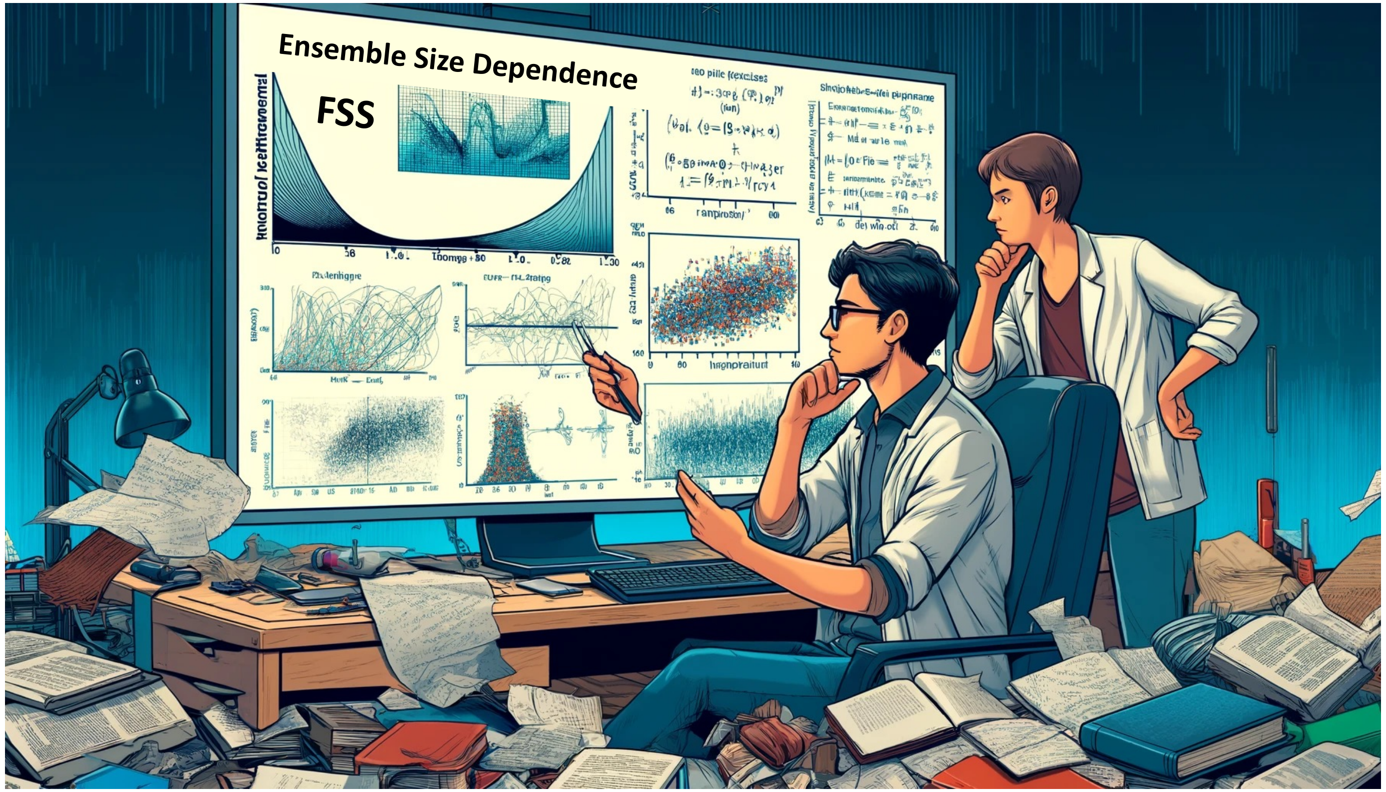
The Fractions Skill Score (FSS) for ensemble verification

Ludwig Wolfgruber, *Tobias Necker, Lukas Kugler, Manfred Dorninger, Martin Weissmann, Stefano Serafin
Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria (*tobias.necker@univie.ac.at)

universität wien

imgw Institut für Meteorologie und Geophysik

EGU General Assembly #EGU24 14 -19 April 2024



DALL·E 2024-04-11 17.32.31 - A comic image showcasing two young scientists sitting in front of a computer, deep in thought and collaboration.

We know: Larger ensembles provide better estimates of forecast probabilities

We wondered: Why does a larger ensemble not yield a higher FSS score?

In 2022, we decided to assess the ensemble size dependence of the FSS

Goals for this talk

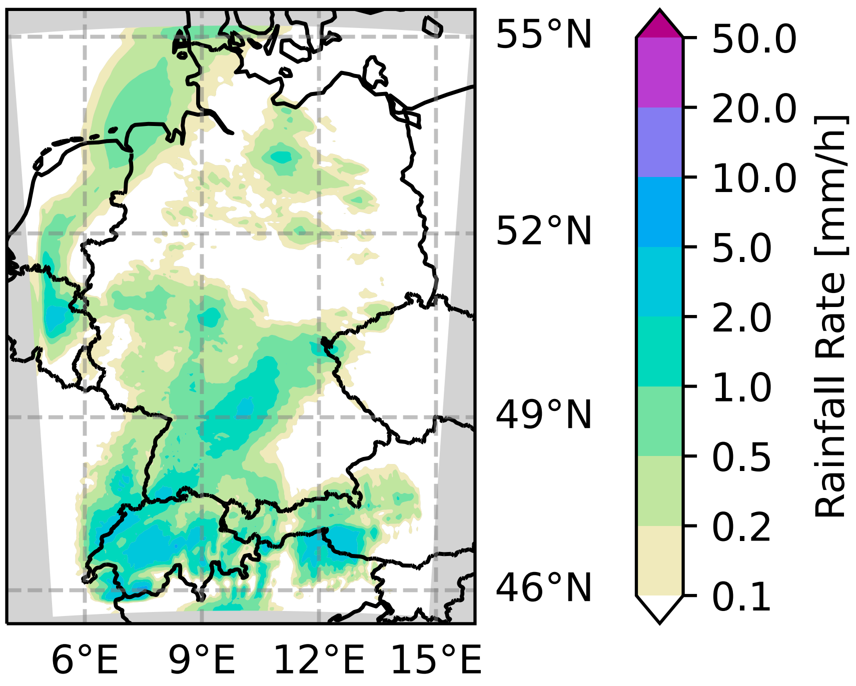
Motivation - Why using the FSS for verification?

Theory - How to compute an ensemble-based FSS?

Results - Comparison of ensemble-based FSS options

Summary - What can you expect in our publication?

Ensemble mean precipitation
960-member ensemble



Motivation: Why using the Fractions Skill Scores (FSS) for verification?

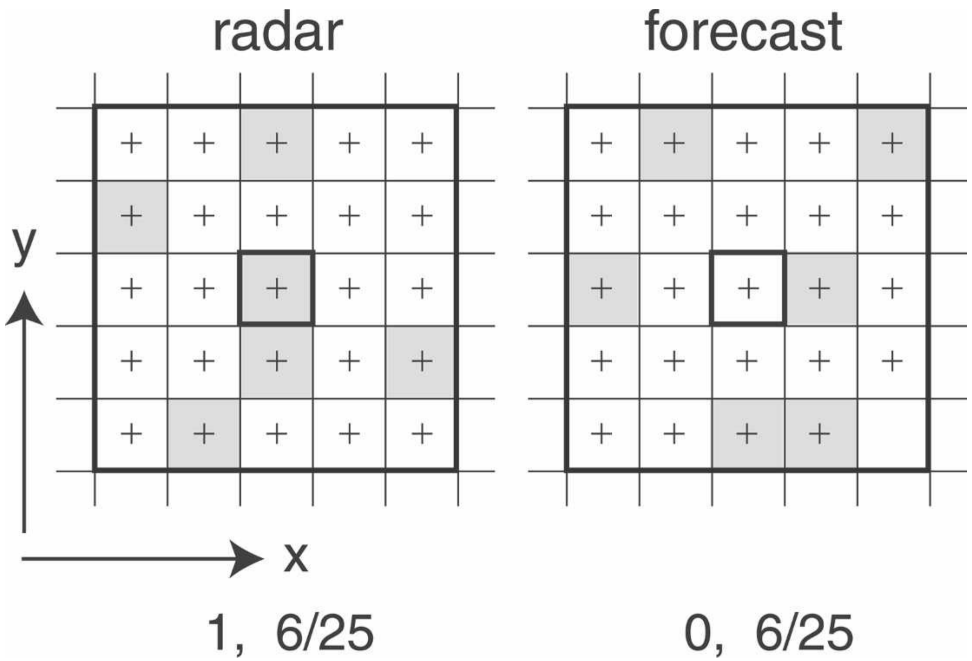
Problem - Double Penalty Errors
- Forecasts often do not overlap on grid scale and are penalized twice for deviations to observations

Solution - Spatial Verification Approaches
- We focus on the FSS - a neighbourhood verification method
- Challenge: How to apply the FSS for ensemble verification?

