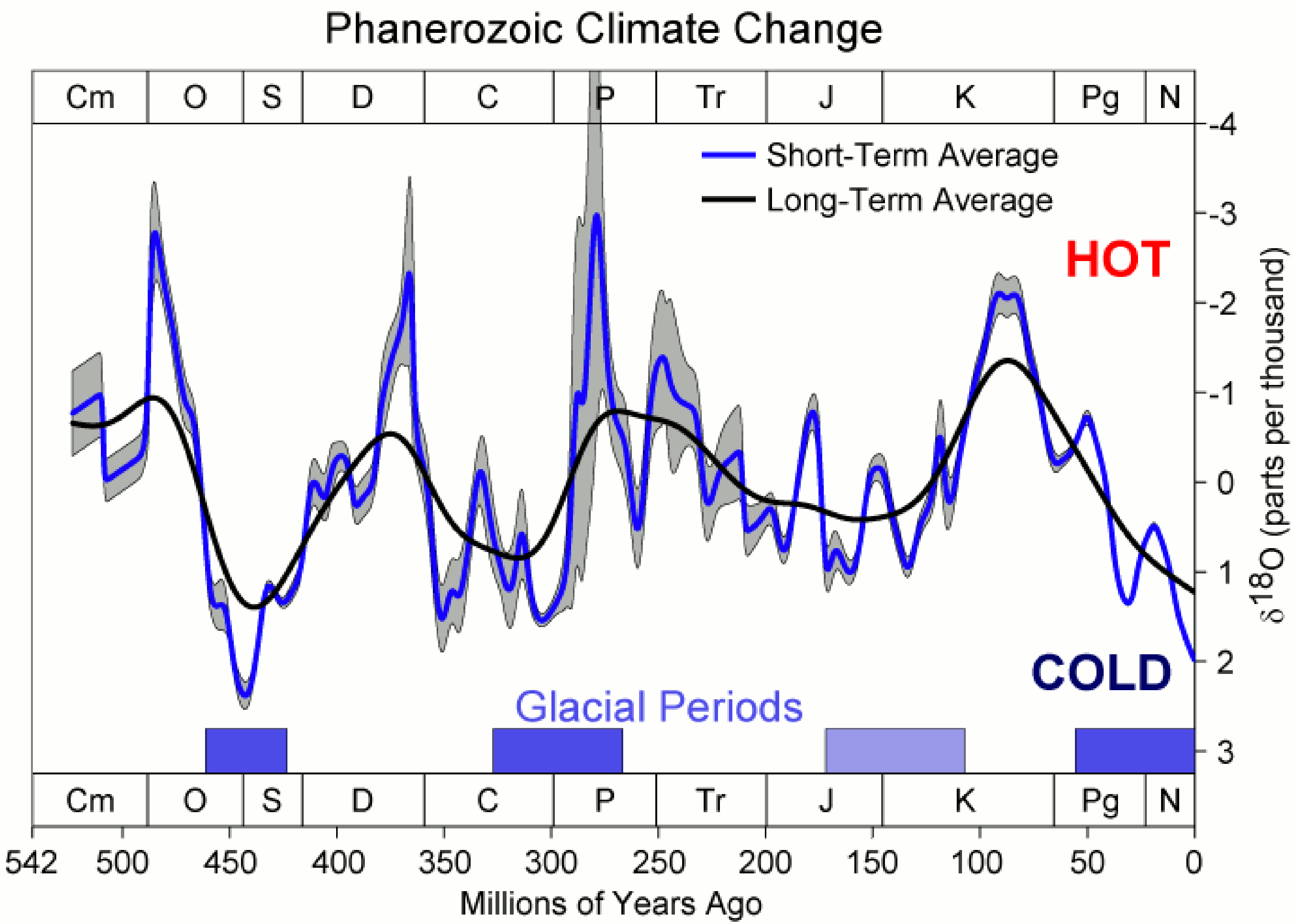
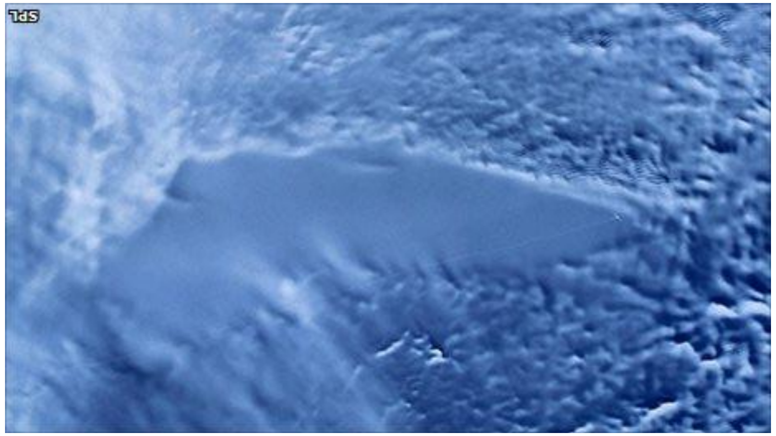



**Figure 3. Cores drilled from the ice** at Camp Century, Greenland, and processed on the spot in 1964 (see photo), revealed ancient climate changes in unprecedented detail. The ratio of oxygen-18 to oxygen-16 isotopes in the annual snow layers serves as a thermometer, as shown in the plot: Part per thousand variations to the right indicate warmer temperatures; those to the left, cooler ones. The large rise in temperature started about 14 000 years ago at the end of the last ice age. The plot also shows 1–2°C temperature leaps even within the one-century resolution of the data, but the authors of the 1971 report barely mentioned them in passing. Their concern was the cycles lasting a few centuries or more, which were remarkable enough. Since only a single site was sampled, none of the changes could confidently be called global, and the leaps could have been artifacts due to flow of the deep ice layers. (Photo by David Atwood, courtesy of US Army-ERDC-Cold Regions Research and Engineering Laboratory. Graph adapted from ref. 18.)

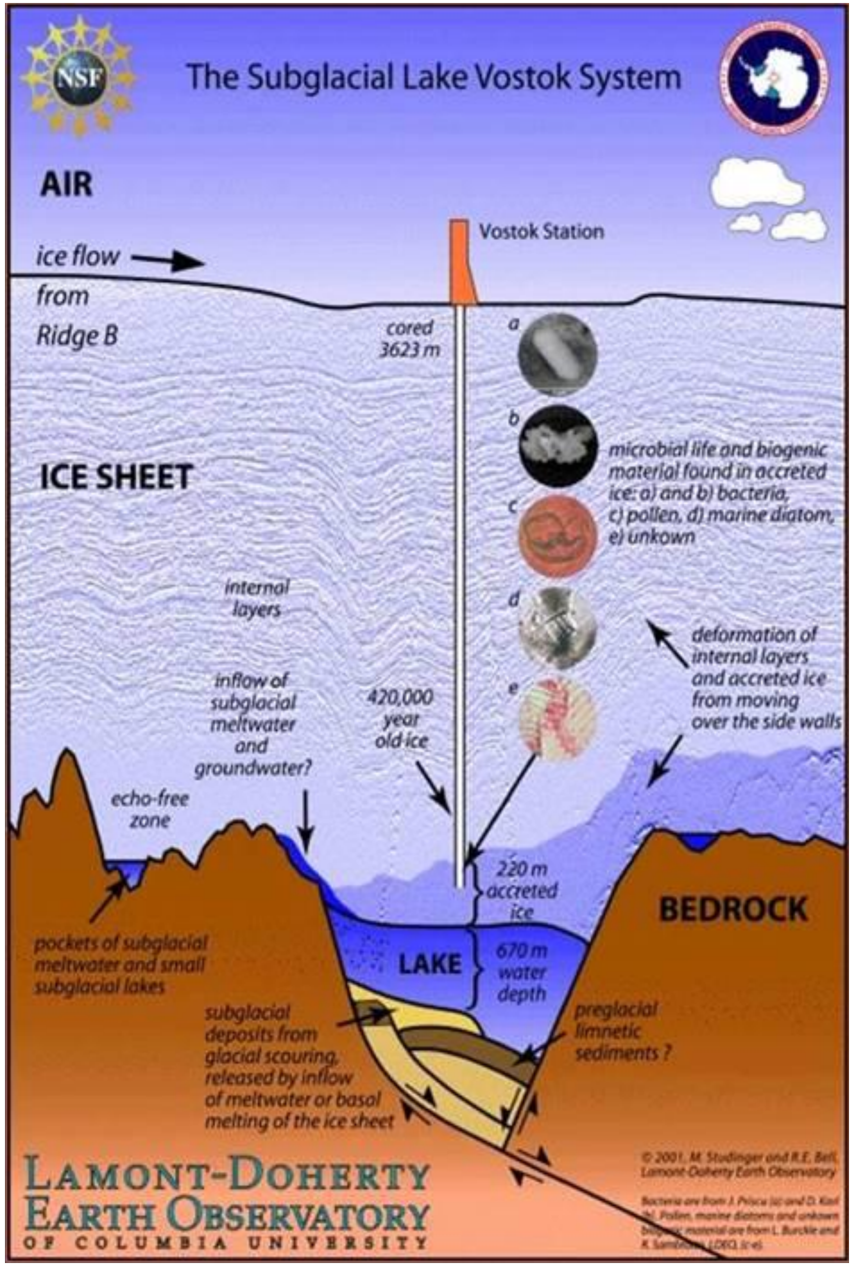
08 May 2024 17:49:43


Phanerozoic Climate Change
Cm
O
S
D
C
P
Tr
J
K
Pg
N
Short-Term Average
Long-Term Average
HOT
COLD
Glacial Periods
-4
-3
-2
-1
0
1
2
3
δ18O (parts per thousand)
542
500
450
400
350
300
250
200
150
100
50
0
Millions of Years Ago

# Ice core data from lake Vostok in Antarctica



Radar satellite image of lake Vostok




Lake Vostok is the largest of Antarctica's 675 known subglacial lakes.

