

## Time-variable surface mapping of exoplanets: an application to Io

FRAN BARTOLIĆ,<sup>1,2</sup> RODRIGO LUGER,<sup>1</sup> AND DANIEL FOREMAN-MACKEY<sup>1</sup>

<sup>1</sup>*Center for Computational Astrophysics, Flatiron Institute, New York, NY*

<sup>2</sup>*Centre for Exoplanet Science, SUPA, School of Physics and Astronomy, University of St. Andrews, St. Andrews, UK*

### ABSTRACT

Jupiter’s moon Io is the most volcanically active body in the Solar System with hundreds of active volcanoes varying in intensity on different timescales. Existing photometric observations in the near infrared taken during occultations by Jupiter and Galilean moons encode a wealth of information about its surface. As such, Io is an ideal testbed for models of time-variable exoplanet surfaces. We build a single probabilistic model of Io’s surface using Nonnegative Matrix Factorization to infer one map per light curve. The model is built on top of the code starry  which can compute occultation light curves and phase curves in both emitted and reflected light. It also enables efficient inference with gradient based samplers. We use the inferred map to study the time variability of prominent volcanoes over a timescale of decades. While we mostly focus on Io, the lessons we learn are directly applicable to mapping exoplanets. We discuss this application towards the end of the paper .

### 1. INTRODUCTION

The surface of Jupiter’s moon Io is littered with hundreds of volcanoes appearing as bright spots in the near infrared and varying in intensity on timescales of days to decades. The volcanic activity is driven by tidal interactions with Jupiter and sustained by the Laplace resonance with Europa and Ganymede (?). Io is a perfect analogue of an extremely volcanically active exoplanet. Such rocky exoplanets, often dubbed *super-Ios* or *lava worlds*, have gathered a lot of interest in recent years because the photometric and spectral signatures of volcanic activity on such worlds will likely be detectable in the near future (???). Studying volcanism on both Io and exoplanets is scientifically valuable because it provides a window into the properties of early Earth and it informs theories of planet formation.

Existing exoplanet detections with potential volcanism include CoRoT-7b (?), the first discovered rocky exoplanet with likely significant tidal heating; 55 Cancri e, whose longitudinal offset in peak surface emission has been attributed to, among other

things, lava flows on the surface (??); several planets in the TRAPPIST-1 system in a Laplace-like resonance with likely volcanic activity (??); and many others. Not all exoplanets will have Io-like volcanism, some will have magma oceans due to their proximity to the star or volcanism caused by nuclear decay similar to early Earth volcanism.

Io has been observed extensively using both space and ground observatories. High resolution images of Io's surface were taken by space missions such as Voyager (?), Galileo (?) and Juno (?). It has also been resolved from the ground in the near infrared using disk resolved imaging (???) and adaptive optics observations (???). Most importantly for this work, Io has been sporadically observed for decades using high cadence near infrared photometry taken during occultations by Jupiter starting with ?. Occultations by Jupiter occur twice every orbit lasting 1.7 days and mutual occultations with Europa, Ganymede or Callisto happen every 6 years when Earth passes through the orbital plane of Galilean satellites. The majority of occultations by Jupiter are observed when Io is in Jupiter's shadow ("in eclipse") while mutual occultations are almost always observed in sunlight because mutual occultations while Io is in eclipse are extraordinarily rare. Only the brightest volcanoes are visible over the reflected light background when Io is illuminated by the Sun(??).

Both kinds of occultations have been used to study volcanic activity on the surface. ? observed several occultations of Io by Europa and detected a major brightening of the most powerful of Io's volcanoes, Loki, relative to previous Voyager observations. ??? used observations from NASA's IRTF telescope to study the long term variability of different volcanic spots finding evidence of periodicity in Loki's eruptions and established the transient nature of the volcanoes (only Loki is persistently active). By studying an occultation of Io by Europa, ? mapped the Loki region to a precision of about 2 kilometers. In addition to observations of occultations in near infrared, several groups have been observing mutual occultations in the optical (???, and references therein) with the purpose of inferring the optical albedo of Io and improving ephemeris precision for Galilean satellites.

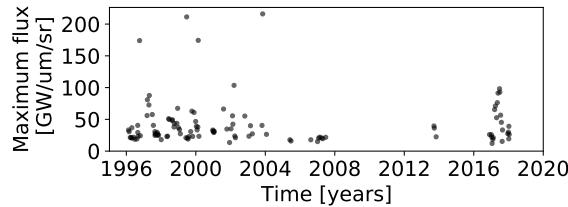
None of the studies of occultations in the infrared attempted to infer a two-dimensional map of Io's surface. The light curves are usually fit independently with the goal of inferring one-dimensional longitudinal variations in brightness on its surface. Strong priors on the locations of individual volcanoes are assumed based on maps from space mission observations. In this work, we build a fully probabilistic model in order to infer a time-dependent two-dimensional *map* of Io's volcanic surface from archival IRTF observations in the near infra red. The model is built on top of the recently developed code starry (?, Luger et al. 2020 in prep) which enables fast analytic computation of occultation light curves and phase curves for objects whose surface features are represented in terms of spherical harmonics. starry can compute phase curves and occultations in both emitted light (for modeling the isotropic volcanic emission from Io's surface) and reflected light (for mapping albedo variations).

Instead of fitting one map per occultation light curve, we reduce the dimensionality of the problem using matrix factorization, assuming a fixed map for a given light curve which is a linear combination of  $K$  basis maps. We use priors on the map features which ensure the maps are physical and prevent overfitting of the data. The model is interpretable, each of the basis maps roughly corresponds to a single spot on the surface, and the coefficients multiplying the basis maps encode information on the time variability.

Our goals for this paper are to both improve on past studies of Io using occultation light curves and to construct a flexible model directly applicable to future efforts on inferring time-dependent maps of variable features on exoplanet surfaces such as volcanoes, magma oceans or clouds. The fact that we have lots of prior information on what the surface of Io should look like enables us to test our model in a realistic setting, testing sensitivity to chosen priors, data quality, the orbital parameters and different kinds of surface features.

The paper is organized as follows. In §2 we describe the IRTF light curves of occultations by Jupiter which we use to infer maps. In §3 we discuss the generative model for the data for the case of a static map, and a dynamic map modeled using a matrix factorization approach and we write down the relevant likelihood functions. In §4 we focus on the question of how to do Bayesian inference given the forward model specifications defined in §3. We discuss and test different priors on map features, the information content of light curves and fitting the models using variational inference (VI) methods and Hamiltonian Monte Carlo (HMC). We present results on simulated data. §5 shows the results on real data. In particular, the inferred dynamic map, the different component maps and accompanying coefficients, the uncertainties of the maps, sensitivity to different choices of priors, comparison of inferred locations and intensities of spots to known values from the literature. Finally, in §6 we discuss applications of the model to future exoplanet observations and test the model predictions on simulated JWST observations of a volcanic Super-Earth.