Knowledge gap:
- Different ensemble-based FSS approaches have been proposed, however, no systematic evaluation has been done
- Missing crucial knowledge of ensemble size dependence

Our goal:
- Better understanding of the FSS in context of probabilistic forecast verification and its ensemble size dependence

Examples of neighborhood fractions for precipitation



from Roberts and Lean (2008)

Theory: How to compute the FSS of a deterministic forecast?

(Roberts and Lean 2008)

Binary Probability: $$\mathrm{BP}_i = \begin{cases} 1 & \text{if } F_i \geq q \\ 0 & \text{otherwise} \end{cases}$$

Threshold / Frequency

"free parameters"

Neighborhood size

Neighborhood Probability: $$\mathrm{NP}_i = \frac{1}{N_b} \sum_{m=1}^{N_b} \mathrm{BP}_{m,i}$$

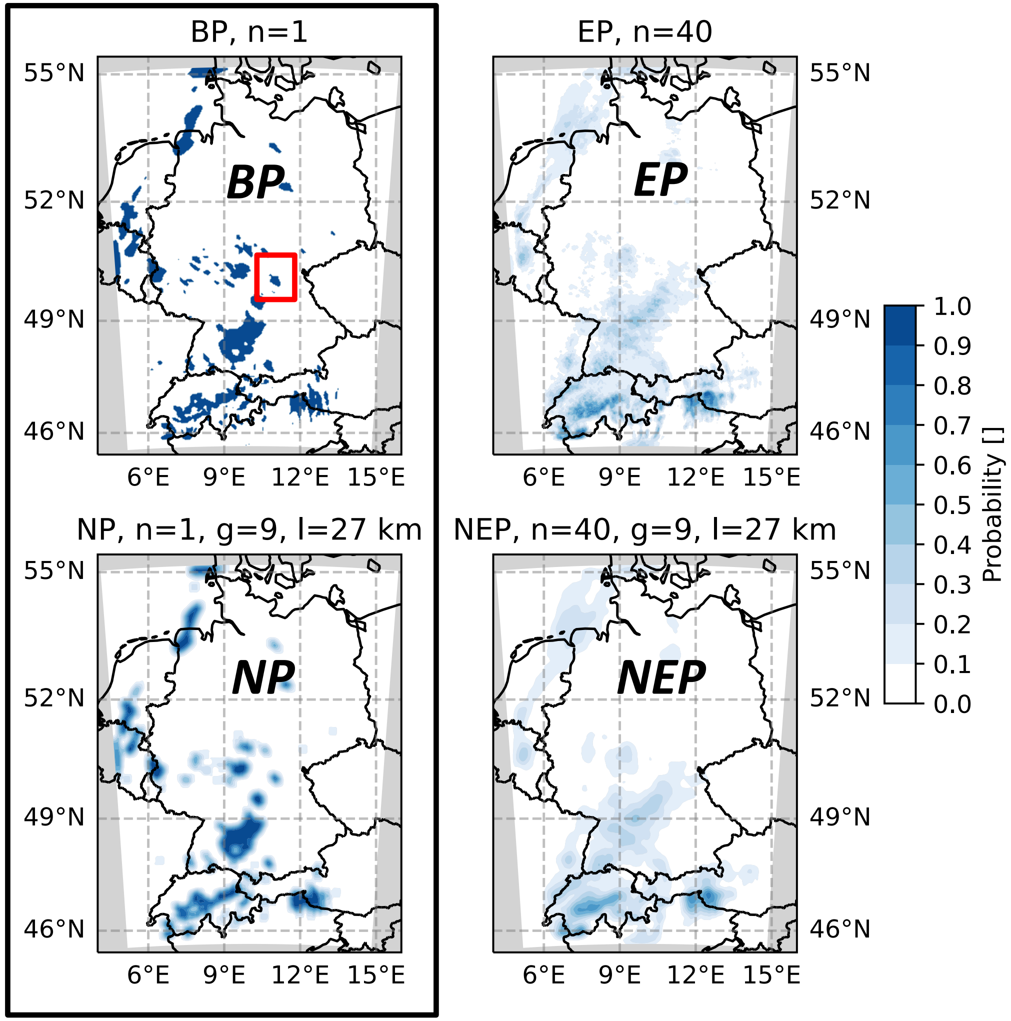
Fractions Brier Score: $$\mathrm{FBS} = \frac{1}{N_v} \sum_{i=1}^{N_v} (\mathrm{NP}_{i,f} - \mathrm{NP}_{i,o})^2$$

Worst FBS: $$\mathrm{WFBS} = \frac{1}{N_v} \left[\sum_{i=1}^{N_v} \mathrm{NP}_{i,f}^2 + \sum_{i=1}^{N_v} \mathrm{NP}_{i,o}^2 \right]$$

Fractions Skill Score: $$\mathrm{FSS} = 1 - \frac{\mathrm{FBS}}{\mathrm{WFBS}}$$

FSS ranges from **0 (no skill)** to **1 (perfect skill)**

Convolutions of precipitation fields



Theory: How to account for additional ensemble dimension?

(Mittermaier MP. 2007, Roberts and Lean 2008, Schwartz et al. 2010, Duc et al. 2013., Dey et al. 2014, Necker et al. 2024)

emFSS

Binary Probability: $$\mathrm{BP}_i = \begin{cases} 1 & \text{if } F_i \geq q \\ 0 & \text{otherwise} \end{cases}$$

pFSS

Neighborhood Probability: $$\mathrm{NP}_i = \frac{1}{N_b} \sum_{m=1}^{N_b} \mathrm{BP}_{m,i}$$

$$\boxed{\mathrm{NEP}_{i,f} = \frac{1}{n} \sum_{k=1}^{n} \mathrm{NP}_{i,k,f}}$$

Fractions Brier Score: $$\mathrm{FBS} = \frac{1}{N_v} \sum_{i=1}^{N_v} (\mathrm{NP}_{i,f} - \mathrm{NP}_{i,o})^2$$