Lake Vostok is located at the southern Pole of Cold, beneath Russia's Vostok Station under the surface of the central East Antarctic Ice Sheet, which is at 3,488 m above mean sea level.

The surface of this freshwater lake is approximately 4,000 m under the surface of the ice, which places it at approximately 500 m below sea level.

Measuring 250 km long by 50 km wide at its widest point, it covers an area of 12,500 km$^2$ making it the 16th largest lake by surface area.

With an average depth of 432 m, it has an estimated volume of 5,400 km$^3$, making it the 6th largest lake by volume.

The lake is divided into two deep basins by a ridge. The liquid water depth over the ridge is about 200 m, compared to roughly 400 m deep in the northern basin and 800 m deep in the southern.

(from Wikipedia, https://en.wikipedia.org/wiki/Lake_Vostok )

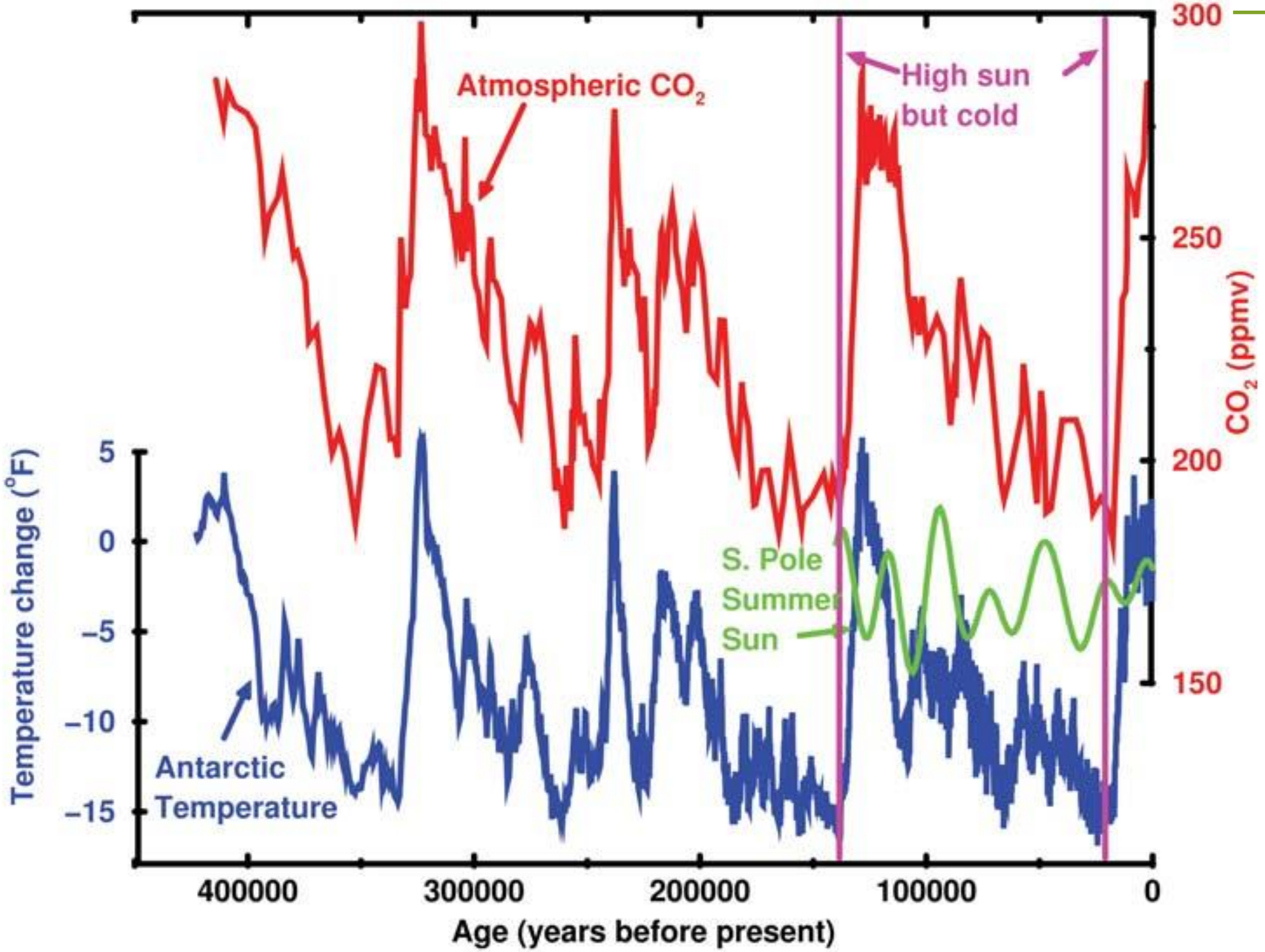
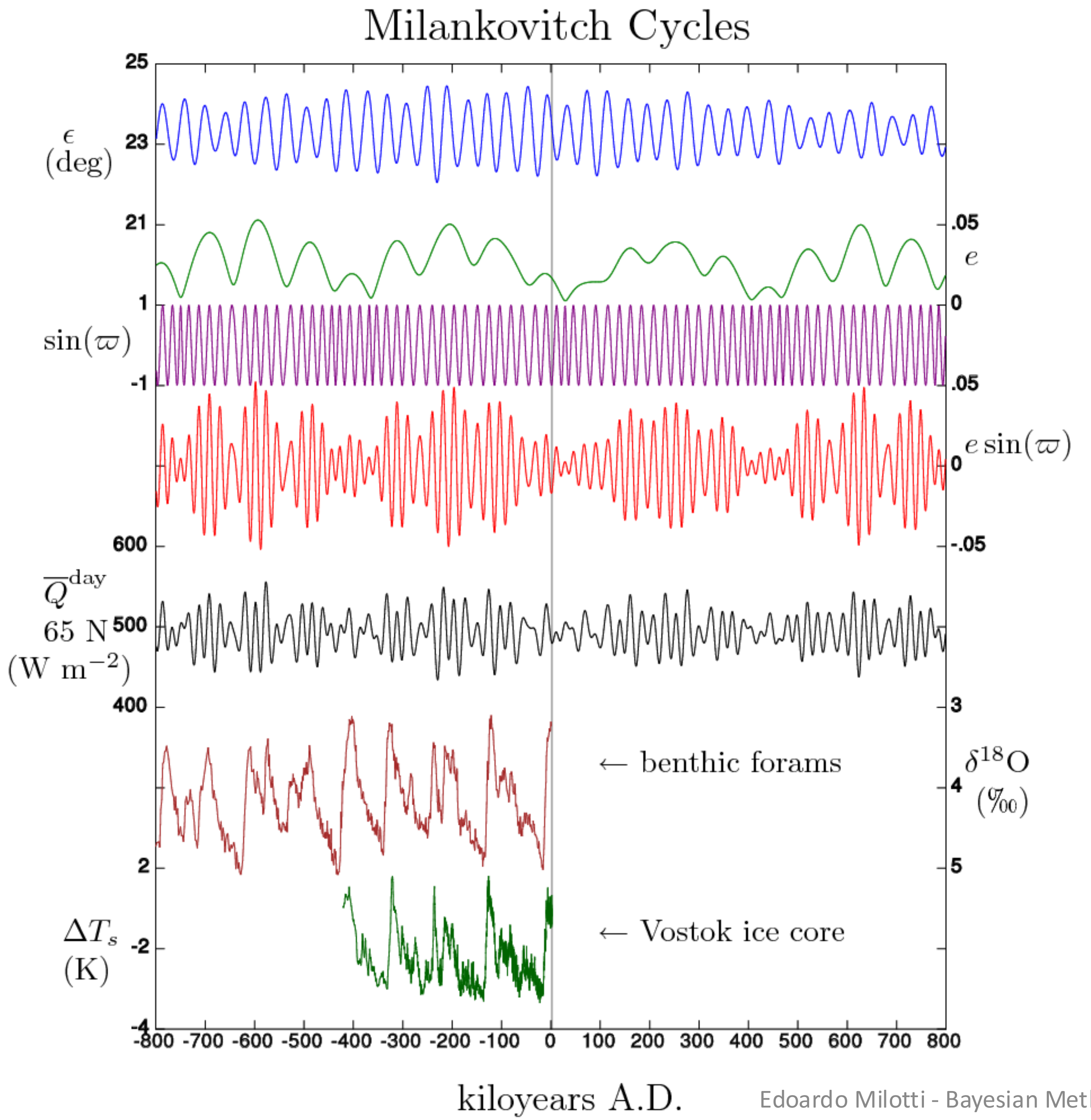



