Hyperspectral infrared machine learning

ITSC-23 25 June 2021

Tim Hultberg, Thomas August

tim.hultberg@eumetsat.int thomas.august@eumetsat.int

Abstract (skipped when presented)

Supervised learning has proven its potential for satellite retrieval and is an important first step in the operational IASI L2 processors of both EUMETSAT and NOAA. Nevertheless, in clear sky cases, these first retrievals are subsequently improved by variational (optimal estimation like) retrievals. This raises the question whether the precision of ML retrieval can be improved, such that it can fully replace optimal estimation for HSIR temperature and water vapour retrieval. We discuss options for improving the ML retrieval precision.

But good retrieval precision alone is not enough. We also need reliable error characterisation and the possibility to exploit a-priori knowledge to guide the retrieval of fine scale structures, which can't be determined by the measurements. That ML retrieval serves both these needs very well, will be demonstrated in detail. The optimal estimation error characterisation, which depends on the configured observation and a-priori error covariance matrices, is often considered to be more "rigorous" than the error estimates associated with ML retrievals. Such a claim is unjustified as the main error source of retrievals is the null space error (aka "smoothing" error), which is determined by the averaging kernel and the statistical distribution of the profiles - both of which can be well characterised by the ML retrieval. Likewise, there is no problem of characterising the impact of the instrument noise on the ML retrievals (which is typically a much smaller component of the total retrieval error). Extensive validation confirms the good quality of the error estimates of the ML piecewise linear regression IASI retrievals. The incorporation of prior knowledge, typically from NWP forecasts, is embarrassingly simple, just add the additional knowledge to the list of predictors. We show that this enables us to retain the fine scale structures of the forecasts while still correcting the broad scale features when they are not compatible with the measurements. As a by-product, for each retrieved quantity, we obtain a weight describing how big a fraction of the retrieval is taken from the a-priori.

How do we find a retrieval function f which relates the parameter of interest x with the measurement y ?

Variational

Forward model $y = F(x)$

$$f(y) = \mathrm{argmin}_x(|F(x)-y|)$$

Inflexible, online optimization

Statistical

Training set $\{(y_i, x_i)\}$

$$f = \mathrm{argmin}_f(|x_i-f(y_i)|)$$

Flexible, offline optimization

Machine learning – formerly known as statistical retrieval

(in both cases) The retrieval error $f(y)-x_t$ is caused by

1. **Systematic errors in the forward model or the training set**
2. **Overfitting!**
3. **Null space error**
4. **Instrument noise**

"Regularization is the process of adding information in order to solve an ill-posed problem or to prevent overfitting." Important in both cases

Could ML fully replace OE for HSIR L2 retrieval?

Supervised learning already proved its value – important first step in operational retrievals (EUMETSAT and NOAA)

Do we need a second step of variational retrieval to improve any of the following?

1. **Retrieval precision**
2. **Error characterisation**
3. **Use of a-priori knowledge**

The error characterisation and ability to incorporate a-priori knowledge are often seen as strong points of optimal estimation and it might come as a surprise that ML is actually very well suited for both.

Retrieval precision

Piecewise linear regression (PWLR) has been instrumental in improving the quality of the operational IASI L2 retrievals, nevertheless the precision is still slightly improved (in clear cases) by a subsequent variational retrieval.

Can we improve the PWLR/ML? → Ongoing experiments

- **Neural network, Support vector machine, K-nearest neighbour, Random forest, PWLR,…**
- **Increasing the number of PWLR regression classes and optimising the classification**

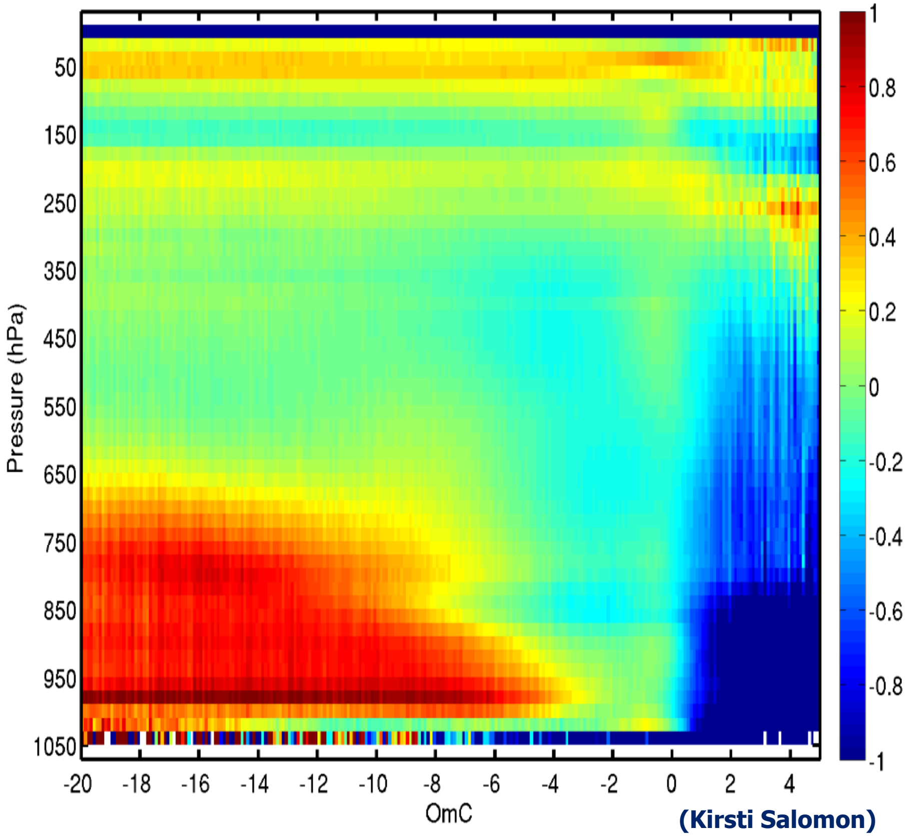
Splitting regression classes for improved precision

The superior performance of PWLR over LR comes from the classification of the data. The details of the classification is key to further improvements.

