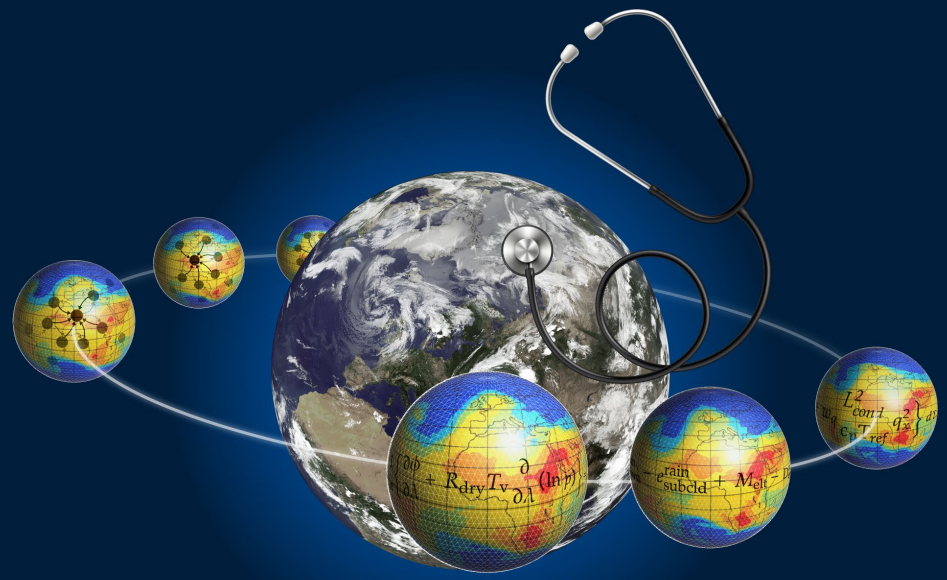
Workshop on Diagnostics for Global Weather Prediction

Reading | 9–12 September 2024

#diagnosticsWS



ECMWF

Scale-dependent evaluation and fair comparison of ensemble systems

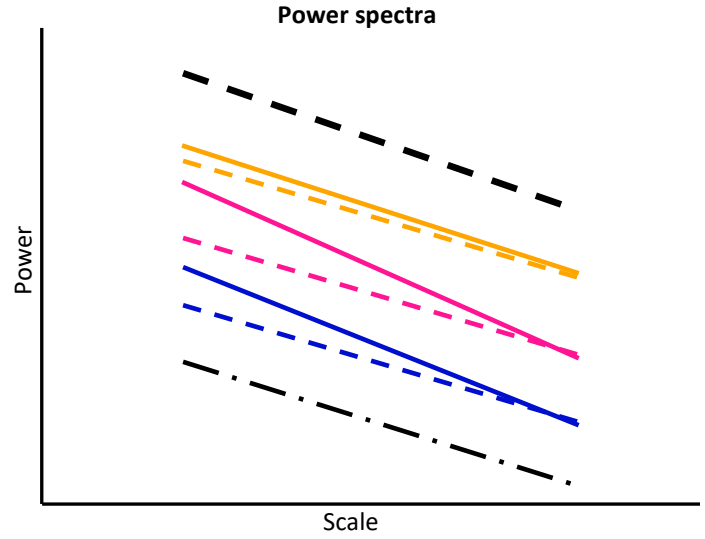
Workshop on Diagnostics for Global Weather Prediction

Mark Rodwell, Mariana Clare, Sarah-Jane Lock, Katrin Lonitz, Matthieu Chevallier

ECMWF

11 September 2024

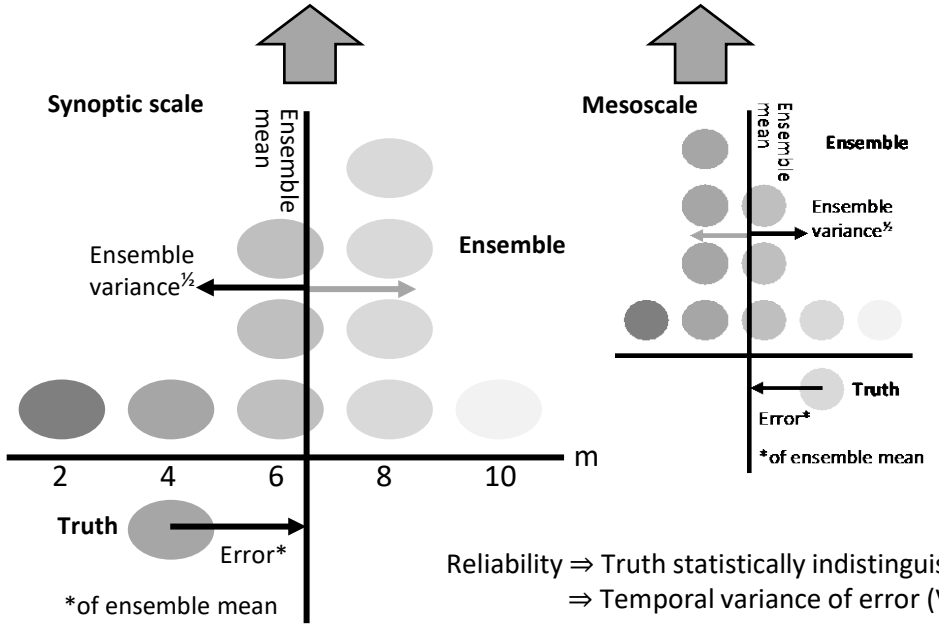
Creating power spectra to show the spatiotemporal evolution of ensemble spread and error

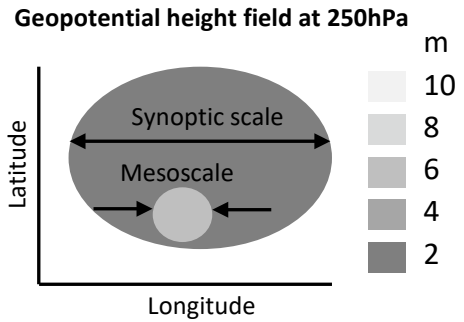


Sum over all scales

DAY	SME	VF	VE	VF+VA
1	4	34	25	37
2	10	148	104	151
5	50	1410	1407	1413
10	327	7367	7522	7370
15	580	11670	11215	11673
20	714	13317	12865	13319
40	794	14302	13600	14305

VA = Mean ensemble variance of analysis
VE = Temporal variance of ensemble-mean error
VF+VA = Mean variance of forecast + mean variance of verifying analysis
VMA = Temporal variance of mean analysis (≈ VE_{max})
SME = Squared temporal-mean of error ("squared bias")

