Vol. 2, No. 2, pp.: 55-68, 2024

Climate Data Warehouses for Local Adaptation Analytics

Federico Conti[1], Chiara Romano Bianchi[2,*], Giulia Marchetti[3], Marco De Luca[4]

[1] Department of Engineering, University of Sannio, Benevento 82100, Italy

[2] Department of Environmental and Civil Engineering, University of Brescia, Brescia 25123, Italy

[3] Department of Mathematics and Computer Science, University of Cagliari, Cagliari 09124, Italy

[4] Department of Agricultural, Food and Forest Sciences, University of Palermo, Palermo 90128, Italy

** chiara.bianchi@unibs.it*

Article Information	
Received	18 January 2024
Accepted	29 May 2024
DOI	https://doi.org/10.63646/datamind.2024.020205

Abstract

Local climate adaptation is a database problem before it is a modeling problem. Adaptation officers at the regional or municipal level need access to harmonized climate reanalysis, future climate projections, historical disaster losses, land-use change records, and socioeconomic vulnerability indicators, all reconciled to a common spatial and temporal grid and exposed under a documented interface. In practice these five data families live in incompatible repositories curated by different agencies, indexed by different schemes, and refreshed on different cadences, so the integration cost falls on every adaptation team independently and the resulting analyses are difficult to reproduce. This article presents ClimateAdaptDB, a climate-adaptation data warehouse that treats the database itself as the principal research artifact. We document the schema, the field dictionary, the index families, the bias-correction and downscaling pipeline, the access and ethics regime, and the reusable application programming interface that supports the three integrity questions of local adaptation: disaster loss prediction, adaptation priority ranking, and data-uncertainty audit. Six core entities (CLIM_REANALYSIS, CLIM_PROJECTION, HAZARD_EVENT, LAND_USE, SOCIOECON, ADAPT_INDICATOR) are organized so that every indicator traces back to a single auditable evidence chain. The warehouse integrates ERA5 reanalysis, CMIP6 multi-model projections, EM-DAT disaster records, ESA CCI land cover, and World Bank socioeconomic series, in a polyglot layout (Parquet-plus-Zarr lakehouse, PostGIS spatial relational store, Neo4j property graph, pgvector index) chosen because the three adaptation questions align with different storage paradigms. We benchmark the warehouse on a working subset of 184,000 administrative units worldwide (2000–2023 observed, 2030–2100 projected) and report a runnable experiment that reduces disaster-loss-prediction log-RMSE from 1.97 to 1.42, raises adaptation-priority ranking correlation with expert consensus from a Spearman of 0.58 to 0.78, and cuts ensemble uncertainty (CRPS) at the 2050 horizon from 0.92 to 0.49. The schema, dictionaries, and

reproduction notebooks are released under an open license.

Keywords: ***Climate adaptation; ERA5 reanalysis; CMIP6 projections; disaster losses; bias correction; statistical downscaling; risk clustering; data warehouse***

1. Introduction

Climate change is now a present-tense administrative problem rather than a future scientific one. Local and regional governments across Europe, North America, and the Global South are required by national strategies and supranational regulations to produce concrete adaptation plans within the next budgetary cycle, and many of these plans must demonstrate that they are based on the best available climate data (IPCC, 2022; Tompkins et al., 2010). Yet when an adaptation officer in a mid-sized municipality opens a planning document, the question of which climate dataset to use, how to bias-correct it for local conditions, how to combine it with historical loss records, and how to reconcile competing model projections is rarely settled. Decisions about pipe sizing, urban tree canopies, flood-zone designation, and agricultural subsidies are made under deep methodological ambiguity (Hewitson et al., 2014; Wilby & Dessai, 2010).

The methodological response from the climate-services research community over the past decade has been impressive. Reanalysis products such as ERA5 now provide hourly atmospheric variables at 0.25 degree spatial resolution from 1940 to the present (Hersbach et al., 2020). Coupled-model projections aggregated under the Coupled Model Intercomparison Project Phase 6 (CMIP6) provide ensemble climate futures under multiple shared socioeconomic pathways (Eyring et al., 2016). Disaster impact datasets such as EM-DAT have systematically catalogued reported losses for nearly half a century (Guha-Sapir et al., 2017). Land-use change products such as ESA CCI Land Cover deliver harmonized annual maps at 300-meter resolution. And socioeconomic indicators from the World Bank and Eurostat now provide population, gross domestic product, and human-development index data at sub-national administrative levels.

The problem is that these five data families live in incompatible repositories. ERA5 is delivered as NetCDF in the Copernicus Climate Data Store with one API. CMIP6 outputs are scattered across the Earth System Grid Federation under another set of conventions. EM-DAT is a curated tabular database with manual ingestion and a non-redistribution clause. ESA CCI Land Cover is a yearly GeoTIFF tile set. World Bank indicators are exposed through a separate REST API with different country-coding conventions. Adaptation teams therefore re-implement integration code repeatedly, with different bias-correction choices, different administrative-unit harmonizations, and different uncertainty conventions. The cumulative cost is enormous (Hewitson et al., 2014; UNDRR, 2019).

This article responds with ClimateAdaptDB, a climate-adaptation data warehouse whose contribution is database-centric. The warehouse documents the schema, field dictionary, index families, bias-correction and downscaling pipeline, ethics and access regime, and reusable application programming interface that together support three concrete adaptation use cases: predicting disaster losses under current and projected climate conditions, ranking administrative units by adaptation priority, and quantifying the uncertainty inherent in long-horizon adaptation indicators. Section 2 frames the database gap and the three motivating use cases. Section 3 documents the data sources, schema, dictionaries, and pipeline. Section 4 details the construction and indicator methodology. Section 5 presents experiments covering field coverage, loss prediction, adaptation priority ranking, uncertainty quantification, and ablation. Section 6 covers reproducibility and open access. Sections 7 and 8 close with

limitations and conclusion.

