Supplementary Information for

Climate-driven impacts in renewable resources alter green hydrogen supply

Jinhui Ren[1], Qianzhi Zhang[1], Shulei Zhang[2,3], Shu Zhang[1], Wenying Chen[1*]

1 Institute of Energy, Environment and Economy, Tsinghua University, 100084 Beijing, China
2 State Key Laboratory of Loess Science, Institute of Earth Environment, Chinese Academy of Sciences, 710061 Xi'an, China
3 University of Chinese Academy of Sciences, 101408 Beijing, China

* Corresponding author: Wenying Chen (chenwy@tsinghua.edu.cn)

Supplementary Note 1. Two-way ANOVA for decomposing variance across techno-economic and meteorological dimensions

We employ a two-way analysis of variance (ANOVA) to decompose the variance of modeled outcomes in the joint sample generated from techno-economic and meteorological sampling. The full simulation dataset is constructed by independently varying (i) techno-economic parameters (e.g., capital expenditures, conversion efficiencies, financing assumptions) and (ii) meteorological inputs (e.g., multi-year wind and solar resource profiles). The Cartesian combination of these two sampling dimensions yields a factorial structure in which each observation corresponds to a unique pair of techno-economic settings and meteorological conditions.

Let each simulation be indexed by a pair (i, j), where $i = 1, \dots, I$ denotes the techno-economic scenario and $j = 1, \dots, J$ denotes the meteorological realization. For each combination (i, j), the model produces an output y_{ij} such as levelized cost of hydrogen (LCOH), energy yield, or financial returns. The two-way ANOVA model is written as

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij},$$

where μ is the grand mean, α_i captures the effect of techno-economic variation, β_j captures the effect of meteorological variation, $(\alpha\beta)_{ij}$ represents potential interactions between the two factor dimensions, and ε_{ij} is the residual error.

The ANOVA decomposition partitions the total variance of y_{ij} into contributions from each of these components. In particular, the sum of squares associated with the meteorological factor β_j quantifies the portion of variability that arises solely from differences in wind and solar resource conditions, conditional on the techno-economic sampling. Similarly, the interaction term captures how meteorological sensitivity depends on techno-economic settings. This decomposition allows us to extract and report the variance component attributable to meteorological drivers in isolation, thereby providing a clear measure of the extent to which project outcomes are shaped by climate-related variability rather than by techno-economic assumptions.

This framework ensures that, after jointly sampling across both dimensions, the meteorological contribution to overall uncertainty can be identified and quantified separately, providing a transparent characterization of the role of weather and climate variability in determining long-term project performance.

Supplementary Note 2. Construction of Feature-Engineering Variables

To systematically characterize the climatic and techno-economic drivers relevant for project-level hydrogen production, we construct a set of engineered features derived from the raw physical and cost inputs. The feature-engineering procedure aims to translate multi-dimensional climate statistics and technology parameters into a compact and interpretable set of metrics that can capture the key mechanisms affecting hydrogen production, variability, and system integration costs. The resulting features fall into five categories: (i) direct technology-cost terms, (ii) wind-resource statistics, (iii) solar-resource statistics, (iv) temperature-related metrics, and (v) interaction terms combining techno-economic parameters with climate variability or firmness. The full feature-construction logic is summarized below.

1. Techno-economic cost variables

Four variables capture the baseline capital cost assumptions for photovoltaic systems, wind turbines, battery storage, and electrolyzers. These direct cost variables (d_cost_pv, d_cost_wind, d_cost_storage, d_cost_elec) are included without transformation, as they enter the techno-economic model linearly and serve as fundamental drivers of levelized hydrogen cost. These terms provide the baseline against which the climatic modifiers operate.

2. Wind-resource statistics

Wind features are derived from summary statistics of the 100-m wind-speed distribution. Each metric is selected to represent a distinct physical property of the wind resource relevant for energy conversion:

- **Typical wind resource level** (wind_level)
  Represented by either the median or mean wind speed. This feature captures the central tendency of the resource and is the primary determinant of expected wind power output.
- **Wind variability** (wind_var)
  Defined as the temporal variance of wind speed. Variability affects both renewable intermittency and storage requirements.
- **Peakiness** (wind_peakiness)
  Constructed as the difference between maximum and mean wind speed. This metric captures the presence of short-duration extreme events and the skewness of wind-speed distributions.
- **Firmness** (wind_firmness)
  The 10th-percentile wind speed serves as a proxy for low-generation conditions that drive backup requirements and influence electrolyzer utilization.

These features collectively summarize the mean–variance–extremes structure of the wind-speed distribution without relying on raw time series.

3. Solar-resource statistics

Solar features mirror the construction used for wind and are derived from broadband surface shortwave radiation (ssrd):

- **Typical solar level** (ssrd_level): median or mean radiation
- **Solar variability** (ssrd_var): variance of irradiance
- **Solar peakiness** (ssrd_peakiness): max – mean
- **Solar firmness** (ssrd_firmness): 10th percentile

These features capture irradiance availability, intermittency, and the presence of short-lived high-irradiance peaks that disproportionately influence PV output.

4. Temperature-related metrics

Temperature affects both electrolyzer efficiency and thermal stress on system components. We extract:

- **Hot-temperature exposure** (t_hot): 90th-percentile temperature
- **Cold-temperature exposure** (t_cold): transformed 10th-percentile temperature (sign-flipped for intuitive interpretation)
- **Temperature spread** (t_spread): q75 – q25, representing the breadth of typical temperatures

These metrics summarize thermal conditions affecting both technology performance and operational reliability.

5. Interaction terms combining technology costs with climatic conditions

To account for the fact that the impact of climate conditions is mediated by technology costs, we construct interpretable interaction features:

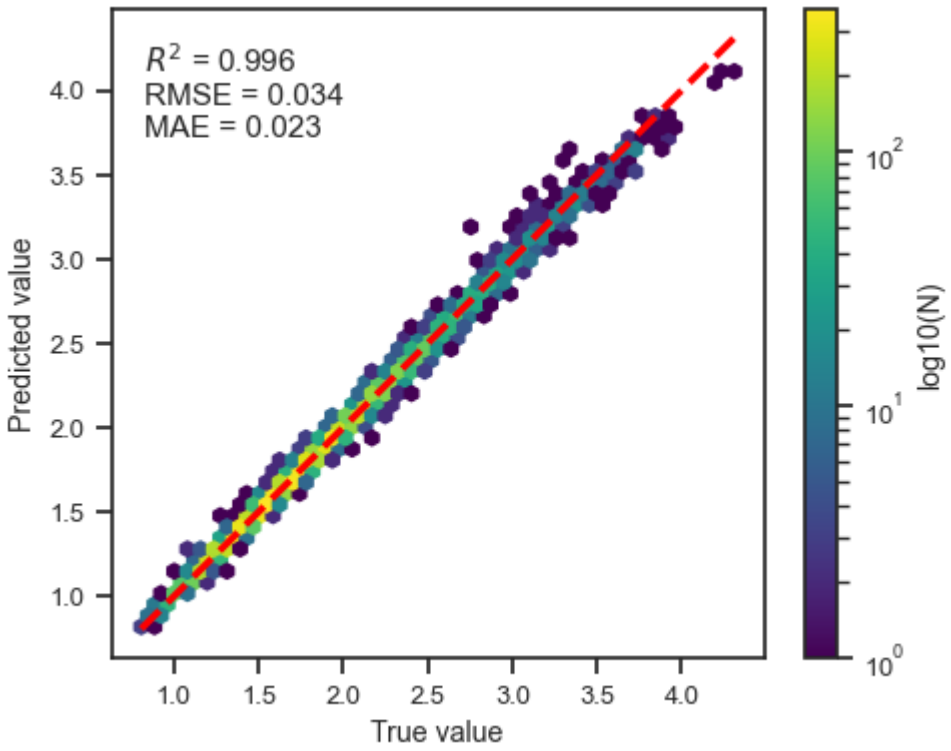
- **PV cost × solar level** (pv_cost_x_level): captures how solar availability interacts with PV cost assumptions.
- **Wind cost × wind level** (wind_cost_x_level): analogous term for wind.
- **Storage cost × variability** (storage_cost_x_variab): quantifies how the economic penalty of variability depends on storage costs.
- **Electrolyzer cost × firmness** (elec_cost_x_firmness): captures the sensitivity of electrolyzer economics to low-output periods.

These terms ensure that models can represent how climate statistics influence the marginal value of each technology under cost uncertainty.

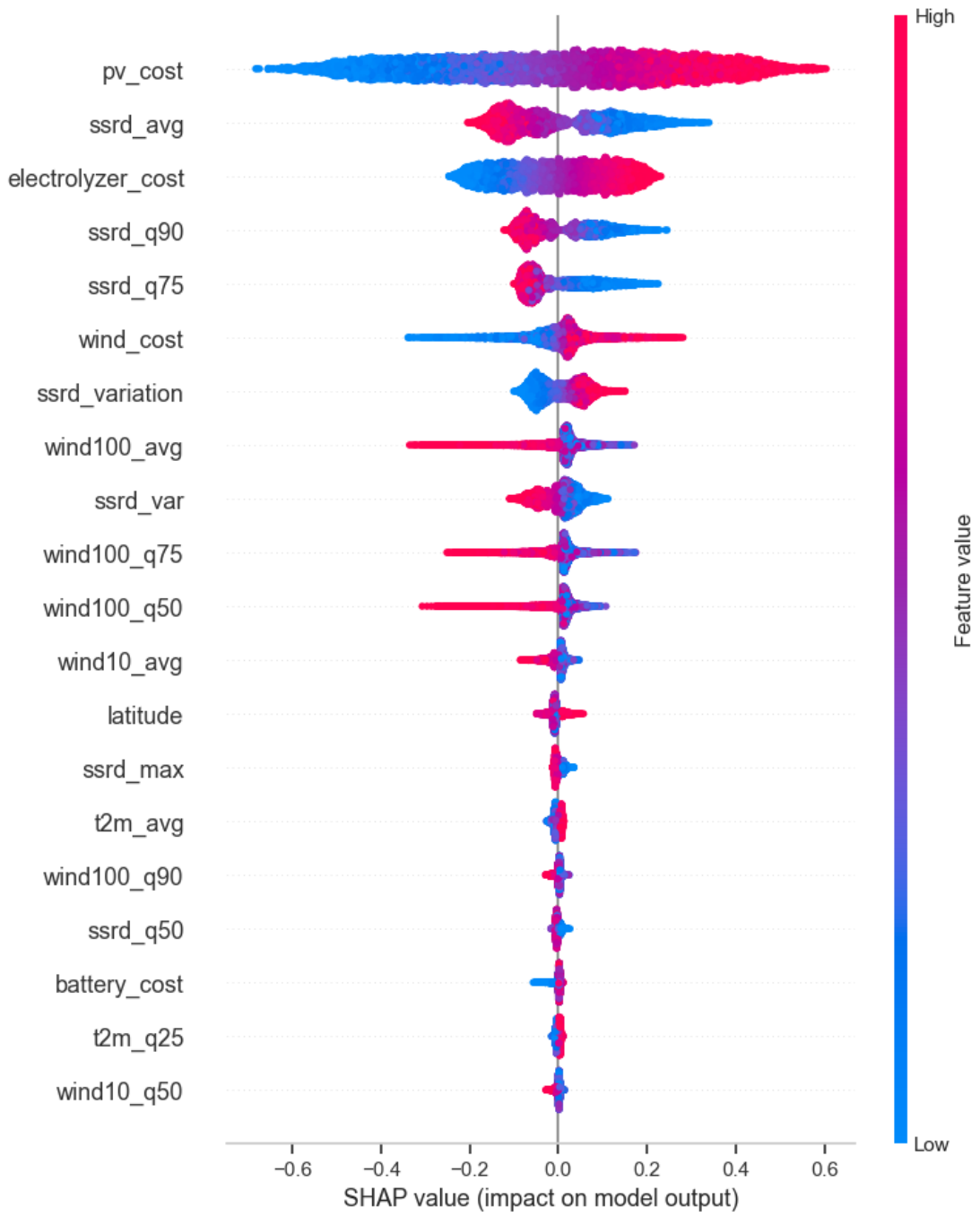
6. Final feature selection and grouping

A minimal, non-redundant feature set is retained by selecting one representative feature for each physical concept (typical level, variability, peakiness, firmness, temperature extremes, and interaction structures). Features are then assigned to groups (Tech-cost, Wind, Solar, Temperature, Interaction) to facilitate grouped SHAP interpretation and structured visualization.

