nature portfolio

Peer Review File

Measurement of reactor neutrino oscillation with the first JUNO data

Corresponding Author: Professor Liangjian Wen

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
Title:
First measurement of reactor neutrino oscillations at JUNO

This manuscript presents the first physics results from the JUNO experiment: a precision measurement of the solar neutrino oscillation parameters using reactor antineutrinos collected during the commissioning phase. The analysis delivers significantly improved determinations of Δm^2_{21} and $\sin^2\theta_{12}$ compared to current world averages, while simultaneously demonstrating that a 20-kton liquid-scintillator detector can achieve the targeted few-percent-level energy resolution. These results not only validate the detector performance and analysis methodology but also establish JUNO as a key player in the emerging precision era of neutrino oscillation physics, with direct implications for tests of the three-flavor paradigm, global oscillation fits, and future determinations of the neutrino mass ordering.

General Assessment
The scientific quality and technical execution of the work are outstanding. The successful commissioning of a 20-kton liquid-scintillator detector with such high energy resolution is a major milestone in experimental neutrino physics. The data quality appears robust, and the analysis is carefully designed and thoroughly cross-checked.
At the same time, for a broad audience journal like Nature, the manuscript currently leans heavily on demonstrating “technical success” (calibration, stability, background rejection), while the “scientific necessity” and broader impact of the new precision regime are somewhat underplayed. The paper would, in my view, be substantially strengthened by streamlining technical details in the main text and expanding the discussion of the physics reach and conceptual implications. Addressing the comments below—particularly by sharpening the narrative around the broader physics goals—would bring the manuscript much closer to meeting the standards of Nature. Although the manuscript is technically correct, the language is often too specialized for Nature’s broad readership (including physicists who are not neutrino-oscillation experts). I recommend simplifying the main-text narrative by reducing jargon and acronyms, adding short qualitative context before technical details, and moving highly implementation-specific discussion to Methods/Extended Data. This would improve readability without compromising rigor.
Major Comments
The authors present their results primarily as numerical improvements over previous measurements. While the reduction in uncertainty is indeed impressive (~1.6% for Δm^2_{21} and ~2.8% for $\sin^2\theta_{12}$), the manuscript would benefit from a more explicit discussion of why moving towards sub-percent precision is transformative for the field.
To make the broader physics motivation more apparent, the authors could more explicitly articulate JUNO’s long-term physics goals. For example, the Introduction and Discussion might briefly organize the narrative around themes such as (the list below is not meant to be exhaustive):
1) From discovery to precision tests of the three-flavor framework.
The present results already show that JUNO is operating at a level that will soon enable sub-percent precision on multiple oscillation parameters. This is more than simply “a better measurement”: it marks a transition of neutrino oscillation studies from a discovery phase to a genuine precision-test regime, where non-trivial internal consistency checks of the three-flavor paradigm across different channels and baselines become possible. Stating this explicitly would help non-specialist readers

appreciate the conceptual importance of achieving such precision.
2) Role in global fits and in breaking parameter degeneracies.
Although this analysis focuses on the solar parameters, the demonstrated detector performance implies that JUNO will soon measure the effective atmospheric mass splitting Δm^2_{ee} with high precision. It would be helpful to spell out how having Δm^2_{21}, $\sin^2\theta_{12}$ and Δm^2_{ee} all measured precisely within a single experiment impacts global oscillation fits. In particular, fixing these parameters independently reduces degeneracies in long-baseline accelerator analyses and indirectly facilitates a more robust determination of the θ_{23} octant and the CP-violating phase δCP, even though these quantities are not directly accessible via reactor antineutrino disappearance.
3) Connection to the mass ordering and to broader questions in fundamental physics.
While the present data set does not yet determine the neutrino mass ordering, the results validate JUNO's capability to do so in future runs. A brief but explicit connection between a future normal/inverted ordering determination and broader questions —such as the interpretation of neutrinoless double-beta decay searches (and the Majorana nature of the neutrino) and cosmological constraints on the sum of neutrino masses—would help convey to the Nature readership why this program is of wide and lasting interest.
4) Clarify oscillation regimes and the origin of mass-ordering sensitivity (p.10)
On p.10 (first column), the discussion of the "first solar oscillation maximum" is potentially confusing because it mixes (i) the slow spectral distortion driven by Δm^2_{21} and $\sin^2\theta_{12}$ with (ii) the fast atmospheric-scale oscillations governed by $\Delta m^2 3_1$ and $\Delta m^2 3_2$, and (iii) the mechanism that provides mass-ordering sensitivity, which comes from the energy-dependent interference between the two close atmospheric frequencies $\Delta m^2 3_1$ vs $\Delta m^2 3_2$), not from the solar maximum itself. I strongly suggest rewriting this paragraph for clarity. A simple figure showing Pee(E) at L ~ 52.5 km (solar-only vs full; and optionally NO vs IO, or NO–IO difference) would immediately remove ambiguity; this could be placed in Extended Data.

5) Solar–reactor consistency as a test for new physics.
A key virtue of high-precision reactor measurements of Δm^2_{21} and $\sin^2\theta_{12}$ is the direct comparison with solar neutrino results, which probe a CPT-conjugate channel and are subject to matter effects in the Sun. Any statistically significant tension between solar and reactor determinations could point to physics beyond the standard three-flavor picture, for example non-standard neutrino interactions or other flavor-dependent matter effects. JUNO is uniquely positioned to perform such consistency tests; the present measurement can be presented as the first step in this program.

6) Consistency of the PMNS framework and unitarity tests.
By precisely pinning down $|U_{e1}|$ and $|U_{e2}|$ through $\sin^2\theta_{12}$ and Δm^2_{21} (and eventually Δm^2_{ee}), JUNO provides crucial input for global tests of the unitarity and internal consistency of the PMNS framework when combined with accelerator and atmospheric data. If the authors wish, a short remark in this direction could further emphasize that JUNO contributes not only isolated measurements but also stringent cross-checks of the overall three-flavor scheme.

7) Refining reactor antineutrino flux models.
Finally, the high-statistics, high-resolution reactor spectrum at JUNO—especially in combination with the TAO near detector —will allow stringent tests of reactor antineutrino flux predictions and may shed light on existing spectral anomalies. Even if detailed flux studies lie beyond the scope of the present analysis, a brief mention of this long-term capability would broaden the perceived impact of the experiment.
These examples are intended as possible directions rather than a fixed list; the authors may well have additional physics goals or synergies in mind that could be highlighted to further strengthen the overarching narrative.

Technical detail and narrative flow
The main text currently devotes substantial space to calibration procedures, energy-scale characterization and background rejection (for example, in the discussion around Figure 2). These technical elements are essential for establishing the robustness of the result, but in their present form they somewhat interrupt the narrative flow for a non-specialist reader.
I suggest that the authors consider moving the most detailed performance plots (such as the energy-scale residuals and certain background-rejection efficiency plots) to the Extended Data or Methods sections and keeping in the main text only a concise, qualitative summary of the calibration strategy and performance, with clear references to the corresponding Extended Data figures. This re-organization would allow the oscillation spectrum and fit (currently Figure 3) to take center stage in the main text and would help refocus the narrative on the physics outcome and its implications. The full technical validation would remain available in the appropriate sections for expert readers who wish to examine the analysis in detail.
While detailed technical material is appropriately placed in the Methods, there are a few conceptual elements currently discussed only there that, in my view, deserve a brief summary in the main text. A short, equation-free treatment would make the physics more transparent for non-specialist readers without compromising the concision of the paper. Concretely, the authors might consider the following:
Qualitative origin of the oscillation signal.
The Methods provide a clear explanation of how the electron antineutrino survival probability is dominated, in this configuration, by the "solar" term driven by Δm^2_{21} and θ_{12}, with a subdominant contribution from the atmospheric term. It would help readers if this were summarized in the main text near the discussion of Fig. 3. For example, the authors could add a short paragraph such as:
"The observed distortion of the reactor antineutrino spectrum is governed predominantly by the 'solar' oscillation term driven by Δm^2_{21} and θ_{12}, which introduces a slow, baseline-scale modulation of the event rate. A subdominant contribution arises from the 'atmospheric' oscillation term, which produces a finer interference pattern that is only weakly relevant for the present analysis (without a brief clarification, non-expert readers may not understand the historical origin of the "solar" and "atmospheric" terms). In this work, the fit is therefore primarily sensitive to the solar parameters, while remaining largely insensitive to the atmospheric mass splitting."

Semi–model-independent treatment of the reactor flux.
The semi–model-independent approach to reactor flux modelling, anchored by Daya Bay near-detector data, is an important ingredient that currently appears only in the Methods. Given the long-standing interest in reactor flux systematics, it would be worth summarizing this strategy in one or two sentences in the main text, together with the statement that the residual model dependence is small. For instance:
"To minimize the dependence on reactor antineutrino flux models, we adopt a semi–model-independent strategy in which the JUNO data are fitted using a segmented spectral function that is shared with the Daya Bay near detectors. In this approach, Daya Bay data effectively anchor the overall normalization and coarse spectral shape, while theoretical calculations provide the fine structure and its uncertainty. After optimization, the residual model dependence of the extracted Δm^2_{21} and $\sin^2\theta_{12}$ is below the 0.3% level, well below the total experimental uncertainty."
Relevance of the current energy resolution for this measurement.
The Methods usefully clarify that the achieved positron energy resolution, while slightly worse than the nominal design value, is not critical for the present analysis of the solar parameters. Since energy resolution is a central performance metric for JUNO, it may help to make this point explicit in the main text to avoid any confusion about the impact on the current result versus future NMO sensitivity. For example:
"The achieved positron energy resolution in this commissioning data set, $\sigma_E/E \approx \ldots$ at 1 MeV, is slightly worse than the nominal design value. However, for the determination of the solar oscillation parameters reported here, the result is driven mainly by the overall spectral distortion over a wide energy range and is only weakly sensitive to small changes in resolution. The present measurement is therefore not limited by the current resolution, whereas the full design performance will be more critical for future determinations of the neutrino mass ordering."
High-level summary of background constraints.
The Methods contain a very complete treatment of how the dominant backgrounds (cosmogenic $^9Li/^8He$, geoneutrinos, antineutrinos from other reactors, etc.) are constrained using control samples, external information and simulations. In the main text, it would be helpful to summarize this logic in a single sentence, emphasizing that backgrounds are both subdominant and independently constrained, and thus unlikely to mimic the oscillatory signal. For instance:
"All major backgrounds, including cosmogenic $^9Li/^8He$, geoneutrinos and antineutrinos from other reactors, are constrained by a combination of data-driven control samples, external measurements and simulations, and remain small compared to the oscillated reactor signal. Within the fitted energy window, these independently constrained background components cannot reproduce the characteristic spectral distortion induced by oscillations, and their uncertainties are subdominant in the error budget on Δm^2_{21} and $\sin^2\theta_{12}$."
Multiple independent analysis chains.
The existence of three independent analysis implementations, with different treatments of the covariance matrix and systematics, is a strong robustness check that is currently described primarily in the Methods. Since this is the first major physics result from JUNO, it would be worthwhile to highlight this more clearly in the main text. A possible formulation could be for instance:
"The analysis framework has been implemented in three independent analysis chains that employ different treatments of the covariance matrix and systematic uncertainties. All three analyses yield mutually consistent best-fit values and uncertainties for Δm^2_{21} and $\sin^2\theta_{12}$; the results quoted in this work correspond to one representative implementation, with the spread among the three providing an additional robustness check."
These additions need only be at the level of one short paragraph or sentence per item; the detailed derivations and numerical inputs can remain in the Methods, where they are entirely appropriate.

Suggestions for simplifying Figure 3
I have a few additional, more concrete suggestions aimed at simplifying Fig. 3 and making the main physics message easier to grasp:
Because the non-oscillated prediction is much higher than the oscillated spectrum, the vertical scale of the main panel is set by this large reference curve. As a result, the individual background components are effectively unreadable in this plot. I would therefore recommend combining all subdominant backgrounds into a single "total background" contribution in Fig. 3 and showing the detailed background breakdown in a separate Extended Data/Methods figure. This would both improve legibility and allow the legend in the main panel to be simplified.
The colored band used to represent the signal makes the main panel visually heavier and slightly harder to read. The authors may wish to consider whether this shading is really necessary, or whether a simpler representation would serve the physics message better.
A thinner continuous line for the non-oscillated prediction, together with slightly reduced line thickness for the background component, could help to place visual emphasis on the best-fit oscillated spectrum. Making the best-fit curve the most visually prominent object in the plot would guide the reader's eye directly to the oscillation effect.
Finally, since the y-axis is labelled "Events / 0.1 MeV" while the bin widths are not uniform (as is apparent from the figure), it might be helpful to add a short remark in the Supplementary/Methods material explaining how to interpret this quantity in the presence of varying bin sizes. A brief clarification would avoid possible confusion for readers who are less familiar with such conventions.
Taken together, these changes would make Fig. 3 cleaner and more immediately interpretable, while all detailed information about backgrounds and binning could still be documented in the Extended Data and Methods.

Comments on Figure 4
Figure 4 is a central physics summary and plays an important role in conveying the impact of the measurement. The content is appropriate, but a few aspects of the presentation could be clarified.
The figure combines 2D confidence regions in the ($\sin^2\theta_{12}$, Δm^2_{21}) plane with 1D confidence intervals shown in the marginal $\Delta\chi^2$ projections. This is reasonable, but the different numbers of degrees of freedom are not fully transparent to a non-expert reader. I see three possible ways to make this clearer:

- explicitly state in the caption that the contours in the central plane correspond to 2D confidence regions (two degrees of freedom, with the appropriate $\Delta\chi^2$ thresholds for 1σ, 2σ, 3σ), while the 1D intervals in the marginal plots use $\Delta\chi^2 = 1, 4, 9$ for one degree of freedom;
- alternatively, keep all genuinely 2D information in the central panel and all purely 1D intervals in the marginal projections, so that each panel uses a single, well-defined convention;
- as a third option, the authors could consider using a one–degree-of-freedom convention for JUNO everywhere ($\Delta\chi^2 = 1, 4, 9$ both in the plane and in the projections), as is sometimes done when profiling; in that case, it would be important to state explicitly that the plotted "1σ, 2σ, 3σ" regions are defined using the 1 d.o.f. $\Delta\chi^2$ thresholds.

In the central panel, the shaded JUNO confidence region is visually very prominent. This makes JUNO's result clearly visible, which is good, but it also makes the plot somewhat heavy. The authors might consider slightly reducing the opacity of the shading or restricting the filled area to the innermost (e.g. 1σ) region and indicating outer confidence levels with contours only. Coloring only the 1σ region would also be more coherent with the way the 1D projections emphasize the 1σ intervals and would give a more uniform visual convention across the different panels.

Since several external measurements are shown, it is important that all curves remain distinguishable and that JUNO's result is still the first thing the eye is drawn to. From a graphical point of view, the authors might consider giving JUNO a clearly visible best-fit marker and slightly thicker contours, and using distinct line styles (solid, dashed, dotted) together with a color-blind–friendly palette for the external constraints. This is not a physics convention, but a standard accessibility/readability recommendation and would make the figure easier to interpret.

Finally, the caption could be slightly more explicit about which contours and shaded regions correspond to JUNO, which line styles/colors correspond to each external experiment, and which confidence-level conventions (number of degrees of freedom and $\Delta\chi^2$ thresholds) are used in the central panel and in the marginal plots. A brief clarification there would remove any ambiguity about how to interpret the contours and intervals.

Comments on the Methods section

The Methods section is comprehensive and technically solid, and it contains all the ingredients that one would expect for a first precision result from JUNO (oscillation formalism, detector and calibration, reconstruction and selection, reactor-flux modelling, background treatment, and statistical analysis). In my view, the main improvements would be in tightening the exposition and making the structure easier to navigate, rather than in adding more technical material.

I have a few concrete suggestions in this direction. It would help if the Methods opened with a very short "roadmap" paragraph that explicitly states how the section is organized (oscillation formalism; detector and calibration; reconstruction and event selection; reactor flux and backgrounds; statistical analysis). Clear subheadings following this logic would make it easier for readers to locate specific parts of the analysis and see how they fit together.

The description of the calibration systems and waveform reconstruction is very detailed, which is valuable for experts. Some of the most technical implementation details (for example, specific algorithmic choices or numerical settings that are not directly propagated into the oscillation fit) could, however, be moved to Supplementary/Extended Data. In the main Methods text, it would be sufficient to emphasize: (i) which calibration inputs determine the energy scale, non-linearity and resolution models used in the fit, and (ii) how the corresponding uncertainties are propagated into the covariance matrix.

For the backgrounds, a short introductory paragraph summarizing how each major component is constrained (control samples, external inputs, simulations) and how its uncertainty enters the fit would improve readability, with the detailed cross-checks left to sub-subsections or Supplementary material.

The statistical framework based on the χ^2, the construction of the covariance matrix, and the treatment of nuisance parameters are central to the credibility of the result. It would be useful to state in one place which classes of systematics are encoded as nuisance parameters with priors, which are represented through covariance matrices, and which are constrained externally. A brief remark on which parameters are profiled in the fit and which are constrained by Gaussian priors would make the methodology more transparent. This would also be a natural place to mention that three independent analysis chains, with different treatments of the covariance matrix and systematics, give mutually consistent results, referring to the corresponding figure in the Methods.

To summarize, the Methods section is comprehensive and technically detailed, which is fully appropriate for a first flagship physics result from JUNO. However, the density of the presentation can make the logical progression from detector description to reconstructed spectrum and finally to the oscillation fit somewhat difficult to follow, especially for readers who are not experts in large liquid-scintillator experiments. The readability could be improved, without removing technical content, by slightly strengthening the structural guidance within this section. In particular, it would be helpful to open the Methods with a short roadmap paragraph that explicitly states how the material is organized (oscillation formalism; detector and calibration; reconstruction and event selection; reactor flux and backgrounds; statistical analysis). In addition, each major subsection could begin with a brief explanatory summary (one or two sentences) that states its purpose, identifies the key ingredients, and explains how its outputs enter the final oscillation analysis. Such introductory statements would assist the reader in navigating the detailed material and make the overall structure and logic of the Methods more transparent.

Comments on the figures

Figure 5 is conceptually valuable: it provides an intuitive, equation-light visualization of the slow ("solar") versus fast ("atmospheric") oscillation components and explains why JUNO's ~52.5 km baseline is special compared to Daya Bay and KamLAND. This is exactly the kind of orienting plot that helps a broad readership understand what sets JUNO apart. From a presentation point of view, the figure is rather small and contains several visual elements at once. As a result, the large colored circles and the experiment labels draw attention away from the red/blue curves, which carry the main physics message. In addition, the logarithmic baseline axis may not be immediately intuitive for non-specialist readers. The plot would be clearer if the "detector-size" circles were made less prominent (e.g. smaller or more transparent) and if the relevant experimental baselines were indicated more directly (for example with simple vertical markers). If the circle-size encoding is kept, it would be helpful to print the baseline value next to each experiment name and to label the two curves directly on the plot (solar-only vs full P_{ee}), so that the reader does not need to rely on the caption to identify them.

Figure 7 is a useful validation plot because it simultaneously supports two key claims: (i) the selected IBD candidates are spatially uniform within the fiducial volume, and (ii) the result is stable across multiple, independent vertex reconstruction algorithms. This directly underpins the robustness of the fiducial-volume definition and the control of edge-related backgrounds. In terms of clarity, the choice of plotting in cubic radius (R^3) is defensible for uniform-density arguments, but it is not immediately intuitive to non-expert readers. A simple improvement would be to add a secondary x-axis in R (meters). The current color choices in Fig. 7 are not fully color-blind friendly; using a color-blind–safe palette (and/or adding distinct line styles or markers) would improve accessibility and ensure all curves remain clearly distinguishable. A similar issue applies to the point colors in Figure 11.
Finally, Given the flagship nature of this first JUNO precision result, it would be helpful if the Data Availability / Code Availability statements clarified which tabulated data products and key analysis inputs will be made publicly available at publication (or provided upon reasonable request), at least at the level needed to reproduce the published fit using the binned likelihood/covariance model.

