



# EvaSuite: A Framework for Model Evaluation and Verification Using Observational and Model Datasets

**SOFTWARE  
METAPAPERS**

**EVGENII CHURIULIN**

**BEATE GEYER**

**CHRISTOPH BRAUN**

**MAY M. BOHMANN**

**ELEONORA CUSINATO**

**SHWETA SINGH**

**SERGEI PETROV**

**PHILIPP S. SOMMER**

**ANDREAS WILL**

**DIVYAJA LAWAND**

**CHRISTIAN BEIER**

**MUHAMMED KARADAN**

**MARTINA SCHUBERT-FRISIUS**

**RONNY PETRIK**

*\*Author affiliations can be found in the back matter of this article*

**u**[ubiquity press

## **CORRESPONDING AUTHOR:** **Evgenii Churiulin**

Institute of Meteorology and  
Climate Research Troposphere  
Research (IMKTRO), Karlsruhe  
Institute of Technology,  
Karlsruhe, Germany

[evgenii.churiulin@kit.edu](mailto:evgenii.churiulin@kit.edu)

## **ABSTRACT**

Evaluation Suite (EvaSuite) is a modular software framework for the standardised evaluation of regional climate model (RCM) simulations against observational and reference datasets, including model sensitivity analysis. It is designed for high-performance computing platforms and addresses the need for transparent, reproducible, and flexible evaluation workflows for RCMs. EvaSuite combines data harmonisation, spatial and temporal remapping, statistical evaluation, and visualisation within a unified configurable workflow. The framework requires only minimal constraints on input datasets, which can be provided in NetCDF format and identifiable spatial and temporal coordinates, and integrates established climate data processing tools with Python-based modules and a modular extension concept, enabling scalable and customisable analyses. EvaSuite is an open-source evaluation framework maintained by the Climate Limited-area Modelling Community (CLM-Community). It supports systematic reuse across different models, variables, and spatial domains, facilitating consistent evaluation of model performance, including the assessment of model tuning within regional climate modelling.

## **KEYWORDS:**

Climate modelling; Regional climate modelling; Model evaluation and benchmarking; Scientific evaluation software; CLM-Community; ICON Climate Limited-area Mode (ICON-CLM); EvaSuite

## **TO CITE THIS ARTICLE:**

Churiulin E, Geyer B, Braun C, Bohmann MM, Cusinato E, Singh S, Petrov S, Sommer PS, Will A, Lawand D, Beier C, Karadan M, Schubert-Frisius M, Petrik R 2026 EvaSuite: A Framework for Model Evaluation and Verification Using Observational and Model Datasets. *Journal of Open Research Software*, 14: 66. DOI: <https://doi.org/10.5334/jors.729>

## (1) OVERVIEW

### INTRODUCTION

Regional Climate Models (RCMs) simulate a limited-area domain of the Earth system and are constrained by prescribed lateral boundary conditions. They are typically configured at high spatial resolution, whereas driving data, such as general circulation models or reanalysis datasets, are available at coarser resolution. This allows RCMs to provide climate information at the spatial and temporal scales required for assessing future climate impacts (1). Evaluating the performance of RCMs is essential for uncertainty assessment, model development, and enabling robust analyses for parameter tuning purposes. RCM simulations often use diverse and high-resolution grid structures and produce multi-variable outputs at different temporal resolutions, while observational and reference datasets can differ in spatial projections, units, and temporal aggregation. In this context, reference datasets can be either observational data or outputs from other simulations (e.g., reanalysis products or alternative model configurations). Therefore, using a transparent, reproducible, and flexible evaluation framework is essential for robust and reliable analysis. To address these challenges, the Evaluation Suite (EvaSuite) is introduced as a modular software system for evaluating RCM simulations against observational or model-based reference datasets and for comparison of different simulation results.

Several software packages are already available for the evaluation and verification of climate models. ESMValTool (2) offers a broad collection of diagnostics and metrics primarily targeted at global climate model evaluation, in particular for the analysis of global, CMIP-type (Coupled Model Intercomparison Project (3)) simulations. This package is designed for standardised analyses of the large global multi-model ensembles, with limited flexibility for, for example, user-defined region-specific analyses. The Model Evaluation Tools (MET (4)), forming a component of the broader METplus framework (5), was originally developed for the quality assessment of numerical weather prediction models and provides advanced verification methods and flexible statistical tools, but is mainly optimised for short-term forecasts rather than long-term climate simulations. Freva (6) emphasises interactive evaluation, metadata-driven data access, and shared analysis histories within a centralised portal and is primarily designed to operate on standardised climate datasets (CMORised<sup>1</sup>) that follow established climate data conventions and project-specific requirements. Other tools, such as RCMES (7), scores (8), and xCDAT (9), focus on specific aspects of the evaluation process and are typically provided as libraries that support individual components of the workflow, rather than

integrated end-to-end evaluation frameworks. Existing evaluation tools are typically designed for global model intercomparison, forecast verification, or portal-based analysis environments. While many of these tools are flexible and can be adapted to different use cases, they are not specifically designed with a focus on systematic RCM tuning and benchmarking workflows. This motivated the development of an evaluation framework aligned with the requirements of the CLM-Community.<sup>2</sup>

EvaSuite is explicitly designed for RCM evaluation on high-performance computing (HPC) systems, supporting automated execution, flexible configuration, and a wide range of evaluation diagnostics. EvaSuite addresses the practical challenges of model evaluation. In this case, it is tailored to RCMs and includes iterative model tuning, grid heterogeneity, and the ability to combine preprocessing, evaluation, and visualisation within a single workflow. This is achieved without the need for prior CMORisation. The current version (v1.0.4) provides a unified workflow that covers data preparation, spatial remapping, temporal resampling, and computation of a wide range of evaluation metrics across user-defined subdomains and time periods, facilitating the integration of heterogeneous observational and reference datasets that often differ in spatial coverage, resolution, and formatting. Multiple grid types are supported, including regular and rotated latitude-longitude grids, Lambert Conformal Conic grids, and single-location observations. EvaSuite enables consistent handling of different resolutions and projections, as well as the harmonisation of variables with differing units and temporal representations (e.g., instantaneous values, fluxes, and accumulated quantities) during data preparation. EvaSuite is user-configurable via YAML (10) configuration files, enabling automated execution of the evaluation pipeline from raw input data to final visual output without manual intervention. Wrapper scripts support the efficient execution of multiple job tasks, each processing, for example, a specific variable, metric, or observational dataset. This design makes the framework suitable for HPC environments with extensive parallelisation capabilities. The framework includes a wide range of statistical metrics, ranging from standard measures (e.g., bias, root mean square error (RMSE), mean absolute error (MAE), linear correlation) to advanced diagnostics such as quantile-based metrics, Kling-Gupta efficiency, and bootstrap-based significance testing. Evaluations can be performed for arbitrary spatial domains defined through user-defined polygons, supporting analyses from continental-scale regions to user-defined subregions (e.g., river catchment areas or small administrative regions). EvaSuite is actively used, maintained, and further developed within the CLM-Community for RCM quality assessments, sensitivity studies, and the evaluation of different model configurations and parameter tuning (11–14).

This paper presents the system design, implementation aspects, description of external packages, and application examples of EvaSuite. Section ‘The EvaSuite framework’ provides an overview of the workflow. Sections ‘Major processing steps’ and ‘Implementation details’ describe the architecture and implementation (software and design). Section ‘Platform integration’ informs about platform implementations, requirements, and support. Configuration options and documentation are described in Section ‘Configuration and documentation’, while modular design and extensions are presented in Section ‘Modular components and extensions’. Application examples illustrating typical evaluation outcomes are given in Section ‘Scientific application of EvaSuite’.

## THE EVASUITE FRAMEWORK

EvaSuite is a community-developed evaluation system for standardised evaluation of RCM simulations against reference datasets. It provides a configurable, unified workflow that automates spatial and temporal data transformations, the computation of statistical metrics, and visualisation, while supporting collaborative evaluation through shared configurations referencing distributed observational and reference datasets. The traceable workflow and its centralised handling ensure transparency, consistency, and reproducibility across evaluations by omitting the need for individual

scripting solutions. This facilitates systematic variation of evaluations across the analysed variables, metrics, and spatial resolutions and reduces the risk of errors.

EvaSuite is implemented as a modular framework which is designed for new and experienced users working with RCM simulation output on HPC systems. The conceptual architecture is shown in [Figure 1](#).

## Major processing steps

EvaSuite automates data harmonisation (Section ‘Data harmonisation’), regridding (Section ‘Regridding’), spatial masking, temporal aggregation, and metric computation (Section ‘Statistics and evaluation metrics’), followed by configurable visualisation (Section ‘Visualisation’) output as described in [Figure 2](#). EvaSuite can be applied in two modes. In [CompareRAW](#) mode, all datasets are compared to each other in terms of absolute values, without calculating metrics. In [EvalRAW](#) mode, one or multiple datasets are evaluated against a reference dataset using the metrics specified by the user.

