



# Catene modellistiche diagnostiche e catene prognostiche

Il ruolo della modellistica numerica nella valutazione e nella previsione dell'inquinamento atmosferico a scala regionale e sub-regionale

Palmanova, 27 febbraio 2013

ARPA FVG – CRMA  
Centro Regionale di Modellistica Ambientale  
[crma@arpa.fvg.it](mailto:crma@arpa.fvg.it)



## Sommario della lezione

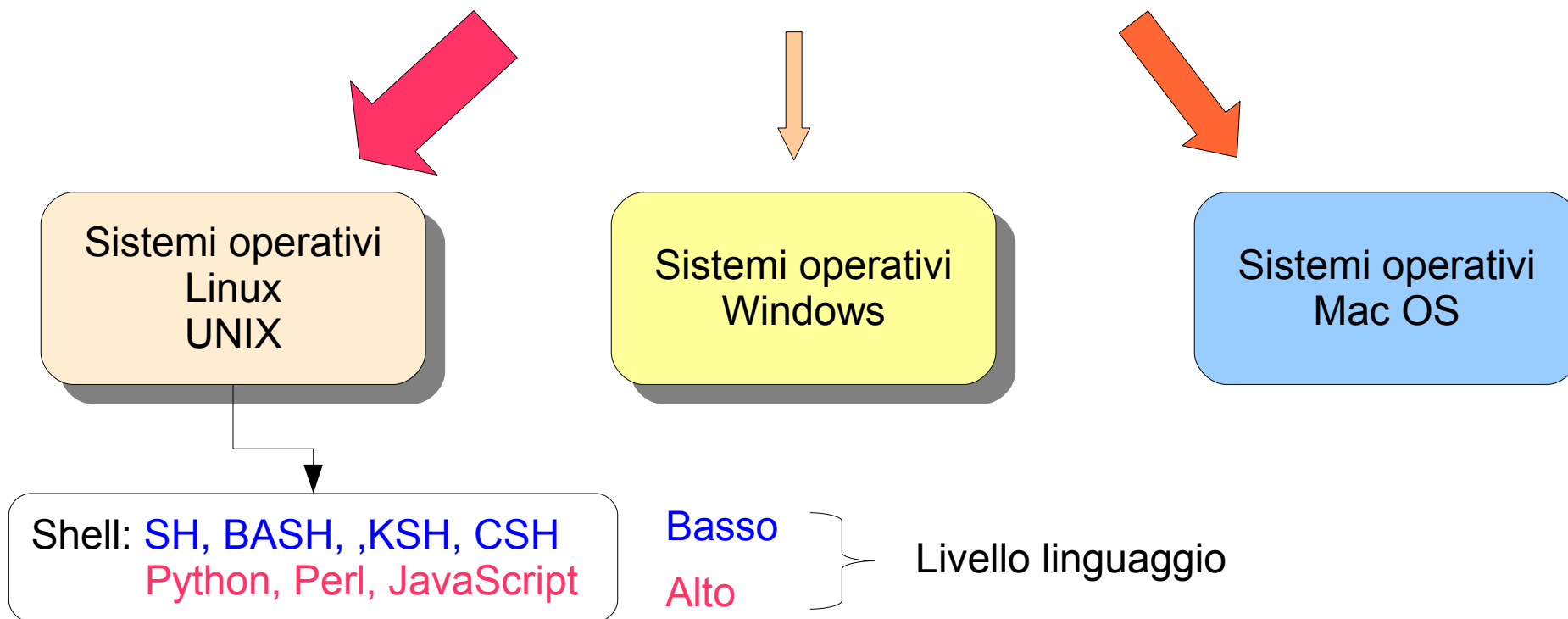


- Requisiti indispensabili per fare modellistica numerica seriamente
- I principali workflow della modellistica della qualità dell'aria
- Aspetti critici della modellistica operativa
- Esempi di catene operative meteorologiche e di qualità dell'aria
- Risorse necessarie ad un centro modellistico per la qualità dell'aria

## Requisiti indispensabili: il sistema operativo e il linguaggio

La modellistica numerica per l'ambiente richiede:

- L'installazione di codici in ambiente High Performance Computing (HPC)
- L'ottimizzazione delle performance dei codici installati
- La verifica e la validazione dei codici installati
- L'esecuzione di operazioni ripetitive ed automatiche
- L'elaborazione di grossi volumi di dati per la sintesi delle informazioni





## Requisiti indispensabili: programmazione e compilazione

### Linguaggio di programmazione per la scrittura e la modifica di codici per il calcolo scientifico

- FORTRAN 90 (preferito)
- C (opzionale)

### Compilatore per la realizzazione degli eseguibili

Compilatori liberi (costo ~ 0 EUR)

