nature portfolio

Peer Review File

Impact of environmental and operational stress on defect formation in China's high-speed rail subgrades

Corresponding Author: Professor Sen Li

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Professor Li,

Your manuscript titled "Operational stress and environmental interactions: unveiling multi-defect patterns in China's high-speed rail subgrades" has now been seen by 3 reviewers, whose comments are appended below. You will see that they find your work of some potential interest. However, they have raised quite substantial concerns that must be addressed. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version that fully addresses these serious concerns.

We hope you will find the reviewers' comments useful as you decide how to proceed. Should additional work allow you to address these criticisms, we would be happy to look at a substantially revised manuscript. Additionally, address properly these editorial thresholds:

***Provide novel and fully supported insight into the factors influencing high-speed rail subgrade defects in China.

****Clarify and support your hypotheses with existing literature and link them to existing theoretical principles.

***Outline your model and method in detail, including all assumptions and computational principles, and justify your choices over the other approaches, provide data preprocessing and validation documentation.

If you choose to take up this option, please either highlight all changes in the manuscript text file, or provide a list of the changes to the manuscript with your responses to the reviewers.

When resubmitting, please provide a point-by-point response to the reviewers' comments. Please submit your responses as a separate file, distinct from your cover letter where you can add responses to the Editors' comments that you do not want to be made available to the reviewers. Word files are preferred.

Important: The response to reviewers must not include any figures, tables or graphs. If you wish to respond to the reviewer reports with additional data in one of these formats, please add them to the main article or Supplementary Information, and refer to them in the rebuttal. Due to current technical limitations, any figures, tables, or graphs embedded in your rebuttal will not be included in the peer review file, if published.

Please bear in mind that we will be reluctant to approach the reviewers again in the absence of substantial revisions.

If the revision process takes significantly longer than three months, we will be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the reviewers' comments with a list of your changes to the manuscript text (which should be in a separate document to any cover letter), a tracked-changes version of the manuscript (as a PDF file) and any completed checklist:

Link Redacted

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

Please do not hesitate to contact us if you have any questions or would like to discuss the required revisions further. Thank you for the opportunity to review your work.

Best regards,

Sadia Ilyas
External Editor
Communications Earth & Environment

Martina Grecequet, PhD
Associate Editor
Communications Earth & Environment
@CommsEarth

EDITORIAL POLICIES AND FORMAT

If you decide to resubmit your paper, please ensure that your manuscript complies with our editorial policies and complete and upload the checklist below as a Related Manuscript file type with the revised article:

Editorial Policy Policy requirements
(Download the link to your computer as a PDF.)

For Manuscripts that fall into the following fields:
• Behavioural and social science
• Ecological, evolutionary & environmental sciences
• Life sciences
An updated and completed version of our Reporting Summary must be uploaded with the revised manuscript
You can download the form here:
https://www.nature.com/documents/nr-reporting-summary.zip

For your information, you can find some guidance regarding format requirements summarized on the following checklist: (https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf) and formatting guide (https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf).

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

In this paper, the structural equation modelling (SEM) was used to complete the factors of common subgrade defects in high speed railways and the mechanisms underlying their co-occurrence risk. The main concern of the reviewer is the novelty and usefulness of this work. Although the topic of this work is interesting, the employing of SEM method for assessing high speed railway subgrade defects is not methodologically innovative and the authors did not propose new models or major improvements to existing theories. Furthermore, modelling approaches in manuscripts that only consider the railway trackbed and its surroundings while ignoring the impact of the track structure are also difficult to be accepted by the reviewers. Consequently, I do not think this manuscript is suitable for publication in Communications Earth & Environment. Some suggestions are given for authors to consider.
My major concerns are the following:
1.The track structure, roadbed and foundation comprise the high-speed railway roadbed structure, but neither the effect of track unevenness nor the effect of train moving load is considered directly in this manuscript, but only the train moving load is defaulted to be the internal variable of the operating frequency. In order to adopt such a complex model but ignore the main influencing factors? This confused the reviewer. In addition, the actual engineering in the construction of railway in poor soil environment, will certainly take foundation reinforcement to ensure the foundation bearing capacity. In this case, the composite foundation cannot be evaluated only by using the geomorphological and geological variables in the manuscript, which is not consistent with the actual situation.

2.What are the assumptions of the SEM model in section 2.3.2 of this paper? And what are the computational principles? Is it just 3 simple linear equations? The assumptions of the models used as well as the principles should be given in concise language when the paper is written, otherwise it will confuse the reader.
3.In section 3 Results part of this paper, why is the R2 not more than 0.35 and how to prove the reliability and significance of the model analysis results? In addition to that, Figure 3 is obviously generated by GPT, which is totally inconsistent with the actual situation. This makes the reviewer sceptical about the results.
4.In this manuscript, the conclusions obtained are qualitative analyses and almost exclusively results that are widely accepted. The authors apply a complex model from the field of statistics, but neither quantitative results nor novel insights are given. How practical is the research content of this manuscript?

Reviewer #2 (Remarks to the Author):

Strengths:
1. Relevance and Timeliness: The study is timely, addressing the pressing issue of infrastructure resilience in the face of climate change and increased HSR operations.
2. Methodology: The use of SEM is appropriate for examining the complex interactions between various factors influencing subgrade defects.
3. Comprehensive Analysis: The study considers both natural and operational factors, providing a holistic view of the issue.
4. Practical Implications: The findings have significant practical implications for maintenance and infrastructure planning, highlighting the need for interventions to enhance resilience.

Weaknesses:
1. Clarity and Specificity: While the abstract provides a broad overview, it lacks specificity in certain areas. For example, the exact nature of the SEM analysis and the specific variables considered could be briefly mentioned.
2. Balance of Information: This could benefit from a more balanced distribution of information. While the impacts of different types of land are mentioned, there is less detail on the operational characteristics of the HSR that influence subgrade defects.

Comments:
1. Citing numerical facts: The authors are suggested to cite wherever anticipated or factual numerical values are reported.
2. Complete definition of DST: The analysis of this study is based on disaster system theory but there is not enough discussion and definition given on this. Authors are suggested to discuss how their approach i) identifies and assess risks within a system; ii) effectively inform on emergency planning; iii) guide recovery and reconstruction efforts by considering the long-term impacts and interactions within the system, and iv) policies that address the complexities of disaster management.
3. Mathematical representations: Disaster system theory could be better explained using i) differential equation state space; ii) network models; iii) agent-based models; iv) stochastic models, and v) objective-based optimisation models. Authors should consider implementing and discussing one of the aforementioned assessments.
4. Figures: Figure 1 is incorrect and its caption should be detailed.
i) Use Internationally Recognized Boundaries: Map to reflect internationally recognized boundaries.
ii) Provide Clear Annotations: If the map is discussing disputed areas, clearly annotate these regions with a note explaining that the area is disputed
5. Clear definition: The authors stated (H1) to (H15) but have not vividly defined it in the manuscript or used anywhere in its model. Use (H[.]) as feature vector components of multi-defect pathways to clearly define the relationship, cause and effect as matrices.
6. Software: Lulutong software lacks integration capabilities and scalability in comparison to SAP TM, Oracle Transport Management, JDA, and others. How do the authors overcome the drawback and ensure their reported results are correct.
7. Conclusion: Conclusion is overclaiming ‘the critical factors for each defect were identified’. The identified factors do not give complete picture and proof of being critical factor. The manuscript should focus more on this. What is the probability that these factors in defect causes, direct impact, and risk occurrence. Decision making should be supported with well-trained data and evidence

Reviewer #3 (Remarks to the Author):

Please refer the attached file. I have given my observations in detail. Kindly refer POINTS to ADDRRESS sections.

Communications Earth & Environment is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as ‘corresponding author’ create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the Manuscript Tracking System by clicking on ‘Modify my Springer Nature account’ and following the instructions in the link below. Please also inform all co-authors that they can add their ORCIDs to their accounts and that they must do so prior to acceptance.
https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research

For more information please visit http://www.springernature.com/orcid

If you experience problems in linking your ORCID, please contact the Platform Support Helpdesk.

Version 1:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Professor Li,

Your revised manuscript titled "Operational stress and environmental interactions: unveiling multi-defect patterns in China's high-speed rail subgrades" has now been seen by 3 reviewers, whose comments are appended below. You will see that they find your work of some potential interest. However, they continue raise quite substantial concerns that must be addressed. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version that fully addresses these serious concerns.

We hope you will find the reviewers' comments useful as you decide how to proceed. Should additional work allow you to
• address these criticisms (that is, either to incorporate the suggestions or provide a compelling argument why the point made by the reviewer is not valid or relevant to the editorial threshold as outlined below)
AND
• meet our editorial thresholds as outlined below,

then we would be happy to look at a revised manuscript.

In the following, we list our minimum requirements for publication.

1. Provide full support for your claims by your data and analysis.
2. Address the poor fit of your model and consider a non-linear fit.
3. Add or clearly justify omissions of processes such as dynamic load and vibration.
4. Address low model accuracy and unexplained variances in the model.

***In addition, please follow our guidelines for using Artificial Intelligence: https://www.nature.com/nature-portfolio/editorial-policies/ai

We hope you will find the reviewers' comments useful as you decide how to proceed. Should additional work allow you to address these criticisms, we would be happy to look at a substantially revised manuscript. If you choose to take up this option, please either highlight all changes in the manuscript text file, or provide a list of the changes to the manuscript with your responses to the reviewers.

When resubmitting, please provide a point-by-point response to the reviewers’ comments. Please submit your responses as a separate file, distinct from your cover letter where you can add responses to the Editors’ comments that you do not want to be made available to the reviewers. Word files are preferred. We recommend that any figures, tables or graphs that are included in the response to reviewers are also included in the main article or Supplementary Information.

Please bear in mind that we will be reluctant to approach the reviewers again in the absence of substantial revisions.

If the revision process takes significantly longer than three months, we will be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the reviewers’ comments with a list of your changes to the manuscript text (which should be in a separate document to any cover letter), a tracked-changes version of the manuscript (as a PDF file) and any completed checklist:

Link Redacted

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

Please do not hesitate to contact us if you have any questions or would like to discuss the required revisions further. Thank you for the opportunity to review your work.

Best regards,

Sadia Ilyas
External Editor
Communications Earth & Environment

Martina Grecequet, PhD
Senior Editor
Communications Earth & Environment

EDITORIAL POLICIES AND FORMAT

If you decide to resubmit your paper, please ensure that your manuscript complies with our editorial policies and complete and upload the checklist below as a Related Manuscript file type with the revised article:

Editorial Policy Policy requirements
(Download the link to your computer as a PDF.)

For Manuscripts that fall into the following fields:
• Behavioural and social science
• Ecological, evolutionary & environmental sciences
• Life sciences
An updated and completed version of our Reporting Summary must be uploaded with the revised manuscript
You can download the form here:
https://www.nature.com/documents/nr-reporting-summary.zip

For your information, you can find some guidance regarding format requirements summarized on the following checklist:
(https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf) and formatting guide
(https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf).

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

All the mentioned issues have been addressed. I recommend that this paper be accepted for publication.

Reviewer #2 (Remarks to the Author):

1) Some contents including images have been generated using Generative Pretrained Transformers.
2) Figure captions lack proper reference to subfigures. It is hard to follow subfigure details.
3) A low R squared value indicates that the fitting is poor. More non-linear fits need to be applied to further increase the goodness of fit.
4) Although the authors have added supplementary Figures, there are issues concerning correct captioning. Referring to Fig 1. and supplementary Figure 2, the caption should include 'disputed regions' and 'based on specifications for the content of public maps in China' since the readers of the article are not limited to a region.
5) Dynamic load and vibration analysis is a major part of operational stress. The article does not cover this. This could be added to manuscript or in limitations and future works.
6) PINN and digital twin can provide a more depth and practical scenario. This could also be included in future scope.
7) The unexplained variances in model needs to be reduced using feature selection and optimisation method to improve model accuracy. This should also include details on handling noise.

Reviewer #3 (Remarks to the Author):

I am happy with the revisions administered by the authors. However, I would like to see sustainable implication and societal or community development by the findings of the research. Keeping a trend for Sustainable development, authors are advised to keep a section for this. It will not only enhance the overall paper's readability but also can gather substantial readership in near future.

So, I would suggest addition of these implications will give a bost to the paper.

Version 2:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Professor Sen Li,

Your manuscript titled "Operational stress and environmental interactions: unveiling multi-defect patterns in China's high-speed rail subgrades" has now been seen by our reviewers, whose comments appear below. In light of their advice we are delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment.

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers. At the same time we ask that you edit your manuscript to comply with our format requirements and to maximize the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

The authors should discuss the limitations of the model and its results, and suggest potential directions for future research in the conclusion section.

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*****Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *****

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at https://www.nature.com/documents/commsj-file-checklist.pdf .

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit https://www.nature.com/commsenv/open-access

At acceptance, you will be provided with instructions for completing the open access license agreement on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:
Link Redacted
** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Sadia Ilyas
External Editor
Communications Earth & Environment

Martina Grecequet, PhD
Senior Editor
Communications Earth & Environment
@CommsEarth

Yann Benetreau, PhD
Consulting Editor, Communications Earth & Environment
Deputy Editor, Communications Sustainability
Nature Portfolio
ORCID: 0000-0002-1897-0887
New York Office

REVIEWERS' COMMENTS:

Reviewer #2 (Remarks to the Author):

The authors have revised the article thoroughly. No further comments.

In this paper, the structural equation modelling (SEM) was used to complete the factors of common subgrade defects in high speed railways and the mechanisms underlying their co-occurrence risk. The main concern of the reviewer is the novelty and usefulness of this work. Although the topic of this work is interesting, the employing of SEM method for assessing high speed railway subgrade defects is not methodologically innovative and the authors did not propose new models or major improvements to existing theories. Furthermore, modelling approaches in manuscripts that only consider the railway trackbed and its surroundings while ignoring the impact of the track structure are also difficult to be accepted by the reviewers. Consequently, I do not think this manuscript is suitable for publication in Communications Earth & Environment. Some suggestions are given for authors to consider.

My major concerns are the following:

1. The track structure, roadbed and foundation comprise the high-speed railway roadbed structure, but neither the effect of track unevenness nor the effect of train moving load is considered directly in this manuscript, but only the train moving load is defaulted to be the internal variable of the operating frequency. In order to adopt such a complex model but ignore the main influencing factors? This confused the reviewer. In addition, the actual engineering in the construction of railway in poor soil environment, will certainly take foundation reinforcement to ensure the foundation bearing capacity. In this case, the composite foundation cannot be evaluated only by using the geomorphological and geological variables in the manuscript, which is not consistent with the actual situation.

2. What are the assumptions of the SEM model in section 2.3.2 of this paper? And what are the computational principles? Is it just 3 simple linear equations? The assumptions of the models used as well as the principles should be given in concise language when the paper is written, otherwise it will confuse the reader.

3. In section 3 Results part of this paper, why is the R2 not more than 0.35 and how to prove the reliability and significance of the model analysis results? In addition to that, Figure 3 is obviously generated by GPT, which is totally inconsistent with the actual situation. This makes the reviewer sceptical about the results.

4. In this manuscript, the conclusions obtained are qualitative analyses and almost exclusively results that are widely accepted. The authors apply a complex model from the field of statistics, but neither quantitative results nor novel insights are given. How practical is the research content of this manuscript?

REVIEW FOR COMMSENV-24-1535

PAPER SUMMARY

investigate the stability of high-speed rail (HSR) subgrades in China under increasing operational frequency and extreme weather events. This study addresses a significant gap in understanding complex interactions and causality between environmental factors and HSR subgrade defects. Given the planned expansion of China's HSR to 50,000 kilometres by 2025, understanding these interactions is crucial for infrastructure resilience and safety. The authors employ structural equation modelling (SEM) to explore the factors contributing to common subgrade defects such as settlement, frost damage, uplift deformation, and mud pumping. The paper aims to identify key influencing variables and mechanisms underlying the co-occurrence of these defects, incorporating geomorphological variables, land use, climate conditions, and operational characteristics of HSR. The main findings indicate that unused lands like sandy areas, Gobi deserts, bare lands, and bare rocks significantly affect the co-occurrence of subgrade defects. Geomorphological variables indirectly influence subgrade defects by impacting HSR operational characteristics, with bare land most significantly affecting settlement and uplift deformation. Further, frost damage and mud pumping are primarily driven by annual freezing days and average annual rainfall. These results highlight the need for integrated management strategies to enhance maintenance and infrastructure resilience. This study advances the understanding of HSR subgrade defects' mechanisms, providing a scientific basis for future HSR maintenance planning.

Review Synopsis

a. The article presents a potentially significant study using structural equation modelling (SEM) to examine the factors influencing high-speed rail (HSR) subgrade defects in China. However, several critical issues must be addressed to ensure the study's validity and rigour.

b. Firstly, the theoretical framework is inadequately integrated. The Disaster System Theory is mentioned but not sufficiently grounded in seminal works such as Mileti et al. (1999). The hypotheses need clearer links to established theoretical principles, supported by a robust literature review that includes alternative hypotheses and justifies chosen pathways.

c. Secondly, the data quality and preprocessing steps lack transparency. The selection criteria, temporal range, and handling of extreme values or outliers are insufficiently detailed. Additionally, the validation and cleaning processes are not described, raising concerns about potential biases and the dataset's reliability. Comprehensive documentation of these steps is crucial.

d. Thirdly, the SEM approach requires improvement. Model specification is under-defined and lacks a comprehensive diagram and detailed equations. The interpretation of model fit indices and path significance is superficial, and diagnostic tests for residuals and multicollinearity are absent. A thorough justification of variable selection, path modifications, and detailed sensitivity analyses are needed to ensure model robustness.

e. The authors should ground their hypotheses in a well-articulated theoretical framework with comprehensive literature support to enhance the study. Detailed documentation of data preprocessing and validation steps is essential. The SEM approach must include exhaustive model specification, fit assessment, and diagnostic testing. These enhancements will significantly improve the study's validity, interpretation, and contribution to the field.

Detail Review

1. Appropriateness of Hypotheses:

Theoretical Framework: The article references Disaster System Theory but fails to incorporate seminal works like Mileti et al. (1999) or subsequently influential papers that thoroughly establish the theory's relevance to HSR subgrade defects.

The manuscript mentions, "This study draws on the Disaster System Theory to analyze the factors influencing HSR subgrade defects in China."

Points to address: Incorporate and cite seminal works and influential papers on Disaster System Theory. Explain how this theory explicitly applies to the hypothesized relationships among geomorphological variables, land use, climate variables, geological variables, and HSR operational characteristics. Elucidate how the mediation pathways logically extend from these theoretical principles.

Literature Review: The literature review inadequately grounds the hypotheses in a comprehensive, up-to-date body of research. Fundamental studies are selectively cited without enough context, and potential alternative hypotheses are not sufficiently acknowledged.

The manuscript mentions, "Prior studies have concentrated on the direct impacts of natural and anthropogenic factors on railway infrastructure, often overlooking their complex interactions and causality."

Points to address: Conduct a thorough literature review that includes a broad spectrum of relevant, high-calibre studies. Ensure accurate interpretation and contextualization of these studies within the field. Acknowledge and discuss potential alternative hypotheses or theories and justify the chosen hypotheses in light of these alternatives.

Specificity and Testability: The hypotheses lack specificity in predicted relationships and expected directionalities, making empirical testing through SEM challenging.

The manuscript mentions, "From a perspective of direct impact, the disaster-prone environment (including climate variables, geomorphological variables, land use, HSR

proximity index) and disaster-causing factors (such as geological variables, HSR operational characteristics) may all potentially have direct effects on HSR subgrade defects."

Points to address: Define each hypothesis with explicit predicted relationships and expected directionalities (positive or negative influences). Ensure that all variables are measurable and articulated for empirical testing. Use precise language to describe how each factor interacts with subgrade defects, facilitating precise modeling in the SEM.

2. ***Quality and Preparation of Data:***

Data Sources and Sampling: The data collection process lacks detail, particularly regarding selection criteria and temporal range, which raises concerns about potential biases or gaps

The manuscript mentions, "Subsequently, 661 georeferenced event records of eight defect types were selected, crossing provincial, municipal, county, township, and smaller scales."

Points to address: Provide detailed information on the selection criteria for event records, including temporal and spatial dimensions. Justify the representativeness of the dataset, ensuring it encompasses a sufficiently diverse set of conditions to mitigate biases. Offer a rationale for data inclusion and discuss any limitations or potential gaps in the dataset.

Data Preprocessing: Moon's method of normalization lacks a thorough rationale, and there is no mention of handling extreme values or outliers.

The manuscript mentions, "We adopted Moon's method of using the normalize module to standardize the factors, eliminating the impact of differing magnitudes on analysis results."

Points to address: Provide a comprehensive rationale for choosing Moon's method over other normalization techniques. Describe how extreme values and outliers are identified and handled. Ensure clear documentation of the ArcGIS processing steps and include justification and verification of the mileage calculations.

Data Validation and Cleaning: The article lacks explicit steps for data validation and cleaning, and there is no evidence of cross-validation with external datasets.

The manuscript mentions, "This critique refers to a general omission in the manuscript. Thus, a direct quote is not provided."

Points to address: Outline and document the steps taken for data cleaning, including handling missing values and inconsistencies. Provide evidence of cross-validation with external datasets or independent sources to ensure data reliability. Include quality control measures, such as error estimates for georeferencing and enhancing dataset robustness.

3. ***Structural Equation Modeling Approach:***

Model Specification: Justification for selecting variables and excluding others is inadequate, and the SEM diagram and equations are under-specified.

The manuscript mentions, "Using SEM, we can significantly enhance the understanding of defect formation mechanisms, accurately identify key influencing variables, and effectively support disaster risk assessment and management decisions."

Points to address: Justify the selection of variables and provide a rationale for excluding others. Include a comprehensive SEM diagram and detailed equations representing the relationships among the variables. Ensure these align with the hypotheses and bolster the model's logical coherence.

Model Fit Indicators: The manuscript lacks a detailed interpretation of model fit indices and does not disclose any path modifications or their justifications.

The manuscript mentions, "In the structural model for subgrade defects, we...retained paths H1, H2, H6, H7, H8, H10, H11, and H13."

Points to address: Report model fit indices explicitly against established benchmarks (e.g., RMSEA < 0.08, CFI > 0.90). Disclose any modifications to path structures, providing justifications and discussing the impact on the overall model fit. Ensure transparency in adjustments made to improve the model.

Path Coefficients and Significance Testing: The significance tests for path coefficients lack detail, and the manuscript does not include sensitivity analyses or robustness checks.

The manuscript mentions "...factors listed in the supplementary materials Table 1."

Points to address: Provide detailed significance levels and confidence intervals for path coefficients. Perform and document sensitivity analyses and robustness checks to demonstrate the stability of the SEM results. Interpret these coefficients in light of their significance and overall model implications.

4. Statistical Analysis and Robustness:

Mediation and Bootstrapping Analysis: The mediation effect testing and bootstrapping procedures lack detailed procedural descriptions and rationale.

The manuscript mentions "Initially, the mediation effect code is tested; subsequently, the Bootstrap method is employed to perform 5000 resamples."

Points to address: Include a detailed explanation of the mediation effect testing procedures and justify the selection of 5000 resamples in the bootstrap method. Describe the VB programming code used in AMOS, ensuring it is accessible for replication. Offer a thorough discussion of the implementation of bias-corrected and percentile methods.

Residuals and Diagnostics: The manuscript does not mention diagnostic tests or residual analysis to detect model reliability issues.

The manuscript mentions "This critique refers to a general omission in the manuscript, thus, a direct quote is not provided."

