Neural Network Approach
to the Inversion of
High Spectral Resolution Observations
for Temperature, Water Vapor and Ozone

- DIFA-University of Basilicata (C. Serio and A. Carissimo)
- IMAA-CNR (G. Masiello, M. Viggiano, V. Cuomo)
- IAC-CNR (I. DeFeis)
- DET-University of Florence (A. Luchetta)

Thanks to EUMETSAT (P. Schlüssel)
Outline

• Neural Network Methodology
• Neural Net vs. EOF Regression (IASI Retrieval exercise)
• Physical inversion methodology
• Physical inversion vs. EOF Regression (a few retrieval examples from IMG)
The Physical forward/inverse scheme for IASI

• \( \varphi \)-IASI is a software package intended for
  • Generation of IASI synthetic spectra
  • Inversion for geophysical parameters:
    1. temperature profile,
    2. water vapour profile,
    3. low vertical resolution profiles of \( \text{O}_3 \), \( \text{CO} \), \( \text{CH}_4 \), \( \text{N}_2\text{O} \).
The $\phi$-IASI family

- $\sigma$-IASI: forward model (available in Fortran with user’s guide)
- $\delta$-IASI: physical inverse scheme (available in Fortran with user’s guide)
- $\nu^2$-IASI: neural network inversion scheme (available in C++ with user’s guide)
- $\varepsilon$-IASI: EOF based regression scheme (available in MATLAB, user’s guide in progress).
Theoretical basis of Neural Network

K. Hornik, M. Stinchcombe, and H. White,
Multilayer feedforward networks are universal approximators
Neural Networks, 2, 359-366, 1989
Basic Mathematical Structure of a generic i-th Neuron

\[ o_i = \psi(y) = \frac{2}{1 + \exp(-\lambda y)} - 1; \ \text{with} \ y = w_i^t x = w_{i1} x_1 + \cdots + w_{iN} x_N \]
Cost Function to determine the weights

\[ E = \frac{1}{2} (d_i - o_i)^2 = \frac{1}{2} (d_i - \psi(y))^2 \]
Multilayer feed-forward Architecture
Simultaneous Architecture for (T, H₂O, O₃) retrieval
Inter-comparison exercise: NN vs. EOF regression

**NEURAL NET**
- Input projected into an EOF basis: 50 PCs retained (Optimized)
- Output projected into EOF basis: 15 PC for T, 10 for H₂O, 15 for O₃

**EOF regression**
- Input projected into an EOF basis: 200 PCs retained (Optimized)
- Output projected into EOF basis: 15 PC for T, 10 for H₂O, 15 for O₃
Rule of the comparison

• Compare the two schemes on a common basis:
  – same a-priori information (training data-set),
  – same inversion strategy: simultaneous
  – same quality of the observations
Definition of the spectral ranges
**Training/Validation** and **Test** Data Sets

<table>
<thead>
<tr>
<th>Air Mass Type</th>
<th>Training/Validation data set</th>
<th>Test data set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tropical</td>
<td>595</td>
<td>221</td>
</tr>
<tr>
<td>Mid-Latitude Summer</td>
<td>305</td>
<td>43</td>
</tr>
<tr>
<td>Mid-Latitude Winter</td>
<td>388</td>
<td>155</td>
</tr>
<tr>
<td>High Latitude Summer</td>
<td>283</td>
<td>80</td>
</tr>
<tr>
<td>High Latitude Winter</td>
<td>740</td>
<td>164</td>
</tr>
</tbody>
</table>

2311 TIGR profiles (TIGR-3 data base)
Tropical Air Mass (RIE test data set)
Tropical Air Mass
Tropical Air Mass
Summary

- NN performs better than EOF Regression provided that they are evaluated on a common basis.
- NN is parsimonious with respect to EOF regression (50 PCs vs. 200 PCs)
Optimization

Eof Regression

Localize training

(Tropics, Mid-Latitude, and so on)
Tropical Air Mass
Tropical Air Mass
Tropical Air Mass
Summary

- EOF regression improves when properly localized
- Neural Net is expected to improve, as well. Results are not yet ready.
Dependency on the Training data set

- One major concern with both the schemes is their critical dependence on the training data set
Comparing N.N. to EOF Reg.
IMG Mid-Latitude Observation
Comparing N.N. to EOF Reg.
IMG Tropical Observation
How to get rid of data-set dependency?
Physical Retrieval
(Initialized by Climatology)

