BIOSIGNATURES IN THE ATMOSPHERES OF OCEAN WORLDS: METHODOLOGICAL CHALLENGES IN DISTINGUISHING BIOTIC FROM ABIOTIC SOURCES OF METHANE AND PHOSPHINE

Aleksei Vladimirovich Sokolov

P.G. Demidov Yaroslavl State University
Yaroslavl, Russian Federation
E-mail: aleksei.vs9@uniyar.ac.ru

Original Scientific Article

Received: 01.05.2026. Accepted: 17.06.2026.

DOI: https://doi.org/10.65932/CAR-2026-1-5 **UDC:** 573.5:523.4

Abstract: The detection of a putative biosignature gas is only as strong as the abiotic explanations it can exclude, and recent history — above all the phosphine controversy at Venus — shows how quickly a claimed sign of life can dissolve into instrument artifact or unremarkable geochemistry. This article addresses the methodological problem directly for the two reduced gases most often proposed as biosignatures on ocean worlds and habitable-zone exoplanets: methane and phosphine. I argue that prevailing frameworks err by asking a gas-first question — is methane, or phosphine, a biosignature? — and by combining lines of evidence through weighted averages that let a strong showing on one axis mask an unaddressed weakness on another. In their place I develop a provenance-first Biotic Provenance framework that scores any candidate detection on three logically independent abiotic-rejection axes: thermodynamic-disequilibrium context, the abiotic-flux ceiling, and spectral and conceptual robustness. Crucially, the framework combines these axes under a weakest-link rule — the Biotic Provenance Index is the minimum of the three axis scores, not their average — on the argument that a false positive is governed by the least-excluded abiotic alternative rather than by the average strength of the case. Applied to three touchstone cases, the framework localizes precisely where each fails: Venus phosphine is limited by spectral robustness, with the detection itself contested; K2-18b methane and dimethyl sulfide are limited by the same axis, resting on marginal spectral features; and Enceladus methane is limited instead by the abiotic-flux ceiling, because serpentinization can supply the observed abundance. None of the three currently clears the threshold for biotic provenance, but each fails for a different, identifiable reason — which is exactly the diagnostic the field needs, and which a weighted-average confidence ladder obscures.

Keywords: *biosignatures, ocean worlds, methane, phosphine, false positives, Enceladus, Europa, exoplanets.*

INTRODUCTION

Few episodes in recent astrobiology have been as instructive as the phosphine affair. In 2020 a team reported phosphine in the clouds of Venus and argued that no known abiotic process could produce it in the observed quantity, so that life became a candidate explanation. Within months the claim was under siege from two directions at once: independent analyses questioned whether the spectral feature was real or an artifact of data processing and confusion with sulfur dioxide (Akins, Lincowski, Meadows & Steffes, 2021; Trompet, Robert, Mahieux, Schmidt, Erwin & Vandaele, 2020), while atmospheric chemists asked whether the abiotic inventory of Venus

© 2026 The Author(s). Published by SAPCRAA. *Open access under CC BY 4.0 — provided the original work is properly cited.*

Cosmological and Astrobiological Review: Journal for the Study of the Universe, Life and the Natural Sciences, Year 2026, Volume 4, Issue 1, Serial number 5, p. 77 – 89.

was understood well enough to exclude non-biological sources at all (Wunderlich, Grenfell & Rauer, 2023). The episode did not settle whether Venus hosts life. It settled something more useful: that a biosignature claim is only as strong as the abiotic alternatives it can rule out, and that ruling them out is harder, and more multi-dimensional, than the gas-by-gas intuition suggests. A recent review in this journal traces the full arc of that debate and its methodological lessons (Davis, 2025).

The problem is about to become acute rather than academic. The Europa Clipper mission is en route to a moon whose subsurface ocean may vent material into space, and Enceladus already presents, in Cassini data, a plume carrying methane, molecular hydrogen and carbon dioxide in proportions that a hungry methanogen would find congenial (Howell & Pappalardo, 2020; Fifer, Catling & Toner, 2022). At the same time, transit spectroscopy of habitable-zone exoplanets has begun returning tentative detections of methane and, more controversially, dimethyl sulfide on worlds such as K2-18b (Madhusudhan, Nixon, Welbanks, Piette & Booth, 2020). In every one of these settings the central question is identical and unforgiving: given a reduced gas that life could make, can we exclude the possibility that geology, photochemistry or the instrument made it instead? Methane and phosphine are the sharpest test cases because both are genuinely producible by life and genuinely producible without it — methane by serpentinization and a dozen other abiotic routes (Klein, Grozeva & Seewald, 2019; Reeves & Fiebig, 2020), phosphine by processes still being mapped (Sousa-Silva, Seager, Ranjan, Petkowski, Zhan, Hu & Bains, 2020).

The central research question of this article is methodological rather than observational: how should a candidate detection of methane or phosphine on an ocean world or exoplanet be evaluated so that the risk of a false positive is minimized, and so that, when a claim fails, we learn exactly why it failed? From this I derive three hypotheses. The first is that the biotic or abiotic provenance of a reduced gas is not a property of the gas but of the full detection in its planetary and instrumental context, so that any gas-first framework is mis-posed. The second is that the several lines of evidence bearing on provenance are logically independent, in the sense that strength on one cannot compensate for weakness on another. The third follows: that these lines of evidence must be combined under a weakest-link rule rather than an average, because a single unexcluded abiotic channel is sufficient to produce a false positive regardless of how strong the rest of the case may be.

