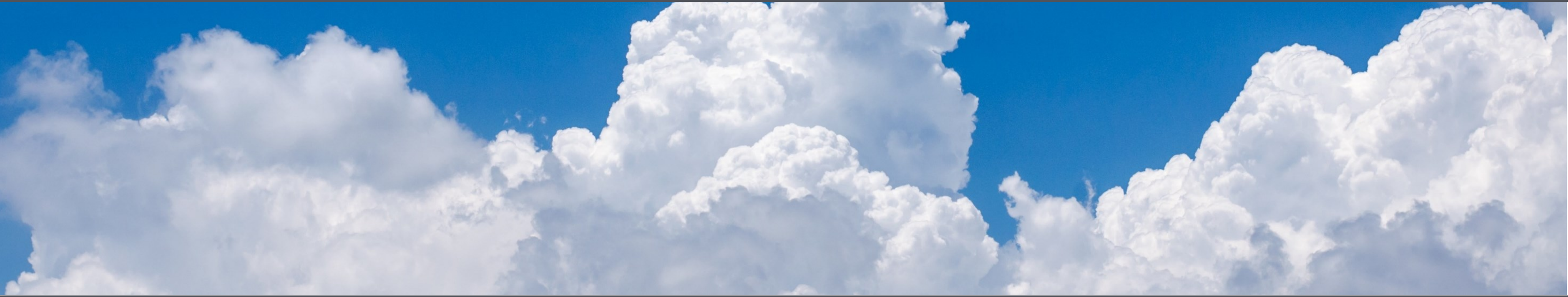
Department of Meteorology
Department of Mathematics & Statistics

Observation uncertainty and information content



Sarah L Dance

Thanks to: Rishabh Bhatt, Massimo Bonavita, Niels Bormann, Alison Fowler, Elias Holm, Guannan Hu, Chris Merchant, Jonathan Mittaz, Stuart Newman, Nancy K. Nichols, Jennifer Scott, David Simonin, Joanne Waller

Outline

- Why should we care about observation uncertainty?
- How can we take account of observation uncertainty? How does observation uncertainty affect analyses?
- What now? Observation impact in the age of ML

Grand- challenge: hazardous weather

- Hazardous events such as intense rainfall, windstorms, fog, snow
  - Physics based: Global km-scale expected to improve predictions (already proven in regional)
  - Deep learning based: high-resolution to avoid smoothing out extremes
- Need high-resolution observation data to initialize these models with detailed information
- Major observing gaps in horizontal spacing, temporal frequency

Capability location	Layer	Accuracy		Horizontal spacing		Vertical resolution		Observation cycle		Timeliness	
		Land Domain									
Surface-based	Near Surface	T	w	T	w			T	w	T	w
		q	P	q	P			q	P	q	P
	PBL	T	w	T	w	T	w	T	w	T	w
		q	iwv	q	iwv	q		q	iwv	q	iwv
		'Cloud free' Domain									
Space-based	**PBL**	T	w	T	w	T	w	T	w		
		q		q		q		q			

Shading	
	Meeting OSCAR goal requirement
	Meeting OSCAR breakthrough requirement
	Meeting OSCAR threshold requirement
	Insufficient information available
	Falling below OSCAR minimum requirement

Thanks to Jacqueline Sugier for this Table.

Requirement from WMO OSCAR RRR

Closing observation gaps

- Too expensive/impractical to close the observation gap with conventional scientific observations alone
- Make better use of the observations we have already
- Exploit crowdsourcing, citizen science and opportunistic data (e.g., Hintz et al, 2019)

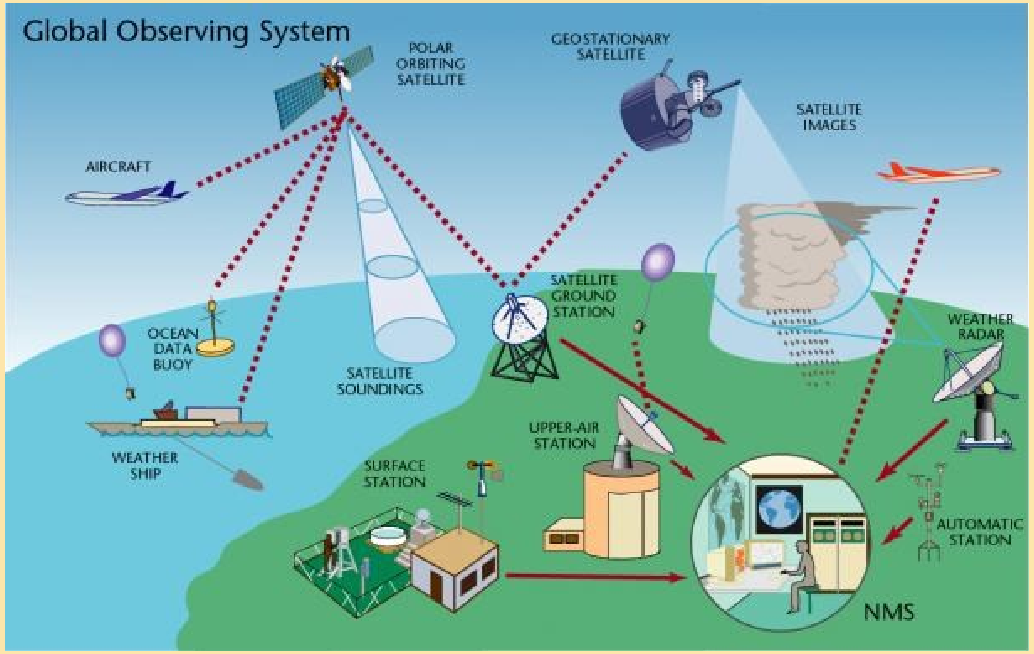
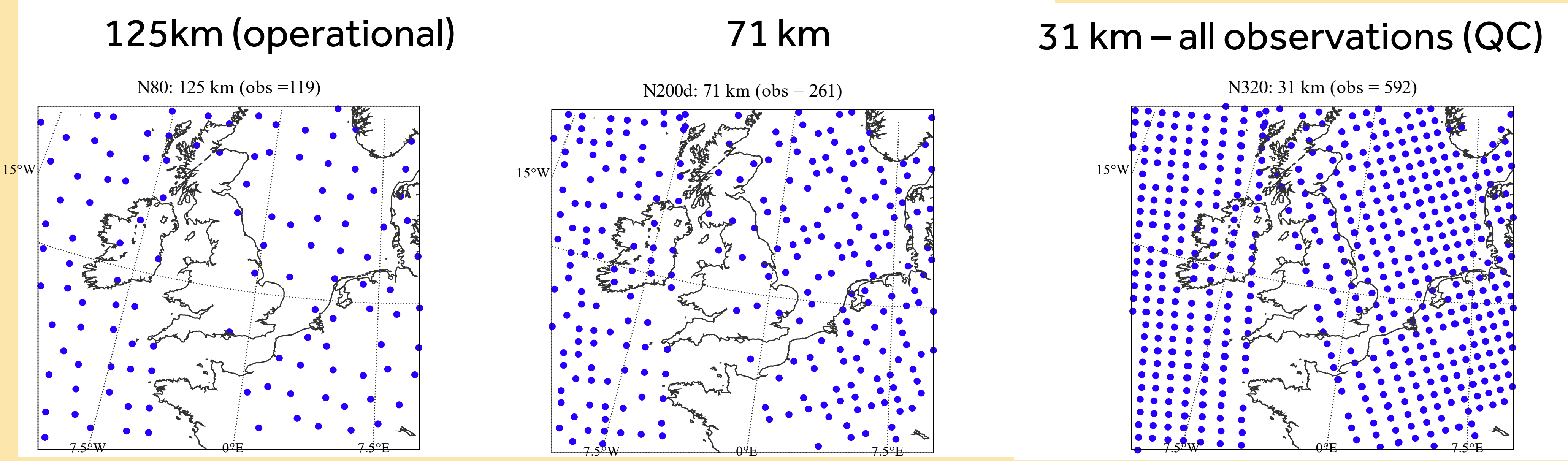


