Bayesian analysis of Enceladus' plume data to assess methanogenesis

SUPPLEMENTARY INFORMATION

Antonin Affholder[1,2], François Guyot[3,4], Boris Sauterey[1], Régis Ferrière[1,5,6*], and Stéphane Mazevet[2*]

[1]Institut de Biologie de l'École Normale Supérieure, ENS, Université Paris Sciences Lettres, Paris, France

[2]Institut de Mécanique Céleste et de Calcul des Éphémérides, Observatoire de Paris, , Université Paris Sciences & Lettres, Paris, France

[3]Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, Muséum National d'Histoire Naturelle, CNRS, Sorbonne Universités , Paris, France

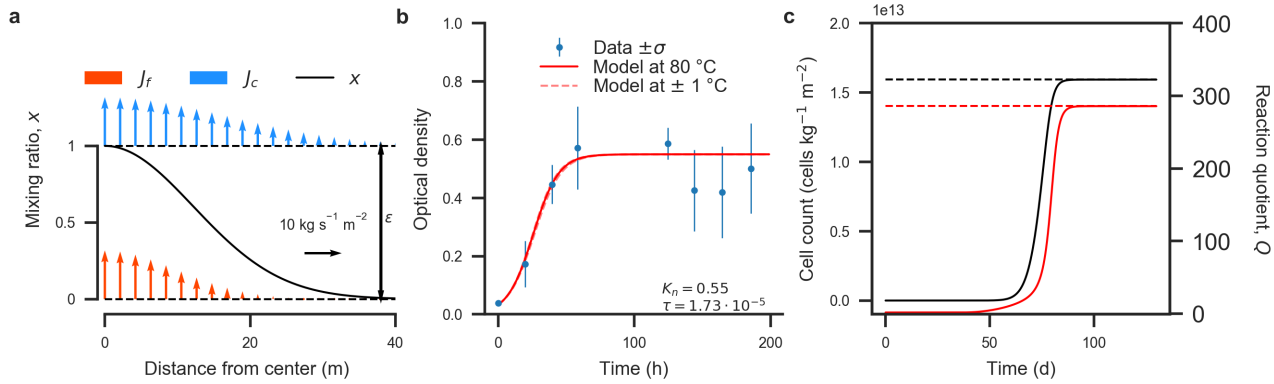
[4]Institut Universitaire de France (IUF)

[5]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, USA

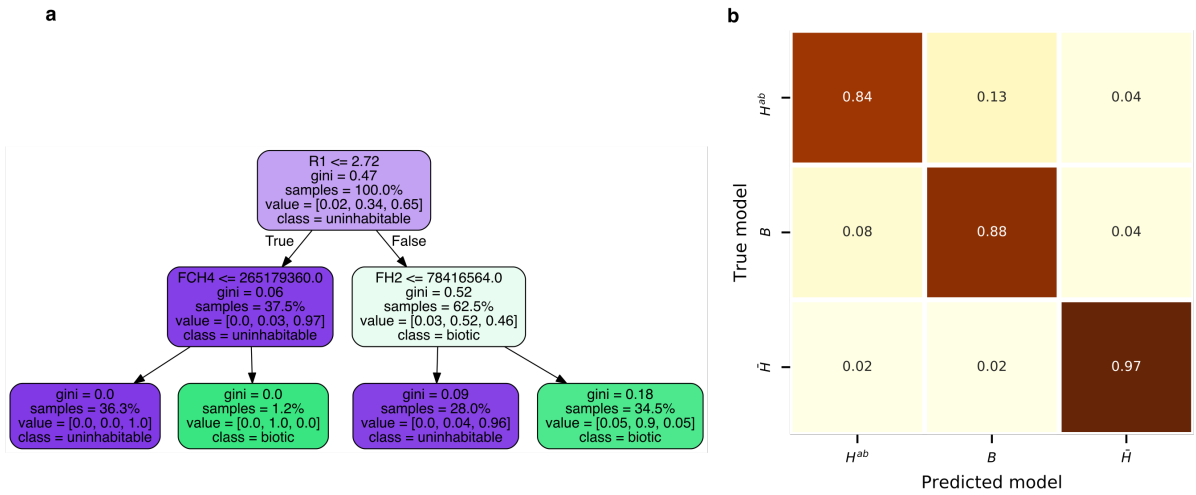
[6]International Research Laboratory for Interdisciplinary Global Environmental Studies (iGLOBES), CNRS, ENS, Université Paris Sciences & Lettres, University of Arizona, Tucson, USA