By construction, linear regression has no bias (w.r.t. the reference data). If the mean retrieval residual varies with one of the predictors (or a function of them), we know that these biases can be removed by further subdivision of regression classes, thereby improving the overall retrieval precision.

Similar to random forests (where each node is a constant value as opposed to a linear model as in PWLR) a tree of regression classes can be build in a divide-and-conquer approach with a systematic search for "good" splits. For execution time reasons, in practice, the evaluation of potential splits must be heuristical and could be based on the variation of the retrieval bias. Here we show an example regression class, in which the retrieval bias varies significantly with the value of the second MHS PC score (synergistic IR+MW retreieval), but very little with the value of the third MHS PC score. Linear combinations of the predictors (for example a first crude linear regression) can also be used to determine splits. Confirming the benefit of a split on independent validation data is important to avoid overfitting. Experiments are ongoing with the goal to improve the retrieval precision compared to the currently operational PWLR.

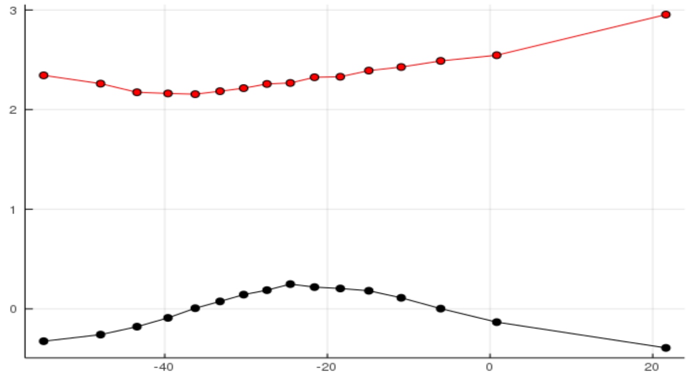
Bias between temperature PWLR retrieval and NWP



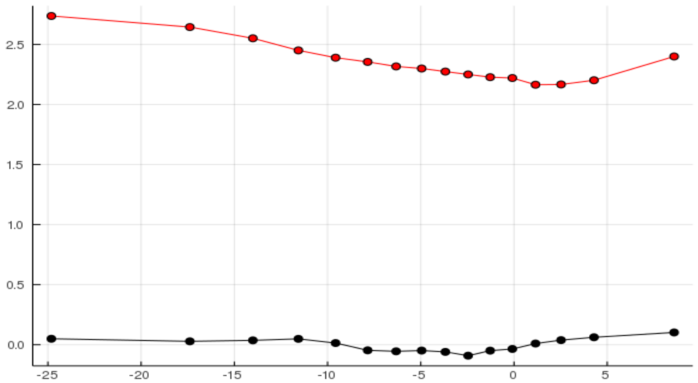
(Kirsti Salomon)

OmC is the retrieved cloud signal (i.e. a function of the measurements)

Mean and standard deviation of surface air temperature residual in a class as a function of two predictors



MHS-PC-2



MHS-PC-3

EUMETSAT

PWLR error characterisation (two options)

I. Using the averaging kernel

Null space error $S_s = (I - A)C_{xx}(I - A)^T$

Retrieval noise $S_n = GS_yG^T$

Where $\mathrm{G} = C_{xy}C_{yy}^{-1}$ and $\mathrm{A} = C_{xy}C_{yy}^{-1}C_{yx}C_{xx}^{-1}$

Total retrieval error $S_s + S_n$

This is applied in each regression class and provides full retrieval error covariance matrix.

II. PWLR retrieval of absolute error

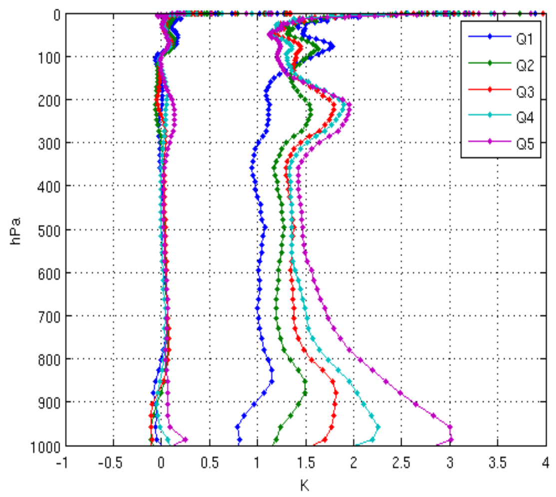
1. Apply PWLR to a training set with reference data
2. Compute the absolute difference between the PWLR retrieval and the reference
3. Train PWLR to retrieve this absolute difference