Referee #2

(Remarks to the Author)
The manuscript reports first results from the JUNO reactor neutrino experiment. We commend the authors for delivering a quick and impactful first result from JUNO. After a (mere) 60 day exposure, the experiment provides world-leading measurements of the solar mixing parameters, sin^2(theta_12) and dm^2(21). The paper describes the experiment and the data analysis, which foreshadows the experiment's ability to achieve its long-term goal of resolving the neutrino mass ordering. This result has been eagerly awaited by neutrino particle physicists, but is of equal interest to a more general population as the first of the next-generation, precision neutrino oscillation experiments that will take the stage over the next 5-10 years. These results carry substantial significance, demonstrating the potential of JUNO to advance our understanding of fundamental particle physics. We believe the manuscript worthy of publication in Nature, contingent upon satisfactory responses to a few issues we will raise.

Major comments:

1. Introduction
1. It would be useful to the non-expert reader to explain how the different oscillation parameters shape the spectrum. Such an explanation early on would help explain why the energy resolution is important to determine the mass ordering (first full paragraph on page 11).

2. Calibration & Reconstruction:
1. Fig. 2a shows the spallation neutron sample response differs from the AmC source by ~1% at small radii, while neutron captures from nuebar show a similar difference at larger radii. However, in Fig. 2b, both show a ~0.5% difference, which we assume is averaged over radius. Would it be more appropriate to assign a conservative systematic uncertainty at the 1% level to cover the full radial range observed across these three neutron sources? If so, how would such an increase propagate to the final measurement?
2. While it is reasonable to describe the reconstruction algorithms in the Methods section, it would be helpful to summarize the achieved performance in the main text. Specifically, please quantify the values (or ranges) for the energy resolution, vertex resolution, and timing resolution.
3. Figure 7 in the Methods demonstrates a uniform distribution of events within the fiducial volume and indicates no significant radial dependence in the reconstructed event rate, except near the detector boundaries, where a relative uncertainty of 1.6% is assigned based on the observed deviation. Do the reconstruction algorithms perform consistently with respect to the spectral shape, particularly as a function of detector radius and incoming neutrino energy? Are any observed variations fully contained within the assigned uncertainty thresholds?
4. In methods, page 18, right column, 3rd paragraph, it says that Channels with low occupancy or timing misalignment were excluded from analysis. How many channels as a percentage of the total were excluded?
5. In methods, the point-like event reconstruction is described on p19. PDFs are brought up in the first paragraph, are these the same PDFs that are described in the following 3 paragraphs for each of the separate reconstruction algorithms? How are the PDFs used in the event reconstruction?
6. In the description of OMILREC, the vertex reconstruction was improved by about 25% to 50%. What do these numbers quantify?

3. Detector Response model
1. I find the description of the fit to the gamma source energy spectra confusing. What are the parameters included in the fit? How are the uncertainties in these fit parameters incorporated into the oscillation fit? What feature of Figure 6 shows that the positron non-linearity uncertainty is 1%?
2. Comparisons show that the instrumental non-linearity for the gamma sources has minimal impact, but on what quantity?

The three analyses agree on what to within 0.5%?

4. Selections:
1. Signal:
1. Page 12, column 2: The term “spatio-temporal coincidence” is too vague to adequately describe the primary signal selection criteria (moreover, it is referred to as prompt-delayed coincident cuts in methods) Fig. 8 clarifies the requirement, is a compelling demonstration of the strength of the cut, and is arguably more informative than Fig. 1 for illustrating the cleanliness of the IBD coincidence signature. The authors might consider promoting that plot to the main body of the paper.
2. Reference to the 10-cm cubes on page 12 column 2, last paragraph is confusing. While it is described in methods, there is not enough context at this point in the paper to understand what that means or why it is relevant.
3. Is figure 7 after all cuts (except the FV cut)?

2. Background:
1. Page 13, Geoneutrinos and world reactor background: The event rates quoted in the text differ from those reported in Table I. This discrepancy presumably arises from the application of the 69.9% detection efficiency correction. Please state this explicitly to avoid confusion.
2. Are all background components efficiency-corrected in the same manner? For geoneutrinos and world-reactor neutrinos, which constitute irreducible IBD signals, applying the same efficiency correction as for the signal appears justified. However, this may not be appropriate for other background categories, which are affected by only a subset of the selection cuts. Please clarify whether efficiency corrections are applied to each background component and, if so, the size of those corrections.

5. Uncertainties:
1. Breaking down the statistical and systematic components of the total uncertainty would help support the statement that these measurements are statistically limited, while also providing an estimate of the relative contribution of each component.
2. Recommend including a dedicated section or paragraph on systematic uncertainties in the main text, as the current discussion is spread across multiple sections.
3. It would be helpful to clarify the impact of various systematics not only on predicted spectra but also on the extracted oscillation parameters. Please include all relevant sources, such as, calibration (energy scale, non-linearity), detector efficiency, number of target protons, various backgrounds, and reactor flux. While Table 2 summarizes some of these for the reactor neutrino rate, it does not indicate how the uncertainty modifies the shape of the non-oscillated spectrum nor how that uncertainty affects the mixing angle or the mass-squared splitting. A more comprehensive table including calibration-related systematics and their effect on both the mixing angle and mass-squared splitting would provide a clearer overview, highlighting similarities or differences in how each systematic impacts the oscillation parameter measurements.
4. Figure 11: It is great to see the consistency of the pulls across the three different analysis strategies. Would it be possible to extend this figure to include all sources of systematic uncertainty, such as energy scale, non-linearity, flux parameters, and oscillation parameters?

6. Spectral fit and Results:
1. Daya Bay near detector data is used as input to the non-oscillated spectrum prediction. More information should be provided as to why this is appropriate and what components of the flux model are common to the two experiments and what parts differ. The uncertainties on the integrated flux are discussed, but how much can the shape of the prediction change under different assumptions? It would be useful to include an error band on the non-oscillated prediction in figure 3.
2. Is the flux model fit at the same time as the oscillation parameters?
3. The same Daya Bay data appears to be used twice — once to constrain the uncertainties on reactor fluxes and overall normalization, and then again to constrain the values on sin^2(2th_13) and dm31. If this is the case, could the authors clarify how double counting is avoided? In particular, why is a simultaneous fit of Daya Bay and JUNO data not needed to properly account for correlations between reactor flux and normalization constraints and the corresponding oscillation parameters?
4. Page 14, Column 2: Daya Bay reports constraints on sin^(2th13), and it is sin^2(2th_13) which enters the JUNO oscillation formula (Eqn 1). Presumably, then the constraint should be applied in sin^2(2*th13) directly rather than sin^2(th_13) as the text suggests?
5. Is the constraint included as a full 2D chi^2 surface, including correlations, or is each constraint included as 1D pull term? Presumably it is the former, but please specify.
6. Daya Bay reports two values of dm31 under normal and inverted mass ordering hypotheses. While it is briefly noted that the results for the inverted ordering are fully compatible, given that JUNO’s primary goal is to determine the neutrino mass ordering, presenting results under only one hypothesis seems incomplete. We would recommend quoting the corresponding numbers and errors for both scenarios, or quantifying what fully compatible means. If fully compatible, does the abstract need the qualifier that says the result is only for the normal mass ordering?
7. There is an additional statement that leaving dm31 unconstrained had a negligible impact on the results – what exactly does this imply? Was dm31 allowed to float freely to both positive and negative values without any bounds? Could the term “negligible” be quantified more precisely, for example as a percentage impact on the final results?
8. Where does the value of the earth’s crust density come from? Both the central value and the error differ from commonly accepted values used in long-baseline experiments. Was the 10% uncertainty quoted on the parameter included in the fit, or was rho held constant? How were the variations in the crust density determined to be negligible?
9. With this updated JUNO measurement, what is the current significance on the difference between the reactor and solar

determination of dm21? Perhaps mention JUNO's solar capabilities to probe this difference. This is a discussion that could potentially make the article more compelling to the non-expert scientist reader.
10. Do the iso-deltachi2 contours provide the correct coverage?

7. References:
1. It would be preferable that references 1 and 2 point to the original literature on oscillations. The appropriate references are included later on as references 7 and 8. References 1 and 2 may be kept, but should also say "and references therein"
2. Reference to the MINOS experiment should be given in addition to Reference 15, the K2K experiment. The appropriate citation is https://doi.org/10.1103/PhysRevLett.97.191801
3. In regards to the measurement of theta_13, it would not hurt to cite the other two short baseline reactor experiments RENO and DoubleChooz as these are included in the PDG average value of theta_13. The authors may also consider citing T2K (https://doi.org/10.1103/PhysRevLett.97.191801) or (https://doi.org/10.1103/PhysRevLett.97.191801) to emphasize the complementarity of reactor and accelerator neutrino experiments.

Copy editing comments:
The following are a few stylistic suggestions that would improve the readability of the manuscript

1. Is the acronym NMO required? It is easily conflated with Normal Mass Ordering
2. The detector description in the JUNO experimental setup includes many acronyms, some of which are not used again in the main body of the paper. It would be easier on the reader if the acronyms were omitted from the main body of the paper, but introduced in the methods section. At the very least, it would be helpful to reintroduce them so the reader doesn't have to flip back and forth over several pages.
3. Out of curiosity, why are 2 different types of large PMTs used?
4. P10 right column, 1st full paragraph, first line, drop "of" before LS
5. P17, first paragraph of Detector filling and commissioning, 5th line, Water->water
6. P20 first full paragraph on left column, 3rd line, "having" should be "have"
7. P21, in "fiducial volume cut" paragraph, 9th line, add "the" before "z axis"
8. P22, 0th paragraph on right hand side, line 10, "vetos" should be "veto"
9. P24, 0th paragraph on left, line 7, add "the" before "alpha"
10. P25, out of curiosity, what does MI stand for in S_MI (eq 3)?

Referee #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #4

(Remarks to the Author)
This paper presents the first results from the JUNO experiment and details a comprehensive data analysis incorporating (i) calibrations of the detector response; (ii) background-suppressing event-selection criteria that enhance IBD purity; and (iii) systematic-uncertainty assessments from detector modeling, event selection, backgrounds, and signal prediction. The final oscillation fit is obtained using a dedicated statistical methodology and a blinded analysis strategy.

Using the first 59.1 live days of data from its 20-kton liquid-scintillator detector, the JUNO Collaboration has achieved the world's most precise measurements of the oscillation parameters $\sin 2\theta_{12}$ and Δm^2_{12}, improving the global precision by a factor of 1.6 in both. These fundamental parameters play a central role in the Standard Model description of lepton mixing, and tightening their uncertainties has broad impact, from testing the three-flavor framework to strengthening global oscillation fits and informing future neutrino-physics programs.

As the world's largest liquid-scintillator detector, JUNO's ability to achieve this level of precision after such a short period of data taking is remarkable. Reaching such performance so early in the experiment's lifetime underscores the excellent quality of the detector performance, calibration strategy, and analysis framework. Equally striking is JUNO's first simultaneous observation of both atmospheric and solar neutrino oscillations, a milestone that highlights the detector's unprecedented sensitivity.

Taken together, these accomplishments demonstrate JUNO's strong potential for major future breakthroughs in neutrino physics and its central role in refining the global three-flavor framework. Given the significance and broad impact of these results, the work is well aligned with the standards of Nature.

At the same time, we believe the paper is suitable for publication once the authors have adequately addressed the following comments.

Major comments for considerations and improvement:

1. In abstract, the authors state: "These results advance the basic understanding of neutrinos, validate the detector's design, and confirm JUNO's readiness for its primary goal of resolving the neutrino mass ordering with a larger dataset.", which is

also repeated in the final "discussion and outlook" section. We find this statement overly strong given the current results, and we recommend that the authors revise it to more accurately reflect what can be supported by the presented results.

Here is a plot illustrate the achieved (taken from https://arxiv.org/abs/2511.14590v1) vs. expected energy resolution (https://arxiv.org/pdf/2405.18008) The resolution, if not improved, would significantly limit JUNO's capability in the NMO determination. Because of this, it is not appropriate to claim "confirm JUNO's readiness for its primary goal of resolving the neutrino mass ordering with a larger dataset".

We understand that the current energy resolution is enough for the θ_{12} and Δm^2_{12} measurement. In addition, further improvements in the calibration and understanding of the detector will likely lead to improvement in the energy resolution, but it is not yet appropriate to claim JUNO's "readiness for mass ordering determination".

2. Related to the previous comment, C-14 pileup is discussed as a potential explanation for the degraded energy resolution observed in the Ge-68 data. If this is indeed the dominant effect, it would be helpful to understand why a similar pileup does not appear to affect the α events, which achieve better than 3% energy resolution at comparable energies.

Alternatively, if the difference primarily arises from the use of an extended source versus a point-like source, this distinction should be clarified, as improving the energy resolution for positron in this case may pose a more fundamental challenge than mitigating pileup effects alone.

3. "Shape uncertainty" appears repeatedly in the oscillation analysis, for example:
i) Page 13: "This comparison constrained the spectral shape uncertainty to 20%";
ii) an overall energy uncertainty of 0.5% was assigned? A single number?
iii) "A 5% shape uncertainty accounts for the two spectral models used [49, 50], the $^{238}U/^{232}Th$ compositional variations";
iv) "This background contributes to 1.26 IBD events per day at the JUNO site, with 10% rate and 5% shape uncertainties."
v) this comparison constrained the spectral shape uncertainty of (Li9/He8) to 20%.

Given the importance of shape uncertainties in the spectral fit and in determining the final oscillation parameters, it would be very helpful for the authors to clarify precisely how the "shape uncertainty" is defined. For instance: What is the actual degree of freedom associated with this uncertainty? What is the uncertainty for each knob?

At present, the description of these uncertainties is more consistent with a rate-only treatment and does not fully specify the additional degrees of freedom relevant for a shape analysis. Providing a summary table, perhaps in the Supplemental Material, detailing the rate and shape uncertainties, their parameterizations, and their implementation in the fit would greatly improve clarity.

4. JUNO's first simultaneous observation of atmospheric and solar neutrino oscillations represents a significant milestone, demonstrating the detector's exceptional sensitivity to multiple oscillation modes. Although a Δm^2_{31} result is not yet being presented, we encourage the authors to provide some discussion of the projected precision on this parameter, beyond simply stating that adding the Daya Bay constraint has no impact on the fit. Since the precision measurement of Δm^2_{31} and compare it with the measurement from the numu disappearance provides another way to determine MO, including this information can help establishing the case towards JUNO's readiness in MO determination.

Additional comments for improvements:
1. In addition to the total uncertainties, it would be useful to provide estimation of statistical and/or systematic uncertainties for the $\sin 2\theta_{12}$ and Δm^2_{12}

2. In Figure 2a, The E_rec appears to show a dip around 13m for AmC source and v_e bar candidate, In addition, the AmC source seems to have an overall lower E_rec than the others. And the muon spallation source continues to decrease beyond 13m.
i) Some discussions regarding what causes all sources going down at 13m. Any insights on the different behaviours among different sources?
ii) Some discussions regarding what causes the overall lower reconstruction of AmC than others can be useful?

3. Page12: "Other selection efficiencies including the muon veto, spallation neutron veto, and multiplicity cut—were derived via live-time analysis using 10-cm cubes", For clarity, it would help to briefly explain what is meant by "10-cm cubes" for a general reader. For example, the author could clarify that the full fiducial volume (FV) is divided into small cubes of size of (10cm)^3 for the efficiency study then somehow average over FV.

4. It was stated that the geometric volume fraction is determined to be 80.6%, and this has been validated by the IBD above 3.5 MeV. What is the precision for the validation? Is it sufficient to validate at 1.6% level? The same question applies to the B-12 validation.
What is the energy dependence on the uncertainty for the fiducial volume?

5. In the “JUNO detector setup” section, it is mentioned that “a very small geometrical inefficiency due to the top chimney …”. It would be useful to be quantitative on this inefficiency.

6. In the “Detector response model” section, it is stated that “systematic effects including energy leakage from source encapsulation and peak stability” were also evaluated. It would be useful to be quantitative on these.

7. In method, Figure 6(a). the terms “non-linearity” and “effective gamma energy” are not clearly defined. Presumably “non-linearity” refers to the ratio between reconstructed energy and true energy, and “effective gamma energy” involves some treatment for multiple-gamma sources, but they should be defined explicitly.

8. In the same section, there is a statement “... show that including or excluding instrumental non-linearity for centrally located gamma …”. What about the sources at off-center locations? Some clarification can be useful.

9. In Figure 7, the binning appears to be coarse. It would be helpful if the author could clarify the binning strategy, and provide more details in the bin near the edge of the FV. In the next bin after FV cut, results from 3 strategies start to separate indicate some inconsistency of the understanding of the event reconstruction. Does this inconsistency happen earlier inside the FV? With such a coarse binning, it is difficult to evaluate this point.

10. In Figure 8, it would be useful to show the Delta_d distribution beyond the 1.5 m.

11. In Section “Backgrounds”, it is stated that the Li9/He8 ratio is taken from Daya Bay. Given the overburden is different between Daya Bay and JUNO, this ratio may be different, since it would depend on the muon deposited energy. Additional discussion on the uncertainty would be useful.

12. In page 23, The expression “We therefore performed a combined fit across four muon energy intervals, accounting for cross-interval correlations where 9Li/8He events from one interval mimic antineutrino signals in others.” is confusing. Figure 9 shows five muon-energy bins, and it is not clear how these relate to the “four intervals” mentioned in the text. In addition, the phrase “^{9}Li/^{8}He events from one interval mimic antineutrino signals in others” needs further clarification. It would be helpful if the author could explain the mechanism of how this happened and how this correlation is modeled in combined fit.

13. Fig.9, the last bin does not have a label, please add.

14. In Equation 4, what is the value of “N”?

Referee #5

(Remarks to the Author)
Firstly, the JUNO Collaboration is to be congratulated for designing, building, and constructing a liquid scintillator detector of unprecedented size and resolution. This paper is the first physics result and therefore an important milestone in the field of neutrino physics. Comments follow:

Main Narrative:
1. Some introduction to the NMO issue in the introduction would put the paper more in context for readers who are not particle physicists and who may not read the detailed “Methods” section.

2. Ref [38] is refered to in several places but has no information on where it is published

3. In Fig, 2(a) is it understood why there a systematic shift in reconstructed energy between the three sources of neutrons? Even though this is taken as a systematic uncertainty this should be noted and discussed.

4. What is the actual muon rate through the detector? I didn't see this number given.

5. The flasher problem made a quality cut necessary. How many flashers were there in this data set? What cuts were made exactly? This would be appropriate for the Methods section.

6. Values of sigma_rel in Table 1 imply units are %, but better to state explicitly for clarity

7. In Table 1 refers to “Multiplicity” but this is not discussed until the section on Methods. There should be a short definition of this factor in the main narrative as there is with other issues

8.A bit more information on the difference between “best-fit” and “pre-fit” (which I take to mean best estimate before spectral fitting) would be useful to the reader not familiar with these terms.

9.Page 14, when discussing external constraints on Delta_m^2_31 and sin^2_theta_13 please state explicitly the values used and their uncertainty. Did both these values come from Ref [46]? If so, the authors should consider using a global best fit values for these instead from a single experiment.

10. Figure 4 is very important and for clarity the caption should state that all error bars from other experiments are 1 sigma.

Methods:
1.If possible add information on the PMT manufacturer, model, and any publications that describe performance parameters

2. “Instead, SPMTs have…” would be better as “Conversely, SPMTs have…”

3. Again, Ref. [38] does not have any information on where it is published

4. Pg 18 in Calibration Strategy, A figure should be included that shows all the locations of the AmC souce in order to show how well the CD volume was covered.

