Occultation Mapping of Io’s Surface in the Near-infrared. I. Inferring Static Maps

Fran Bartoli1,2, Rodrigo Luger2, Daniel Foreman-Mackey2, Robert R. Howell3, and Julie A. Rathbun4
1 Centre for Exoplanet Science, SUPA, School of Physics and Astronomy, University of St. Andrews, St. Andrews, UK; fbar90@st-andrews.ac.uk
2 Center for Computational Astrophysics, Flatiron Institute, New York, NY, USA
3 Department of Geology & Geophysics, University of Wyoming, Laramie, WY, USA
4 Planetary Science Institute, Tucson, AZ, USA

Received 2021 March 5; revised 2021 August 16; accepted 2021 September 17; published 2022 March 25

Abstract

With hundreds of active volcanoes varying in intensity on different timescales, Jupiter’s moon Io is the most volcanically active body in the solar system. Io has been observed from Earth using high-cadence near-infrared photometry during occultations by Jupiter and other Galilean moons since the 1980s. These observations encode a wealth of information about the volcanic features on its surface. We built a generative model for the observed occultation light curves using the code starry, which enables fast, analytic, and differentiable computation of occultation light curves in emitted and reflected light. Using this model, we are able to recover surface thermal emission maps of Io containing known volcanic hot spots without having to make assumptions about the locations, shapes, or number of hot spots. Our model is also directly applicable to the problem of mapping the surfaces of stars and exoplanets.

Unified Astronomy Thesaurus concepts: Io (2190); Volcanoes (1780); Volcanism (2174); Occultation (1148); Bayesian statistics (1900); Exoplanets (498)

1. Introduction

The surface of Io is covered with hundreds of volcanoes that appear as bright spots in the near-infrared; their intensities vary on timescales ranging from days to decades. The global heat flow on its surface is about 40 times larger than Earth’s (Breuer & Moore 2007; Davies & Davies 2010), and about 50% of the heat flow emanates from only 1.2% of Io’s surface (Veeder et al. 2012). This intense volcanic activity cannot be explained by radioactive decay and residual heat from formation, the mechanisms that drive volcanism on Earth. Instead, the volcanism on Io is driven by tidal interactions with Jupiter and sustained by the Laplace resonance with Europa and Ganymede (Peale et al. 1979). Contrary to Earth, Io lacks plate tectonics. The heat transport mechanism most likely operating on Io is heat-pipe volcanism, a process in which magma is transported to the surface through localized vents from the lithosphere (the outermost shell of a terrestrial body; O’Reilly & Davies 1981). Heat-pipe volcanism likely occurred in the early history of terrestrial planets in the solar system (Moore & Webb 2013; Moore et al. 2017), most notably on early Earth prior to the onset of plate tectonics.

Besides providing a window into early volcanic activity in the solar system, Io is also in many ways an analog of a volcanically active exoplanet. Volcanic exoplanets, sometimes dubbed “super-Ios” or “lava worlds,” have gathered a lot of interest in recent years. Photometric and spectral signatures of volcanic activity on such worlds will likely be detectable in the near future with telescopes such as the James Webb Space Telescope (JWST) and LUVOIR (Kaltenegger et al. 2010; Henning et al. 2018; Oza et al. 2019; Chao et al. 2021). Existing exoplanet detections of planets with potential volcanic activity include CoRoT-7b (Barnes et al. 2010), the first rocky exoplanet discovered and one that is likely heated by strong tidal forces; 55 Cancri e, whose inferred longitudinal offset in peak surface emission has been attributed to (among other things) lava flows on the surface (Demory et al. 2016a, 2016b; Hammond & Pierrehumbert 2017); several planets in the TRAPPIST-1 system that are in a Laplace-like resonance and likely exhibiting volcanic activity (Kislyakova et al. 2017; Dobos et al. 2019); and many others. Not all such exoplanets are expected to have Io-like volcanism. Some will have magma oceans because of their proximity to the star, and others will have volcanism powered by nuclear decay and residual heat, similar to volcanism on present-day Earth.

Io has been extensively observed using both space and ground observatories. High-resolution images of Io’s surface were taken by space missions such as Voyager (Smith et al. 1979), Galileo (Belton et al. 1996), and Juno (Mura et al. 2020). The surface has also been resolved from ground-based observations in the near-infrared using disk resolved imaging (Howell & McGinn 1985; Simonelli & Veverka 1986; Spencer et al. 1990) and adaptive optics observations (Marchis et al. 2000, 2005; de Kleer & de Pater 2016a). Most importantly for this work, starting with Spencer et al. (1990), Io has been sporadically observed over a time span of decades using high-cadence near-infrared photometry taken during occultations by Jupiter. Occultations by Jupiter occur once every orbit of ~1.7 days, whereas occultations by Europa, Ganymede, or Callisto (so-called “mutual occultations”) happen every ~6 yr when Earth passes through the orbital plane of the Galilean satellites. Multiple occultations are then observable over a course of approximately 1 yr. 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. Only the brightest volcanoes are visible over the reflected-light background when Io is illuminated by the Sun (Veeder et al. 1994; de Kleer & de Pater 2016b).

Both kinds of occultations have been used to study volcanic activity on the surface. Spencer et al. (1994) 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. Rathbun et al. (2002) and Rathbun & Spencer (2006, 2010) used observations from NASA’s Infrared Telescope Facility (IRTF) to study the long-term variability of different volcanic spots, finding evidence of periodicity in Loki’s eruptions and establishing the transient nature of observed emission from most volcanoes. By studying a (spatially resolved) occultation of Io by Europa, de Kleer et al. (2017) mapped the Loki Patera (a patera is a type of irregular crater) region to a precision of about 2 km. In addition to observing occultations in the near-infrared, several groups have been observing mutual occultations in the optical for decades with the purpose of inferring the optical albedo of Io and improving ephemeris precision for Galilean satellites (Arlot et al. 1974; Lainey et al. 2009; Morgado et al. 2016; Saquet et al. 2018 and references therein). Understanding the detailed albedo distribution of Io’s surface is crucial to constraining the ephemeris of Io to a very high precision.

Most studies of (unresolved) occultations of Io fit multiple light curves independently with the goal of inferring one-dimensional longitudinal variations in brightness on its surface (notable exceptions are Spencer et al. 1994 and de Kleer et al. 2017), assuming relatively strong priors on the locations of individual volcanoes. In this work, we build a fully probabilistic model in order to infer a two-dimensional map of Io’s surface thermal emission using archival IRTF observations of Io in the near-infrared. We assume that the map we fit to a given set of light curves is static (up to an overall amplitude), an assumption that is justified if the spatial structure and intensity of hot spots does not change substantially between the times of observation. In Paper II in this series, we will relax this assumption by constructing a model that can be used to infer a time-variable map of Io from any number of light curves observed at arbitrary times.

The model relies on the recently developed code *starry* (Luger et al. 2019a, 2021a), which enables fast analytic computation of occultation light curves and phase curves for objects whose surface features can be represented in terms of spherical harmonics. The *starry* code can compute phase curves and occultations in both emitted light (for modeling the isotropic thermal emission from Io’s surface) and reflected light (for mapping albedo variations). It is many orders of magnitude faster and more accurate than pixel-based algorithms, and it computes the exact gradients of the flux with respect to all parameters through automatic differentiation. It also comes with extensive tools for visualizing spherical harmonic maps and simulating data.

Our main goal in this paper is to develop a general model for mapping the surfaces of occulted bodies with sparse, high-contrast features and apply it to observations of Io. Because of the existence of the high-resolution resolved imaging data of Io, we have some knowledge of the ground truth, so we are able to validate our model. The paper is organized as follows. In Section 2, we describe the IRTF light curves of occultations by Jupiter that we use to infer maps. In Section 3, we discuss the generative model for the data, and we write down the likelihood function. In Section 4, we focus on the question of how to do Bayesian inference given the forward model specifications defined in Section 3. We discuss and test the use of different priors on map features and quantify the information content of the occultation light curves. We then fit the model with realistic simulated data using Hamiltonian Monte Carlo (HMC). In Section 6, we show the results for IRTF data for two separate pairs of light curves, one pair observed in 1998 and another in 2017. We show the inferred maps and the parameters quantifying the location and intensity of the hot spots, discuss the sensitivity of the results to different choices of priors, and plot the inferred hot spots overlaid on a high-resolution optical map of Io constructed from Galileo observations. Finally, in Section 7, we summarize the main results of the paper and discuss potential applications of our model to observations of volcanic exoplanets.

