A new framework for calibrating COVID-19 SEIR models with spatial-/time-varying coefficients using genetic and sliding window algorithms

Huan Zhou[1], Ralf Schneider[1] , Sebastian Klüsener[2] and Andreas Backhaus[2]

WSSP‘35, HLRS, Germany, 14.04.2023

[1] High Performance Computing Center Stuttgart (HLRS), Germany
[2] Federal Institute for Population Research (BiB)

Outline

- Problem statement
- Solution
  - CoSMic
  - OSW: Segmentation and automation
  - PGA: Segmented calibration
- Results
  - Calibration accuracy

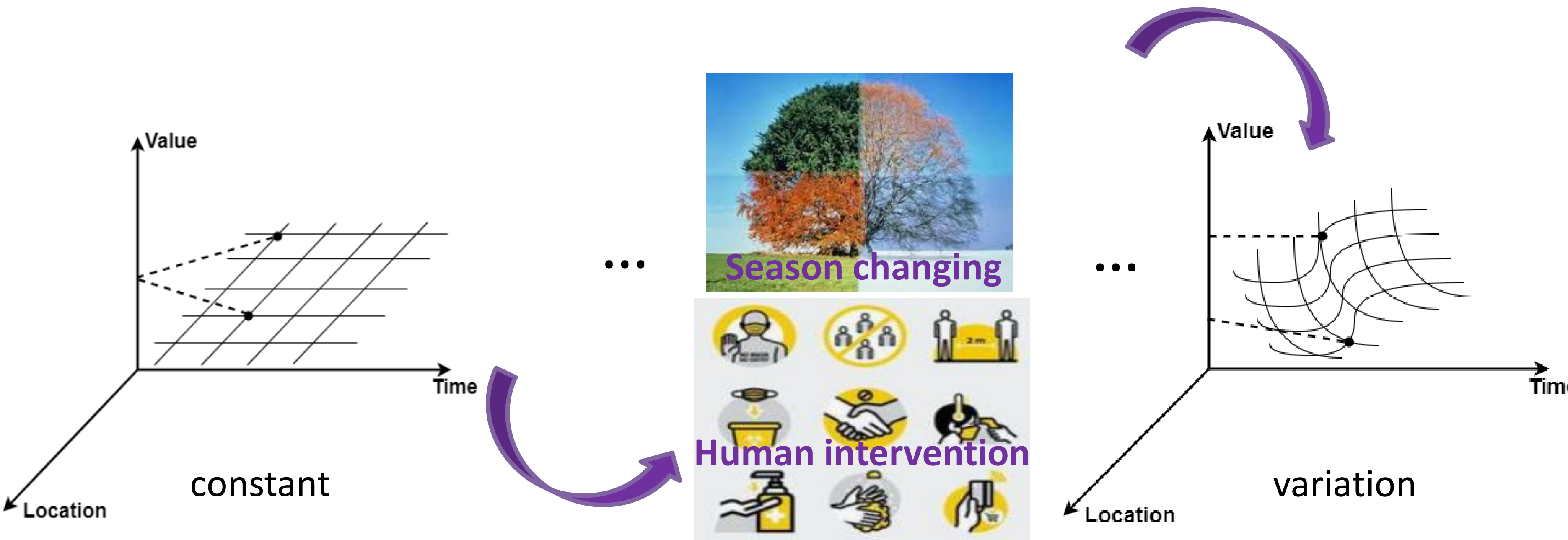
Problem statement

- Model calibration (outputs → inputs): improve the simulation output and be closer to available observations
  - Identify the uncertain input parameters
  - Estimate the identified input parameters containing uncertainties

Focus: estimation

Problem statement

- The estimated parameters are spatial-/time-varying; they are common in the epedemical (i.e., SEIR) model
  - Capture the historical trend over time, but spatially the trends differ from location to location



Problem statement

- **Challenge**: fast and accurate estimation of the spatial-/time-varying parameters
  - Segmentation
    - Segment a complex calibration over a long lifespan into smaller ones over different time slices, which are solved sequentially. The sub-solutions are refined for accuracy and then combine to determine the best trajectory of the parameter
  - Optimization methods
    - Each segmented calibration is prone to facing a high-dimensional search space due to spatial variation
    - The applied optimization method should be easy to be parallelized for fast estimation

Outline

- Problem statement
- Solution
  - CoSMic
  - OSW: Segmentation and automation
  - PGA: Segmented calibration
- Results
  - Calibration accuracy

Proposed calibration approach

- Our calibration approach
  - Segmentation: Overlapping Sliding Window (OSW)
    - The sub-solutions are refined progressively as the window moves forward
  - Optimization method: Parallel Genetic Algorithm (PGA)
    - GA is a population-based metaheuristic inspired by the theory of natural evolution
    - Hybrid MPI+OpenMP

Combine OSW and PGA

Outline

- Problem statement
- Solution
  - CoSMic
  - OSW: Segmentation and automation
  - PGA: Segmented calibration
- Results
  - Calibration accuracy

CoSMic

- Use CoSMic model[1] to exemplify and test our calibration approach
- CoSMic
  - Extended Susceptible-Exposed-Infectious-Removed (SEIR) model for pandemic COVID-19 spread
  - Features a region-/weekly-varying coefficient (aka. parameter μ) impacting the reproduction number
    - Calibrated against the ICU data on NUTS-2 level

$<\vec{\mu}_1, \vec{\mu}_2, \ldots, \vec{\mu}_N>$ is a stream of data points over N weeks

A vector of optimized values of μ across all NUTS-2 regions at the second COVID-19 week

[1]https://github.com/hpcralf/CoSMic

Outline

- Problem statement
- Solution
  - CoSMic
  - OSW: Segmentation and automation
  - PGA: Segmented calibration
- Results
  - Calibration accuracy

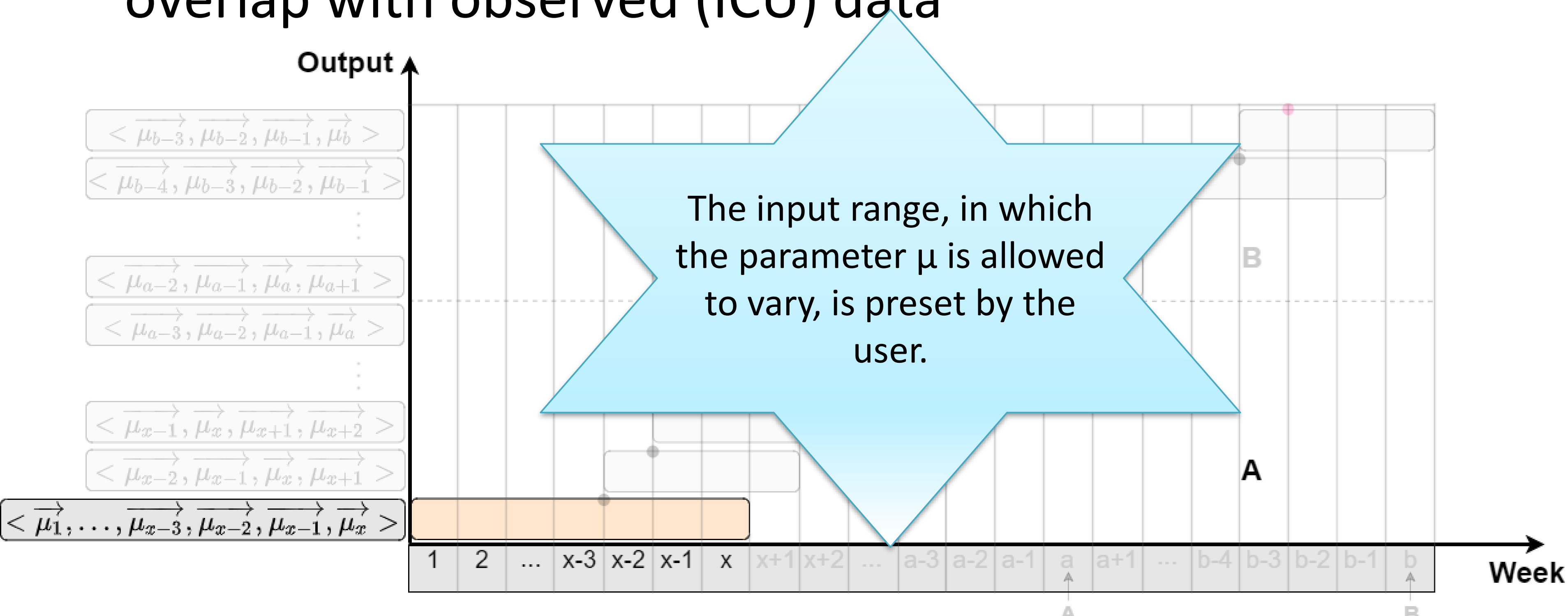
OSW: Segmentation and automation

- Case study with two consecutive calibration executions named A and B, respectively
  - Window size: 4; shift factor: 1