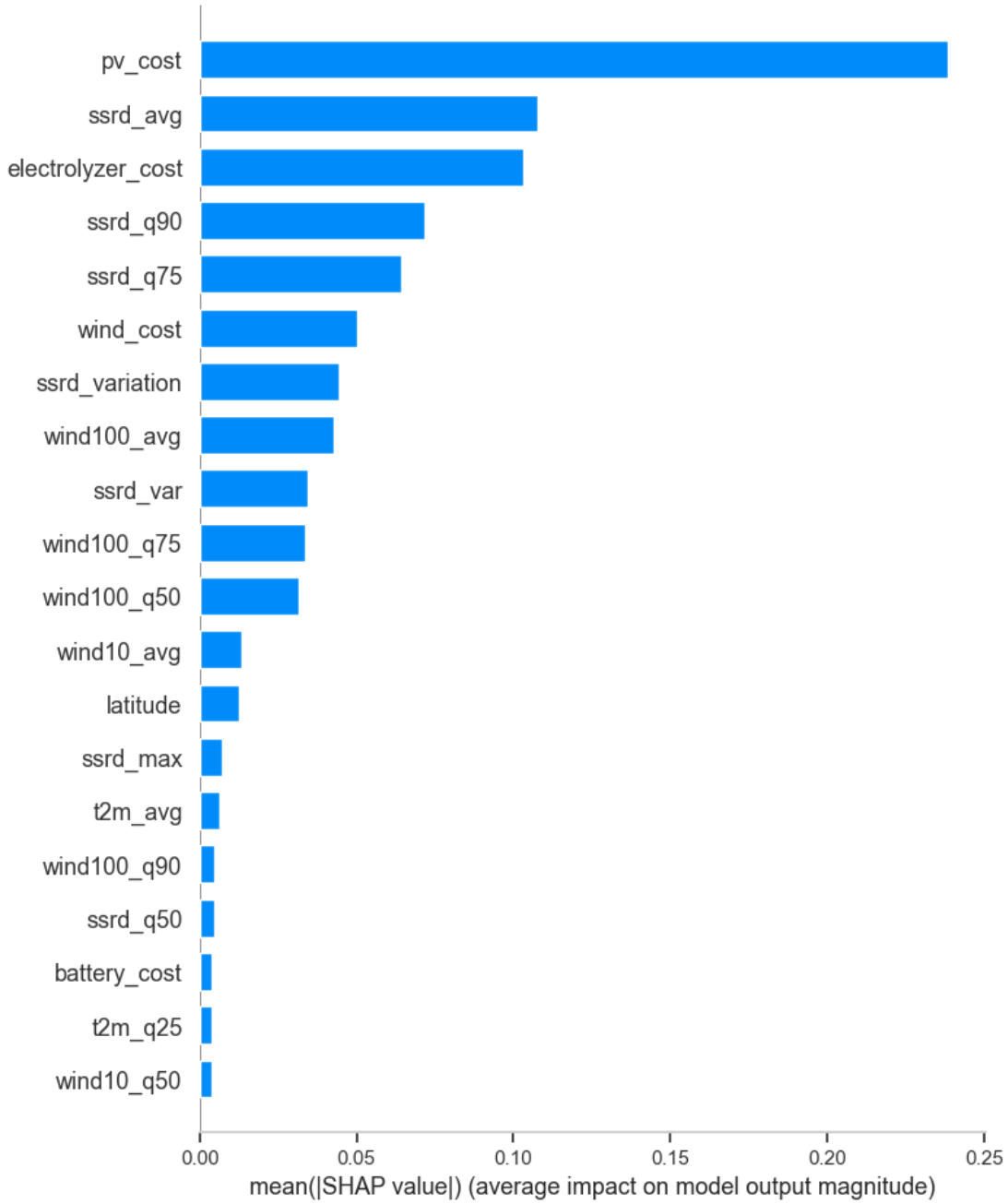
The final engineered dataset therefore provides a compact but physically interpretable representation of both climatic and techno-economic determinants of hydrogen-system performance. This structured feature set enables transparent attribution analysis, supports model generalization across climate and cost scenarios, and preserves interpretability while substantially reducing the dimensionality of the raw inputs.



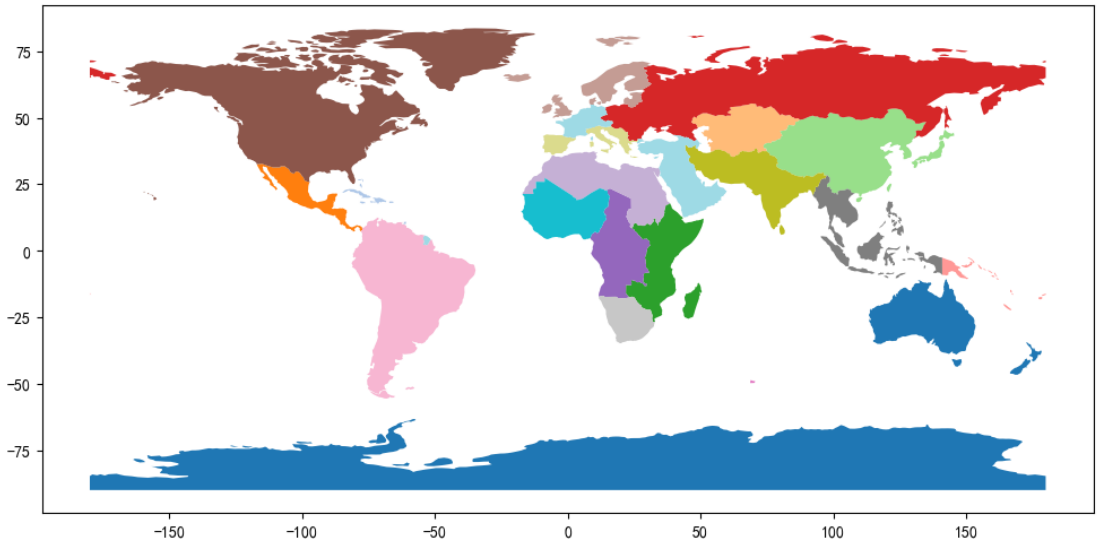
Supplementary Fig. 1 | Model performance evaluation results on the test set. Density scatter plot comparing the predicted versus actual (optimized) values of LCOH. The color gradient reflects the density of data points, where colors closer to yellow indicate a higher concentration of samples. The red dashed line represents the $y = x$ line. Key performance metrics, including the coefficient of determination (R^2) and root mean square error (RMSE), are reported in the top-left inset.