## Data harmonisation

In a first step, EvaSuite extracts user-defined variables and time periods from the specified datasets based on mappings provided in the YAML configuration file. This allows harmonisation of differing variable names and units across datasets, such as 2 m air temperature (e.g., **tas**, **T\_2m**, or **tg** in K or °C) or surface pressure (e.g., in hPa or Pa), as well as standardisation of temporal

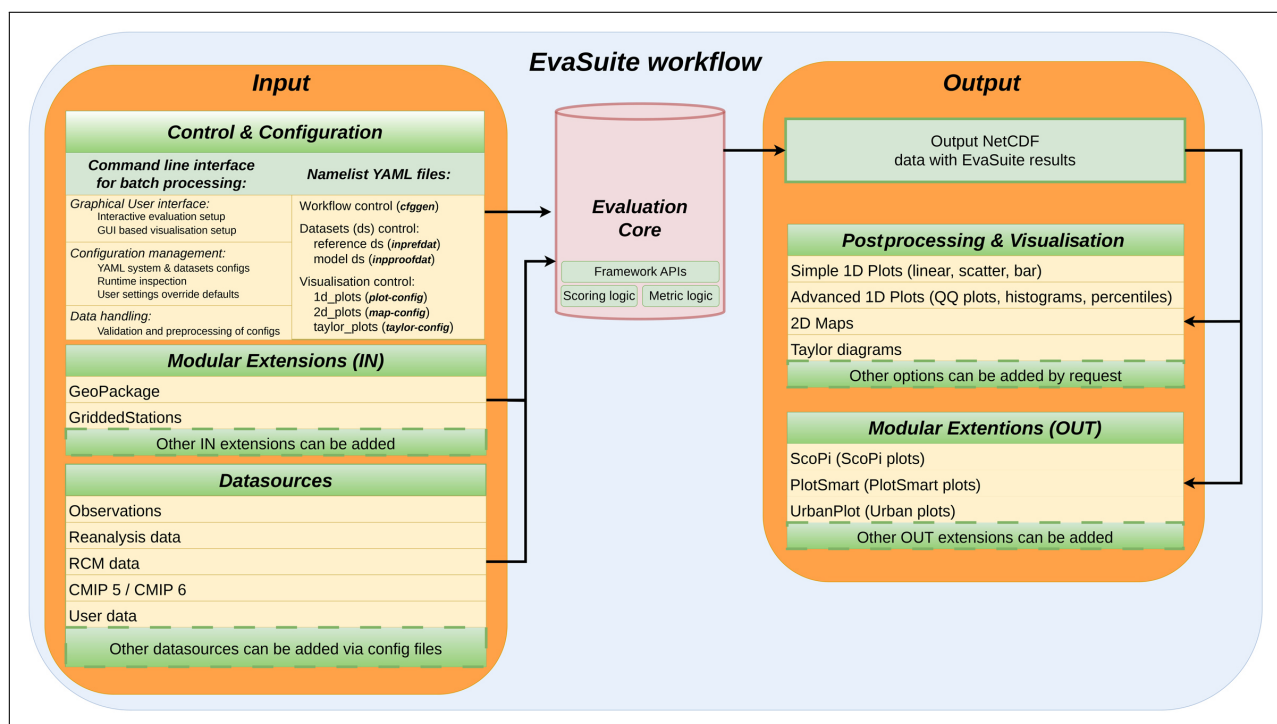


**Figure 1** Conceptualised architecture of the EvaSuite workflow. The control and configuration layer provides the user interface and interacts via the evaluation core (centre) with the processing, regridding, computing, and reporting modules. Modular extensions (IN/OUT) and heterogeneous data sources are integrated through defined interfaces and positioned laterally within the architecture. The detailed block scheme of the evaluation core is presented in [Figure 2](#).

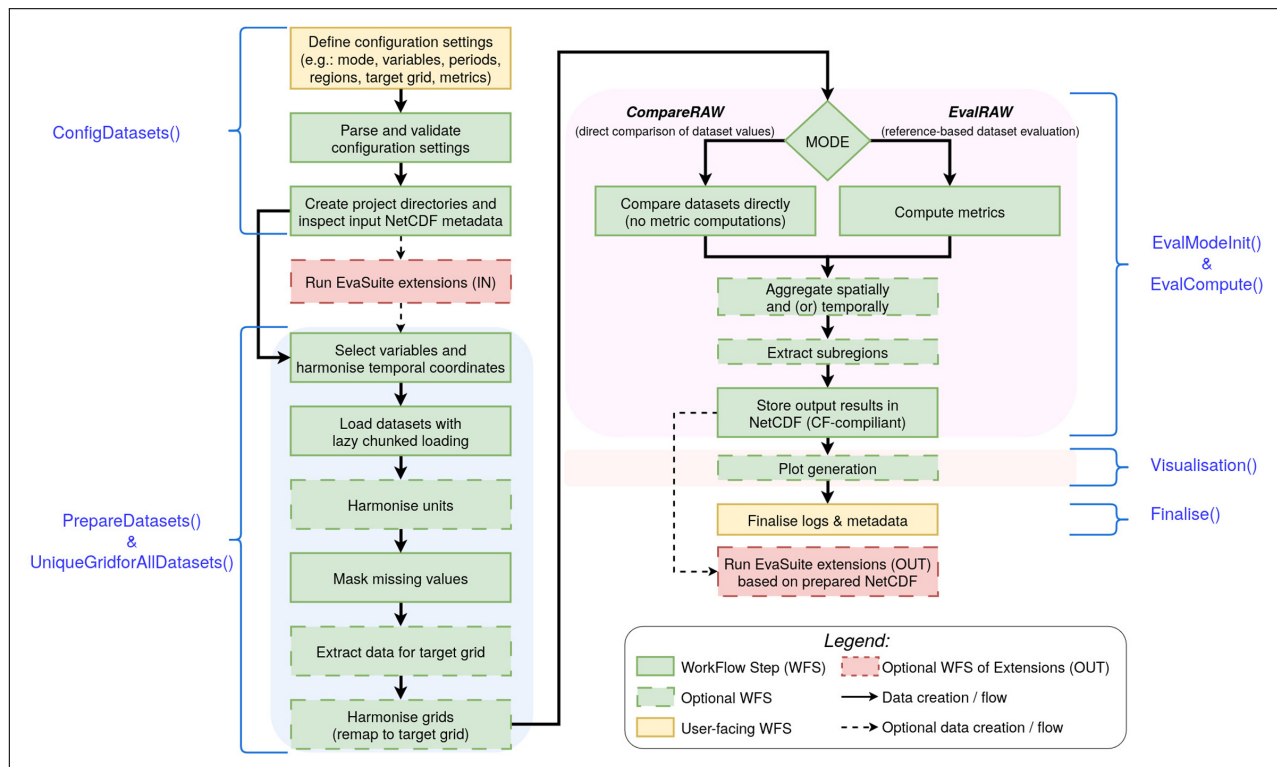


**Figure 2** Block scheme of the EvaSuite core. Blue annotations indicate its high-level methods.

coordinates according to user-defined rules. To ensure consistency of the datasets used, input datasets must comply with the NetCDF Climate and Forecast (CF) Metadata Conventions (15).

### Regridding

Regridding enables model, observational, and reanalysis datasets to be transformed onto a common horizontal and vertical grid to ensure comparability. Horizontal interpolation is performed using Climate Data Operators (CDO) (16), supporting bilinear, nearest-neighbour, distance-weighted, and conservative (first- and second-order) remapping methods. These methods can be applied using reference grid descriptions or precomputed weight files to optimise remapping. Vertical interpolation is supported through linear interpolation to target pressure or height levels and is suitable for atmospheric evaluation. All regridding operations, including target grid selection, interpolation method, vertical level definition, and masking, are defined in the configuration file, allowing consistent integration of datasets with different grids, spatial resolutions, and coordinate systems. As a limitation of the current implementation, the availability of individual processing methods can depend on the version of the external tools used by EvaSuite. For example, second-order conservative remapping relies on the CDO **remapcon2** operator, which is no longer available from CDO version 2.3.0 onwards. Changes in external dependencies such as CDO, NCO, and NCSYS therefore need to be considered when maintaining the workflow.

### Statistics and evaluation metrics

To quantify differences between model simulations and reference datasets, EvaSuite provides a comprehensive set of evaluation metrics supporting flexible temporal and/or spatial aggregation, including user-defined aggregation intervals prior to metric computations. All metrics are configured in the configuration file, allowing flexible selection of aggregation dimensions, datasets, and significance levels within a standardised workflow. Implemented metrics include:

- Bias-based metrics (absolute and normalised);
- Error measures such as absolute error, median absolute deviation, and RMSE;
- Correlation metrics;
- Distribution-based diagnostics, including quantile-based measures and probability density function-based skill scores;
- Skill and contingency scores, including Perkins Skill Score, Kling-Gupta efficiency, and binary event-based scores (e.g., accuracy, probability of detection, false alarm ratio);
- Statistical significance tests, including Student's t-tests, bootstrap-based significance tests, and confidence-interval-based approaches for comparing metrics between experiments.

### Visualisation

Visualisation plays an important role in presenting the evaluation results in a clear and scientifically meaningful way. The framework supports automatic generation of

multiple plot types, including time series, spatial maps, Taylor diagrams, scatter plots, box plots, histograms, and distribution plots, using standardised configuration files, ensuring reproducibility and consistent appearance. Visual styles, colour schemes, axis limits, thresholds, geographical projections, zoom levels, and land-sea masking can be customised. Figures can be exported as raster images or as publication-ready PDF files. Representative examples of the default visualisation capabilities are provided in Sections ‘Spatial diagnostic and map-based evaluation example’, ‘Station-based evaluation example’, ‘Distribution-based diagnostics: QQ plots and percentiles’, and ‘Multi-metric summary: Taylor diagram’, while additional visualisation options are available through extensions (Section ‘Modular components and extensions’).