The original contribution of this article is a provenance-first Biotic Provenance framework that operationalizes these hypotheses. It scores a candidate detection on three logically independent abiotic-rejection axes — thermodynamic-disequilibrium context, the abiotic-flux ceiling, and spectral and conceptual robustness — and combines them into a single Biotic Provenance Index defined as the minimum, not the mean, of the axis scores. Unlike existing confidence ladders and weighted assessments, which aggregate evidence additively and thereby permit a strong axis to mask a weak one, the minimum rule enforces the logic of false positives: a claim is only as credible as its least-excluded abiotic alternative. I want to be candid that the axis scores I assign are heuristic estimates drawn from the literature rather than outputs of a unified model; their purpose is to make the structure of the argument explicit and its verdicts disputable, not to certify a probability of life.

The value of the framework is diagnostic as much as evaluative. Applied to Venus phosphine, Enceladus methane, and K2-18b, it does not merely rank the three by strength; it identifies the specific axis on which each currently fails, and those axes turn out to differ. That is the payoff of separating the lines of evidence and combining them by minimum rather than average: the field learns not only that a case is weak but where to push to make it strong. A brief clarification of scope: I am concerned here with reduced-gas biosignatures interpreted from remote spectroscopy or plume sampling, not with in situ metabolic or morphological detection, and I make no

claim that the three axes exhaust the problem — only that they are necessary, independent, and jointly more false-positive-robust than the averaged alternatives now in use.

The article proceeds through a combined literature review and methodology section that reconstructs the relevant research strands and defines the three axes and the minimum rule; a results section that presents the axis scores for the three cases without interpretation; three analytical sections treating in turn the disequilibrium axis, the abiotic-flux axis, and the robustness axis with the synthesis they imply; and a conclusion that answers the hypotheses, restates the contribution, and marks the framework's limits.

LITERATURE REVIEW AND METHODOLOGY

Literature Review

The first strand of relevant work concerns the logic of biosignature assessment itself, which has matured considerably since the naive era of single-gas detection. Thompson, Krissansen-Totton, Wogan, Telus and Fortney (2022) make the case for atmospheric methane as an exoplanet biosignature precisely by embedding it in context: methane is compelling not on its own but when it coexists with carbon dioxide and an absence of carbon monoxide on a planet where abiotic methane fluxes should be small. Their argument is a template for the provenance-first view, because it locates the biosignature in a pattern rather than a molecule. Wogan and Catling (2020) provide the thermodynamic backbone, asking when chemical disequilibrium in an atmosphere is a biosignature and when it is not, and showing that disequilibrium alone is ambiguous until the abiotic maintenance of that disequilibrium is ruled out. On the assessment side, Pohorille and Sokolowska (2020) survey how biosignatures should be evaluated for life detection, and Vickers, Cowie, Dick, Gillen, Jeancolas, Rothschild and McMahon (2023) raise the deepest worry — the problem of unconceived alternatives, the abiotic explanations we have not yet thought of — which is the epistemic hazard the robustness axis is meant to track.

The second strand documents just how productive abiotic chemistry is at making the very gases we hope are biotic, and it is sobering. Methane is generated abiotically by serpentinization, the reaction of water with ultramafic rock, at rates and in settings that overlap extensively with habitable environments (Klein, Grozeva & Seewald, 2019; Reeves & Fiebig, 2020). High-pressure serpentinization extends the abiotic methane factory to still deeper settings (Boutier, Vitale Brovarone, Martinez, Sissmann & Mana, 2021). The consequence for ocean worlds is direct: a moon with a rocky seafloor in contact with liquid water is an abiotic methane generator by default, so methane in its plume cannot be assumed biological. Isotopic tools offer a partial escape, since biological and abiotic methane can differ in their carbon and hydrogen isotope systematics and, more powerfully, in their clumped-isotope signatures; Labidi, McCollom, Giunta, Sherwood Lollar, Leavitt and Young (2024) show that clumped isotopes carry information about the formation temperature and mechanism of methane, which can in principle discriminate sources. The catch is that these measurements are demanding even in terrestrial laboratories and far harder to obtain remotely or from a fast plume flyby.

The third strand is specific to ocean worlds and their plumes, the settings where in situ sampling might one day break the ambiguity that spectroscopy cannot. Lorenz (2019) frames plume-based biosignature detection in explicitly Bayesian terms, insisting that any detection be weighed against priors and alternatives rather than read off as a positive. Fifer, Catling and Toner (2022) model the chemical fractionation that occurs as Enceladus material erupts, warning that the plume composition is not a transparent window onto the ocean; and Glein and Truong (2025) use phosphates to constrain the ocean's pH, tightening the geochemical context in which any

methane signal must be read. The feasibility literature is frank about instrumentation: New, Kazemi, Price, Cole, Spathis, Mathies and Butterworth (2020) and Mathies, New, Golozar and Butterworth (2021) examine whether a spacecraft can actually make an informative biosignature measurement while flying through a plume at kilometers per second, and the answer is a qualified and hard-won yes. Robare and Shock (2025) address the discriminating question head-on, asking how potential organic biosignatures on ocean worlds can be distinguished from abiotic background — the exact problem this article generalizes into a framework.

