Advancements and limitations in aerosol remote sensing:

inversion, inversion and inversion

Questions:

- *How to derive maximum information? What does it mean quantitatively?*

- *What to do if information in the measurements is not sufficient?*

Oleg Dubovik

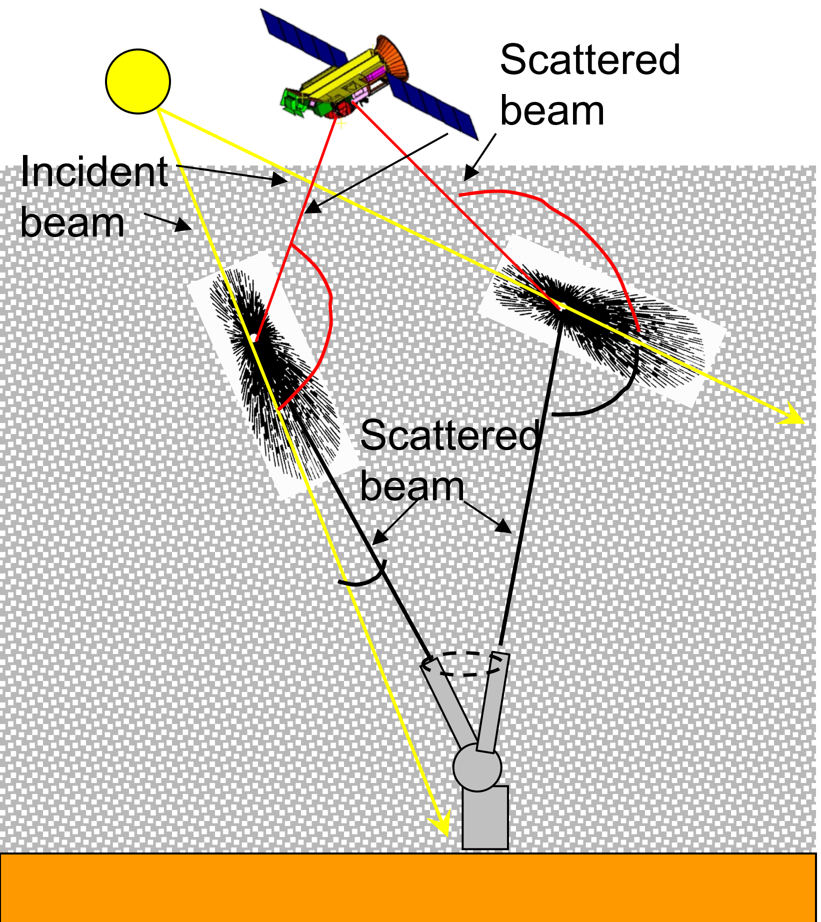
Laboratoire d'Optique Atmosphérique, CNRS/University of Lille, France

Funded by the European Union

ATRRI Workshop, Limassol, Cyprus, 6–7 March 2025

Inversion is an inherent part of any remote sensing approach

Light scattering measured from ground and space

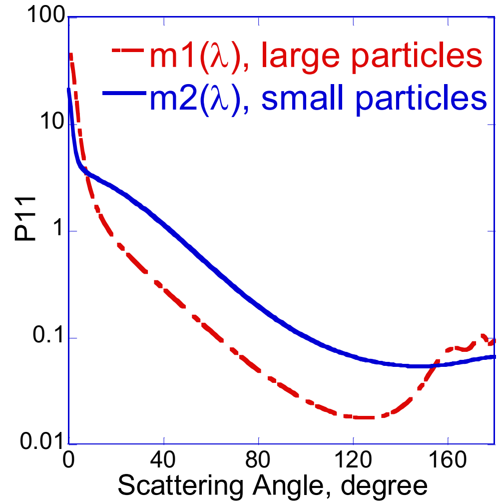


Diffuse light

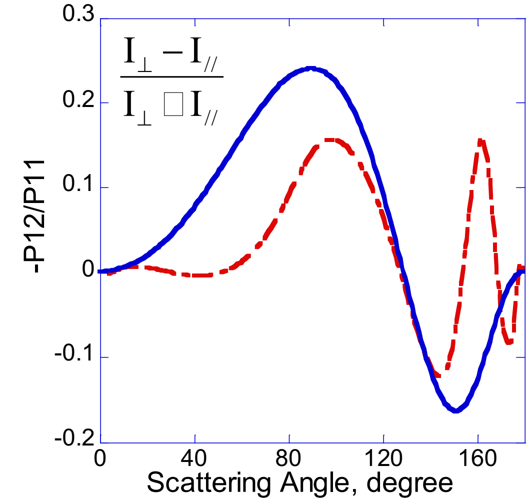
$$\Theta \neq \Theta_0 (!!!)$$

$$F(\lambda) \approx F_0(\lambda)\, e^{-\left(\frac{\tau(\lambda)}{\cos(\Theta_0)}\right)} P(\Theta,\lambda) + \ldots$$

Phase Function P(Θ,λ)



Polarization

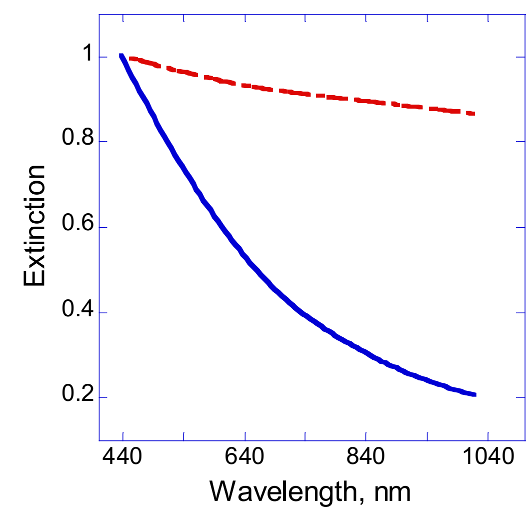


Direct light

$$\Theta = \Theta_0 (!!!)$$

$$F(\lambda) \approx F_0(\lambda)\, e^{-\left(\frac{\tau(\lambda)}{\cos(\Theta_0)}\right)} + \ldots$$

Extinction τ(λ) →



GRASP: Generalized Retrieval of Atmosphere and Surface Properties

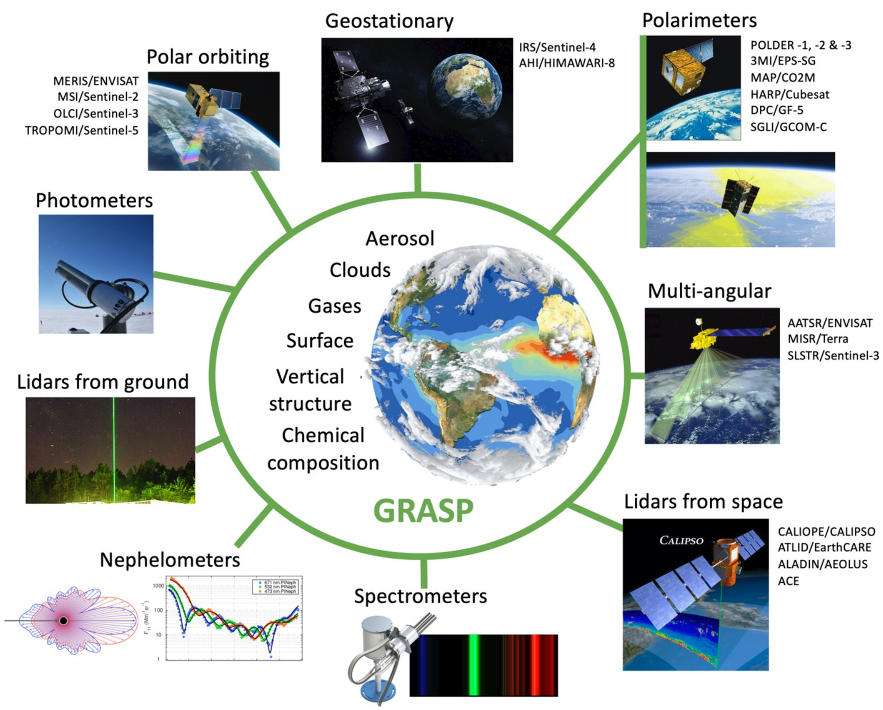
GRASP is advanced algorithm for retrieval of aerosol, gas and surface properties from diverse remote sensing observations and any combination of them based on:

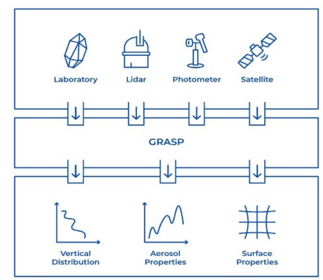
Forward Model for rigorous simulation of atm. radiation.

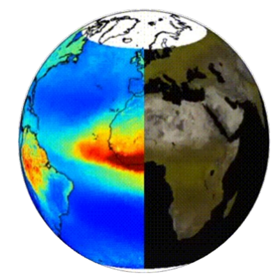
+

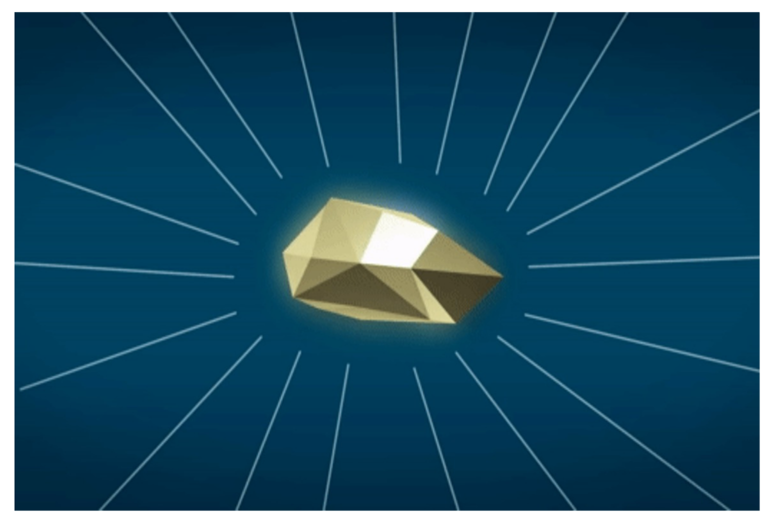
Inversion with *a*pplying multiple *a priori* constraints

Dubovik et al. "A Comprehensive Description of Multi-Term LSM for Applying Multiple a Priori Constraints in Problems of Atmospheric Remote Sensing: GRASP Algorithm, Concept, and Applications", Front. Remote Sens., 2021