Figure 2. The last 440,000 years of climate in central East Antarctica, from the Vostok ice core. Today is on the right, and 440,000 years ago on the left. The lower curve shows the history of temperature estimated from the isotopic composition of the ice. The large, approximately 100,000-year cycle of ice ages is evident. This basic pattern is also evident in most climate records obtained from anywhere on Earth. Also shown is the variation in local sunshine in Antarctica over the best-dated and more recent part of the record, calculated from knowledge of orbital physics. Peaks in Antarctic sunshine are spaced about 20,000 years apart, and occur when northern sunshine was especially low, including the Antarctic peak in sunshine about 20,000 years ago when Antarctica was especially cold. The only explanation of this behavior that "works" is that the carbon-dioxide concentration of the atmosphere followed northern sunshine, as shown by the upper curve and that, in turn, carbon dioxide was more important for southern temperature than was southern sunshine.

from Alley, 2004

Milankovitch cycles describe the collective effects of changes in the Earth's movements on its climate over thousands of years. The term was coined and named after the Serbian geophysicist and astronomer Milutin Milanković.

In the 1920s, he hypothesized that variations in **eccentricity, axial tilt, and precession** combined to result in cyclical variations in the intra-annual and latitudinal distribution of solar radiation at the Earth's surface, and that this orbital forcing strongly influenced the Earth's climatic patterns.



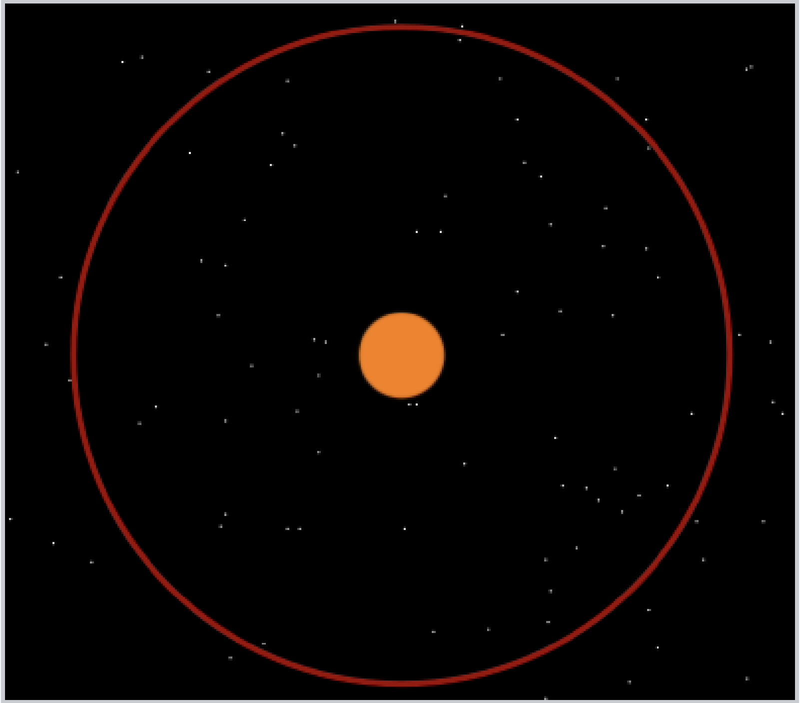

The semi-analytic planetary theory VSOP (French: Variations Séculaires des Orbites Planétaires) is a mathematical model describing long-term changes (secular variation) in the orbits of the planets Mercury to Neptune. The earliest modern scientific model considered only the gravitational attraction between the Sun and each planet, with the resulting orbits being unvarying Keplerian ellipses. In reality, all the planets exert slight forces on each other, causing slow changes in the shape and orientation of these ellipses. Increasingly complex analytical models have been made of these deviations, as well as efficient and accurate numerical approximation methods.

VSOP was developed and is maintained (updated with the latest data) by the scientists at the Bureau des Longitudes in Paris. The first version, VSOP82, computed only the orbital elements at any moment. An updated version, VSOP87, computed the positions of the planets directly at any moment, as well as their orbital elements with improved accuracy.

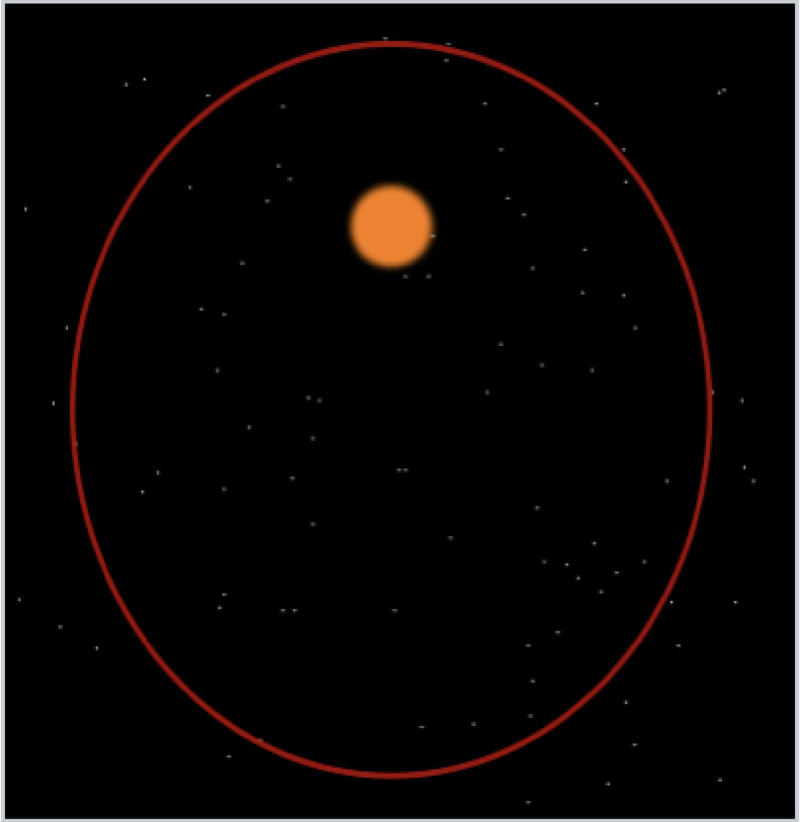
(https://en.wikipedia.org/wiki/VSOP_model)

Past and future Milankovitch cycles via VSOP model

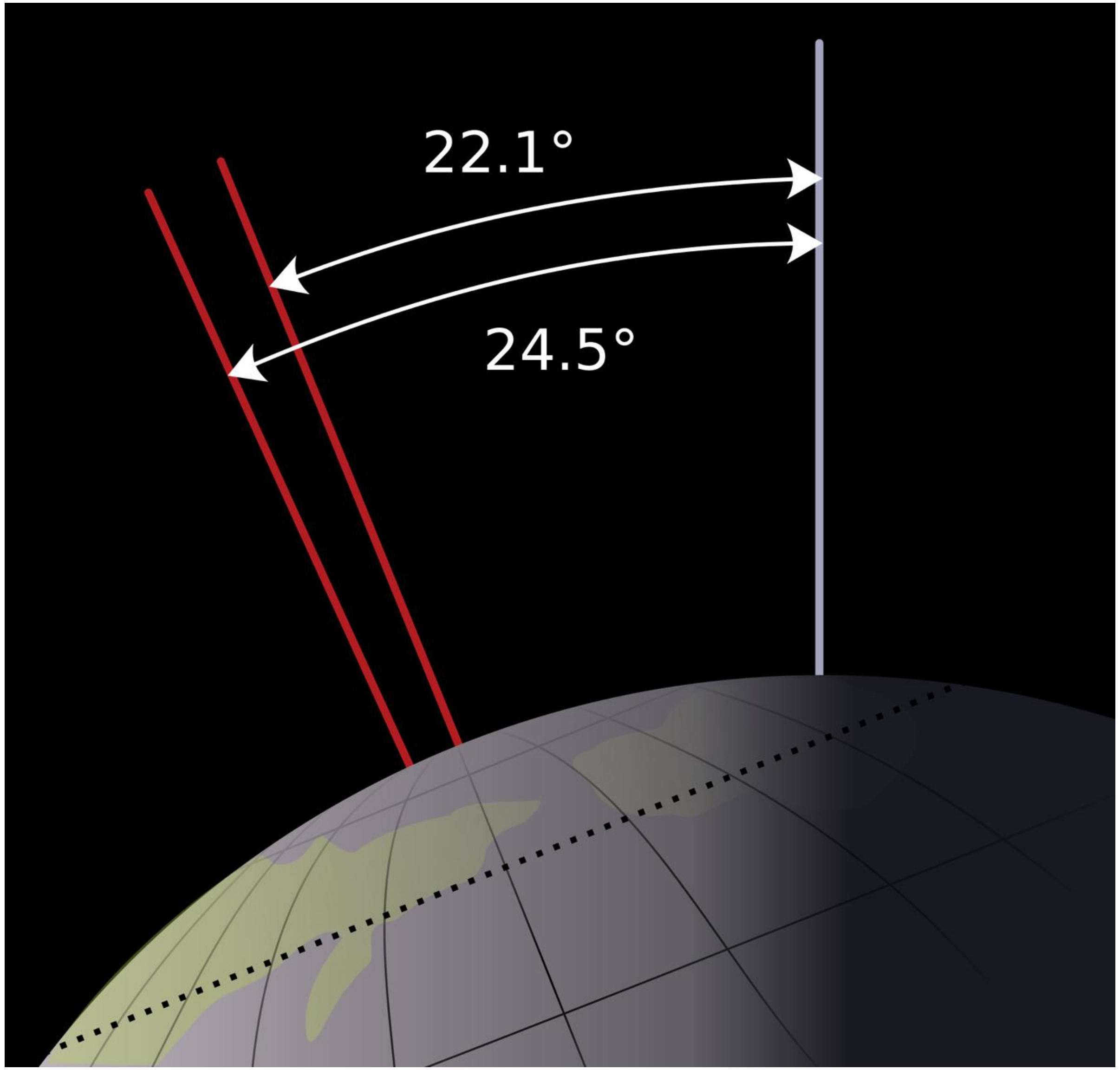
- Graphic shows variations in five orbital elements:
  - Axial tilt or obliquity (ε).
  - Eccentricity (*e*).
  - Longitude of perihelion ($\sin(\varpi)$).
  - Precession index ($e\sin(\varpi)$)
- Precession index and obliquity control insolation at each latitude:
  - Daily-average insolation at top of atmosphere on summer solstice ( $\overline{Q}^{\text{day}}$ ) at 65° N
- Ocean sediment and Antarctic ice strata record ancient sea levels and temperatures:
  - Benthic forams (57 widespread locations)
  - Vostok ice core (Antarctica)
- Vertical gray line shows present (2000 CE)



