Application of Statistical Methods to Wind Forecasting for Wind Power Generation

September 2001

Alex Barnett

Lyman Laboratory of Physics, Harvard University

Collaboration with:

- Michael Brower (TrueWind, Andover, MA)
- John Zack (MESO, Albany, NY)
- Radford Neal (U. of Toronto CS/Statistics)

Funding: TrueWind Solutions LLC

Computing resources: Heller Group at Harvard Physics, MESO

Outline of today's talk

- Motivation
- Existing forecasting methods
- Data, variables & timescales
- Bayesian neural networks
- Wind velocity results
- Wind-to-power relation
- Power results
- Future directions & conclusions

Motivation: why forecast for wind power?

Electricity from wind...

- now most cost-effective renewable, safe
- worldwide installed capacity doubling every ~ 2 years
- wind fraction of total capacity: *e.g.* Denmark 30%, US $\sim 1\%$

Intermittent **resource:** increased wind fraction $\rightarrow$ problems

- Utilities keep reserve non-renewable capacity running (waste)
- *Trading* on electricity markets 1 day ahead
- Increased fluctuations in load (grid balancing problems)

Value of accurate daily forecasts, 0–2 days ahead:

e.g. $\sim$ \$20m per year to a larger US utility

Future penetration of wind power: improved forecasts of

- wind speed and direction
- power output of turbine or cluster of turbines

Existing forecasting methods

Two approaches:

1. **Extrapolation** of wind speed time series measured at site
   - no physical modelling
   - few seconds–1 hour ahead only
   - linear (*e.g.* Kalman filters) and nonlinear (neural net)
     → *small* improvement over persistence

2. **Hydrodynamic weather model simulation**
   - initialized daily from observations (weather services)
   - useful up to ∼ 4 days ahead (good models)
   - computer-intensive → ∼ 4 km min spatial resolution
   - are *systematic errors*:

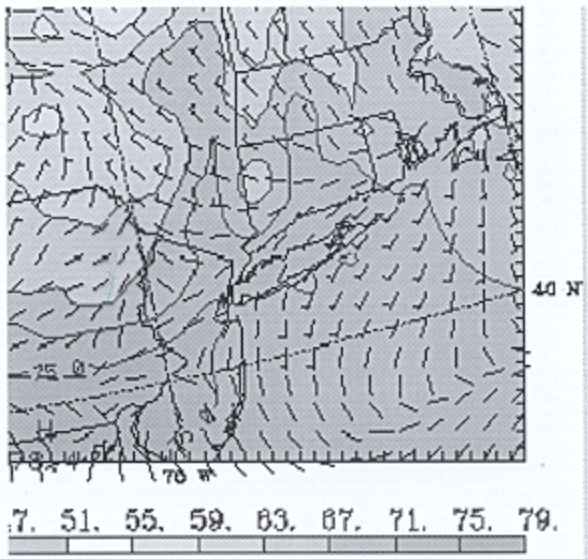
     model variables $\xrightarrow{\text{transfer function}}$ observed at site
   - TrueWind (USA), Risø(Denmark): linear error correction

Why systematic errors?

- local topography (< 4 km)
- turbines at windy spots!

Correcting errors
≡ learning transfer function
≡ **'Downscaling'** from grid to local site

'mesoscale' numerical model



This talk: BETTER DOWNSCALING FOR WIND POWER

Complex flows, thresholding effects ⇒ nonlinear methods

Model variables

Many, often highly correlated. At each grid point have...