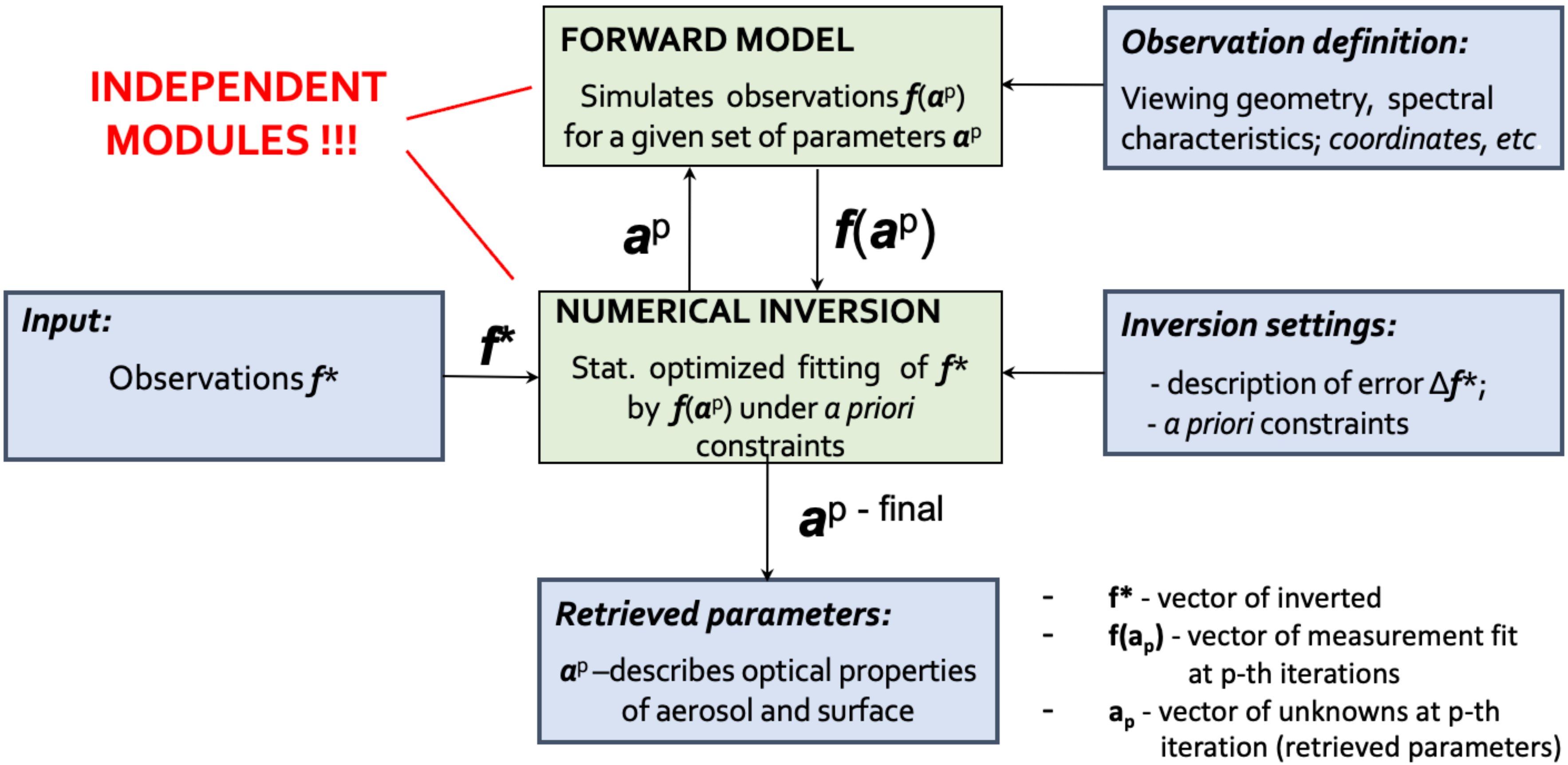






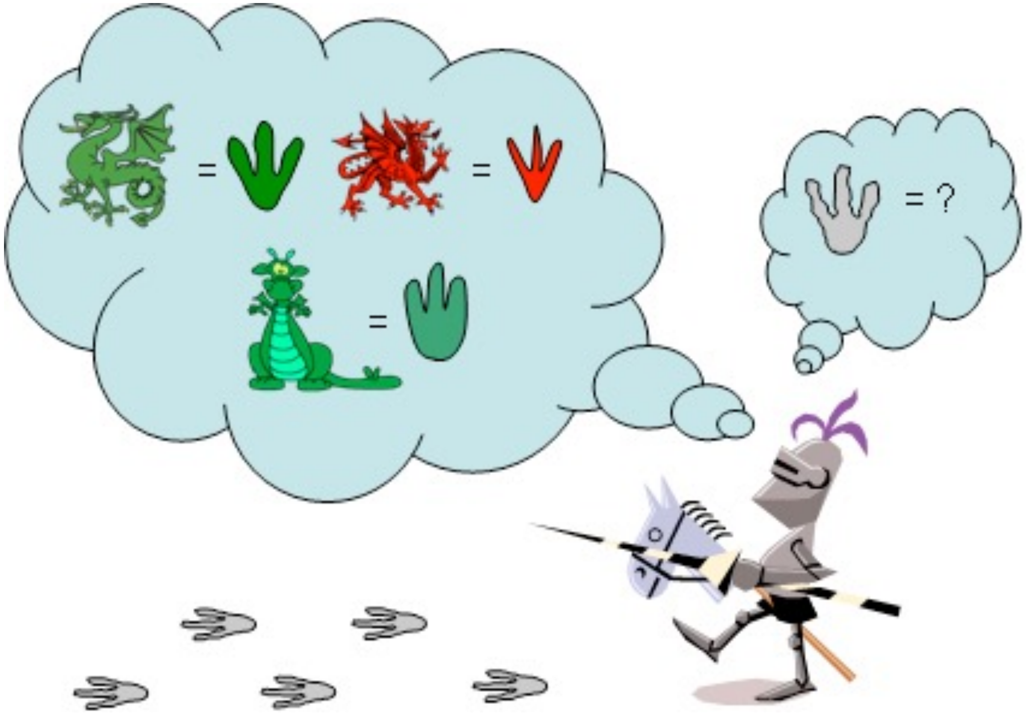
ATRRI Workshop, Limassol, Cyprus, 6–7 March 2025

General structure of the GRASP algorithm

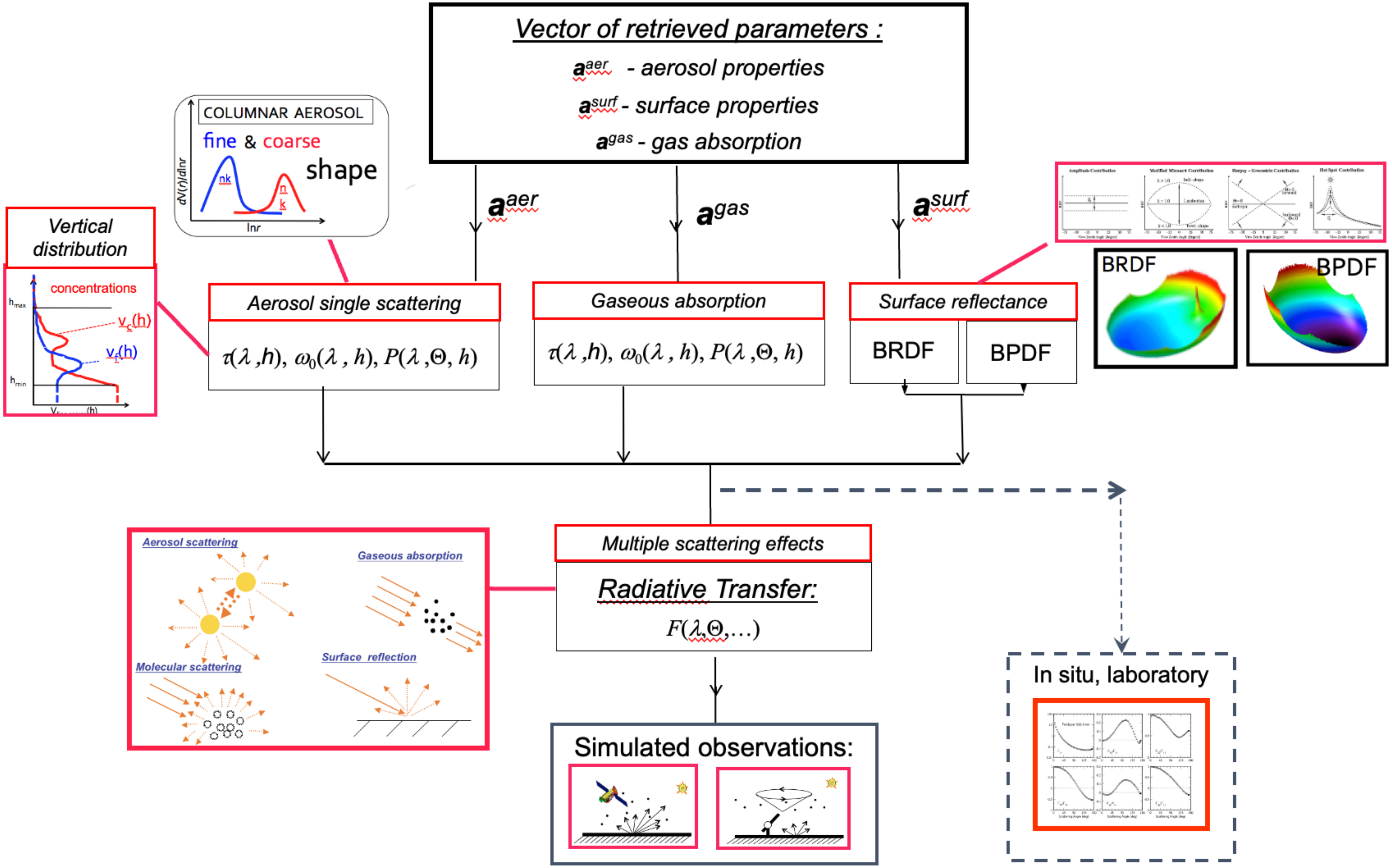


Forward problem: knowing the footprint that a particular species of dragon leaves behind - easy !

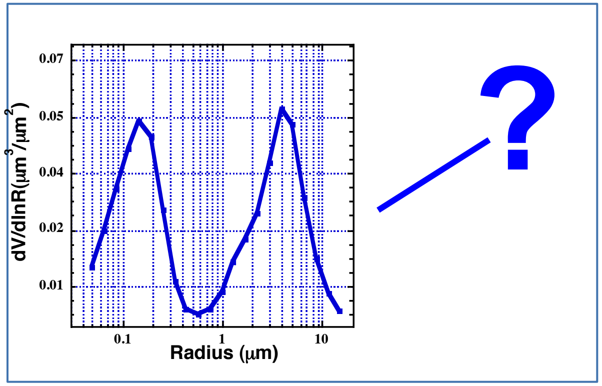
Inverse problem: inferring the species of dragon from the footprint – difficult !