### Implementation details

The main components are implemented in Python (17), while auxiliary scripts are written in shell (18). Operations are implemented using documented functions with type hints and descriptive documentation strings. Some functions are currently covered by unittests (19), and unittest coverage is planned to be extended to all functions in future releases. The software is developed and distributed as a Python package, with additional dependencies installed via pip (20). Modular components and Extensions are also written in Python (with the exception of GriddedStations, which is implemented in R (21)) and are integrated through a modular architecture.

The main EvaSuite source code is located in the **hereon\_evasuite** Python package, with **EvaSuite.py** serving as the main application entry point. After installation, the workflow is invoked through the **evasuite** command, which calls the **main()** function in **EvaSuite.py**. This function processes the command-line arguments, loads the required configuration files, initialises the **EvaSuite** class, and starts the workflow execution. The main processing stages and their corresponding high-level methods are shown in Figure 2. Dataset configuration is handled by **ConfigDatasets()**, followed by **PrepareDatasets()** and **UniqueGridforAllDatasets()** for preprocessing and grid harmonisation. The evaluation stage is controlled by **EvalModeInit()** and **EvalCompute()**, with the choice between the **EvalRAW()** and **CompareRAW()** modes determined by the user-defined configuration settings. At the visualisation stage, graphical output is generated by **Visualisation()**. The workflow is then completed by **Finalise()**, where the associated metadata are collected.

Each processing step is logged, including the configuration context, input files, applied methods, runtime warnings, and generated output. This supports workflow reconstruction, anomaly tracing, and comparison across experiments and computing environments. Metadata integrity is maintained

through standardised repair routines that preserve the original attributes, units, and variable names during transformations. The built-in validation mechanisms check user-defined settings, detect combinations of unsupported parameters, and generate informative exception messages without prematurely terminating execution. Consistency checks across datasets, metrics, and spatial domains help to identify mismatches early and reduce error propagation, thereby enhancing the credibility of evaluation results.

### Platform integration

EvaSuite is designed to operate on Linux-based IT platforms (22). A Python environment is required for deployment. For offline HPC systems, pre-built Conda environments or container-based setups can be packaged and deployed using the provided helper scripts. Development and testing were performed on HPC systems at the German Climate Computing Center (DKRZ – Levante<sup>3</sup>), the High-Performance Computing Center Stuttgart (HLRS – Hawk<sup>4</sup>), the Deutscher Wetterdienst (DWD – SX-Aurora<sup>5</sup>), and the Swiss National Supercomputer Center (CSCS – Alps<sup>6</sup>), where the required external tools (CDO, NCO (23) and NetCDF utilities (24)) are available. Installation on local workstations is technically possible, but full functionality has been validated only on HPC platforms.

### Configuration and documentation

All user interaction is based on a flexible configuration system driven entirely by YAML configuration files, allowing evaluation workflows to be defined without modifying the source code. Six command-line arguments can be used to specify the corresponding YAML configuration files: **cfggen** for general workflow control, **inprefdat** and **inpproofdat** for reference and model dataset descriptions, and **map-config**, **plot-config**, and **taylor-config** for spatial plots, time series and statistical plots, and Taylor diagrams. Visualisation configuration files are required only when the corresponding user-defined visualisation settings are requested. Otherwise, EvaSuite uses the available default settings. This clear separation supports reproducibility, reuse, and collaborative development across projects (e.g., COPAT<sup>7</sup>, NUKLEUS<sup>8</sup> and UDAG<sup>9</sup>). In particular, the reference and model dataset description files enable a community to work efficiently with distributed observational and model data and their version control. The complexity of the evaluation can be adjusted from basic settings for rapid testing to advanced options for fine-tuning interpolation methods, domain masking, or plot layout.

The execution of EvaSuite is recommended to be handled via a shell-based workflow using resource management tools, for example, Simple Linux Utility for Resource Management (SLURM) (25).

### Documentation

Comprehensive documentation is available via GitLab pages (26). It outlines the scope, structure, and workflow of EvaSuite and provides detailed guidance on installation, configuration, and execution. The documentation includes a step-by-step introduction to using EvaSuite on HPC systems, such as DKRZ, based on example use cases, comprising 10 simple, 7 close-to-reality predefined test cases and publicly available test datasets. The test cases are adapted to resource handling via SLURM. The stand-alone documentation is complemented by Markdown-based content hosted in the EvaSuite GitLab repository, providing online access to up-to-date user and developer guidance.

### Modular components and extensions

Extensions are developed as independent Git repositories within the CLM-Community GitLab and are linked to EvaSuite via Git submodules. This structure allows extensions to be shared and maintained collaboratively with the community. For long-term accessibility and reproducibility, these extensions are archived and published through Zenodo. All components are version-controlled using Git (27) and are briefly described in the EvaSuite documentation. The modular design allows components to be developed and maintained independently while remaining fully integrated into the overall workflow.

### GeoPackage

GeoPackage supports the preparation and management of spatial region definitions within the EvaSuite workflow, providing standardised polygon geometries for dataset subsettings via spatial masking. The GeoPackage module contains collections of predefined subregions (28) for EvaSuite in GPKG format (29). These polygon definitions serve as static spatial inputs and are applied consistently across experiments and model configurations. Furthermore, GeoPackage provides two interfaces for exploring, modifying, and extending the set of predefined regions: a command-line interface integrated into EvaSuite and a Jupyter notebook that facilitates visual inspection of existing polygons and creation of custom user-defined regions.

### GriddedStations

The GriddedStations extension enables the preprocessing of observational station data. Time series from meteorological stations, such as those provided by DWD (30, 31), are converted into NetCDF format. Station coordinates are used to map the observational data onto a target model grid, including rotated and curvilinear grids such as the EURO-CORDEX grid (32). This mapping allows station-based observations to be compared with the gridded RCM output. Quality control is integrated into

the preprocessing workflow to ensure data reliability. As part of this step, stations are retained only if they meet a minimum data availability threshold (e.g., 90% of expected observations per year). This criterion is applied independently of the evaluation period and serves solely to exclude incomplete or sparsely sampled time series from further analysis.

### Significance tests and score points of evidence (ScoPi)

The recently developed ScoPi metric supports the evaluation and visualisation of statistically significant differences between model experiments (12). It operates on EvaSuite evaluation results and applies significance testing to quantify coherent improvements or degradations relative to a reference experiment across metrics, variables, and subregions with respect to observational datasets. The resulting scores for each simulation enable standardised comparison on the basis of EvaSuite's regional analysis results and are visualised using dedicated ScoPi plots. This approach supports the flexible comparison of multiple model configurations. A representative example is presented in Section 'Multi-metric summary: ScoPi plots'.

### PlotSmart

PlotSmart focuses on the generation of high-quality, publication-ready figures from NetCDF datasets produced by EvaSuite. It provides flexible control over subplots, titles, colour scales, and colour bars, enabling consistent visualisation of multiple variables, models, and spatial domains within a single figure. In contrast to the standard EvaSuite visualisation routines, which focus on predefined diagnostic output, PlotSmart supports visualisation requirements that commonly arise during the publication process, such as customised multi-panel layouts, simultaneous display of multiple parameters, and adaptation to journal-specific formatting standards. A representative example is presented in the PlotSmart subsection of 'Scientific Application of EvaSuite'.

### UrbanPlot

UrbanPlot provides high-resolution, urban-scale visualisation of output from RCM simulations. It is specifically designed for rendering ICON and preprocessed EvaSuite NetCDF datasets. UrbanPlot combines dynamic model fields with static OpenStreetMap (OSM) data, including administrative boundaries, buildings, roads, parks, and hydrography. Spatial buffering in a projected coordinate system ensures geometrically correct domain selection and simplification prior to visualisation. The layout is created once and reused across frames, enabling efficient batch production of time-resolved figures and animations. A representative example is presented in the UrbanPlot subsection of 'Scientific Application of EvaSuite'.

## SCIENTIFIC APPLICATION OF EVASUITE

This section illustrates the practical application of EvaSuite and its extensions (ScoPi, PlotSmart, UrbanPlot) using representative evaluation examples. We are using real simulation data with their simulation IDs, including RCM name, forcing data abbreviations, or test case numbers. All of these names are not important for understanding and, therefore, are not explained further but are used to show the functionality of the EvaSuite.

### Spatial diagnostic and map-based evaluation example

EvaSuite supports spatially explicit diagnostics that allow detailed inspection of model performance patterns. [Figure 3](#) presents an example of temporally aggregated bias maps for near-surface air temperature (**tas**) over Europe for a selected evaluation period. The figure shows bias fields for two model configurations relative to a reference dataset, as well as their difference. Such spatial diagnostics enable identification of regionally coherent model biases and systematic differences between experiments.

### Station-based evaluation example

EvaSuite is not limited to gridded model output but also supports the evaluation of station-based observational datasets within the same workflow. [Figure 4](#) shows an example application using downward shortwave radiation (**rsds**) observations from DWD stations in Germany. The monthly mean values derived from station observations are evaluated alongside corresponding values extracted from multiple RCM simulations using the grid cells closest to the station locations. The station coordinates are mapped to the model grid cells based

on the minimum distance to the centre of the grid cell. The resulting time series enable a direct comparison of the seasonal cycle between observations and model configurations. Differences in amplitude and timing between datasets are clearly visible, providing an intuitive first-order assessment of model performance.

