

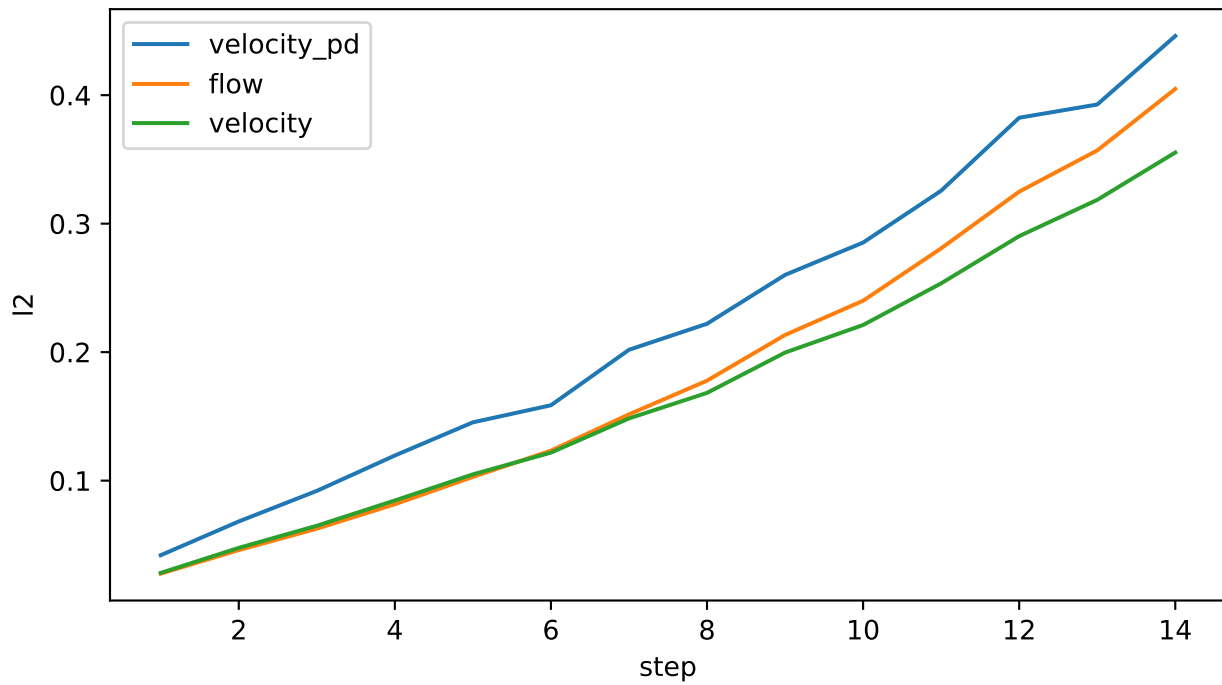
Evaluation Report

Models: velocity and flow velocity

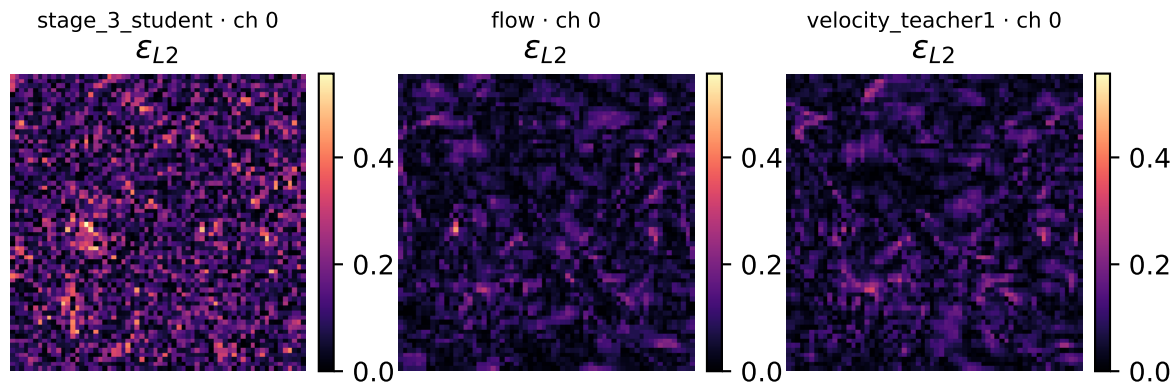
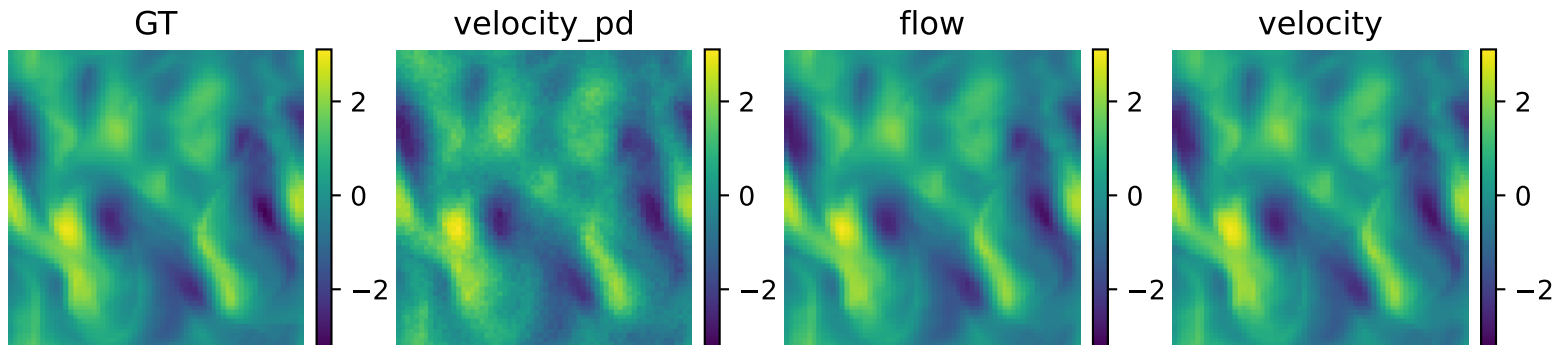
ns2d | 2025_09_12T04_48_16

mode	model	metric	mean	median	std	count
one_step	flow	l2	0.02839	0.02482	0.01278	128
one_step	velocity	l2	0.02922	0.02549	0.01318	128
one_step	velocity_pd	l2	0.04278	0.03966	0.01175	128
rollout	flow	l2	0.1853	0.1617	0.1247	1386
rollout	velocity	l2	0.1719	0.1518	0.1105	1386
rollout	velocity_pd	l2	0.2244	0.2064	0.1343	1386