Points to address: Perform and document diagnostic tests, including residual plots, variance inflation factors (VIF), and other measures to detect heteroscedasticity, autocorrelation, or multi-collinearity. Discuss the results of these diagnostics and their implications for the model's robustness.

Reproducibility: The study lacks sufficient detail for replication, including methodological steps, programming scripts, or data sources.

The manuscript mentions "The results of this study will provide a scientific basis for the maintenance of the railway system, thereby enhancing the safety and efficiency of HSR."

Points to address: Provide comprehensive methodological documentation, including clear descriptions of all steps and programming scripts used in the analysis. Ensure datasets and codes are made accessible through supplementary materials or repositories, facilitating independent replication.

5. Interpretation of Results and Generalizability:

Contextual Interpretation: The results are moderately interpreted within the specific conditions of China's HSR, but practical implications for HSR maintenance and operational strategies are underexplored.

The manuscript mentions "These findings enhance our understanding of the complexity of HSR subgrade defects and underscore the necessity for integrated management strategies."

Points to address: Delve deeper into the practical implications for HSR maintenance and operational strategies. Revisit the theoretical framework in the discussion to link the findings back to foundational principles and previous research, highlighting how this study advances understanding in the field.

Generalizability: The study does not adequately address the broader applicability of the findings, particularly to regions outside China or similar infrastructure systems.

The manuscript mentions "This critique refers to a general omission in the manuscript, thus, a direct quote is not provided."

Points to address: Discuss the geographic and temporal scope of the dataset and its broader applicability. Evaluate the potential for extrapolation to other regions or similar infrastructure systems and acknowledge limitations regarding generalizability. Propose future research directions to confirm and expand upon these findings.

Implications and Recommendations: Policy recommendations lack empirical grounding, and the discussion on infrastructure resilience and future trends is brief and underdeveloped.

The manuscript mentions "The findings underscore the necessity for integrated management strategies."

Points to address: Expand on the policy recommendations with a detailed exploration of the empirical evidence presented. Discuss infrastructure resilience and future climate and usage trends in-depth, providing actionable, data-driven recommendations. Highlight the anticipated impact on HSR safety, efficiency, and maintenance, supported by robust empirical findings.

Misc. Points to revise

1. The comprehensiveness and representativeness of the 661 georeferenced event records are inadequately addressed. The study does not offer a detailed data collection methodology, including selection criteria and temporal range. There is insufficient assessment of geographical diversity to ensure a broad representation of environmental and operational conditions. Potential biases in the data collection process are neither identified nor mitigated.

The manuscript mentions "Subsequently, 661 georeferenced event records of eight defect types were selected, crossing provincial, municipal, county, township, and smaller scales."

Points to address: Provide a detailed methodology section specifying the criteria for selecting the 661 event records.- Ensure a broad temporal range is covered to represent various environmental conditions over time.- Justify the geographical diversity of the selected records to cover a wide range of environmental and operational conditions.- Identify and discuss potential biases in the data collection process and the steps taken to mitigate them.

2. The normalization process using Moon's method lacks detailed justification. There is no information on how extreme values or outliers were managed, which could potentially distort relationships among variables. Details on the preprocessing steps, mainly the split intersecting lines function in ArcGIS, are insufficiently documented, raising concerns about reproducibility.

The manuscript mentions "We adopted Moon's method of using the normalize module to standardize the factors, eliminating the impact of differing magnitudes on analysis results."

Points to address: Justify the choice of Moon's method over other normalization techniques suitable for this dataset. Explicitly describe how extreme values or outliers in the dataset were identified and managed. Provide a detailed step-by-step account of the preprocessing procedures, ensuring that all methods, including the ArcGIS split intersecting lines function, are accurately documented and reproducible.

3. The article does not outline steps for validating and cleaning the data. There is no evidence of procedures to handle missing values or inconsistencies. Additionally, the manuscript lacks proof of robust quality control measures, such as cross-validation with external datasets or triangulation from multiple sources, which are crucial for ensuring data reliability and accuracy.

The manuscript mentions "This critique refers to a general omission in the manuscript, thus, a direct quote is not provided."

Points to address: Outline the steps taken to validate and clean the dataset. Describe approaches for handling missing values and inconsistencies within the dataset. Implement and document robust quality control measures, such as cross-validation with external datasets and triangulation from multiple data sources, to ensure the dataset's reliability and accuracy. Provide examples or references to established methodologies used in these processes.

4. The SEM includes a range of latent variables, but the justification for choosing specific variables over others remains insufficient. The specification of observed variables is not comprehensively discussed, which raises concerns about their ability to capture the constructs of interest altogether.

The manuscript mentions "Using SEM, we can significantly enhance the understanding of defect formation mechanisms, accurately identify key influencing variables and effectively support disaster risk assessment and management decisions."

Points to address: Clearly articulate the rationale for selecting the particular latent and observed variables used. Provide a more detailed SEM diagram to visualize all the relationships and pathways considered in the model. Ensure that the selected variables comprehensively capture all key constructs of interest. Explain any exclusions explicitly and justify them based on theory or prior research.

5. The article reports model fit indices but lacks thorough interpretation against established thresholds. There is no mention of any modifications made to improve the model fit, nor is there a discussion of their justifications and impact.

The manuscript mentions "We eliminated the latent variable of geological factors as they had minimal impact on the occurrence of subgrade defects and did not improve the model fit."

Points to address: Report the model fit indices explicitly against established benchmarks (e.g., Chi-square, RMSEA < 0.08, CFI > 0.90).- Discuss any modifications made to the model, providing clear justifications and evaluating their impact on the overall model fit.- Include the interpretation of each fit statistic and explain why they indicate that the model is a good fit or what additional improvements could be made to meet the desired thresholds.

6. The significance tests for path coefficients are mentioned but not detailed enough to evaluate thoroughly. There is also a lack of sensitivity analyses or robustness checks to ensure the stability of the SEM results.

The manuscript mentions "In the structural model for subgrade defects, we estimated and tested the hypothesized paths...while retaining paths H1, H2, H6, H7, H8, H10, H11, and H13 which all met the significance criterion (P<0.05)."

Points to address: Provide detailed significance levels and confidence intervals for each path coefficient. Perform and document sensitivity analyses and stability checks to demonstrate the robustness of the results across different model specifications and assumptions. Include interpretations of the direct, indirect, and total effects, ensuring that the significance tests are correctly applied and explained within the context of the structural model.

7. The article mentions the use of mediation effect testing and bootstrapping with 5000 resamples, but lacks detailed procedural descriptions and rationale, casting doubt on the appropriateness and correctness of implementation.

The manuscript mentions "This paper uses VB programming within AMOS to write mediation effect testing code...subsequently, the Bootstrap method is employed to perform 5000 resamples."

Points to address: Provide a comprehensive description of the mediation effect testing procedure, including the specific mediation models tested and how VB programming was utilized in AMOS. Justify the choice of 5000 resamples in bootstrapping, explaining why this number is sufficient to ensure statistical robustness.- Include detailed results of the bias-corrected and percentile methods, ensuring the Lower-Upper interval does not include zero for significant pathways. To illustrate these results, offer examples or visualizations (e.g., confidence interval plots).

8. There is no discussion on residuals or diagnostic statistics. Hence, potential issues like heteroscedasticity, autocorrelation, or multi-collinearity are unaddressed, compromising the model's robustness.

The manuscript mentions "This critique refers to a general omission in the manuscript, thus, a direct quote is not provided."

Points to address: Conduct residual analysis, including residual plots, to check for heteroscedasticity or autocorrelation patterns. Calculate and report variance inflation factors (VIF) to assess and address multi-collinearity. Include any additional diagnostic tests used in SEM to evaluate the goodness-of-fit and robustness of the model. Discuss the implications of these diagnostics and any corrective measures taken.

9. The article lacks sufficient replication details, including clear descriptions of methodological steps and access to programming scripts or datasets.

The manuscript mentions "Thus, SEM synthesizes theory and data to analyze the direct and indirect effects between variables systematically."

Points to address: Provide exhaustive methodological documentation that details every step of the SEM analysis, from data preprocessing to final outputs.- Ensure all AMOS scripts, VB code, and any other programming used are included as supplementary materials or housed in an accessible repository.- Offer access to the datasets analyzed, or at least detailed descriptions of how the data can be acquired or constructed. This transparency is crucial for verifying results through independent replication.

10. The results are somewhat interpreted within China's HSR environment, yet the practical implications for maintenance and operational strategies are not thoroughly explored. The discussion often remains high-level without drilling into actionable insights that maintenance engineers or policymakers could directly apply.

The manuscript mentions "These findings enhance our understanding of the complexity of HSR subgrade defects and underscore the necessity for integrated management strategies."

Points to address: Expand the discussion to include specific, actionable insights into HSR subgrade maintenance and operational strategies. Relate findings directly back to operational decisions such as scheduling, maintenance intervals, and design modifications. Draw explicit connections to the theoretical framework and previous research to ground these practical implications.

11. The study fails to thoroughly examine the generalizability of the findings beyond the specific dataset and contextual conditions of China's HSR. There is little discussion on whether these results can be applied to other geographic regions or similar types of railway infrastructure systems.

The manuscript mentions "This critique refers to a general omission in the manuscript, thus, a direct quote is not provided."

Points to address: Discuss the potential for extrapolating findings to other regions or similar HSR systems. Include a geographic, temporal, and operational scope analysis, specifying how these factors might influence defect patterns elsewhere. Acknowledge the study's transferability limitations and suggest future research to test these findings in different contexts, ensuring a clearer understanding of their broader applicability.

12. Although policy recommendations are presented, they are not sufficiently grounded in empirical evidence. The discussion on infrastructure resilience and future trends is superficial and lacks in-depth analysis.

The manuscript mentions "The findings underscore the necessity for integrated management strategies."

Points to address: Provide more empirically grounded and detailed policy recommendations. Include specific strategies for infrastructure resilience, such as predictive maintenance schedules based on detailed environmental risk assessments. Discuss the impact of future climate changes and increased HSR usage on subgrade defects more comprehensively. Use data-driven insights to support these recommendations and detail how these strategies could be practically implemented in HSR operational plans and maintenance protocols.

So, in my observation, I recommend MAJOR REVISION.

Response letter

The authors would like to thank the three reviewers for their time and constructive comments on the manuscript. The manuscript entitled ‘Operational stress and environmental interactions: unveiling multi-defect patterns in China’s high-speed rail subgrades’ was submitted for consideration in *Communications Earth & Environment* (COMMSENV-24-1535A).

In general, several major changes have been made:
- We have extended and strengthened our reference on the Disaster System Theory (DST) to explicitly explain how the theory applies to the hypothesized relationships between variables and defects.
- We have provided a detailed overview of the modeling methods, including model construction methodology, high-speed rail (HSR) mileage calculation methods, model fit indices, mediation effect validation, and testing methods.
- We have now improved our descriptions of data analytic steps including methods applied for handling data outliers, extreme values, and missing values.
- We have reexamined and addressed factor multicollinearity, making adjustments to the independent defect models.
- We have expanded the discussion on the application of the Structural Equation Modeling (SEM) in different infrastructure or regional contexts.

We have also edited the text thoroughly to correct language errors and improve the flow of arguments.

Detailed response to each comment of the three reviewers is attached below.

In response to Reviewer #1’s comments

In this paper, the structural equation modelling (SEM) was used to complete the factors of common subgrade defects in high-speed railways and the mechanisms underlying their co-occurrence risk. The main concern of the reviewer is the novelty and usefulness of this work. Although the topic of this work is interesting, the employing of SEM method for assessing high speed railway subgrade defects is not methodologically innovative and the authors did not propose new models or major improvements to existing theories. Furthermore, modelling approaches in manuscripts that only consider the railway trackbed and its surroundings while ignoring the impact of the track structure are also difficult to be accepted by the reviewers. Consequently, I do not think this manuscript is suitable for publication in Communications Earth & Environment. Some suggestions are given for authors to consider.
My major concerns are the following:
[Response 1.1]: Thank you for your thorough review and valuable feedback on our paper. We appreciate your time and the insights you’ve provided. In response to the primary concerns you raised, we have made several significant revisions to enhance the novelty, practicality, and comprehensiveness of our study:

First, we have now enhanced novelty and methodological contribution. While the SEM model is not a new method in statistical analysis, its application to large-scale empirical data analysis of high-speed railway subgrade defects is, to our knowledge, unprecedented. Our study uniquely integrates multiple environmental and operational factors, such as climate variables, geomorphology, land use types, and operational characteristics, into a single SEM framework. This holistic approach allows us to uncover complex interactions and indirect effects that have not been explored in previous studies. Moreover, we introduced previously underrepresented variables, such as the impact of unused land types (e.g., sandy lands, bare lands, Gobi deserts), on the co-occurrence of subgrade defects. This brings a new perspective to understanding how specific land use categories affect HSR subgrade stability. Furthermore, our analysis reveals indirect relationships, such as how geomorphological features influence subgrade defects through their impact on operational characteristics like train speed and frequency. This adds depth to the theoretical understanding of defect formation mechanisms.

Second, we have already incorporated track structure into the study. We acknowledge the importance of considering the track structure's impact on subgrade defects. Our SEM included variables representing track structure conditions, i.e., track unevenness. Our dataset includes 182 records specifically related to track structure unevenness out of 661 total defect records. These records were based on strict guidelines from China's High-speed Railway Design Specifications (TB 10621-2017) and Railway Technical Management Regulations, reflecting the potential impact of track irregularities on subgrade defects. Accordingly, to avoid any misunderstanding, we have added a description about track structure in §2 (Theoretical framework) of the main text. Additionally, we discussed in detail the factors influencing train moving loads, such as train speed and environmental conditions (e.g., temperature variations and rainfall), all of which are reflected in the model, thereby enhancing its comprehensiveness and practical applicability.

Third, to enhance the applicability and expand the scope of the study, we have provided a detailed discussion on the potential for applying our model in different regions and infrastructure types. We elaborated on how geomorphological features and land use can influence the occurrence of subgrade defects and explained that even with foundation reinforcement, these factors could still be the fundamental causes of such defects. Additionally, we included real-world engineering case studies, such as the use of machine learning algorithms on the Zamudio railway bridge in the Basque region of Spain, which further demonstrates the practical significance and potential application of our research. This expanded discussion highlights the broader applicability of our findings to diverse geographical regions and infrastructure systems, while also acknowledging the limitations and suggesting future research directions to validate and extend these findings in varied contexts.

Fourth, to enhance data transparency and reproducibility, we have uploaded the complete defect dataset to the figshare Generalist repository and provided a DOI link for access. This move not only enhances the credibility of our research but also provides foundational data support for further studies by other researchers.

With these improvements and additions, we believe that the novelty and practicality of the paper have been significantly enhanced, which better addresses your concerns. If you have any further comments or suggestions, we are more than willing to make additional improvements. Thank you again for your review and feedback!

The track structure, roadbed and foundation comprise the high-speed railway roadbed structure, but neither the effect of track unevenness nor the effect of train moving load is considered directly in this manuscript, but only the train moving load is defaulted to be the internal variable of the operating frequency. In order to adopt such a complex model but ignore the main influencing factors? This confused the reviewer. In addition, the actual engineering in the construction of railway in poor soil environment, will certainly take foundation reinforcement to ensure the foundation bearing capacity. In this case, the composite foundation cannot be evaluated only by using the geomorphological and geological variables in the manuscript, which is not consistent with the actual situation.
[Response 1.2]: Thank you for your valuable suggestions. We have indeed considered factors related to track unevenness in our study. Specifically, we analyzed 661 records of HSR subgrade defects, of which 182 records specifically pertain to track structure unevenness. These records were incorporated into the dataset following strict guidelines from China's High-speed Railway Design Specifications (TB 10621-2017) and Railway Technical Management Regulations. This reflects the potential impact of track irregularities on subgrade defects. Accordingly, to avoid any misunderstanding, we have added a description in §2 (Theoretical framework) of the main text. As described, "We compiled an extensive georeferenced dataset of historical HSR subgrade defects, selecting 661 event records that span various administrative levels and involve elements of the track structure, roadbed, and foundation. This dataset includes defects directly related to track unevenness, enabling us to analyze the influence of track structure on subgrade defects."

Since the track structure, subgrade, and foundation jointly bear the upper train load, unevenness in the subgrade's geometric dimensions may also lead to track irregularities. The defects we recorded, which directly affect the subgrade, can also be seen as early indicators of damage to the track structure. Moreover, during our statistical analysis, we identified multiple causes of track irregularities, including but not limited to settlement of the subgrade and foundation structures, as well as track aging and wear due to factors such as temperature changes, rainfall, and load. These causes form the basis of our model's assumptions, thereby enhancing its explanatory power and scientific rigor. To promote transparency and facilitate data reuse, we have uploaded the complete dataset to the figshare Generalist repository, which can be accessed via the DOI link: https://doi.org/10.6084/m9.figshare.24086016.

Regarding the effects of train moving load, we considered important factors such as train speed, environmental conditions, cumulative frequency effects, and the stiffness and elasticity of the track foundation. In practice, train speed is a dynamic variable, with each acceleration or deceleration affecting moving loads. We used design speed as a proxy for train speed to account for these variations. Additionally, higher design speeds typically require stiffer and more elastic *track foundations, indirectly reflecting the standards of the track foundation. Environmental factors, like temperature variations and rainfall, can also influence moving loads. For instance, temperature changes may cause track materials to expand or contract, leading to additional stress (corresponding to the factor "Number of days with maximum temperature exceeding 35 degrees Celsius"), while rainfall could reduce the stiffness of the track and subgrade, increasing the impact of moving loads (corresponding to the factor "Consecutive 5-day rainfall"). As for the train mass and structure, since it is determined by axle load and passenger capacity, railway authorities adjust train frequency based on actual ridership. Consequently, daily train loads remain relatively stable, so we did not include this factor separately in the model.*

Finally, regarding the evaluation of the bearing capacity of composite foundations, we agree that in actual engineering practice, foundation reinforcement is implemented in poor soil environments to ensure adequate bearing capacity. Our study does not directly use geomorphological and geological variables to assess the bearing capacity of composite foundations. Instead, these variables are used to evaluate the likelihood of subgrade defects, even when foundation reinforcement measures are in place. During the construction of railways, designers adopt specific foundation reinforcement measures based on the geological conditions to ensure subgrade bearing capacity. However, our analysis of HSR subgrade defect records revealed numerous defect occurrences. This suggests that despite bearing capacity being ensured, factors such as geology, geomorphology, and environmental conditions may still play significant roles in the development of subgrade defects. Our model aims to identify and quantify these influences to better understand the underlying causes of defects that persist despite foundation reinforcement.

What are the assumptions of the SEM model in section 2.3.2 of this paper? And what are the computational principles? Is it just 3 simple linear equations? The assumptions of the models used as well as the principles should be given in concise language when the paper is written, otherwise it will confuse the reader.

[Response 1.3]: Thank you for your valuable suggestions. In §2.3.2 (multi-defect model for relationships between factors and defect co-occurrence) of the main text, we have added the basic assumptions of the Structural Equation Model (SEM) and provided a description of the variables. As described, "Before applying the SEM model, certain assumptions must be met: the data used should have no outliers, the causal relationship between endogenous and exogenous variables must adhere to a time sequence, the sample size should be sufficient, and the error terms in the model should be independent of one other[1]. SEM consists of a measurement model and a structural model. The measurement model associates observed variables (such as average annual rainfall, consecutive 5-day rainfall) with latent theoretical variables (such as climate variables), representing relationships between endogenous and exogenous variables. The structural model investigates the causal relationships among these latent variables (such as paths H1-H15), analyzing their interactions."

We have added a description of the principles of the Structural Equation Model (SEM) in §2.3.2 (Multi-defect model for relationships between factors and defect co-occurrence) of the main *text. As described, "Equations (1) and (2) represent the measurement model, which links observed variables with latent variables. This process helps identify measurement errors and explains how the data reflect the underlying theoretical constructs. Specifically, ξ and η represent the exogenous and the endogenous latent variable, respectively. Λ_x and Λ_y denote the relationships between observed exogenous variables and exogenous latent variables, and between observed endogenous variables and endogenous latent variables, respectively. δ and ε represent the errors in the measurement model for x and y, respectively. Equation (3) is the structural model, which analyzes the causal relationships among latent variables and assess their impact on the overall model. In this model, B is the coefficient matrix representing the impact of endogenous latent variables on exogenous latent variables, Γ denotes the coefficient matrix for the direct effects of exogenous latent variables on endogenous latent variables, and ζ is the error vector for the structural equations."*

While the core equations of the model can be simplified into three linear equations, SEM is far more than just a simple combination of these equations. It systematically integrates disaster system theory with high-speed railway data to analyze complex variable relationships. SEM combines confirmatory factor analysis and multiple regression analysis, utilizing maximum likelihood estimation to optimize the model. The model's validity is then assessed through fit indices such as RMSEA and SRMR.

[Ref 1] Ullman, J.B. (2012). Structural Equation Modeling. In Handbook of Psychology, Second Edition (eds I. Weiner, J.A. Schinka and W.F. Velicer).

In section 3 Results part of this paper, why is the R^2 not more than 0.35 and how to prove the reliability and significance of the model analysis results? In addition to that, Figure 3 is obviously generated by GPT, which is totally inconsistent with the actual situation. This makes the reviewer sceptical about the results.

[Response 1.4]: Thank you for your valuable feedback. Regarding the R^2 value not exceeding 0.35, this outcome is primarily due to the complexity of the systems involved in our research, which encompass multiple layers of influencing variables. The intricate relationships among these factors, as well as the potential exclusion of some latent variables, naturally result in a lower R^2 value. It should be noted that while R^2 is an important measure of a model's explanatory power, it does not fully capture the model's overall effectiveness or significance of the model, especially in complex systems analysis. In the field of Structural Equation Modeling (SEM), it is not uncommon to encounter lower R^2 values[1-2], often below 0.25, particularly in cases involving multifaceted variables. For our model, certain paths achieved an R^2 of 0.329, which is relatively high compared to similar studies, indicating meaningful explanatory power within this context.

To further validate the reliability and scientific significance of our model, we used a variety of goodness-of-fit indices, including SRMR, RMSEA, GFI, AGFI, CFI, TLI, NFI, IFI, and RFI. By comparing these indices with established benchmark values, we were able to demonstrate how well the model fits the data and confirm its overall reliability. Moreover, the path coefficients in the model are statistically significant and align with theoretical expectations, indicating that *the causal relationships and interactions between variables are robust and credible. Finally, our model is grounded in a strong theoretical framework, with each path based on established theories and research hypotheses, thereby ensuring that the scientific significance of these relationships remains valid, even with a lower R^2 value.*

Regarding Figure 3, we understand the concern raised about its presentation. To enhance the clarity and visual appeal of the chart, we initially used image generation software to create the schematic of the defects. However, we recognize that this approach may have led to skepticism. As a result, we have now revised Figure 3 by removing the schematic of the defects in the center, ensuring that it more accurately reflects the actual data and research context. Additionally, we assure that all our data are well-documented and fully reproducible. If further clarification is needed regarding any of the figures or data, we welcome any additional questions or requests for information.

[Ref 1] Radulescu, A. T. et al. Impact of factors that predict adoption of geomonitoring systems for landslide management. Land 12, 752 (2023).
[Ref 2] Dai, Z. et al. Metallic micronutrients are associated with the structure and function of the soil microbiome. Nat. Commun. 14, 8456 (2023).

In this manuscript, the conclusions obtained are qualitative analyses and almost exclusively results that are widely accepted. The authors apply a complex model from the field of statistics, but neither quantitative results nor novel insights are given. How practical is the research content of this manuscript?