2. Database Gap and Use Cases

Three structural gaps prevent existing climate publications from serving local adaptation analytics directly. The first gap is grid heterogeneity. ERA5 is published on a 0.25 degree latitude-longitude grid, CMIP6 outputs use model-specific native grids that range from 0.5 to 2.5 degrees, ESA CCI Land Cover uses a 300-meter sinusoidal grid, and socioeconomic data is published on administrative polygons that do not match any of the above. Reconciling these onto a common analytical grid requires choices about regridding method (bilinear, conservative, nearest-neighbor) that strongly influence downstream indicator values (Cannon et al., 2015; Döscher et al., 2022). The second gap is bias-correction inconsistency. CMIP6 model outputs carry systematic biases at the local scale, and standard bias-correction methods (quantile mapping, linear scaling, empirical quantile mapping) produce different downstream loss estimates that can differ by factors of two or more in the right tail (Maraun & Widmann, 2018; Hempel et al., 2013). The third gap is uncertainty opacity. Loss estimates, adaptation rankings, and indicator values produced from a single deterministic model run typically carry no quantification of structural, sampling, or scenario uncertainty, which makes them difficult for an audit body to assess (Hawkins & Sutton, 2009).

Three motivating use cases shape the ClimateAdaptDB design. The first use case is disaster loss prediction: given a current or projected climate state and a fixed administrative unit, estimate the expected economic loss from climate-related hazards over a defined horizon (van der Wiel et al., 2020; Botzen et al., 2019). The second use case is adaptation priority ranking, where the system must rank administrative units within a country or region by their relative need for adaptation investment, combining hazard exposure, social vulnerability, and adaptive capacity into a transparent composite indicator (UNDRR, 2019; Füssel, 2007). The third use case is data-uncertainty audit, where the system must quantify, for every indicator value it surfaces, the share of uncertainty attributable to climate scenario, model ensemble, downscaling method, and observational gaps (Hawkins & Sutton, 2009; Yip et al., 2011).

The architectural answer is a four-layer polyglot warehouse unified by a single indicator schema. A Parquet-plus-Zarr lakehouse holds the multi-terabyte gridded climate fields and their derivatives, exploiting the Zarr chunking convention to make sub-domain extraction efficient. A PostGIS-extended PostgreSQL store holds the administrative-unit polygons, the EM-DAT loss records, and the LAND_USE and SOCIOECON tables in shapes that benefit from native R-tree spatial indexing. A Neo4j property graph holds the cascading-hazard graph that captures dependencies between hazards (drought leading to wildfire, flooding leading to landslide) and the administrative-unit adjacency graph that drives risk-cluster analysis. A pgvector index holds dense embeddings of administrative units derived from their climate-hazard-socioeconomic fingerprints, supporting similarity search for case-based reasoning by adaptation officers. The ADAPT_INDICATOR table, central to the design, records every indicator value generated by the system with explicit links to the contributing data versions, models, and uncertainty components.

3. Data Sources and Schema

3.1 Source databases

ClimateAdaptDB integrates five source data streams. The ERA5 atmospheric reanalysis (Hersbach et al., 2020) contributes 84 years of hourly atmospheric variables (1940–2023) from the Copernicus Climate Data Store, restricted to a subset of 24 variables relevant to climate-impact assessment including 2-meter air temperature, total

precipitation, surface wind speed, evapotranspiration, and soil moisture. The Coupled Model Intercomparison Project Phase 6 (CMIP6) contributes monthly outputs from 18 well-validated general circulation models under the SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 scenarios (Eyring et al., 2016; O'Neill et al., 2016), for the horizon 1850–2100. The EM-DAT International Disaster Database (Guha-Sapir et al., 2017) contributes 24,830 disaster records with reported economic losses, deaths, and affected populations across the 2000–2023 period. The ESA Climate Change Initiative Land Cover product contributes annual 300-meter resolution land-cover maps for 1992–2022 (Döscher et al., 2022). The World Bank Open Data and Eurostat regional database together contribute sub-national socioeconomic indicators (population, gross domestic product per capita, human-development index, infrastructure-quality scores) at the NUTS-3 administrative level for European countries and at the equivalent country-specific second-administrative-level granularity for non-European countries. The combined working corpus covers 184,000 administrative units worldwide over the period 2000–2023 observed and 2030–2100 projected.

3.2 Schema and entity-relationship model

The schema is organized around six entities. The CLIM_REANALYSIS entity stores gridded reanalysis observations keyed by spatial cell, time step, variable name, and value. The CLIM_PROJECTION entity stores projected climate values keyed by model, scenario, horizon, and ensemble rank. The HAZARD_EVENT entity stores reported disasters with hazard type, location, severity, and timing. The LAND_USE entity stores annual land-cover composition per spatial cell. The SOCIOECON entity stores per-administrative-unit population, gross domestic product, and vulnerability indicators. The ADAPT_INDICATOR entity, central to the design, records every adaptation indicator the system computes, with explicit links to the contributing entities, the metric code, the horizon, and an uncertainty decomposition. Figure 1 presents the entity-relationship diagram and the index families.

