Evaluation Report

ns2d | 2025_09_11T23_35_37

Models: velocity and velocity

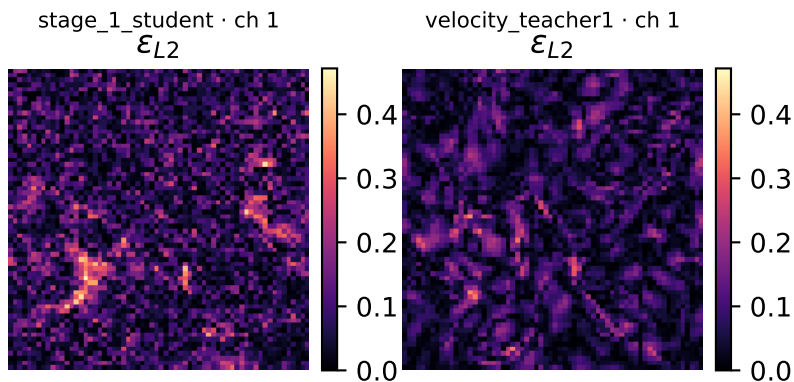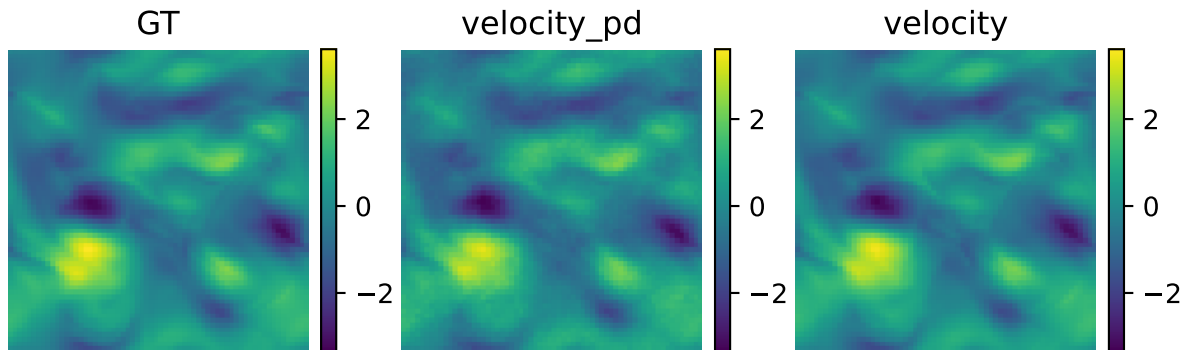| mode	model	metric	mean	median	std	count
one_step	velocity	l2	0.02776	0.02442	0.01246	32
one_step	velocity_pd	l2	0.03435	0.03208	0.01295	32
rollout	velocity	l2	1.166	1.218	0.4948	256
rollout	velocity_pd	l2	1.124	1.156	0.4718	256

Rollout error curves (l2)