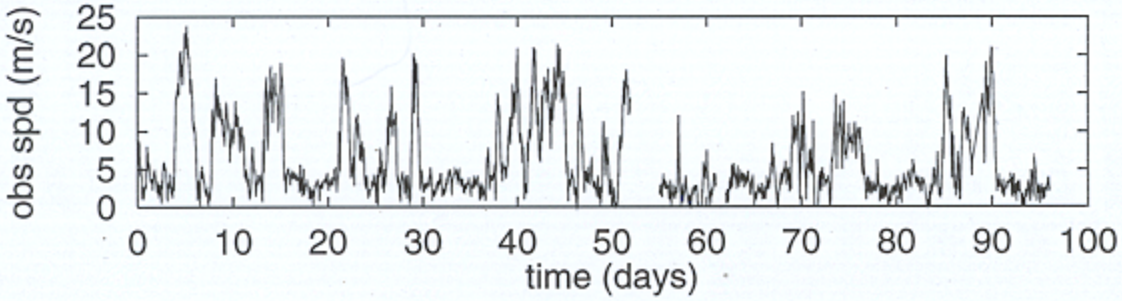
1 C	850 mb Temperature	(1)
1 kg/kg	850 mb H2O Vapor Mixing Ratio	(2)
1 m/s	850 mb U Wind Component	(3)
1 m/s	850 mb V Wind Component	(4)
1 m	850 mb Geopotential Height	(5)
1 1/sec	850 mb Absolute vorticity	(6)
1 C	500 mb Temperature	(7)
1 kg/kg	500 mb H2O Vapor Mixing Ratio	(8)
1 m/s	500 mb U Wind Component	(9)
1 m/s	500 mb V Wind Component	(10)
1 m	500 mb Geopotential Height	(11)
1 1/sec	500 mb Absolute vorticity	(12)
1 F	Surface Air Temperature	(13)
1 F	2 m Dew Point Temperature	(14)
1 m/s	10 m Wind Speed	(15)
1 F	Surface Dew Point Temperature	(16)
1 m/s	Surface U Wind Component	(17)
1 m/s	Surface V Wind Component	(18)
1 m/s	10 m U Wind Component	(19)
1 m/s	10 m V Wind Component	(20)
1 m/s	25 m U Wind Component	(21)
1 m/s	25 m V Wind Component	(22)
1 m/s	40 m U Wind Component	(23)
1 m/s	40 m V Wind Component	(24)
1 m/s	50 m U Wind Component	(25)
1 m/s	50 m V Wind Component	(26)
1 m/s	70 m U Wind Component	(27)
1 m/s	70 m V Wind Component	(28)
1 m2s2	10 m TKE	(29)
1 mb	Surface Altimeter Setting	(30)
1 F	Surface Skin Temperature	(31)
1 W/m2	Surface Sensible Heat Flux	(32)
1 W/m2	Surface Latent Heat Flux	(33)
1 F	2 m Temperature	(34)
1 frac	INTG Cloud Cover	(35)
1 frac	900 Low Clouds	(36)
1 frac	625 Mid Clouds	(37)
1 frac	275 High Clouds	(38)
1 kg/m2	INTG Liquid Water Path	(39)
1 W/m2	Model Top Outgoing Infrared Radiation	(40)
1 K	500 mb Lifted Index	(41)
1 %	Sfc-500mb Mean RH	(42)
1 m	1000-500mbThickness	(43)
1 in	Surface Total Precipitation	(44)
1 mm	Surface Convective Precipitation	(45)
1 kg/kg	850 mb Cloudwater Mixing Ratio	(46)
1 kg/kg	500 mb Cloudwater Mixing Ratio	(47)
1 kg/kg	300 mb Cloudwater Mixing Ratio	(48)
1 kg/m**2	SFC-VILFW	(49)
1 kg/m**2	SFC-Total Cloudwater	(50)
1 kg/m**2	SFC-Total Cloud Ice	(51)
1 kg/m**2	SFC-Total Rain water	(52)
1 kg/m**2	SFC-Total snow water	(53)
1 in	Surface Snow on Ground	(54)
1 in	Surface New Snow	(55)
1 mm	Surface Rain	(56)
1 in	Surface Snow	(57)
1 mm	Surface Freezing Rain	(58)
1 mm/hr	Surface Rainfall rate	(59)
1 in	Surface Snowfall rate	(60)
1 vol frac	Surface Shal Soil Moist	(61)

wind variables at various altitudes

Which variables are significant in downscaling to observed wind?

A look at some data

Wind speed at site in California (Nov 2000 — Feb 2001):

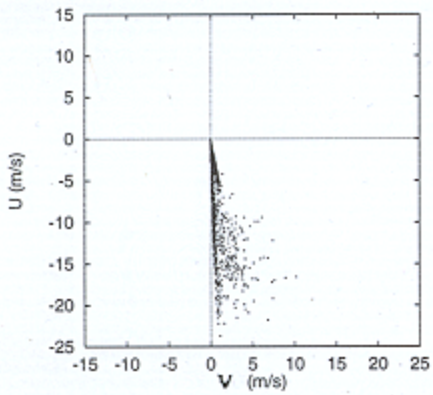


- flips between 'calm' and 'windy' (rarer events) states.
- no clear daily seasonality.