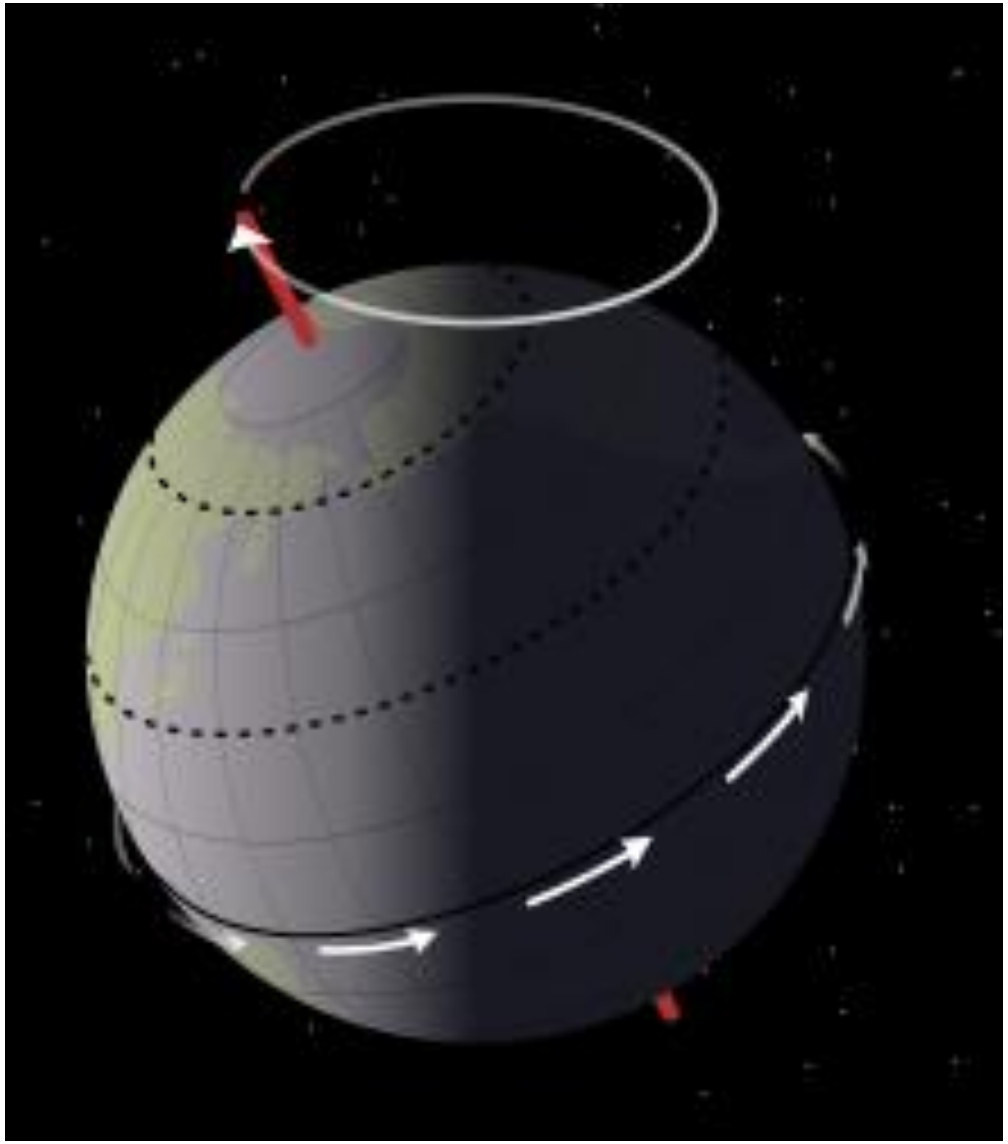
Circular orbit, no eccentricity



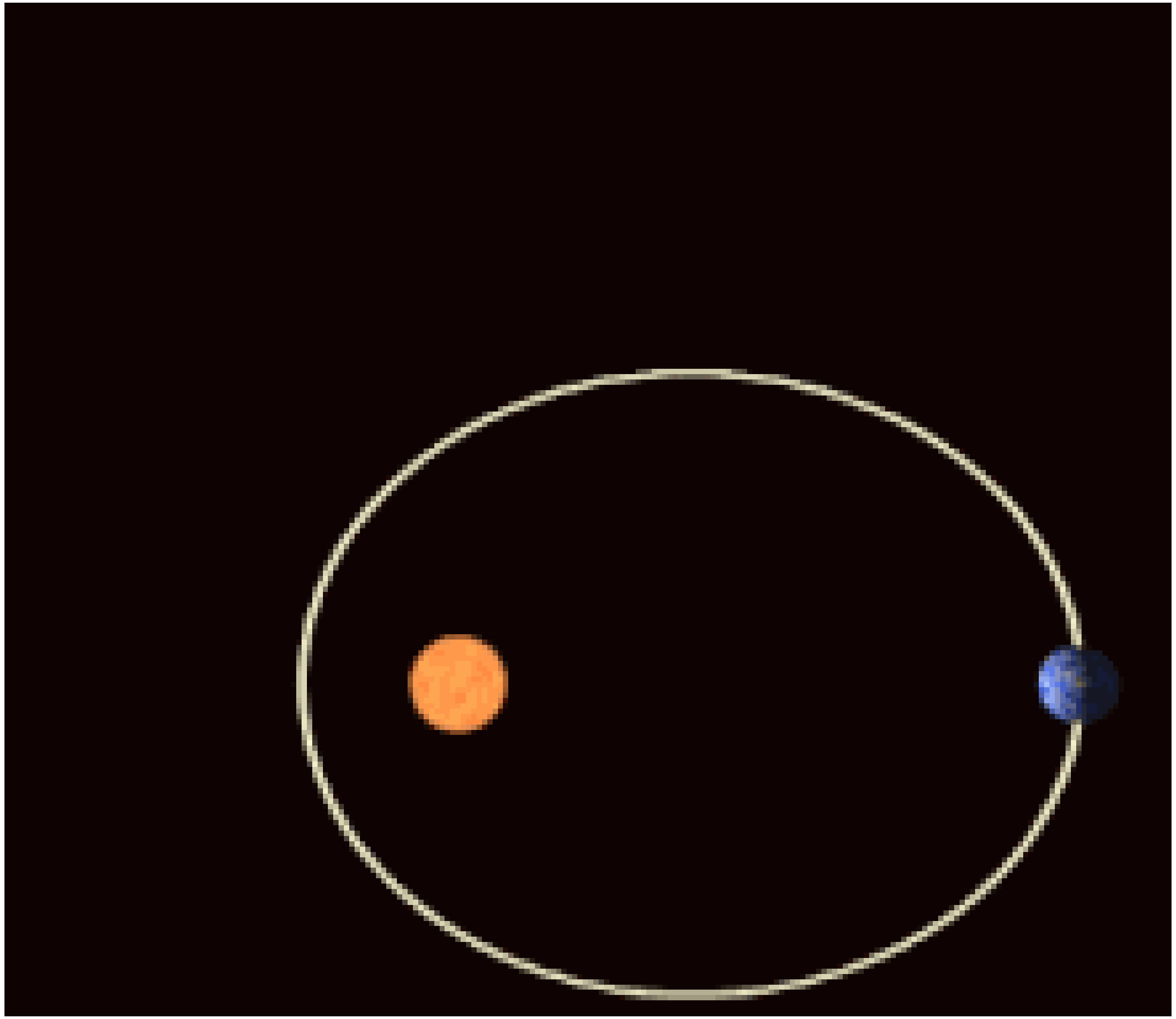
Orbit with 0.5 eccentricity, exaggerated for illustration; Earth's orbit is only slightly eccentric