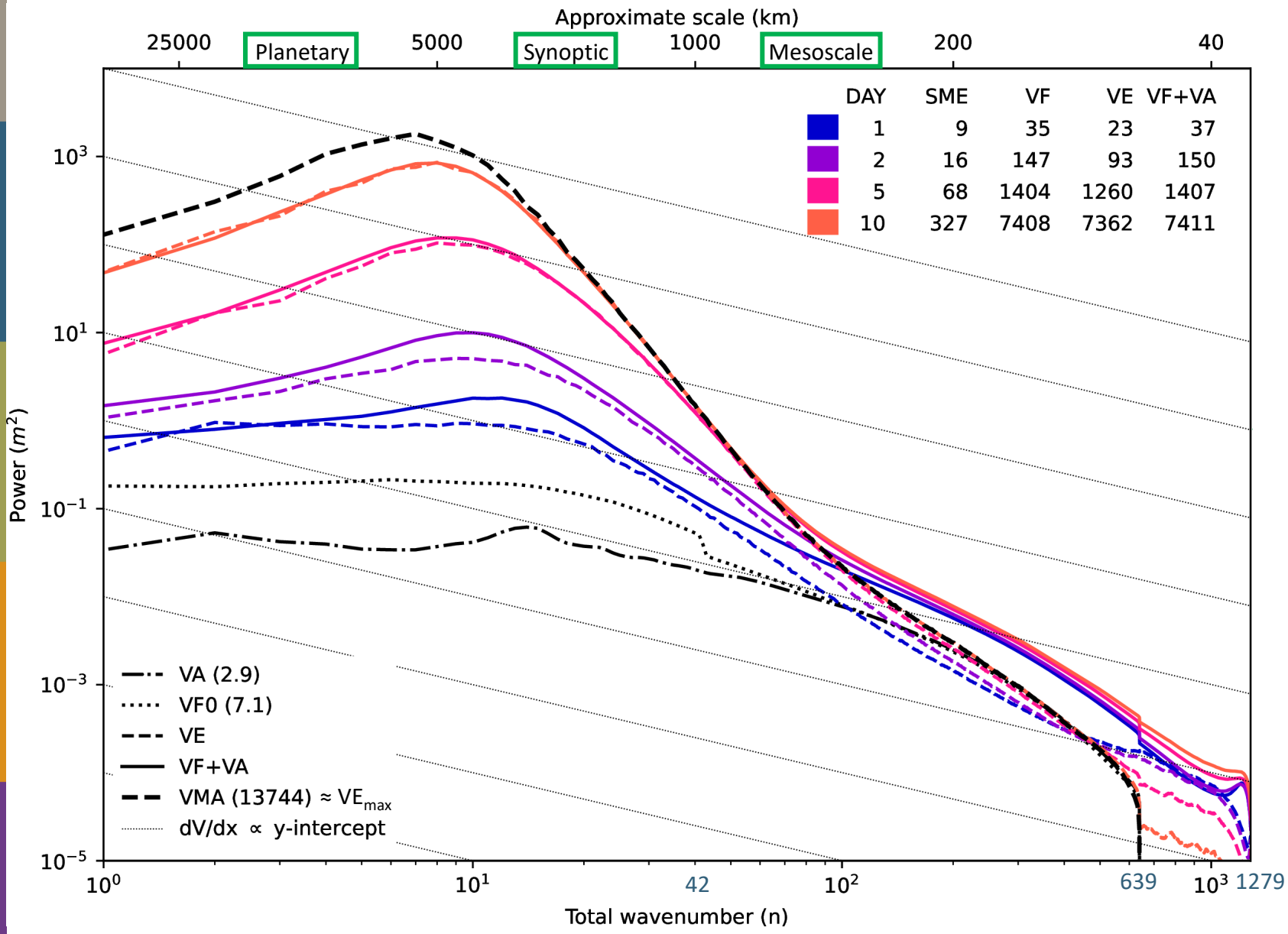




Reliability ⇒ Truth statistically indistinguishable from an ensemble member
⇒ Temporal variance of error (VE) = mean variance of forecast (VF) { +mean variance of verifying analysis (VA) }

Fields are decomposed into spherical harmonics. Power spectra show the variances of fields as a function of total wavenumber

Operations: Z250 power spectra of medium-range ensemble: Dec 2023 – Feb 2024 (DJF24)



Log-log plot

- Wavenumbers (n) more tightly packed towards right
- Diagonal contours indicate contribution to total variance per unit linear distance on x-axis
- Contour value is proportional to the y-intercept

EDA

- Maximum variance contribution at scales ≈ 1000km
- Global average total variance = 2.9m^2
- Spectral resolution n = 639 (≈60km), 50 members
- Curious wobbles at synoptic scales

ENS initial conditions

- A few hours forecast variance growth
- Singular vector (SV) perturbations at n ≤ 42
- Maximum Variance Contribution at ≈ 2000km

ENS days 1—10

- Spectral resolution n = 1279 (≈30km), 50 members

Synoptic scales

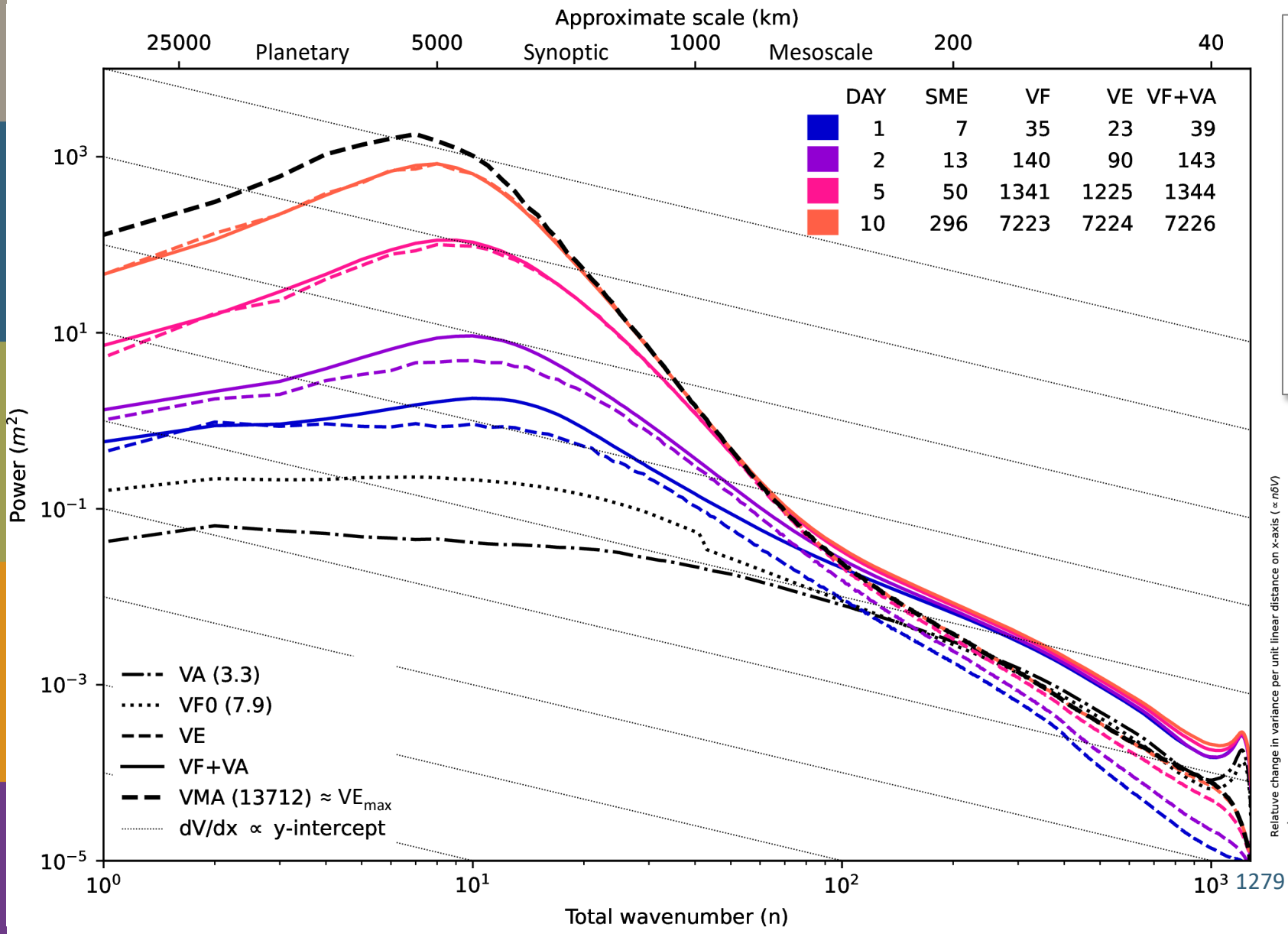
- Dominate medium-range variance (role of SVs?)
- Over-dispersive (in stormtracks) at days 1,2
- Better reliability at days 5,10
- Maximum error not reached by day 10

Smaller scales

- Errors quickly converge to theoretical maximum
- Forecast variance saturates at higher level
- Assumptions in spectral error-spread rel^n less valid?
- Disregard error and reliability for n > 639 as no EDA

