ECMWF
Copernicus
Procurement

Invitation to Tender

Copernicus Climate Change Service

Development of Machine Learning-based Methods for Downscaling the ERA5 Global Reanalysis to European and Arctic Regions

Volume II: Specification of Requirements

ITT Ref: C3S2_361b
ISSUED BY: ECMWF Administration Department Procurement Section
Date: 29 July 2024
Version: Final

Funded by the European Union

Implemented by ECMWF

Table of Contents

1 Introduction

Copernicus is the European Union's flagship Earth-observation programme created to achieve operational monitoring of the atmosphere, oceans, and continental surfaces. It aims to provide reliable, validated information services for a range of environmental and security applications. The Copernicus Climate Change Service (C3S, https://climate.copernicus.eu) responds to environmental and societal challenges associated with climate change. The service gives access to information for monitoring and predicting climate change and thus helps support adaptation and mitigation. C3S produces and brokers a wide range of data and products describing the past, present and future of the climate system. This includes global and regional reanalyses, Essential Climate Variables (ECVs), near-term climate predictions, climate projections and a variety of sectoral climate information. The data are offered to users through the C3S Climate Data Store (CDS, https://cds.climate.copernicus.eu).

2 Context

The Copernicus Climate Data Store (CDS) is the C3S infrastructure underpinning user access to its wide range of climate data. The CDS catalogue includes a large set of reanalysis datasets including global (ERA5) and regional (CERRA and CARRA) reanalyses.

ECMWF as the Entrusted Entity for the Copernicus Climate Change Service (C3S) invites Tenders for services related to machine learning-based downscaling of global reanalysis datasets with **the main goal of the quick timely updates of the European (CERRA) and Arctic (CARRA) regional reanalyses.**

Within its first phase (Cop1: 2015 – 2020), the Service consolidated many years of preparatory research and development to deliver a range of operational services. In its present phase (Cop2: 2021 – 2028) these services are further consolidated, improved and expanded to address all the existing and emerging requirements.

Regarding regional reanalyses, C3S supported the development of CERRA (Copernicus European Regional Reanalysis) and CARRA (Copernicus Arctic Regional Reanalysis) and their associated services including their Timely Updates (TU). The TU services are typically running by a latency of 2-4 months with respect to real time. This service started for CARRA in February 2023, while the CERRA TU service is planned to be deployed during 2024.

More information on the CERRA and CARRA regional reanalyses can be found at https://climate.copernicus.eu/copernicus-regional-reanalysis-europe-cerra and https://climate.copernicus.eu/copernicus-arctic-regional-reanalysis-service, respectively.

The provided datasets in the CDS are available at https://cds.climate.copernicus.eu/#!/search?text=cerra and https://cds.climate.copernicus.eu/#!/search?text=carra, respectively.

The procured work could take inspiration or even build upon the outcomes of the work performed in the ECMWF's 2023 Code4Earth challenge. Particularly, the results obtained in these challenges can be used to set the quality targets of the dataset to be provided by Machine Learning (ML) algorithms. Details are available at https://codeforearth.ecmwf.int/previous-editions (see particularly the "TesseRugged" and "DeepR: Deep reanalysis" challenges).

3 Contract Summary

The main objective of the present Invitation to Tender (ITT) is to add to the regional reanalyses datasets already provided by C3S with additional Machine Learning (ML)–based products, which can enhance the existing service elements of C3S on regional reanalysis. The production element of the outcome of the present work would be a Timely Update (TU) Service for the existing CARRA and CERRA reanalysis products (in the exact same form as it is in the present products) in a near real time fashion (5 days behind real time or so). This would include at least two basic variables (2m temperature and total precipitation amount), but possibly more variables in a consistent fashion. A desirable element of the work would be also the provision

of uncertainty estimates (ensemble) to the provided variables. This work would be complemented by an outlook how in the future dynamical and ML-based methods can be used (combined) to provide regional reanalyses.

In general, the regional reanalysis systems are required to generate high-resolution, physically consistent, gridded estimates of GCOS (Global Climate Observing System) Essential Climate Variables that will complement the information available from lower-resolution global reanalyses. The dataset generated by ML-based downscaling techniques for the provided variables should also comply with these requirements.

ECMWF shall take ownership of the software developed during the contract lifetime, as well as of all regional reanalysis products delivered by this contract. The contract might include support for computing services for training of the ML methodology. The software delivering the production data should be ported to ECMWF's computers since the TU delivery suites need to be run at ECMWF. Permanent archiving of the data products is included in the contract with preference of the use of ECMWF's MARS archiving system. The computing and storage requirements need to be clearly specified including what resources are needed in ECMWF's High Performance Computing Facility (HPCF) and in any external computer or cloud resources. Ideally, the ECMWF HPCF computing resource needs should be expressed in System Billing Units (SBU), if possible. Tenderers shall provide all information required to publish the provided reanalysis products via the CDS. This includes a dataset overview and adequate documentation such as hands-on user guides and detailed technical and scientific specifications.