5. Detection efficiency is a function of the photon location on a PMT. What uncertainty was assigned to cover these geometrical effects if there were only central measurements used for the relative efficiency? Were there measurements off-center to validate the central measurement and associated uncertainty?

6. What fraction of channels were excluded from the analysis in the LPMTs due to low occupancy or timing misalignment?

7. It would be informative to see a plot of the vertex differences event-by-event for the 3 methods used. Were these differences used to put a lower limit on the vertex reconstruction resolution?

8. Detector response model: how does the effective light yield change of off-center? A plot of this would be very informative. Also, if 1785 p.e./MeV (4 significant figures) is used, it should not be described as “approximate”?

9. The beta decay shape for 12-B decay to 12-C is an important validation for the energy non-linearity model. Please give information on how the theoretical shape was calculated, including how the electron screening was handled for the lower energy electrons. What model(s) were used and what uncertainly was assigned to this screening on the theoretical shape? In other words, I gather that Birk’s Law formula was used and fit to the data based on departure from the theoretical curve, but to what uncertainty is the theoretical curve known at lower energies where screening is important?

10. Same for 11-C screening, but in addition the 11-C decay is not pure G-T like the 12-B decay is. This could (or could not) have implications for the shape and should be considered.

11. In addition, the 11-C decay positron spectrum only goes to 0.96 MeV while the positron curve in Figure 6(d) goes to 10 MeV. Does this imply that the positron data below 0.96 MeV plus the expected contribution from annihilation gammas was used to fit Birk’s Law and then extrapolated to 10 MeV? If so, it would seem that the positron curve would how more uncertainty at higher energy, but it would appear from the figure to have less uncertainty than the electrons as energy increases. More information on this would make a stronger case. They authors should also consider making Fig. 6(d) logarithmic on the y-axis to make it more visible.

12. Pg 21 says an alternative analysis used 17.m meters for a FV on delayed events. Was a FV cut used in the main analysis? Since one is not mentioned I would infer that such a cut was not used, but best to state this explicitly.

13. Pg 22, why was a FV efficiency of 1.6% used based on comparing different fitters rather than the 1.8% from source deployments near the edge of the FV? Different fitters can have correlated systematics which always result in underestimated systematic uncertainties, whereas the source deployment does not suffer from this issue.

14.In the Methods section on 9-Li/8-He it would be appropriate to show exactly how the 34% uncertainty waws assigned to the rate. It is the largest background but after nearly a whole page of discussion it is still not clear to me exactly where this value for this uncertainty comes from.

15.In Fig 10 it is not clear to me whether the red line prediction includes the off times mention or does not. If it does, then there is no need to dwell on the details of individual reactors turning on and off as it implies it is not already covered.

R.Svoboda

Referee #6

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
Dear Editor,

Thank you for forwarding the authors' rebuttal and the revised manuscript.

Let me state very clearly at the outset that the JUNO effort is exceptional. This experiment is technically extremely demanding, and the collaboration's capability in delivering such a dataset and analysis is, in my view, fundamental for the future of neutrino physics. The result reported here is important and timely.

From the perspective of a neutrino physicist, the paper was technically sound and publishable already in its original form, and it is even more clearly so in the revised version, following the authors' efforts to address the comments from all referees. I have no scientific objections, and I do not identify any issues that would prevent publication.

My initial suggestions focused on helping the manuscript speak to a broader Nature readership—i.e., improving accessibility, narrative focus, and contextual framing, in line with the very detailed guidance Nature provides to referees. If, however, the editorial view from the start was that the result's importance and urgency justified publication essentially in its original structure and level of specialization, it would have been perfectly reasonable to state that explicitly at the outset, rather than orienting the review process around particularly strict presentation-style expectations.

At this stage, I do not think it is useful to add further suggestions. The experimentalists in JUNO undoubtedly have more important things to do than iterating on stylistic reshaping, and I assume the same is true for the editorial office. I would be happy for you to proceed with publication in the form you consider most appropriate.

With best regards

Referee #2

(Remarks to the Author)
We thank the authors for their detailed answers to our comments and questions. The authors' answers were satisfactory and the revised text greatly improved our understanding.

We would like to recommend that in specific cases, the answers to our questions be included in the text as we think those responses would strengthen the paper. In particular:

The distinction between using the data from Daya Bay ND for the flux constraint and primarily the Daya Bay FD (or ratio of FD/ND) for the oscillation parameters is an important clarification and should be included in the narrative.

The facts that the Daya Bay information on dm2_13 and sin2theta_13 is included as 2, 1d constraints and the correlation between the two is small is worth mentioning in the methods when details of the fit are described. The main text does mention the 1d constraints, but doesn't motivate why that is a valid implementation.

We better understand the limitations on extending Fig. 11. However, the pulls currently shown could appear cherry-picked to demonstrate consistency. It would be useful to clarify that these correspond to the largest pulls from the nominal values that are common across the three analyses.

Finally, we have one small suggestion to help with lingering confusion over the numbers in table 1. The authors added a sentence to the caption of table 1 in response to one of our comments, but that extra sentence added confusion. We would suggest dropping that sentence in the caption and add to description of geoneutrinos and world reactor nu backgrounds on page 14 that the numbers quoted in the text must be further scaled down to account for the selection efficiency. Fundamentally, the reader wants to know how many of the 2379 selected events are expected to be background. Perhaps provide that number explicitly in the table for the reader and don't make them figure it out for themselves.

Less critically, there are a few typos:

page 12, right column, 0th paragraph, 7th line there is an extra "." after the parenthetical

p17 the comma at the end of eq. 1 should be a full stop.

p18, left, first full paragraph, 2nd line, a comma is needed after WP

Otherwise, we are happy to recommend the paper for publication.

Referee #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #4

(Remarks to the Author)
We want to thank the authors for their responses and updated manuscript.

These are regarding the responses provided by the authors
Regarding "These results advance the basic understanding of neutrinos, validate the detector's design, and confirm JUNO's readiness for its primary goal of resolving the neutrino mass ordering with a larger dataset." in the abstract.
The underlying facts are clear, but the the current wording is not appropriate, and should be improved for clarity
Suggest "…and demonstrate encouraging progress toward JUNO's primary goal of resolving the neutrino mass ordering with a larger dataset." or "…and validate key aspects of the detector performance, supporting JUNO's path toward its primary goal of resolving the neutrino mass ordering."
Regarding the "B-12" validation of the geometric volume fraction, the newly added clarification is good. It is better to add that the validation is done above 4 MeV energy.
Thanks for clarifying the meaning of "N" in modeling the reactor neutrino spectrum. The actual value of "N" used in the analysis should be provided.

Some additional comments regarding the revised manuscripts:
CP-violating phase → CP phase. Note this phase can lead to CP violation at certain values. Because of this, it is better to call it CP phase.
In the text above Fig. 4, "... parameters is statistically dominated …" → "statistically limited". Given the current breakdown of statistical and systematic uncertainties, the data do not support the statement that the precision is statistically dominated. It would be more accurate to describe the measurement as statistically limited.
The newly added text in the "Discussion and outlook" regarding the 1.5 sigma difference between the reactor and solar neutrino dm^2_21 is an overmotivation. Suggest to revise the text as "Although the current difference between our reactor-based \Delta m^2_{21} result and the solar neutrino value [50] is only at the \sim1.5\sigma level, it motivates continued independent probes of this parameter."

Referee #5

(Remarks to the Author)
I am satisfied with the responses to my review and have no further comments

Referee #6

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Dear Editor and Referees,

We would like to thank the editor and the reviewers for their valuable time and professional feedback on our manuscript. Their insightful comments have helped us improve the quality of our work. We have revised the manuscript accordingly and resubmitted it for your consideration. Below is our point-by-point response (in blue color) to the reviewers' reports.

Referees' comments:

Referee #1 (Remarks to the Author):
Title:
First measurement of reactor neutrino oscillations at JUNO
This manuscript presents the first physics results from the JUNO experiment: a precision measurement of the solar neutrino oscillation parameters using reactor antineutrinos collected during the commissioning phase. The analysis delivers significantly improved determinations of Δm^2_{21} and $\sin^2\theta_{12}$ compared to current world averages, while simultaneously demonstrating that a 20-kton liquid-scintillator detector can achieve the targeted few-percent-level energy resolution. These results not only validate the detector performance and analysis methodology but also establish JUNO as a key player in the emerging precision era of neutrino oscillation physics, with direct implications for tests of the three-flavor paradigm, global oscillation fits, and future determinations of the neutrino mass ordering.

We thank our referee for very positive comments on the physical significance of the first JUNO results.

General Assessment
The scientific quality and technical execution of the work are outstanding. The successful commissioning of a 20-kton liquid-scintillator detector with such high energy resolution is a major milestone in experimental neutrino physics. The data quality appears robust, and the analysis is carefully designed and thoroughly cross-checked.
At the same time, for a broad audience journal like Nature, the manuscript currently leans heavily on demonstrating "technical success" (calibration, stability, background rejection), while the "scientific necessity" and broader impact of the new precision regime are somewhat underplayed. The paper would, in my view, be substantially strengthened by streamlining technical details in the main text and expanding the discussion of the physics reach and conceptual implications. Addressing the comments below—particularly by sharpening the narrative around the broader physics goals—would bring the manuscript much closer to meeting the standards of Nature. Although the manuscript is technically correct, the language is often too specialized for Nature's broad readership (including physicists who are not neutrino-oscillation experts). I recommend simplifying the main-text narrative by reducing jargon and acronyms, adding short qualitative context before technical details, and moving highly implementation-specific discussion to Methods/Extended Data. This would improve

readability without compromising rigor.

Major Comments

The authors present their results primarily as numerical improvements over previous measurements. While the reduction in uncertainty is indeed impressive (~1.6% for Δm^2_{21} and ~2.8% for $\sin^2\theta_{12}$), the manuscript would benefit from a more explicit discussion of why moving towards sub-percent precision is transformative for the field. To make the broader physics motivation more apparent, the authors could more explicitly articulate JUNO's long-term physics goals. For example, the Introduction and Discussion might briefly organize the narrative around themes such as (the list below is not meant to be exhaustive):

1) From discovery to precision tests of the three-flavor framework.

The present results already show that JUNO is operating at a level that will soon enable sub-percent precision on multiple oscillation parameters. This is more than simply "a better measurement": it marks a transition of neutrino oscillation studies from a discovery phase to a genuine precision-test regime, where non-trivial internal consistency checks of the three-flavor paradigm across different channels and baselines become possible. Stating this explicitly would help non-specialist readers appreciate the conceptual importance of achieving such precision.

2) Role in global fits and in breaking parameter degeneracies.

Although this analysis focuses on the solar parameters, the demonstrated detector performance implies that JUNO will soon measure the effective atmospheric mass splitting Δm^2_{ee} with high precision. It would be helpful to spell out how having Δm^2_{21}, $\sin^2\theta_{12}$ and Δm^2_{ee} all measured precisely within a single experiment impacts global oscillation fits. In particular, fixing these parameters independently reduces degeneracies in long-baseline accelerator analyses and indirectly facilitates a more robust determination of the θ_{23} octant and the CP-violating phase δCP, even though these quantities are not directly accessible via reactor antineutrino disappearance.

3) Connection to the mass ordering and to broader questions in fundamental physics.

While the present data set does not yet determine the neutrino mass ordering, the results validate JUNO's capability to do so in future runs. A brief but explicit connection between a future normal/inverted ordering determination and broader questions—such as the interpretation of neutrinoless double-beta decay searches (and the Majorana nature of the neutrino) and cosmological constraints on the sum of neutrino masses—would help convey to the Nature readership why this program is of wide and lasting interest.

4) Clarify oscillation regimes and the origin of mass-ordering sensitivity (p.10)

On p.10 (first column), the discussion of the "first solar oscillation maximum" is potentially confusing because it mixes (i) the slow spectral distortion driven by Δm^2_{21} and $\sin^2\theta_{12}$ with (ii) the fast atmospheric-scale oscillations governed by $\Delta m^2 3_1$ and $\Delta m^2 3_2$, and (iii) the mechanism that provides mass-ordering sensitivity, which comes from the energy-dependent interference between the two close atmospheric frequencies $\Delta m^2 3_1$ vs $\Delta m^2 3_2$), not from the solar maximum itself. I strongly suggest rewriting this paragraph for clarity. A simple figure showing Pee(E) at L ~ 52.5 km (solar-only vs full; and optionally NO vs IO, or NO–IO difference) would immediately remove ambiguity; this could be placed in Extended Data.

5) Solar–reactor consistency as a test for new physics.
A key virtue of high-precision reactor measurements of Δm^2_{21} and $\sin^2\theta_{12}$ is the direct comparison with solar neutrino results, which probe a CPT-conjugate channel and are subject to matter effects in the Sun. Any statistically significant tension between solar and reactor determinations could point to physics beyond the standard three-flavor picture, for example non-standard neutrino interactions or other flavor-dependent matter effects. JUNO is uniquely positioned to perform such consistency tests; the present measurement can be presented as the first step in this program.

6) Consistency of the PMNS framework and unitarity tests.
By precisely pinning down $|U_{e1}|$ and $|U_{e2}|$ through $\sin^2\theta_{12}$ and Δm^2_{21} (and eventually Δm^2_{ee}), JUNO provides crucial input for global tests of the unitarity and internal consistency of the PMNS framework when combined with accelerator and atmospheric data. If the authors wish, a short remark in this direction could further emphasize that JUNO contributes not only isolated measurements but also stringent cross-checks of the overall three-flavor scheme.

7) Refining reactor antineutrino flux models.
Finally, the high-statistics, high-resolution reactor spectrum at JUNO—especially in combination with the TAO near detector—will allow stringent tests of reactor antineutrino flux predictions and may shed light on existing spectral anomalies. Even if detailed flux studies lie beyond the scope of the present analysis, a brief mention of this long-term capability would broaden the perceived impact of the experiment.
These examples are intended as possible directions rather than a fixed list; the authors may well have additional physics goals or synergies in mind that could be highlighted to further strengthen the overarching narrative.

We greatly appreciate our referee for providing a comprehensive summary of the significance of the JUNO results and for valuable suggestions on how to improve the manuscript. In light of the broad interest expected for our first result, we agree that expanding the introduction and conclusions to better emphasize the impact of our findings and the future direction of JUNO would be beneficial. In the revised version of our manuscript, we have added in the second paragraph of the section 'Discussion and outlook' several sentences accordingly to stress the impacts of JUNO results on the future neutrino oscillation program, the neutrinoless double-beta decays and the sum of neutrino masses for cosmological observations.

Technical detail and narrative flow
The main text currently devotes substantial space to calibration procedures, energy-scale characterization and background rejection (for example, in the discussion around Figure 2). These technical elements are essential for establishing the robustness of the result, but in their present form they somewhat interrupt the narrative flow for a non-specialist reader.
I suggest that the authors consider moving the most detailed performance plots (such

as the energy-scale residuals and certain background-rejection efficiency plots) to the Extended Data or Methods sections and keeping in the main text only a concise, qualitative summary of the calibration strategy and performance, with clear references to the corresponding Extended Data figures. This re-organization would allow the oscillation spectrum and fit (currently Figure 3) to take center stage in the main text and would help refocus the narrative on the physics outcome and its implications. The full technical validation would remain available in the appropriate sections for expert readers who wish to examine the analysis in detail.

While detailed technical material is appropriately placed in the Methods, there are a few conceptual elements currently discussed only there that, in my view, deserve a brief summary in the main text. A short, equation-free treatment would make the physics more transparent for non-specialist readers without compromising the concision of the paper. Concretely, the authors might consider the following:

Qualitative origin of the oscillation signal.

The Methods provide a clear explanation of how the electron antineutrino survival probability is dominated, in this configuration, by the “solar” term driven by Δm^2_{21} and θ_{12}, with a subdominant contribution from the atmospheric term. It would help readers if this were summarized in the main text near the discussion of Fig. 3. For example, the authors could add a short paragraph such as:

“The observed distortion of the reactor antineutrino spectrum is governed predominantly by the ‘solar’ oscillation term driven by Δm^2_{21} and θ_{12}, which introduces a slow, baseline-scale modulation of the event rate. A subdominant contribution arises from the ‘atmospheric’ oscillation term, which produces a finer interference pattern that is only weakly relevant for the present analysis (without a brief clarification, non-expert readers may not understand the historical origin of the “solar” and “atmospheric” terms). In this work, the fit is therefore primarily sensitive to the solar parameters, while remaining largely insensitive to the atmospheric mass splitting.”

Semi–model-independent treatment of the reactor flux.

The semi–model-independent approach to reactor flux modelling, anchored by Daya Bay near-detector data, is an important ingredient that currently appears only in the Methods. Given the long-standing interest in reactor flux systematics, it would be worth summarizing this strategy in one or two sentences in the main text, together with the statement that the residual model dependence is small. For instance:

“To minimize the dependence on reactor antineutrino flux models, we adopt a semi–model-independent strategy in which the JUNO data are fitted using a segmented spectral function that is shared with the Daya Bay near detectors. In this approach, Daya Bay data effectively anchor the overall normalization and coarse spectral shape, while theoretical calculations provide the fine structure and its uncertainty. After optimization, the residual model dependence of the extracted Δm^2_{21} and $\sin^2\theta_{12}$ is below the 0.3% level, well below the total experimental uncertainty.”

Relevance of the current energy resolution for this measurement.

The Methods usefully clarify that the achieved positron energy resolution, while slightly worse than the nominal design value, is not critical for the present analysis of the solar parameters. Since energy resolution is a central performance metric for JUNO, it may help to make this point explicit in the main text to avoid any confusion about the impact on the current result versus future NMO sensitivity. For example:

"The achieved positron energy resolution in this commissioning data set, σ_E/E ≈ … at 1 MeV, is slightly worse than the nominal design value. However, for the determination of the solar oscillation parameters reported here, the result is driven mainly by the overall spectral distortion over a wide energy range and is only weakly sensitive to small changes in resolution. The present measurement is therefore not limited by the current resolution, whereas the full design performance will be more critical for future determinations of the neutrino mass ordering."

High-level summary of background constraints.

The Methods contain a very complete treatment of how the dominant backgrounds (cosmogenic $^{9}Li/^{8}He$, geoneutrinos, antineutrinos from other reactors, etc.) are constrained using control samples, external information and simulations. In the main text, it would be helpful to summarize this logic in a single sentence, emphasizing that backgrounds are both subdominant and independently constrained, and thus unlikely to mimic the oscillatory signal. For instance:

"All major backgrounds, including cosmogenic $^{9}Li/^{8}He$, geoneutrinos and antineutrinos from other reactors, are constrained by a combination of data-driven control samples, external measurements and simulations, and remain small compared to the oscillated reactor signal. Within the fitted energy window, these independently constrained background components cannot reproduce the characteristic spectral distortion induced by oscillations, and their uncertainties are subdominant in the error budget on Δm^2_{21} and $\sin^2\theta_{12}$."

Multiple independent analysis chains.

The existence of three independent analysis implementations, with different treatments of the covariance matrix and systematics, is a strong robustness check that is currently described primarily in the Methods. Since this is the first major physics result from JUNO, it would be worthwhile to highlight this more clearly in the main text. A possible formulation could be for instance:

"The analysis framework has been implemented in three independent analysis chains that employ different treatments of the covariance matrix and systematic uncertainties. All three analyses yield mutually consistent best-fit values and uncertainties for Δm^2_{21} and $\sin^2\theta_{12}$; the results quoted in this work correspond to one representative implementation, with the spread among the three providing an additional robustness check."

These additions need only be at the level of one short paragraph or sentence per item; the detailed derivations and numerical inputs can remain in the Methods, where they are entirely appropriate.