One-Step Snapshot #0 — ch 0

GT
velocity_pd
velocity

stage_1_student · ch 0
ε_{L2}

velocity_teacher1 · ch 0
ε_{L2}

One-Step Snapshot #0 — ch 1

GT   velocity_pd   velocity

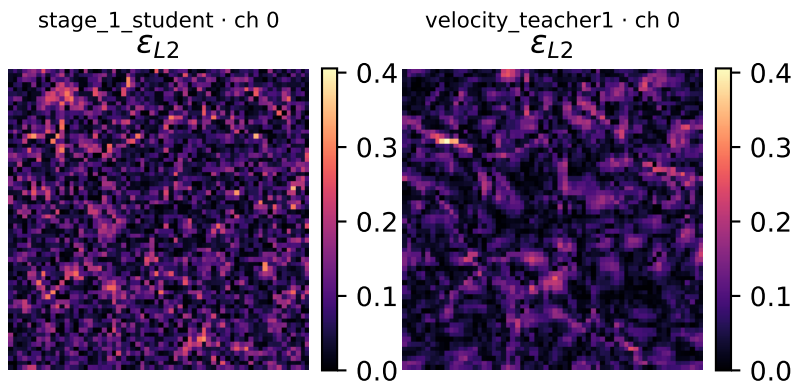stage_1_student · ch 1
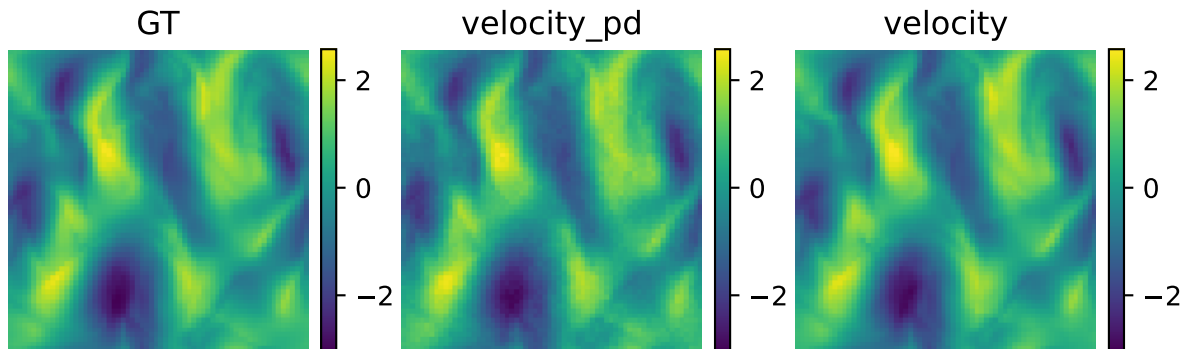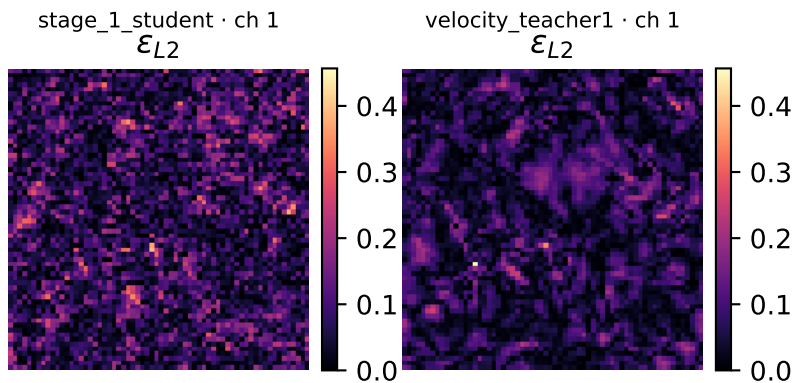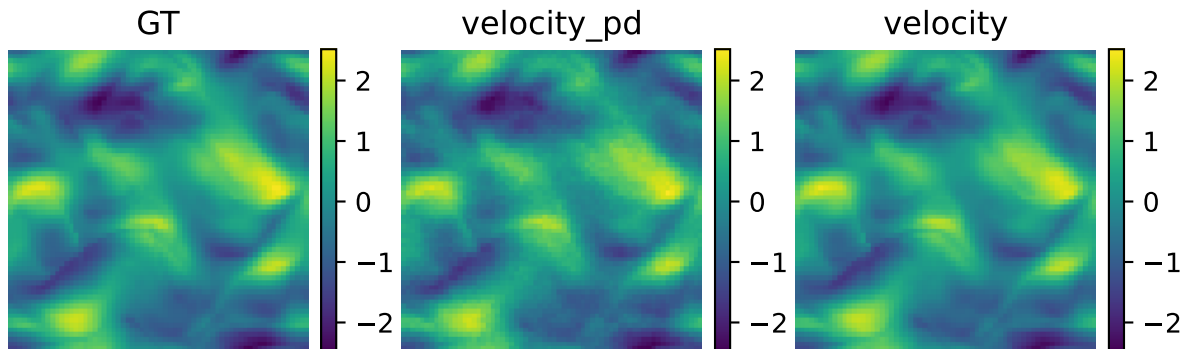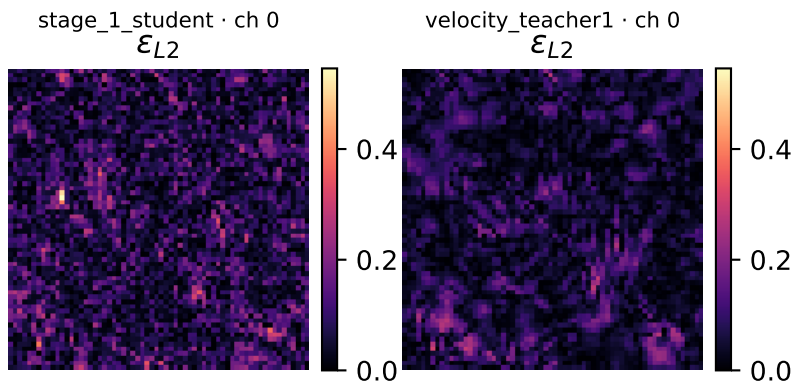ε_{L2}