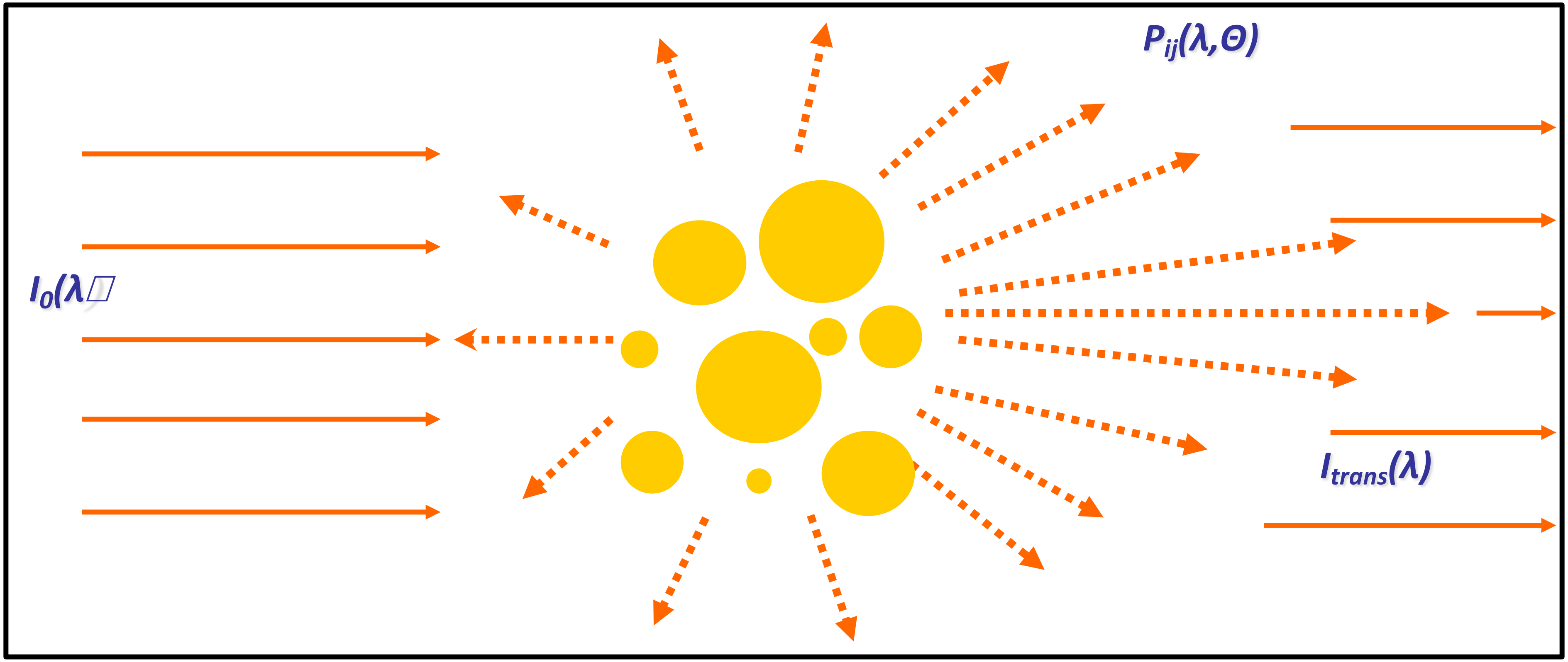
GRASP forward model



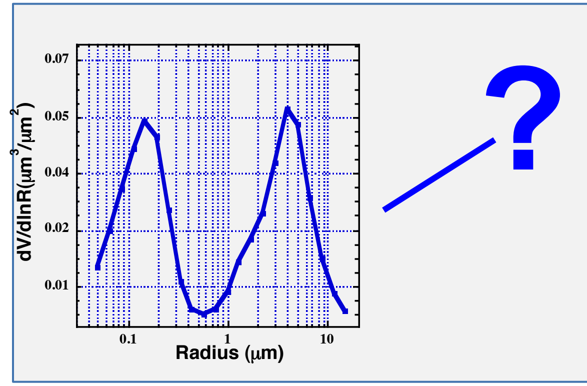
Inverse Problem: Retrieval of particle size distribution from light scattering



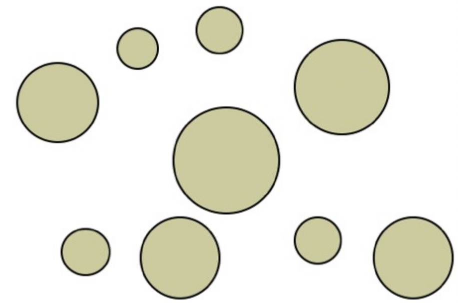
$$P(\lambda;\Theta)=\int_{r_{\min}}^{r_{\max}} K\left(\lambda;\Theta;n,k,\ldots\right)\frac{dV\left(r\right)}{dr}dr$$



Inverse Problem: Retrieval of particle size distribution from light scattering



Fredholm integral equation of first kind:



$$P(\lambda;\Theta)=\int_{r_{min}}^{r_{max}} K\left(\lambda;\Theta;n,k,..\right)\frac{dV\left(r\right)}{dr}dr$$

↓

$$\mathbf{F}\boldsymbol{a}=\boldsymbol{f}^{*}$$

=

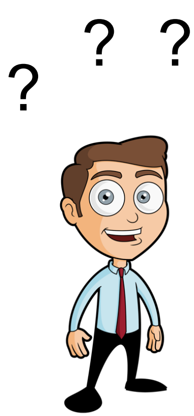
$$\begin{pmatrix} f_1^{*} \\ f_2^{*} \end{pmatrix}=\begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

- ***to solve ?***

Goal: **Objectively** the best retrieval ?

user of inversion

esa
European Space Agency
Copernicus
Europe's eyes on Earth
EUMETSAT
HEALTH AND PEOPLE
air quality
Army
etc...
RESEARCH



accuracy

clarity

Exact science

In practice !!!

speed

simplicity

flexibility

robustness

accessibility

completeness

etc...

Which approach to use?

$$P(\Delta \hat{f}^*) = P(f(\hat{a}) - f^*) = P(f(\hat{a}) \mid f^*) = max.$$

- MML

$$\hat{a} = \left(F_1^T C_1^{-1} F_1\right)^{-1} F_1^T C_1^{-1} f_1^*$$

- LSM

$$\hat{a} = \left(F^T C_f^{-1} F + C_a^{-1}\right)^{-1} \left(F^T C_f^{-1} f^* + C_a^{-1} a^*\right)$$

- « Optimal estimations », C. Rodgers

$$\hat{a} = a^* - C_{a^*} F^T (C + F C_{a^*} F^T)^{-1} (F a^* - f^*)$$

- Kalman filter

$$\hat{a} = \left(F^T C_f^{-1} F + \gamma S^T S\right)^{-1} \left(F^T C_f^{-1} f^*\right)$$

- Phillips-Tikhonov-Twomey

$$\hat{a} = \left(F^T F + \gamma I\right)^{-1} F^T f^*$$

- Tikhonov Regularization

$$a^{p+1} = a^p - t_p F_p^T C^{-1} (f(a^p) - f^*)$$

- Steepest Desent Method

$$a_i^{p+1} = a_i^p \prod_{j=1}^{N_f} \left(1 + \left(f_j^* / f_j^p - 1\right) \tilde{F}_{ji}\right)$$

- Twomey-Chahine

$$a_i^{p+1} = a_i^p \left(f_i^* / f_i^p\right)$$

- Chahine

Assimilation, 4DVR

SVD, gradient methods, etc.

$$\mathbf{F}\boldsymbol{a} = \boldsymbol{f}^*$$

Basic idea of inversion

$$\begin{pmatrix} f_1^* \\ f_2^* \end{pmatrix} = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \Rightarrow \boxed{\hat{\boldsymbol{a}} = (\mathbf{F})^{-1} \boldsymbol{f}^*}$$

$\boldsymbol{f}^*$

$\mathbf{F}$ – square

$\boldsymbol{a}$ - parameters of size distribution

$$\begin{pmatrix} f_1^* \\ f_2^* \\ f_3^* \end{pmatrix} = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \\ F_{31} & F_{32} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \Rightarrow \hat{\boldsymbol{a}} = ???$$