Image from
https://public.wmo.int/en/programmes/global-observing-system

Problem: observation thinning

- We currently only use a small percentage of some observation types



NOAA-18 AMSU-A (Nov 2023)

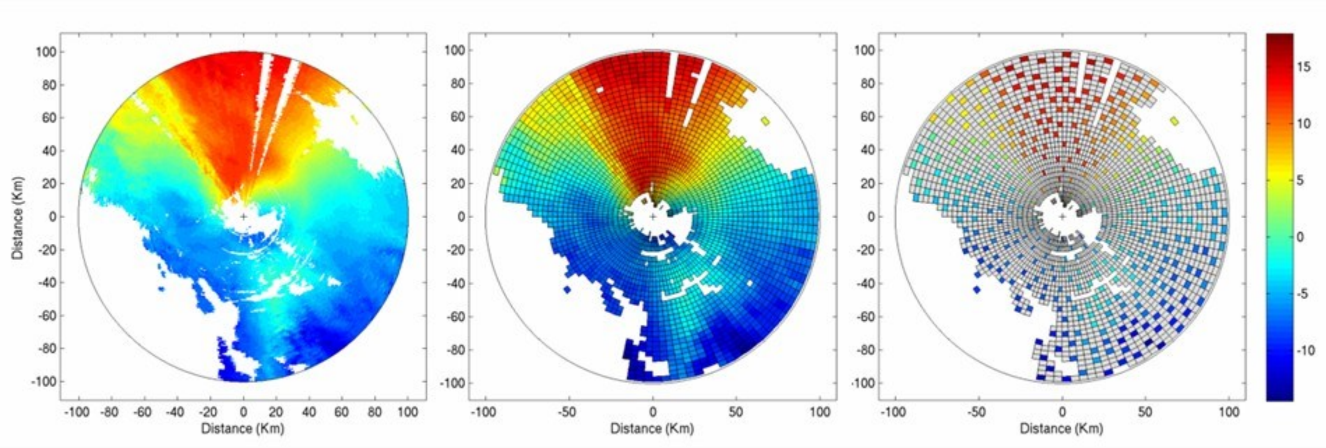
ECMWF system

- Data are thinned due to lack of treatment of spatial/temporal correlations, to avoid overfitting
- We now have an approximate way to estimate the observation error structures (e.g., Waller et al 2016)
- But computational cost for large matrices (10^7 x 10^7) is enormous!

Taking proper account of observation error correlations

- Use higher proportion of observations

Raw (~600m) Superobbed (~3km) Thinned (~6km)



Example: Doppler radar winds (*old* Met Office system from Jo Waller/David Simonin)

- Improve analysis accuracy and forecast skill (e.g., Stewart et al. 2013; Weston et al., 2014)
- Changes to scales of observation information content in analysis depending on both the prior and observation error correlations (Fowler et al, 2018)