[Response 1.5]: Thank you for your feedback on our research. While the conclusions may align with widely accepted findings, our research offers significant quantitative analysis, as illustrated in Figures 2 and 3. These figures highlight the direct and indirect effects of various variables (e.g., HSR operation frequency, rainfall) on the occurrence and co-occurrence of subgrade defects. For instance, in Figure 2, the standardized effect value between HSR operation characteristics and defects is 0.24, meaning that for every 1 standard deviation increase in HSR operation characteristics, defects increase by 0.24 standard deviations. This result indicates that HSR operation characteristics have a significant impact on the co-occurrence of subgrade defects. Through this quantitative analysis, we not only identified the strength of the impact of HSR operation characteristics on defects but also revealed its key role in defect formation.

While previous studies have relied on laboratory model tests and numerical simulations to analyze subgrade defects, these methods often have limited scope and tend to overlook the long-term and broader impacts of environmental factors and operational characteristics. Our research moves beyond these traditional approaches by employing a large-scale empirical analysis combined with SEM, which enables us to explore complex causal relationships in a more holistic and practically relevant manner. Additionally, disaster system theory suggests that when assessing the relationship between natural disasters and environmental evolution over medium- and long-term time scales, smaller spatial scales tend to produce results that deviate more from actual conditions. Therefore, this study is the first to apply large-scale *empirical analysis and SEM modeling, providing a more comprehensive and practically relevant insight into the factors influencing subgrade defects.*

The practical value of our research lies in its potential to inform and improve real-world maintenance strategies. Our research not only validates that using disaster system theory can provide effective solutions for complex causal relationships of subgrade defects, but also offers corresponding policy recommendations. These research outcomes are significant for preventing and addressing future risks of subgrade defects due to climate change and increased HSR operation pressures. With these insights, relevant authorities can develop more effective maintenance strategies and preventive measures in practice, thereby enhancing the overall resilience and safety of the HSR system.

Furthermore, SEM's advantages over laboratory tests and numerical simulations make it particularly suited for large-scale data-driven analyses. For instance, China's high-speed railway currently covers 95% of major cities, but in complex environments like Tibet, high-speed rail has yet to be implemented. In Tibet, annual rainfall can reach up to 2000 millimeters, with extreme temperature variations—up to 40°C in summer and down to -20°C in winter. These unique climatic conditions and complex geological environments pose greater challenges to subgrade stability. Therefore, studying the failure mechanisms of subgrades under the combined effects of train loads and environmental factors is crucial. However, most current research only considers the separate effects of train loads or freeze-thaw cycles[1]. Our research offers a comprehensive solution for the complex causal relationships of such subgrade defects. Additionally, our study has broad international application value, providing insights not only for other countries' high-speed railway systems but also for the analysis and optimization of other infrastructure.

[Ref 1] Wu, Y., Fu, H., Bian, X., & Chen, Y. (2023). Impact of extreme climate and train traffic loads on the performance of high-speed railway geotechnical infrastructures. J. Zhejiang Univ., Sci., A. 24(3), 189-205. doi:10.1631/jzus.A2200341

In response to Reviewer #2's comments

Reviewer #2 (Remarks to the Authors):

Relevance and Timeliness: The study is timely, addressing the pressing issue of infrastructure resilience in the face of climate change and increased HSR operations.

[Response 2.1]: Thank you for the positive comments and for acknowledging the relevance and timeliness of our research.

Methodology: The use of SEM is appropriate for examining the complex interactions between various factors influencing subgrade defects.

[Response 2.2]: Thank you for the positive comments and for recognizing our methodology.

Comprehensive Analysis: The study considers both natural and operational factors, providing a holistic view of the issue.

[Response 2.3]: Thank you for the positive comments and for acknowledging our comprehensive analysis.

Practical Implications: The findings have significant practical implications for maintenance and infrastructure planning, highlighting the need for interventions to enhance resilience.

[Response 2.4]: Thank you for the positive comments and for recognizing the practical application value of our research.

Clarity and Specificity: While the abstract provides a broad overview, it lacks specificity in certain areas. For example, the exact nature of the SEM analysis and the specific variables considered could be briefly mentioned.

[Response 2.5]: Thank you for your valuable suggestions. We have revised the abstract to incorporate a more specific description of the SEM analysis, as requested. Due to the200-word limit, we carefully selected important variables to ensure the abstract remains concise while effectively communicating the core aspects of the study. In the revised version, we have now highlighted important variables such as sandy lands, gobi, bare lands, bare rock, geomorphological variables, operational characteristics, annual freezing days, and average annual rainfall.

The abstract has now been revised as read "As high-speed rail (HSR) operations increase and extreme weather events intensify, HSR subgrade stability is increasingly at risk. Prior studies have concentrated on the direct impacts of natural and anthropogenic factors on railway infrastructure, often overlooking their complex interactions and causality. This study employs structural equation modeling to examine the causal relationships between climate, geomorphology, geology, proximity, land use, operational characteristics, and economic factors affecting subgrade defects in China's HSR and the mechanisms underlying their co-occurrence. Our findings indicate that unused land types, such as sandy lands, gobi, bare lands, and bare rock, significantly impacts the co-occurrence of subgrade defects. Geomorphological variables and HSR operational characteristics also play critical roles, with the former indirectly affecting subgrade defects by influencing the latter. Bare land has the greatest influence on settlement and uplift deformation, while frost damage and mud pumping are primarily affected by annual freezing days and average annual rainfall. Projections based on climate trends and HSR operation plans suggest an increased risk of subgrade settlement, mud pumping, and uplift deformation is likely to increase in the future, underscoring the need for resilience interventions. This study advances the understanding of the mechanisms and potential causality of subgrade defects."

Balance of Information: This could benefit from a more balanced distribution of information. While the impacts of different types of land are mentioned, there is less detail on the operational characteristics of the HSR that influence subgrade defects.

[Response 2.6]: Thank you for your valuable feedback. We have addressed the need for a more balanced distribution of information by expanding on the operational characteristics of HSR.

Specifically, we have added Supplementary Figure 2: "Distribution of design speeds for China's HSR," which provides a detailed breakdown of the distribution of different HSR design speeds (100-199 Km/h, 200-299 Km/h, and 300-350 Km/h) across various provinces in China. Correspondingly, we have added relevant descriptions in the main text under §2.3.1 (Data preparation and processing steps), as described: "The distribution of design speeds for China's HSR system is shown in Supplementary Figure 2."

In addition, we have included further details in the Supplementary Methods section to clarify the HSR Mileage Calculation process, as described: "Given that longer mileage increases the likelihood of defects, we calculated the HSR mileage within each district and county. First, we validated the shapefiles (.shp) of the HSR lines against the official Chinese HSR map to ensure data accuracy. Then, in ArcGIS, we overlaid the map of China, divided by district and county boundaries, with the HSR line map. Using the split intersecting lines function, we segmented the HSR lines within each district and county. Afterward, we applied ArcGIS's zonal statistics function to calculate the HSR line lengths within each district and county. After calculating the mileage, we compared the results with pre-processed data to verify the accuracy of the segmentation. Additionally, we cross-referenced the HSR mileage data for each district or county with official records and conducted manual measurements for randomly selected counties to further validate data accuracy and consistency."

These additions ensure a more balanced discussion, emphasizing how both HSR design speeds and operational mileage contribute to subgrade defects, complementing the previously mentioned factors such as land types. This broader analysis provides a more comprehensive understanding of the interplay between natural factors and HSR operational characteristics, thus enhancing the overall robustness of the study.

Citing numerical facts: The authors are suggested to cite wherever anticipated or factual numerical values are reported.

[Response 2.7]: Thank you for your valuable suggestion. We have added citations wherever anticipated or factual numerical values are reported in the main text. Specifically, these include: " Fourteenth Five-Year Plan", it is anticipated that by 2025, China's operational HSR mileage will reach 50,000 kilometers, covering more than 95% of major cities[1]; and According to the "Global Climate Change Report 2015-2019" released by the World Meteorological Organization[45], the "2022 China Climate Bulletin" published by the China Meteorological Administration[46], and the "2016 China Medium and Long-term Railway Network Plan" report by the National Development and Reform Commission[47], it is indicated that in the future, temperatures will gradually rise, the frequency of HSR operations in China will increase.

[Ref 1] State Council of the People's Republic of China. Notice on the "14th Five-Year Plan for Modern Comprehensive Transportation System Development". Central People's Government of the People's Republic of China https://www.gov.cn/xinwen/2022-01/18/content_5669136.htm (2021).

[Ref 45] World Meteorological Organization. Global climate in 2015-2019: climate change accelerates. World Meteorological Organization https://wmo.int/news/media-centre/global-climate-2015-2019-climate-change-accelerates (2019).
[Ref 46] China Climate Committee. 2022 China Climate Bulletin. China Meteorological Administration https://www.cma.gov.cn/zfxxgk/gknr/qxbg/202102/t20210224_27562 50.html (2023).
[Ref 47] National Development and Reform Commission. Medium and Long-Term Railway Network Plan. Central People's Government of the People's Republic of China https://www.gov.cn/xinwen/2016-07/20/content_5093165.htm (2016).

Complete definition of DST: The analysis of this study is based on disaster system theory but there is not enough discussion and definition given on this. Authors are suggested to discuss how their approach i) identifies and assess risks within a system; ii) effectively inform on emergency planning; iii) guide recovery and reconstruction efforts by considering the long-term impacts and interactions within the system, and iv) policies that address the complexities of disaster management.
[Response 2.8]: Thank you for your valuable feedback. We have added the definition and description of Disaster System Theory (DST) in §2 (Theoretical framework) of the main text. As described, "DST provides a theoretical framework designed to systematically analyze and manage complex disasters. It seeks to explain the mechanisms of disaster occurrence, the pathways of disaster impact, and how systematic management can mitigate the negative impacts of disasters. The theory views disasters as the result of complex interactions among the natural environment, engineering structures, and human activities[1-4]."

Regarding point (i): In §4.1 (Characteristics and trends of defect co-occurrence) and §4.2 (Comparison of defect occurrence patterns) of the discussion, we utilized DST as a theoretical framework to identify and quantitatively assess various risk factors associated with HSR subgrade defects. For example, in Figure 2, the standardized effect value between HSR operational characteristics and defects is 0.24, indicating that a 1 standard deviation increase in HSR operational characteristics corresponds to a 0.24 standard deviation increase in defects. This result demonstrates that HSR operational characteristics significantly impact the co-occurrence of subgrade defects. Through this quantitative analysis, we not only clarified the risk posed by HSR operational characteristics on defects but also revealed their critical role in defect formation. Additionally, we have added a description in the text under §4 (Discussion), stating, "Identifying the most important factors chains helps prevent subgrade defects and reduce the overall system risk for HSR."

Regarding point (ii): We have revised §4.4 (Policy recommendations) in the discussion part of the main text to provide suggestions for high-speed railway (HSR) emergency planning and empirical support based on the important factors we identified related to subgrade defects. As described, "Our study shows that extreme temperatures have a significant impact on the stability of high-speed rail subgrades during operation. Specifically, the number of days with maximum temperature exceeding 35°C leads to the co-occurrence of subgrade defects, such as settlement, uplift, and mud pumping. Therefore, it is recommended to reduce train frequency *during high-temperature periods by strengthening monitoring and early warning systems, in order to alleviate the load on the subgrade and track. Additionally, a high-temperature emergency response mechanism should be established to ensure that timely measures can be taken under extreme weather conditions to maintain railway safety. Based on long-term climate change forecasts, it is suggested that localized modifications be implemented in high-risk areas (with extended periods of high temperatures) to reduce subgrade deformation and structural fatigue under high-temperature conditions. Furthermore, our research indicates that an increase in the annual freezing days significantly raises the risk of subgrade frost damage (standardized total effect value of 0.651). Therefore, it is recommended to use materials with stronger anti-frost heave properties in permafrost regions (such as the Harbin-Dalian HSR) to ensure the long-term stability and safety of infrastructure in cold climates. For example, the Madrid-Cádiz railway line in Spain uses platforms to monitor foundation degradation in extreme climate conditions, with alarm thresholds were set trigger speed and frequency adjustments[5].”*

Regarding point (iii): We have revised §4.4 (Policy recommendations) in the discussion part of the main text to guide recovery and reconstruction efforts by considering the long-term impacts and interactions of factors within Disaster System Theory (DST) and provided support with practical cases. As described, "Our study shows that adverse land use and geomorphological variables have standardized direct impact values of 0.320 and 0.383, respectively, on the co-occurrence of HSR subgrade defects. Therefore, when planning new railway routes, areas prone to subgrade problems, such as low-coverage grasslands, gobi deserts, saline-alkaline lands, and bare lands should be avoided. During the planning and construction phases, comprehensive geological and environmental assessments should be conducted to identify high-risk areas, such as high-altitude lines like the Lanzhou-Xinjiang HSR. In areas with unfavorable land use, climate-appropriate vegetation and soil improvement techniques can enhance stability. For Gobi or saline-alkaline areas, geotextiles and windbreaks can be used to mitigate the effects of weathering and salinization. For example, a study in Old Dalby, UK, using Electrical Resistivity Tomography (ERT), showed that grass cover can improve the stability of railway embankments[6].”

Regarding point (iv): We have revised §4.4 (Policy recommendations) in the discussion part of the main text to address the complexities of disaster management by strengthening defect management through the integration of causal analysis and data-driven methods, such as machine learning, to better predict risks. This revision is supported by practical cases. As described, "Our research shows that causal analysis can identify key factor chains influencing subgrade defects (e.g., geomorphology affecting HSR operational characteristics, which in turn impact subgrade defects). Therefore, it is recommended to combine causal analysis with data-driven management by regularly collecting data on high-speed rail operations, environmental conditions, and infrastructure. By integrating these datasets with the key factors identified through causal analysis, data-driven methods (such as machine learning) can be applied to more accurately identify and predict potential subgrade defect risks. This integration not only improves the understanding of the interactions between various factors through causal chains, but also enhances the transparency and interpretability of data-driven management, avoiding *the "black box" issue often associated with models. Additionally, optimizing predictions can enhance subgrade predictive maintenance, improving the efficiency and safety of railway infrastructure management while reducing unnecessary costs. For example, on the Zamudio railway bridge in the Basque region of Spain, machine learning algorithms were used to monitor railway infrastructure. Compared to traditional monitoring methods, this approach reduced costs and improved the efficiency of detecting structural defects[7]."*

[Ref 1] Mileti, D. A Sustainability Framework for Natural and Technological Hazards Disasters by Design: A Reassessment of Natural Hazards in the United States 17-19 (1999)
[Ref 2] Kates, R.W. et al. Environment and development -: Sustainability science. Science. 292, 641-642 (2001)
[Ref 3] Blaikie, P., Cannon, T., Davis, I., & Wisner, B. Disaster Pressure And Release Model At Risk: Natural Hazards, People's Vulnerability and Disasters 19-23 (1994).
[Ref 4] Hewitt, K. Risk and damaging events Regions of Risk: A Geographical Introduction to Disasters 10-15 (1997).
[Ref 5] Sanz Bobi, J. D., Garrido Martinez-Llop, P., Rubio Marcos, P., Solano Jimenez, A., & Fernandez, J. G. (2024). Prediction of degraded infrastructure conditions for railway operation. Sensors (Basel), 24(8). doi:10.3390/s24082456
[Ref 6] Holmes, J. et al. (2022). 4D electrical resistivity tomography for assessing the influence of vegetation and subsurface moisture on railway cutting condition. Engineering Geology, 307. doi:10.1016/j.enggeo.2022.106790
[Ref 7] Armijo, A., & Zamora-Sanchez, D. (2024). Integration of railway bridge structural health monitoring into the internet of things with a digital twin: A case study. Sensors (Basel), 24(7). doi:10.3390/s24072115

Mathematical representations: Disaster system theory could be better explained using i) differential equation state space; ii) network models; iii) agent-based models; iv) stochastic models, and v) objective-based optimisation models. Authors should consider implementing and discussing one of the aforementioned assessments.

[Response 2.9]: Thank you for your valuable suggestions. We appreciate the recommendation to use mathematical representation methods (such as differential equation state-space models, network models, agent-based models, stochastic models, and objective-based optimisation models) to better explain Disaster System Theory (DST). In this study, we chose Structural Equation Modeling (SEM) as the primary analytical tool. Although SEM does not fully belong to any of the five mathematical models mentioned, it does share some similarities with network models by effectively exploring and quantifying complex interactions among multiple variables within a disaster system.

SEM is particularly advantageous for its ability to analyze both direct and indirect effects, revealing pathways that are crucial for understanding the intricate dynamic relationships within a disaster system. This capability is instrumental in illustrating causal relationships and interactions between variables, especially in the context of HSR subgrade defects.

We acknowledge that other mathematical models, such as stochastic models or goal-based *optimization models, may offer significant advantages in specific application scenarios. We plan to explore the potential of these methods in future research to enhance our understanding and management of disaster systems.*

Additionally, to further enhance the explanatory power of DST, we have added Supplementary Figure 1, which links important variables to the disaster system theory, providing a clearer representation of the roles and interrelationships of these variables within the disaster system.

Figures: Figure 1 is incorrect and its caption should be detailed.
i) Use Internationally Recognized Boundaries: Map to reflect internationally recognized boundaries.
ii) Provide Clear Annotations: If the map is discussing disputed areas, clearly annotate these regions with a note explaining that the area is disputed
[Response 2.10]: Thank you for your valuable suggestions. We have corrected the title of Figure 1 to "Spatial distribution of HSR subgrade defect records across various land use types in China." Additionally, we have reviewed Figure 1 following the requirements of the "Specifications for the content of public maps in China." The content of the image complies with these specifications.

Clear definition: The authors stated (H1) to (H15) but have not vividly defined it in the manuscript or used anywhere in its model. Use (H[.]) as feature vector components of multi-defect pathways to clearly define the relationship, cause and effect as matrices.
[Response 2.11]: Thank you for your valuable feedback on our research. We have addressed your concerns by adding Supplementary Table 3, "Detailed explanation of pathways (multi-defect model)," to the supplementary materials. This table provides a comprehensive explanation for each pathway from H1 to H15, including: Pathways: The corresponding pathway hypothesis number (e.g., H1, H2, etc.); Cause: The influencing factors (e.g., geomorphological variables, land use); Effect: The affected outcomes (e.g., subgrade defects); Hypothetical relationship: The direction of the hypothetical relationship (positive or negative); Brief introduction: A concise description of each hypothesis; References: Supporting literature references. We hope this addition clarifies the relationships, causes, and effects in our multi-defect model and enhances the overall understanding of our study.

Thank you for your valuable feedback. We have added Supplementary Table 2: Path coefficients for the structural model (multi-defect model) in the supplementary materials, using (H.) to denote the final structural model. Additionally, in Supplementary Figure 1: Hypothesized structural equation model for High-Speed Rail (HSR) subgrade defects, we have used H1-H15 to represent each hypothesis. We also utilized pathways (e.g., H10-H11-H6) to express mediation effects in Supplementary Table 4: Standardized bootstrap mediation effects test for composite subgrade defects over 5000 iterations (multi-defect model). In the main text, section §3.1 (Socio-environmental causalities underpinning the occurrence of defects), (H.) has been used to express the results of the multi-defect model.

We have added Supplementary Table 3, titled "Detailed explanation of pathways (multi-defect *model),” where each pathway’s causal relationships and definitions are thoroughly detailed. We believe that presenting the model structure visually through Supplementary Table 3 and Supplementary Figure 1, “Hypothesized structural equation model for high-speed rail (HSR) subgrade defects,” is more appropriate, as they directly illustrate the causal relationships and interactions between variables. Additionally, the SEM pathway principles are already explained with equations in §2.3.2 (Multi-defect model for relationships between factors and defect co-occurrence). Therefore, including matrix representations might unnecessarily complicate the manuscript. We are confident that the existing figures and text effectively communicate the core models and findings of the study.*

Software: Lulutong software lacks integration capabilities and scalability in comparison to SAP TM, Oracle Transport Management, JDA, and others. How do the authors overcome the drawback and ensure their reported results are correct.

[Response 2.12]: Thank you for your valuable feedback. We acknowledge the limitations of Lulutong software in terms of integration and scalability compared to systems like SAP TM, Oracle Transport Management, or JDA. However, Lulutong provides specific advantages in handling Chinese high-speed rail operational data, particularly in supporting timetables and operational data specific to China’s public transport system, which those other systems typically do not cover. To ensure the accuracy of our results, we have implemented a rigorous validation process. This includes cross-referencing the frequency data obtained from Lulutong with official sources such as the China Railway Customer Service Center (www.12306.cn). By doing so, we have ensured that our data is consistent and accurate.

Conclusion: Conclusion is overclaiming ‘the critical factors for each defect were identified’. The identified factors do not give complete picture and proof of being critical factor. The manuscript should focus more on this. What is the probability that these factors in defect causes, direct impact, and risk occurrence. Decision making should be supported with well-trained data and evidence

[Response 2.13]: Thank you for your valuable suggestions. We have carefully revised the conclusion section to address your concern. Specifically, we have now replaced the term “critical factors” with “important factors” to more accurately reflect our findings. We acknowledge that while these factors have shown significant impacts under certain conditions, they may not independently serve as definitive factors. This revision aims to more precisely convey the role of each factor in our study and to avoid overstatements.

To provide a clearer understanding of the probability and impact of these factors on defect causes, direct impact, and risk occurrence, we have included standardized total effect values for all variables in Figures 2 and 4. For instance, in Figure 2(b), the direct effect value of geomorphological variables on defects is 0.382, with an indirect effect value of -0.116, leading to a standardized total effect value of 0.266. This indicates that a one standard deviation increase in geomorphological variables results in a 0.266 standard deviation increase in defects. Similarly, in Figure 4(a), the standardized total effect value of HSR operational frequency on settlement is 0.173, indicating that a one standard deviation increase in operational frequency leads to a 0.173 standard deviation increase in defects.

We have ensured the sufficiency of data and evidence, strictly adhering to scientific research standards to support the decision-making process. Our data sources include subgrade defect data and variable data. The subgrade defect data have undergone peer review and are published in the journal Scientific Data. The dataset can be accessed via the following DOI link: https://doi.org/10.6084/m9.figshare.24086016, ensuring its scientific validity and applicability. The complete introduction to the subgrade defect data can be found in Reference 17 of the main text. Variable data are sourced from authoritative open-source databases such as the National Earth System Science Data Center, National Bureau of Statistics, and the Geological Survey of China. These data have been extensively validated and applied. In the manuscript, we thoroughly introduce these variables (Climate variables, Geomorphological variables, Geological variables, Land use, High-speed rail proximity index) in Reference 26.

Additionally, in §4.4 (Policy recommendations), we have added three typical international railway cases to support decision-making. Examples include: "For example, a study in Old Dalby, UK, using Electrical Resistivity Tomography (ERT), showed that grass cover can improve the stability of railway embankments[1]."
"For example, the Madrid-Cádiz railway line in Spain uses platforms to monitor foundation degradation in extreme climate conditions, with alarm thresholds were set trigger speed and frequency adjustments[2]."
"For example, on the Zamudio railway bridge in the Basque region of Spain, machine learning algorithms were used to monitor railway infrastructure. Compared to traditional monitoring methods, this approach reduced costs and improved the efficiency of detecting structural defects[3]."