$\boldsymbol{f}^*$

$\mathbf{F}$ – rectangular

$\boldsymbol{a}$

$$\mathbf{F}\boldsymbol{a} = \boldsymbol{f}^*$$

Basic idea of inversion

$$\begin{pmatrix} f_1^* \\ f_2^* \end{pmatrix} = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \Rightarrow \quad \hat{\boldsymbol{a}} = (\mathbf{F})^{-1} \boldsymbol{f}^*$$

$\boldsymbol{f}^*$ $\mathbf{F}$ – square a - parameters of size distribution

$$\begin{pmatrix} f_1^* \\ f_2^* \\ f_3^* \end{pmatrix} = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \\ F_{31} & F_{32} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \Rightarrow \quad \hat{\boldsymbol{a}} = \left(\mathbf{F}^T \mathbf{F}\right)^{-1} \mathbf{F}^T \boldsymbol{f}^*$$

$\boldsymbol{f}^*$ $\mathbf{F}$ – rectangular a

Least Square Method - LSM

$$\begin{pmatrix} f_1^* \\ f_2^* \\ f_3^* \end{pmatrix} = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \\ F_{31} & F_{32} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \end{pmatrix}$$

noise

system is redundant
⬇
noise can be accounted

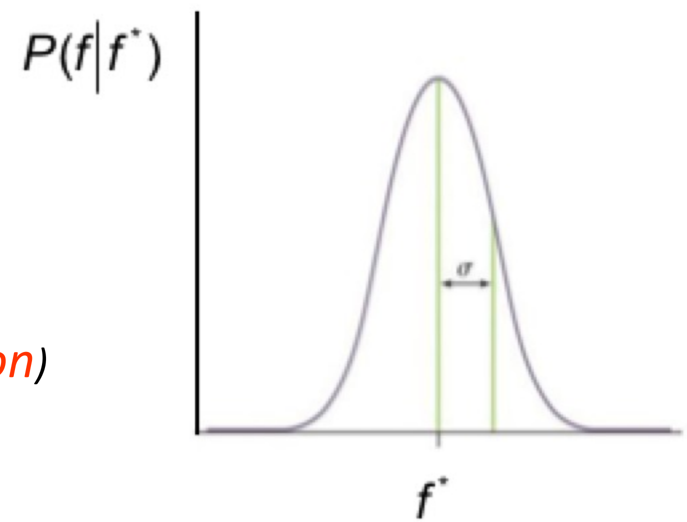
$$\mathbf{F}\,\boldsymbol{a} = \boldsymbol{f}^* = \boldsymbol{f} + \Delta \boldsymbol{f} \rightarrow$$

Random value

$$P(f \mid f^*)$$ — *PDF (Probability Density Function)*

$$\hat{\mathbf{a}} = \mathbf{a}_{true} + \Delta\mathbf{a}$$

$$P(f(a) \mid f^*)$$ — *PDF(Likelihood function)*

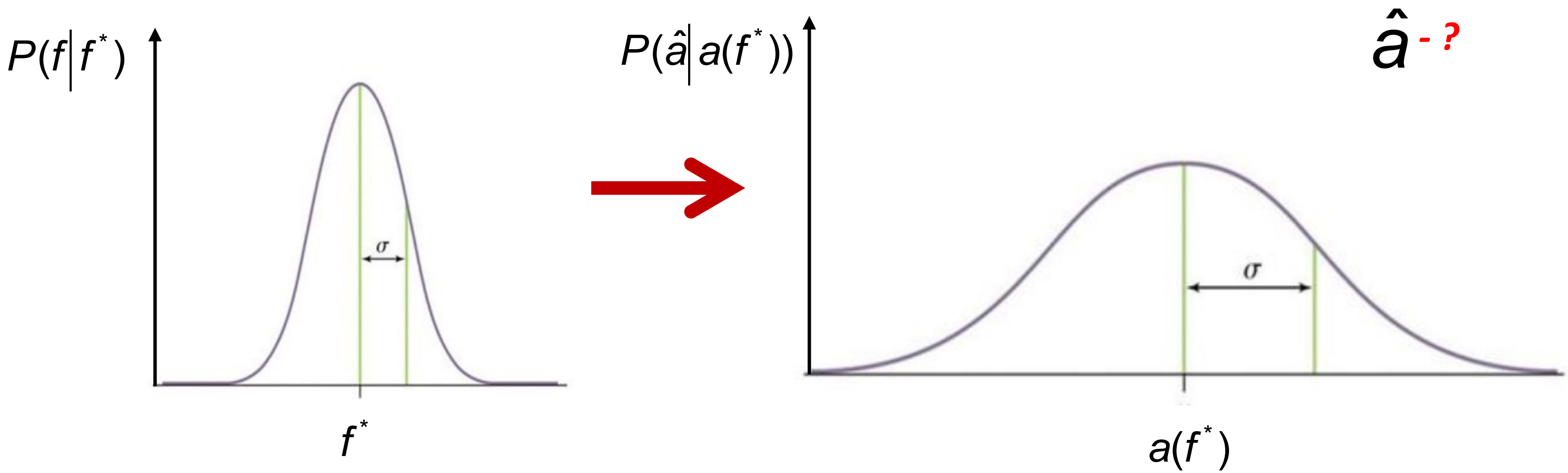


Statistical estimation - optimization

$$P(f(a)|f^*) = P(f(a) - f^*) = P(\vartriangle f(a))$$

$$P(f|f^*) \longrightarrow P(\hat{a}|a(f^*)) = P(\hat{a}|f^*)$$

Need to derive maximum information



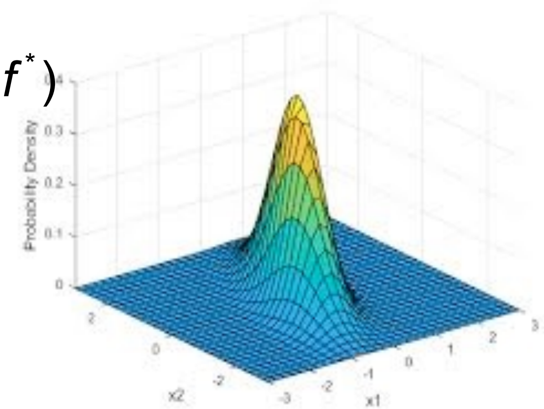
Statistical estimation - optimization

$$P(f(a)|f^*) = P(f(a) - f^*) = P(\triangle f(a))$$

$$P(f|f^*) \longrightarrow P(\hat{a}|a(f^*)) = P(\hat{a}|f^*)$$

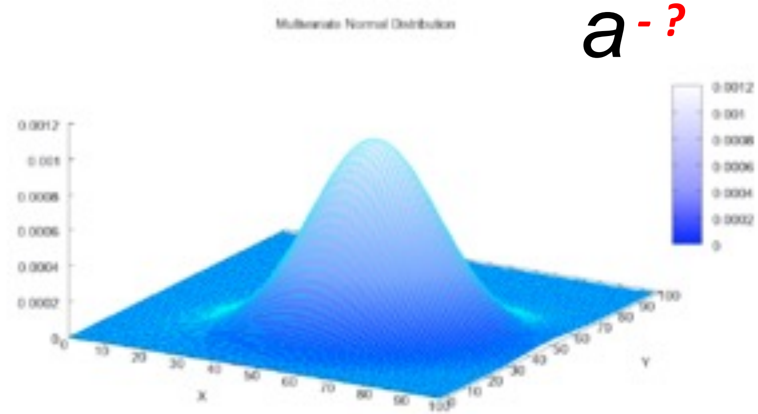
Need to derive maximum information

$P(f|f^*)$



$P(\hat{a}|a(f^*))$

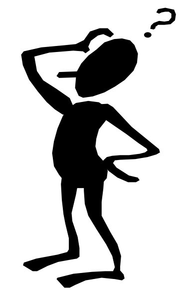
$\hat{a}$ - ?



f^* 2 parameters $a(f^*)$

Information quantity -?