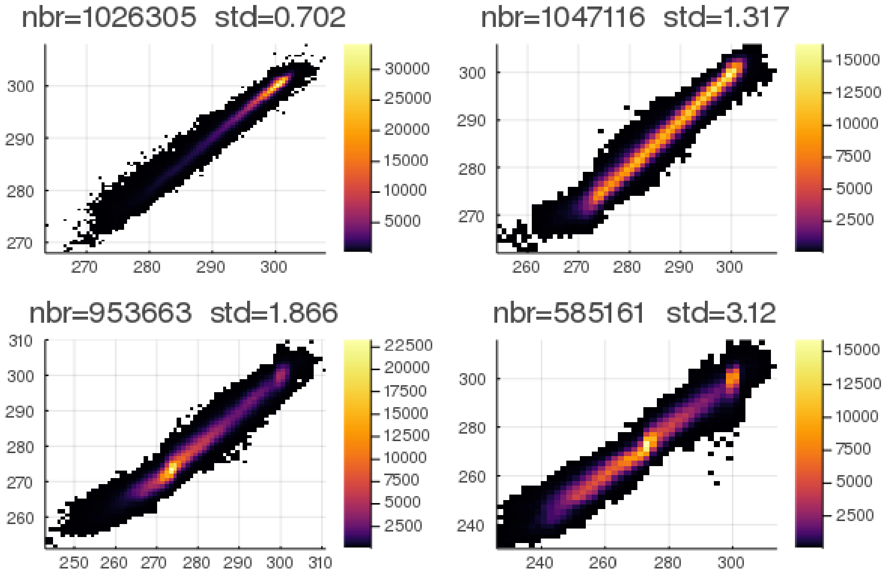
$$y \approx \bar{y} + R(x - \bar{x})$$

$$\left|\bar{y} + R(x - \bar{x}) - y\right| \approx \overline{\left|\bar{y} + R(x - \bar{x}) - y\right|} + R^E(x - \bar{x})$$

The error estimates found by I. apply to a regression class whereas the ones found by II. apply to individual pixels. As the number of regression classes increase and the classes become more homogeneous, the two estimates approach each other.

Independent retrieval statistics separated according to the value of the error estimate confirm the good quality of the error characterisation.