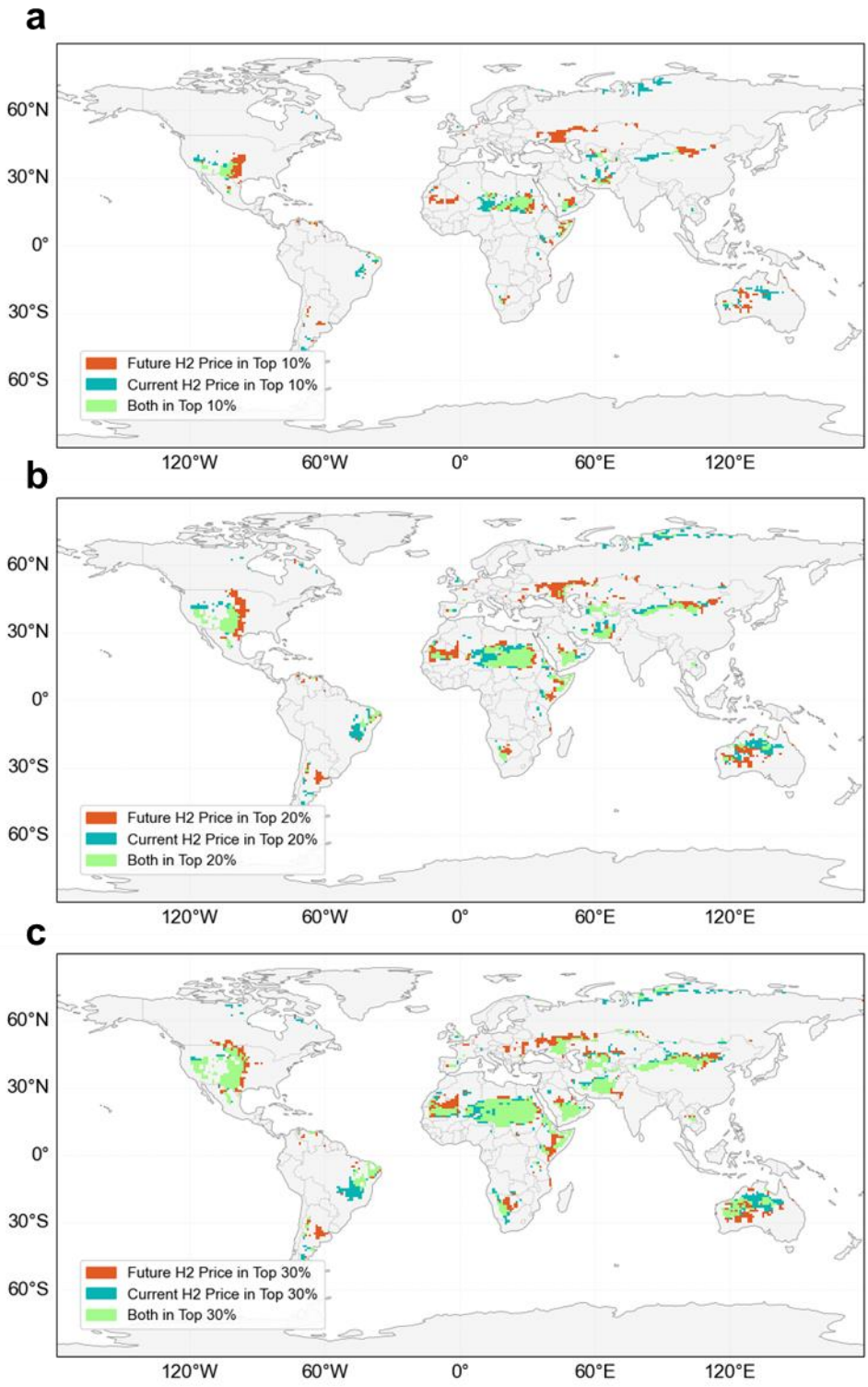
Supplementary Fig. 2 | SHAP analysis of key drivers. The summary plot illustrates the distribution of SHAP values for the top features, representing their respective contributions to the LCOH. The horizontal axis indicates the impact on the model output: positive SHAP values correspond to an increase in LCOH, while negative values indicate a reduction. Features are ranked by their global importance.



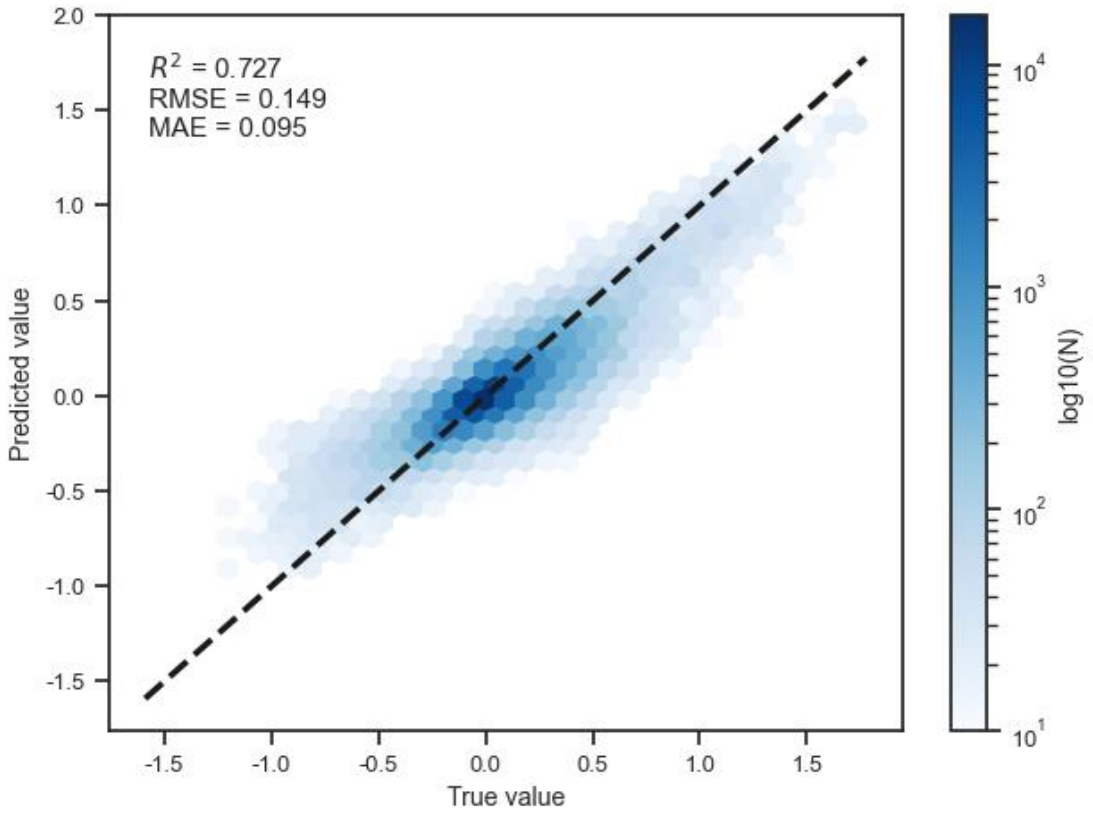
Supplementary Fig. 3 | Global feature importance ranking. The bar chart displays the top input variables sorted by their mean absolute SHAP values. A longer bar indicates a greater average impact of the feature on the predicted LCOH magnitude.