Suggestions for simplifying Figure 3

I have a few additional, more concrete suggestions aimed at simplifying Fig. 3 and making the main physics message easier to grasp:

Because the non-oscillated prediction is much higher than the oscillated spectrum, the vertical scale of the main panel is set by this large reference curve. As a result, the

individual background components are effectively unreadable in this plot. I would therefore recommend combining all subdominant backgrounds into a single "total background" contribution in Fig. 3 and showing the detailed background breakdown in a separate Extended Data/Methods figure. This would both improve legibility and allow the legend in the main panel to be simplified.
The colored band used to represent the signal makes the main panel visually heavier and slightly harder to read. The authors may wish to consider whether this shading is really necessary, or whether a simpler representation would serve the physics message better.
A thinner continuous line for the non-oscillated prediction, together with slightly reduced line thickness for the background component, could help to place visual emphasis on the best-fit oscillated spectrum. Making the best-fit curve the most visually prominent object in the plot would guide the reader's eye directly to the oscillation effect.
Finally, since the y-axis is labelled "Events / 0.1 MeV" while the bin widths are not uniform (as is apparent from the figure), it might be helpful to add a short remark in the Supplementary/Methods material explaining how to interpret this quantity in the presence of varying bin sizes. A brief clarification would avoid possible confusion for readers who are less familiar with such conventions.
Taken together, these changes would make Fig. 3 cleaner and more immediately interpretable, while all detailed information about backgrounds and binning could still be documented in the Extended Data and Methods.

Comments on Figure 4
Figure 4 is a central physics summary and plays an important role in conveying the impact of the measurement. The content is appropriate, but a few aspects of the presentation could be clarified.
The figure combines 2D confidence regions in the ($\sin^2\theta_{12}$, Δm^2_{21}) plane with 1D confidence intervals shown in the marginal $\Delta\chi^2$ projections. This is reasonable, but the different numbers of degrees of freedom are not fully transparent to a non-expert reader. I see three possible ways to make this clearer:
- explicitly state in the caption that the contours in the central plane correspond to 2D confidence regions (two degrees of freedom, with the appropriate $\Delta\chi^2$ thresholds for 1σ, 2σ, 3σ), while the 1D intervals in the marginal plots use $\Delta\chi^2$ = 1, 4, 9 for one degree of freedom;
- alternatively, keep all genuinely 2D information in the central panel and all purely 1D intervals in the marginal projections, so that each panel uses a single, well-defined convention;
- as a third option, the authors could consider using a one–degree-of-freedom convention for JUNO everywhere ($\Delta\chi^2$ = 1, 4, 9 both in the plane and in the projections), as is sometimes done when profiling; in that case, it would be important to state explicitly that the plotted "1σ, 2σ, 3σ" regions are defined using the 1 d.o.f. $\Delta\chi^2$ thresholds.
In the central panel, the shaded JUNO confidence region is visually very prominent. This makes JUNO's result clearly visible, which is good, but it also makes the plot

somewhat heavy. The authors might consider slightly reducing the opacity of the shading or restricting the filled area to the innermost (e.g. 1σ) region and indicating outer confidence levels with contours only. Coloring only the 1σ region would also be more coherent with the way the 1D projections emphasize the 1σ intervals and would give a more uniform visual convention across the different panels.

Since several external measurements are shown, it is important that all curves remain distinguishable and that JUNO's result is still the first thing the eye is drawn to. From a graphical point of view, the authors might consider giving JUNO a clearly visible best-fit marker and slightly thicker contours, and using distinct line styles (solid, dashed, dotted) together with a color-blind–friendly palette for the external constraints. This is not a physics convention, but a standard accessibility/readability recommendation and would make the figure easier to interpret.

Finally, the caption could be slightly more explicit about which contours and shaded regions correspond to JUNO, which line styles/colors correspond to each external experiment, and which confidence-level conventions (number of degrees of freedom and $\Delta\chi^2$ thresholds) are used in the central panel and in the marginal plots. A brief clarification there would remove any ambiguity about how to interpret the contours and intervals.

Comments on the Methods section

The Methods section is comprehensive and technically solid, and it contains all the ingredients that one would expect for a first precision result from JUNO (oscillation formalism, detector and calibration, reconstruction and selection, reactor-flux modelling, background treatment, and statistical analysis). In my view, the main improvements would be in tightening the exposition and making the structure easier to navigate, rather than in adding more technical material.

I have a few concrete suggestions in this direction. It would help if the Methods opened with a very short "roadmap" paragraph that explicitly states how the section is organized (oscillation formalism; detector and calibration; reconstruction and event selection; reactor flux and backgrounds; statistical analysis). Clear subheadings following this logic would make it easier for readers to locate specific parts of the analysis and see how they fit together.

The description of the calibration systems and waveform reconstruction is very detailed, which is valuable for experts. Some of the most technical implementation details (for example, specific algorithmic choices or numerical settings that are not directly propagated into the oscillation fit) could, however, be moved to Supplementary/Extended Data. In the main Methods text, it would be sufficient to emphasize: (i) which calibration inputs determine the energy scale, non-linearity and resolution models used in the fit, and (ii) how the corresponding uncertainties are propagated into the covariance matrix.

For the backgrounds, a short introductory paragraph summarizing how each major component is constrained (control samples, external inputs, simulations) and how its uncertainty enters the fit would improve readability, with the detailed cross-checks left to sub-subsections or Supplementary material.

The statistical framework based on the χ^2, the construction of the covariance matrix,

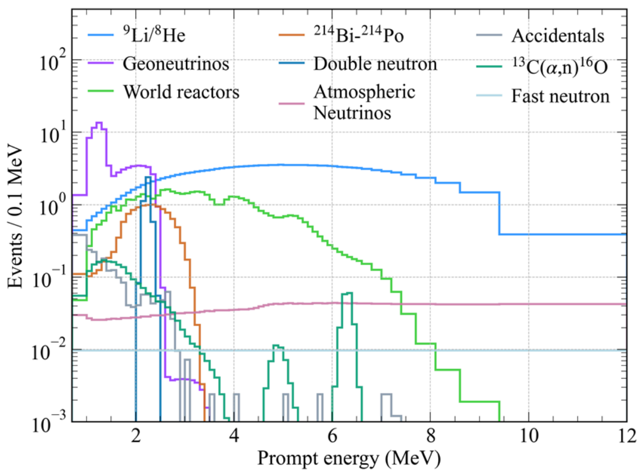
and the treatment of nuisance parameters are central to the credibility of the result. It would be useful to state in one place which classes of systematics are encoded as nuisance parameters with priors, which are represented through covariance matrices, and which are constrained externally. A brief remark on which parameters are profiled in the fit and which are constrained by Gaussian priors would make the methodology more transparent. This would also be a natural place to mention that three independent analysis chains, with different treatments of the covariance matrix and systematics, give mutually consistent results, referring to the corresponding figure in the Methods.
To summarize, the Methods section is comprehensive and technically detailed, which is fully appropriate for a first flagship physics result from JUNO. However, the density of the presentation can make the logical progression from detector description to reconstructed spectrum and finally to the oscillation fit somewhat difficult to follow, especially for readers who are not experts in large liquid-scintillator experiments. The readability could be improved, without removing technical content, by slightly strengthening the structural guidance within this section. In particular, it would be helpful to open the Methods with a short roadmap paragraph that explicitly states how the material is organized (oscillation formalism; detector and calibration; reconstruction and event selection; reactor flux and backgrounds; statistical analysis). In addition, each major subsection could begin with a brief explanatory summary (one or two sentences) that states its purpose, identifies the key ingredients, and explains how its outputs enter the final oscillation analysis. Such introductory statements would assist the reader in navigating the detailed material and make the overall structure and logic of the Methods more transparent.

Comments on the figures
Figure 5 is conceptually valuable: it provides an intuitive, equation-light visualization of the slow ("solar") versus fast ("atmospheric") oscillation components and explains why JUNO's ~52.5 km baseline is special compared to Daya Bay and KamLAND. This is exactly the kind of orienting plot that helps a broad readership understand what sets JUNO apart. From a presentation point of view, the figure is rather small and contains several visual elements at once. As a result, the large colored circles and the experiment labels draw attention away from the red/blue curves, which carry the main physics message. In addition, the logarithmic baseline axis may not be immediately intuitive for non-specialist readers. The plot would be clearer if the "detector-size" circles were made less prominent (e.g. smaller or more transparent) and if the relevant experimental baselines were indicated more directly (for example with simple vertical markers). If the circle-size encoding is kept, it would be helpful to print the baseline value next to each experiment name and to label the two curves directly on the plot (solar-only vs full P_{ee}), so that the reader does not need to rely on the caption to identify them.

Figure 7 is a useful validation plot because it simultaneously supports two key claims: (i) the selected IBD candidates are spatially uniform within the fiducial volume, and (ii) the result is stable across multiple, independent vertex reconstruction algorithms. This directly underpins the robustness of the fiducial-volume definition and the control of edge-related backgrounds. In terms of clarity, the choice of plotting in cubic radius (R^3) is defensible for uniform-density arguments, but it is not immediately intuitive to non-expert readers. A simple improvement would be to add a secondary x-axis in R (meters). The current color choices in Fig. 7 are not fully color-blind friendly; using a color-blind–safe palette (and/or adding distinct line styles or markers) would improve accessibility and ensure all curves remain clearly distinguishable. A similar issue applies to the point colors in Figure 11.

Finally, Given the flagship nature of this first JUNO precision result, it would be helpful if the Data Availability / Code Availability statements clarified which tabulated data products and key analysis inputs will be made publicly available at publication (or provided upon reasonable request), at least at the level needed to reproduce the published fit using the binned likelihood/covariance model.

We are grateful to our referee for careful reading of the manuscript and for all the pertinent comments and valuable suggestions on the context and figures. In the revised version, we have updated the caption of Figure 4 to specify the $\Delta\chi^2$ thresholds corresponding to the 1σ, 2σ, and 3σ 2D confidence regions.



Regarding Figure 3, we have prepared a separate plot showing the detailed background breakdown (see above). However, since the dominant backgrounds are 9Li/8He, geoneutrinos, and world reactors, we believe including this plot in the Methods section would add limited additional insight, thus we just provide it for the referee's reference.

As for Figures 5 and 7, we have considered the referee’s suggestion but, after careful evaluation, have decided to retain the current plot style.

We greatly appreciate the referee’s valuable suggestions on rephrasing the text. We have carefully reviewed them and considered their intent, and believe that the current wording effectively conveys the intended scientific meaning. Therefore, in order to maintain consistency and clarity in the current narrative, we have opted to retain the existing text.

Referee #2 (Remarks to the Author):

The manuscript reports first results from the JUNO reactor neutrino experiment. We commend the authors for delivering a quick and impactful first result from JUNO. After a (mere) 60 day exposure, the experiment provides world-leading measurements of the solar mixing parameters, sin^2(theta_12) and dm^2(21). The paper describes the experiment and the data analysis, which foreshadows the experiment's ability to achieve its long-term goal of resolving the neutrino mass ordering. This result has been eagerly awaited by neutrino particle physicists, but is of equal interest to a more general population as the first of the next-generation, precision neutrino oscillation experiments that will take the stage over the next 5-10 years. These results carry substantial significance, demonstrating the potential of JUNO to advance our understanding of fundamental particle physics. We believe the manuscript worthy of publication in Nature, contingent upon satisfactory responses to a few issues we will raise.

Major comments:

1. Introduction
1. It would be useful to the non-expert reader to explain how the different oscillation parameters shape the spectrum. Such an explanation early on would help explain why the energy resolution is important to determine the mass ordering (first full paragraph on page 11).

We have added one sentence in the 3rd paragraph on page 10:

"Determining the neutrino mass ordering hinges on resolving a fine-scale oscillation ripple (driven by Δm^2_{31}) superimposed on a slow, large-amplitude wave (driven by Δm^2_{21}), with the ordering signature encoded in the phase of this fine-scale structure."

2. Calibration & Reconstruction:
1. Fig. 2a shows the spallation neutron sample response differs from the AmC source by ~1% at small radii, while neutron captures from nuebar show a similar difference at larger radii. However, in Fig. 2b, both show a ~0.5% difference, which we assume is averaged over radius. Would it be more appropriate to assign a conservative systematic uncertainty at the 1% level to cover the full radial range observed across these three neutron sources? If so, how would such an increase propagate to the final measurement?

The assigned 0.5% energy scale uncertainty corresponds to the volume-weighted average across the entire fiducial volume defined for IBD candidate selection. It is not intended to only account for local variations at specific radii. The error bars in Fig 2b. incorporate both spatial non-uniformity and temporal variation observed across the entire detector and dataset. Meanwhile, the average energy scale for the three bulk neutron samples is also well controlled at the 0.5% level.

Conversely, the observed energy scale difference between uniformly distributed sources, such as spallation neutrons, IBD neutrons and the AmC calibration source, is primarily attributed to the finite theta and phi coverage of calibration systems.

To conclude, we consider that assigning 0.5% uncertainty is consistent with observations in the current dataset, while 1% uncertainty would be too conservative.

2. While it is reasonable to describe the reconstruction algorithms in the Methods section, it would be helpful to summarize the achieved performance in the main text. Specifically, please quantify the values (or ranges) for the energy resolution, vertex resolution, and timing resolution.

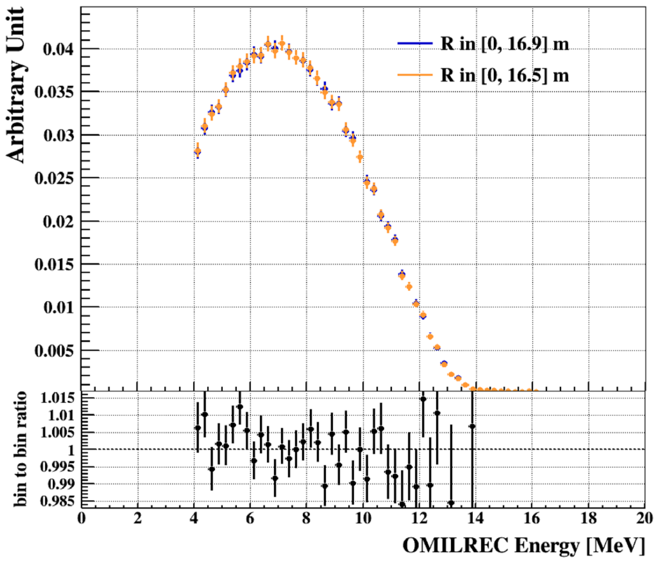
Given the inherent complexity of reconstruction performance (which depends on particle energy, position, type, and algorithm) and the limited calibration data available in this first result (only at dozens of specific points along the z-axis and at a few off-axis positions), a simple numerical summary is inadequate. To address this, we conducted a comprehensive evaluation of multiple algorithms using not only calibration sources but also natural radioactivity, muon-induced spallation neutrons, and cosmogenic isotopes. Full details are provided in Ref.[38] ("Initial performance results of the JUNO detector"), and it is cited when needed in the Main text and Methods sections.

Upon the suggestion, we have added following description into the Main text: "for ^{68}Ge deployed along the z-axis with |z|<16.5m, the vertex bias is within [-13, +10]cm, meanwhile the vertex resolution ranges from 13 cm to 27 cm."

3. Figure 7 in the Methods demonstrates a uniform distribution of events within the fiducial volume and indicates no significant radial dependence in the reconstructed event rate, except near the detector boundaries, where a relative uncertainty of 1.6% is assigned based on the observed deviation. Do the reconstruction algorithms perform consistently with respect to the spectral shape, particularly as a function of detector radius and incoming neutrino energy? Are any observed variations fully contained within the assigned uncertainty thresholds?

To examine whether the reconstruction algorithms perform consistently with respect to the spectral shape, we performed a data-driven validation using the high-statistics cosmogenic control samples ^{12}B. The reconstructed energy spectra obtained with OMILREC were compared between two radial regions: ^{12}B events within the FV (0<R<16.5m) and a sample extended to outside the FV (0<R<16.9m). As shown in the figure below, no significant energy-dependent deviation is observed. Any residual variation is constrained to the percent level within the assigned uncertainty.

To clarify this, we have added the following sentence in italics to the Methods:
"The spread among the different reconstruction algorithms near the fiducial boundary yields a 1.6% relative FV efficiency uncertainty. *Possible energy dependence of this FV-related uncertainty is examined using high-statistics cosmogenic ^{12}B events, and no significant energy-dependent effect is observed within the assigned uncertainty.*"



4. In methods, page 18, right column, 3rd paragraph, it says that Channels with low occupancy or timing misalignment were excluded from analysis. How many channels as a percentage of the total were excluded?

Low occupancy in approximately 2%-3% of channels resulted from automatic HV-off due to the flashing effect, whereas timing misalignment issues were present in <0.1% of channels.

Thus, **we have added the quantitative information into the revised text:**
"Channels with low occupancy (about 2%-3%) or timing misalignment (<0.1%) were excluded from analysis."

5. In methods, the point-like event reconstruction is described on p19. PDFs are brought up in the first paragraph, are these the same PDFs that are described in the following 3 paragraphs for each of the separate reconstruction algorithms? How are the PDFs used in the event reconstruction?

The PDFs for the residual hit time, introduced in the first paragraph in p19, are the same ones used in the OMILREC and JVertex reconstruction algorithms. In contrast, the VTREP algorithm uses a different strategy that does not rely on PDFs for vertex reconstruction.

As described in the first paragraph, after subtracting the time-of-flight of photons (t_{tof}) and the event start time (t_0), the residual hit times of all fired PMTs are required to conform these PDFs. A time-based likelihood function is constructed as follows:

$$L(t_{1,res}, t_{2,res}, \ldots t_{n,res} | x,y,z) = \prod_{i}^{n} p(t_{i,res} | x,y,z), t_{i,res} = t_i - t_{tof}(x,y,z) - t_0$$

the vertex (x, y, z) that maximizes this likelihood function is taken as the reconstructed event vertex. It should be noted that, while the time likelihood functions in OMILREC and JVertex are rather similar, but their underlying PDFs are constructed in slightly different ways. More details can be found in Ref.[75].

We have revised the context as: "Its time likelihood calculation utilizes the aforementioned residual hit-time PDFs, "
An additional sentence is also added at the end of the JVertex paragraph: "Similar to OMILREC, the vertex is determined by maximizing the likelihood function."

6. In the description of OMILREC, the vertex reconstruction was improved by about 25% to 50%. What do these numbers quantify?

There was a typo. **The text was modified as follows:** "the method significantly improved vertex resolution by about 25% to 50% [38], particularly for events near the detector edge, with respect to utilizing timing information only."

[Figure redacted]

[Redacted] Both JVertex and VTREP reconstruct the vertex using only PMT timing information. In contrast, OMILREC incorporates an additional constraint from the PMT charge distribution, which is also sensitive to the event vertex and particularly enhances reconstruction accuracy for events near the detector edge.

3. Detector Response model
1. I find the description of the fit to the gamma source energy spectra confusing. What are the parameters included in the fit? How are the uncertainties in these fit parameters incorporated into the oscillation fit? What feature of Figure 6 shows that the positron non-linearity uncertainty is 1%?

To accurately extract the reconstructed energy peak and the with (energy resolution) for each gamma source, we conducted a detailed spectral fitting analysis. A Gaussian function was employed to model the full-energy peak, with its amplitude, centroid and sigma as free parameters. All decay pathways as documented in nuclear databases were taken into account. Furthermore, we accounted for the energy leakage caused by the source encapsulation, which appears as a shoulder-like feature in the spectrum, using either Monte Carlo simulations or empirical polynomial functions. The potential influence of backgrounds such as C14 pileup was also evaluated and incorporated into the fitting.

We have revised the context in the "Energy scale non-linearity" part.