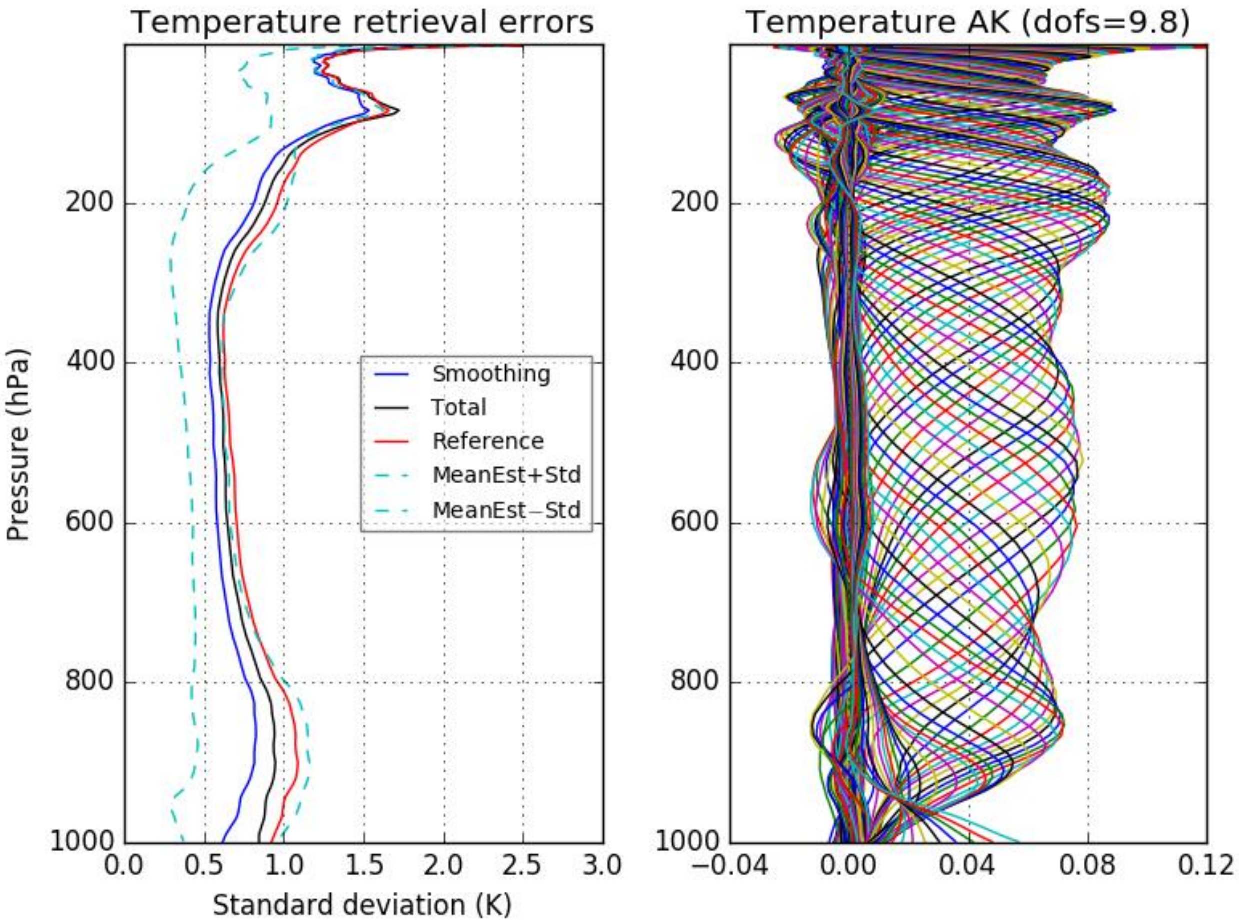




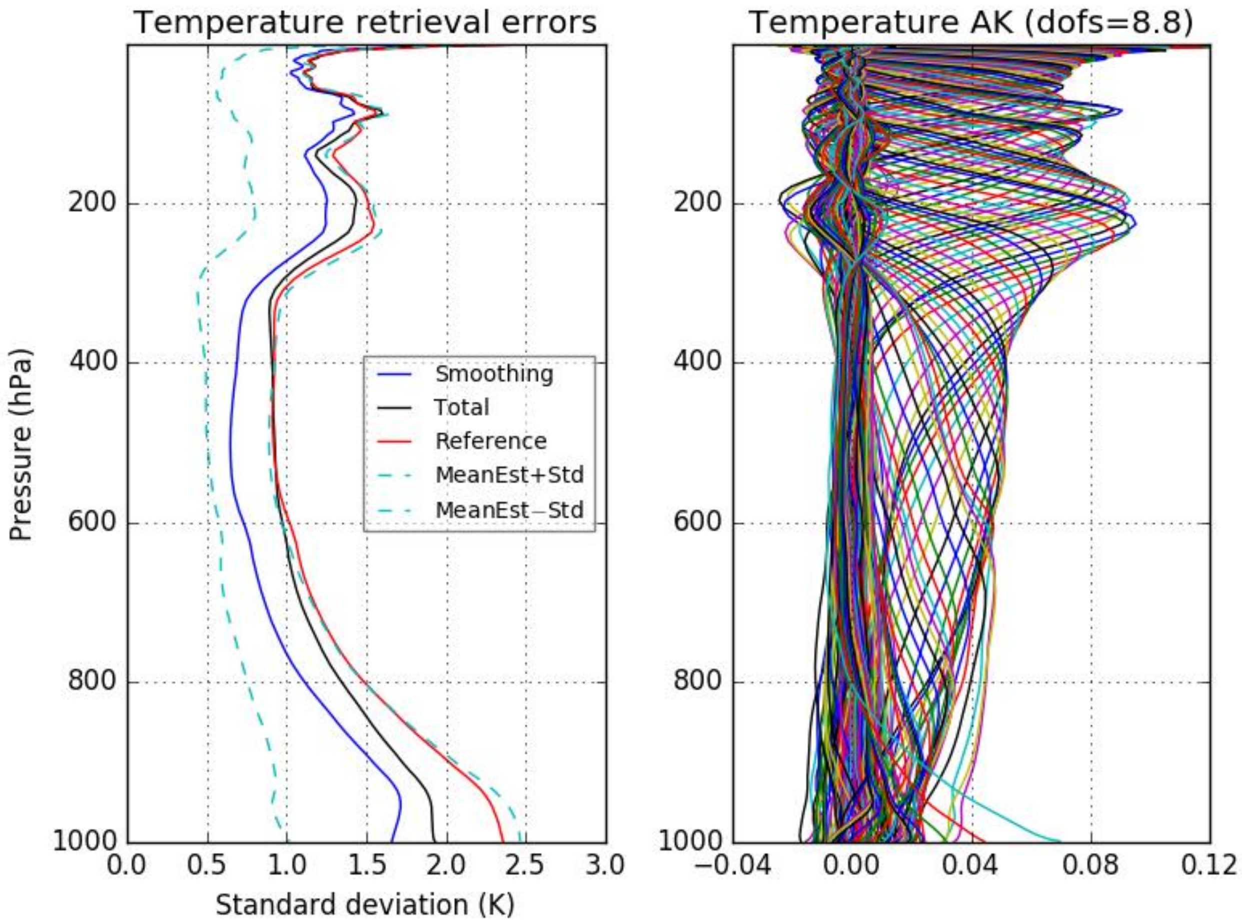
EUMETSAT

Good agreement between error estimates and actual errors

A clear sky regression class



A cloudy regression class



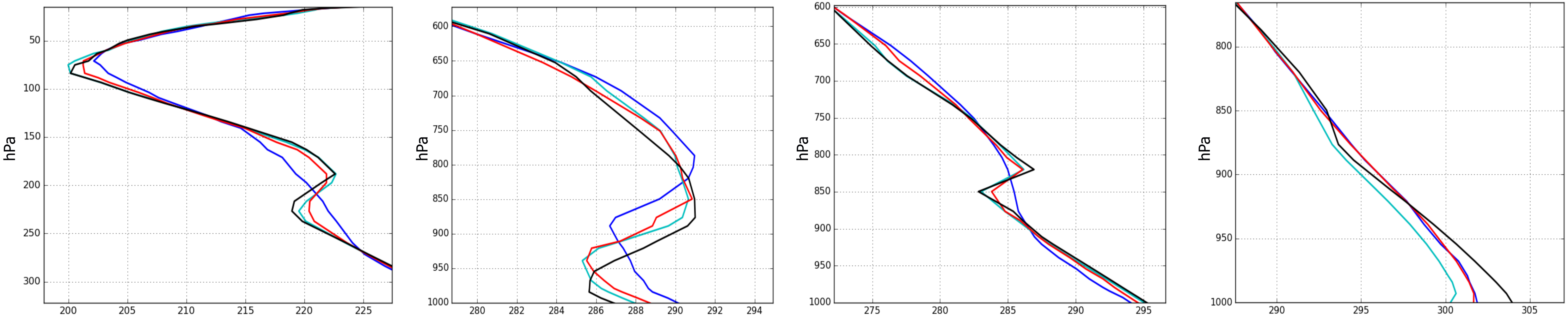
EUMETSAT

Use of a-priori knowledge

A-priori information is easily incorporated by adding it to the predictors y
- y now consists of both measurements and forecast
- x consists of analysis

Learning phase determines relative weights of measurement and a-priori ➔ No need for explicit error covariance of the a-priori

Temperature profile examples **Black: ERA5** **Cyan: FCT** **Blue: PWLR** **Red: PWLR with FCT as prior**

