Warm-Start Randomized SVD for Streaming Data
Compression Fidelity on Hurricane Isabel

Ahmer Nadeem Khan

HDPS Lab

8 May 2026

Outline

Outline

The Streaming Low-Rank Approximation Problem

- Many applications produce a **sequence of large matrices** $A_1, A_2, \ldots, A_T$ that evolve slowly over time.
- At each timestep t, we want a rank-k approximation

$$A_t \approx U_t \,\mathrm{diag}(\mathbf{s}_t) V_t^T, \qquad U_t \in \mathbb{R}^{m \times k}$$

- **Cold-start**: compute from scratch every step — ignores prior structure.
- **Warm-start**: reuse U_{t-1} to guide the sketch — exploits assumed temporal structure.

Central question: Does reusing U_{t-1} improve **compression fidelity**?

Cold-Start rSVD

Algorithm Cold-Start Randomized SVD

Require: $A \in \mathbb{R}^{m \times n}$, rank k, oversampling p_c
Ensure: $U \in \mathbb{R}^{m \times k}$, $\mathbf{s} \in \mathbb{R}^k$, $V^T \in \mathbb{R}^{k \times n}$
1: Draw $\Omega \in \mathbb{R}^{n \times (k+p_c)} \sim \mathcal{N}(0,1)$ [sketch width $= k + p_c$]
2: $Y \leftarrow A\Omega$ [range approximation]
3: $Q, \_ \leftarrow \mathrm{QR}(Y)$
4: $B \leftarrow Q^T A$
5: $\hat{U}, \mathbf{s}, V^T \leftarrow \mathrm{SVD}(B)$
6: $U \leftarrow Q\hat{U}$; truncate to first k columns

Warm-Start rSVD

Algorithm Warm-Start Randomized SVD

Require: $A \in \mathbb{R}^{m\times n}$, **prior basis** $U_{\text{prev}} \in \mathbb{R}^{m\times k}$, rank k, p_w
Ensure: $U \in \mathbb{R}^{m\times k}$, $\mathbf{s} \in \mathbb{R}^{k}$, $V^T \in \mathbb{R}^{k\times n}$

1: $G \leftarrow A^T U_{\text{prev}}$ [warm projection]
2: $Y_1 \leftarrow AG$ [exploit prior subspace, width k]
3: Draw $\Omega \in \mathbb{R}^{n\times p_w} \sim \mathcal{N}(0,1)$
4: $Y_2 \leftarrow A\Omega$ [fresh exploration, width p_w]
5: $Y \leftarrow [Y_1 \mid Y_2]$ [sketch width $= k + p_w < k + p_c$]
6: $Q, \_ \leftarrow \mathrm{QR}(Y)$; $B \leftarrow Q^T A$; SVD; lift; truncate

Approximation Quality Metrics

Relative Frobenius Error

$$\varepsilon_F = \frac{\|A - U\,\mathrm{diag}(\mathbf{s})V^T\|_F}{\|A\|_F} = \frac{\sqrt{\|A\|_F^2 - \|\mathbf{s}\|^2}}{\|A\|_F}$$

Fraction of energy *not* captured by the rank-k approximation. Measures average reconstruction quality across all grid points.

Max Element-wise Error

$$\varepsilon_{\max} = \max_{i,j} \left|A_{ij} - (U\,\mathrm{diag}(\mathbf{s})V^T)_{ij}\right|$$

Worst-case reconstruction error at any single grid point. Critical for compression: a small Frobenius error can hide large localised spikes.

Compression Quality Metrics

PSNR (Peak Signal-to-Noise Ratio)

$$\mathrm{PSNR} = 10\,\log_{10}\left(\frac{\mathrm{peak}^2}{\mathrm{MSE}}\right)\ \mathrm{dB}, \quad \mathrm{peak} = \max|A|, \quad \mathrm{MSE} = \frac{1}{mn}\sum_{i,j}(A_{ij} - \hat{A}_{ij})^2$$

Standard lossy compression metric. Higher = better. Normalises by the variable's dynamic range ⇒ comparable across variables with different physical units and magnitudes.

Percentile Errors (99th, 99.9th)

$$P_{99} = \text{99th percentile of } \{|A_{ij} - \hat{A}_{ij}|\}$$

Robust tail measures — less sensitive to isolated outliers than raw max, but still capture whether large errors are widespread or confined to a few grid points.