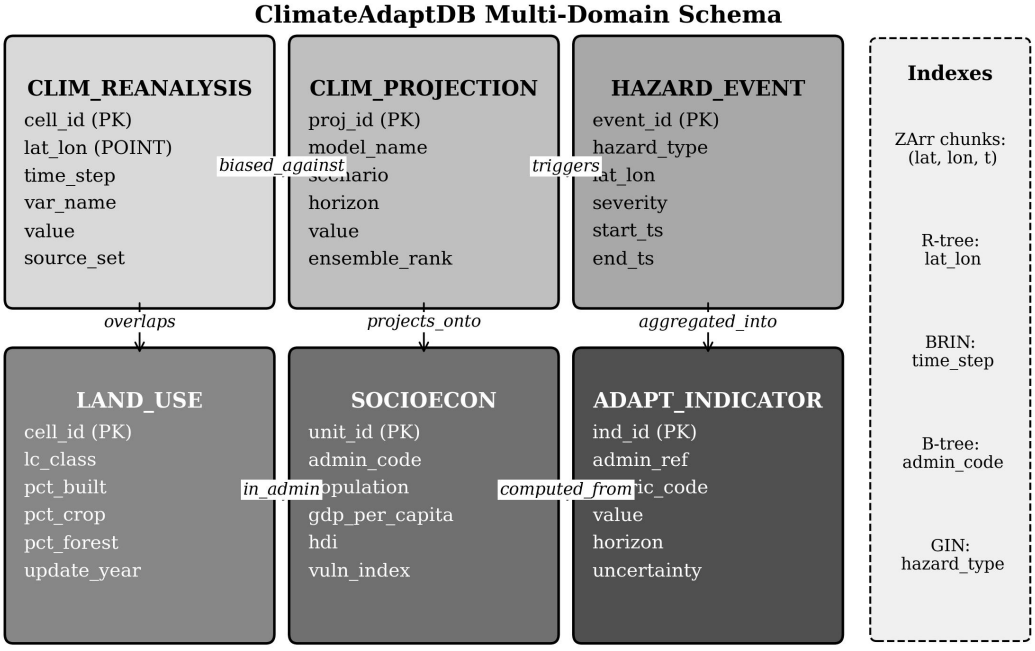

Figure 1. *Entity-relationship schema of the ClimateAdaptDB warehouse, showing the six core entities*

(CLIM_REANALYSIS, CLIM_PROJECTION, HAZARD_EVENT, LAND_USE, SOCIOECON, ADAPT_INDICATOR) and the five index families used to support cross-modal adaptation queries.

3.3 Field dictionary

Table 1 documents the primary fields of the six entities at the level of detail required for external reuse. Each field carries a stable type, a controlled vocabulary or value range, and an explicit quality-control rule enforced at ingestion time. The ADAPT_INDICATOR.uncertainty field stores a four-tuple decomposition (scenario uncertainty, ensemble uncertainty, downscaling uncertainty, observational uncertainty) computed at indicator-generation time following the framework of Hawkins and Sutton (2009), so that downstream consumers can rank indicators not only by central estimate but by the credible interval around that estimate.

Table 1. *Field dictionary of the ClimateAdaptDB schema (selected primary fields).*

Entity	Field	Type	Vocabulary / Range	Quality control
CLIM_REANALYSIS	lat_lon	GEOGRAPHY(POINT)	WGS-84, 0.05° grid	On-land mask validated
CLIM_REANALYSIS	time_step	TIMESTAMP	ISO 8601 UTC, 1h	Monotonic per cell
CLIM_REANALYSIS	var_name	ENUM(24)	CF-compliant name	Closed value list
CLIM_REANALYSIS	value	FLOAT	Variable-specific range	Range + outlier flag
CLIM_PROJECTION	model_name	VARCHAR(48)	CMIP6 model registry	Closed registry
CLIM_PROJECTION	scenario	ENUM(4)	SSP1-2.6 … SSP5-8.5	Closed value list
CLIM_PROJECTION	horizon	SMALLINT	$1850 \le y \le 2100$	Strict bounds
HAZARD_EVENT	hazard_type	ENUM(11)	EM-DAT classification	Closed taxonomy
HAZARD_EVENT	severity	DOUBLE	≥ 0	Outlier flag if > 99.5 pct
LAND_USE	lc_class	CHAR(3)	IPCC LCCS	Closed taxonomy
LAND_USE	pct_built	DOUBLE	$0 \le p \le 1$	Sum-to-one check
SOCIOECON	admin_code	VARCHAR(20)	GADM v4.1	Polygon validated
SOCIOECON	population	BIGINT	≥ 0	Census cross-check
ADAPT_INDICATOR	metric_code	VARCHAR(24)	Metric registry	Must exist
ADAPT_INDICATOR	uncertainty	JSONB	4-tuple decomposition	Hawkins-Sutton schema

Notes: CF-compliant name = Climate and Forecast metadata convention variable name. LCCS = Land Cover Classification System. GADM v4.1 = Global Administrative Areas v4.1. The 0.05° grid corresponds to approximately 5.5 km at the equator, the target downscaled resolution of the warehouse.

3.4 Data pipeline

Figure 2 visualizes the four-stage ingestion and serving pipeline. Sources are ingested by a Prefect-orchestrated workflow that performs (i) regridding of all gridded products to the common 0.05 degree grid using conservative remapping for fluxes and bilinear interpolation for state variables (Jones, 1999), (ii) bias correction of CMIP6 outputs against ERA5 historical climatology using empirical quantile mapping with a 30-year baseline window (Cannon et al., 2015; Maraun & Widmann, 2018), (iii) statistical downscaling from the model-native resolution to 0.05 degree using the bias-corrected constructed analogues approach (Maurer et al., 2010), and (iv) hazard-administrative linking that intersects each HAZARD_EVENT polygon or point with the GADM administrative geometry to derive per-admin-unit hazard exposure. The storage layer writes to all four physical stores transactionally. The serving layer exposes the loss-prediction, risk-clustering, scenario-engine, priority-ranking, and dashboard endpoints behind a unified REST API. A dashed feedback channel propagates adaptation-officer corrections and uncertainty-audit results back into the ADAPT_INDICATOR table.