While the uncertainty in the gamma energy peak derived from the calibration source spectral fitting is small (<0.1%), the overall uncertainty in the energy scale is determined by incorporating additional contributions from factors such as residual detector non-uniformity and temporal stability. The combined analysis results in a total energy uncertainty of approximately 0.5%. Further details are provided in the "Calibration and Reconstruction" section and Fig. 2. In the oscillation fit, this energy scale uncertainty is included as a nuisance parameter.

Fig. 6 shows the fitted non-linear energy responses for positrons (green), electrons (blue), and gammas (red) as solid lines, with the surrounding uncertainty bands indicating the 68% confidence-level range compatible with the calibration data. For positrons, the band boundaries fall within **±0.6%** of the central value, as indicated in the legend of the Fig.6(d). **We revised the context accordingly.**

2. Comparisons show that the instrumental non-linearity for the gamma sources has minimal impact, but on what quantity? The three analyses agree on what to within 0.5%?

We primarily examined the impact of instrumental non-linearity on the derived positron non-linearity curves. Comparisons show that including or excluding this effect for centrally located gamma calibration sources has a minimal influence; the resulting discrepancy is only about 0.2% at 12 MeV. This level of variation aligns with the broader 0.5% agreement observed among the three independent analyses on the same curves.

We have revised the context accordingly.

4. Selections:
1. Signal:
1. Page 12, column 2: The term "spatio-temporal coincidence" is too vague to adequately describe the primary signal selection criteria (moreover, it is referred to as prompt-delayed coincident cuts in methods) Fig. 8 clarifies the requirement, is a compelling demonstration of the strength of the cut, and is arguably more informative than Fig. 1 for illustrating the cleanliness of the IBD coincidence signature. The authors might consider promoting that plot to the main body of the paper.

In this context, the spatio-temporal coincidence is the most essential feature characterizing reactor antineutrino signals. We have presented this concept schematically (and in a cartoon-style illustration) in Fig. 1 to render it more accessible to the broad readership of Nature. Other selection criteria, such as the fiducial volume cut, muon veto and multiplicity cut, are primarily employed to suppress background events. These points are elaborated in the subsequent paragraph. Regarding the potential relocation of Fig. 8 to the main text, we prefer to retain it in the Methods section due to the strict limit on the number of figures permitted in the main text of a Nature article.

2. Reference to the 10-cm cubes on page 12 column 2, last paragraph is confusing. While it is described in methods, there is not enough context at this point in the paper to understand what that means or why it is relevant.

The first sentence "Other selection efficiencies—including the muon veto, spallation-neutron veto, and multiplicity cut—were derived via livetime analysis using 10-cm cubes." **has been revised as:**

"The efficiencies for the remaining selection criteria (muon veto, spallation-neutron veto, and multiplicity cut) were obtained through livetime analysis, performed by segmenting the detector into 10-cm cubes and summing the livetimes in each segment. This segmentation method addresses the complexity arising from overlapping vetoed volumes and their varying veto durations, which preclude a straightforward analytical calculation of the total livetime."

3. Is figure 7 after all cuts (except the FV cut)?

Yes, Figure 7 uses events selected with all cuts other than the fiducial volume's radial requirement (R<16.5m).

2. Background:

1. Page 13, Geoneutrinos and world reactor background: The event rates quoted in the text differ from those reported in Table I. This discrepancy presumably arises from the application of the 69.9% detection efficiency correction. Please state this explicitly to avoid confusion.

We have added one sentence in the caption: "The background rates are prior to efficiency correction and have all selection cuts applied."

In addition, **add one sentence at the end of the paragraph that discusses World Reactor background:** "Note that the rates for the World Reactor (and geoneutrino) background in Table 1 are scaled by the total $\overline{\nu}_{e}$ selection efficiency."

Also see the answer to the next question.

2. Are all background components efficiency-corrected in the same manner? For geoneutrinos and world-reactor neutrinos, which constitute irreducible IBD signals, applying the same efficiency correction as for the signal appears justified. However, this may not be appropriate for other background categories, which are affected by only a subset of the selection cuts. Please clarify whether efficiency corrections are applied to each background component and, if so, the size of those corrections.

In Table 1, all quoted background rates are prior to efficiency correction but have all selection cuts applied. Among them, the rates of geoneutrinos and world reactor background are originally estimated via calculation, as explained in the text; therefore, their values shown in the table are scaled by the total reactor electron antineutrino selection efficiency.

5. Uncertainties:

1. Breaking down the statistical and systematic components of the total uncertainty would help support the statement that these measurements are statistically limited, while also providing an estimate of the relative contribution of each component.

An uncertainty breakdown is provided for the referee. For $\sin^2\theta_{12}$, the 2.81% total uncertainty consists of 1.95% (stat.) and 2.02% (syst.). For Δm^2_{21}, the 1.55% total uncertainty consists of 1.24% (stat.) and 0.93% (syst.).

As estimated in the above breakdown, the statistical uncertainty is not dominant for $\sin^2\theta_{12}$. This is because, in the present analysis, we adopted a conservative systematic uncertainty based mainly on the initial performance of the JUNO detector (arXiv:2511.14590, "Initial performance results of the JUNO detector"). For instance, the fiducial volume cut uncertainty constitutes a dominant systematic uncertainty for $\sin^2\theta_{12}$. This uncertainty can be reduced with increased statistics of inverse beta decay (IBD) events and improved vertex reconstruction. Moreover, the systematic uncertainty includes the uncertainties in the backgrounds, whose precision will improve with definite statistics. This is particularly the case for the dominant background, Li9/He8, which is entirely estimated from data. Therefore, it is true that the current measurements are statistically limited, and that acquiring more data will improve not only the statistical uncertainty but also the systematic one.

We prefer not to provide a detailed breakdown of uncertainties at this stage because simply scaling the current statistical uncertainty without accounting for the expected improvements in systematic uncertainty could lead to misleading projections for mid-term and long-term results. Moreover, there are correlations between the uncertainties, and it is extremely difficult to fully separate the statistical and systematic components for the reasons explained above. Finally, in the revised version ("Spectral fit and results" section), **we have added qualitative discussions on the dominant systematic uncertainties.**

2. Recommend including a dedicated section or paragraph on systematic uncertainties in the main text, as the current discussion is spread across multiple sections.

We thank the referee for their suggestion to include a dedicated section on systematic uncertainties.

While a consolidated systematic uncertainties section can be valuable for specialists who have deep expertise in the details of a neutrino experiment, the present manuscript is intended for the broad readership of Nature, where extended technical discussions may not be the most accessible format.

Moreover, the systematic uncertainties in this oscillation analysis span multiple aspects—including event selection efficiencies, backgrounds, detector energy response, and reactor flux and spectrum predictions. Each has a different origin, evaluation method, and impact. We believe it is more natural and transparent to discuss each uncertainty within its relevant technical context rather than consolidating them into a single section. In the current manuscript, we have already highlighted the dominant systematic effects in their respective sections.

After careful consideration, we have therefore decided to maintain the current organization, which we feel best serves the clarity and flow of the paper for a general scientific audience.

3. It would be helpful to clarify the impact of various systematics not only on predicted spectra but also on the extracted oscillation parameters. Please include all relevant sources, such as, calibration (energy scale, non-linearity), detector efficiency, number of target protons, various backgrounds, and reactor flux. While Table 2 summarizes some of these for the reactor neutrino rate, it does not indicate how the uncertainty modifies the shape of the non-oscillated spectrum nor how that uncertainty affects the mixing angle or the mass-squared splitting. A more comprehensive table including calibration-related systematics and their effect on both the mixing angle and mass-squared splitting would provide a clearer overview, highlighting similarities or differences in how each systematic impacts the oscillation parameter measurements.

The primary reason we do not provide a detailed breakdown of systematic effects on individual oscillation parameters in this paper is the challenge of cleanly separating them in the presence of current statistical fluctuations. For example, a negative contribution may appear because of fluctuation. Please also refer to the answer to the question 5.1 in the above.

As the referee pointed out the importance of such a breakdown for understanding shape distortions, we have in fact performed this type of analysis in our physics projection paper (Chin. Phys. C 46 (2022) 12, 123001) for the systematics-dominated regime expected in the future, e.g., with 6 years statistics.

A dedicated breakdown table, as suggested by the referee, is planned for a future publication once increased exposure reduces statistical correlations and allows for a clearer, more robust decomposition of the systematic uncertainties.

After careful consideration, we prefer not to provide a detailed breakdown of systematic effects on individual oscillation parameters in this paper.

4. Figure 11: It is great to see the consistency of the pulls across the three different analysis strategies. Would it be possible to extend this figure to include all sources of systematic uncertainty, such as energy scale, non-linearity, flux parameters, and oscillation parameters?

We appreciate the suggestion to extend Fig. 11. However, a direct graphical comparison of all systematic pulls across the three analyses is challenging, due to different treatments. Some details are listed below:

- The analyses have a very large number of pulls. It is just impractical to show them all. The ones that we decided to show are the ones that have the most impact and/or that show the largest deviation from the nominal. The great majority of the other pulls sit very close to the nominal expectation, simply because our data does not have enough power to constrain them.

- The three analyses used different statistical treatments (covariance matrix vs. pull terms). For instance, the energy scale and non-linearity uncertainties were handled via a covariance matrix in one analysis, which does not produce a plottable pull parameter.
- For non-linearity, the treatments of non-linearity differ mainly in how the energy non-linearity response is modeled and how the B12 spectrum shape uncertainty is handled. In the energy response, two analyses treat instrumental non-linearity as an additional source alongside scintillation non-linearity, while the third removes it beforehand through single-channel PMT charge correction. Regarding the B12 spectrum, one analysis includes its shape uncertainty, whereas the other two currently treat it as secondary and omit it.
- For the absolute flux uncertainty, it was incorporated into the segment parameters of the reactor neutrino spectrum model. This uncertainty is constrained by the Daya Bay spectral measurement and its covariance matrix, such that the overall normalization and shape of the reactor spectrum are allowed to vary within a correlated manner consistent with the external data. It's difficult to graphically illustrate this uncertainty.
- Finally, the most important cross-check is the agreement on the oscillation parameters themselves, which are consistent to better than 0.5 sigma. To maintain clarity, we report only one set of oscillation parameters.

Therefore, we prefer not to extend Fig. 11 with more pulls.

6. Spectral fit and Results:
1. Daya Bay near detector data is used as input to the non-oscillated spectrum prediction. More information should be provided as to why this is appropriate and what components of the flux model are common to the two experiments and what parts differ. The uncertainties on the integrated flux are discussed, but how much can the shape of the prediction change under different assumptions? It would be useful to include an error band on the non-oscillated prediction in figure 3.

The strategy on the flux prediction and constraints is discussed in the Method part [Reactor flux prediction and uncertainties, and Statistical methods].

The Daya Bay near-detector spectrum provides a direct experimental measurement of the non-oscillated reactor antineutrino flux, appropriate due to the use of similar reactor types and a negligible oscillation effect at its short baseline. Using this as an input for our prediction is the most data-driven approach available and minimizes reliance on ab initio theoretical flux models. Our prediction uses the same underlying per-fission isotope spectra but re-normalizes to the JUNO experiment's conditions:

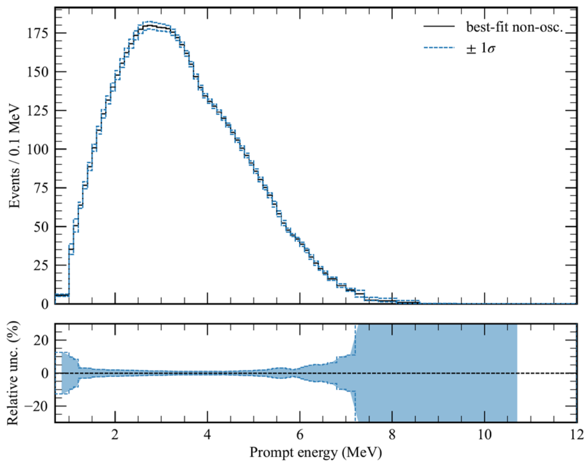
Common components: The fundamental isotopic fission yields (from U-235, Pu-239, U-238, Pu-241) and the associated reactor antineutrino emission spectra per fission are identical. The evolution of these fission fractions over a reactor cycle is also modeled similarly.

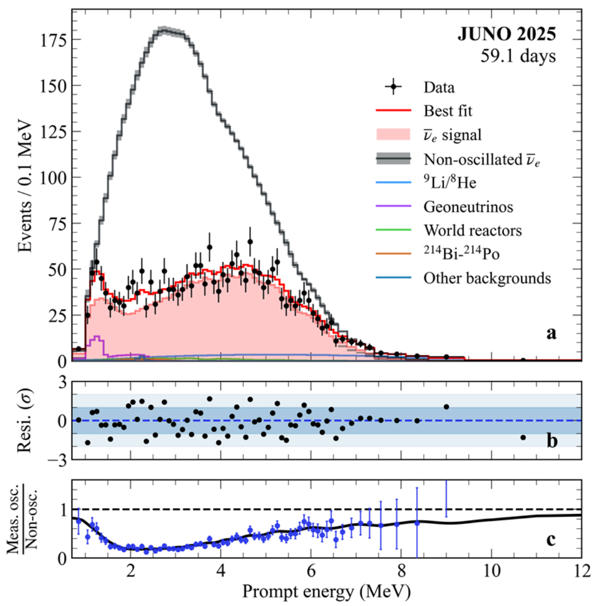
Different components: The primary difference lies in the absolute normalization, accounting for the site-specific differences in time-averaged fuel composition (i.e., fission fractions, 1% difference between Daya Bay and JUNO for the current 59.1 days data), spent fuel and non-equilibrium corrections, as well as small baseline-dependent oscillation effects. Those effects are subdominant.

In addition, no correlation is considered between JUNO and Daya Bay detector systematical uncertainties, which is a conservative choice.

The uncertainty on the spectral shape, propagated from the Daya Bay measurement covariance matrix, is approximately 1%-10% (relative, point-to-point) across the oscillation analysis energy range, see below.

An error band representing this shape uncertainty has been added to the non-oscillated prediction in Figure 3a, also see below. Relevant text is also added to describe the error band.





2. Is the flux model fit at the same time as the oscillation parameters?

Yes, the freely varying segment parameters are fitted together with oscillation parameters.

3. The same Daya Bay data appears to be used twice — once to constrain the uncertainties on reactor fluxes and overall normalization, and then again to constrain the values on sin^2(2th_13) and dm31. If this is the case, could the authors clarify how double counting is avoided? In particular, why is a simultaneous fit of Daya Bay and JUNO data not needed to properly account for correlations between reactor flux and normalization constraints and the corresponding oscillation parameters?

To constrain the flux in the joint fit we use only the data from the DYB near detectors, while the information about θ_{13} and Δm^2_{31} comes overwhelmingly from the far detectors, whose data are not used directly.

For θ_{13} and Δm^2_{31} constraints: these oscillation parameters are extracted from the relative comparison between Daya Bay's near and far detectors. The residual dominant uncertainties are the statistical uncertainty of the far detectors and the near-far relative systematic uncertainties.

For reactor flux prediction at JUNO: the data from the Daya Bay near detectors was used to anchor a common reactor spectrum, supplemented by additional corrections specific to JUNO. These corrections account for core fission numbers, fission fractions, spent nuclear fuel, and non-equilibrium effects. Consequently, the uncertainties in the reactor flux predictions for JUNO arise from both the constraints provided by Daya Bay and the specific reactor systematics associated with JUNO. It is important to note that, no correlation between the systematics of Daya Bay and JUNO is assumed, in order to adopt a more conservative approach.

Therefore, in this paper, for simplicity, we did not perform a simultaneous fit of Daya Bay and JUNO data. Our current approach, which treats the θ_{13} and Δm^2_{31} as independent external inputs, is conservative and robust.

4. Page 14, Column 2: Daya Bay reports constraints on sin^(2th13), and it is sin^2(2th_13) which enters the JUNO oscillation formula (Eqn 1). Presumably, then the constraint should be applied in sin^2(2*th13) directly rather than sin^2(th_13) as the text suggests?

We also use $\cos^2\theta_{13}$ in the oscillation fitting and so $\sin^2\theta_{13}$ is more convenient. We have already propagated the uncertainties to $\sin^2\theta_{13}$. We have done a test which shows the results are identical for using $\sin^2 2\theta_{13}$ or $\sin^2\theta_{13}$. Therefore, we keep to use $\sin^2\theta_{13}$ in this analysis.

5. Is the constraint included as a full 2D chi^2 surface, including correlations, or is each constraint included as 1D pull term? Presumably it is the former, but please specify.

For the present analysis, we implemented the external constraints using separate 1D penalty terms, not the full 2D correlated surface. This simplification is justified because the correlation between Δm^2_{31} and $\sin^2\theta_{13}$ in Daya Bay is weak (see Fig. 1 in https://arxiv.org/pdf/2211.14988), and the constraints themselves have a very small influence on the final oscillation parameters.

This is demonstrated by our cross-check in the Answer to question #7 below, where the difference between using one or two such 1D constraints was found to be negligible. Consequently, any further difference introduced by moving from two independent 1D terms to the fully correlated 2D constraint would be even smaller and immaterial to the current results. In future analyses as precision increases, a full 2D implementation can be adopted.

For clarification, we make revisions "which incorporates statistical and systematic uncertainties of both signal and backgrounds **as well as two individual constraints on oscillation parameters, Δm^2_{31} and $\sin^2\theta_{13}$."** in the first paragraph of page 15.

6. Daya Bay reports two values of dm31 under normal and inverted mass ordering hypotheses. While it is briefly noted that the results for the inverted ordering are fully compatible, given that JUNO's primary goal is to determine the neutrino mass ordering, presenting results under only one hypothesis seems incomplete. We would recommend quoting the corresponding numbers and errors for both scenarios, or quantifying what fully compatible means. If fully compatible, does the abstract need the qualifier that says the result is only for the normal mass ordering?

We did the analysis in inverted ordering case, and the results are $\sin^2\theta_{12}$=0.30920±0.008698 and Δm^2_{21}=(7.498±0.1166)x10^{-5} eV^2, respectively. The difference is negligible compared with $\sin^2\theta_{12}$=0.30917±0.008691 and Δm^2_{21}=(7.498±0.1165)x10^{-5} eV^2 in the normal ordering case. The published digits are identical.

We modified the context in page 15 as " The results for the inverted mass ordering scenario are identical to those for the normal mass ordering at the quoted numerical precision. " (Original: "The results obtained for the inverted mass ordering scenario are fully compatible.")

7. There is an additional statement that leaving dm31 unconstrained had a negligible impact on the results – what exactly does this imply? Was dm31 allowed to float freely to both positive and negative values without any bounds? Could the term "negligible" be quantified more precisely, for example as a percentage impact on the final results?

About fitting range: Δm^2_{31} is a free parameter bounded within [1.5, 3.0]x10^{-3} eV^2, a range much wider than the physically allowed range. Fits were performed separately for each mass ordering scenario.

About the "negligible" statement: We quantify this by comparing fits with and without the external Δm^2_{31} constraint.

- Without Δm^2_{31} constraint, the results are $\sin^2\theta_{12}$=0.3094±0.0087 and Δm^2_{21}=(7.50±0.12)x10^{-5} eV^2.

- With Δm^2_{31} constraint (as reported in the paper), the results are $\sin^2\theta_{12}$=0.3092+/-0.0087 and Δm^2_{21} =(7.50+/-0.12)x10^{-5} eV^2. The fractional difference for Δm^2_{21} is 0.06% which is negligible compared to the uncertainty (2.8%).

We modified the text to be "Leaving Δm^2_{31} unconstrained had a negligible (<0.06%) impact on the results. "

8. Where does the value of the earth's crust density come from? Both the central value and the error differ from commonly accepted values used in long-baseline experiments. Was the 10% uncertainty quoted on the parameter included in the fit, or was rho held constant? How were the variations in the crust density determined to be negligible?