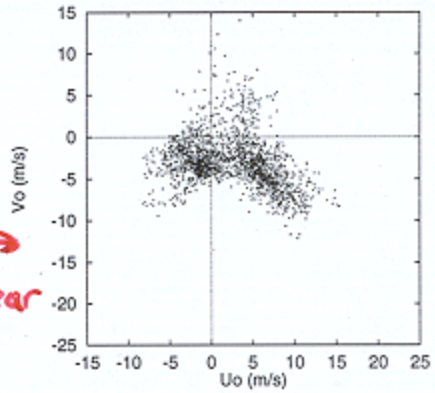
Scatter of wind **velocity vector**:

OBSERVED (u, v) at site

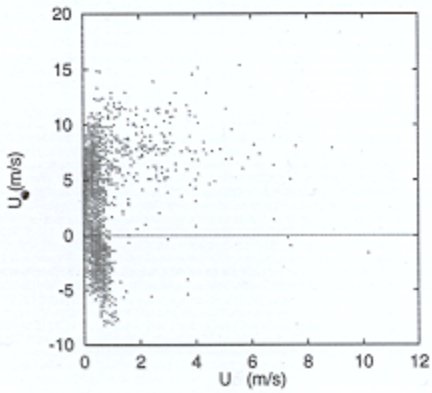
MODEL (u_0, v_0) zero altitude

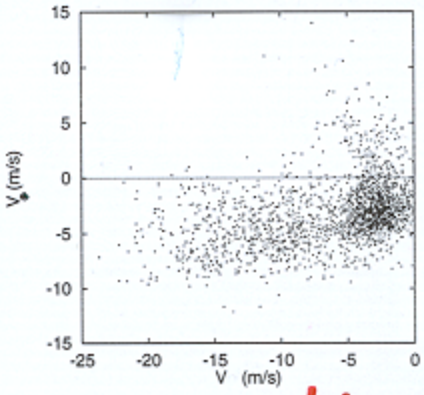


nonlinear ?



Noisy, linear correlations small:



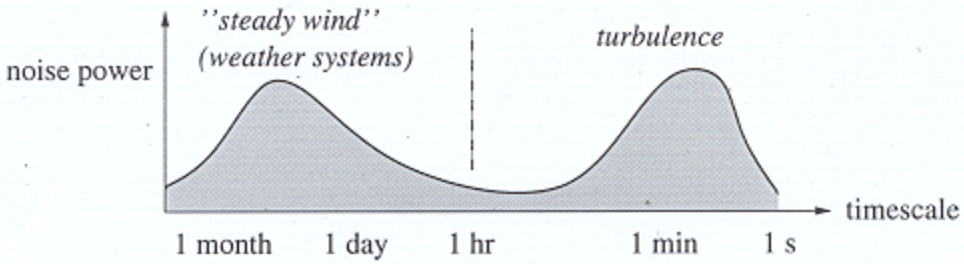


bimodal

Mountain pass !

Timescales & historical data

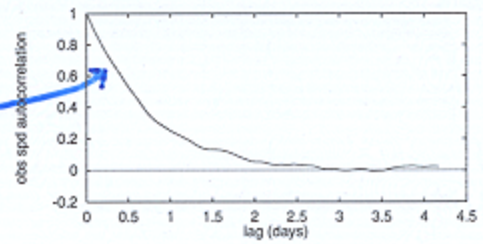
Wind fluctuations (in $\sim$ 1 km boundary layer) 2 timescales...



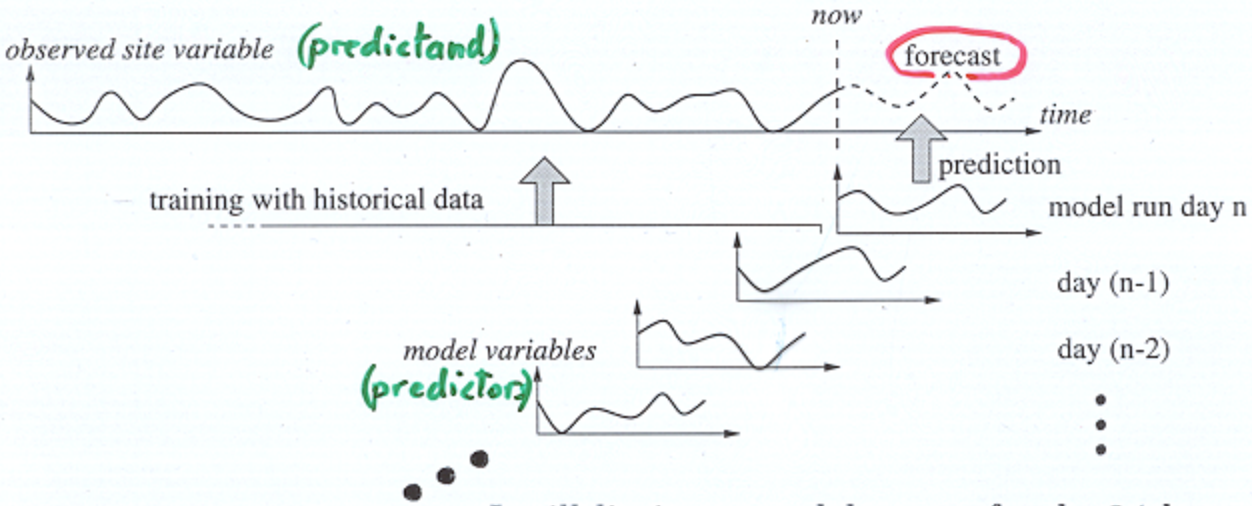
- care only about "steady wind"
- observed (u, v), generated power : 1-hour averages
- model variables (u_0, v_0), etc : hourly samples