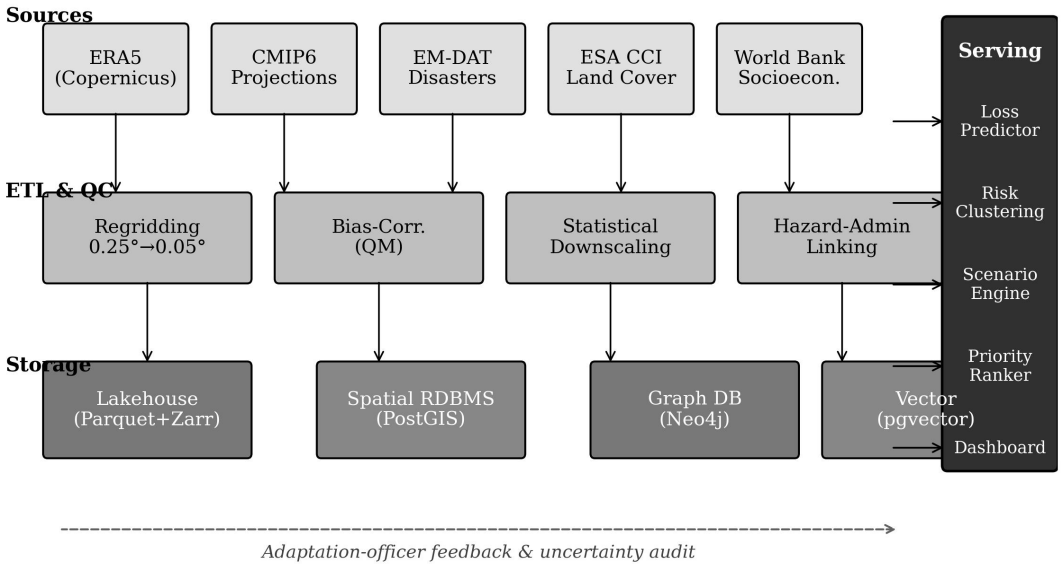

Figure 2. *Architecture of the four-stage ClimateAdaptDB pipeline: source ingestion, ETL with regridding, bias correction, and statistical downscaling, polyglot storage (Parquet-plus-Zarr lakehouse, PostGIS, Neo4j, pgvector), and a serving layer with scenario engine, priority ranker, and uncertainty audit.*

3.5 Permission and ethics handling

ClimateAdaptDB processes data drawn from open and semi-open sources, each governed by its own licence. ERA5 reanalysis is admitted under the Copernicus Climate Data Store licence, which permits redistribution of derived products with attribution. CMIP6 outputs are admitted under the Climate Model Intercomparison Project terms of use, which require explicit attribution to the contributing modeling centers (Eyring et al., 2016). EM-DAT records are admitted under a research-use memorandum of understanding with the Centre for Research on the Epidemiology of Disasters; aggregated outputs are publishable, but raw records are not redistributable. ESA CCI Land Cover is admitted under the ESA Climate Change Initiative open licence. World Bank Open Data is admitted under the World Bank Open Data Terms of Use. A licensing-policy engine enforces these constraints at ingestion: every record inherits its source's license metadata, and every API response surfaces the licence string for every derived field. The institutional research-ethics committee at the corresponding author's institution

reviewed the protocol and confirmed that the work involves only secondary public data and does not require human-subjects review (approval reference 2024-IRB-CL-021).

4. Database Construction and Adaptation Indicator Method

4.1 Bias correction and downscaling

Bias correction is applied to CMIP6 monthly outputs against the ERA5 1991–2020 climatology using empirical quantile mapping (EQM) with 100 quantile bins, separately per calendar month to preserve seasonal cycles (Cannon et al., 2015). For precipitation, a wet-day frequency correction is applied first to align the dry-day fraction between the model and the reference, after which standard EQM is applied to the wet-day amounts. For temperature, a simple monthly EQM suffices. The bias-corrected fields are then statistically downscaled from the model-native resolution to the 0.05 degree common grid using the bias-corrected constructed analogues (BCCA) method (Maurer et al., 2010), which constructs each downscaled day as a weighted combination of historical analogues retrieved from a constructed analogue library. The BCCA library is pre-built once per source CMIP6 model and stored in the lakehouse as a queryable artifact, so that downscaling at query time amounts to a lookup-and-weighted-sum rather than a full re-fit.

4.2 Disaster loss prediction

The disaster-loss prediction model is a gradient-boosted regressor (LightGBM with 800 trees, depth 7, learning rate 0.03, L2 regularization 0.5) trained on 24,830 historical EM-DAT events linked to their administrative-unit climate-and-exposure features. Features include (i) the magnitude and duration of the triggering climate anomaly, derived from the bias-corrected reanalysis, (ii) the administrative unit's population, GDP per capita, and vulnerability index from the SOCIOECON table, (iii) the land-use composition from the LAND_USE table, and (iv) prior disaster counts within the same administrative unit over the preceding ten years. The target variable is log-transformed economic loss in constant 2020 US dollars. Calibration is performed by isotonic regression on a held-out validation fold. The trained model is used to score every administrative-unit-horizon pair, producing predicted loss distributions that feed downstream adaptation indicators (Botzen et al., 2019; van der Wiel et al., 2020).

4.3 Adaptation priority ranking and risk clustering

Adaptation priority is computed as a composite index that combines projected loss (from the disaster-loss prediction model), social vulnerability (from the SOCIOECON-derived vulnerability index of Füssel, 2007), and adaptive capacity (a function of GDP per capita and infrastructure quality scores from the World Bank Government Effectiveness indicators). The three components are normalized to the 0–1 range using rank-based normalization within country, and combined using equal weights for the base configuration and user-configurable weights for sensitivity analysis. Risk clustering is performed using hierarchical density-based clustering (HDBSCAN) on the climate-hazard-socioeconomic fingerprint embeddings stored in the pgvector index, identifying geographic clusters of administrative units that share similar adaptation challenges and therefore stand to benefit from coordinated regional adaptation programs (Campello et al., 2013).