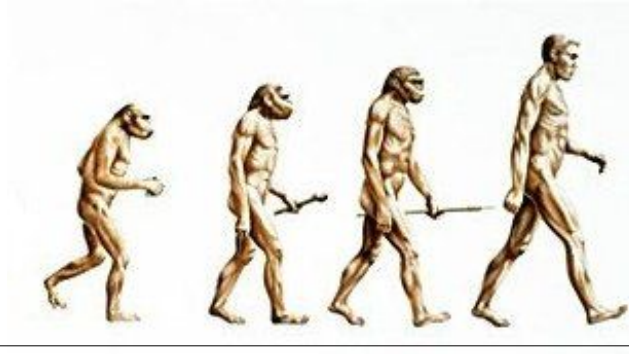
- GCC (GNU Compiler Collection)

Compilatori a pagamento (costo ~ k EUR)

- PGI® Optimizing Fortran, C and C++ Compilers & Tools (Portland Group)
- Intel® Compilers compiler, C++ compiler, and Fortran (Intel Group)
- Cray Fortran Compiler (CRAY company)

## Abilità nella programmazione per la scrittura e la modifica di codici dedicati al calcolo scientifico:

- da Australopiteco a Homo habilis (1 mese di apprendimento intensivo)
- da Homo habilis a Homo erectus (6 mesi di applicazione su un problema specifico)
- da Homo erectus a Homo sapiens (alcuni anni di applicazione e rivisitazione)



Calcolo di base  
Foglio di calcolo

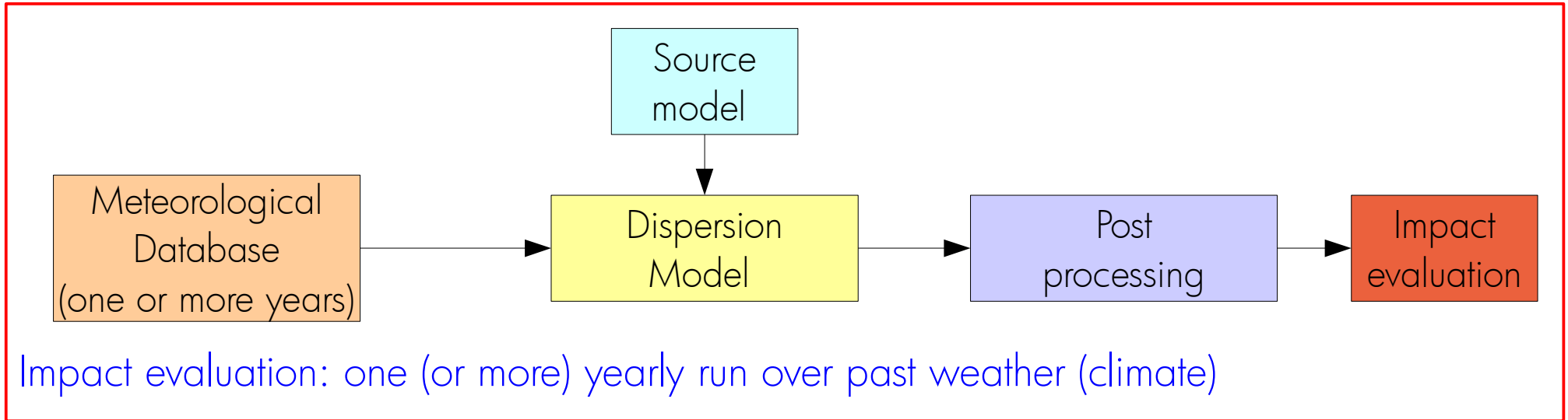
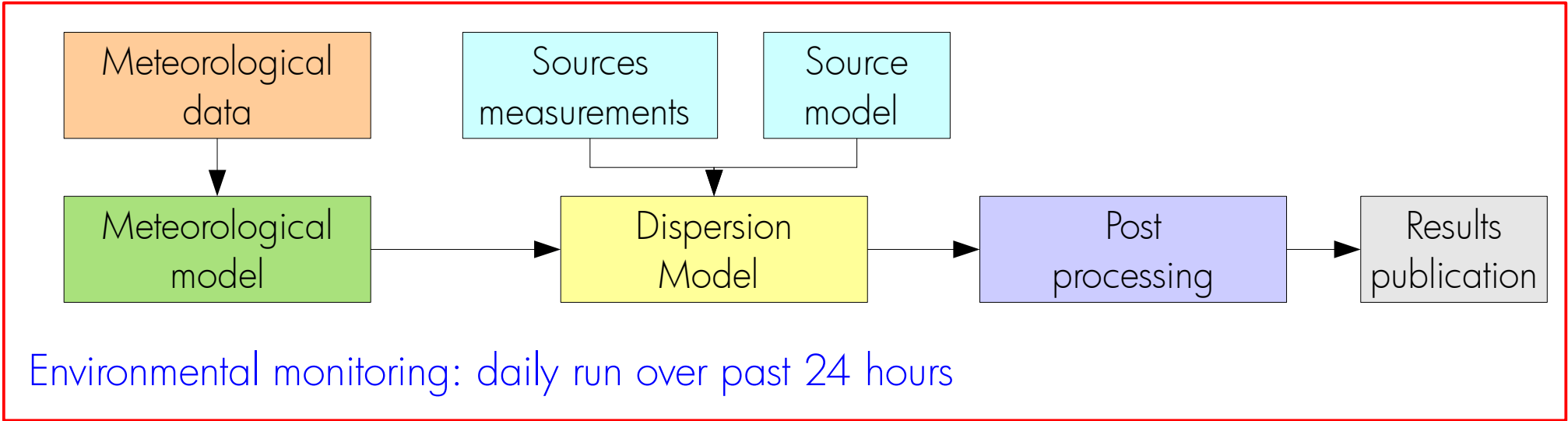
Calcolo scientifico  
High Performance Computing