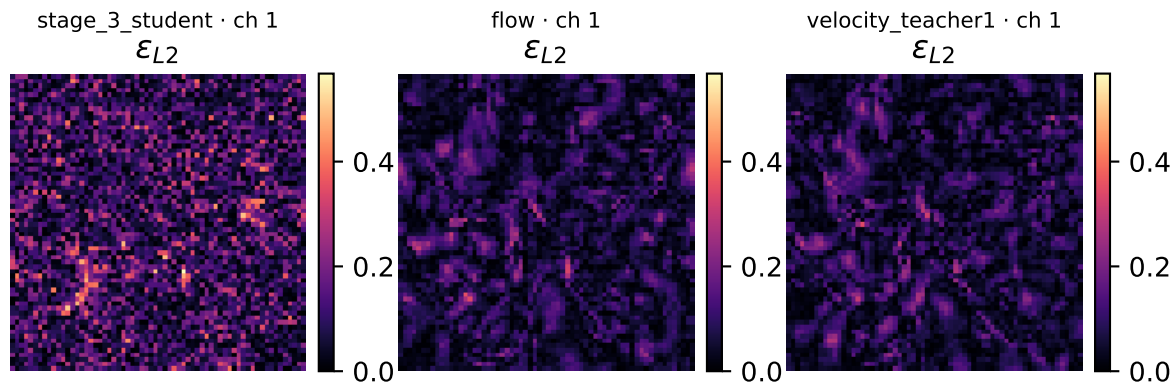
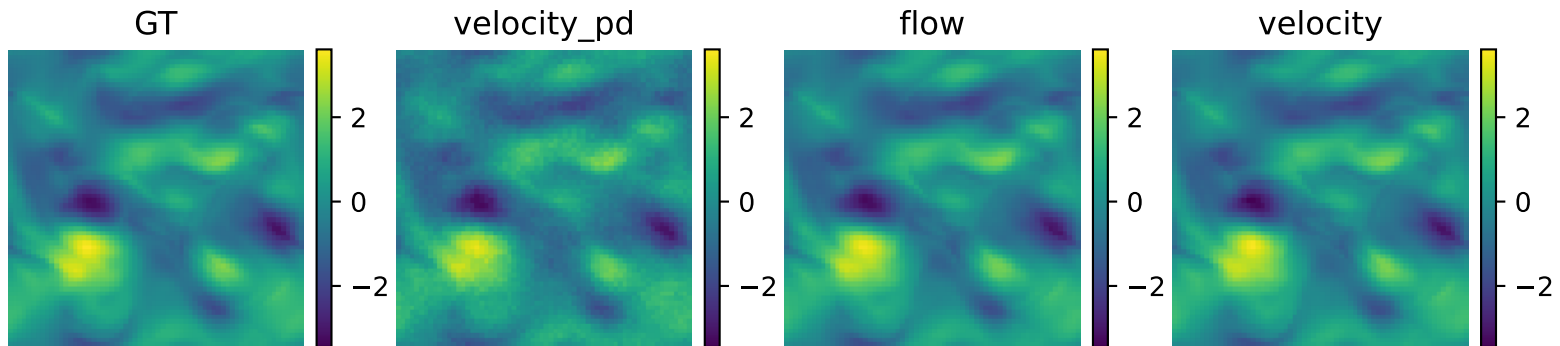
Rollout error curves (l2)



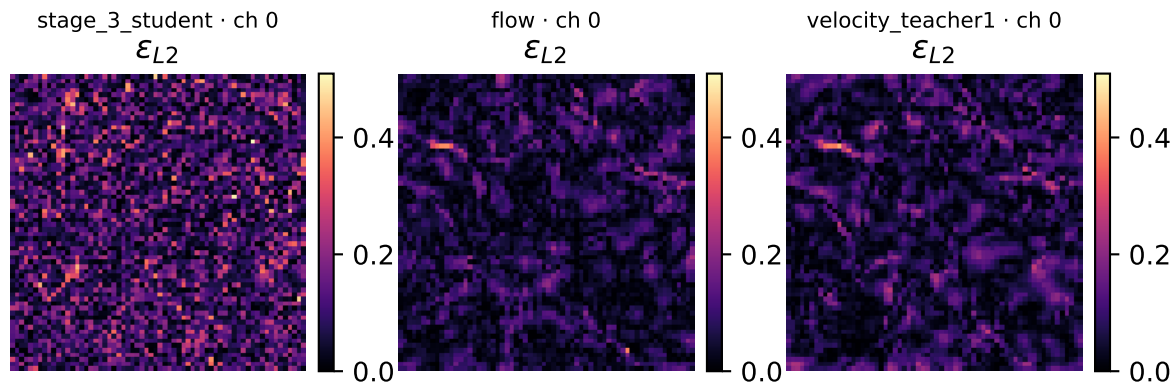
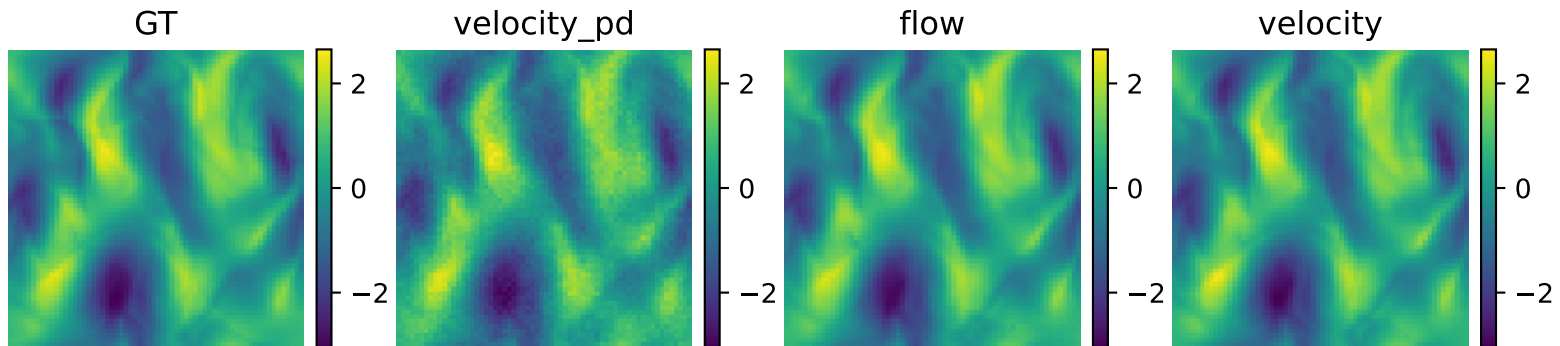
One-Step Snapshot #0 — ch 0