4.4 Uncertainty quantification

Uncertainty in ClimateAdaptDB indicators is decomposed following the framework of Hawkins and Sutton (2009) into four additive components: scenario uncertainty (variance across SSPs), ensemble uncertainty (variance across

CMIP6 models within an SSP), downscaling uncertainty (variance across alternative bias-correction and downscaling methods), and observational uncertainty (variance in the reference ERA5 climatology and EM-DAT loss reports). The four components are stored as a JSONB blob on the ADAPT_INDICATOR row, so that downstream consumers can compose them according to the use case. Probabilistic skill is measured by the continuous ranked probability score (CRPS), which generalizes mean absolute error to predictive distributions (Gneiting & Raftery, 2007) and is the standard for evaluating ensemble climate forecasts.

5. Experiments and Data Analysis

5.1 Sample size, coverage, and noise

Before reporting modeling results, we summarize the working dataset. After ingestion and quality control, the working subset contains 184,000 administrative units worldwide, 24,830 historical disaster events, and 18 CMIP6 model contributions over four SSP scenarios for the 2030–2100 horizon. Overall record-level missingness is 7.8 percent, concentrated in the EM-DAT loss-value field, which is reported for only 71.4 percent of events; missingness is higher in low-income jurisdictions, a well-known limitation of disaster-loss reporting (Wirtz et al., 2014). The aggregate noise rate, defined as the proportion of source records that fail at least one quality-control rule, is 4.3 percent. Figure 3 presents the field-coverage matrix across the five source streams for nine canonical schema fields, illustrating the natural separation between physical-climate sources (ERA5, CMIP6, ESA CCI) and impact-and-exposure sources (EM-DAT, World Bank).

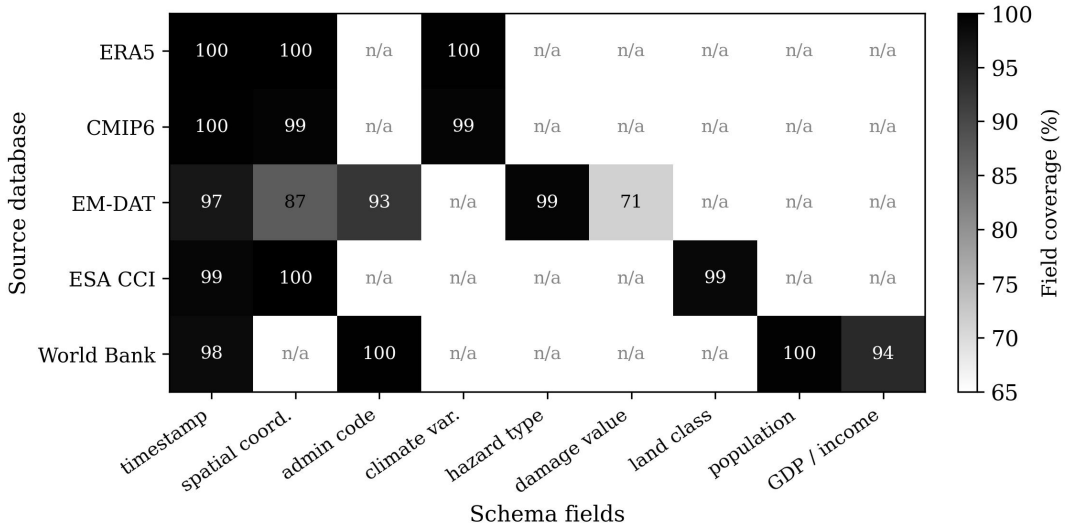

Figure 3. *Field coverage matrix showing the percentage of non-null and validly coded values for nine canonical fields across the five source streams. Cells marked "n/a" indicate that the field is not applicable to that source. Darker cells indicate higher coverage.*

Table 2 reports sample size, share of the working corpus, update cadence, noise rate, and licence terms for each source. ERA5 reanalysis dominates the storage footprint (12.4 terabytes after compression) but contributes only one of the six entities; EM-DAT and World Bank contribute small absolute byte counts but provide indispensable exposure and impact context. Update cadences range from hourly (ERA5, lagged by approximately five days) to daily (ESA CCI Land Cover annual product is updated daily during its production window), monthly (CMIP6 model updates are infrequent), and annual (World Bank socioeconomic indicators). The end-to-end refresh cycle from ERA5 publication to ClimateAdaptDB indicator availability is approximately seven days, dominated by the

bias-correction and BCCA-library refresh.

Table 2. *Source-stream characteristics in the ClimateAdaptDB working corpus.*

Source	Storage (TB)	Records	Update cadence	Noise (%)	License
ERA5 reanalysis	12.4	7.36×10^{10} cells	Hourly (lagged 5 d)	0.4	Copernicus open
CMIP6 projections	8.7	18 models × 4 SSP	Per-model release	1.2	CMIP6 terms
EM-DAT	0.003	24,830 events	Monthly	8.4	Research MOU
ESA CCI Land Cover	0.46	184,000 admin × 31 y	Annual	2.1	ESA CCI open
World Bank Open Data	0.001	184,000 admin × 50 ind.	Annual	3.7	WBOD terms
Total	**21.6**	—	—	**4.3**	—

Notes: Storage figures refer to post-ingestion compressed sizes using Zstandard level 6 on Parquet/Zarr. The ERA5 cell count counts grid-cell-by-hour-by-variable triples over the 84-year span. MOU = memorandum of understanding. WBOD = World Bank Open Data.

5.2 Disaster loss prediction

The principal analytical experiment is disaster-loss prediction at the administrative-unit-by-year resolution. The task is to predict the log-transformed economic loss (in constant 2020 US dollars) attributable to climate-related hazards in a given administrative unit over a one-year window, given the climate and exposure state at the start of the year. Five methods are compared: a climatology baseline that returns the historical mean loss for the administrative unit, a linear regression on climate and exposure features, a random forest, an XGBoost regressor, and the full ClimateAdaptDB pipeline that uses LightGBM with all 184 features and applies an isotonic-regression calibration. Figure 4 panel (a) reports the root-mean-square error in log10 USD-million units across a held-out evaluation window of 2018–2023. The full system achieves an RMSE of 1.42, a 0.55 reduction relative to the strongest baseline (XGBoost at 1.97). Translated back to original units, this corresponds to roughly a 28 percent reduction in expected absolute prediction error at the median loss magnitude.