agFSS

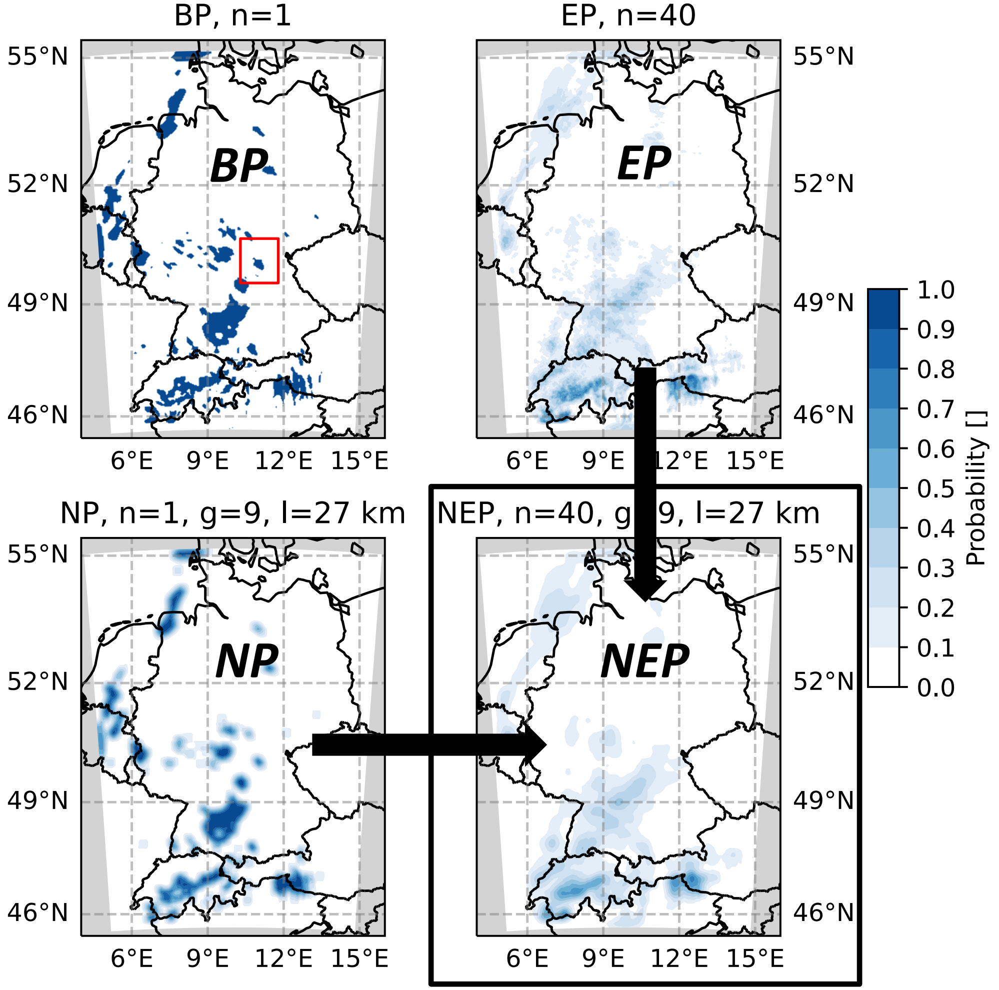
Worst FBS: ***Ensemble Pooling*** $$\mathrm{WFBS} = \frac{1}{N_v} \left[\sum_{i=1}^{Nv} \mathrm{NP}_{i,f}^2 + \sum_{i=1}^{Nv} \mathrm{NP}_{i,o}^2 \right]$$

avFSS

Fractions Skill Score: $$\mathrm{FSS} = 1 - \frac{\mathrm{FBS}}{\mathrm{WFBS}}$$

We expect all four approaches to behave differently with ensemble size given non-linear operations in the FSS computation, e.g. thresholding!

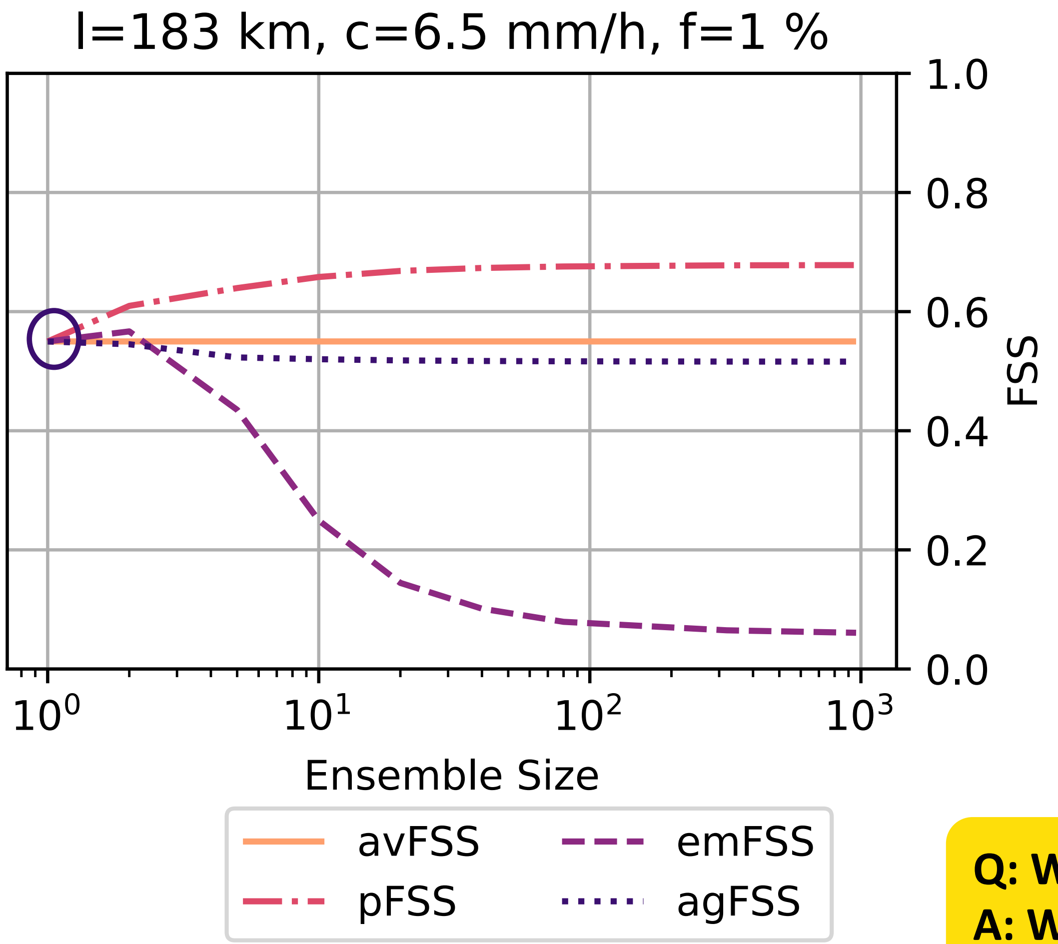
Convolutions of precipitation fields



NEP: Neighborhood Ensemble Probability

Result: Different FSS formulations reveal distinct behaviour with ensemble size

FSS approaches as function of ensemble size

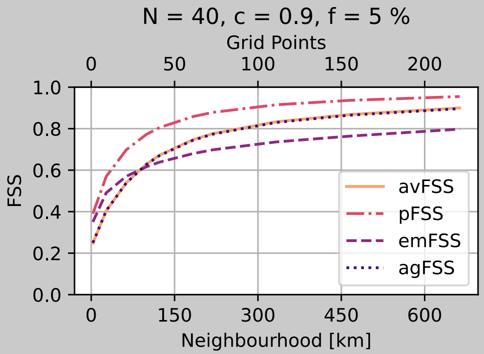


Conclusions:

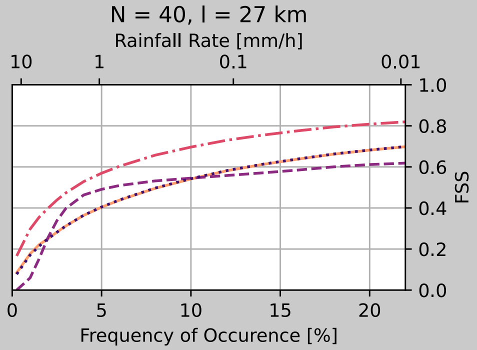
- Agreement: All approaches converge to the same deterministic solution
- **averaged (avFSS)** or **aggregated (agFSS) FSS**:
  - almost constant with ensemble size
- **ensemble mean FSS (emFSS)**:
  - Non-linear / sub-optimal
- **probabilistic FSS (pFSS) using NEP**:
  - Increases with ensemble size, well-behaved, and able to exploit the prob. information
  - Reveals similarities to Brier Skill Score (BSS)

Q: Why did our larger ensemble not yield a higher FSS score?
A: We computed the "agFSS", which is almost constant with ens. size