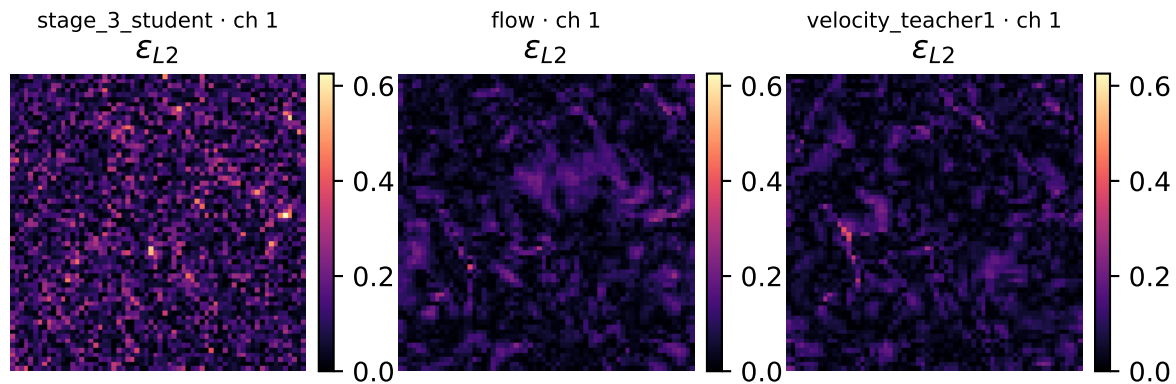
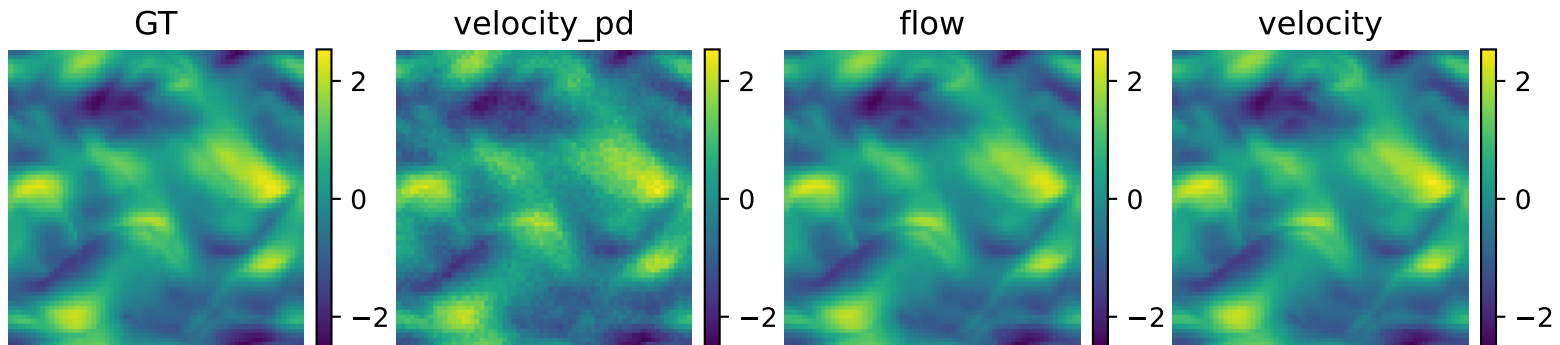
One-Step Snapshot #0 — ch 1