The code associated with this paper is publicly hosted in a GitHub repository. All of the figures in this paper were autogenerated using the Azure Pipelines continuous integration service. Icons next to each of the figures link to the exact script used to generate them to ensure the reproducibility of our results.

## 2. Data

The observations of Io we use in this work were first described in Spencer et al. (1994), who began the observing campaign in 1989 and were joined by others over the following years (Stansberry et al. 1997; Rathbun et al. 2002; Rathbun & Spencer 2006, 2010; Tate et al. 2020). These data are currently being organized to be posted on the Planetary Data System (PDS; J. Rathbun, personal communication). We used an early version of this data archive. The final version of the archive will contain on the order of 100 separate light curves of occultations of Io by Jupiter, observed at times when Io was in Jupiter’s shadow (“in eclipse”). The observations for each year consist of a sequence of only ingress (egress) light curves followed by only egress (ingress) light curves. This is because the geometry of the occultations of Io by Jupiter is such that depending on the time of the year, Io is in eclipse during either the ingress or the egress of the occultation.

Since we have restricted the scope of this work to fitting static maps of Io’s thermal emission, we selected light curves that are closely spaced in time so that Io’s surface emission does not drastically change between the observations. In addition, we selected observations of both the ingress and egress of an occultation because Jupiter’s limb sweeps across Io’s projected disk at different angles at ingress and egress; this means that with both kinds of light curves, we can break the degeneracy in the position of the hot spots. The final set of light curves consists of a single pair of ingress/egress observations from 1998 observed 94 days apart and another pair of ingress/egress observations from 2017 observed 54 days apart. We fit one map for each of the two pairs of light curves. In Paper II, we plan to simultaneously fit nearly all light curves in the archive simultaneously.

The 1998 occultations were observed with the NSFCam 1–5 μm camera (Shure et al. 1994), while the 2017 occultations were observed with the SpeX 0.8–5.5 μm spectrograph/imager (Rayner et al. 2003). The pair of observations from 1998 is shown in Figure 1; the observations from 2017 are qualitatively similar. Each light curve covers a time span of ~4 minutes, which is the duration of the ingress (egress) part of the complete occultation. The observing cadence for each light curve was 3:67 μt, assuming relatively strong priors on the locations of individual volcanoes.
curve is about a second, which sets a lower limit for the size of the inferred spatial features (∼15 km). The steplike features visible in Figure 1 correspond to volcanic hot spots coming in or out of view during the course of the occultation, so the light curves clearly encode information about emission features on the surface of Io.

Because the early version of the data archive that is eventually going to be uploaded on the PDS does not contain estimates of error bars for the light curves, we work around this limitation by estimating the error bars from the data (see Section 6.1 for more details). As with all ground-based photometry, the observations are influenced by atmospheric variability, which results in the presence of some correlated noise that needs to be accounted for in the model.

3. Model

In this section, we describe the generative model for the data presented in Section 2. To simulate the observed light curves, we first need to specify the geometry of each occultation event (Section 3.1). We then need to specify what the surface emission of Io looks like in a spherical harmonic basis and compute the theoretical flux at the times of observation using starry (Section 3.3). Finally, we need to specify a noise model for the observed data (Section 3.4).

3.1. Orbital Parameters

To compute the geometry of every occultation, we use the JPL Horizons database, which uses the latest ephemeris for the position of Jupiter (Folkner et al. 2014) and its satellites (Jacobson & Brozovic 2015). The ephemeris is usually accurate to a few kilometers on Io’s surface. We access JPL Horizons through the Python package astroquery (Ginsburg et al. 2019). To compute the relative position of Jupiter and Io, we need the R.A. and decl., and to fix the orientation of Io, we need the longitude and latitude at the center of Io’s disk as seen from Earth (Ob-lon and Ob-lat in astroquery) and the counterclockwise angle between the celestial north pole unit vector projected onto the plane of the sky and Io’s north pole (NP.ang). All longitudes provided by Horizons are positive in the direction of west. Horizons provides all ephemerides with a minimum cadence of 1 minute; we interpolate these values so that we can evaluate them at arbitrary times.

The coordinate system in starry is defined to be right-handed, such that the $\hat{z}$-axis points toward 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 $\text{obl}$ between the $\hat{y}$-axis and the north pole of the sphere; the inclination angle $\text{inc}$, which is set to 90° if the north pole is aligned with the $\hat{y}$-axis; and the phase angle $\theta$, which determines the rotation of the sphere around the $\hat{y}$-axis in the eastward direction. Expressed in terms of Horizons variables, these angles are given by $\text{obl} = \text{NP.ang}$, $\text{inc} = 90^\circ - \text{Ob-lat}$, and $\theta = \text{Ob-lon}$. The occultor position relative to the occulted object is given by

$\text{xo}/\gamma = -\Delta\alpha \cos \delta$, \hspace{1cm} (1)

$\text{yo}/\gamma = \Delta\delta$, \hspace{1cm} (2)

$\text{zo}/\gamma = 1$, \hspace{1cm} (3)

where $\Delta\alpha$ and $\Delta\delta$ are differences in R.A. and decl., respectively, relative to the occulted object, and $\gamma$ is the angular radius of the occulted sphere.
3.2. Jupiter’s Effective Radius

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 greater than its polar radius by thousands of kilometers. Because starry does not (yet) support occultations for nonspherical objects, and because the effect is negligible at the resolution of our maps, we assume that Jupiter is locally spherical at the point of an occultation. We estimate an effective radius from the measurements of Jupiter’s shape from Pioneer and Voyager radio occultation data (Figure 7 of Lindal et al. 1981). When computing the occultation latitude, we fix Jupiter’s latitude to the value at the center of Io’s disk because the variation of Jupiter’s effective radius along Io’s disk is negligible.

Second, Jupiter is gaseous, so it does not have a well-defined boundary. In principle, we should compute an effective radius of Jupiter at different altitudes (pressure) in the atmosphere and model an occultation of Io by a fuzzy occultor. Although this is possible with starry, it is unnecessary for our models because the characteristic scale height of Jupiter is around 27 km, which is below the uncertainty of our inferred maps. We instead follow the approach in Spencer et al. (1990) and compute the effective radius of Jupiter at about 2.2 mbar, the pressure and the associated effective radius) at which a bright source on the surface of Io fades by 50% due to differential refraction during the course of an occultation.

Third, the information on the effective radius in Lindal et al. (1981) is provided only at a fixed pressure of 100 mbar. To adjust the values for a lower pressure of 2.2 mbar, we assume an exponential pressure profile \( P = P_0 e^{-\Delta H/r} \), 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 to 2.2 mbar, we need to add the factor \( -H \ln(2.2/100) \), which is assumed to be constant in the \( \pm 21^\circ \) latitude range in which the occultations occur. In addition to refractive absorption in Jupiter’s atmosphere, there is also slight additional molecular absorption due to methane, which is estimated by a vector of spherical harmonic coefficients \( Y_0^0, Y_2^0, Y_2^{-2}, \ldots \) of Jupiter’s atmosphere, which results in a smaller projected limb of Jupiter on the surface of Io than would be the case for straight propagation. 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 Jupiter planetocentric latitude at which Io’s disk disappears or reappears behind the limb, 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. The shape profile data are quite old, and it is not known if the structure of Jupiter has remained constant since the 1980s. The shape profile also depends on the wind velocity structure and temperature, which we do not take into account. In addition to uncertainties on the atmospheric structure, there are uncertainties associated with digitizing the data shown in Figure 7 in Lindal et al. (1981) because they are not available in table form.

Problems with Galileo’s high-gain antenna prevented it from obtaining more recent radio occultations of the neutral atmosphere that could be used to improve the shape measurements, and Juno’s orbit has thus far avoided such occultations. However, over the 2023–2025 period, an extended Juno mission may be able to obtain radio occultations that would refine our understanding of Jupiter’s size and shape (\( r \)). Since our main focus in this work is the map model, we leave a detailed investigation of the various sources of error that go into the radius estimate for future work.

3.3. Linear Model for the Flux

Given the geometry of an occultation event at the times of observation, computing the predicted flux with starry is straightforward. The starry code computes the integrated flux of an unocculted or occulted sphere analytically given an expansion of the surface map in spherical harmonics \( Y_m^\ell(\theta, \phi) \) up to a certain degree \( J \). The map is defined by a vector of spherical harmonic coefficients \( y \) that is dotted into the spherical harmonic basis vector \( (Y_0^0, Y_1^0, Y_1^{-1}, Y_2^0, Y_2^{-2}, \ldots) \). The total number of spherical harmonic coefficients is \( (J+1)^2 \). 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 in which the occulted object is illuminated by a distant light source (Luger et al. 2021a). In that case, the coefficient vector \( y \) represents spherical albedo. For observations of Io in sunlight in a wavelength range where both the thermal and reflected sunlight components of the observed flux are comparable, it would be necessary to simultaneously fit for two vectors of spherical harmonic coefficients, one for the albedo distribution and the other for thermal emission. In this paper, we only use observations of Io in eclipse, so we do not need to model the reflected-light component.