[Ref 1] Holmes, J. et al. (2022). 4D electrical resistivity tomography for assessing the influence of vegetation and subsurface moisture on railway cutting condition. Engineering Geology, 307. doi:10.1016/j.enggeo.2022.106790
[Ref 2] Sanz Bobi, J. D., Garrido Martinez-Llop, P., Rubio Marcos, P., Solano Jimenez, A., & Fernandez, J. G. (2024). Prediction of degraded infrastructure conditions for railway operation. Sensors (Basel), 24(8). doi:10.3390/s24082456
[Ref 3] Armijo, A., & Zamora-Sanchez, D. (2024). Integration of railway bridge structural health monitoring into the internet of things with a digital twin: A case study. Sensors (Basel), 24(7). doi:10.3390/s24072115

In response to Reviewer #3's comments

Reviewer #3 (Remarks to the Authors):
Investigate the stability of high-speed rail (HSR) subgrades in China under increasing operational frequency and extreme weather events. This study addresses a significant gap in understanding complex interactions and causality between environmental factors and HSR subgrade defects. Given the planned expansion of China's HSR to 50,000 kilometres by 2025, understanding these interactions is crucial for infrastructure resilience and safety. The authors employ structural equation modelling (SEM) to explore the factors contributing to common

subgrade defects such as settlement, frost damage, uplift deformation, and mud pumping. The paper aims to identify key influencing variables and mechanisms underlying the co-occurrence of these defects, incorporating geomorphological variables, land use, climate conditions, and operational characteristics of HSR. The main findings indicate that unused lands like sandy areas, Gobi deserts, bare lands, and bare rocks significantly affect the co-occurrence of subgrade defects. Geomorphological variables indirectly influence subgrade defects by impacting HSR operational characteristics, with bare land most significantly affecting settlement and uplift deformation. Further, frost damage and mud pumping are primarily driven by annual freezing days and average annual rainfall. These results highlight the need for integrated management strategies to enhance maintenance and infrastructure resilience. This study advances the understanding of HSR subgrade defects' mechanisms, providing a scientific basis for future HSR maintenance planning.

[Response 3.1]: Thank you for your insightful review and feedback. We appreciate your recognition of the importance of investigating the stability of HSR subgrades in China, especially given the planned expansion of the network by 2025. In response to your comments, we have further addressed the complex interactions between environmental factors and subgrade defects in more detail, while also refining our structural equation modelling (SEM) approach. Our revisions have aimed to better emphasize the critical mechanisms identified, provide stronger empirical evidence, and suggest actionable strategies to improve HSR maintenance and resilience.

The article presents a potentially significant study using structural equation modelling (SEM) to examine the factors influencing high-speed rail (HSR) subgrade defects in China. However, several critical issues must be addressed to ensure the study's validity and rigour.

Firstly, the theoretical framework is inadequately integrated. The Disaster System Theory is mentioned but not sufficiently grounded in seminal works such as Mileti et al. (1999). The hypotheses need clearer links to established theoretical principles, supported by a robust literature review that includes alternative hypotheses and justifies chosen pathways.

[Response 3.2]: Thank you for your valuable feedback. We have strengthened the integration of the theoretical framework by grounding it more thoroughly in seminal works, particularly including Mileti et al. (1999), and other key contributions to Disaster System Theory (DST), c.f. [Response 3.6]. In §2 (Theoretical framework), we now clearly link our hypotheses to established principles from DST, ensuring a more cohesive and theoretically robust foundation for our study. Furthermore, we have expanded our literature review to incorporate and evaluate alternative hypotheses such as similarity theory, soil mechanics, and stress analysis theory, and justify the selection of DST based on these alternatives. In [Response 3.7], we provide a detailed discussion on these alternatives to justify the selection of DST as suitable framework for the study. By doing so, we ensure that the theoretical basis for our study is both comprehensive and well-supported by the literature.

Secondly, the data quality and preprocessing steps lack transparency. The selection criteria, temporal range, and handling of extreme values or outliers are insufficiently detailed. Additionally, the validation and cleaning processes are not described, raising concerns about potential biases and the dataset's reliability. Comprehensive documentation of these steps is crucial.

[Response 3.3]: Thank you for your valuable suggestions. In response, we have taken several steps to enhance the transparency and rigor of our data handling and preprocessing procedures. Firstly, we have now clarified the data sources, selection criteria, temporal range, and potential limitations in [Response 3.9] and [Response 3.11], ensuring the reliability of our dataset. Secondly, we have improved the documentation of how extreme values and outliers were handled in [Response 3.10]. We have now offered a clearer explaination of the methodological processes, providing a comprehensive theoretical foundation for normalization. Additionally, we have now detailed the preprocessing steps in ArcGIS, including the calculation and validation of high-speed rail mileage, to enhance the transparency and robustness of our methodology.

Thirdly, the SEM approach requires improvement. Model specification is under-defined and lacks a comprehensive diagram and detailed equations. The interpretation of model fit indices and path significance is superficial, and diagnostic tests for residuals and multicollinearity are absent. A thorough justification of variable selection, path modifications, and detailed sensitivity analyses are needed to ensure model robustness.

[Response 3.4]: Thank you for your valuable feedback. Based on your comments, we have undertaken several improvements to enhance the robustness and clarity of our SEM approach：

(1) We provided a more detailed rationale for the selection and exclusion of variables in [Response 3.12]. We have also expanded this explanation in the supplementary methods. A comprehensive SEM diagram was added to illustrate the relationships clearly. Additionally, we have added a description of the principles and basic assumptions of the Structural Equation Model (SEM), along with a detailed explanation of the variables, in §2.3.2 (Multi-defect model for relationships between factors and defect co-occurrence) of the main text. As mentioned, "Before applying the SEM model, certain assumptions must be met: the data used should have no outliers, the causal relationship between endogenous and exogenous variables must adhere to a time sequence, the sample size should be sufficient, and the error terms in the model should be independent of one other. SEM consists of a measurement model and a structural model. The measurement model associates observed variables (such as average annual rainfall, consecutive 5-day rainfall) with latent theoretical variables (such as climate variables), representing relationships between endogenous and exogenous variables. The structural model investigates the causal relationships among these latent variables (such as paths H1-H15), analyzing their interactions. Thus, SEM synthesizes theory and data to analyze the direct and indirect effects between variables systematically. It is calculated according to equations (1)-(3)."

(2) While we recognize the importance of detailed equations, we have opted for visual clarity to avoid overcomplicating the manuscript. We have included a series of figures and tables (Fig 2, Fig 3, Supplementary Figure 1, and Supplementary Table 3) that clearly represent the relationships and directional paths between variables. These visual representations, combined with detailed descriptions, effectively communicate the structure and findings of the model

without the need for overly complex mathematical equations. We believe this approach strikes a balance between thoroughness and readability.

(3). Regarding the interpretation of model fit indices, in [Response 3.13], we have now enhanced the discussion, including comparisons to benchmark values for RMSEA, CFI, etc. We have also provided a more detailed explanation of the improvements made following path adjustments, which ensures a clearer interpretation of model fit and path significance. This helps to clarify the robustness of our model and the relevance of the paths specified.

(4). Regarding diagnostic test for multicollinearity and residuals, we have incorporated Variance Inflation Factor (VIF) analyses for each single-defect model, see [Response 3.16], and added a residual covariance plot in the supplementary materials. These enhancements ensure a more robust and transparent SEM approach, improving the model's validity and interpretability.

The authors should ground their hypotheses in a well-articulated theoretical framework with comprehensive literature support to enhance the study. Detailed documentation of data preprocessing and validation steps is essential. The SEM approach must include exhaustive model specification, fit assessment, and diagnostic testing. These enhancements will significantly improve the study's validity, interpretation, and contribution to the field.
[Response 3.5]: Thank you for your insightful feedback. Based on your comments, we have made several key improvements: (1). Theoretical framework: As explained in [Response 3.6], we have strengthened the foundation by integrating Disaster System Theory (DST), citing seminal works such as Mileti et al. (1999). This provides a robust theoretical basis for our hypotheses. (2). Data preprocessing and Validation: Detailed explanations of data selection, preprocessing, and validation have been provided in [Response 3.9] and [Response 3.11]. This includes handling missing values, standardizing variables, and the methodology for outlier identification. (3). SEM model specification and Fit: As addressed in [Response 3.12], we included a comprehensive SEM diagram and explained variable selection, exclusions, and the theoretical paths. In [Response 3.13], we added detailed interpretations of model fit indices and documented modifications made to improve fit, all benchmarked against standard criteria like RMSEA, CFI, and SRMR. (4). Diagnostic testing: In [Response 3.16], we documented the diagnostic tests conducted, such as residual covariance plots and multicollinearity checks, further enhancing model robustness. These enhancements ensure a more detailed and scientifically rigorous presentation, addressing the areas you highlighted. We believe these revisions greatly strengthen the study's clarity, validity, and contribution to the field.

Theoretical Framework: The article references Disaster System Theory but fails to incorporate seminal works like Mileti et al. (1999) or subsequently influential papers that thoroughly establish the theory's relevance to HSR subgrade defects.
The manuscript mentions, “This study draws on the Disaster System Theory to analyze the factors influencing HSR subgrade defects in China.”
Points to address: Incorporate and cite seminal works and influential papers on Disaster System Theory. Explain how this theory explicitly applies to the hypothesized relationships among

geomorphological variables, land use, climate variables, geological variables, and HSR operational characteristics. Elucidate how the mediation pathways logically extend from these theoretical principles.

[Response 3.6]: Thank you for your thorough review and valuable suggestions. We have added seminal works and influential papers on Disaster System Theory (DST) to the references in our manuscript as references 19, 20, 21, and 22. Additionally, we have also revised §2 (Theoretical framework) to explain in greater depth how DST explicitly applies to geomorphological variables, land use, climate variables, geological variables, and HSR operational characteristics. As described, "DST provides a theoretical framework designed to systematically analyze and manage complex disasters. It seeks to explain the mechanisms of disaster occurrence, the pathways of disaster impact, and how systematic management can mitigate the negative impacts of disasters. The theory views disasters as the result of complex interactions among the natural environment, engineering structures, and human activities.

We compiled an extensive georeferenced dataset of historical HSR subgrade defects, selecting 661 event records that span various administrative levels and involve elements of the track structure, roadbed, and foundation. This dataset includes defects directly related to track unevenness, enabling us to analyze the influence of track structure on subgrade defects. The distribution of HSR subgrade defect records and land use is illustrated in Fig 1. Based on the analysis of these records, the causes of the defects were categorized according to different attributes and associations into factors such as land use, climate, geology, geomorphology, proximity to HSR, HSR operational characteristics, and economic factors. As DST defines the disaster-prone environment as encompassing both natural and human environments, factors related to the natural environment, such as land use, climate, geology, and geomorphology, along with those related to the human environment, like economic factors, can be integrated into the disaster-prone environment category. According to DST, disaster-causing factors can be classified into natural and human-induced categories. In this study, HSR operational characteristics (human-induced) and proximity to HSR (natural-induced) are identified as disaster-causing factors."

We have clarified in §2 (Theoretical framework) how the mediating pathways are derived from Disaster System Theory. As described, "It posits that different disaster-causing factors arise from different environmental systems, allowing these factors to serve as mediating variables between the disaster-prone environment and the disaster-bearing body (i.e., the HSR subgrade). Additionally, due to the Earth's inherent self-regulating abilities, internal factors within the disaster-prone environment may influence each other through mediating effects."

[Ref 19] Mileti, D. A Sustainability Framework for Natural and Technological Hazards Disasters by Design: A Reassessment of Natural Hazards in the United States (ed. Mileti, D.)17-19 (Joseph Henry Press, 1999)

[Ref 20] Kates, R.W. et al. Environment and development -: Sustainability science. Science. 292, 641-642 (2001)

[Ref 21] Blaikie, P., Cannon, T., Davis, I., & Wisner, B. Disaster Pressure And Release Model At Risk: Natural Hazards, People's Vulnerability and Disasters (ed. Blaikie.) 19-23 (Routledge, 1994).

[Ref 22] Hewitt, K. Risk and damaging events Regions of Risk: A Geographical Introduction to Disasters (ed. Hewitt, K.) 10-15 (Routledge, 1997).

Literature Review: The literature review inadequately grounds the hypotheses in a comprehensive, up-to-date body of research. Fundamental studies are selectively cited without enough context, and potential alternative hypotheses are not sufficiently acknowledged.
The manuscript mentions, "Prior studies have concentrated on the direct impacts of natural and anthropogenic factors on railway infrastructure, often overlooking their complex interactions and causality."
Points to address: Conduct a thorough literature review that includes a broad spectrum of relevant, high-calibre studies. Ensure accurate interpretation and contextualization of these studies within the field. Acknowledge and discuss potential alternative hypotheses or theories and justify the chosen hypotheses in light of these alternatives.
[Response 3.7]: Thank you for your thorough review and valuable suggestions. In response, we have significantly expanded the literature review to ensure a more comprehensive and balanced grounding of our hypotheses in the existing body of research.

First, we have added a new "Supplementary Table 3: Detailed explanation of pathways (multi-defect model)", where each hypothesis is detailed with supporting references, totaling 31 citations. Additionally, in §1 (Introduction), we provided literature discussing the interaction of variables to ensure a balanced citation of fundamental studies. As described, *"Wu et al. (2023) explored how high-frequency railway operations interact with various environmental loads, such as rainwater infiltration, seasonal wet-dry cycles, and freeze-thaw cycles, leading to the degradation of infrastructure performance[1]. Sundell et al. (2019) examined the increased risk of subgrade settlement in high-speed rail systems due to groundwater extraction around the railway, highlighting the complex environmental factors influencing railway infrastructure stability[2]."*

We have also addressed the suggestion to acknowledge and discuss potential alternative hypotheses or theoretical frameworks. During the course of this study, we also considered similarity theory, soil mechanics theory, and stress analysis theory as potential alternatives to Disaster System Theory (DST). These theories have already formed the foundation for laboratory physical model tests and numerical simulations, which we briefly overviewed in the Introduction section. Similarity theory is used for designing indoor physical model tests that replicate real-world conditions, while soil mechanics theory provides a critical analytical framework for assessing the deformation and failure mechanisms of subgrade under various stress conditions. Stress analysis theory, often combined with Finite Element Analysis (FEA), helps explain how stress distribution and transfer within the railway subgrade, induced by train loads and geological conditions, may lead to the formation of subgrade defects.

Although these theories offer important perspectives for analyzing individual factors, we found that they have limitations when addressing the complex interactions between multiple variables. In contrast, DST offers a more holistic and systematic approach by considering the combined effects of environmental factors, operational characteristics, and potential disaster factors on *subgrade defects. This comprehensive perspective enables us to better understand the mechanisms of HSR subgrade defects, thereby enabling the development of more effective prevention and mitigation strategies. This comprehensive perspective is why we selected DST as the primary theoretical framework for our study. Moreover, DST incorporates relevant principles of soil mechanics theory and stress analysis theory, providing a unified framework to explain the causes and complex interactions of HSR subgrade defects. We believe that this systematic approach, grounded in DST, offers a more robust explanation of subgrade defects and provides better scientific support for the long-term maintenance and management of HSR infrastructure.*

[Ref 1] Wu, Y., Fu, H., Bian, X., & Chen, Y. (2023). Impact of extreme climate and train traffic loads on the performance of high-speed railway geotechnical infrastructures. J. Zhejiang Univ., Sci., A. 24(3), 189-205. doi:10.1631/jzus.A2200341
[Ref 2] Sundell, J., Haaf, E., Tornborg, J. & Rosen, L. Comprehensive risk assessment of groundwater drawdown induced subsidence. Stoch Env Res Risk A. 33, 427-449 (2019).

Specificity and Testability: The hypotheses lack specificity in predicted relationships and expected directionalities, making empirical testing through SEM challenging.
The manuscript mentions, "From a perspective of direct impact, the disaster-prone environment (including climate variables, geomorphological variables, land use, HSR proximity index) and disaster-causing factors (such as geological variables, HSR operational characteristics) may all potentially have direct effects on HSR subgrade defects."
Points to address: Define each hypothesis with explicit predicted relationships and expected directionalities (positive or negative influences). Ensure that all variables are measurable and articulated for empirical testing. Use precise language to describe how each factor interacts with subgrade defects, facilitating precise modeling in the SEM.
[Response 3.8]: Thank you for your valuable suggestions. We have now added "Supplementary Table 3: Detailed explanation of Pathways (multi-defect model)" in the supplementary materials, where each hypothesis is defined with explicit predicted relationships and expected directionalities (positive or negative influences). This table clarifies how each factor interacts with subgrade defects, facilitating precise modeling in SEM. Specifically, the table includes: (1)Pathways: The corresponding hypothesis number (e.g., H1, H2); (2)Cause: Influencing factors (e.g., geomorphological variables, land use); (3)Effect: The impacted outcome (e.g., subgrade defects); (4)Hypothetical relationship: The directional influence (positive or negative); (5)Brief introduction: A brief explanation of each hypothesis, describing the positive or negative influence, and; (6)References: Supporting literature. For example, for H1 (the impact of geomorphological variables on subgrade defects), we explicitly describe how geomorphological variables (such as slope orientation and gradient) influence surface water flow and subgrade temperature, thereby increasing the likelihood of subgrade defects. This refined and quantified hypothesis definition will help in accurately testing and validating the hypothesized relationships in the SEM model. We believe these improvements will further enhance the scientific rigor and practical applicability of the study.

Data Sources and Sampling: The data collection process lacks detail, particularly regarding

selection criteria and temporal range, which raises concerns about potential biases or gaps.
The manuscript mentions, “Subsequently, 661 georeferenced event records of eight defect types were selected, crossing provincial, municipal, county, township, and smaller scales.”
Points to address: Provide detailed information on the selection criteria for event records, including temporal and spatial dimensions. Justify the representativeness of the dataset, ensuring it encompasses a sufficiently diverse set of conditions to mitigate biases. Offer a rationale for data inclusion and discuss any limitations or potential gaps in the dataset.
[Response 3.9]: Thank you for your thorough review and valuable suggestions. We have now provided more comprehensive information on the data collection process, including the selection criteria for event records and considerations regarding temporal and spatial dimensions, to enhance the representativeness of the data and address potential biases, c.f. details in our recent data paper, reference 17 (Wang et al., 2024, Scientific Data 11: 293). Specifically, the event records were selected based on two main criteria: (1) Temporal span: The dataset covers records from 1999 to 2023, ensuring that the temporal dimension reflects dynamic changes in defects under various climate and operational conditions. (2) Spatial coverage: The records encompass all high-speed railways constructed in China within this time span, ensuring that the spatial dimension is broad and representative of different geographical regions. Each defect record includes latitude and longitude coordinates, ensuring spatial accuracy, and the specific timing of each event is documented to reflect the dynamic and temporal characteristics of the defects. We have also highlighted a potential limitation: if certain subgrade defects do not meet the standards specified in China's “High-Speed Railway Design Specification” (TB10621-2014), they may not have been recorded, which could lead to underreporting of minor defects. To enhance transparency and facilitate further research, we have provided the Figshare link to download the complete dataset: https://doi.org/10.6084/m9.figshare.24086016.

Data Preprocessing: Moon's method of normalization lacks a thorough rationale, and there is no mention of handling extreme values or outliers.
The manuscript mentions, “We adopted Moon's method of using the normalize module to standardize the factors, eliminating the impact of differing magnitudes on analysis results.”
Points to address: Provide a comprehensive rationale for choosing Moon's method over other normalization techniques. Describe how extreme values and outliers are identified and handled. Ensure clear documentation of the ArcGIS processing steps and include justification and verification of the mileage calculations.
[Response 3.10]: Thank you for your thorough review and valuable suggestions. We have now added detailed explanations in §2.3.1 (Data preparation and processing steps) of the main text to justify our chosen methods. As described, “We adopted Moon's min-max normalization method to standardize the factors, thereby minimizing the impact of differing magnitudes on analysis results. Min-max normalization was chosen over other methods because it scales all variables to a [0,1] range, ensuring equal contribution to the analysis without introducing biases due to differences in units or magnitudes. This approach is particularly suitable for the SEM analysis conducted in this study, as it better preserves the intrinsic relationships between variables.”

We have also elaborated on the process of identifying and handling outliers and extreme values in the Supplementary Methods. As described, "To prevent errors in the SEM analysis caused by extreme values and outliers, we examined the data distribution of each variable and identified values outside the Interquartile Range (IQR). During this process, no significant outliers were found that would substantially impact the analysis. Additionally, we conducted a comprehensive check for missing values and found none. Therefore, no alterations or removals of data points were necessary."

Furthermore, we have detailed the process of calculating the HSR mileage using ArcGIS in the Supplementary Methods section of the supplementary materials. As described, "High-speed rail mileage calculation: Given that longer mileage increases the likelihood of defects, we calculated the HSR mileage within each district and county. First, we validated the shapefiles (.shp) of the HSR lines against the official Chinese HSR map to ensure data accuracy. Then, in ArcGIS, we overlaid the map of China, divided by district and county boundaries, with the HSR line map. Using the split intersecting lines function, we segmented the HSR lines within each district and county. Afterward, we applied ArcGIS's zonal statistics function to calculate the HSR line lengths within each district and county. After calculating the mileage, we compared the results with pre-processed data to verify the accuracy of the segmentation. Additionally, we cross-referenced the HSR mileage data for each district or county with official records and conducted manual measurements for randomly selected counties to further validate data accuracy and consistency."

Data Validation and Cleaning: The article lacks explicit steps for data validation and cleaning, and there is no evidence of cross-validation with external datasets.
The manuscript mentions, "This critique refers to a general omission in the manuscript. Thus, a direct quote is not provided."
Points to address: Outline and document the steps taken for data cleaning, including handling missing values and inconsistencies. Provide evidence of cross-validation with external datasets or independent sources to ensure data reliability. Include quality control measures, such as error estimates for georeferencing and enhancing dataset robustness.
[Response 3.11]: Thank you for your valuable suggestions on our manuscript. During the data processing phase, we conducted a thorough check for missing values. Since our analysis was conducted at the district and county level, with a relatively broad spatial resolution, the data resolution used in this study was sufficient to cover all analysis areas fully, ensuring that no data gaps were present. We have now documented this process in the Supplementary Methods. As described, "Additionally, we conducted a comprehensive check for missing values and found none."

The data sources we used all come from authoritative open-source databases or peer-reviewed papers, ensuring both reliability and widespread applicability in scientific research. The subgrade defect data, which have been peer-reviewed and published in the journal Scientific Data, are available via this DOI link: https://doi.org/10.6084/m9.figshare.24086016. This dataset includes detailed information on each defect point, such as coordinates, references, time, and HSR line, to ensure its scientific validity and applicability. The complete introduction *to the defect data can be found in Reference 17 of the main text.*

Additionally, the variable data were obtained from reputable sources like the National Earth System Science Data Center, the National Bureau of Statistics, and the Geological Survey of China, all of which have been extensively validated and widely used in various scientific studies. In Table 1 "Detailed description of variables," we provided a detailed introduction to these factors (Climate variables, geomorphological variables, geological variables, land use, High-speed rail proximity index) to support their reliability.

Model Specification: Justification for selecting variables and excluding others is inadequate, and the SEM diagram and equations are under-specified.

The manuscript mentions, "Using SEM, we can significantly enhance the understanding of defect formation mechanisms, accurately identify key influencing variables, and effectively support disaster risk assessment and management decisions."

Points to address: Justify the selection of variables and provide a rationale for excluding others. Include a comprehensive SEM diagram and detailed equations representing the relationships among the variables. Ensure these align with the hypotheses and bolster the model's logical coherence.

[Response 3.12]: Thank you for your detailed review and valuable suggestions. We have extended the methodology for both the multi-defect model and the single-defect model, providing clear justifications for the selection and exclusion of variables. These details can be found in the Supplementary Methods section.