4 Technical Specifications

Regional reanalysis data products are needed for C3S to provide consistent information on a variety of ECVs for the purpose of monitoring the state of the climate and to support development of climate-related information products for users, including those with economic interests in the given region.

Machine Learning based regional reanalysis data shall be produced using as input the existing reanalysis information provided by C3S. These are the ERA5 (ERA5-Land), CERRA and CARRA datasets, all available in the CDS. The methodology shall be based on Machine Learning techniques to provide regional reanalysis data based on training using the abovementioned datasets. When run operationally close to real time, the systems should take the ERA5 initial release called ERA5T (see details of ERA5T at https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation#heading-Dataupdatefrequency) as input, optionally using other static fields, and provide analyses and predictions for CERRA and CARRA domains. The output dataset shall be a consistent set of variables being at least the 2m temperature and total accumulated precipitation amount. These variables will be produced and made available in near to real time (typically 5 days behind real time soon after ERA5T is ready) with the aim of providing a quick snapshot of the CERRA and CARRA regional reanalyses. Consistent set of additional variables and also provision of uncertainty estimates need to be explored for the production. Emphasis should be placed on capturing the consistency between variables and across spatial and temporal length scales and reliable uncertainty information possibly in the form of ensembles. The evaluation methodology and the quality requirements of the new ML-based dataset shall be agreed prior to the training of the methodology and production of the data. The dataset shall be provided in the exact same format as it is done now for the existing CERRA and CARRA data. This means the same geographical domains, same set of analyses and forecasts, grid resolution, projection, data format (e.g. GRIB2), MARS archiving etc. The details of the present single level datasets offered in the CDS are available at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-cerra-single-levels for CERRA and at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-carra-single-levels for CARRA.

The new dataset will be tentatively called CERRA-TU-ML and CARRA-TU-ML (TU refers to Timely Updates and ML for Machine Learning), respectively. The timely update products will act as an initial guess of the CERRA/CARRA regional reanalysis prior to their dynamically based updates 2-4 months behind real time.

In parallel with respect to the production work a thorough evaluation methodology shall be developed, particularly comparing the ML-based products (CERRA-TU-ML, CARRA-TU-ML) to the dynamics-based existing datasets. In this aspect the results of the Code4Earth challenges (see references above) can be considered as reference when setting up the quality targets for the datasets produced by the ML algorithms. The outputs of these challenges will be provided to the Successful Tenderer and the minimum quality targets will be established considering these outputs as baseline. The details of such evaluation will be discussed and agreed with ECMWF in the early stages of the contract execution. The system shall also include quality assessment during production and post-production, e.g. also possibly based not only on comparison to the available dynamics-based systems, but also on comparison with alternative C3S ECV products derived from observations and/or global reanalyses. The strengths and weaknesses of this new dataset shall be documented as addition to the information already available for the users in the Copernicus Knowledge Base (CKB) on the two regional reanalyses services. The operational exploitation of the monthly updates needs to be ensured and the data shall be available for the C3S users through the Copernicus Climate Data Store (CDS).

The final data products shall:

- include estimates of multiple ECVs (at least 2m temperature and total precipitation amount) that are consistent in time and space;
- ideally include consistent and reliable estimates of uncertainty, for example via ensembles;
- accurately represent observed climate variability and change.

It is expected to have an outlook on how dynamics-based and ML-based products can be used together in the next generation regional reanalyses.

4.1 Specification of Work

The following work packages are recommended. This list should be considered as a guideline and not as a strict formal requirement. The cost of the administrative reporting (quarterly and annual reports, implementation plans etc.) shall be limited to a maximum of 10% of that of the total human resources.

4.1.1 Work package 1: Development of ML-based regional reanalysis downscaling system

This work package would deal with the design of the ML-based data-driven system to realise the near real-time production updates of the CARRA/CERRA products with comparable quality to the dynamical system already in use. This work package would also include the selection of variables, which can be provided in a dynamically consistent manner. This shall include at least 2m temperature and total precipitation amount, but other consistent set of variables needs to be also explored. Uncertainty estimation would be also a desirable feature. This work package will also deal with the optimal settings of the training methodology.