The fourth strand covers the exoplanet and Venus cases that supply the framework's hardest tests. For exoplanets, the assessment literature has grown careful about which gases are even worth chasing: Huang, Seager, Petkowski, Ranjan and Zhan (2022) assess ammonia, Zhan, Huang, Seager, Petkowski and Ranjan (2022) argue that organic carbonyls are poor biosignatures, and Glidden, Seager, Petkowski and Ono (2023) explore whether isotopologues could help — each paper narrowing the space of credible claims. Lustig-Yaeger, Meadows and Lincowski (2019) deliver a cautionary classic, showing that Venus-like clouds could masquerade as a habitable signal and generate a statistical false positive, which is the atmospheric analogue of the spectral confusion that later afflicted the Venus phosphine claim itself. Bixel and Apai (2021) supply the survey-level statistics for how many such tests we can expect to run and with what power. And the Venus phosphine literature — the detection attempts and upper limits (Trompet et al., 2020), the ALMA complications (Akins et al., 2021), the photochemical uncertainty (Wunderlich et al., 2023), and the broader question of an aerial biosphere (Patel, Mason, Nordheim & Dartnell, 2022) — is less a body of results than a live demonstration of every failure mode a biosignature claim can suffer. The phosphine biosignature concept itself, laid out by Sousa-Silva and colleagues (2020), remains sound in principle even as its first application faltered in practice, and that gap between principle and practice is precisely what a provenance framework must formalize.

What the literature offers, then, is a rich inventory of individual failure modes and individual discriminants, but no single structure that says how they combine. The dominant metaphor is the confidence ladder, which aggregates evidence toward higher rungs. My contention, developed below, is that aggregation by averaging is the wrong combination rule for a problem whose logic is disjunctive — where any one unexcluded alternative suffices to defeat the claim — and that a minimum rule is both more faithful to that logic and more useful diagnostically.

Research Methodology

The study is conceptual and comparative. It constructs an evaluative framework and applies it to three documented cases, drawing all parameters from peer-reviewed work published from 2019 onward. The framework rests on three abiotic-rejection axes, each scored from 0 to 1, where a score of 1 means the corresponding class of abiotic explanation has been essentially excluded and a score near 0 means it remains wide open.

The first axis, thermodynamic-disequilibrium context, scores whether the candidate gas participates in a chemical disequilibrium that abiotic processes cannot easily maintain, following the logic of Wogan and Catling (2020) and the contextual argument of Thompson and colleagues (2022). A high score requires not merely that the gas be present but that its coexistence with other species be hard to sustain without a biological flux. The second axis, the abiotic-flux ceiling, scores whether known abiotic sources — serpentinization, photochemistry, volcanism, radiolysis — are demonstrably unable to supply the observed abundance, drawing on the abiotic-methane literature (Klein et al., 2019; Boutier et al., 2021) and on isotopic discriminants (Labidi et al., 2024). A high score requires that the abiotic ceiling lie below the observed level. The third axis, spectral and conceptual robustness, scores whether the detection itself is secure against

instrumental artifact, spectral confusion, and unconceived alternatives (Akins et al., 2021; Lustig-Yaeger et al., 2019; Vickers et al., 2023). A high score requires that the feature be real and its interpretation not obviously hostage to an overlooked possibility.

The combination rule is the framework's distinctive commitment. Rather than averaging the three axes, I define the Biotic Provenance Index as their minimum: BPI = min(A1, A2, A3). The justification is that a false positive occurs whenever any single abiotic route is left open, so the credibility of a biotic interpretation is bounded above by the weakest-excluded alternative, not by the average across alternatives. Averaging would allow a spectrally shaky detection with a strong disequilibrium argument to score respectably, which is exactly the error the Venus phosphine episode exemplified; the minimum rule forbids it. I recognize that the minimum is a severe rule, and I discuss in the analysis where a softened version might be defended, but I adopt the strict minimum here because the cost of a false positive in this field is uniquely high.

The scoring rubric is explicit so that it can be contested. An axis receives a value near 1 when the literature provides direct evidence that the relevant abiotic class is excluded for the case at hand, near 0.5 when the exclusion is partial or contested, and near 0 when the abiotic route is unconstrained or the detection insecure. I assign the scores myself, from the cited literature, and I treat this single-author scoring as the method's principal weakness, to which I return in the limitations. A sensitivity check varies the scores within plausible bands and confirms that the identity of each case's limiting axis — the diagnostic output that matters — is stable. The numbers are ordered estimates meant to structure disagreement, not measurements of a probability of life.

RESEARCH RESULTS

Applying the Biotic Provenance framework to three touchstone cases yields the axis scores and indices collected in Table 1. The three cases are the reported phosphine at Venus, the methane in the Enceladus plume, and the methane and dimethyl sulfide reported for the exoplanet K2-18b.

For Venus phosphine, the disequilibrium axis scores low-to-moderate (A1 ≈ 0.40). Phosphine in an oxidized atmosphere is indeed a disequilibrium species, which is what motivated the original claim, but the atmospheric inventory of Venus is poorly enough characterized that the strength of the disequilibrium cannot be firmly established (Wunderlich et al., 2023). The abiotic-flux axis is low (A2 ≈ 0.35): although phosphine is difficult to produce abiotically in bulk, the abiotic phosphorus chemistry of Venus is not mapped well enough to place a confident ceiling below the claimed abundance. The robustness axis is very low (A3 ≈ 0.15), because the detection itself is contested — independent analyses attribute much or all of the signal to sulfur dioxide and data-reduction artifacts (Akins et al., 2021; Trompet et al., 2020). Under the minimum rule the index is BPI ≈ 0.15, and the limiting axis is robustness: the Venus case fails first and foremost because it is not clear the feature is real.