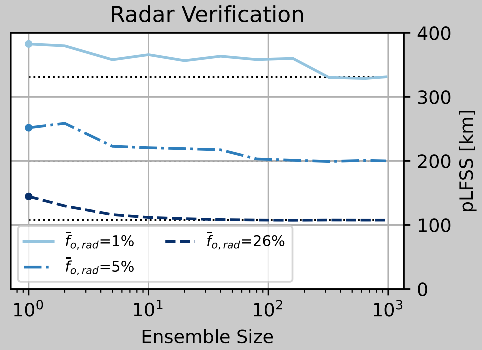
Further main results from our paper



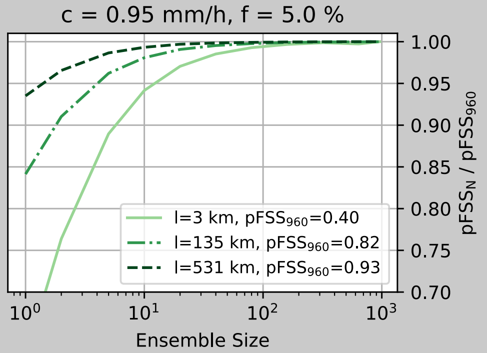
Dependence on neighbourhood size



Dependence on frequency of occurrence



Probabilistic skillful spatial scales

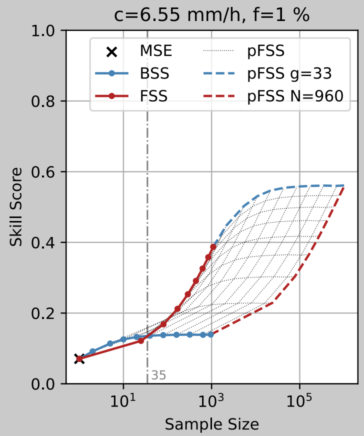


Saturation of the probabilistic FSS (pFSS) with ensemble size

$$BSS_N = BSS_\infty \left(1 - \frac{1}{N}\right) + \frac{BSS_1}{N}$$

$$pFSS_N = pFSS_\infty \left(1 - \frac{1}{N^\alpha}\right) + pFSS_1 \left(\frac{1}{N^\alpha}\right)$$

Behaviour of pFSS with ensemble size



Comparison of scores:

- **MSE**
- **BSS**
- **FSS**
- **pFSS**

Dependence of forecast skill on neighbourhood or ensemble size

Main Research Questions We Answer In Our Paper

doi

How do different ensemble-based FSS approaches behave with ensemble size, neighbourhood size, and frequency of occurrence?

How does the probabilistic FSS (pFSS) depend on ensemble size, and how to predict its behaviour?

How do the ensemble and neighbourhood size influence the forecast skill?

Our recommendation: **Use the probabilistic FSS (pFSS) for ensemble forecast verification!**

Preprint available online: Necker et al. 2024: The fractions skill score for ensemble forecast verification. *Authorea;* February 23, 2024. DOI: 10.22541/au.169169008.89657659/v2 (Submitted to QJRMS)

References

Mittermaier MP. **2007**. Improving short-range high-resolution model precipitation forecast skill using time-lagged ensembles. Quart. J. Roy. Meteor. Soc. 133(627): 1487–1500, doi:10.1002/qj.135

Roberts NM, Lean HW. **2008**. Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. Mon.Wea. Rev. 136(1): 78–97, doi:10.1175/2007MWR2123.1

Schwartz CS, Kain JS, Weiss SJ, Xue M, Bright DR, Kong F, Thomas KW, Levit JJ, Coniglio MC, Wandishin MS. **2010**. Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. Wea. Forecasting 25(1): 263–280, doi:10.1175/2009WAF2222267.

Duc L, Saito K, Seko H. **2013**. Spatial-temporal fractions verification for high-resolution ensemble forecasts. Tellus A Dynamic Meteorology and Oceanography 65(1): 18 171, doi:10.3402/tellusa.v65i0.18171

Dey SRA, Leoncini G, Roberts NM, Plant RS, Migliorini S. **2014**. A Spatial View of Ensemble Spread in Convection Permitting Ensembles. Mon.Wea. Rev. 142(11): 4091–4107, doi:10.1175/MWR-D-14-00172.1

Necker T, Geiss S, Weissmann M, Ruiz J, Miyoshi T, Lien GY. **2020**. A convective-scale 1,000-member ensemble simulation and potential applications. Quart. J. Roy. Meteor. Soc. 146(728): 1423–1442, doi:10.1002/qj.3744

Necker T, Wolfgruber L, Kugler L, Dorninger M, Weissmann M, Serafin S, **2024**: The fractions skill score for ensemble forecast verification. Authorea; February 23, 2024. DOI: 10.22541/au.169169008.89657659/v2

Backup 0: Comparison of four different ensemble-based FSS approaches

1. **Probabilistic FSS (pFSS)**: Using Neighborhood Ensemble Probabilities (NEPs)

$$\mathrm{pFSS} = 1 - \frac{\sum_{i=1}^{I}\left(\sum_{n=1}^{N} \mathrm{NP}_{in,f} - \mathrm{NP}_{i,o}\right)^2}{\sum_{i=1}^{I}\left(\sum_{n=1}^{N} \mathrm{NP}_{in,f}\right)^2 + \sum_{i=1}^{I} \mathrm{NP}_{i,o}^2} \qquad \mathrm{NEP}_{i,f} = \frac{1}{JN}\sum_{j=1}^{J}\sum_{n=1}^{N} \mathrm{BP}_{ijn,f}$$

2. **Ensemble mean FSS (emFSS)**: Evaluating the ensemble mean forecast

$$\mathrm{EM}_i = \frac{1}{N}\sum_{n=1}^{N} F_{in}$$

3. **Ensemble aggregated FSS (agFSS)**: Aggregation as ensemble pooling method

$$\mathrm{agFSS} = 1 - \frac{\sum_{n=1}^{N}\sum_{i=1}^{I}\left(\mathrm{NP}_{in,f} - \mathrm{NP}_{in,o}\right)^2}{\sum_{n=1}^{N}\sum_{i=1}^{I} \mathrm{NP}_{in,f}^2 + \sum_{i=1}^{I} \mathrm{NP}_{in,o}^2}$$

4. **Ensemble averaged FSS (avFSS)**: Computing the FSS for each member separately

$$\mathrm{avFSS} = 1 - \frac{1}{N}\sum_{n=1}^{N} \frac{\sum_{i=1}^{I}\left(\mathrm{NP}_{in,f} - \mathrm{NP}_{in,o}\right)^2}{\sum_{i=1}^{I} \mathrm{NP}_{in,f}^2 + \sum_{i=1}^{I} \mathrm{NP}_{in,o}^2}$$

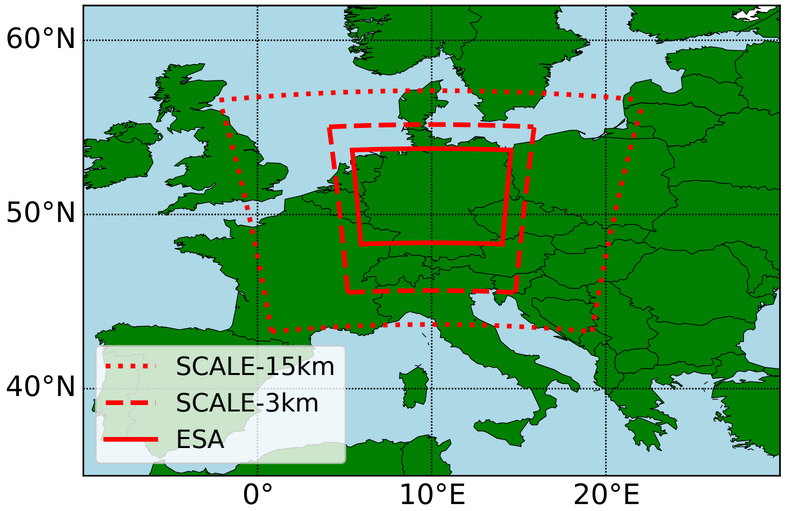
Four options exist to account for ensemble dimension "*n*"