Assimilation with spatial correlations

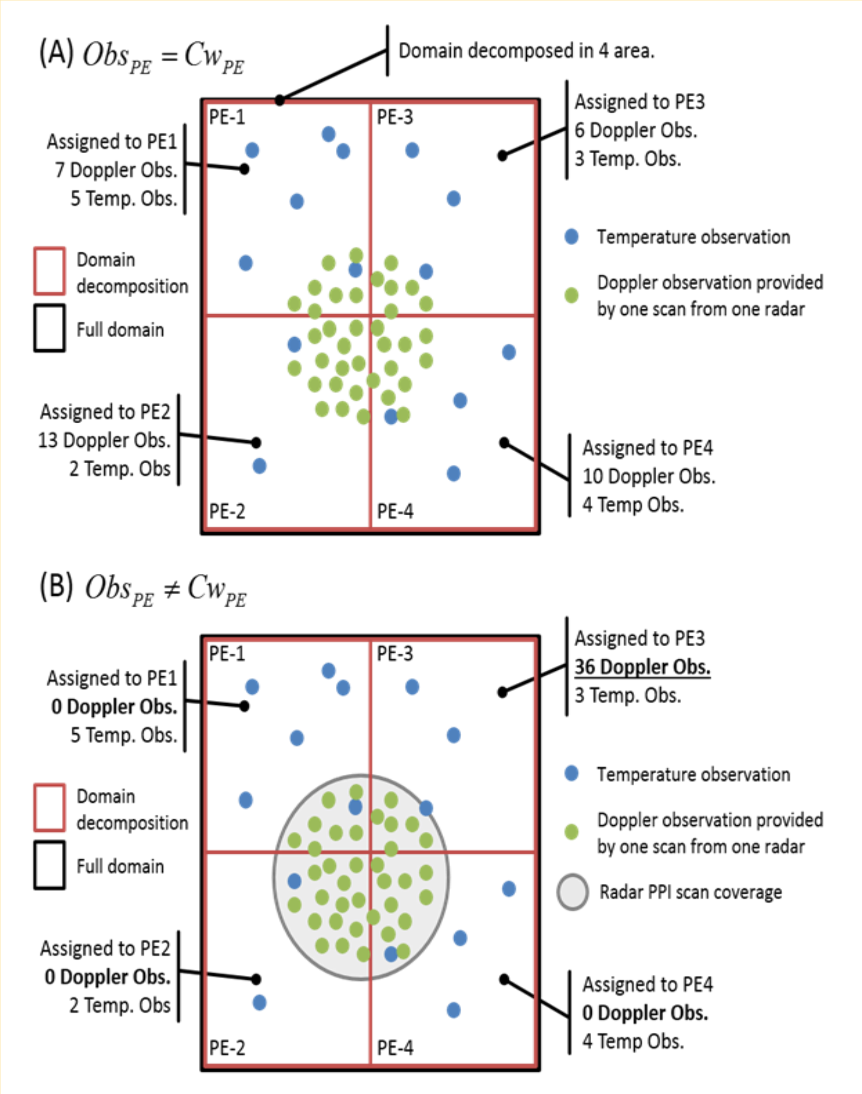
University of Reading

Minimize

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^\top \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + \frac{1}{2}(H(\mathbf{x})-\mathbf{y})^\top \boxed{\mathbf{R}^{-1}(H(\mathbf{x})-\mathbf{y}),}$$

Need new methods – time taken is critical in operations

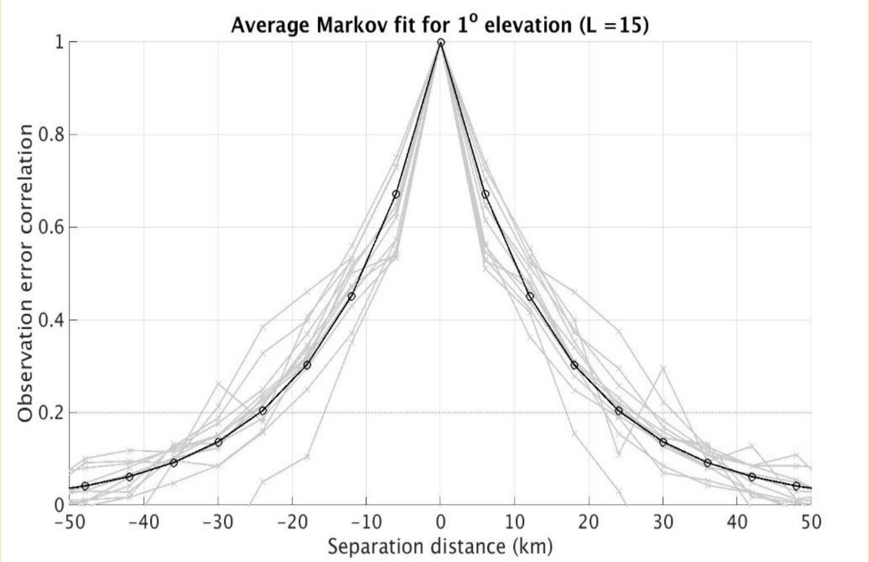
- Usually approximated as diagonal
- Full matrices introduce communication/load imbalance costs using parallel computing techniques
- Inverse matrices change each cycle (QC etc)
- Convergence of whole problem changes (Tabeart et al. 2021, Goux et al. 2024)



Treating correlated observation error for Doppler radar winds (Simonin et al, 2019)

- Current 6km thinning reduced to 3km by treating spatial observation error correlations
- Spatial correlations estimated offline using Desroziers et al (2005) method
- New parallelization and smart load balancing