EOF Regression

Introducing Physical Inversion

Circles: ecmwf;
Line: retrieval
Physical Retrieval
(Initialized by Climatology)

EOF Regression

Introducing Physical Inversion
Statistical Regularization

Tikhonov/Twomey Regularization

Data constrained Optimisation
The subscript $g$ stands for a suitable background atmospheric state!

\[
(\tilde{v} - v_g)^t L (\tilde{v} - v_g) \quad \text{MIN!}
\]

\[
(y - Kx)^t S^{-1} (y - Kx) \leq \chi^2_{\alpha}
\]

$S$ : Obs. Cov. Matrix

$L$ : Smoothing Operator
About $L$

- Twomey’s approach

\[ L \equiv \int_0^+ \left| \frac{d^n x}{dh^n} \right|^2 dh; \]

- Rogers’ approach

- $L$ is intrinsically a covariance operator, $B$ in his notation

- $n=0, 1, 2$
Physical Consistency of the L-norm

\[(\widetilde{v} - v_g)^t L(\widetilde{v} - v_g)\]

Twomey’s L is lacking dimensional consistency! it attempts to Sum unlike quantity, e.g, (K+g/kg)

Rodgers’ L ensures dimensional consistency

\[L = B^{-1}\]

which makes the norm above dimensionless
Our Choice

• Since we are interested in simultaneous inversion which involves unlike quantities such as Temperature, water vapour concentration and so on we choose
  • $L = B^{-1}$

• However, the methodology we are going to discuss still hold for any Twomey’s $L$
Finding the solution through Lagrange multiplier method

\[ \hat{x} = \hat{v} - v_g = (\gamma B^{-1} + K'S^{-1}K)^{-1} K'S^{-1}y \]

\( \gamma = 1 \) gives the usual Statistical Regularization;

\[ S_y = (\gamma B^{-1} + K'S^{-1}K)^{-1} (\gamma^2 B^{-1} + K'S^{-1}K)(\gamma B^{-1} + K'S^{-1}K)^{-1}; \]

\[
\begin{cases}
S_y = (B^{-1} + K'S^{-1}K)^{-1}; & \text{for } \gamma = 1 \\
S_y = (K'S^{-1}K)^{-1}; & \text{for } \gamma \to 0 \\
S_y = B; & \text{for } \gamma \to \infty
\end{cases}
\]
Uncovering the elemental constituent of regularization

\[ B = B^{2}B^{2} \]

\[ (\gamma B^{-\frac{1}{2}}B^{-\frac{1}{2}} + J'J)\widehat{x} = J'z; \text{ with } J = S^{-\frac{1}{2}}K, \quad z = S^{-\frac{1}{2}}y; \]

\[ B^{-\frac{1}{2}}(\mathcal{I} + B^{-\frac{1}{2}}J'JB^{-\frac{1}{2}})B^{-\frac{1}{2}}\widehat{x} = J'z; \]

\[ G = JB^{2} \quad \text{and} \quad \widehat{u} = B^{-\frac{1}{2}}\widehat{x}; \]

\[ (\mathcal{I} + G'G)\widehat{u} = G'z \quad \text{Ridge Regression} \]
Continued

- The same decomposition may obtained for Twomey regularization by putting
  - $L = M^t M$
- $M$ may be obtained by Cholesky decomposition for any symmetric full rank matrix $L$
- Twomey’s $L$ is typically singular. Nevertheless the above decomposition may be obtained by resorting to GSVD (Hansen, *SIAM Review, Vol. 34*, pp. 561, (1992))
Summary

• The RIDGE regression is the paradigm of any regularization method,

• The difference between the various methods is:
  ➢ the way they normalize the Jacobian
  ➢ the value they assign to the Lagrange multiplier
- **Levenberg-Marquardt:**
  - $\gamma$ is assigned alternatively a small or a large value, the Jacobian is not normalized, that is $L=I$.
- **Thikonov**
  - $\gamma$ is a free-parameter (chosen by trial and error), the Jacobian is normalized through a mathematical operator.
- **Rodgers:**
  - $\gamma=1$, the Jacobian is normalized to the a-priori covariance matrix. It is the method which enables dimensional consistency.
- **Our Approach**
  - Rodgers approach combined with an optimal choice of the $\gamma$ parameter ($L$-curve criterion).
A simple numerical exercise
Statistical Regularization
1st Iteration