Note: We expect that all four approaches behave differently with ensemble size given non-linear operations in the FSS computation:

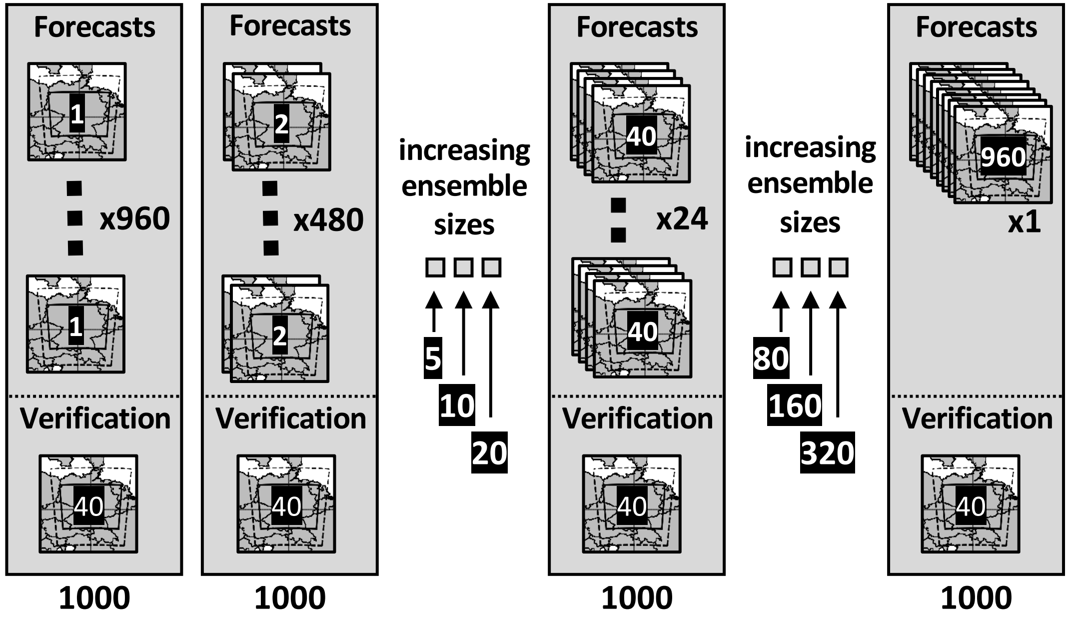
- Thresholding
- Taking second power
- Taking quotient

Backup 1: Experimental setup and subsampling

- **Model**: convective-scale 1000-member ensemble forecasts using SCALE-RM (Necker et al. 2020)
- **Domain**: 3 km grid spacing, 350x250 grid points with 30 levels centered over Germany
- **Period**: 5 days high impact weather in Mai/June 2016
- **Verification of precipitation:**
  - 10 forecasts at 3, 6, 9, 12 lead time
    → yields 40 verification time steps
  - 40 independent random members as „truth“
    → yields 1600 verification scenarios
- **Subsampling approach**:
  - Sub-samples randomly drawn without replacement from the pool of 960 members
  - Same information digested for each ensemble size

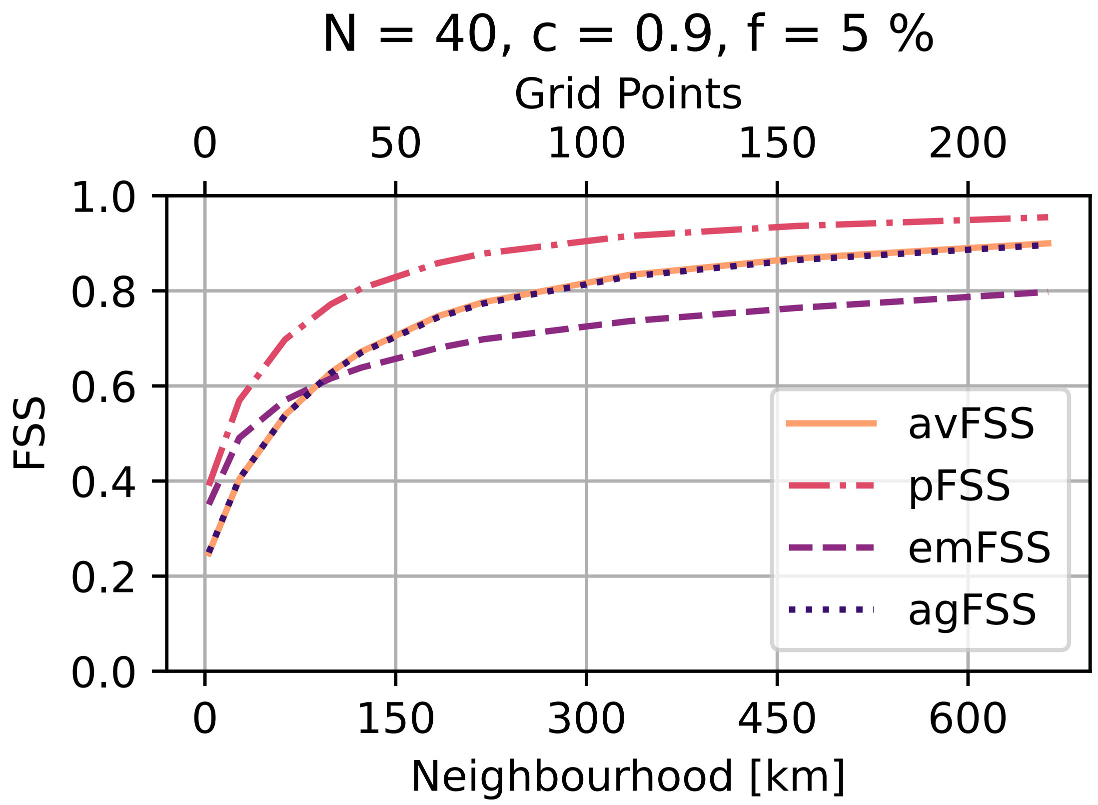


Subsampling strategy and model domain



Backup 2 Comparison: Behaviour of FSS approaches with neighborhood size

FSS approaches as function of neighborhood size



Theory: FSS is expected to asymptote to a specific FSS value (AFSS), which is determined by the frequency bias (FB):

$$\mathrm{AFSS} = \frac{2\mathrm{FB}}{\mathrm{FB}^2 + 1} = \frac{2 f_o f_f}{f_o^2 + f_f^2}$$

Behaviour with neighborhood size:

- Ensemble-based FSS approaches follow the theoretical expectation
- Approaches exhibit different AFSS values as the frequency biases changes with FSS formulation
- emFSS particularly sensitive to the frequency bias

Backup 3 Formula for ensemble size dependence of pFSS

$$\mathrm{BSS}_N = \mathrm{BSS}_\infty\left(1-\frac{1}{N}\right)+\frac{\mathrm{BSS}_1}{N}$$

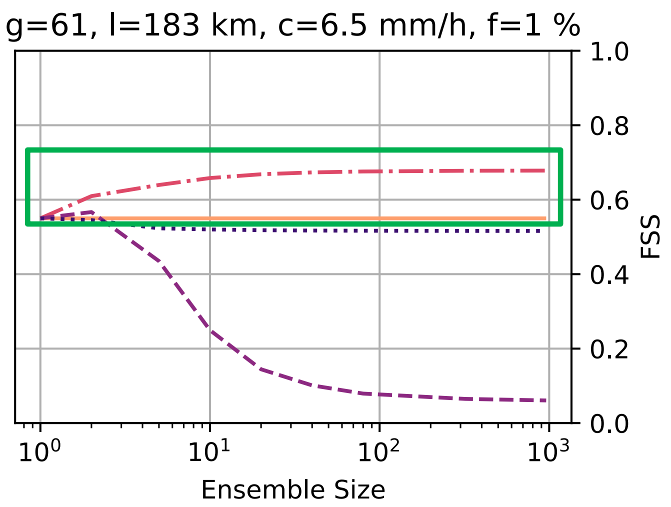
$$\boxed{\mathrm{pFSS}_N = \mathrm{pFSS}_\infty\left(1-\frac{1}{N^\alpha}\right)+\mathrm{pFSS}_1\left(\frac{1}{N^\alpha}\right)}$$

n: ensemble size | α: exponent / slope parameter
FSS^1: det. FSS | FSS$^\infty$: infinity FSS