velocity_teacher1 · ch 1
ε_{L2}
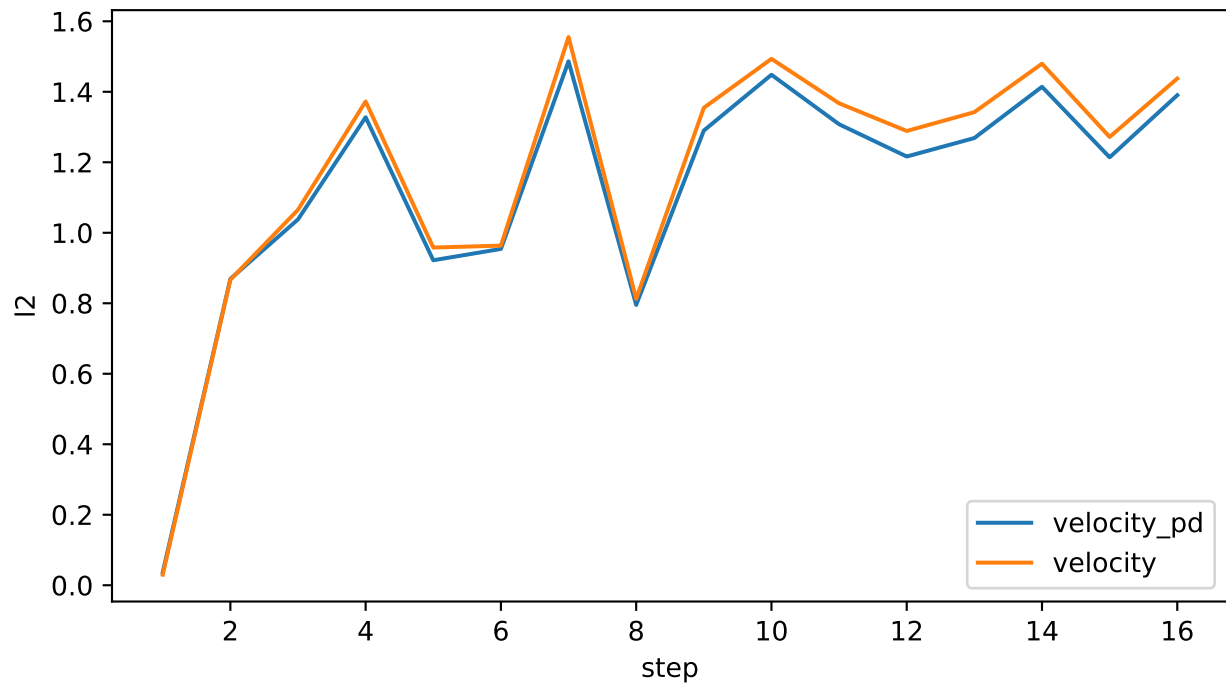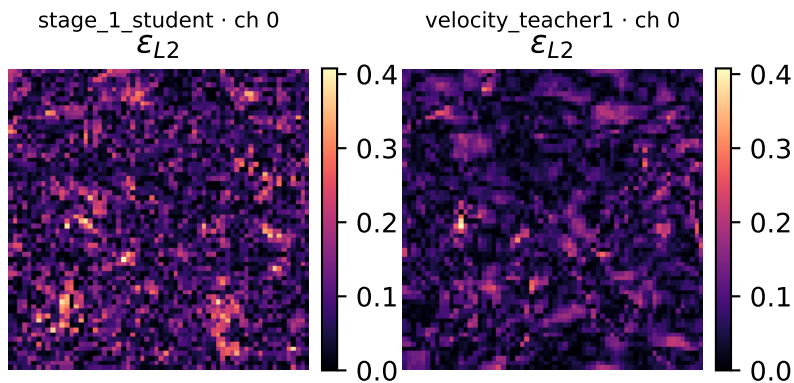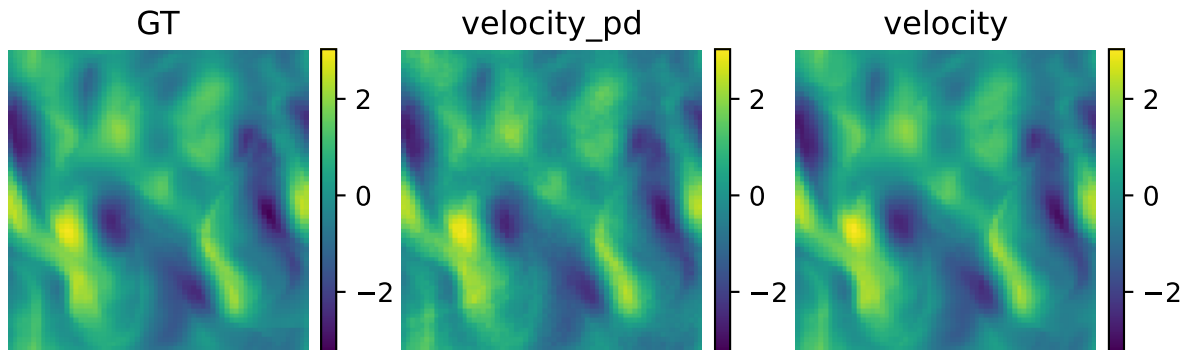
One-Step Snapshot #1 — ch 0