Execution	Start week	End week
A	1	a
B	a-2	b

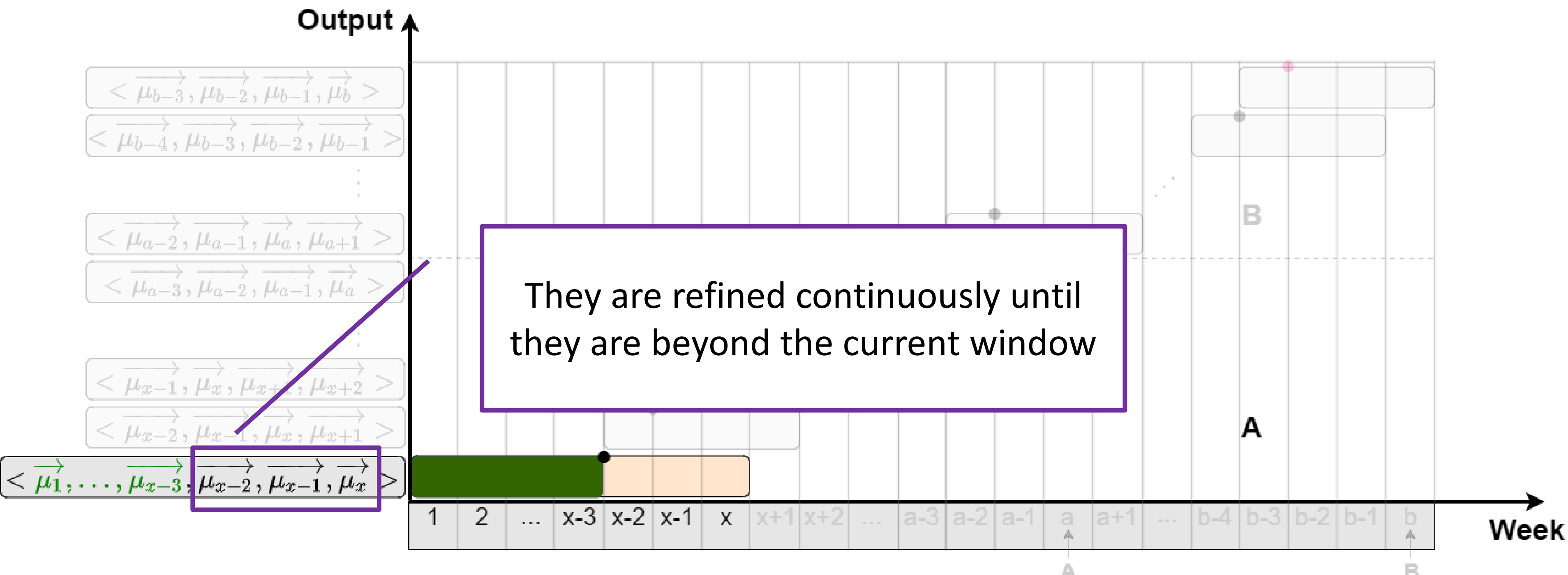
OSW: Segmentation and automation – execution A

- Segment a stream of time series data <μ⃗1, μ⃗2, ..., μ⃗a >, the initial window ends at the x-th week to minimally overlap with observed (ICU) data



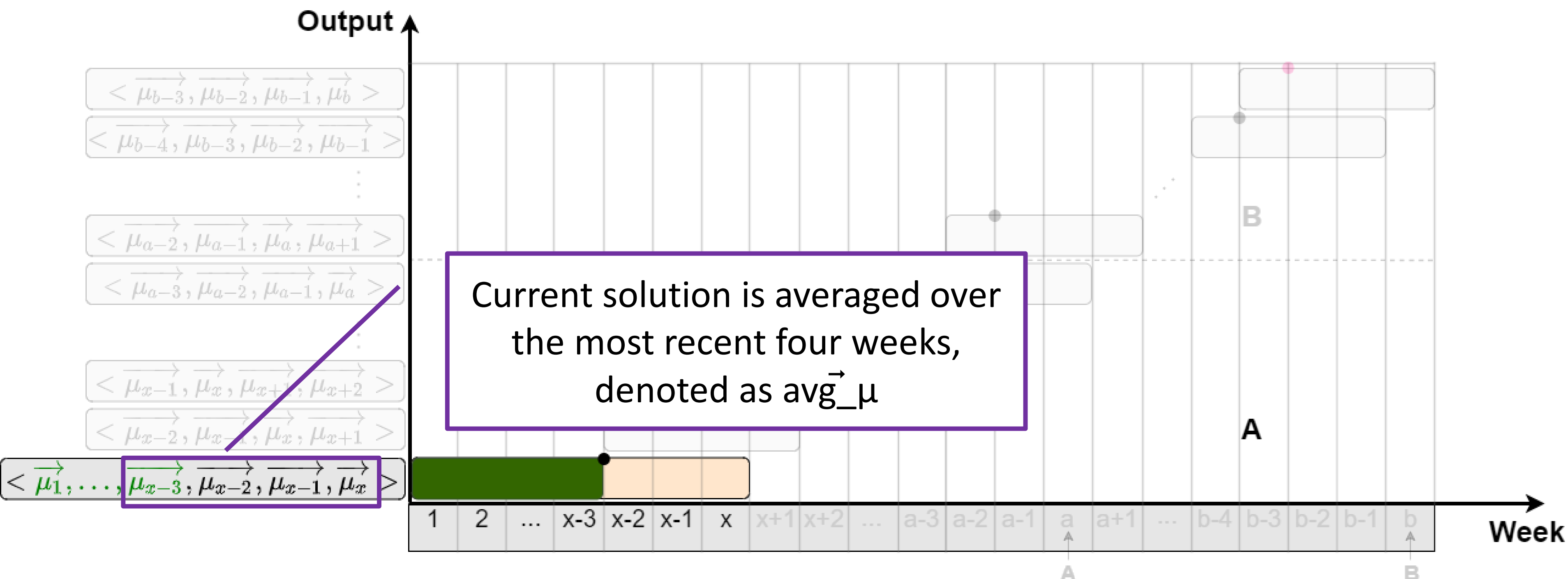
OSW: Segmentation and automation – execution A

- Use the estimated < $\vec{\mu}_1$, $\vec{\mu}_2$, ..., $\vec{\mu}_{x-3}$> to calculate the simulation status at *(x-2)*-th week (temporal restart point) from which the next segment starts



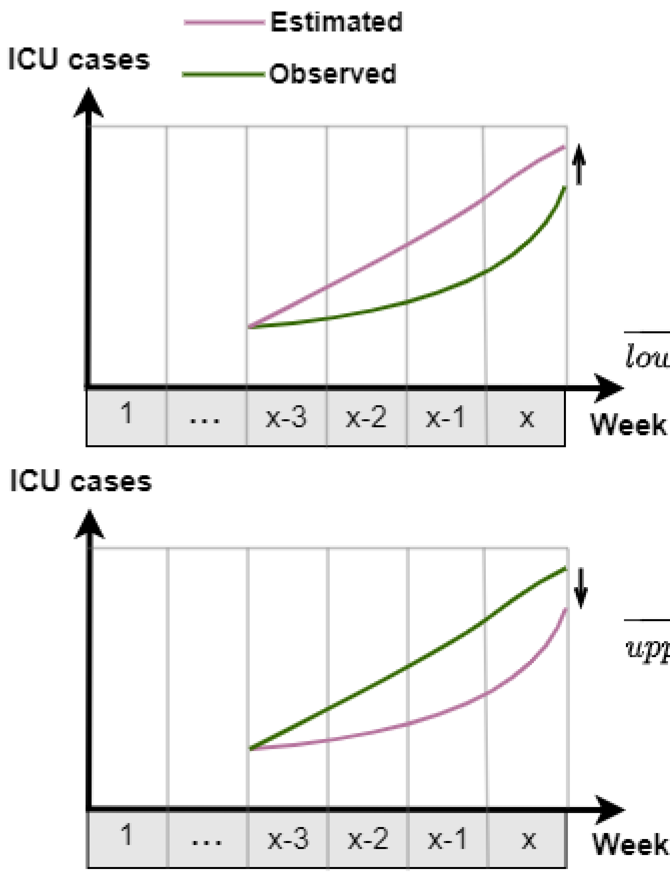
OSW: Segmentation and automation – execution A

- Adjust the input range of parameter μ for the next segmented calibration based on the current solution and the difference between the estimated and observed ICU cases at the last day of week x



OSW: Segmentation and automation – execution A

- How the input range is adjusted?



$$\overrightarrow{upper\_bound} = \overrightarrow{avg\_\mu}$$

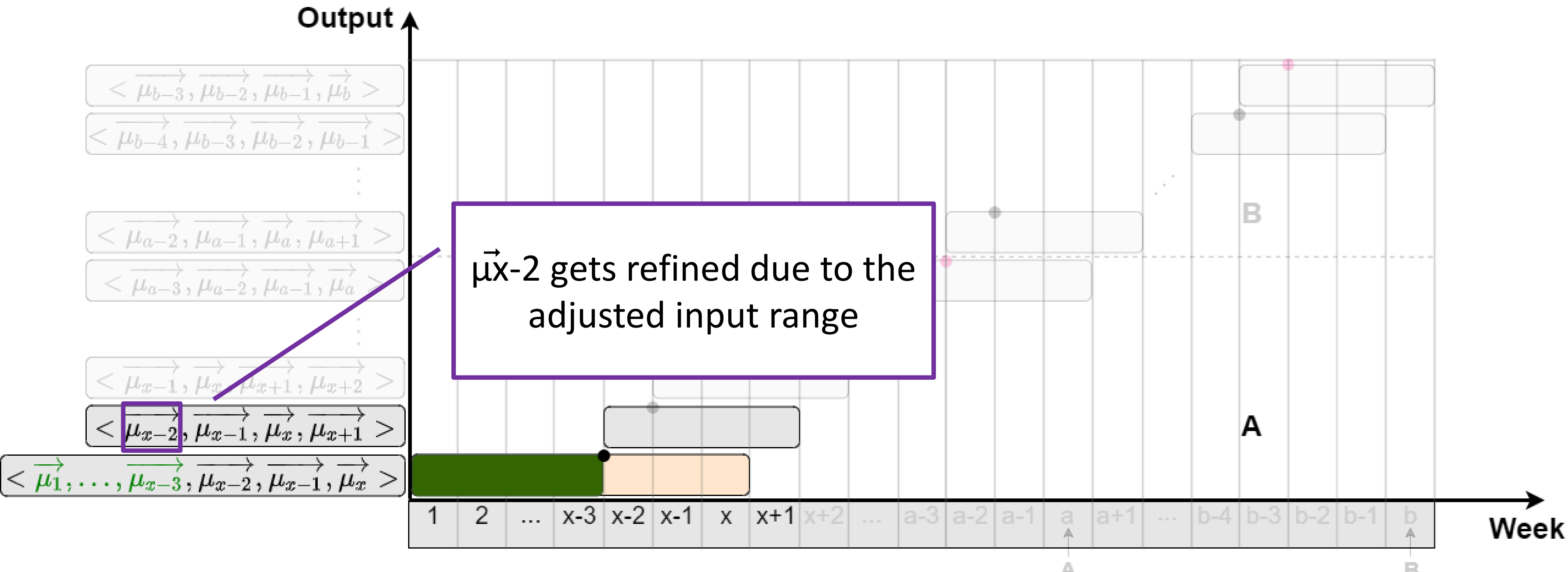
$$\overrightarrow{lower\_bound} = \overrightarrow{avg\_\mu} - coeff$$