[*]These authors jointly supervised this work

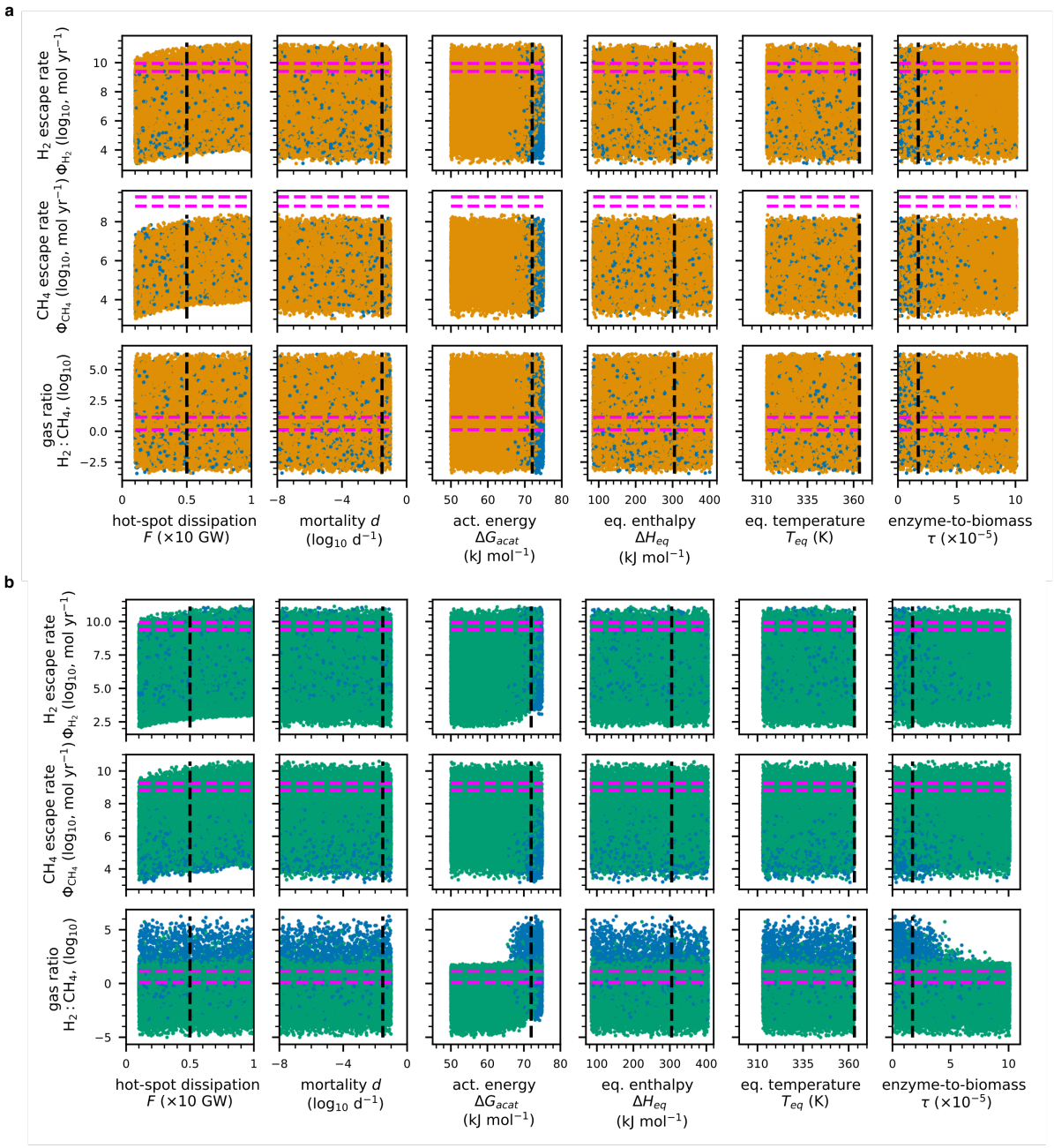
Supplementary Figures



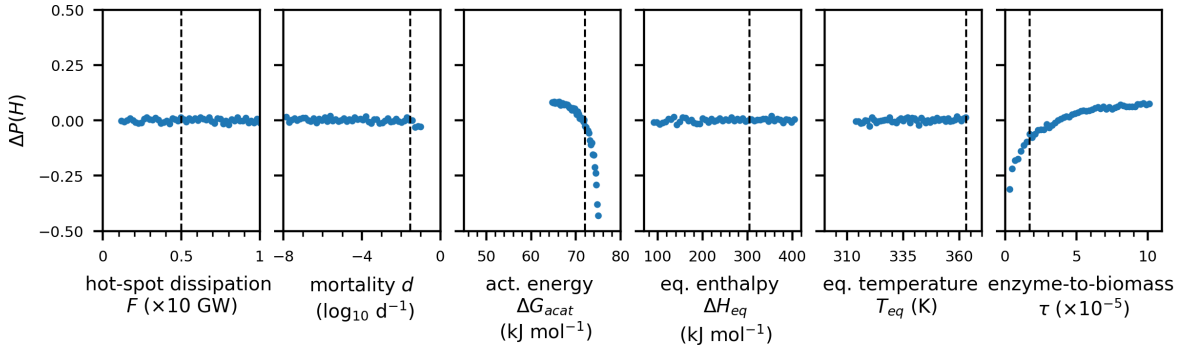
Supplementary Figure 1 | Mixing layer model and population dynamics outputs. a Structure of the mixing layer. x is the mixing ratio of hydrothermal fluid over ocean waters. ε is the thickness of the mixing layer (m). Vertical arrows indicate the advection velocities of water masses, with scale indicated by the black horizontal arrow. **b** Confronting the bio-ecological model to experimental data. Red: logistic growth curve using the experimental population density at $t = 0$ as initial state $n_0 = 0.03$, an optical density $K_n = 0.55$ as carrying capacity and the growth rate predicted by the model using $\tau = 1.73 \cdot 10^{-5}$ in experimental conditions ($T = 353\ K$, $[H_2]_0 = 1.5$ mmol kg^{-1}, $[DIC]_0 = 7.7$ mmol kg^{-1} $[CH_4]_0 = 10^{-10}$ mol kg^{-1}). Red dotted lines are logistic curves using growth rates obtained with ± 1 temperature difference, corresponding to the temperature range maintained during