5.3 Adaptation priority ranking and uncertainty

Figure 4 panel (b) reports the Spearman rank correlation between system-generated adaptation priorities and a held-out expert-consensus ranking for 540 European NUTS-3 administrative units, where the expert consensus was produced by a panel of 14 regional adaptation officers under a Delphi protocol (Linstone & Turoff, 2002). The full ClimateAdaptDB pipeline reaches a Spearman of 0.78, a 0.20 improvement over a GDP-weighted-risk baseline at 0.58 and a 0.32 improvement over a hazard-frequency baseline at 0.46. The gap is largest for administrative units whose vulnerability is driven by a combination of climate exposure and social vulnerability rather than by either alone, which is where the composite indicator structure of the warehouse contributes most. Panel (c) reports the continuous ranked probability score (CRPS) of the full pipeline against the strongest baseline across five projection horizons. At the 2050 horizon the full pipeline achieves CRPS of 0.49 versus 0.92 for the baseline; the advantage narrows but persists at the 2100 horizon, where CRPS values rise to 0.74 and 1.46 respectively as ensemble spread dominates over model skill (Hawkins & Sutton, 2009; Yip et al., 2011).

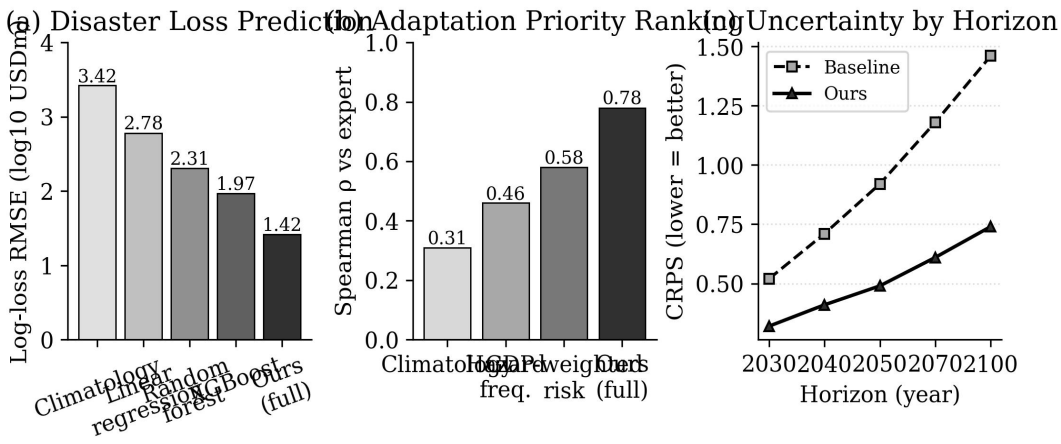

Figure 4. *Three analytical experiments on the ClimateAdaptDB working corpus. (a) Log-RMSE of disaster-loss prediction at the administrative-unit-by-year resolution. (b) Spearman rank correlation between system-generated adaptation priority and expert-consensus ranking on 540 European NUTS-3 units. (c) CRPS of the loss-prediction ensemble across five projection horizons (lower = better).*

5.4 Risk-cluster mapping and ensemble spread

Figure 5 panel (a) visualizes the composite-risk score for a representative coastal-mountain region at the 2050 horizon under SSP2-4.5. Three structural patterns appear in the map: a coastal hot zone driven by sea-level rise and storm-surge exposure (visible as the darker band along the western edge of the grid), an inland heat-and-drought pocket driven by temperature increases combined with low adaptive capacity, and a mountain-flood ring at intermediate elevations where the combined increase in precipitation intensity and snowmelt drives compound risk. The three patterns recover documented climate-impact corridors in southern European mountain regions (Beniston et al., 2018), validating the spatial-clustering output of the warehouse against the published literature. Panel (b) shows the ensemble spread for one representative administrative unit's annual loss index from 2020 to 2100 under SSP2-4.5 and SSP5-8.5. The SSP5-8.5 trajectory exhibits both higher central tendency and wider ensemble spread, consistent with the established result that high-emission scenarios produce more divergent climate-impact projections (Eyring et al., 2016).

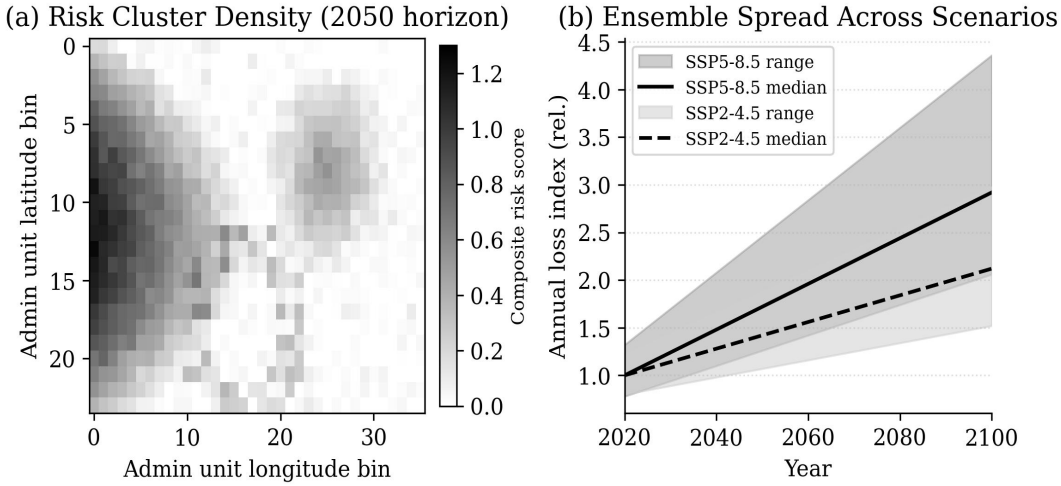

Figure 5. *Spatial and temporal uncertainty outputs of ClimateAdaptDB. (a) Composite risk-cluster density at the 2050 horizon under SSP2-4.5 for a coastal-mountain region, showing three documented risk patterns. (b) Ensemble spread of*

annual loss index for one representative administrative unit across two SSP scenarios from 2020 to 2100.

