Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

Community climatology

Combining official observations, crowd-sourced observations and covariates to obtain high-resolution gridded climate data sets

Jouke de Baar, Gerard van der Schrier, Theo Brandsma (KNMI)

Budapest, 11 May 2023 (online)

Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

Introduction

Methods

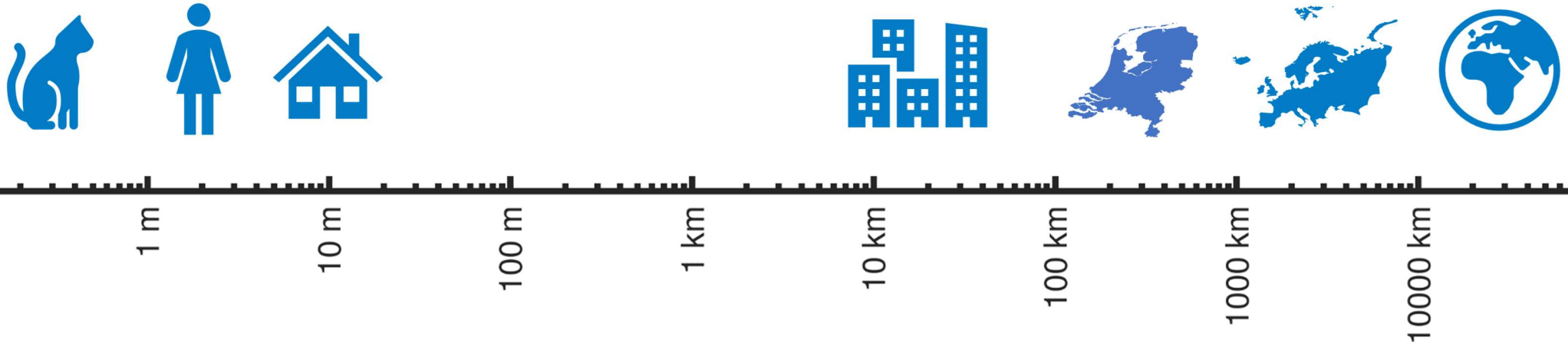
Results

Next steps

Introduction

The scale we live on

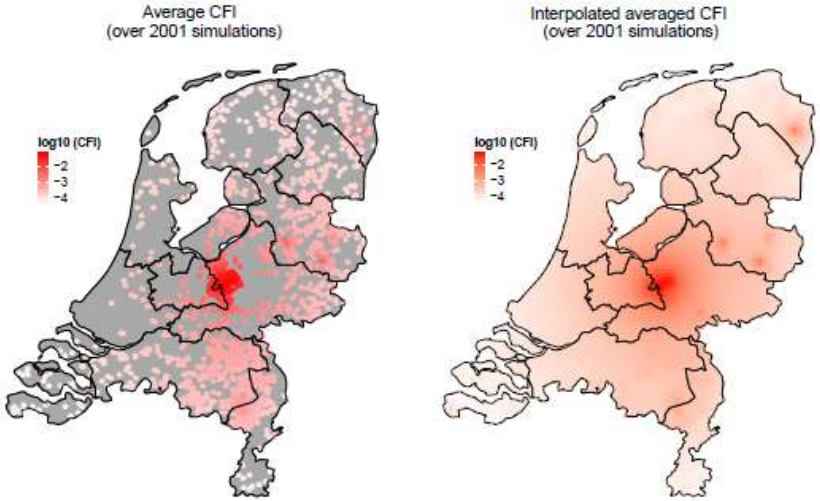
weather & impact warnings



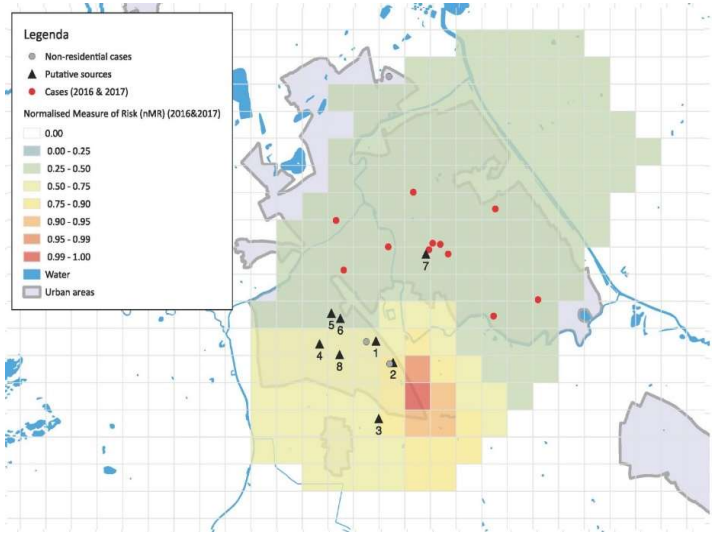
Introduction

Question from: National Institute for Public Health and the Environment

Can KNMI provide historical weather maps at 1 km X 1 km X 1 hour resolution?



Risk mapping for airborne infectious diseases



Source detection of airborne infectious diseases

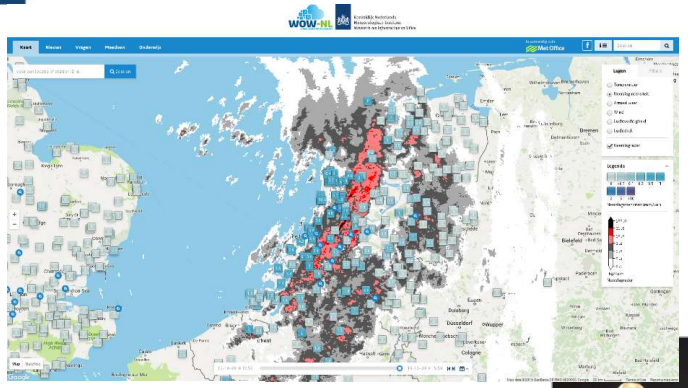
Source: RIVM, https://doi.org/10.3201%2Feid2410.180906

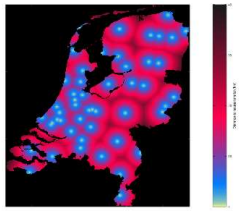
Introduction

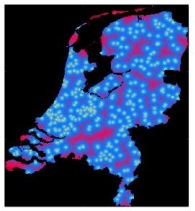
- Initiated by UK Met Office
- WOW-NL portal launched 2015
- Teaching Material and Outreach
- Smart Cities – Senshagen Zwolle
- User support by back office
- Exchange of best practices with other WOW partners

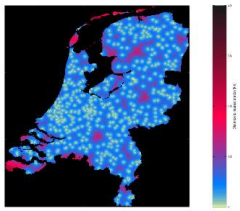
We are very grateful
for all the efforts from
the WOW community
in providing data!