Conditioned on fixed parameters specifying the geometry of the occultations, the starry model is linear in the spherical harmonic coefficients. Luger et al. (2019a) achieved this by representing all operations needed to compute the flux as linear transformations. Since a sequence of linear mappings is also linear, the predicted flux can be written as

\[ f = A y, \]

where the column vector \( f \) of shape \( (T, 1) \) is the predicted flux for different values of the occultor position, and \( A \) is the design matrix (see Appendix B.1. in Luger et al. 2021a) of shape \( (T, N) \) (with \( N = (J+1)^2 \)), which encodes all operations needed for computing the integrated flux at the different viewing angles of the occulted sphere: relative positions and radii of the occultor and occulted object, rotations, basis transformations, integrals over the visible disk of the occulted object, and the cosine emission law. If the geometry is not known precisely, \( A \) is not fixed, and the model is, in principle, no longer linear, although one can still use the fact that it is linear when conditioned on a particular value of the nonlinear parameters to speed up inference.
The characteristic angular size of the features that can be represented with a given map is set by the degree of the map, and it is approximately equal to 180°/l. For reference, 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 offset from the substellar point), 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). The starry code can handle occultations up to l ≈ 20 before numerical instabilities kick in (Luger et al. 2019a) that correspond to a minimum resolution of 9°, which means that we are not able to constrain the physical size of the spot. We discuss the implication of this resolution limit in Sections 6 and 7.

Although starry was built around the idea of expanding surface features in a spherical harmonic basis in which we can compute all fluxes analytically, this basis may not be ideal for doing inference because it can be difficult to encode assumptions (priors) on what we expect the map to look like. For example, the most important constraint on the map we would like to incorporate in the model is that the intensity (power per unit area) of the map at any given point is positive. This constraint is important not only because we want to avoid having unphysical regions in inferred maps but also because it imposes a very strong prior on the map, which substantially reduces the difficulty of inference when the data are not particularly informative.

There are two ways of enforcing positivity that we are aware of, both of which involve evaluating the map intensity on a fixed grid of pixels (intensity evaluated at fixed set of points on the sphere) and not allowing negative values of those pixels because they are unphysical. The first is to fit for the spherical harmonic coefficients y, evaluate the pixel grid at each Markov Chain Monte Carlo (MCMC) step, and reject all samples of the coefficient vector y that result in pixels with negative intensity. The issue with this procedure is that rejecting samples in this way implies a prior probability distribution on y that cannot be mapped to a parameter space with infinite support, and HMC does not work well with constrained parameter spaces. The second approach is to dispense with the spherical harmonics entirely and compute the full model using a pixelated intensity map. This is very difficult to do in practice because we would need a very high resolution grid to compute the light curves accurately with minimal discretization noise. The starry code uses spherical harmonics as a basis precisely to avoid this problem. Instead, we opt for a hybrid approach in which we fit for a vector of pixel intensities, but at each MCMC step, we convert the pixels to spherical harmonic coefficients to compute the light curve.

To implement the hybrid approach, we need to be able to switch back and forth between spherical harmonics and pixels. Given a vector of pixel intensities p evaluated on an equal area Mollweide grid, the transformation from spherical harmonics y to p is given by a linear operator P:

\[ p = P y. \]  

Each row of P contains values of each of the spherical harmonic coefficients at a given point on the grid. We choose to use the Mollweide grid so that each area element on the sphere has approximately the same number of pixels. The grid needs to be fine enough to ensure that the intensity is positive over most of the sphere.

Switching from pixels to spherical harmonics is somewhat more complicated because P is not generally a square matrix, so we cannot compute its inverse to obtain the inverse transform. Instead, we can compute an approximate inverse (a pseudoinverse) \( P^\dagger \) by solving the linear system \( PP^\dagger = I \), where I is the identity matrix. The solution is given by

\[ P^\dagger = (P^T P + \lambda I)^{-1} P^T, \]  

where \( \lambda \) is a small regularization parameter and I is the identity matrix. The mapping from pixels to spherical harmonics is then given by

\[ y \approx P^\dagger p. \]  

Both P and \( P^\dagger \) can be precomputed to speed up inference. When using pixels to impose a positivity constraint on the spherical harmonic map, we need to make sure that the number of pixels is greater than the number of spherical harmonic coefficients by a factor of a few to ensure positivity approximately everywhere on the sphere. In practice, we find that we need to use at least four times as many pixels as spherical harmonics, which means that the computational cost of this model is larger than for a pure spherical harmonic model.

To summarize, in our hybrid model, we first construct a fixed high-resolution pixel grid in latitude and longitude, then we use Equation (7) to convert a vector of pixel intensities to spherical harmonic coefficients, and finally, we use Equation (4) to evaluate the model. Although we fit for the pixels p, we store the spherical harmonic coefficient vectors y as the final product of the inference. Figure 2 illustrates the transformation from the pixel basis to the spherical harmonic basis via \( P^\dagger \). On the left, we show the pixel map, where each pixel was independently drawn from an exponential prior. On the right is the same pixel map transformed to a spherical harmonic basis via \( P^\dagger \). The histograms under each map show the distribution of intensities. Since the pixel map is higher resolution than the spherical harmonic map, it can only be approximately represented at a finite order of the spherical harmonic expansion, so the map on the right appears to be smoother, and the intensity distribution is more similar to a skewed Gaussian with a heavy right tail than an exponential distribution. Nevertheless, we find that setting priors on pixels is a far better solution than fitting the spherical harmonic coefficients directly and is worth the extra computational cost that comes with the increased dimensionality of the parameter space (see Section 4.2 for a demonstration).

### 3.4. The Likelihood

Finally, we have to specify the noise model, which means we have to define a likelihood function. Assuming we have a single light curve with T data points and a map defined by the pixels p, the (Gaussian) log likelihood is given by

\[ \ln L = -\frac{1}{2} [f_{\text{obs}} - f] \Sigma^{-1} [f_{\text{obs}} - f], \]  

where \( f_{\text{obs}} \) is the observed light curve, \( \Sigma \) is the data covariance matrix, and f is the predicted flux given by

\[ f = A' p + b 1_T, \]  

where \( A' = AP \), and \( b \) is a constant flux offset parameter that we have added to account for stray flux that cannot be
attributed to Io. Depending on the observation, this flux is most often residual light from Jupiter.

To model the data covariance $\Sigma$, we use use a Gaussian process, and we compute the likelihood using the fast Celerite method (Foreman-Mackey et al. 2017) as implemented in the celerite2 package (Foreman-Mackey et al. 2017; Foreman-Mackey 2018). We use the simple (approximate) Matérn 3/2 kernel function that is parameterized by two values, a standard deviation parameter $\sigma_{\text{GP}}$ and a characteristic timescale parameter $\rho_{\text{GP}}$. The Matérn 3/2 kernel is defined by

$$k(\tau; \sigma_{\text{GP}}, \rho_{\text{GP}}) = \sigma_{\text{GP}}^2 \left( 1 + \frac{\sqrt{3} \tau \rho_{\text{GP}}^-}{\sigma_{\text{GP}}^2} \right) \exp \left( -\frac{\sqrt{3} \tau \rho_{\text{GP}}^-}{\sigma_{\text{GP}}^2} \right),$$

where $\tau = |t_n - t_m|$. A single element of the data covariance matrix $\Sigma$ is then

$$\Sigma_{nm} = \sigma_n^2 \delta_{nm} + k(\tau; \sigma_{\text{GP}}, \rho_{\text{GP}}),$$

where $\sigma_n$ is the error bar for the $n$th data point (a free parameter). Since we fit multiple independent light curves, the total log likelihood is the sum of the individual likelihoods defined in Equation (8).

4. The Inverse Problem

Having defined a probabilistic model that describes how to compute a realistic light curve for an occultation of Io in the previous section, in this section, we discuss the inverse problem of inferring a surface map by fitting a set of occultation light curves in a Bayesian framework.