Indicative list of required deliverables:

- Description of the ML methodology design and practical installation for the CERRA-TU-ML and CARRA-TU-ML systems including the training details.
- Portable and documented code capable of training the operational versions of each of the delivered ML systems.
- Description of the choices of consistent set of variables and outlook for the creation of a full list of variables (with respect to the ones already available from the present CERRA/CARRA system).
- Description of the uncertainty estimation method and the details of the provided uncertainty information or an outlook on the perspectives of providing uncertainty information to the proposed variables.
- Outlook for the methodological perspectives to create full regional reanalysis datasets in the future based on combination of dynamical and ML techniques.

4.1.2 Work package 2: Evaluation methodology

Key element of the development work is to ensure that the output datasets meet the C3S quality requirements, i.e. they have comparable quality to the present CERRA and CARRA products and also the ones developed for the Code4Earth challenges. The minimum quality requirements will be established at the beginning of the work. A thorough evaluation methodology should be built to compare the ML-based CARRA/CERRA products to their dynamical counterparts. This needs to be discussed and agreed prior to any training and production work. Methodology should be proposed to measure accuracy, statistical calibration (if uncertainty estimates are provided), spatial and temporal consistency. This would also possibly include the use of other proxies for the reference ("truth") in order to be able to assess which products (i.e. ML-based or dynamics-based) provide better quality data. The quality assessment results need to be reflected in the user guidance materials (see Work package 4).

Indicative list of required deliverables:

- Minimum quantitative quality requirements for the ML-generated datasets.
- Detailed description of the evaluation methodology and underlying reference datasets used.
- Report on the accuracy of CERRA-TU-ML and CARRA-TU-ML using the agreed evaluation methodology.

4.1.3 Work package 3: Operational production of the dataset

This work package is responsible for realising the production aspects of the work particularly the near real-time element of the service (CERRA-TU-ML and CARRA-TU-ML). Scripts and procedures need to be ported to the ECMWF computer system and automatic procedures with daily production need to be established to create the new set of data and provide it to the CDS. This work will also include the data preparation and archiving (in ECMWF's MARS archiving system) and the interface of the data to the CDS. The training of the ML-methodology might be performed in external computers, if needed. This work package will also deal with the monitoring of the production in terms of technical and data quality performance.

Indicative list of required deliverables:

- Technical description of the CERRA-TU-ML and CARRA-TU-ML production suites including handover details for the ECMWF team.
- Description of the monitoring of the production suites in terms of technical and quality assurance details.

4.1.4 Work package 4: User guidance and support

It is essential that the C3S (CDS) users are well informed on the new type of data to be provided by this contract. For this we need detailed documentations how the data was computed and particularly what are the strengths and weaknesses of the products with respect to the dynamics-based present solution. This work would also include specific information related to the new CDS catalogue entries like overview, variable definition and documentation information (standard set of information used for every CDS entry).

Indicative list of required deliverables:

- User guidance documents (strengths and weaknesses) on CERRA-TU-ML and CARRA-TU-ML with special emphasis on compatibility of the new dataset with respect to the original dynamics-based ones.
- Overview and documentation information for the new CERRA-TU-ML and CARRA-TU-ML CDS catalogue entries.

4.1.5 Work package 0: Management and coordination

This work package includes overall responsibility for day-to-day service management and coordination.

The following management aspects shall be briefly described in the technical proposal:

Meetings:

- Kick-off meeting
- ECMWF will organise regular progress review meetings (by videoconference).
- ECMWF organises annual C3S General Assemblies. The Contractor (1-3 team members) is expected to attend these meetings and contribute to discussions related to the topic of this ITT.
- Tenderers can propose additional project internal meetings, as they deem needed, as part of their response.

Quality assurance and control: The timely delivery as well as final quality check of the deliverables shall be ensured by the prime contractor (in terms of content, use of ECMWF reporting templates for deliverables and reports (Microsoft Word), format, deliverable numbering and naming, spelling and typos…); all reports and deliverables in this project shall be submitted in English. Unless otherwise specified the specific contract Deliverables shall be made available to ECMWF in electronic format, via the relevant deliverable repository system.

Communication management (incl. external and internal communication). Any external communication activity must be agreed with the ECMWF Copernicus Communication team in advance. This includes, but not exhaustively, communication planning, branding and visual style, media outreach, website and social media activity, externally facing text and graphical content and events. Such agreed activity would also need to be evaluated and reported on once complete so that success measures and KPIs could be provided to the European Commission (cf. Clause 2.4.6 of the Framework Agreement).