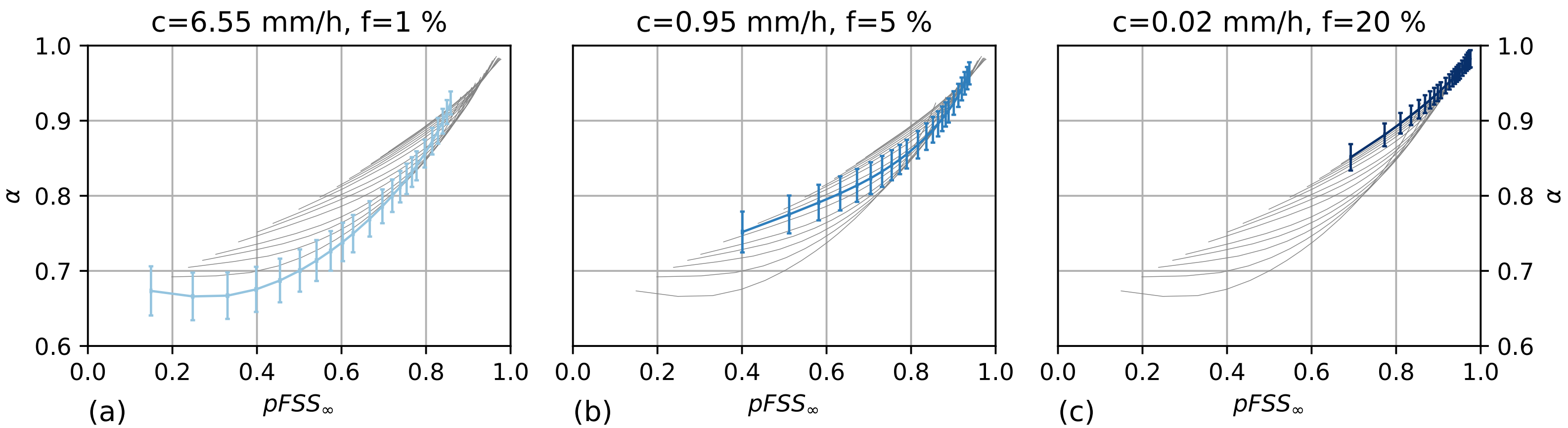
→ We can reformulated our **equation** to estimate the expected **FSS for reaching an infinite large ensemble**:

$$\mathrm{pFSS}_\infty = \frac{N^\alpha \mathrm{pFSS}_N - \mathrm{pFSS}_1}{N^\alpha - 1}$$

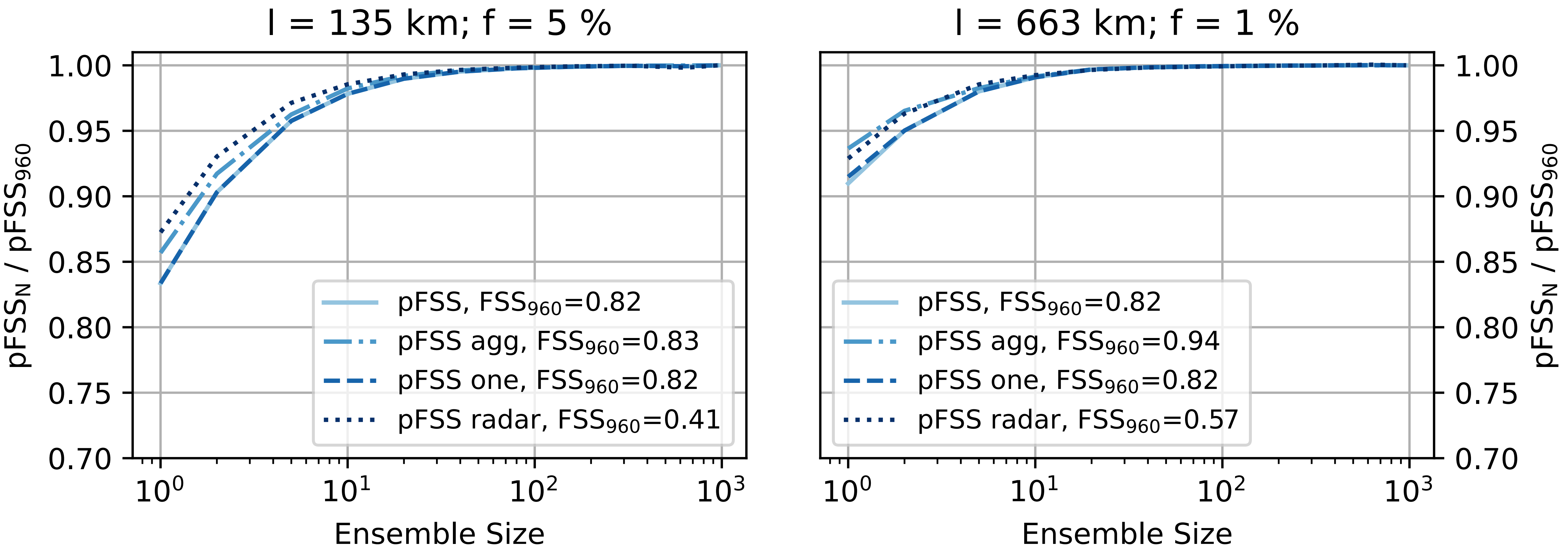
→ **Saturation of pFSS** with ensemble size depends on **slope parameter α**



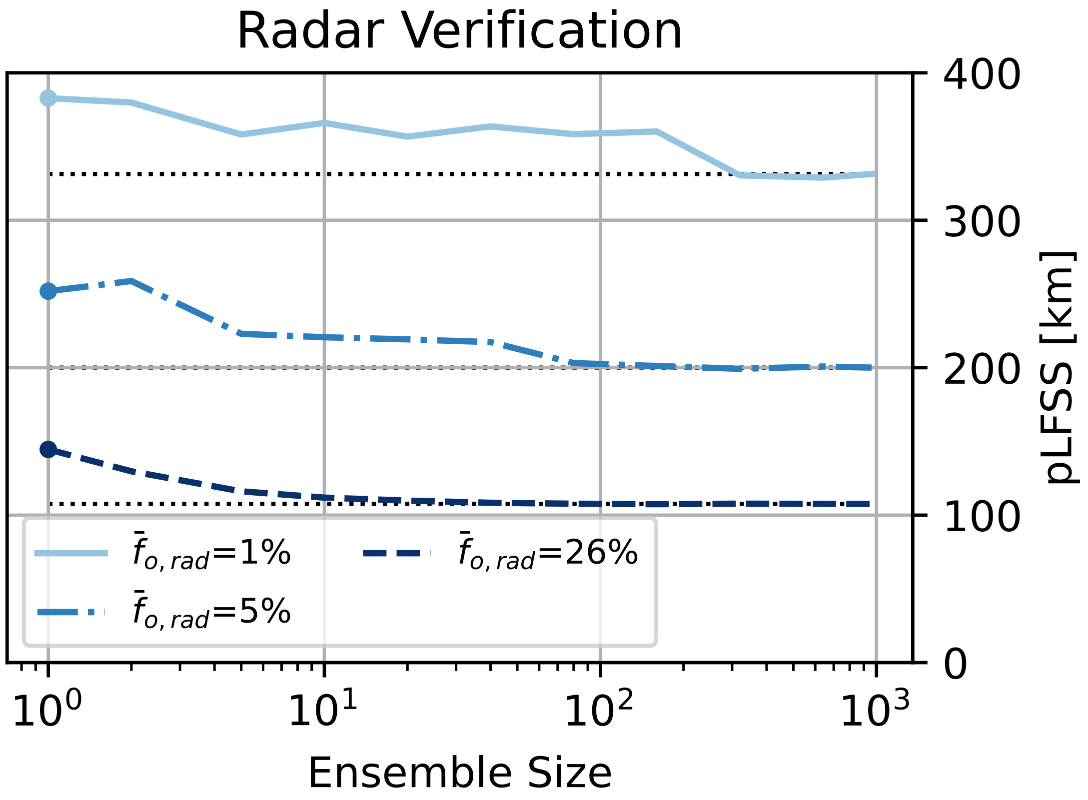
Exponent α as function of FSS$^\infty$, frequency of occurrence f (%), and neighborhood size

