PWLR⁺ Exploiting horizontal correlation in PieceWise Linear Regression

TIM HULTBERG, THOMAS AUGUST, MARC CRAPEAU
PWLR\(^3\) (Pea-vee-el-are cube)

All sky retrieval using IASI, AMSU and MHS

- Trained with real measurements co-located with ECMWF analysis
- **PieceWise Linear Regression**
- **High performance retrieval in its own right and a-priori for the optimal estimation in clear sky**
- To replace the PWLR in EUMETSAT’s IASI L2 processor version 6.2 (end of May 2016)
- **IASI Level 2 Regional Service: high timeliness dissemination on EUMETCAST (hdf5 format)**
- Temperature, water vapour and ozone profiles, surface pressure, skin temperature and emissivity as well as cloud-signal and error estimates
- **Cube:** single field of view retrievals using data from four fields of view (exploiting horizontal correlation)
Why regression?

Rodgers page 113:
• “is identical to the maximum a posteriori solution”
• “implementing the MAP solution without having to characterise the instrument with great care and without having to implement an accurate forward model”

“features can be retrieved which comes from historical correlations and not from the measurements”

But only if they are in the null space anyway

Any biases in the reference data will be preserved in the retrievals

But forward model based retrievals also struggle with biases
For each of the four FOVs:
\[ \exp\left(-\frac{z}{7000}\right) \] (where \( z \) in the surface elevation in meters)

200 PC scores representing the IASI measurements in the four FOVs derived from 1200 operational PC scores \((90+120+90)\) times four

26 PC scores representing the MW measurements in the four FOVs derived from 15 AMSU channels and four times 5 MHS channels

Simultaneous retrieval of individual profiles in four adjacent pixels
IASI EFOV eigenvector scores

IASI PC score #2
Common features

IASI PC score #3
Inter-pixel differences
IASI EFOV eigenvectors

Some eigenvectors capture common features (#1, 2, 5) and others capture differences between IFOVs (#3, 4, 6)
State-vector representation

For each of the four FOVs in an EFOV
- \( Ps \) (hPa) Surface pressure
- \( Ta \) (K) Surface air temperature
- \( Wa \) (K) Surface air WV dew point temperature
- \( Ts \) (K) Surface skin temperature
- \( OmC \) (K) Cloud-signal (predicted Obs minus Calc)

100 PC scores for T and W profiles in all four FOVs
- \( T \) (K) at 137 model levels
- \( W \) (K) dew point temperature at 137 model levels

20 PC scores for ozone profiles in all four FOVs
- \( O \) (K) “dew point temperature” at 137 model levels

20 PC scores for surface emissivity in all four FOVs

For each of the four FOVs in an EFOV
- \( QP \) (hPa) absolute error of surface pressure
- \( QT \) (K) absolute error of temperature profile
- \( QW \) (K) absolute error of dew point profile
- \( QO \) (K) absolute error of ozone “dew point” profile
- \( QTs \) (K) absolute error of surface skin temperature
PWLR retrieved surface pressure

Used for interpolation of model levels to fixed pressure grid and as FRTM input in OE
LST correction

Any systematic biases in reference data (ECMWF analysis) are retained

→ Problem for ECMWF LST over arid regions at daytime

LST in training set is replaced with 1DVAR LST, if at daytime, it is higher and there is no inversion
Cloud signal

Experiment to predict the window channel OBS minus CALC brightness temperature (average of five channels)

Traditionally OBS-CALC(NWP FCT) used for cloud screening

- OBS-CALC(PWLR$^3$) is better
- PWLR$^3$ (OBS-CALC(PWLR$^3$)) is faster and better
Training with real measurements and ECMWF

Training data based on co-located ECMWF analysis from 23 days (1st and 17th of each month from July 2013 to June 2014) ~ 28 million cases

15*2*16*4 regression classes with individual regression coefficients
- 15 (symmetric) scan positions
- 2 Day/Night
- 16 K-means clustering on normalised leading MW and IR PC scores
- 4 instances, different number of PC scores used for clustering

Final retrieval obtained as average of 4 ‘ensemble’ retrievals, to reduce random noise

Special versions of the IASI EFOV eigenvectors are available for the case that the IASI measurements are flagged as bad in one or more of the four FOVs - but the regression coefficients stay the same
All sky products ➔ new possibilities

Total water vapour column

Total ozone column

Temperature and water vapour transect plots
Improvements achieved by
- exploiting horizontal correlation
- a better classification (based on k-means clustering)
- averaging four individual retrievals with partially uncorrelated errors
Surface emissivity
Trained with UW Baseline Fit Emissivity Database

Modis Emissivity (1204.8 cm⁻¹) 20120701

lasi Emissivity (1204.8 cm⁻¹) 20120701
Emissivity difference wrt training atlas

Consistent differences between PWLR\textsuperscript{3} retrieval and training atlas day and night

→ Do not simply get the atlas back

? But which emissivities are best?