Results based on 0UTC runs only for compatibility with experimental system

New IFS cycle: Z250 power spectra of the medium-range ensemble: DJF24 0Z



EDA

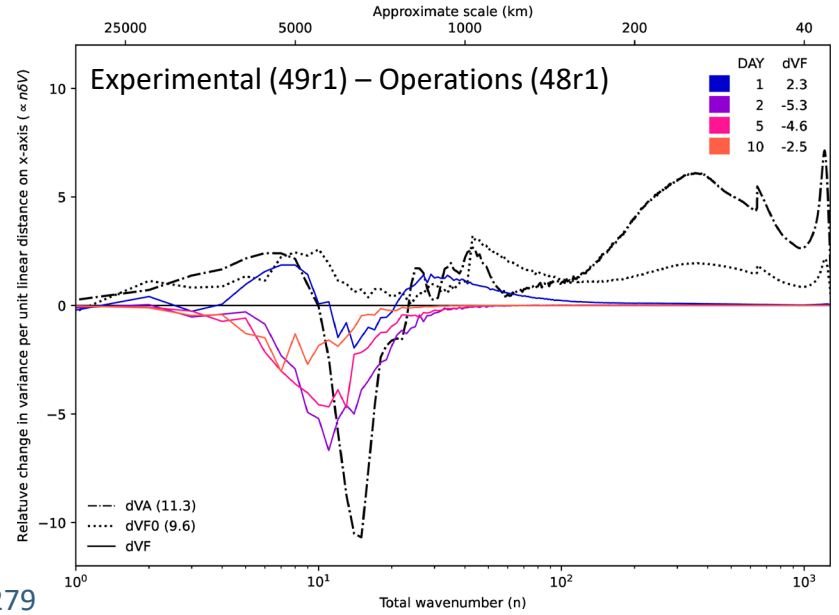
- The higher resolution EDA looks much better
- Reduced synoptic-scale wobbles
- Less drop-off in variance at scales 300 – 40 km)

ENS initial conditions

- Initialisation adds slightly more variance

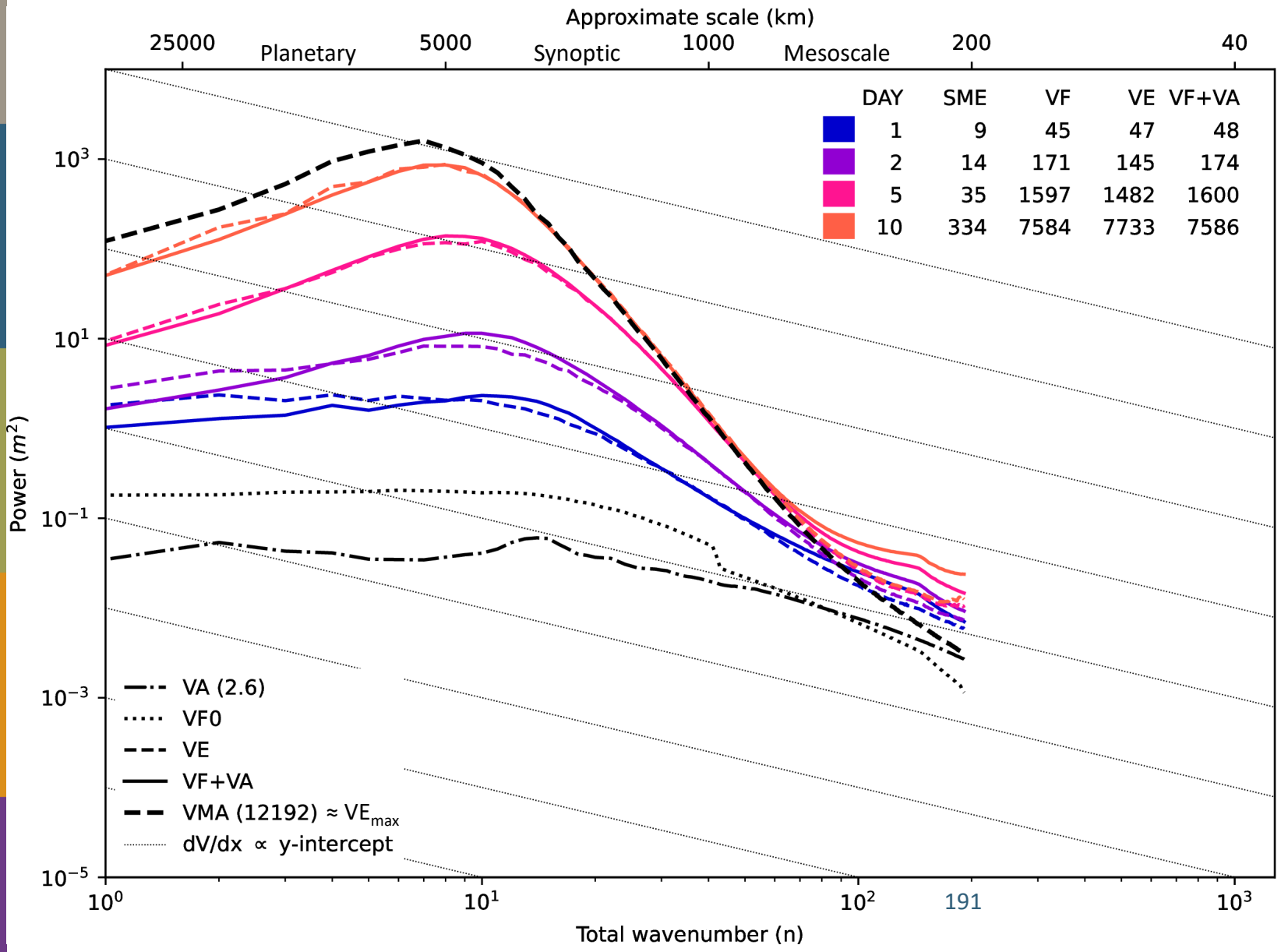
ENS days 1—10

- Reduction in synoptic-scale over-dispersion (due to major change in stochastic physics: SPPT ⟶ SPP?)
- Global budgets: reduced bias, error variance and better match between VE and VF+VA



In next cycle EDA has same resolution as ENS and looks much improved. ENS reliability and sharpness improves a little (due to new stoch. phys.?)

AIFS: Z250 power spectra of experimental Artificial Intelligence ensemble: DJF24 0/12 Z



AIFS ensemble (experimental; not real-time version)

- Resolution ~1° ⟶ T191 (≈210km)
- Timestep 6h
- Ensemble size = 8
- Trained on ERA5 1979—2017
- Optimisation based on proper scoring (CRPS)
- Retains random component ("stochastic physics")

ENS days 1—10

Planetary scales

- Under-dispersive at days 1,2 (worse than IFS)

Synoptic scales

- Over-dispersive at days 1,2 (better than IFS)

Smaller scales

- VE > VE_{max} (as VMF is not negligible)