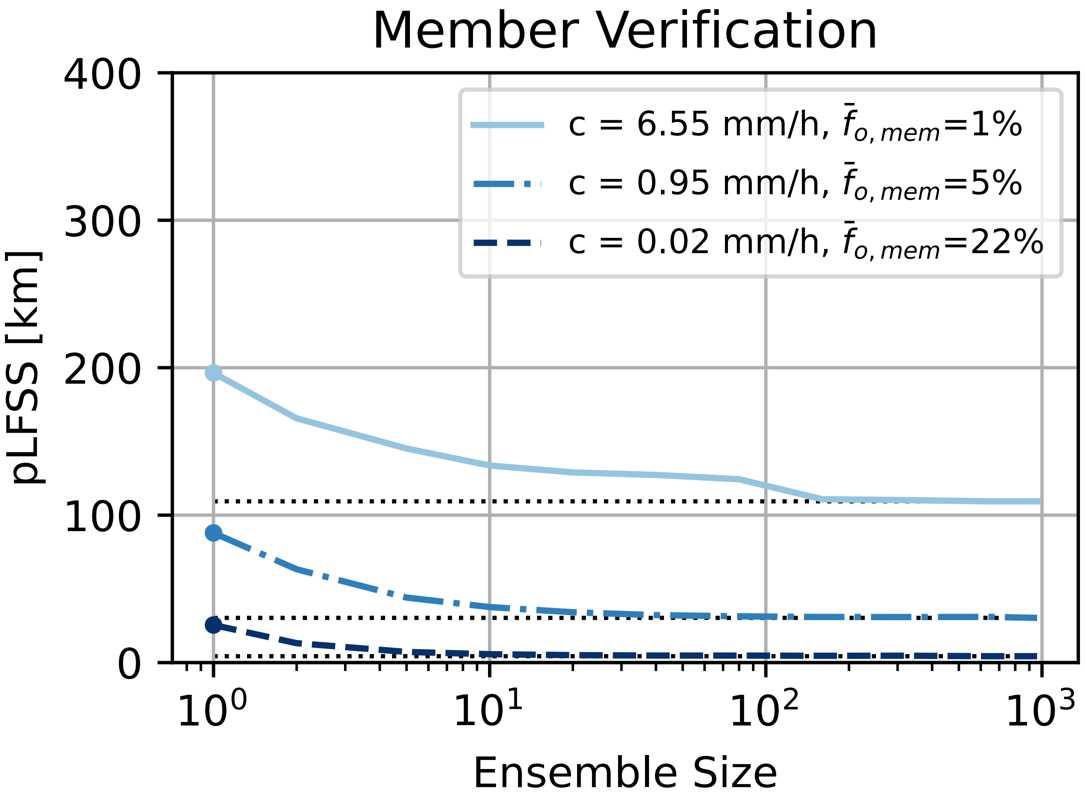


Backup 4 Sensitivity studies



Backup 5 Probabilistic Skillful Spatial Scales

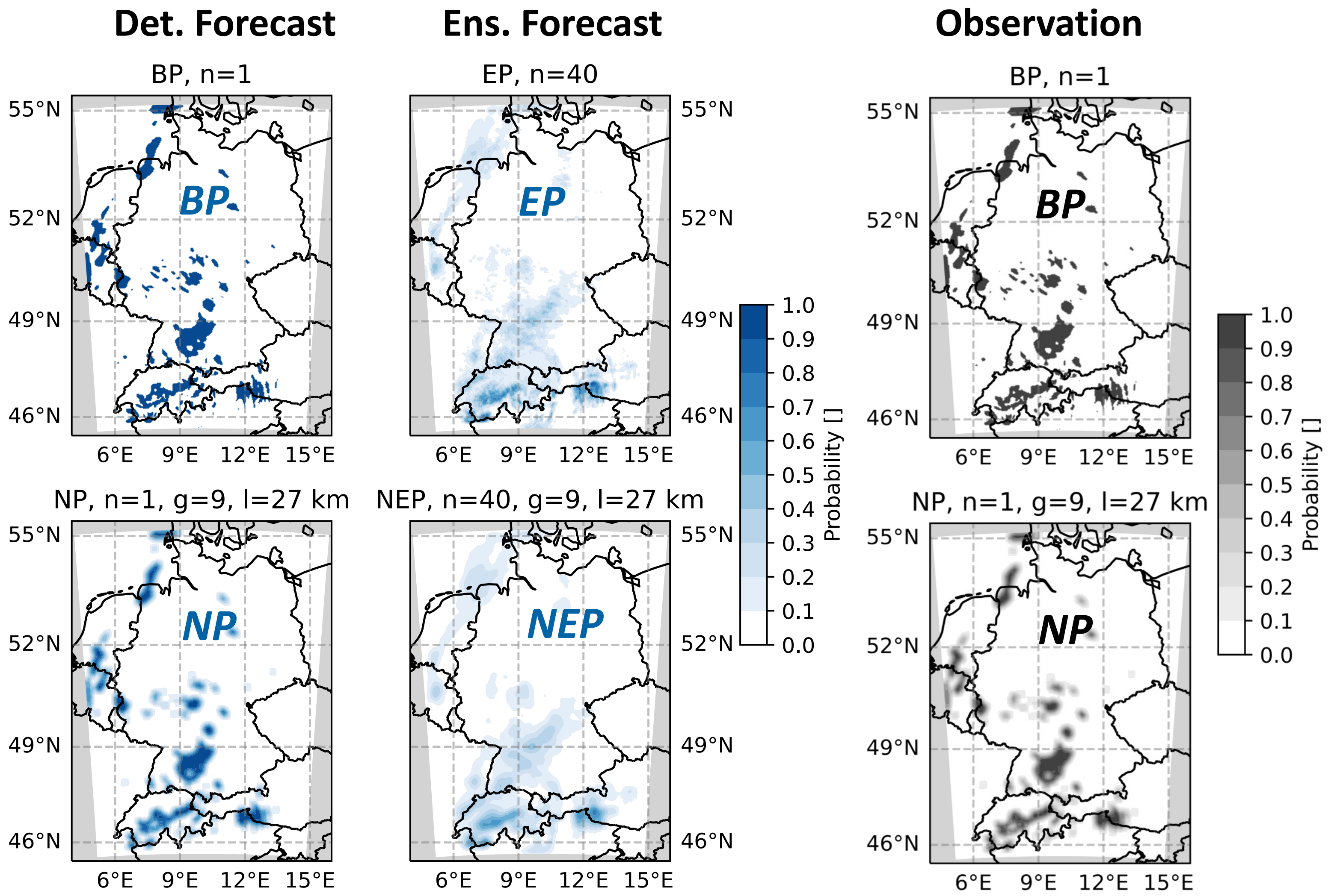




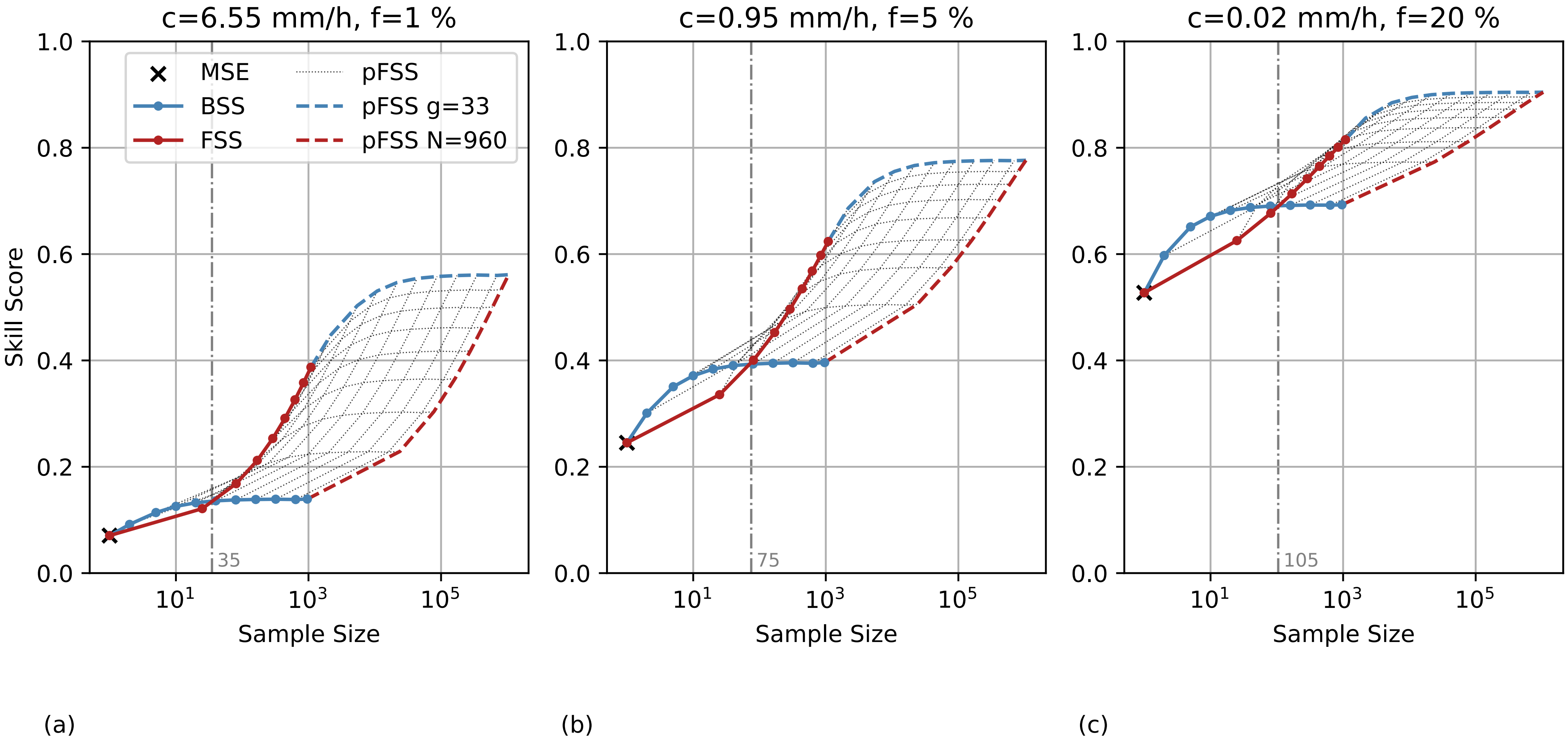
Backup 6 Comparison of different Skill Scores and underlying probabilities

Compared on next slide

Skill Score	Forecast	Observation
MSE	BP	BP
BSS	EP	BP
FSS	NP	NP
pFSS	NEP	NP