### Distribution-based diagnostics: QQ plots and percentiles

Beyond mean-based metrics, EvaSuite provides distribution-based diagnostics through built-in quantile-quantile (**QQ**) analysis. [Figure 5a](#) illustrates the application of these diagnostics to daily minimum 2 m air temperature (**tasmin**) in a user-selected domain. QQ-based methods enable the assessment of agreement across the full value range and support the identification of systematic deviations within the distribution. In addition, EvaSuite enables the generation of percentile curves ([Figure 5b](#)), which facilitate a structured comparison of extremes and distribution tails between selected datasets.

### Multi-metric summary: Taylor diagram

EvaSuite supports multi-metric model evaluation using the Taylor diagram approach. The Taylor diagram provides a compact statistical summary of model performance by jointly representing standard deviation, correlation coefficient, and the centred root mean square difference relative to a reference dataset ([34](#)). This visualisation enables efficient comparison of multiple RCMs, sensitivity experiments, and variables within a single plot. [Figure 6](#) shows the Taylor diagram that illustrates the performances of different RCMs' sensitivity experiments for 2 m temperature over the selected domain.

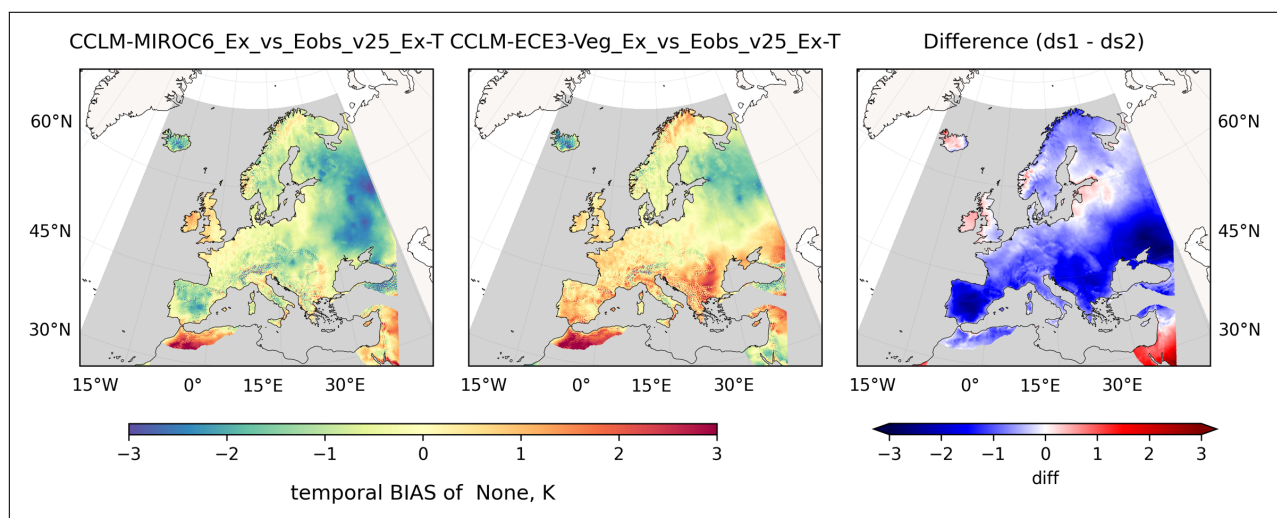


**Figure 3** Monthly mean bias of **tas** over Europe for July 2014, relative to the EObs (33) observational dataset. Positive (negative) values indicate overestimation (underestimation) of **tas**. The left panel shows the bias for the CCLM\_MIROC6 simulation, the middle panel for CCLM\_EC\_Earth3\_Veg, and the right panel shows the difference between the two bias fields (CCLM\_MIROC6 minus CCLM\_EC\_Earth3). All panels are shown on the EObs reference grid.

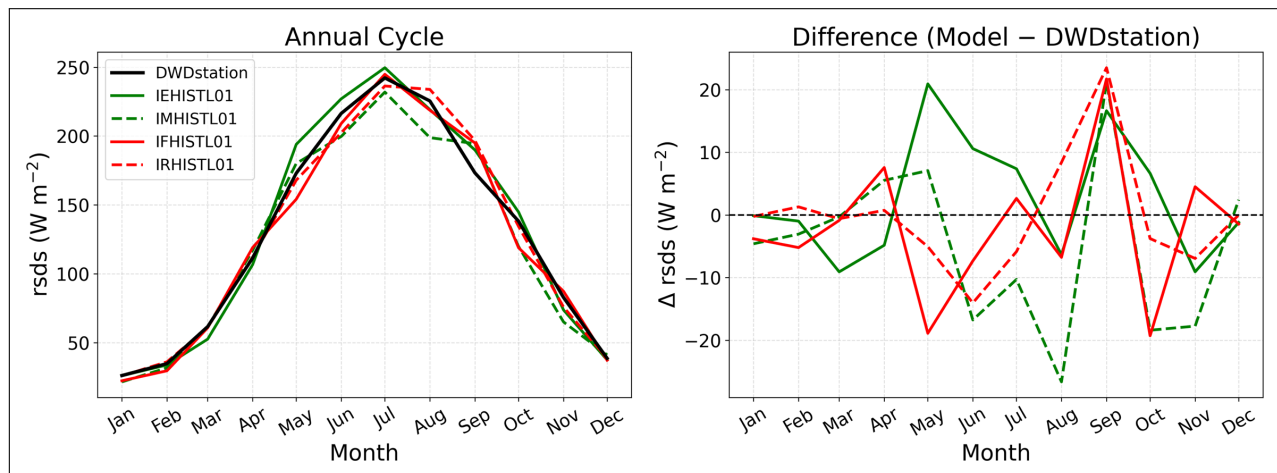


**Figure 4** Monthly mean **rsds** from 2003 to 2008, averaged over Germany's DWD station locations: DWD observational data (black) and several RCM simulation data (coloured). The left figure shows the annual cycle derived from monthly averages, and the right figure highlights differences in the amplitude and seasonal evolution between observational station data and model results.

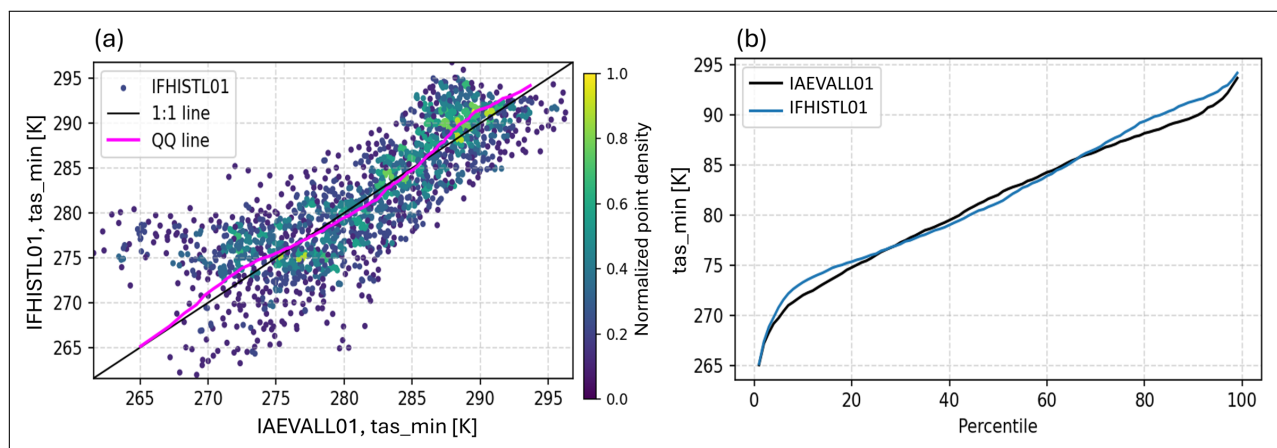


**Figure 5** Distribution-based evaluation of daily **tasmin** over the Mid-Europe region. The left panel shows a scatter plot of model values (IFHISTL01) against the reference dataset (IAEVALL01), including the 1:1 line and the QQ line. Point colours indicate the normalised data density. The right panel displays the corresponding percentile curves for the model simulation and the reference. Deviations from the 1:1 and QQ lines and differences between percentile curves highlight systematic biases and distributional differences across the full temperature range.

### Multi-metric summary: ScoPi plots

The recently introduced ScoPi metric for aggregated simulation evaluation is explained in detail by (12). Figure 7 gives an example of the relative performance of the selected model configurations in COSMO-CLM and ICON-CLM<sup>10</sup> through ScoPi scores. The box plots indicate the  $\text{ScoPi}_{\text{region}}$  values computed across selected climate variables, here **tas**, **tasmax**, **tasmin**, **rsds**, **psl**, for the period 2003–2007 against EOBS data, shown separately for each PRUDENCE region (13). The score  $\text{ScoPi}_{\text{simulation}}$  given on the y-axis shows the weighted sum over all PRUDENCE regions. In this example, the COSMO-CLM setup C2C316 serves as the reference (on the x-axis). C2I200c and C2I301c refer to two ICON-CLM simulation setups that were generated during a systematic model optimisation process. The final ICON-CLM configuration C2I301c, which attains the highest ScoPi values, indicates the greatest overall improvement relative to C2C316.

### PlotSmart

Figure 8 demonstrates how PlotSmart enables consistent visualisation of multi-variable climate model biases across different spatial domains within a single, coherent subplot framework. More real PlotSmart examples are available in (13).