For Enceladus methane, the profile differs markedly. The disequilibrium axis scores moderate-to-high (A1 ≈ 0.60), because Cassini measured methane together with hydrogen and carbon dioxide in a combination that represents genuine disequilibrium and that terrestrial methanogens exploit (Fifer et al., 2022). The robustness axis is high (A3 ≈ 0.70): the plume species are securely detected, and future in situ sampling could raise this further. But the abiotic-flux axis is only moderate (A2 ≈ 0.45), and this is the binding constraint, because serpentinization on Enceladus's rocky seafloor can plausibly supply the observed methane without any biology (Klein et al., 2019; Reeves & Fiebig, 2020). Under the minimum rule the index is BPI ≈ 0.45, and the limiting axis is the abiotic-flux ceiling: the Enceladus case fails not because the signal is doubtful but because geology can account for it.

For K2-18b, the disequilibrium axis scores moderate (A1 ≈ 0.50), reflecting a claimed methane-plus-carbon-dioxide pattern of the kind Thompson and colleagues (2022) identify as promising, though the planet's nature as a possible hycean world complicates the interpretation. The abiotic-flux axis is moderate (A2 ≈ 0.40), since abiotic routes to methane and to dimethyl sulfide on such a world are not excluded. The robustness axis is low (A3 ≈ 0.20), because the dimethyl sulfide feature in particular rests on marginal spectral evidence at the edge of the instrument's capability. Under the minimum rule the index is BPI ≈ 0.20, and the limiting axis is again robustness, though for a different reason than at Venus: not a confusing confounder but a signal too weak to trust.

Two patterns emerge. First, none of the three cases clears a threshold for biotic provenance — if one sets a passing bar at, say, 0.7, all three fall well short, and they do so under the minimum rule even though one or two axes score respectably in each case. Averaging would have flattered all three: Enceladus, for instance, would average to about 0.58 and Venus to about 0.30, obscuring the fact that each is defeated by a specific open alternative. Second, and more usefully, the limiting axis differs across cases — robustness for Venus and K2-18b, abiotic flux for Enceladus — which tells each investigation where to concentrate effort. The sensitivity analysis confirms that these limiting-axis identifications are stable: varying every score by ±0.1 leaves Venus and K2-18b robustness-limited and Enceladus flux-limited in all combinations tested, because in each case the limiting axis is separated from the next by more than the perturbation.

Abiotic-rejection axis (0–1)	**Venus phosphine**	**Enceladus methane**	**K2-18b (CH_4 / DMS)**
A1 — Disequilibrium context	0.40	0.60	0.50
A2 — Abiotic-flux ceiling	0.35	0.45	0.40
A3 — Spectral / conceptual robustness	0.15	0.70	0.20
BPI = min(A1, A2, A3)	**0.15**	**0.45**	**0.20**
Limiting (failed) axis	A3 (detection contested)	A2 (serpentinization)	A3 (marginal spectrum)

Table 1. Biotic Provenance Index (BPI) for three touchstone detections. Each abiotic-rejection axis is scored 0–1 (1 = abiotic alternative essentially excluded); the BPI is the minimum (weakest-link) of the three axes.

Note. Values are reasoned estimates drawn from the cited literature; the scoring procedure is described in the Research Methodology section. The BPI uses the minimum rather than the mean of the three axes, because a false positive is governed by the least-excluded abiotic alternative. The limiting axis — the diagnostic output — differs across cases and is stable under ±0.1 perturbation of the scores.

THE DISEQUILIBRIUM AXIS: NECESSARY CONTEXT, INSUFFICIENT PROOF

The thermodynamic-disequilibrium axis is where the modern biosignature concept is strongest and where, on its own, it still falls short. The insight that life betrays itself by holding an atmosphere or ocean away from chemical equilibrium is genuinely powerful, because equilibrium is the default that physics enforces in the absence of a continuous free-energy source (Wogan & Catling, 2020). A biosphere is such a source, and the coexistence of species that should react away — oxygen and methane on Earth, or methane and hydrogen and carbon dioxide in a plume — is a real fingerprint of ongoing chemistry that something must be driving. This is why the axis belongs in the framework and why, for Enceladus in particular, it scores well: the plume's redox pairing is not what a quiescent, equilibrated ocean would present (Fifer et al., 2022).

But disequilibrium is necessary, not sufficient, and the reason is that abiotic processes can also drive and maintain disequilibrium. Serpentinization is itself a free-energy source; it produces hydrogen and methane far from equilibrium with the surrounding rock and water (Klein et al.,

2019). Volcanism, photochemistry and radiolysis all inject reactive species that keep a planetary envelope out of balance without any biology. The disequilibrium axis therefore cannot, by itself, distinguish a biosphere from a geosphere; it can only establish that some free-energy source is at work. Thompson and colleagues (2022) understand this, which is why their methane argument is contextual rather than thermodynamic alone — they require the disequilibrium to sit in a setting where the abiotic free-energy sources are demonstrably too weak. That requirement is really a hand-off to the second axis, and the structure of their argument vindicates the claim that the axes are independent and must all be satisfied.