For the multi-defect model, we based our selection on Disaster System Theory (DST) hypotheses (H1–H15). The process of variable selection followed a systematic approach: "(1) Initial variable selection: All initial variable subclasses and variables were imported into AMOS to construct a preliminary measurement model. At this stage, we screened the variables based on their statistical significance to the target defects. Specifically, variables with a P-value greater than 0.05 were excluded, as they were not statistically significant at the 95% confidence level. This step was essential to ensure that only statistically significant factors were retained, thereby improving the model's accuracy and explanatory power. The variables excluded during this process included: silty clay, clay loam, loam, sand, effective water storage capacity -50 mm/m, effective water storage capacity-0 mm/m, soil depth-10cm, drainage class -well drained, drainage class -somewhat poorly drained, drainage class -poorly drained, drainage class -very poorly drained, distance to fault, drylands, forested lands, shrublands, vegetable woodlands, other forested lands, canals, lakes, reservoirs, mudflats, beach land, urban land, rural residential, other unused land (alpine desert, tundra), marine, and frost damage. This step was essential to ensure that only statistically significant factors were retained, thereby improving the model's accuracy and explanatory power. (2) Path fitting and further screening: After the initial screening, the remaining variables were connected according to the hypothesized theoretical paths to construct a preliminary structural model. We excluded paths with a P-value greater than 0.05 to ensure statistical significance. Variables with a significance level greater than 0.1 and lacking significant causal relationships with other variables were also removed. During this phase, we excluded the entire category of geological variables due *to their minimal impact on defects and weak correlations with other variables. (3) Model optimization: After the initial path fitting, we further optimized the model by evaluating its fit indices (e.g., SRMR, RMSEA, GFI, CFI, TLI, NFI, IFI, RFI). Variables that were difficult to integrate into the model during the fitting process, such as wind speed and consecutive 5-day rainfall, were excluded. The final set of variables that were statistically significant and contributed to the model fit are listed in Supplementary Table 1." This systematic approach ensured that the final model was both statistically robust and highly explanatory, thereby enhancing the validity of the findings.*

For the single-defect model, we followed a similarly structured approach: "(1) Detection and addressing of multicollinearity: During the construction of the independent defect models, we first used variance inflation factors (VIF) to detect and address multicollinearity, excluding factors with VIF greater than 10 to improve model stability and accuracy. (2) Initial model building and variable screening: We then built preliminary models by associating each defect with relevant variable subclasses. During this phase, we used AMOS software for an initial screening, excluding variable subclasses with a P-value greater than 0.05. (3) Categorization and path fitting: Based on Disaster System Theory, candidate variables were categorized into two groups: those that could not establish causal relationships with variables having P-values less than 0.05 and those that could. Variables unable to establish such relationships were excluded. Remaining candidate variables were fitted into path models based on causal relationships, and paths with P-values greater than 0.05 were removed. (4) Model optimization: Finally, we optimized the model using fit indices such as SRMR, RMSEA, GFI, CFI, TLI, NFI, IFI, and RFI. Variables that were difficult to integrate into the model during this process were excluded. This structured approach ensured the final models were both statistically robust and explanatory."

Regarding SEM diagram and equations, we have added a comprehensive SEM diagram to the supplementary material (Supplementary Figure 1. Hypothesized structural equation model for high-speed rail (HSR) subgrade defects.), which includes all paths of the SEM integrated model and fully incorporates the hypotheses based on disaster system theory. This approach aims to help readers better understand the model structure and the complex relationships among variables. However, as you noted, the large number of variables may make the SEM diagram overly complex. Therefore, we recommend that readers refer to both Supplementary Figure 1 and Table 1. Table 1 provides detailed descriptions of all variables and their corresponding paths in the SEM diagram, making it easier to understand each variable's role and its specific relationships within the model.

We appreciate your suggestion to include detailed mathematical equations in the manuscript. After careful consideration, we believe that the revised model specifications, along with the figures and tables (Fig 2, Fig 3, Supplementary Figure 1, and Supplementary Table 3), sufficiently explain the relationships and directional paths between variables. Therefore, we chose to use SEM diagrams to visually present the model structure, as they more effectively demonstrate the causal relationships and interactions between variables. Additionally, the path coefficients and significance levels in the diagram clearly convey the strength of associations *and the logical coherence of the model. Including detailed mathematical equations might make the manuscript unnecessarily lengthy and complex, so we opted for clear visual representations to effectively communicate the model's structure and findings.*

Model Fit Indicators: The manuscript lacks a detailed interpretation of model fit indices and does not disclose any path modifications or their justifications.

The manuscript mentions, "In the structural model for subgrade defects, we...retained paths H1, H2, H6, H7, H8, H10, H11, and H13."

Points to address: Report model fit indices explicitly against established benchmarks (e.g., RMSEA < 0.08, CFI > 0.90). Disclose any modifications to path structures, providing justifications and discussing the impact on the overall model fit. Ensure transparency in adjustments made to improve the model.

[Response 3.13]: Thank you for your thorough review and valuable suggestions. We have carefully addressed your comments by providing a detailed explanation of the model fit indices and explicitly disclosing the modifications made to the path structures, including their justifications and the impact on the overall model fit.

Regarding model fit indicators, we have now provided a detailed interpretation of the fit indices in the Supplementary Methods, ensuring that they are reported in relation to established benchmarks. Specifically, we included the following indicators:

(1) Standardized Root Mean Square Residual (SRMR): SRMR is a standardized measure of the difference between observed data and model predictions[1]. It reflects the gap between the model's predicted mean square residuals and the standardized residuals' root mean square. The smaller the SRMR value, the better the model fits the observed data. Generally, SRMR < 0.1 indicates an acceptable model fit, while SRMR < 0.05 suggests an excellent fit.

(2) Root Mean Square Error of Approximation (RMSEA): RMSEA measures the degree of discrepancy between the model and an ideal model, considering model complexity and adjusting for sample size[2]. Smaller RMSEA values indicate a closer fit to the ideal model. Typically, RMSEA < 0.1 is considered an acceptable fit, and RMSEA < 0.05 indicates a model that fits the data very well.

(3) Goodness-of-Fit Index (GFI): GFI measures the proportion of variance in the observed data that is explained by the model. Higher GFI values indicate better model performance in explaining data variance. GFI > 0.9 is generally considered a good fit, while GFI > 0.95 indicates an excellent model fit.

(4) Adjusted Goodness-of-Fit Index (AGFI): AGFI is an adjusted version of GFI that takes model degrees of freedom into account, adjusting for model complexity. Higher AGFI values suggest better model performance in explaining the data. AGFI > 0.8 is usually seen as acceptable, while AGFI > 0.9 is considered a good fit.

(5) Comparative Fit Index (CFI): CFI is used to compare the fit of the proposed model to an independent baseline model[3]. A CFI value closer to 1 indicates a better fit. Typically, CFI > 0.9 indicates a good fit, and CFI > 0.95 suggests an excellent model fit.

(6) Tucker-Lewis Index (TLI): TLI compares the fit of the proposed model to an independent baseline model while considering model complexity. Higher TLI values indicate better model fit. TLI > 0.9 is considered good, and TLI > 0.95 indicates an excellent model fit.

(7) Normed Fit Index (NFI): NFI compares the fit of the proposed model to a null model. Higher NFI values indicate greater improvement over the null model. Generally, NFI > 0.9 is considered good, while NFI > 0.95 suggests an excellent model fit.
(8) Incremental Fit Index (IFI): IFI measures the improvement in model fit relative to an independent baseline model, accounting for model complexity. Higher IFI values suggest better model fit. IFI > 0.9 is considered good, and IFI > 0.95 indicates an excellent model fit.
(9) Relative Fit Index (RFI): RFI compares the fit of the proposed model to an independent baseline model. Higher RFI values indicate better model fit. Typically, RFI > 0.9 is considered good, and RFI > 0.95 indicates an excellent model fit.

Regarding path modifications and justifications, we have disclosed the changes to the path structures in Supplementary Table 3 "Detailed explanation of pathways (multi-defect model)" and assessed their impact on the overall model fit. As described, "The significance of the structural model was evaluated after fitting the paths according to Supplementary Figure 1. Some paths showed low significance (H3 at 0.817, H4 at 0.852, H5 at 0.699, H12 at 0.681, H15 at 0.995, and H9 at 0.062) and were sequentially removed. After removing these paths, the significance of H14 decreased from 0.02 to 0.232, leading to its exclusion as well. The remaining paths all met the significance criterion (H8 changed from 0.061 to 0.022 and was retained). Removing these paths significantly improved the model's fit indices. For example, the NFI increased from 0.853 to 0.878, the RFI from 0.821 to 0.856, the IFI from 0.858 to 0.884, the TLI from 0.827 to 0.862, and the RMSEA decreased from 0.122 to 0.109. These improvements indicate that the overall model fit was notably enhanced after path optimization. The final retained paths were H1, H2, H6, H7, H8, H10, H11, and H13, all of which met the significance standard (P < 0.05). Please note, this P-value does not represent the final significance of the paths (for the final path significance, please refer to Supplementary Table 2)."

[Ref 1] McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. Psychological Methods, 7(1), 64–82. https://doi.org/10.1037/1082-989X.7.1.64
[Ref 2] Raykov, T., Tomer, A. & Nesselroade, J. R. Reporting structural equation modeling results in Psychology and Aging: some proposed guidelines. Psychol. Aging 6, 499-503 (1991).
[Ref 3] Schreiber J. B., Nora, A., Stage, F. K., Barlow, E.A. & King, J. Reporting structural equation modeling and confirmatory factor analysis results: A review. J. Educ. Res. 99, 323-337 (2006).

Path Coefficients and Significance Testing: The significance tests for path coefficients lack detail, and the manuscript does not include sensitivity analyses or robustness checks.
The manuscript mentions "...factors listed in the supplementary materials Table 1."
Points to address: Provide detailed significance levels and confidence intervals for path coefficients. Perform and document sensitivity analyses and robustness checks to demonstrate the stability of the SEM results. Interpret these coefficients in light of their significance and overall model implications.
[Response 3.14]: Thank you for your detailed review and valuable suggestions. We have revised

the manuscript to include the requested details regarding path coefficients, significance testing, and robustness checks.

Regarding path coefficients and significance testing, we have added the confidence intervals at the 95% level for the measurement model in Supplementary Table 1 and provided more detailed significance levels. Additionally, we have introduced Supplementary Table 2 in the supplementary materials, which provides the 95% confidence intervals and path significance for the structural model.

Each path coefficient is now interpreted with detailed explanations of its significance and its implications for the overall model. For example, in the manuscript, we discuss the relationships between specific variables such as train frequency and subgrade defects, highlighting both the direction and strength of these associations. We have provided a detailed explanation of the coefficients in Supplementary Table 1, addressing their significance and overall model implications. As described, "(1) Path coefficients: These represent the estimated strength and direction of the relationship between two variables in a structural equation model. A positive coefficient indicates a positive relationship, while a negative coefficient indicates a negative relationship. For example, if the path coefficient between train frequency and subgrade defects is positive, it suggests that increasing train frequency directly contributes to a higher occurrence of subgrade defects. (2) 95% Confidence interval: The range within which the true path coefficient is expected to lie 95% of the time. The lower and upper bounds define the limits of this interval. A narrower confidence interval indicates the robustness of the model. (3) Standard Error: Measures the uncertainty of the estimate[1]. Smaller standard errors indicate more reliable estimates. (4) Critical ratio: The ratio of the estimate to its standard error, equivalent to the t-value[2-3], used for significance testing. Generally, a C.R. value greater than 1.96 indicates significance at the 95% confidence level. (5) P-value: Indicates the significance level of the path coefficient, with $P < 0.05$ generally considered significant. Significant P-values help to determine the main pathways of defect occurrence ($p < .05$, **$p < .01$, ***$p < .001$)."*

Regarding sensitivity analysis, it may not be necessary for this study due to the nature of Structural Equation Modeling (SEM), SEM is a statistical modeling technique based on observed data, with parameter estimates directly derived from the data rather than assumptions or external conditions. Unlike System Dynamics models (SD), SEM results are typically robust without requiring sensitivity analysis, provided the model structure is clear, and the data is stable. Moreover, the parameter estimates and path coefficients within SEM have already undergone rigorous statistical testing within the model, ensuring their robustness. Therefore, we did not perform additional sensitivity analyses.

Although sensitivity analyses are typically not required in structural equation modeling (SEM), we acknowledge the reviewer's concern regarding the importance of result robustness. Therefore, we have conducted a detailed robustness check using the Bootstrap sampling method. Specifically, in the 5000 Bootstrap samples, 0 samples were discarded due to a singular covariance matrix[4]. This indicates that the model handled the covariance matrix well, which is a key indicator of the model's robustness. Additionally, 0 Bootstrap samples were unused due *to a failure to find a solution, meaning that the model was able to converge and find a solution successfully in all samples. This further confirms the robustness of the model. Moreover, the target of 5000 Bootstrap samples was fully achieved, with no samples discarded or failed estimations due to errors. This ensures that all path coefficient standard errors and confidence intervals were reliably estimated with sufficient data support. Based on these results, we can confirm that the model's parameter estimates are stable and consistent across different sample conditions, supporting the significance of the path coefficients and the interpretability of the model.*

[Ref 1] Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P., Ray, S. An introduction to structural equation modeling. In: Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R. Classroom Companion: Business. Springer, Cham. https://doi.org/10.1007/978-3-030-80519-7_1 (2021).
[Ref 2] Bollen, K. A., & Davies, W. R. (2009). Causal indicator models: Identification, estimation, and testing. Structural Equation Modeling, 16(3), 498–522.
[Ref 3] Henseler, J., Hubona, G.S., & Ray, P.A. (2016). Using PLS path modeling in new technology research: updated guidelines. Ind. Manag. Data Syst., 116, 2-20.
[Ref 4] Yu, L., Zhao, P. & Zhou, W. Testing the number of common factors by bootstrapped sample covariance matrix in high-dimensional factor models. J. Am. Stat. Assoc. Early Access, DOI: 10.1080/01621459.2024.2346364 (2024).

Mediation and Bootstrapping Analysis: The mediation effect testing and bootstrapping procedures lack detailed procedural descriptions and rationale.
The manuscript mentions “Initially, the mediation effect code is tested; subsequently, the Bootstrap method is employed to perform 5000 resamples.”
Points to address: Include a detailed explanation of the mediation effect testing procedures and justify the selection of 5000 resamples in the bootstrap method. Describe the VB programming code used in AMOS, ensuring it is accessible for replication. Offer a thorough discussion of the implementation of bias-corrected and percentile methods.
[Response 3.15]: Thank you for your thorough review and valuable suggestions. We have added the mediation effect testing procedure in the Supplementary Methods, providing a detailed description of the process. As described, “Mediation Effect Testing Steps: (1) Model setup: Begin by drawing the final multi-defect and single-defect models in AMOS. (2) Effect output: In the “Analysis Properties” menu, select “Output” and check the “Indirect, direct & total effects” option to display these effects. (3) Bootstrap settings: Under the “Bootstrap” tab, enable “Perform bootstrap” and set the number of bootstrap samples to 5000, balancing computational time with precision. Check “Bias-corrected confidence intervals” and set the confidence level to 95%. (4) Custom formulas: Input custom formulas for indirect effects, such as “StdIndA = StdH10×StdH11×StdH6” represents the product of paths H10, H11, and H6. Here, StdH10 represents the path from geomorphological variables to economic variables, StdH11 indicates the path from economic variables to HSR operational characteristics, and StdH6 refers to the path from HSR operational characteristics to subgrade defects. (5) Mediation analysis: AMOS calculates bias-corrected bootstrap confidence intervals for mediation effects, providing lower (BCLI) and upper (BCUI) bounds. If the interval includes

zero, the mediation effect is not significant. If the interval excludes zero, the mediation effect exists."

Regarding sampling frequency justification, the purpose of the bootstrap analysis is to estimate the distribution of a statistic through resampling. Theoretically, the more resamples, the more stable and accurate the results. However, setting a high number of resamples increases computation time and resource consumption. To verify the stability of the statistical results with 5000 bootstrap resamples, we conducted additional tests on the multi-defect model with 4000, 5000, 6000, and 7000 resamples, comparing the respective confidence intervals. The corresponding results are presented in Supplementary Table 5: Distribution of 95% CI values under different bootstrap sampling times (multi-defect model). The validation results indicate that, starting from 4000 resamples, the variation in the confidence intervals generated by the Bias-corrected and Percentile methods becomes very small, and both the standard errors and effect values tend to stabilize. This demonstrates that 5000 resamples are sufficient to provide robust estimates, and additional resamples would not significantly improve the precision of the results. Therefore, 5000 resamples are a reasonable and efficient choice for our sample. Accordingly, we have added an explanation about the reasonableness of the resampling frequency in the Supplementary Methods: "To confirm the validity of using 5000 resamples, we tested with 4000, 6000, and 7000 resamples to observe the changes in the Lower and Upper bounds of the confidence intervals in the Bias-corrected method and the Percentile method. Based on Supplementary Table 5, resampling with 4000, 5000, 6000, and 7000 iterations showed slight variation in the confidence interval results, verifying the reasonableness of the chosen number of iterations." Additionally, we have added a description of the multiple resampling results in §3.1 (Socio-environmental causalities underpinning the occurrence of defects) of the main text: "Additional tests with different resample counts (4000, 5000, 6000, and 7000), shown in Supplementary Table 5, indicated consistent confidence intervals, confirming that 5000 resamples provided sufficient robustness."

Regarding description of VB code used, to enhance the readability of the manuscript and address your requests, we have reformatted the original code into equations to clearly present the mediation effect calculations. For example, in the multi-defect model, the mediation effects are calculated as follows: StdIndA=StdH10×StdH11×StdH6; StdIndB=StdH13×StdH2; StdIndC=StdH8×StdH2; StdIndD=StdH7×StdH6. Here, StdIndA represents the product of the paths H10, H11, and H6, where StdH10 refers to the path from geomorphological variables to economic variables, StdH11 to the path from economic variables to HSR operational characteristics, and StdH6 to the path from HSR operational characteristics to subgrade defects. Additionally, we used the AMOS Analysis Properties interface to select the bias-corrected and percentile methods with a 95% confidence level and set the number of bootstrap samples to 5000, which directly calculates the mediation effects in the SEM.

Regarding bias-corrected and percentile methods, we have thoroughly explained the implementation of bias-corrected and percentile methods in the Supplementary Methods section. As described, "(1) bias-corrected method: The bias-corrected method in Bootstrap analysis is used to adjust for the asymmetry of confidence intervals caused by skewed sample distributions.

This adjustment results in more accurate confidence intervals, especially important in mediation analysis where effect distributions may not be symmetrical[1]. AMOS software automatically applies this method, providing bias-corrected 95% confidence intervals. (2) percentile method: The percentile method, one of the simplest Bootstrap approaches, calculates confidence intervals based on the distribution of bootstrap samples[2]. It directly determines the upper and lower bounds of effect estimates, making it particularly useful for cases with unclear distributions or small sample sizes. AMOS includes this method, offering Percentile 95% confidence intervals. Combining the bias-corrected and percentile methods ensures the robustness of mediation effect estimates. The bias-corrected method adjusts for distributional biases, while the percentile method provides a straightforward confidence interval. Together, they offer a comprehensive understanding of the mediation effect's range and stability, ensuring reliable results across different distribution assumptions."

[Ref 1] Streukens, S., Leroi-Werelds, S. Bootstrapping and PLS-SEM: A step-by-step guide to get more out of your bootstrap results. Eur. Manag. J. 34, 618-632 (2016).
[Ref 2] Cheung, M. W.-L. Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. Struct. Equ. Modeling. 14, 227-246 (2007).

Residuals and Diagnostics: The manuscript does not mention diagnostic tests or residual analysis to detect model reliability issues.
The manuscript mentions "This critique refers to a general omission in the manuscript, thus, a direct quote is not provided."
Points to address: Perform and document diagnostic tests, including residual plots, variance inflation factors (VIF), and other measures to detect heteroscedasticity, autocorrelation, or multi-collinearity. Discuss the results of these diagnostics and their implications for the model's robustness.
[Response 3.16]: Thank you for your valuable suggestions. We fully recognize the importance of residual analysis in evaluating model fit and identifying potential issues in Structural Equation Modeling (SEM).

Regarding residual analysis in SEM, unlike traditional statistical models, SEM focuses on modeling the covariance matrix of observed variables rather than individual observations. As a result, traditional residual plots (e.g., the difference between observed and predicted values) are not applicable in SEM. To ensure model reliability, we have included Supplementary Figure 3, the residual covariance plot, in the supplementary materials. The residual covariance matrix shows the difference between the model-predicted covariance matrix and the observed covariance matrix. A well-constructed model will have residual covariance values close to zero. Analyzing the residual covariance plot helps us further evaluate the model's fit quality and robustness. Additionally, we have added an explanation of the residual covariance plot in §3.1(Socio-environmental causalities underpinning the occurrence of defects) of the main text. as follows: "As shown in Supplementary Figure 3, the residual covariance values are close to zero, suggesting that the model was well-constructed and that the predicted covariances aligned with the observed data."

Regarding multicollinearity and Variance Inflation Factor (VIF) diagnostics, we have expanded our discussion of VIF: One of the key advantages of Structural Equation Modeling (SEM) over traditional regression analysis is its inherent ability to address multicollinearity. SEM uses multiple indicators to describe latent constructs (unobserved variables), which prevents multicollinearity since each latent construct represents a distinct concept[1]. Therefore, we did not perform multicollinearity tests in the multi-defect model with latent variables. However, for the single-defect models, where latent variables are absent, we recognized the need to perform multicollinearity diagnostics. Following the methodology outlined in the Supplementary Methods (Methodology for single-defect model construction), we re-modeled the four independent defect models, removing variables with VIF values greater than 10. The updated results are reflected in the revised Figures 3 and 4, and the corresponding mediation effect values have been adjusted in Supplementary Table 8. We have also updated §3.2 (Varying effects of factor categories on different defects) and §4.2 (Comparison of defect occurrence patterns) of the main text to reflect these adjustments. Additionally, we have included a new Supplementary Table 7, which shows the VIF values retained in the final set for the four single-defect models, demonstrating that multicollinearity has been appropriately addressed. This ensures the robustness and accuracy of the independent defect models.

[Ref 1] Suhr, Diana D. "The Basics of Structural Equation Modeling." (2006).

Reproducibility: The study lacks sufficient detail for replication, including methodological steps, programming scripts, or data sources.
The manuscript mentions "The results of this study will provide a scientific basis for the maintenance of the railway system, thereby enhancing the safety and efficiency of HSR."
Points to address: Provide comprehensive methodological documentation, including clear descriptions of all steps and programming scripts used in the analysis. Ensure datasets and codes are made accessible through supplementary materials or repositories, facilitating independent replication.
[Response 3.17]: Thank you for your valuable suggestions. We recognize the importance of providing sufficient detail to ensure the study's reproducibility. In response, we have made several enhancements to the manuscript and supplementary materials to facilitate independent replication of our findings.

To provide a clearer and more comprehensive overview of our methodology, we have added a "Supplementary Methods" section in the supplementary materials, which provides comprehensive documentation of our methodology. This includes detailed steps for model construction, such as high-speed rail mileage calculation, methodology for multi-defect model construction, methodology for single-defect model construction, Structural Equation Modelling fit indicators, mediation effect testing steps, bias-corrected and percentile methods, and validation of resampling iterations for mediation effect. These sections supplement §2.3 (Methods) in the main text.

