Gap-filling of meteorological data sets

Amber Jacobs

03/07/2023

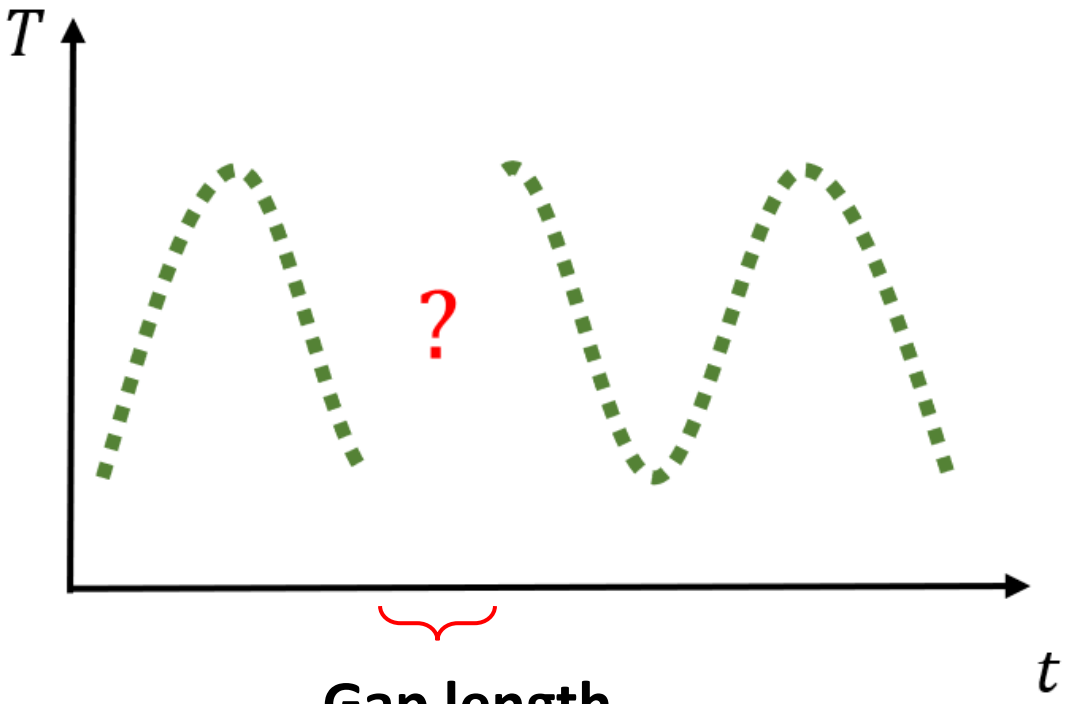
GHENT UNIVERSITY

Content

1. What is a gap?
2. Basic gap-filling techniques
3. ERA5
4. Gap-filling with ERA5
5. Best gap-filling technique + outlook

1. What is a gap?

Definition of a gap



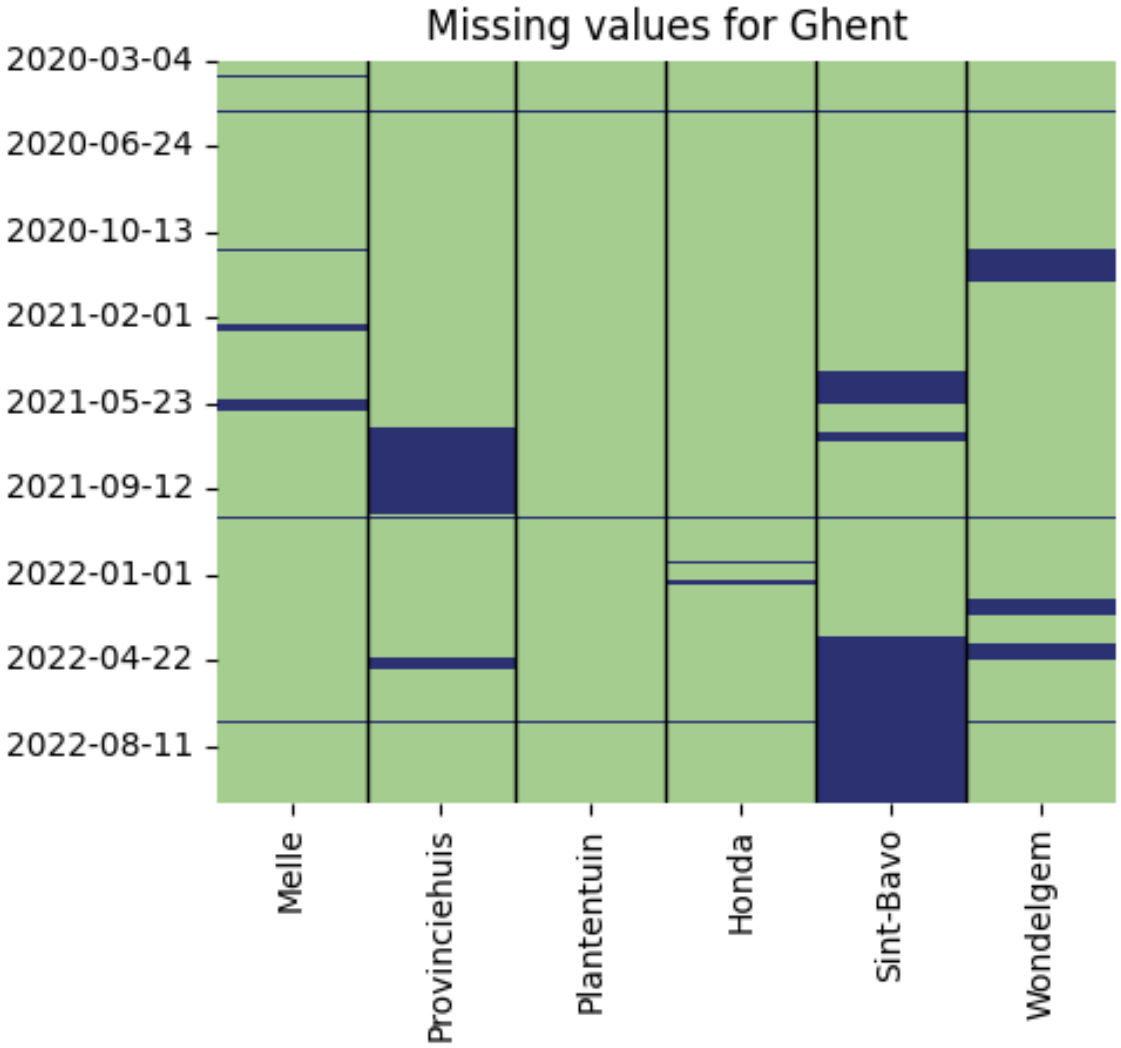
Gap length

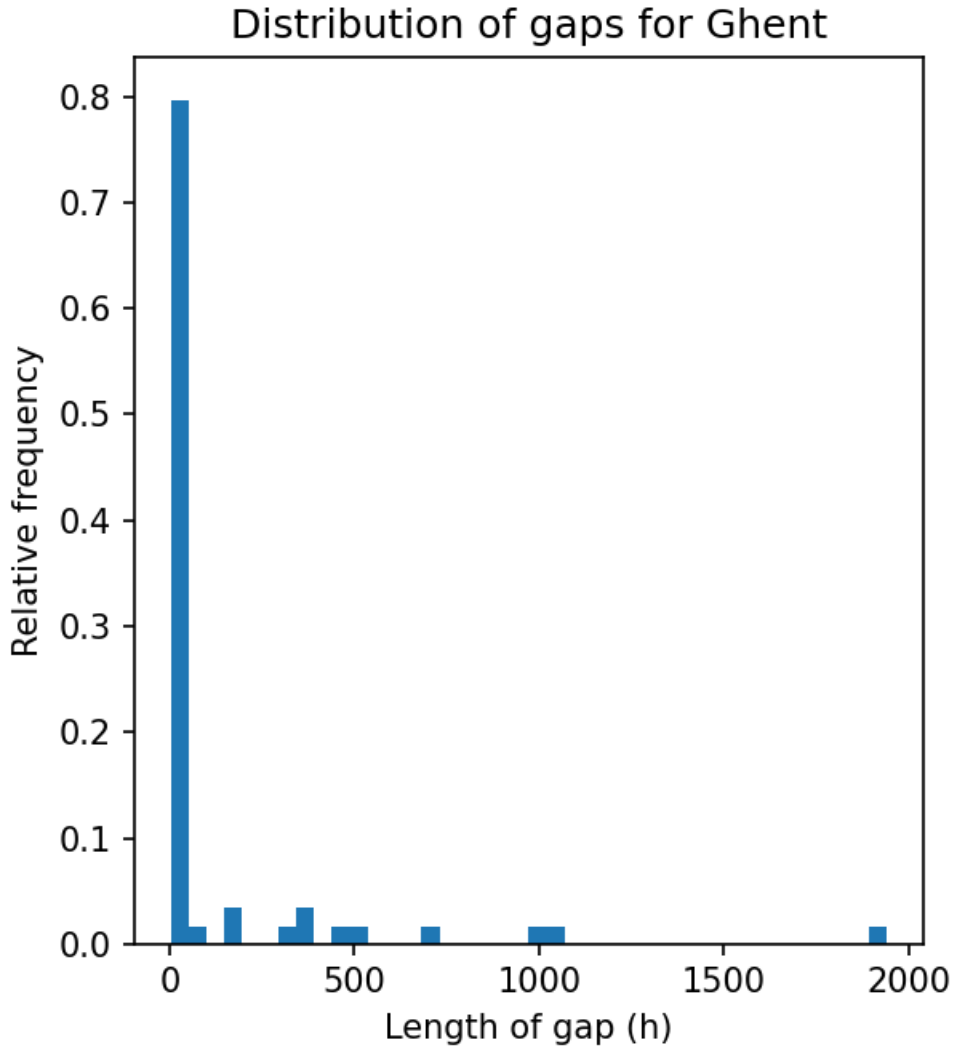
Can be expressed in:

- Number of missing values
- Time

1. What is a gap?

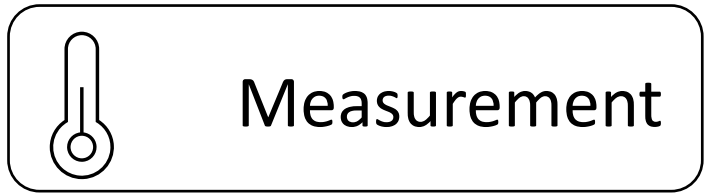
Definition of a gap



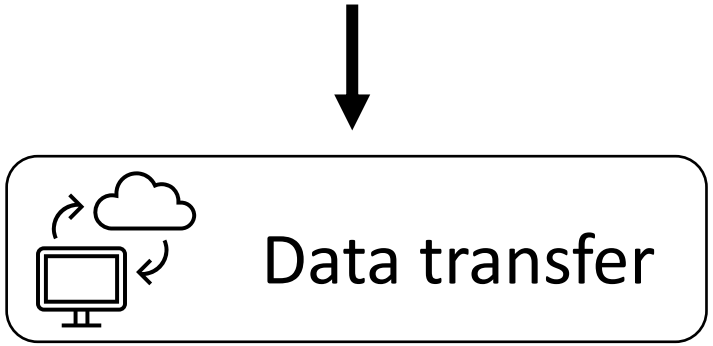


1. What is a gap?

Origin of a gap



- Malfunctioning of device



- Miscommunication

- Delete wrong values

1. What is a gap?

Representation of a gap

Missing timestamp

Datum	Tijd (UTC)	Temperatuur
2022-09-01	0:00:00	18.8
2022-09-01	0:05:00	18.8
2022-09-01	0:10:00	18.8
2022-09-01	0:15:00	18.7
2022-09-01	0:20:00	18.7
2022-09-01	0:25:00	18.7
2022-09-01	0:30:00	18.7
2022-09-01	0:35:00	18.6
2022-09-01	0:40:00	18.6
2022-09-01	0:45:00	18.6
2022-09-01	0:50:00	18.5
2022-09-01	0:55:00	18.4
2022-09-01	5:00:00	15.2
2022-09-01	5:05:00	15.1
2022-09-01	5:10:00	15
2022-09-01	5:15:00	14.9

NaN-value

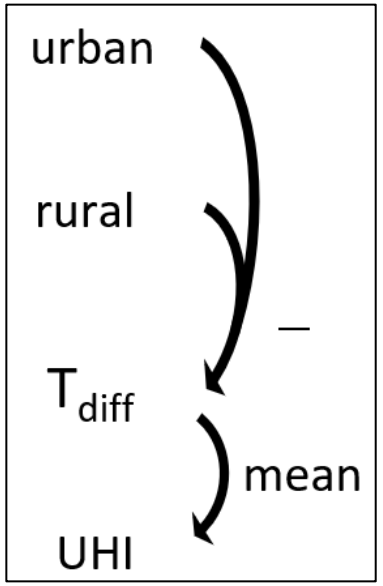
Datum	Tijd (UTC)	Temperatuur	Vochtigheid
2022-09-01	0:00:00	18.8	65
2022-09-01	0:05:00	18.8	65
2022-09-01	0:10:00	18.8	65
2022-09-01	0:15:00	18.7	65
2022-09-01	0:20:00	18.7	65
2022-09-01	0:25:00	18.7	65
2022-09-01	0:30:00	18.7	65
2022-09-01	0:35:00	18.6	65
2022-09-01	0:40:00		65
2022-09-01	0:45:00		65
2022-09-01	0:50:00		65
2022-09-01	0:55:00		65
2022-09-01	1:00:00		65
2022-09-01	1:05:00	18.3	66
2022-09-01	1:10:00	18.2	66
2022-09-01	1:15:00	18.1	66
2022-09-01	1:20:00	18	66
2022-09-01	1:25:00	18	65

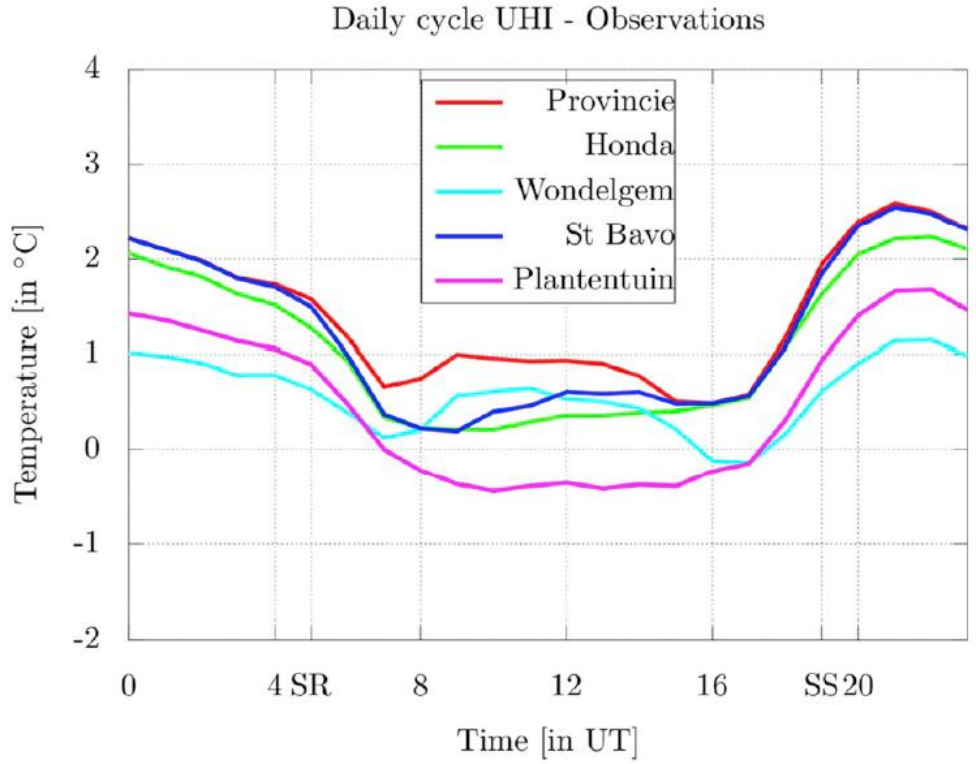
1. What is a gap?