How come quantity of information something to do with PDF



1. Fisher Information

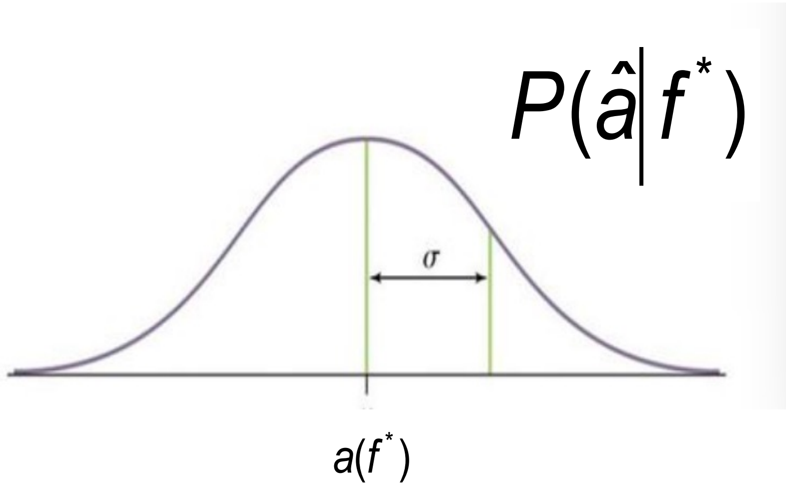
$$h\left(P\left(\hat{a}\right)\right)=\int\left(-\frac{\partial^2 \ln P\left(\hat{a}\middle|f\right)}{\partial \hat{a}^2}\right)P\left(\hat{a}\middle|f\right)df$$

2. Shanon Information

$$h\left(P\left(\hat{a}\right)\right)=\int\left(-\log_2 P\left(\hat{a}\middle|f\right)\right)P\left(\hat{a}\middle|f\right)df=N_{bits}=f(-\log_2(\sigma))$$

$$N_{symbols}\sim 2^{N_{bits}}$$

number of bits (binary digits) needed to represent the number of distinct estimates that could have be obtained



$$P(a^{true} = a') - P(\hat{a})$$ **- minimum**

⇊↓

$$lnP(a^{true} = a') - lnP(\hat{a})$$ **- minimum**

$$\boldsymbol{P}_{1,2,3} = \boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{P}_3 \ldots \longrightarrow ln\boldsymbol{P}_{1,2,3} = ln\boldsymbol{P}1 + ln\boldsymbol{P}_2 + ln\boldsymbol{P}_3 \ldots$$

$$\ln P(a) = \ln P(a') + \left.\frac{\partial \ln P}{\partial a}\right|_{a=a'} (a-a') + \frac{1}{2}\left.\frac{\partial^2 \ln P}{\partial a^2}\right|_{a=a'} (a-a')^2 + \ldots (a-a')^3 + \ldots$$

$$\ln P(a) - \ln P(a') = \left.\frac{\partial \ln P}{\partial a}\right|_{a=a'} (a-a') + \frac{1}{2}\left.\frac{\partial^2 \ln P}{\partial a^2}\right|_{a=a'} (a-a')^2 + \ldots (a-a')^3 + \ldots$$

0

$$\left.\frac{\partial \ln P(a)}{\partial a}\right|_{a=a'} = 0$$

MML (Method of Maximum Likelihood)

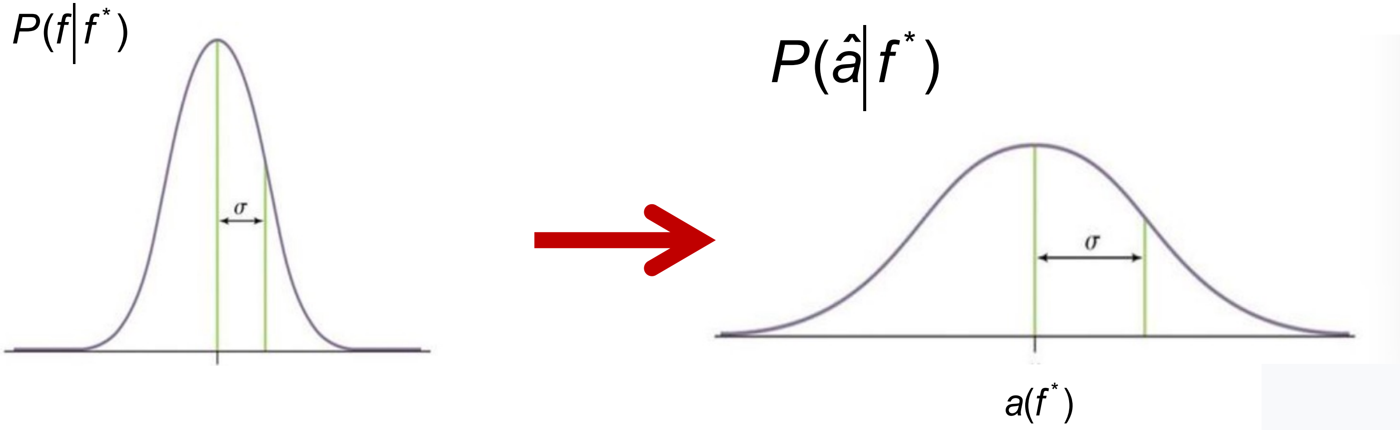
$$\ln P(a) \approx \ln P(a') + \frac{1}{2}\frac{\partial^2 \ln P}{\partial a^2}\bigg|_{a=a'} (a-a')^2$$

$$P(a) \sim e^{\left(\frac{1}{2}\left(\frac{\partial^2 \ln P}{\partial a^2}\right)(a-a')^2\right)} = e^{-\frac{1}{2}\frac{(a-a')^2}{\left(-\frac{\partial^2 \ln P}{\partial a^2}\right)^{-1}}} = e^{-\frac{1}{2}\frac{(a-a')^2}{\sigma_a^2}}$$

Normal Distribution

$$\int\left(-\frac{\partial^2 \ln P}{\partial a^2}\right)P\,da = \int\left(-\frac{\partial \ln P}{\partial a}\right)^2 P\,da = \frac{1}{\sigma_a^2}$$

- Fisher Information



The narrower *P(a|f*)* around *a* the better

$$I_a = \frac{1}{\sigma^2_{smallest}} = \int \left(-\frac{\partial^2 \ln P}{\partial a^2} \right) P\, da = \int \left(-\frac{\partial \ln P}{\partial a} \right)^2 P\, da$$

