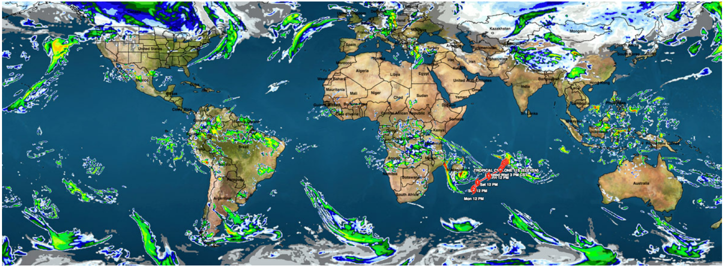
Overview on the weather forecasting

Kirill Semkin

October 9, 2025

What is the weather

Weather is the state of the atmosphere at a specific time and place. The atmosphere works as a giant fluid system which transfers mass and energy around the globe. The fluid is a complex mixture, but the main components are air (N_2, O_2) and water vapour (H_2O). The system is affected by the "external" forces: space (e.g. solar radiation) and the planet's surface (e.g. water sources, terrain heat, mountains). The overall dynamics are described by the fundamental laws of physics.



The main variables

The weather can be characterized by its motion, thermodynamics, components' phase states.

Symbol	**Meaning**
(u, v, w)	Zonal, meridional, and vertical velocity components
p	Pressure
T	Temperature
Φ	Geopotential (related to gravitational potential energy)
ρ	Air density
q	Specific humidity
$f = 2\Omega \sin \phi$	Coriolis parameter, depends on latitude ϕ
R	Gas constant for dry air
c_p	Specific heat at constant pressure

Preliminaries to the Governing dynamics

1. Every physical model has its limits. The chosen equations have to be adequate to the real observations.
2. The equations must describe how the air masses are moved, how different types of energy are moved, and how proportions of different air components are changed.
3. Introduce *the material derivative*

$$\frac{\mathrm{D}y}{\mathrm{D}t} \equiv \frac{\partial y}{\partial t} + \mathsf{u} \cdot \nabla y,$$

It basically corresponds to the total variation of y in time if the y is associated with the moving flow u. In our case u will be the moving air masses.

The Governing dynamics

1. The air in the atmosphere is quite sparse so we can use the **ideal gas** model as a state equation

$$p = \rho R T. \tag{1}$$

2. **Continuity Equation** describes mass conservation

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathrm{v} = 0. \tag{2}$$

3. Because the atmosphere is thin, vertical acceleration is negligible, so we have balanced forces - **Hydrostatic balance**:

$$\frac{\partial p}{\partial z} = -\rho g. \tag{3}$$

   This links pressure and altitude — a key simplification in global models.

The Governing dynamics

4. The horizontal acceleration of the masses is described by the **Navier–Stokes equations**. It is basically the Newtone's second law for viscous fluid

$$\frac{Du}{Dt} - fv = -\frac{1}{\rho}\frac{\partial p}{\partial x} + F_u, \tag{4}$$

$$\frac{Dv}{Dt} + fu = -\frac{1}{\rho}\frac{\partial p}{\partial y} + F_v, \tag{5}$$

   where F_u, F_v are frictional forces (like viscosity force $= \nu \Delta \boldsymbol{v}$, ν is kinetic viscosity).
5. **Thermodynamic Energy Equation** describes differential change of the heat in the "small" air parcel.

$$\frac{DT}{Dt} - \frac{1}{\rho \cdot c_p}\frac{Dp}{Dt} = \frac{Q}{c_p}, \tag{6}$$

   where Q is diabatic heating (radiation, latent heat, etc.). The diffusion term $\nabla \cdot (k \nabla T)$ is negligible relative to convection, work of expansion and external sources.

6. **Moisture Conservation**

$$\frac{Dq}{Dt} = S_q, \tag{7}$$

where S_q includes sources/sinks (condensation, evaporation, precipitation).

Numerical integration - classical weather prediction

The basic computational scheme is:

1. The Earth is covered by the grid (Cubed-Sphere, Icosahedral).
2. Apply finite differences to the derivatives (time and space)
3. Set grid-specific boundary conditions and initial conditions from the meteostations, satellites, etc.
4. Integrate forward

Do we have scheme's numerical approximation?