### UrbanPlot

Figure 9 shows the **tas** bias between EOBS, regridded to the 1 km ICON domain, and the ICON-CLM simulation output over the Paris metropolitan area. Due to its higher resolution, the ICON-CLM simulation shows greater intra-urban variability than the EOBS dataset. In the Paris region, EOBS is strongly influenced by observations from airport stations (e.g., Orly and Le Bourget), whose surroundings differ substantially from those of the urban core and are therefore not fully representative of inner-city conditions. In the city centre, the EOBS dataset shows

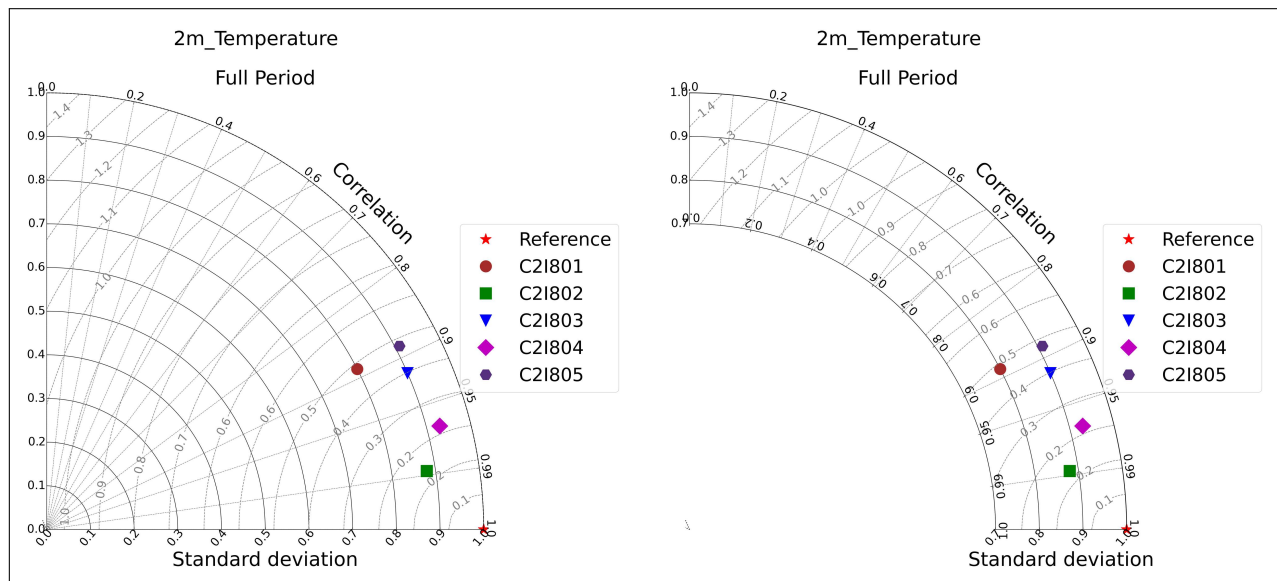


**Figure 6** Taylor diagram illustrating performances of different sensitivity experiments for **tas**. Left: standard quarter-circle representation. Right: arc-shaped representation focusing on a restricted standard deviation range.

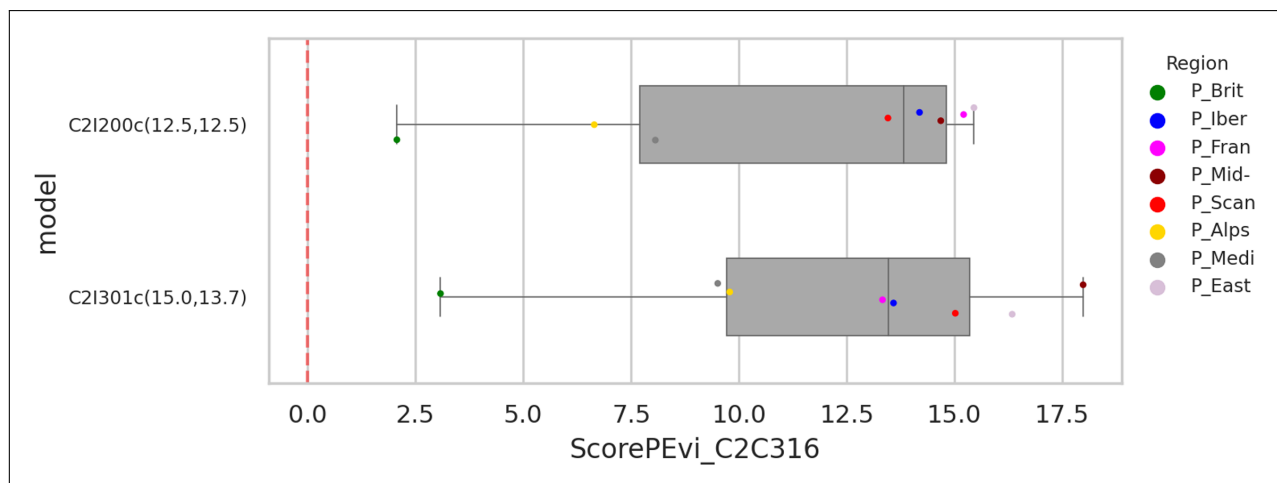


**Figure 7**  $\text{ScoPi}_{\text{region}}$  based on the differences in the mean BIAS of standard climate variables (**tas**, **tasmax**, **tasmin**, **rsds**, **psl**) between the observations and each simulation of selected configurations of ICON-CLM (C2I200c, C2I301c), against the COSMO-CLM setup C2C316. The colours indicate the eight PRUDENCE analysis regions. Small vertical offsets are applied to the regional data points within each configuration to avoid overlap and improve visibility. They have no quantitative meaning. The numbers given on the y-axis labels in brackets are the  $\text{ScoPi}_{\text{simulation}}$  representing the averages over all eight PRUDENCE regions, weighted by the distance to Mid-Europe (first value) and by their area (second value).

higher **tas** values than the model output. This indicates an underestimation of the temperatures in the ICON-CLM simulation for the chosen month.

## DISCUSSION AND CONCLUSION

This paper presents EvaSuite, a modular and automated evaluation framework designed to address key challenges in the assessment of RCM simulations. EvaSuite is a practical response to the increasingly complex requirements of RCM evaluation workflows. Rather than aiming for maximal generality, the framework prioritises automation, configurability, and reproducibility in environments where RCM simulation evaluation is performed iteratively. This

focus is particularly relevant for tuning activities, where simulations with varying configurations have to be evaluated for different variables, metrics, and regions consistently and efficiently.

A key design feature of EvaSuite is the tight integration of preprocessing, evaluation, and visualisation within a single workflow. Many existing tools treat these steps as loosely coupled components, often requiring users to bridge gaps through custom scripting. Although such flexibility can be advantageous, it also introduces opportunities for inconsistency and undocumented methodological divergence. EvaSuite addresses this by enforcing standardised preprocessing and metric computation through configuration driven workflows,