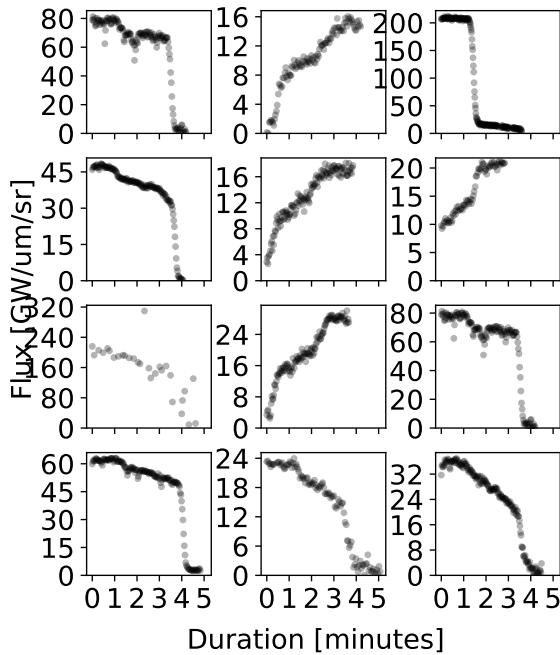
## 2. DATA



**Figure 1.** Maximum flux for each curve taken using NASA’s IRTF telescope during an occultation of Io by Jupiter.

The dataset we use to infer maps of Io’s surface consists of 112 observations of Io in the near infrared taken from 1996 until 2018 using various instruments at NASA’s IRTF observatory. Figure 1 shows the maximum flux in  $\text{GW}/\mu\text{m}/\text{sr}$  for each light

curve plotted as a function of time which approximately corresponds to the disc integrated flux of Io at the beginning of ingress or the end of egress prior to it being occulted by Jupiter. Two features of this plot are apparent. First although the observations span decades the cadence is non-uniform with only a few observations taken between 2008 and 2016. Second, the baseline brightness is varying stochastically by a large amount and there are several notable events of increased volcanic activity. All of the observations were taken while Io was *in eclipse*, meaning that it was in Jupiter's shadow and all of the observed emission from the surface is due to thermal radiation. Observations of Io *in sunlight* on the other hand probe both the thermal (volcanic) emission and the surface albedo variations in the near infrared.



**Figure 2.** A selection of sample light curves taken during occultations of Io by Jupiter in our dataset. The step-like morphology of the light curves is due to bright volcanoes on Io's surface coming in or out of view during an occultation. These light curves thus visibly encode information about the features on the surface. 

In Figure 2 we display a random subset of all the light curves in the dataset. All occultations last for  $\sim 4$  min and the cadence for each light varies but it is on the order of a second. It follows that over the course of a single exposure Jupiter's limb move by about 15km which provides a lower bound for the size of the features on the surface that we can reliably estimate. The shapes of light curves strongly deviate from the smooth variability one would expect assuming a homogeneous distribution of thermal emission on the surface. Especially prominent are light curves with clear step-like features which are present when bright spots come in or out of view during the course of an occultation. The fact that these features are so clearly visible means that even individual light curves encode a wealth of information about the surface.

The photometric quality varies from year to year because multiple instruments were used over the years. An additional issue is that estimated errorbars aren't provided. As with all ground based photometry, the observations are influenced by atmospheric variability which induces correlated noise in the light curves. Because of this the flux isn't always monotonically increasing or decreasing as one would expect. All of these issues need to be accounted for in the final model.

### 3. MODEL

#### 3.1. Orbital parameters

To model the occultations we need to know the projected separation between Jupiter and Io and the orientation of Io at any given time. We use the [JPL Horizons database](#) and the latest planetary ephemeris DE430 (?) which are accurate to a few kilometers on Io's surface. Specifically, we use the Right Ascension and Declination, the angular (equatorial) diameter (`Ang-diam`), the visibility flag indicating whether Io is occulted or not and whether it is in partial or total eclipse (`Ang-sep/v`), the longitude and latitude at the center of Io's disc as seen from Earth (`Ob-lon`, `Ob-lat`), the sub-solar point longitude and latitude which specify the direction pointing towards the Sun on Io's surface (`Sl-lon`, `Sl-lat`), the distance from Io to the Sun (`r`) and the counterclockwise angle between the Celestial North Pole unit vector projected onto the plane of the sky and the Io's north pole (`NP.ang`). All longitudes are positive to the west. **HORIZONS** provides all ephemeris with a minimum cadence of 1 second so we interpolate all values such that we can evaluate them for arbitrary times.

The coordinate system in **starry** is defined to be right handed such that the  $\hat{z}$  axis points towards the observer and the  $\hat{x}$  axis points to the right on the plane of the sky. The radius of the occulted sphere is fixed to 1 and the orientation is specified by three angles, the counterclockwise obliquity angle `obl` between the  $\hat{y}$  axis and the North Pole of the sphere, the inclination angle `inc` which is set to  $90^\circ$  if the North Pole is aligned with the  $\hat{y}$  axis and the phase angle `theta` rotates the sphere around the  $\hat{y}$  axis in the eastward direction. We have  $\text{obl} = \text{NP.ang}$ ,  $\text{inc} = 90^\circ - \text{Ob-lat}$  and  $\text{theta} = \text{Ob-lon}$ . The occultor position relative to the occulted object is given by

$$x_0/\gamma = -\Delta\alpha \cos\delta \quad (1)$$

$$y_0/\gamma = \Delta\delta \quad (2)$$

$$z_0/\gamma = 1 \quad (3)$$

where  $\Delta\alpha$  are the  $\Delta\delta$  differences in Right Ascension and Declination respectively relative to the occulted object and  $\gamma$  is the angular radius of the occulted sphere. If the occulted object is illuminated by the Sun then we also need to specify the

coordinates of the light source which are given by

$$xs = r \cos \theta \sin \phi \quad (4)$$

$$ys = r \sin \theta \quad (5)$$

$$zs = r \cos \theta \cos \phi \quad (6)$$

where  $\theta$  is obtained by subtracting **Ob-lat** from **Sl-lat** and similarly  $\phi$  is obtained by subtracting **Ob-lon** from **Sl-lon**.  $r$  is equal to the heliocentric distance of the occulted object  $r$ .

### 3.2. Effective radius of Jupiter

Although the sky position of Jupiter relative to Io is known to a precision of a few kilometers, several complications arise when attempting to compute Jupiter's radius. First, Jupiter is not spherical and its equatorial radius is larger than the polar radius by thousands of kilometers. Because **starry** does not support occultations by non-spherical occultors we assume that Jupiter is locally spherical at the point of an occultation and estimate an effective radius from measurements of Jupiter's shape which are based on Voyager radio occultation measurements ([Lindal et al. 1981](#)). These data come in the form of a plot of effective radius as a function of planetocentric latitude at a fixed pressure of 100 mbar (Fig. 7 in [Lindal et al. 1981](#)). When computing the occultation latitude we ignore the variation in effective radius along Jupiter's limb projected on Io's surface and instead fix Jupiter's latitude to the value corresponding to the center of Io's disc.

The second complication is that Jupiter is gaseous which means that it does not have a well defined boundary. In principle, we should compute an effective radius of Jupiter at different locations (pressure) in the atmosphere and model an occultation of Io by a fuzzy occultor. Although this can be done with **Starry**, it is not necessary because the characteristic scale height of Jupiter is around 27 km (CITE) which is below the uncertainty of our inferred maps. Instead, we follow [Spencer et al. \(1990\)](#) and compute the effective radius of Jupiter at about 2.2 mbar, this is the pressure (and the associated effective radius) at which during an occultation a bright source on the surface of Io fades by 50% due to differential refraction.

Since the effective radius information is provided only at a fixed pressure of 100 mbar, to adjust those values for a lower pressure of 2.2 mbar, we assume an exponential pressure profile  $P = P_0 e^{-\Delta r/H}$  where  $H$  is the scale height and  $\Delta r$  is the height difference between the two pressure levels. It follows that to convert the shape profile at 100 mbar to 2.2 mbar we need to add the factor  $-H \ln(2.2/100)$  which is assumed to be constant in the  $\pm 21$  deg latitude range in which the occultations occur. Besides refractive absorption there is also a slight additional methane molecular absorption which [Spencer et al. \(1990\)](#) estimate to be equal to around 12% in their filter, we choose to ignore this because it is far below the resolution of our maps.

Finally, we have to account for the fact that the light from Io is getting significantly bent at the point of half reflective intensity in Jupiter’s atmosphere which results in a smaller shadow of Jupiter on the surface of Io then would be the case if the light didn’t get bent. It is zero at the beginning of the disappearance of a hot spot when there is no refraction, increasing to one scale height  $H$  at the half intensity point and then increasing further to a large value when Io disappears behind Jupiter’s limb. We ignore the variation in the bending and adopt a fixed value of one scale height for this effect which we subtract from the value of the effective radius.

