Methodological Challenges for PTFs in Earth System Models

Attila Nemes

NIBIO, Norwegian Institute of Bioeconomy Research, Ås, Norway

A very frequently seen pattern in submissions:

- ...we have collected data of 67 samples from an experimental site...

- ...we developed MLR and ANN (*just an example*) PTFs...

- ...we used R^2, RMSR to test the performance...

- ...ANNs are better than MLR...and...

- ...RMSR is 0.04327928966, which is better than...

- ...our model outperforms HYPRES and ROSETTA...

- ...our findings confirm the findings of...

Assessment and advice by…

Laj Ahuja	Attila Nemes
Niels Batjes	Yakov Pachepsky
Ary Bruand	David Radcliffe
Johan Bouma	Nunzio Romano
Wim Cornelis	Marcel Schaap
Daniel Gimenez	Navin Twarakavi
Gerrit Hoogenboom	Rien van Genuchten
Allan Lilly	Harry Vereecken
Alex McBratney	Larry West
Budiman Minasny	Henk Wösten

- Insufficient amount/quality of data (89%)

- Lack of novelty (67%)

- Lack of innovation (67%)

- Lack of depth in the analysis (67%)

- Inappropriate/insufficient use of statistical evaluation (67%)

- Presentation/Discussion too specific (cannot generalize, i.e. not useful to most readers) (56%)

- Lack of discussion (33%)

- … (country of origin, language, manuscript style – 0%)

- Do you think PTF studies still have a future in main-stream journals and conferences? YES (100%)
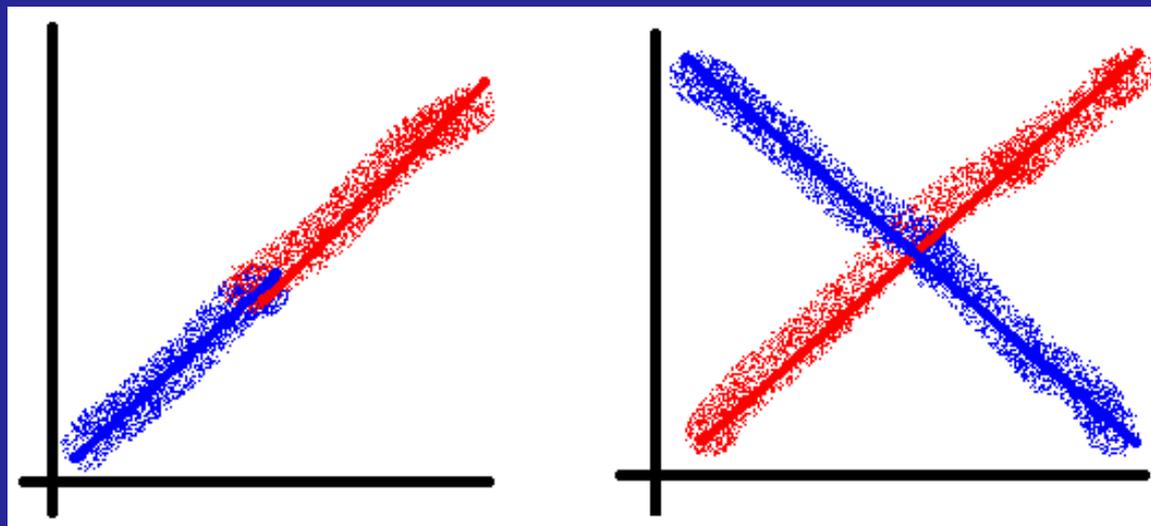
Challenge: Extrapolation…

Risk: PTF is applied beyond its *area of development*

What does this mean really?

Despite their areas being generally covered, knowledge gaps exist for specific soil types (e.g. saline, Andosols, peats, shrink-swell, etc.)

…data in different range can show similar correlations
…data in the same range can show different correlations

There are so many ways the data may not be compatible!

Evaluate the reasons, communicate them, and be critical!
- derived vs. original data (method of aggregation in e.g. SSURGO?)
- particle size by laser vs. pipette/hydrometer/PARIO (PTF for this now…)
- BD measurement protocol (what moisture? core vs. clod?)

Risk: bias in the estimations

Is it looked at sufficiently? Reported sufficiently?

Is it possible to control it?

Communication of metadata is of absolute necessity!
- Field and lab methods
- Management, landuse, landcover
- Sampling time, method, circumstances…

How can you tell if the data domains are similar or different?

Before…
- Basic diagnostics (range, correlation structure, identifying patterns…)
- Understand the significance of differences and look for alternatives

During…
- Characterize differences…(derive or intercept difference metrics (k-NN)
- Apply penalty for a sample being outside the domain (fuzzy k-means)
- Use it as flag or as a propagated measure of uncertainty? (especially if combined with ensemble estimations)

After…
- Conduct detailed examination of error distribution, variance of estimates, etc.