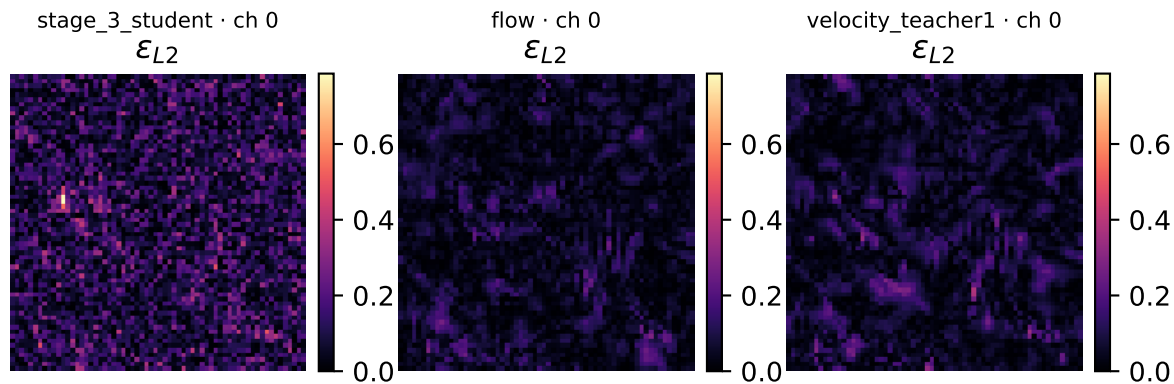
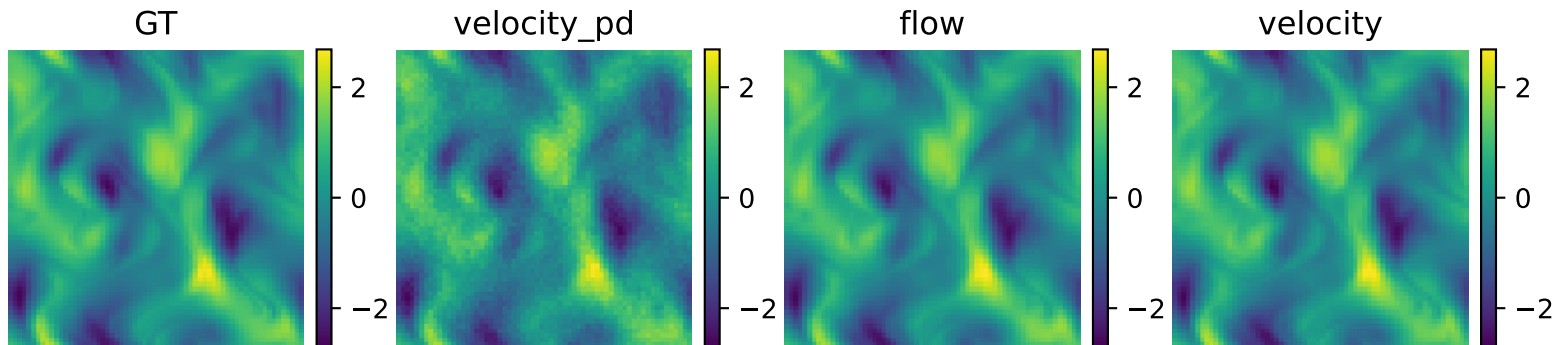
One-Step Snapshot #1 — ch 0



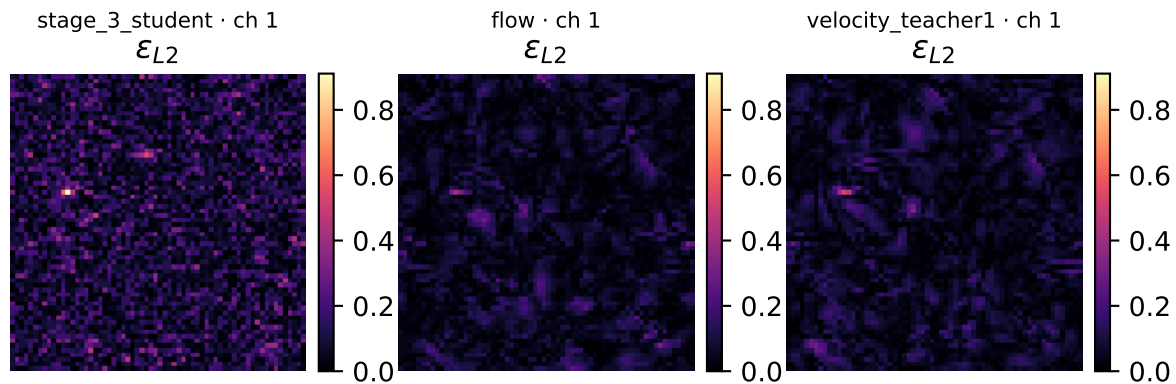
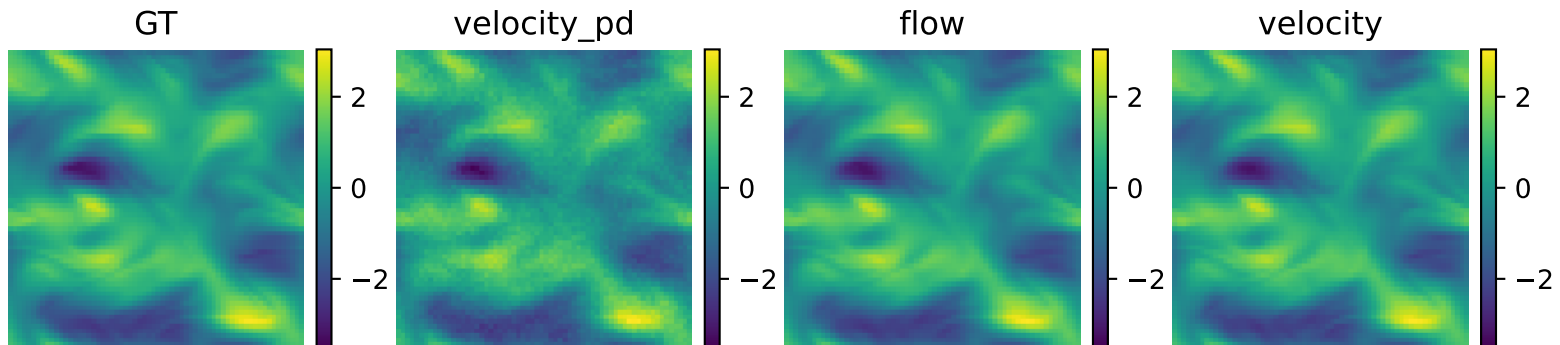
One-Step Snapshot #1 — ch 1