Regarding data sources, our data sources include subgrade defect data and variable data. The subgrade defect data is listed in Reference 17 of the main text and can be accessed via the DOI *link: https://doi.org/10.6084/m9.figshare.24086016. The factor data, which includes Climate Variables, Geomorphological Variables, Geological Variables, Land Use, and High-Speed Rail Proximity Index. The sources and processing methods for Economic Variables and High-Speed Rail Operational Characteristics are comprehensively described in §2.3.1 (Data preparation and processing steps).*

Regarding programming scripts, detailed descriptions of the programming scripts have been provided in response [3.15]. These scripts, along with the methodological steps, ensure that all analyses can be independently replicated.

Contextual Interpretation: The results are moderately interpreted within the specific conditions of China's HSR, but practical implications for HSR maintenance and operational strategies are underexplored.

The manuscript mentions "These findings enhance our understanding of the complexity of HSR subgrade defects and underscore the necessity for integrated management strategies."

Points to address: Delve deeper into the practical implications for HSR maintenance and operational strategies. Revisit the theoretical framework in the discussion to link the findings back to foundational principles and previous research, highlighting how this study advances understanding in the field.

[Response 3.18]: Thank you for your valuable suggestions. We have revisited the theoretical framework and linked our findings to foundational principles and previous research. In §4.4 (Policy recommendations) of the main text, we have further explored the practical implications for HSR maintenance and operational strategies.

Our study highlights the importance of incorporating land use patterns and geomorphological characteristics into the planning process to mitigate subgrade defects, advancing current understanding in HSR infrastructure design. As described in §4.4 (Policy recommendations), "Our study shows that adverse land use and geomorphological variables have standardized direct impact values of 0.320 and 0.383, respectively, on the co-occurrence of HSR subgrade defects. Therefore, when planning new railway routes, areas prone to subgrade problems, such as low-coverage grasslands, gobi deserts, saline-alkaline lands, and bare lands should be avoided. During the planning and construction phases, comprehensive geological and environmental assessments should be conducted to identify high-risk areas, such as high-altitude lines like the Lanzhou-Xinjiang HSR. In areas with unfavorable land use, climate-appropriate vegetation and soil improvement techniques can enhance stability. For Gobi or saline-alkaline areas, geotextiles and windbreaks can be used to mitigate the effects of weathering and salinization. For example, a study in Old Dalby, UK, using Electrical Resistivity Tomography (ERT), showed that grass cover can improve the stability of railway embankments."

In addition to land use and geomorphology, managing extreme temperatures is another crucial factor for ensuring HSR subgrade stability. As described in §4.4 (Policy recommendations), "Our study shows that extreme temperatures have a significant impact on the stability of high-speed rail subgrades during operation. Specifically, the number of days with maximum *temperature exceeding 35°C leads to the co-occurrence of subgrade defects, such as settlement, uplift, and mud pumping. Therefore, it is recommended to reduce train frequency during high-temperature periods by strengthening monitoring and early warning systems, in order to alleviate the load on the subgrade and track. Additionally, a high-temperature emergency response mechanism should be established to ensure that timely measures can be taken under extreme weather conditions to maintain railway safety. Based on long-term climate change forecasts, it is suggested that localized modifications be implemented in high-risk areas (with extended periods of high temperatures) to reduce subgrade deformation and structural fatigue under high-temperature conditions. Furthermore, our research indicates that an increase in the annual freezing days significantly raises the risk of subgrade frost damage (standardized total effect value of 0.651). Therefore, it is recommended to use materials with stronger anti-frost heave properties in permafrost regions (such as the Harbin-Dalian HSR) to ensure the long-term stability and safety of infrastructure in cold climates. For example, the Madrid-Cádiz railway line in Spain uses platforms to monitor foundation degradation in extreme climate conditions, with alarm thresholds were set trigger speed and frequency adjustments."*

Alongside climate management strategies, integrating causal analysis with data-driven methods offers an effective approach to enhancing HSR subgrade maintenance. As described in §4.4 (Policy recommendations), "Our research shows that causal analysis can identify key factor chains influencing subgrade defects (e.g., geomorphology affecting HSR operational characteristics, which in turn impact subgrade defects). Therefore, it is recommended to combine causal analysis with data-driven management by regularly collecting data on high-speed rail operations, environmental conditions, and infrastructure. By integrating these datasets with the key factors identified through causal analysis, data-driven methods (such as machine learning) can be applied to more accurately identify and predict potential subgrade defect risks. This integration not only improves the understanding of the interactions between various factors through causal chains, but also enhances the transparency and interpretability of data-driven management, avoiding the "black box" issue often associated with models. Additionally, optimizing predictions can enhance subgrade predictive maintenance, improving the efficiency and safety of railway infrastructure management while reducing unnecessary costs. For example, on the Zamudio railway bridge in the Basque region of Spain, machine learning algorithms were used to monitor railway infrastructure. Compared to traditional monitoring methods, this approach reduced costs and improved the efficiency of detecting structural defects."

Generalizability: The study does not adequately address the broader applicability of the findings, particularly to regions outside China or similar infrastructure systems.
The manuscript mentions "This critique refers to a general omission in the manuscript, thus, a direct quote is not provided."
Points to address: Discuss the geographic and temporal scope of the dataset and its broader applicability. Evaluate the potential for extrapolation to other regions or similar infrastructure systems and acknowledge limitations regarding generalizability. Propose future research directions to confirm and expand upon these findings.
[Response 3.19]: Thank you for your valuable suggestion. To enhance the generalizability of

our study, we have added a new paragraph in §4.3 (Limitations, future improvements and potential applications), which elaborates on the potential of applying our methodology in different regions or infrastructure types. As described, "While the research focused on China, its findings may be applicable to other regions with similar climates and geographical conditions, such as Southeast Asia and parts of Europe. The methods used can contribute to infrastructure maintenance and operational strategies globally. Moreover, as climate change was predicted to lead to more frequent extreme weather events[1], the study's insights on the impact of climate variables on subgrade defects offer valuable guidance for future infrastructure planning. In addition to HSR, these methods could also be adapted to assess other types of infrastructure, such as highways, bridges, and tunnels, under various environmental pressures. However, regional differences in geological and economic conditions may limit the direct applicability of these findings, indicating a need for further testing in diverse contexts. This study provides a theoretical foundation for improving global infrastructure resilience and informs strategies for managing the challenges posed by climate change."

Thank you again for your insightful recommendations. Future research will focus on risk assessments of subgrade defects under future climate conditions, aligning with your suggestions.

*[Ref 1] Stott, P. How climate change affects extreme weather events Research can increasingly determine the contribution of climate change to extreme events such as droughts. Science, **352**, 1517-1518 (2016)*

Implications and Recommendations: Policy recommendations lack empirical grounding, and the discussion on infrastructure resilience and future trends is brief and underdeveloped.

The manuscript mentions "The findings underscore the necessity for integrated management strategies."

Points to address: Expand on the policy recommendations with a detailed exploration of the empirical evidence presented. Discuss infrastructure resilience and future climate and usage trends in-depth, providing actionable, data-driven recommendations. Highlight the anticipated impact on HSR safety, efficiency, and maintenance, supported by robust empirical findings.

[Response 3.20]: Thank you for your feedback. We have expanded the discussion in §4.4 (Policy recommendations) by incorporating additional content on infrastructure resilience and adaptation to future climate conditions, along with recommendations for data-driven approaches.

Regarding infrastructure resilience and future climate conditions, as described in § 4.4 (Policy recommendations), "Our study shows that extreme temperatures have a significant impact on the stability of high-speed rail subgrades during operation. Specifically, the number of days with maximum temperature exceeding 35°C leads to the co-occurrence of subgrade defects, such as settlement, uplift, and mud pumping. Therefore, it is recommended to reduce train frequency during high-temperature periods by strengthening monitoring and early warning systems, in order to alleviate the load on the subgrade and track. Additionally, a high-temperature emergency response mechanism should be established to ensure that timely *measures can be taken under extreme weather conditions to maintain railway safety. Based on long-term climate change forecasts, it is suggested that localized modifications be implemented in high-risk areas (with extended periods of high temperatures) to reduce subgrade deformation and structural fatigue under high-temperature conditions. Furthermore, our research indicates that an increase in the annual freezing days significantly raises the risk of subgrade frost damage (standardized total effect value of 0.651). Therefore, it is recommended to use materials with stronger anti-frost heave properties in permafrost regions (such as the Harbin-Dalian HSR) to ensure the long-term stability and safety of infrastructure in cold climates."*

Regarding actionable, data-driven recommendations, as described in § 4.4 (Policy recommendations), " Our research shows that causal analysis can identify key factor chains influencing subgrade defects (e.g., geomorphology affecting HSR operational characteristics, which in turn impact subgrade defects). Therefore, it is recommended to combine causal analysis with data-driven management by regularly collecting data on high-speed rail operations, environmental conditions, and infrastructure. By integrating these datasets with the key factors identified through causal analysis, data-driven methods (such as machine learning) can be applied to more accurately identify and predict potential subgrade defect risks. This integration not only improves the understanding of the interactions between various factors through causal chains, but also enhances the transparency and interpretability of data-driven management, avoiding the "black box" issue often associated with models."

We have further expanded §4.4 (Policy recommendations) by incorporating empirical evidence to support each recommendation. As described, "For example, a study in Old Dalby, UK, using Electrical Resistivity Tomography (ERT), showed that grass cover can improve the stability of railway embankments[1]."
"For example, the Madrid-Cádiz railway line in Spain uses platforms to monitor foundation degradation in extreme climate conditions, with alarm thresholds were set trigger speed and frequency adjustments[2]."
"For example, on the Zamudio railway bridge in the Basque region of Spain, machine learning algorithms were used to monitor railway infrastructure. Compared to traditional monitoring methods, this approach reduced costs and improved the efficiency of detecting structural defects[3].

[Ref 1] Holmes, J. et al. (2022). 4D electrical resistivity tomography for assessing the influence of vegetation and subsurface moisture on railway cutting condition. Engineering Geology, 307. doi:10.1016/j.enggeo.2022.106790
[Ref 2] Sanz Bobi, J. D., Garrido Martinez-Llop, P., Rubio Marcos, P., Solano Jimenez, A., & Fernandez, J. G. (2024). Prediction of degraded infrastructure conditions for railway operation. Sensors (Basel), 24(8). doi:10.3390/s24082456
[Ref 3] Armijo, A., & Zamora-Sanchez, D. (2024). Integration of railway bridge structural health monitoring into the internet of things with a digital twin: A case study. Sensors (Basel), 24(7). doi:10.3390/s24072115

The comprehensiveness and representativeness of the 661 georeferenced event records are inadequately addressed. The study does not offer a detailed data collection methodology, including selection criteria and temporal range. There is insufficient assessment of geographical diversity to ensure a broad representation of environmental and operational conditions. Potential biases in the data collection process are neither identified nor mitigated.

The manuscript mentions “Subsequently, 661 georeferenced event records of eight defect types were selected, crossing provincial, municipal, county, township, and smaller scales.”

Points to address: Provide a detailed methodology section specifying the criteria for selecting the 661 event records.- Ensure a broad temporal range is covered to represent various environmental conditions over time.- Justify the geographical diversity of the selected records to cover a wide range of environmental and operational conditions.- Identify and discuss potential biases in the data collection process and the steps taken to mitigate them.

[Response 3.21]: Thank you for your thorough review and valuable suggestions. We acknowledge the importance of providing a clear explanation of our data collection methodology and ensuring the comprehensiveness and representativeness of the dataset. In response to your concerns, we have elaborated on the data collection process in [Response 3.9], which includes detailed criteria for selecting the 661 event records, considerations for temporal and spatial dimensions, and a discussion on the representativeness and potential biases in the dataset, along with the steps taken to mitigate these biases.

The normalization process using Moon's method lacks detailed justification. There is no information on how extreme values or outliers were managed, which could potentially distort relationships among variables. Details on the preprocessing steps, mainly the split intersecting lines function in ArcGIS, are insufficiently documented, raising concerns about reproducibility.

The manuscript mentions “We adopted Moon's method of using the normalize module to standardize the factors, eliminating the impact of differing magnitudes on analysis results.”

Points to address: Justify the choice of Moon's method over other normalization techniques suitable for this dataset. Explicitly describe how extreme values or outliers in the dataset were identified and managed. Provide a detailed step-by-step account of the preprocessing procedures, ensuring that all methods, including the ArcGIS split intersecting lines function, are accurately documented and reproducible.

[Response 3.22]: Thank you for your thorough review and valuable feedback. We acknowledge the need to justify our methodological choices and enhance the transparency of our process. In response, we have expanded on our rationale for selecting Moon's normalization method over other techniques in [Response 3.10]. We also detailed how extreme values and outliers were identified and managed to prevent distortion of variable relationships. Furthermore, we provided a step-by-step description of the preprocessing procedures, including the ArcGIS split intersecting lines function, to ensure that our study is fully transparent and reproducible.

The article does not outline steps for validating and cleaning the data. There is no evidence of procedures to handle missing values or inconsistencies. Additionally, the manuscript lacks proof of robust quality control measures, such as cross-validation with external datasets or triangulation from multiple sources, which are crucial for ensuring data reliability and accuracy.

The manuscript mentions “This critique refers to a general omission in the manuscript, thus, a direct quote is not provided.”

Points to address: Outline the steps taken to validate and clean the dataset. Describe approaches for handling missing values and inconsistencies within the dataset. Implement and document robust quality control measures, such as cross-validation with external datasets and triangulation from multiple data sources, to ensure the dataset's reliability and accuracy. Provide examples or references to established methodologies used in these processes.

[Response 3.23]: Thank you for your thorough review and valuable feedback. We have addressed the concerns related to data validation and cleaning in [Response 3.11], where we provided a comprehensive explanation of the steps taken for data validation, including the cleaning process and quality control measures implemented to ensure data reliability. Additionally, since our analysis was conducted at the district and county level with relatively large spatial resolution, the data used was sufficient to fully cover all analysis areas, with no missing values or inconsistencies. This has been elaborated on in [Response 3.11].

The SEM includes a range of latent variables, but the justification for choosing specific variables over others remains insufficient. The specification of observed variables is not comprehensively discussed, which raises concerns about their ability to capture the constructs of interest altogether.

The manuscript mentions "Using SEM, we can significantly enhance the understanding of defect formation mechanisms, accurately identify key influencing variables and effectively support disaster risk assessment and management decisions."

Points to address: Clearly articulate the rationale for selecting the particular latent and observed variables used. Provide a more detailed SEM diagram to visualize all the relationships and pathways considered in the model. Ensure that the selected variables comprehensively capture all key constructs of interest. Explain any exclusions explicitly and justify them based on theory or prior research.

[Response 3.24]: Thank you for your careful review and valuable feedback. We have addressed the concerns in [Response 3.12], providing a detailed explanation for the selection and exclusion of variables. Additionally, we have added "Supplementary Figure 1. Hypothesized structural equation model for high-speed rail (HSR) subgrade defects" to better illustrate the relationships and pathways among the variables. Given the large number of variable subclasses, including all of them would have made the SEM diagram overly complex. Therefore, we recommend that readers refer to both Supplementary Figure 1 and Table 1 together. Table 1 and Supplementary Figure 1 provide detailed descriptions of all the variables and their corresponding paths in the SEM diagram, allowing readers to clearly understand the role of each variable and its specific relationships within the model. We have ensured that the selected variables comprehensively capture the key constructs of interest and have explicitly justified any exclusions based on theory or prior research.

The article reports model fit indices but lacks thorough interpretation against established thresholds. There is no mention of any modifications made to improve the model fit, nor is there a discussion of their justifications and impact.

The manuscript mentions "We eliminated the latent variable of geological factors as they had minimal impact on the occurrence of subgrade defects and did not improve the model fit."

Points to address: Report the model fit indices explicitly against established benchmarks (e.g., Chi-square, RMSEA < 0.08, CFI > 0.90).- Discuss any modifications made to the model, providing clear justifications and evaluating their impact on the overall model fit.- Include the interpretation of each fit statistic and explain why they indicate that the model is a good fit or what additional improvements could be made to meet the desired thresholds.

[Response 3.25]: Thank you for your thorough review and valuable feedback. This comment aligns with the points already addressed in [Response 3.13], where we have provided a detailed explanation of the selection and comparison of model fit indices against established benchmarks (e.g., Chi-square, RMSEA < 0.08, CFI > 0.90). Additionally, we have explained the modifications made to improve the model fit in Supplementary Table 3 "Detailed explanation of pathways (multi-defect model)" and assessed their impact on the overall model fit. As described, "The significance of the structural model was evaluated after fitting the paths according to Supplementary Figure 1. Some paths showed low significance (H3 at 0.817, H4 at 0.852, H5 at 0.699, H12 at 0.681, H15 at 0.995, and H9 at 0.062) and were sequentially removed. After removing these paths, the significance of H14 decreased from 0.02 to 0.232, leading to its exclusion as well. The remaining paths all met the significance criterion (H8 changed from 0.061 to 0.022 and was retained). Removing these paths significantly improved the model's fit indices. For example, the NFI increased from 0.853 to 0.878, the RFI from 0.821 to 0.856, the IFI from 0.858 to 0.884, the TLI from 0.827 to 0.862, and the RMSEA decreased from 0.122 to 0.109. These improvements indicate that the overall model fit was notably enhanced after path optimization. The final retained paths were H1, H2, H6, H7, H8, H10, H11, and H13, all of which met the significance standard (P < 0.05). Please note, this P-value does not represent the final significance of the paths (for the final path significance, please refer to Supplementary Table 2)."

The significance tests for path coefficients are mentioned but not detailed enough to evaluate thoroughly. There is also a lack of sensitivity analyses or robustness checks to ensure the stability of the SEM results.

The manuscript mentions "In the structural model for subgrade defects, we estimated and tested the hypothesized paths...while retaining paths H1, H2, H6, H7, H8, H10, H11, and H13 which all met the significance criterion (P<0.05)."

Points to address: Provide detailed significance levels and confidence intervals for each path coefficient. Perform and document sensitivity analyses and stability checks to demonstrate the robustness of the results across different model specifications and assumptions. Include interpretations of the direct, indirect, and total effects, ensuring that the significance tests are correctly applied and explained within the context of the structural model.

[Response 3.26]: Thank you for your thorough review and valuable feedback. This comment closely relates to [Response 3.14]. To address your concerns, we have provided our response as follows:

(1) Significance testing and confidence intervals: We have added the confidence intervals at the 95% level for the measurement model in Supplementary Table 1 and provided more detailed significance levels. Additionally, we have introduced Supplementary Table 2 in the supplementary materials, which provides the 95% confidence intervals and path significance for the structural model.

(2) Robustness and sensitivity analyses: Regarding sensitivity analysis, we have provided a detailed explanation in [Response 3.14]. For robustness analysis, we have conducted a robustness check using the Bootstrap sampling method. Specifically, in the 5000 Bootstrap samples, 0 samples were discarded due to a singular covariance matrix. This indicates that the model handled the covariance matrix well, which is a key indicator of the model's robustness. Additionally, 0 Bootstrap samples were unused due to a failure to find a solution, meaning that the model was able to converge and find a solution successfully in all samples. This further confirms the robustness of the model. Moreover, the target of 5000 Bootstrap samples was fully achieved, with no samples discarded or failed estimations due to errors. This ensures that all path coefficient standard errors and confidence intervals were reliably estimated with sufficient data support. Based on these results, we can confirm that the model's parameter estimates are stable and consistent across different sample conditions, supporting the significance of the path coefficients and the interpretability of the model.

(3) Additional bootstrap iterations for robustness verification of mediation effects:
To further verify the robustness of our results, we conducted multiple Bootstrap iterations (5,000 samples) to test the stability of the mediation variables across different random samplings. This provides further confidence that the results are stable under different model specifications and assumptions. The testing steps and results for this analysis have been detailed in [Response 3.15].

(4) Interpretation of direct, indirect, and total effects:
we have clarified the interpretation of the direct, indirect, and total effects within the structural model. As illustrated in Fig. 2, we have clarified the interpretation of direct, indirect, and total effects as: "The total effect value is the sum of the direct and indirect effects. The direct effect is represented by the path coefficients between variables, while the indirect effect is calculated by multiplying the coefficients of multiple paths[1]."

[Ref 1] Moi, D. A. et al. Human pressure drives biodiversity-multifunctionality relationships in large Neotropical wetlands. Nat. Ecol. Evol. 6, 1279-1285 (2022).

The article mentions the use of mediation effect testing and bootstrapping with 5000 resamples, but lacks detailed procedural descriptions and rationale, casting doubt on the appropriateness and correctness of implementation.
The manuscript mentions "This paper uses VB programming within AMOS to write mediation effect testing code...subsequently, the Bootstrap method is employed to perform 5000 resamples."
Points to address: Provide a comprehensive description of the mediation effect testing procedure, including the specific mediation models tested and how VB programming was utilized in AMOS. Justify the choice of 5000 resamples in bootstrapping, explaining why this number is sufficient to ensure statistical robustness.- Include detailed results of the bias-corrected and percentile methods, ensuring the Lower-Upper interval does not include zero for significant pathways. To illustrate these results, offer examples or visualizations (e.g., confidence interval plots).

[Response 3.27]: Thank you for your valuable suggestions. This review point is similar to the concerns addressed in [Response 3.15]. We have added the mediation effect testing steps in the Supplementary Methods, providing a detailed description of the process. As described, "Mediation Effect Testing Steps: (1) Model setup: Begin by drawing the final multi-defect and single-defect models in AMOS. (2) Effect output: In the "Analysis Properties" menu, select "Output" and check the "Indirect, direct & total effects" option to display these effects. (3) Bootstrap settings: Under the "Bootstrap" tab, enable "Perform bootstrap" and set the number of bootstrap samples to 5000, balancing computational time with precision. Check "Bias-corrected confidence intervals" and set the confidence level to 95%. (4) Custom formulas: Input custom formulas for indirect effects, such as "StdIndA = StdH10×StdH11×StdH6" represents the product of paths H10, H11, and H6. Here, StdH10 represents the path from geomorphological variables to economic variables, StdH11 indicates the path from economic variables to HSR operational characteristics, and StdH6 refers to the path from HSR operational characteristics to subgrade defects. (5) Mediation analysis: AMOS calculates bias-corrected bootstrap confidence intervals for mediation effects, providing lower (BCLI) and upper (BCUI) bounds. If the interval includes zero, the mediation effect is not significant. If the interval excludes zero, the mediation effect exists."

Regarding the issue of setting the number of bootstrap resamples, we have provided a detailed explanation in [Response 3.15]. we conducted additional tests on the multi-defect model with 4000, 5000, 6000, and 7000 resamples, comparing the respective confidence intervals. The corresponding results are presented in Supplementary Table 5: Distribution of 95% CI values under different bootstrap sampling times (multi-defect model). The validation results indicate that, starting from 4000 resamples, the variation in the confidence intervals generated by the Bias-corrected and Percentile methods becomes very small, and both the standard errors and effect values tend to stabilize. Therefore, 5000 resamples were a reasonable and efficient choice for our sample. Accordingly, we have added an explanation about the reasonableness of the resampling frequency in the Supplementary Methods: "To confirm the validity of using 5000 resamples, we tested with 4000, 6000, and 7000 resamples to observe the changes in the Lower and Upper bounds of the confidence intervals in the Bias-corrected method and the Percentile method. Based on Supplementary Table 5, resampling with 4000, 5000, 6000, and 7000 iterations showed slight variation in the confidence interval results, verifying the reasonableness of the chosen number of iterations." Additionally, we have added a description of the multiple resampling results in §3.1 (Socio-environmental causalities underpinning the occurrence of defects) of the main text: "Additional tests with different resample counts (4000, 5000, 6000, and 7000), shown in Supplementary Table 5, indicated consistent confidence intervals, confirming that 5000 resamples provided sufficient robustness."