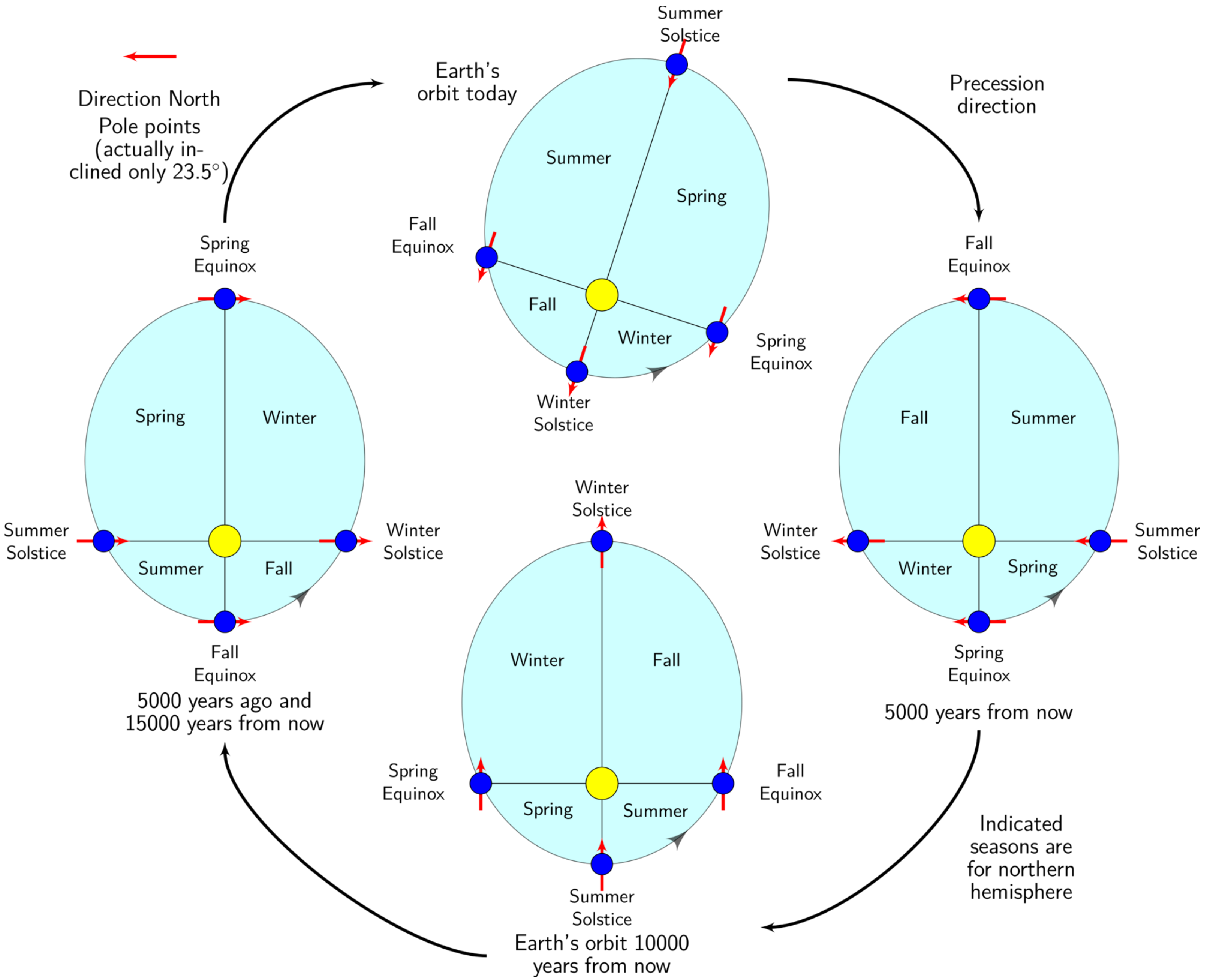

22.1–24.5° range of Earth's obliquity.

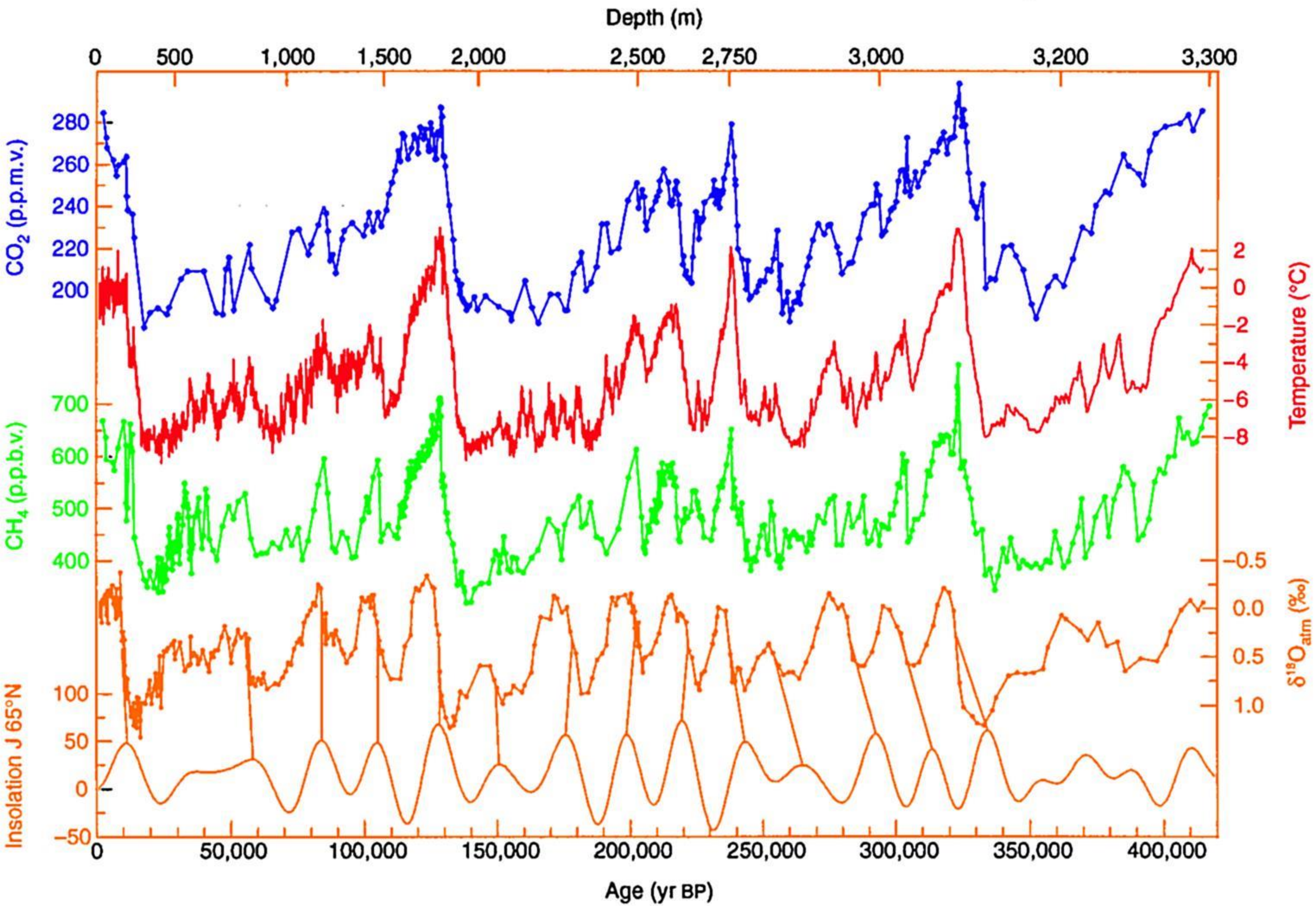


Axial precessional movement.