Hurricane Isabel: Streaming Setup

- We focus on **5 structured variables** with temporally stable subspaces:
  - Wind: U_f (east–west), V_f (north–south)
  - Temperature: TC_f
  - Moisture: $QVAPOR_f$, $QSNOW_f$
- Each variable is processed **independently** as a stream of 48 matrices.
- $t = 1$: **cold start only** — no prior basis available.
- $t \geq 2$: both cold and warm rSVD run on the **same** A_t; all metrics are averaged over the **47 warm-active** timesteps.

Experiment Design: C++ Benchmark

Parameter	Value
Target rank k	20
Cold oversampling p_c	10
Warm oversampling p_w	5
Sketch width (cold)	30
Sketch width (warm)	25
Power iterations q	0
Timesteps per variable	48
Data type	float32

Compression Ratio: SVD as Lossy Compressor

Storage: original vs. rank-k SVD

Original: $m \times n \times 4$ bytes. SVD stores $U \in \mathbb{R}^{m \times k}$, $\boldsymbol{\sigma} \in \mathbb{R}^{k}$, $V \in \mathbb{R}^{n \times k}$.

$$\text{ratio} = \frac{m \cdot n}{k\,(m+n+1)} \approx \frac{n}{k} \quad (\text{since } m \gg n,\ U \text{ dominates storage})$$

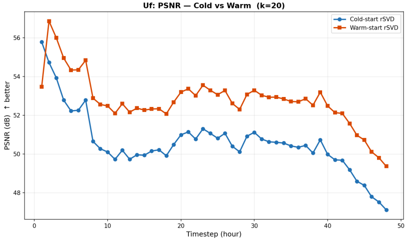
k	Compressed	Ratio	bpe
5	5.0 MB	20×	1.6
10	10.0 MB	10×	3.2
20	**20.0 MB**	**5**×	**6.4**
50	50.0 MB	2×	16

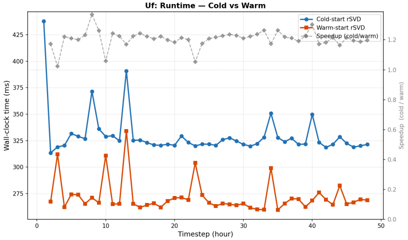
bpe = bits per element (original: 32 bpe)

At $k = 20$: **5**× compression at 6.4 bpe — comparable to scientific lossy compressors (SZ, ZFP target 1–16 bpe), but without pointwise error guarantees.

Outline

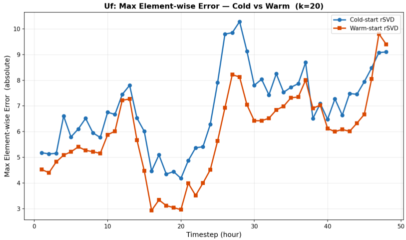
U_f (East–West Wind): PSNR & Runtime

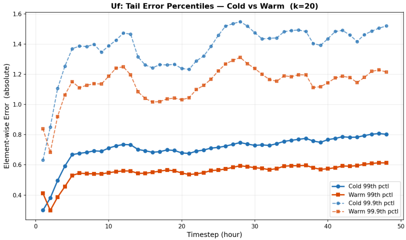




Warm PSNR consistently 2–3 dB above cold (mean +2.3 dB). Speedup: 1.21×.

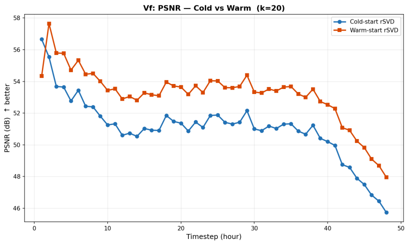
U_f: Max Element-wise Error & Tail Percentiles

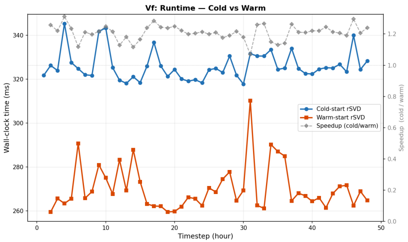




Left: Max error reduced ∼15%; warm below cold at 94% of steps.
Right: 99th pctl ∼0.56 vs 0.72 m/s; 99.9th pctl confirms errors are spatially concentrated (∼2× the 99th).