Set of Key Performance Indicators (KPIs) suitable for monitoring various aspects of contract and service performance. The proposed KPIs shall be SMART (specific, measurable, actionable, realistic and time bound). All KPIs shall be regularly reviewed and updated together with ECMWF, during the contract. The Contractor shall report to ECMWF on these KPIs as part of the quarterly progress review meetings, as well as part of the Quarterly and Annual Implementation Reports. The template to be used by the Tenderers to describe the KPIs is included in Volume IIIB of the ITT "Template for Tenderers".

Risk Management: The proposal shall include a risk register that describes identified risks for each work package, along with a mitigation strategy for each of the identified risks. This mitigation strategy shall be composed by both preventive and corrective measures. The risk register shall be updated regularly by the Contractor, and any update (related to new risks, likelihood or impact) shall be reported during the progress review meeting, as well as part of the quarterly and annual implementation reports.

Resources planning and tracking using the appropriate tools.

Subcontractor management, including conflict resolution, e.g. the prime contractor is responsible for settling disagreements, although advice/approval from ECMWF may be sought on the subject. A list of subcontractors describing their contribution and key personnel shall be provided, as well as backup names for all key positions in the contract. Tenderers shall describe how the Framework Agreement; in particular Clause 2.9 on Sub-contracting has been flowed down to all their subcontractors.

Management of personal data and how this meets the requirements of Clause 2.8 on Personal Data Protection and Annex 6 of the Framework Agreement.

List of minimum deliverables and milestones required as part of WP0, covering the contractual and financial reporting obligations towards ECMWF in line with the Terms and Conditions of the Framework Agreement (cf. Clause 2.3 and Annex 5):

WP0 Deliverables		
Deliverable #	Title	Due
D361b.0.1.1-YYYY.QQ	Quarterly Implementation Report QQ YYYY (QQ YYYY being the previous quarter)	On 15/04, 15/07 and 15/10

D361b.0.1.2-YYYY	Annual Implementation Report Part 1 YYYY (YYYY being the Year n-1) This includes: 1) Quarterly implementation Report for the previous quarter Q4 YYYY 2) Preliminary financial form YYYY (YYYY being the Year n-1)	Annually on 15/01
D361b.0.1.3-YYYY	Annual Implementation Report Part 2 YYYY (YYYY being the Year n-1)	Annually on 28/02
D361b.0.1.4	Final implementation report	end of the contract
D361b.0.2.1-YYYY	Annual Implementation Plan YYYY (YYYY being the Year n+1)	Annually on 30/09
D361b.0.3.1-YYYY	Copy of prime contractor's general financial statements and audit report YYYY (YYYY being the Year n-1)	Annually (no-cost associated)

WP0 Milestones			
Milestone #	Title	Means of verification	Due
M361b.0.1.1.QX	Progress review meetings with ECMWF	Minutes of meeting	Quarterly
M361b.0.1.2	Kick-off meeting	Minutes of meeting	By M1

5 General requirements

5.1 Implementation Schedule

ECMWF intends to award a single Framework Agreement for a period of 24 months, which shall be implemented via a single Service Contract expected to commence in Q1 2025. The Tenderer shall provide a detailed implementation plan of the proposed activities.

5.2 Deliverables and Milestones

Deliverables should be consistent with the technical requirements specified in section 4. A deliverable is a substantial, tangible or intangible good or service produced as a result of the contract. In other words, a deliverable is an outcome produced in response to the specific objectives of the contract. Deliverables are subject to acceptance by the technical and contract management officers at ECMWF. (for deliverables quality assurance and control requirements, please see also section 4.1.5)

Each Deliverable shall be listed in the Deliverable List tab of Volume IIIA with an associated resource allocation (in person-months) and relevant price. The total of these allocated resources shall amount to the requested budget associated with payroll (in cost and prices tab of Volume IIIA).

Milestones should be designed as markers of demonstrable progress in service development and/or quality of service delivery, as applicable. They should not duplicate deliverables, neither have associated resource allocation, contrary to the deliverables.

Tenderers shall complete the relevant table in Volume IIIA as part of their bid, including there the details, schedule and concrete delivery/completion due date of all deliverables and milestones under each work package.

5.3 Communication

The Successful Tenderer shall support ECMWF in its communication activities for the C3S services, where they are related to the activities described in this ITT. Additional activities such as C3S website news items, C3S brochures and flyers, may be discussed on a case-by-case basis during the contract implementation. For communication management requirements, please see also section 4.1.5.

5.4 Data and IPR

It is a condition of EU funding for Copernicus that ownership of any datasets/software developed with Copernicus funding passes from the suppliers to the European Union via ECMWF. Ownership will pass from

the date of creation of the datasets/software. Suppliers will be granted a non-exclusive licence to use the datasets/software which they have provided to Copernicus for any purpose.