$$\overrightarrow{upper\_bound} = \overrightarrow{avg\_\mu} + coeff$$

$$\overrightarrow{lower\_bound} = \overrightarrow{avg\_\mu}$$

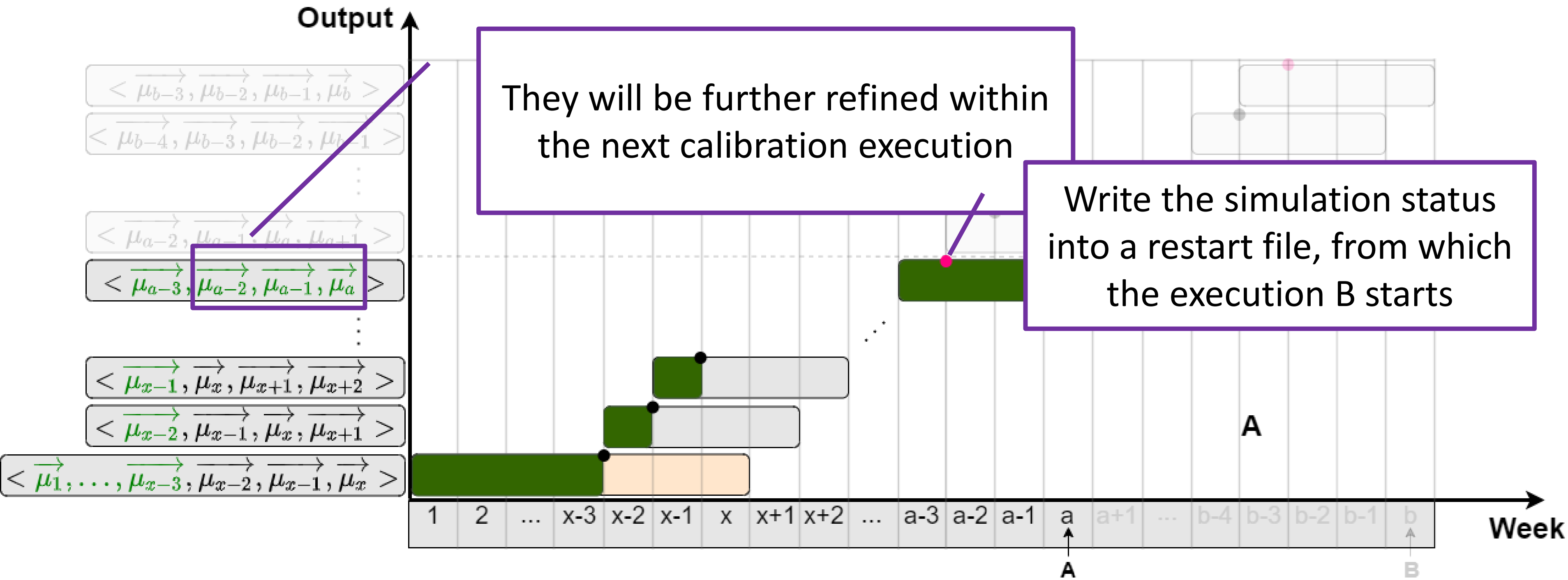
OSW: Segmentation and automation – execution A

- Slide the window by one week and continue to estimate the next segment with the standard window size (i.e.,4)



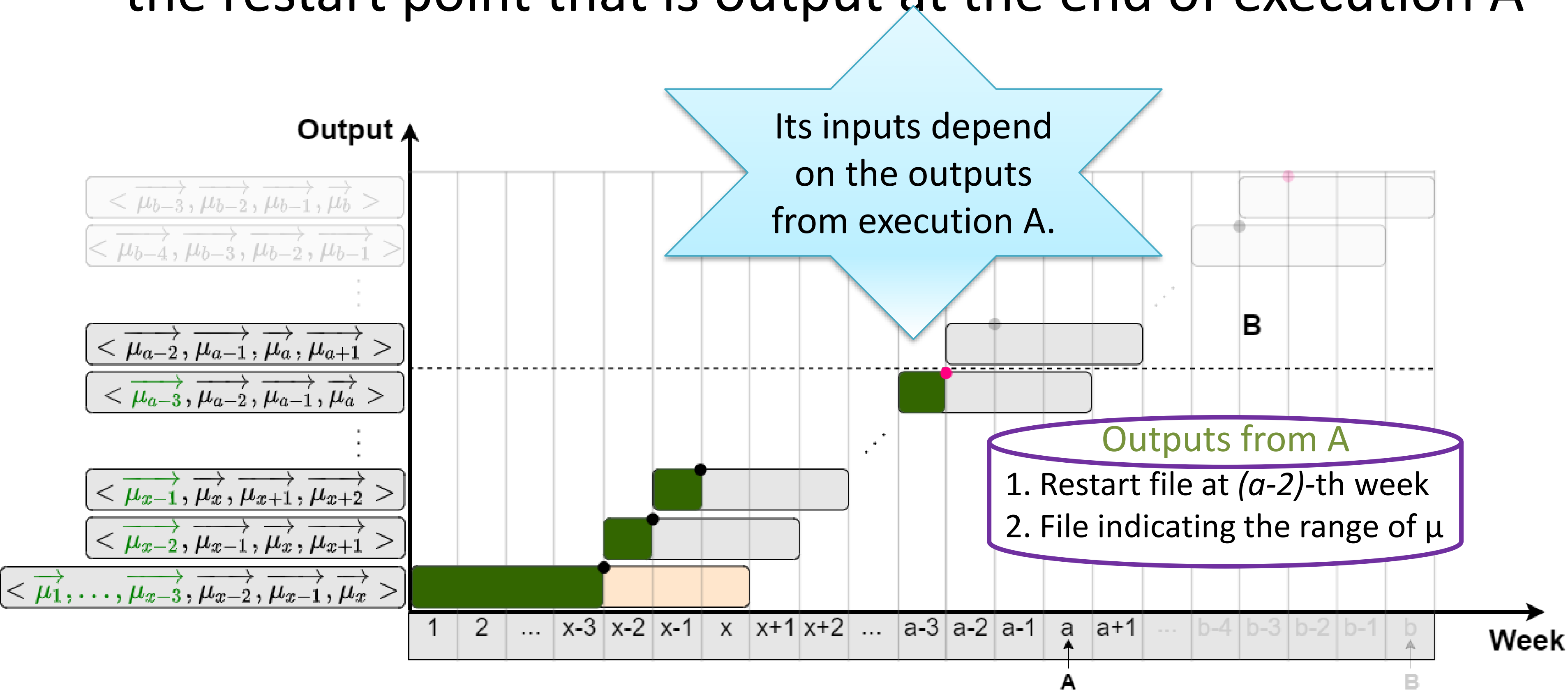
OSW: Segmentation and automation – execution A

- Execution A proceeds until the *a*-th week is approached
  - – Combine the segmented (colored with green) outputs
    - $< \vec{\mu}_1, \vec{\mu}_2, \ldots, \vec{\mu}_{a-1}, \vec{\mu}_a >$



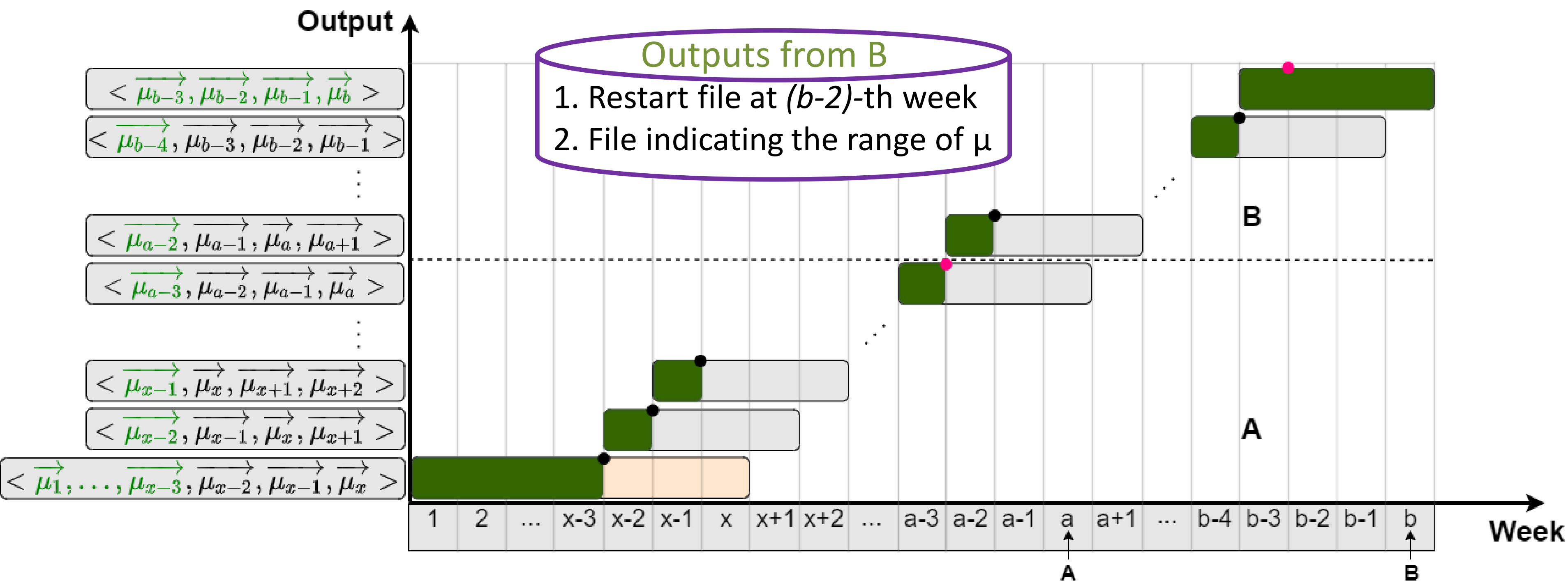
OSW: Segmentation and automation – execution B