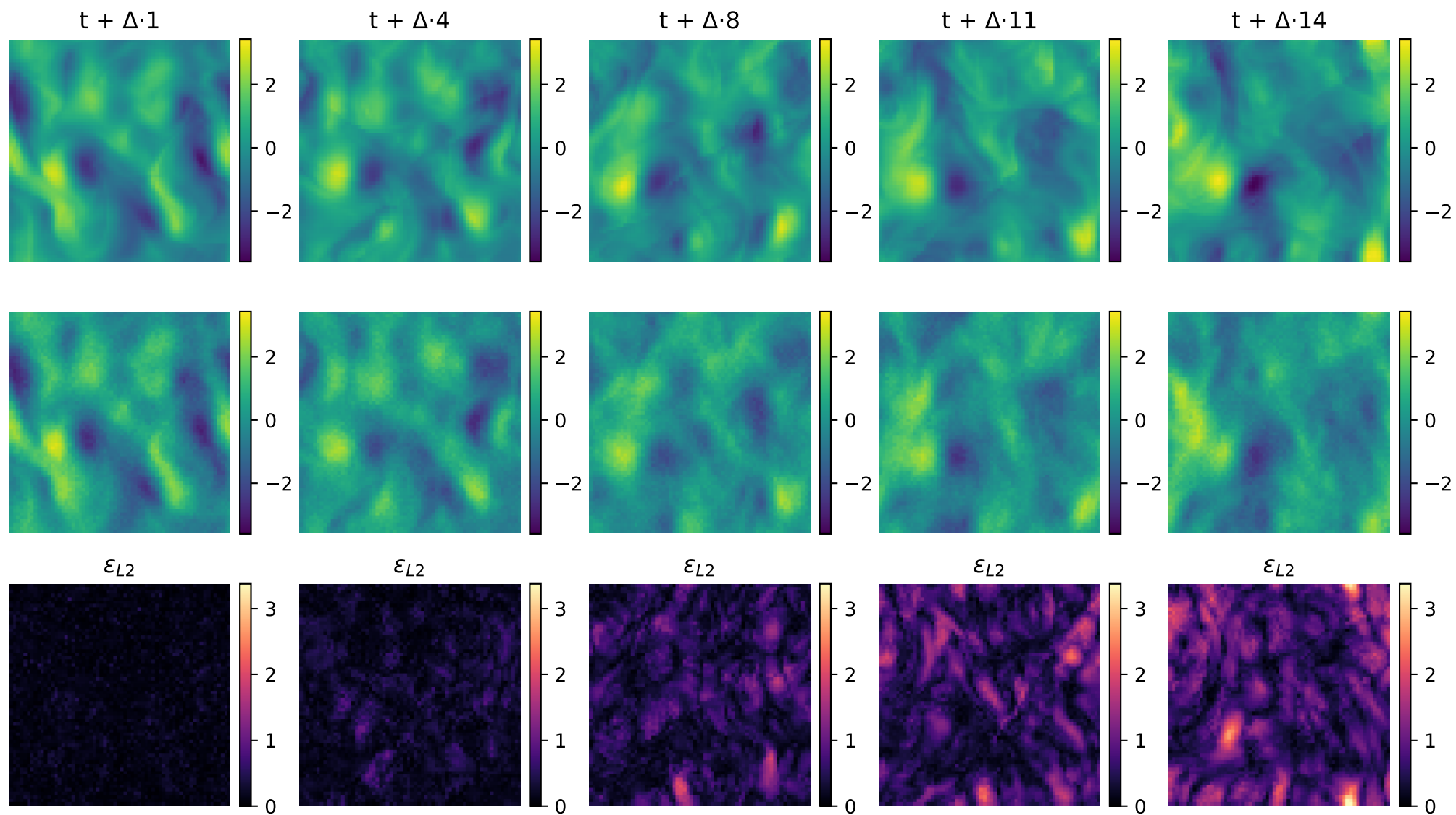
One-Step Snapshot #2 — ch 0



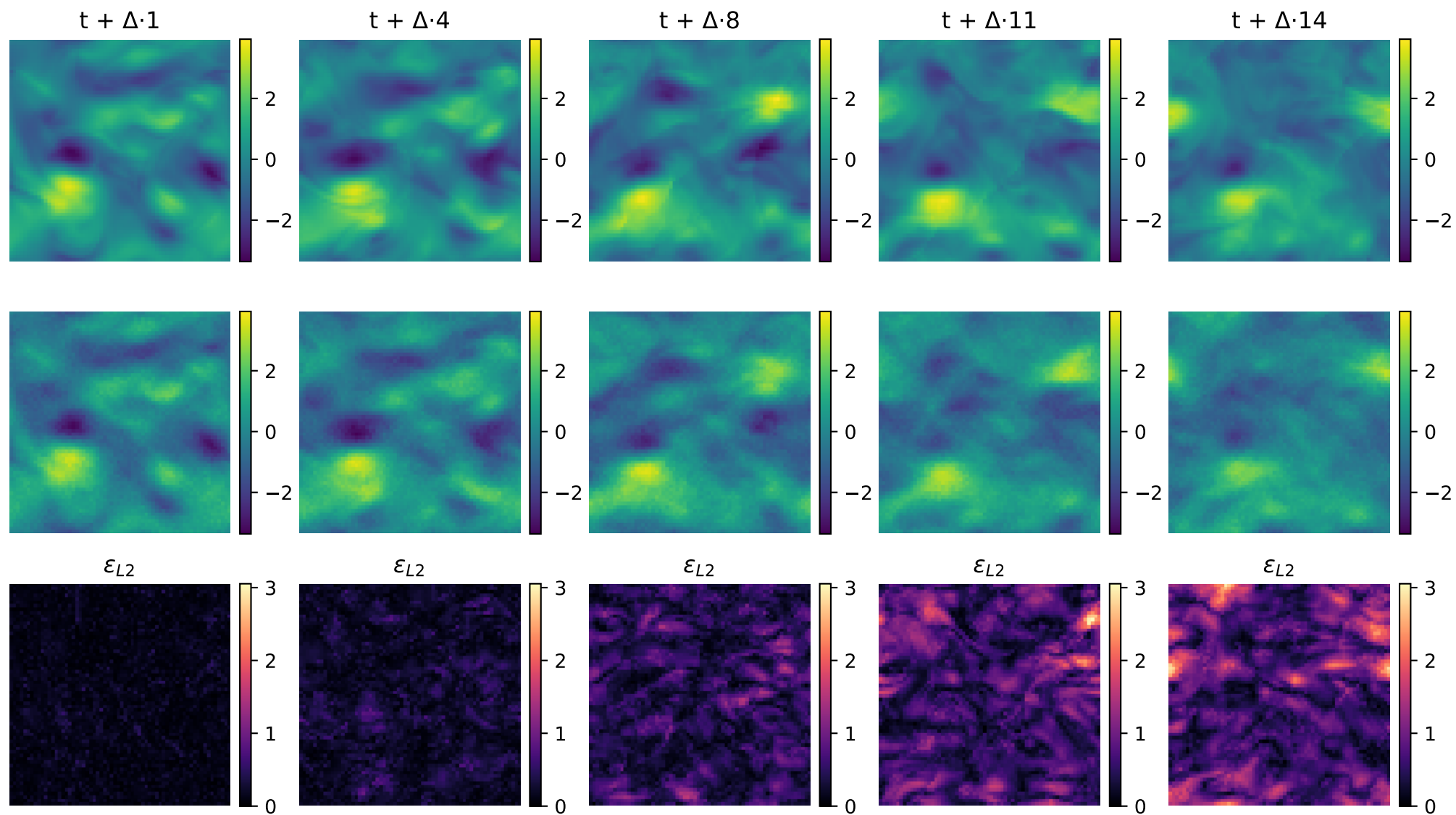
One-Step Snapshot #2 — ch 1