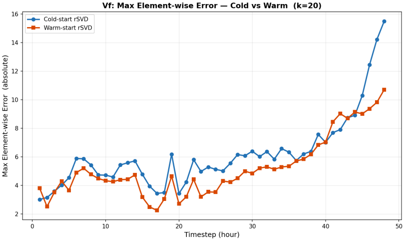
V_f (North–South Wind): PSNR & Runtime

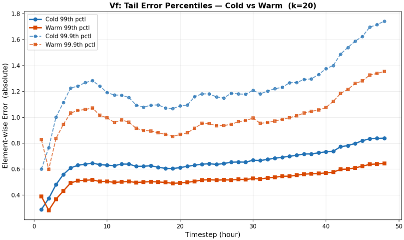




Near-identical profile to U_f; wind components share the same subspace structure. PSNR gain: $+2.2$ dB. Speedup: $1.21\times$.

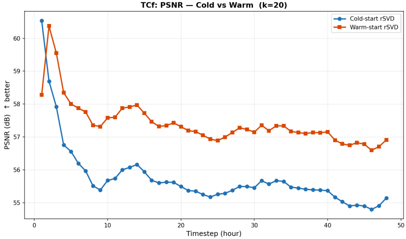
V_f: Max Element-wise Error & Tail Percentiles

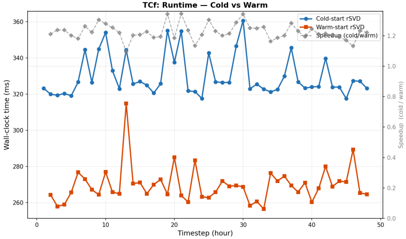




Max error reduced ∼16%. 99th pctl: 0.53 vs 0.67 m/s.

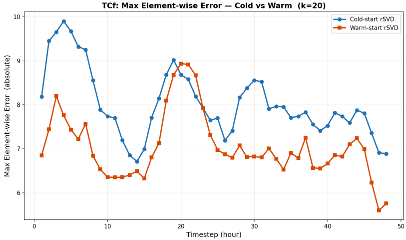
TC_f (Temperature): PSNR & Runtime

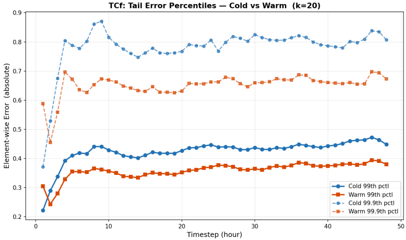




Most stable subspace of the five variables. PSNR gain: +1.8 dB. Highest speedup: 1.22×.

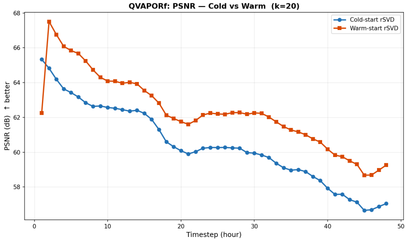
TC_f: Max Element-wise Error & Tail Percentiles

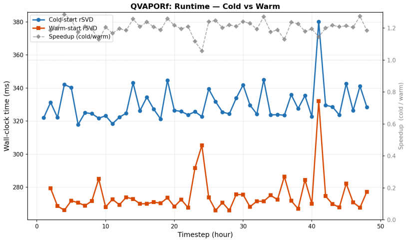




Max error ∼8°C (localised near eyewall); reduced 12% by warm-start.
99th pctl: 0.36 vs 0.43°C.

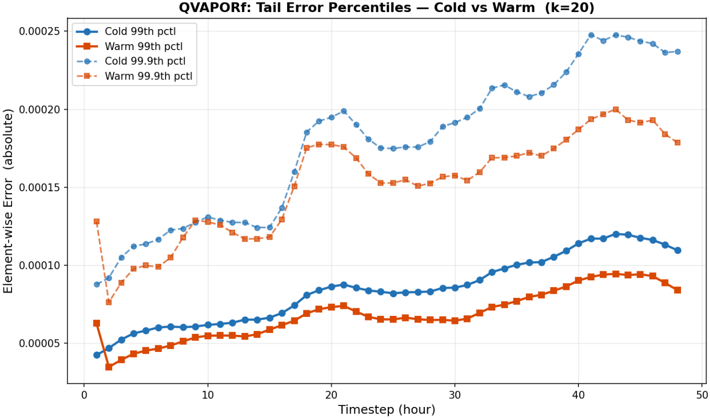
$QVAPOR_f$ (Water Vapour): PSNR & Runtime





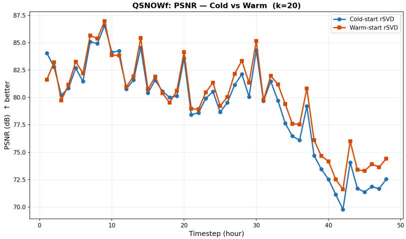
PSNR gain: +2.0 dB. Speedup: 1.20×. Largest max-error reduction of all five variables: −25.5%.