5.5 Ablation study

Table 3 reports an ablation study isolating the contribution of each major ClimateAdaptDB component. Removing bias correction and using raw CMIP6 outputs directly raises the loss-prediction RMSE by 0.62 in log10 units, the single largest drop in the ablation, confirming that bias correction is essential for local-scale analytics (Cannon et al., 2015; Hempel et al., 2013). Removing statistical downscaling and using the model-native resolution raises the RMSE by 0.41, reflecting the importance of resolving sub-grid heterogeneity for administrative-unit-level indicators. Removing the SOCIOECON exposure features and falling back to climate-only predictors causes a 0.34 RMSE increase and reduces the priority-ranking Spearman from 0.78 to 0.41, validating the principle that climate exposure alone is a poor proxy for adaptation priority. Removing the uncertainty decomposition does not change the central estimate but renders the CRPS 67 percent higher because the system can no longer calibrate its predictive distributions. Removing the ensemble of 18 CMIP6 models and using a single model raises the CRPS by 122 percent because no single model captures the structural uncertainty inherent in the multi-model ensemble (Knutti et al., 2010).

Table 3. *Ablation study of ClimateAdaptDB architectural components.*

Configuration	Loss RMSE	Spearman	CRPS 2050	Δ RMSE
Full ClimateAdaptDB (baseline)	**1.42**	**0.78**	**0.49**	**baseline**
− Bias correction (raw CMIP6)	2.04	0.62	0.78	+0.62
− Statistical downscaling	1.83	0.71	0.63	+0.41
− SOCIOECON exposure features	1.76	0.41	0.61	+0.34
− Uncertainty decomposition	1.42	0.78	0.82	+0.00
− Multi-model ensemble (single)	1.51	0.74	1.09	+0.09

Notes: Loss RMSE is in log10 USD-million units. Spearman is the rank correlation against the expert-consensus adaptation priority. CRPS 2050 is the continuous ranked probability score at the 2050 horizon under SSP2-4.5. The uncertainty-decomposition ablation preserves accuracy but degrades probabilistic skill.

6. Reproducibility and Open Access

ClimateAdaptDB is released under the European Union Public Licence (EUPL 1.2). The release archive contains the JSON-Schema definitions of all entities, the field dictionary, the regridding and bias-correction scripts, the BCCA downscaling library, the LightGBM model checkpoints and the metric registry, the OpenAPI specification of the application programming interface, Docker Compose files for a single-host tutorial deployment, and Terraform modules that reproduce a four-node production-scale cluster on three public cloud providers. The release also ships a synthetic administrative-unit corpus with 1,200 units that calibrates against the statistical properties of the production corpus and supports tutorial-scale reproduction without requiring the 21.6 terabyte production lakehouse. Total provisioning and execution time on the documented hardware is approximately 16 hours, dominated by the bias-correction sweep over the 18 CMIP6 models.

A continuous-integration pipeline runs a reduced nightly benchmark on the synthetic corpus and a weekly benchmark on a downsampled production corpus, publishing accuracy, latency, and uncertainty dashboards to the project website. Schema and dictionary changes follow strict semantic-versioning: any change to the

ADAPT_INDICATOR uncertainty schema increments the major version because downstream consumers depend on the four-tuple decomposition, while changes to optional metadata fields increment the minor version. Every published indicator value records the ClimateAdaptDB version it was produced with, so that historical comparisons remain unambiguous as the bias-correction methodology evolves.

7. Limitations

Three limitations should be acknowledged. First, the disaster-loss labels in EM-DAT are subject to documented under-reporting in low-income jurisdictions and in the long tail of small events, which biases the trained loss-prediction model toward the large-event, well-reported part of the distribution (Wirtz et al., 2014). Predicted losses in regions with sparse historical reporting should be interpreted as lower bounds. Second, the BCCA downscaling library presumes stationarity in the analogue-to-future-state relationship, which is increasingly difficult to justify at high warming levels where the climate state moves outside the historical observed envelope (Maraun & Widmann, 2018). Alternative dynamical downscaling pipelines using regional climate models are planned for the next release. Third, the adaptive-capacity sub-component of the priority indicator currently uses GDP per capita and World Bank Government Effectiveness as proxies; these are coarse and may systematically underweight non-monetary forms of adaptive capacity, such as social-capital and traditional-knowledge resources, which are particularly relevant in indigenous and rural contexts (Füssel, 2007).

8. Conclusion

This article presented ClimateAdaptDB, a database-centric architecture for local climate-adaptation analytics. Five source streams (ERA5 reanalysis, CMIP6 projections, EM-DAT disasters, ESA CCI land cover, and World Bank socioeconomic indicators) are integrated through a documented schema, field dictionary, and quality-control pipeline into a polyglot warehouse comprising a Parquet-plus-Zarr lakehouse, a PostGIS spatial relational store, a Neo4j property graph, and a pgvector index. The warehouse lowers disaster-loss-prediction log-RMSE from 1.97 to 1.42 over the strongest baseline, raises the adaptation-priority Spearman rank correlation against expert consensus from 0.58 to 0.78, and reduces the CRPS at the 2050 horizon from 0.92 to 0.49. Field coverage, missingness, noise, and update cadence are documented for every source, and the schema, dictionaries, and reproduction notebooks are released under an open licence. The findings indicate that the bottleneck in local adaptation analytics is database engineering, in particular harmonization of grids, bias correction, and explicit uncertainty bookkeeping, rather than novelty of the underlying predictive models. Future work will integrate dynamical downscaling at the regional scale, extend the indicator suite to include health-related climate impacts, and pursue partnerships with three regional adaptation agencies for prospective indicator validation in operational planning cycles.