the experiment in Taubner et al., 2018 [1]. Blue: data from Taubner et al., 2018 [1], error bars are standard deviation of triplicates. **c** Numerical solving of model trajectory in 1 m^3 of water for population size (N, cells kg^{-1}, black) and local reaction quotient (Q, red), and predicted steady-state (dashed lines). Initial concentrations used in this example : $[H_2] = 22$ mmol kg^{-1}, $[DIC] = 18$ mmol kg^{-1}, $[CH_4] = 0.03\ \mu$mol kg^{-1}. Temperature is 329 K. These conditions correspond to the composition found 5 m from the center of a hot spot generated with random parameters sampled in our prior distributions (described in Extended Data Table 1) : $[H_2]_o = 8.0 \times 10^{-8}$ mol kg^{-1}, $[H_2]_f = 5.0 \times 10^{-6}$ mol kg^{-1}, $[CO_2]_o = 5.7 \times 10^{-7}$ mol kg^{-1}, $[CO_2]_f = 6.6 \times 10^{-8}$ mol kg^{-1}, $[CH_4]_o = 1.0 \times 10^{-4}$ mol kg^{-1}, $[CH_4]_f = 1.1 \times 10^{-5}$ mol kg^{-1} and $T_f = 421.5$ K. See Methods for model equations. Model parameters used for this example are default values given in Extended Data Table 2.