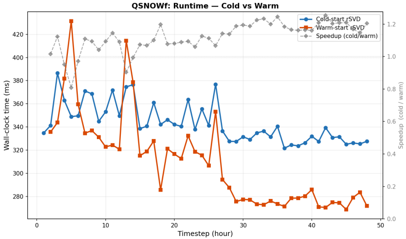
$QVAPOR_f$: Tail Percentiles



99th pctl reduced 20%; 99.9th reduced 15%. Clear and consistent warm advantage across all 48 hours.

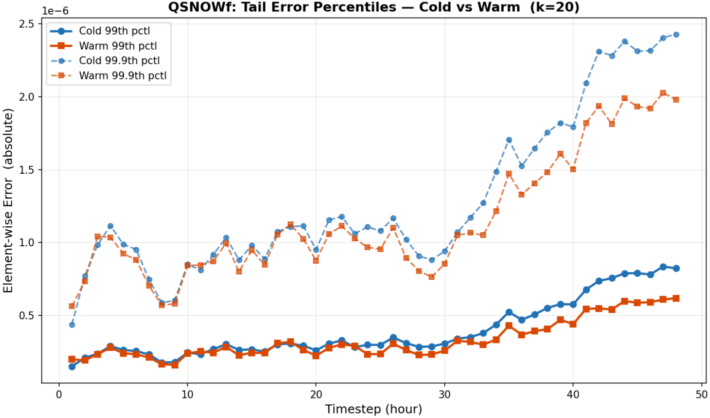
$QSNOW_f$ (Snow): PSNR & Runtime





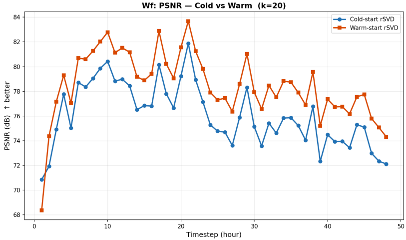
PSNR gain: +0.8 dB (∼79–80 dB, near-perfect reconstruction).
Moderate speedup: 1.12×.

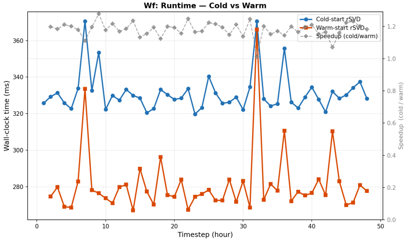
$QSNOW_f$: Tail Percentiles



Despite the noisy max, the 99th percentile clearly favours warm (-17%). Absolute errors are tiny ($\sim 10^{-7}$) because $\max|A| \sim 0.001$.

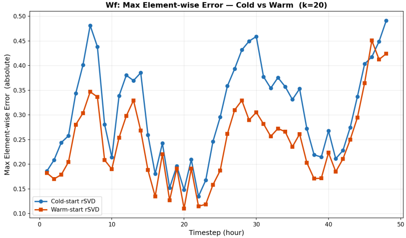
W_f (Vertical Wind): PSNR & Runtime

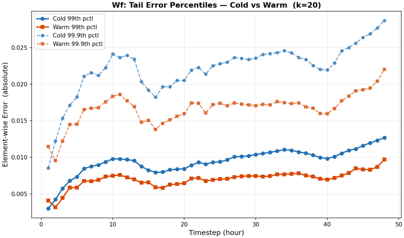




Largest PSNR gain of all variables: $+2.5$ dB (76.2 vs 78.7 dB). Speedup: $1.18\times$.

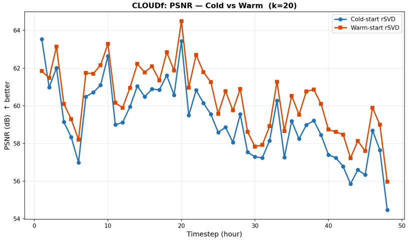
W_f: Max Element-wise Error & Tail Percentiles

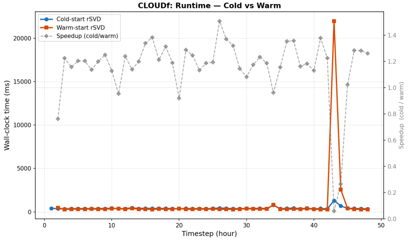




Max error reduced 21%; warm wins 98% of timesteps. 99th pctl: 0.0071 vs 0.0096 m/s (−26%).