→ Analysis of obs – calc for different emissivity sources gives the answer
Standard deviation of residual (obs minus calc) with different emissivity sources
Emissivity conclusions

New emissivity databases available – to be tested

Any advantage of using retrievals rather than atlases for training?

Emissivity retrieved globally (i.e. also over sea and under cloud) at 10 wavelengths.

Any set of emissivity eigenvectors can be used to reconstruct full spectral resolution emissivity.
PWLR$^3$ Quality indicators

Regression coefficients to predict the absolute value of the retrieval error used to derive quality indicators for all retrieved parameters

\[ y \approx \bar{y} + R(x - \bar{x}) \]
\[ |y + R(x - \bar{x}) - y| \approx |y + R(x - \bar{x}) - y| + R^E(x - \bar{x}) \]

Cloud signal 2013.07.17

QT (Temperature error) 2013.07.17
Global temperature retrieval minus ECMWF statistics

Quality classes defined by value of the quality indicator for Ta

<table>
<thead>
<tr>
<th></th>
<th>MWIR</th>
<th></th>
<th></th>
<th>IRON</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>Q4</td>
<td>Q5</td>
<td></td>
</tr>
<tr>
<td>MWIR</td>
<td>15.6%</td>
<td>15.6%</td>
<td>15.6%</td>
<td>15.6%</td>
<td>37.9%</td>
<td></td>
</tr>
<tr>
<td>IRON</td>
<td>5.7%</td>
<td>8.6%</td>
<td>7.8%</td>
<td>8.6%</td>
<td>69.3%</td>
<td></td>
</tr>
</tbody>
</table>
Further improvements / Premium SST

Trained for global / all-sky maximum performance
- (clear sky) improvements by further refinement of retrieval classes
- use of predicted SST for class separation

Analyse biases in order to determine opportunities for class splitting

Plot show that bias does not vary with scan position (not surprising as each scan position already has separate class)
AVHRR inhomogeniety

Std(AVHRR radiance)/Mean(AVHRR radiance) [average of channel 4 and 5]

SST bias varies with the scene inhomogeniety
Cloud signal

SST bias relatively independent of cloud strength, except for positive cloud signal.
Validation in state space, NPROVS by NOAA

Comparison with sondes. Reprocessed PWLR^3 profiles to be ingested
Validation in radiance space, Band 1

Clear sky ocean +-60 lat, 2016.03.15

OBS – CALC(ECMWF FCT)
OBS – CALC(ECMWF ANA)
OBS – CALC(IASI PWLR³)

Ozone
Strat.
Temp.

Wavenumber [cm⁻¹]

stddev [mW/m²/sr/cm⁻¹]
Validation in radiance space

Clear sky ocean + -60 lat, 2016.03.15

\[ \text{stddev [mW/m}^2/\text{sr/cm}^{-1}] \]

- OBS – CALC(ECMWF FCT)
- OBS – CALC(ECMWF ANA)
- OBS – CALC(IASI PWLR^3)
Validation in radiance space

Clear sky land and sea high latitudes only |lat| > 50 degree, 2016.03.15

- OBS – CALC(ECMWF ANA)
- OBS – CALC(IASI PWLR$^3$)

**Ozone**

stddev [mW/m$^2$/sr/cm$^{-1}$]
Validation in radiance space, Band 2

Clear sky ocean +-60 lat, 2016.03.15

OBS – CALC(ECMWF FCT)
OBS – CALC(ECMWF ANA)
OBS – CALC(IASI PWLR³)

stddev [mW/m²/sr/cm⁻¹]
PWLR³ averaging kernels

PWLR³ smoothing error

PWLR³ retrieval noise

Combined averaging kernels for optimal estimation using PWLR³ as a-priori
Conclusion

PWLR³ works well!

• Synergistic use of IR and MW
• Comes with reliable quality indicators
• Trained with real measurements → Handles features not modelled by the RTM
• Trained with BIG datasets → Insensitive to random errors in reference data
• Easy and efficient way to handle non linear response
• Further improvement possible by increasing number of classes
• Coefficient size currently about 1.2 GB for MWIR and IRON versions
• Improves optimal estimation when used as a-priori
• Trained with ECMWF analysis, but retrievals do not use forecast
• No minimisation of residuals involved → validation in radiance space possible