One-Step Snapshot #1 — ch 1

GT

velocity_pd

velocity

stage_1_student · ch 1
ε_{L2}

velocity_teacher1 · ch 1
ε_{L2}

One-Step Snapshot #2 — ch 0

GT velocity_pd velocity

stage_1_student · ch 0
ε_{L2}

velocity_teacher1 · ch 0
ε_{L2}

One-Step Snapshot #2 — ch 1

GT

velocity_pd

velocity

stage_1_student · ch 1
ε_{L2}

velocity_teacher1 · ch 1
ε_{L2}
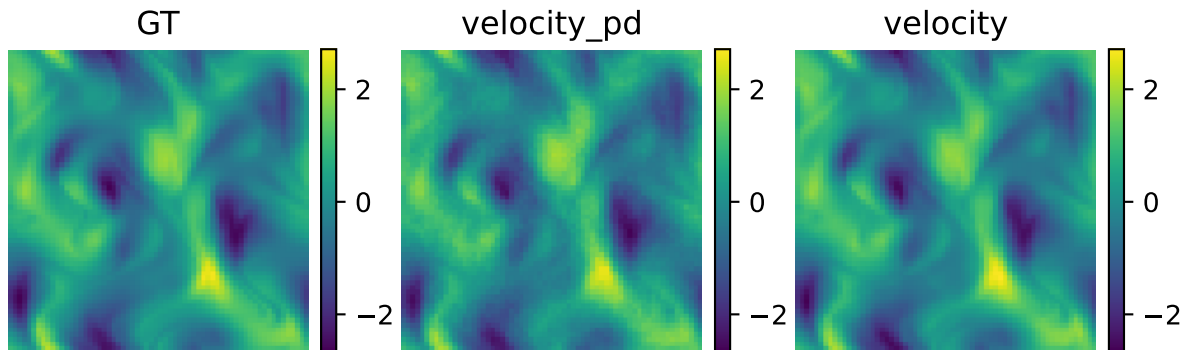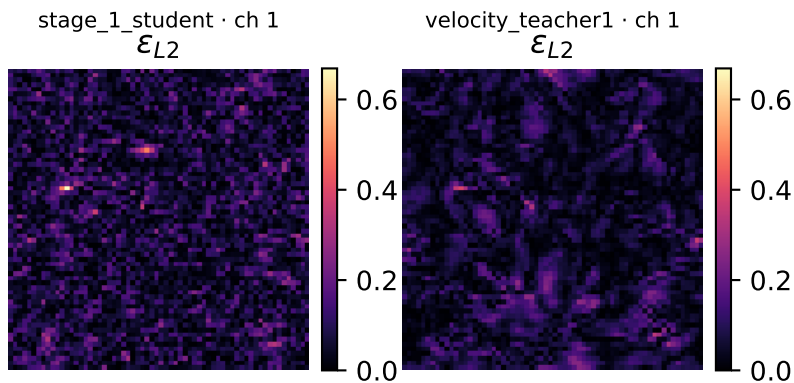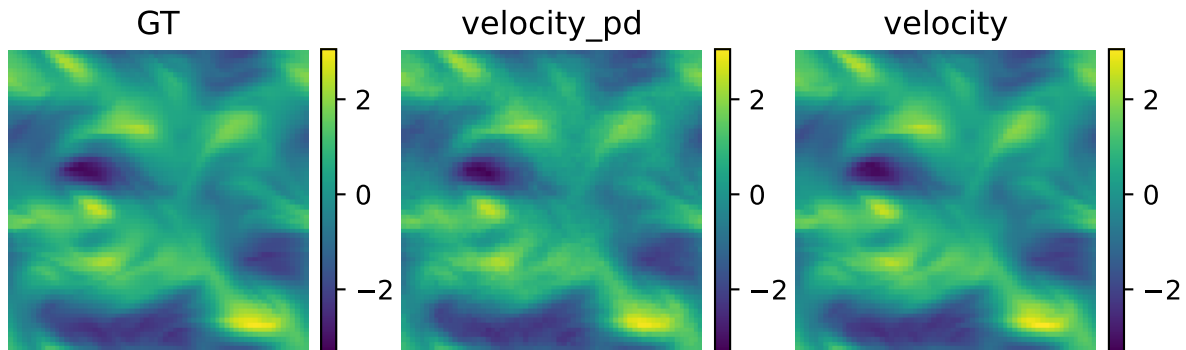
Rollout Snapshot #0 -- ch 0 -- GT (top row) vs stage_1_student (middle row)
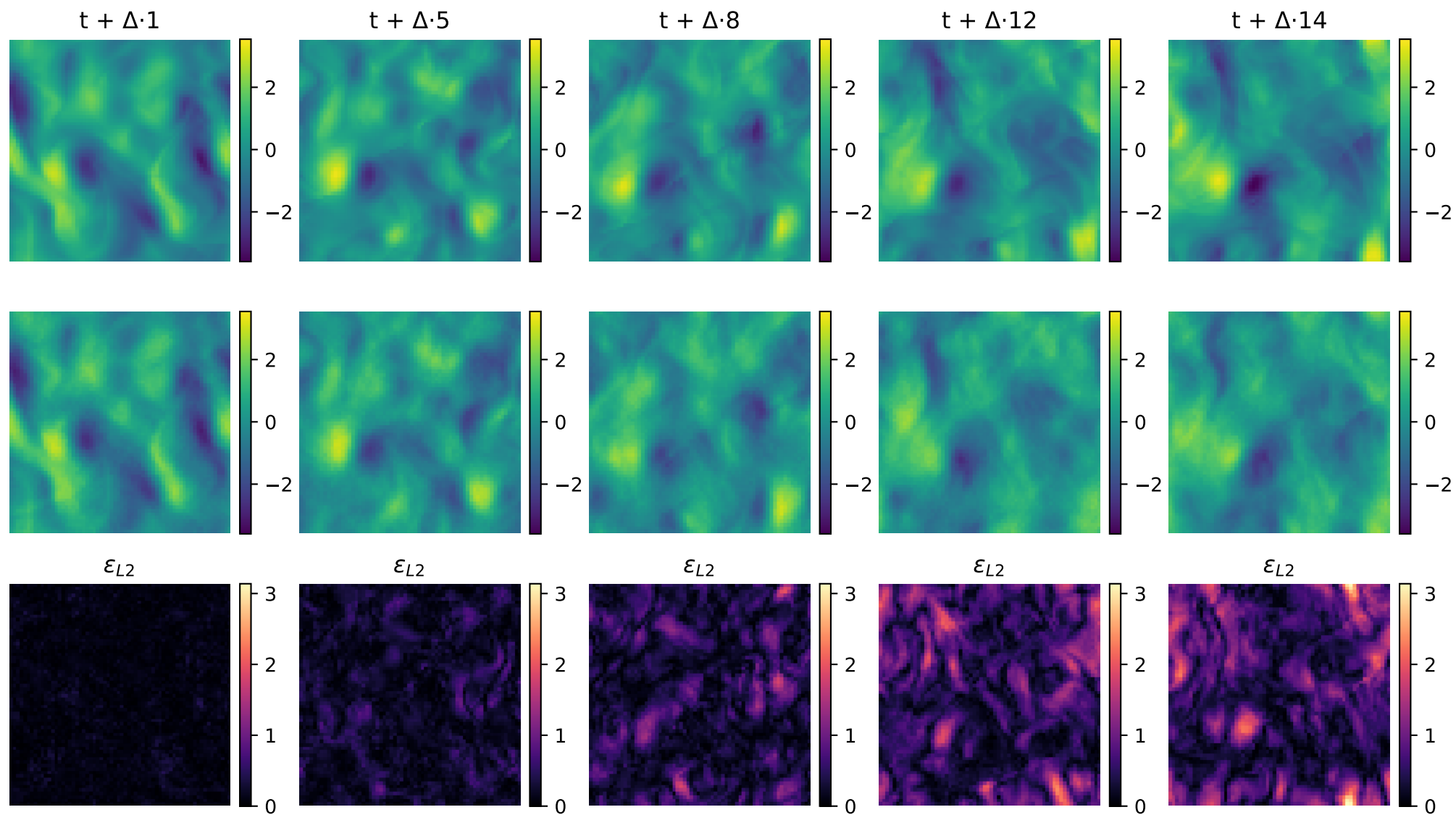
Rollout Snapshot #0 -- ch 1 -- GT (top row) vs stage_1_student (middle row)