⇒ can assimilate 4x as many observations without an increase in wall-clock time



Joint work using Met Office UKV

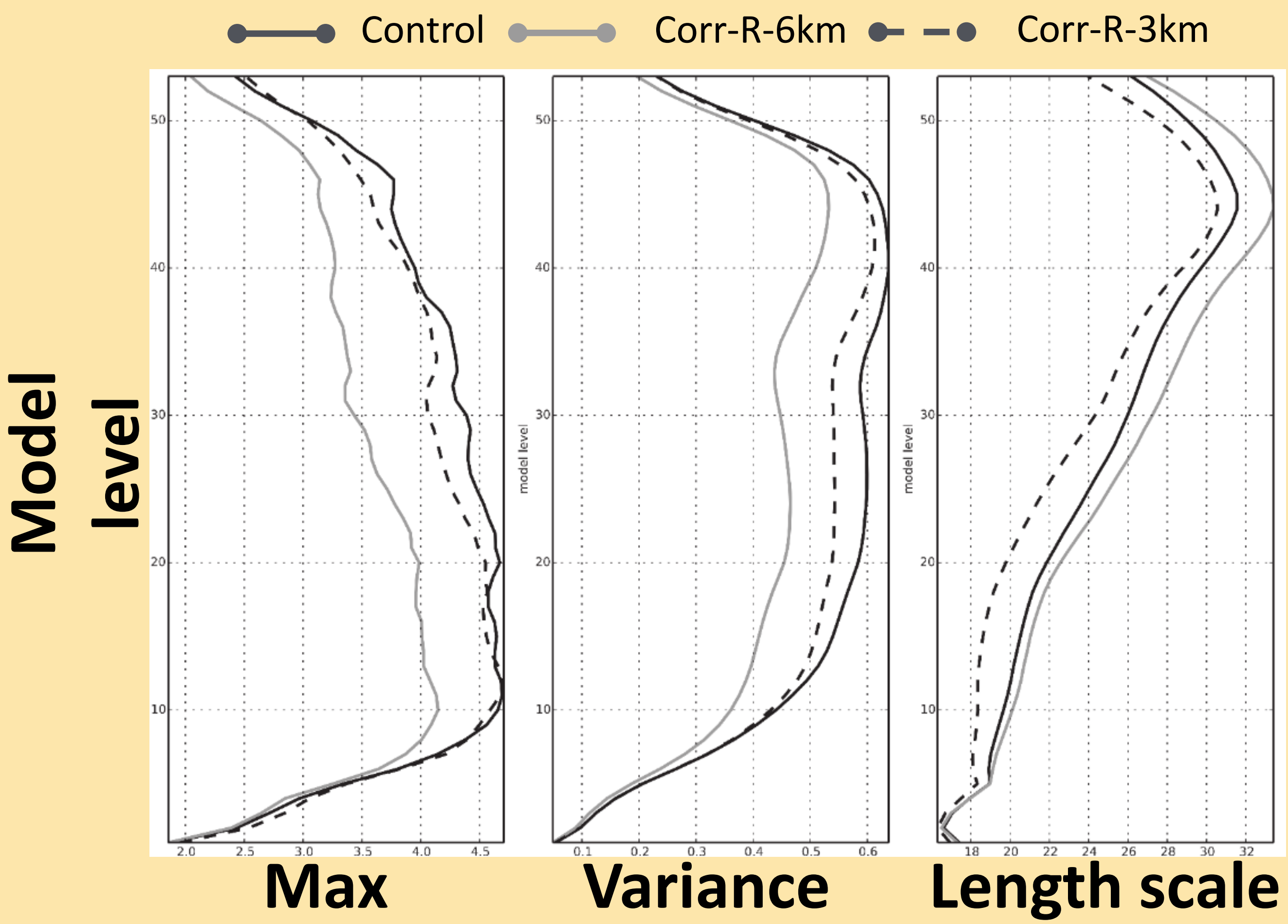
Effect on wind increments (Simonin et al 2019)

Compared to the Control with diagonal **R** and 6km thinning

The **Corr-R-3km** wind's increments show more small scale features with smaller range

Compares well with theory (Fowler et al 2018)

We also found improvements in forecast skill – particularly for convective rainfall.

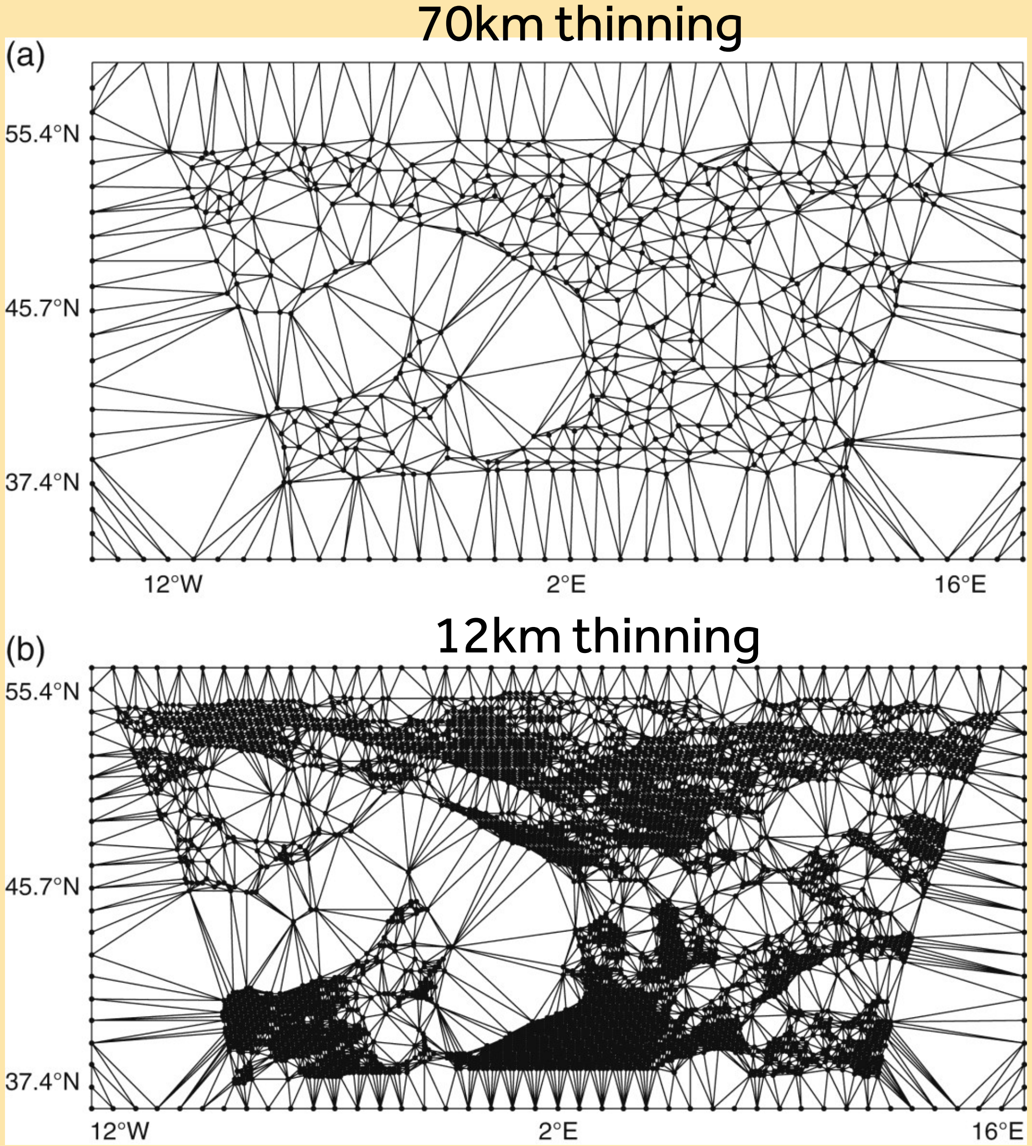


Geostationary data (Guillet et al 2019)

- Previous approach not suitable when matrices are very large

- Guillet et al proposed modelling spatial error correlations using a diffusion operator and fast meshing techniques (see Olivier Goux's poster)
- Mesh locations for SEVIRI with Meteo France AROME configuration.