All software and products used by the Successful Tenderer to produce the Copernicus datasets/software will remain the property of the Successful Tenderer, except for those components which are acquired or created specifically for Copernicus purposes, with Copernicus funding, and which are separable and useable in isolation from the rest of the Successful Tenderers' production system. The identity and ownership of such exceptional components will be passed to the European Union annually. The Successful Tenderer will be granted a non-exclusive licence to use them for any purpose.

5.5 Payment Plan

Tenderers can propose a Payment Plan in ITT Volume IIIA "Pricing and deliverables" (cf. Excel spreadsheet "Payment Plan preparation"):

The Payment Milestones should relate to the deliverables and milestones delivered during the corresponding Payment Milestone period (e.g. the payment covering the period January-June would only relate to the deliverables and milestones whose due dates are part of the same period). The recommended frequency of payments is on semestrial basis (i.e. two payments per year).

In case of request for a payment at contract signature, please note that this should be duly substantiated (e.g. in terms of necessary investment prior to implementation or during first weeks/months for ensuring the initial set up of the project). It is necessary to relate this payment to activities subject to other Payment Milestones.

6 Tender Format and Content

General guidelines for the Tender are described in Volume IIIB of this ITT. This section describes specific requirements to prepare the proposal for this particular Tender, along with guidelines for minimum content expected to be included in the proposal, additional to the content described in the general guidelines of Volume IIIB. This is not an exhaustive description and additional information may be necessary depending on the Tenderer's response.

6.1 Page Limits

As a guideline, it is expected that individual sections of the Tenderer's response do not exceed the page limits listed below. These are advisory limits and should be followed wherever possible, to avoid excessive or wordy responses.

Section	*Page Limit*
Executive Summary	2
Track Record	2 (for general) and 2 (per entity)
Quality of resources to be Deployed	2 (excluding Table 1 in Volume IIIB and CVs with a maximum length of 2 pages each)
Technical Solution Proposed	2 + 3 per Work package (Table 2 in Volume IIIB, the section on references, publications, patents and any pre-existing IPR is excluded from the page limit and has no page limit)
Management and Implementation	6 (excluding Table 4 and Table 5 in Volume IIIB) + 2 per each Work package description (Table 3 in Volume IIIB)
Pricing Table	No limitation

Table 1: *Page limits*

6.2 Specific additional instructions for the Tenderer's response

The following is a guide to the minimum content expected to be included in each section, additional to the content described in the general guidelines of Volume IIIB. This is not an exhaustive description and additional information may be necessary depending on the Tenderer's response.

6.2.1 Executive Summary

The Tenderer shall provide an executive summary of the proposal, describing the objectives, team and service level.

6.2.2 Track Record

The Tenderer shall demonstrate for itself and for any proposed subcontractors that they have experience with relevant projects in the public or private sector at national or international level. ECMWF may ask for evidence of performance in the form of certificates issued or countersigned by the competent authority.

6.2.3 Quality of Resources to be deployed

The Tenderer shall propose a team that meets at least the following requirements:

- A senior team member with more than 5 years of experience in managing activities related to this ITT (referred to as Principal Investigator). This person will be the point of contact on technical matters.
- A team member with experience of managing projects and contracts of this type and size (referred to as Service Manager). This person will be the main point of contact for administrative matters.
- Team members with demonstrated experience in performing activities related to the various aspects of this ITT.

These team members shall be involved in the activities of this ITT at a minimum level of 10% of their total working time.

6.2.4 Technical Solution Proposed

The Tenderer is expected to provide a short background to the proposed technical solution, to facilitate understanding of the solution proposed, as well as a clear and detailed description of the proposed technical solution and its organisation into work packages.

Estimates as to the compute resources required both to train the final ML model and perform daily reanalyses and predictions to form the operational system should be provided.

6.2.5 Management and Implementation Plan

As part of the general project management description, and in addition to the guidance provided in Volume IIIB, Tenderers shall consider the elements described in section 4.1.5 above. Note that costs associated with fulfilling WP0 requirements shall not exceed 10% of the total price of the Tender.

Furthermore, should any sub-contractors be proposed in the Tender, in order to ensure a comprehensive and realistic proposal, it is desirable for the Tenderer to actively involve all such sub-contractors in the development of the proposal. This involvement should include, but is not limited to, collaborative planning, clear communication of project timelines, and agreement on deliverables and deadlines. The Tenderer must provide documented evidence of this collaboration, demonstrating that each sub-contractor has been consulted and has agreed to their respective roles, responsibilities, and deadlines as outlined in the proposal. This requirement is instituted to promote a cohesive and feasible project plan, reflecting a true and committed partnership among all participating entities.