NEP-BSS	NEP	BP
BP-FSS	NP	BP

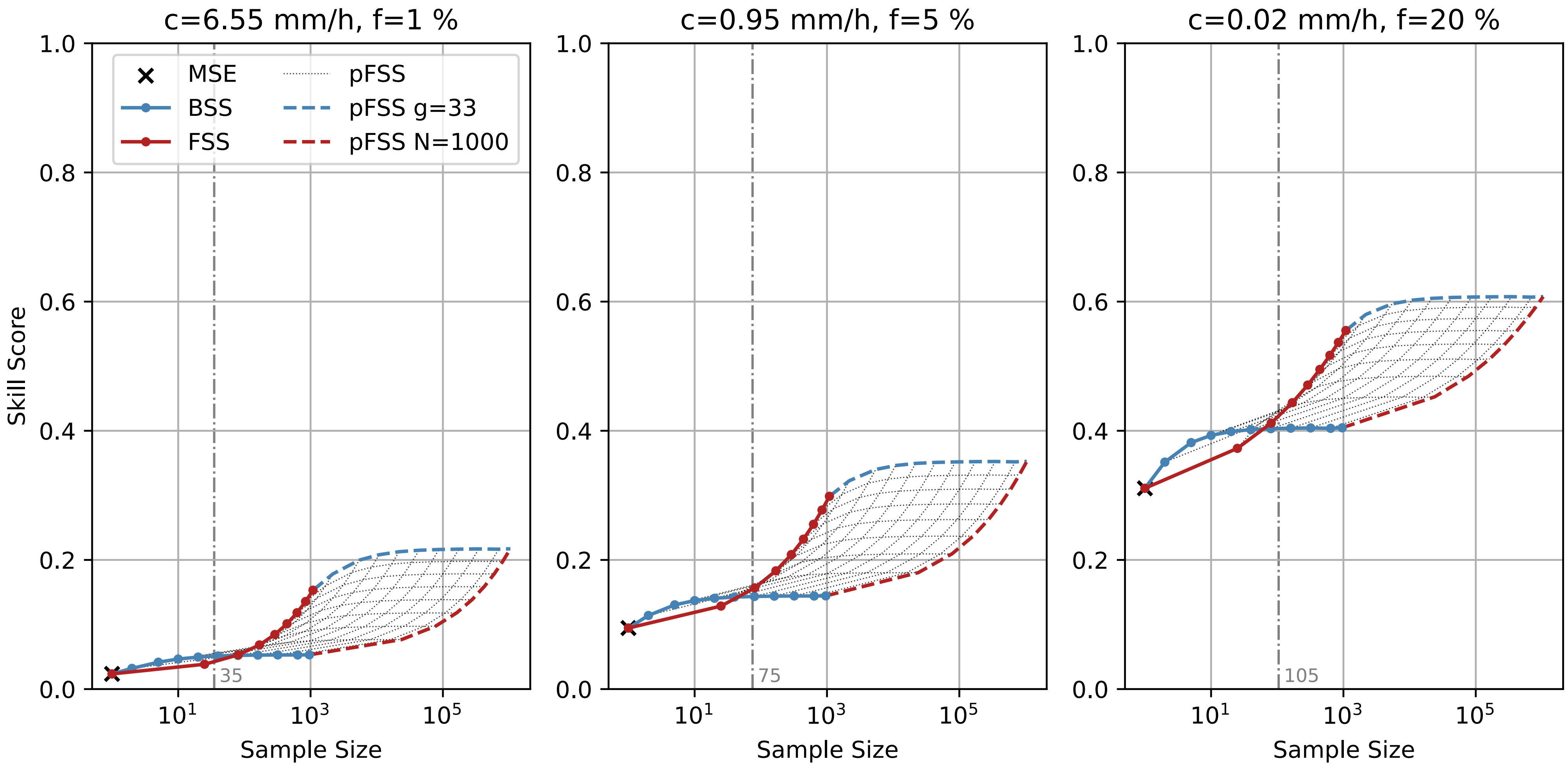


Backup 7 Comparison of MSE, BSS, FSS, and pFSS (member verification)



Sample size (x-axis) refers to the number of samples used to compute BP, EP, NP, or NEP

Backup 8 Comparison of MSE, BSS, FSS, and pFSS (radar verification)



Sample size (x-axis) refers to the number of samples used to compute BP, EP, NP, or NEP