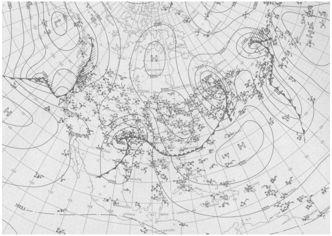
Improving Numerical Weather Forecasts by Clustering

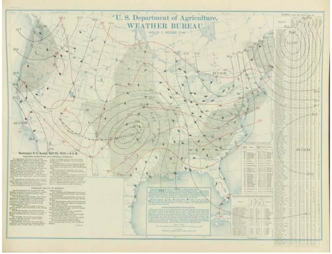
Pavel Krč

Institute of Computer Science
Academy of Sciences of the Czech Republic

Nov 6, 2014

Traditional (judgmental) forecasting

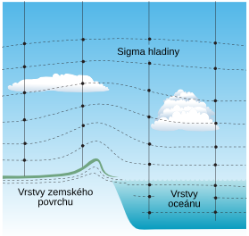




Forecast made by expert(s) using solely measurements and own knowledge/experience.

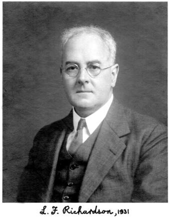
- Analyzing and modifying synoptic maps by hand
- Spot observations from previous day(s)
- Mostly general circulation
- Typical outputs:
    - Greater areas (hundreds of km)
    - Daily averages and/or extremes

Numerical weather prediction (NWP)



source: Australian Boureau of Meteorology

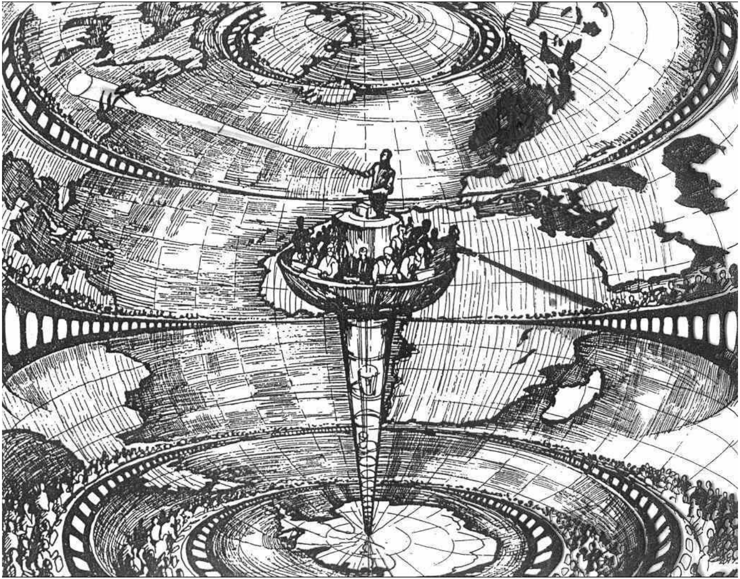
- State variables on a grid
- Navier–Stokes equations, thermodynamics
- L. F. Richardson, 1920's:



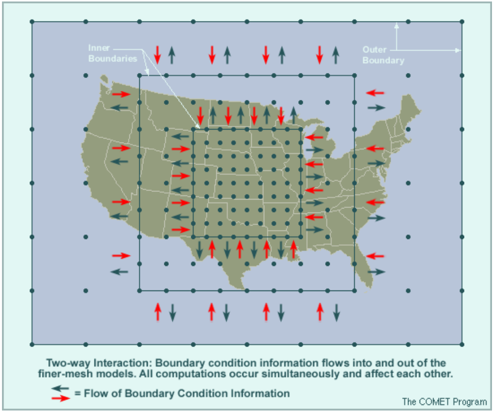
2 points, 6 hours forecast ⇒ 6 weeks computation

- Nowadays: resolution in order of km, weeks ahead

Richardson's NWP



Nested domains



1. Global models
   - data assimilation
   - probabilistic forecast
2. Mesoscale models (LAM)
   - grid nudging, nestdown and nesting

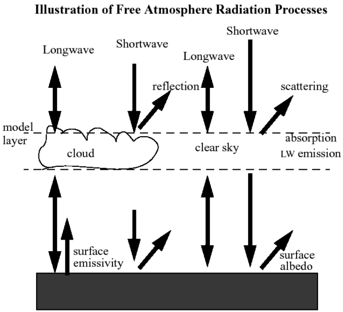
Parametrizations

- Parametrization = (statistical) simplification
- Processes of too small scale
- Modular
- Configuration and tuning

Example

Convection, PBL, ground surface, microphysics, radiative transfer.