Problem of a gap

- Occurrence of gaps can cause difficulties for analysis:
  - Calculate average: bias can occur
  - Prevent calculation of certain quantity
  - Use data for numerical model

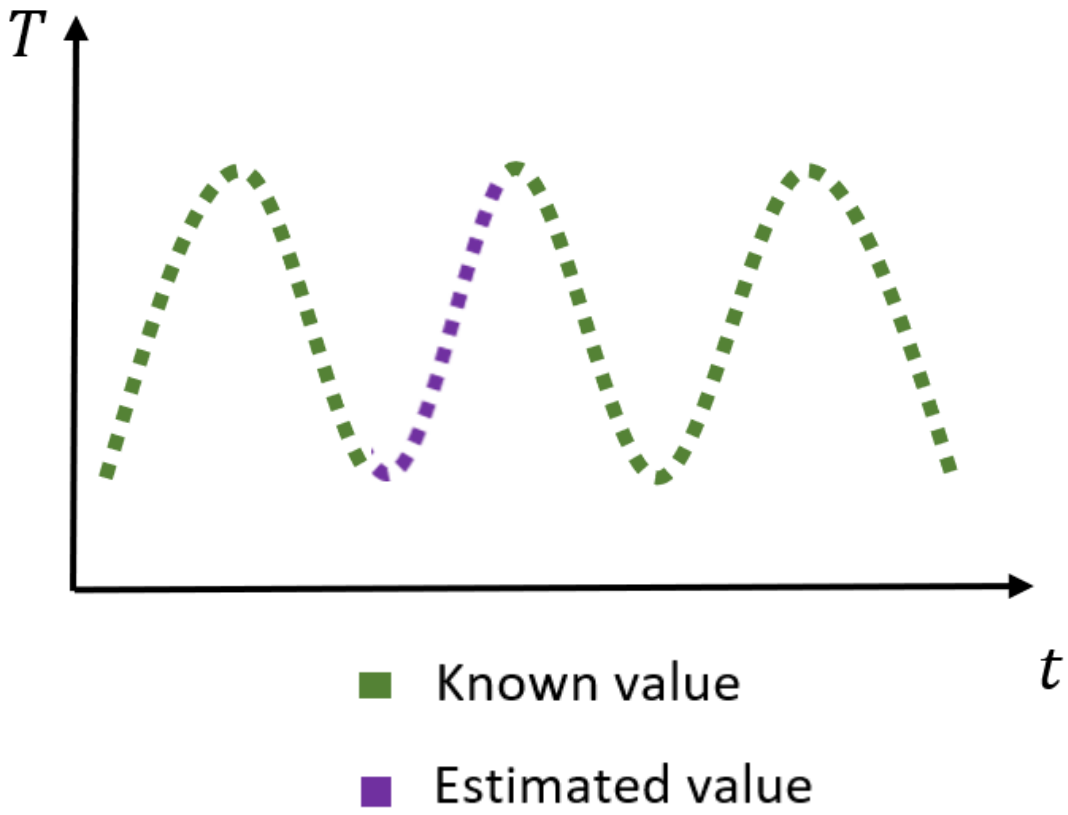
- UHI-effect
  - Urban and rural needed
  - Use same time period

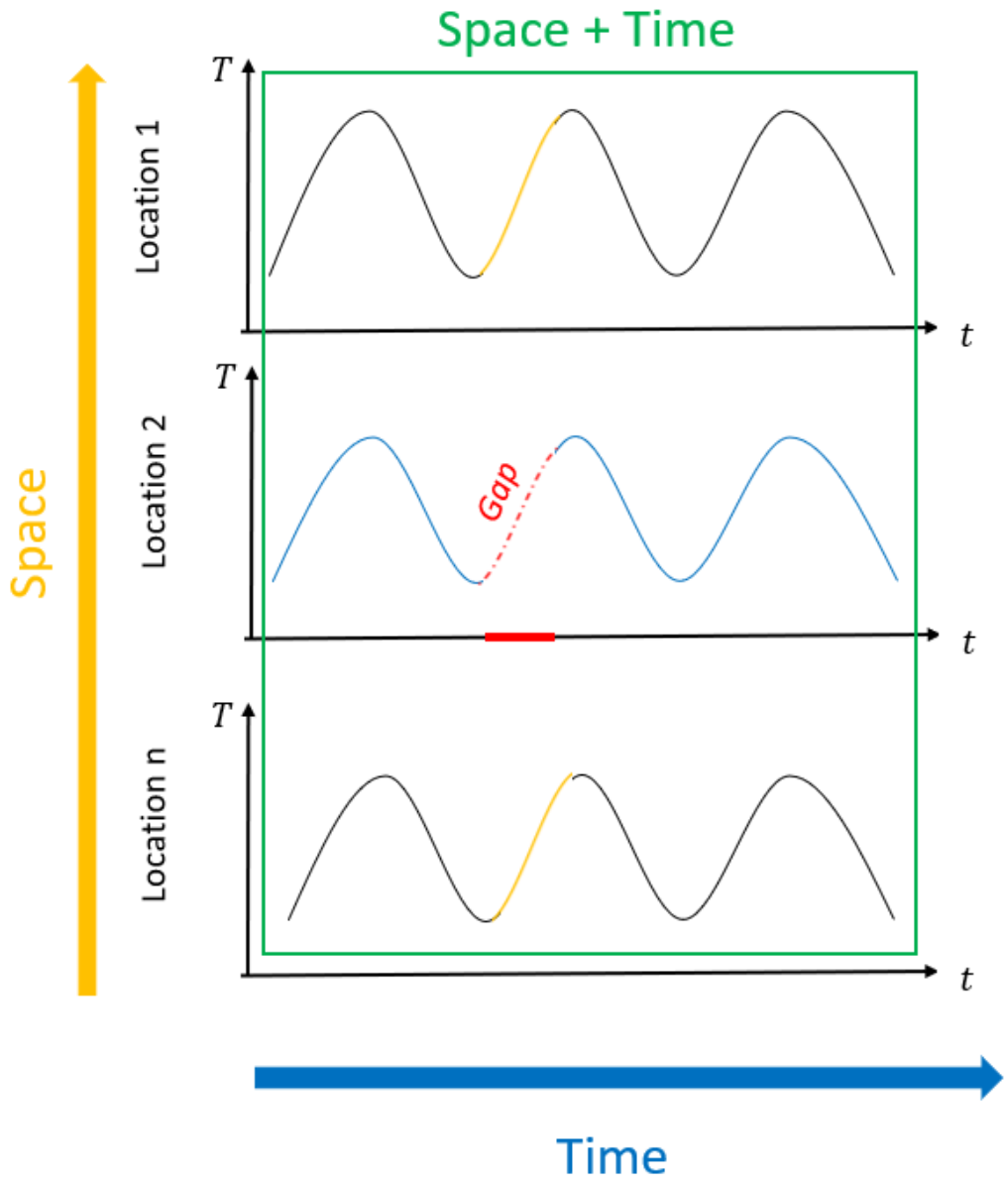




(Caluwaerts et al., 2020)