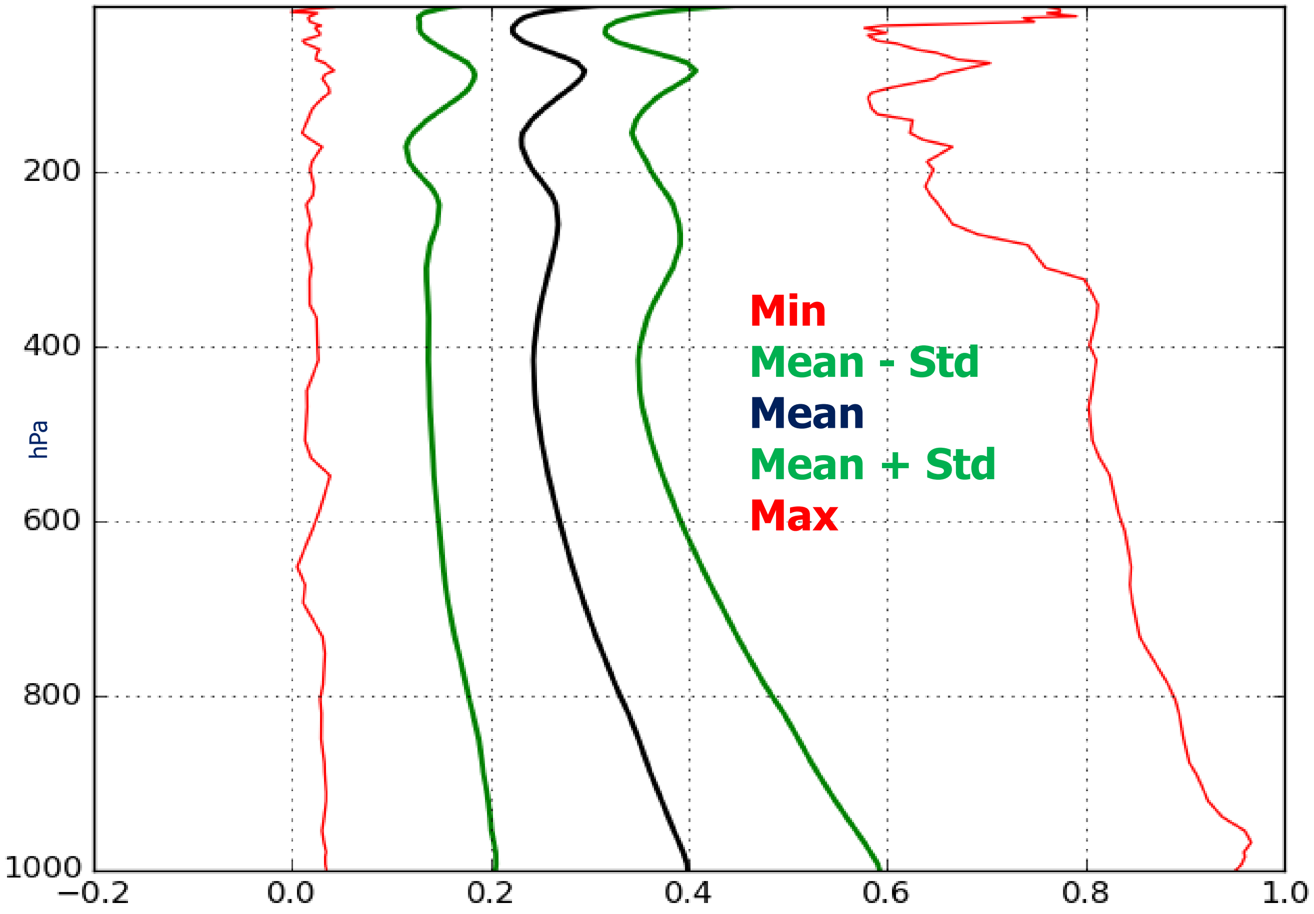


The first three cases show temperature profiles with fine scale structures which are not seen by IASI and are therefore not captured by the standard PWLR. Adding information from the forecast we are able to retain these features in the retrieval. In these cases the retrieval follows the forecast relatively closely. In the last case the forecast seems to be too cold, but IASI provides enough information, not to be led astray by the erroneous forecast.

How much information comes from the a-priori (FCT)?

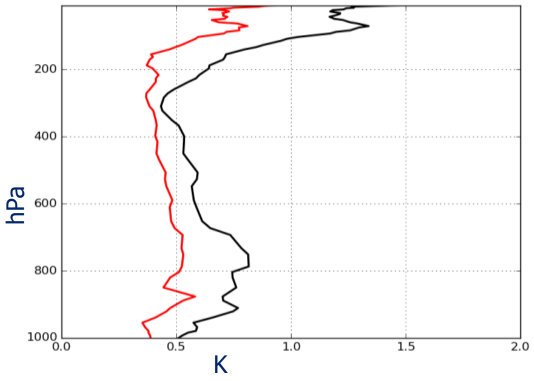
- **The weight of the a-priori for each retrieval is readily available.**
- **On average about 1/3 of each retrieval is taken from the forecast (with considerable variation)**
- **(as expected) the weight of the a-priori is always between 0 and 1**

