

Allegato 2 – Aree tematiche per la presentazione dei Progetti

Il presente Allegato descrive le aree tematiche definite per i Bandi a Cascata dello Spoke 4 del Centro Nazionale ICSC.

Spoke 4 - Centro Euro-Mediterraneo sui Cambiamenti Climatici CMCC

N. area	Area tematica	Dimensione minima (k€)	Dimensione massima (k€)
1	Attività di supporto alla ottimizzazione e parallelizzazione dei codici numerici e al loro porting sulle nuove architetture HPC messe a disposizione dal CN.	100.000 €	400.000€
2	Aumento della exploitation dei dati e creazione una più forte connessione dei servizi sviluppati dallo Spoke con l'infrastruttura cloud dell'hub	100.000 €	400.000€
3	Sviluppo e implementazione di servizi e realizzazione di prodotti, principalmente finalizzati a migliorare l'utilizzo dei risultati dello Spoke 4 nell'ambito delle attività di adattamento ai cambiamenti climatici.	100.000 €	400.000€
4	Miglioramento della capacità delle catene modellistiche utilizzate nell'ambito dello Spoke 4 al fine di produrre dati e informazioni sugli impatti in settori socio-economici (es., agro-alimentare, risorse idriche, produzione e domanda di energia, salute) di rilevanza per il sistema italiano.	100.000 €	400.000€

Dotazione finanziaria totale: 3.200.000 €

Minima dotazione finanziaria destinata a soggetti con sedi operative nelle regioni in "Quota Sud": 1.600.000 €



2. Descrizione scientifica delle Tematiche

Tematica 1- Attività di supporto alla ottimizzazione e parallelizzazione dei codici numerici e al loro porting sulle nuove architetture HPC messe a disposizione dal CN.

Dotazione Finanziaria: 800.000€

L'attività si deve concentrare sulla parallelizzazione e ottimizzazione di codici software di simulazione del sistema Terra e delle sue componenti. L'obiettivo generale è quello di rendere i codici più efficienti aumentando le loro performance computazionali e trasferirli sulle nuove architetture accelerate (GPU)

I progetti dovranno realizzare almeno uno dei seguenti output:

1. Porting e ottimizzazione di modelli climatici e componenti di modelli climatici riportati nell'allegato tecnico A sulla partizione GPU di Leonardo.
2. Analisi delle prestazioni, analisi di scalabilità, profiling e identificazione dei colli di bottiglia a livello di singole routine e moduli dei modelli climatici di interesse dello Spoke 4 (**vedi Appendice Tecnica**) individuando i limiti prestazionali se dovuti (i) all'inefficiente uso della gerarchia di memoria; (ii) al poco utilizzo delle capacità di calcolo (scarso livello di vettorizzazione, etc.); (iii) ad un uso inefficiente del sottosistema I/O; (iv) inefficienze inerenti la parallelizzazione di codici già paralleli (eccessivo overhead parallelo, load unbalancing, etc.). Individuare inoltre possibili ottimizzazioni da applicare per migliorare le prestazioni e dimostrare l'efficacia delle trasformazioni proposte valutando il miglioramento prestazionale ottenibile su un kernel estratto dal modello in esame.
3. Modularizzazione, ottimizzazione, porting (incluso implementazione di user-oriented interface) del codice delle parametrizzazioni in sviluppo (**vedi Appendice Tecnica**), per l'utilizzo su infrastrutture CPU e ibride (CPU/GPU); il lavoro di ottimizzazione e porting è anche esteso a tutti i tool necessari per la creazione delle condizioni al contorno eventualmente richieste dalle suddette parametrizzazioni.
4. Ottimizzazione, parallelizzazione e modularizzazione di librerie, possibilmente standalone e comunque MPI-parallele, per specifici tasks di data assimilation come horizontal covariance modelling (2D-3D parallel Gaussian filtering) o ensemble generation/sampling che possano essere usate da diversi schemi paralleli di assimilazione e che se computazionalmente conveniente prevedano porting su infrastrutture di tipo GPU.
5. Aggiornamento del software preposto alla gestione della comunicazione tra le varie componenti del sistema modellistico ("coupler") descritto nell'**Appendice Tecnica**. Profiling dei codici numerici delle varie componenti modellistiche con particolare attenzione alla parte di I/O per ottimizzare le performance eliminando eventuali colli di bottiglia.