Planets orbiting the Sun follow elliptical (oval) orbits that rotate gradually over time (apsidal precession). The eccentricity of this ellipse, as well as the rate of precession, are exaggerated for visualization


Direction North
Pole points
(actually in-
clined only 23.5°)
Earth's
orbit today
Summer
Solstice
Summer
Spring
Fall
Equinox
Fall
Winter
Spring
Equinox
Winter
Solstice
Precession
direction
Spring
Equinox
Spring
Winter
Summer
Solstice
Summer
Fall
Winter
Solstice
Fall
Equinox
5000 years ago and
15000 years from now
Fall
Equinox
Fall
Summer
Winter
Solstice
Winter
Spring
Summer
Solstice
Spring
Equinox
5000 years from now
Winter
Solstice
Winter
Fall
Spring
Equinox
Spring
Summer
Fall
Equinox
Summer
Solstice
Earth's orbit 10000
years from now
Indicated
seasons are
for northern
hemisphere


Depth (m)
0
500
1,000
1,500
2,000
2,500
2,750
3,000
3,200
3,300
CO2 (p.p.m.v.)
280
260
240
220
200
Temperature (°C)
2
0
−2
−4
−6
−8
CH4 (p.p.b.v.)
700
600
500
400
δ18Oatm (‰)
−0.5
0.0
0.5
1.0
Insolation J 65°N
100
50
0
−50
0
50,000
100,000
150,000
200,000
250,000
300,000
350,000
400,000
Age (yr BP)

# Lomb-Scargle periodograms

$$S(\omega) = \frac{1}{N^2} \left| \sum_n s_n e^{-i\omega t_n} \right|^2$$

**Schuster periodogram**: rough approximation of the power spectral density for unequally spaced data

$$s_n^{(\mathrm{th})} = a \sin \omega t_n + b \cos \omega t_n$$

$$S = \sum_n \left( s_n - s_n^{(\mathrm{th})} \right)^2$$

**Lomb-Scargle method**: spectral estimates equivalent to linear regressions

$a$, $b$ are obtained by minimizing $S$

$$a = \frac{\sum_n s_n \sin \omega t_n \sum_n \cos^2 \omega t_n - \sum_n s_n \cos \omega t_n \sum_n \sin \omega t_n \cos \omega t_n}{\sum_n \sin^2 \omega t_n \sum_n \cos^2 \omega t_n - \left(\sum_n \sin \omega t_n \cos \omega t_n\right)^2}$$

$$b = \frac{\sum_n s_n \cos \omega t_n \sum_n \sin^2 \omega t_n - \sum_n s_n \sin \omega t_n \sum_n \sin \omega t_n \cos \omega t_n}{\sum_n \sin^2 \omega t_n \sum_n \cos^2 \omega t_n - \left(\sum_n \sin \omega t_n \cos \omega t_n\right)^2}$$

*Astrophysics and Space Science* **39** (1976) 447–462.

**LEAST-SQUARES FREQUENCY ANALYSIS OF UNEQUALLY SPACED DATA**

N. R. LOMB
*School of Physics, University of Sydney, N.S.W., Australia*

(Received 15 May, 1975)

**Abstract.** The statistical properties of least-squares frequency analysis of unequally spaced data are examined. It is shown that, in the least-squares spectrum of gaussian noise, the reduction in the sum of squares at a particular frequency is a $\chi_2^2$ variabl.. The reductions at different frequencies are not independent, as there is a correlation between the height of the spectrum at any two frequencies, $f_1$ and $f_2$, which is equal to the mean height of the spectrum due to a sinusoidal signal of frequency $f_1$, at the frequency $f_2$. These correlations reduce the distortion in the spectrum of a signal affected by noise. Some numerical illustrations of the properties of least-squares frequency spectra are also given.

THE ASTROPHYSICAL JOURNAL, **343**:874–887, 1989 August 15


STUDIES IN ASTRONOMICAL TIME SERIES ANALYSIS. III. FOURIER TRANSFORMS, AUTOCORRELATION FUNCTIONS, AND CROSS-CORRELATION FUNCTIONS OF UNEVENLY SPACED DATA

JEFFREY D. SCARGLE
Theoretical Studies Branch, Space Science Division, NASA-Ames Research Center
*Received 1988 August 8; accepted 1989 January 24*

ABSTRACT

This paper develops techniques to evaluate the discrete Fourier transform (DFT), the autocorrelation function (ACF), and the cross-correlation function (CCF) of time series which are not evenly sampled. The series may consist of quantized point data (e.g., yes/no processes such as photon arrival). The DFT, which can be inverted to recover the original data and the sampling, is used to compute correlation functions by means of a procedure which is effectively, but not explicitly, an interpolation. The CCF can be computed for two time series not even sampled at the same set of times. Techniques for removing the distortion of the correlation functions caused by the sampling, determining the value of a constant component to the data, and treating unequally weighted data are also discussed. FORTRAN code for the Fourier transform algorithm and numerical examples of the techniques are given.

Demonstration Jupyter notebook and material at these links

https://github.com/prappleizer/prappleizer.github.io/tree/master/Tutorials/MCMC
https://prappleizer.github.io/Tutorials/MCMC/MCMC_Tutorial.html

… unfortunately, full of unsolved issues …

Is **`emcee`** the ultimate MCMC? Certainly not!

*Aust. N. Z. J. Stat.* **64**(1), 2022, 1–26 doi: 10.1111/anzs.12358

# Properties of the affine-invariant ensemble sampler's 'stretch move' in high dimensions

David Huijser[1*], Jesse Goodman[1] and Brendon J. Brewer[1]

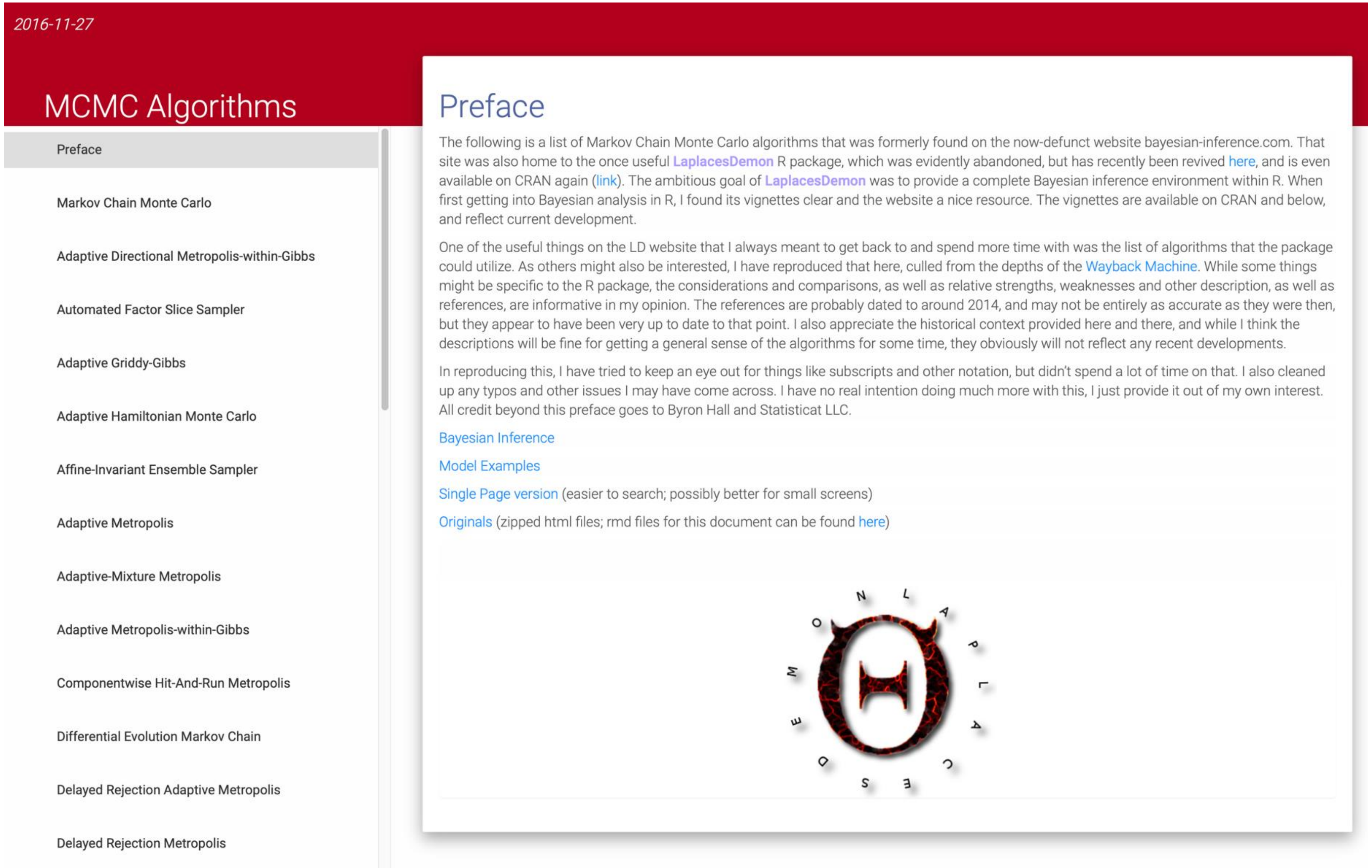
*Department of Statistics, University of Auckland*

**Summary**

We present theoretical and practical properties of the affine-invariant ensemble sampler Markov Chain Monte Carlo method. In high dimensions, the sampler's 'stretch move' has unusual and undesirable properties. We demonstrate this with an $n$-dimensional correlated Gaussian toy problem with a known mean and covariance structure, and a multivariate version of the Rosenbrock problem. Visual inspection of a trace plots suggests the burn-in period is short. Upon closer inspection, we discover the mean and the variance of the target distribution do not match the known values, and the chain takes a very long time to converge. This problem becomes severe as $n$ increases beyond 50. We also applied different diagnostics adapted to be applicable to ensemble methods to determine any lack of convergence. The diagnostics include the Gelman–Rubin method, the Heidelberger–Welch test, the integrated autocorrelation and the acceptance rate. The trace plot of individual walkers appears to be useful as well. We therefore conclude that the stretch move should be used with caution in moderate to high dimensions. We also present some heuristic results explaining this behaviour.