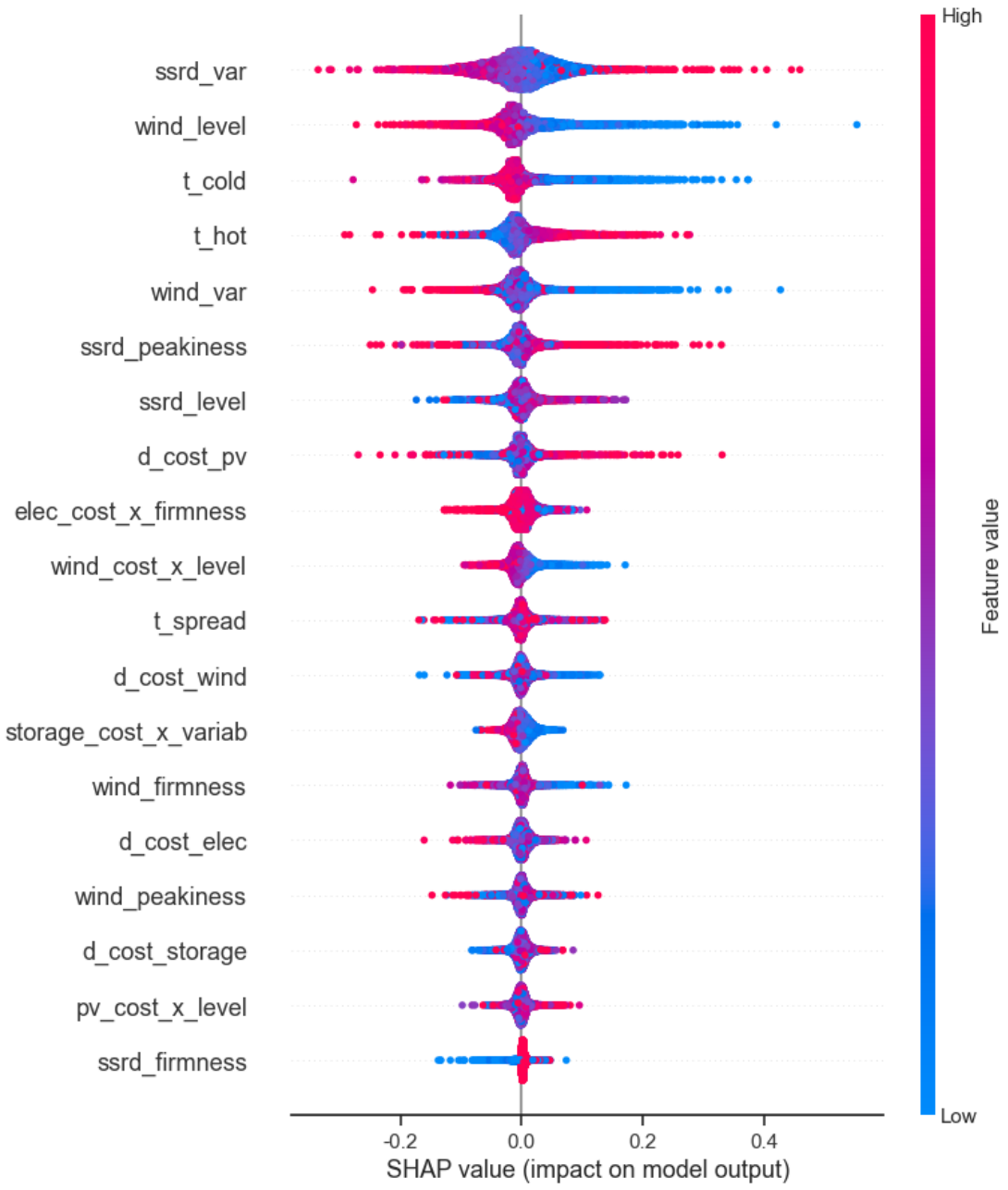
Supplementary Fig. 4 | The 22 United Nations-defined statistical subregions. The map illustrates the geographical boundaries of the subregions used for regional aggregation and analysis in this study.