6. Realizzazione di una libreria di modelli basata su tecnologie container ampiamente diffuse (i.e. Docker, Singularity) da integrare all'interno degli ambienti (cloud ed HPC) messi a disposizione da parte del Centro Nazionale ICSC e quelli delle infrastrutture HPC dei partner dello Spoke 4. Per la lista esaustiva dei modelli da "containerizzare" si faccia riferimento all'**Appendice Tecnica**.

Tematica 2 - Aumento dell'exploitation dei dati e creazione di una più forte connessione dei servizi sviluppati dallo Spoke con l'infrastruttura cloud dell'hub.

Dotazione Finanziaria: 800.000€

L'attività è finalizzata a realizzare nuovi strumenti digitali che migliorino l'accesso e lo sfruttamento diretto dei dati prodotti dallo Spoke 4 e delle loro elaborazioni da parte di una più vasta platea di utenti. L'obiettivo generale è quello di aumentare i possibili utilizzi dei dati prodotti dallo Spoke e la creazione di nuovi prodotti che rispondano ad esigenze nuove o emergenti.

I progetti dovranno realizzare almeno uno dei seguenti output/prodotti:

1. Realizzazione di un "Data Node" presso CINECA per lo Spoke 4 per la ricerca, l'accesso e la visualizzazione dei dati. I progetti proposti dovranno contenere:
 - 1.1. Sviluppo, setup e mantenimento operativo di uno "Spoke 4 Data Node" presso CINECA che ospiti e permetta l'accesso a tutti i dati prodotti dai vari partner dello Spoke 4. I datasets dovranno essere disponibili tramite servizi di comunità (e.g. OPeNDAP, THREDDS); laddove applicabile, i datasets dovranno essere offerti anche tramite interfaccia ESGF, che prevede il set up di uno specifico software stack ("ESGF Data Node"). I progetti dovranno inoltre, garantire l'integrazione dello "Spoke 4 Data Node" con ambiente cloud messo a disposizione da parte del Centro Nazionale ICSC al fine di permettere uno sfruttamento efficace ed efficiente dei dati anche a livello di calcolo da parte dei partner dello Spoke 4. La proposta dovrà prevedere un processo di sviluppo completo, ivi incluse fasi di raccolta di requisiti, feedback e validazione con i membri dello Spoke 4 e dei fornitori dell'infrastruttura.
 - 1.2. Progettazione, implementazione e mantenimento operativo di un'applicazione web e della corrispettiva versione mobile per la ricerca, l'accesso e la visualizzazione dei dati presenti all'interno dello "Spoke 4 Data Node" (riportato al punto 1.1) che ne massimizzi la disseminazione e lo sfruttamento. La proposta dovrà prevedere un processo di sviluppo completo, ivi incluse fasi di raccolta di requisiti, feedback e validazione con i membri dello Spoke 4 e dei fornitori dell'infrastruttura.

- 1.3. Progettazione, implementazione e mantenimento operativo di un ambiente di Interactive Computing HPC (su infrastruttura messa a disposizione da parte del Centro Nazionale ICSC) basato su Jupyter Notebook e co-locato con lo “Spoke 4 Data Node” (riportato al punto 1.1) che supporti gli utenti nello sviluppo di applicazioni per l’analisi e la visualizzazione dei dati. La proposta dovrà prevedere un processo di sviluppo completo, ivi incluse fasi di raccolta di requisiti, feedback e validazione con i membri dello Spoke 4 e dei fornitori dell’infrastruttura.

Tematica 3 - Sviluppo e implementazione di servizi e realizzazione di prodotti, principalmente finalizzati a migliorare l'utilizzo dei risultati dello Spoke 4 nell'ambito delle attività di adattamento ai cambiamenti climatici.

Dotazione Finanziaria: 800.000€

L’attività è finalizzata a migliorare l’utilizzabilità dei dati creati dallo Spoke per pianificare e implementare attività di adattamento, sia attraverso il miglioramento delle metodologie di utilizzo dati sia attraverso lo sviluppo di nuovi strumenti digitali per la pianificazione e implementazione delle pratiche di adattamento ai cambiamenti climatici.