# A nice list of MCMC samplers (not quite up to date, however and with many broken links)

https://m-clark.github.io/docs/ld_mcmc/#preface



2016-11-27

MCMC Algorithms

- Preface
- Markov Chain Monte Carlo
- Adaptive Directional Metropolis-within-Gibbs
- Automated Factor Slice Sampler
- Adaptive Griddy-Gibbs
- Adaptive Hamiltonian Monte Carlo
- Affine-Invariant Ensemble Sampler
- Adaptive Metropolis
- Adaptive-Mixture Metropolis
- Adaptive Metropolis-within-Gibbs
- Componentwise Hit-And-Run Metropolis
- Differential Evolution Markov Chain
- Delayed Rejection Adaptive Metropolis
- Delayed Rejection Metropolis

## Preface

The following is a list of Markov Chain Monte Carlo algorithms that was formerly found on the now-defunct website bayesian-inference.com. That site was also home to the once useful LaplacesDemon R package, which was evidently abandoned, but has recently been revived here, and is even available on CRAN again (link). The ambitious goal of LaplacesDemon was to provide a complete Bayesian inference environment within R. When first getting into Bayesian analysis in R, I found its vignettes clear and the website a nice resource. The vignettes are available on CRAN and below, and reflect current development.

One of the useful things on the LD website that I always meant to get back to and spend more time with was the list of algorithms that the package could utilize. As others might also be interested, I have reproduced that here, culled from the depths of the Wayback Machine. While some things might be specific to the R package, the considerations and comparisons, as well as relative strengths, weaknesses and other description, as well as references, are informative in my opinion. The references are probably dated to around 2014, and may not be entirely as accurate as they were then, but they appear to have been very up to date to that point. I also appreciate the historical context provided here and there, and while I think the descriptions will be fine for getting a general sense of the algorithms for some time, they obviously will not reflect any recent developments.

In reproducing this, I have tried to keep an eye out for things like subscripts and other notation, but didn't spend a lot of time on that. I also cleaned up any typos and other issues I may have come across. I have no real intention doing much more with this, I just provide it out of my own interest. All credit beyond this preface goes to Byron Hall and Statisticat LLC.

Bayesian Inference

Model Examples

Single Page version (easier to search; possibly better for small screens)

Originals (zipped html files; rmd files for this document can be found here)

Mon. Not. R. Astron. Soc. **403,** 731–747 (2010) doi:10.1111/j.1365-2966.2009.16233.x

# A Bayesian periodogram finds evidence for three planets in 47 Ursae Majoris

Philip C. Gregory[1★] and Debra A. Fischer[2,3★]

[1] *Physics and Astronomy Department, University of British Columbia, 6224 Agricultural Rd., Vancouver, BC V6T 1Z1, Canada*
[2] *Department of Physics and Astronomy, San Francisco State University, San Francisco, CA 9413, USA*
[3] *Department of Astronomy, Yale University, New Haven, CT 06520, USA*

Accepted 2009 December 7. Received 2009 December 3; in original form 2009 August 27

**ABSTRACT**

A Bayesian analysis of 47 Ursae Majoris radial velocity data confirms and refines the properties of two previously reported planets with periods of 1079 and 2325 d. The analysis also provides orbital constraints on an additional long-period planet with a period of ∼10 000 d. The three-planet model is found to be $10^5$ times more probable than the next most probable model which is a two-planet model. The non-linear model fitting is accomplished with a new hybrid Markov chain Monte Carlo (HMCMC) algorithm which incorporates parallel tempering, simulated annealing and genetic crossover operations. Each of these features facilitate the detection of a global minimum in $\chi^2$. By combining all three, the HMCMC greatly increases the probability of realizing this goal. When applied to the Kepler problem, it acts as a powerful multiplanet Kepler periodogram.

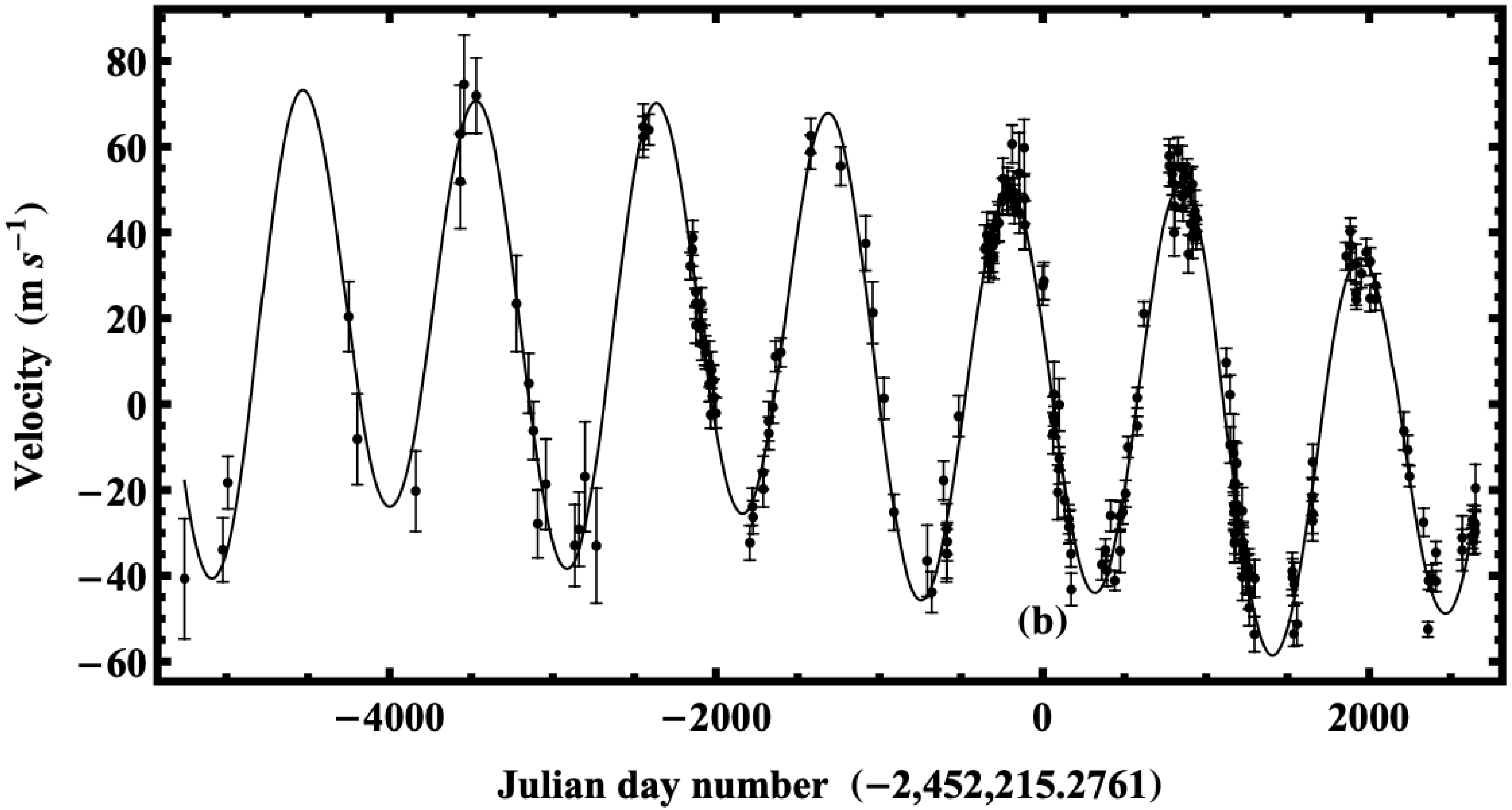
The latest version of the algorithm, Gregory (2009), incorporates a genetic crossover operation into the MCMC algorithm. The new adaptive hybrid MCMC (HMCMC) algorithm incorporates PT, simulated annealing and genetic crossover operations. Each of these techniques was designed to facilitate the detection of a global minimum in $\chi^2$. Combining all three in an adaptive HMCMC greatly increases the probability of realizing this goal.