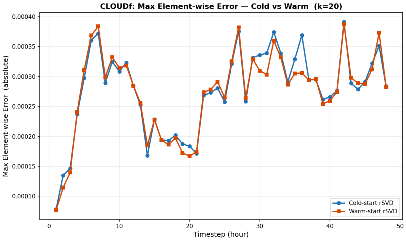
$CLOUD_f$ (Cloud Fraction): PSNR & Runtime

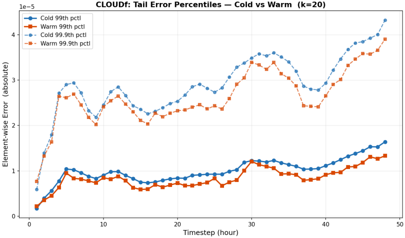




PSNR gain: +1.2 dB (59.1 vs 60.3 dB); warm wins 100% of steps.
Speedup: 1.15×.

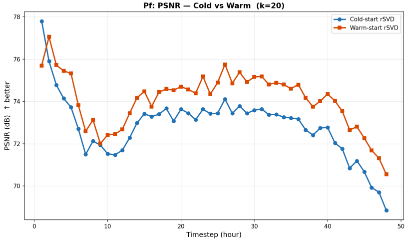
$CLOUD_f$: Max Element-wise Error & Tail Percentiles

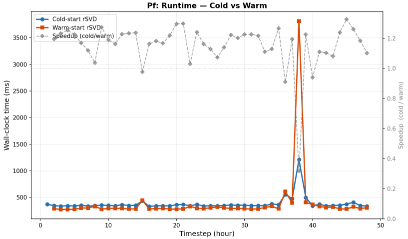




Max error nearly identical (warm wins only 53%); but 99th pctl reduced 17%.

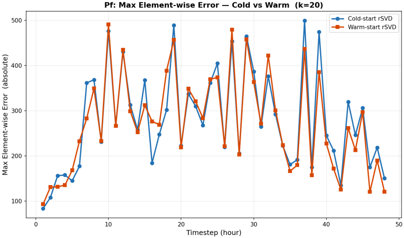
P_f (Pressure): PSNR & Runtime

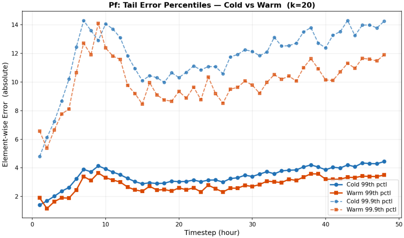




PSNR gain: +1.3 dB (72.7 vs 74.1 dB); warm wins 100% of steps.
Speedup: 1.15×.

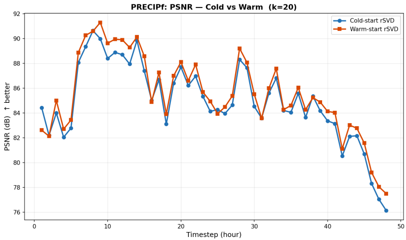
P_f: Max Element-wise Error & Tail Percentiles

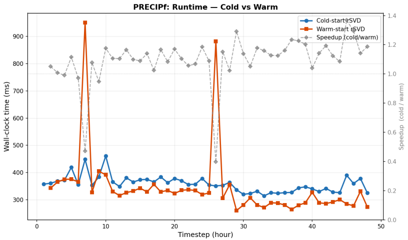




Max error ∼280 Pa (warm wins 60%); 99th pctl reduced 18% (∼3.5 vs 2.8 Pa).

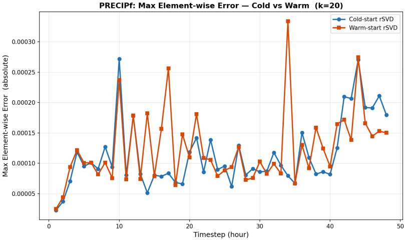
$PRECIP_f$ (Precipitation): PSNR & Runtime

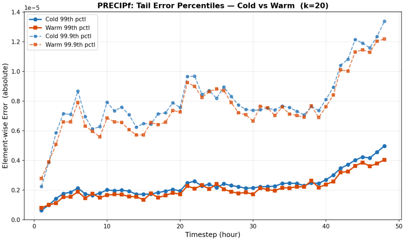




Modest PSNR gain: +0.7 dB (84.9 vs 85.6 dB); warm wins 87% of steps.
Speedup: 1.10×. Subspace less stable than wind/temperature fields.