- Fisher Information

Ronald Fisher
FRS

Fisher in 1913

Born	Ronald Aylmer Fisher 17 February 1890 London, England, UK
Died	29 July 1962 (aged 72) Adelaide, SA, Australia

$$\nabla \ln P(\boldsymbol{a})\Big|_{\hat{\boldsymbol{a}}} = \boldsymbol{0}$$

MML for several unknowns

$$P(\hat{\boldsymbol{a}}_{MML}) \sim e^{-\frac{1}{2}(\Delta \boldsymbol{a})^T \mathbf{C}_a^{-1} (\Delta \boldsymbol{a})} = e^{-\frac{1}{2}(\Delta \boldsymbol{a})^T \mathbf{I}_{Fisher} (\Delta \boldsymbol{a})}$$

$$\left\{\mathbf{I}_{Fisher}\right\}_{ij} = -\int \left(\frac{\partial \ln P(\boldsymbol{a})}{\partial a_i} \frac{\partial \ln P(\boldsymbol{a})}{\partial a_j} \right) P\left(\hat{a} \middle| f\right) df$$

Fisher Information Matrix

$$(\mathbf{I}_{\text{Fisher}})^{-1} = \mathbf{C}_{\boldsymbol{a}\,\text{MML}}$$

MML (Method of Maximum Likelihood):

$$P\left(\hat{\mathbf{a}} \middle| \mathbf{f}^*\right) = \max$$

~

$$\left.\frac{\partial \ln P(a)}{\partial a}\right|_{a=a'} = 0$$

For several Parameters:

$$\nabla \ln P(\boldsymbol{a})\big|_{\hat{\mathbf{a}}} = \boldsymbol{0}$$

Optimality of MML:

a_{MML} - asymptotically Normally distributed vector

a_{MML} - asymptotically jointly effective (most accurate!)

$$\mathbf{C}_{a,\,MML} = \left(\mathbf{I}_{Fisher}\right)^{-1}$$

a_{MML} - asymptotically **the best**

Conditions:

$\boldsymbol{f}^* = \boldsymbol{f}\left(\boldsymbol{a}\right) + \Delta_f$ - physical function

Conditions:

$$\left.\frac{d\boldsymbol{f}\left(\boldsymbol{a}\right)}{d\boldsymbol{a}}\right|_{\boldsymbol{a}}$$ — *derivatives exist and limited in whole range of variability*

$a \in \left]-\infty;+\infty\right[$ $\quad$ $f \in \left]-\infty;+\infty\right[$

Cramer-Rao inequality

g - a characteristic linearly dependent on **a**

i.e. $g = g_1 a_1 + g_2 a_{2+\ldots} = \mathbf{g}^T \mathbf{a}$, **g** - a vector of coefficients)

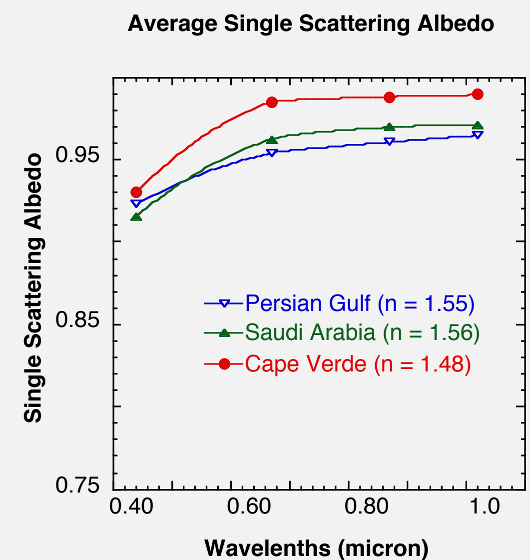
$$\sigma_g^2 = \mathbf{g}\ \mathbf{C}_a \mathbf{g}^T \geq \mathbf{g}\ \mathbf{C}_{a,MML} \mathbf{g}^T = \mathbf{g}\left(\mathbf{I}_{Fisher}\right)^{-1}\mathbf{g}^T$$

Smallest possible

$\mathbf{a}_{MML}=(a_1, a_2,\ldots)^T$ – ***jointly effective !!!***

Very important ***in practice***

For example, ***g*** *=* ***SSA***



Cramer-Rao inequality

g - a characteristic linearly dependent on **a**

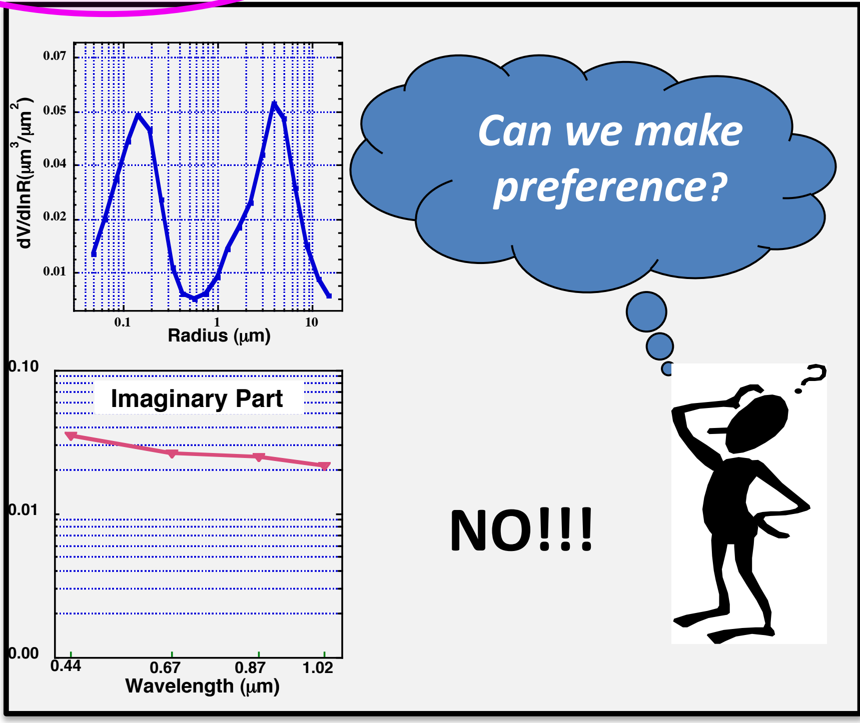
i.e. $g = g_1 a_1 + g_2 a_{2+\dots} = \mathbf{g}^T \mathbf{a}$, **g** - a vector of coefficients)

$$\sigma_g^2 = \mathbf{g}\,\mathbf{C}_a \mathbf{g}^T \geq \mathbf{g}\,\mathbf{C}_{a,MML} \mathbf{g}^T = \mathbf{g}\left(\mathbf{I}_{Fisher}\right)^{-1} \mathbf{g}^T$$

Smallest possible

$\mathbf{a}_{MML} = (a_1, a_2, \dots)^T$ – ***jointly effective !!!***

Very important ***in practice***

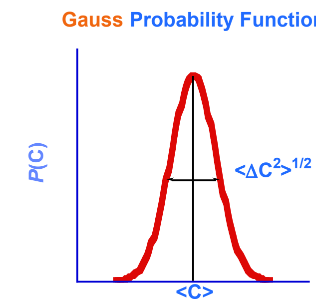


Central Limit Theorem (CLT)

random variables

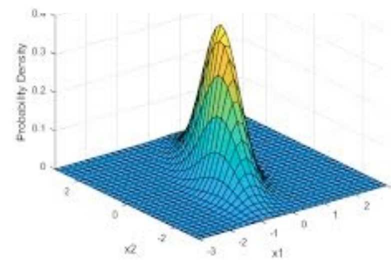
$$s = \Delta_1 + \Delta_2 + \ldots + \Delta_N$$

$$\langle s \rangle = \langle \Delta_1 \rangle + \langle \Delta_2 \rangle + \ldots + \langle \Delta_N \rangle \qquad \langle s^2 \rangle = \langle \Delta_1^2 \rangle + \langle \Delta_2^2 \rangle + \ldots + \langle \Delta_N^2 \rangle$$



$$\frac{s - \sum_{i=1,\ldots,N} \langle \Delta_i \rangle}{\sqrt{\sum_{i=1,\ldots,N} \langle \Delta_i^2 \rangle}} \to N(0,1)$$

is normal distribution with mean 0 and variance 1



CLT: the infinite addition of the independent errors independently of their distributions would result to the normally distributed error