Retrieved and test temperature profiles

Retrieved and test water vapour profiles
Statistical Regularization
2nd Iteration

Retrieved and test temperature profiles

Retrieved and test water vapour profiles
L-Curve, 1st Iteration

Retrieved and test temperature profiles

Retrieved and test water vapour profiles
L-Curve, 2nd Iteration

Retrieved and test temperature profiles

Retrieved and test water vapour profiles
Convergence example based on an IMG real spectrum

- Inversion strategy:
  - 667 to 830 cm\(^{-1}\) simultaneous for (T, H\(_2\)O)
  - 1100 to 1600 cm\(^{-1}\) (super channels) sequential for H\(_2\)O alone
  - 1000 to 1080 cm\(^{-1}\) sequential for Ozone
Temperature
Water Vapor
Ozone

Retrieved Columnar Ozone: 237 Dobson
Toms/Adeos: 236 Dobson
\( \chi^2 \)-constraint
Physical Inversion vs. EOF regression: Exercise based on Real Observations (IMG)

- **EOF regression:**
- Training data set: a set of profiles from ECMWF analyses
Physical Inversion
Based on Climatology

EOF Regression

Circles: ECMWF Analysis
Exercise for tomorrow

• Inter-compare
  – Physical Inversion
  – EOF Regression
  – Neural Net

• With NAST-I data (work supported by EUMETSAT)

• With our AERI-like BOMEM FTS (work supported by Italian Ministry for the research)
Research program to speed up physical inversion (next future)
Develop the RTE in EOF-basis

\[ r = K_T x_T + K_w x_w + \text{h.o.t.}; \]

\[ r = R - R_o; \quad x_T = T - T_o; \quad x_w = w - w_o \]

\[ E = (R_1, R_2, ..., R_M); \quad C = \frac{1}{M} E^T E; \quad S_R = \text{diag } (C); \]

\[ S_R^{-\frac{1}{2}} r = S_R^{-\frac{1}{2}} K_T S_T \frac{1}{2} S_T^{-\frac{1}{2}} x_T + S_R^{-\frac{1}{2}} K_w S_w \frac{1}{2} S_w^{-\frac{1}{2}} x_w \]

\[
\begin{align*}
    y &= S_R^{-\frac{1}{2}} r \\
    t &= S_T^{-\frac{1}{2}} x_T \\
    z &= S_w^{-\frac{1}{2}} x_w \\
    A_T &= S_R^{-\frac{1}{2}} K_T S_T \frac{1}{2}; \quad A_w = S_R^{-\frac{1}{2}} K_w S_w \frac{1}{2}
\end{align*}
\]

\[ y = A_T t + A_w z \]
EOF decomposition of the linearized RTE

$$
\begin{align*}
\mathbf{U}_R ; \quad & \mathbf{U}_T ; \quad \mathbf{U}_w \\
\{ c_y & = \mathbf{U}_R^T \mathbf{y} \\
\{ c_t & = \mathbf{U}_T^T \mathbf{t} \\
\{ c_w & = \mathbf{U}_w^T \mathbf{z} \\
\mathbf{U}_R^T \mathbf{y} & = \mathbf{U}_R^T \mathbf{A}_T \mathbf{U}_T \mathbf{U}_T^T \mathbf{t} + \mathbf{U}_R^T \mathbf{A}_w \mathbf{U}_w \mathbf{U}_w^T \mathbf{w} \\
\end{align*}
$$

\[ c_y = \mathbf{G}_T c_t + \mathbf{G}_w c_w \]

\[ \begin{align*}
\mathbf{G}_T &= \mathbf{U}_R^T \mathbf{A}_T \mathbf{U}_T \\
\mathbf{G}_w &= \mathbf{U}_R^T \mathbf{A}_w \mathbf{U}_w 
\end{align*} \]
Conclusions

- The inversion tools developed within the ISSWG activities by the DIFA-IMAA-IAC groups have been presented.
- A comparison has been provided of the relative performance of the various methods (although more work is needed).
- A tentative list for now sees at the top the:
  1. Physical inversion (not suitable for operational end-uses)
  2. Neural Network (very fast, still complex to train, its dependence on the training data set has to be assessed)
  3. EOF Regression (appealing for its simplicity, the training needs to be localized, does not seem to provide reliable results for $H_2O$)