The Earth's crust density in our analysis comes from a site-specific geological model (arXiv:2511.15702, Ref.69). The density value differs from that often used in long-baseline experiments because reactor neutrinos travel a much shallower path, which includes lower-density sedimentary layers.

In our oscillation fit, we used a fixed average density (the central value from the model), and the estimated ~10% uncertainty on this average was included as an independent systematic uncertainty and marginalized over. The justification for treating density variations as negligible is also provided in the referenced work (arXiv:2511.15702), which shows that even large local density fluctuations affect the oscillation probability at a level far below our current experimental sensitivity.

9. With this updated JUNO measurement, what is the current significance on the difference between the reactor and solar determination of dm21? Perhaps mention JUNO's solar capabilities to probe this difference. This is a discussion that could potentially make the article more compelling to the non-expert scientist reader.

The significance of our measured reactor-solar Δm^2_{21} difference keeps this tension at the 1.5 σ level, similar to the global picture prior to this result.

Following the referee's suggestion to enhance the discussion, we have added a sentence in the outlook: "The current ~1.5σ difference between our reactor-based Δm^2_{21} and the solar neutrino value [ref] highlights the importance of an independent probe. JUNO's upcoming measurement of ^{8}B solar neutrinos [ref] will provide a direct and competitive measurement of Δm^2_{21} from a second source within the same experimental setup."

10. Do the iso-deltachi2 contours provide the correct coverage?

Yes, they have the correct coverage. The Feldman-Cousins approach has been employed to validate Wilks' theorem.

We add one sentence in the Methods section (page 28).

7. References:

1. It would be preferable that references 1 and 2 point to the original literature on

oscillations. The appropriate references are included later on as references 7 and 8. References 1 and 2 may be kept, but should also say "and references therein"

2. Reference to the MINOS experiment should be given in addition to Reference 15, the K2K experiment. The appropriate citation is

https://doi.org/10.1103/PhysRevLett.97.191801

3. In regards to the measurement of theta_13, it would not hurt to cite the other two short baseline reactor experiments RENO and DoubleChooz as these are included in the PDG average value of theta_13. The authors may also consider citing T2K (https://doi.org/10.1103/PhysRevLett.97.191801)
or (https://doi.org/10.1103/PhysRevLett.97.191801)
to emphasize the complementarity of reactor and accelerator neutrino experiments.

We thank the referee for these valuable suggestions regarding the references. We agree that citing the original literature provides important historical context, and that referencing complementary experiments (MINOS, RENO, Double Chooz, T2K) gives a more complete picture of the field. **We have implemented all suggestions as follows:**

- In the Introduction, references to the original oscillation literature (now Refs. 7 & 8) have been updated to include "and references therein" to better guide readers to the foundational work.
- The suggested references for the MINOS experiment and the other short-baseline reactor experiments (RENO, Double Chooz) have been added to the Methods section, where we provide a more detailed account of the relevant experimental results. The T2K reference has also been included to highlight reactor–accelerator complementarity.
- To maintain the concise style expected for a Nature introduction, we have kept the citation list there focused on first discoveries and milestone results, while the full set of supporting references is now provided in the Methods. This approach is consistent with another vey recent Nature paper in neutrino field: "Joint neutrino oscillation analysis from the T2K and NOvA experiments" (https://doi.org/10.1038/s41586-025-09599-3).

We believe these changes address the referee's points directly and improve the references' completeness.

Copy editing comments:
The following are a few stylistic suggestions that would improve the readability of the manuscript

1. Is the acronym NMO required? It is easily conflated with Normal Mass Ordering
2. The detector description in the JUNO experimental setup includes many acronyms, some of which are not used again in the main body of the paper. It would be easier on the reader if the acronyms were omitted from the main body of the paper, but introduced in the methods section. At the very least, it would be helpful to reintroduce

them so the reader doesn't have to flip back and forth over several pages.
3. Out of curiosity, why are 2 different types of large PMTs used?
4. P10 right column, 1st full paragraph, first line, drop "of" before LS
5. P17, first paragraph of Detector filling and commissioning, 5th line, Water->water
6. P20 first full paragraph on left column, 3rd line, "having" should be "have"
7. P21, in "fiducial volume cut" paragraph, 9th line, add "the" before "z axis"
8. P22, 0th paragraph on right hand side, line 10, "vetos" should be "veto"
9. P24, 0th paragraph on left, line 7, add "the" before "alpha"
10. P25, out of curiosity, what does MI stand for in S_MI (eq 3)?

For 1, we have replaced the acronym "NMO" with the full term "neutrino mass ordering" throughout the text.

For 2, we follow the suggestion to remove the acronym in the main body, and re-introduce in the Methods section if necessary.

For 3-9, Done.

For 10, MI stands for "model independent", we have added a parentheses notation in the text.

Referee #3 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #4 (Remarks to the Author):

This paper presents the first results from the JUNO experiment and details a comprehensive data analysis incorporating (i) calibrations of the detector response; (ii) background-suppressing event-selection criteria that enhance IBD purity; and (iii) systematic-uncertainty assessments from detector modeling, event selection, backgrounds, and signal prediction. The final oscillation fit is obtained using a dedicated statistical methodology and a blinded analysis strategy.

Using the first 59.1 live days of data from its 20-kton liquid-scintillator detector, the JUNO Collaboration has achieved the world's most precise measurements of the oscillation parameters $\sin2\theta_{12}$ and Δm^2_{12}, improving the global precision by a factor of 1.6 in both. These fundamental parameters play a central role in the Standard Model description of lepton mixing, and tightening their uncertainties has broad impact, from testing the three-flavor framework to strengthening global oscillation fits and informing future neutrino-physics programs.

As the world's largest liquid-scintillator detector, JUNO's ability to achieve this level of precision after such a short period of data taking is remarkable. Reaching such performance so early in the experiment's lifetime underscores the excellent quality of the detector performance, calibration strategy, and analysis framework. Equally striking is JUNO's first simultaneous observation of both atmospheric and solar neutrino oscillations, a milestone that highlights the detector's unprecedented sensitivity.

Taken together, these accomplishments demonstrate JUNO's strong potential for major future breakthroughs in neutrino physics and its central role in refining the global three-flavor framework. Given the significance and broad impact of these results, the work is well aligned with the standards of Nature.

At the same time, we believe the paper is suitable for publication once the authors have adequately addressed the following comments.

Major comments for considerations and improvement:
1. In abstract, the authors state: "These results advance the basic understanding of neutrinos, validate the detector's design, and confirm JUNO's readiness for its primary goal of resolving the neutrino mass ordering with a larger dataset.", which is also repeated in the final "discussion and outlook" section. We find this statement overly strong given the current results, and we recommend that the authors revise it

to more accurately reflect what can be supported by the presented results.

Here is a plot illustrate the achieved (taken from https://arxiv.org/abs/2511.14590v1) vs. expected energy resolution (https://arxiv.org/pdf/2405.18008) The resolution, if not improved, would significantly limit JUNO's capability in the NMO determination. Because of this, it is not appropriate to claim "confirm JUNO's readiness for its primary goal of resolving the neutrino mass ordering with a larger dataset".
We understand that the current energy resolution is enough for the θ_{12} and Δm^2_{12} measurement. In addition, further improvements in the calibration and understanding of the detector will likely lead to improvement in the energy resolution, but it is not yet appropriate to claim JUNO's "readiness for mass ordering determination".

We appreciate the reviewer's point of view. The JUNO detector is currently performing at or above design specifications for all key metrics—with the sole exception of energy resolution, which is actively being improved. Our focused efforts on calibration and reconstruction are on track to deliver significant improvements in resolution within months. Additionally, well-defined plans are in place to enhance the signal-to-background ratio (via further excluding $^9Li/^8He$ backgrounds by using muon track information) and to increase statistics (via fiducial volume optimization). Given this progress and roadmap, we believe it is reasonable and justified to state that the JUNO detector's performance demonstrates readiness for mass ordering determination.

We have expanded the "Discussion and Outlook" section in the revised version to include prospects for further improvements in upcoming analyses.

2. Related to the previous comment, C-14 pileup is discussed as a potential explanation for the degraded energy resolution observed in the Ge-68 data. If this is indeed the dominant effect, it would be helpful to understand why a similar pileup does not appear to affect the α events, which achieve better than 3% energy resolution at comparable energies.
Alternatively, if the difference primarily arises from the use of an extended source versus a point-like source, this distinction should be clarified, as improving the energy resolution for positron in this case may pose a more fundamental challenge than mitigating pileup effects alone.

Thanks for pointing out this. The energy resolution is crucial to the future determination of neutrino mass ordering. The collaboration is working hard to improve it.

The C-14 pileup affects all physics events. In addition, different ionization quenching and Cherenkov processes for gammas (e.g., from 68Ge positron annihilation) and alphas (e.g., from ^{214}Po) also impact the energy resolution, even when the reconstructed energies of these particles are comparable. Ongoing Monte Carlo studies have indicated additional contributions to the energy resolution from the energy deposit process, that were not captured in our previous Geant4 simulations, such as positron in-flight annihilation and positronium formation. All these effects are currently under careful evaluation to better understand the energy resolution for electrons, positrons, gammas, and alphas.

3. “Shape uncertainty” appears repeatedly in the oscillation analysis, for example:
i) Page 13: “This comparison constrained the spectral shape uncertainty to 20%”;
ii) an overall energy uncertainty of 0.5% was assigned? A single number?
iii) “A 5% shape uncertainty accounts for the two spectral models used [49, 50], the $^{238}U/^{232}Th$ compositional variations”;
iv) “This background contributes to 1.26 IBD events per day at the JUNO site, with 10% rate and 5% shape uncertainties.”
v) this comparison constrained the spectral shape uncertainty of (Li9/He8) to 20%.
Given the importance of shape uncertainties in the spectral fit and in determining the final oscillation parameters, it would be very helpful for the authors to clarify precisely how the “shape uncertainty” is defined. For instance: What is the actual degree of freedom associated with this uncertainty? What is the uncertainty for each knob?
At present, the description of these uncertainties is more consistent with a rate-only treatment and does not fully specify the additional degrees of freedom relevant for a shape analysis. Providing a summary table, perhaps in the Supplemental Material, detailing the rate and shape uncertainties, their parameterizations, and their implementation in the fit would greatly improve clarity.

The term "shape uncertainty" in this paper refers to all systematic uncertainties associated with spectral variations. In the current analysis, shape uncertainties are included using three different methods:

1. The systematics related to energy resolution, energy scale, non-linearity, and matter effects are described using appropriate parameterizations. The variations of these parameters are propagated to define the variations in the predicted energy spectrum. Meanwhile, these parameters are constrained by the pull terms in the χ^2 function.

2. The systematics of background shape variations are treated as uncorrelated between different energy bins and are included as bin-to-bin uncorrelated uncertainties. These are added to the statistical uncertainty; for a clear definition, please refer to the methods section.

3. The reactor neutrino spectral uncertainty is constrained by the Daya Bay data. We first employ a segmented function with free segment variations to connect the non-oscillated reactor energy spectrum of JUNO with that of the Daya Bay near detector. Then, a combined fit of JUNO and Daya Bay data is used to constrain the segment variations to within a range of 1%-10%.

We have modified the context in page 13-14 and page 24-26 relevant with the spectral shape uncertainties.

In addition, we prefer not to provide a summary table to avoid duplications with current tables and content. We believe it is more natural and transparent to discuss each uncertainty within its relevant technical context.

4. JUNO's first simultaneous observation of atmospheric and solar neutrino oscillations represents a significant milestone, demonstrating the detector's exceptional sensitivity to multiple oscillation modes. Although a Δm^2_{31} result is not yet being presented, we encourage the authors to provide some discussion of the projected precision on this parameter, beyond simply stating that adding the Daya Bay constraint has no impact on the fit. Since the precision measurement of Δm^2_{31} and compare it with the measurement from the numu disappearance provides another way to determine MO, including this information can help establishing the case towards JUNO's readiness in MO determination.

Due to the currently limited statistics, it has been known that the delta-χ^2 profile on Δm^2_{31} would have multiple local minima, while this was not the case for the Asimov analysis in our previous sensitivity paper ("Sub-percent precision measurement of neutrino oscillation parameters with JUNO", *Chin.Phys.C* 46 (2022) 12, 123001).

Such statistical fluctuations would gradually disappear as the statistics accumulates. Thus, it is not straightforward to discuss its projected precision at current stage. **Furthermore, given that the measurement of Δm^2_{31} is not the scope of this paper, we prefer not to distract the readers. We'll elaborate the precision in our future measurement to Δm^2_{31}.**

Additional comments for improvements:
1. In addition to the total uncertainties, it would be useful to provide estimation of statistical and/or systematic uncertainties for the $\sin 2\theta_{12}$ and Δm^2_{12}

Another referee (Referee #2) asked similar question. For $\sin^2\theta_{12}$, the 2.81% total uncertainty consists of 1.95% (stat.) and 2.02% (syst.). For Δm^2_{21}, the 1.55% total uncertainty consists of 1.24% (stat.) and 0.93% (syst.). Meanwhile, the systematic uncertainties mentioned here primarily arise from data-driven methods, which will significantly decrease with the accumulation of more statistics.

It is true that the current measurements are statistically limited, and acquiring more data will improve not only the statistical uncertainty but also the systematic uncertainty. However, as shown, the statistical uncertainty is not dominant for $\sin^2\theta_{12}$. This is because, in the present analysis, we adopted a conservative systematic uncertainty based mainly on the initial performance of the JUNO detector (arXiv:2511.14590, "Initial performance results of the JUNO detector"). For instance, the fiducial volume cut uncertainty constitutes a dominant systematic uncertainty for $\sin^2\theta_{12}$. This uncertainty can be reduced with increased statistics of inverse beta decay (IBD) events and improved vertex reconstruction. We prefer not to provide a detailed breakdown of uncertainties at this stage because simply scaling the current statistical uncertainty without accounting for the expected improvements in systematic uncertainty could lead to misleading projections for mid-term and long-term results. Once a mature and refined estimation of systematic uncertainties is established, a detailed uncertainty breakdown will be performed, which will be highly valuable for projecting future uncertainties.

Therefore, we prefer not to provide the statistical and systematic components of the total uncertainty in this paper. Instead, in the revised version, we qualitatively have discussed the dominant systematic uncertainties to the current measurement in the "Spectral fit and results" section in page 15.

2. In Figure 2a, The E_rec appears to show a dip around 13m for AmC source and v_e bar candidate, In addition, the AmC source seems to have an overall lower E_rec than the others. And the muon spallation source continues to decrease beyond 13m.
i) Some discussions regarding what causes all sources going down at 13m. Any insights on the different behaviours among different sources?
ii) Some discussions regarding what causes the overall lower reconstruction of AmC than others can be useful?

In Figure 2a, the observed differences in the reconstructed energy scale among the AmC source, IBD candidates, and muon spallation neutrons can be attributed mainly to two factors:

- Limited sampling points (AmC) vs. Uniform data (IBD, spallation neutron)

The AmC source provides calibration only at discrete fixed positions. The off-axis calibration was initially performed at a specific azimuthal angle. This is a significant constraint because the detector geometry is not perfectly symmetric in the azimuthal direction, due to internal structures such as the supports of the acrylic vessel. In contrast, IBD candidates and spallation neutrons are distributed throughout the detector volume. Their energy scale is averaged over complete spherical shells, which naturally smooths out variations in both the polar and azimuthal coordinates.

- Muon-induced late afterpulses (mainly affect spallation neutron)

The spallation neutron sample is selected within a muon-veto window. For muons producing tremendous amount of photons, PMTs are observed to generate late afterpulses (tens to hundreds of microseconds), as confirmed by JUNO muon data and standalone measurements. These afterpulses contaminate spallation neutron signals and bias their reconstructed energy. For neutrons near the LS edge, the parent muon passes close to the boundary, creating a more asymmetric afterpulse pattern that distorts the spatial dependence of reconstructed energy. Although we optimized spallation neutron selection to reduce afterpulse effects, residual influence at several ~0.1% level persists. In Figure 2, the OMILREC algorithm was used. As shown in our detector paper (arXiv:2511.14590), IBD neutrons and ^{214}Po exhibit consistent spatial dependence, as neither is affected by muon-induced afterpulses.

3. Page12: "Other selection efficiencies including the muon veto, spallation neutron veto, and multiplicity cut—were derived via live-time analysis using 10-cm cubes", For clarity, it would help to briefly explain what is meant by "10-cm cubes" for a general reader. For example, the author could clarify that the full fiducial volume (FV) is divided into small cubes of size of (10cm)^3 for the efficiency study then somehow average over FV.

Same response for the Comment 4.1.2 from Referee 2. The sentence "Other selection efficiencies—including the muon veto, spallation-neutron veto, and multiplicity cut—were derived via livetime analysis using 10-cm cubes." **has been revised to**:

"The efficiencies of the remaining selection criteria (muon veto, spallation-neutron veto, and multiplicity cut) were determined through a livetime analysis performed by segmenting the detector volume into 10-cm cubes and summing the livetime accumulated in each segment. This segmentation method addresses the complexity arising from overlapping vetoed volumes and their varying veto durations, which preclude a straightforward analytical calculation of the total livetime."

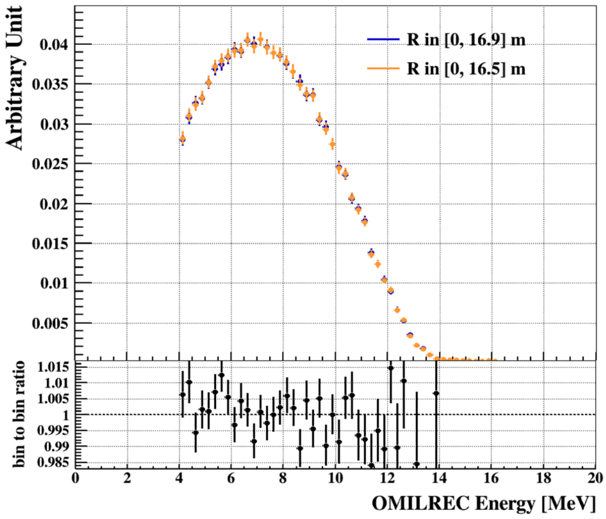
4. It was stated that the geometric volume fraction is determined to be 80.6%, and this has been validated by the IBD above 3.5 MeV. What is the precision for the validation? Is it sufficient to validate at 1.6% level? The same question applies to the B-12 validation.
What is the energy dependence on the uncertainty for the fiducial volume?

In the validation by the IBD candidates above 3.5 MeV, 3481 events are selected within the FV and 4270 events without the FV cut. The spallation-neutron veto was not applied in order to increase the statistics. This yields a fiducial-volume efficiency of 3481/4270=81.52%. A correction accounting for neutron spill-out at the fiducial boundary, evaluated using dedicated neutron Monte Carlo simulations, is implemented as a 2.4% increase of the denominator, resulting in a corrected efficiency of 79.6%. This value is consistent with the geometric expectation of 80.6% within the assigned 1.6% relative uncertainty.

The same validation strategy is applied using high-statistics cosmogenic control samples ^{12}B. Due to the correlation of ^{12}B production with parent muon tracks, its spatial distribution is not expected to be strictly uniform, which introduces systematic biases in the FV fraction. As a result, the effective validation precision of the ^{12}B-based FV fraction is limited by these systematic effects and is at the few-percent level, and is therefore less constraining than the IBD-based validation. Nevertheless, the high statistics of the ^{12}B sample make it well suited for probing possible energy-dependent FV effects.

To clarify the precision for the validation, we have revised the texts in the Methods as below: "We performed independent estimations using a ^{12}B sample and a high-energy (>3.5 MeV) IBD sample by calculating the fraction of these events reconstructed inside the FV relative to the total. *For the high-energy IBD sample, this validation is consistent with a precision comparable to the assigned FV-related uncertainty. For the ^{12}B sample, the effective validation precision is limited by systematic effects associated with its non-uniform spatial distribution, which is correlated with parent muon tracks, and is therefore less constraining for the absolute FV efficiency. Overall, these results are consistent with the nominal method within the quoted uncertainty*."