There is no discussion on residuals or diagnostic statistics. Hence, potential issues like heteroscedasticity, autocorrelation, or multi-collinearity are unaddressed, compromising the model's robustness.
The manuscript mentions "This critique refers to a general omission in the manuscript, thus, a direct quote is not provided."
Points to address: Conduct residual analysis, including residual plots, to check for

heteroscedasticity or autocorrelation patterns. Calculate and report variance inflation factors (VIF) to assess and address multi-collinearity. Include any additional diagnostic tests used in SEM to evaluate the goodness-of-fit and robustness of the model. Discuss the implications of these diagnostics and any corrective measures taken.

[Response 3.28]: Thank you for your valuable suggestions. This review point aligns with the concerns addressed in [Response 3.16], where we provided a detailed explanation that covers residual analysis and multicollinearity diagnostics. To address heteroscedasticity and autocorrelation, we assessed fit of the multi-defect model using a residual covariance plot included in the supplementary Figure 3. Additionally, we assessed multi-collinearity by calculating variance inflation factors (VIF) analysis in the single-defect models without latent variables, removing variables with VIF values greater than 10. All diagnostic results and their impact on model robustness have been documented in the supplementary materials and updated figures.

The article lacks sufficient replication details, including clear descriptions of methodological steps and access to programming scripts or datasets.

The manuscript mentions "Thus, SEM synthesizes theory and data to analyze the direct and indirect effects between variables systematically."

Points to address: Provide exhaustive methodological documentation that details every step of the SEM analysis, from data preprocessing to final outputs.- Ensure all AMOS scripts, VB code, and any other programming used are included as supplementary materials or housed in an accessible repository.- Offer access to the datasets analyzed, or at least detailed descriptions of how the data can be acquired or constructed. This transparency is crucial for verifying results through independent replication.

[Response 3.29]: Thank you for your valuable suggestions. This comment closely relates to [Response 3.17]. To address your concerns, we have added a "Supplementary Methods" section in the supplementary materials, which includes comprehensive documentation of our methodology. It covers detailed steps for model construction, such as high-speed rail mileage calculation, methodology for multi-defect model construction, methodology for single-defect model construction, Structural Equation Modelling fit indicators, mediation effect testing steps, bias-corrected and percentile methods, and validation of resampling iterations for mediation effect. These sections supplement the "2.3 Methods" in the main manuscript.

Moreover, we explained how to access the datasets used in the study. Specifically, the subgrade defect data is accessible via the following DOI link: https://doi.org/10.6084/m9.figshare.24086016. The complete introduction to the defect data can be found in Reference 17 of the main text. Additionally, the factor data is detailed and supplemented in §2.3.1 of the main text. These references provide comprehensive guidance on how to acquire or construct the datasets, ensuring transparency and facilitating independent replication of the results.

The results are somewhat interpreted within China's HSR environment, yet the practical implications for maintenance and operational strategies are not thoroughly explored. The discussion often remains high-level without drilling into actionable insights that maintenance

engineers or policymakers could directly apply.
The manuscript mentions "These findings enhance our understanding of the complexity of HSR subgrade defects and underscore the necessity for integrated management strategies."
Points to address: Expand the discussion to include specific, actionable insights into HSR subgrade maintenance and operational strategies. Relate findings directly back to operational decisions such as scheduling, maintenance intervals, and design modifications. Draw explicit connections to the theoretical framework and previous research to ground these practical implications.
[Response 3.30]: Thank you for your valuable feedback. As mentioned in [Response 3.18] and [Response 3.20], we have thoroughly addressed the issue in the revised manuscript, expanding on the practical implications for HSR maintenance and operational strategies. To ensure the results have practical utility, we now highlight specific recommendations for HSR maintenance engineers and policymakers, including optimizing route planning, strengthening predictive maintenance, and reducing train frequency during high-risk periods. These practical insights are directly linked to the findings of our structural equation model and grounded in our theoretical framework. We explicitly tie them back to the relationships uncovered between important factors (e.g., environmental conditions, operational characteristics, and subgrade integrity) and draw on previous research to validate our recommendations. By integrating empirical evidence and theory, this ensures that the results are both relevant and applicable to real-world HSR operations and maintenance.

The study fails to thoroughly examine the generalizability of the findings beyond the specific dataset and contextual conditions of China's HSR. There is little discussion on whether these results can be applied to other geographic regions or similar types of railway infrastructure systems.
The manuscript mentions "This critique refers to a general omission in the manuscript, thus, a direct quote is not provided."
Points to address: Discuss the potential for extrapolating findings to other regions or similar HSR systems. Include a geographic, temporal, and operational scope analysis, specifying how these factors might influence defect patterns elsewhere. Acknowledge the study's transferability limitations and suggest future research to test these findings in different contexts, ensuring a clearer understanding of their broader applicability.
[Response 3.31]: Thank you for your valuable suggestion. As mentioned in [Response 3.19], we have addressed these issues in the revised manuscript. Below is a summary of the specific changes made:

(1) Geographic extrapolation analysis: While this study focuses on China's high-speed rail (HSR) system, we believe the findings and methodologies have geographic transferability. In particular, regions with similar climates and geographical features, such as Southeast Asia and parts of Europe, can benefit from the insights gained in this study. For example, the significant impact of unused land types like bare land and saline-alkaline land on subgrade defects observed in this study may similarly affect HSR infrastructure in regions with comparable environmental conditions. As described in §4.3 (Limitations, future improvements and potential applications): "While the research focused on China, its findings may be

applicable to other regions with similar climates and geographical conditions, such as Southeast Asia and parts of Europe. The methods used can contribute to infrastructure maintenance and operational strategies globally."

(2) Temporal extrapolation analysis: We believe the findings also have temporal generalizability. As global climate change intensifies, the frequency and severity of extreme weather events, such as high temperatures and extreme rainfall, are expected to increase in the future. The results of this study provide forward-looking insights for managing infrastructure under the evolving conditions brought about by climate change in other regions. As described in §4.3 (Limitations, future improvements and potential applications): "Moreover, as climate change was predicted to lead to more frequent extreme weather events, the study's insights on the impact of climate variables on subgrade defects offer valuable guidance for future infrastructure planning."

(3) Operational scope analysis: We believe the methods and findings of this study can also be applied to other types of infrastructure, such as highways, bridges, and tunnels. These infrastructures face similar risks from environmental pressures, and by adjusting the input variables, the model used in this study can more accurately predict the risks specific to these different infrastructures. This would provide scientific support for their design, construction, and maintenance. As described in §4.3 (Limitations, future improvements and potential applications): "In addition to HSR, these methods could also be adapted to assess other types of infrastructure, such as highways, bridges, and tunnels, under various environmental pressures."

(4) Transferability limitations: We acknowledge that significant differences in geological conditions, climate characteristics, and levels of economic development between regions may influence the manifestation and mechanisms of subgrade defects. Therefore, future research should further test whether the findings of this study are applicable to different environments and operational conditions. Cross-regional comparative studies are needed to deepen the understanding of how these factors operate in diverse contexts. As described in §4.3 (Limitations, future improvements and potential applications): "However, regional differences in geological and economic conditions may limit the direct applicability of these findings, indicating a need for further testing in diverse contexts."

Although policy recommendations are presented, they are not sufficiently grounded in empirical evidence. The discussion on infrastructure resilience and future trends is superficial and lacks in-depth analysis.
The manuscript mentions "The findings underscore the necessity for integrated management strategies."
Points to address: Provide more empirically grounded and detailed policy recommendations. Include specific strategies for infrastructure resilience, such as predictive maintenance schedules based on detailed environmental risk assessments. Discuss the impact of future climate changes and increased HSR usage on subgrade defects more comprehensively. Use data-driven insights to support these recommendations and detail how these strategies could be

practically implemented in HSR operational plans and maintenance protocols.

[Response 3.32]: Thank you for your valuable suggestions. In response to your comments and as detailed in [Response 3.20], we have significantly expanded on the policy recommendations by integrating empirical evidence.

(1) Infrastructure resilience strategies: We recommend that when planning new railway routes, a detailed environmental risk assessment should be conducted. For existing railways, preventive maintenance should be increased in high-risk environments. As described in §4.4 (Policy recommendations): "Our study shows that adverse land use and geomorphological variables have standardized direct impact values of 0.320 and 0.383, respectively, on the co-occurrence of HSR subgrade defects. Therefore, when planning new railway routes, areas prone to subgrade problems, such as low-coverage grasslands, gobi deserts, saline-alkaline lands, and bare lands should be avoided. During the planning and construction phases, comprehensive geological and environmental assessments should be conducted to identify high-risk areas, such as high-altitude lines like the Lanzhou-Xinjiang HSR. In areas with unfavorable land use, climate-appropriate vegetation and soil improvement techniques can enhance stability. For Gobi or saline-alkaline areas, geotextiles and windbreaks can be used to mitigate the effects of weathering and salinization."

(2) Impact of future climate change and increased HSR usage on subgrade defects: We have proposed operational strategies to mitigate these risks and recommend establishing an emergency response mechanism. As outlined in §4.4 (Policy recommendations): "Our study shows that extreme temperatures have a significant impact on the stability of high-speed rail subgrades during operation. Specifically, the number of days with maximum temperature exceeding 35°C leads to the co-occurrence of subgrade defects, such as settlement, uplift, and mud pumping. Therefore, it is recommended to reduce train frequency during high-temperature periods by strengthening monitoring and early warning systems, in order to alleviate the load on the subgrade and track. Additionally, a high-temperature emergency response mechanism should be established to ensure that timely measures can be taken under extreme weather conditions to maintain railway safety. Based on long-term climate change forecasts, it is suggested that localized modifications be implemented in high-risk areas (with extended periods of high temperatures) to reduce subgrade deformation and structural fatigue under high-temperature conditions. Furthermore, our research indicates that an increase in the annual freezing days significantly raises the risk of subgrade frost damage (standardized total effect value of 0.651). Therefore, it is recommended to use materials with stronger anti-frost heave properties in permafrost regions (such as the Harbin-Dalian HSR) to ensure the long-term stability and safety of infrastructure in cold climates."

(3) Data-driven insights supporting recommendations: We propose integrating causal analysis with data-driven methods (such as machine learning) to enhance defect management. As outlined in §4.4 (Policy recommendations): "Our research shows that causal analysis can identify key factor chains influencing subgrade defects (e.g., geomorphology affecting HSR operational characteristics, which in turn impact subgrade defects). Therefore, it is recommended to combine causal analysis with data-driven management by regularly collecting

data on high-speed rail operations, environmental conditions, and infrastructure. By integrating these datasets with the key factors identified through causal analysis, data-driven methods (such as machine learning) can be applied to more accurately identify and predict potential subgrade defect risks. This integration not only improves the understanding of the interactions between various factors through causal chains, but also enhances the transparency and interpretability of data-driven management, avoiding the "black box" issue often associated with models. Additionally, optimizing predictions can enhance subgrade predictive maintenance, improving the efficiency and safety of railway infrastructure management while reducing unnecessary costs."

In addition to the responses provided above in sections (1), (2), and (3) of this review, we have also included quantitative data and real case studies in §4.4 (Policy recommendations) of the main text to further support our policy recommendations.

So, in my observation, I recommend MAJOR REVISION.
[Response 3.33]: Thank you again for your detailed and constructive feedback which helped to improve the manuscript significantly.

Response letter

The authors would like to thank the three reviewers for their time and constructive comments on the manuscript. The manuscript entitled ‘Operational stress and environmental interactions: unveiling multi-defect patterns in China’s high-speed rail subgrades’ was submitted for consideration in *Communications Earth & Environment* (COMMSENV-24-1535B).

In general, we have made the following major changes to address the concerns raised:
- We have added spatial lag variables to account for spaital auto-correlation into SEM to improve R-squared values and explanatory power.
- To capture non-linear relationships and validate the SEM findings, we firther employed Random Forest (RF) method as a complementary analytical approach.
- An explanation has been added to justify the omission of dynamic load and vibration analysis in this study. Addtionally, we have proposed future research directions to address these aspects comprehensively.

We have also edited the text thoroughly to correct language errors and improve the flow of arguments.

Detailed response to each comment of the three reviewers is attached below.

In response to Reviewer #1’s comments

Reviewer #1 (Remarks to the Author):

All the mentioned issues have been addressed. I recommend that this paper be accepted for publication.
[Response 1.1] Thank you very much for your positive feedback and recommendation for acceptance. We truly appreciate your time and efforts in reviewing our paper and for your helpful comments, which have greatly improved the quality of our work.

In response to Reviewer #2’s comments

Reviewer #2 (Remarks to the Author):

1) Some contents including images have been generated using Generative Pretrained Transformers.
[Response 2.1] Thank you for your valuable feedback. Previously, to enhance the visual appeal and readability of the image summaries, we utilized image generation tools. However, upon review, we have thoroughly reviewed all image summaries and removed any images created using generative software. These images have been replaced with content generated through conventional methods, ensuring that all visuals meet the standards for scientific accuracy and traceability.

2) Figure captions lack proper reference to subfigures. It is hard to follow subfigure details.

[Response 2.2] Thank you for your valuable suggestions. We have made revisions to the captions in the main text to better describe the subfigures for clarity and to help readers follow the details of each subfigure more easily. Specifically, we have revised the captions to include detailed descriptions of standardized path coefficients, significance levels, correlations, and the role of each subfigure, thereby highlighting the key findings.

*For Fig. 2, the revised caption is as follows: "Fig 2. Variables analysis of the multi-defect model. (a). multi-defect pathways on the HSR subgrade: This subfigure depicts the final SEM assessing the impact of geomorphological variables, land use composition, climate variables, geological variables, economic variables, spatial lag, HSR operational characteristics, and the HSR proximity index on subgrade defects. Standardized path coefficients are indicated next to the one-sided directional arrows, representing the hypothesized causal relationships. Significant path coefficients are marked with asterisks, denoting levels of significance (*p < .05, **p < .01, ***p < .001). Gray arrows represent non-significant paths. The small double-sided arrows near the endogenous variables denote the covariance, and the large arrows between latent variables indicate correlation coefficients. (b). Standardized effect values of various variables on subgrade defects: This subfigure summarizes the total, direct, and indirect effect values for key variables influencing subgrade defects. The total effect value is the sum of the direct and indirect effects. The direct effect is represented by the path coefficients between variables, while the indirect effect is calculated by multiplying the coefficients of multiple paths."*

*For Figs. 3, we explicitly referenced subfigures using labels (e.g., (a), (b), (c), etc.) and provided detailed explanations for each subfigure. As described: "Fig 3 Subgrade single-defect model occurrence pathways: (a) Subgrade settlement, (b) Frost damage, (c) Uplift deformation, and (d) Mud pumping. Standardized path coefficients are displayed alongside the directional arrows, quantifying the hypothesized influences, where the arrow's thickness represents the relationship's strength. Notably, significant path coefficients are marked with asterisks, indicating levels of statistical significance (*p < .05, **p < .01, ***p < .001)."*

Finally, we added a detailed description for the newly included Supplementary Figure 5, as follows: "Supplementary Figure 5. Evaluation of the Single-defect model using Random Forest with AUC values and ROC curves: (a) Subgrade settlement, (b) Frost damage, (c) Uplift deformation, and (d) Mud pumping. The confidence intervals in the legends represent the standard deviations of all AUC values for each subfigure. Each "Individual ROC Curves" in the subfigures represents the aggregation of the ROC curves from the 5-fold cross-validation."

We hope these detailed revisions address your concerns and enhance the clarity and readability of our figures and their captions. Thank you again for your thoughtful feedback.

3) A low R squared value indicates that the fitting is poor. More non-linear fits need to be applied to further increase the goodness of fit.

[Response 2.3] We have carefully considered your suggestion to improve the goodness of fit and have implemented several strategies to enhance the model's performance. Following measures have been taken in response to your comment:

(1) Introducing spatial lag variables to improve R-squared values.

To enhance the R-squared values of the model, we re-examined the spatial characteristics *of defect data. Using the Spatial Autocorrelation (Moran's I) tool in ArcGIS, we found that defect occurrences and co-occurrences exhibited significant spatial autocorrelation. This spatial dependency could lead to biases in the SEM model, making it difficult to accurately reflect the relationships between variables. To address this issue, we incorporated spatial lag variables into the model to eliminate the interference caused by spatial autocorrelation. The results showed that the R-squared value of the multi-defect model increased from 0.205 to 0.914. For the single-defect models, the R-squared values for settlement, frost heave, swelling uplift, and mud pumping improved from 0.091, 0.329, 0.140, and 0.067 to 0.122, 0.496, 0.260, and 0.204, respectively. Compared to similar studies[1-2], these R-squared values are at an excellent level, demonstrating that the inclusion of spatial lag variables significantly enhanced the model's explanatory power and robustness.*

We have provided additional details on spatial autocorrelation analysis in the Supplementary Methods, as described: "Spatial autocorrelation calculation: HSR subgrade defects often exhibit spatial clustering, meaning that defects in one location are not independent of those in nearby locations. This spatial dependence, known as spatial autocorrelation, can lead to biased estimates in traditional statistical models, including SEM. To address this issue and accurately capture the spatial characteristics of subgrade defects, we performed a spatial autocorrelation analysis using the Spatial Autocorrelation (Moran's I) tool in ArcGIS. The conceptualization of spatial relationships was defined as Inverse Distance, with the distance method set to Euclidean Distance. This configuration ensures that nearby features have a stronger influence on each other, aligning with the spatial nature of defect data. Finally, the analysis generated Moran's I indices, along with corresponding Z-scores and P-values, for each type of defect: settlement, frost damage, uplift deformation, and mud pumping. These indices serve as the basis for determining whether spatial lag analysis should be performed."

Additionally, we provided details on the calculation of spatial lag values in the Supplementary Methods, as described: "Spatial lag value calculation: To capture spatial dependence in the subgrade defects, we constructed a spatial lag model. The process involved the following steps: (1) Coordinate system transformation: To ensure the accuracy of boundary distance calculations, the coordinate system of the geospatial data was reprojected into the Albers Equal-Area Projection (suitable for the China region, with units in meters). This projection minimizes distortion in distance measurements across the study area. (2) Spatial weight matrix construction: To quantify spatial relationships, a spatial weight matrix was constructed. A threshold of 100 meters was set to define spatial neighbors, meaning that two polygons were considered neighbors if the minimum distance between them was less than or equal to this threshold. For neighboring polygons, a binary weight of 1 was assigned, while non-neighboring polygons were assigned a weight of 0. To address variations in the number of neighbors across polygons and ensure comparability among observations, the spatial weight matrix was row-standardized. This approach ensures that the spatial relationships captured in the matrix are both meaningful and consistent. (3) Spatial lag model estimation: Each defect type (settlement, frost damage, uplift deformation, and mud pumping) was incorporated into the spatial lag model. Using the Maximum Likelihood Estimation (MLE) method, the spatial lag values were calculated, accounting for the influence of neighboring regions. (4) Model outputs: The model produced key outputs for each defect type, including the spatial lag coefficient (rho), along with its Z-score and P-value. These outputs were used to evaluate the *strength and statistical significance of the spatial lag effect."*

We also supplemented Supplementary Table 9 with the results of spatial autocorrelation (Moran's I index, Z-score, and P-value) and spatial lag analysis (spatial lag coefficient ρ, Z-score, and P-value) for subgrade defects.

Accordingly, we updated the main text as follows:

In §2.1 (Hypotheses on impact pathways), we added a description of spatial lag as a hypothesized pathway: "Moreover, it is important to recognize that spatial lag effects, as subgrade defects in regions with similar environmental, geological, and operational conditions, may potentially impact neighboring areas, thereby amplifying the overall risk of defect occurrence and exacerbating defect propagation[3]. Therefore, the potential impact of spatial lag on defect pathways corresponds to hypothesis H16."

In §2.3.1 (Data preparation and processing steps), we elaborated on the source of spatial lag data: "The spatial lag variables were obtained through the application of Moran's I measurement in ArcGIS to determine whether defects exhibited spatial autocorrelation. If spatial autocorrelation was confirmed, a spatial weight matrix was constructed, and the spatial lag values for each defect were calculated, as detailed in the Supplementary Methods."

In §3.1 (Socio-environmental causalities underpinning the occurrence of defects), we added the standardized path coefficients of spatial lag for subgrade defects: "Notably, in path H16, spatial lag factors exhibited a strong positive influence on subgrade defects, with a standardized path coefficient of 0.91."

In §3.1 (Socio-environmental causalities underpinning the occurrence of defects), we described the improvement in R-squared values for the multi-defect model after incorporating spatial lag: "By introducing the spatial lag variable, the model's R^2 increased from 0.205 to 0.914, indicating that the model explains 91.4% of the variance."

In §3.2 (Varying effects of factor categories on different defects), we described the improvement in R-squared values for the single-defect models: "By incorporating the spatial lag variable, the R^2 for the models of settlement, frost damage, uplift deformation, and mud pumping defects increased from 0.091, 0.329, 0.140, and 0.067 to 0.122, 0.496, 0.260, and 0.204, respectively."

In §4.1 (Characteristics and trends of defect co-occurrence), we added a discussion on the impact of spatial lag: "Spatial lag effects were found to have a particularly significant impact on the co-occurrence of subgrade defects, indicating that the influence of defects is not confined to a single location but is propagated spatially, leading to the expansion and co-occurrence of defects in neighboring areas. Spatial lag refers to the influence of one location's conditions on its surroundings, highlighting spatial dependencies that can reveal risks and patterns. While spatial lag itself cannot be eliminated, its effects, such as the spread and co-occurrence of defects, should be mitigated through proactive measures like enhanced monitoring, preventive reinforcement, and the integration of digital twin technology. These strategies can effectively manage spatial lag impacts, enhance infrastructure resilience, and reduce the risk of defect expansion."

In addition, we updated the corresponding figures in the main text and supplementary materials based on the latest model results, including Fig. 2 Variables analysis of the multi-defect model, Fig. 3 Subgrade single-defect model occurrence pathways, Fig. 4 Total effect values of the SEM for subgrade single-defect, and revised the corresponding result descriptions.

(2) Validation of model goodness-of-fit

In addition to evaluating the R-squared value, we further assessed the goodness-of-fit (GOF) indices of the SEM model, which are considered more reliable for evaluating model fit in statistical modeling. These indices can be categorized into two types: (1) Residual-based measures, such as SRMR, which is a standardized measure of the difference between observed data and model predictions, and RMSEA, which quantifies the degree of discrepancy between the model and an ideal model, considering model complexity and sample size; and (2) Comparative measures, including indices like GFI and AGFI, which evaluate the proportion of variance in the observed data explained by the model, with AGFI adjusting for model complexity, as well as CFI, TLI, NFI, IFI, and RFI, which compare the fit of the proposed model to baseline or null models, where higher values indicate better model performance. The validation results demonstrated that the goodness-of-fit indices for the SEM model all met statistical requirements, as detailed in Supplementary Table 6, indicating that the model exhibits excellent overall fit. While the R-squared value provides additional information about the model's explanatory power, it is not the decisive factor for evaluating model fit. Instead, the goodness-of-fit indices, which comprehensively assess residuals and comparative performance, are more robust and reliable for determining the quality of the model.

(3) Attempts with non-linear structural equation models

To further improve the R-squared value of the single-defect model, we constructed a non-linear structural equation model (Non-linear SEM) using R, focusing on exploring non-linear interactions between factors. Specifically, we introduced quadratic, cubic, and higher-order terms, and used an iterative approach to compare the changes in R-squared values before and after adding interaction terms to identify interactions that could significantly improve the R-squared value. For example, in the single-defect model, adding the quadratic interaction term between operating frequency and clay loam increased the R-squared value for settlement from 0.122 to 0.536. However, this improvement came at the cost of a substantial drop in the model's goodness-of-fit index (GFI), which decreased from 0.902 to 0.231, likely due to the increased complexity of the non-linear model, which reduced its ability to fit the observed data.