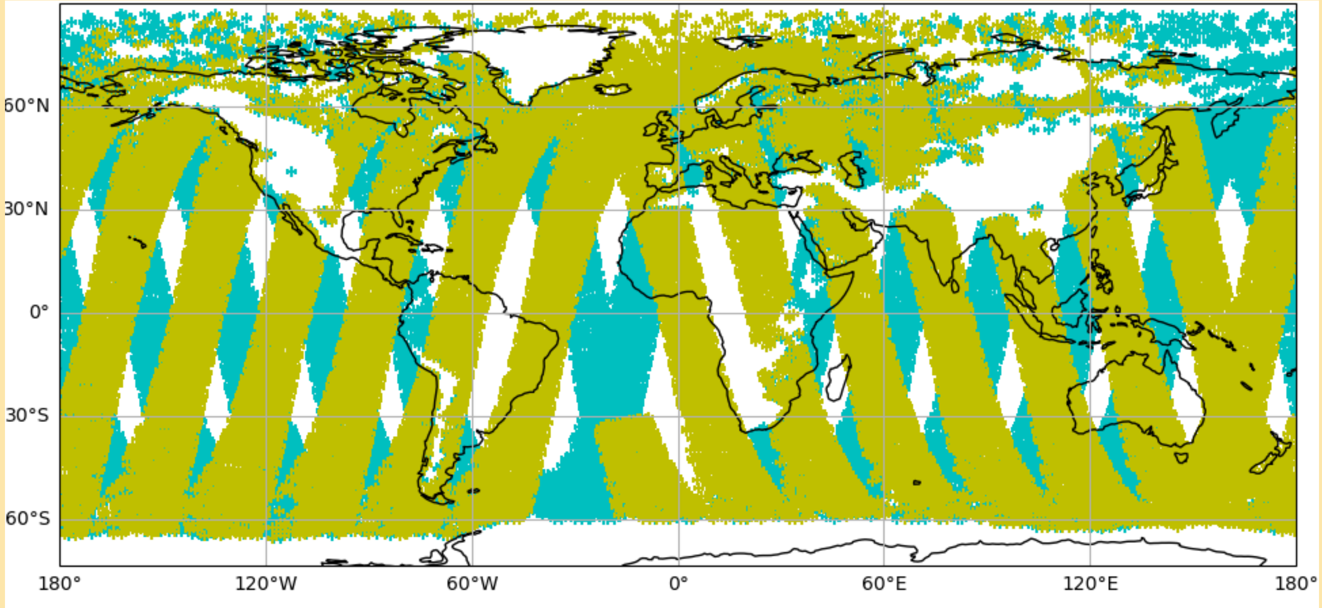
- How to deal with inter-channel error correlations at the same time?



Polar orbiter – AMSU-A

- Our new project – HiSCORE: High resolution data assimilation with Spatially and temporally Correlated ObseRvation Errors

- Step 1 – estimate error correlations and understand their structures

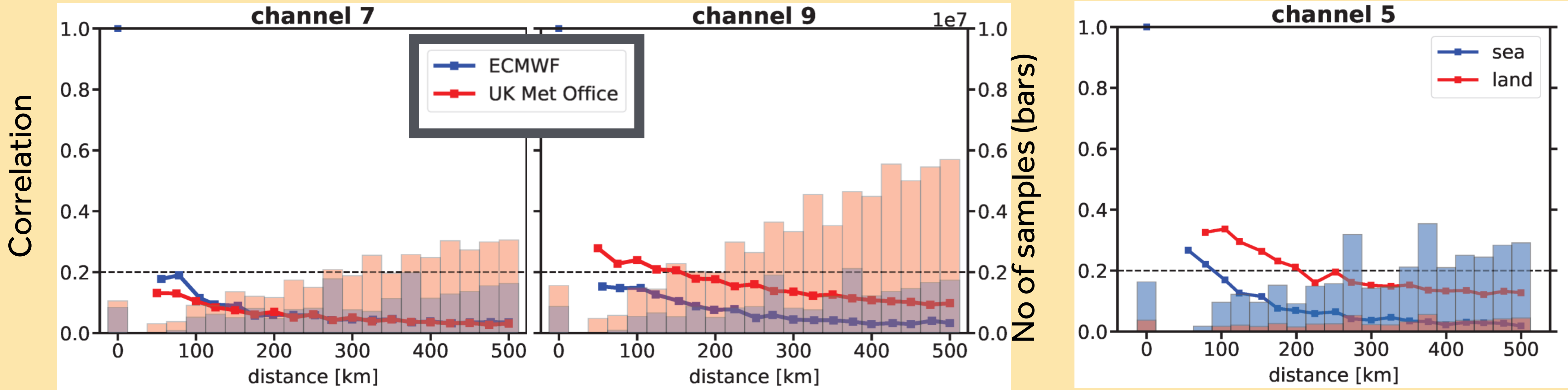


+ 00 – 12 UTC
+ 12 – 24 UTC

AMSU-A **(Advanced Microwave Sounding Unit)**

Chosen as high impact, all-sky data, but correlations expected to be simple(ish)

AMSU-A on MetOp-C (Bhatt et al, in prep)



- Some channels could be assimilated with diagonal R at higher density as correlations are small (e.g., channel 7 - LEFT)
- Some evidence that some correlations are due to – imperfect bias correction (e.g. MO channel 9- MIDDLE), issues with surface emissivity over land (channel 5 RIGHT) . Can we fix some of these problems?