In summary, to compute an effective radius of Jupiter for a given occultation of Io we first have to compute the latitude at which the occultation is occurring, use a modified shape profile from [Lindal et al. \(1981\)](#) to get an effective radius and finally subtract one scale height due to light bending. There are substantial uncertainties in each of these steps, most notably the shape profile data are quite old and it is not clear that the structure of Jupiter’s atmosphere stayed the same since 1980s. The shape profile also depends on the wind velocity structure and temperature which we ignore. In addition to uncertainties about the atmospheric structure, there are uncertainties associated with digitizing the data shown in Fig. 7 in [Lindal et al. \(1981\)](#) because it is not available in table form. We digitizing this figure by reading off the data points “by eye” and then interpolating them using cubic spline interpolation. Finally we add the pressure conversion factor and subtract one scale height due to light bending.

Throughout this paper the highest resolution maps we fit to the data are of order  $l = 25$  which corresponds to a characteristic length scale of around 230 km on the surface of Io. Quantifying the uncertainty in each of these steps is beyond the scope of this paper but we are confident that the errors in the radius estimate are below the typical resolution of our inferred maps. We provide justification for this assumption in section ?? where we infer the location of Loki and it ends up matching well with the high resolution geological map of Io.

### 3.3. *Static map model*

Given the geometry of on occultation event at an arbitrary time, computing the predicted flux with `starry` is straightforward. `starry` computes the integrated flux of an unocculted or an occulted sphere analytically assuming that the surface map can be expanded in terms of spherical harmonics  $Y_{lm}(\theta, \phi)$  up to a certain *degree*  $l$ . The map is thus defined by a vector of spherical harmonic coefficients  $\mathbf{y}$  multiplying the spherical harmonic basis  $\tilde{\mathbf{y}}(x, y) = (Y_{0,0} Y_{1,-1} Y_{1,0} Y_{1,1} Y_{2,-2} Y_{2,-1} Y_{2,0} Y_{2,1} Y_{2,2} \dots)^T$  and the total number of coefficients is  $(l + 1)^2$ . In addition to computing *thermal* phase curves and occultation light curves, `starry` also solves the considerably more complex problem of computing reflected light phase curves and occultations when the sphere is illuminated by a distant light source (Luger et al. 2020 in prep ) which is complicated by the presence of the terminator line. In the case of a reflected light map, the coefficient vector  $\mathbf{y}$  represents spherical albedo in a given wavelength range. Most importantly,

conditional on the parameters specifying the geometry of the occultations being fixed, the **starry** model is *linear* for both emitted and reflected light maps. This is achieved by representing all rotations, changes of basis and integrals with complicated boundaries needed to compute the flux as linear transformations. Since a sequence of linear mappings is also linear, the predicted flux can be written as

$$\mathbf{f} = \mathbf{A} \mathbf{y} \quad (7)$$

where the column vector  $\mathbf{f}$  of shape  $(T, 1)$  is the predicted flux for different values of the occultor position and optionally the direction of the illumination source and  $\mathbf{A}$  is the design matrix of shape  $(T, N)$  (with  $N = (l + 1)^2$ ) which encodes information all operations needed to compute the integrated flux. If the geometry isn't known precisely then the matrix  $\mathbf{A}$  is not fixed and the model is no longer linear.

The characteristic angular scale of features that can be represented with a given map is set by the degree of the map and it is approximately equal to  $\frac{180^\circ}{l}$ . State of the art inferences involving phase curves and secondary eclipses of exoplanets are able to constrain features of order  $l = 1$  (inferring a bright spot on single hemisphere) but for Io we need to fit much higher order maps because the typical scale of volcanic spots is on the order of tens of kilometers (a few degrees). **starry** can handle computations with maps up to  $l \approx 20$  before numerical instabilities kick in (TODO: make this statement more precise).

Although **starry** is built around the idea of expanding surface features in a spherical harmonic basis, this is not the ideal basis for doing inference because it makes it difficult to ensure that the intensity is positive everywhere on the map. The positivity constraint is important not only because we want to avoid having unphysical regions on inferred maps but because it also imposes a strong constraint on the map which reduces the complexity of the inference problem. There is no analytic way to determine whether a given spherical harmonic map  $\mathbf{y}$  is positive-valued everywhere on the sphere. The only way of circumventing this problem that we are aware of is to evaluate the intensity on a (sufficiently fine) discrete grid of *pixels*  $\mathbf{p}$  on the sphere and then either fit for  $\mathbf{y}$  while rejecting inference steps where any of the pixels  $\mathbf{p}$  end up being negative, or we can treat the pixels as model parameters but still compute the actual light curve model in the  $Y_{lm}$  basis where everything is fast. We opt for the latter approach because it is much easier to implement. There exists a linear operator  $\mathbf{P}$  which maps  $\mathbf{y}$  it to pixels:

$$\mathbf{p} = \mathbf{P} \mathbf{y} \quad (8)$$

Each row of  $\mathbf{P}$  contains values of each of spherical harmonic coefficients at a point on the grid. To construct the grid we use an equal area Molleweide projection so that we don't end up with more pixels in some areas of the sphere than others. The grid needs to be fine enough to ensure positivity over most of the sphere, we find that we need to have more pixels than spherical harmonics by a factor of at least 4. Although we can't invert  $\mathbf{P}$  to obtain the inverse transform we can still compute the

Moore-Penrose pseudoinverse  $\mathbf{P}^\dagger$  by solving the linear system  $\mathbf{P} \mathbf{P}^\dagger = \mathbf{I}$  where  $\mathbf{I}$  is the identity matrix. The solution is given by

$$\mathbf{P}^\dagger = (\mathbf{P}^\top \mathbf{P} + \lambda \mathbf{I})^{-1} \mathbf{P}^\top \quad (9)$$

where  $\lambda$  is a small regularization parameter and  $\mathbf{I}$  is the identity matrix. We then have

$$\mathbf{y}' = \mathbf{P}^\dagger \mathbf{p}' \quad (10)$$

Both matrices can be easily pre-computed with `starry` before doing inference.

The flux can then be computed as a linear operation on the pixels:

$$\mathbf{f} = \mathbf{A} \mathbf{P}^\dagger \mathbf{p}' \quad (11)$$

We use the symbol  $\mathbf{p}'$  to denote pixels which are mapped to spherical harmonic coefficients using  $\mathbf{P}^\dagger$  and  $\mathbf{p}$  for those generated from  $\mathbf{y}$  using  $\mathbf{P}$ . In the rest of the paper we call the former kind pixels *schmixels*. In Section 4.2 we test differences between fitting a model for spherical harmonic coefficients vs. fitting for schmixels.

### 3.4. Nonnegative Matrix Factorization

In the case when we cannot assume a static map, the situation is considerably more complicated. In principle, the surface map is different for each *data point*. Of course, fitting one map per data point is intractable so we might instead fit a single map per light curve, generated by Eq. 11. Even then, we would need to fit on the order of 1k parameters! This is both because the number of spherical coefficients scales as  $(l + 1)^2$  and because when fitting in pixel space we need many more pixels than spherical harmonics to ensure positivity. Although fitting a model of such high dimensionality for a single IRTF light curve of high signal to noise ratio is tractable thanks to exact gradients provided by autodifferentiation in `starry`, we don't want to do that for a hundred or so separate light curves. Such a model would be very difficult to fit and it would require strong regularization between successive maps in time. It also wouldn't directly provide much physical insight into the volcanic activity, we would have to conduct a separate analysis on the inferred maps to constrain the time variability of the volcanoes. Instead, we need an approach which reduces the dimensionality of the problem. One way of accomplishing that is to expand the spherical harmonic coefficients (or pixels) in a Taylor series in time about a certain point. This is the approach ? took to model a map of Earth's albedo in reflected light from stray Earthshine in the aperture of the TESS space telescope. This issue with this approach and similar series expansions such as the Fourier series is that with the very long baseline of IRTF data and the observations being scattered throughout the years sporadically, it is unclear which point should be the origin of the expansion. More importantly, there likely doesn't exist a well defined global time-averaged map of Io so requiring too much smoothness between successive maps wouldn't work.