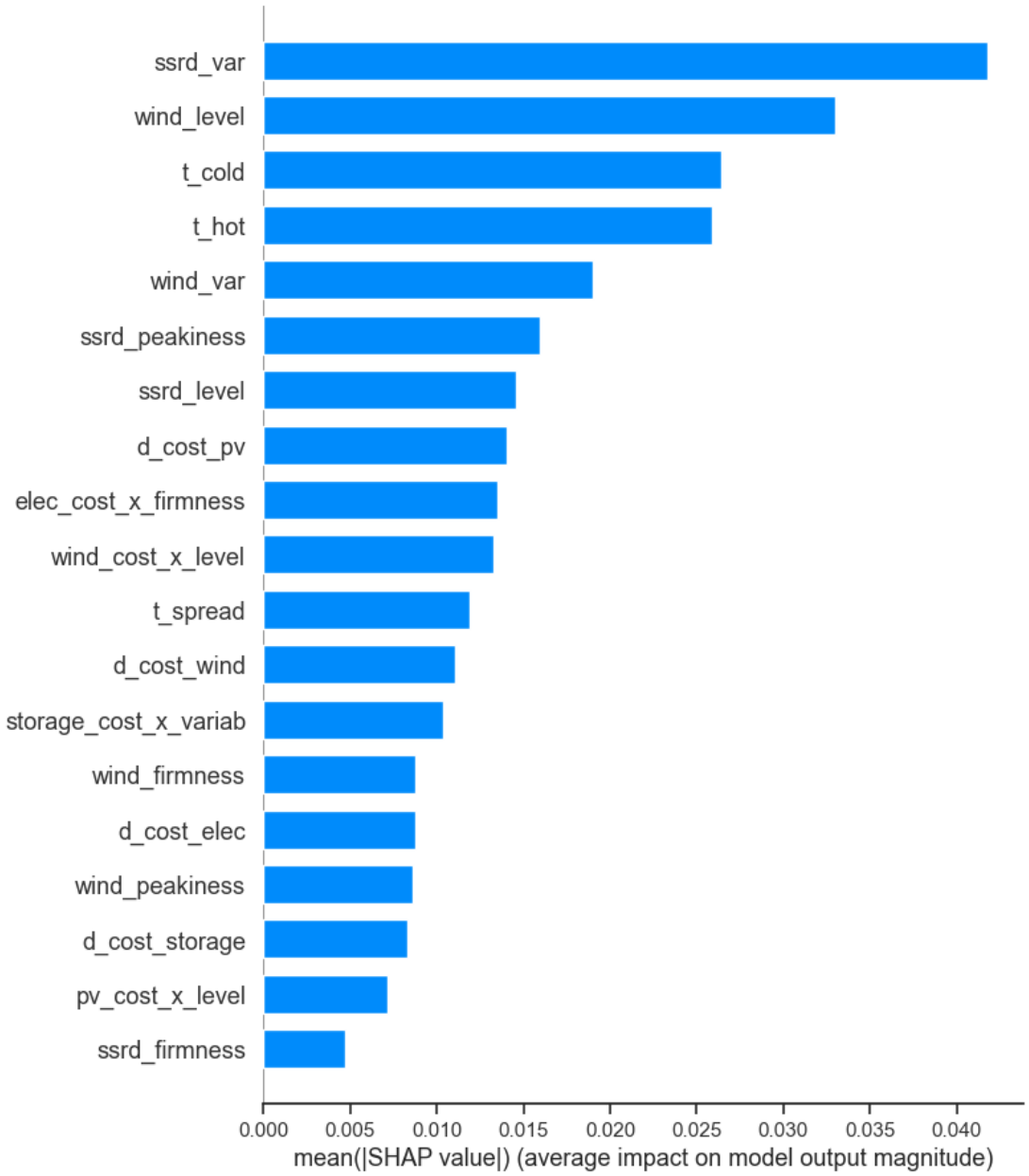
Supplementary Fig. 5 | Global distribution of the lowest-cost hydrogen regions under current and future climate conditions. The maps identify the United Nations-defined statistical subregions with the most competitive levelized cost of hydrogen (LCOH). **a**–**c**, Spatial extent of subregions ranking in the top 10% (**a**), 20% (**b**), and 30% (**c**) of the lowest costs, respectively.



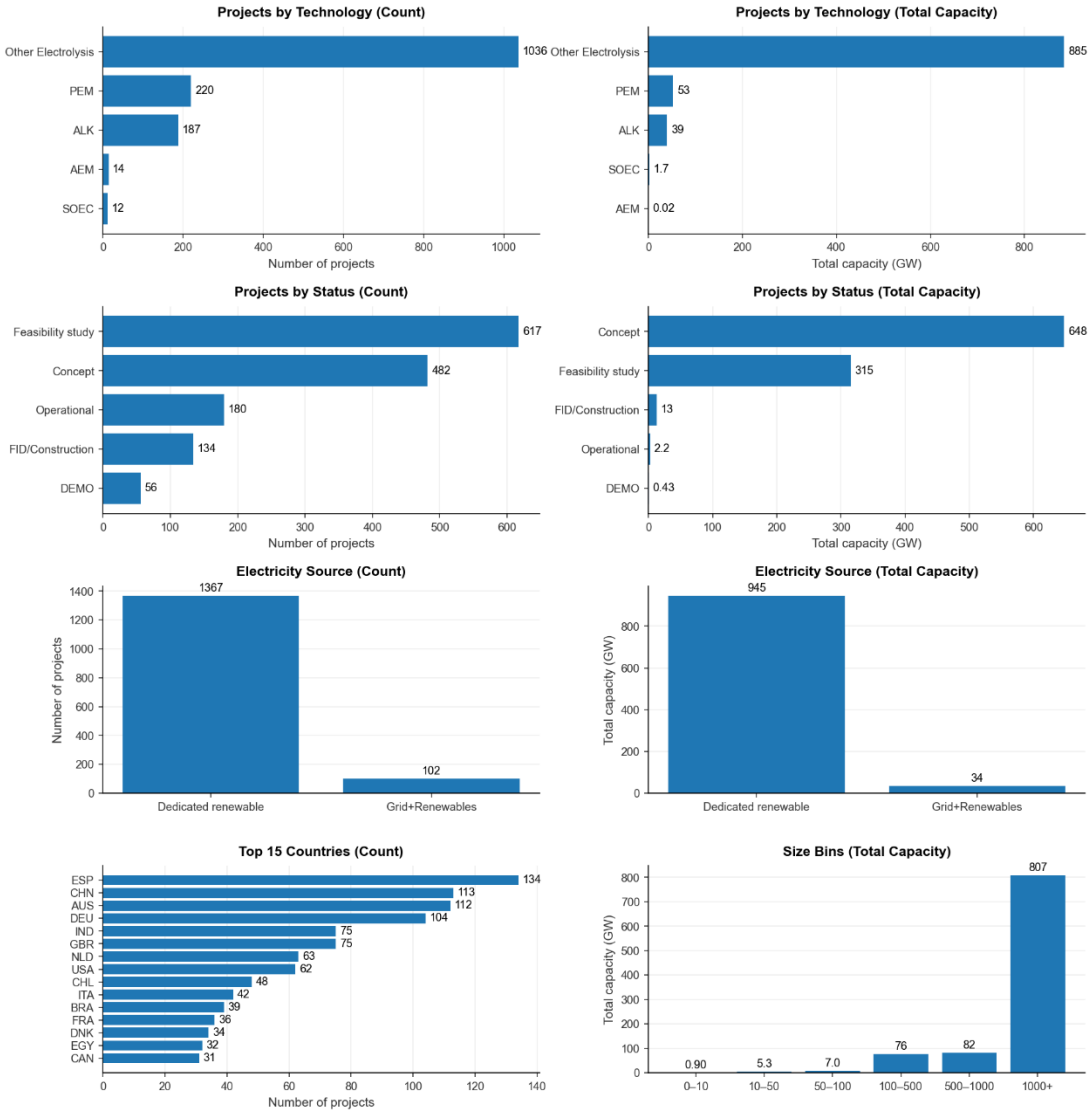
Supplementary Fig. 6 | Evaluation of the attribution model. Density scatter plot comparing the predicted versus actual values. Darker colors indicate a higher concentration of samples. The black line represents the $y = x$ line. Key performance metrics, including R^2 and RMSE, are reported in the top-left inset.