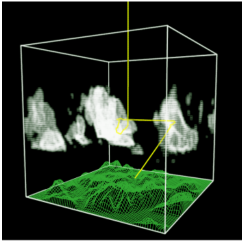
Modeling radiative transfer



Main processes

- Absorption
- Scattering
- Reflection

- Subgrid level



Important inputs

Solar geometry, albedo, atmospheric water, aerosols, ozone

Deficiencies of all–purpose NWPs

Model variables

state (prognostic) variables × derived (diagnostic) variables

- Not all model variables are regularly and properly evaluated
- Even derived variables might be tuned for model internals (e.g. radiation → heat fluxes)
- Well-captured overall weather state doesn't imply precise prediction of each variable

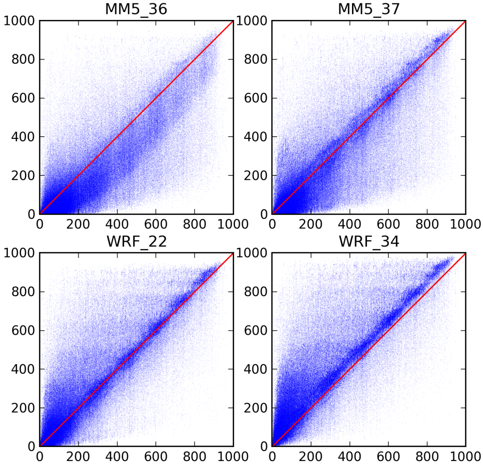
Postprocessing

- Measured error in many variables can be reduced by statistical postprocessing
  - local bias (spatial representativity)
  - similar weather patterns
- Shortwave irradiance errors: stratification (E. Lorenz et al.)
  - solar zenith angle ϑ_t
  - clearness (clear-sky) index k_t

$$GHI_t = ETR_t \cdot k_t \cdot \cos \vartheta_t$$

Multi–model prediction

- Combination of models with uncorrelated errors = strong improvement
  - different initial & boundary conditions
- Some improvement even with strongly correlated models
- MM5 and WRF: various radiation schemes and model versions



Weather classification and clustering

- Localized weather classification
  - Pasquill–Gifford stability class
- Synoptic classification
  - standardized, yet based subjective similarity
- NWP based clustering
  - analog ensembles (probabilistic forecast)

Experiment design

Idea: Combine weather type clustering, postprocessing and multi–model forecasting

Goals: Explore relative advantages of different NWPs under different weather conditions, outline possible model improvement

Conditions:

- Clustering input variables based on loosely defined target (expected relative model performance)
- High dimensionality of input data
- Explanatory result expected – shouldn't use e.g. PCA

Method

1. Use CART regression tree (forest) learning with vector of model deviations as a target variable
2. Identify clusters defined by leaf nodes
3. Optimize weights for linear combination of models within each cluster
    - bias removed directly
    - relative model performance expected indirectly (depth and branching limits)

Input data

Measurements:

- Hourly averages of global horizontal irradiation (GHI)
- 15 professional weather stations operated by CHMI

Forecasts:

- Four NWP models: MM5 3.6, MM5 3.7, WRF 2.2, WRF 3.4
- Identical nested domains
  - 9×9 km horizontal resolution covering Czech Republic
  - 27×27 km covering central Europe
  - NCEP GFS initial and boundary conditions
- Simulated operational regime
  - new forecast every 6 hours, 48 hour horizon
  - May 2011 to April 2012

Split by date: 20 × 18-day segments (15 days training + 3 days validation)

Meteorological variables for clustering

Meteorological variables:

- Air temperature at 2 m
- Air humidity at 2 m
- Wind at 10 m
- Precipitation amount
- Surface air pressure
- Planetary boundary layer height
- Water vapor & different precipitation types
  - low, medium and high altitude

Other predictors:

- Clearness index
- Extraterrestrial solar radiation
- Solar geometry
- Season indicator
- Latitude and longitude
- Terrain height

Experiment results and control results

Model	RMSE $[W \cdot m^{-2}]$ in-sample	RMSE $[W \cdot m^{-2}]$ out-sample
MM5 3.6		158.0
MM5 3.6 ⊙	140.4	144.0
MM5 3.7		146.5
MM5 3.7 ⊙	141.4	149.3
WRF 2.2		140.7
WRF 2.2 ⊙	123.1	121.6
WRF 3.4		167.9
WRF 3.4 ⊙	131.4	133.7
Simple combination	115.1	115.4
Combination of ⊙	114.3	115.7
Stratified combination	113.6	115.8
Random forests	106.3	113.2

⊙ = bias correction on stratified data

Further research

- Complex spatial information
  - only point forecasts used so far
  - spatial averages, differences and variances: metalearning of sizes and (relative) positions of areas
- Purposely excluded: localized learning
  - usually huge improvement against local measurements
  - not so usable for analyzing behavior