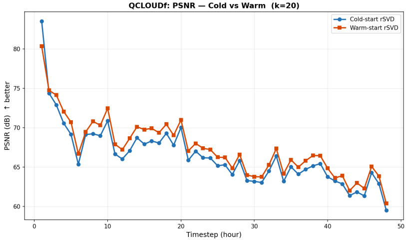
$PRECIP_f$: Max Element-wise Error & Tail Percentiles

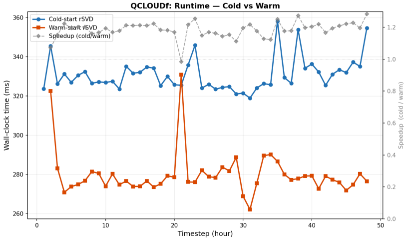




Max error slightly worse for warm (−9%); but 99th pctl still reduced 12%.

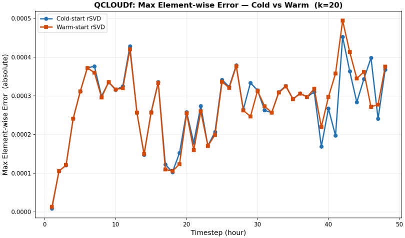
$QCLOUD_f$ (Cloud Water): PSNR & Runtime

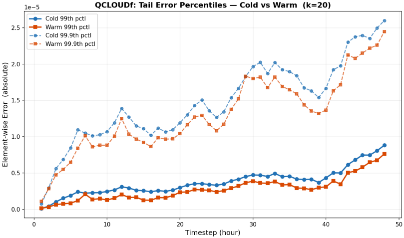




PSNR gain: +1.1 dB (66.1 vs 67.1 dB); warm wins 100% of steps.
Speedup: 1.18×.

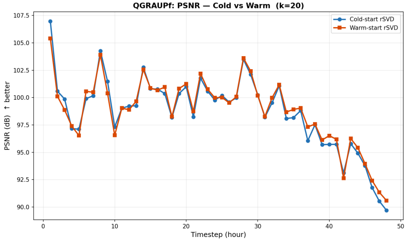
$QCLOUD_f$: Max Element-wise Error & Tail Percentiles

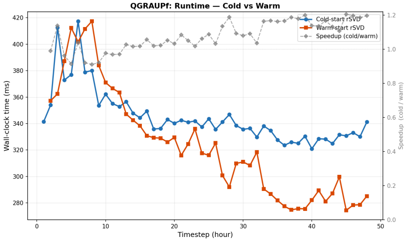




Max error comparable (warm wins 62%); 99th pctl strongly reduced (−26%).

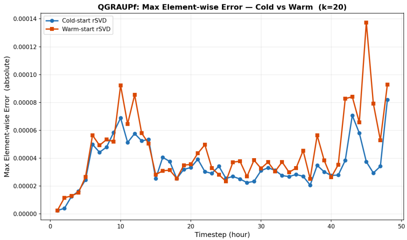
$QGRAUP_f$ (Graupel): PSNR & Runtime

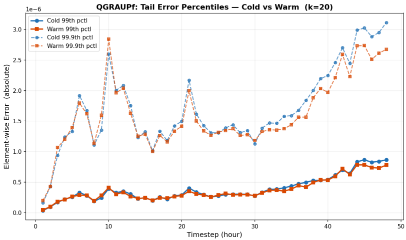




Near-perfect reconstruction: ∼98.5 dB; marginal warm gain (+0.2 dB, wins 72%). Speedup: 1.08×. Very low effective rank — little room to improve.