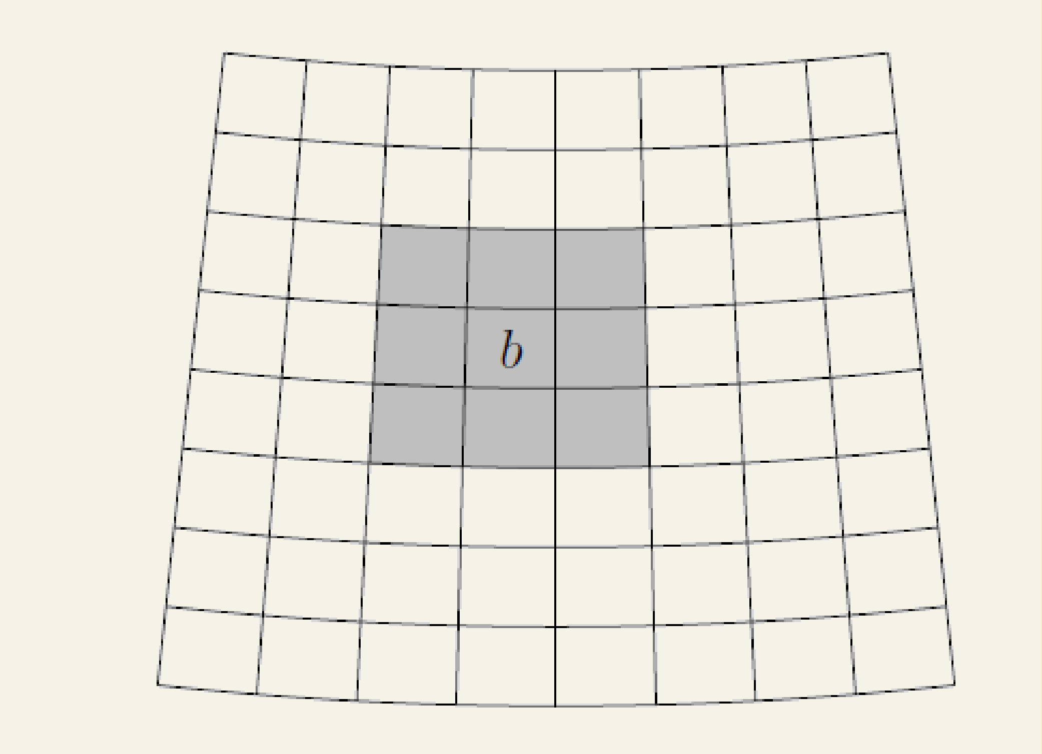
What to do about remaining correlations?

- We need a fast, flexible matrix-vector product approach…

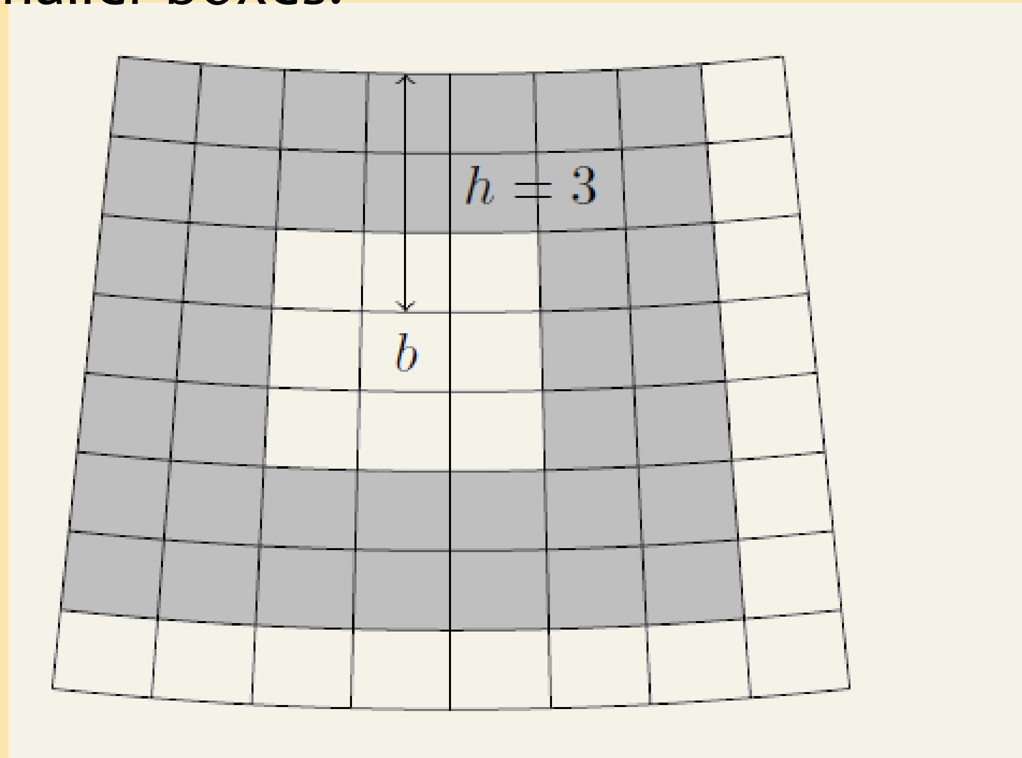
The local SVD-FMM (Hu & Dance, 2024)

Domain localization + Singular value decomposition (SVD) approach of the fast multipole method (FMM; Gimbutas and Rokhlin, 2003)

Step 1: Divide the observation domain into smaller boxes.