2. Basic gap-filling techniques

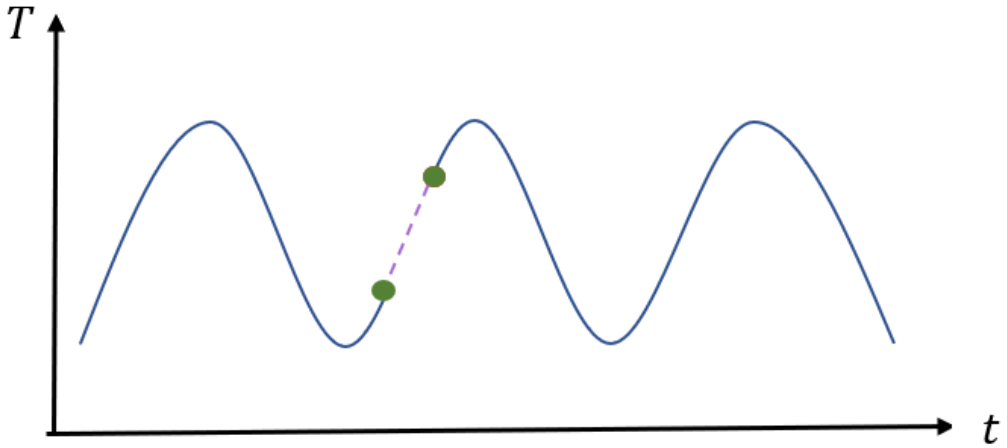


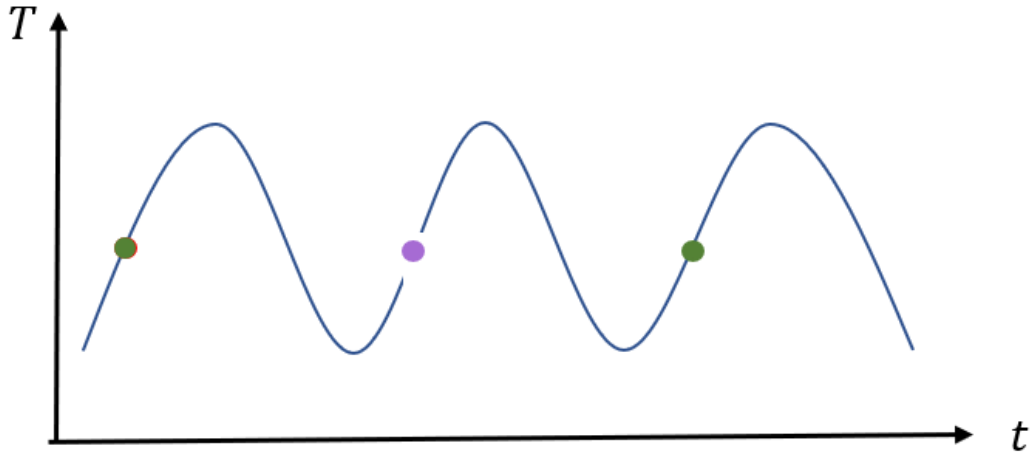


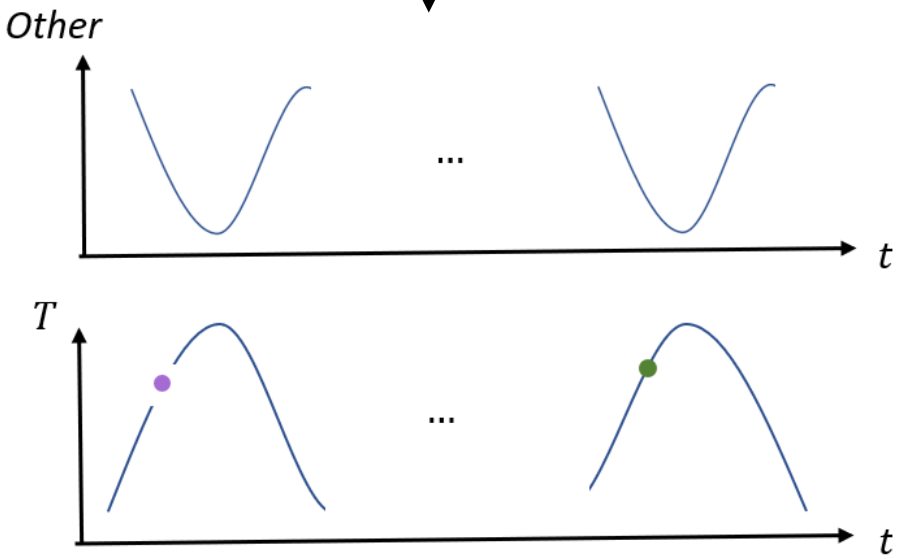
2. Basic gap-filling techniques

Temporal

- Linear interpolation
- Polynomial interpolation
- Diurnal interpolation
- Marginal Distribution Sampling
- …



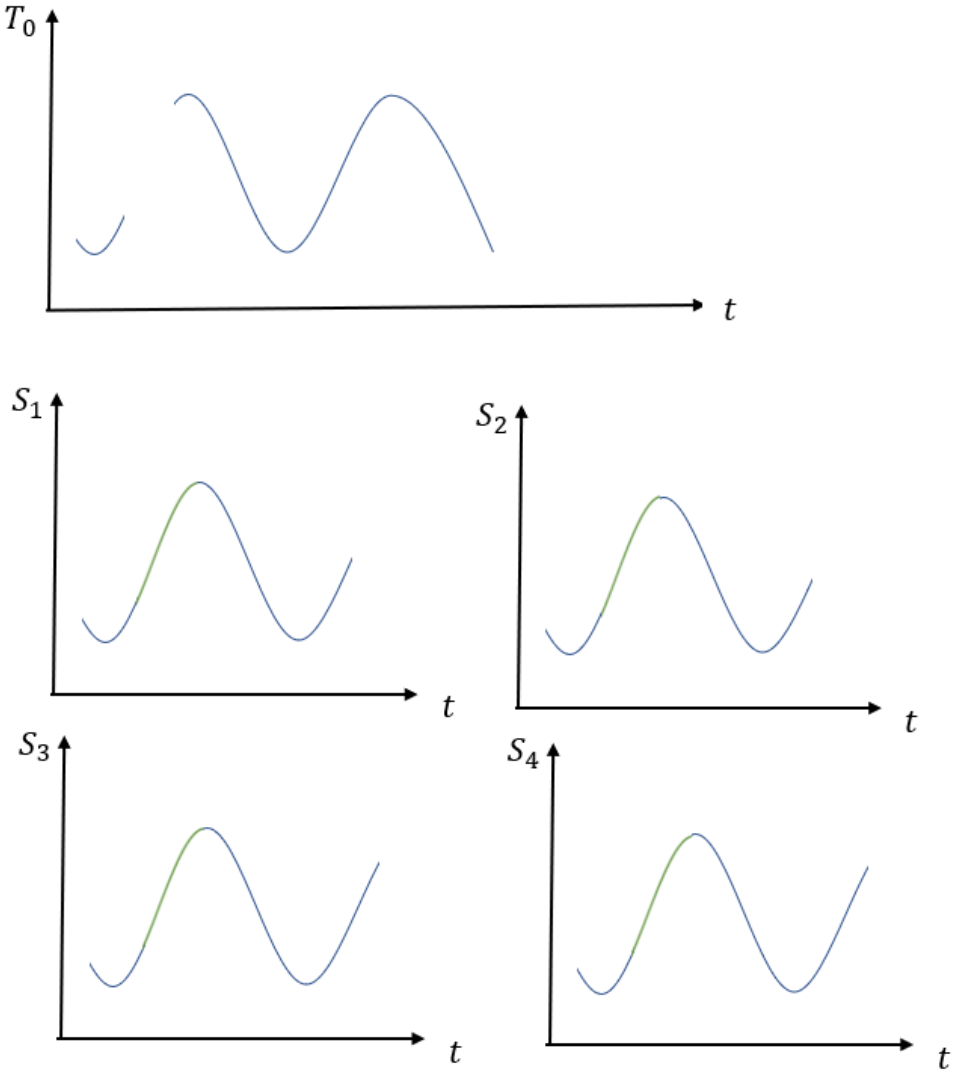




2. Basic gap-filling techniques

Spatial

- Inverse-distance weighting
- Thin-plate splines
- Kriging
- Multiple regressions
- …