Weather Observation Website: WOW





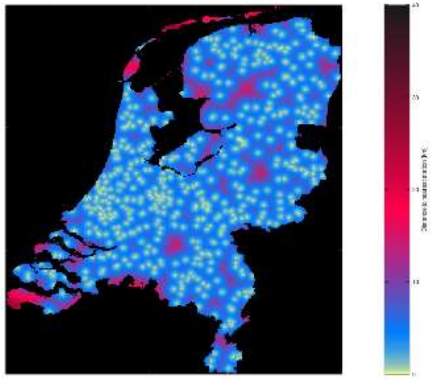


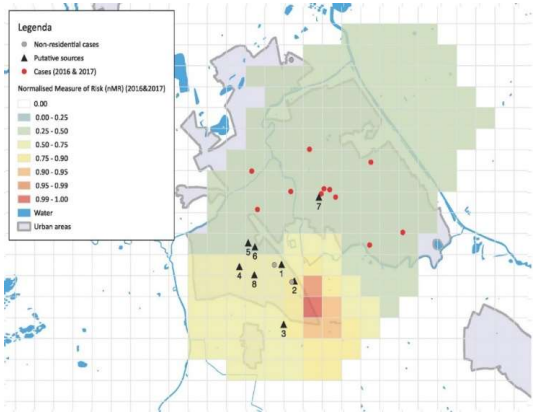


Introduction

data by community,

results for community





KNMI

Source: RIVM, https://doi.org/10.3201%2Feid2410.180906

Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

Introduction

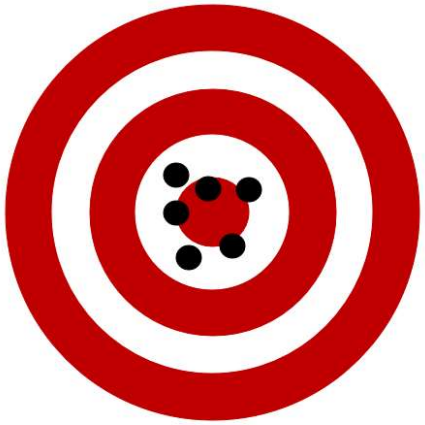
Methods

Results

Next steps

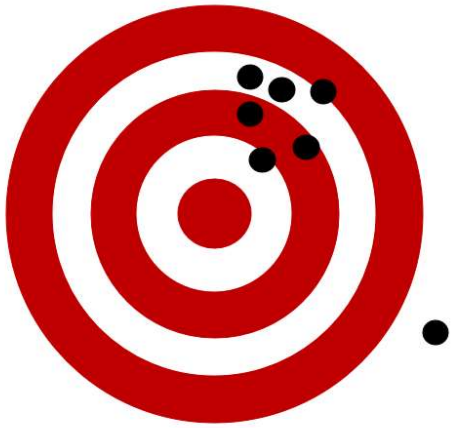
Methodology

"accurate"
(≠ reality)



noise

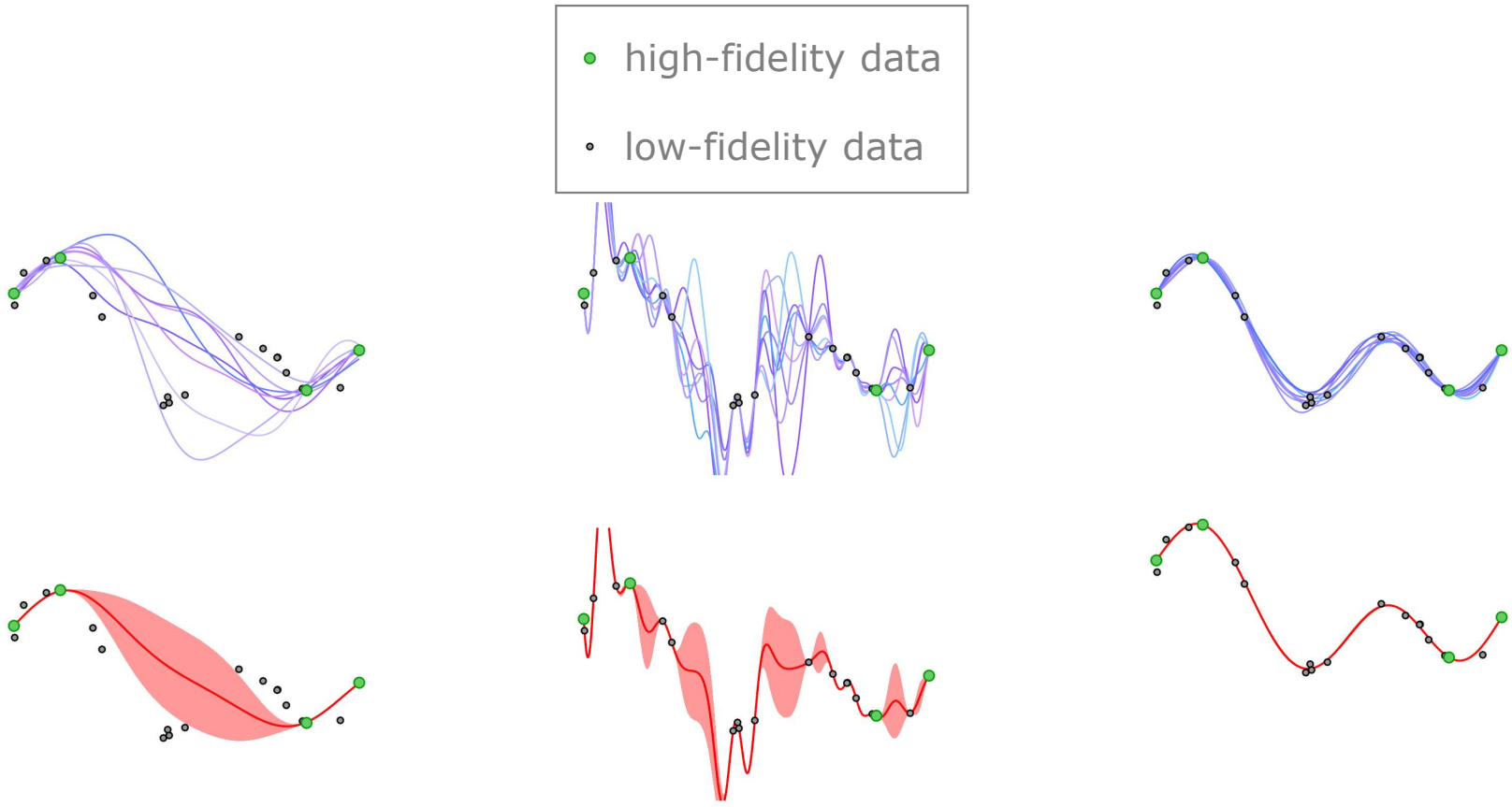
bias + noise



bias + noise
+ outlier

Methodology

Multi-fidelity Gaussian process regression



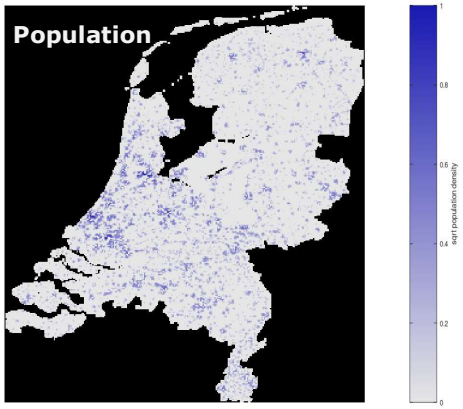
ignoring low-fi data

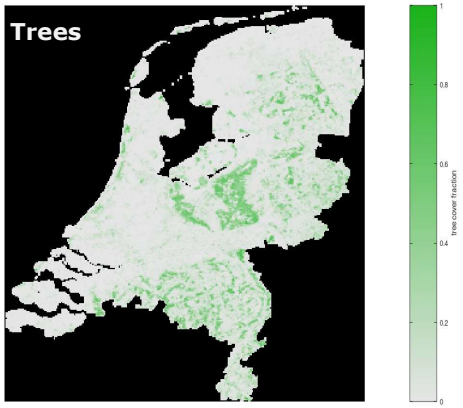
exact interpolation

regression with noise treatment

Methodology

High-resolution covariates





Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

Introduction

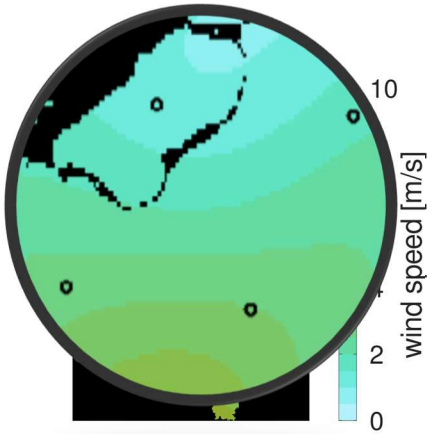
Methods

Results

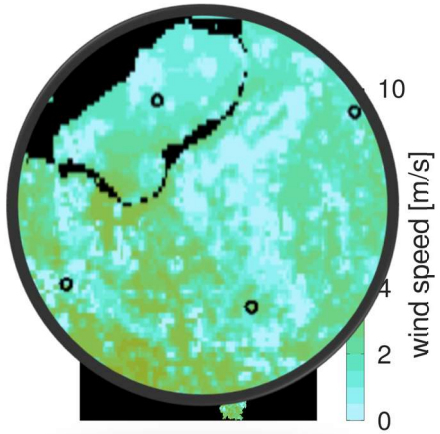
Next steps

Results

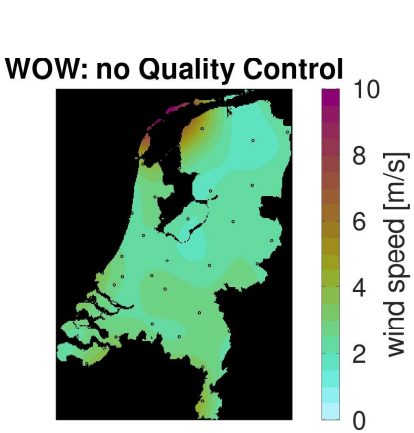
Wind speed (29 June 2021, 8am-9am)