Near field of box b



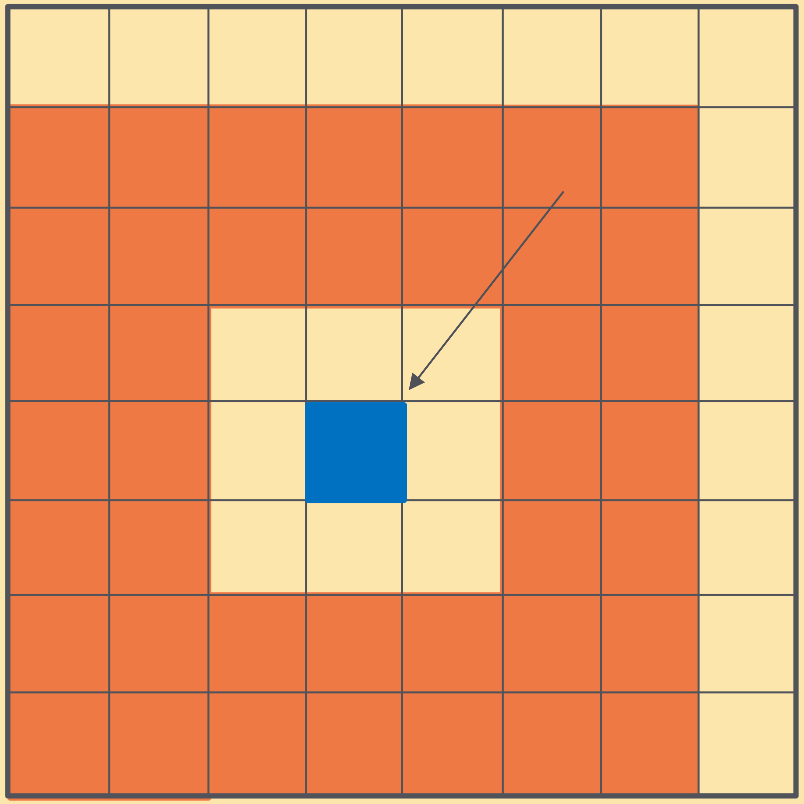
Interaction and non-interaction fields (non-dimensional interaction length h)

Step 2 Separate the calculation of matrix-vector products

$$\mathbf{q} = \sum_b \mathbf{q}^b = \sum_b \mathbf{E}^b \mathbf{d}$$

$$= \sum_b \underbrace{\boxed{\mathbf{E}^b_{near}\mathbf{d}^b_{near}}}_{\text{Exact}} \underbrace{\boxed{+\, \mathbf{E}^b_{int}\mathbf{d}^b_{int}}}_{\text{Approximate}} \underbrace{+\, \mathbf{E}^b_{nint}\mathbf{d}^b_{nint}}_{\text{Discard}}$$

- Non-interaction-field products ($\mathbf{E}^b_{nint}\mathbf{d}^b_{nint}$) are discarded.
- Information ($\mathbf{d}^{\mathrm{b}}$) in each box is **compressed**, which reduces the **algorithmic complexity** and the **size of message** in parallel computing.



$\mathbf{E}^b_{int}\mathbf{d}^b_{int}$: The blue box requires information from orange boxes (which are in its interaction field).

DA experiments with the local SVD-FMM (efficiency)

Experimental design

$$\nabla J(\delta \mathbf{z}) = \left(\mathbf{I} + \mathbf{B}^{1/2}\mathbf{H}^{\top}\mathbf{E}\mathbf{H}\mathbf{B}^{1/2}\right)\delta \mathbf{z} - \mathbf{B}^{1/2}\mathbf{H}^{\top}\mathbf{E}\mathbf{d}$$

- Minimization: conjugate gradient method
- $\mathbf{B}$: $15{,}904 \times 15{,}904$
- $\mathbf{E}$: 3976×3976
- $\mathbf{H}$: linear
- p: number of singular vectors
- h: localization length

Results

Matrix-vector multiplication	CPU times (s)		
	Min	Avg	Max
Standard	279.81	298.65	306.94
$p = 4, h = 2$	$0.77 + 9.21$	$0.91 + 10.96$	$1.64 + 12.81$
$p = 4, h = 3$	$1.19 + 8.76$	$1.30 + 9.69$	$1.96 + 10.86$
$p = 10, h = 3$	$2.30 + 8.91$	$2.48 + 9.72$	$2.71 + 10.88$

*The local SVD-FMM: initialization time (calculation of SVDs, etc.) + iteration time (CG)

**100 realizations of $\mathbf{d}$

Still working on it…

- SVD initialization is expensive with large-matrices
  - Can we use low rank updates and down-dates, or precomputed approximations?

- Current version of method assumes $\boldsymbol{R}^{-1}$ is already available
  - Can we use this hierarchical idea with a Cholesky decomposition?
  - Would another modern numerical analysis approach be better?

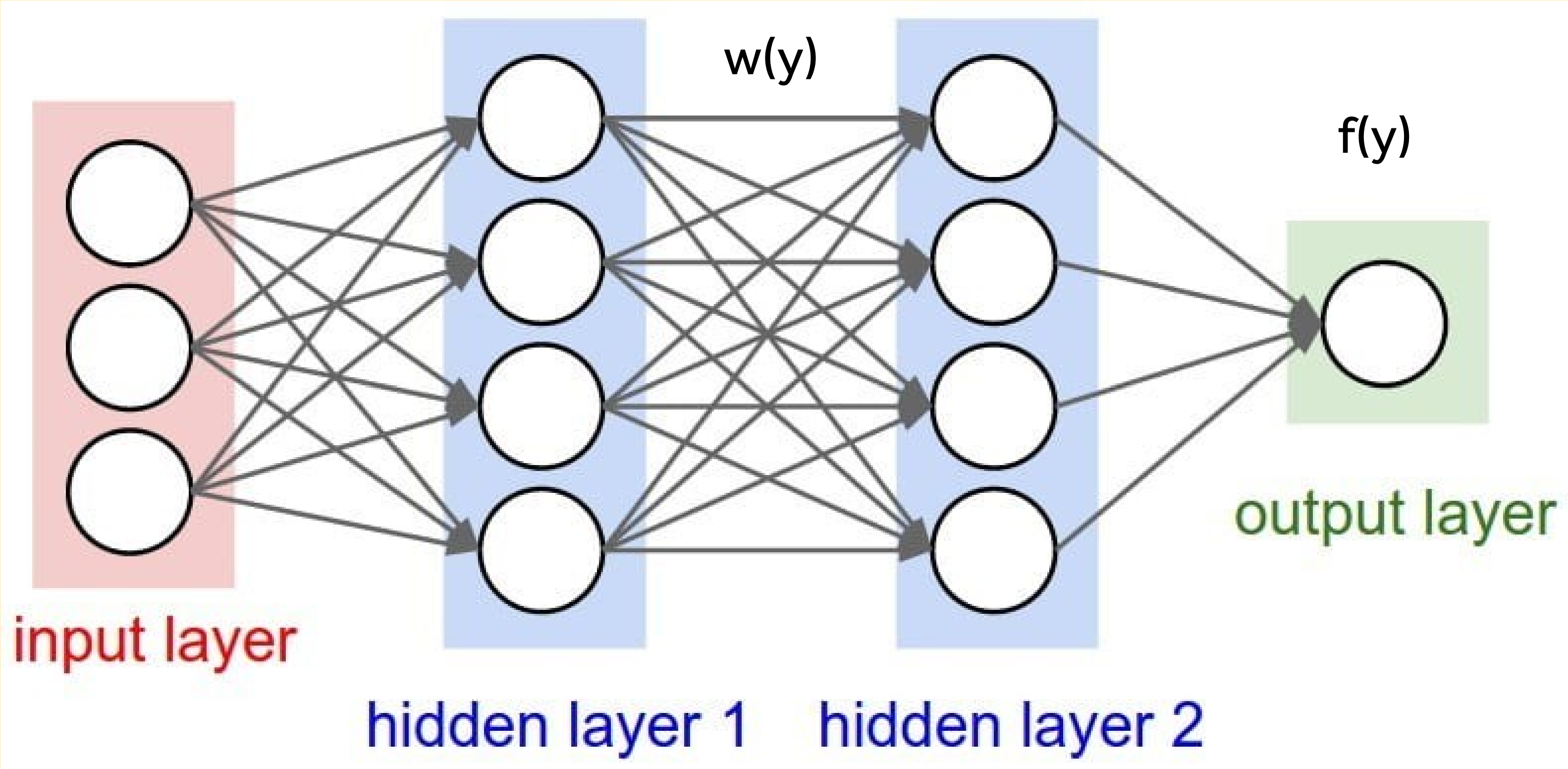
What now? Observation impacts in physics-based systems