Rollout Snapshot #1 -- ch 0 -- GT (top row) vs stage_1_student (middle row)
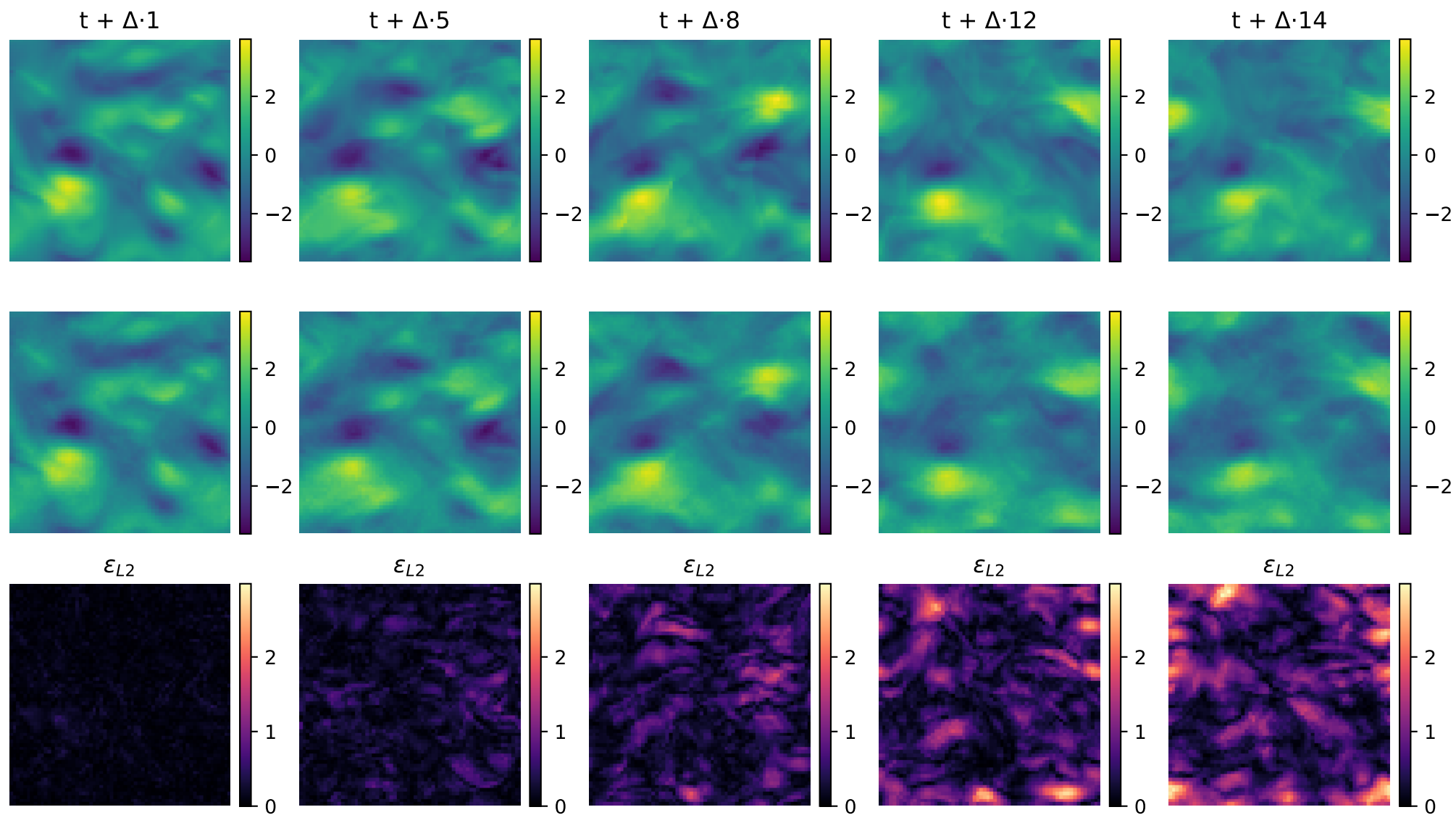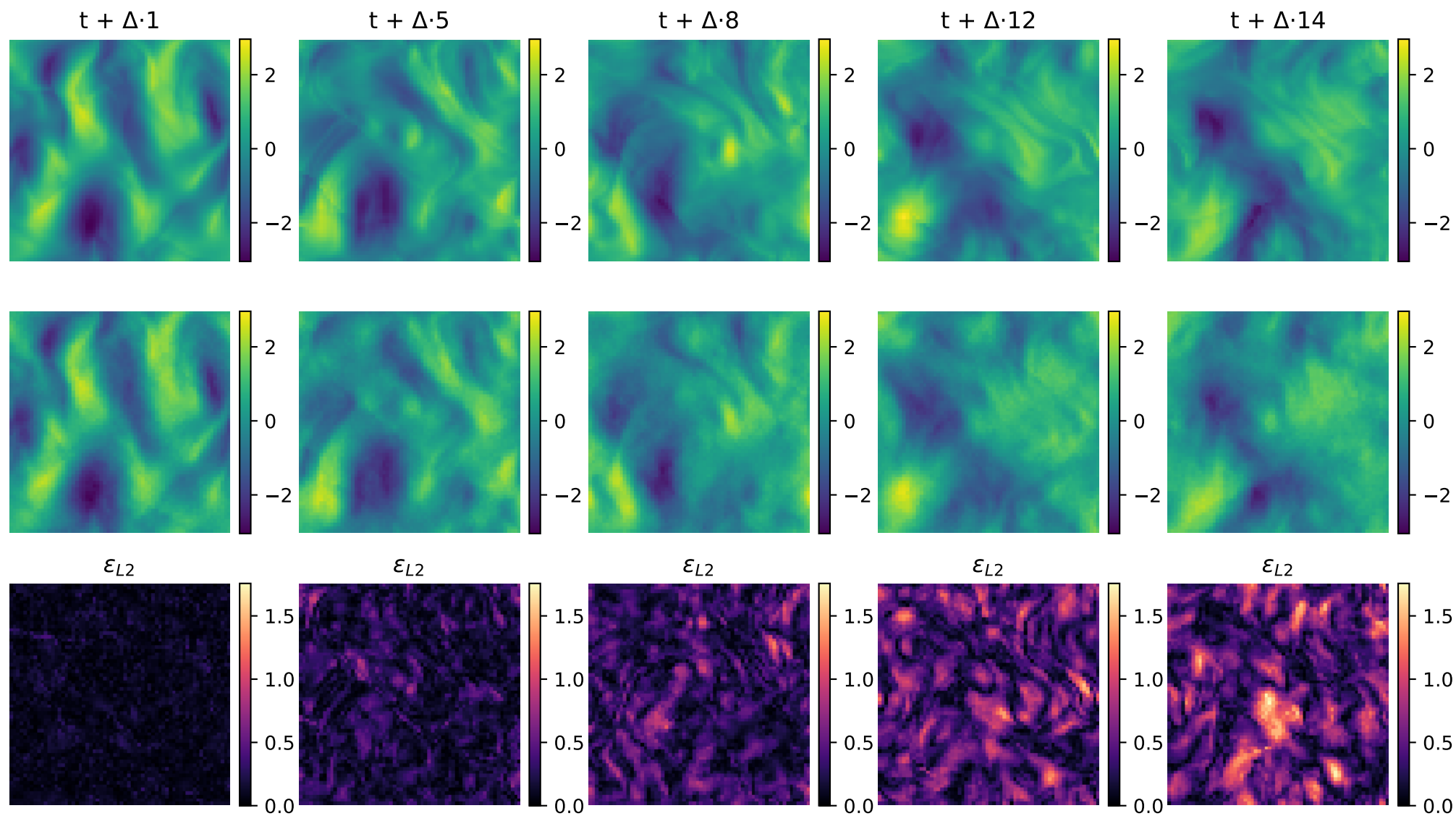