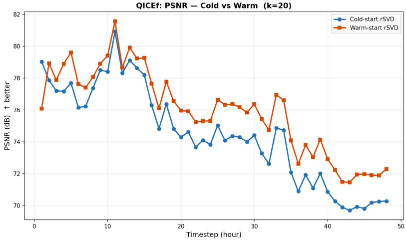
$QGRAUP_f$: Max Element-wise Error & Tail Percentiles

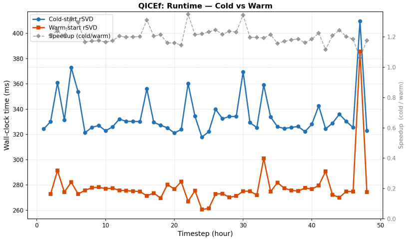




Max error 28% *worse* for warm (wins only 19%); 99th pctl nearly tied (-5%). At $\sim$99 dB, both methods are near machine-precision limits.

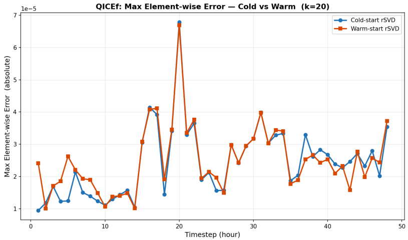
$QICE_f$ (Ice): PSNR & Runtime

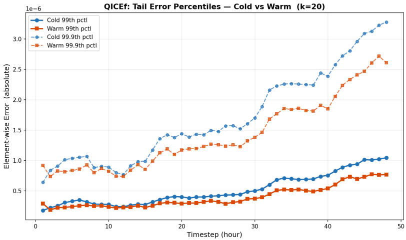




PSNR gain: +1.6 dB (74.4 vs 75.9 dB); warm wins 100% of steps.
Speedup: 1.21×.

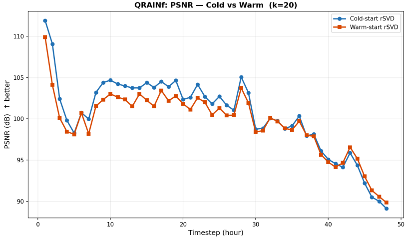
$QICE_f$: Max Element-wise Error & Tail Percentiles

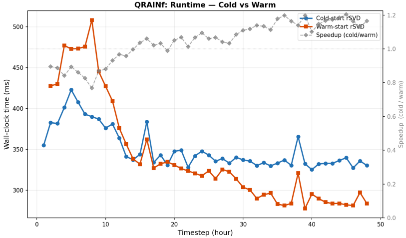




Max error comparable (warm wins 45%); 99th pctl reduced 24%.

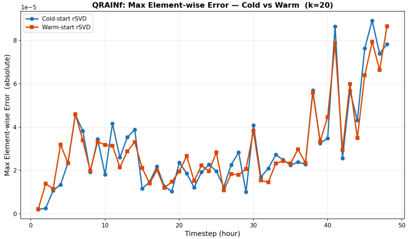
$QRAIN_f$ (Rain Water): PSNR & Runtime

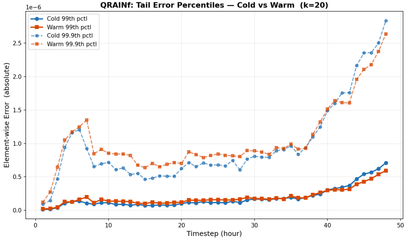




Only variable where warm is worse: PSNR -0.9 dB (warm wins only 21%). Speedup: $1.05\times$. Very low effective rank ($\sim$100 dB); subspace too unstable for warm-start.

$QRAIN_f$: Max Element-wise Error & Tail Percentiles





All metrics nearly tied or slightly worse for warm. Confirms $QRAIN_f$ has an unstable subspace that warm-start cannot exploit.

Outline

PSNR: What the Numbers Mean

PSNR Range	Quality
< 30 dB	Poor — visible artefacts
30–40 dB	Acceptable for many applications
40–50 dB	Good — errors barely perceptible
> 50 dB	Excellent — near-lossless

Our results (rank $k = 20$, $5\times$ compression):

- Wind (U_f, V_f, W_f): 51–79 dB. Horizontal winds hardest; vertical easiest.
- Cloud/moisture ($CLOUD_f$, $QCLOUD_f$, $QICE_f$, $QVAPOR_f$): 59–75 dB.
- Low-rank fields ($QSNOW_f$, $PRECIP_f$, $QGRAUP_f$, $QRAIN_f$): 79–100 dB.
- Pressure (P_f): $\sim$73 dB. Temperature (TC_f): $\sim$57 dB.

All 13 variables exceed 50 dB — near-lossless regime — with
warm start +0.2 to +2.5 dB better for 12/13 variables.