2. Basic gap-filling techniques

Spatiotemporal

- Empirical Orthogonal Functions
- …

Stations

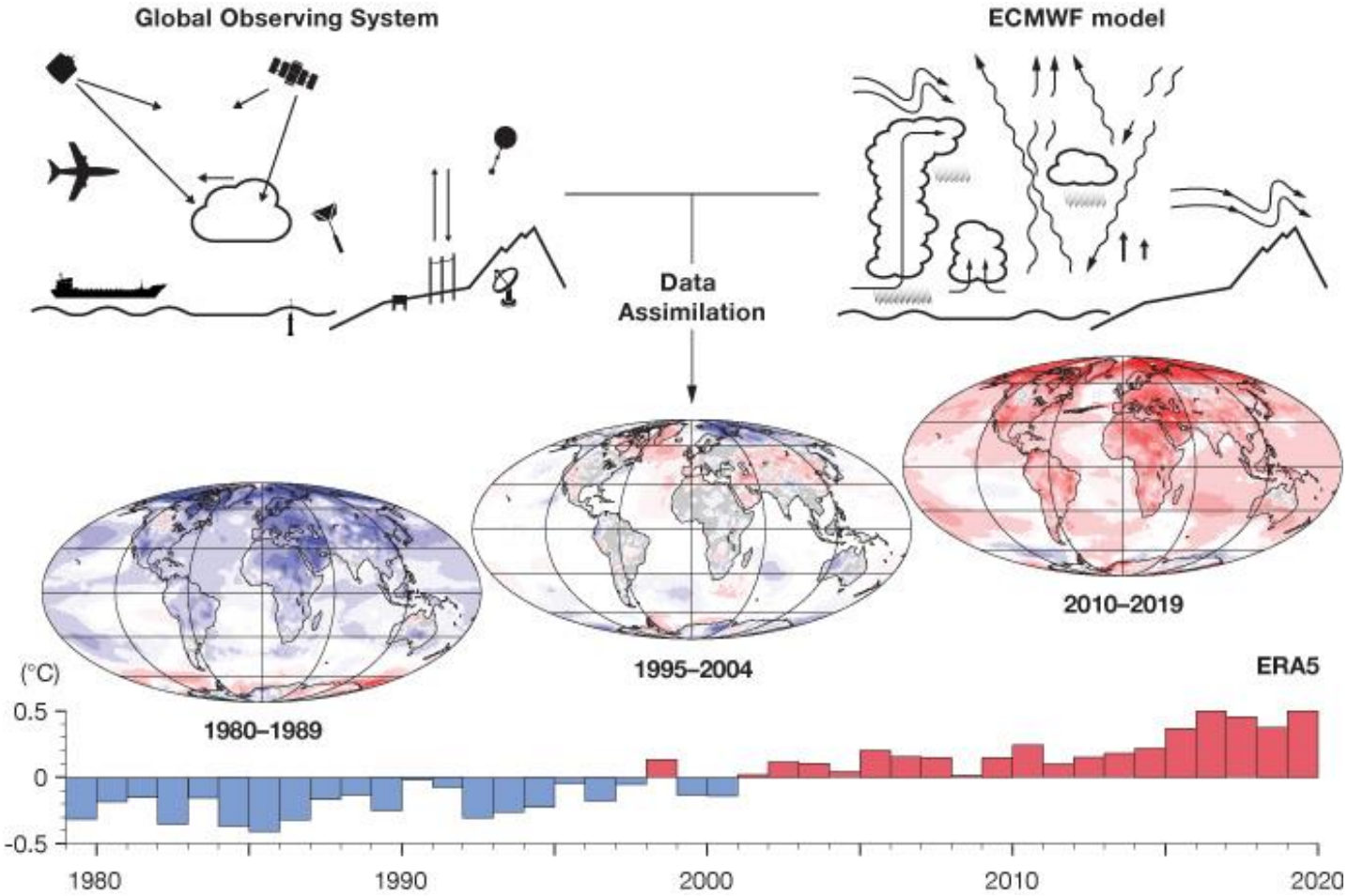
$$A = \begin{bmatrix} T_{11} & \cdots & T_{1s} \\ \vdots & \ddots & \vdots \\ T_{t1} & \cdots & T_{ts} \end{bmatrix}$$

Time

3. ERA5

What is ERA5?

- Created by ECMWF
- Complete
- Homogeneous
  - Hourly
  - Grid (0.25° x 0.25°)
- Bias present



ECMWF, https://www.ecmwf.int/en/about/media-centre/focus/2020/fact-sheet-reanalysis

3. ERA5

What is ERA5?

- Created by ECMWF
- Complete
- Homogeneous
  - Hourly
  - Grid (0.25°)
- Bias present
  - Especially for urban locations!

3. ERA5

Bias of ERA5 temperature

- Observations vs ERA5
- 2 errors:

$$MBE = \sum_{t=1}^{N} \frac{\mathrm{ERA5}_t - \mathrm{OBS}_t}{N}$$

$$MSE = \sum_{t=1}^{N} \frac{(\mathrm{ERA5}_t - \mathrm{OBS}_t)^2}{N}$$

- Bias depends on:
  - Hour
  - Season
  - Daily temperature range

3. ERA5

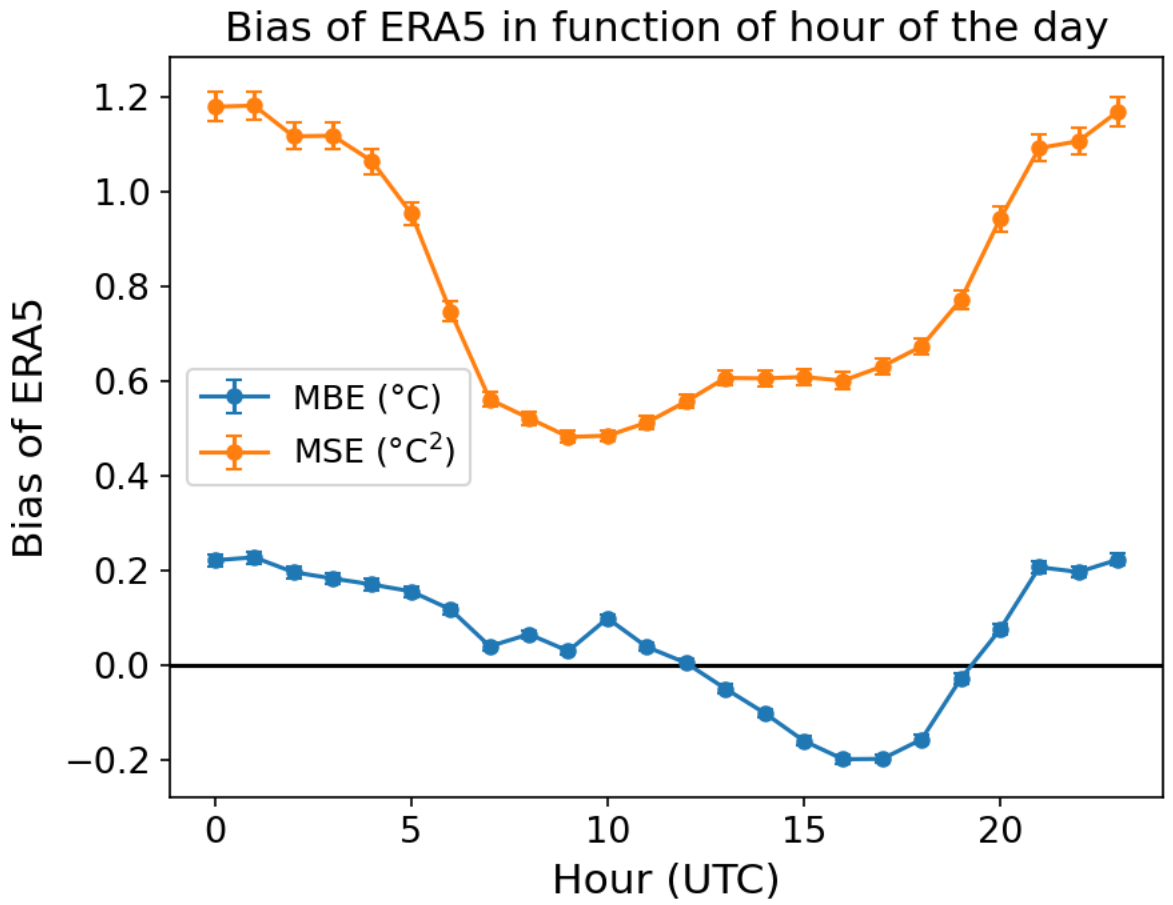
Bias of ERA5 temperature

- Observations vs ERA5
- 2 errors:

$$MBE = \sum_{t=1}^{N} \frac{\mathrm{ERA5}_t - \mathrm{OBS}_t}{N}$$

$$MSE = \sum_{t=1}^{N} \frac{(\mathrm{ERA5}_t - \mathrm{OBS}_t)^2}{N}$$

- Bias depends on:
  - **Hour**
  - Season
  - Daily temperature range



3. ERA5

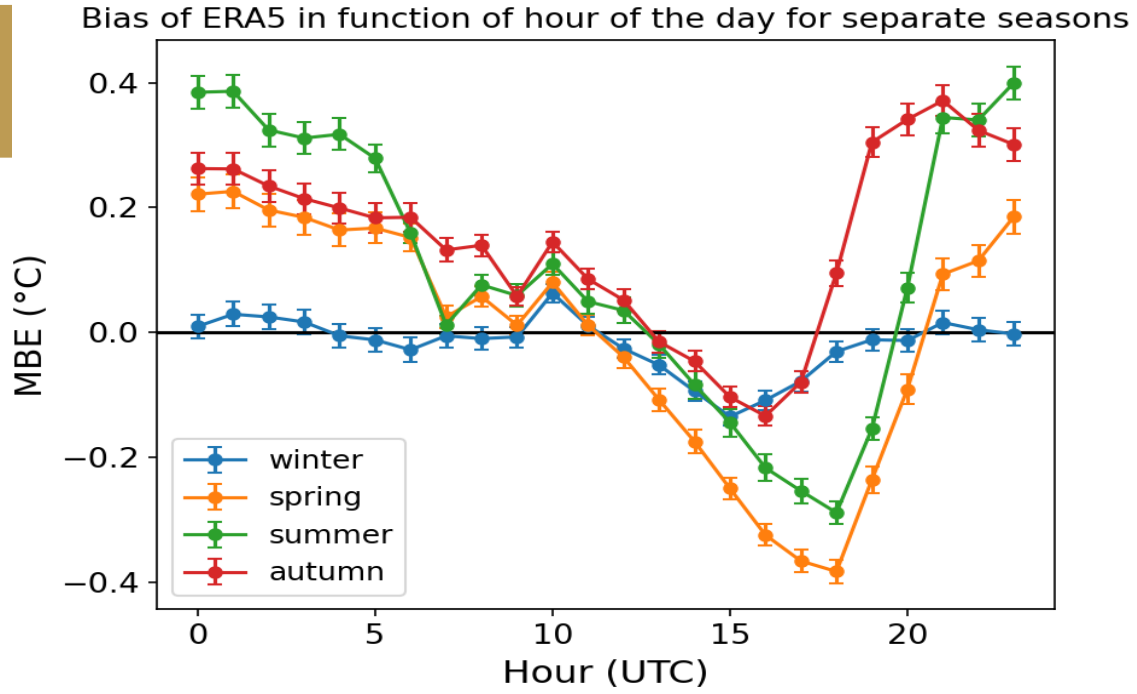
Bias of ERA5 temperature

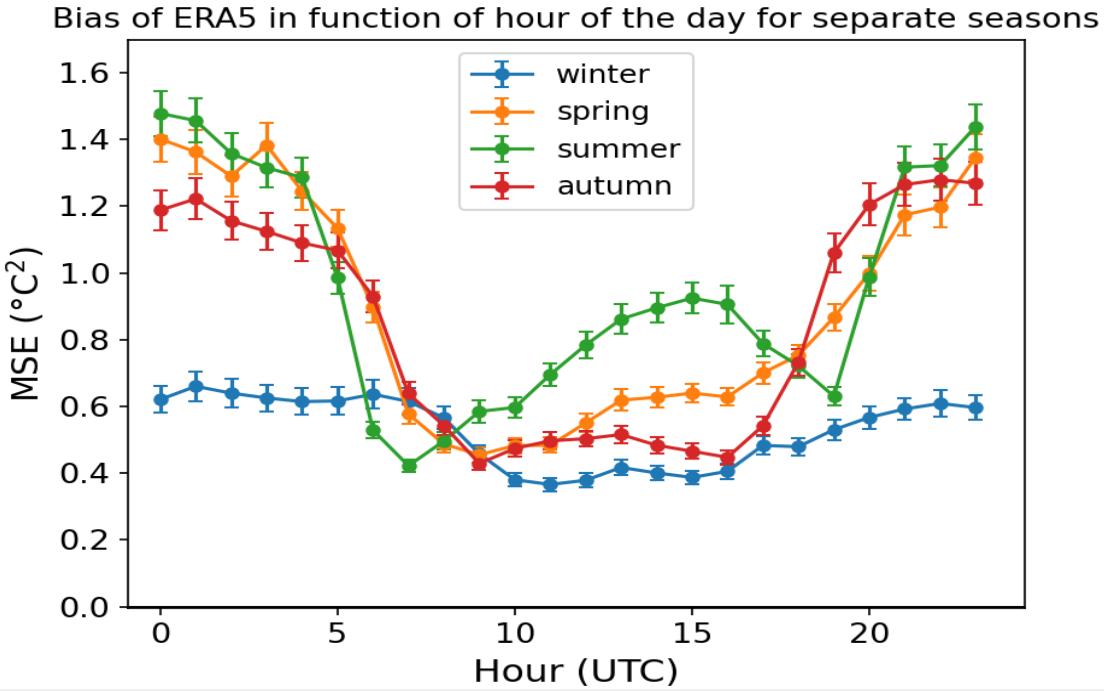
- Observations vs ERA5
- 2 errors:

$$MBE = \sum_{t=1}^{N} \frac{\mathrm{ERA5}_t - \mathrm{OBS}_t}{N}$$

$$MSE = \sum_{t=1}^{N} \frac{(\mathrm{ERA5}_t - \mathrm{OBS}_t)^2}{N}$$

- Bias depends on:
  - Hour
  - **Season**
  - Daily temperature range





3. ERA5

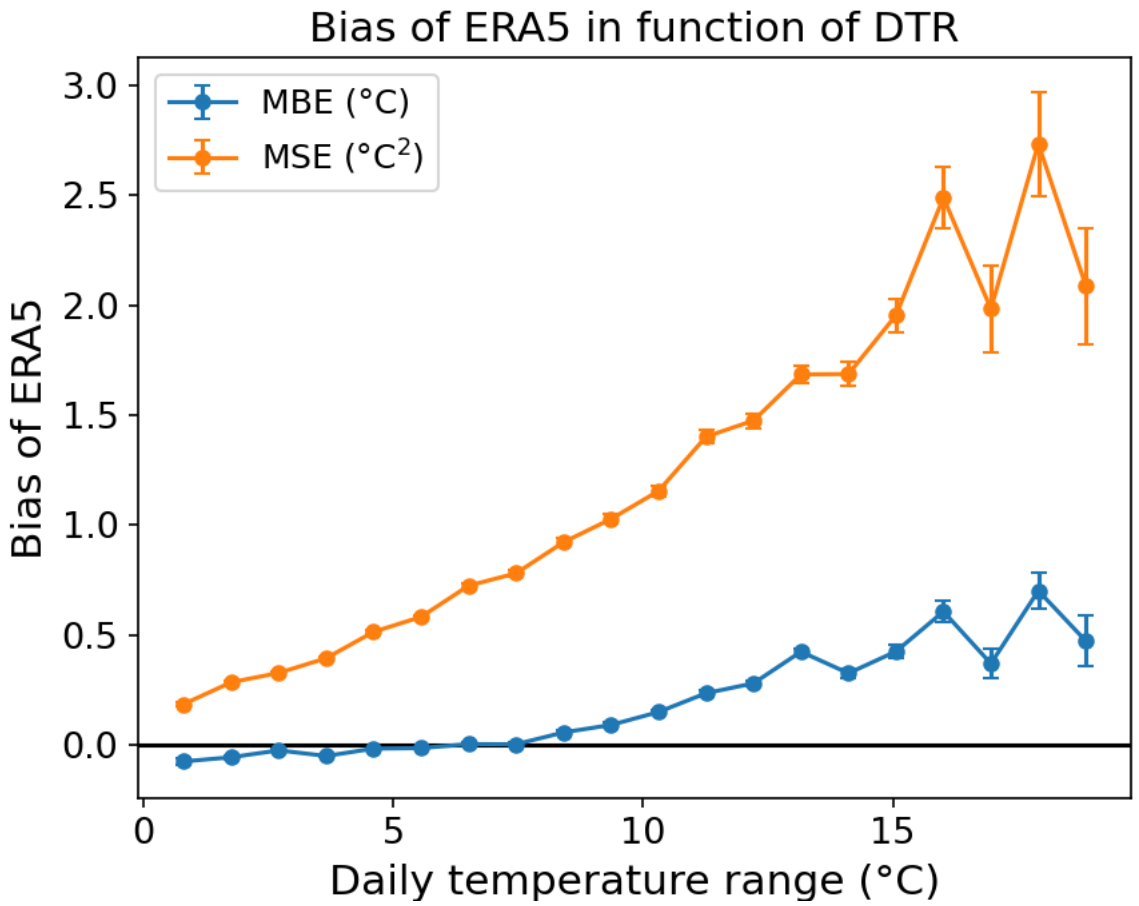
Bias of ERA5 temperature

- Observations vs ERA5
- 2 errors:

$$MBE = \sum_{t=1}^{N} \frac{\mathrm{ERA5}_t - \mathrm{OBS}_t}{N}$$

$$MSE = \sum_{t=1}^{N} \frac{(\mathrm{ERA5}_t - \mathrm{OBS}_t)^2}{N}$$

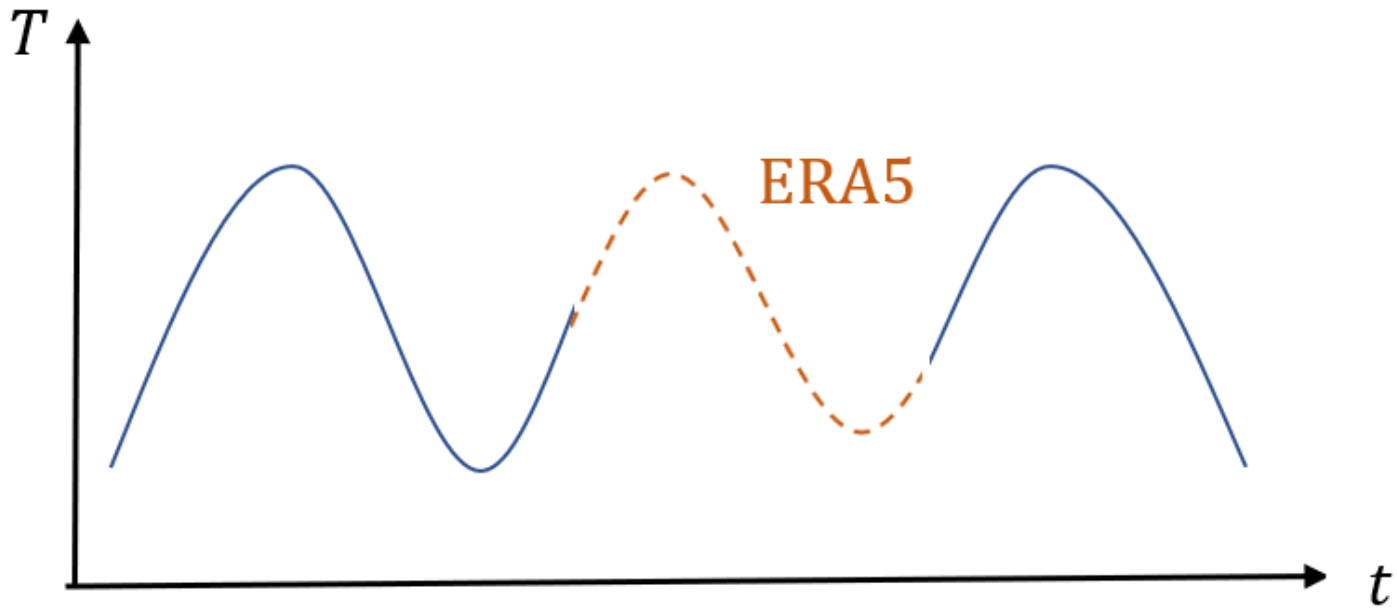
- Bias depends on:
  - Hour
  - Season
  - **Daily temperature range**



4. Gap-filling with ERA5

Different techniques

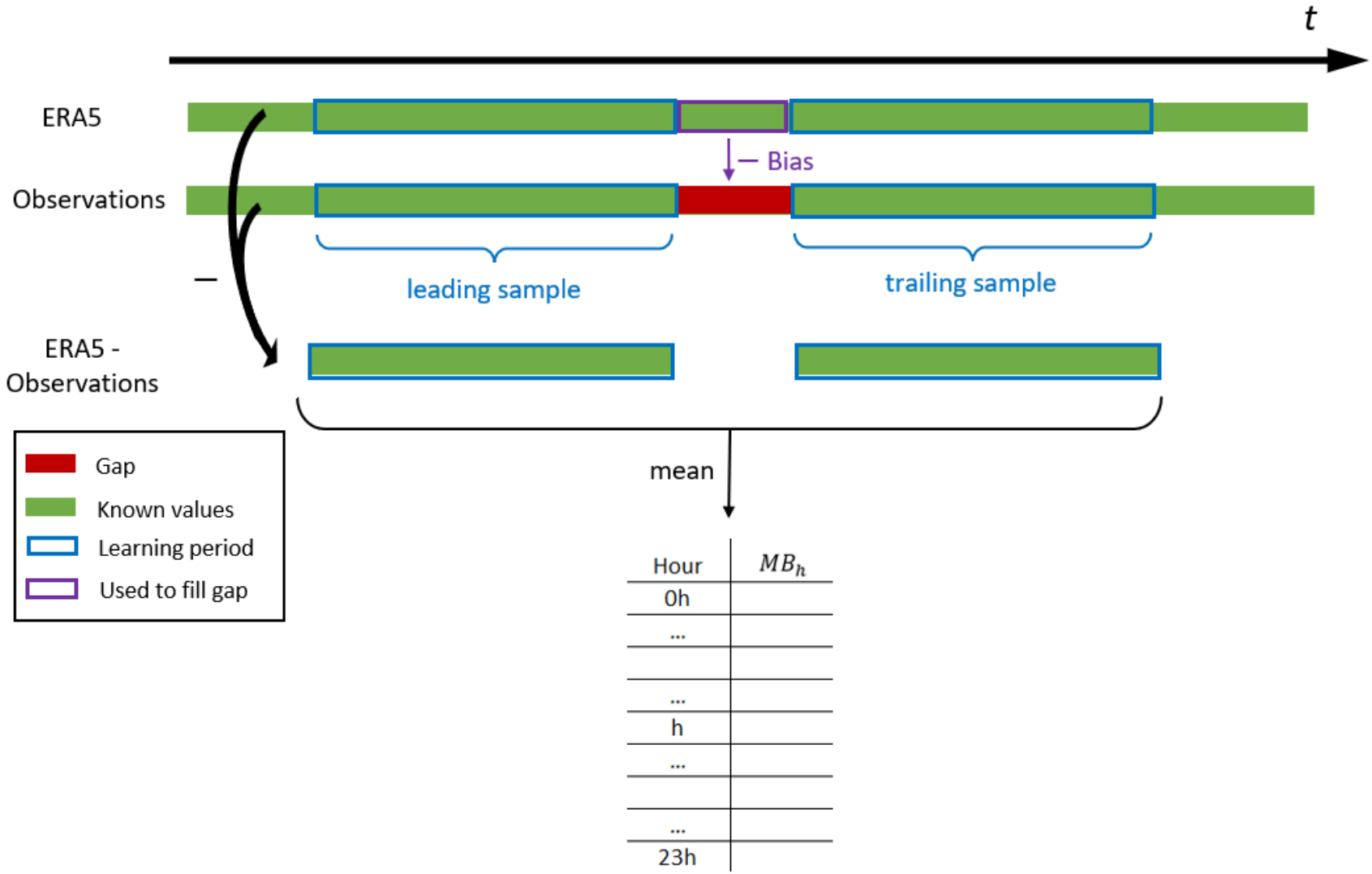
1. Original ERA5
2. Debiased ERA5: mean bias