industry timescale

Autocorrelation dies $\sim$ 1 day:
(random walk at high frequencies)



Learning transfer function from history:



I will limit to model runs of only 24 hrs

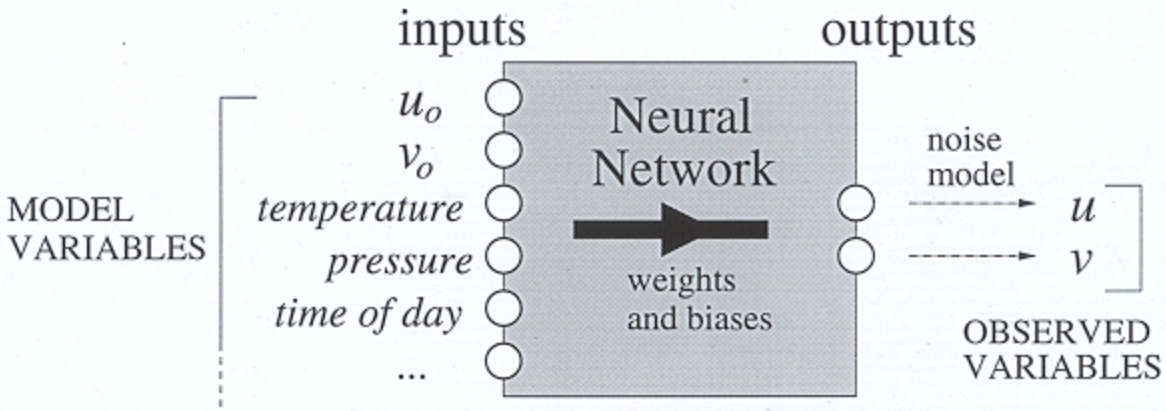
1 m/s = 1 hr to cross grid cell $\Rightarrow$ suspect system has short 'memory'

$\Rightarrow$ Start with memory-less transfer function

Bayesian Neural Networks

MacKay used BNNs:

- forecast building's energy use given 4 other variables
- won prediction competition (1993)



Architectures: compared...

- no hidden layers (≡ linear regression)
- 1 hidden layer of 8 units with ARD (Neal & Mackay) nonlinear

Noise model: found t-distribution (long tails) beats gaussian.

Bayes: $\begin{cases} \text{Regularization (weight decay)} \rightarrow \textit{prior} \text{ on weights} \\ \text{Training} \rightarrow \text{finding } \textit{posterior} \text{ given training data} \end{cases}$

Sampling the posterior with MCMC (Neal's `fbm` software).
→ predictive distribution

Measuring success:

- compare MAE (*mean absolute error*) – for NN predictions
- Assessments:

  a) split time series *randomly* into training & test sets

  b) simulated forecasting using only data from *past*

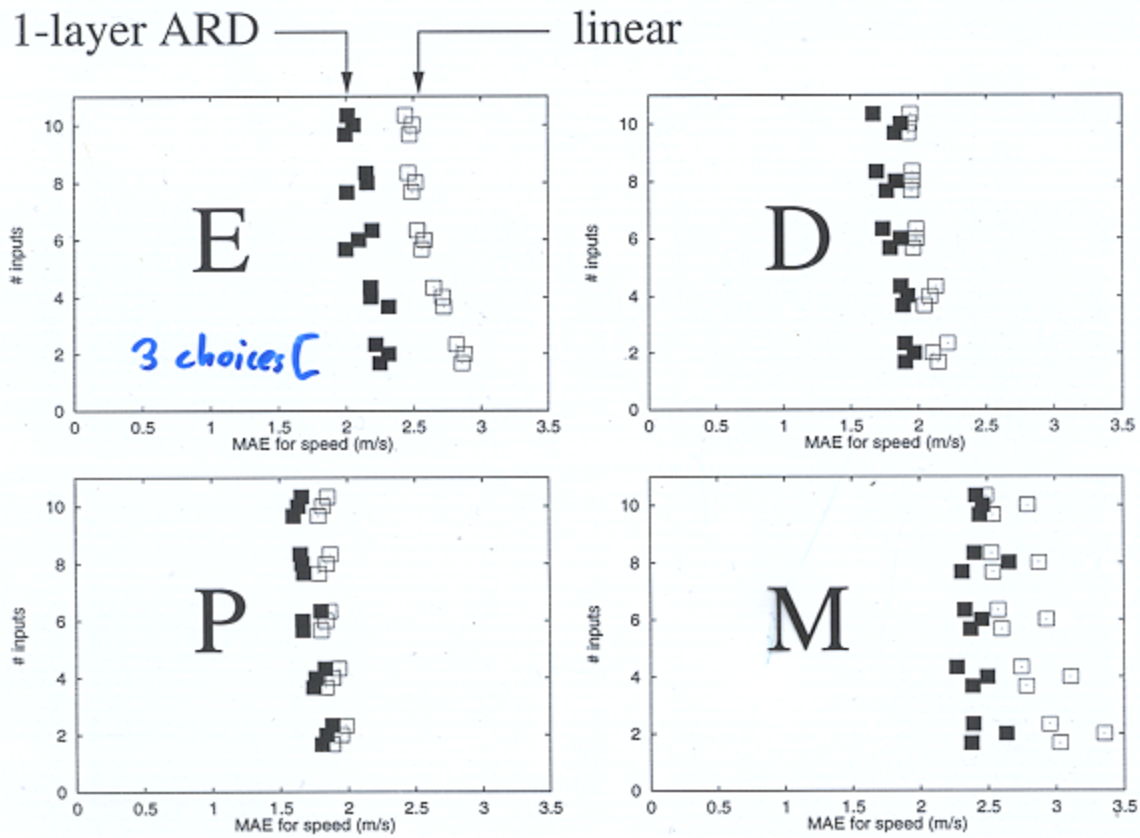
a) Learning wind velocity. Results I

Random split assessment (no regard to time order)

- $\sim$ 2000 cases: 1/4 training, 3/4 test
- try different split choices—judge performance fluctuations (*not* quantifiable statistical significance)
- medium dataset size dominated by rare events $\rightarrow$ cannot discern small changes in performance.

Not enough computer time to use $\sim$ 100 variables as inputs...

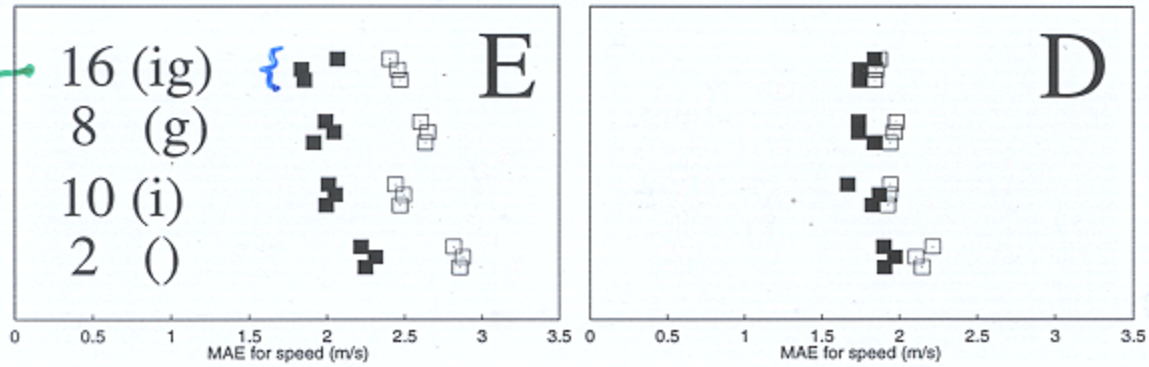
Increasing # inputs: (4 sites E,D,P,M). ⇒ pick 'by hand'



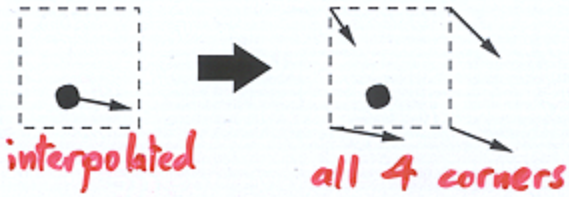
- site variation (E-D close, P-M close, < 1 grid cell)
- M problem : nonlinear advantage lost as input # ↑

a) Learning wind velocity. Results II

Expanding input space: (u_o, v_o) from **4** grid points...