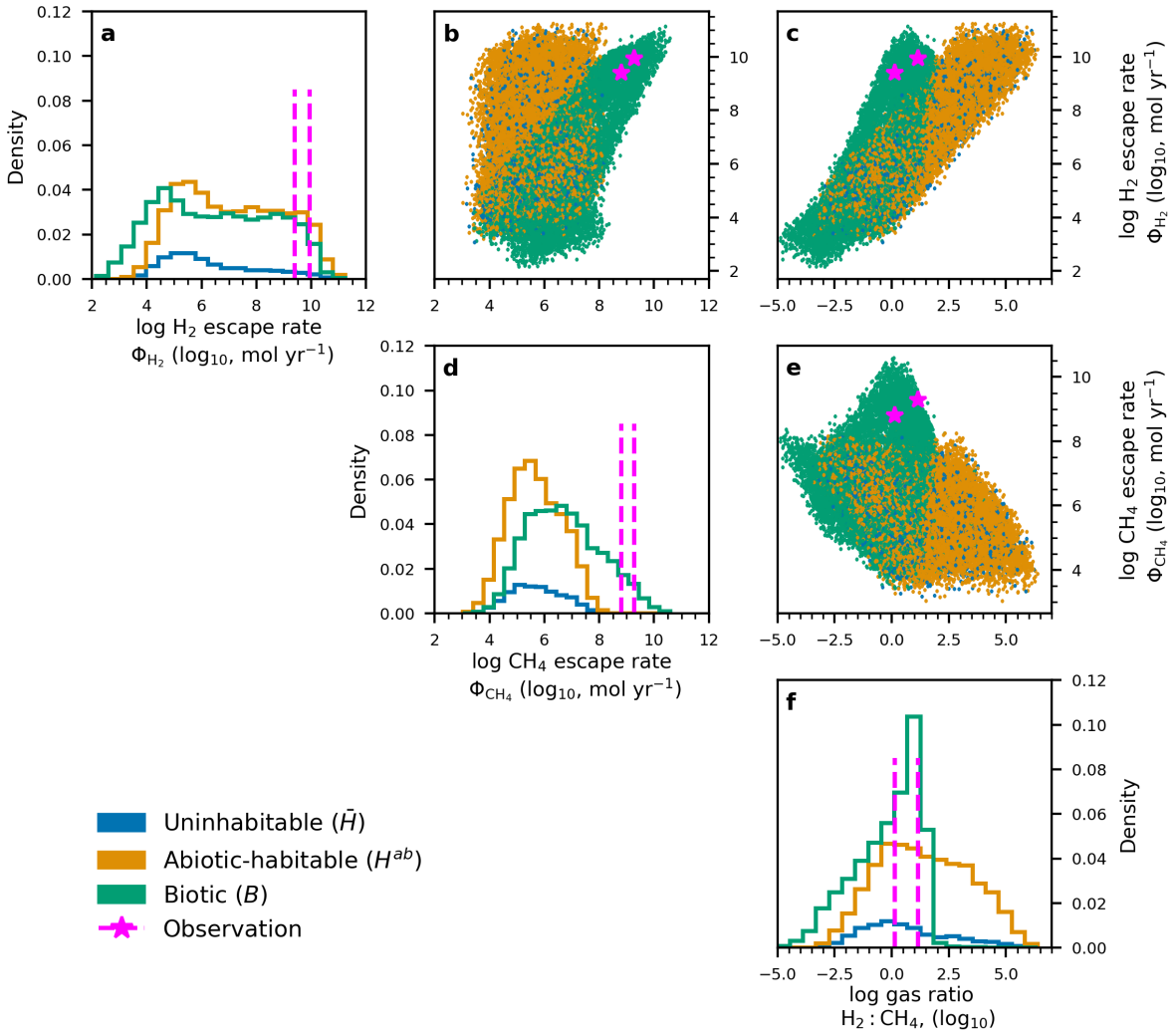
Supplementary Figure 2 | Approximated Bayesian Computation - Random Forest (ABC-RF) classifier. **a** Example of a decision tree in a decision forest fitted to pseudo-data generated by our model. At each node, the left arrow is followed if the test of the node is true, the right arrow is followed if the test is false. Here, the allowed maximum depth of the tree is 2, meaning that the tree has two levels of nodes, making it easy to read. "Gini" stands for the Gini impurity, which was minimized to fit the tree to the data (Methods). "Samples" is the proportion of samples filtered at each node in the training dataset and "value" contains the probability associated with every class. In the tree, $R_1 = \mathrm{H_2} : \mathrm{CH_4}$. **b** Confusion matrix of the ABC-RF classifier for our serpentinization-only dataset in which $P(B|H) = 0.5$. The confusion matrix evaluates the performance of the classifier on a set of pseudo-data on which the random forest was not fitted. The values and colors in the squares (darker means higher value) denote the proportion of simulations in the "True model" class that were classified into "Predicted model". See Methods for model equations. Model parameters used are given in Extended Data Tables. 2 and 3.