Rollout Snapshot #1 -- ch 1 -- GT (top row) vs stage_1_student (middle row)
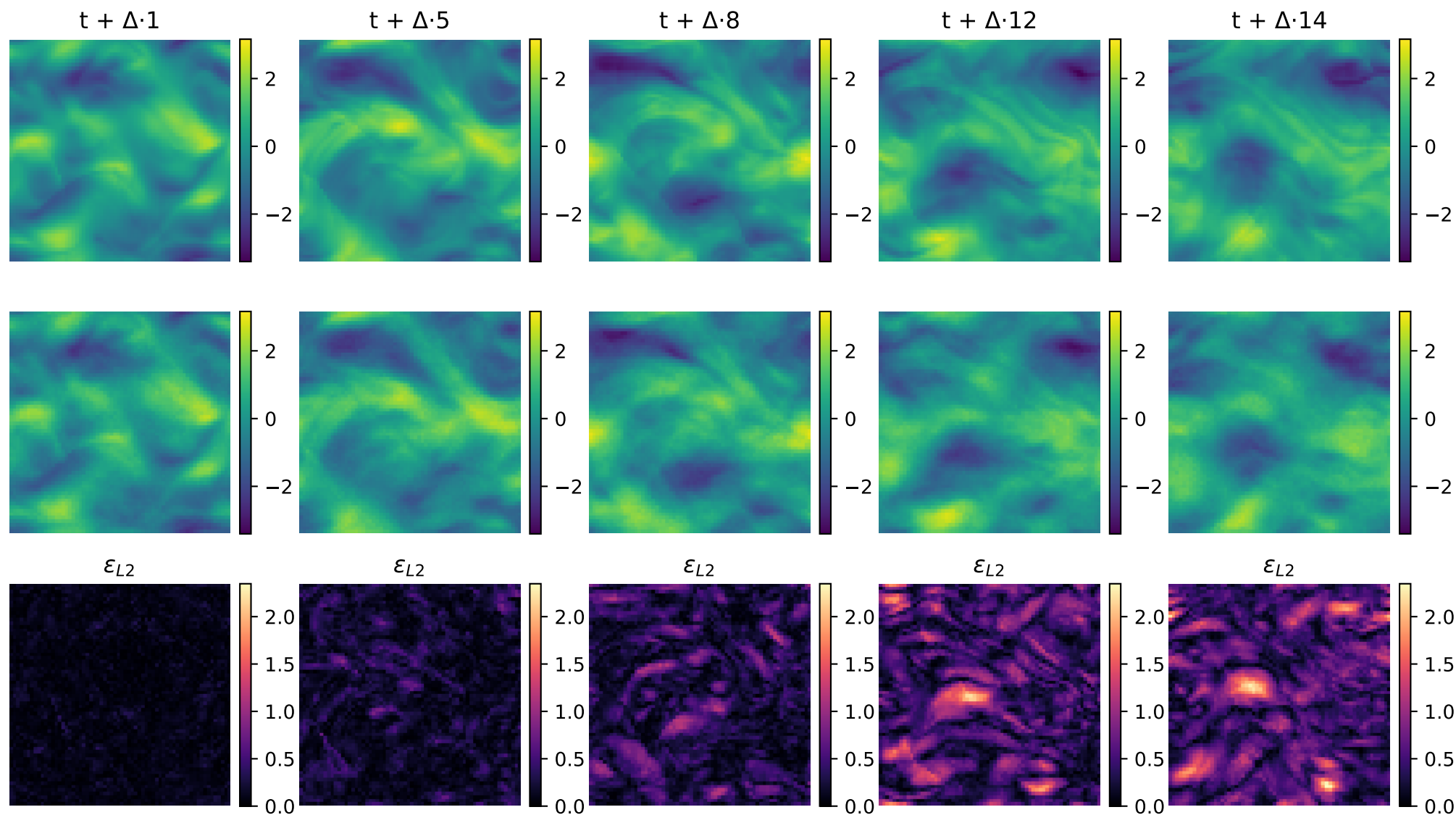