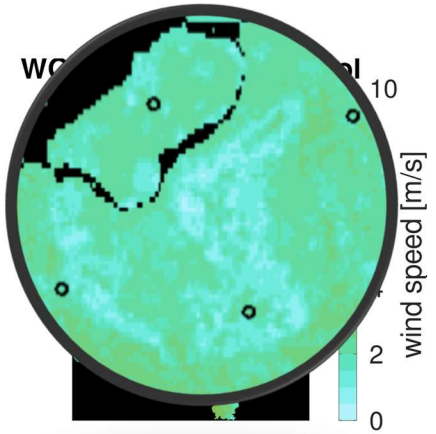
KNMI



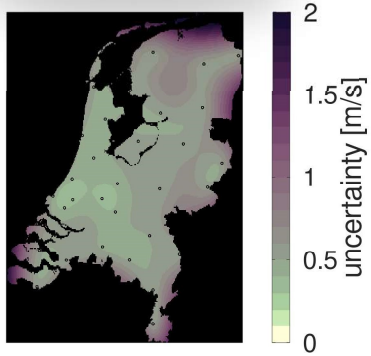
KNMI + cov



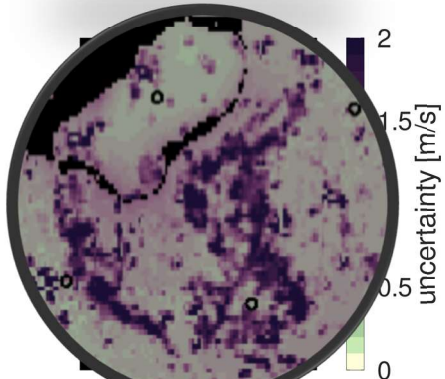
KNMI + WOW



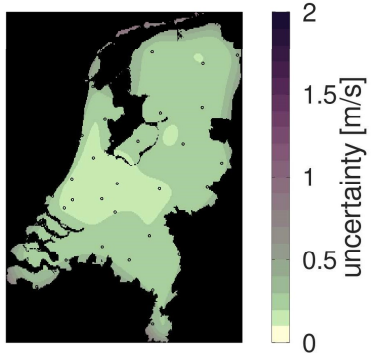
KNMI + WOW + cov



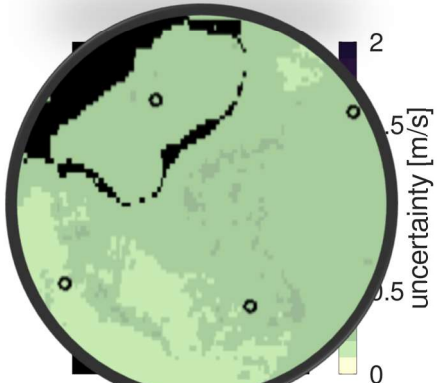
KNMI



KNMI + cov



KNMI + WOW

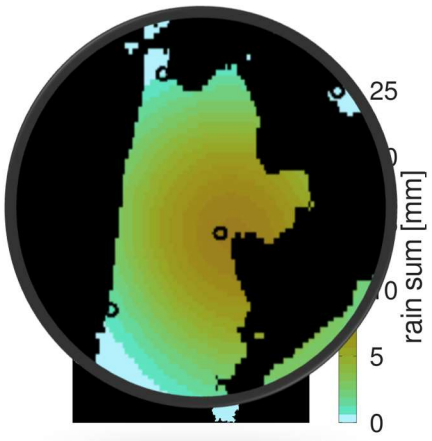


KNMI + WOW + cov

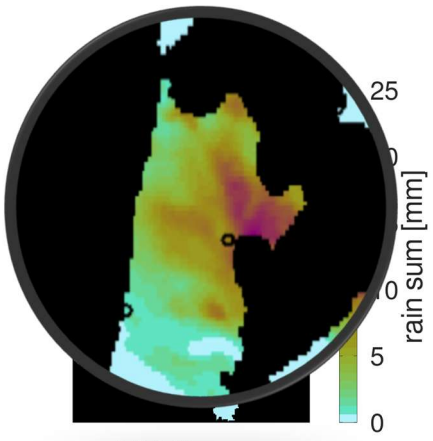
Source: De Baar, Garcia-Marti, Van der Schrier, EMS2022-53

Results

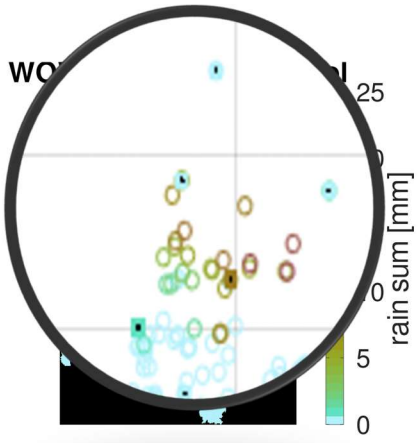
Rain (29 June 2021, 8am-9am)



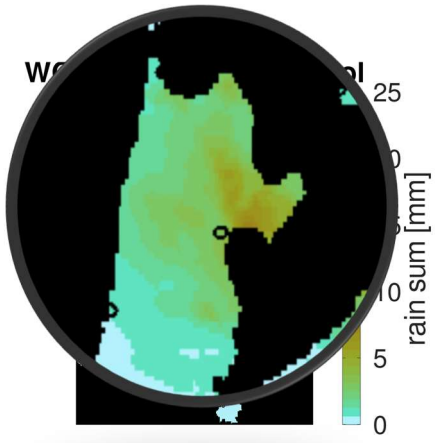
KNMI



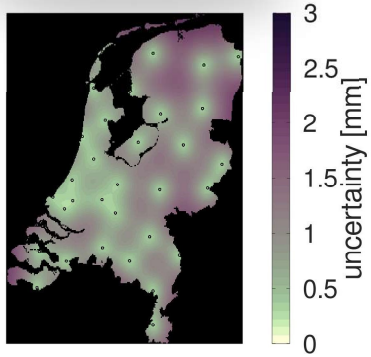
KNMI + cov



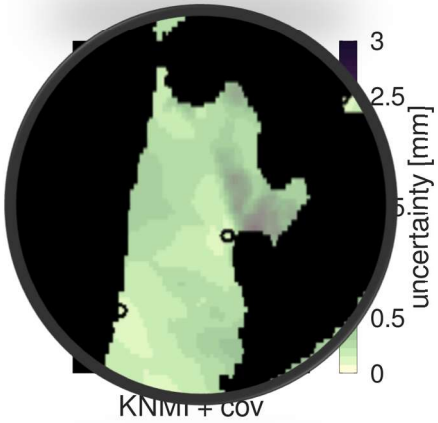
KNMI + WOW



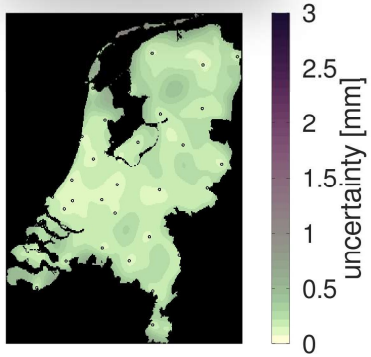
KNMI + WOW + cov



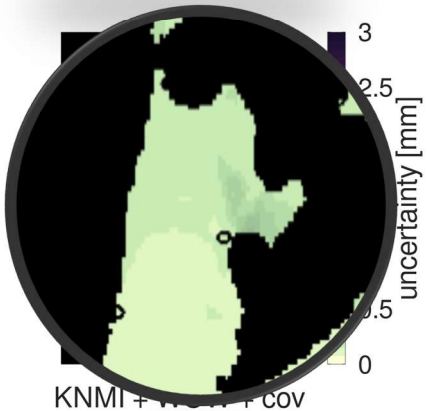
KNMI



KNMI + cov



KNMI + WOW

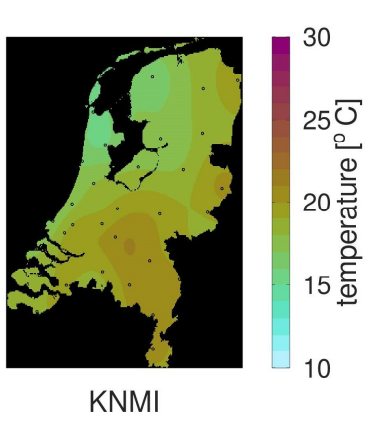


KNMI + cov

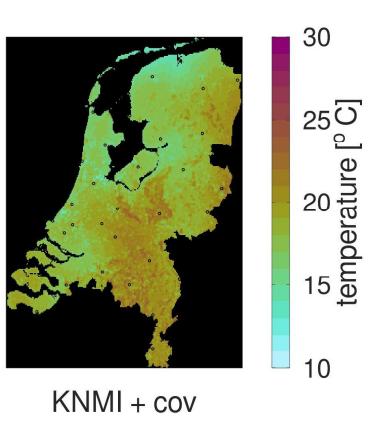
Source: De Baar, Garcia-Marti, Van der Schrier, EMS2022-53