Compression Summary: All 13 Variables

	PSNR (dB)		Max Elem. Err		99th Pctl		
Variable	**Cold**	**Warm**	**Cold**	**Warm**	**Cold**	**Warm**	**Speedup**
U_f	50.6	**52.8**	6.95	**5.93**	0.716	**0.558**	1.21×
V_f	51.0	**53.2**	6.22	**5.24**	0.666	**0.526**	1.21×
TC_f	55.7	**57.4**	8.02	**7.05**	0.429	**0.360**	1.22×
$CLOUD_f$	59.1	**60.3**	2.8e-4	2.8e-4	1.0e-5	**8.5e-6**	1.15×
$QVAPOR_f$	60.4	**62.3**	3.0e-3	**2.2e-3**	8.5e-5	**6.8e-5**	1.20×
$QCLOUD_f$	66.1	**67.1**	2.8e-4	2.8e-4	3.8e-6	**2.8e-6**	1.18×
P_f	72.7	**74.1**	284	**277**	3.47	**2.83**	1.15×
$QICE_f$	74.4	**75.9**	2.5e-5	2.5e-5	5.2e-7	**4.0e-7**	1.21×
W_f	76.2	**78.7**	0.311	**0.246**	9.6e-3	**7.1e-3**	1.18×
$QSNOW_f$	79.1	**79.9**	4.2e-5	4.8e-5	4.0e-7	**3.3e-7**	1.12×
$PRECIP_f$	84.9	**85.6**	1.2e-4	1.3e-4	2.4e-6	**2.1e-6**	1.10×
$QGRAUP_f$	98.5	**98.7**	3.6e-5	4.6e-5	4.0e-7	3.8e-7	1.08×
$QRAIN_f$	**100.2**	99.4	3.1e-5	3.1e-5	1.9e-7	2.0e-7	1.05×

Warm-start improves PSNR for **12/13** variables ($+0.2$ to $+2.5$ dB). Only $QRAIN_f$ is worse (-0.9 dB) — unstable subspace at $\sim$100 dB. 99th pctl reduced 5–26% for 11/13 variables.

Tail Error Distribution: 99th vs 99.9th Percentile

	99th Percentile		99.9th Percentile	
Variable	**Cold**	**Warm**	**Cold**	**Warm**
U_f	0.716	**0.558**	1.389	**1.143**
V_f	0.666	**0.526**	1.246	**1.007**
TC_f	0.429	**0.360**	0.790	**0.653**
$CLOUD_f$	1.0e-5	**8.5e-6**	3.0e-5	**2.7e-5**
$QVAPOR_f$	8.5e-5	**6.8e-5**	1.8e-4	**1.5e-4**
$QCLOUD_f$	3.8e-6	**2.8e-6**	1.5e-5	**1.3e-5**
P_f	3.47	**2.83**	11.96	**10.1**
$QICE_f$	5.2e-7	**4.0e-7**	1.7e-6	**1.4e-6**
W_f	9.6e-3	**7.1e-3**	2.2e-2	**1.7e-2**
$QSNOW_f$	4.0e-7	**3.3e-7**	1.3e-6	**1.2e-6**
$PRECIP_f$	2.4e-6	**2.1e-6**	8.1e-6	**7.7e-6**
$QGRAUP_f$	4.0e-7	3.8e-7	1.7e-6	1.6e-6
$QRAIN_f$	1.9e-7	2.0e-7	9.9e-7	1.1e-6

- 99.9th pctl $\approx$ 2–3$\times$ the 99th everywhere — errors are spatially concentrated.
- Warm-start reduces tail errors for 11/13 variables (5–26% on 99th pctl).

Outline

Key Findings

1. **Warm-start improves compression fidelity for 12/13 variables.**
   - PSNR gain: $+0.2$ to $+2.5$ dB. Only $QRAIN_f$ is worse (-0.9 dB).
   - 99th percentile error reduced 5–26% for 11/13 variables.
   - Benefit scales with subspace stability: strongest for wind/temperature, weakest for sporadic fields ($QGRAUP_f$, $QRAIN_f$).
2. **All 13 variables achieve near-lossless quality.**
   - $>$50 dB PSNR at rank $k = 20$ (5$\times$ compression ratio).
   - 99% of grid points have error below 1 m/s even for the hardest wind fields.
3. **Warm-start is faster, not slower.**
   - 5–22% speedup across all 13 variables.
   - Better fidelity *and* lower cost — no accuracy–speed trade-off.