I progetti dovranno realizzare almeno uno dei seguenti prodotti:

1. Sviluppo di un sistema di downscaling statistico di ensemble forecasts sulla penisola italiana e nei mari circostanti. Il prodotto finale di questo progetto consisterà nello sviluppo e implementazione di algoritmi e codici software (preferibilmente in linguaggio Python), che sulla base di un ensemble forecast, quale per es. la previsione stagionale C3S, dovranno essere in grado ottimizzare la performance dell’ensemble forecast al fine di produrre previsioni ad alta risoluzione (almeno ~ 25 km, preferibilmente ~10 km o meno) di variabili meteorologiche di largo interesse per applicazioni e studi di impatto, quali, per es. temperatura a 2 metri, precipitazione, venti a 10 metri, ecc. ecc.
2. Produzione di un’analisi dei rischi per le infrastrutture determinati dai cambiamenti nei parametri meteo-climatici sul territorio italiano o porzioni di esso (per esempio, regioni). Il prodotto finale di questo progetto consisterà in un database, visualizzabile ed esplorabile per mezzo di una piattaforma web, rappresentante il rischio al quale infrastrutture strategiche (per esempio, autostrade e principali arterie viarie, reti ferroviarie, centrali di produzione e distribuzione dell’energia, impianti industriali, ecc. ecc...) saranno esposte in conseguenza ai cambiamenti climatici, sul territorio italiano o per specifiche regioni. L’analisi del rischio sarà basata su dati meteo-climatici ad altissima risoluzione derivati dai dati ottenuti da proiezioni di cambiamento climatico prodotte o messe a disposizione dallo Spoke 4.



3. Produzione di un'analisi finalizzata alla valutazione degli impatti dei cambiamenti climatici in termini di "land suitability". Il prodotto finale consisterà in un database, visualizzabile ed esplorabile per mezzo di una piattaforma web, rappresentante le variazioni di "land suitability" cui saranno esposte produzioni colturali di rilevanza al sistema agro-alimentare italiano (es. ulivi, viticoltura, ecc.), in conseguenza ai cambiamenti climatici, sul territorio italiano o per specifiche regioni. L'analisi dovrà essere basata su dati meteo-climatici ad altissima risoluzione derivati dai dati ottenuti da proiezioni di cambiamento climatico prodotte o messe a disposizione dallo Spoke 4.

Tematica 4 - Miglioramento della capacità delle catene modellistiche utilizzate nell'ambito dello Spoke 4 di produrre dati e informazioni sugli impatti in settori socio-economici (es., agro-alimentare, risorse idriche, produzione e domanda di energia, salute) di rilevanza per il sistema italiano.

Dotazione Finanziaria: 800.000€

L'attività si dovrà focalizzare sullo sviluppo di catene modellistiche finalizzate alla diretta quantificazione degli impatti della variabilità e dei cambiamenti climatici in settori di grande rilevanza per il sistema socio-economico italiano. A titolo di esempio: sviluppo di sistemi di accoppiamento di modelli di resa colturale ("crop models"), modelli idrologici, ecc., da accoppiare ai modelli climatici utilizzati nello Spoke 4, che consentano una migliorata produzione di dati e informazioni sulle previsioni o le proiezioni climatiche di resa agricola, di condizioni idrologiche, ecc., sulla penisola italiana o altre aree e regioni di interesse.

Il progetto dovrà raggiungere almeno uno dei seguenti obiettivi specifici/prodotti:

1. Accoppiamento di modelli numerici di produzione colturale, per esempio modello di produzione di grano, con modelli climatici per la realizzazione di previsioni o proiezioni di produzione della coltura sulla penisola italiana o su un'area di produzione della coltura considerata. Il modello di produzione colturale dovrà essere implementato in una catena modellistica in grado di produrre informazioni e dati sulla produzione della coltura considerata a diverse scale temporali nel futuro, per esempio sub-stagionali, stagionali, decennali o multi-decennali. Di particolare interesse saranno considerate le proposte che prevedono catene modellistiche che riguardano produzioni agricole di interesse e rilevanza per la filiera agro-alimentare italiana, quali grano, vino, olio, caffè, ecc.
2. Accoppiamento di modelli numerici utilizzati per rappresentare specie marine di interesse commerciale con modelli fisici e biogeochimici per la realizzazione di proiezioni di produzione in funzione di scenari climatici e di pesca. Poiché esistono diversi modelli a supporto della gestione della pesca (da singola specie -stock assessment a multispecie e modelli di ecosistema marino) è possibile immaginare e



proporre diverse forme di catena modellistica, in grado di produrre informazioni e simulazioni utili a valutare i trade-off inevitabili della gestione integrata. I processi e le aree di analisi potranno essere analizzati a scale sia spaziali che temporali diverse. Di particolare interesse saranno considerate le proposte che prevedono catene modellistiche che riguardano produzioni ittiche di interesse e rilevanza per la filiera alimentare italiana.