Typical size of the grid cell is 10-25 km (because of the enormous size of the equations' system). Not only do we have no guarantees for numerical convergence, but we also miss small-scale processes which are smaller than the grid cell. *That's one reason why we don't have precise weather forecast.*

Weather is a chaotic dynamical system

This is the second reason why accurate predicting is impossible. For illustration we will use classical *Lorenz system*:

$$\frac{dX}{dt} = \sigma(Y - X),$$
$$\frac{dY}{dt} = X(\rho - Z) - Y,$$
$$\frac{dZ}{dt} = XY - \beta Z.$$

These equations can be derived from the previous physics laws using several simplifications. The variables here are:

- X: The rate of convection (how intensely the fluid is rolling). A positive X means clockwise rotation, negative means counterclockwise.
- Y: The temperature difference between the ascending and descending fluid currents.
- Z: The distortion of the vertical temperature profile from linearity (how much the temperature graph deviates from a straight line).

Weather is a chaotic dynamical system

$$\frac{dX}{dt} = \sigma(Y - X),$$
$$\frac{dY}{dt} = X(\rho - Z) - Y,$$
$$\frac{dZ}{dt} = XY - \beta Z.$$

The parameters σ, ρ, and β are dimensionless numbers related to the physical properties of the fluid:

- σ: The Prandtl number (ratio of viscous diffusion to thermal diffusion).
- ρ: The Rayleigh number (ratio of buoyancy to viscous forces), normalized by a critical value. This is the primary "heating" parameter.
- β: A geometric factor related to the aspect ratio of the convection cells.

Weather is a chaotic dynamical system

When $\rho = 28, \sigma = 10, \beta = 8/3$ the Lorenz system has chaotic solutions (but not all solutions are chaotic). Almost all initial points will tend to an invariant set - the *Lorenz attractor* – a strange fractal attractor. System's trajectories heavily depends on the initial conditions.

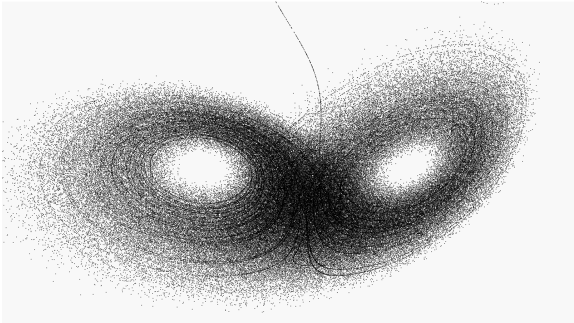


Figure: The Lorenz attractor

Numerical integration - Parametrization

The small-scale processes such as

1. Individual clouds or raindrops,
2. Turbulent eddies in the boundary layer,
3. Vertical convection (thunderstorms, updrafts, downdrafts),
4. Radiative transfer through clouds and gases,
5. Surface fluxes of heat and moisture,

can't be restored from the grid solution. These are crucial factors for the weather formation and can't be ignored. To embed them into the system we use *parametrization* - replacing the explicit (unresolvable) physical process with a simplified, statistical, or empirical formula that expresses its average effect on the larger-scale (resolved) flow.

Example: Convective heating rate Q_c

It describes vertical heat transfer by the clouds. For some function f, Q_c can be approximated from the gird as

$$Q_c = f(T, q, p).$$

ML models

Main idea - just solve regression problem for the physical variables using powerful models and large dataset.

1. *Architectures*: RNN, CNN, Transformers
2. *Input*: Gridded atmospheric data at the current time (e.g., geopotential, temperature, wind at different heights)
3. *Output*: The same grid, but at 6, 12, 24 hours, etc. into the future
4. *Pros*: High forecast speed, high accuracy (short, mid-term)
5. *Cons*: Data dependence, difficulty with extreme events, long-term stability issues

The demonstartion code - ResNet for global temperature forecasting.