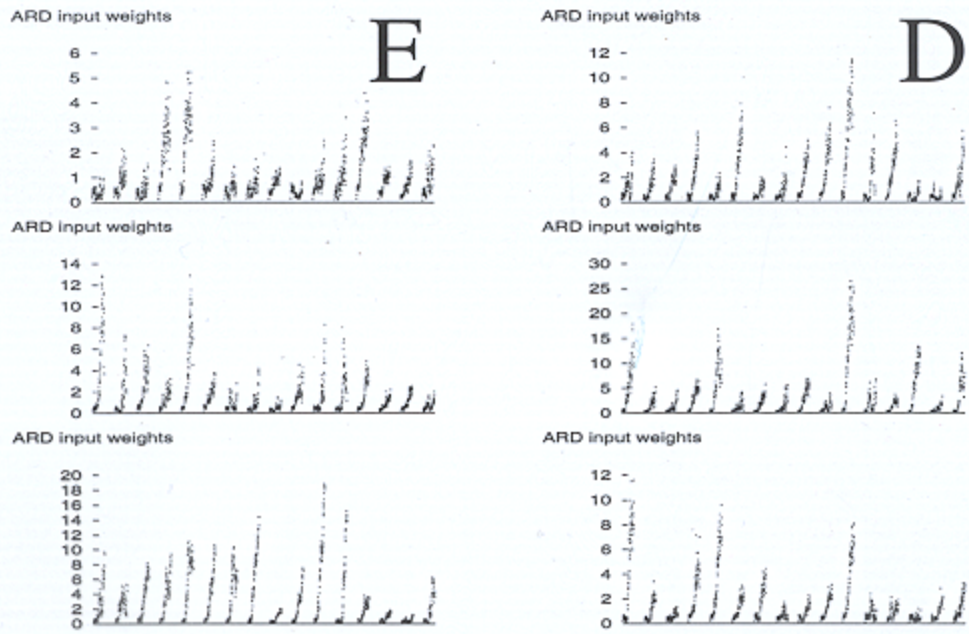
i = include 8 extra inputs
g = use u_o, v_o at 4 grid points:



- i → ig: no significant improvement

ARD—little consistency in which inputs relevant:

split choice



(weights often continue to grow during MCMC). $\leftarrow v_{70}$

posterior distributions of weights grouped by input

b) Forecasting wind velocity results

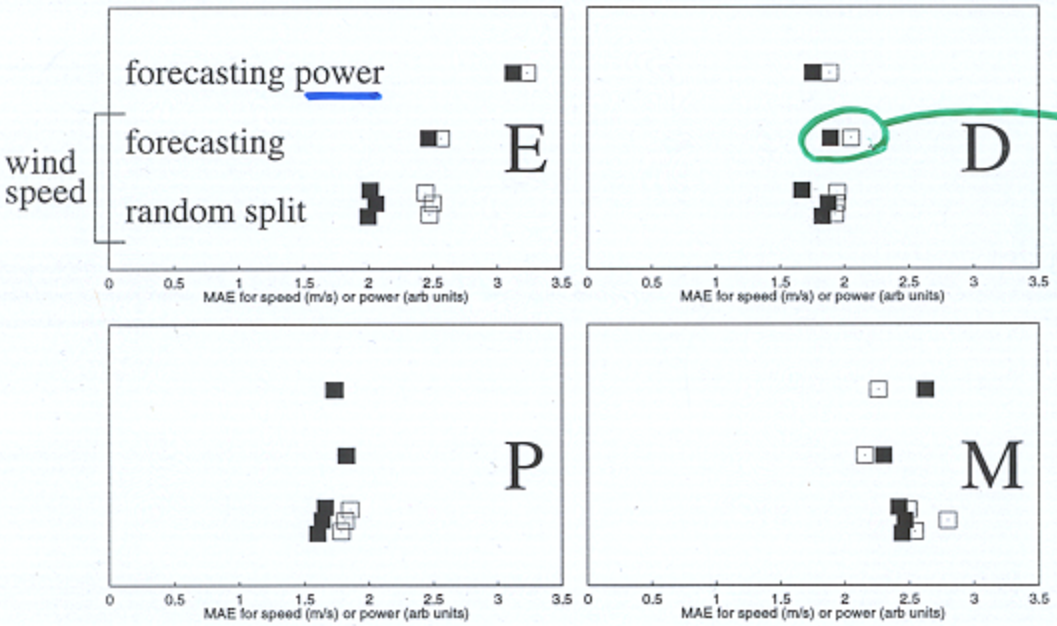
Test 9 consecutive weeks, each using only *past* cases to train.