3. Accoppiamento di modelli idrologici con modelli climatici per la realizzazione di previsioni o proiezioni di disponibilità della risorsa idrica e relative condizioni potenzialmente favorevoli al verificarsi di siccità o alluvioni, a diverse scale temporali nel futuro, per esempio sub-stagionali, stagionali, decennali o multi-decennali e per diversi bacini idrografici. Di particolare interesse saranno considerate le proposte che prevedono catene modellistiche che riguardano la disponibilità di risorsa idrica italiana nei principali bacini italiani.

Appendice tecnica per le specifiche dei modelli e dei codici numerici di parametrizzazione per i quali effettuare ottimizzazioni e porting.

Specifiche tecniche dei codici (relative al porting e alle ottimizzazioni relative ai progetti dell'Area Tematica 1)

Nome del modello o del codice	linguaggio utilizzato (per es.: fortran, C, ...)	Stima sul numero complessivo di righe di codice presenti	Schemi numerici utilizzati (per es.: elementi finiti, volumi finiti, differenze finite, spettrale ...)	Schema numerico di integrazione temporale (per es.: leap frog, runge kutte, implicito, esplicito...)	Librerie o tool a corredo del modello (per es.: PETSc, XIOS, NETCDF, ...)	Architecture sulle quali il software è già in funzione (per es.: Intel, AMD, PowerPC ...)	Eventuali parallelizzazione già esistenti (per es.: MPI, OpenMP, ...)	Note (per es., richieste di ottimizzazione di tutto il codice o solo routine specifiche).
CMCC-CM3	Fortran, Python, xml	2.700.000 (CIME 69k, CAM 1500k, CLM 500k, NEMO 230k, HYDROS 12k, CICE 210k, CMEPS 58k, CDEPS 39k)	Volumi Finiti	Esplicito (CAM), leap frog (NEMO parte non diffusiva), implicito (NEMO vertical diffusion)	NetCDF, XIOS, ESMF,PIO, Parallel NetCDF,	Intel	IMPI (Intel MPI)	Si richiede l'ottimizzazione della componente atmosferica CAM).
ocean/sea ice (unstructured grid)	Fortran, Python, xml		elementi finiti	semi-implicito	perl, metis, NetCDF, Parallel NetCDF, XIOS	Intel	Intel MPI	possibile ottimizzazione della componente

								dinamica del ghiaccio
ENEA-REG (WRF+MI TGCM+CaMa-Flood+RegESM)	Fortran	3000000	Differenze finite	Runge kutta+implicit integration for acoustic waves	ESMF+NUOP, NETCDF, Parallel-IO, Adios2	-CRESCO6 (Intel Xeon Platinum 8160 2.10GHz) -Leonardo (Intel Xeon 8358 2.6 GHz)	MPI, OpenMP	Si richiede l'ottimizzazione del coupler per velocizzare l'I/O della componente atmosferica (WRF)
SHYFEMCM	Fortran	300000	elementi finiti	semi-implicito	Metis, PETSc, NetCDF, XIOS	Intel	MPI, OpenMP	Ottimizzazione e Porting del modello su GPU
Globo	Fortran	60000 (ecRad 40k)	volumi finiti	split-explicit	netcdf, eccodes, openmpi	AMD, Intel	MPI	Ottimizzazione e porting su GPU di tutto il codice.
3DVAR	Fortran90	110000	assimilazione variazionale	non applicabile	NetCDF, ARPACK, LAPACK	Intel, AMD	MPI+OpenMP	Ottimizzazione, parallelizzazione e porting su GPU di tutto il codice

Informazioni di accesso ai codici e ai modelli

Nome del modello o del codice	link al repository del codice software	link ad un repository con la configurazione e i file ausiliari necessari all'esecuzione di un caso di test
CMCC-CM3	https://github.com/CMCC-Foundation/CMCC-CM	https://esmci.github.io/cime/versions/master/html/index.html (Guide) https://svn-ccsm-inputdata.cgd.ucar.edu/trunk/inputdata/ (data)
ENEA-REG	WRF: https://zenodo.org/records/4392230 MITGCM: https://github.com/uturuncoglu/MITgcm CaMa-Flood: https://github.com/global-hydrodynamics/CaMa-Flood_v4 Coupler (RegESM): https://github.com/uturuncoglu/RegESM	http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_v4/contents.html http://www2.mmm.ucar.edu/wrf/users/download/get_source.html
SHYFEMCM	https://github.com/shyfemcm/shyfemcm	
Globo	https://zenodo.org/records/8087495	Un manuale d'uso è disponibile insieme al codice del modello; istruzioni più dettagliate e file ausiliari necessari alla esecuzione di un caso di test saranno disponibili, in fase di sviluppo, tramite accesso alla piattaforma GitLab interna ad ISAC.