All scales

- *AIFS more reliable but at higher variances than IFS*

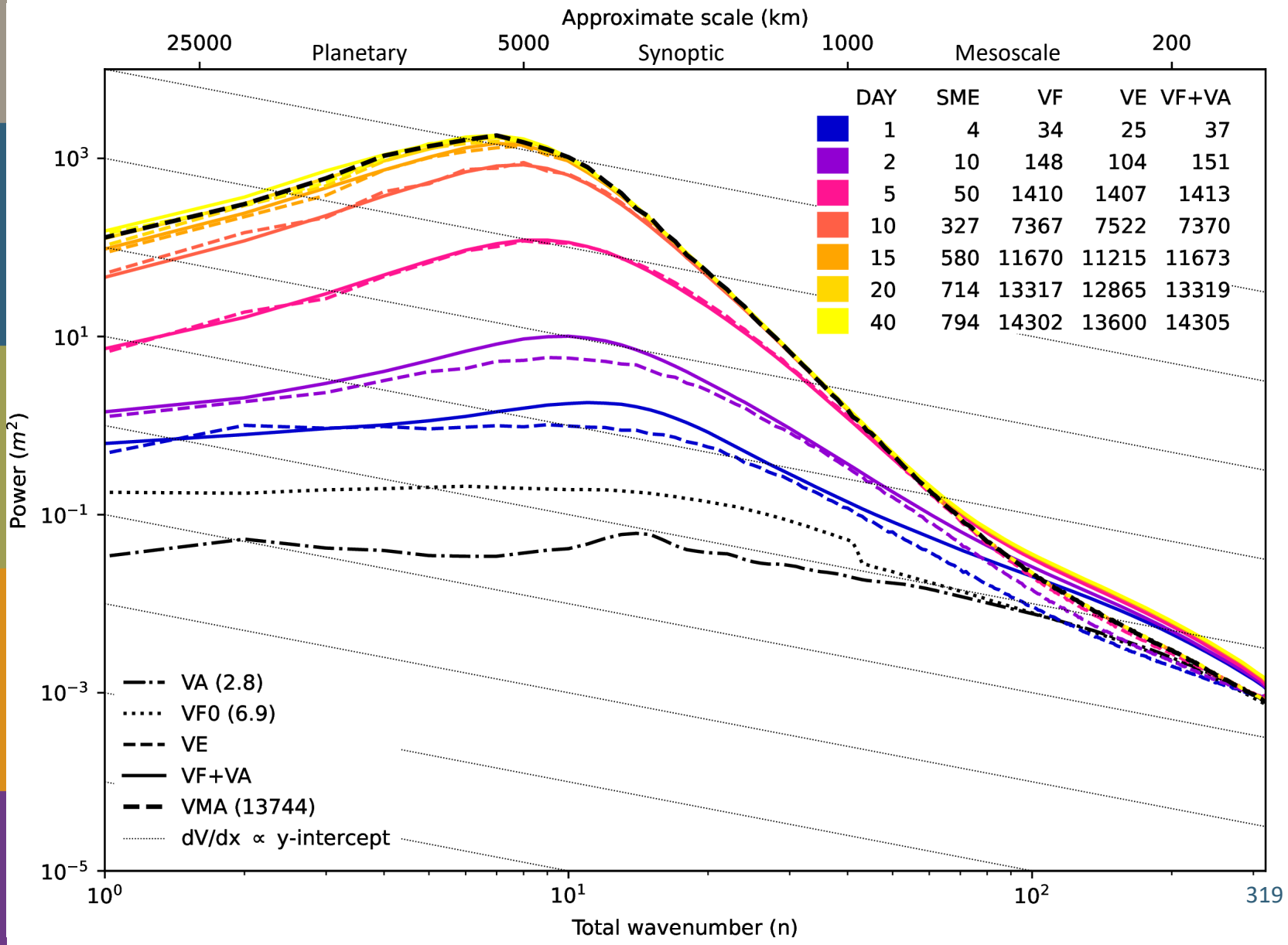
Corresponding table for IFS (truncated to T191)

DAY	SME	VF	VE	VF+VA
1	7	33	23	36
2	15	144	92	146
5	65	1370	1218	1372
10	312	7259	7149	7262

All curves and table values are scaled here to give unbiased estimators for VF

AIFS ensemble spread and error are larger than IFS, but more reliable overall. Curious errors at small scales

Sub-seasonal ranges: Z250 power spectra of operational ensemble: DJF24 0Z



DAY	SME	VF	VE	VF+VA
1	4	34	25	37
2	10	148	104	151
5	50	1410	1407	1413
10	327	7367	7522	7370
15	580	11670	11215	11673
20	714	13317	12865	13319
40	794	14302	13600	14305

ENS days 1—40

- Spectral resolution n = 319 (≈125km), 100 members
- Similar spatiotemporal behaviour to T1279 ENS
- Errors almost completely saturate at VMA by day 40
- T1279 ENS *error* comparison ⇒ Resolution matters

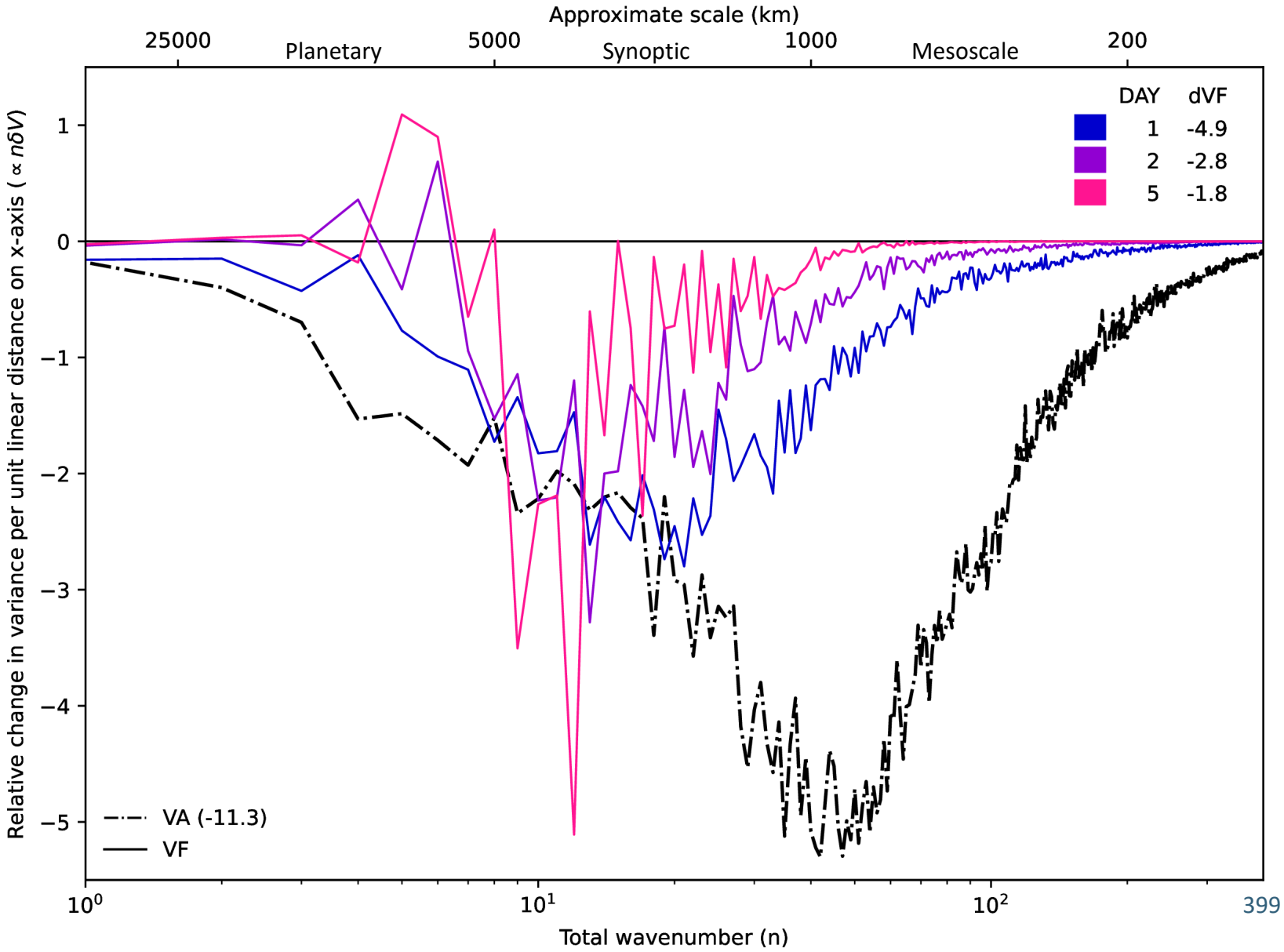
Note that the interhemispheric mode in Z250 displays additional temporal variance associated with the seasonal cycle. This was removed with a quadratic fit. (Just affects Z_1^0, i.e. wavenumber 1 = 10^0).