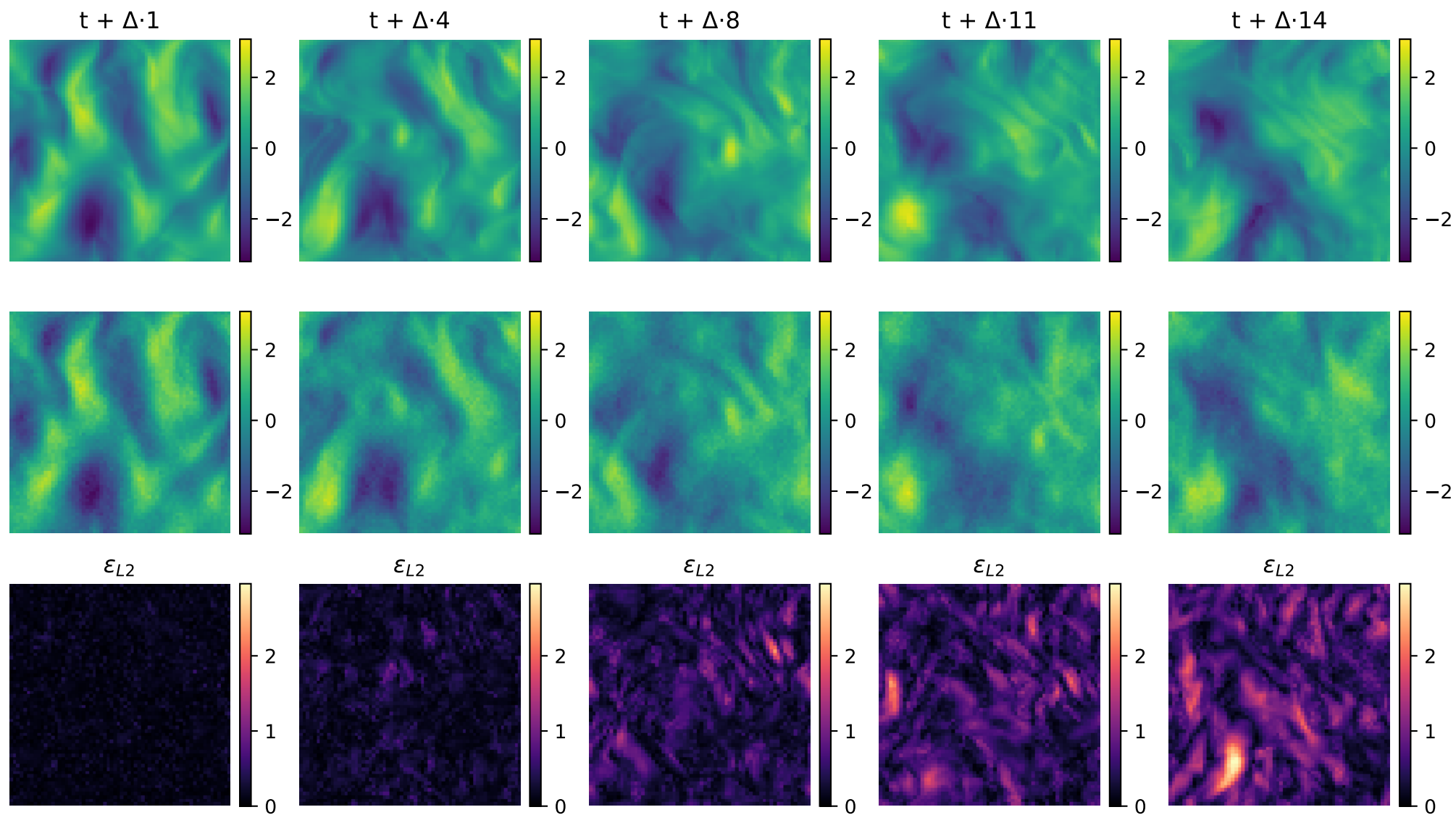
Rollout Snapshot #0 -- ch 0 -- GT (top row) vs stage_3_student (middle row)