4. Gap-filling with ERA5

Different techniques

1. Original ERA5
2. Debiased ERA5: mean bias



4. Gap-filling with ERA5

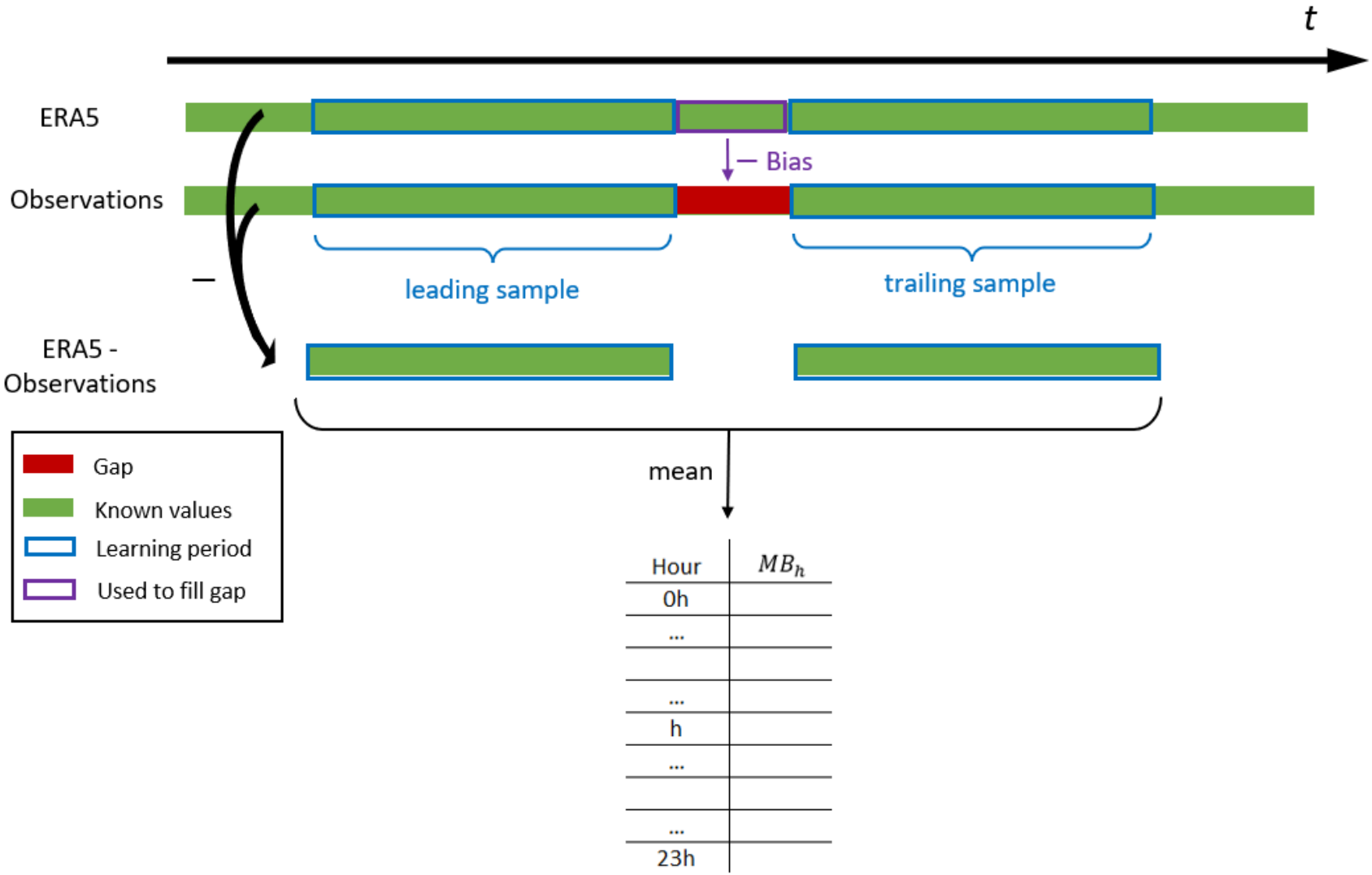
Different techniques

1. Original ERA5
2. Debiased ERA5: mean bias

Other:

- Debiased ERA5: linear regression
- Debiased ERA5: machine learning

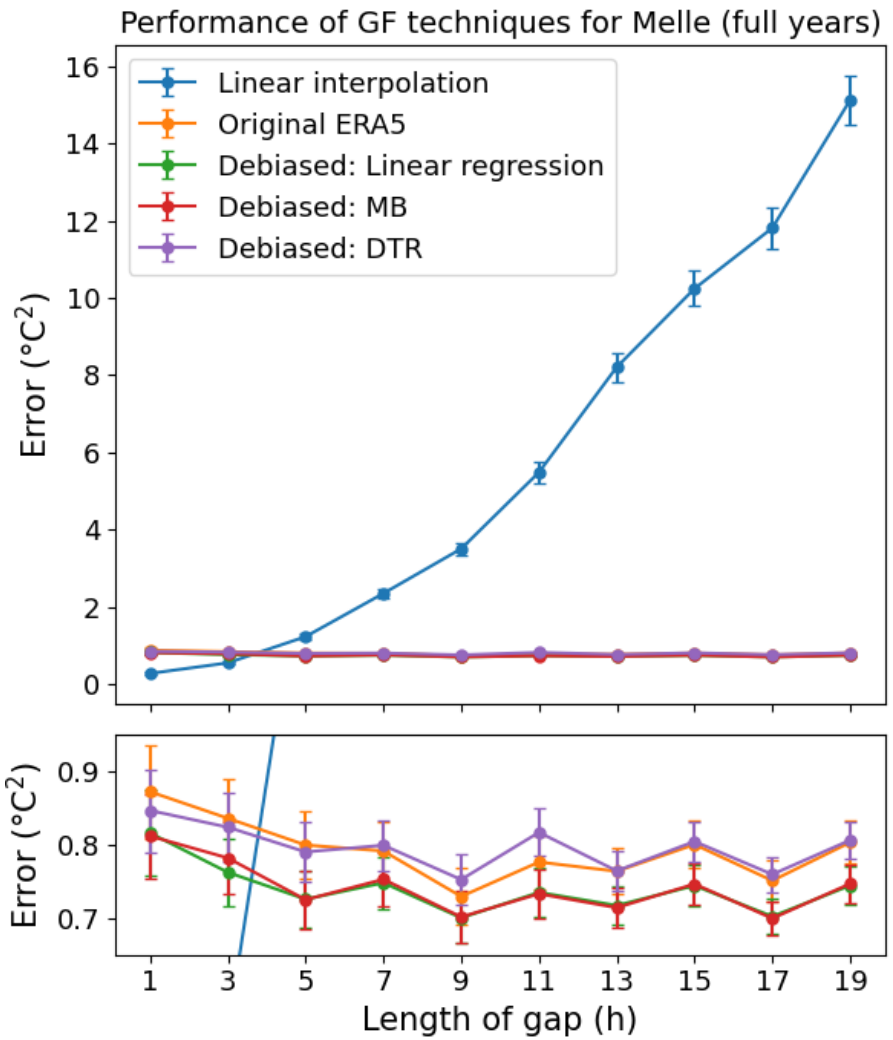
…



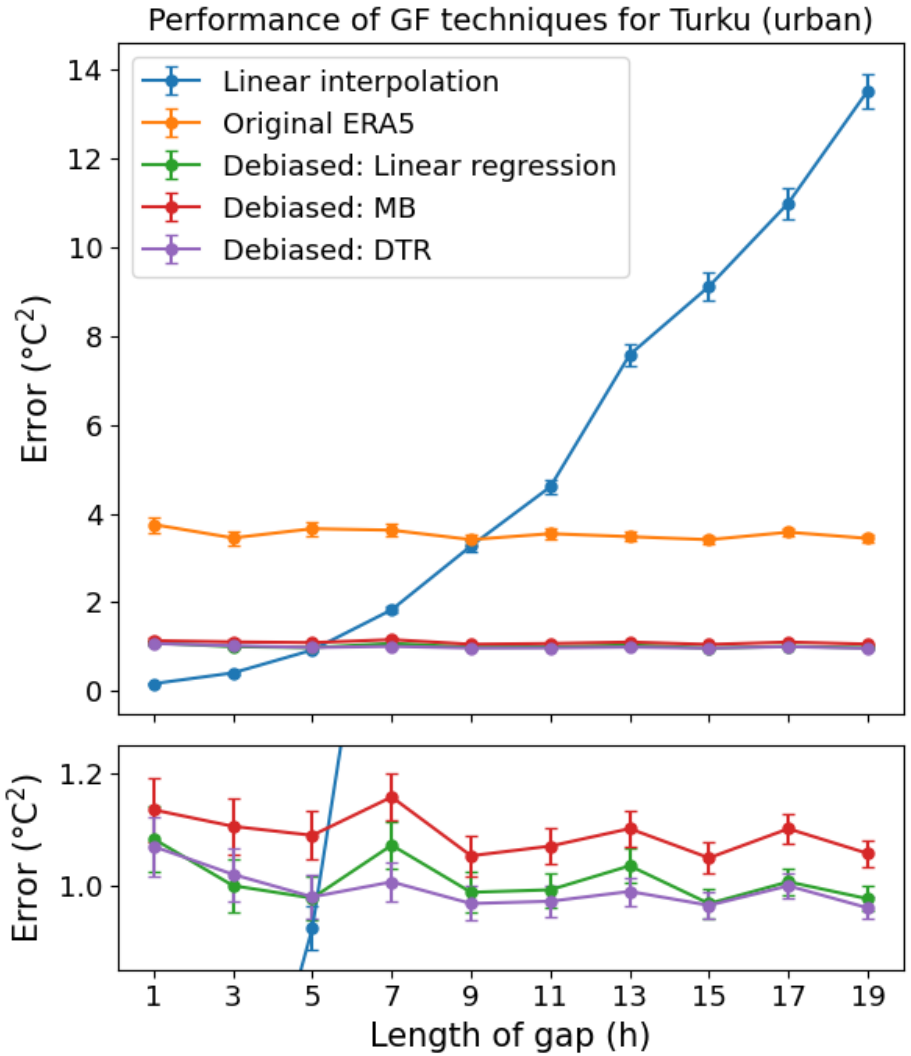
4. Gap-filling with ERA5

Performance evaluation

RURAL



URBAN



5. Best gap-filling technique + outlook

Best gap-filling technique depends on:

- Available observations
- Characteristics of data
  - Urban vs rural
- Characteristics of the gap
  - Gap length

During the exercise:

- Selection of gap-filling techniques:
  1. Linear interpolation
  2. Debiased ERA5: mean bias
  3. Hybrid (mix of previous two)
- Best gap-filling technique for small/large gaps?