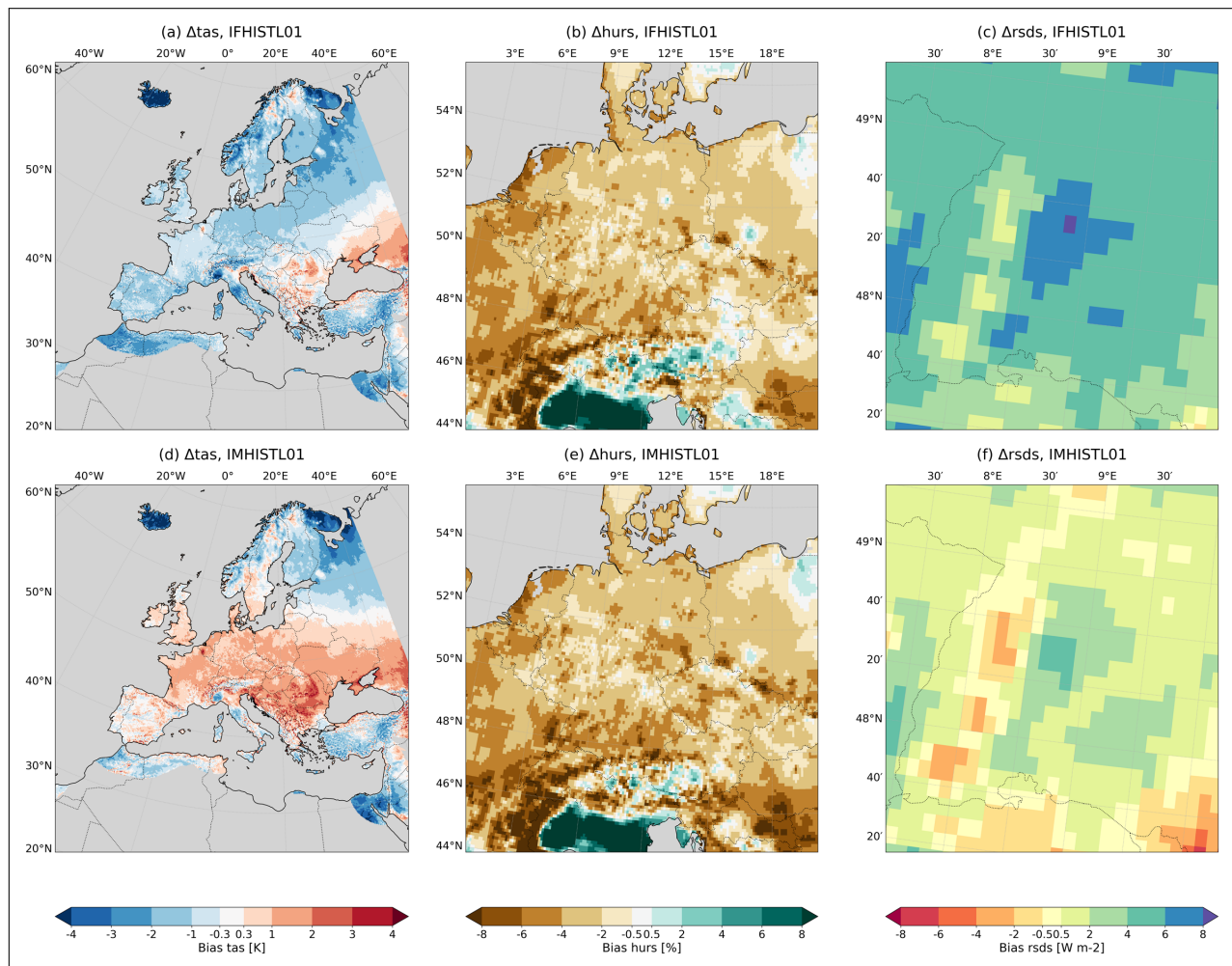


**Figure 8** Biases of **tas** (left column) over Europe, relative humidity (**hurs**, middle column) over Central Europe, and **rsds** (right column) over the Baden-Württemberg region in south-western Germany, shown at progressively zoomed spatial scales. Each row represents an ICON-CLM simulation dynamically downscaling CMIP6 global climate data. Biases are computed with respect to the observational dataset (EOBS) and over the period 1961–1990 for the winter season. Positive (negative) values indicate an overestimation (underestimation) of the considered variable.

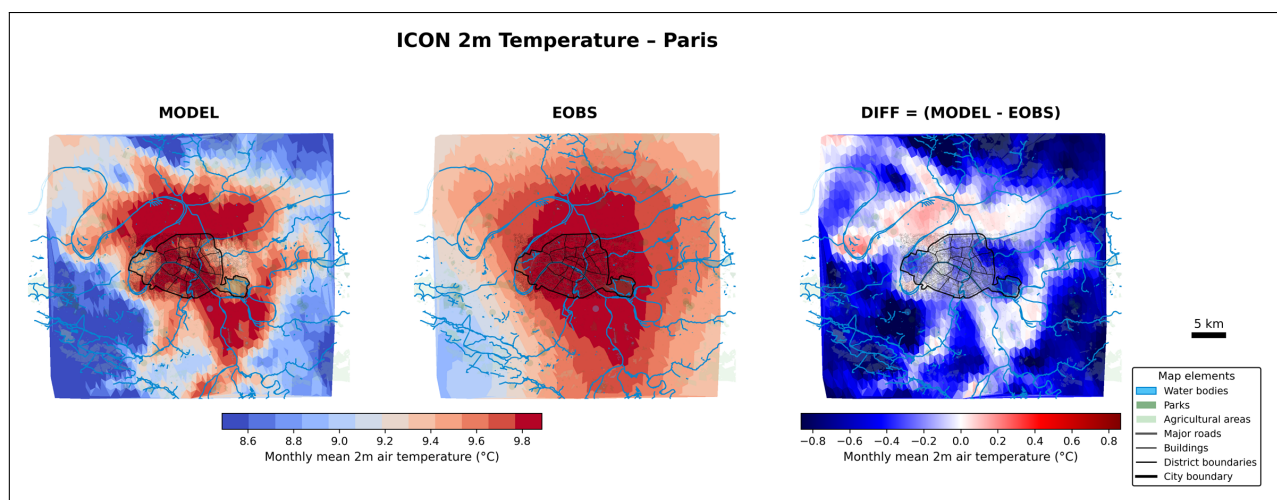


**Figure 9** Monthly mean **tas** over Paris metropolitan area for April 2023. Left: ICON-CLM simulation (0.009° grid spacing), Middle: EOBS (0.1° resolution), Right: bias (ICON-CLM – EOBS).

thereby reducing ambiguity in evaluation outcomes. This approach is especially beneficial in cross-institutional collaborations, where comparable diagnostics between

users and projects are essential. Moreover, the end-to-end approach of EvaSuite prevents not only from undocumented methodological divergence but also

from storing a lot of intermediate results – only the original reference dataset and the final results remain.

The reliance on YAML configuration files represents another deliberate trade-off in flexibility. By externalising workflow control, EvaSuite lowers the barrier for non-developer users and supports reproducibility across computing environments. At the same time, this design constrains ad hoc experimentation compared to interactive analysis frameworks. EvaSuite favours systematic evaluation over exploratory analysis, a choice that aligns with its primary application in structured tuning and benchmarking rather than rapid prototyping.

EvaSuite's focus on regional climate evaluation distinguishes it from tools primarily designed for global model evaluation or operational forecast verification. Features such as GeoPackage-based region handling, gridded station support, and distribution-based diagnostics enable targeted assessment of model behaviour across heterogeneous spatial domains and observational sources. The ScoPi extension further extends this capability by enabling statistically grounded comparison of experiments, addressing a common challenge in interpreting multi-metric evaluation results. The framework's utilisation of established external tools such as CDO and NCO ensures robustness and performance but also ties EvaSuite to Linux-based HPC infrastructures. While this aligns with most RCM simulation and analysis environments, it limits immediate portability to more lightweight or cloud-based platforms. Similarly, the current focus on NetCDF CF-compliant datasets streamlines processing but may require additional preprocessing for non-standard data sources. This disadvantage will disappear as soon as all data providers comply with these guidelines, which are essential for scientific work.

EvaSuite provides a structured and reproducible approach to RCM evaluation that is explicitly tailored to the demands of HPC based, community driven modelling workflows. By prioritising end-to-end automation, transparent configuration, and region-focused diagnostics, it addresses gaps not addressed by more general evaluation frameworks. The modular design of the software allows it to evolve alongside emerging evaluation needs while maintaining methodological consistency. As such, EvaSuite facilitates systematic RCM assessment and model development, particularly in the context of iterative tuning and coordinated evaluation efforts.

## QUALITY CONTROL

EvaSuite has been successfully applied in the NUKLEUS<sup>9</sup> and UDAG<sup>10</sup> projects producing a number of publications (11–14). The framework has also been tested by several installations on different HPC environments

(Section 'Platform integration'). A functional installation and test workflow is described in the subchapter of the documentation (35). The current unittest suite comprises 55 test files covering input and configuration validation, evaluation metrics, general utility functions, and CDO/NCO/Xarray-based processing routines. Unittests are executed automatically through the GitLab continuous integration (CI) pipeline. Future extensions of the test suite will focus on broader testing of preprocessing, core and postprocessing components, synthetic and real end-to-end workflows, additional edge and failure cases, and comparisons against independently defined reference results.

## (2) AVAILABILITY

### OPERATING SYSTEM

Linux (tested on: Levante (DKRZ), Hawk (HLRS), SX-Aurora (DWD), and Alps (CSCS) machines).

### PROGRAMMING LANGUAGE

Python (3.10) and BASH (must exist)

### ADDITIONAL SYSTEM REQUIREMENTS

SLURM scheduler (required for the supported HPC workflows), environment modules (optional), memory  $\geq$  6GB RAM and 30 GB of disk space, Git, CDO, NCO, NetCDF system utilities (NCSYS), and optionally GDAL.

### DEPENDENCIES

EvaSuite uses a predefined Python environment with fixed dependency versions to support a reproducible evaluation workflow. The complete Python dependency specification is provided in **pyproject.toml**. Core dependencies include NumPy, SciPy, Pandas, Xarray, NetCDF4, Dask, GeoPandas, Fiona, Regionmask, Matplotlib, Cartopy, Shapely, PyProj, Statsmodels, h5py, and xskillscore. External software dependencies include CDO, NCO, and NCSYS. GDAL may additionally be required by geospatial Python packages such as Fiona, depending on the installation environment. For the current EvaSuite release, the reference environment validated on DKRZ Levante uses Python 3.10 together with CDO 2.0.5, NCO 5.0.6, and NCSYS 4.8.1. The GriddedStation extension additionally requires R 4.1.2 but is not required for the main EvaSuite workflows or other extensions.

### LIST OF CONTRIBUTORS

The EvaSuite framework has been developed through the collaborative efforts of the CLM-Community and associated research partners. Ronny Petrik, Beate Geyer, and Burkhardt Rockel (affiliated at Helmholtz-Zentrum Hereon GmbH) developed the basic core of EvaSuite and used it for the first evaluation results.

## SOFTWARE LOCATION

### Archive

**Name:** Zenodo

**Persistent identifier:** <https://doi.org/10.5281/zenodo.17130605>

**Licence:** Apache-2.0

**Publisher:** Beate Geyer

**Version published:** v1.0.4

**Date published:** 19/10/2026

### Code repository

**Name:** GitLab

**Identifier:** <https://gitlab.dkrz.de/clm-community/public/evasuite>

**Licence:** Apache-2.0

**Date published:** 14/10/2026

## LANGUAGE

English

## (3) REUSE POTENTIAL

EvaSuite is intended for reuse by the regional climate modelling community as an adaptable framework for the evaluation of RCM simulations. It can be applied across a broad range of gridded model and observational datasets, supporting both routine model assessment and iterative tuning activities. Analyses based on EvaSuite can easily be tailored to custom domains as well as extended to metrics beyond the currently available set. The modular structure of EvaSuite allows users to adapt the framework to new scientific questions and technical environments. Typical extensions include the addition of new evaluation metrics or statistical tests, the inclusion of further spatial region definitions, the use of alternative data storage systems, or the replacement of external processing tools. EvaSuite can be linked to existing data management platforms, for example, by embedding it in Freva (6) through automated configuration file generation.