Corresponding table for T1279 ENS (truncated to T319)

DAY	SME	VF	VE	VF+VA
1	9	34	23	37
2	16	147	93	150
5	68	1403	1260	1406
10	327	7407	7362	7410
15	625	11733	11286	11735

Errors finally appear to saturate at ~ day 40. Remarkable agreement

Impact of new observations: Z250 % change in power spectra



ENS days 0—5

- 30,000 new Radio Occultation observations each analysis cycle (from IROWG ROMEX initiative)
- Error and spread reduced
- Informative to see spatiotemporal impact on ENS
- 11% reduction in EDA variance, centred ~ 1000km
- Biggest impact seen at ~ 4000km by day 5
- Bigger impact in northern midlatitudes (not shown)

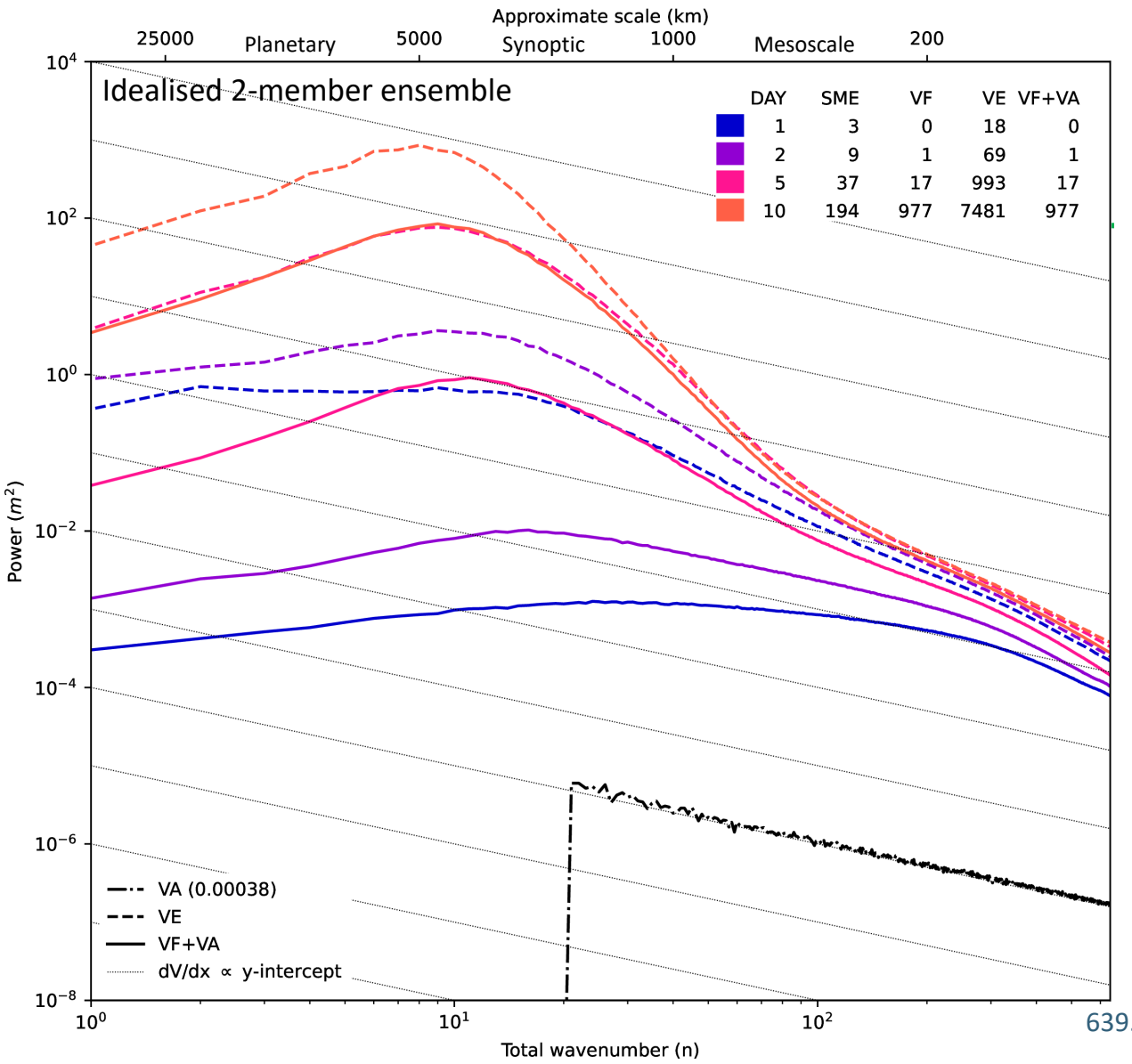
Experimental details

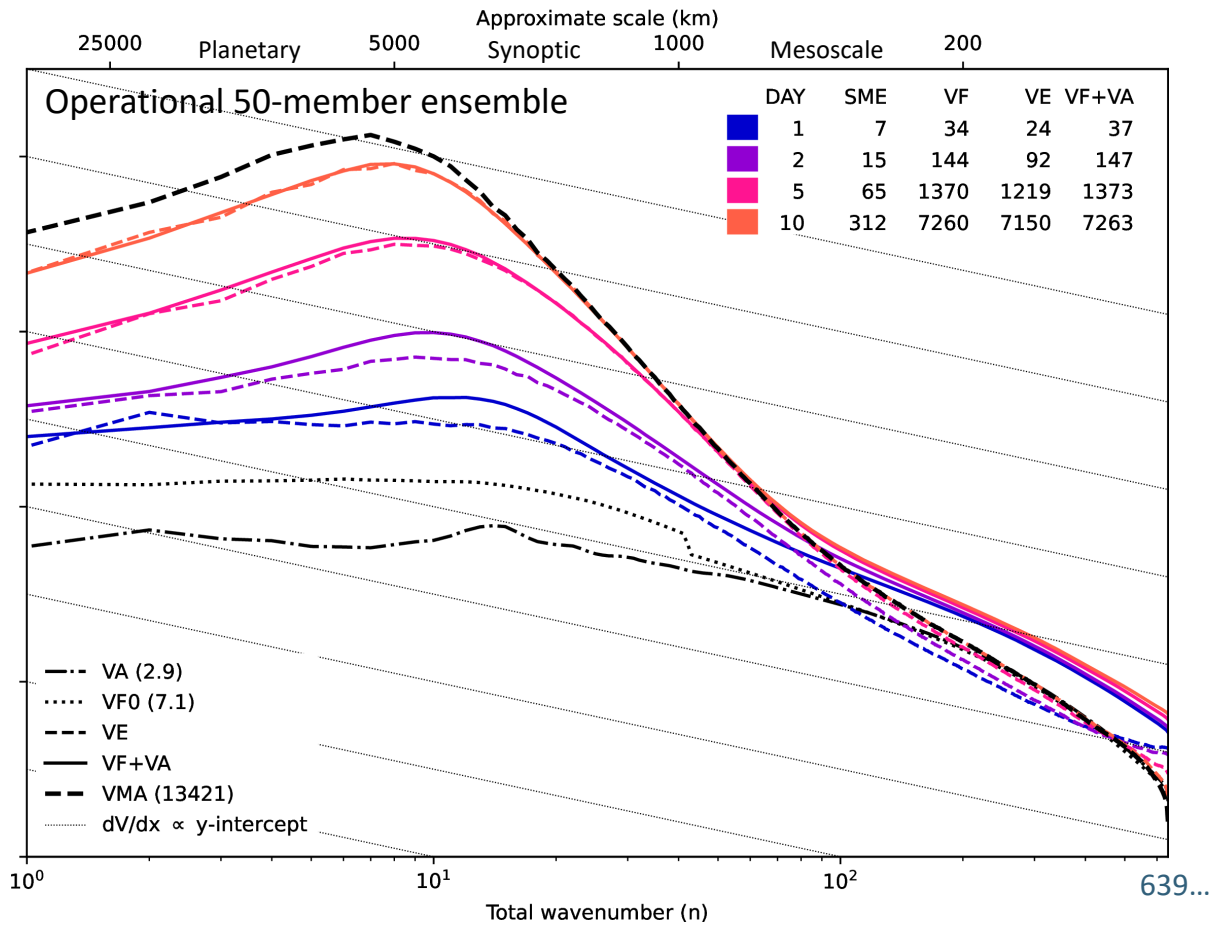
- Cycle 48r1
- Spectral resolution n = 399 (≈100km), 10 members
- No initialisation step
- Sept 2022

IROWG=International Radio Occultation Working Group
ROMEX=Radio Occultation Modeling Experiment

Spatiotemporal impact of observations seems informative – How would this look like for, *e.g.*, dense but geographically limited observations?