## Abilità nell'uso del compilatore per la realizzazione degli eseguibili, uso degli accessori per il debug e la profilazione dei codici:

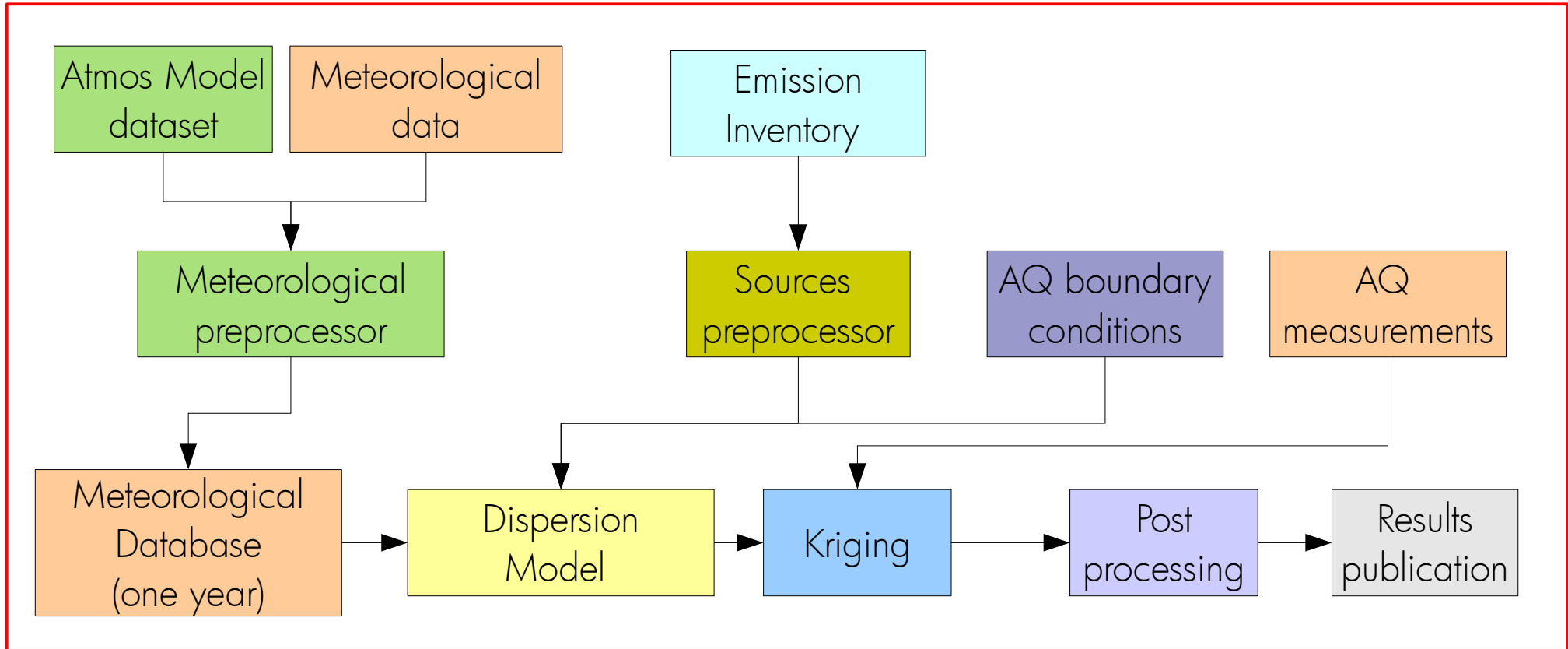
- da Australopiteco a Homo habilis (1 ora di apprendimento)
- da Homo habilis a Homo erectus (1 settimana di applicazione su un problema)
- da Homo erectus a Homo sapiens (alcuni anni di applicazione e rivisitazione)