Results

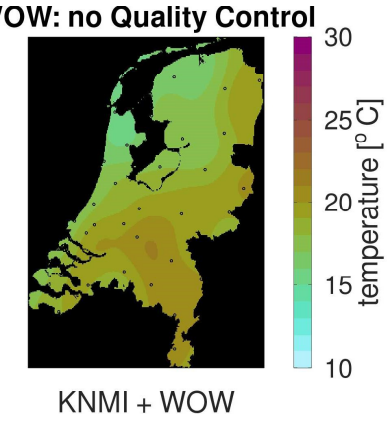
Temperature (29 June 2021, 8am-9am)



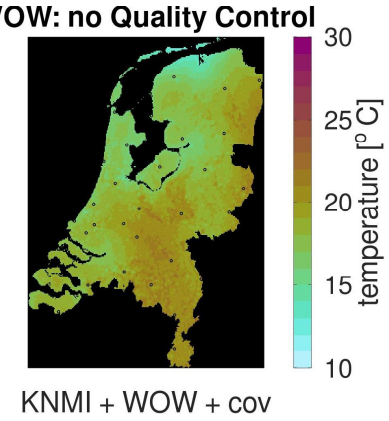
KNMI



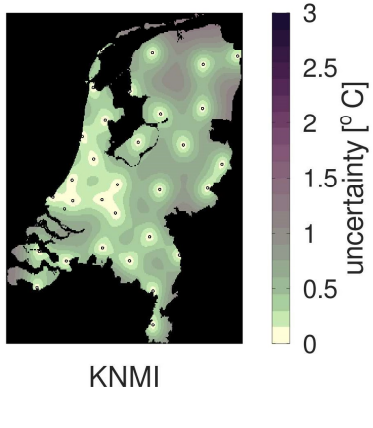
KNMI + cov



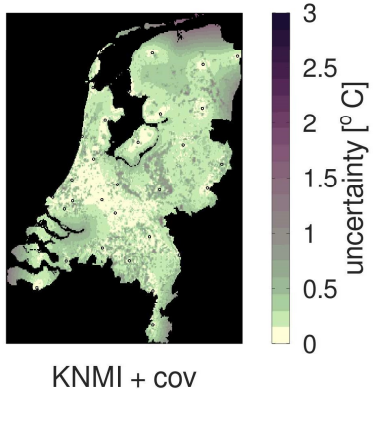
KNMI + WOW



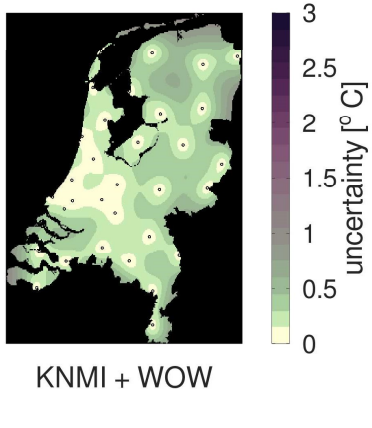
KNMI + WOW + cov



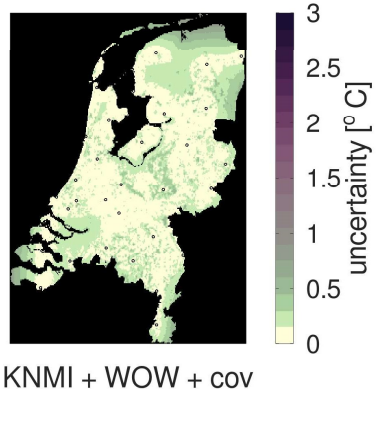
KNMI



KNMI + cov



KNMI + WOW

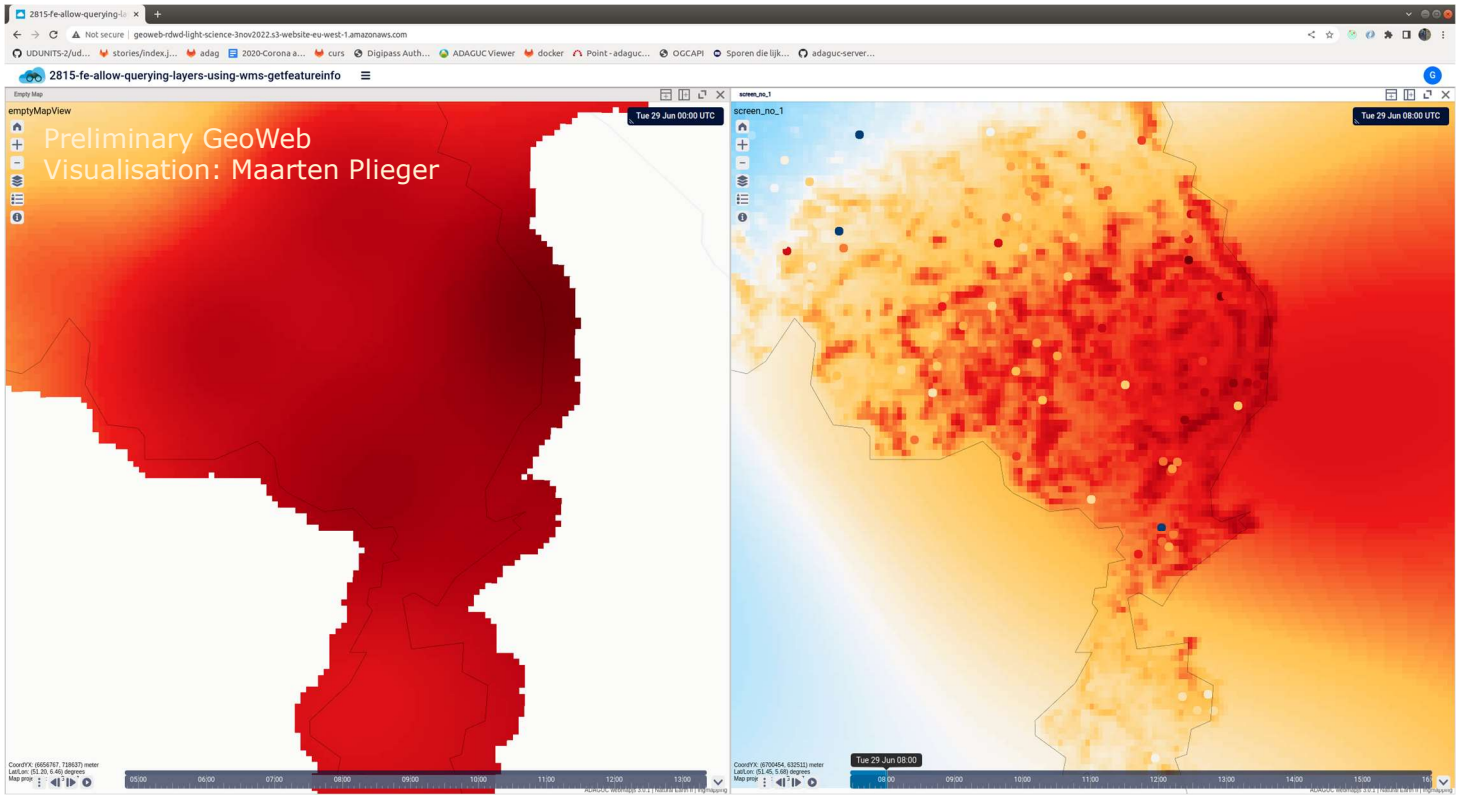


KNMI + WOW + cov

Source: De Baar, Garcia-Marti, Van der Schrier, EMS2022-53

Results

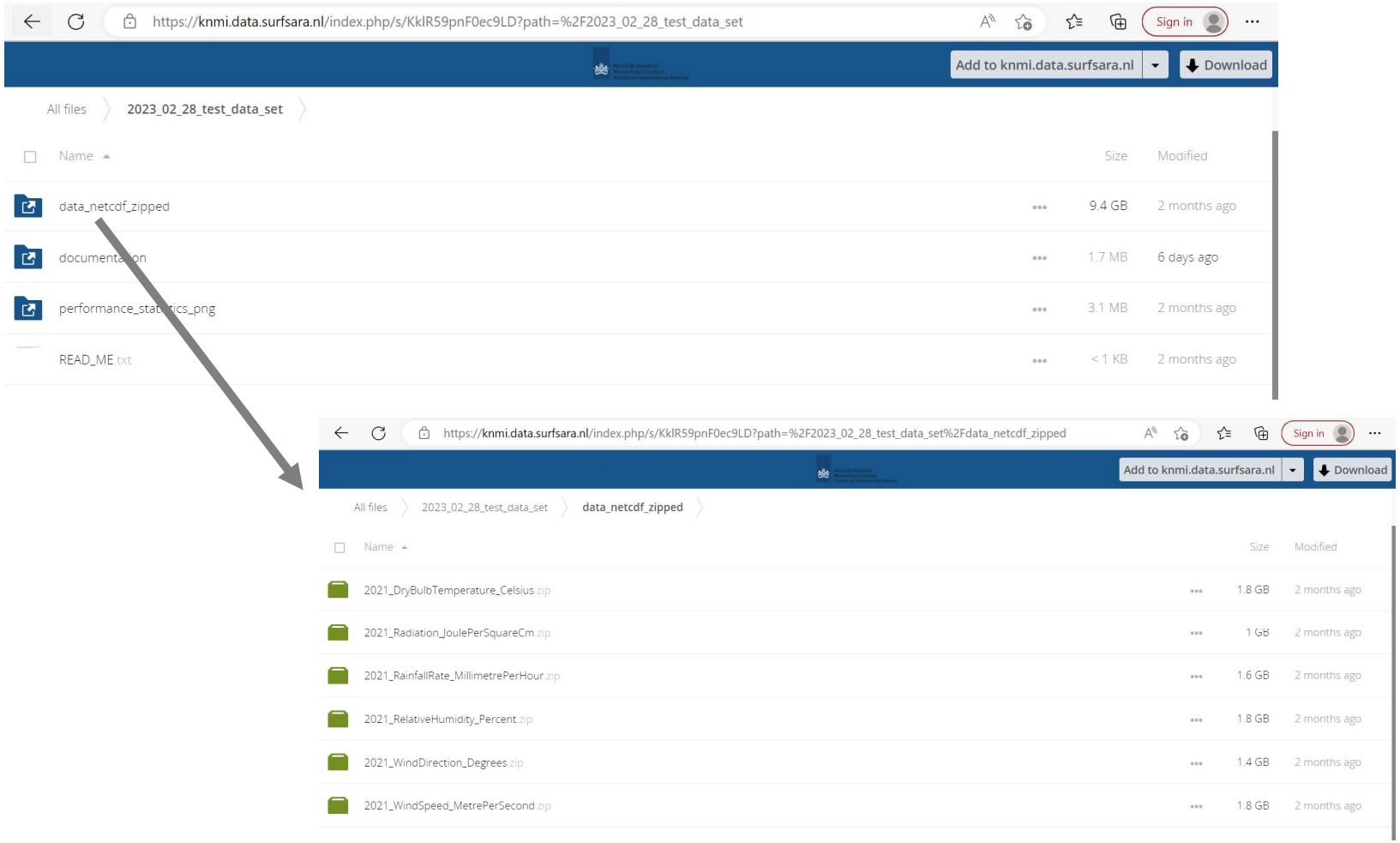
Temperature (29 June 20212, 24h)