- Segment <$\vec{\mu}_{a-2}$, $\vec{\mu}_{a-1}$, …, $\vec{\mu}_b$ >; execution B begins from the restart point that is output at the end of execution A



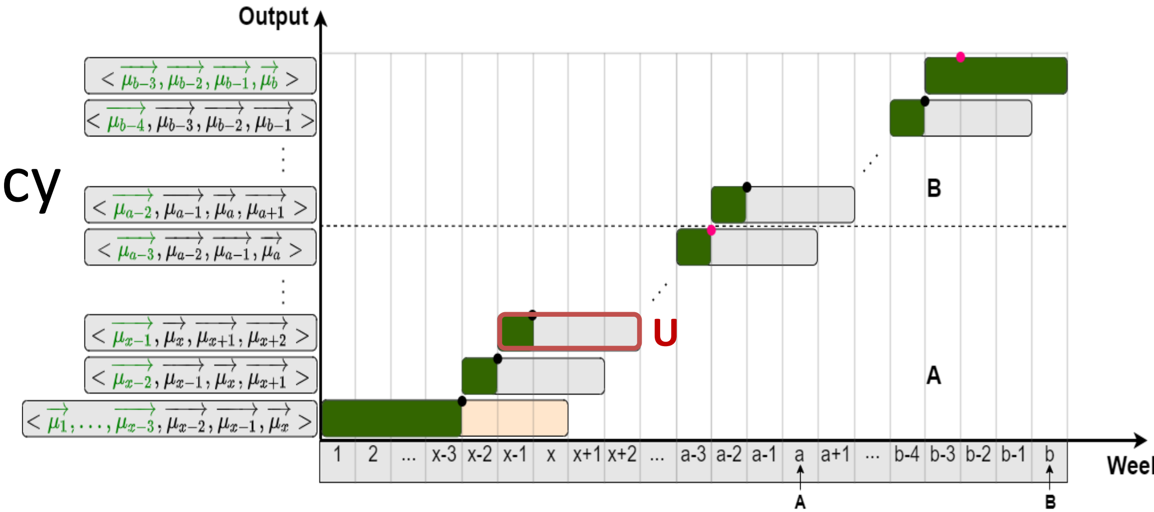
OSW: Segmentation and automation – execution B

- Execution B ends and combines the segmented outputs
  - < $\vec{\mu}_{a-2}$, $\vec{\mu}_{a-1}$, ..., $\vec{\mu}_{b-1}$, $\vec{\mu}_{b}$>



Outline

- Problem statement
- Solution
  - CoSMic
  - OSW: Segmentation and automation
  - PGA: Segmented calibration
- Results
  - Calibration accuracy



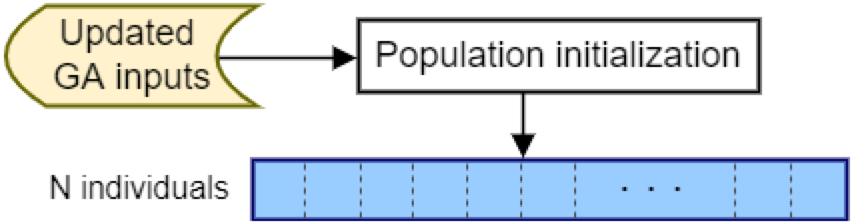
PGA: Segmented calibration

- GA is a population-based metaheuristic inspired by the theory of natural evolution
  - Fitter individuals are more likely to be selected and reproduce offspring of the next generation

- Five components in a basic GA
  - Initial population
  - Fitness evaluation
  - Selection
  - Crossover
  - Mutation

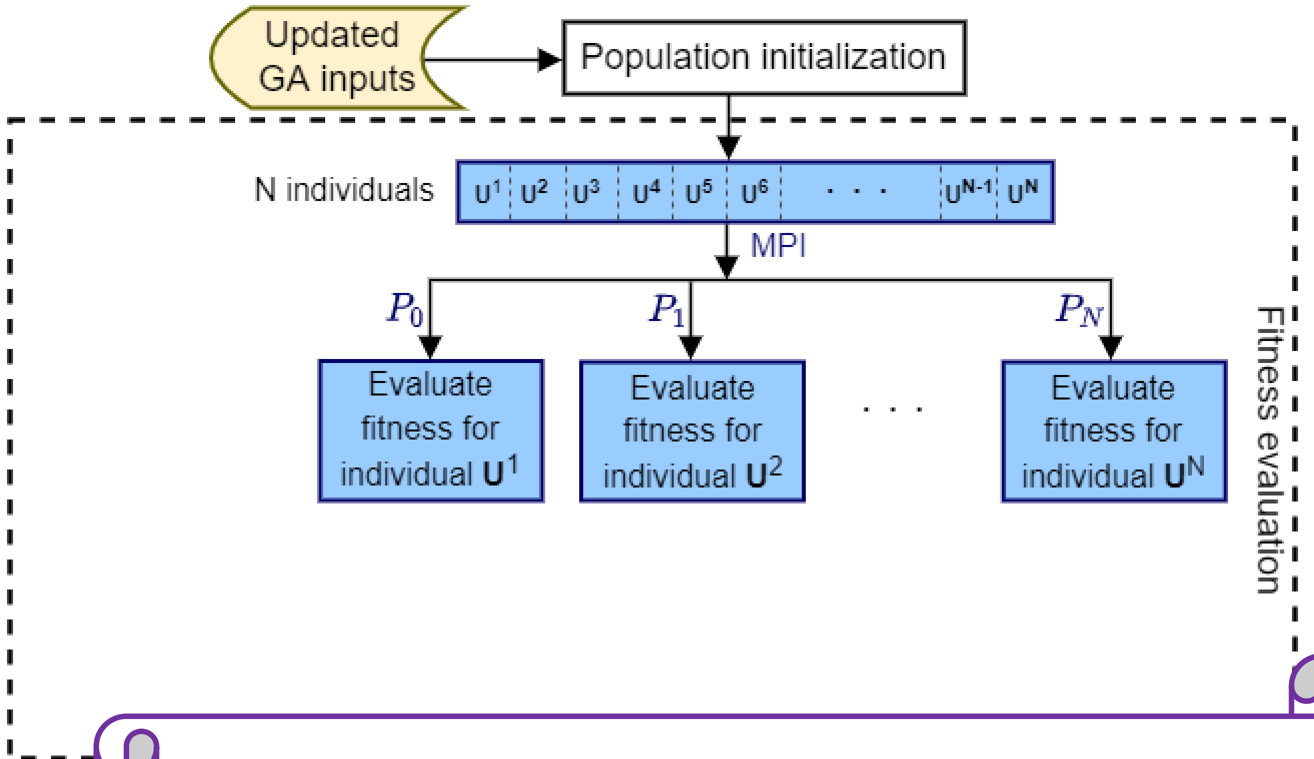
- Master-slave parallelization
  - Fitness evaluation
    - Combines with the problem of interest (CoSMic) → expensive
    - Done in parallel → independency between individuals

PGA: Segmented calibration – workflow of applying PGA to segmented calibration of CoSMic



- GA inputs
  - ✓ Population/generation size
  - ✓ Lower/Upper bound of the estimated parameter
- Population initialization
  - ✓ Comprise N candidate solutions/individuals
  - ✓ Each individual is generated randomly