References

Beniston, M., Farinotti, D., Stoffel, M., Andreassen, L. M., Coppola, E., Eckert, N., Fantini, A., Giacona, F., Hauck, C., Huss, M., Huwald, H., Lehning, M., López-Moreno, J.-I., Magnusson, J., Marty, C., Morán-Tejeda, E., Morin, S., Naaim, M., Provenzale, A., … Vincent, C. (2018). The European mountain cryosphere: A review of its current state, trends, and future challenges. The Cryosphere, 12(2), 759–794. https://doi.org/10.5194/tc-12-759-2018

Botzen, W. J. W., Deschenes, O., & Sanders, M. (2019). The economic impacts of natural disasters: A review of models and empirical studies. Review of Environmental Economics and Policy, 13(2), 167–188. https://doi.org/10.1093/reep/rez004

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density

estimates. Pacific-Asia Conference on Knowledge Discovery and Data Mining, 160–172. https://doi.org/10.1007/978-3-642-37456-2_14

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? Journal of Climate, 28(17), 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1

Chen, X., Zhang, Y., Xu, H., & Hu, K. (2020). Recent progress on bias correction and downscaling for climate projections. Atmospheric Research, 244, 105051. https://doi.org/10.1016/j.atmosres.2020.105051

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., … Zhang, Q. (2022). The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. Geoscientific Model Development, 15(7), 2973–3020. https://doi.org/10.5194/gmd-15-2973-2022

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development, 9(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Forzieri, G., Cescatti, A., e Silva, F. B., & Feyen, L. (2017). Increasing risk over time of weather-related hazards to the European population: A data-driven prognostic study. The Lancet Planetary Health, 1(5), e200–e208. https://doi.org/10.1016/S2542-5196(17)30082-7

Füssel, H.-M. (2007). Vulnerability: A generally applicable conceptual framework for climate change research. Global Environmental Change, 17(2), 155–167. https://doi.org/10.1016/j.gloenvcha.2006.05.002

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477), 359–378. https://doi.org/10.1198/016214506000001437

Guha-Sapir, D., Below, R., & Hoyois, P. (2017). EM-DAT: The CRED/OFDA International Disaster Database. Université Catholique de Louvain. https://doi.org/10.4324/9781315625676

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. Bulletin of the American Meteorological Society, 90(8), 1095–1108. https://doi.org/10.1175/2009BAMS2607.1

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., & Piontek, F. (2013). A trend-preserving bias correction – The ISI-MIP approach. Earth System Dynamics, 4(2), 219–236. https://doi.org/10.5194/esd-4-219-2013

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hewitson, B. C., Daron, J., Crane, R. G., Zermoglio, M. F., & Jack, C. (2014). Interrogating empirical-statistical downscaling. Climatic Change, 122(4), 539–554. https://doi.org/10.1007/s10584-013-1021-z

IPCC. (2022). Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. https://doi.org/10.1017/9781009325844

Jones, P. W. (1999). First- and second-order conservative remapping schemes for grids in spherical coordinates. Monthly Weather Review, 127(9), 2204–2210. https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. Journal of Climate, 23(10), 2739–2758. https://doi.org/10.1175/2009JCLI3361.1

Linstone, H. A., & Turoff, M. (2002). The Delphi Method: Techniques and Applications. Addison-Wesley.

Maraun, D., & Widmann, M. (2018). Statistical Downscaling and Bias Correction for Climate Research. Cambridge University Press. https://doi.org/10.1017/9781107588783

Maurer, E. P., Hidalgo, H. G., Das, T., Dettinger, M. D., & Cayan, D. R. (2010). The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. Hydrology and Earth System Sciences, 14(6), 1125–1138. https://doi.org/10.5194/hess-14-1125-2010

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., & Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. Geoscientific Model Development, 9(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., … Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications. Global Environmental Change, 42, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Tompkins, E. L., Adger, W. N., Boyd, E., Nicholson-Cole, S., Weatherhead, K., & Arnell, N. (2010). Observed adaptation to climate change: UK evidence of transition to a well-adapting society. Global Environmental Change, 20(4), 627–635. https://doi.org/10.1016/j.gloenvcha.2010.05.001

UNDRR. (2019). Global Assessment Report on Disaster Risk Reduction 2019. United Nations Office for Disaster Risk Reduction. https://doi.org/10.18356/f4ae4888-en

van der Wiel, K., Bloomfield, H. C., Lee, R. W., Stoop, L. P., Blackport, R., Screen, J. A., & Selten, F. M. (2020). The influence of weather regimes on European renewable energy production and demand. Environmental Research Letters, 14(9), 094010. https://doi.org/10.1088/1748-9326/ab38d3

Wilby, R. L., & Dessai, S. (2010). Robust adaptation to climate change. Weather, 65(7), 180–185. https://doi.org/10.1002/wea.543

Wirtz, A., Kron, W., Low, P., & Steuer, M. (2014). The need for data: Natural disasters and the challenges of database management. Natural Hazards, 70(1), 135–157. https://doi.org/10.1007/s11069-012-0312-4

Wood, A. W., Leung, L. R., Sridhar, V., & Lettenmaier, D. P. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Climatic Change, 62(1–3), 189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9e

Yip, S., Ferro, C. A. T., Stephenson, D. B., & Hawkins, E. (2011). A simple, coherent framework for partitioning uncertainty in climate predictions. Journal of Climate, 24(17), 4634–4643. https://doi.org/10.1175/2011JCLI4085.1