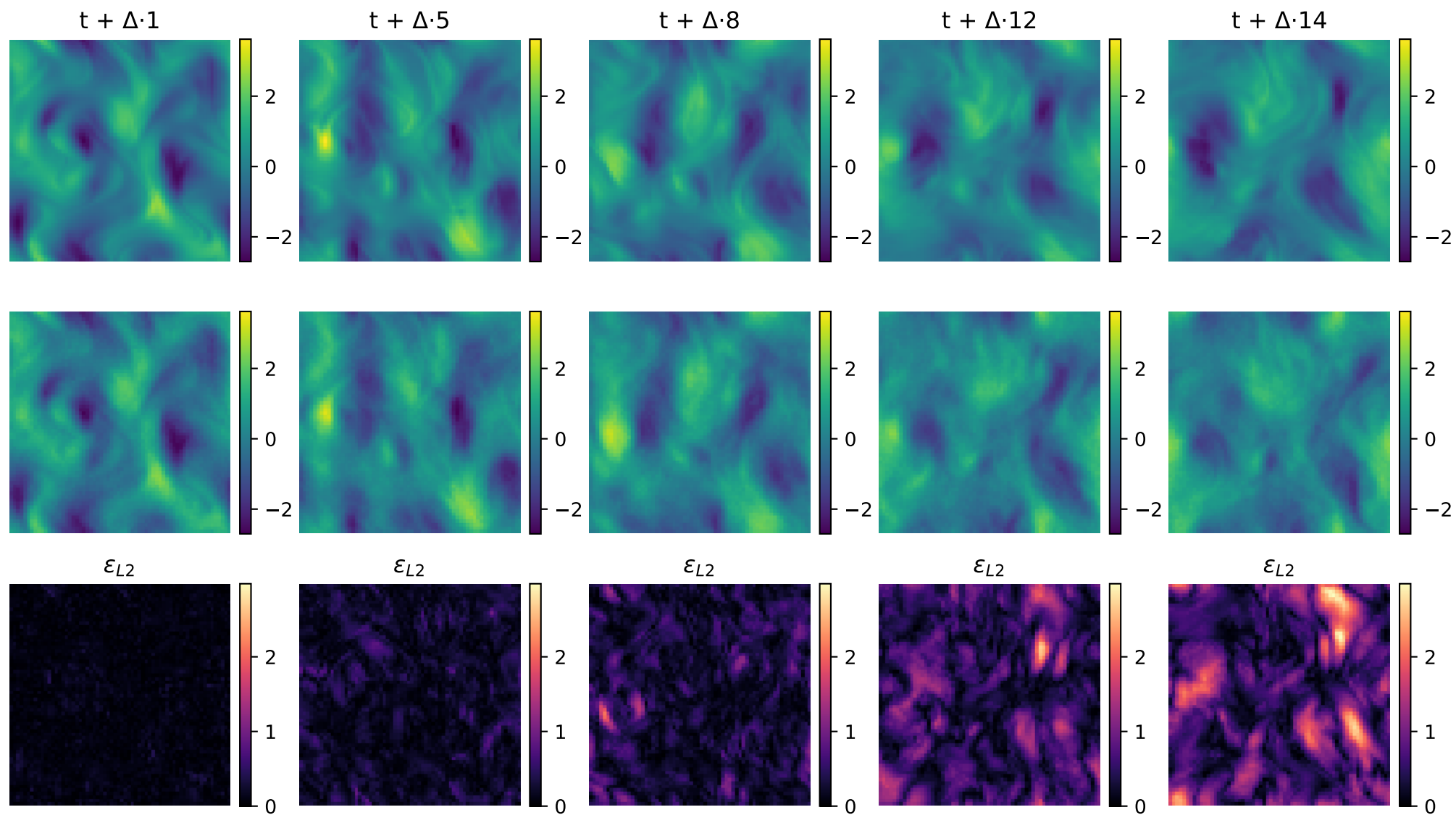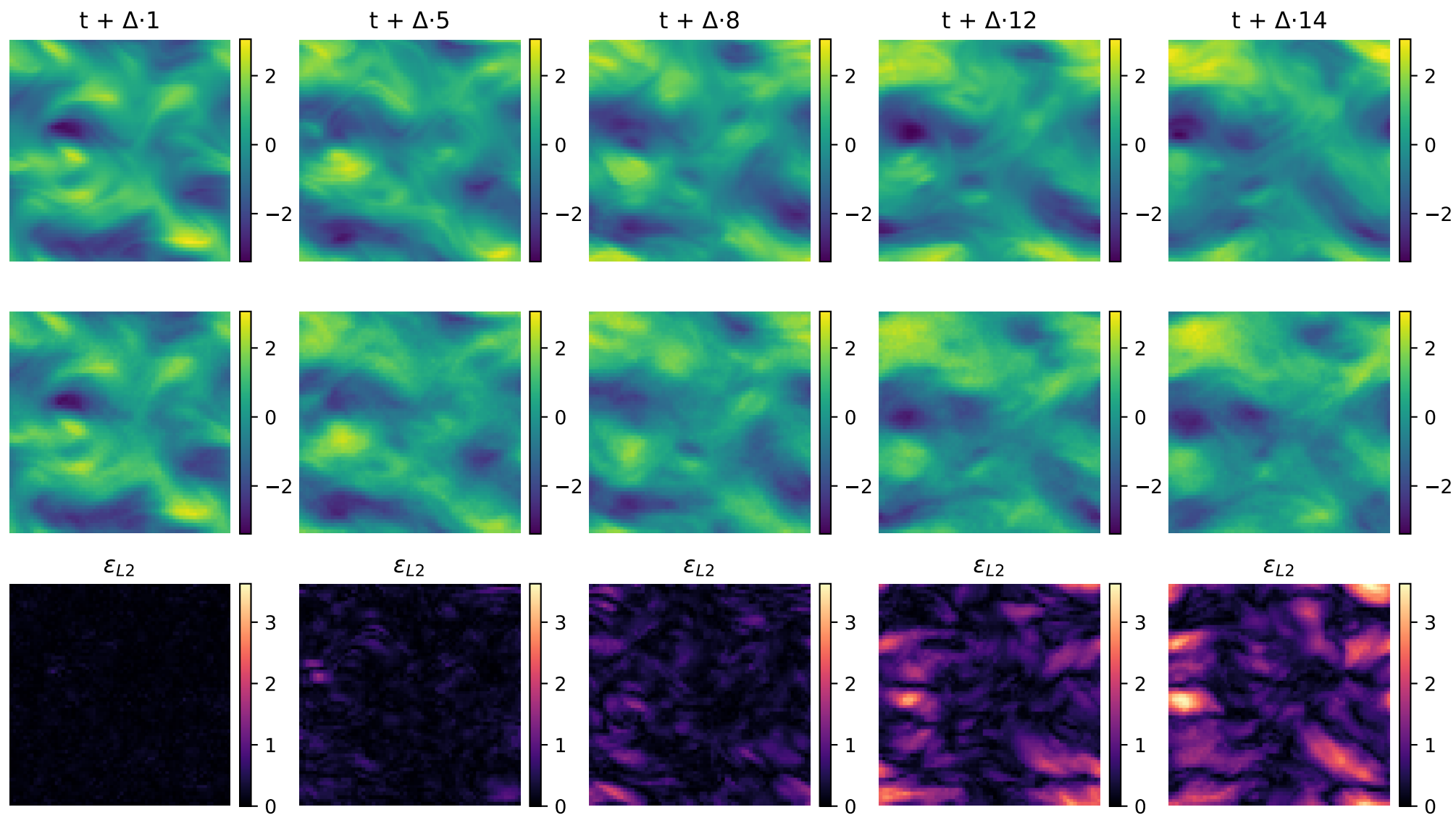
Rollout Snapshot #2 -- ch 0 -- GT (top row) vs stage_1_student (middle row)

