One-Week PINN Project — Checklist & Repo Layout (Denmark Soil Moisture)

A tight, printable 1-page checklist to run a full-time PINN MVP in 5 days. No code—just the playbook.

Snapshot

- **Goal:** Fit a PINN that reconstructs θ(z,t) at one Danish site (multi-depth sensors) with 1-D Richards + van Genuchten–Mualem; free-drainage bottom; flux top (rain − evap).
- **Outputs:** Cleaned dataset + data card, diagnostics, bucket baseline, PINN fit & learned params, 5–7 page report.
- **Cadence:** Hourly (preferred) or daily; Europe/Copenhagen timezone.

Day-by-Day

Day 1 — Data & EDA - [ ] Select station & season (≥3 depths, ≥3–5 rain events). - [ ] Acquire θ(z,t)+QC (ISMN/HOBE), meteo (DMI), optional soils (SoilGrids/OpenLandMap). - [ ] Harmonize: timezone, cadence, units (θ m^3/m^3; precip mm/step; depth m). - [ ] QC: drop bad flags; mask long θ gaps; smooth met spikes only. - [ ] Plots: θ(z,t) heatmap; event-aligned Δθ; rough water balance ($P - ET_0$ vs ΔS).

Day 2 — Problem Spec - [ ] Freeze physics (1-D Richards; van G-M; bare soil or fixed shallow uptake). - [ ] Boundary conditions: top flux (rain − evap), bottom unit gradient; initial θ profile. - [ ] Non-dimensionalization (z, t, head); write scales in spec. - [ ] Decide learned params: (α, n, Ks, θr, θs) global; (optional) two-layer. - [ ] Losses: data, PDE residual, BC/IC, parameter-prior penalties. - [ ] Collocation sampling plan (densify near surface & rain windows). - [ ] Time splits (Train 70% / Val 15% / Test 15%, keep events intact).

Day 3 — First Training Pass - [ ] Initialize params within pedotransfer bounds. - [ ] Curriculum: start data-heavy; ramp PDE/BC weights. - [ ] Track: per-depth RMSE (Train/Val), water balance, parameter ranges. - [ ] Early stop on Val; save best checkpoint.

Day 4 — Tighten & Compare - [ ] Build simple bucket baseline; calibrate 1–2 knobs. - [ ] Add small top-flux correction term (L2-penalized) if met forcing imperfect. - [ ] Ablation: remove parameter priors once; note Δ in metrics. - [ ] Choose best PINN variant by Test RMSE + physical plausibility.

Day 5 — Package & Report - [ ] Compute deep drainage q_bottom(t); cumulative drainage. - [ ] Compare learned (α, n, Ks, θr, θs) to soil-class ranges. - [ ] Quick LOEO (leave-one-event-out) sanity check. - [ ] Finalize artifacts (data, figs, config) + short report.

Data Prep & QC (Quick Checks)

- [ ] θ within [θr, θs]; no long saturation plateaus unless justified.
- [ ] Rain events list (thresholds: ≥2 mm/hr or ≥5 mm/day) for sampling/plots.
- [ ] ET_0 (FAO-56) computed once; used for sanity only.
- [ ] Soil texture by depth extracted; pedotransfer → **wide** bounds for (α, n, Ks, θr, θs).

Training & Evaluation

- [ ] Non-dimensional inputs/targets; gradient clipping enabled.
- [ ] Collocation resampling where PDE residual high (fronts, surface).
- [ ] Metrics: RMSE per depth; event timing (wetting-front lag); mass balance drift.
- [ ] Compare vs bucket baseline on Val/Test.

Success Criteria

- **Fit:** Test RMSE ≤ 0.03–0.05 m^3/m^3 per depth; plausible wetting-front timing.
- **Physics:** No θ>θs; reasonable drainage; learned params within soil-class ranges.
- **Beat baseline:** PINN improves RMSE and event timing vs bucket.

Risks & Fast Mitigations

- **Training instability:** data-first curriculum; normalize; clip grads.
- **Bad top flux:** enable small trainable correction (strongly penalized).
- **Heterogeneity:** defer layering unless Day 3 already stable.
- **Time sink:** keep MLP+tanh (5–8 layers); one file + config.

Repo Layout (barebones)

```
pinndk/
├─ README.md                  # What/why; quickstart; reproduction steps
├─ config.yaml                # Paths, cadence, hyperparams, loss weights
├─ spec.md                    # Physics, BCs, scales, losses, splits
├─ data/
│  ├─ raw/                    # Original downloads (ISMN, DMI, soils)
│  ├─ clean/                  # siteX.parquet/CSV + event_list.csv
│  └─ meta/                   # soil_priors.yaml; station_metadata.json
├─ figs/                      # Heatmaps, event stacks, fits, balance, tables
├─ src/
│  ├─ pipeline.py|jl          # Single entry: load→QC→EDA→train→eval→export
```

```
│  └─ utils.py|jl              # Minimal helpers (I/O, plots). Keep tiny.
└─ report/
   ├─ report.md                 # 5–7 pages w/ figures (or .tex/.docx)
   └─ tables.xlsx               # Metrics, learned parameters, comparisons
```

“Data Card” Template (fill once)

- **Site & Period:** name, lat/lon, start–end dates, timezone.
- **Sensors:** depths (m), sampling cadence, QC flags used, outages.
- **Meteo:** source(s), variables, cadence, missing intervals.
- **Soils:** texture by depth; source & date; pedotransfer used; parameter bounds.
- **Preprocessing:** resampling rules, unit conversions, filters.
- **Events:** detection thresholds; number of events; notable extremes.
- **Splits:** train/val/test dates; rationale.
- **Known Issues:** rain gauge gaps, sensor drift, site notes.

Notes

- Keep all plots legible in grayscale for printing.
- Log every assumption in `spec.md`; do not change mid-week without noting.