The Venus phosphine case shows the axis failing in a subtler way, through ignorance rather than through an abiotic driver. Phosphine in Venus's oxidized atmosphere is a disequilibrium species, so in principle the axis should score high. It does not, because the atmospheric chemistry of Venus is so incompletely known that we cannot quantify how large the disequilibrium is or how hard it would be to maintain abiotically (Wunderlich et al., 2023). Here the axis is limited not by a competing abiotic source we can name but by the coarseness of our model of the planet. This is worth flagging as a general hazard: a disequilibrium argument is only as good as the inventory against which the disequilibrium is computed, and for poorly characterized worlds that inventory is thin. I initially scored Venus higher on this axis and revised downward once I took seriously how much of the Venusian phosphorus and sulfur chemistry remains unconstrained; the honest score reflects our ignorance, not a positive abiotic explanation.

There is a deeper methodological point here that motivates the whole framework. If the disequilibrium axis were sufficient, biosignature science would be a matter of measuring one number and comparing it to an equilibrium calculation. It is not, because the same disequilibrium can be written by life or by rock, and telling the two apart requires leaving the thermodynamic axis entirely and asking a different question: not whether something drives the disequilibrium, but whether that something must be biological. That question is the domain of the abiotic-flux ceiling, and the independence of the two questions is why averaging them is a category error. A high disequilibrium score cannot buy down a low flux-ceiling score, because they are answers to different questions, and a false positive lives in whichever question was left unanswered.

THE ABIOTIC-FLUX CEILING: WHY ENCELADUS IS HARD AND WHY GEOLOGY USUALLY WINS THE TIE

The abiotic-flux axis is the one on which ocean-world methane claims most often founder, and Enceladus is the exemplary case. The plume's methane is real, its disequilibrium context is favourable, and the instrumentation to sample it is within reach — yet the case does not clear the framework, because the axis that asks whether geology alone could supply the methane returns an uncomfortable "yes, plausibly." Serpentinization of a rocky seafloor in contact with liquid water is not an exotic possibility to be argued for; it is the expected state of affairs for a small icy moon with a differentiated interior, and it produces methane abiotically as a matter of course (Klein et al., 2019; Reeves & Fiebig, 2020). The abiotic ceiling, in other words, sits at or above the observed methane abundance, and until it can be pushed below that abundance the biotic interpretation cannot be credited.

This is a genuinely hard problem, and it will not be solved by measuring more methane. What is needed is a discriminant that separates methane made by cells from methane made by rock at the same abundance, and the most promising such discriminant is isotopic. Biological and abiotic methane can differ in bulk carbon and hydrogen isotopes, and more diagnostically in their clumped-isotope signatures, which encode formation temperature and mechanism (Labidi et al., 2024). If a plume-sampling spacecraft could measure the clumped-isotope composition of

Enceladus methane with sufficient precision, it could in principle place the methane on the biotic or abiotic side of the line. But the qualifier is heavy: clumped-isotope measurements are exacting in a terrestrial mass-spectrometry laboratory, and performing them on the trace methane captured during a hypervelocity plume transit is a formidable instrumental challenge (New et al., 2020; Mathies et al., 2021). The framework's verdict is therefore not that Enceladus is uninteresting but that its interest is contingent on an isotopic measurement that has not yet been made, and it names that measurement as the specific thing that would raise the abiotic-flux score.

There is a general lesson in the Enceladus case about the asymmetry between biology and geology in tie-breaking. When an abiotic process can supply the observed abundance, the burden of proof lies entirely on the biological interpretation, because parsimony favours the mechanism that requires no biosphere. This is not a bias against life detection; it is the correct default, and it is why Robare and Shock (2025) frame the ocean-world problem as one of distinguishing organics from an abiotic background rather than of detecting organics at all. The framework encodes the asymmetry by demanding a high abiotic-flux score — an actual exclusion of the geological route — before crediting provenance, rather than treating a plausible biological story as sufficient. A reader might object that this sets the bar impossibly high, since abiotic methane is so ubiquitous. My response is that the bar is high because the problem is hard, and that pretending otherwise is how false positives are manufactured; the honest move is to state the bar clearly and identify, as the framework does, exactly what measurement would clear it.

The exoplanet cases share the structure but shift the balance. On K2-18b the abiotic-flux axis is also unsatisfied, because abiotic routes to methane and especially to dimethyl sulfide on a hycean world are not excluded and are barely constrained. But for K2-18b the abiotic-flux problem is not even the binding one, because the detection itself is shakier than the geochemistry — which brings the argument to the third axis and to the reason the minimum rule matters.

THE ROBUSTNESS AXIS AND THE WEAKEST-LINK RULE: WHY AVERAGES MISLEAD

The robustness axis asks the most basic and most easily overlooked question: is the detection real, and is its interpretation secure against confounders we can name and alternatives we have not conceived? It is the axis on which the Venus phosphine claim ultimately foundered, and its history is a compact argument for why this axis cannot be traded against the others. The original phosphine detection had a respectable disequilibrium rationale, but the spectral feature proved fragile — attributable in substantial part to sulfur dioxide and to the aggressive data reduction used to extract it (Akins et al., 2021; Trompet et al., 2020). No strength on the other axes could rescue a signal that might not be there. A biosignature built on an artifact is not a weak biosignature; it is not a biosignature at all, and this is the qualitative fact that a weighted average cannot represent.