To examine the energy dependence on the uncertainty for the fiducial volume, we performed a data-driven validation using the high-statistics cosmogenic control samples ^{12}B. The reconstructed energy spectra obtained with OMILREC were compared between two radial regions: ^{12}B events within the FV (0<R<16.5m) and a sample extended to outside the FV (0<R<16.9m). As shown in the figure below, no significant energy-dependent deviation is observed. Any residual variation is constrained to the percent level within the assigned uncertainty.



We have added the following sentence in italics to the Methods:

"The spread among the different reconstruction algorithms near the fiducial boundary yields a 1.6% relative FV efficiency uncertainty. *Possible energy dependence of this FV-related uncertainty is examined using high-statistics cosmogenic ^{12}B events, and no significant energy-dependent effect is observed within the assigned uncertainty.*"

5. In the "JUNO detector setup" section, it is mentioned that "a very small geometrical inefficiency due to the top chimney …". It would be useful to be quantitative on this inefficiency.

[Figure redacted]

[redacted] he muon rate and muon tagging efficiency were monitored across the data set, also see in the above. The inefficiency of the water Cherenkov detector to tag CD muons is less than 0.02%. Detailed analyses show that these untagged muons are passing through the CD top chimney and track length in water is short.

We've added the inefficiency value in the revised manuscript.

6. In the "Detector response model" section, it is stated that "systematic effects including energy leakage from source encapsulation and peak stability" were also evaluated. It would be useful to be quantitative on these.

Gammas from calibration sources can deposit energy in the source encapsulation material, affecting the Compton contributions. The potential mis-modeling of this shape due to the imperfect modeling of source encapsulation might bias the peak position. We modeled the Compton contribution with both a polynomial and an MC-based quenched deposited energy, respectively, the variations in the fitted peak were <0.1%.

For peak stability, we used three types of full-detector samples (alphas from Po210, Po214 and spallation neutrons). The variation through the dataset period for all three sources are within 0.2%.

We have added the quantitative information in the revised manuscript (page 20).

7. In method, Figure 6(a). the terms "non-linearity" and "effective gamma energy" are not clearly defined. Presumably "non-linearity" refers to the ratio between reconstructed energy and true energy, and "effective gamma energy" involves some treatment for multiple-gamma sources, but they should be defined explicitly.

In Fig.6a, the effective gamma energy of a single gamma event is its true energy. If there are multiple gammas in one event, such as ^{68}Ge, ^{60}Co and n^{12}C, the effective gamma energy refers to the average visible energy of each gamma after taking into account the branching ratio and scintillator non-linearity.

$$E_{effective}^{^{68}Ge} = \frac{0.511 \times NL^{\gamma}_{E^{\gamma}_{true}=0.511MeV} + 0.511 \times NL^{\gamma}_{E^{\gamma}_{true}=0.511MeV}}{NL^{\gamma}_{E^{\gamma}_{true}=0.511MeV} + NL^{\gamma}_{E^{\gamma}_{true}=0.511MeV}}$$

$$E_{effective}^{^{60}Co} = \frac{1.173 \times NL^{\gamma}_{E^{\gamma}_{true}=1.173MeV} + 1.333 \times NL^{\gamma}_{E^{\gamma}_{true}=1.333MeV}}{NL^{\gamma}_{E^{\gamma}_{true}=1.173MeV} + NL^{\gamma}_{E^{\gamma}_{true}=1.333MeV}}$$

$$E_{effective}^{n^{12}C} = \frac{68\% \times 4.945 \times NL^{\gamma}_{E^{\gamma}_{true}=4.945MeV} + 32\% \times 3.683 \times NL^{\gamma}_{E^{\gamma}_{true}=3.683MeV} + 32\% \times 1.261 \times NL^{\gamma}_{E^{\gamma}_{true}=1.261MeV}}{68\% \times NL^{\gamma}_{E^{\gamma}_{true}=4.945MeV} + 32\% \times NL^{\gamma}_{E^{\gamma}_{true}=3.683MeV} + 32\% \times NL^{\gamma}_{E^{\gamma}_{true}=1.261MeV}}$$

We have added the following clarification in the manuscript: "As used in Fig. 6a, the effective gamma energy is defined as the true gamma energy for a single gamma-ray. For calibration sources emitting multiple gammas (e.g., from ^{68}Ge, ^{60}Co, or n 12C capture), the effective gamma energy represents the branching-ratio-weighted average of the visible energy per gamma, after accounting for the best-fit scintillator non-linearity. "

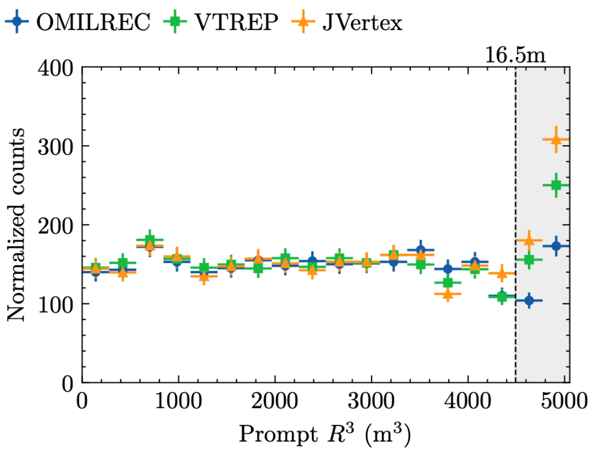
8. In the same section, there is a statement "... show that including or excluding instrumental non-linearity for centrally located gamma …". What about the sources at off-center locations? Some clarification can be useful.

There is no evidence of non-linearity at off-center positions deviating from center locations. The sources placed at ±14 m along the center axis shows difference up to 0.4% across gamma sources compared to the center location as shown in the figure below. This is well within the residual non-uniformity and properly taken into account in the energy scale systematics.

We have added one sentence in the revised manuscript: "Non-linearity measurements at off-center positions (z=±14 m) show no significant deviation from the center, with variations <0.4% included in the overall energy scale systematics."

9. In Figure 7, the binning appears to be coarse. It would be helpful if the author could clarify the binning strategy, and provide more details in the bin near the edge of the FV. In the next bin after FV cut, results from 3 strategies start to separate indicate some inconsistency of the understanding of the event reconstruction. Does this inconsistency happen earlier inside the FV? With such a coarse binning, it is difficult to evaluate this point.

The bins inside the FV and the next bin after the FV cut have the same bin width, defined uniformly in R^3. The narrower bins at larger R^3 are only a visual choice to accommodate the large dynamic range on the y axis. For referee only, we provide an alternative version of Figure 7 with doubled bin-size, in which all bins have the same width, defined uniformly in R^3.

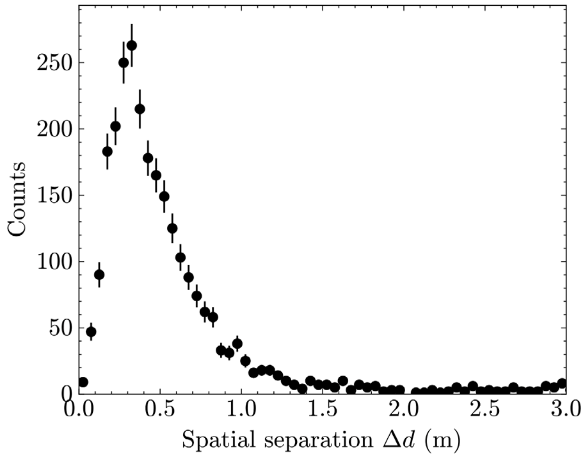


As seen, no systematic separation is observed inside the FV (R<16.5m). A visible divergence appears only in the first bin immediately outside the FV cut. This behavior has been included in the assigned FV efficiency uncertainty.

For clarity of presentation, we retain the original figure in the manuscript, as the alternative binning leads to the same conclusion.

10. In Figure 8, it would be useful to show the Delta_d distribution beyond the 1.5 m.

We have provided the extended *Δd* distribution up to 3.0 m for referee only, as shown below. Events with *Δd*>1.5m are mainly due to accidental coincidences.



11. In Section "Backgrounds", it is stated that the Li9/He8 ratio is taken from Daya Bay. Given the overburden is different between Daya Bay and JUNO, this ratio may be different, since it would depend on the muon deposited energy. Additional discussion on the uncertainty would be useful.

In Daya Bay's first measurement of cosmogenic 8He ([2402.05383] First measurement of the yield of 8He isotopes produced in liquid scintillator by cosmic-ray muons at Daya Bay), the measured Li9 yields and He8 yields were fitted with a power-law function, shown in below.

The predicted ^{8}He-to-^{9}Li ratio in JUNO using the average muon energy of 207 GeV with the power derived power law in [2402.05383] is (1.43±0.23)% for the β-n sample. We took a much more conservative uncertainty of this ratio, i.e. varying the ^{8}He-to-^{9}Li ratio from 1.43% to 10%, which changes the fitted central values of ^{8}He-to-^{9}Li ratio only by 2%. This 2% is accounted for in the final total ^{9}Li/^{8}He rate uncertainty.

We also added relevant clarifications in the methods part for the impact of ^{8}He-to-^{9}Li ratio: " A systematic uncertainty arises as the muon selection may miss the true parent of ^{9}Li/^{8}He, as assessed by varying the joint CD-WP muon tagging requirements which changed the fitted rate by 6%. Varying the ^{8}He-to-^{9}Li ratio by up to 10% shifts the central value by only 2%. The difference between the residual ^{9}Li/^{8}He rates in data and Monte Carlo is 11%. Combining this with the statistical uncertainty, an overall 34% uncertainty is assigned to the ^{9}Li/^{8}He background rate."

[Figure redacted]

12. In page 23, The expression “We therefore performed a combined fit across four muon energy intervals, accounting for cross-interval correlations where 9Li/8He events from one interval mimic antineutrino signals in others.” is confusing. Figure 9 shows five muon-energy bins, and it is not clear how these relate to the “four intervals” mentioned in the text. In addition, the phrase “^{9}Li/^{8}He events from one interval mimic antineutrino signals in others” needs further clarification. It would be helpful if the author could explain the mechanism of how this happened and how this correlation is modeled in combined fit.

Thank the reviewer for pointing this out. The combined fit was performed over the last four bins in Figure 9. The first bin (muon energies below 0.6 GeV) was not included in the combined fit since it is difficult to reliably fit the corresponding Li9/He8 contribution due to the very low production yield and high muon rate. Therefore, prediction from MC was shown in the first bin.

To avoid confusion, we revised the sentence as: “We therefore performed a combined fit across four muon energy intervals (all with visible energy above 0.6 GeV and similar muon rates), explicitly accounting for correlations arising from ^{9}Li/^{8}He events produced in one interval and appearing as antineutrino candidates in others."

Further explanation to “^{9}Li/^{8}He events from one interval mimic antineutrino signals in others”: For time-since-last-muon fit, the muons were divided into four energy ranges. For each muon energy range, we take all IBD-candidates and calculate the time-since-last-muon in this particular energy range. For Li9/He8 produced by muons not in this energy range, their time-since-last-muon distribution is uncorrelated with muon, just the same as real IBD events. Events in different muon energy slices are not independent. Each muon energy slice contains the same types of random coincidence events, in particular IBD and Li9/He8 events induced by muons from other energy slices. To properly account for these correlations, a combined fit that incorporates the relationships among different muon energy slices is applied.

13. Fig.9, the last bin does not have a label, please add.

Fig.9 has been revised.

We also noticed that, in the original plot, the points had been corrected for the IBD selection efficiency, which was inconsistent with the values reported in Table.1. We have therefore revised the plot accordingly. **The caption now specifies**: "The residual $^{9}Li/^{8}He$ rates are prior to efficiency correction but have all selection cuts applied.".

14. In Equation 4, what is the value of “N”?

N defines the granularity of the parameterization, i.e., how many segments are allowed for reactor neutrino flux variations in the fit. We revised the context to clarify this.

Referee #5 (Remarks to the Author):

Firstly, the JUNO Collaboration is to be congratulated for designing, building, and constructing a liquid scintillator detector of unprecedented size and resolution. This paper is the first physics result and therefore an important milestone in the field of neutrino physics. Comments follow:

Main Narrative:
1. Some introduction to the NMO issue in the introduction would put the paper more in context for readers who are not particle physicists and who may not read the detailed "Methods" section.

Thank the referee's suggestion. **We have added a bit more discussion about the neutrino mass ordering issue in the introduction.**

2. Ref [38] is refered to in several places but has no information on where it is published

The arXiv eprint link is added.

3. In Fig, 2(a) is it understood why there a systematic shift in reconstructed energy between the three sources of neutrons? Even though this is taken as a systematic uncertainty this should be noted and discussed.

Referee #4 raised similar question. **We added one sentence in the calibration and reconstruction part:** "... around 0.45%. The slight residual discrepancy in energy scale among different neutron samples could result from the limited set of calibration points and muon induced afterpulses.".
Same replies are copied below:

In Figure 2a, the observed differences in the reconstructed energy scale among the AmC source, IBD candidates, and muon spallation neutrons can be attributed mainly to two factors:

- Limited sampling points (AmC) vs. Uniform data (IBD, spallation neutron)

The AmC source provides calibration only at discrete fixed positions. The off-axis calibration was initially performed at a specific azimuthal angle. This is a significant constraint because the detector geometry is not perfectly symmetric in the azimuthal direction, due to internal structures such as the supports of the acrylic vessel. In contrast, IBD candidates and spallation neutrons are distributed throughout the detector volume. Their energy scale is averaged over complete spherical shells, which naturally smooths out variations in both the polar and azimuthal coordinates.

- Muon-induced late afterpulses (mainly affect spallation neutron)

The spallation neutron sample is selected within a muon-veto window. For muons producing tremendous mount of photons, PMTs are observed to generate late afterpulses (tens to hundreds of microseconds), as confirmed by JUNO muon data and standalone measurements. These afterpulses contaminate spallation neutron signals and bias their reconstructed energy. For neutrons near the LS edge, the parent muon passes close to the boundary, creating a more asymmetric afterpulse pattern that distorts the spatial dependence of reconstructed energy. Although we optimized spallation neutron selection to reduce afterpulse effects, residual influence at several ~0.1% level persists. In Figure 2, the OMILREC algorithm was used. As shown in our detector paper (arXiv:2511.14590), IBD neutrons and Po-214 exhibit consistent spatial dependence, as neither is affected by muon-induced afterpulses.

4. What is the actual muon rate through the detector? I didn't see this number given.

In the JUNO detector paper Ref [38], Fig. 21 shows that the muon rate passing through CD is about 5 Hz, while passing through the WCD is about 9 Hz. **We have added the numbers in the revised version, page 12.**

[Figure redacted}

5. The flasher problem made a quality cut necessary. How many flashers were there in this data set? What cuts were made exactly? This would be appropriate for the Methods section.

The identified flasher event rate within [0.7, 12] MeV is 16.8±0.06 mHz in the analysis dataset. And none were found to form any IBD-like pairs. The flasher events were rejected using the standard deviation of hit time and counts across all fired PMTs with the following requirement.

$$\varepsilon = \left(\frac{\sigma_{N_{\mathrm{hit}}} - 0.55}{0.45}\right)^2 + \left(\frac{\sigma_{t_{\mathrm{hit}}} - 170}{80}\right)^2 < 1$$

The flasher event is efficiently vetoed by their distinctive feature and had a negligible impact on selection. **We believe the current level of descriptions (impact, variable used for cut) in the Methods are adequate.**

6. Values of sigma_rel in Table 1 imply units are %, but better to state explicitly for clarity

Yes, the unit of % is already mentioned after the words of "Selection Efficiencies".

7. In Table 1 refers to “Multiplicity” but this is not discussed until the section on Methods. There should be a short definition of this factor in the main narrative as there is with other issues

We have added a brief definition of Multiplicity in the main text: “The multiplicity cut rejects IBD candidates for which additional energy deposits are observed within a defined time window surrounding the prompt–delayed pair, thereby avoiding ambiguities in event association and suppressing accidental and correlated backgrounds. ”

8.A bit more information on the difference between “best-fit” and “pre-fit” (which I take to mean best estimate before spectral fitting) would be useful to the reader not familiar with these terms.

As suggested, in the caption of Table 1, we've added a brief explanation for "pre-fit" in the revised manuscript: "pre-fit (best estimate before spectral fitting)".

9.Page 14, when discussing external constraints on Delta_m^2_31 and sin^2_theta_13 please state explicitly the values used and their uncertainty. Did both these values come from Ref [46]? If so, the authors should consider using a global best fit values for these instead from a single experiment.

Yes, the two external constraints take the 1-D chi2 profiles from the Daya Bay 2023 result (Ref. 41 in the revised version). In the paper, we used the values and uncertainties of the two constraints.

Using the global fitting value for $\sin^2\theta_{13}$ has a negligible impact on our results, as it is dominated by Daya Bay data. For the Δm^2_{31} constraint, the global fit value includes contributions from both long-baseline accelerator and short-baseline reactor experiments and carries a slight dependence on the neutrino mass ordering. To avoid this ambiguity, we opted to use the Daya Bay fitting results as external constraints. We would also like to clarify that using these external constraints from the outset has a very minor impact on our findings. Incorporating Δm^2_{31} constraints from other experiments will be considered in future work.

10. Figure 4 is very important and for clarity the caption should state that all error bars from other experiments are 1 sigma.

Thanks for the suggestion. We added the statement in the caption.

Methods:
1.If possible add information on the PMT manufacturer, model, and any publications that describe performance parameters

Details of the PMT characteristics and performance are reported in the JUNO detector paper (arXiv: 2511.14590).

2. “Instead, SPMTs have…” would be better as “Conversely, SPMTs have…”

Corrected.

3. Again, Ref. [38] does not have any information on where it is published

The arXiv eprint link is added.

4. Pg 18 in Calibration Strategy, A figure should be included that shows all the locations of the AmC souce in order to show how well the CD volume was covered.

The JUNO detector paper (Initial performance results of the JUNO detector) is referenced, and we suggest the reviewer to refer to Figure 3 there.

[Figure redacted]

5. Detection efficiency is a function of the photon location on a PMT. What uncertainty was assigned to cover these geometrical effects if there were only central measurements used for the relative efficiency? Were there measurements off-center to validate the central measurement and associated uncertainty?

The angular dependence of the PMT photon detection efficiency (PDE) was mapped in our prior characterization study (Eur. Phys. J. C 82 (2022) 12, 1168). For the <75° zenith angles relevant to our fiducial volume (R < 16.5 m), the PDE variation is <10%, see the plot below. This effect is currently not corrected for in the reconstruction. These geometry effects will be carefully calibrated and tested in the reconstruction, and we'll continue to improve the reconstruction at the detector edge in the future.

[Figure redacted]

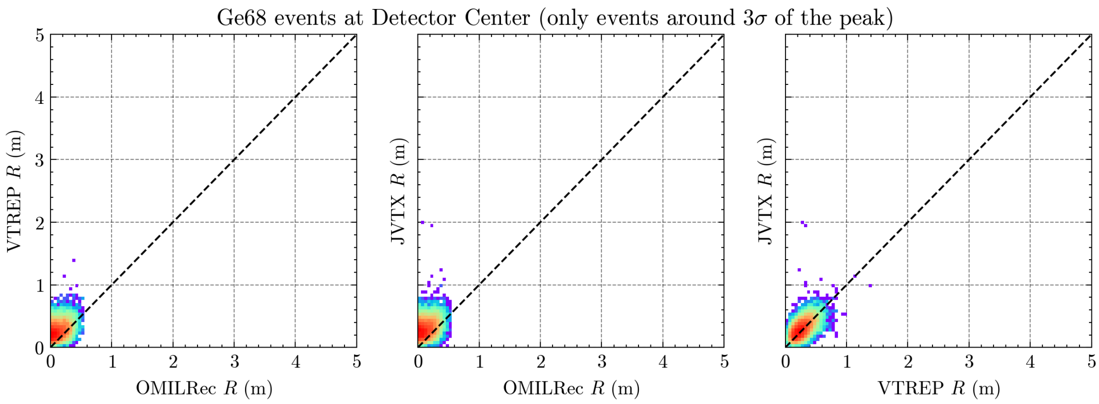
6. What fraction of channels were excluded from the analysis in the LPMTs due to low occupancy or timing misalignment?