4.1. The Information Content of a Light Curve

The mapping problem is famously ill posed, meaning that specific linear combinations of spherical harmonic coefficients will be in the null-space of the linear mapping $A$ in Equation (4) (Luger et al. 2021b). This means that in general, even if we had noiseless observations, it would still be impossible to recover certain features on the surface. To recover the most information about the surface, we need to have a mechanism that breaks the various degeneracies. For example, with phase curves, we can recover primarily longitudinal variations in emission. Occultations are substantially better because the limb of the occultor sweeps across the surface of the occulted sphere, thereby exposing or blocking light from different points on the surface. An ideal set of observations would consist of phase curves together with observations of multiple occultations by a small occultor at different latitudes and phases. Phase curves and occultations in reflected light are even more informative because of the nonuniform illumination profile of the incident radiation and the presence of a day/night terminator line (Luger et al. 2021a). In some cases, for reflected-light observations (the phase curves of an inclined planet, for example), there can even be no null-space at all low spherical harmonic degrees.

We can reformulate these statements on how useful given observations are more precisely by computing a measure of their information content. Given that our model is linear, assuming Gaussian priors on the spherical harmonic coefficients with covariance $\Lambda_\nu$ and a Gaussian likelihood, the posterior can be computed analytically and its mean is given by

$$\hat{y} = \Sigma_\beta (A^\top \Sigma_f^{-1} A + \Lambda_\nu^{-1})^{-1} A^\top \Sigma_f^{-1} f + \Lambda_\nu^{-1} \mu_\nu, \quad (12)$$

where the posterior covariance matrix $\Sigma_\beta$ is

$$\Sigma_\beta = (A^\top \Sigma_f^{-1} A + \Lambda_\nu^{-1})^{-1}. \quad (13)$$

We define the information content as the variance reduction of the posterior relative to the prior, which is called posterior shrinkage. The posterior shrinkage $S$ is defined as (Betancourt 2018; Luger et al. 2021a)

$$S \equiv 1 - \lim_{\sigma_0^2 \to \infty} \frac{\sigma_0^2}{\sigma^2}, \quad (14)$$

where $\sigma_0^2$ is the prior variance and $\sigma^2$ is the posterior variance for a given spherical harmonic coefficient. It tells us how well we can constrain a particular coefficient in the limit of infinite signal-to-noise ratio ($S/N$) observations. Posterior shrinkage of
Figure 3. Posterior shrinkage for different kinds of observations of Io as a function of spherical harmonic degree (angular scale), averaged across all m modes. Posterior shrinkage of 1 represents maximum information gain in updating from the prior to the posterior, while zero represents no information gain. The posterior variance has been computed 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).

1 indicates that the data provide perfect information on the parameters, while zero indicates no gain in information relative to the prior.

In order to compute the posterior shrinkage, we first compute the design matrices $A$ using starry for different kinds of observations of Io for the period starting on 2009 January 1 and ending on 2010 May 1. We chose this period because it covers the full season of mutual occultations that occur every 6 yr. We compute $A$ for all observable occultations of Io by Galilean moons during that period, occultations of Io by Jupiter, and phase-curve observations. The purpose of this is to determine the upper bound on what we can learn about the surface. We take the ephemeris from JPL Horizons and assume that it is known exactly; we also assume that all observations are observations of thermal emission independent of whether Io is in sunlight or eclipse because we are interested in constraining the volcanic emission rather than the albedo.

Figure 3 shows the posterior shrinkage as a function of $l$ (averaged over all $m$ modes) for phase-curve observations (blue lines), occultations by Jupiter (orange), the former two combined (green), and mutual occultations by other Galilean moons (red). As expected, the mutual occultations of Io by other Galilean moons are by far the most informative, with posterior shrinkage of unity at all angular scales considered. Occultations by Jupiter are less informative because we only see one side of Io during an occultation. Shrinkage for phase curves at odd degrees above $l = 2$ is exactly zero because these coefficients are in the null-space for objects rotating about an axis perpendicular to the line of sight and therefore cannot be constrained using only phase curves (Luger et al. 2021b). Although observations of mutual occultations most easily break the degeneracies because of the varying impact parameters and different sizes of occultors, the drawback of these types of observations is that they only happen every 6 yr, and they almost never happen while Io is in eclipse, which means that only the brightest volcanoes are visible above the reflected sunlight. This fact is not captured in Figure 3.

Figure 3 gives us some idea about which kinds of observations are most informative, but it does not really tell us how well we can constrain the bright spotlike features we expect to see on Io. To answer this question, we have to create a simulated data set and conduct the whole inference process.

4.2. Pixels versus Spherical Harmonics

We use starry to generate a single simulated light curve of an occultation of Io by Jupiter from an $l = 30$ map with known coefficients. The simulated map consists of a spherical harmonic expansion of a bright spot with a Gaussian profile, which we add to a uniform brightness map 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°N latitude and 51°E longitude, and we set the diameter of the spot (2$\Delta \theta$) to 5° and the amplitude of the spot such that the total luminosity of the map increases by 50% with the addition of the spot. We generate two light curves with 150 data points each, one for the duration of the ingress of the full occultation and the other for the egress. Because the limb of the occultor sweeps over the disk of Io at different angles during ingress and egress, this makes it possible to break most degeneracies in the map and recover the location of the simulated spot. We set the phase of the simulated map to be 10°E at the beginning of ingress and 10°W at the end of egress. We assume that the geometry of the occultation is known exactly, and we set the error bars such that $S/N = 50$, where the signal is defined to be the maximum value of the computed flux. We use this data set to test the difference between setting a prior in the spherical harmonic basis $y$ and in the pixel basis $p$ by fitting an $l = 20$ map to the data set.

In the spherical harmonic model, we place a Gaussian prior on $y$ with covariance $A = \text{diag}(1^2, 0.5^2, \ldots, 0.5^2)$. Since the model is linear and the prior is Gaussian, the posterior probability distribution is also Gaussian, and we can solve for the posterior mean (Equation (12)) and covariance (Equation (13)) analytically. In the hybrid pixel model, we place a positive exponential prior on the pixels that are defined on a Mollweide grid. The purpose of the exponential prior is to favor sparser solutions for the map because the exponential distribution pushes most pixels
toward zero intensity. We use four times as many pixels as spherical harmonics. Although the model is also linear in this case, the posterior distribution for the pixels is not analytic because of the non-Gaussian prior, so we sample the posterior using MCMC instead. As the end product of inference, we save the spherical harmonic coefficients \( y = P^T p \) rather than the pixels themselves. The coefficients \( y \) can then be used to evaluate the map on a pixelated grid of arbitrary resolution via the matrix \( P \).

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 5000 samples from the prior on \( y \) and evaluate \( P y \) for each; we then compute the standard deviation of these pixels and use that as the scale parameter in the exponential prior. We implement the pixel model in the probabilistic programming language NumPyro (Phan et al. 2019), which is built on top of the JAX (Bradbury et al. 2018) library, and we fit it using HMC with the No-U-Turn Sampler (NUTS; Hoffman & Gelman 2014). JAX is a numpy-like library that supports automatic differentiation, parallelization, and GPUs. We run the chains for 1000 tuning steps and 2000 final steps, monitoring divergences (Betancourt & Girolami 2013) and the R-hat diagnostic (Gelman & Rubin 1992) to check for convergence. Since all models we fit in this paper have hundreds if not thousands of parameters, it would be extremely challenging to sample the posterior without the use of automatic differentiation and HMC (at least for non-Gaussian priors).

To visualize the inferred maps, we plot the heat map of the median intensity at each point on the map computed from posterior samples. Results are shown in Figure 4. The first row shows the simulated map, and beneath it, we show the inferred maps in Mollweide projection for the two models (second row), the data and posterior flux samples (third row, orange lines), and the residuals with respect to the median flux (fourth row). The difference between the two models is striking. The left map has an elongated feature that does not resemble the spot in the simulated map, while the map on the right is nearly identical (except for a difference in intensity) to the simulated map. We should emphasize here that the fact that the pixel model results in a spotlike map is mostly a consequence of the exponential prior, which favors sparse solutions in pixel space. When we compared the spherical harmonic model to the pixel model using a prior with a lighter tail, such as a half Gaussian (Gaussian truncated at zero), we obtained a more elongated feature similar to that shown in the left map in Figure 4, but the map was still noticeably less complex because the positivity constraint substantially reduces the space of maps that fit the data well. Thus, using the pixel model makes it easy to impose the positivity constraint on the map, as well as other constraints, such as sparsity.

An important issue with the map on the right is that there is a series of concentric rings around the spot that result in a wavelike pattern in the predicted flux and residuals. This ringing pattern arises because representing spotlike features requires constructive interference between different spherical harmonic modes inside the spot and destructive interference elsewhere. The pattern is more pronounced when we fit a low-resolution map (approximately 180/20 = 9° in this case) and the model tries to represent a feature below the resolution of the map.