Challenge: PTF Uncertainty…

For real applications, it is absolutely desirable, and growingly possible.

- Should be propagated into subsequent estimations and applications
- It can amount to be a very significant factor

How?

- Using an ensemble of (already published) PTFs (with different structure, theory, data, inputs – But how to best weigh them?)
- Resampling (even this can yield a different model structure (regr.tree)
- Inference system (minimize propagated variance 'downstream' in subsequent estimations)
- ?

Challenge: Model integration…

- Which single model to use? Often hard to justify!

- Ensembles can be useful – common trends are amplified, others are suppressed

- Their output can be interpreted into a measure of uncertainty (hurricane forecast cone?)

- Little concerned about the co-application of different equations to get e.g. MvG curves, for example. (If the estimated parameters or processes are expected to be correlated, their independent, sequential estimation may not be optimal.) (see later…)

- Functional evaluation?

Functional evaluation!

- PTFs are not the ultimate goal by themselves

- Differences in an application may be suppressed by other factors in the application/model – or the environmental conditions that are simulated.

Theta_residual   *Norway*

Theta_saturated   *Saudi Arabia*

Functional evaluation…contd.

- What are they applied for?

- They can cause surprises, so get to know them!

- What if you are using them to simulate the effect of soil compaction, but you don't have BD properly represented. How will the wet part of WRC and k be represented?

- What if you were to evaluate the impact of increased OM, as a result of more efficient sequestration, on Ks for example?

- What if the interpretation of the results is odd? (Norway: FC based criterion for workability.)

What are they used for?

- Research vs. application?

- Not all PTF methods are equal (*except their RMSE…*)

- Continuous vs. categorical (class)
- Ease of application….
- etc…

	Class PTFs	MLR, GLM, GAM	(Boosted) CART	CHAID	GMDH	(Fuzzy) k-Means	ANN	k-NN	SVM
Predictive Power	red	orange/red	yellow	yellow	green-yellow	yellow	green	green	green
Transparency (of algorithm)	green	red/green	green	green	red	orange	red	green	red
Ease of Use	green	yellow/green	green	green	yellow	orange	red	yellow/green	green
Interpretability (of results)	green	red/green	green	green	red	green	red	green	red
Transferability (of model)	green	red/green	green	green	yellow	yellow	red	red	red
Mixed Inputs (qual./quan.)	red	yellow/green	green	green	red	orange/green	red	green	orange
Uncertainty estimate	red	red/green	green	yellow/green	red	green-yellow	yellow	green	yellow
Data grouping	red	yellow	yellow	yellow	red	red	red	red	red
Missing Data	red	red	orange/green	green	red	red	red	orange/green	red
Non-linearity	red	red/green	green	green	green	green	green	green	green
Efficiency on Large Datasets	green	green-yellow	green	green	green	green	yellow	green	green
Variable selection	red	green	green	green	green	red	red	red	red

Partially adapted from:

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer Series in Statistics. Springer-Verlag, NY.

Minasny, B. 2007. Predicting Soil Properties. Jurnal Ilmu Tanah dan Lingkungan 7(1): 54-67

Challenge: Upscaling…

- For mapping purposes, an upscaling dilemma is at what level to aggregate? Aggregate predictors at sub-unit level and estimate for their average or estimate at sub-unit level and then average?

    (SSURGO: USDA-NRCS does not aggregate, but provide aggregation tools for the user to decide.)

- Simultaneous vs. sequential estimation of potentially correlated parameters or processes?

    Pringle et al (2007): correlation and variance of observed and predicted variables should be checked, and error patterns should be examined in both spatial and non-spatial contexts

- PTFs are typically derived from point-scale data, with no spatial relationships, or correlations assumed in any direction (X,Y,Z).

- Be aware that PTFs are smoothing tools – generally underestimate variance in the predicted data (whether spatial or non-spatial)

Challenge: Upscaling… contd.

- What if available data types are limiting us from progress?

    What if your PTF at smaller scale does not include variables that are important at the larger scales of target?

    Necessary input is scale dependent (e.g. topography at larger scales, climate, land use/cover…   pore structure at small scales…)

- Experimentation with hypothetically important factors is necessary at every scale.

    - Soil structure in general

    -Hyperspectral information

    - X-Ray Tomography

    - Other sensing techniques…

Challenge: Temporal variability…

- A number of factors can be influential on 'effective' soil properties at shorter or longer temporal scales   (e.g. tillage, land-use and land management, freezing-thawing, crack development…)

- Need for metadata
- Need for greater variability in collected data to help reflect those conditions (if for nothing else, but validate theoretical approaches)

Earlier today…

Questions on Rosetta 3:

- Why not include OM? strong correlation of Ks and OM?
- Tropical soils are underrepresented
- Split the model into 2 portions – bigger database for WRC than K?
- Where is it valid and where is it not (area/purpose)?

Thank You!

attila.nemes@nibio.no