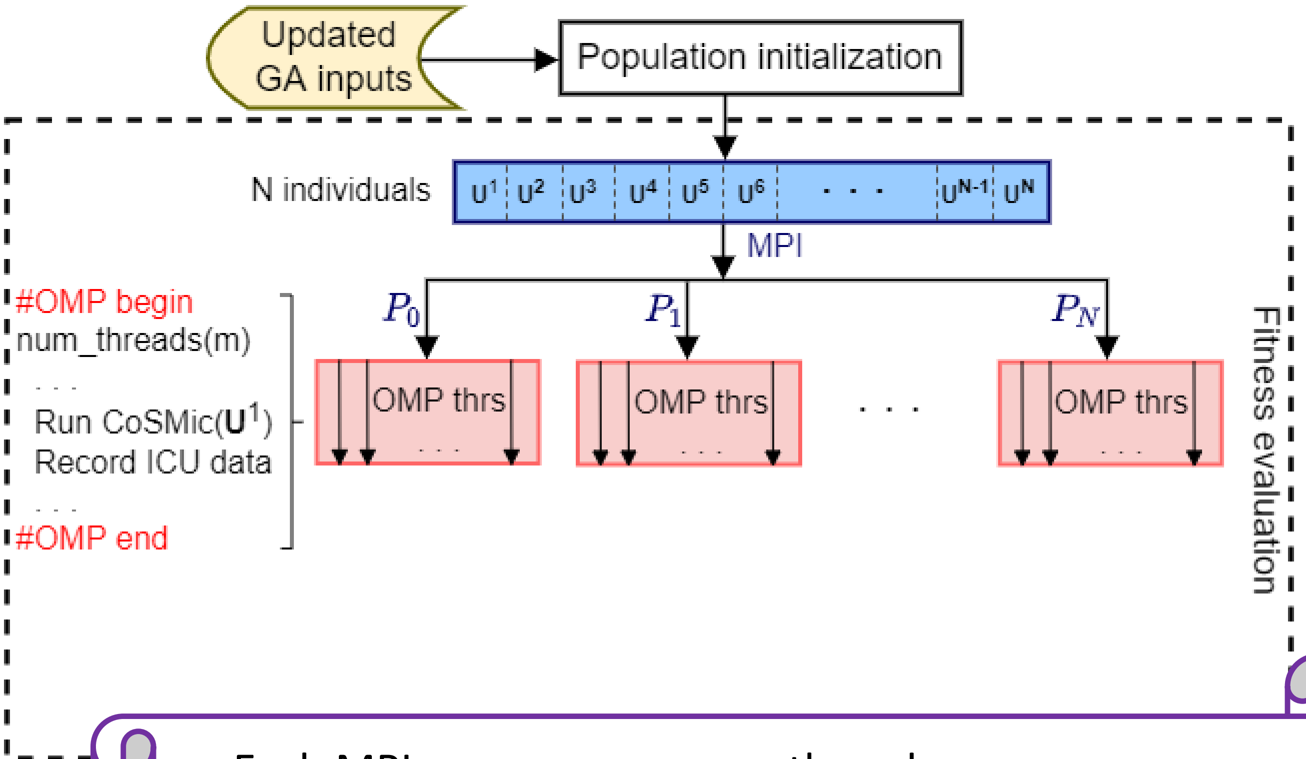
PGA: Segmented calibration – workflow of applying PGA to segmented calibration of CoSMic



- Fitness evaluation
  - ✓ The individuals are evaluated by N MPI processes in parallel
  - ✓ P_0: master; $P_{1\text{-}N}$: slaves

! MPI serves the distributed memory parallelism

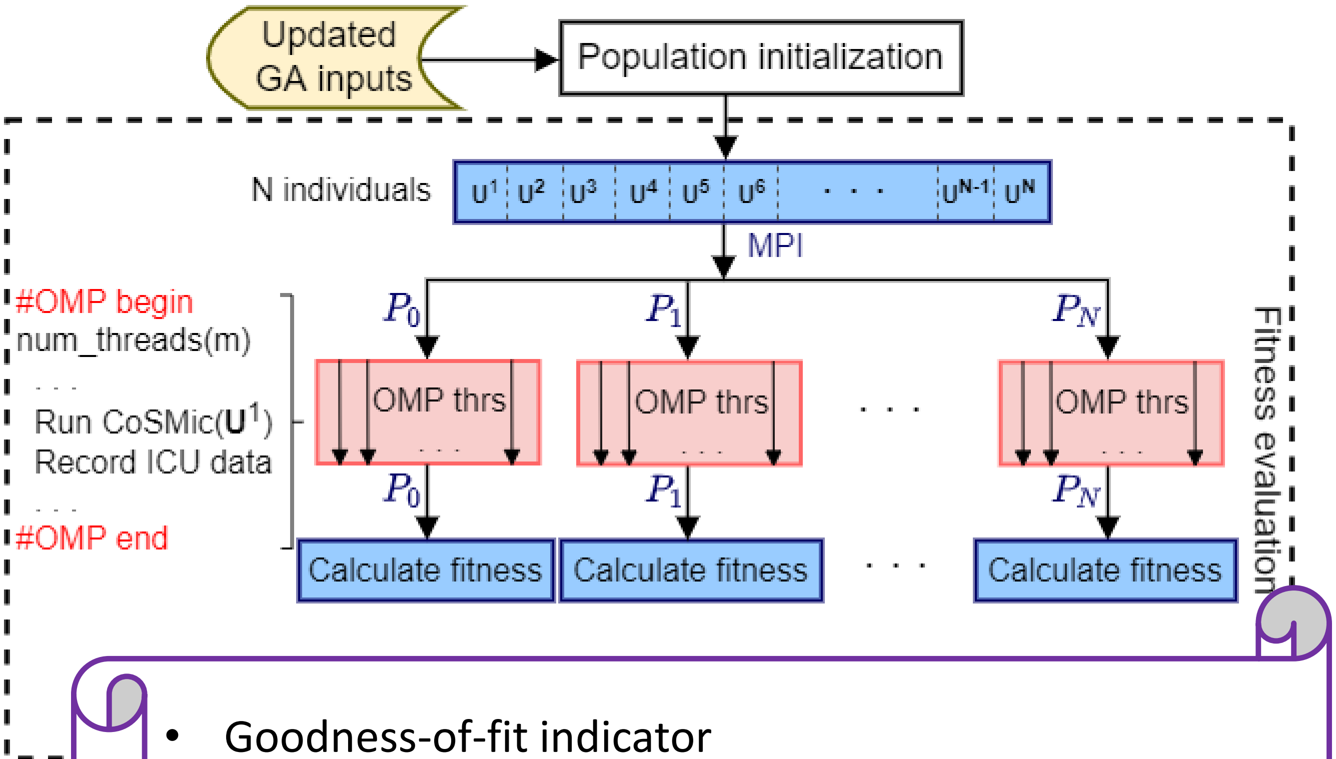
PGA: Segmented calibration – workflow of applying PGA to segmented calibration of CoSMic



- Each MPI process spawns m threads
  - ✓ Concurrently handle the multiple iterations of simulation in running the CoSMic model
  - ✓ Output: simulated ICU data

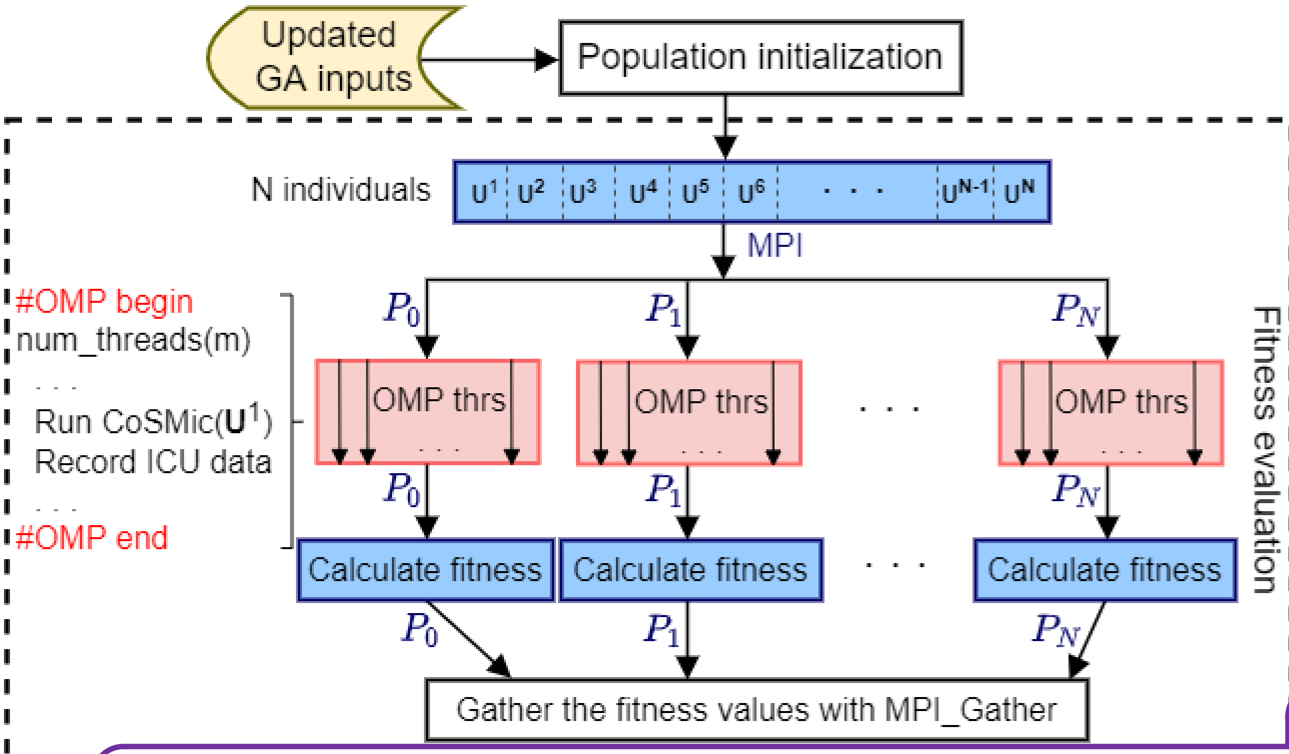
! OpenMP serves the shared-memory parallelism

PGA: Segmented calibration – workflow of applying PGA to segmented calibration of CoSMic



- Goodness-of-fit indicator
  - ✓ Compute root-mean-square Error (RMSE) by comparing the simulated and observed ICU cases
- Fitness = -RMSE = $-\sqrt{\sum_{k=0}^{n}\binom{n}{k}\left(ICU_{sim}^{k} - ICU_{obs}^{k}\right)^{2}/n}$
  - ✓ Fitter individuals have higher fitness values