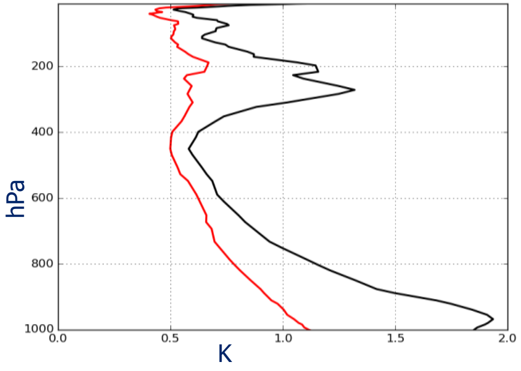
Weight of FCT in the retrieval



Examples of PWLR null space error estimates

Black: FCT free PWLR **Red: FCT as prior in PWLR**

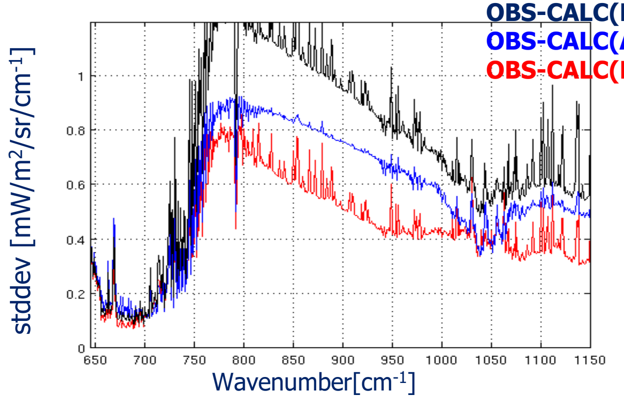


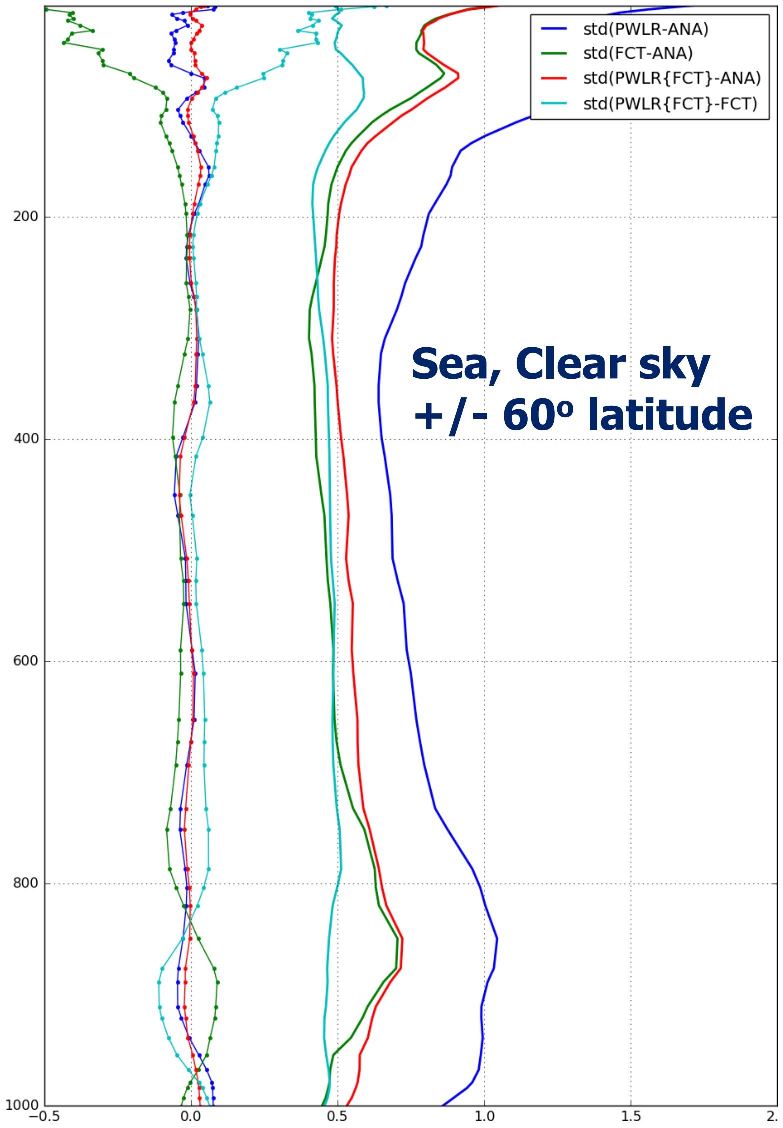


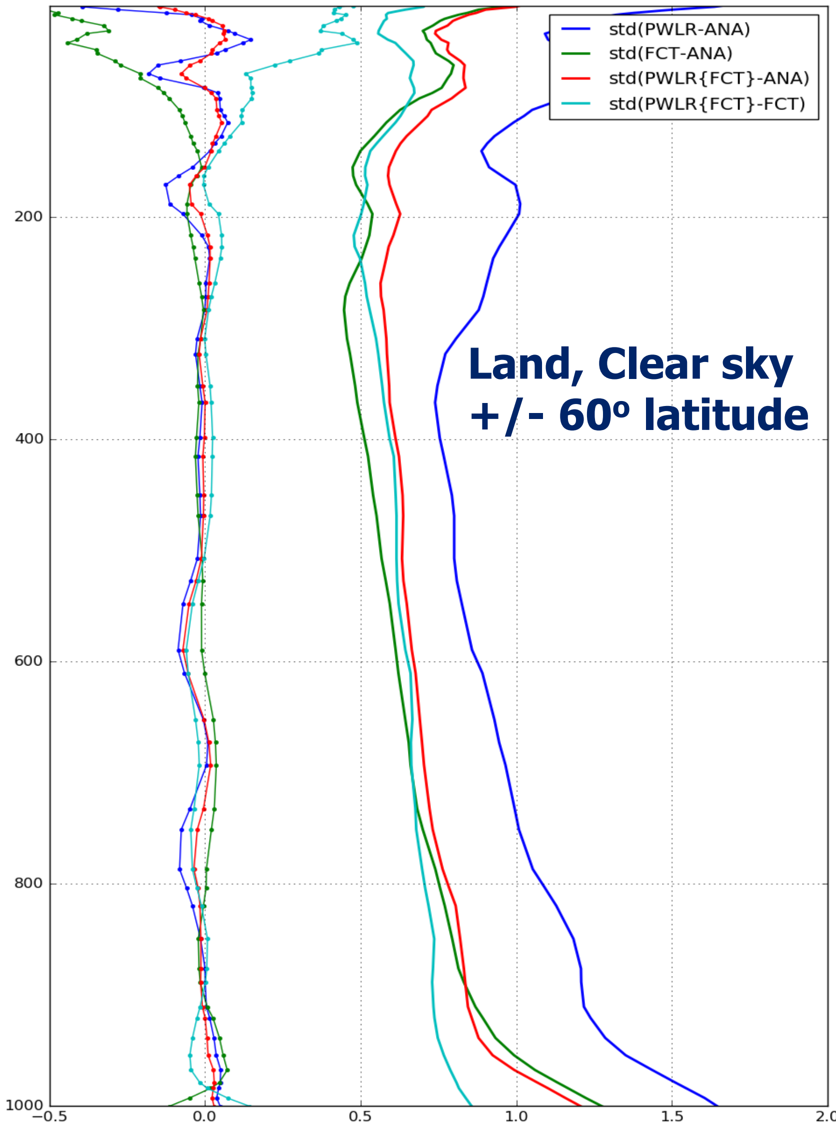
EUMETSAT

Use of FCT a-priori improves retrieval performance

- **Standard deviation of the retrieval error (assessed against analysis) improves between 0.1K and 0.5K depending on the level**
- **Standard deviation of the difference between the a-priori and the retrieval is about 0.5K (a bit higher over land)**