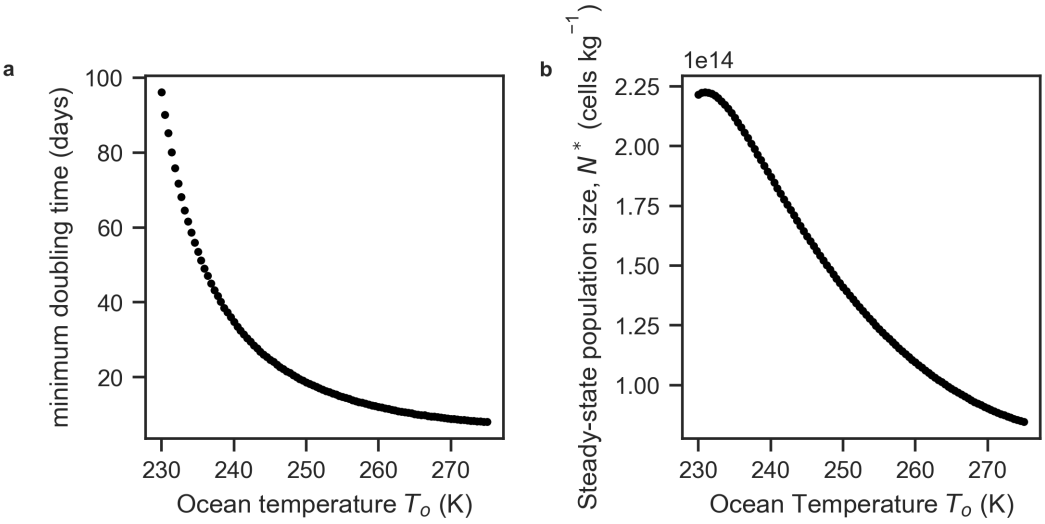
Supplementary Figure 3 | Model outputs as a function of model parameters. a Outputs of the purely abiotic model. Blue dots indicate 'uninhabitable' simulations ($\bar{H}$), orange dots indicate 'abiotic-habitable' simulations (H^{ab}).**b** Outputs of the model including biological activity. Green dots indicate habitable simulations in which biological methanogenesis occurs (B). F is the power dissipated by the hotspot ($\times 10$ GW), d is the baseline death rate of the population (in d^{-1}, $\log_{10}$ scale), ΔG_{acat}, ΔH_{eq} and T_{eq} are parameters governing the shape of the catabolic rate as a function of temperature (in kJ, kJ and K respectively), τ is the active biomass scaling factor, in $\times 10^{-5}$ mol_{enzyme} $mol_{C_x}^{-1}$ (see Methods for further detail). Black dashed lines (vertical) indicate default parameter values. Magenta dashed lines (horizontal) indicate Cassini observations[2]. Internal parameters (HF and OW compositions, HF temperature) are sampled from log-uniform distributions in ranges given in Extended Data Table. 1. See Methods for model equations. Parameters F, d, ΔG_{acat}, ΔH_{eq}, T_{eq} and τ vary as described in Supplementary Methods.