Comparison of the approaches

Table: Traditional vs. ML Approaches

Criterion	**Traditional Physical Models**	**Pure ML Models**
Basis	Physics equations	Historical data (re-analyses)
Speed	Slow (hours of computation)	Very fast (seconds/minutes)
Interpretability	High (physically based)	Low ("black box")
Accuracy	Very high, but with systematic errors	Comparable to or exceeds best models at medium ranges
Extreme event representation	Good, through physics	Data-dependent, may underestimate
Main cost	Computation (supercomputer)	Model training (very expensive), inference is cheap

PINNs: ML model $\to$ Classical

The ML model approximate solution function for some variable $\hat{u} = f(t, \boldsymbol{x})$. It learns by maximizing empirical risk function (like MSE) $\mathcal{L}_{\text{data}}$ on the data. We can regularize the risk using the discrepancy of $\hat{u}$ when substituted into the physics equations:

$$\mathcal{L}_{\text{physics}} = \|g(u, \partial_t u, \nabla u, \ldots)\|^2,$$

where g is the physics equations (Naive-Stocks, Thermodynamics, etc.). The final loss is a linear combination:

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda \mathcal{L}_{\text{physics}} \tag{8}$$

The approach is very similar to the *penalty method* in the constraint optimization.

The demonstration on the Taylor-Green vortex is available.

Neural parametrizations: Classical → ML model

NNs can be used in the classical models as more accurate parametrization of sub-grid processes instead of hand-made heuristics. They also can adjust outputs of the classical models e.g. correcting temperature or precipitation probability for a specific location.

Publications on the topic:

1. Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., and Gentine, P. (2021). *Enforcing analytic constraints in neural networks for modeling physical processes with irregular outputs*. Journal of Advances in Modeling Earth Systems (JAMES), 13(2).
2. Subramanian, S., Chattopadhyay, A., Gunzburger, M., and Balajewicz, M. (2023). *Physics-informed neural networks for atmospheric dynamics*. Geoscientific Model Development (GMD), 16(3), 943–960.

Current state of affairs: WeatherBench

	Geopotential (500hPa geopotential RMSE [kg²/m²])					Temperature (850hPa temperature RMSE [K])					Humidity (700hPa specific humidity RMSE [g/kg])					Wind Vector (850hPa wind vector RMSE [m/s])				
Lead time [days]	1	3	5	7	10	1	3	5	7	10	1	3	5	7	10	1	3	5	7	10
IFS HRES	41	136	306	521	809	0.63	1.19	1.84	2.62	3.62	0.53	0.96	1.28	1.55	1.84	1.60	3.23	5.15	7.07	9.13
IFS ENS (mean)	40	132	278	439	621	0.62	1.09	1.60	2.14	2.76	0.50	0.82	1.05	1.22	1.38	1.57	2.93	4.40	5.70	6.90
ML / hybrid models: Pangu-Weather (oper.)	45	138	304	517	799	0.68	1.12	1.76	2.55	3.57	0.53	0.87	1.18	1.46	1.78	1.69	3.03	4.88	6.81	8.91
GraphCast (oper.)	38	120	270	465	747	0.56	0.95	1.55	2.31	3.36	0.47	0.75	1.01	1.28	1.60	1.46	2.68	4.41	6.28	8.47
GenCast (oper.) (mean)	39	124	264	421	607	0.56	0.95	1.47	2.04	2.69	0.48	0.76	0.98	1.15	1.33	1.49	2.70	4.16	5.48	6.77
Aurora (oper.)	38	119	262	442	668	0.56	0.95	1.50	2.15	2.97	0.47	0.77	1.01	1.21	1.43	1.43	2.67	4.21	5.68	7.14
FGN (oper.) (mean)	32	104	233	388	583	0.51	0.86	1.34	1.91	2.59	0.44	0.70	0.91	1.10	1.30	1.33	2.43	3.82	5.20	6.60
Pangu-Weather	44	134	295	503	789	0.62	1.05	1.71	2.51	3.56	0.53	0.86	1.16	1.45	1.76	1.65	2.98	4.80	6.71	8.85
GraphCast	39	122	270	459	732	0.51	0.93	1.54	2.30	3.36	0.46	0.77	1.03	1.28	1.58	1.40	2.71	4.42	6.20	8.27
Climatology	810	810	810	810	810	3.41	3.41	3.41	3.41	3.41	1.59	1.59	1.59	1.59	1.59	7.88	7.88	7.88	7.88	7.88

−50 −20 −10 −5 −2 −1 1 2 5 10 20 50

Better ⟵ % difference in RMSE vs IFS HRES ⟶ Worse

Figure: Models' metrics compared to the IFS model on different prediction tasks