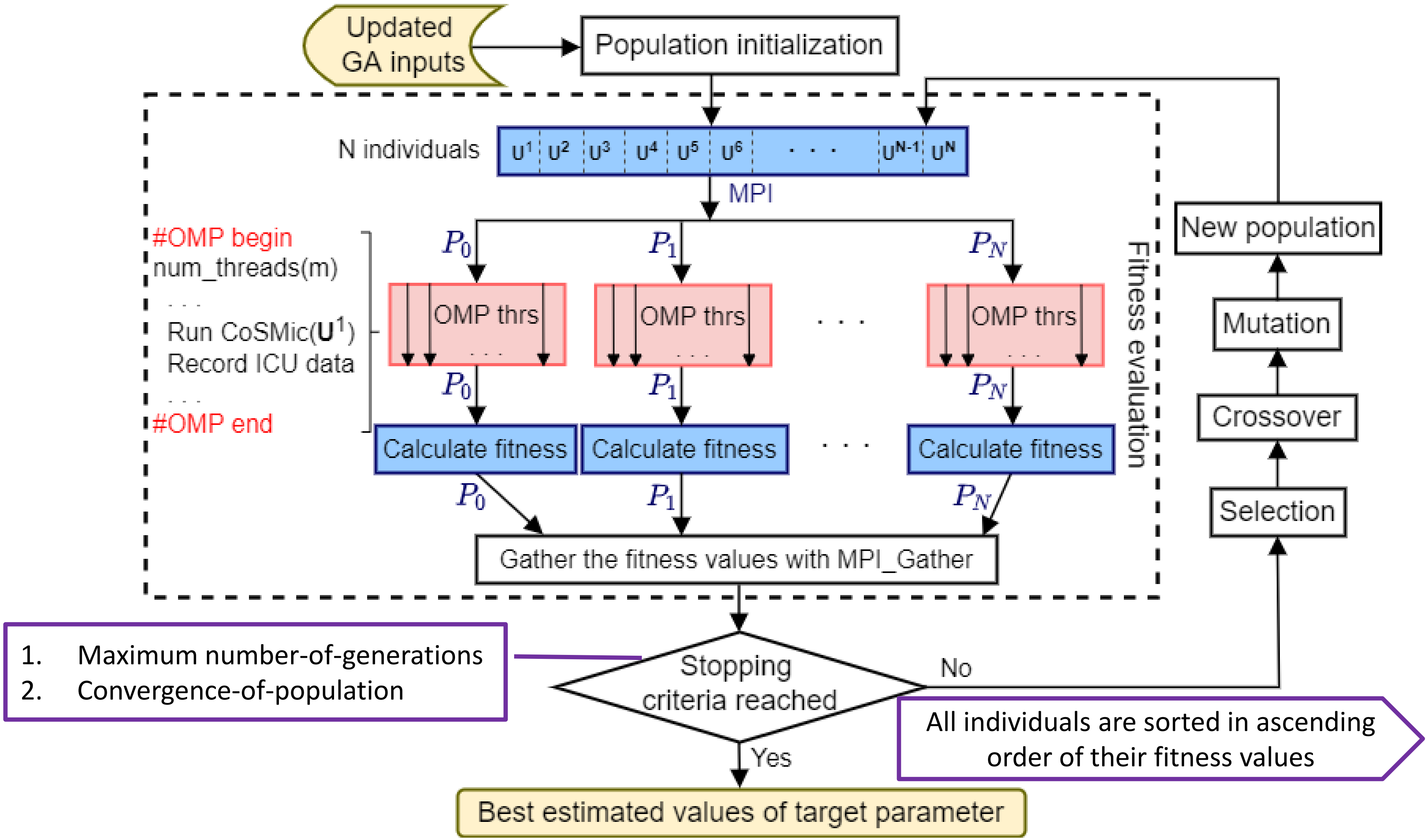
PGA: Segmented calibration – workflow of applying PGA to segmented calibration of CoSMic



- MPI communication
  - ✓ Slaves receive individuals from the master (broadcast)
  - ✓ Master (P0) gathers all the evaluated fitness values from slaves

! Hybrid MPI+OpenMP method

PGA: Segmented calibration – workflow of applying PGA to segmented calibration of CoSMic



Outline

- Problem statement
- Solution
  - CoSMic
  - OSW: Segmentation and automation
  - PGA: Segmented calibration
- Results
  - Calibration accuracy

Experimental Environment

- **Vulcan (NEC cluster)**
  - Infiniband (network)
  - 2x Intel Xeon E5-2660v3 processor
  - 20 cores per node
  - OpenMPI

Results – calibration accuracy

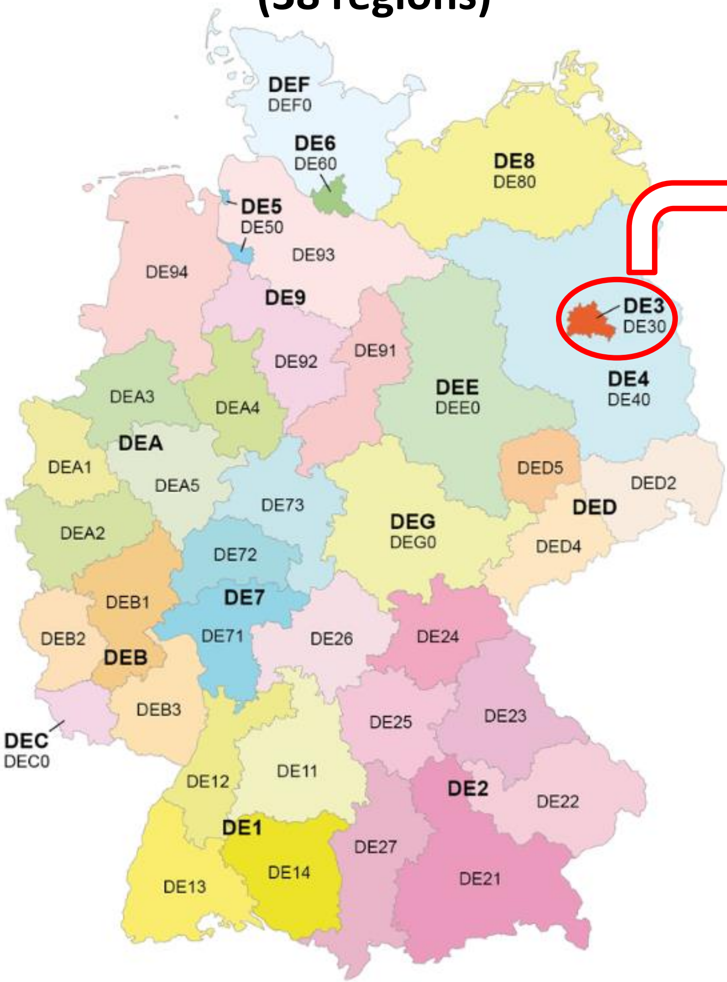
- Application benchmark: specification.

Sample of German population	GA inputs		Iterations of simulation
100% (83,237,124)	Population: 80	Generation: 10	40

- Compare the simulated and observed ICU data of all the German NUTS-2 regions over 2+ years
  - Hybrid version on 1600 cores
    - 80 MPI processes; 20 threads per MPI process

Results – calibration accuracy

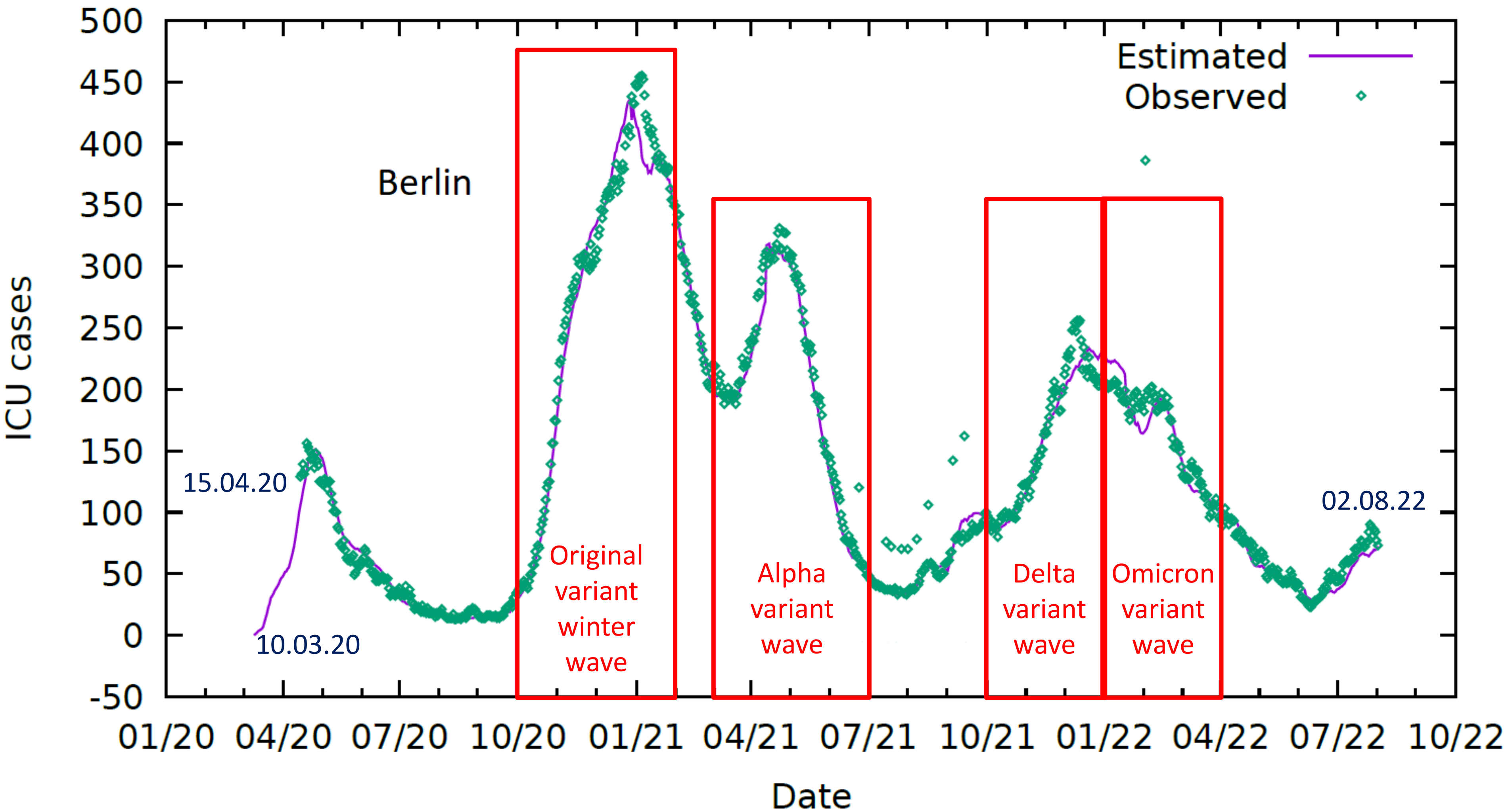
German NUTS-2 map
(38 regions)