Supplementary Figure 4 | Change in local density of the prior probability of habitability as a function of model parameters. F is the power dissipated by the hotspot ($\times$10 GW), d is the baseline death rate of the population (in d^{-1}, $\log_{10}$ scale), ΔG_{acat}, ΔH_{eq} and T_{eq} are parameters governing the shape of the catabolic rate as a function of temperature (in kJ, kJ and K respectively), τ is the active biomass scaling factor, in $\times 10^{-5}$ mol_{enzyme} $mol_{C_x}^{-1}$ (see Methods for further detail). For each point, the x coordinate is the center of a range of a model parameter's value in which the proportion of habitable simulations is computed. The y coordinate is the difference between the proportion of habitable simulations in the whole set and the local proportion. Black dashed lines (vertical) indicate default parameter values. Internal parameters (HF and OW compositions, HF temperature) are sampled from log-uniform distributions in ranges given in Extended Data Table 1. Parameters F, d, ΔG_{acat}, ΔH_{eq}, T_{eq} and τ vary as described in Supplementary Methods.



Supplementary Figure 5 | Cassini observations and distributions of model outputs over the space of observables when model parameters vary. a, d, f Density distributions of pseudo-data generated by simulation of observables in which model parameters are drawn at random. Blue indicates the subset of simulations in which no biological methanogenesis is possible ('uninhabitable', $\bar{H}$). Orange indicates the subset of simulations in which biological methanogenesis is possible but not taken into account ('abiotic-habitable', H). Green indicates the subset of habitable simulations in which biological methanogenesis is included to generate the pseudo-data ('biotic', B). Magenta dashed lines indicate Cassini observations[2]. Note the $\log_{10}$ scale on the horizontal axis. **b, c, e** Joint distributions of observable pseudo-data generated by the model. Magenta stars indicate Cassini observations. Note the $\log_{10}$ scale on both axes. See Fig. 2 for more details, Methods for model equations and Extended Data Tables 1 and 2 for prior internal parameter ranges and model parameter values. Parameters F, d, ΔG_{acat}, ΔH_{eq}, T_{eq} and τ vary as described in Supplementary Methods.



Supplementary Figure 6 | Effect of lower ocean freezing temperature on a population of methanogens. **a** Minimum doubling time as a function of ocean temperature. **b** Steady-state population size as a function of temperature. All data points were obtained by numerically solving the differential equation system for the concentrations and population size (equation (19) in Methods) localized 5 m from the center of a hydrothermal vent with $T_f = 334\ K$, $[H_2]_f = 0.02$ mol kg^{-1}, $[DIC]_f = 0.02$ mol kg^{-1}, $[CH_4]_f = 2.9\ 10^{-8}$ mol kg^{-1}, $[H_2]_o = 3.35\ 10^{-8}$ mol kg^{-1}, $[DIC]_o = 2.8\ 10^{-6}$ mol kg^{-1} and $[CH_4]_o = 6.2\ 10^{-8}$ mol kg^{-1}. We varied T_o from 230 to 273 K to explore the consequences of a lower freezing point of water in Enceladus' ocean. Parameters for the methanogenic physiology are the standard values in Extended Data Table 2.

Supplementary Methods

Sensitivity analysis

We conducted a sensitivity analysis of our model by varying the parameters T_{eq}, ΔG_a, ΔH_{eq}, τ, d and F (see Extended Data Table 2 for default values). We generated a set of 50,000 simulations using the prior distributions of internal parameters and drawing model parameters at random in the following distributions. T_{eq}, ΔG_a and ΔH_{eq}, the parameters governing the catabolic rate function, are drawn from uniform laws over ranges given in Daniel et al. (2010)[3]:

$$\begin{aligned} T_{eq} &\sim \mathbf{U}(313,364) \quad \text{K} \\ \Delta G_a &\sim \mathbf{U}(50,75) \quad \text{kJ} \\ \Delta H_{eq} &\sim \mathbf{U}(85,420) \quad \text{kJ} \end{aligned} \tag{1}$$