The current EvaSuite version also has several limitations that should be considered when reusing the framework. It was primarily developed and validated for Linux-based, SLURM-managed HPC environments. Local execution is possible and is supported by the updated launcher scripts, but requires the user to install and configure external dependencies such as CDO, NCO, and NCSYS. The documentation holds installation instructions for such a case based upon the Conda package installer, but full functionality cannot be guaranteed, and HPC environments are preferred. The availability of some processing methods can additionally depend on the versions of these external tools. EvaSuite also relies on a predefined Python environment with fixed dependency versions. While this limits flexibility in combining EvaSuite with alternative or

newer package versions, this approach was chosen to meet the requirements of scientific users for reproducible evaluation results and to provide a consistent software environment across different supported platforms and computing systems. This is particularly important for long-term scientific reproducibility, as it enables users to reproduce analyses performed with a specific EvaSuite release even several years later and to maintain comparability with previously generated results. Input datasets are expected to be provided in NetCDF format with CF-compliant metadata and identifiable spatial and temporal coordinates. Datasets that do not fulfil these requirements may therefore require additional preprocessing. Furthermore, the current unittest suite covers only part of the framework, and broader coverage of preprocessing, core, postprocessing, and end-to-end workflows is planned for future releases.

EvaSuite shares commonly used evaluation concepts, statistical metrics, and graphical diagnostics with existing climate evaluation tools. Its reuse potential should be understood primarily in terms of how these established evaluation approaches are organised within the workflow rather than as an alternative implementation of individual metrics. In particular, EvaSuite combines preprocessing, evaluation, visualisation, user-defined regional analysis, and model sensitivity assessment within a configuration driven workflow tailored to iterative RCM evaluation on HPC systems. EvaSuite is developed openly, and contributions in the form of code, extensions, or feature suggestions are welcome via the project repository (36). Bug reports, feature requests, and other technical issues can be submitted through the GitLab issue tracker, while code contributions can be proposed through merge requests. For closer scientific collaboration or larger extensions, interested users are encouraged to contact the authors via [imktro-climate-modelling@lists.kit.edu](mailto:imktro-climate-modelling@lists.kit.edu). The codebase is modular and documented to support long-term reuse.

## NOTES

- 1 [https://tutorials.dkrz.de/tutorial\\_cdocmor.html](https://tutorials.dkrz.de/tutorial_cdocmor.html) Accessed Sep 2026.
- 2 <https://www.clm-community.eu/> Accessed Sep 2026.
- 3 <https://www.dkrz.de/de/systeme/hpc> Accessed Sep 2026.
- 4 <https://www.hirs.de/solutions/systems/hpe-apollo-hawk> Accessed Sep 2026.
- 5 [https://www.dwd.de/DE/derdwd/it\\_functions/Teasergroup/datenverarbeitung.html](https://www.dwd.de/DE/derdwd/it_functions/Teasergroup/datenverarbeitung.html) Accessed Sep 2026.
- 6 <https://www.cscs.ch/computers/alps> Accessed Sep 2026.
- 7 [https://luv.dkrz.de/media/abstracts/abstract\\_1155.pdf](https://luv.dkrz.de/media/abstracts/abstract_1155.pdf) Accessed Sep 2026.
- 8 <https://www.fona.de/de/massnahmen/foerdermassnahmen/RegiKlim/nukleus.php> Accessed Sep 2026.
- 9 [https://www.dwd.de/EN/research/projects/udag/udag\\_node.html](https://www.dwd.de/EN/research/projects/udag/udag_node.html) Accessed Sep 2026.
- 10 <https://www.clm-community.eu/ICON-CLM/> Accessed Sep 2026.

## ACKNOWLEDGEMENTS

All active members of the CLM-Community and associated research partners should be acknowledged. The authors thank Burkhardt Rockel, Ivonne Anders, and Peter Berg for their work on developing the first versions of EvaSuite. We used DeepL and Copilot to improve the English. Note that these two tools did not add content. The authors acknowledge the DKRZ, HLRS, DWD, and CSCS computer centres for the use of resources in terms of computing time and storage capacity.

## FUNDING STATEMENT

This work was supported by funds from the German Federal Ministry of Research, Technology and Space through the project “Updating the data basis for adaptation to climate change in Germany (UDAG)” (grant no. 01LP2326D). The basic core of the EvaSuite was developed within the project openFRED (“Erstellung von Einspeisezeitreihen der fluktuierenden Erneuerbaren Energien auf Basis einer offenen Datenbank”, FKZ 0324006) of the Federal Ministry for Economic Affairs and Energy. This basic core was used to produce the first evaluation results.

## COMPETING INTERESTS

The authors have no competing interests to declare.

## AUTHOR CONTRIBUTIONS

**Evgenii Churiulin:** Wrote the manuscript, developed the software, contributed to the development of external packages, and developed the documentation.

**Beate Geyer:** Led the conceptualisation and coordination of the study, managed the data, prepared external datasets, developed external packages, contributed to the development of the documentation, and participated in writing the manuscript.

**Christoph Braun:** Developed the documentation and participated in writing the manuscript.

**May M. Bohmann:** Performed model verification and analysed the results.

**Eleonora Cusinato:** Performed model verification and analysed the results.

**Shweta Singh:** Prepared external datasets, analysed and discussed the results, and participated in writing the manuscript.

**Sergei Petrov:** Developed external packages and analysed the results.

**Philipp S. Sommer:** Served as a consultant.

**Andreas Will:** Performed model verification and analysed the results.

**Divyaja Lawand:** Performed model verification and analysed the results.

**Christian Beier:** Developed external packages and analysed the results.

**Muhammed Karadan:** Analysed the results.

**Martina Schubert-Frisius:** Developed external packages and analysed the results.

**Ronny Petrik:** Developed the software, contributed to the development of external packages, and developed the documentation.

## AUTHOR AFFILIATIONS

**Evgenii Churiulin**  [orcid.org/0000-0003-4527-2125](https://orcid.org/0000-0003-4527-2125)

Institute of Meteorology and Climate Research Troposphere Research (IMKTRO), Karlsruhe Institute of Technology, Karlsruhe, Germany

**Beate Geyer**  [orcid.org/0000-0001-8017-3136](https://orcid.org/0000-0001-8017-3136)

Institute of Coastal Systems, Helmholtz-Zentrum Hereon, Geesthacht, Germany

**Christoph Braun**  [orcid.org/0000-0002-8939-6501](https://orcid.org/0000-0002-8939-6501)

Institute of Meteorology and Climate Research Troposphere Research (IMKTRO), Karlsruhe Institute of Technology, Karlsruhe, Germany

**May M. Bohmann**  [orcid.org/0000-0002-9125-7922](https://orcid.org/0000-0002-9125-7922)

Institute of Meteorology and Climate Research Troposphere Research (IMKTRO), Karlsruhe Institute of Technology, Karlsruhe, Germany

**Eleonora Cusinato**  [orcid.org/0000-0002-5895-7169](https://orcid.org/0000-0002-5895-7169)

Institute of Meteorology and Climate Research Troposphere Research (IMKTRO), Karlsruhe Institute of Technology, Karlsruhe, Germany

**Shweta Singh**  [orcid.org/0000-0003-0752-1952](https://orcid.org/0000-0003-0752-1952)

Deutscher Wetterdienst, Offenbach am Main, Germany

**Sergei Petrov**  [orcid.org/0009-0006-8880-9390](https://orcid.org/0009-0006-8880-9390)

Institute of Coastal Systems, Helmholtz-Zentrum Hereon, Geesthacht, Germany

**Philipp S. Sommer**  [orcid.org/0000-0001-6171-7716](https://orcid.org/0000-0001-6171-7716)

Institute of Coastal Systems, Helmholtz-Zentrum Hereon, Geesthacht, Germany

**Andreas Will**  [orcid.org/0009-0009-6964-1766](https://orcid.org/0009-0009-6964-1766)

Fachgebiet Atmosphärische Prozesse, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus, Germany

**Divyaja Lawand**  [orcid.org/0009-0007-8442-0166](https://orcid.org/0009-0007-8442-0166)

Fachgebiet Atmosphärische Prozesse, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus, Germany

**Christian Beier**  [orcid.org/0009-0008-7267-2232](https://orcid.org/0009-0008-7267-2232)

Fachgebiet Atmosphärische Prozesse, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus, Germany

**Muhammed Karadan**  [orcid.org/0000-0002-6666-9028](https://orcid.org/0000-0002-6666-9028)

Fachgebiet Atmosphärische Prozesse, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus, Germany