- **Agreement with analysis is similar for retrieval and forecast (but agreement with IASI is much better for the retrieval)**







EUMETSAT

Limitations / challenges

Outliers (Identification and removal of outliers, Robust regression)

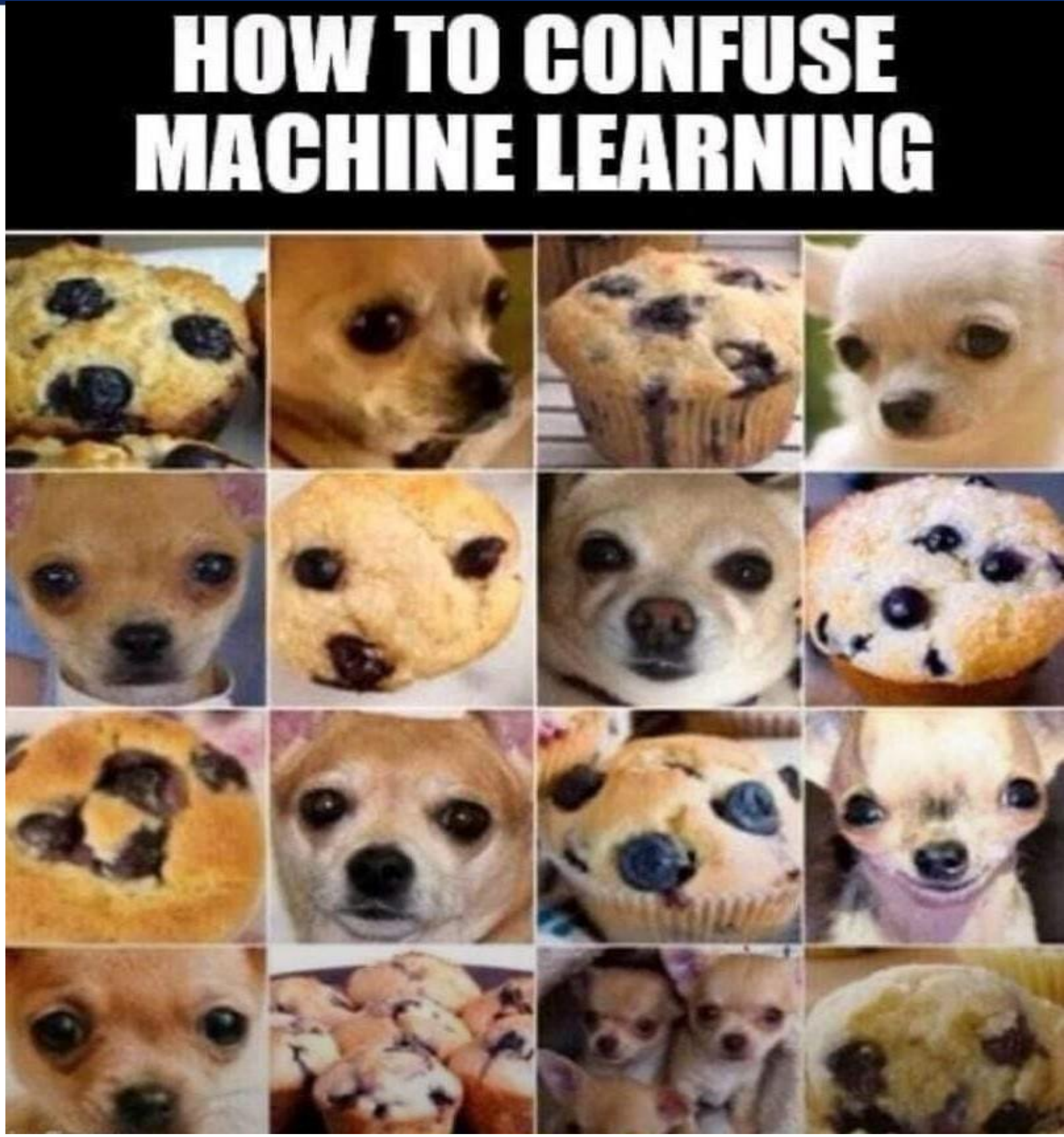
Generalisation I (IASI-B ➔ IASI-C, IASI ➔ IRS)

Generalisation II (unusual spectra – Raikoke, increasing CO2)

- Some very bad retrievals associated with Raikoke eruption – well identified by the error characterisation
- Trained CO2 retrieval with reference data from 2016.07 -2017.06 and applied for 2008 to 2018. Obtained global CO2 increase of 0.4 ppm/year (correct value is about 2.3 ppm/year)

Lack of reference data (N_2O)