The deeper hazard on this axis is the one Vickers and colleagues (2023) name: unconceived alternatives, the abiotic explanations that have not yet occurred to anyone. By construction these cannot be scored directly, but their standing threat is a reason to treat any interpretation that rests on a single fragile line as insecure. Lustig-Yaeger, Meadows and Lincowski (2019) supply a concrete instance of a conceived-just-in-time alternative — Venus-like clouds producing a statistical false positive that mimics a habitable signal — and the lesson generalizes: the space of abiotic mimics is larger than our current catalogue of them, so robustness must include a discount for the alternatives we may be missing. This is why, for K2-18b, the robustness axis scores low even though no specific artifact has been proven: the dimethyl sulfide feature sits at the margin of

detectability, exactly where both instrumental artifacts and unconceived abiotic chemistries are most able to hide (Madhusudhan et al., 2020).

This axis is where the framework's combination rule earns its keep, and it is worth stating the argument for the minimum plainly. A false positive is a disjunctive failure: it occurs if the detection is spurious, or if geology can supply the gas, or if the disequilibrium is abiotically maintained, or if an unconceived alternative applies. Any one of these suffices. The credibility of the biotic interpretation is therefore bounded above by the least-excluded of these alternatives — by the minimum axis score — and averaging, which lets a high score on one axis compensate for a low score on another, systematically overstates credibility precisely when one alternative has been neglected. The Venus case is the cautionary instance: a weighted average would have assigned it a middling score and thereby dignified a claim whose spectral basis was in doubt, whereas the minimum rule correctly reports that the case is only as strong as its weakest, contested link. I acknowledge that the strict minimum is severe and that a softened rule — say, the minimum with a small bonus when the other axes are uniformly high — might be defensible for well-characterized targets; but for the frontier cases that dominate this field, where our models are thin and our detections marginal, the strict minimum is the safer discipline, and I would rather err toward demanding too much exclusion than toward crediting too little.

The diagnostic payoff of separating the axes and combining them by minimum is the framework's most practical result. Because the three cases fail on different axes, the framework tells each investigation something specific: Venus needs a secure, independently confirmed detection before anything else is worth debating; K2-18b needs higher-fidelity spectroscopy to establish that the dimethyl sulfide feature is real; and Enceladus, whose detection is secure and whose disequilibrium is favourable, needs an isotopic measurement that can push the abiotic-flux ceiling below the observed methane. A single averaged confidence number could not have delivered these three distinct prescriptions. It would have ranked the cases; the framework, by contrast, tells us what to do about each — which is what a methodology, as opposed to a scoreboard, is for.

CONCLUSION

This article set out to formalize how a detection of methane or phosphine on an ocean world or exoplanet should be evaluated so that false positives are minimized and, when a claim fails, the reason for its failure is legible. The three hypotheses that structured the inquiry are supported by the analysis. The first, that provenance is a property of the detection in context rather than of the gas, is borne out by the fact that the same molecule scores very differently across the three cases depending on planetary and instrumental circumstances. The second, that the lines of evidence are logically independent, is what licenses treating the three axes separately and is illustrated by cases that are strong on one axis and fatally weak on another. The third, that the axes must be combined by a weakest-link rule rather than an average, is the article's central methodological claim, and the Venus phosphine episode is its standing proof: only a minimum rule reports that a claim resting on a contested detection is weak regardless of its thermodynamic rationale.

The principal original contribution is the provenance-first Biotic Provenance framework and its minimum-rule index, which scores a candidate reduced-gas detection on three independent abiotic-rejection axes — disequilibrium context, abiotic-flux ceiling, and spectral and conceptual robustness — and credits biotic provenance only to the extent of the least-excluded abiotic alternative. Where existing confidence ladders aggregate evidence additively and thereby allow a strong axis to mask a weak one, the minimum rule enforces the disjunctive logic of false positives and, applied to Venus phosphine, Enceladus methane, and K2-18b, localizes the specific axis on

which each currently fails. That localization — robustness for Venus and K2-18b, abiotic flux for Enceladus — is a diagnostic a single aggregate score cannot provide, and it is, I think, the framework's most useful product.

The limitations are real and I state them without softening. The axis scores are heuristic, assigned by a single author from the literature rather than computed from a unified model, so their exact values carry an unquantified subjectivity; only the limiting-axis identifications, which the sensitivity analysis shows to be stable, should be leaned on heavily. The three axes, while I argue they are necessary and independent, may not be exhaustive — a fourth axis, perhaps for temporal variability or spatial context, could be warranted and would slot naturally into the minimum rule. The strict minimum is deliberately severe and may be too unforgiving for exceptionally well-characterized targets, where a softened rule could be justified. And the framework evaluates remote and plume-sampled detections only; it says nothing about in situ metabolic or morphological evidence, which follows a different logic. Each of these is a boundary of the present treatment rather than a flaw within it, but each deserves explicit acknowledgment.

Three directions for future work follow. The first is to replace the heuristic scores with quantitative, model-derived ones: a disequilibrium score from an actual free-energy calculation, an abiotic-flux score from a serpentinization or photochemical model with isotopic constraints, and a robustness score from a formal detection-significance and confounder analysis. The second is to apply the framework prospectively to the data Europa Clipper will return, using it before the fact to specify which measurements would most raise each axis, so that the mission's biosignature yield is designed rather than hoped for. The third is to test the minimum rule against the softened alternatives on a corpus of past claims, asking which rule would have best tracked the eventual scientific consensus. Whether the strict minimum proves to be exactly right is an open question I cannot settle here; but that the combination rule matters at least as much as the individual measurements is, I believe, the durable lesson, and it is one the phosphine affair taught the field at considerable cost.