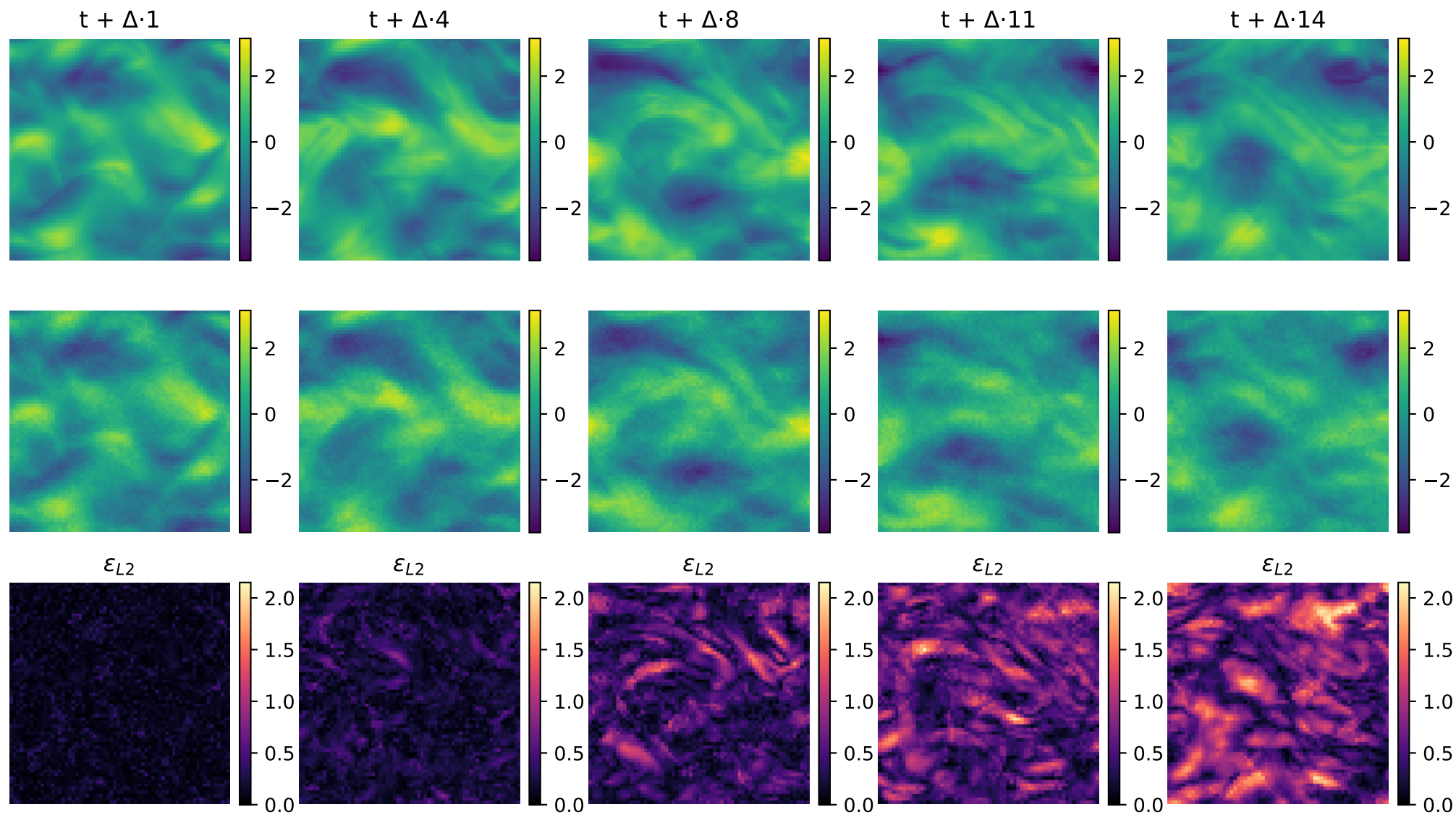
Rollout Snapshot #0 -- ch 1 -- GT (top row) vs stage_3_student (middle row)



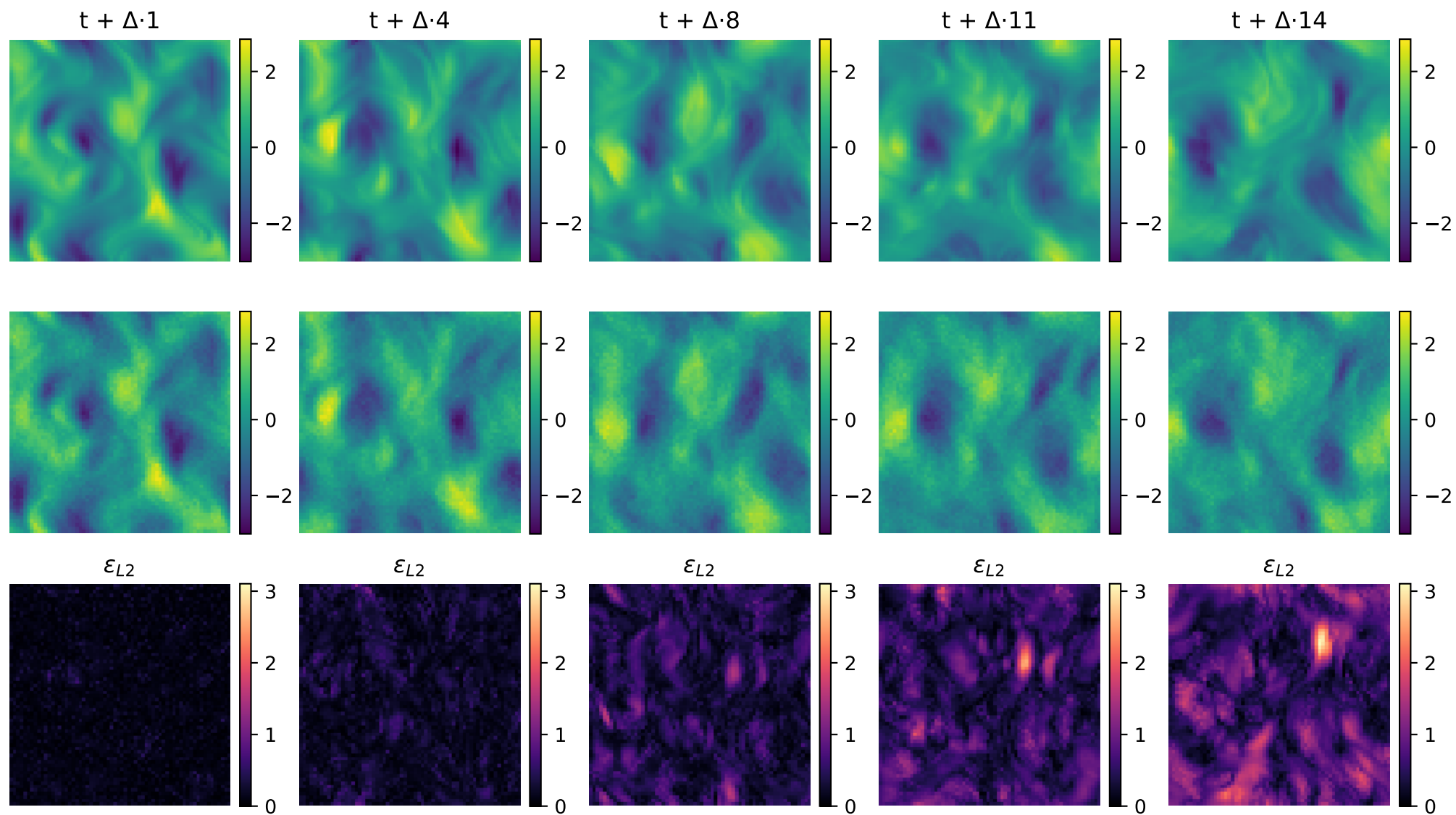
Rollout Snapshot #1 -- ch 0 -- GT (top row) vs stage_3_student (middle row)