- The observational process (ie bias and noise) is an intrinsic part of the regression
- Even without any QC on the crowd-sourced data, we can get reasonable results
- However, in the project we now do a first step of basic QC

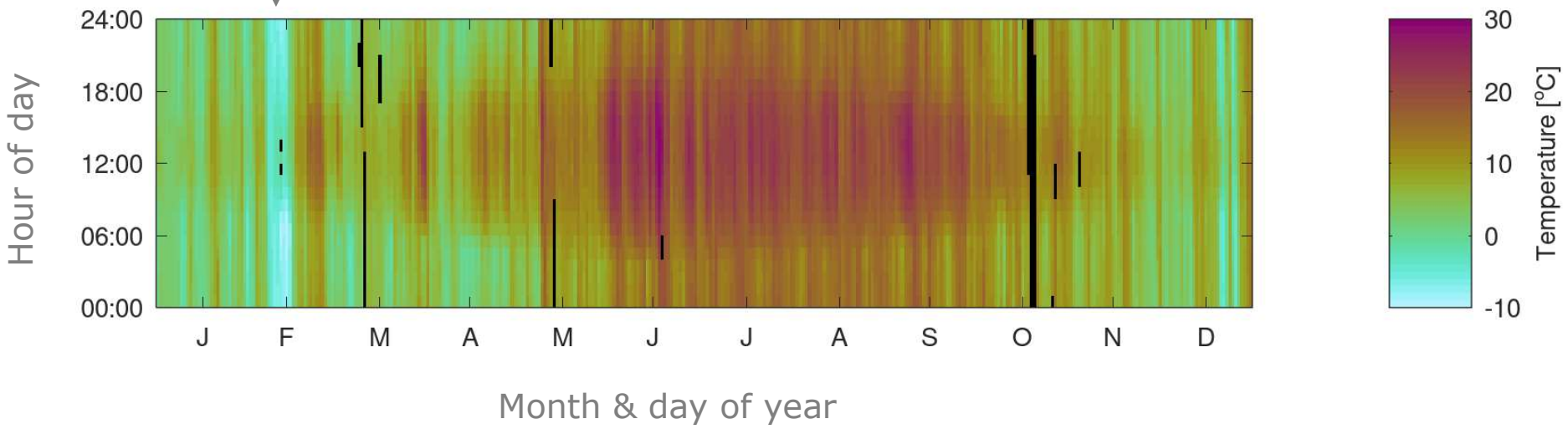
Results

Prototype data set: Year 2021



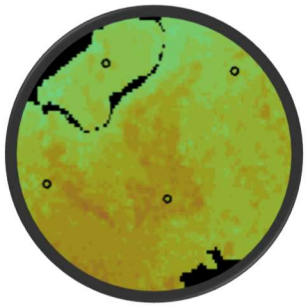
Results

Prototype data set: Year 2021

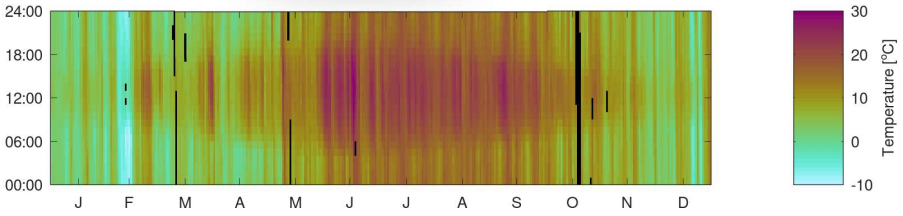


Results for prototype data set

The x-validation only represents official KNMI locations; however, we expect the largest improvements in non-official locations!