Another way of reducing the dimensionality of the problem would be a parametric approach where we place a few spots on the sphere and then fit only for the locations and intensities of each spot. This is assuming that we know a priori that the surface features are spot like and roughly how many we should expect. While such a strong assumption might work for studying individual spots in the context of Io, it certainly isn't justified for exoplanets. In Section 4 we do assume stronger priors on the locations of and shapes of surface features to better constrain the variability of individual volcanoes, but we still fit for pixels, at least in principle allowing the data to override our assumptions.

The approach opted for is to treat the problem of inferring a single map per light curve as a probabilistic matrix factorization problem. The idea is to assume that a map for any given light curve can be expressed as a linear combination of  $K$  "basis maps". Consider  $L$  lightcurves, the model for the  $l$ -th light curve is

$$\mathbf{f}_l = \mathbf{A}_l \mathbf{P}^\dagger \mathbf{p}'_l \quad (12)$$

We can stack the column vectors  $\mathbf{p}_l$  into a matrix  $\mathbf{Y}$  of shape  $(N_p, L)$  where  $N_p$  is the number of pixels for each light curve. We then assume that the  $\mathbf{Y}$  can be decomposed into a product of two matrices  $\mathbf{B}$  and  $\mathbf{Q}$  as

$$\mathbf{Y} = \mathbf{B} \mathbf{Q} \quad (13)$$

where  $\mathbf{B}$  has shape  $(N_p, K)$  and  $\mathbf{Q}$  has shape  $(K, L)$ . The model for all light curves can be written as

$$\mathbf{f} = \mathbb{A} \text{vec}(\mathbf{Y}) \quad (14)$$

Where  $\mathbf{f}$  is a tall column vector consisting of predictions for all light curves stacked together,  $\mathbb{A}$  is a block diagonal matrix with matrices  $\mathbf{A}_l \mathbf{P}^\dagger$  on the diagonal for  $l = 1, \dots, L$  and the vec operator stacks the columns of  $\mathbf{Y}$  into a tall vector. The model is *bilinear* which means that is linear in the matrices  $\mathbf{B}$  and  $\mathbf{Q}$  separately. The interpretation of Equation 13 is simple, each of the  $K$  columns of  $\mathbf{B}$  represents pixels of a basis map and the columns of  $\mathbf{Q}$  determine how those maps add together to produce a final map for the  $l$ -th light curve. For this reason we call  $\mathbf{B}$  the *basis matrix* and the matrix  $\mathbf{Q}$  the *encoding matrix*. Ideally, we want the basis maps to be physically meaningful and the coefficients encode the time variability. A nice feature of this model is that there is no requirement that successive maps are smoothly varying between different light curves.

The matrix factorization problem as written in Equation 13 is highly degenerate and practically intractable a probabilistic framework with a small dataset. Since each of the columns of  $\mathbf{B}$  are pixels representing emitted light or albedo, we can substantially reduce the ambiguity in the decomposition by requiring that both matrices are strictly positive. This matrix factorization problem is known as *Nonnegative Matrix Factorization* (NMF) (??) and it has a rich history across many different fields. It is

commonly used for decomposing physical signals in the spectral or the time domain. Even though NMF is more tractable than unconstrained matrix factorization, it is still an NP hard problem (?) and it requires additional constraints or priors to be tractable. Simultaneously transforming  $\mathbf{B} \leftarrow \mathbf{B} \mathbf{S}^{-1}$  and  $\mathbf{Q} \leftarrow \mathbf{S} \mathbf{Q}$  with a nonsingular matrix  $\mathbf{S}$  doesn't change the value of the objective function. The objective function is also invariant to a simultaneous permutation of the columns of  $\mathbf{B}$  and rows of  $\mathbf{Q}$  and rescaling either of the two matrices by a scalar. Nevertheless, getting around these degeneracies in practice is possible, especially when we have a lot of prior knowledge on the problem. For a recent review article with a detailed analysis of these degeneracies and common algorithms, see ?.

Broadly, there are two approaches to fitting NMF models, the constrained optimization approach and the probabilistic approach. In the optimization approach we optimize  $\mathbb{Y}$  in a least-squares sense under a set of constraints on either one or both of the matrices (see ????, for recent examples from the astronomical literature) with the goal of obtaining a point estimate for the two matrices. The probabilistic approach introduces priors and the inference is usually done with variational inference in order to obtain an approximate posterior distribution over the matrices. In this work we opt for the probabilistic approach because we care about the uncertainties on the map features a lot.

One difference between most applications of NMF in the literature and NMF within the context of inferring surface maps from 1D light curves is that in the majority of the work in the literature (except ?) the matrix  $\mathbb{Y}$  is assumed to be directly observed up to a simple noise term. In our case the problem is considerably more challenging because we observe the light curves  $\mathbb{A} \text{vec}(\mathbb{Y})$  instead of  $\mathbb{Y}$  directly, a process in which some information is lost because, depending on the geometry of the occultations, certain combinations of spherical harmonic coefficients will be in the nullspace. ? are using NMF to model phase variations in directly imaged exoplanets with the goal of simultaneously inferring a surface map and a spectral decomposition of the map into several components, a case in which the matrix which is to be decomposed is also unobserved. Although we don't attempt to solve the problem of spectrally decomposing the maps, our approach is directly applicable to that problem as well. (TODO: should I add more details here?) One could also imagine inferring a multi spectral component and time dependent map by solving two NMF problems simultaneously. We leave this application for future work.

### 3.5. *The likelihood*

Given Eq. 13 we can write down the likelihood for our data and in Section 4 we discuss in detail the priors which make the problems of inferring static and dynamic maps tractable. Assuming a Gaussian noise process for the data with a dense covariance matrix, the (log) likelihood is

$$\ln \mathcal{L} = -\frac{1}{2} [\mathbf{f}_{\text{obs}} - (\mathbf{f} + \mathbf{b})]^{\top} \boldsymbol{\Sigma}^{-1} [\mathbf{f}_{\text{obs}} - (\mathbf{f} + \mathbf{b})] \quad (15)$$

where  $\mathbf{f}_{\text{obs}}$  is the vector of stacked observed light curves,  $\Sigma$  is the data covariance matrix and  $\mathbf{b}$  is a fixed flux offset per light curve accounting for stray flux not attributed to Io (mostly due to Jupiter). We model the data covariance with a Gaussian Process using the fast Celerite method(?) as implemented in the `exoplanet` package (TODO: cite) to capture correlations imposed by the seeing plus an additional white noise term which is different for each light curve. The IRTF data isn't provided with any uncertainties so we estimate the uncertainties by filtering the data with the Savitzky Golay filter as implemented in `SciPy` (TODO: CITE) and estimating the variance of the residuals. We assume identical errors for all data points in a given light curve but allow for a rescaling factor when fitting the model. In the next section, we add priors to this likelihood and fit the model on simulated data using optimization and Variational Inference (VI).

#### 4. THE INVERSE PROBLEM

Having defined the forward models in the previous section, in this section we focus on choosing priors, doing inference and testing the models on simulated data. We compute the information content of different kinds of occultation and phase light curves and , we test the differences between fitting in the  $Y_{lm}$  basis vs. fitting in the pixel basis, prior choices for pixel maps and priors for NMF. To fit the models, we use variational inference and Hamiltonian Monte Carlo.

##### 4.1. *The information content of a light curve*

The mapping problem is famously ill posed. In other words, certain linear combinations of spherical harmonic coefficients will be in the nullspace of the linear mapping  $\mathbf{A}$  in Eq. 7. This means that even if we had noiseless observations, certain features would still be impossible to recover. To recover surface features from observed flux integrated over the disc of the object we need to have some mechanism which breaks the various degeneracies. For phase curves in emitted light, we can only recover longitudinal variations in intensity. Occultations are a lot better thanks to the limb of the occultor sweeping across the surface of the occulted sphere. Even better are grazing occultations with an occultor of a small diameter sweeping across the disc at different latitudes and therefore breaking the latitudinal degeneracies. Observing phase curves and occultations in reflected light improves the model even more because of the nonuniform illumination profile of the incident radiation and the presence of a day/night terminator line (?, Luger et al. 2020 in prep). In some cases for reflected light observations (phase curves of an inclined planet for example) there is no nullspace at all at low degrees of the map.