BIBLIOGRAPHY

Akins, A. B., Lincowski, A. P., Meadows, V. S., & Steffes, P. G. (2021). Complications in the ALMA detection of phosphine at Venus. *The Astrophysical Journal Letters*, *907*(2), L27. https://doi.org/10.3847/2041-8213/abd56a

Bixel, A., & Apai, D. (2021). Bioverse: A simulation framework to assess the statistical power of future biosignature surveys. *The Astronomical Journal*, *161*(5), 228. https://doi.org/10.3847/1538-3881/abe042

Boutier, A., Vitale Brovarone, A., Martinez, I., Sissmann, O., & Mana, S. (2021). High-pressure serpentinization and abiotic methane formation in metaperidotite from the Alpine subduction zone. *Lithos*, *396–397*, 106190. https://doi.org/10.1016/j.lithos.2021.106190

Davis, C. (2025). Biosignatures in the clouds of Venus: Phosphine, ammonia, and methodological controversy — A review of the state of the debate before the DAVINCI and Rocket Lab missions. *Cosmological and Astrobiological Review*, *3*(1), 63–77. https://doi.org/10.65932/CAR-2025-1-4

Fifer, L. M., Catling, D. C., & Toner, J. D. (2022). Chemical fractionation modeling of plumes indicates a gas-rich, moderately alkaline Enceladus ocean. *The Planetary Science Journal*, *3*(8), 191. https://doi.org/10.3847/psj/ac7a9f

Glein, C. R., & Truong, N. (2025). Phosphates reveal high pH ocean water on Enceladus. *Icarus*, *441*, 116717. https://doi.org/10.1016/j.icarus.2025.116717

Glidden, A., Seager, S., Petkowski, J. J., & Ono, S. (2023). Can isotopologues be used as biosignature gases in exoplanet atmospheres? *Life*, *13*(12), 2325. https://doi.org/10.3390/life13122325

Howell, S. M., & Pappalardo, R. T. (2020). NASA's Europa Clipper—a mission to a potentially habitable ocean world. *Nature Communications*, *11*(1), 1311. https://doi.org/10.1038/s41467-020-15160-9

Huang, J., Seager, S., Petkowski, J. J., Ranjan, S., & Zhan, Z. (2022). Assessment of ammonia as a biosignature gas in exoplanet atmospheres. *Astrobiology*, *22*(2), 171–191. https://*doi.org/10.1089/ast.2020.2358*

Klein, F., Grozeva, N. G., & Seewald, J. S. (2019). Abiotic methane synthesis and serpentinization in olivine-hosted fluid inclusions. *Proceedings of the National Academy of Sciences*, *116*(36), 17666–17672. https://doi.org/10.1073/pnas.1907871116

Labidi, J., McCollom, T. M., Giunta, T., Sherwood Lollar, B., Leavitt, W. D., & Young, E. D. (2024). Clumped isotope signatures of abiotic methane: The role of the composition of the reacting fluids. *Journal of Geophysical Research: Solid Earth*, *129*(7), e2023JB028194. https://doi.org/10.1029/2023jb028194

Lorenz, R. D. (2019). A Bayesian approach to biosignature detection on ocean worlds. *Nature Astronomy*, *3*(6), 466–467. https://doi.org/10.1038/s41550-019-0810-3

Lustig-Yaeger, J., Meadows, V. S., & Lincowski, A. P. (2019). A mirage of the cosmic shoreline: Venus-like clouds as a statistical false positive for exoplanet habitability. *The Astrophysical Journal Letters*, *887*(1), L11. https://doi.org/10.3847/2041-8213/ab5965

Madhusudhan, N., Nixon, M. C., Welbanks, L., Piette, A. A. A., & Booth, R. A. (2020). The interior and atmosphere of the habitable-zone exoplanet K2-18b. *The Astrophysical Journal Letters*, *891*(1), L7. https://doi.org/10.3847/2041-8213/ab7229

Mathies, R. A., New, J. S., Golozar, M., & Butterworth, A. L. (2021). On the feasibility of informative biosignature measurements using an Enceladus plume organic analyzer. *The Planetary Science Journal*, *2*(4), 163. https://doi.org/10.3847/psj/ac0e9b

New, J. S., Kazemi, B., Price, M. C., Cole, M. J., Spathis, V., Mathies, R. A., & Butterworth, A. L. (2020). Feasibility of Enceladus plume biosignature analysis: Successful capture of organic ice grains in hypervelocity impacts. *Meteoritics & Planetary Science*, *55*(8), 1936–1948. https://doi.org/10.1111/maps.13554

Patel, M. R., Mason, J. P., Nordheim, T. A., & Dartnell, L. R. (2022). Constraints on a potential aerial biosphere on Venus: II. Ultraviolet dosimetry. *Icarus*, *373*, 114796. https://doi.org/10.1016/j.icarus.2021.114796

Pohorille, A., & Sokolowska, J. (2020). Evaluating biosignatures for life detection. *Astrobiology*, *20*(10), 1236–1250. https://doi.org/10.1089/ast.2019.2151