- In conventional data assimilation, the description of observation uncertainty plays a key role in the influence of those observations on the analysis
  - Magnitude of increments
  - Lengthscales of increments
- There are a number of metrics that we use to measure observation influence and impact (e.g., DFS, FSOI)
- Types of experiment – OSSEs, data denial, ensemble sensitivity…

What now? Observation impacts in the age of ML

- We need to be able to assess observation influence on
  - training (network weights)
  - output (a given prediction)

Observations
y

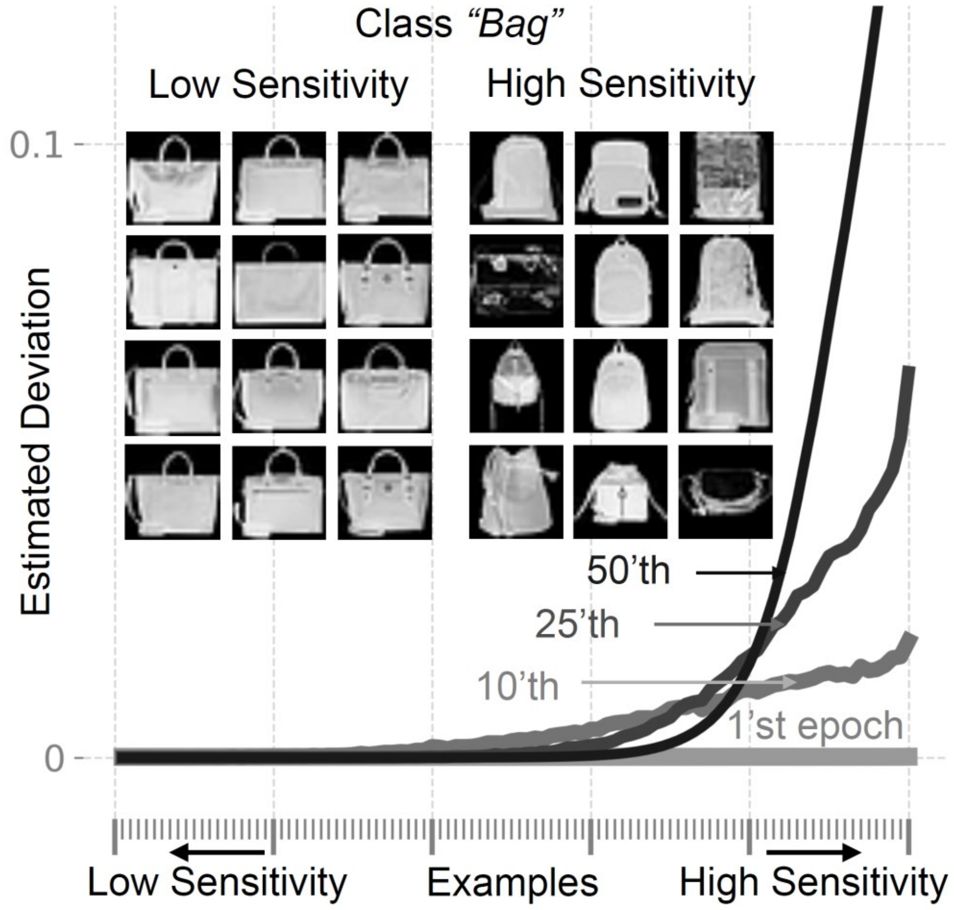


Value of each sample

University of Reading

- Related but different to explainable AI
  - We want to evaluate the informativeness for the final weights of a subset of training samples,
  - Feature selection aims to quantify the informativeness for the task variable of a subset of features
- Most literature on ML classifiers – unusual data has most influence on weights
- Information extracted depends on the network architecture, initialization, and the training procedure (i.e. the training algorithm).

Nickl et al 2023



(b) Evolution of sensitivities during training

Conclusions and questions

- Properly accounting for observation error correlations in data assimilation can result in more accurate analyses and analysis information on appropriate scales.

- Large, full, matrix-vector products required – effective example for Doppler radar winds, but how can we compute these efficiently more generally?

- How observations influence ML weights (training) and predictions is an open question – how can we ensure that we maximize information content?