Data Assimilation Questions

1. Which are the general key capabilities of a fully-fledged data assimilation system?

2. How can the ill-conditioned optimisation problem be addressed in variational data assimilation?

3. Which optimality criterion and error distribution is the underlying assumption for 4D-var and Kalman filtering?

4. Which retrieval states of remote sensing techniques can be assimilated?

5. What consideration determines the choice of an analysis parameter (e.g. initial value, surface forcings)?
Answer to Question 1:
Objective of atmospheric data assimilation

- "is to produce a regular, physically consistent four dimensional representation of the state of the atmosphere
- from a heterogeneous array of in situ and remote instruments
- which sample imperfectly and irregularly in space and time.

Data assimilation
- extracts the signal from noisy observations (filtering)
- interpolates in space and time (interpolation) and
- reconstructs state variables that are not sampled by the observation network (completion).“ (Daley, 1997)
Answer to question 2: Transformed cost function

\[ d_i := y_i^o - H \text{M}(t_i, t_0)x(t_0) \]

\[ \delta x(t_0) := x^b - x(t_0) \]

transformation

\[ v := B^{-1/2} \delta x_b \]

transformed cost function

\[ J(v) = 1/2 v^T v + 1/2 \sum_{m=0}^{N} d_m^T R^{-1} d_m \]

transformed gradient of the cost function

\[ \nabla_v J(v) = v + B^{1/2} \sum_{m=0}^{N} M^T(t_m, t_0) H^T R^{-1} d_m \]

Pro transformation:  
minimisation problem is better conditioned  
Contra:  
strictly positive definite approximation to B required

Computation of inverse B, square root B, inverse square root B by (Sca)LAPACK eigenpair decomposition, but better choices available.
Answer to Question 3:

• The least square distance, weighted by observation error and background/forecast error covariance matrices is the optimality criterion for 4D-var and Kalman filtering.

• Errors are assumed to have Gaussian distribution, uncorrelated in time (white noise).
Answer to question 4: At which retrieval level to assimilate remote sensing data?

- **Assimilation of retrieved products from space agencies or research institutes (“level 2”)**
  - Pro: most simple for assimilation (assimilation “like in situ observations”)
  - Con: inexact as \( \mathbf{x}_b \) and \( \mathbf{B} \) are inconsistent and mostly of poorer quality as available at NWP centres

- **Locally produced or “1D-Var” or “Averaging Kernel” retrievals**
  - Pro: \( \mathbf{x}_b \) and \( \mathbf{B} \) much better known (e.g. fronts featured and consistent with model)
  - Con: \( \mathbf{x}_b \) and \( \mathbf{B} \) used twice with the subsequent assimilation: \( \mathbf{y} \) and \( \mathbf{x}_b \) correlated

- **Direct assimilation of radiances (“level 1”)**
  - Pro: retrieval step is essentially incorporated within the main analysis by finding the model variables that minimize a cost function measuring the departure between the analysed state and both the background and available observations.
  - Con: Fast linearized radiative transfer scheme and its adjoint is needed.
Answer to Question 5:

Selection of optimisation parameter
Atmospheric chemistry treatment of parameters with type: ("paucity of knowledge*importance") is high:

For meteorology:
- initial values, in future soil parameters

For atmospheric chemistry:
- emission rates,
- deposition velocities,
- PBL structure