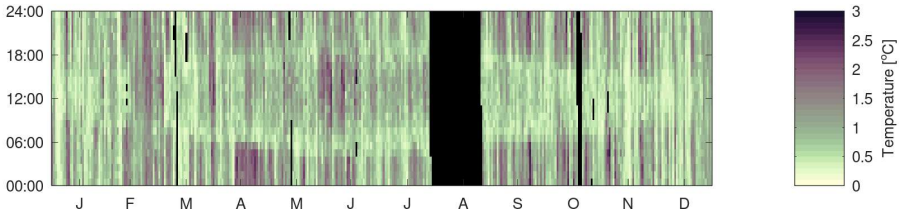


Air temperature

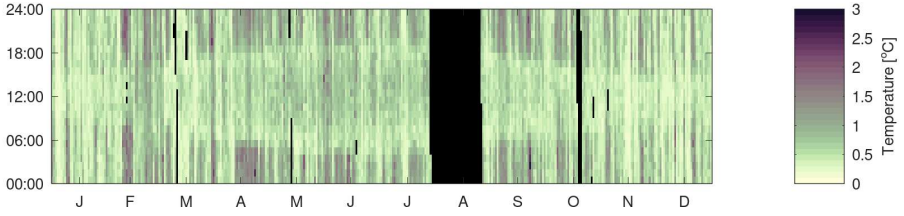


Cross-validation: air temperature

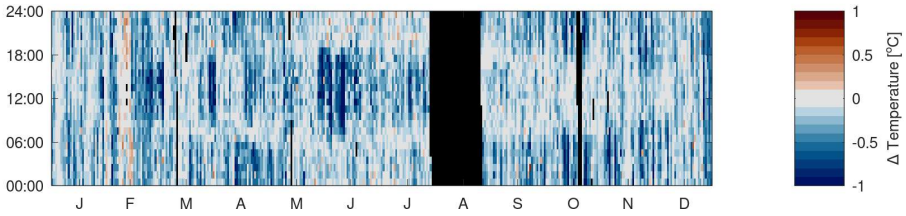
x-Val RMSE: KNMI



x-Val RMSE: KNMI + WOW + cov

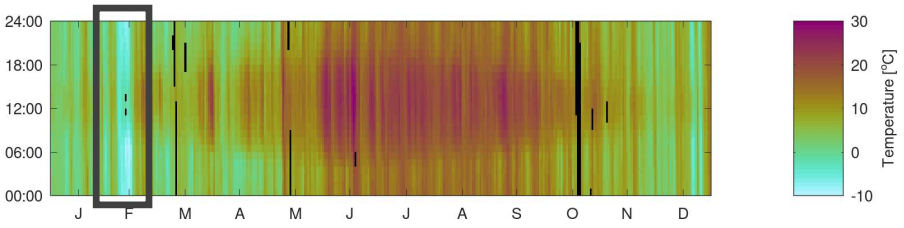


x-Val difference (blue = better)

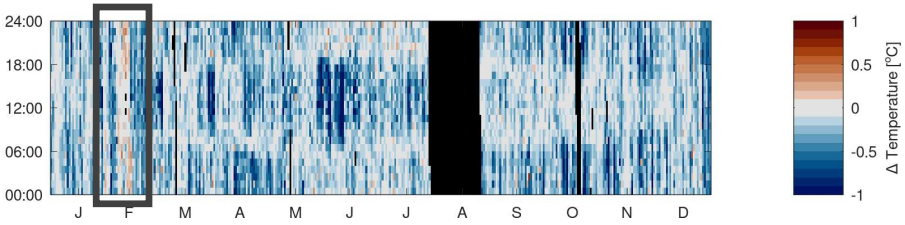


Results for prototype data set

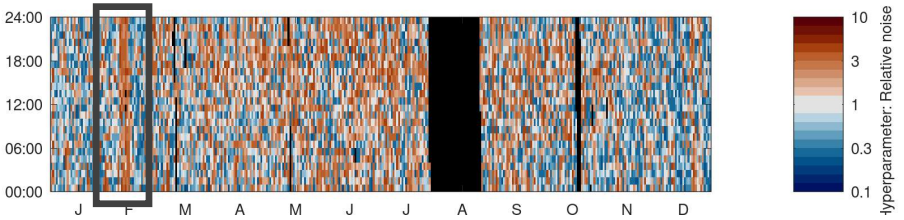
Air temperature



x-Val difference (blue = better)

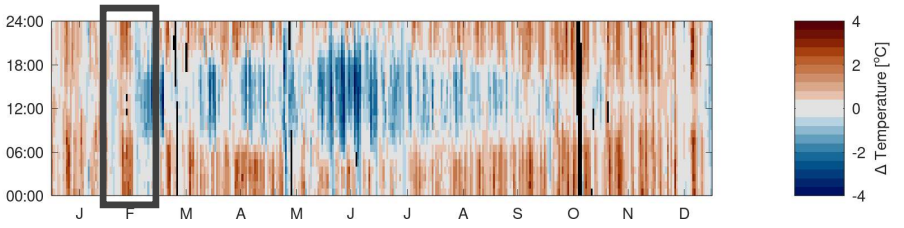


Hyperparameter: estimated noise level

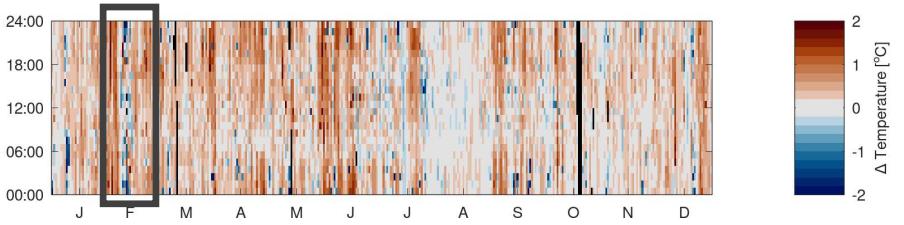


Issue with noise estimation & over-fitting

Coastal effect? (10% closest to water)

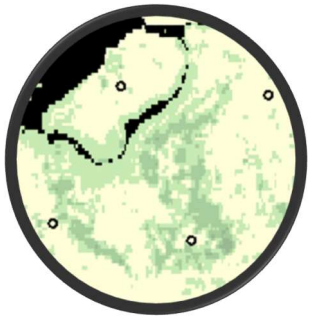


Population effect? (10% highest population)

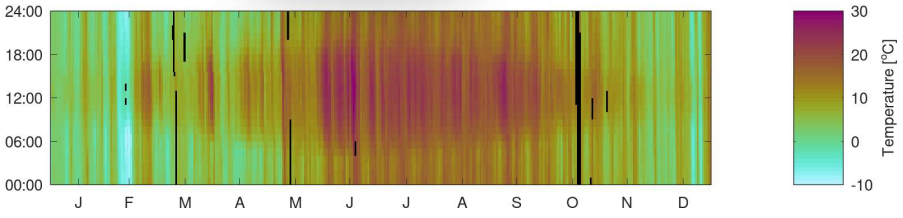


Results for prototype data set

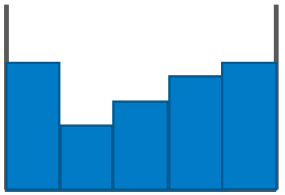
Looking for ways to improve the uncertainty maps!