Further attempts to optimize the model, such as adding covariance connections between variables, successfully restored GFI from 0.231 to 0.901, but this may have weakened the clarity of the path relationships, resulting in a decrease in the R-squared value to 0.120. These results highlighted a fundamental trade-off between goodness-of-fit and explanatory power in the non-linear SEM, making it challenging to balance these two aspects effectively. Moreover, the ability of non-linear SEM to capture complex non-linear relationships appeared to be limited.

Additionally, the construction of the single-defect model focused solely on the direct causal relationships between observed variables, without accounting for abstract concepts or unobservable variables hidden in the background (i.e., latent variables were excluded during model construction). Furthermore, collinearity filtering was applied during the development of the model. As a result, the single-defect model's explanatory power for observed variables was inherently weaker compared to the multi-defect model, which incorporated a more comprehensive structure, including latent variables. Consequently, the R-squared values of the *single-defect model should be interpreted with caution and regarded as a reference rather than a definitive measure of explanatory power.*

Given these limitations, we ultimately decided not to adopt non-linear SEM as a solution. Instead, we introduced the random forest method, which is well-suited for capturing non-linear interactions between variables and addressing the limitations observed in the SEM approach.

(4) Supplementary analysis using the RF method:

To validate the robustness of the structural equation model (SEM) and capture potential non-linear interactions that may have been missed, we employed the random forest method as a supplementary analysis. This approach served two purposes: (1) to evaluate the predictive power of the variables identified by SEM, and (2) to explore non-linear relationships that SEM could not fully capture. First, random forest models were constructed using the variables selected by SEM, and their predictive ability was assessed. The average AUC values of the random forest models were 0.722 (multi-defect model), 0.798 (settlement), 0.976 (frost damage), 0.832 (uplift deformation), and 0.894 (mud pumping) (see Supplementary Figure 4 and Supplementary Figure 5). These results demonstrate that the variables identified by SEM exhibited strong predictive performance across both single-defect and multi-defect models. Next, the importance of variables in the random forest models was analyzed and compared with the total effect values obtained from SEM. The variable importance rankings in the random forest models (Supplementary Figure 6 and Supplementary Figure 7) showed good consistency with the SEM results, further confirming the rationality of SEM's theoretical assumptions, pathway design, and total effect calculations. This consistency underscores that the SEM results are robust and reliable. At the same time, RF complemented SEM by effectively capturing non-linear relationships and complex interaction effects that SEM, constrained by its linear modeling framework, could not represent. By identifying these relationships, RF expanded the explanatory power of the model, providing insights beyond those captured by SEM.

Correspondingly, we added an explanation in §2.3.3 (Single-defect model to distinguish the driving mechanisms of each defect), emphasizing the use of RF as a nonlinear complement to SEM, as described: "Additionally, since the SEM relies on linear fitting methods, it may overlook nonlinear relationships and complex interaction effects between variables. Therefore, we further employed Random Forest (RF) to validate variable importance and explore potential nonlinear relationships."

Correspondingly, we have provided the details of random forest model construction in the Supplementary Methods, as described: "Random forest model construction: (1) Data preparation: To ensure the effectiveness of model training and validation, the dataset was randomly split into a training set and a validation set in a 7:3 ratio. The training set was used to build the model, while the validation set was employed to evaluate the model's generalization performance. The dependent variable of the multi-defect model was obtained by aggregating single defects, including subgrade settlement, uplift deformation, and mud pumping. In addition, our study adopted a class imbalance control sampling design, ensuring a 1:3 ratio between presence and pseudo-absence samples. This process was repeated 30 times using guided resampling techniques, ensuring robust and unbiased results. To eliminate the influence of feature scaling, all input features were standardized using StandardScaler to have a mean of 0 and a standard deviation of 1. This ensured comparability among features during model *training. (2) Hyperparameter optimization: To improve the performance of the RF models, the Optuna framework was employed for hyperparameter optimization. Optuna is a Bayesian optimization-based automated tuning tool that efficiently searches the hyperparameter space and identifies the optimal parameter combinations. The optimized hyperparameters included the number of trees (n_estimators), tree depth (max_depth), node split and leaf sample requirements (min_samples_split and min_samples_leaf), the number of features considered for splits (max_features), and whether bootstrap sampling was used (bootstrap). The optimization objective was to maximize the mean AUC on the training set using 5-fold cross-validation, ensuring the selection of robust parameter configurations. A total of 300 trials were conducted to identify the best-performing hyperparameters, further improving the model's predictive performance. (3) Model training and evaluation: To evaluate the model's performance and reduce the risk of overfitting, two separate 5-fold cross-validations were conducted to ensure robust optimization and evaluation. The first 5-fold cross-validation, as described in step (2) Hyperparameter optimization, was used to identify the best-performing parameter configurations, with the mean AUC score as the evaluation metric. The second 5-fold cross-validation was employed to assess the generalization ability of the optimized models, where the training set was divided into five subsets, with four subsets used for training and the remaining subset for validation in each iteration. This process was repeated five times to ensure that each subset served as a validation set once. For each fold, the AUC (Area Under the Curve) was calculated, and ROC curves (Receiver Operating Characteristic curves) were plotted to visually assess the model's classification performance. Finally, the Random Forest model was retrained on the entire training set using the optimal hyperparameters to fully utilize all available data. This approach not only optimized the model's configuration but also ensured a comprehensive evaluation of its predictive performance and generalizability. (4) Feature importance analysis: After training the model, the importance of each feature was evaluated using the feature_importances_ attribute of the RF algorithm, which calculates the Gini importance for each feature. To ensure robustness, the feature importance values were averaged across 30 independently constructed models, resulting in the final ranking of the features. To visualize the results, a lollipop chart was used to display the importance of each feature."*

Correspondingly, we added a section in the main text under Section 3.2 Varying effects of factor categories on different defects to describe the results of the random forest analysis, as follows: *"Furthermore, factors included in SEM were analyzed using RF for 30 sampling analyses, which display the receiver operating characteristic (ROC) curves with average AUC values of 0.722 (multi-defect), 0.798 (settlement), 0.976 (frost damage), 0.832 (uplift deformation), and 0.894 (mud pumping) (Supplementary Figures 4 and 5). In the multi-defect RF model, results revealed spatial lag and HSR operational characteristics as the most critical factors driving the co-occurrence of subgrade defects (Supplementary Figure 6), consistent with SEM findings. In the single-defect model, RF identified spatial lag as the primary driver for all defect types, aligning closely with SEM. Both methods agreed on the importance of individual factors, such as construction length for settlement defects and annual freezing days for frost damage defects. However, RF assigned higher importance to certain variables, such as "Elevate" in uplift deformation defects and operating frequency in mud pumping defects, reflecting its strength in modeling variable interactions (Supplementary Figure 7). This consistency indicates that the factors identified through SEM not only capture key relationships* *within the dataset but also serve as reliable inputs for machine learning models, effectively improving the predictive performance and interpretability of these models (Supplementary Results and Discussion)."*

Accordingly, we included a section in the Supplementary Results & Discussion to analyze the differences between the results of RF and SEM. as follows: *"The complementary role of SEM and RF in defect modeling: Since RF can capture the complex nonlinearities and interaction effects that SEM fails to model, we introduced RF as a complement to SEM. In the multi-defect model, the importance analysis results of RF (see Supplementary Figure 6) indicate that spatial lag and HSR operational characteristics are the most critical variables driving the co-occurrence of subgrade defects. These results are highly consistent with the findings of SEM, further validating the significant role of spatial lag and operational characteristics in the co-occurrence of defects. However, RF also revealed that climate variables play a more prominent role. In SEM, the effects of climate variables are mainly transmitted to subgrade defects through indirect pathways (e.g., via land use composition), which weakens their direct effects in the total effect calculation. In contrast, RF directly evaluates the overall contribution of variables to the target outcome, highlighting the importance of climate variables. This result indicates that RF effectively supplements SEM in capturing direct effects.*

In the single-defect model, the analysis results of RF indicate that spatial lag is the primary driving factor for all types of defects, which is highly consistent with the conclusions of SEM. This consistency suggests that spatial lag has a strong direct effect that can be consistently identified by both methods. Furthermore, RF and SEM exhibit a high degree of agreement in the importance rankings of individual factors. For instance, in the case of settlement defects, both SEM and RF identified construction length as a significant factor. Similarly, for frost damage defects, SEM identified annual freezing days as a key factor, and RF ranked it as the second most important factor after spatial lag. However, for uplift deformation defects, RF assigned a higher importance to Elevate (0.177), while SEM assigned it a lower importance. This discrepancy arises because SEM, through its path decomposition, found that the direct effect of Elevate was weak, while its indirect effects, mediated by operating frequency, exerted a negative impact, thus reducing its total effect value. RF, on the other hand, captures the interactions between elevate and other variables (e.g., spatial lag or climate variables), thereby increasing its importance. Additionally, for mud pumping, RF also highlighted the significant role of operating frequency (0.235), further reflecting the strength of RF in modeling complex variable relationships.

In summary, the results of SEM and RF exhibit a high degree of consistency. Simultaneously, RF addresses SEM's limitations in modeling direct effects by capturing nonlinearities and interaction effects, whereas SEM provides a clear explanation of causal mechanisms through path decomposition. This complementary relationship not only enhances the predictive performance of the models but also deepens the theoretical understanding of defect mechanisms, offering a more comprehensive perspective for the integrated modeling and practical application of subgrade defect analysis."

As a result, the complementary use of SEM and the RF method demonstrated their respective strengths, with SEM providing a solid theoretical foundation and random forest effectively capturing non-linear interactions that SEM could not represent. This combined approach not only validated the robustness of the SEM model but also addressed its limitations *in handling non-linear relationships, thereby significantly enhancing the overall credibility and interpretability of our research conclusions.*

[Ref 1] Dai, Z. et al. Metallic micronutrients are associated with the structure and function of the soil microbiome. Nat. Commun. ***14****, 8456 (2023).*
[Ref 2] Radulescu, A. T. et al. Impact of factors that predict adoption of geomonitoring systems for landslide management. Land ***12****, 752 (2023).*
*[Ref 3]. Deng, T.-T., Wang, D.-D., Yang, Y., & Yang, H. Shrinking cities in growing China: Did high-speed rail further aggravate urban shrinkage? Cities.****86****, 210-219 (2019).*

4) Although the authors have added supplementary Figures, there are issues concerning correct captioning. Referring to Fig 1. and supplementary Figure 2, the caption should include 'disputed regions' and 'based on specifications for the content of public maps in China' since the readers of the article are not limited to a region.
[Response 2.4] Thank you for your valuable feedback. In response to your suggestion, we have revised the captions to include the source of the map boundaries, ensuring transparency for all readers. Specifically, the title of Fig. 1 has been updated to: "Fig 1. Spatial distribution of HSR subgrade defect records across various land use types in China (Map boundaries follow China's "Regulations on the Content Representation of Public Maps[1]")", and the title of Supplementary Figure 2 has been updated to: "Supplementary Figure 2. Distribution of design speeds for China's HSR (Map boundaries follow China's "Regulations on the Content Representation of Public Maps[1]")". Additionally, we have included citations for the Regulations on the Content Representation of Public Maps to ensure clarity regarding the source and rationale for the boundaries depicted.

We have made every effort to ensure that the maps are appropriately captioned and aligned with the article's focus, while adhering to the legal guidelines for map usage and presentation in our region. We appreciate your understanding on this matter.

[Ref 1]. Ministry of Natural Resources of the People's Republic of China. Notice of the Ministry of Natural Resources on Issuing the "Specifications for the Content Representation of Public Maps". The Central People's Government of the People's Republic of China. https://www.gov.cn/gongbao/content/2023/content_5752310.htm (2023).

5) Dynamic load and vibration analysis is a major part of operational stress. The article does not cover this. This could be added to manuscript or in limitations and future works.
[Response 2.5] Thank you for your valuable suggestion regarding dynamic load and vibration analysis. These aspects are recognized as critical components of operational stress, but were not included in the current study due to the relatively large research units, as the analysis is conducted at the district or county level. Acquiring highly localized and detailed data poses significant challenges.

To address this limitations, we have a discussion in §4.3 (Limitations, future improvements, and potential applications) and outlined their importance for future research. As described: "Lastly, this study does not explicitly incorporate the effects of dynamic load and vibration, despite their recognized influence on subgrade performance. This omission arises primarily *from the scope of this research, which is conducted at a county-scale level and encompasses a large spatial area, whereas dynamic load and vibration effects are typically localized phenomena that require highly detailed and specific data, which are challenging to obtain comprehensively at such a scale.*

Future research could address this limitation by focusing on specific local regions, integrating real-time monitoring data of dynamic load characteristics (e.g., train speed and axle load) and vibration features (e.g., frequency and amplitude). Utilizing numerical simulation methods, such as finite element analysis (FEA) and multibody dynamics modeling (MBD), would allow for quantifying the physical impact of dynamic loads and vibrations on subgrade performance. Combined with SEM, this approach could provide a more comprehensive understanding of the effects and causal pathways of multiple factors, including dynamic loads, environmental conditions, and material properties, on subgrade defects."

We agree that incorporating dynamic load and vibration analysis into future studies would significantly improve the accuracy, comprehensiveness, and predictive power of subgrade defect models. By integrating real-time data and advanced simulation methods, future research could more effectively quantify the physical impacts of dynamic loads and vibrations and their interactions with other factors influencing subgrade performance.

We appreciate your insight and believe that this proposed direction will add significant value to future work on HSR subgrade infrastructure. Thank you again for your helpful recommendation.

6) PINN and digital twin can provide a more depth and practical scenario. This could also be included in future scope.

[Response 2.6] Thank you for your valuable suggestion regarding PINN and digital twin technologies. We appreciate the potential of these technologies to provide deeper insights and practical applications for subgrade performance analysis. Recognizing their importance, we have identified the integration of advanced methodologies such as PINN into a digital twin framework as a promising direction for future research, as highlighted in §4.3 (Limitations, future improvements, and potential applications). As described: "Furthermore, physics-informed neural networks (PINN) could integrate physical laws and monitoring data to compute critical variables such as dynamic stress distribution, which can be incorporated as potential variables in SEM to analyze the complex interactions between dynamic loads and environmental factors affecting subgrade defects. These technologies could eventually be extended into a digital twin framework, enabling real-time monitoring and lifecycle prediction of subgrade performance. This would provide a scientific foundation for optimizing design and maintenance strategies." These technologies build on the findings of our current study by combining SEM variables with advanced physical modeling and real-time monitoring capabilities. They not only address the complexities of dynamic interactions but also enable predictive and real-time analysis, enhancing the practical applicability and depth of subgrade defect research.

Moreover, in §4.4 (Policy Recommendations), we recommend the establishment of a digital twin platform as a practical application of our findings on spatial lag effects. As described: "Establishing a digital twin platform: This study reveals that spatial lag variables significantly impact subgrade defects, with their spatial propagation characteristics potentially *leading to the expansion of defects in neighboring areas. Building on this finding, we recommend the deployment of sensors in high-risk zones to monitor key data in real time, such as train operations (e.g., speed and axle load), subgrade deformation, and environmental conditions. The collected data can then be integrated into a digital twin platform. This platform aligns with the United Nations Sustainable Development Goals (SDGs), particularly SDG 9 (Industry, Innovation, and Infrastructure) and SDG 11 (Sustainable Cities and Communities), by improving the reliability of high-speed rail operations, minimizing disruptions caused by subgrade defects, and enhancing connectivity in remote and underdeveloped areas. While this study focuses on the theoretical aspects of defect spatial propagation, the proposed digital twin platform provides a practical framework for future research and intelligent high-speed rail management. For example, a digital twin system was established on the Varzea Bridge in the Aveiro district of Portugal, improving the efficiency of fatigue assessment during operations[1]."*

By addressing these technologies across both policy recommendations and future research, we aim to provide a comprehensive roadmap for advancing subgrade defect analysis. While digital twin platforms enable real-time monitoring and dynamic simulation, the integration of PINN offers the potential to incorporate physical laws and predict dynamic stress behaviors, further enhancing the practical applicability and depth of subgrade defect research.

[Ref 1]. Nhamage, I.A.A. et al. Performing fatigue state characterization in railway steel bridges using digital twin models. Appl. Sci-Basel. ***13****, 6741 (2023).*

7) The unexplained variances in model needs to be reduced using feature selection and optimisation method to improve model accuracy. This should also include details on handling noise.

[Response 2.7] Thank you for your constructive feedback on the need to reduce unexplained variances using feature selection and optimization methods, as well as handling noise in the data. Below, we summarize the measures we have taken to address these issues:

(1) Reducing unexplained variances through feature selection and optimization

To minimize unexplained variances, we implemented a systematic approach during model construction:

-Initial screening: Variables with P-values greater than 0.05 were excluded to ensure statistical significance at the 95% confidence level. This process eliminated variables with low relevance or predictive power, such as certain soil types (e.g., silty clay, loam) and secondary land use categories (e.g., mudflats, reservoirs), as detailed in Supplementary Methods.

-Iterative path testing: During path model construction, we iteratively tested and excluded hypothesized paths with P-values exceeding 0.05. Variables with weak causal relationships or significance levels above 0.1 were also removed to enhance model clarity and parsimony.

-Feature selection based on model fit indices: To ensure model accuracy, we evaluated variables using fit indices such as SRMR, RMSEA, GFI, AGFI, CFI, TLI, NFI, IFI, and RFI. Variables that caused significant declines in these indices during the fitting process were excluded from the final model.

-Multicollinearity management: For single-defect models, we calculated the Variance Inflation Factor (VIF) to detect and exclude variables with VIF values greater than 10. This step ensured that only statistically independent variables were retained, improving model

stability and explanatory power.

These steps helped refine the model by focusing on the most statistically relevant factors, thereby reducing unexplained variances and enhancing the model's overall accuracy and explanatory capability.

(2) Handling noise in the data

To ensure data quality and mitigate noise, we implemented a combination of data preprocessing, statistical validation, and spatial noise handling:

-Data preprocessing: We conducted a thorough examination of all variables to identify potential outliers using the Interquartile Range (IQR) method, which detects values falling outside 1.5 times the IQR above the third quartile or below the first quartile. Upon review, no significant outliers were identified that could influence the results. Additionally, a comprehensive check for missing data revealed no gaps, eliminating the need for imputations or exclusions. These steps ensured the dataset's completeness, reliability, and suitability for accurate analysis.

-Bootstrap analysis: To validate mediation effects and minimize the impact of sampling noise, we applied a 4000-iteration Bootstrap resampling method, calculating both bias-corrected and percentile confidence intervals to adjust for distribution skewness and ensure reliable estimates. We further tested different iteration counts (3000, 4000, 5000, and 6000) to evaluate result stability, with confidence intervals showing minimal variation across these iterations, as detailed in Supplementary Table 5. This confirmed the robustness and reliability of the selected parameters.

-Addressing spatial noise: Using Moran's I analysis in ArcGIS, we identified significant spatial clustering of subgrade defects, highlighting spatial dependencies where defects in one location were influenced by neighboring regions. To address these dependencies, spatial lag variables were incorporated into the SEM model, quantifying the impact of neighboring regions and reducing biases caused by unmodeled spatial relationships, thereby improving the model's accuracy and explanatory power.

-Through these measures, the R^2 values of the multi-defect model improved significantly, increasing from 0.205 to 0.914. Likewise, the R^2 values for single-defect models showed notable enhancements: settlement increased from 0.091 to 0.122, frost damage from 0.329 to 0.496, swelling uplift from 0.140 to 0.260, and mud pumping from 0.067 to 0.204. These improvements are illustrated in Fig. 2 (Variables analysis of the multi-defect model) and Fig. 3 (Subgrade single-defect model occurrence pathways). Furthermore, the SEM fit indices met all statistical requirements, as shown in Supplementary Table 6 (Goodness-of-fit measures [GOFs] for the structural equation models). These results demonstrate the effectiveness of the feature selection and optimization methods in reducing unexplained variances and enhancing the model's accuracy and explanatory power.

(3) Verification and future directions for model improvement

While these improvements are substantial, some unexplained variances remain due to the inherent limitations of linear modeling and the complexity of interactions among factors. To address these limitations, we employed the random forest method as a supplementary analysis. This approach served two main purposes: (1) to validate the predictive power of the variables *identified by SEM, and (2) to explore non-linear relationships that SEM may not fully capture due to its linear framework. The random forest models demonstrated strong predictive performance, with average AUC values of 0.722 (multi-defect model), 0.798 (settlement), 0.976 (frost damage), 0.832 (uplift deformation), and 0.894 (mud pumping) (see Supplementary Figure 4 and Supplementary Figure 5). Additionally, the variable importance rankings in the random forest models were consistent with the total effect values identified by SEM (see Supplementary Figure 6 and Supplementary Figure 7), confirming the robustness of the theoretical assumptions and pathways used in SEM. These findings highlight the complementary value of random forest methods in capturing non-linear interactions and providing additional explanatory power beyond what SEM alone can achieve.*

Building on these insights, future research aims to address the remaining gaps by incorporating additional dynamic variables, such as train speed, axle load, and vibration characteristics, which are critical for understanding operational stress on subgrade performance. Furthermore, advanced modeling approaches, such as physics-informed neural networks (PINN), could integrate physical laws and real-time monitoring data to compute critical variables (e.g., dynamic stress distribution), enabling a deeper analysis of the interactions between dynamic loads and environmental factors. By combining these methods with SEM, future studies could refine the understanding of complex relationships and causal pathways, ultimately reducing unexplained variances and enhancing predictive accuracy.

In response to Reviewer #3’s comments

Reviewer #3 (Remarks to the Author):

I am happy with the revisions administered by the authors. However, I would like to see sustainable implication and societal or community development by the findings of the research. Keeping a trend for Sustainable development, authors are advised to keep a section for this. It will not only enhance the overall paper's readability but also can gather substantial readership in near future.

So, I would suggest addition of these implications will give a bost to the paper.

[Response 3.1] Thank you very much for your efforts in reviewing our work and for your valuable suggestion. In response, we have included a dedicated discussion in §4.4 Policy Recommendations to highlight the practical and societal implications of our findings, explicitly linking them to the United Nations Sustainable Development Goals (SDGs). As described, "Establishing a digital twin platform: This study reveals that spatial lag variables significantly impact subgrade defects, with their spatial propagation characteristics potentially leading to the expansion of defects in neighboring areas. Building on this finding, we recommend the deployment of sensors in high-risk zones to monitor key data in real time, such as train operations (e.g., speed and axle load), subgrade deformation, and environmental conditions. The collected data can then be integrated into a digital twin platform. This platform aligns with the United Nations Sustainable Development Goals (SDGs), particularly SDG 9 (Industry, Innovation, and Infrastructure) and SDG 11 (Sustainable Cities and Communities), by improving the reliability of high-speed rail operations, minimizing disruptions caused by subgrade defects, and enhancing connectivity in remote and underdeveloped areas. While this

study focuses on the theoretical aspects of defect spatial propagation, the proposed digital twin platform provides a practical framework for future research and intelligent high-speed rail management. For example, a digital twin system was established on the Varzea Bridge in the Aveiro district of Portugal, improving the efficiency of fatigue assessment during operations[1]*."*

[Ref 1]. Nhamage, I.A.A. et al. Performing fatigue state characterization in railway steel bridges using digital twin models. Appl. Sci-Basel. ***13****, 6741 (2023).*