To compute how much information can in principle be obtained by observing occultations of Io by Jupiter and other moons or phase curves we need to compute some measure of how much does observing Io at certain times reduce our prior uncertainty on the map features. Given that the model for a static map (Eq. 7) is linear, under Gaussian priors with covariance  $\Lambda$  on the spherical harmonic coefficients and a

Gaussian likelihood, the posterior is analytic and the mean is given by

$$\hat{\mathbf{y}} = \Sigma_{\hat{\mathbf{y}}} (\mathbf{A}^\top \Sigma_{\mathbf{f}}^{-1} \mathbf{f} + \Lambda_{\mathbf{y}}^{-1} \boldsymbol{\mu}_{\mathbf{y}}) \quad (16)$$

where the posterior covariance matrix  $\Sigma_{\hat{\mathbf{y}}}$  is

$$\Sigma_{\hat{\mathbf{y}}} = (\mathbf{A}^\top \Sigma_{\mathbf{f}}^{-1} \mathbf{A} + \Lambda_{\mathbf{y}}^{-1})^{-1} \quad (17)$$

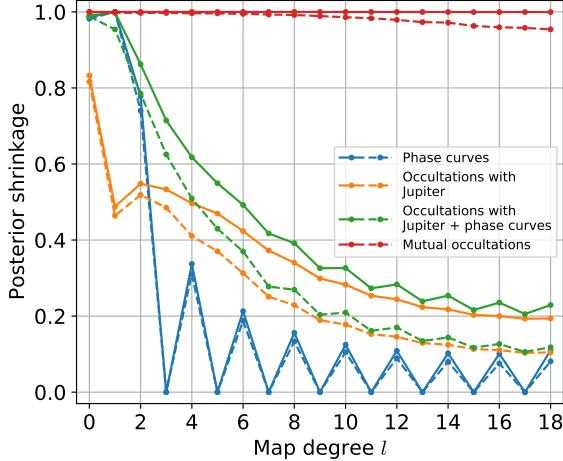
The posterior covariance matrix depends on the design matrix  $\mathbf{A}$ , the prior covariance  $\Lambda$  and the data covariance  $\Sigma_{\mathbf{f}}$ . We first compute the design matrices using `starry` for different kinds of events  $a$  over the course of a single year: occultations by Jupiter, occultations by Europa, Ganymede and Callisto and phase curves.

We take the ephemeris data from **JPL HORIZONS** and assume that it is known exactly, we also assume all observations are of the emitted light component independent of whether Io is in sunlight or in eclipse. This is because we are interested in constraining the volcanic emission rather than the albedo. Given the fixed design matrices, we can compute the variance reduction of the posterior relative to the prior known as *posterior shrinkage*, defined as

$$1 - \frac{\sigma_{\text{post}}^2}{\sigma_{\text{prior}}^2} \quad (18)$$

where  $\sigma_{\text{prior}}^2$  is the prior and  $\sigma_{\text{post}}^2$  is the posterior variance for a given parameter. To compute this quantity for each spherical harmonic coefficient in  $\mathbf{y}$  we divide the diagonal elements of matrices  $\Sigma_{\hat{\mathbf{y}}}$  and  $\Lambda$ , divide them, and average over all  $m$ -modes. We need to assume some variance for the Gaussian prior on the  $Y_{l,m}$  coefficients so we choose a broad prior  $\Lambda = \text{diag}(1^2, 0.1^2, \dots, 0.1^2)$ . Although the posterior shrinkage depends on this prior the main results do not change for a different choice of the prior. The result is plotted in Fig. 3. The plot shows posterior shrinkage for phase curve observations (blue lines), occultations by Jupiter (orange), the previous two combined (green) and mutual occultations by other Galilean moons (red). Solid lines correspond to SNR=100 while dashed lines correspond to SNR=10 where the signal is defined as the min-max difference of the predicted flux.

It is clear that observations of mutual occultations of Io (red lines) are by far the most informative and most useful for constraining surface features. The posterior shrinkage is near unity at all angular scales  $l$  up to  $l = 18$  owing to degeneracy breaking features of these occultations. Occultations by Jupiter (orange lines) are far less informative because we only see one side of Io during an occultation (hence the kink in the curve at  $l = 1$ ) and the position of Jupiter's limb on the surface of Io does not vary significantly so latitudinal features are only weakly constrained. The posterior shrinkage is exactly zero for phase curves at all odd degrees above  $l = 2$  because phase curves for objects rotating about an axis perpendicular to the line of sight encode zero information on these coefficients because they are in the nullspace.



**Figure 3.** The plot shows the information content of different kinds of observations of Io as a function of spherical harmonic degree (angular scale of surface features). On the y axis is posterior shrinkage (the variance ratio between the prior and the posterior distribution) for each spherical harmonic coefficient at a given degree  $l$  on the x axis, averaged across all  $m$  modes. Posterior shrinkage of 1 represent maximum information gain in updating from the prior to the posterior while a shrinkage of 0 represents no information gain. We compute the posterior variance for different kinds of simulated observations of Io over the course of a single year: phase curves (blue), occultations by Jupiter (orange), combined phase curves and occultations by Jupiter (green) and occultations of Io by other Galilean moons (red lines). Solid lines represent observations at SNR=100 while dashed lines represent observations at SNR=10.

Although observations of mutual occultations most easily break the degeneracies, they only occur every 6 years and more importantly they almost never occur while Io is in eclipse. This means that only the brightest volcanoes will be visible above the reflected light background emission due to the Sun. Figure 3 does not reflect this fact. When modeling the observations taken while Io is in sunlight, we have to take into account the reflected light component. Although `starry` supports computing occultations and phase curves in reflected light, we only fit in eclipse observations in this work.

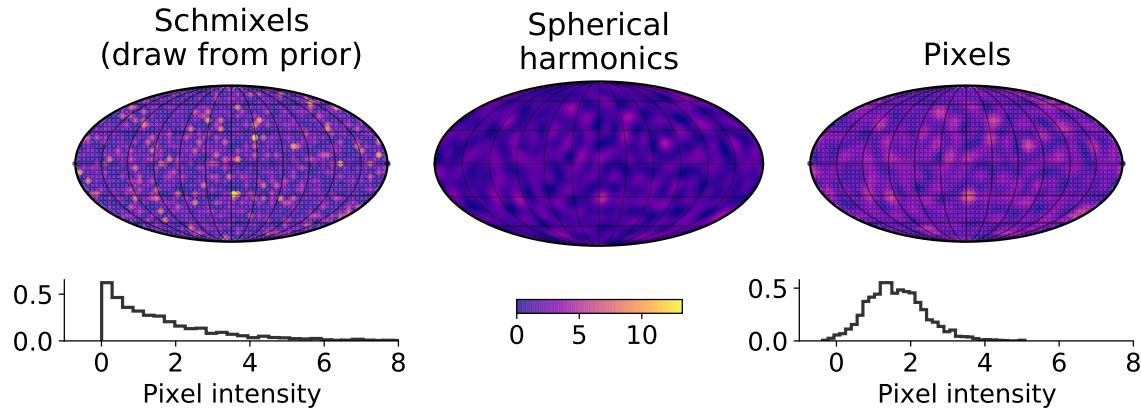
#### 4.2. Schmixels vs. spherical harmonics

To test inference with the static map model defined in Section 3.3 we use `starry` to generate a single simulated light curve of an occultation of Io by Jupiter from an  $l = 30$  map. The simulated map consists of a spherical harmonic expansion of a bright spot with a Gaussian profile which we add using the built in `add_spot` function in `starry`. The expansion is in the quantity  $\cos(\Delta\theta)$  where  $\Delta\theta$  is the angular separation between the center of the spot and another point on the surface of the sphere. We place the spot at 13 degrees latitude and 51 degrees East longitude, set the diameter of the spot ( $2\Delta\theta$ ) to 5 degrees and the amplitude to 55% of the total luminosity of a spotless map. We take ephemeris parameters (assumed to be fixed) from a real occultation and assume an observational cadence of 1s and errorbars equal to 2% of the total flux, similar to high quality IRTF light curves. We use this dataset to test

the difference between fitting a model in the spherical harmonic basis (Eq. 7) and in the schmixel basis (Eq. 11) by fitting for an  $l = 20$  map.