# Catene modellistiche: i principali workflow

## Le simulazioni di tipo Off-line

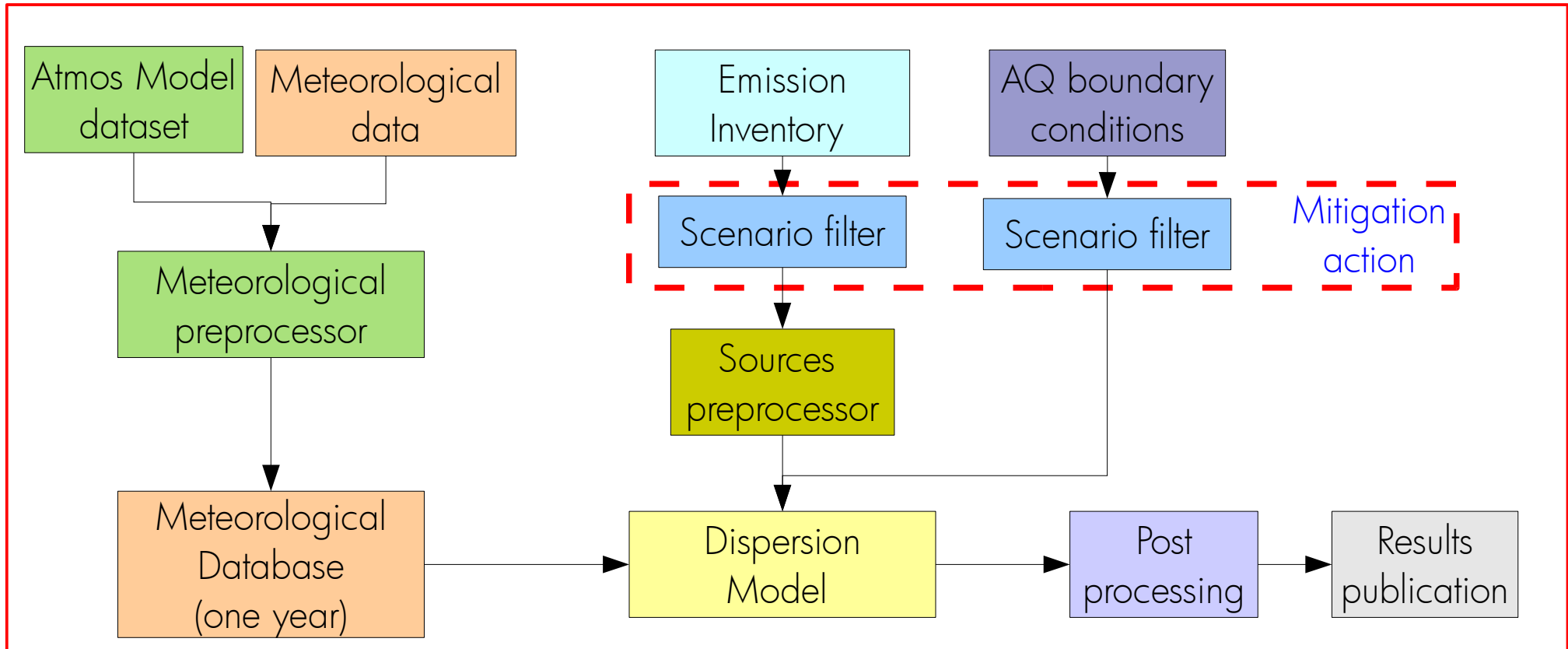


## Workflow usato al CRMA nell'ambito del progetto ADE – run di riferimento QA





## Workflow usato al CRMA nell'ambito del progetto ADE – scenari



## Aspetti critici delle previsioni numeriche operative

Si debbono risolvere equazioni differenziali alle derivate parziali, quindi sono necessarie le seguenti informazioni:

- Condizioni iniziali
- Condizioni al contorno
- Sorgenti

$$\frac{du}{dt} - \frac{uv \tan(\phi)}{R} + \frac{uw}{R} = 2\Omega v \sin(\phi) - 2\Omega w \cos(\phi) - \frac{1}{\rho} \frac{\partial p}{\partial x} + (\nu \Delta \mathbf{v})_x \quad (2.5)$$

$$\frac{dv}{dt} + \frac{u^2 \tan(\phi)}{R} + \frac{vw}{R} = -2\Omega u \sin(\phi) - \frac{1}{\rho} \frac{\partial p}{\partial y} + (\nu \Delta \mathbf{v})_y \quad (2.6)$$

$$\frac{dw}{dt} - \frac{u^2 + v^2}{R} = 2\Omega u \cos(\phi) - \frac{1}{\rho} \frac{\partial p}{\partial z} - g + (\nu \Delta \mathbf{v})_z \quad (2.7)$$

The state equation:

$$p = \rho RT \quad (2.8)$$

The continuity equation:

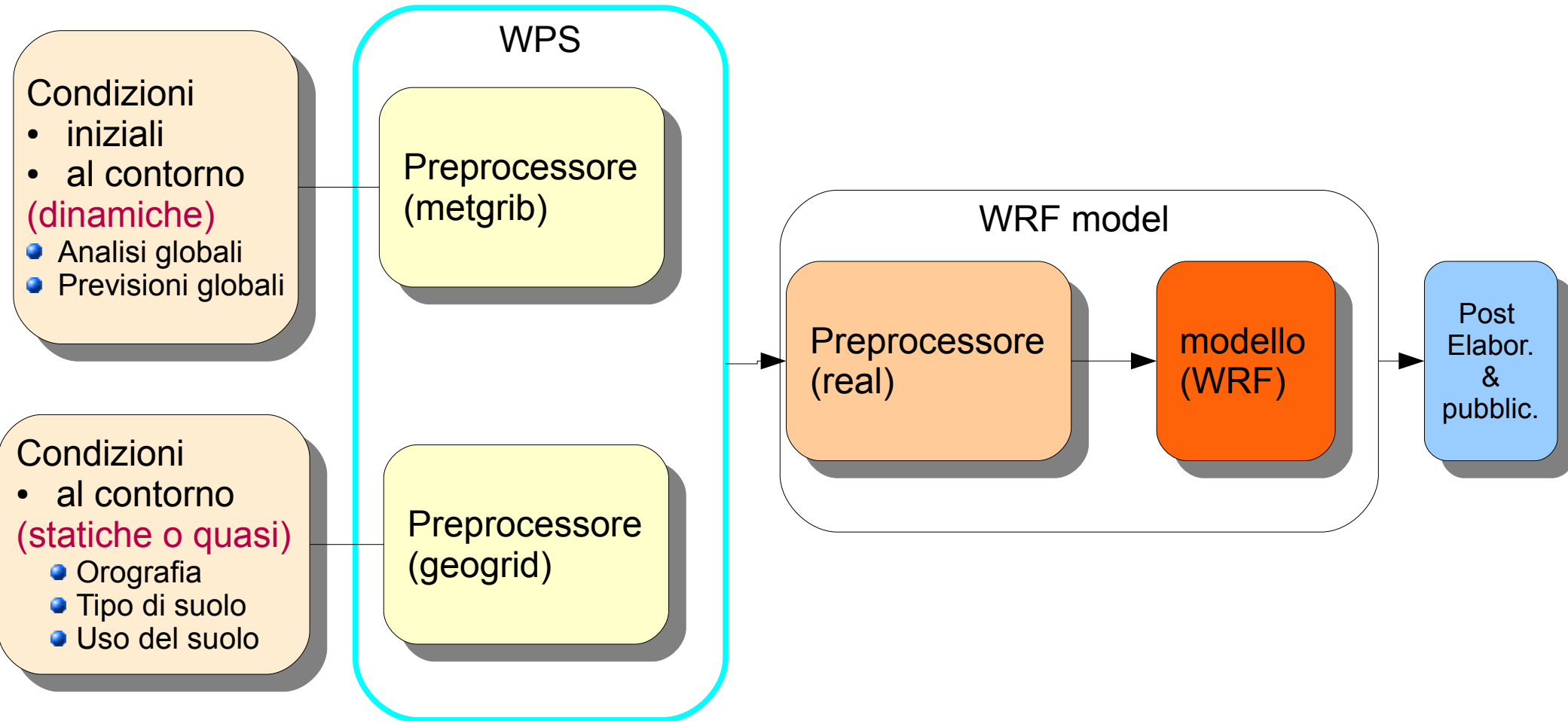
$$\frac{d\rho}{dt} + \rho \nabla \cdot \mathbf{v} = 0 \quad (2.9)$$

The energy conservation:

$$\frac{d\eta}{dt} = C_p \frac{dT}{dt} - \frac{1}{\rho} \frac{dp}{dt} \quad (2.10)$$

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x} \left( D_x \frac{\partial C}{\partial x} \right) + \frac{\partial}{\partial y} \left( D_y \frac{\partial C}{\partial y} \right) + \frac{\partial}{\partial z} \left( D_z \frac{\partial C}{\partial z} \right) + \frac{\partial E}{\partial t} + \frac{\partial D}{\partial t} \quad 10$$

## Esempio del modello meteorologico WRF – senza assimilazione dei dati locali



## Esempio del modello meteorologico WRF – acquisizione delle condizioni iniziali e al contorno

Condizioni  
 • iniziali  
 • al contorno  
 (dinamiche)

- Analisi globali
- Previsioni globali

Per scopi **prognostici**:

- ECMWF – Accesso tramite provider nazionale (AM)
  - Costo di rilascio
  - A disposizione dei centri funzionali della protezione civile
- NOAA – NCEP (GFS) Model – Accesso libero tramite FTP  
<http://www.nco.ncep.noaa.gov/pmb/products/gfs/>  
**Frequenza di acquisizione: quotidiana e sub quotidiana**

Per scopi **diagnostici**:

- ECMWF – Accesso con autorizz. del provider nazionale (AM)
    - Nessun costo, solo autorizzazione all'accesso ECMWF
- Frequenza di acquisizione: a piacere**

Condizioni  
 • al contorno  
 (statiche o quasi)

- Orografia
- Tipo di suolo
- Uso del suolo

Database globali: NOAA, USGS (WRF web site)  
 Database continentali: EEA - CORINE LAND COVER  
 Database locali: Regione  
**Frequenza di acquisizione: una tantum**

Altre possibili fonti di condizioni iniziali e al contorno **non globali**

Per scopi **prognostici** e **diagnostici**:

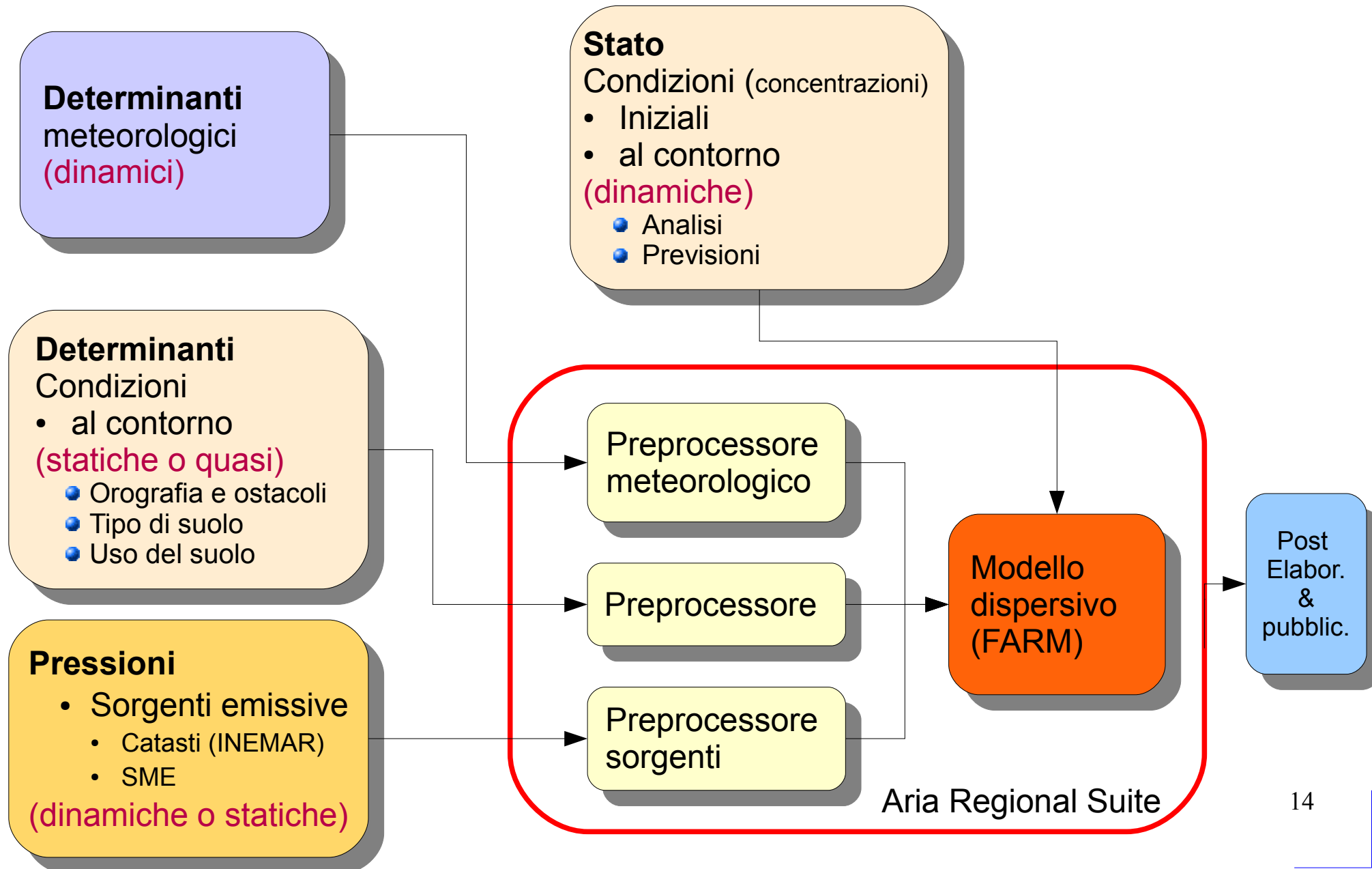
- ALADIN  
su convenzione o accordo
- COSMO  
su partecipazione al consorzio
- ecc.

Frequenza di acquisizione:  
quotidiana e sub quotidiana - **prognostici**  
secondo necessità - **diagnostici**