Ringing is also the reason why the inferred spot for the pixel model appears to be noticeably dimmer than the simulated spot. There is nonnegligible leakage of total flux from the spot into the rings surrounding it, meaning that if we were to integrate the map intensity over a region encompassing the brightest part of the inferred spot, it would be an underestimate of the total emitted flux from the spot within the same area. Ringing is also 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 in some cases when we fit low-degree maps, the model will place a bright spot on the unobserved side of Io in order to produce a ringing artifact on the observed side to explain an increase or decrease in brightness in the light curve. Fortunately, convolving the map with a spatial smoothing filter prior to evaluating the flux fixes this issue. We describe how to apply the smoothing filter in the following section.

4.3. Smoothing Out Spurious Features

To suppress ringing artifacts that appear around inferred spots such as the one shown in Figure 4, we apply a spatial smoothing filter to the spherical harmonic coefficients. 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 use a Gaussian-like kernel function given by

\[
B(\theta) = \frac{\exp(-\theta^2/2\sigma_x^2)}{2\pi\sigma_x^2},
\]

where \( \sigma_x \) is a parameter that sets the characteristic scale of the smoothing. This function can be expanded in terms of spherical harmonics as

\[
B(\theta) = \sum_{l=0}^{\infty} \left( \frac{2l + 1}{4\pi} \right) B_l P_l(\cos \theta),
\]

where \( B_l \) are the spherical harmonic coefficients and \( P_l \) are the associated Legendre polynomials. They depend only on \( l \) because all nonzero \( m \) modes vanish due to azimuthal symmetry. For \( \sigma_x \ll 1 \), \( B_l \) can be approximated as (White & SREDNICKI 1995; Seon 2007)

\[
B_l \approx \exp\left[-\frac{1}{2}l(l + 1)\sigma_x^2\right].
\]

The effect of this filter is to exponentially suppress features on scales smaller than \( l \sim \sigma_x^{-1} \).

Figure 5 shows the effect of (Gaussian) smoothing on a spherical harmonic expansion of a spot with a Gaussian intensity profile. We place the spot at 0° latitude and longitude and set the size of the spot \( \Delta \theta \) to 5°. All three panels show the exact profile of the spot (black line) and expansions up to three different orders \( l \) (colored lines). The panel on the left shows the expansion with no smoothing (\( \sigma_x = 0 \)), in which case the symmetric ringing around the center of the spot is clearly visible even at relatively high order \( l = 20 \). The middle panel shows an intermediate level of smoothing with \( \sigma_x = 0.1 \), meaning that all features on scales above \( l \approx 10 \) are exponentially suppressed. 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. For \( \sigma_x = 0.2 \) (right), there
is practically no ringing, but the expansions at \( l = 10, 15, \) and 20 result in a spot of the same size because all coefficients above \( l = 5 \) are significantly suppressed.

Thus, there is a trade-off between smoothing and the ability to resolve smaller-scale features in maps. In principle, we can always get rid of ringing by fitting sufficiently high-order maps. In practice, the analytic integrals computed in \textsc{starry} become computationally unstable above \( l \approx 20 \), so instead of going to very high order, we apply some smoothing to mitigate the ringing. We find that setting \( \sigma_s = 2/l \), where \( l \) is the order of expansion of the map, is a good default setting for \( \sigma_s \).

5. Results: Simulated Data

5.1. Fitting Simulated Ingress/Egress Light Curves

In this section, we generalize the example from Section 4.2 such that the simulated map includes an additional faint hot spot located at \(-15^\circ\text{N}\) latitude and \(-40^\circ\text{E}\) longitude with a
diameter of 5°. We set the amplitude of the “faint” spot to 30% of the luminosity of the featureless map. As before, we assume that we know the geometry of the occultation exactly. We generate the light curves from an $l = 30$ map and fit an $l = 20$ map, and we apply a Gaussian smoothing filter to both the simulated map and the inferred map with the smoothing parameter set to $\sigma_s = 2/l = 0.1$. We fit the model using NUTS with 1000 warm-up steps and 2000 final steps.

We found that in cases where the simulated map consists of a bright and a faint spot, the pixel model with an exponential prior such as the one we used in Section 4.2 does not recover the faint spot. Instead, we use a different heavy-tailed prior called the regularized horseshoe prior7 (also known as the “Finnish horseshoe”; Piironen & Vehtari 2017). This prior is specifically designed for use in Bayesian sparse linear regression. It is an improvement on the horseshoe prior introduced in Carvalho et al. (2010). The key idea behind both kinds of horseshoe priors is to set the scale for each regression coefficient (pixel) to a product of a global scale $\tau$ and a local scale $\lambda_i$ (where $i$ indexes all of the pixels), and we marginalize over these scales by setting a prior for each. For clarity, we omit the discussion of the horseshoe priors here and refer the reader to the Appendix.

The results for a high-S/N light curve ($S/N = 50$) are shown in Figure 6; the model recovers both spots. The figure shows the simulated map (first row), the median posterior estimate of the inferred map (second row), the inferred map as seen by the observer during the occultation (small circles), the data and posterior samples of flux (orange lines), and the residuals with respect to a median estimate of the flux (fifth row). The locations of the simulated spots on the inferred map are marked with gray crosses. The bright spot is nearly indistinguishable from the simulated spot; the fainter spot is somewhat less well constrained, but the error in position for both is at most a few degrees. Ringing artifacts are minimal because of the smoothing filter, and there are no discernible patterns in the residuals. We found that without the horseshoe priors, the model was not able to capture the first step in the light curve, which is due to the fainter spot coming in or out of view; it would only recover the brighter spot. In Figure 7, we show the output of the same model assuming we have data of worse quality ($S/N = 10$). In this case, the model still recovers both spots, but the error in the position of the spots is greater.

5.2. Comparison to a Parametric Model

It is instructive to compare the model developed in the previous section to a parametric model, in which we assume that the map contains some fixed number of spots and fit for their positions, amplitudes, and sizes. A parametric spot model might be useful if we want to fit for a small number of spots and we know what the map should look like. To test how the parameterized spot model compares to the model defined in Section 5.1, we fit it to the light curves shown in Figure 6. The model consists of four parameters: the latitude, longitude, amplitude, and size of the spot. We place uniform priors on the angles and positive Gaussian priors on other parameters. We find that if the number of modeled spots matches the number of simulated spots, the model easily converges to the true solution. If the number of modeled spots is larger or smaller than the true number of spots, the model does not converge to the true solution due to pathologies in the posterior distribution.

Although a parametric approach would likely have been sufficient for modeling the light curves shown in this work, it does not work well when the number of spots is unknown or if the features are not spots. There is no need to restrict ourselves to a parametric model because the pixel model gives the same results with only weak assumptions about the global structure of the map. The computational advantages of the parametric spot model relative to the pixel model are minimal. Even though the spot model has only five parameters per spot and the pixel model has thousands of parameters, the runtime for the pixel model is longer only by a factor of a few.

---

7 To be precise, we use the truncated version of the regularized horseshoe prior, where the coefficients are required to be positive.
6. Results: IRTF Light Curves

6.1. 1998 Pair of Occultations by Jupiter

Having demonstrated that our model works well on simulated data, we turn to fitting observations of actual occultations of Io observed with the IRTF. We selected two pairs of high-quality ingress/egress light curves relatively closely spaced in time so that the assumption that the surface map is identical for both the ingress and egress light curves is at least approximately correct, although we allow for a difference in the overall amplitude of the maps between the two occultations. We fit a pair of events from 1998, an ingress occultation observed on August 27 and an egress occultation observed 94 days later on November 29. As a reminder, the period of Loki’s variability is around 540 days (Rathbun et al. 2002). For comparison, we also fit a pair of events observed nearly two decades later in 2017.

The predicted fluxes for the ingress and egress of the occultation are given by

$$f_I = A_I p_I + b_I 1_{N_I}$$

(18)
and

\[ f_E = A_E' p_E + b_E \mathbf{1}_{N_E}, \]

where \( f_I \) is the predicted flux for the ingress light curve with \( N_I \) data points, \( f_E \) is the predicted flux for the egress light curve with \( N_E \) data points, and \( b_I \) and \( b_E \) are the constant flux offset parameters. Since we assume that the map is the same for both light curves up to an overall amplitude difference, we have

\[ p_E = a p_I, \]

where \( a \) is a dimensionless parameter.