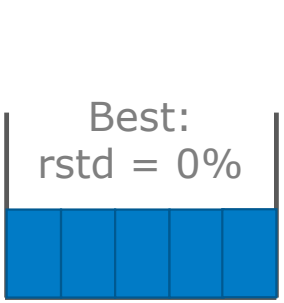


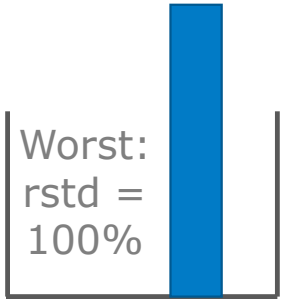
Air temperature



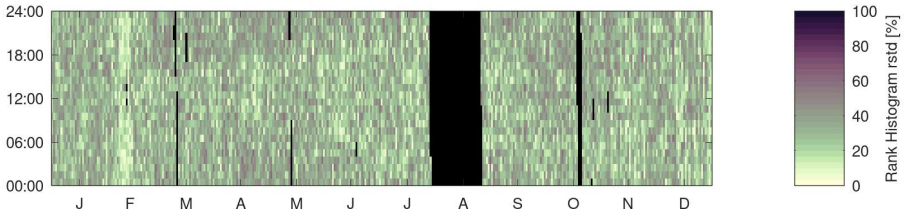
Cross-validation: rank histogram







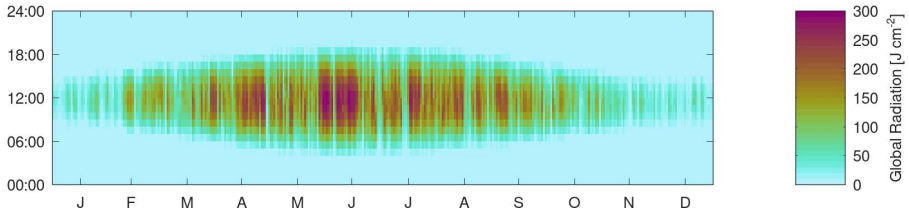
x-Val rank histogram rstd



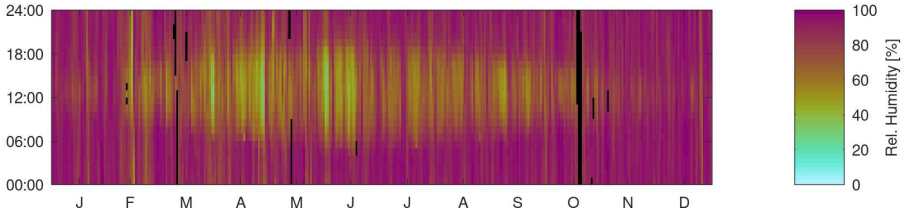
Results

Prototype data set: Year 2021

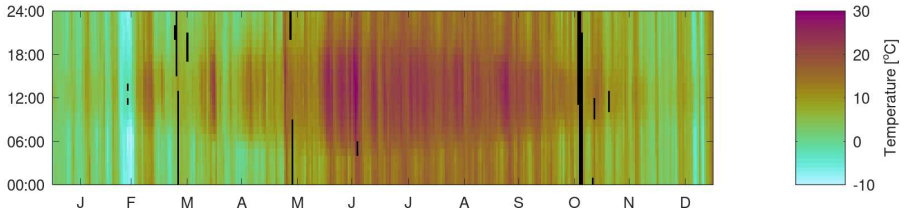
Global radiation



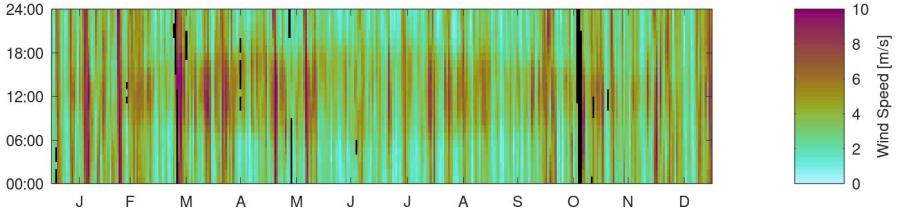
Relative humidity



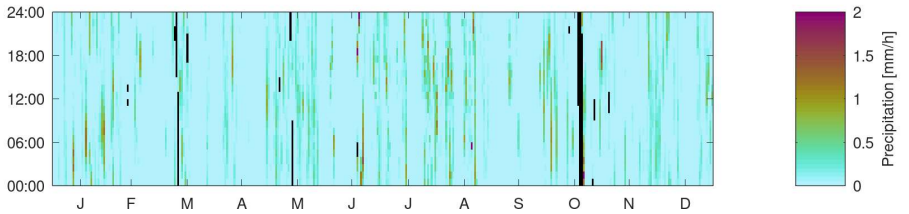
Air temperature



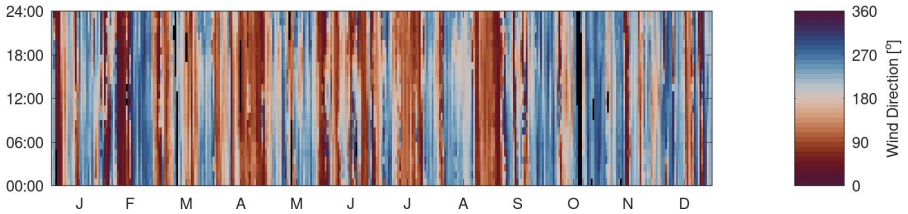
Wind speed



Precipitation



Wind direction



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

Introduction

Methods

Results

Next steps

Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

Next steps

- Collect & implement feedback from user group

- Correct issues

- Think about verification & validation

- Provide data set for full period (2015-2022)

Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

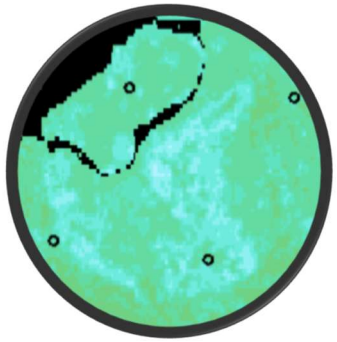
Thank you very much for your kind attention

Jouke de Baar, Gerard van der Schrier,Theo Brandsma (KNMI)

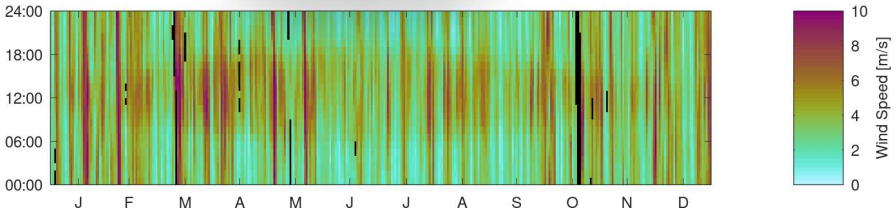
Budapest, 11 May 2023 (online)

Question & answer slide

The x-validation only represents official KNMI locations; however, we expect the largest improvements in non-official locations!

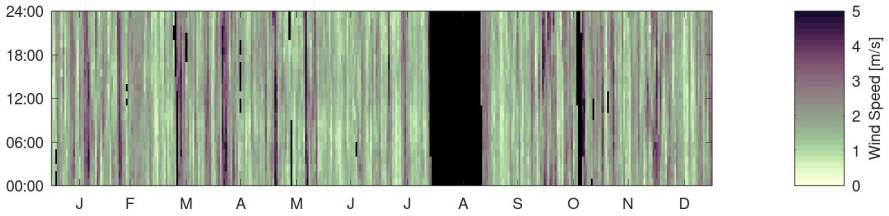


Wind speed

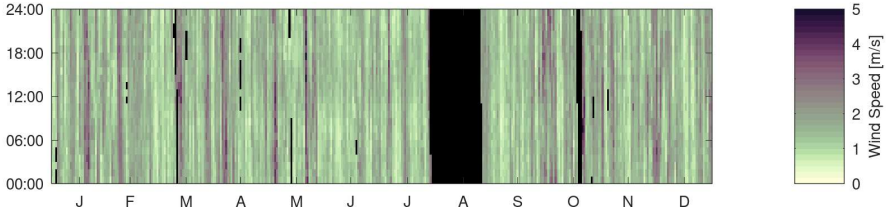


Cross-validation: wind speed

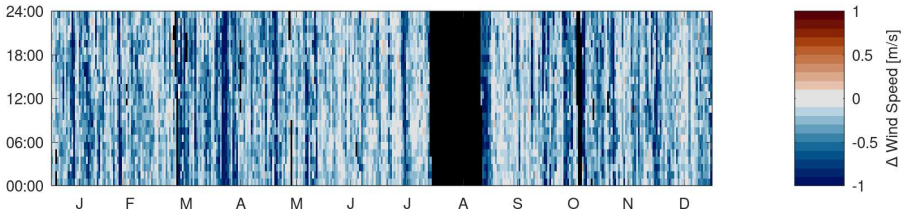
x-Val RMSE: KNMI



x-Val RMSE: KNMI + WOW + cov



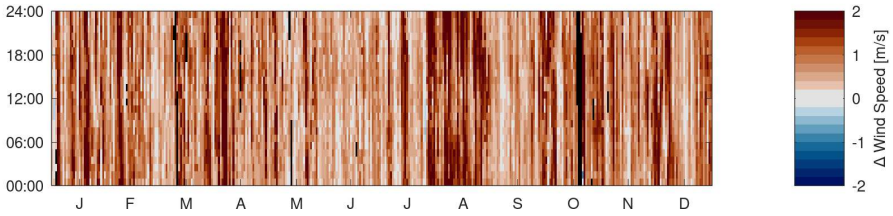
x-Val difference (blue = better)



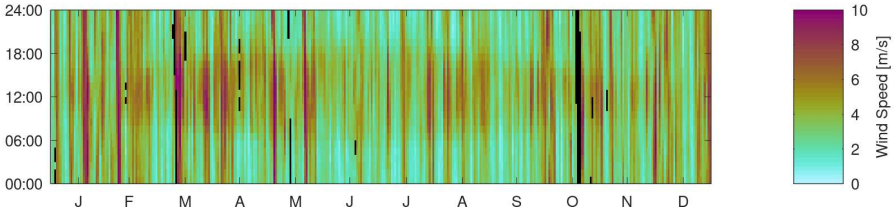
Question & answer slide

Local wind speed effects?