The power dissipated by the hotspot is taken in the range proposed in Choblet et al. (2017)[4] :

$$F \sim \mathbf{U}(1,10) \text{ GW} \tag{2}$$

The active biomass scaling factor is taken in a range allowing it to vary of one order of magnitude:

$$\tau \sim \mathbf{U}(0.1,10) \cdot 10^{-}5 \text{ mol}_{enzyme} \cdot \text{mol}_{C_x}^{-1} \tag{3}$$

It has been argued[5] that the basal death rate d can be virtually 0, hence a range of very small values:

$$d \sim \log_e \mathbf{U}(10^{-8}, 10^{-1}) \text{ d}^{-1} \tag{4}$$

Modeling the consequences of a lower freezing temperature of water

We address separately the possibility that ammonia in Enceladus' ocean might lead to a lower freezing point of water and thus a lower ocean temperature. We vary T_o linearly between 230 and 275 K and model a hot-spot with otherwise fixed parameters $T_f = 334\ K$, $[\text{H}_2]_f = 0.02$ mol kg^{-1}, $[\text{DIC}]_f = 0.02$ mol kg^{-1}, $[\text{CH}_4]_f = 2.9\ 10^{-8}$ mol kg^{-1}, $[\text{H}_2]_o = 3.35\ 10^{-8}$ mol kg^{-1}, $[\text{DIC}]_o = 2.8\ 10^{-6}$ mol kg^{-1} and $[\text{CH}_4]_o = 6.2\ 10^{-8}$ mol kg^{-1}. We numerically solve the population dynamics in a localized place around the hot-spot (5 m) and measure the minimum doubling time as

$$\min(\text{doubling time}) = \frac{N(\text{argmax}\left\{\frac{dN}{dt}\right\})}{\max\left\{\frac{dN}{dt}\right\}} \times \log_e 2 \tag{5}$$

We identify the steady-state population size as the population at the end of the integration interval of 1 million days. Results are presented in Supplementary Fig. 6 and discussed in Supplementary Discussion below.

Supplementary Results and Discussion

Sensitivity to model parameters

Perturbing the model by using randomly drawn parameter values (with distributions described in Supplementary Section) shows little influence of parameter values on simulated observables (Supplementary Fig. 3). We find that parameters ΔG_{acat} (governing the shape of the catabolic rate as a function of temperature) and τ (scaling cellular growth with enzyme function) significantly influence the distribution of observables in simulations run with parameters that allow methanogens to exist (Supplementary Fig. 3).

Overall, variation in model parameters, and physiological parameters in particular, results in a higher fraction of habitable simulated environments ($P(H) = 0.91$, instead of $P(H) = 0.27$ with default parameter values). The ΔG_{acat} parameter is found to have great influence on $P(H)$, and higher values for this parameter, including the default value, lead to much lower $P(H)$ than the average (Supplementary Fig. 4).

ΔG_{acat} accounts for the activation energy of catabolic enzyme and, all else being equal, increasing ΔG_{acat} limits the catabolic rate (equation (15) in main text). As a consequence, lower values of ΔG_{acat} allow methanogens to function in harsher environments due to increased q_{cat} and unchanged q_m (see Methods in main text).

Although the perturbation of model parameters decreases the overall classification power, it does not alter our inference qualitatively, that methanogenesis at Enceladus' seafloor is plausible (Supplementary Fig. 5g). Methanogenesis under varying parameters still explains well the observed high value of methane flux, but the median value of biotic methane flux is shifted to lower values in comparison with pseudo-data produced with fixed parameters. Altogether, the higher *a priori* probability of habitability and the lower median biotic methane production might be explained by a lower value of ΔG_{acat}, allowing harsher environments to be habitable, while colder environments, being poor in hydrothermal fluid, are also poor in nutrients, thus limiting methane production. The Φ_{CH_4} distribution might then be skewed by pseudo-data points corresponding to low availability of H_2, rendered inhabitable by the low value of ΔG_{acat} and thus producing less methane.