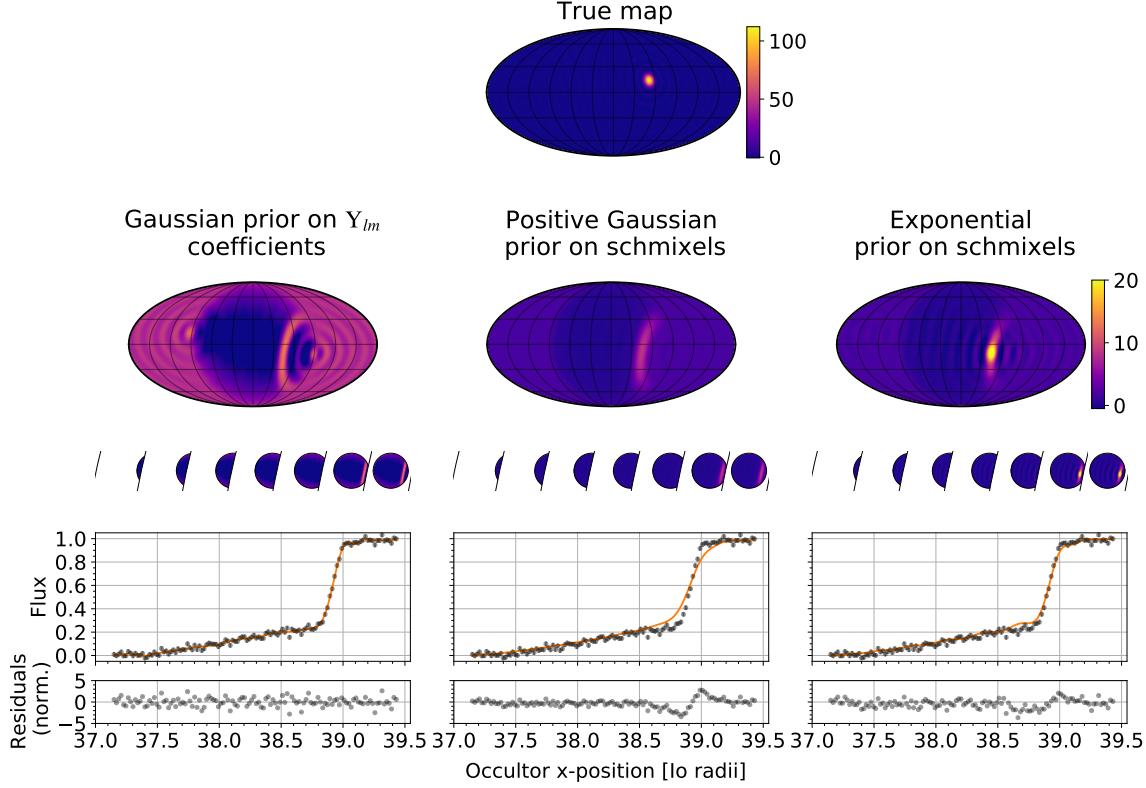
In the first case we place a Gaussian prior on  $Y_{lm}$  coefficients with covariance  $\Lambda = \text{diag}(1^2, 0.1^2, \dots, 0.1^2)$  and optimize for the values of the coefficients. In the second case we fit for schmixel  $\mathbf{p}'$  evaluated on a fixed (latitude, longitude) grid with two sets of positive priors, a Half Gaussian prior with zero mean and an Exponential prior. To ensure we have a grid of high enough resolution we use 4 times as many schmixels as spherical harmonic coefficients.

To evaluate the flux (Eq. 11) we map the schmixels to a spherical harmonic coefficient vector  $\mathbf{y}$  using the matrix  $\mathbf{P}^\dagger$  at each evaluation of the posterior. As the end product of the inference we do not use the schmixels  $\mathbf{p}'$  but rather the spherical harmonic coefficients  $\mathbf{y} = \mathbf{P}^\dagger \mathbf{p}'$  generated from the schmixels. The coefficients  $\mathbf{y}$  can then be used to evaluate the map on a pixelated grid of arbitrary resolution via the matrix  $\mathbf{P}$ . Although the priors on schmixels  $\mathbf{p}'$  are strictly positive, the pixels  $\mathbf{P} \mathbf{y}$  don't completely respect those priors because some information is lost in going from schmixels to spherical harmonics and there will still be some regions of the map with negative intensity. To visualize the effect of the transformation from schmixels to spherical harmonics, in Figure 4 we plot a single draw of schmixels from an Exponential prior (left), convert them to spherical harmonics (middle) and then generate another pixelated map from the spherical harmonics on a grid of the same size as the schmixel grid. The middle map is visibly smoother than the prior map and based on the histograms shown in black, it seems that the effective prior is closer to a Gaussian than an Exponential. There is also minimal leakage into negative values.



**Figure 4.** Mapping from a pixelated map in which we set our priors (left) to the spherical harmonic basis which we use to compute the model (middle) and then back to pixels (right). The mapping from spherical harmonics to pixels is exact but the reverse is not true so to differentiate between pixels mapped to spherical harmonics and pixels constructed from spherical harmonics we call the former "schmixels". Since there are many more schmixels than there are spherical harmonics coefficients the map drawn from an Exponential prior on the left can only be approximately represented in the  $Y_{lm}$  basis and the effective prior on pixels ends up looking more like a Gaussian with a heavier right tail than an Exponential.





**Figure 5.** Maximum A Posteriori (MAP) estimates of  $l = 20$  maps (second panel from the top) produced by fitting a simulated light curve (bottom) which was generated from an  $l = 30$  map consisting of a single bright spot (top). The small circles beneath each inferred map show the progression of the occultation. The map on the left corresponds to a model in which we directly fit for spherical harmonic coefficients  $\mathbf{y}$  with a Gaussian prior. The map in the middle comes from a model in which we instead fit for schmixels  $\mathbf{p}'$  with a Half Gaussian prior but still use  $\mathbf{y}$  to compute the flux. The map on the right is the same as the middle one except we use an Exponential prior. The scale parameter of the Half Gaussian and the Exponential prior is set to a value which approximates the scale of Gaussian prior on  $\mathbf{y}$  to ensure fair comparison between the solutions. While all models fit the data reasonably well, the ones in which we fit for schmixels result in more plausible and sparse maps. The location of the true spot is not recovered correctly because a single light curve does not contain enough information to break the degeneracy in position along the limb of the occultor.

To ensure that the difference in the inferred maps is not in part due to a difference in the scale of the priors, we take 10000 prior samples of the  $\mathbf{y}$  in the spherical harmonic model, compute the standard deviation and use that as the scale parameter of the Half Gaussian and the Exponential prior. We fit all models for the Maximum A Posteriori (MAP) estimate of the parameters using the BFGS algorithm as implemented in `scipy.optimize.minimize` in the `scipy` package (CITE) and the exact gradients of the model with respect to all parameters provided by `Theano`. The results are shown in Figure 5. The top subplot shows the true map with a localized bright spot and the inferred maps are shown in the second row from the top together with the progression of the occultation (miniature maps). Bottom subplots show the simulated light curves, the predicted flux from the model and the normalized residu-

als. The difference between the inferred maps is substantial despite the fact that the predicted flux is similar in all three cases.