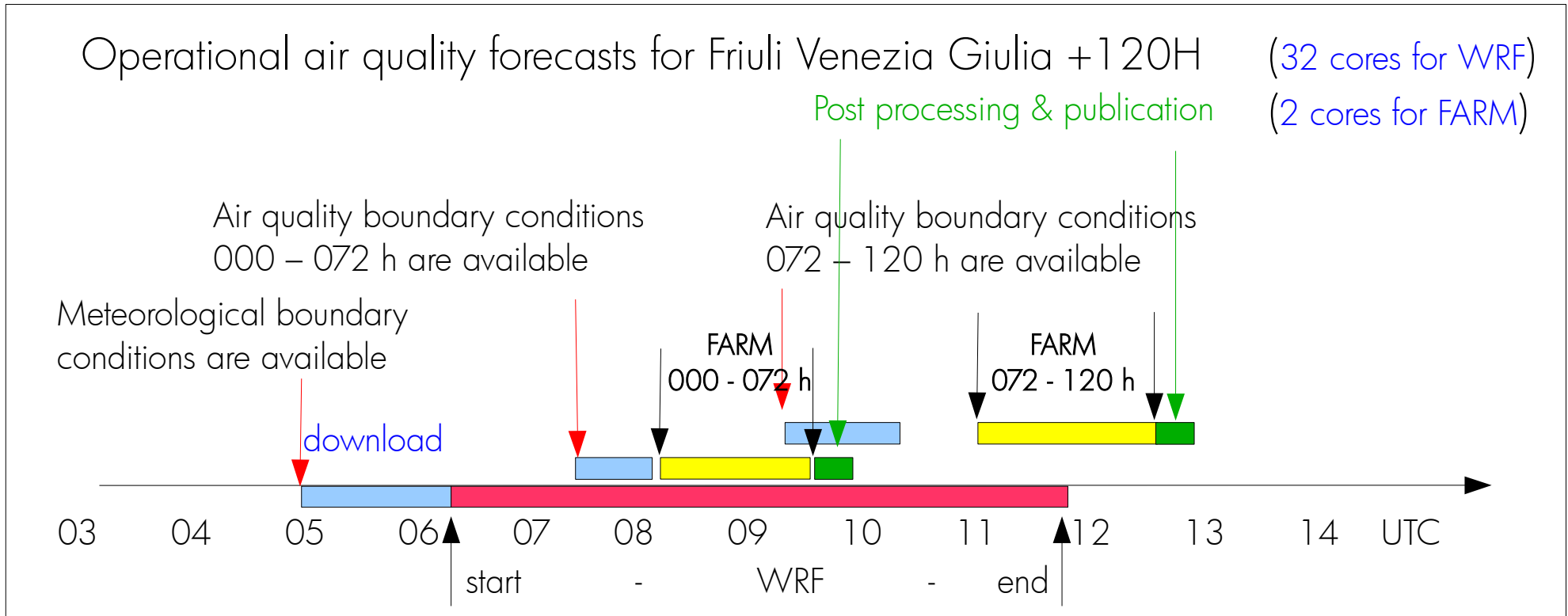
Condizioni

- Iniziali
- al contorno  
(dinamiche)
- Analisi globali
- Previsioni globali

Esempio del modello dispersivo off line FARM – senza assimilazione dei dati locali



## Esempio: catena operativa per la previsione della qualità dell'aria in regione FVG



**Criticità: Reti veloci – HPC – workflow managers**  
[http://en.wikipedia.org/wiki/High-performance\\_computing](http://en.wikipedia.org/wiki/High-performance_computing)

- **Scaricare solo il necessario: usare curl e i suggerimenti proposti dal NCEP**  
[ftp://ftp.cpc.ncep.noaa.gov/wd51we/fast\\_downloading\\_grib/](ftp://ftp.cpc.ncep.noaa.gov/wd51we/fast_downloading_grib/)
- **Verificare che lo scaricamento sia andato a buon fine oppure ritentare**
- **Monitorare il workflow - Utilizzare software che gestisca il workflow**  
 ECFLOW <https://software.ecmwf.int/wiki/display/ECFLOW/Home>  
 Kepler <https://kepler-project.org/>

### Criticità: HTC – spazio disco - workflow managers

(HTC High Throughput Computing [http://en.wikipedia.org/wiki/High-throughput\\_computing](http://en.wikipedia.org/wiki/High-throughput_computing))

Air quality scenarios Friuli Venezia Giulia (computation only)

- 1/3 hour-core for one simulation day
- 365 days of simulation
- 2 hour-core for one scenario post-proccesing
- 15 scenarios

About 1 800 hours-core

Air quality scenarios Friuli Venezia Giulia (storage only)