We stress that in order to incorporate varying parameters into the Bayesian inference, it is necessary to constrain the distribution from which parameters are drawn, so that they bring relevant information to the inference. Variation in biological parameters that measure cell 'traits' may be driven by individual plasticity or population adaptation. Our knowledge of the mechanistic relationships that may constrain plastic or adaptive variation, such as physiological and genetic trade-offs between traits, is, however,

limited. Therefore we chose to interpret parameter variation as random perturbations of the model, rather than attempt at a process-based biological exploration of the parameter space.

Lower freezing temperature of water

If the freezing temperature point is lower, *e.g* due to high ammonia content in the ocean, the ocean temperature might be lower than 275 K. In the preliminary analysis described in Supplementary Methods, we find that lower ocean temperatures lead to potentially much higher doubling times in the modeled population of methanogens (Supplementary Fig. 6a). When we generate the sets of simulations used in our statistical approach, we only use the steady state of the population dynamics, regardless of the time required to reach it (see Methods). Thus, our results are unaffected by the fact that populations of methanogens might grow exceedingly slowly in an ocean colder that 275 K. In addition, our simulations show that, in spite of extremely slow growth in a cold ocean, populations eventually reach a much larger size at steady-state (Supplementary Fig. 6b). This may result from lower ocean temperatures increasing the volume of interfaces between hydrothermal fluid and ocean water that are cooler than the upper thermal limit of life, thus enhancing the growth of the whole population. This outcome (*i.e.* larger steady-state population size in a colder ocean) highlights the fact that ecosystems that are predicted viable at 275 K may not be imperiled by colder temperature. Furthermore, larger populations at steady state might translate in a stronger environmental impact of biological methanogenesis, which might even increase the likelihood of the observations under the inhabited model.

It is worth noting that the preliminary analysis presented here uses physiological parameters of a hyperthermophilic methanogen (*e.g.* $T_{eq} = 90\ ^{\circ}C$). In contrast, psychrophilic methanogens may have physiological adaptations that allow them to grow faster than the predicted growth of thermophilic methanogens exposed to a cold environment. For all these reasons, we expect a lower freezing point of ocean water to have limited effects on our results. This, however, warrants further investigation.

Sustainability of favorable conditions

Back-tracking rates of conversion of H_2 and DIC to methane by the modeled methanogens yields mean values of the order of 10^8 mol of H_2 and DIC consumed per year by methanogens. Waite et al. (2017)[2] estimates that the stock of minerals in Enceladus' core could support the observed outgassing rate $\approx 10^9$ mol yr^{-1} for billions of years. Estimation of DIC concentration in Glein et al. (2015)[6] (10^{-3} to 10^{-1} mol kg^{-1}) yields a total stock of DIC in Enceladus' ocean (which mass is computed from ocean dimensions in Choblet et al. (2017)[4]) of the order of 10^{18} mol, which amounts to 50 to 500 Myr of modern outgassing rates and consumption by methanogens. Given that these estimates correspond to the modern stock of DIC, stocks in the past were probably larger. A potential methanogenic biosphere might therefore be limited in the future if no carbon recycling process is present.

References

[1] Taubner, R.-S. et al. Biological methane production under putative Enceladus-like conditions. *Nature Communications* **9**, 748 (2018).

[2] Waite, J. H. et al. Cassini finds molecular hydrogen in the Enceladus plume: evidence for hydrothermal processes. *Science* **356**, 155–159 (2017).

[3] Daniel, R. M. et al. The molecular basis of the effect of temperature on enzyme activity. *Biochemical journal* **425**, 353–360 (2010).

[4] Choblet, G. et al. Powering prolonged hydrothermal activity inside Enceladus. *Nature Astronomy* **1**, 841 (2017).

[5] Connolly, J. P. & Coffin, R. B. Model of carbon cycling in planktonic food webs. *Journal of Environmental Engineering* **121**, 682–690 (1995).

[6] Glein, C. R., Baross, J. A. & Waite Jr, J. H. The pH of Enceladus' ocean. *Geochimica et Cosmochimica Acta* **162**, 202–219 (2015).