The adaptive HMCMC is a very general Bayesian non-linear model fitting program. After specifying the non-linear model, data and priors, Bayes theorem dictates the target joint probability distribution for the model parameters which can be very complex. To compute the marginals for any subset of the parameters, it is necessary to integrate the joint probability distribution over the remaining parameters. In high dimensions, the principal tool for carrying out the integrals is MCMC based on the Metropolis algorithm. The greater efficiency of an MCMC stems from its ability, after an initial burn-in period, to generate samples in parameter space in direct proportion to the joint target probability distribution. In contrast, straight Monte Carlo (MC) integration randomly samples the parameter space and wastes most of its time sampling regions of very low probability.

An important feature that prevents the HMCMC from becoming stuck in a local probability maximum is PT. Multiple MCMC chains are run in parallel. The joint probability density distribution for the parameters ($X$) of model $M_i$, for a particular chain, is given by

$$p(X|D, M_i, I, \beta) \propto p(X|M_i, I) \times p(D|X, M_i, I)^\beta. \quad (1)$$

Each MCMC chain corresponding to a different $\beta$, with the value of $\beta$ ranging from 0 to 1. When the exponent $\beta = 1$, the term on the left-hand side of the equation is the target joint probability distribution for the model parameters, $p(\boldsymbol{X}|D, M_i, I)$. It is the posterior probability of a particular choice of parameter vector, $\boldsymbol{X}$, given the data represented by $D$, the model choice $M_i$ and the prior information $I$. In general, the model parameter space of interest is a continuum so $p(\boldsymbol{X}|D, M_i, I)$ is a probability density distribution. The first term on the right-hand side of the equation, $p(\boldsymbol{X}|M_i, I)$, is the prior probability density distribution of $\boldsymbol{X}$, prior to the consideration of the current data $D$. The second term, $p(D|\boldsymbol{X}\ M_i, I)$, is called the likelihood and it is the probability that we would have obtained the measured data $D$ for this particular choice of parameter vector $\boldsymbol{X}$, model $M_i$ and prior information $I$. At the very least, the prior information, $I$, must specify the class of alternative models (hypotheses) being considered (hypothesis space of interest) and the relationship between the models and the data (how to compute the likelihood). For further details of the likelihood function for this problem see Gregory (2005b). In many situations, the log of the likelihood is simply proportional to the familiar $\chi^2$ statistic. If we later acquire another data set $D'$ then the new prior, $p(\boldsymbol{X}|M_i, I')$, is equal to our previous posterior, $p(\boldsymbol{X}|D, M_i, I)$, i.e. $I' = I, D$. An exponent $\beta = 0$, yields a broader joint probability density equal to the prior. The reciprocal of $\beta$ is analogous to a temperature, the higher the temperature the broader the distribution.




best three-planet model fit compared to the data

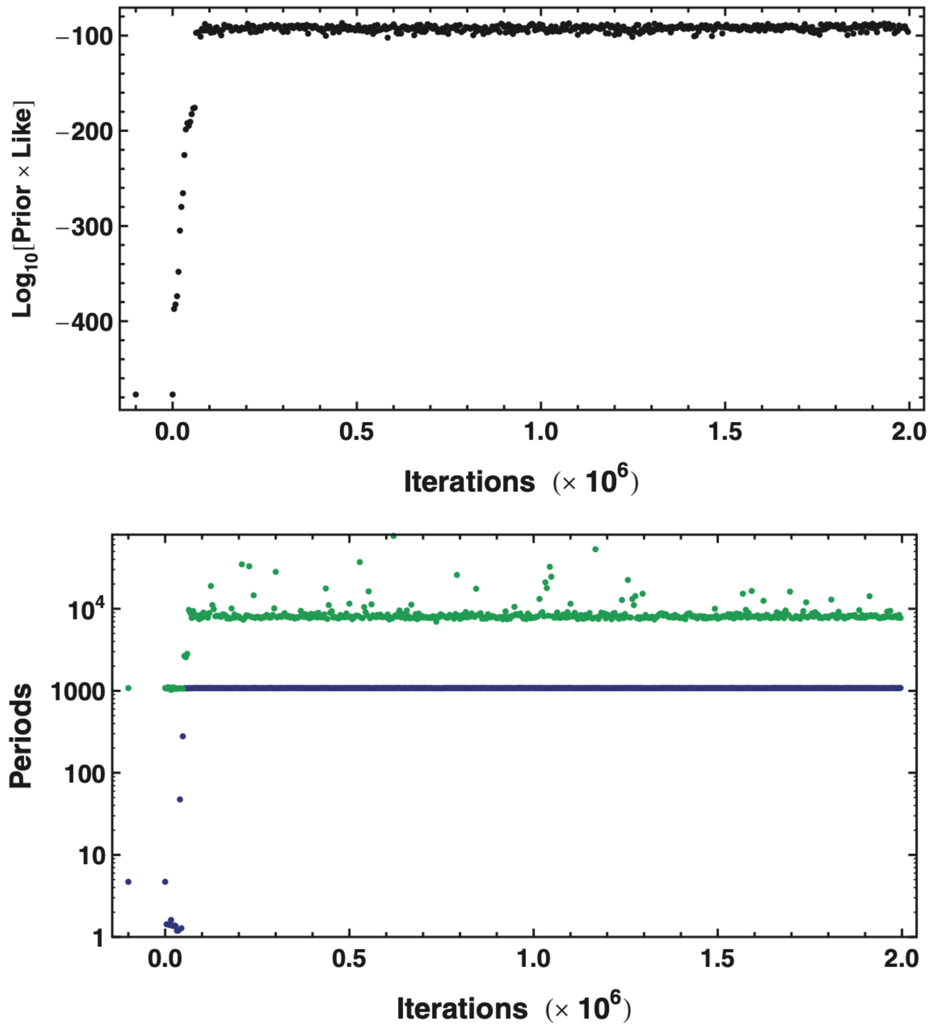



**Figure 7.** Plot of $\text{Log}_{10}$[Prior $\times$ Likelihood] (upper) and period (lower) versus MCMC iteration for a two-planet model.

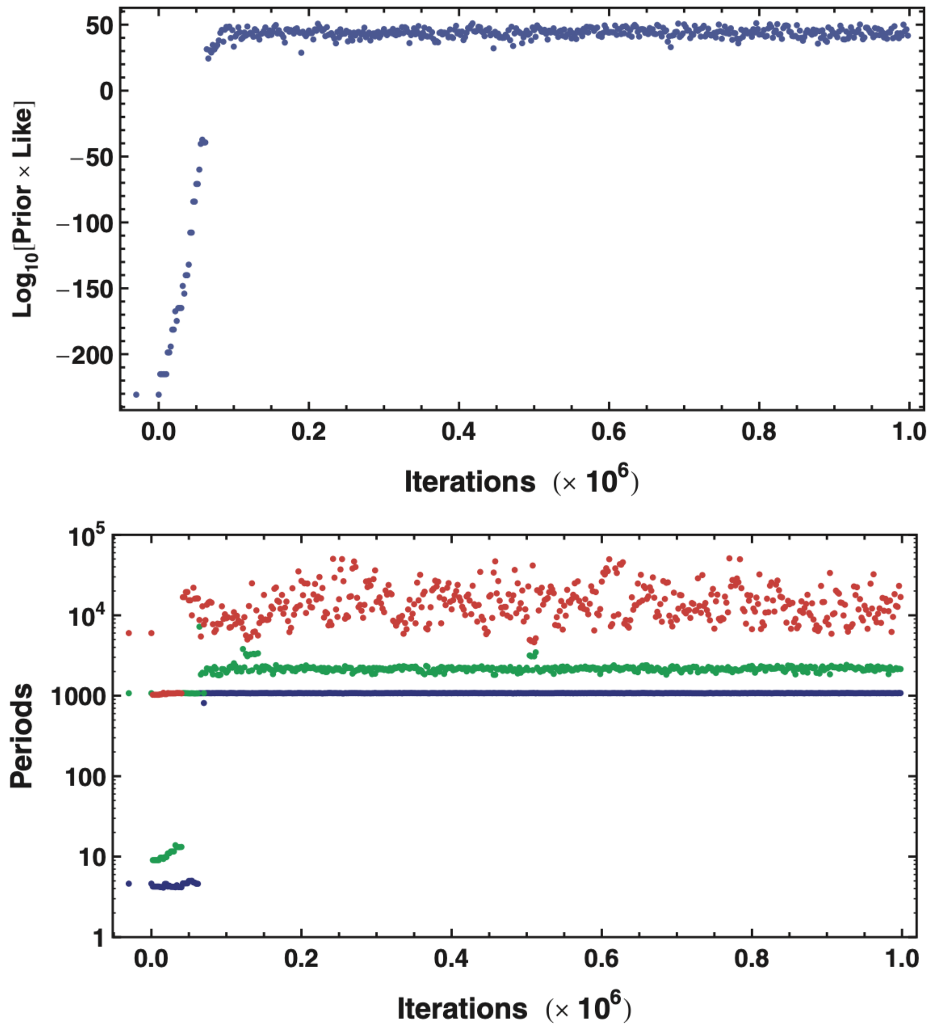



**Figure 9.** Plot of $\text{Log}_{10}$[Prior $\times$ Likelihood] (upper) and period (lower) versus HMCMC iteration for a three-planet fit.