Same response for Comment 2.4 from Referee 2.

Low occupancy in approximately 2% of channels resulted from automatic HV-off due to the flashing effect, whereas timing misalignment issues were present in <0.1% of channels. The values have been added in the revised manuscript.

7. It would be informative to see a plot of the vertex differences event-by-event for the 3 methods used. Were these differences used to put a lower limit on the vertex reconstruction resolution?

The comparison below, based on Ge-68 gamma source events at the detector center, illustrates the differences in the reconstructed radius among the three methods. However, these inter-algorithm differences do not set a lower limit on the vertex resolution. Each of our independent analysis uses a distinct reconstruction algorithm, and the resolution for a given method must be derived from its own internal precision and calibration—not from its divergence from other methods.



We refer the referee to our detailed detector performance paper (arXiv:2511.14590, excerpt its Fig.14 below) for an extended comparison of vertex reconstruction methods.

[Figure redacted]

8. Detector response model: how does the effective light yield change of off-center? A plot of this would be very informative. Also, if 1785 p.e./MeV (4 significant figures) is used, it should not be described as "approximate"?

The detailed discussion on light yield is in the detector performance paper (arXiv:2511.14590), and its Fig.13 is also shown below. A boundary effect, primarily caused by total reflection at the acrylic surface, becomes significant beyond 15 m. Within our defined fiducial volume, the resulting light yield variation is contained within 15%. The "approximate" word is removed in the revised version.

[Figure redacted]

9. The beta decay shape for 12-B decay to 12-C is an important validation for the energy non-linearity model. Please give information on how the theoretical shape was calculated, including how the electron screening was handled for the lower energy electrons. What model(s) were used and what uncertainly was assigned to this screening on the theoretical shape? In other words, I gather that Birk's Law formula was used and fit to the data based on departure from the theoretical curve, but to what uncertainty is the theoretical curve known at lower energies where screening is important?

The uncertainty in the theoretical β-decay shape of B12 is incorporated into our energy non-linearity model, following the methodology established in the Daya Bay energy-response studies (Nucl.Instrum.Meth.A 940 (2019) 230-242). The B12 decay is an allowed Gamow-Teller transition. Our spectral prediction is based on the formalism described in Phys. Rev. C 84 (2011) 024617 and Phys. Rev. C 83 (2011) 054615.

Specifically, we include shape corrections for the screening effect, electromagnetic and weak-interaction finite-size effects, and the weak magnetism effect. Radiative corrections are omitted because virtual-photon exchange does not affect the β-shape, and emitted real photons are detected within the liquid scintillator.

The dominant theoretical uncertainty originates from the weak magnetism correction, which introduces a linear energy-dependent term with a coefficient of (0.48 ± 0.24)%/MeV, as illustrated in the referenced Daya Bay work (also shown below). This uncertainty is treated as a nuisance parameter in the global fit of the energy non-linearity model, thereby propagating the theoretical uncertainty in the B12 spectrum—particularly relevant at lower energies where screening is important—into our final response model.

[Figure redacted]

We have added one sentence in page 21: "…, the treatment of systematic uncertainties follows the strategy of Ref. [39]".

10. Same for 11-C screening, but in addition the 11-C decay is not pure G-T like the 12-B decay is. This could (or could not) have implications for the shape and should be considered.

Our analysis uses C11 and B12 decays to constrain the energy non-linearity model across different energy regions. The C11 spectrum (~1 MeV) provides a low-energy complement to gamma calibration sources, while B12 (up to ~13 MeV) dominantly constrains the high-energy region.

Implementation of atomic and Coulomb corrections:

For the low-energy C11 spectrum, atomic screening effects are implemented following the Dirac–Hartree–Fock–Slater approach described by Hayen et al. (2018). Other Coulomb-related corrections follow the established prescriptions of Wilkinson (1989, 1993) and Huber (2011). These corrections are applied universally and are independent of the nuclear transition type.

Treatment of sub-dominant nuclear effects for C11:

The C11 decay is an allowed mixed Fermi–Gamow–Teller transition. Two nuclear structure-dependent effects are considered:

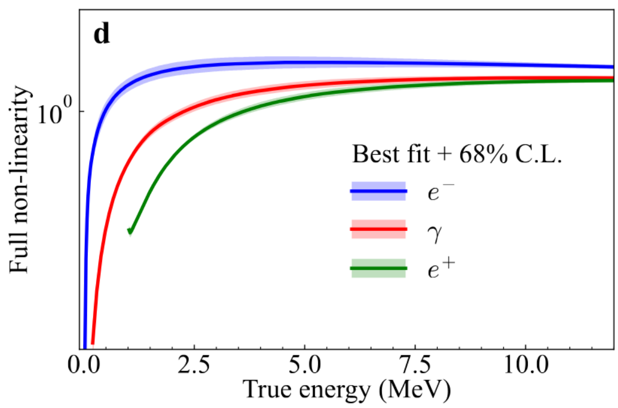
1. Recoil correction: Associated with the Fermi–Gamow–Teller mixing, this effect modifies the spectral shape at approximately the 0.05% level (Wilkinson 1989), an

order of magnitude below the ~0.5% uncertainty of our energy non-linearity model, and is therefore neglected.

2. Shape factor: Evaluated within the Gamow–Teller formalism using form factors constrained by experimental data (Severijns et al. 2023), its magnitude is about 0.1–0.2%. Given that this is also well below the model uncertainty and that C11 provides only a sub-dominant constraint, this evaluation is sufficient for the present analysis. We note that existing formalisms (Hayen et al. 2018; Huber 2011; Wilkinson 1993) treat shape factors separately for pure transitions and do not provide an explicit prescription for mixed transitions like C11, thus we only consider the shape factor within the Gamow-Teller formalism.

11. In addition, the 11-C decay positron spectrum only goes to 0.96 MeV while the positron curve in Figure 6(d) goes to 10 MeV. Does this imply that the positron data below 0.96 MeV plus the expected contribution from annihilation gammas was used to fit Birk's Law and then extrapolated to 10 MeV? If so, it would seem that the positron curve would how more uncertainty at higher energy, but it would appear from the figure to have less uncertainty than the electrons as energy increases. More information on this would make a stronger case. They authors should also consider making Fig. 6(d) logarithmic on the y-axis to make it more visible.

Our modeling of the energy non-linearity mainly follows the established strategy from the Daya Bay experiment (*NIMA 940 (2019) 230-242*). The model is fundamentally built on the electron response. The response to gammas is derived from the summed response of all secondary electrons and positrons they produce (simulated via Geant4), while the positron response is treated as an electron of the same kinetic energy plus the contribution from its annihilation gammas. The ^{12}B spectrum provides a direct, high-precision measurement of the electron response up to ~13 MeV, which directly anchors the model for positrons of the same kinetic energy. This explains why the positron uncertainty band in Fig. 6(d) remains comparable to or even smaller than that for electrons at high energy—it benefits from the constraints provided by these other, statistics-rich calibration samples. In summary, the model is constrained across the entire energy range by a combined fit to all available data: for the region below 2 MeV is primarily constrained by ^{11}C positrons and low-energy γ sources; for the region above 2 MeV is primarily constrained by ^{12}B electrons and high-energy gamma sources.



We tried the logarithmic y-axis for Fig.6(d), see above, actually the uncertainty band is less visible, so we prefer to keep the original version.

12. Pg 21 says an alternative analysis used 17m meters for a FV on delayed events. Was a FV cut used in the main analysis? Since one is not mentioned I would infer that such a cut was not used, but best to state this explicitly.

Yes, **we have applied a FV cut in the main analysis as described in the Methods (Page 21):** "Prompt positron candidates are required to have radius R < 16.5 m and |z| < 15.5 m to strongly suppress accidental and cosmogenic neutron backgrounds and to minimize the impact of biases in the reconstruction algorithms."

In addition to the FV cut on the prompt event, the alternative analysis mentioned on Page 21 applies an additional FV requirement on the delayed event (e.g. R<17.2 m).

13. Pg 22, why was a FV efficiency of 1.6% used based on comparing different fitters rather than the 1.8% from source deployments near the edge of the FV? Different fitters can have correlated systematics which always result in underestimated systematic uncertainties, whereas the source deployment does not suffer from this issue.

The 1.8% uncertainty from source deployments is derived solely from calibration data taken along the detector's z-axis, and therefore does not provide a full three-dimensional constraint on reconstruction effects related to the fiducial volume. In contrast, the 1.6% uncertainty adopted in the main analysis is obtained from the spread among different reconstruction algorithms, evaluated near the fiducial boundary using the three-dimensional distribution of IBD events.

As noted in the text: "The spread among the different reconstruction algorithms near the fiducial boundary yields a 1.6% relative FV efficiency uncertainty". This spread reflects not only the differences among reconstruction algorithms, but also the residual non-uniformity of the reconstructed vertex distribution near the FV boundary relative to the inner detector region.

Therefore, **potential correlated systematic effects among reconstruction algorithms are expected to be minor in this uncertainty estimation.**

14.In the Methods section on 9-Li/8-He it would be appropriate to show exactly how the 34% uncertainty was assigned to the rate. It is the largest background but after nearly a whole page of discussion it is still not clear to me exactly where this value for this uncertainty comes from.

The residual $^{9}Li/^{8}He$ rate in the IBD sample is estimated by fitting the time distribution of the prompt signal of IBD-like event to the most recent preceding muon. These preceding muons were required to be tagged by both the CD and WP detectors. The dominant uncertainty is statistical. The systematics were also considered, such as the parent muon tagging efficiency, ^{8}He-to-^{9}Li ratio, Monte Carlo and data agreement, etc: 1) The preceding muon selection may miss the true parent muon of $^{9}Li/^{8}He$. By varying the criteria of joint CD-WP muon tagging, the fitted residual $^{9}Li/^{8}He$ rate changed ~6%;

2) We varied ^{8}He-to-^{9}Li ratio in the fit, the changes in central values were found to be 2% level; 3) The data/MC difference in residual ^{9}Li/^{8}He rate is 11%. Taken these systematic errors into account, the overall assigned uncertainty is 35%, totally dominated by statistical uncertainty in the current dataset.

In the original manuscript, the quoted uncertainty was 34%. Upon recalculation of its components, this value was updated to 35%, a change primarily attributed to numerical rounding.

We have revised the manuscript accordingly in page 24: "Rate uncertainties were derived from the covariance matrix of the interval counts and validated using a bootstrap method that accounts for correlations within the dataset. A systematic uncertainty arises as the muon selection may miss the true parent of ^{9}Li/^{8}He, as assessed by varying the joint CD-WP muon tagging requirements which changed the fitted rate by 6%. Varying the ^{8}He-to-^{9}Li ratio by up to 10% shifts the central value by only 2%. The difference between the residual ^{9}Li/^{8}He rates in data and Monte Carlo is 11%. Combining these with the statistical uncertainty, an overall 35% uncertainty is assigned to the ^{9}Li/^{8}He background rate."

15.In Fig 10 it is not clear to me whether the red line prediction includes the off times mention or does not. If it does, then there is no need to dwell on the details of individual reactors turning on and off as it implies it is not already covered.

The red-line prediction is calculated using the actual thermal power data provided by the nuclear power plants. The dominant contribution comes from the antineutrinos emitted by the two Taishan reactors and six Yangjiang reactors. The thermal power of a reactor is set to zero when it is offline. This on/off information is critical, as it explains the corresponding variations in the predicted rate. For instance, a rough estimate based on the total thermal power change indicates that the predicted rate would decrease by approximately 11% or 17% if one Yangjiang or one Taishan reactor is turned off, respectively. Thus, we provide the on/off status of individual reactors to help readers understand the variations in the predicted rate.

Referee #6 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Dear Editor and Referees,

We would like to thank again the editor and the reviewers for their valuable time and professional feedback on our revised manuscript #2025-11-30874A. We have revised the manuscript accordingly and resubmitted it (tracking ID #2025-11-30874B) for your consideration, with all revisions marked in magenta. Our point-by-point responses to the reviewers' reports are provided below in blue.

Referees' comments:

Referee #1 (Remarks to the Author):

Dear Editor,

Thank you for forwarding the authors' rebuttal and the revised manuscript.

Let me state very clearly at the outset that the JUNO effort is exceptional. This experiment is technically extremely demanding, and the collaboration's capability in delivering such a dataset and analysis is, in my view, fundamental for the future of neutrino physics. The result reported here is important and timely.

From the perspective of a neutrino physicist, the paper was technically sound and publishable already in its original form, and it is even more clearly so in the revised version, following the authors' efforts to address the comments from all referees. I have no scientific objections, and I do not identify any issues that would prevent publication.

My initial suggestions focused on helping the manuscript speak to a broader Nature readership—i.e., improving accessibility, narrative focus, and contextual framing, in line with the very detailed guidance Nature provides to referees. If, however, the editorial view from the start was that the result's importance and urgency justified publication essentially in its original structure and level of specialization, it would have been perfectly reasonable to state that explicitly at the outset, rather than orienting the review process around particularly strict presentation-style expectations.

At this stage, I do not think it is useful to add further suggestions. The experimentalists in JUNO undoubtedly have more important things to do than iterating on stylistic reshaping, and I assume the same is true for the editorial office. I would be happy for you to proceed with publication in the form you consider most appropriate.

With best regards

Thank you for the excellent summary of the physics implications of the measurement, which has helped us extend the scope of the manuscript.

Referee #2 (Remarks to the Author):

We thank the authors for their detailed answers to our comments and questions. The authors' answers were satisfactory and the revised text greatly improved our understanding.

We would like to recommend that in specific cases, the answers to our questions be included in the text as we think those responses would strengthen the paper. In particular:

The distinction between using the data from Daya Bay ND for the flux constraint and primarily the Daya Bay FD (or ratio of FD/ND) for the oscillation parameters is an important clarification and should be included in the narrative.

To address this point, we have added the following clarification in the "Statistical methods" section:
"To ensure a rigorous treatment of external inputs, we distinguish between the data sources used for flux and oscillation constraints. In the fit, the reactor flux is constrained exclusively using Daya Bay near detector data, while the constraints on Δm^2_{31} and $\sin^2\theta_{13}$ are derived predominantly from the near-far relative measurements, without directly incorporating the underlying raw spectral data."
Also, the context has been adjusted accordingly.

The facts that the Daya Bay information on dm2_13 and sin2theta_13 is included as 2, 1d constraints and the correlation between the two is small is worth mentioning in the methods when details of the fit are described. The main text does mention the 1d constraints, but doesn't motivate why that is a valid implementation.

Based on the above revision, we have also restructured the explanation regarding the implementation of the χ^2_{osc} term, and added a clarifying sentence: "This is achieved using two separate one-dimensional penalty terms, an approach validated by the negligible correlation between the parameters."

We better understand the limitations on extending Fig. 11. However, the pulls currently shown could appear cherry-picked to demonstrate consistency. It would be useful to clarify that these correspond to the largest pulls from the nominal values that are common across the three analyses.

We have modified the sentence to clarify this point: "The distribution of nuisance-parameter pulls shown in Extended Data Fig. 7 (*Note: the old Fig.11*) represents the largest deviations from nominal values among the common parameters most strongly correlated with the fitted solar neutrino oscillation parameters, illustrating the stability of the fit and the agreement among the three analyses."

Finally, we have one small suggestion to help with lingering confusion over the numbers in table 1. The authors added a sentence to the caption of table 1 in response to one of our comments, but that extra sentence added confusion. We would suggest dropping that sentence in the caption and add to description of geoneutrinos and world reactor nu backgrounds on page 14 that the numbers quoted in the text must be further scaled down to account for the selection efficiency. Fundamentally, the reader wants to know how many of the 2379 selected events are expected to be background. Perhaps provide that number explicitly in the table for the reader and don't make them figure it out for themselves.

We thank the reviewer for this helpful suggestion to improve clarity. We have removed the confusing sentence from the Table 1 caption as requested, and in the main text there is one sentence to clarify that the rates for world reactor and geoneutrino backgrounds are scaled down by the selection efficiency.
Furthermore, we have added a row in Table 1 to show the total background rate. This would allow readers to directly compare the background with the observed signal rate (2379/59.1 ~ 40 cpd) .

Less critically, there are a few typos:

page 12, right column, 0th paragraph, 7th line there is an extra "." after the parenthetical

p17 the comma at the end of eq. 1 should be a full stop.

p18, left, first full paragraph, 2nd line, a comma is needed after WP

Otherwise, we are happy to recommend the paper for publication.

The typos have been corrected in the revised version.

Referee #3 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #4 (Remarks to the Author):

We want to thank the authors for their responses and updated manuscript.

These are regarding the responses provided by the authors

Regarding ”These results advance the basic understanding of neutrinos, validate the detector’s design, and confirm JUNO’s readiness for its primary goal of resolving the neutrino mass ordering with a larger dataset.” in the abstract.
The underlying facts are clear, but the current wording is not appropriate, and should be improved for clarity
Suggest “…and demonstrate encouraging progress toward JUNO’s primary goal of resolving the neutrino mass ordering with a larger dataset.” or “…and validate key aspects of the detector performance, supporting JUNO’s path toward its primary goal of resolving the neutrino mass ordering.”

The second-last sentence in the abstract has been revised following the suggestion: “These results advance the basic understanding of neutrinos, validate the detector’s design, and **indicate JUNO’s readiness for resolving the neutrino mass ordering** with a larger dataset.”

Regarding the “B-12” validation of the geometric volume fraction, the newly added clarification is good. It is better to add that the validation is done above 4 MeV energy.
Thanks for clarifying the meaning of “N” in modeling the reactor neutrino spectrum. The actual value of “N” used in the analysis should be provided.

The context has been revised into “an independent cosmogenic ^{12}B sample with energies above 4 MeV”.
The actual value of “N=25” has been provided in the Methods section.

Some additional comments regarding the revised manuscripts:
CP-violating phase → CP phase. Note this phase can lead to CP violation at certain values. Because of this, it is better to call it CP phase.
In the text above Fig. 4, “... parameters is statistically dominated …” → “statistically limited”. Given the current breakdown of statistical and systematic uncertainties, the data do not support the statement that the precision is statistically dominated. It would be more accurate to describe the measurement as statistically limited.
The newly added text in the “Discussion and outlook” regarding the 1.5 sigma difference between the reactor and solar neutrino dm^2_21 is an overmotivation. Suggest to revise the text as “Although the current difference between our reactor-based \Delta m^2_{21} result and the solar neutrino value [50] is only at the \sim1.5\sigma level, it motivates continued independent probes of this parameter.”

The context has been revised according to the suggestions, whereas we prefer to keep the “CP-violating phase” as it is more commonly used.

Referee #5 (Remarks to the Author):

I am satisfied with the responses to my review and have no further comments

Referee #6 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

This email has been sent through the Springer Nature Manuscript Tracking System NY-610A-SN&MTS

Beyond the referee's comments, we have noted a minor discrepancy in the reported uncertainty for Δm^2_{21}. While the unrounded values yield 1.55%, the relative uncertainty based on the quoted value $(7.50 \pm 0.12) \times 10^{-5}$ eV^2 is exactly 1.6%. We have therefore corrected this to 1.6% in the main text to maintain consistency. This change does not alter our conclusions.