Rollout Snapshot #2 -- ch 1 -- GT (top row) vs stage_1_student (middle row)

Evaluation Configuration

```
{
    "common": {
        "device": "cuda",
        "seed": 42,
        "samples": 16,
        "batch_size": 32
    },
    "models": [
        "velocity_pd",
        "velocity"
    ],
    "viz": {
        "enabled": true,
        "n_samples": 3,
        "channels": [
            0,
            1
        ],
        "colorbar": true,
        "rollout_steps": 5
    },
    "one_step": {
        "steps": {
            "velocity": 50,
            "flow": 1,
            "velocity_pd": 25
        },
        "method": {
            "velocity": "midpoint",
            "velocity_pd": "midpoint"
        }
    },
    "rollout": {
        "length": 16,
        "latent_policy": "new_latent",
        "steps": {
            "velocity": 50,
            "flow": 1,
            "velocity_pd": 25
        },
        "method": {
            "velocity": "midpoint",
            "velocity_pd": "midpoint"
        },
        "use_sequence_gt_for_metrics": true
    },
    "metrics": {
        "enabled": [
            "l2"
        ],
        "params": {
            "sob": {
                "alpha": 1.0,
                "beta": 0.0
            }
        },
        "use_denormalized": false
    },
    "artifacts": {
        "pdf": {
            "filename": "eval_report.pdf",
            "dpi": 150,
            "max_figs_per_model": 6,
            "include_rollout_curves": true
        },
        "save_csv": true,
        "save_json": true
    },
    "experiment": "ns2d"
}
```