The inferred map from the spherical harmonic model (left) has near zero intensity on the observer facing side of Io except for a bright arc tracing the projected limb of Jupiter at the location of the true spot. The flux on the unobserved side of Io is nearly uniform and significantly higher than the flux on the observer facing side which results in an unrealistic transition between the bright and the dark region at  $\pm 90$  degrees longitude. A very clear "ringing" pattern is visible in the with high intensity peaks and low (in some regions negative) intensity troughs. These features are a consequence of the finite order spherical harmonic expansion of the map and we could in principle get rid of them by fitting a higher order map but at  $l = 20$  we are already pushing the limits of what `starry` can do. Besides avoiding unphysical negative intensities, ringing is undesirable because we want to avoid situations in which the model uses the rings to explain the data instead of just placing a spot directly. For example, we find that for some fits at low  $l$ , the model would place a bright spot on the unobserved side of Io in order to produce a ringing artefact on the observed side to explain an increase or decrease in brightness in the light curve. This is happening to a lesser extent in the shown map where a small spot is visible at the right edge of the observed hemisphere in order to amplify the visible bright arc through ringing.

The schmixel model with Half Gaussian priors produces a more sensible map in which the unobserved hemisphere has much lower intensity but the arc is less sharply defined and more extended and as a result the predicted flux does not rise steeply enough compared to the light curve. The map from a model with an Exponential prior on the schmixels provides the best fit. It is the sparsest of the three maps, having a single bright spot at the center of the projected limb of Jupiter at the location of the true spot with extended tails along the limb. In principle, the sparsity is a consequence of the Exponential prior which has significant probability mass near zero, meaning that it pushes most schmixels towards zero intensity, and a heavy tail towards large intensities. It favors parse maps with near zero intensity everywhere except for a few high intensity regions which is exactly the kind of prior information we want to assume for Io because we expect spot like features on its surface. In practice, as shown in Figure 4, the Exponential prior implies a more Gaussian like prior in spherical harmonic space and yet it still gives noticeably different looking map than a Gaussian (NOT CONVINCING).

#### 4.3. Smoothing out spurious features

To suppress ringing which results in regions of negative intensity surrounding a spot like feature as in Figure 5, we choose to apply a spatial smoothing filter to the spherical harmonics. Mathematically, the filtering operation is a convolution between the map and some kernel function  $B(\theta, \phi)$ . Assuming both the map and the kernel function are expanded in terms of spherical harmonics, the convolution operation is

simply a multiplication between the two sets of spherical harmonic coefficients. We choose a simple azimuthally symmetric Gaussian kernel function given by

$$B(\theta) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp(-\theta^2/2\sigma_s^2) \quad (19)$$

where  $\sigma_s$  sets the characteristic scale of the smoothing. This function can be expanded in terms of spherical harmonics as

$$B(\theta) = \sum_{l=0}^{\infty} \left( \frac{2l+1}{4\pi} \right) B_l \mathcal{P}_l(\cos \theta) \quad (20)$$

where  $B_l$  are the spherical harmonic coefficients and  $\mathcal{P}_l$  are the associated Legendre polynomials. They depend only on  $l$  because all nonzero  $m$  modes vanish due to azimuthal symmetry. For  $\sigma_s \ll 1$   $B_l$  can be approximated as (Seon 2007; White & Srednicki 1995)

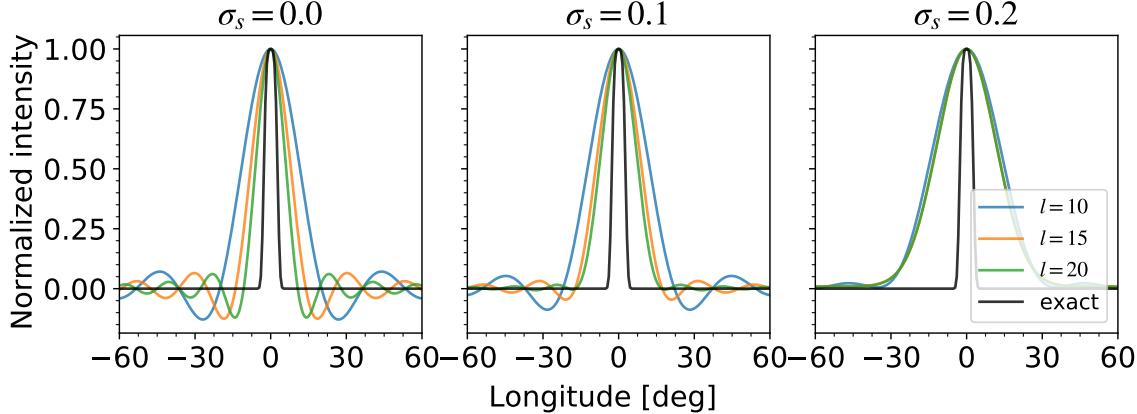
$$B_l \simeq \exp \left[ -\frac{1}{2} l(l+1)\sigma_s^2 \right] \quad (21)$$

The Gaussian filter results in an exponential cutoff for scales smaller than  $l \sim \sigma_s^{-1}$ .

In Figure 6 we show the effect of (Gaussian) smoothing on a spherical harmonic expansion of a spot with a Gaussian intensity profile. The expansion is in the quantity  $\cos \Delta\theta$  where  $\Delta\theta$  is the angular separation between the center of the spot and another point on the surface of the sphere. We place the spot at zero degrees latitude and longitude and set the size of the spot  $\Delta\theta$  to 5 degrees. All three subplots show the exact profile of the spot (black line) and expansions up to three different orders  $l$  (colored lines). The leftmost subplot shows the expansion with no smoothing ( $\sigma_s = 0$ ) in which case the symmetric ringing around the center of the spot is clearly visible even at relatively high order ( $l = 20$ ). The middle subplot shows an intermediate level of smoothing where  $\sigma_s = 0.1$ , meaning that all features on scales above  $l \approx 10$  are exponentially suppressed. In this case the negative ringing is a lot less visible albeit at the cost of having a slightly larger spot because suppressing higher order harmonics necessarily means that we lose some ability to represent smaller scale features. This is most apparent in the rightmost subplot where  $\sigma_s = 0.2$ . Although there is practically no ringing, expansions at  $l = 10$ ,  $l = 15$  and  $l = 20$  look almost the same because all coefficients above  $l = 5$  are significantly suppressed.

Thus, there is a trade-off between smoothing and resolution of the maps. In principle we can always get rid of ringing by fitting sufficiently high order maps. In practice, computing the analytic integrals computed by `starry` becomes computationally unstable above  $l \approx 25$  so instead of going to very high order we apply some smoothing to mitigate the ringing.

#### 4.4. Fitting simulated ingress/egress light curves



**Figure 6.** Normalized intensity of a spherical harmonic expansion of a Gaussian spot placed at zero degrees latitude and longitude. The expansion is in the quantity  $\cos \Delta\theta$  where  $\Delta\theta$  is the angular separation between the center of the spot and another point on the surface of the sphere, it is set to 5 degrees. The three insets show the spot profile at different strengths of the smoothing kernel where the leftmost spot has no smoothing ( $\sigma_s = 0$ ) and the rightmost spot is strongly smoothed. The lines of different colors correspond to spot expansions up to certain order and the black line is the expansion at infinite order. The purpose of smoothing is to taper the higher order spherical harmonic coefficients. No smoothing results in significant "rings" where the intensity goes negative, even at high order. High levels of smoothing reduce the ringing but increase the spot size and eliminate differences between expansions above a certain order. 

In Figure 7 we test a model with the smoothing filter by simulating data as in Figure 5, except this time we simulate two light curves, an ingress light curve generated from real ephemeris of Io disappearing behind the limb of Jupiter and egress light curve generated when Io reappears.

