Understanding the performance of neural network models for short-term predictions applied to geomagnetic indices

Swedish Institute of Space Physics
Peter Wintroft and Magnus Wik
The Origin of Magnetic Storms.

By Arthur Schuster, F.R.S.

(Received January 9,—Read January 26, 1911.)

1. Lord Kelvin,* in discussing the origin of magnetic storms, came to the conclusion that they could not be due to a direct solar action on account of the enormous energy which would have to be supplied by the sun. This verdict was generally accepted until recently, when the theory of a direct solar action has been revived in a form, which is assumed to be free from the objection raised, the magnetic action being supposed to be due to a swarm of electrified corpuscles ejected by the sun. The question of energy has not, so far as I know, been discussed in this case, and it seems to be taken for granted that the total energy of the magnetic field due to a swarm of corpuscles is equal to the sum of the energies of each, calculated as if the others were not present. If the corpuscles are sufficiently far apart, this is approximately correct; but in that case the magnetic field itself would have to be negligible, except within molecular distance from each particle. How far we may go wrong by treating the energy as if it could be obtained by a process of addition becomes apparent when we consider that such treatment would lead to coefficients of self-induction which are proportional to the length of a circuit and independent of its shape.
• 30 model settings: 3D MHD models; kinetic models; specification models.

• 4 storms: Aug 2001 (-40 nT), Oct 2003 (-353 nT), Aug 2005 (-131 nT), Dec 2006 (-139 nT)

Neural networks

• Data …

• Data gaps (esp. hard for time-series).

• Division of datasets used for
  • training,
  • validation, and
  • testing.

• Training set => parameter estimation

• Validation set => hyper-parameter search

• Test set => final estimate of performance

A lot of other considerations:
  Learning algorithm
  Input parameters
  Transformation of inputs
  Network architecture
“Multilayer feedforward networks are universal approximators” (Hornik et al., 1989).
“Multilayer feedforward networks are universal approximators” (Hornik et al., 1989).
“Multilayer feedforward networks are universal approximators” (Hornik et al., 1989).
"Multilayer feedforward networks are universal approximators" (Hornik et al., 1989).
Multilayer feedforward networks are universal approximators" (Hornik et al., 1989).
Multilayer feedforward networks are universal approximators” (Hornik et al., 1989).
Dynamic NNs

• Time delays
• Recurrent connections
• More advanced: LSTM

\[ Dst^*(t + 1) = Dst^*(t) + [Q(t) - \lambda(t)Dst^*(t)] \]


Recurrent Neural Networks are universal approximators, Schäfer and Zimmermann, Int J Neural Syst. 2007 Aug;17(4):253-63.

3-hour averages

\[ Kp(t+3) = f(Bz(t), Bz(t-3), ..., n(t), n(t-3), ..., V(t), V(t-3), ...) \]
Derivation of Kp

- Sensitivity to sub-3-hour variations.
- High-pass filtered storm dynamics

Derivation, meaning, and use of geomagnetic indices, P. N. Mayaud, AGU, 22, 1980.
Performance in low-density state-space

Points with Kp>6 from training set

Trace of a storm from the test set

Observed Kp
Median of pred.

Solar wind

NN

NN

Median

Kp

Ensemble of networks

Increasing Kp prediction lead time?
BIAS = 0.54
RMSE = 0.83
CORR = 0.92
R2 = 0.71

http://lund.irf.se/forecast/kp2017/

Will be included at ESA SSA ESC-G
Operational forecasts of the geomagnetic Dst index, H. Lundstedt and H. Gleisner and P. Wintoft
Range of Dst from NN model

- Outputs from final hidden layer are limited to [-1, +1].
- Sum |weights| + bias => possible range of Dst.

[-650, 190] nT

75%: [-500, 140] nT

\[ R_{\text{quick}} = 0.4 \mu \omega^{1/2} m_p^{1/2} \sin^2(\theta/2) C^{-1/2} n_{sw}^{1/2} v_{sw}^2 (1 + \beta_s)^{-3/4} \]
5,10,1 (71 weights)

5,10,10,1 (181 weights)

C = 0.989
R^2 = 0.978

C = 0.997
R^2 = 0.995
NN hidden output

5 inputs, 10 hidden, 1 hidden, 4 recurrent, 1 output

C = 0.90
Node 1

Node 2

Rquick

sqrt(Pressure)
Some events belong to the “training set”!!

**Figure 5.** Relationship between $\Delta Dst_{\text{min}}/|Dst_{\text{min}}|$ and $|Dst_{\text{min}}|$ for the TL model. Here the point symbols represent the four interplanetary structures. Super storms ($Dst \leq -200$ nT) are divided by the dotted vertical line.


Figure 9. Correlation between peak Dst and interplanetary parameters: peak Bs, Ey, By+, By−, Bp, Pdyn, Np, and Vsw.
• Storms identified using MODWT filtering.
• Determine storm minimum.


Extreme value analysis of the time derivative of the horizontal magnetic field and computed electric field, P. Wintoft and A. Viljanen and M. Wik, Annales Geophysicae 34 485--491 (2016)

Related

- Support Vector Machines:
  - Both classification and regression
  - Global minima
  - Support vectors drawn from training vectors
  - Nonlinearity through transformation to higher-dimensional space, but the space need not to be known only that there exist a dot-product => kernel trick (e.g. RBF corresponds to an infinite-dimensional space)

- Nonlinear Auto-Regressive Moving Average with eXogenous inputs (NARMAX):
  - A “library” of terms (e.g. polynomials or other functions) are ranked and combined including time lags.
  - Will reveal a function that is more assessable for interpretation than SVM or NN.
Summary

• The selection of training, validation, and test data is important but can also be challenging:

  • Data gaps (esp. hard for time-series).
  
  • Division of data based on output distribution, but what about multidimensional inputs?
  
• Both NN and SVM will be limited by the range of the training data.

  • Use them in their valid regimes.
  
  • Is extrapolation improved by handling known non-linearities?
  
• Should simpler models be used at the extremes?

  • How to determine extreme?