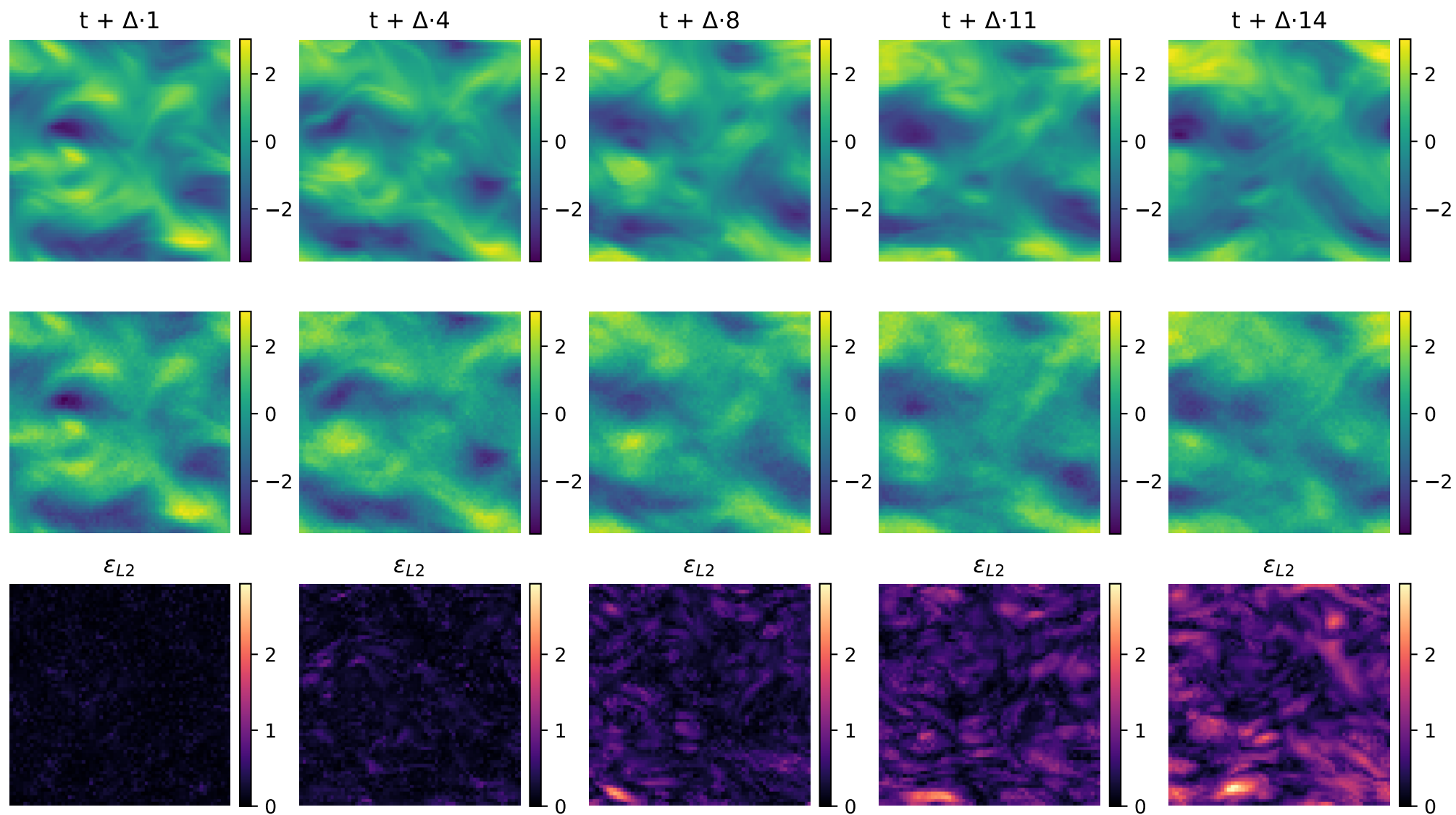
Rollout Snapshot #1 -- ch 1 -- GT (top row) vs stage_3_student (middle row)



Rollout Snapshot #2 -- ch 0 -- GT (top row) vs stage_3_student (middle row)



Rollout Snapshot #2 -- ch 1 -- GT (top row) vs stage_3_student (middle row)



Evaluation Configuration

```
{
  "common": {
    "device": "cuda",
    "seed": 42,
    "batch_size": 32,
    "samples": 99
  },
  "models": [
    "velocity_pd",
    "flow",
    "velocity"
  ],
  "viz": {
    "enabled": true,
    "n_samples": 3,
    "channels": [
      0,
      1
    ],
    "colorbar": true,
    "rollout_steps": 5
  },
  "one_step": {
    "steps": {
      "velocity": 50,
      "flow": 1,
      "velocity_pd": 6
    },
    "method": {
      "velocity": "midpoint",
      "velocity_pd": "midpoint"
    }
  },
  "rollout": {
    "length": 14,
    "latent_policy": "new_latent",
    "steps": {
      "velocity": 50,
      "flow": 1,
      "velocity_pd": 6
    },
    "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"
}
```