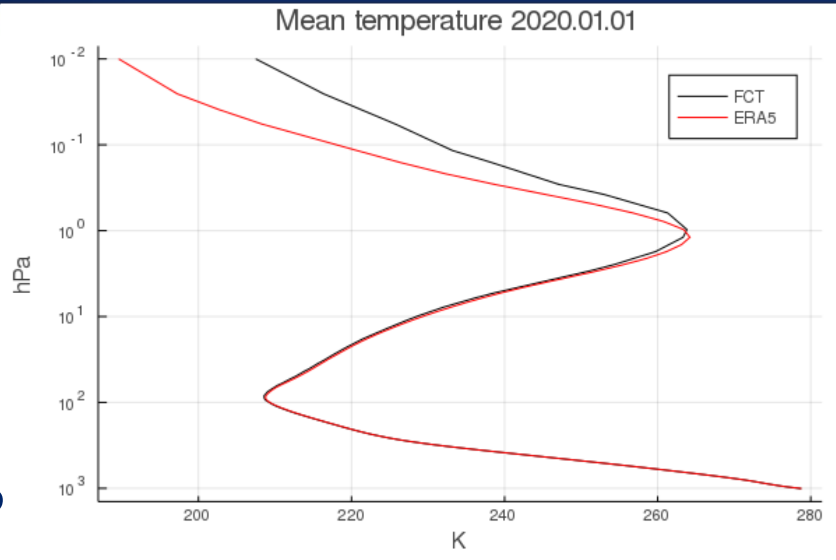
Biases in the reference data, inherited by the training



Stratospheric temperature biases

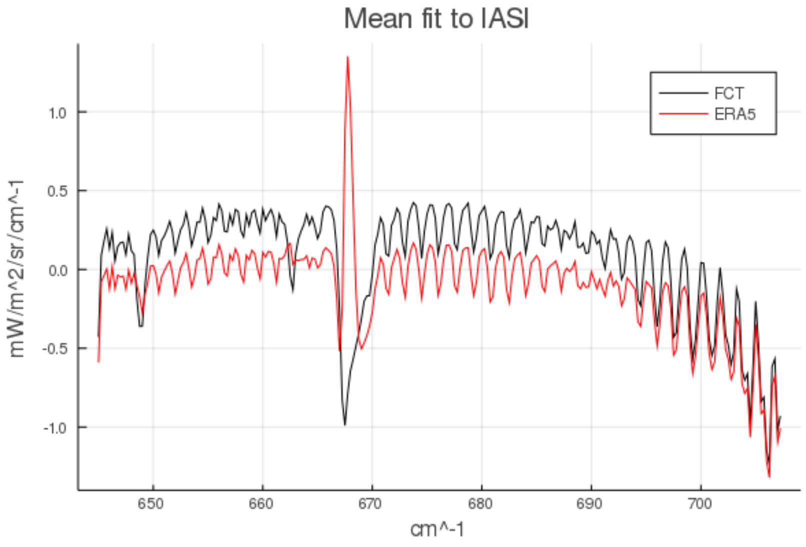
Two different opinions on stratospheric temperature

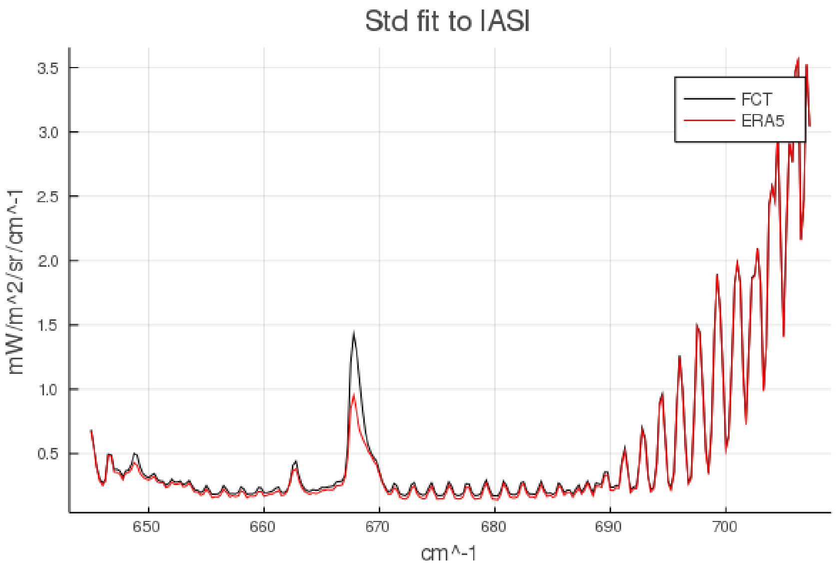
Let's ask IASI who is right?



ERA5 fits IASI better in terms of standard deviation, but for training data the mean is more important. Here there is no clear winner and we have opposite biases around 667.75 cm^{-1}, indicating that an average of the two might fit IASI better.

In a variational retrieval experiment we got excellent fit to IASI with relatively small modifications of the temperature, no matter if we used FCT or ERA5 as background – the IASI null space allows for big variations in the stratospheric temperature. ML and OE alike need external guidance.





EUMETSAT

Conclusions

- Attempts to improve 1DVar generally relies on statistics ➔ Often better to use statistics directly to characterise the retrieval function

- The key to good retrievals is to lower the errors in the null space. How? The measurements alone can't help you, but the 'historical' correlations in the reference data can.

- PWLR/ML retrievals are in very good agreement with sondes/analysis **and** with the measured radiances (despite radiance residuals not being minimised explicitly)

- The null space error can only be estimated with knowledge of the underlying statistics. These are determined as an integrated part of the ML training rather than postulated upfront as in OE ➔ ML error estimates tend to be more reliable than error estimates from variational retrieval

- PWLR/ML is very well suited for incorporation of a-priori knowledge and this can improve the overall retrieval quality (on the other hand the usefulness of adding fine-scale structures in the instrument null space is debatable. A low rank representation of the profiles retaining only the observable broad scale features might be a better choice and would be suitable for L2 assimilation)

- Variational retrieval could still play an important role as a final refinement of the ML retrieval for better generalisation of the retrieval to situations which have not been included in the training set (for example to cope with atmospheric trends)