#### 4.5. Fitting IRTF ingress/egress light curves

#### 4.6. Fitting a dynamic map

### 5. RESULTS

#### 5.1. Individual events

#### 5.2. The time-variable map

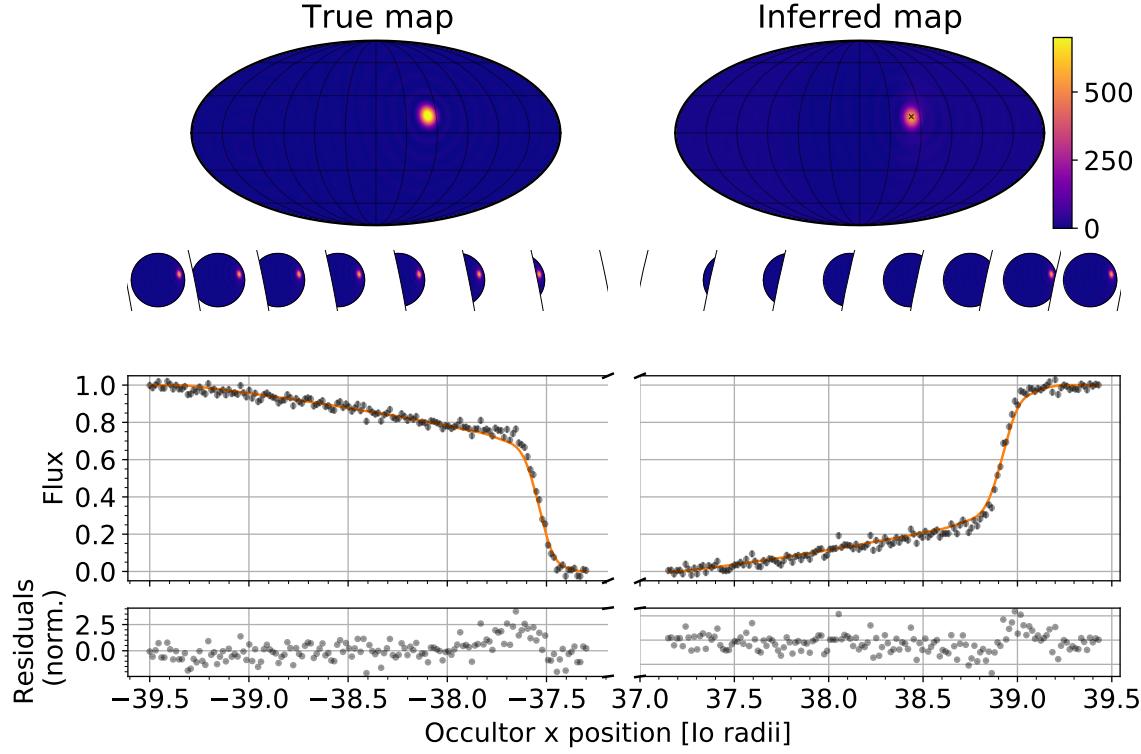
#### 5.3. Variability of known hotspots

### 6. MAPPING VOLCANIC EXOPLANETS

### 7. CONCLUSIONS

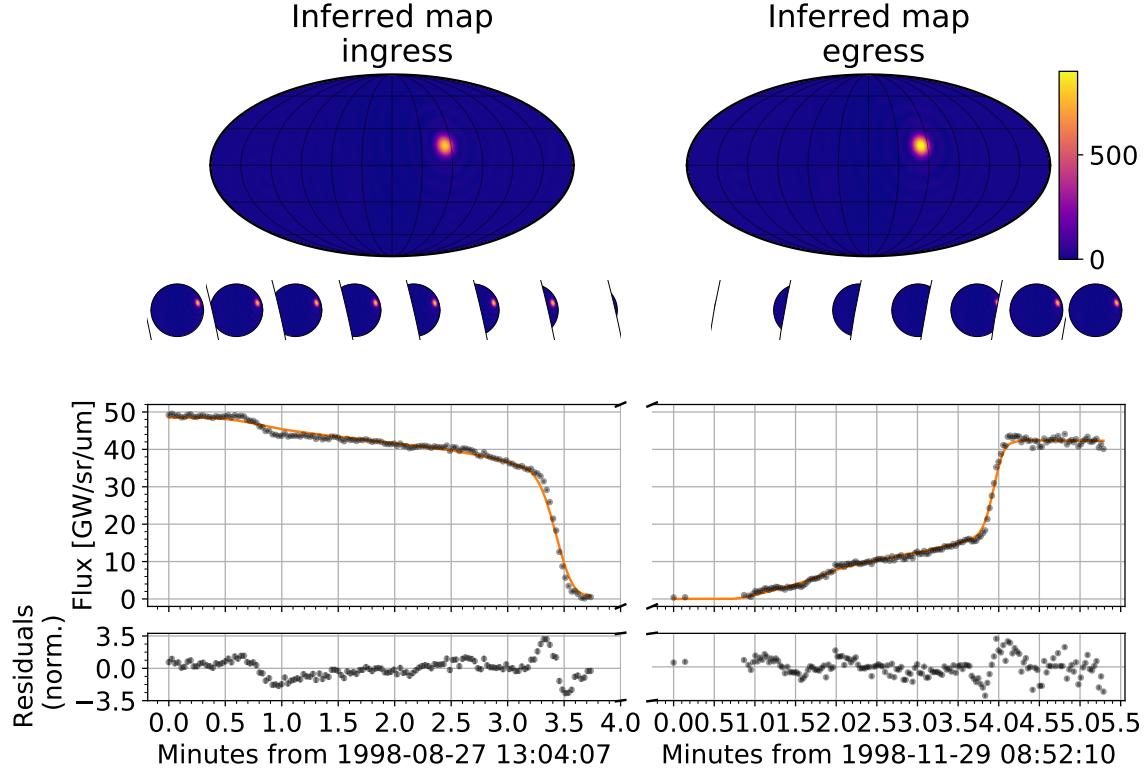
### REFERENCES

- Lindal, G. F., et al. 1981, J. Geophys. Res., 86, 8721. <http://doi.wiley.com/10.1029/JA086iA10p08721>
- Seon, K.-I. 2007, arXiv:astro-ph/0703168, arXiv: astro-ph/0703168. <http://arxiv.org/abs/astro-ph/0703168>
- Spencer, R., et al. 1990, 348, 4

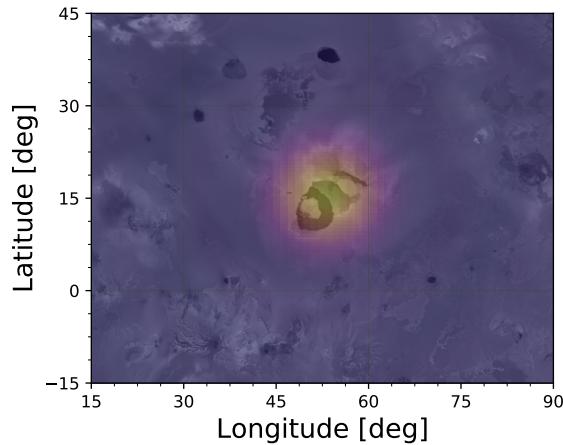


**Figure 7.** Maximum A Posteriori (MAP) estimate of an  $l = 25$  map (right panel from the top) produced by fitting a pair of light curves simulated from the true map which is shown on the left. The simulated light curves are shown in the third panel from the top, an ingress light curve on the left and an egress light curve on the right. The small circles above each of the light curves show the progression of the occultation over the inferred map. The precise location of the true spot is not recovered for either of the two models which is not surprising because a single light curve is not sufficient to break the degeneracies.

White, M., & Srednicki, M. 1995, The  
Astrophysical Journal, 443, 6.  
<http://adsabs.harvard.edu/abs/1995ApJ...443....6W>



**Figure 8.** Maximum A Posteriori (MAP) estimate of two  $l = 25$  maps produced by simultaneously fitting two IRTF light curves of an occultation of Io by Jupiter. The light curve on the left corresponds to an ingress of Io while the one on the right fitting a pair of light curves simulated from the true map which is shown on the left. The simulated light curves are shown in the third panel from the top, an ingress light curve on the left and an egress light curve on the right. The small circles above each of the light curves show the progression of the occultation over the inferred map. The precise location of the true spot is not recovered for either of the two models which is not surprising because a single light curve is not sufficient to break the degeneracies. [🔗](#)



**Figure 9.** This is a plot of a pretty function. And at the end of this caption is a symbol with a link to the *exact* script that generated it, hosted on GitHub. [🔗](#)