Predictability study: power spectra of idealised ensemble with tiny initial condition uncertainty





- ENS Cntl & "HRES" have identical resolution with tiny initial condition differences
- Opportunity for a 2-member predictability experiment
- Idealised day 10 VF < Operational day 5 VF. 48r1 model ⇒ +5 days possible
- Initial conditions are very unreliable, so VE ≫ VF+VA
- Validity of potential gain? Model is approx^{n} (& lacks stochastic physics)

All curves and table values are scaled here to give unbiased estimators for VF

DJF24 0/12Z

Suggestion for a further 5 days predictability – How model-dependent is this?

Summary on the uses of power spectra in ensemble forecasting

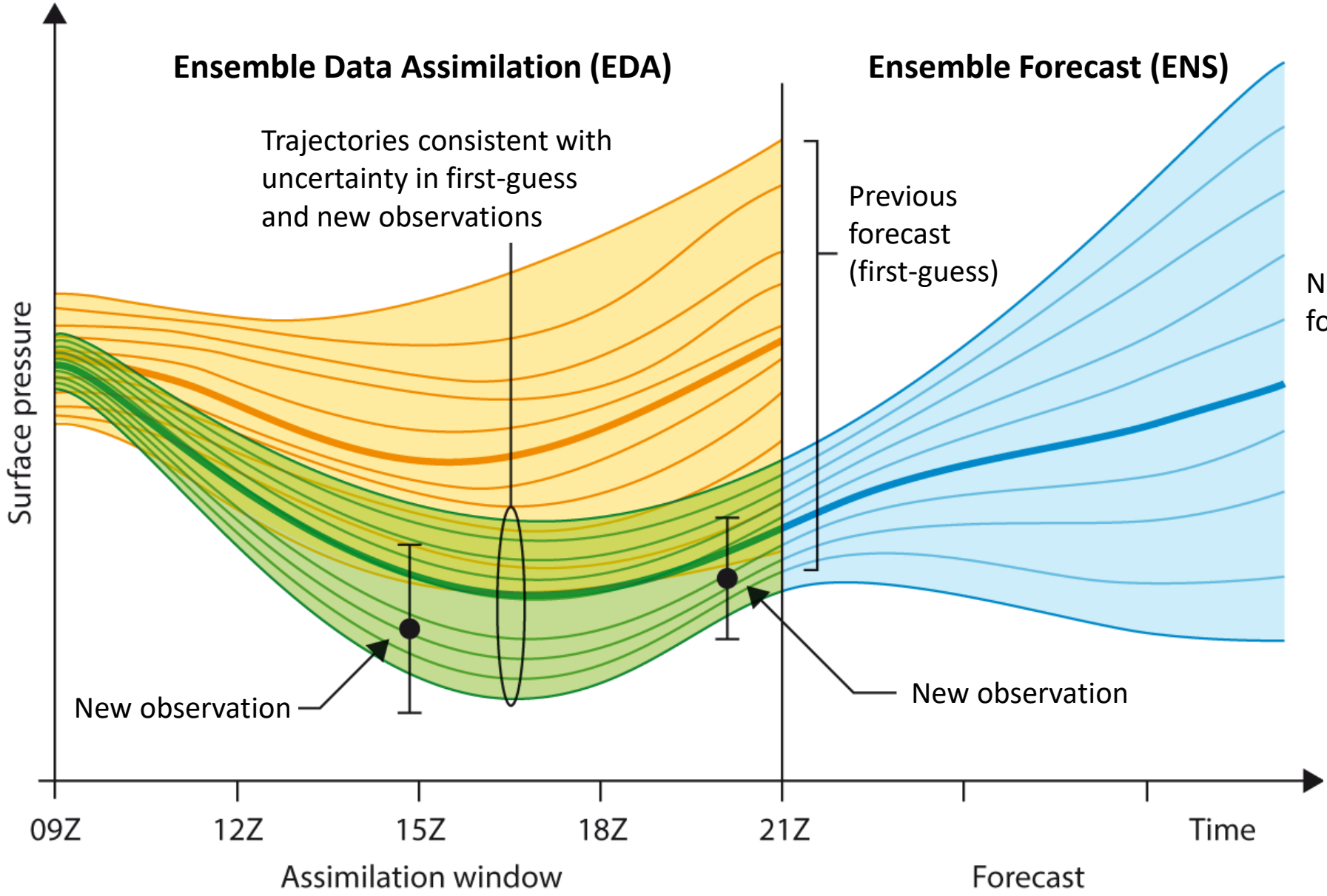
Can:

- Examine spatiotemporal evolution of the ensemble forecast
- Quantify predictive skill as a function of scale
- Chart impact of a new model cycle
- Chart impact of new observations
- Compare ensembles run at different resolutions (inc. AI)

5 more days …?

… We just need to make it reliable!

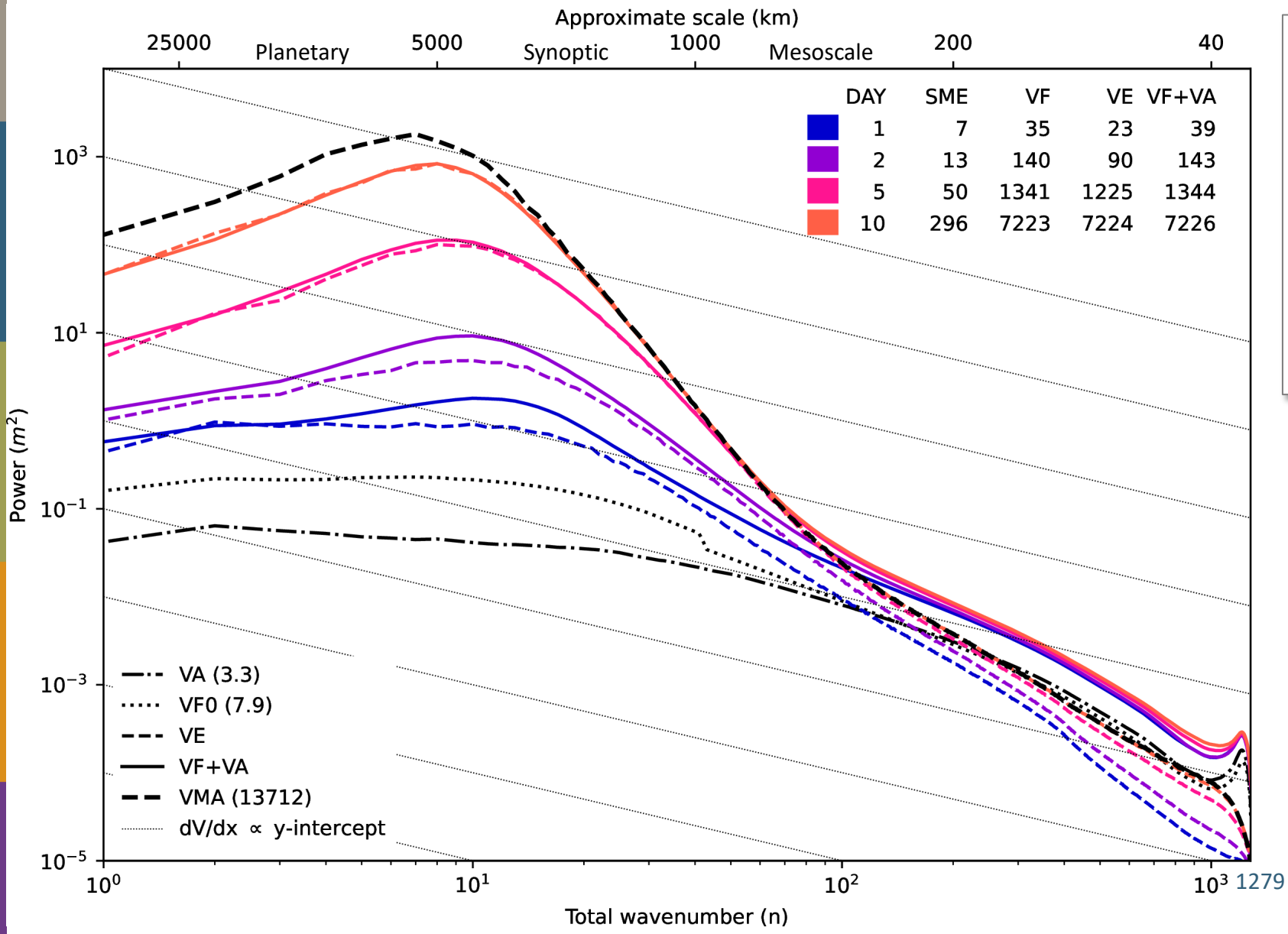
Ensemble Weather Prediction (ECMWF)