**Martina Schubert-Frisius**  [orcid.org/0009-0000-1873-9389](https://orcid.org/0009-0000-1873-9389)

GERICS, Helmholtz-Zentrum Hereon, Geesthacht, Germany

**Ronny Petrik**  [orcid.org/0000-0003-1750-5534](https://orcid.org/0000-0003-1750-5534)

German Navy Headquarters, Directorate Ops, MGeo Division, Rostock, Germany

## REFERENCES

- Jacob D, Petersen J, Eggert B, Alias A, Christensen OB, Bouwer LM, et al. EURO-CORDEX: New high-resolution climate change projections for European impact research. *Reg Environ Change*. 2014;14:563–578. DOI: <https://doi.org/10.1007/s10113-013-0499-2>
- Righi M, Andela B, Eyring V, Lauer A, Predoi V, Schlund M, Vegas-Regidor J, et al. Earth System Model Evaluation Tool (ESMValTool) v2.0 – Technical overview. *Geosci Model Dev*. 2020;13(3):1179–1199. DOI: <https://doi.org/10.5194/gmd-13-1179-2020>
- Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci Model Dev*. 2016;9(5):1937–1958. DOI: <https://doi.org/10.5194/gmd-9-1937-2016>
- Brown B, Jensen T, Gotway JH, Bullock R, Gilleland E, Fowler T, et al. The Model Evaluation Tools (MET): More than a decade of community-supported forecast verification. *Bull Am Meteorol Soc*. 2021;102(4):E782–E807. DOI: <https://doi.org/10.1175/bams-d-19-0093.1>
- Harrold M, Prestopnik J, Opatz J, Halley Gotway J, Vigh J, Kalb C, et al. *The METplus version 6.2.1 user's guide*. Accessed Sep 2026. Available from: <https://github.com/dtcenter/METplus/releases>
- Kadow C, Illing S, Lucio-Eceiza EE, Bergemann M, Ramadoss M, Sommer PS, et al. Introduction to Freva – A free evaluation system framework for earth system modeling. *J Open Res Softw*. 2021;9(1):1–13. DOI: <https://doi.org/10.5334/jors.253>
- Lee H, Goodman A, McGibbney L, Waliser DE, Kim J, Loikith PC, et al. Regional climate model evaluation system powered by Apache Open Climate Workbench v1.3.0: An enabling tool for facilitating regional climate studies. *Geosci Model Dev*. 2018;11(11):4435–4449. DOI: <https://doi.org/10.5194/gmd-11-4435-2018>
- Leeuwenburg T, Loveday N, Ebert EE, Cook H, Khanarmuei M, Taggart RJ, et al. scores: A Python package for verifying and evaluating models and predictions with xarray. *J Open Source Softw*. 2024;9(99):6889. DOI: <https://doi.org/10.21105/joss.06889>
- Vo T, Po-Chedley S, Boutte J, Lee J, Zhang C. xCDAT: A Python package for simple and robust analysis of climate data. *J Open Source Softw*. 2024;9(98):6426. DOI: <https://doi.org/10.21105/joss.06426>
- Wikipedia Contributors. YAML [Internet]. Accessed Sep 2026. Available from: <https://en.wikipedia.org/wiki/YAML>
- Anders I, Brienens S, Demuzere M, Buchignani E, Ferrone A, Geyer B, et al. Evaluation report COSMO-CLM 5.0.
- Geyer B, Russo E, Petrik R, Keuler KG, Adinolfi M, Feldmann H, et al. A novel evaluation metric to determine an optimal regional climate model configuration applied for COSMO-CLM 6.0. *Earth and Space Science*. 2026. DOI: <https://doi.org/10.22541/essoar.15004388>
- Geyer B, Campanale A, Churiulin E, Feldmann H, Goergen K, Hagemann S, et al. Optimisation of ICON-CLM for the EURO-CORDEX domain: Developments, sensitivities, tuning. *Geosci Model Dev*. 2026;2026(12):5439–5490. DOI: <https://doi.org/10.5194/gmd-19-5439-2026>
- Petrov S, Will A, Geyer B. Linear meta-model optimization for regional climate models (LiMMo version 1.0). *Geosci Model Dev*. 2025;18(18):6177–6194. DOI: <https://doi.org/10.5194/gmd-18-6177-2025>
- Hassell D, Gregory J, Blower J, Lawrence BN, Taylor KE. A data model of the Climate and Forecast metadata conventions (CF-1.6) with a software implementation (cf-python v2.1). *Geosci Model Dev*. 2017;10(12):4619–4646. DOI: <https://doi.org/10.5194/gmd-10-4619-2017>
- Max Planck Institute for Meteorology. *Climate Data Operators (CDO), Version 2.5.4* [Internet]. Accessed Sep 2026. Available from: <https://code.zmaw.de/projects/cdo>
- Python Software Foundation. *Python* [Internet]. Accessed Sep 2026. Available from: <https://www.python.org/>
- Fox B, GNU Project. *Bash – Unix shell and command language* [Internet]. Accessed Sep 2026. Available from: <https://www.gnu.org/software/bash/>
- Python Software Foundation. *unittest — unit testing framework* [Internet]. Accessed Sep 2026. Available from: <https://docs.python.org/3/library/unittest.html>
- Python Packaging Authority. *pip – the Python package installer* [Internet]. Accessed Sep 2026. Available from: <https://pip.pypa.io/>
- R Core Team. *R: A language and environment for statistical computing* [Internet]. Accessed Sep 2026. Available from: <https://www.r-project.org/>
- Linux Kernel Community. *Linux kernel* [Internet]. Accessed Sep 2026. Available from: <https://www.kernel.org/>
- The NCO Project. *NCO – netCDF operators* [Internet]. Accessed Sep 2026. Available from: <https://nco.sourceforge.net/>
- Unidata. *NetCDF – Network common data form* [Internet]. Accessed Sep 2026. Available from: <https://www.unidata.ucar.edu/software/netcdf/>
- Yoo AB, Jette MA, Grondona M. SLURM: Simple Linux utility for resource management. In Feitelson DG, Rudolph L, & Schwiegelshohn U (Eds.) *Job scheduling strategies for parallel processing*. Berlin, Heidelberg: Springer; 2003. pp. 44–60. DOI: [https://doi.org/10.1007/10968987\\_3](https://doi.org/10.1007/10968987_3)
- EvaSuite Development Team. *EvaSuite documentation repository* [Internet]. Accessed Sep 2026. Available from: <https://clm-community.gitlab-pages.dkrz.de/public/evasuite/>
- Hamano J, Torvalds L, et al. *Git – Version control system* [Internet]. Accessed Sep 2026. Available from: <https://git-scm.com/>
- EvaSuite Development Team. *EvaSuite predefined spatial subregions (GeoPackage)* [Internet]. Accessed Sep 2026.

- Available from: [https://swift.dkrz.de/v1/dkrz\\_c0725fe8741c474b97f291aac57f268f/clmcom/spice/EvaSuite\\_allpolygons.v1.gpkg](https://swift.dkrz.de/v1/dkrz_c0725fe8741c474b97f291aac57f268f/clmcom/spice/EvaSuite_allpolygons.v1.gpkg)
29. **Open Geospatial Consortium.** *OGC GeoPackage encoding standard* [Internet]. Accessed Sep 2026. Available from: <https://www.geopackage.org/>
  30. **Climate Data Center.** *Daily solar observations – Germany (CDC)* [Internet]. Accessed Sep 2026. Available from: [https://opendata.dwd.de/climate\\_environment/CDC/observations\\_germany/climate/daily/solar/](https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/solar/)
  31. **Climate Data Center.** *Description of the DWD climate daily solar dataset* [Internet]. Accessed Sep 2026. Available from: [https://opendata.dwd.de/climate\\_environment/CDC/observations\\_germany/climate/daily/solar/DESCRIPTION\\_obsgermany-climate-daily-solar\\_en.pdf](https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/solar/DESCRIPTION_obsgermany-climate-daily-solar_en.pdf)
  32. **Sørland SL, Brogli R, Pothapakula PK, Russo E, Van de Walle J, Ahrens B,** et al. COSMO-CLM regional climate simulations in the Coordinated Regional Climate Downscaling Experiment (CORDEX) framework: A review. *Geosci Model Dev.* 2021;14(8):5125–5154. DOI: <https://doi.org/10.5194/gmd-14-5125-2021>
  33. **Cornes RC, van der Schrier G, van den Besselaar EJM, Jones PD.** An ensemble version of the E-OBS temperature and precipitation data sets. *J Geophys Res Atmos.* 2018;123(17):9391–9409. DOI: <https://doi.org/10.1029/2017JD028200>
  34. **Taylor KE.** Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos.* 2001;106(D7):7183–7192. DOI: <https://doi.org/10.1029/2000JD900719>
  35. **EvaSuite Development Team.** *EvaSuite documentation repository – Chapter installation* [Internet]. Accessed Sep 2026. Available from: [https://clm-community.gitlab-pages.dkrz.de/public/evasuite/05\\_installation/index.html](https://clm-community.gitlab-pages.dkrz.de/public/evasuite/05_installation/index.html)
  36. **EvaSuite Development Team.** *EvaSuite master repository* [Internet]. Accessed Sep 2026. Available from: <https://gitlab.dkrz.de/clm-community/public/evasuite>

---

#### TO CITE THIS ARTICLE:

Churiulin E, Geyer B, Braun C, Bohmann MM, Cusinato E, Singh S, Petrov S, Sommer PS, Will A, Lawand D, Beier C, Karadan M, Schubert-Frisius M, Petrik R. 2026 EvaSuite: A Framework for Model Evaluation and Verification Using Observational and Model Datasets. *Journal of Open Research Software*, 14: 66. DOI: <https://doi.org/10.5334/jors.729>

**Submitted:** 27 March 2026    **Accepted:** 15 October 2026    **Published:** 30 October 2026

#### COPYRIGHT:

© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <https://creativecommons.org/licenses/by/4.0/>.

*Journal of Open Research Software* is a peer-reviewed open access journal published by Ubiquity Press.