Reeves, E. P., & Fiebig, J. (2020). Abiotic synthesis of methane and organic compounds in Earth's lithosphere. *Elements*, *16*(1), 25–31. https://doi.org/10.2138/gselements.16.1.25

Robare, J. A., & Shock, E. L. (2025). Distinguishing potential organic biosignatures on ocean worlds from abiotic sources. *Icarus*, *429*, 116431. https://doi.org/10.1016/j.icarus.2024.116431

Sousa-Silva, C., Seager, S., Ranjan, S., Petkowski, J. J., Zhan, Z., Hu, R., & Bains, W. (2020). Phosphine as a biosignature gas in exoplanet atmospheres. *Astrobiology*, *20*(2), 235–268. https://doi.org/10.1089/ast.2018.1954

Thompson, M. A., Krissansen-Totton, J., Wogan, N., Telus, M., & Fortney, J. J. (2022). The case and context for atmospheric methane as an exoplanet biosignature. *Proceedings of the National Academy of Sciences*, *119*(14), e2117933119. https://doi.org/10.1073/pnas.2117933119

Trompet, L., Robert, S., Mahieux, A., Schmidt, F., Erwin, J., & Vandaele, A. C. (2020). Phosphine in Venus' atmosphere: Detection attempts and upper limits above the cloud top. *Astronomy & Astrophysics*, *645*, L4. https://doi.org/10.1051/0004-6361/202039932

Vickers, P., Cowie, C., Dick, S. J., Gillen, C., Jeancolas, C., Rothschild, L. J., & McMahon, S. (2023). Confidence of life detection: The problem of unconceived alternatives. *Astrobiology*, *23*(11), 1202–1212. https://doi.org/10.1089/ast.2022.0084

Wogan, N. F., & Catling, D. C. (2020). When is chemical disequilibrium in Earth-like planetary atmospheres a biosignature versus an anti-biosignature? *The Astrophysical Journal*, *892*(2), 127. https://doi.org/10.3847/1538-4357/ab7b81

Wunderlich, F., Grenfell, J. L., & Rauer, H. (2023). Uncertainty in phosphine photochemistry in the Venus atmosphere prevents a firm biosignature attribution. *Astronomy & Astrophysics*, *676*, A135. https://doi.org/10.1051/0004-6361/202142548

Zhan, Z., Huang, J., Seager, S., Petkowski, J. J., & Ranjan, S. (2022). Organic carbonyls are poor biosignature gases in exoplanet atmospheres. *The Astrophysical Journal*, *930*(2), 133. https://doi.org/10.3847/1538-4357/ac64a8

BIOPOTPISI U ATMOSFERAMA OKEANSKIH SVJETOVA: METODOLOŠKI IZAZOVI RAZLIKOVANJA BIOTIČKIH OD ABIOTIČKIH IZVORA METANA I FOSFINA

Aleksej Vladimirovič Sokolov

Jaroslavski državni univerzitet „P. G. Demidov"
Jaroslavlj, Ruska Federacija
E-mail: aleksei.vs9@uniyar.ac.ru

Originalni naučni članak

Primljeno: 01.05.2026. Prihvaćeno: 17.06.2026.

DOI: https://doi.org/10.65932/CAR-2026-1-5 **UDK**: 573.5:523.4

Sažetak: Detekcija pretpostavljenog biopotpisnog gasa jaka je tek onoliko koliko su jake abiotičke alternative koje uspijeva isključiti, a nedavna istorija — prije svega kontroverza oko fosfina na Veneri — pokazuje kako se navodni znak života brzo raspline u instrumentalni artefakt ili običnu geohemiju. Ovaj rad se tim metodološkim problemom bavi neposredno, za dva reducirana gasa koja se najčešće predlažu kao biopotpisi na okeanskim svjetovima i egzoplanetama u nastanjivoj zoni: metan i fosfin. Tvrdim da postojeći okviri griješe postavljajući pitanje „od gasa" — je li metan, ili fosfin, biopotpis? — i kombinujući dokaze ponderisanim prosjekom koji dopušta da snaga na jednoj osi prikrije neriješenu slabost na drugoj. Umjesto toga razvijam okvir biotičkog porijekla „od porijekla", koji svaku detekciju ocjenjuje na tri logički nezavisne ose isključivanja abiotičkog: termodinamički kontekst neravnoteže, gornja granica abiotičkog fluksa, i spektralna i konceptualna robusnost. Presudno, okvir kombinuje ose pravilom najslabije karike — Indeks biotičkog porijekla je minimum triju ocjena, a ne njihov prosjek — jer je lažno pozitivan rezultat određen najmanje isključenom abiotičkom alternativom, a ne prosječnom snagom slučaja. Primijenjen na tri ključna slučaja, okvir tačno locira gdje svaki od njih pada: fosfin na Veneri ograničen je spektralnom robusnošću, uz spornu samu detekciju; metan i dimetil-sulfid na K2-18b ograničeni su istom osom, počivajući na rubnim spektralnim odlikama; a metan na Enceladusu ograničen je granicom abiotičkog fluksa, jer serpentinizacija može isporučiti opaženu količinu. Nijedan od tri slučaja trenutno ne prelazi prag biotičkog porijekla, ali svaki pada iz drugog, prepoznatljivog razloga.

Ključne riječi: *biopotpisi, okeanski svjetovi, metan, fosfin, lažno pozitivni rezultati, Enceladus, Europa, egzoplanete, spektroskopija.*