- ECMWF's Integrated Forecast System (IFS) includes a global ensemble, which is run with a variety of numbers of members and resolutions.
- The Ensemble of (4D Var) Data Assimilations (EDA) ingests millions of observations each assimilation cycle.

ENS spread arises from chaotic growth of EDA uncertainty, model uncertainty and singular vector perturbations

New IFS cycle: Z250 power spectra of the medium-range ensemble: DJF24 0Z



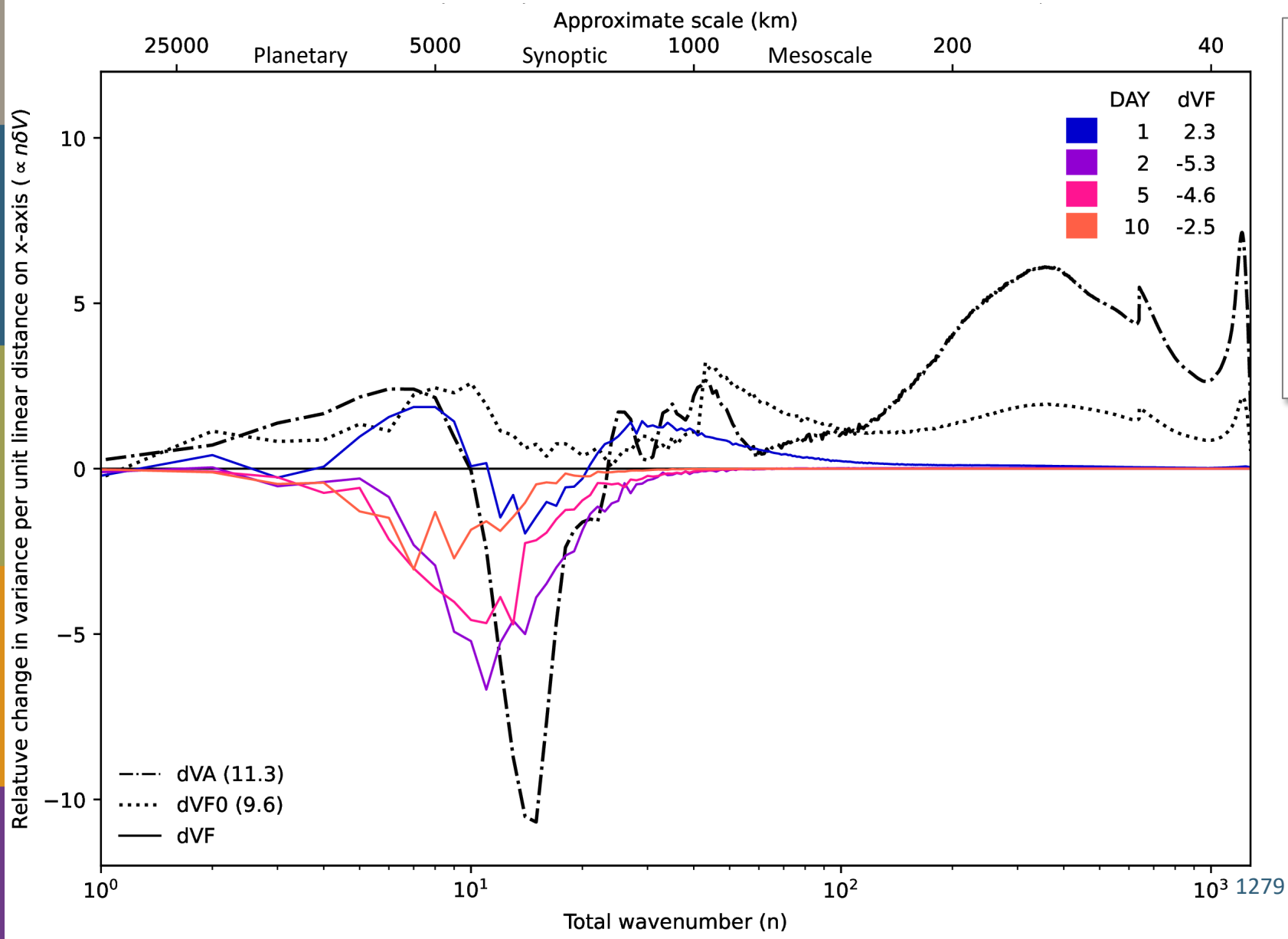
EDA
- The higher resolution EDA looks much better
- Reduced synoptic-scale wobbles
- Less drop-off in variance at scales 200 – 40 km)

ENS initial conditions
- Initialisation adds slightly more variance

ENS days 1—10
- Reduction in synoptic-scale over-dispersion (due to major change in stochastic physics: SPPT ⟶ SPP?)
- Global budgets: reduced bias, error variance and better match between VE and VF+VA

The next operational cycle sees EDA with same resolution as ENS and smoother. Change in stochastic physics. Reliability and sharpness improved

New-Current IFS cycle: Z250 power spectra of the medium-range ensemble: DJF24 0Z



EDA
- The higher resolution EDA looks much better
- Reduced synoptic-scale wobbles
- Less drop-off in variance at scales 200 – 40 km)