How far back in time for training?

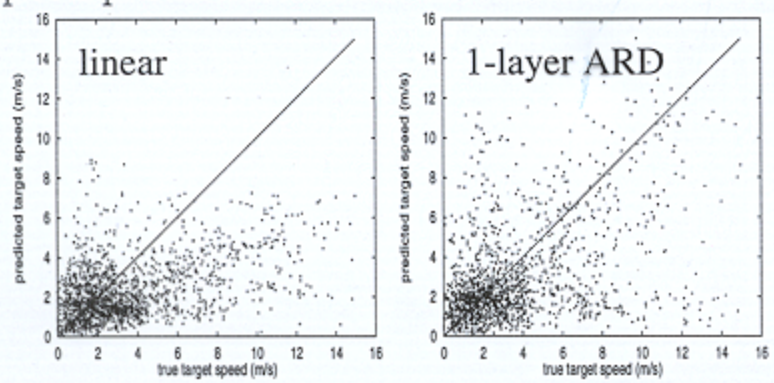
- long enough for many cases, capture *rare events*.
- short enough to react to seasonal changes.
- performance decreased below 30 days (≈ used for weather).

+ TrueWind

Results for 10 inputs:



Examine speed prediction errors for site D:

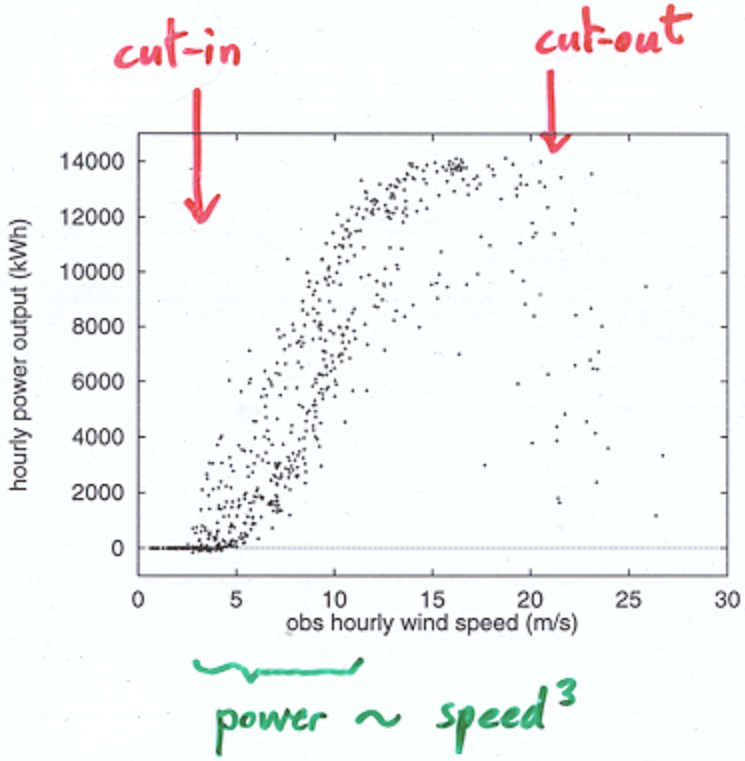


D

most high-speeds predicted much better!

Power prediction would differ due to non-linear relation...

The wind-to-power relation: 'power curve'



Scatter due to

- hourly averaging of turbulence via nonlinear relation
- variable turbine performance (+ rapid direction changes)
- (u, v) often not measured exactly at turbine

Nonlinear shape affects predictive importance:

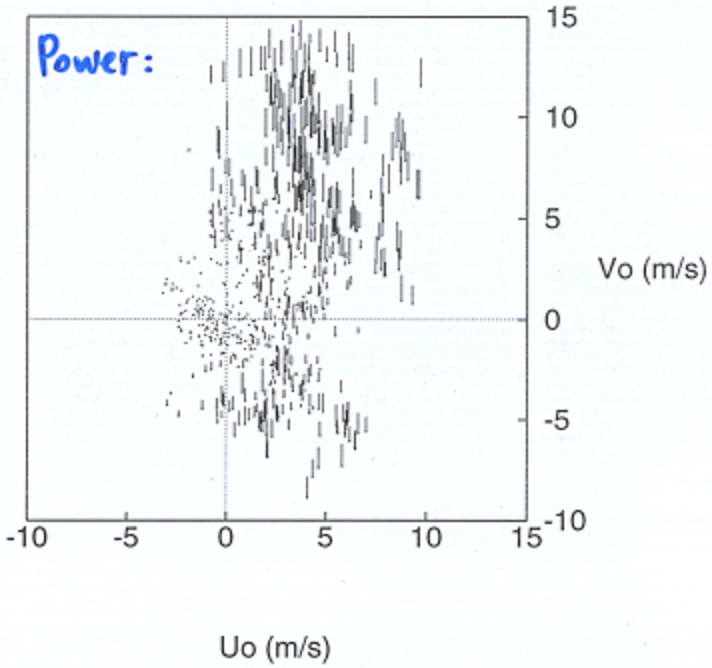
- 'under 4 m/s' and 'over 15 m/s' are binary categories
- speed accuracy important only in intermediate region

.... this type of curve used on last slide

Learning power output directly—problems

PDF of output power is what utilities care about
$\Rightarrow$ predict power directly from model variables?

Patchy coverage of input space.



Reparam. model winds
$(u, v) \rightarrow$ (speed,cos(θ),sin(θ))

- Reflects radial-symm 'bowl'
- gave large improvement

Results: 1-layer net has signatures of *overfitting*

- Increasing # inputs $\rightarrow$ worse performance
- TrueWind's hand-tuned 9 inputs: gives worse than linear fitting
- Later MCMC samples generalize worse

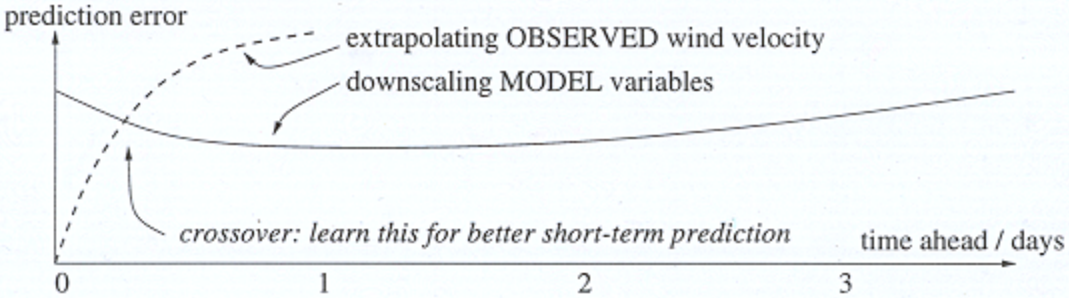
Noise varies drastically as function of wind speed.

- bad noise model can cause overfitting
- remove small-noise region $\rightarrow$ but still overfits
  (speed < 4 m/s)

cf near-optimal performance on dummy data (ideal, uncorrelated, constant noise).

Future directions

- Try more available inputs.
  - Time-filtered versions of (some) inputs: learning transfer function with memory (*e.g.* MacKay 1993)
  - An optimal method would use both model variables and *recently observed* windspeed as inputs:



  - Automated search (feature selection) from $> 10^2$ inputs.
- Better use of historical training data (including climateology) to focus on learning rare events (high speeds).
- Other training methods (GDES), architectures.
- Dominant noise is on *inputs* $\rightarrow$ better noise model?
- Tests with large machine-generated time series, add real-world features:
  - Independent training sets $\rightarrow$ rigorous measures of statistical significance.
  - uncover causes of MCMC overfitting.
  - ways to handle non-constant noise.

Conclusions

1. Neural nets for wind velocity forecasting:

   - good random-split results did not carry over to forecasting.
   - small improvement ($\sim 5\%$) over linear fitting same inputs.
   - depends heavily on site $\rightarrow$ not reliable.
   - now testing by TrueWind (not yet used).
   - maybe there is little nonlinear correlation?

2. Direct neural net learning of turbine output power:

   - no better than hand-tweaked linear methods.
   - plagued by overfitting $\leftarrow$ non-constant noise?

3. Need:

   - lots of time + data to discern relative performance.
   - automated ways to search/reduce large input space.

4. 'Virgin territory': very little statistics/AI expertees in wind power... many directions yet to be tried!