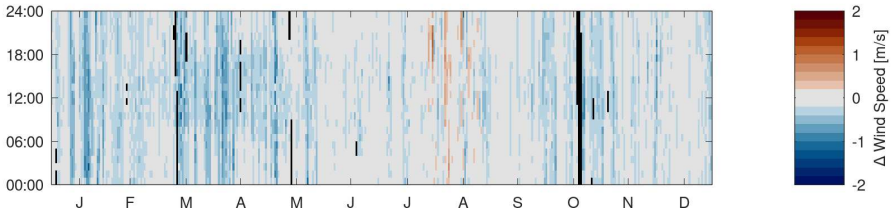
Coastal effect? (10% closest to water)



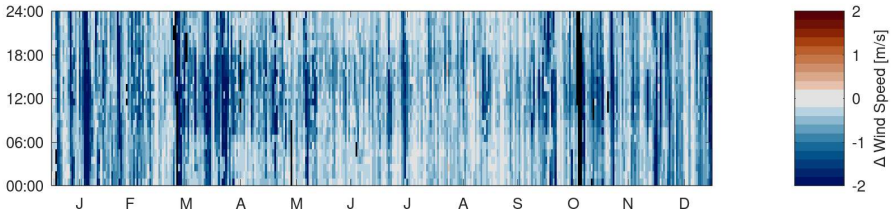
Wind speed



Population effect? (10% highest population)

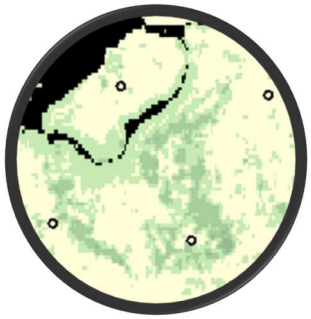


Tree effect? (10% highest tree cover)

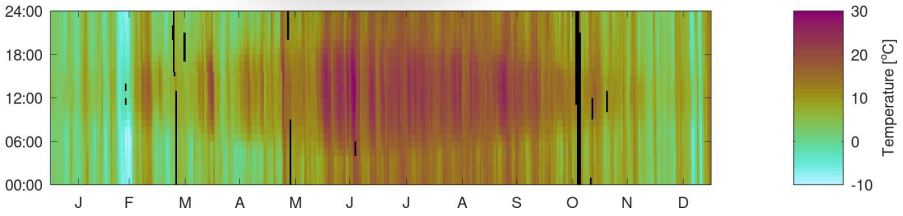


Question & answer slide

Looking for ways to improve the uncertainty maps!



Air temperature



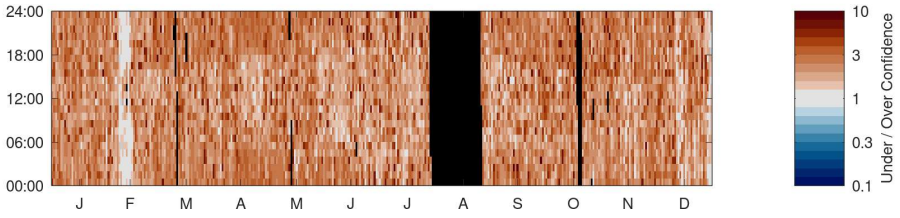
Cross-validation: confidence

- Compare x-validation errors with the local x-validation uncertainty map:
- Error < Uncertainty estimate:

  "Under-confident"
- Error > Uncertainty estimate:

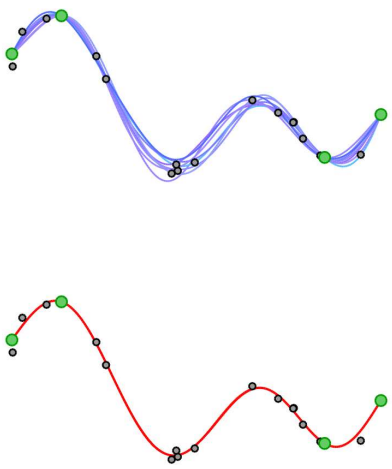
  "Over-confident"

x-Val under/over confidence



Question & answer slide

Equations



regression with noise treatment

Prior $$p(\boldsymbol{X}) = \mathcal{N}(Q\hat{\boldsymbol{\beta}}_Q, P)$$

Likelihood $$p(\boldsymbol{Y}|\boldsymbol{x}) = \mathcal{N}(H\boldsymbol{x} + B\boldsymbol{\beta}_B, R)$$

$$\sqrt{R} = N\hat{\boldsymbol{\beta}}_N I_{n\times n}$$

Posterior $$p(\boldsymbol{X}|\boldsymbol{y}) = \mathcal{N}(\hat{\boldsymbol{x}}, \hat{C})$$

Source: De Baar, Garcia-Marti, Van der Schrier, EMS2022-53 / Proceedings asr-2023-5

Question & answer slide

Equations

Standard GPR

$$E[f(x)] = m(x) + k(x,\mathbf{x'},\theta)\, K(\mathbf{x'},\mathbf{x'},\theta)^{-1} \{ \mathbf{y} - \mathbf{m}(\mathbf{x'}) \},$$

$$var[f(x)] = k(x,x,\theta) - k(x,\mathbf{x'},\theta)^{T}\, K(\mathbf{x'},\mathbf{x'},\theta)^{-1}\, k(x,\mathbf{x'},\theta),$$

With external drift

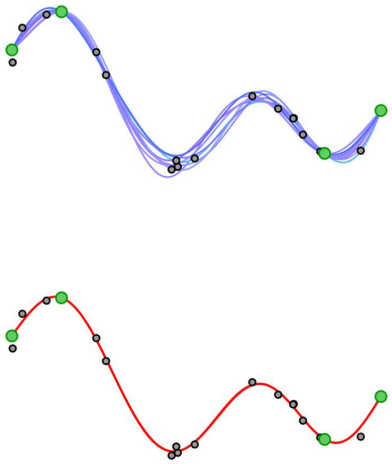
$$E[f(x)] = M(x)\,\boldsymbol{\beta} + k(x,\mathbf{x'},\theta)\, K(\mathbf{x'},\mathbf{x'},\theta)^{-1} \{ \mathbf{y} - \mathbf{M}(\mathbf{x'})\,\boldsymbol{\beta} \},$$

$$var[f(x)] = k(x,x,\theta) - k(x,\mathbf{x'},\theta)^{T}\, K(\mathbf{x'},\mathbf{x'},\theta)^{-1}\, k(x,\mathbf{x'},\theta) + C_M(x,x,\theta,\boldsymbol{\beta}),$$

With bias and noise

$$E[f(x)] = M(x)\,\boldsymbol{\beta} + k(x,\mathbf{x'},\theta) \{ K(\mathbf{x'},\mathbf{x'},\theta) + R(\mathbf{x'},\mathbf{x'},\boldsymbol{\varepsilon}) \}^{-1} \{ \mathbf{y} - \mathbf{M}(\mathbf{x'})\,\boldsymbol{\beta} \},$$

$$var[f(x)] = k(x,x,\theta) - k(x,\mathbf{x'},\theta)^{T} \{ K(\mathbf{x'},\mathbf{x'},\theta) + R(\mathbf{x'},\mathbf{x'},\boldsymbol{\varepsilon}) \}^{-1}\, k(x,\mathbf{x'},\theta) + C_M(x,x,\theta,\boldsymbol{\beta}).$$



regression with noise treatment

Source: De Baar, Garcia-Marti, NATO STO-MP-AVT, 2022, Varna