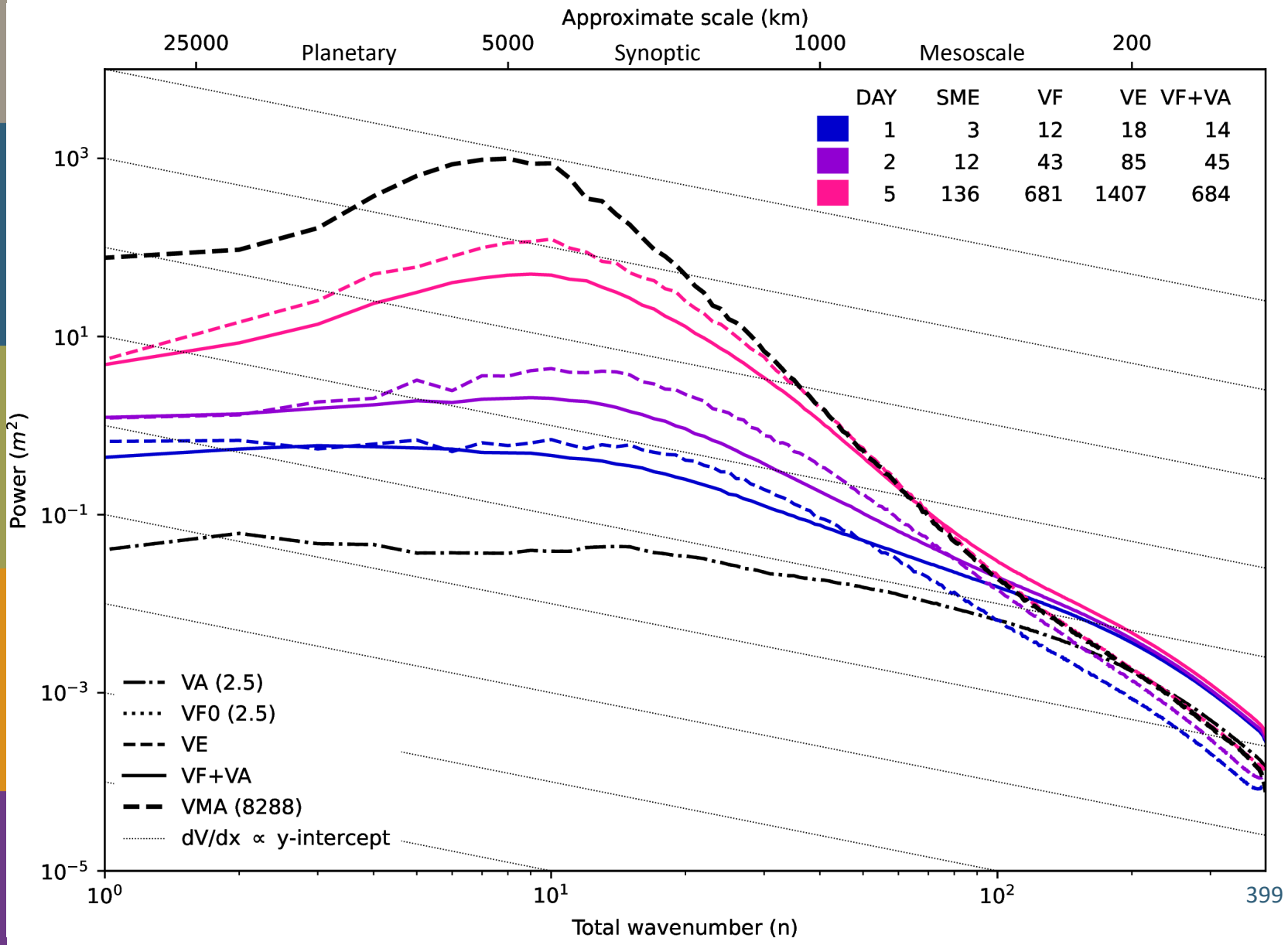
ENS initial conditions
- Initialisation adds slightly more variance

ENS days 1—10
- Reduction in synoptic-scale over-dispersion (due to major change in stochastic physics: SPPT ⟶ SPP?)
- Global budgets: reduced bias, error variance and better match between VE and VF+VA

Experimental (49r1) – Operations (48r1)

The next operational cycle sees EDA with same resolution as ENS and smoother. Change in stochastic physics. Reliability and sharpness improved

Seamless Initialisation: Z250 power spectra of control experiment with no SVs: Sept 2022



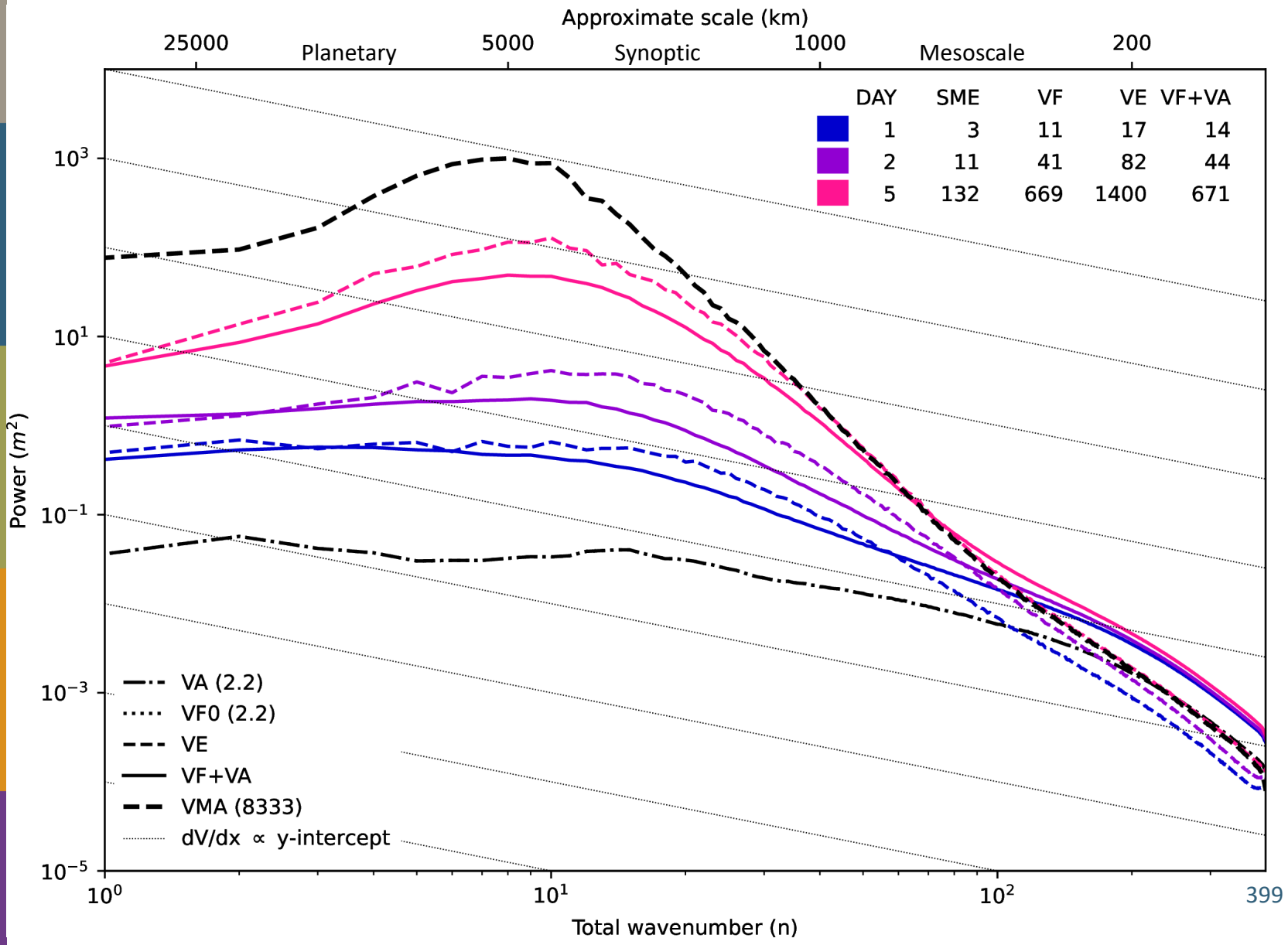
ENS days 1—5

- Cycle 48r1
- Spectral resolution n = 399 (≈100km), 10 members
- No initialisation step, so VF0=VA
- Less of a jump in scales EDA ⟶ Day 1
- ENS is now under-dispersive

Results based on 0/12 UTC forecasts verifying 10–30 September 2022

One might expect fast growing modes to develop naturally within the EDA background, but we see that SVs still serve a useful purpose

Additional data: Z250 power spectra of experiment with 30,000 new occultation observations



ENS days 1—5

- Introduction of 30,000 new Radio Occultation observations each analysis cycle (made available for evaluation as part of the IROWG ROMEX initiative)
- Extra observations reduce error and spread

IROWG=International Radio Occultation Working Group
ROMEX=Radio Occultation Modeling Experiment

The extra observations help constrain the EDA and subsequent ENS