- 20 GB for one scenario
- 15 scenarios

About 300 GB

#### Suggerimenti:

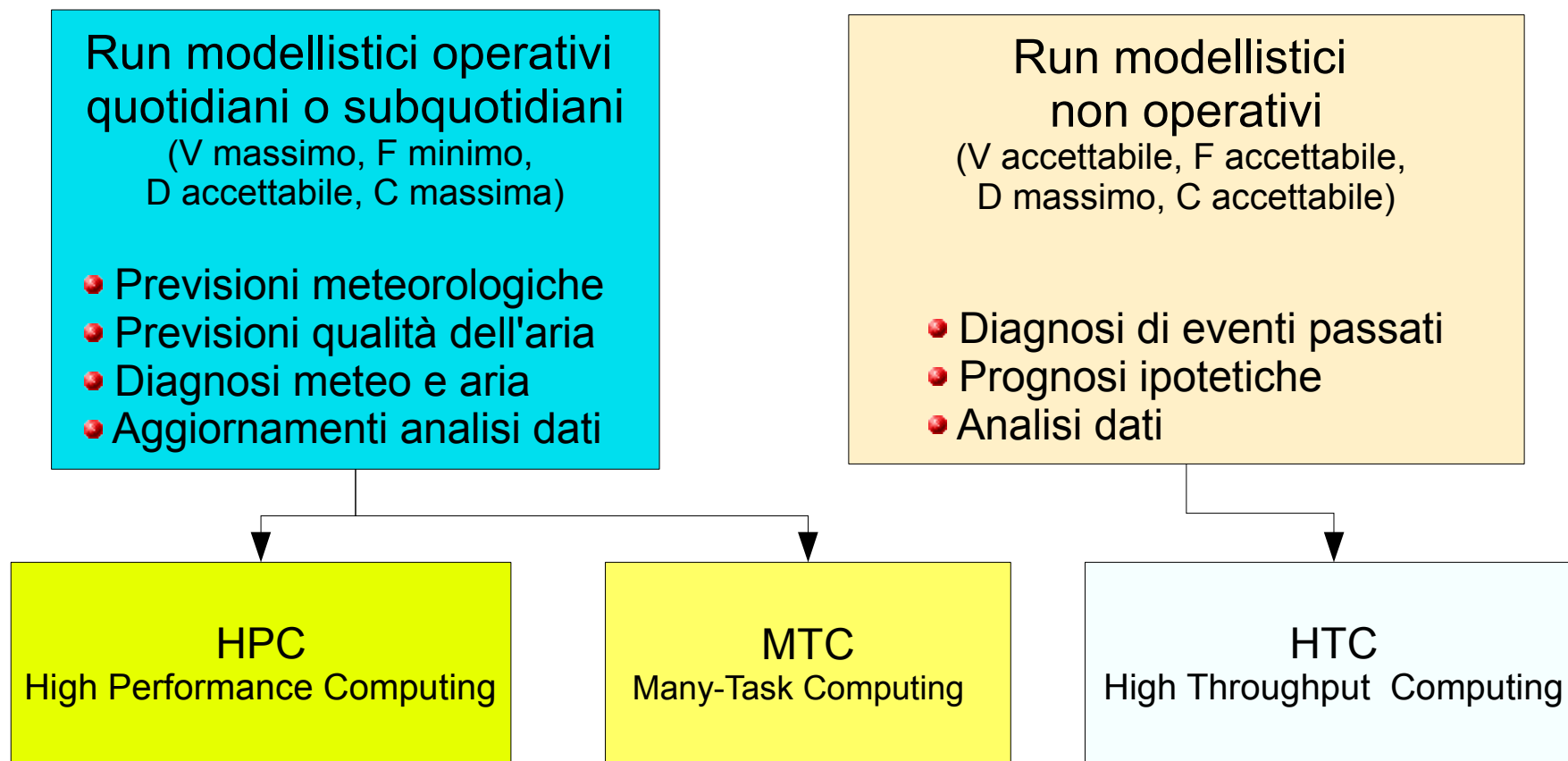
- progettare archiviazione e rimozione automatica degli output (efficienza)
- definire una nomenclatura univoca delle simulazioni (ordine)
- usare links simbolici ogni volta ciò è possibile (economia)



## Catene modellistiche: sintesi I

Ci sono due classi di simulazioni che si distinguono per le seguenti caratteristiche:

- V = velocità di esecuzione
- F = probabilità di fallimento
- D = spazio archiviazione
- C = connessioni esterne e interne



HPC [http://en.wikipedia.org/wiki/High-performance\\_computing](http://en.wikipedia.org/wiki/High-performance_computing)

HTC [http://en.wikipedia.org/wiki/High-throughput\\_computing](http://en.wikipedia.org/wiki/High-throughput_computing)

MTC [http://en.wikipedia.org/wiki/Many-task\\_computing](http://en.wikipedia.org/wiki/Many-task_computing)

## Catene modellistiche: alcuni esempi di risorse necessarie

Alcuni esempi dei sistemi modellistici e le risorse richieste per le seguenti caratteristiche:

- V = velocità di esecuzione
- D = spazio archiviazione
- F = probabilità di fallimento
- C = connessioni esterne e interne

### HPC

High Performance Computing

90 ore cpu/run giornaliero  
(180 ore cpu/run giornaliero)

Fallimenti: 1-2 anno  
Connessioni 1GB/ora

Spazio archivio 100MB/giorno

Previsioni meteorologiche  
Previsioni qualità aria online

### MTC

Many-Task Computing

150 ore cpu/run giornaliero  
(250 ore cpu/run giornaliero)

Fallimenti: 1-2 anno  
Connessioni 1GB/ora

Spazio archivio 1GB/giorno

Previsioni qualità aria offline  
Diagnosi operativa  
Analisi dati operativa

### HTC

High Throughput Computing

5000 ore cpu/run annuale  
(40000 ore cpu/run annuale)

Fallimenti: 10-50 anno  
Connessioni 5TB/anno

Spazio archivio 1TB/anno

Diagnosi eventi passati  
Prognosi ipotetiche  
Analisi dati

Gli approcci e le strategie modellistiche possono diventare estremamente più efficienti se inserite in un ambiente ricco non solo di risorse computazionali, ma anche di competenze complementari a supplementari a quelle del gruppo di modellisti

Gestione efficiente dei flussi dati in ingresso ed in uscita

Controllo dei flussi di calcolo MTC e HTC e gestione dei fallimenti

Disponibilità dei sistemi modellistici a utenti non modellisti

Adeguati sistemi di accesso ed esplorazione dei dati per i diversi utenti

Risorse di calcolo e di archiviazione dati

Gestione archivi dati molto voluminosi

**e-Infrastructure**