The total log likelihood is the sum of the log likelihoods for individual light curves (Equation (8)). To compute it, we need to specify the data covariance matrix defined in Equation (11). Each element of the covariance matrix consists of a white noise component (error bars) and a correlated noise component (Gaussian process). Because the IRTF light curves from the preliminary archive of occultation light curves do not contain robust estimates of the error bars, and the number of data points per light curve is quite small, we choose to treat each error bar as a free parameter and fit for all error bars using a hierarchical model. We assume that there is a common global scale \( l \) for all data points in a given light curve and that the error bar for each data point is drawn from a half-normal distribution with a scale parameter equal to \( l \). In other words, we have

\[ \sigma_{i,t} \sim \text{half} - \mathcal{N}(l_t), \quad i = 1, \ldots, N_t \]

Figure 7. Same as Figure 6 but for light curves with S/N = 10.
and

\[ \sigma_{E,i} \sim \mathcal{N}(l_E), \quad i = 1, \ldots, N_{E}, \]  

(22)

where \( \sigma_{i,E} \) are the error bars for the ingress light curve, \( \sigma_{E,i} \) are the error bars for the egress light curve, \( l_I \) is the common scale for the ingress light-curve error bars, and \( l_{E} \) is the scale for the egress light-curve error bars.

By fitting for error bars in this way, we do not have to assume any parametric forms for the dependence of the variance of the data on the measured flux, and we are also able to account for outliers and heteroscedasticity in the data. Although this means that we are introducing hundreds of additional parameters into the model, this is not a problem for the HMC sampler when the parameters are well constrained by the data and the model is regularized. The noise model as defined above is extremely flexible; a given feature in the light curve can be accounted for by inflating the error bars or varying the timescale or the variance of the Gaussian process with the Matérn 3/2 kernel.

All model parameters and their associated priors are listed in Table 1. To fit the model, we sample the posterior using the NUTS sampler for 1000 warm-up steps and 3000 final steps, monitoring divergences and the R-hat statistic to ensure convergence. The inferred map for the 1998 pair of events is shown in Figure 8, and the median and 16th and 84th percentile estimates of the model parameters from the posterior samples are shown in Table 2. The inferred map has two distinct hot spots: a bright hot spot in the eastern hemisphere and a faint hot spot in the western hemisphere. The orange lines visible in the third row of panels are posterior samples for flux from the complete model (map model plus the Gaussian process noise model).

The bottom part of Figure 8 shows the residuals with respect to the median flux, where each of the data points is shown with the posterior median estimate of its error bar. Both the global scale for the error bars and the individual error bars are well constrained by the data, and the outlier points are naturally accounted for in our model because those points end up having higher variance. Despite the fact that our noise model is very flexible, it does not seem to overfit the data, because the two major steps in the light curves do not end up being absorbed by the noise model. The Gaussian process captures the variation in the data that is due to atmospheric variability, the emission due to surface features that are too faint to be well constrained by the physical model, and the limitation of the physical model to capture the true size of the spot. There is a trade-off between the explanatory power of the noise model and the physical model. In general, we expect that a given feature in the light curve will be accounted for by the physical model if the physical model provides a better explanation for the data than the noise model. We will return to this point shortly.

To obtain some measure of the location of the inferred spots and its uncertainty, for each posterior sample of the spherical harmonic coefficients \( y \), we find the local maximum of intensity in a region around each of the spots and then compute percentile estimates of the spot latitude and longitude. In addition to the locations of the spots, we also compute the total power emitted within a 15° range in latitude and longitude around the (inferred) center of the spots. Both of these quantities are listed in Table 3. Figure 9 shows a contour map of both hot spots overlaid on the U.S. Geological Survey’s map of the surface of Io (Williams et al. 2011), which was constructed from observations by the Galileo satellite. The plot shows the error for the inferred latitude and longitude for each spot (white lines) and a contour map computed from the median posterior estimate of the map. The contour lines correspond to the 5th, 50th, and 95th percentiles of intensity above an arbitrarily defined intensity of the “background” region around the spot.

The location of the bright hot spot is \( 16.39^{+0.03}_{-0.01} N \) latitude and \( 306.90^{+0.01}_{-0.00} W \) longitude, which corresponds to the location of Loki Patera. The faint hot spot is at \( -16.84^{+0.04}_{-0.02} N \) latitude and \( 37.4^{+0.3}_{-0.2} W \) longitude, which is the Kanehekili Fluctus lava flow. The error for the inferred position of peak intensity of the Loki Patera hot spot is only a fraction of a degree, which is much smaller than the minimum resolution of the map features set by the degree of the map. This corresponds to an uncertainty of less than a kilometer on the surface of Io. Thus, despite the fact that at \( l = 20 \), the resolution of the maps is limited to about 9°, we can constrain the centroid of the spots much more precisely. We should note, however, that this uncertainty does not include a possible systematic shift in position due to a wrong estimate of Jupiter’s effective radius (see Section 3.2) or an error in the ephemeris data.
Figure 8. Inferred \( I = 20 \) maps obtained by fitting a pair of observations of occultations of Io by Jupiter in 1998. The observations were made several months apart with the NASA IRTF. We fit a single map to both observations simultaneously, although we allow for a difference in the overall amplitude of the map between ingress and egress. The model includes a Gaussian process to account for correlated noise caused by atmospheric variability and the fact that our limited-resolution map cannot fully capture the sharp steps in data. We treat all error bars as random variables and plot the median posterior estimates of those error bars. The plot shows the inferred maps (first row), the same maps from the perspective of the observer during the occultation (small circles), the light curves and posterior samples of the flux including the Gaussian process (orange lines), and the residuals with respect to a median flux estimate. The maps show two hot spots; the bright one is emission from Loki Patera, and the faint one is most likely emission from Kanehekili. A detailed view of the two hot spots is shown in Figure 9.

| Table 2 | Derived Properties of the Two Hot Spots Visible in Figure 8 |
|---------|---------------------------------|
| Parameter | Value | Unit |
| Spot 1 latitude | 16.30 ± 0.01 | deg |
| Spot 1 (west) longitude | 306.90 ± 0.01 | deg |
| Spot 1 power ingress | 52.1 ± 1 | GW/um |
| Spot 1 power egress | 60.1 ± 1 | GW/um |
| Spot 2 latitude | 16.4 ± 0.6 | deg |
| Spot 2 (west) longitude | 37.4 ± 0.4 | deg |
| Spot 2 power ingress | 5.4 ± 0.8 | GW/um |
| Spot 2 power egress | 6.2 ± 0.8 | GW/um |

Note. The latitude and longitude of the spots are estimated by finding the peak intensity of each spot for each posterior sample. The power of the spot is defined as the total emission from within a 15° circle centered at the inferred location of the spot.

Table 2

Inferred Parameters for the 1998 Pair of Occultations Observed Using the IRTF

| Parameter | Description | Value | Unit |
|-----------|-------------|-------|------|
| \( \tau \) | Global pixel scale | 0.9 ± 0.2 | GW/um |
| \( c \) | Slab scale | 4000 ± 2000 | GW/um |
| \( a \) | Relative change in map amplitude | 1.15 ± 0.01 | Dimensionless |
| \( b_I \) | Flux offset ingress | 0.01 ± 0.01 | GW/um/sr |
| \( b_E \) | Flux offset egress | 0.3 ± 0.4 | GW/um/sr |
| \( l_I \) | Error bar scale ingress | 0.25 ± 0.04 | GW/um/sr |
| \( l_E \) | Error bar scale egress | 0.46 ± 0.04 | GW/um/sr |
| \( \sigma_{GP,I} \) | GP standard deviation ingress | 0.63 ± 0.08 | GW/um/sr |
| \( \sigma_{GP,E} \) | GP standard deviation egress | 0.45 ± 0.05 | GW/um/sr |
| \( \rho_{GW,I} \) | GP timescale ingress | 0.08 ± 0.01 | Minutes |
| \( \rho_{GW,E} \) | GP timescale egress | 0.01 ± 0.02 | Minutes |
To see how the above results change if we do not include a Gaussian process in the noise model, we fit the same pair of occultations from 1998 using the same priors; the results are shown in Figure 10. The main difference in the inferred map compared to Figure 8 is that there are two extra spots visible. The two spots in the eastern hemisphere are the result of the model struggling to explain the main feature in the light curve at around 3.2 minutes in the ingress light curve and 3.8 minutes in the egress light curve that is due to Loki Patera coming out and into view during the occultation. The model inflates the error bars around those times because it is unable to make the spot small enough; this results in the single hot spot at the location of Loki Patera shown in Figure 8 morphing into two spots. While this feature is almost certainly spurious, the other spot in the northwestern hemisphere is more likely to be real. Looking at the miniature maps in Figure 10, this spot is in view only at egress, and it corresponds to a small step in the light curve starting at 0.9 minutes. The estimated location of this spot is approximately ∼20°N latitude and ∼69°W longitude.