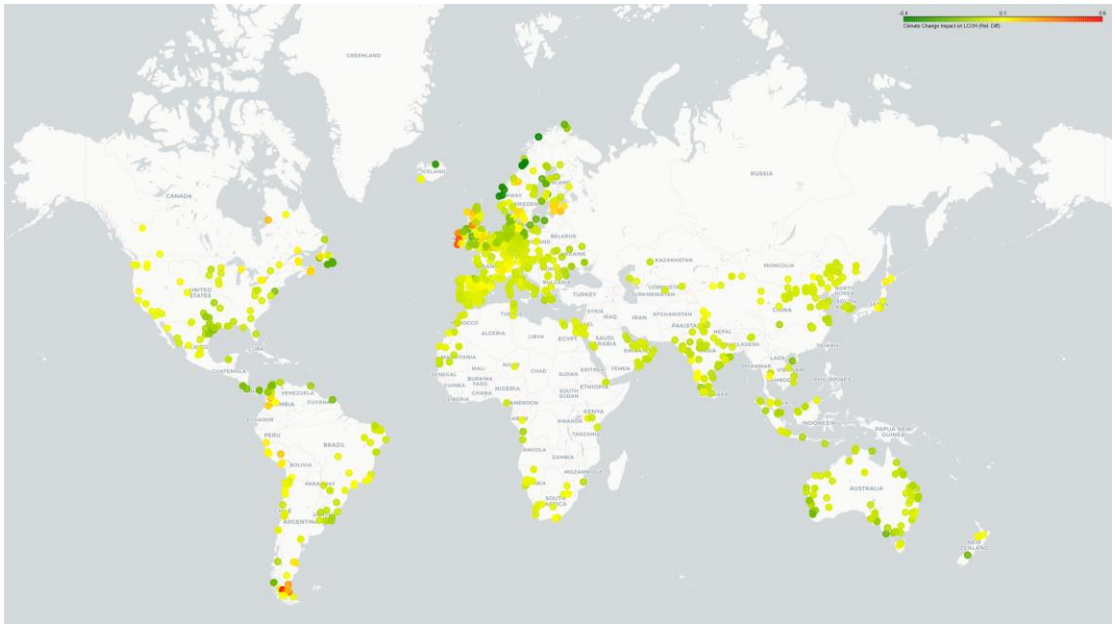
Supplementary Fig. 7 | SHAP analysis of key drivers. The summary plot illustrates the distribution of SHAP values for the top features, representing their respective contributions to the ΔLCOH. The horizontal axis indicates the impact on the model output: positive SHAP values correspond to an increase in ΔLCOH, while negative values indicate a reduction. Features are ranked by their global importance.



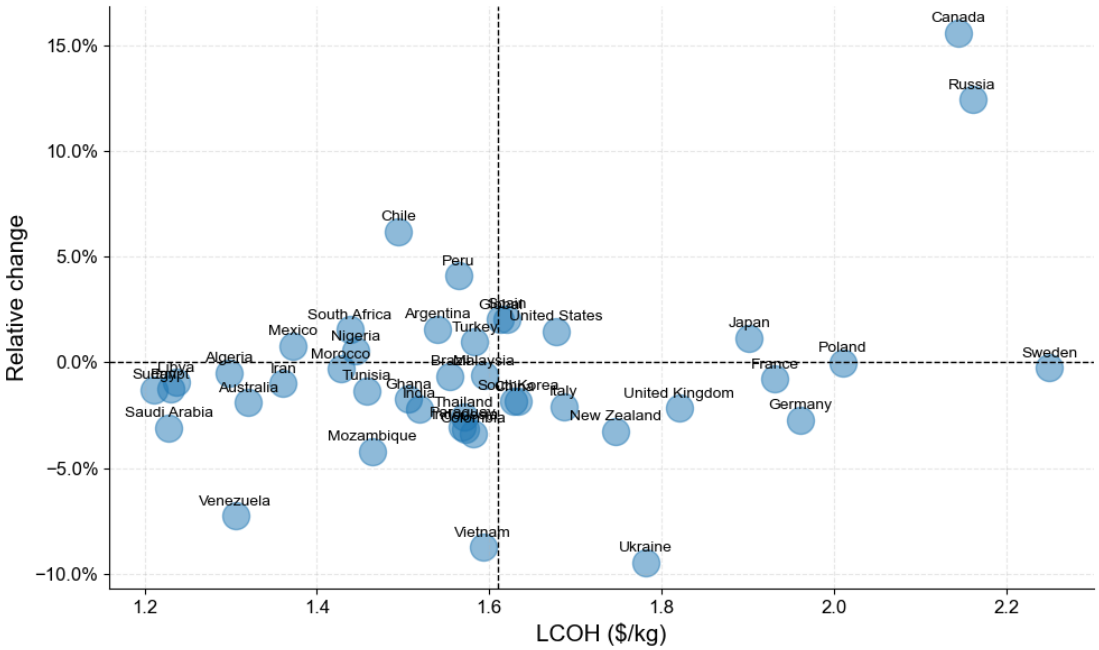
Supplementary Fig. 8 | Global feature importance ranking. The bar chart displays the top input variables sorted by their mean absolute SHAP values. A longer bar indicates a greater average impact of the feature on the predicted ΔLCOH magnitude.



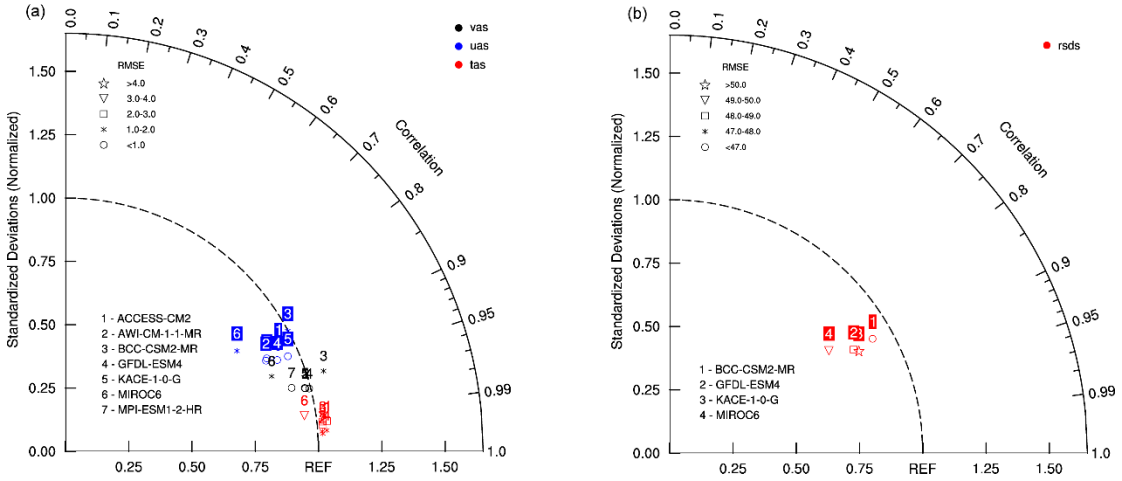
Supplementary Fig. 10 | Overview of selected projects (MWel) from the IEA Hydrogen Projects Database.