Berlin

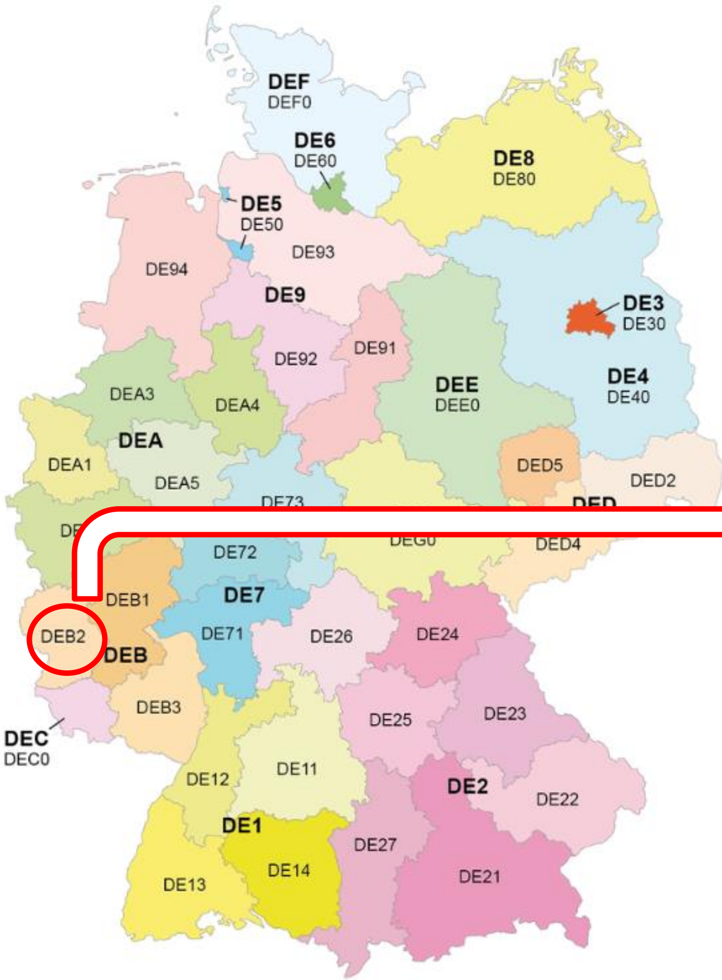
High density: 4,126/km2

Results – calibration accuracy



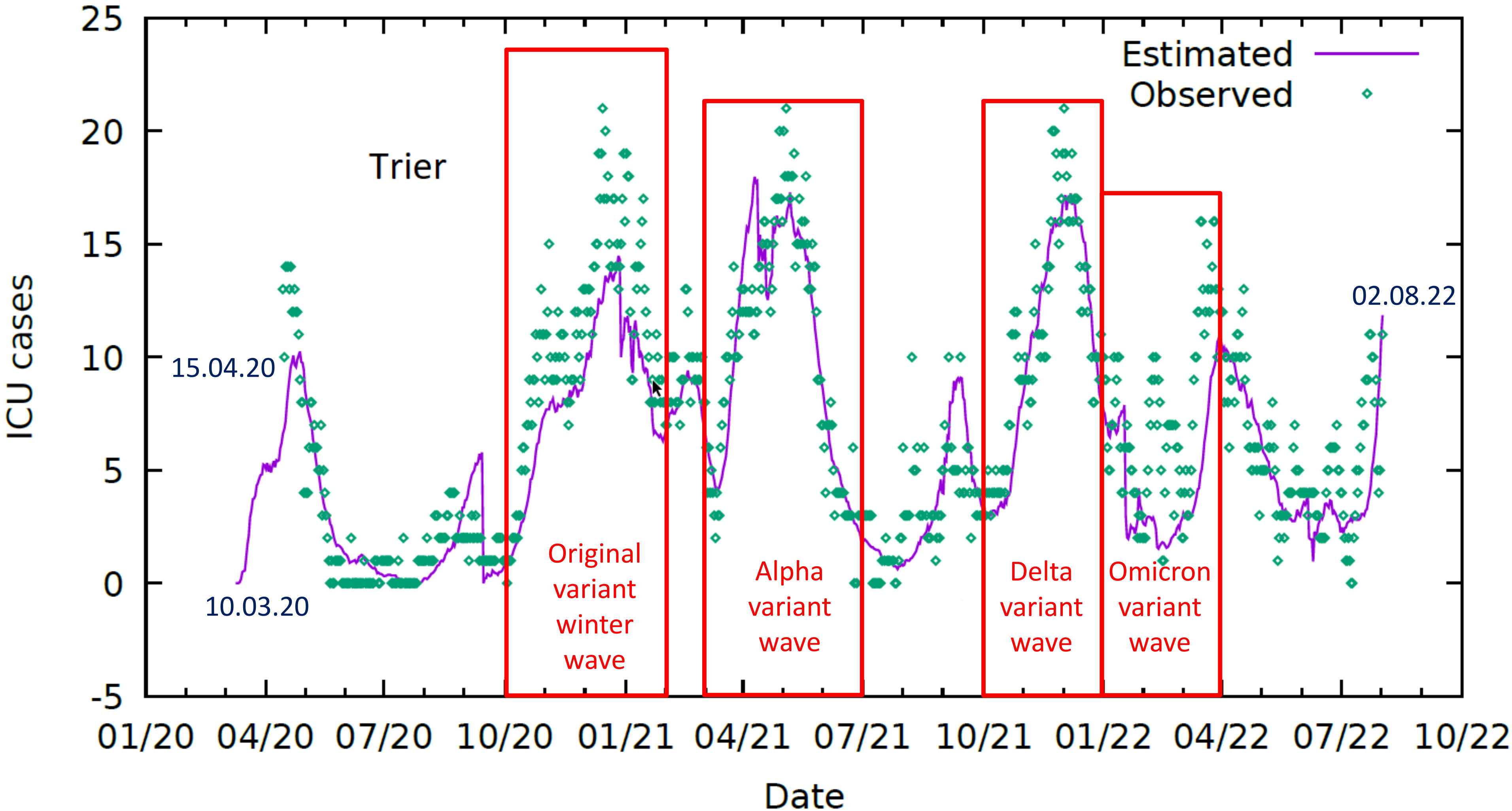
Results – calibration accuracy

German NUTS-2 map



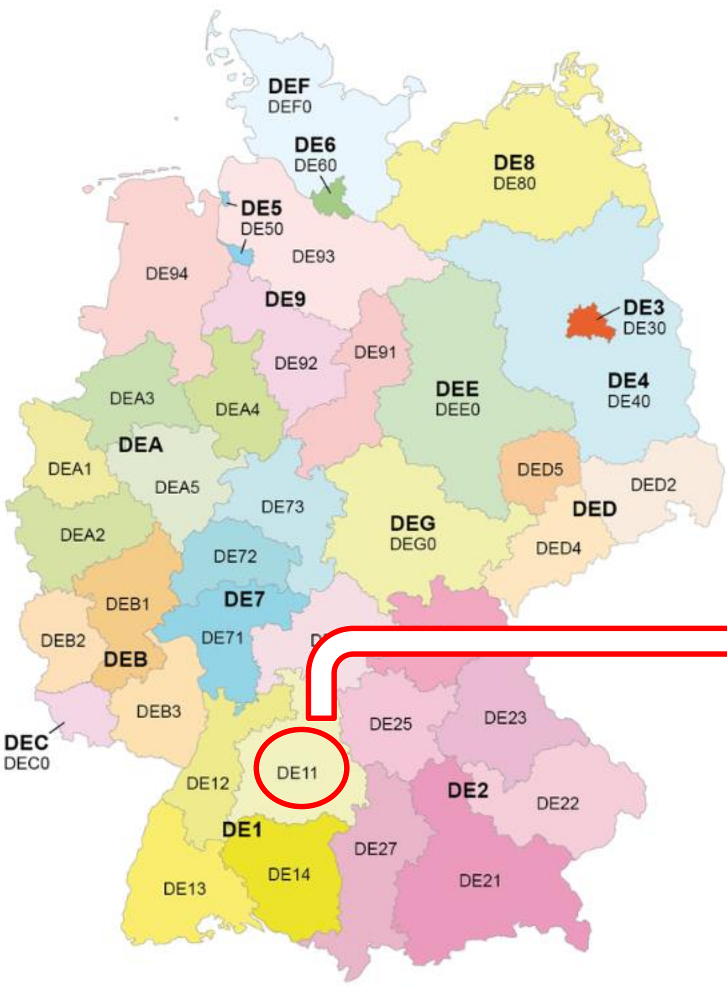
Trier
Low density: 100/km2

Results – calibration accuracy



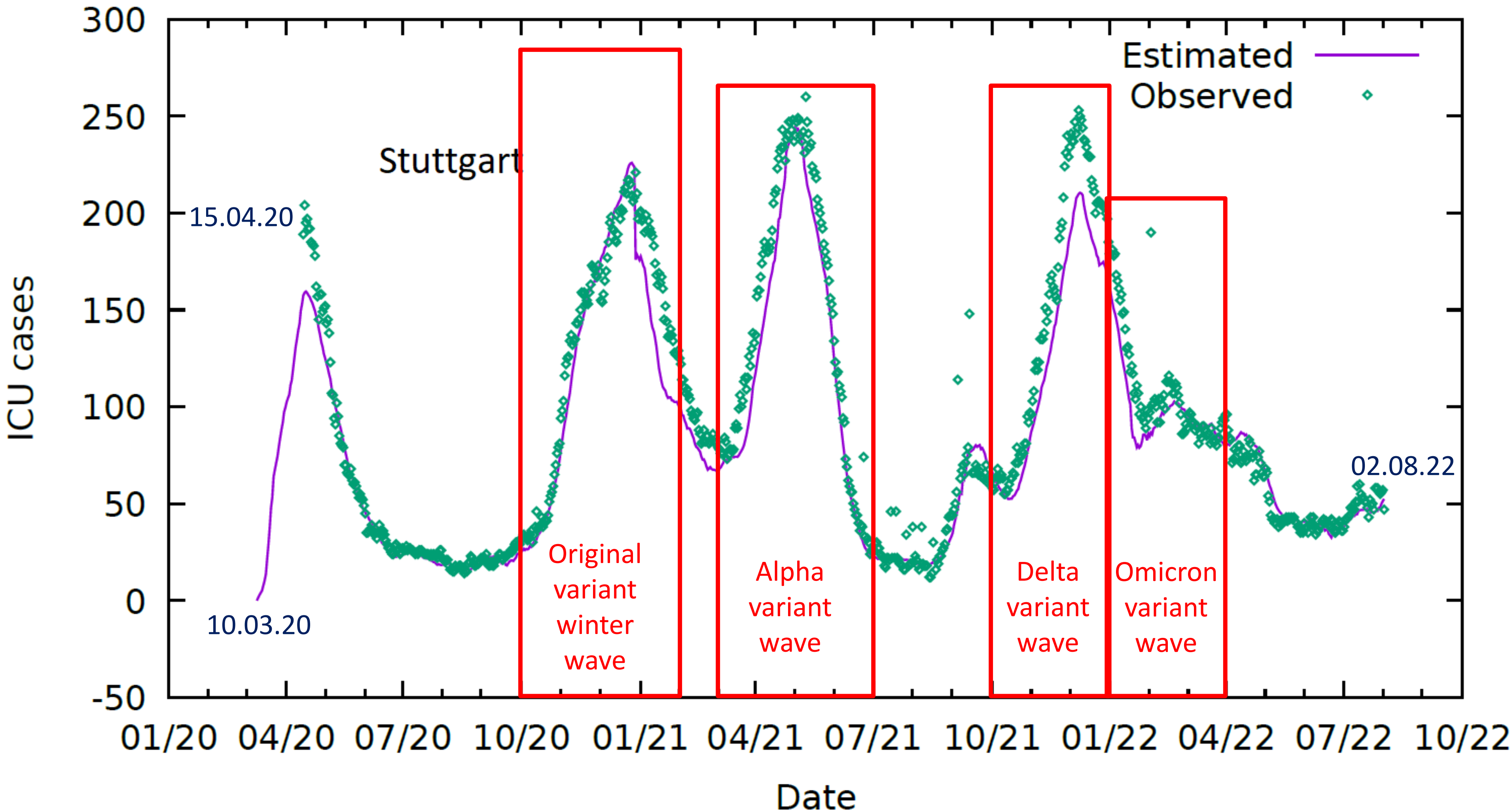
Results – calibration accuracy

German NUTS-2 map

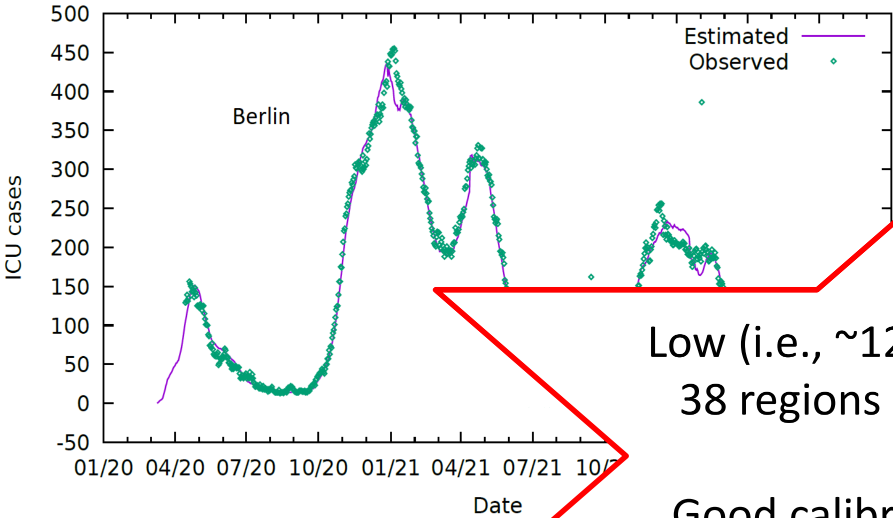


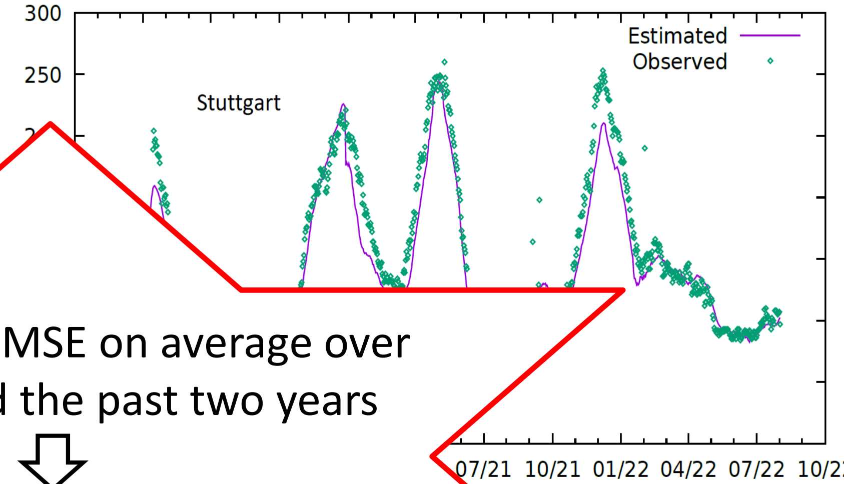
Stuttgart
moderate density: 410/km2

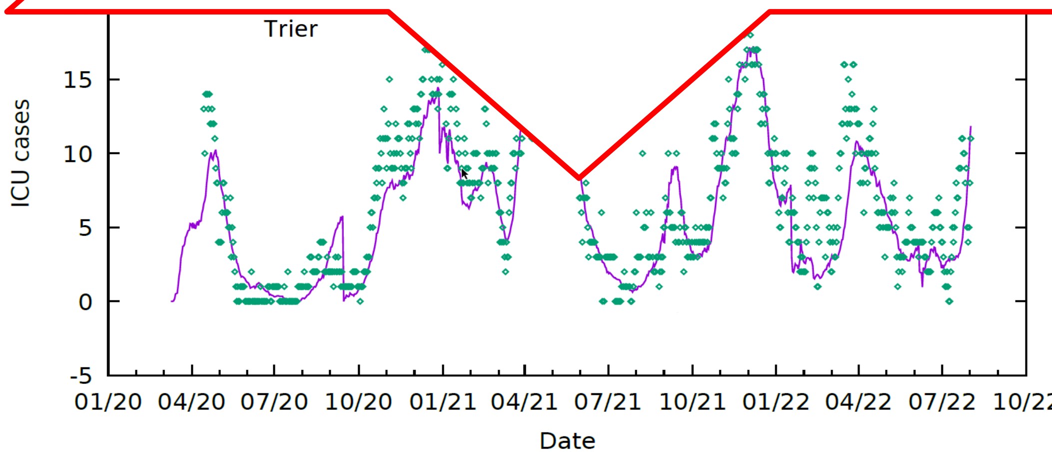
Results – calibration accuracy



Results – calibration accuracy







Low (i.e., ~12) RMSE on average over 38 regions and the past two years

⇩

Good calibration accuracy within a reasonable time – 1 day and half

Conclusion

- Epidemical models containing spatial-/time-varying input parameters are difficult to be calibrated
  - Time-varying → Complexity
    - Segmentation
  - Spatial-varying → High-dimensionality
    - Powerful optimization method

- An approach combining Overlapping Sliding Window (OSW) technique and a hybrid MPI+OpenMP version of the Genetic Algorithm (GA)
  - OSW: segment complex calibration
  - GA: implement the segmented calibration
  - Calibration efficiency is guaranteed
    - Satisfactory calibration accuracy within a reasonable time

Thank You

Questions?

Supported through the project CIRCE
www.hlrs.de/projects/detail/circe