To test how the above results change if we do not include a Gaussian process in the noise model, we fit the same pair of occultations from 1998 using the same priors; the results are shown in Figure 10. The main difference in the inferred map compared to Figure 8 is that there are two extra spots visible. The two spots in the eastern hemisphere are the result of the model struggling to explain the main feature in the light curve at around 3.2 minutes in the ingress light curve and 3.8 minutes in the egress light curve that is due to Loki Patera coming out and into view during the occultation. The model inflates the error bars around those times because it is unable to make the spot small enough; this results in the single hot spot at the location of Loki Patera shown in Figure 8 morphing into two spots. While this feature is almost certainly spurious, the other spot in the northwestern hemisphere is more likely to be real. Looking at the miniature maps in Figure 10, this spot is in view only at egress, and it corresponds to a small step in the light curve starting at 0.9 minutes. The estimated location of this spot is approximately ∼20°N latitude and ∼69°W longitude.

This location corresponds to the southern end of the mountain Mongibello Mons, but there are no known persistent hot spots at that location, so we cannot say if the hot spot is real or not without independently detecting it in other light curves. Overall, we conclude that including the Gaussian process prevents the physical model from overfitting the data, although it might occasionally pick up a feature that is due to a faint real hot spot. With a particular scientific goal in mind, it is straightforward to experiment with different noise models to test how the properties of any given spot change with different assumptions.

One other notable feature in Figure 10 is that the first step in the ingress light curve at around 0.6 minutes is not fully accounted for by the model. Since the model had no trouble accounting for a similar step for the simulated data shown in Figure 6, the reason it did not do so in this case is likely that doing so would result in a poorer fit for the egress light curve. The two occultations were observed months apart, so we expect that our assumption that the map has not changed except for the overall amplitude is wrong in detail.

6.2. 2017 Pair of Occultations by Jupiter

To test how our model generalizes to a different data set, we fit a pair of observations from 2017, an ingress occultation observed on March 31 and an egress occultation observed 41 days later on May 11. We use the exact same model as the one we used to produce Figure 8. The results are shown in Figure 11, and the inferred parameters are listed in Table 4. A contour plot of the two spots overlaid on a surface map of Io from Galileo observations is shown in Figure 12, and the derived spot parameters are listed in Table 5. The inferred map shows two spots, one of which is Loki Patera. Figure 12 shows that the peak emission from Loki Patera is again constrained very precisely, but the location of the peak appears to have shifted southward since 1998. The location of the other hot spot corresponds to Janus Patera.

7. Discussion

7.1. Occultation Mapping of Io

We have presented a novel method for mapping the volcanic emission on Io using occultation light curves. The method
relies on the *starry* algorithm, which enables fast analytic computation of occultation light curves by expanding the surface emission map in spherical harmonics. Our method is different from past work because we do not assume that we know where the volcanic features are located on the surface of Io, how many there are, or what they look like. Instead, we place weaker assumptions on the global structure of the map by requiring that the inferred map have positive intensity everywhere and, most importantly, that it is sparse.\(^8\) Besides the model for the surface map, our method also incorporates a sophisticated noise model that includes a Gaussian process and a hierarchical model for the error bars. As a result of the sparsity and positivity constraints and the flexibility of the noise model, our model is parsimonious—it places features on the map only if the data provide strong evidence for the existence of those features. This property is not always desirable, but it works very well for Io, a moon whose surface is covered with small, highly localized, bright volcanic hot spots. Our model is substantially more flexible than parametric methods that assume some fixed number of spots on the surface of Io and parameterize the spot properties because we do not need to make such strong assumptions about the surface features. The computational cost is small, and it is comparable to parametric models.

To test our method, we first fit a simulated data set and then observations of real occultations of Io by Jupiter observed using NASA’s IRTF. We choose two pairs of light curves to demonstrate that the model can recover known hot spots. Each pair consists of an ingress observation and an egress observation of an occultation. The two observations are sufficient to break the degeneracy in the position of the spots because Jupiter’s limb sweeps across the projected disk of Io at different angles at ingress and egress. From the 1998 observations, we infer a map consisting of two spots whose locations correspond to well-known hot spots on the surface of Io, the major volcano Loki and the lava flow Kanehekili. We also find circumstantial evidence for a third hot spot. Because our model is fully probabilistic, we can derive uncertainties on the location of the inferred hot spots and the total flux emitted from each spot. In addition to the pair of observations from 1998, we test the model on another pair of observations from 2017. We again find two hot spots, one of which is Loki Patera (although the peak of the

\(^8\) Aizawa et al. (2020) also found the sparsity assumption useful in the context of exoplanet mapping.

---

**Figure 10.** Same as Figure 8, except this figure shows the output of a model that does not include a Gaussian process to account for correlated structure in the data. As a result, the model tries to capture the correlations in the data by placing two extra spots on the map and inflating the error bars. At least one of the two extra spots is artificial. \(^8\)
intensity shifted relative to the 1998 map), and the other is Janus Patera.

The main limitation of our model, besides the fact that we do not yet account for time-dependent maps, is the limited resolution of the spherical harmonic maps for which starry can compute occultation light curves. Although we can still constrain the peak intensity of hot spots (or some other measure of the center of emission) with much higher precision, we lose the ability to resolve two spots close to each other and constrain the actual size of the spots. It is possible that there is a way around this problem if we compute the various integrals in starry numerically using high-precision arithmetic. Since all of these operations are contained in the design matrix $A$ (assuming that the ephemeris is fixed), we would only need to do the expensive computation once for each occultation. To be able to constrain the size of the Loki Patera hot spot, for example, we would need to fit maps on the order of $l \approx 50$. However, the

### Table 4

| Parameter      | Description          | Value       | Unit          |
|----------------|----------------------|-------------|---------------|
| $\tau$         | Global pixel scale   | 1.0$^{+0.2}_{-0.2}$ | GW/um        |
| $c$            | Slab scale           | 6000$^{+2000}_{-2000}$ | GW/um        |
| $a$            | Relative change in map amplitude | 1.43$^{+0.01}_{-0.01}$ | Dimensionless |
| $b_I$          | Flux offset ingress  | 0.03$^{+0.02}_{-0.02}$ | GW/um/sr     |
| $b_E$          | Flux offset egress   | 1.0$^{+0.3}_{-0.3}$ | GW/um/sr     |
| $l_I$          | Error bar scale ingress | 0.97$^{+0.06}_{-0.06}$ | GW/um/sr     |
| $l_E$          | Error bar scale egress | 0.52$^{+0.09}_{-0.09}$ | GW/um/sr     |
| $\sigma_{GP,I}$ | GP standard deviation ingress | 0.13$^{+0.04}_{-0.04}$ | GW/um/sr     |
| $\sigma_{GP,E}$ | GP standard deviation egress | 0.60$^{+0.08}_{-0.08}$ | GW/um/sr     |
| $\rho_{GP,I}$  | GP timescale ingress | 2.0$^{+1}_{-2}$ | Minutes       |
| $\rho_{GP,E}$  | GP timescale egress  | 0.02$^{+0.01}_{-0.01}$ | Minutes       |
Figure 12. Contour plot of the hot spots shown in Figure 11 overlaid on the U.S. Geological Survey’s map of the surface of Io, which was constructed from observations by the Galileo spacecraft. The left hot spot is centered around Loki Patera, and the right hot spot is centered around Janus Patera. The contour lines show the 5th, 50th, and 95th percentiles of intensity above an arbitrarily defined intensity of the “background” region around the spot. The asymmetric white cross in the center of the contours in each panel shows the uncertainty in the inferred position of the peak intensity of the hot spot (barely visible in the left panel).

### Table 5

| Parameter                  | Value       | Unit   |
|----------------------------|-------------|--------|
| Spot 1 latitude            | 12.19°±0.01 | deg    |
| Spot 1 (west) longitude     | 307.49°±0.01| deg    |
| Spot 1 power ingress       | 73.4°±0.8   | GW/um  |
| Spot 1 power egress        | 105°±1      | GW/um  |
| Spot 2 latitude            | -4°±5      | deg    |
| Spot 2 (west) longitude     | 41°±4       | deg    |
| Spot 2 power ingress       | 6°±5        | GW/um  |
| Spot 2 power egress        | 8°±3        | GW/um  |

**Note.** The latitude and longitude of the spots are derived by computing the centroid of points around the spots that are in the 90th percentile of intensity. The power is defined as the total emission from within a 15° circle centered at the inferred location of the spot.