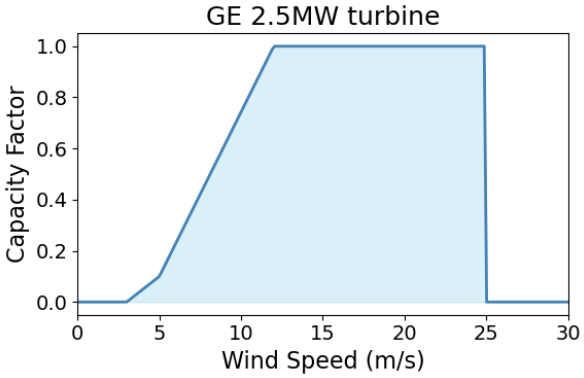
Supplementary Fig. 10 | Global project-level climate impacts on renewable-powered green hydrogen production, presented as an interactive web application (https://github.com/onismyh/climate-impacts-hydrogen).



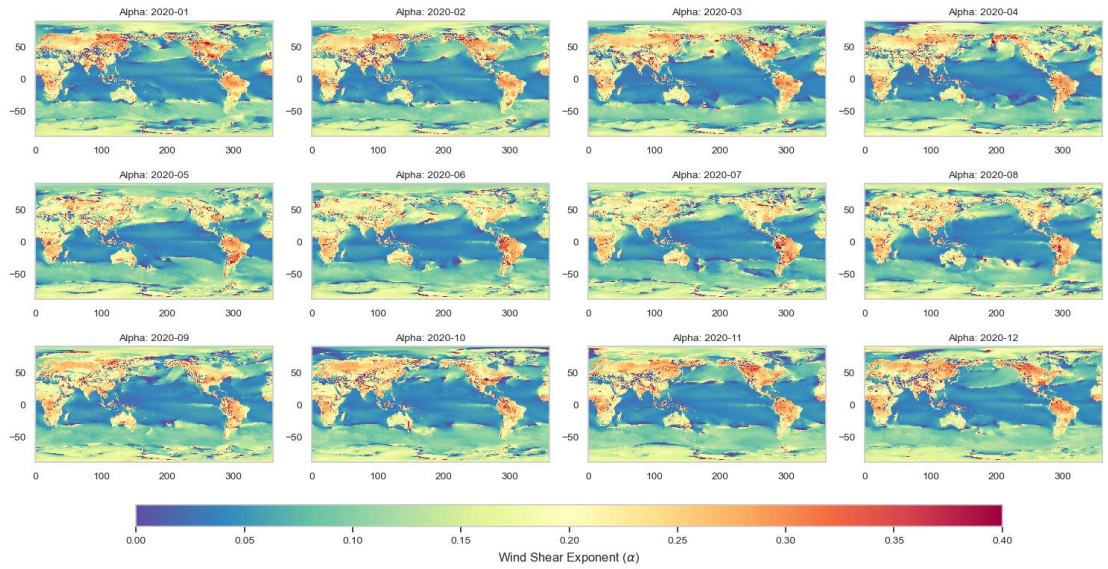
Supplementary Fig. 11 | Distribution of future hydrogen costs and climate change impacts across 42 major countries. The figure illustrates the projected LCOH and the magnitude of climate-induced variations for 42 major economies. Dashed lines indicate reference baselines: the line representing no climate change impact (zero change) and the line corresponding to the global average LCOH.



Supplementary Fig. 12 | Performance evaluation of simulated climate variables. Standardized Taylor diagrams quantifying the ability of seven models to reproduce global climate patterns for the period 2005–2014. **a**, Near-surface air temperature (*tas*) and 10-meter wind speed components (*uas*, *vas*). **b**, Surface downward shortwave radiation (*ssrd*).



Supplementary Fig. 13 | Typical wind power capacity factor curve.



Supplementary Fig. 14 | Global grid-level monthly wind speed scaling factors.

Supplementary Table 1 | Summary of Global Climate Models (GCMs) used in this study.

Model	Institution, Country	Resolutions (Lat*Lon)	Member
ACCESS-CM2	CSIRO, Australia	1.25°×1.875°	r1i1p1f1, r4i1p1f1, r5i1p1f1
AWI-CM-1-1-MR	AWI, Germany	1.067°×1.067°	r1i1p1f1
BCC-CSM2-MR	BCC, China	1.125°×1.125°	r1i1p1f1
GFDL-ESM4	NOAA-GFDL, USA	1°×1°	r1i1p1f1
KACE-1-0-G	NIMS-KMA, Korea	1.25°×1.875°	r1i1p1f1, r2i1p1f1, r3i1p1f1
MIROC6	JAMSTEC, Japan	1.4°×1.4°	r1i1p1f1, r2i1p1f1, r3i1p1f1
MPI-ESM1-2-HR	MPI-M, Germany	1.067°×1.067°	r1i1p1f1, r2i1p1f1

Supplementary Table 2 | Cost distribution ($/kW) in 2050 and key Monte Carlo parameters.

Technology	Distribution	Mean	SD	P5	P10	Mid	P90	P95
PV	Log-normal	283	52.33	205.82	220.0	278.28	352.0	376.25
Wind	Log-normal	904.55	156.70	671.70	715.0	891.27	1111.0	1182.62
Li-ion battery	Log-normal	309.99	131.55	146.12	169.4	285.36	480.7	557.29
Electrolyzer	Log-normal	353.75	65.42	257.28	275.0	347.85	440.0	470.31

Supplementary Table 3 | The selected 42 major countries across the world.

Continents	Involved countries
Oceania	New Zealand, Australia
Asia	China, India, Indonesia, Iran, Japan, Malaysia, Saudi Arabia, Thailand, South Korea, Vietnam
Europe	Turkey, Russia, France, Germany, Italy, Sweden, Ukraine, United Kingdom, Spain, Poland
Americas	United States, Canada, Brazil, Mexico, Venezuela, Argentina, Chile, Colombia, Paraguay, Peru
Africa	Algeria, Egypt, Ghana, Libya, Morocco, Mozambique, Nigeria, South Africa, Sudan, Tunisia