Kind of detailed mapping of the Loki Patera magma lake with a precision of a few kilometers done by de Kleer et al. (2017) would not be possible with a spherical harmonic model because it would require maps of extremely large degrees ($l \gtrsim 700$).

### 7.2. Relevance to Mapping of Exoplanets

Hot super-Earths with (likely) molten surfaces, such as 55 Cancri e, are prime targets for future observations with JWST (Samuel et al. 2014; Henning et al. 2018). The model we have presented in this paper can easily be applied to observations of exoplanets. In fact, one of the motivations for this work was to test methods developed largely for the purpose of mapping exoplanets in a context where the ground truth is more easily accessible because Io is in the solar system. To apply our model to exoplanet observations, we simply need to compute the design matrix $A$ given a specification of the planet’s orbit, which is very straightforward in starry. Since the observations of the secondary eclipses (occultations) of exoplanets will never be as high quality as observations of Io, the limited resolution of the spherical harmonic maps is more than sufficient for modeling exoplanet observations. The question of what kind of features on the surfaces of volcanic exoplanets it will be possible to constrain with secondary eclipse observations using JWST is one that can be answered using the framework presented in this paper.

Besides being useful for modeling secondary eclipses of volcanic super-Earths, our model can be used for modeling surfaces of gaseous exoplanets with sparse features. These will be easier to observe with JWST than super-Earths because of their larger size and higher surface temperatures. Even if the sparsity assumption is not valid for these types of planets, the hybrid pixel/spherical harmonic model can be easily used with a different set of priors more appropriate for these observations. For example, a sensible prior for the surface of a fast-rotating gaseous exoplanet might be requiring some degree of azimuthal symmetry for the inferred map.

### 7.3. Future Work

This paper is a first step in a series of papers dedicated to probabilistic modeling of Io’s surface. A major issue with the model presented in this work is that we cannot naturally account for the time variability of surface emission. In Paper II in this series, we aim to fit an ensemble of light curves (occultations by Jupiter, mutual occultations, and phase curves) with a generalized model that will enable us to infer a time-dependent map of the entire surface and quantify the time variability over timescales of decades. A spatiotemporal map of volcanic emission spanning such a long timescale will help us better understand the global evolution of volcanism on Io. In addition to modeling the global
properties of volcanism on Io, it would also be interesting to
constrain the time evolution of peak emission in Loki Patera with
high accuracy in order to better understand the resurfacing
process.

Finally, it should be straightforward to extend our model for
use in fitting resolved observations of Io (either occulted or
not), such as the adaptive optics observations done by de Kleer & de Pater (2016b, 2016a). Since we would not need to
compute the integrated flux over complicated boundaries in that
case, we could fit maps of much higher order.

We would like to thank Will Farr, Katherine de Kleer, David
W. Hogg, and the Astronomical Data Group at the Center for
Computational Astrophysics for their help and thought-provoking
discussions. F.B. acknowledges support and funding from the
2020 Flatiron Institute Center for Computational Astrophysics
predoctoral program, which made this paper possible.

Appendix

Horseshoe Priors

The regularized horseshoe prior (Piironen & Vehtari 2017) is
specifically designed for use in Bayesian sparse linear regression.
It is a generalization of the horseshoe prior introduced in Carvalho
et al. (2010). The idea behind the horseshoe prior is to set the scale
for each regression coefficient (pixel) to a product of a global scale
τ and a local scale λ, where i indexes all of the pixels. The
horseshoe prior is defined hierarchically as

\[ \tau \sim \text{half} - \mathcal{C}(0, \tau_0) \]
\[ \lambda_i \sim \text{half} - \mathcal{C}(0, 1), \]
\[ p_i \sim \mathcal{N}(0, \tau \lambda_i) \]

where \( p_i \) are the pixels and \( \text{half} - \mathcal{C} \) is the half-Cauchy distribution. Each local scale parameter is drawn from a
unit-scale heavy-tailed half-Cauchy distribution that allows for very
large values of the pixels. The global scale parameter is also a free
parameter, drawn from a half-Cauchy distribution with a scale
equal to \( \tau_0 \). The horseshoe prior is closely related to the spike-and-slab
prior, which is a mixture of a delta function prior at zero
(spike) and some prior elsewhere (slab).

The regularized horseshoe prior adds another level to
Equation (A1) in order to allow fine-tuned control of sparsity and
regularize very large values of coefficients in cases where the
data are only weakly constraining. Piironen & Vehtari (2017)
showed that the regularized horseshoe prior can be
considered as a continuous counterpart of the spike-and-slab
prior with a finite slab width \( c \), whereas the horseshoe prior resembles a spike-and-slab prior with a slab of infinite width.
The prior is defined by

\[ \tau \sim \text{half} - \mathcal{C}(0, \tau_0) \]
\[ c^2 \sim \text{inv} - \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2} s^2\right) \]
\[ \lambda_i \sim \text{half} - \mathcal{C}(0, 1) \]
\[ \tau = \frac{c \lambda_i}{\sqrt{c^2 + \tau^2 \lambda_i^2}} \]
\[ p_i \sim \mathcal{N}(\tau \lambda_i) \]

where \( \text{inv} - \mathcal{G} \) is the inverse gamma distribution and
\( \mathcal{N}(\tau \lambda_i) \) is the half-normal distribution. Integrating out the
slab scale \( c \) implies a marginal Student \( t(\nu, 0, s) \) prior for pixels
far from zero. When the pixels \( p_i \) are close to zero \( (\tau^2 \lambda_i^2 < c^2) \),
we have \( \lambda_i^2 \rightarrow \lambda_i^2 \), and the prior approaches the original
horseshoe. When the pixels are far from zero \( (\tau^2 \lambda_i^2 \gg c^2) \),
then \( \lambda_i^2 \rightarrow c^2/\tau^2 \), and the prior approaches \( \mathcal{N}(0, c) \).

Piironen & Vehtari (2017) suggested the following expression
to set the scale parameter \( \tau_0 \), which is an estimate of the global scale of the pixels

\[ \tau_0 = \frac{p_0}{D - p_0 \sqrt{n}}, \]

where \( p_0 \) is our prior guess for the number of significant pixels
that are sufficiently far above zero, \( D \) is the total number of
pixels, \( n \) is the number of data points, and \( \sigma \) is the standard
deviation of the data points (error bars). Thus, we only need to
specify \( p_0 \), the degree-of-freedom parameter \( \nu \), and the slab
width \( c \).

When using the regularized horseshoe prior in a small data
regime, it is often necessary to use the noncentered parameterization
to avoid funnels in the posterior that are often present in
hierarchical models.9 The purpose of this reparameterization is
to replace the priors in Equation (A2) with zero mean and unit
variance priors and rescale them with deterministic transforms as follows:

\[ \tau \sim \text{half} - \mathcal{C}(0, 1) \]
\[ c^2 \sim \text{inv} - \mathcal{G}\left(\frac{\nu}{2}, 1\right) \]
\[ \lambda_i \sim \text{half} - \mathcal{C}(0, 1) \]
\[ p_i \sim \mathcal{N}(0, \tau \lambda_i) \]
\[ \tau = \frac{c \lambda_i}{\sqrt{c^2 + \tau^2 \lambda_i^2}} \]
\[ p_i = \tau \lambda_i \frac{p_i}{\lambda_i} \]  

We find that without using the noncentered parameterization,
the sampling is problematic, and there are many divergences in the
gradients of the parameters; with the noncentered parameterization,
there are no problems with sampling.

ORCID iDs

Fran Bartolić @ https://orcid.org/0000-0001-8630-9794
Rodrigo Luger @ https://orcid.org/0000-0002-0296-3826
Daniel Foreman-Mackey @ https://orcid.org/0000-0002-9328-5652
Robert R. Howell @ https://orcid.org/0000-0003-4859-2060
Julie A. Rathbun @ https://orcid.org/0000-0001-7619-652X

References

Aizawa, M., Kawahara, H., & Fan, S. 2020, ApJ, 896, 22
Arlot, J. E., Cahichel, H., Link, F., et al. 1974, A&A, 35, 115
Barnes, R., Raymond, S., Greenberg, R., Jackson, B., & Kaib, N. A. 2010,
ApJL, 709, L95

9 https://mc-stan.org/docs/2_26/stan-users-guide/reparameterization-section.html