Least Square Method-LSM

MML: $P\left(\hat{\mathbf{a}}\middle|\mathbf{f}^*\right) \sim \exp\left(-\frac{1}{2}\left(\mathbf{f}^* - \mathbf{f}(\mathbf{a})\right)^T \mathbf{C}^{-1}\left(\mathbf{f}^* - \mathbf{f}(\mathbf{a})\right)\right) = \max$

Normal distribution

$$\Psi\left(\mathbf{a}\right) = \frac{1}{2}\left(\mathbf{f}\left(\mathbf{a}\right) - \mathbf{f}^*\right)^T \mathbf{C}^{-1}\left(\mathbf{f}\left(\mathbf{a}\right) - \mathbf{f}^*\right) = \min$$

$$\nabla\Psi\left(\mathbf{a}\right) = \frac{\partial\Psi\left(\mathbf{a}\right)}{\partial a_i} = \mathbf{0}, \qquad (i = 1,..,N_a)$$

$$\nabla\Psi(\hat{a}) = \mathbf{F}^T\mathbf{C}_f^{-1}\mathbf{F} - \mathbf{F}^T\mathbf{C}_f^{-1}f^* = 0$$

$$\hat{a} = \left(\mathbf{F}_1^T\mathbf{C}_1^{-1}\mathbf{F}_1\right)^{-1}\mathbf{F}_1^T\mathbf{C}_1^{-1}f_1^*$$

$$\mathbf{C}_{MML} = \left(\mathbf{F}_1^T\mathbf{C}_1^{-1}\mathbf{F}_1\right)^{-1}$$

Δf – Normal => Δa - Normal

Optimality of LSM:

for **Δf** is Normal: $\mathbf{a}_{LSM} = \mathbf{a}_{MML}$

Moreover: it is optimum in most of practical situation

Potential issues:

Optimality of MML:

$\mathbf{a}_{MML}$ - *asymptotically jointly effective (most accurate!)*

$$\mathbf{C}_{a,\,MML} = \left(\mathbf{I}_{Fisher}\right)^{-1} \quad — \quad \sigma_g^2 = \boldsymbol{g}\ \mathbf{C}_a \boldsymbol{g}^T$$ — ***Smallest possible***

$$g = g_1 a_1 + g_2 a_{2+\dots} = \boldsymbol{g}^T \boldsymbol{a}$$

What if:

$\sigma_g^2 = \boldsymbol{g}\ \mathbf{C}_a \boldsymbol{g}^T{}_{MML}$ too large? or $\det\left(\mathbf{I}_{Fisher}\right) \rightarrow 0$?

↓

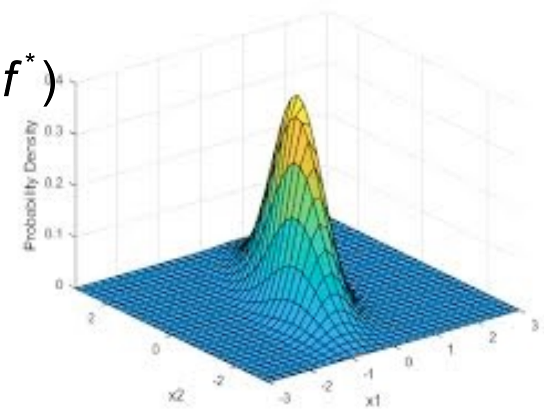
Additional constraints are needed!

Even MML and LSM **may need constraints !**

$$P(f(a)|f^*) = P(f(a) - f^*) = P(\triangle f(a))$$

$$P(f|f^*) \longrightarrow P(\hat{a}|a(f^*)) = P(\hat{a}|f^*)$$

$P(f|f^*)$

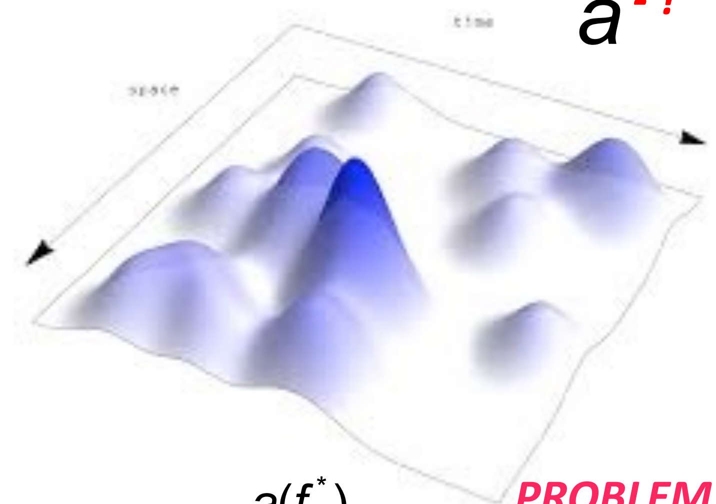


f^*

$\longrightarrow$ $P(\hat{a}|a(f^*))$

2 parameters

$\hat{a}$ - ?



$a(f^*)$

PROBLEM !!!

Need constraints !

$$\mathbf{F}\boldsymbol{a} = \boldsymbol{f}^*$$

Basic idea of inversion

$$\begin{pmatrix} f_1^* \\ f_2^* \end{pmatrix} = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \Rightarrow \boxed{\hat{\boldsymbol{a}} = (\mathbf{F})^{-1} \boldsymbol{f}^*}$$

$\boldsymbol{f}^*$ $\mathbf{F}$ – square; a - parameters of size distribution

$$\begin{pmatrix} f_1^* \\ f_2^* \\ f_3^* \end{pmatrix} = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \\ F_{31} & F_{32} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \Rightarrow \boxed{\hat{\boldsymbol{a}} = \left(\mathbf{F}^T\mathbf{F}\right)^{-1} \mathbf{F}^T \boldsymbol{f}^*}$$

Least Square Method - LSM

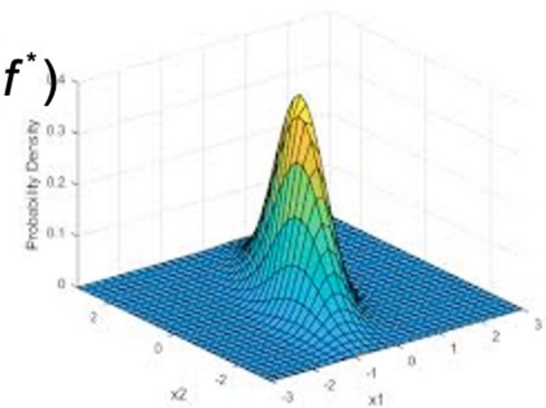
$\boldsymbol{f}^*$ $\mathbf{F}$ – rectangular; a

Basic idea of constrained inversion

$$C_{MML} = \left(\mathbf{F}^T\mathbf{F}\right)^{-1}$$

$$\det\left(\mathbf{F}^T\mathbf{F}\right) \rightarrow 0 \quad \Rightarrow \quad \left(\mathbf{F}^T\mathbf{F}\right)^{-1} - ???$$

$P(f|f^*)$

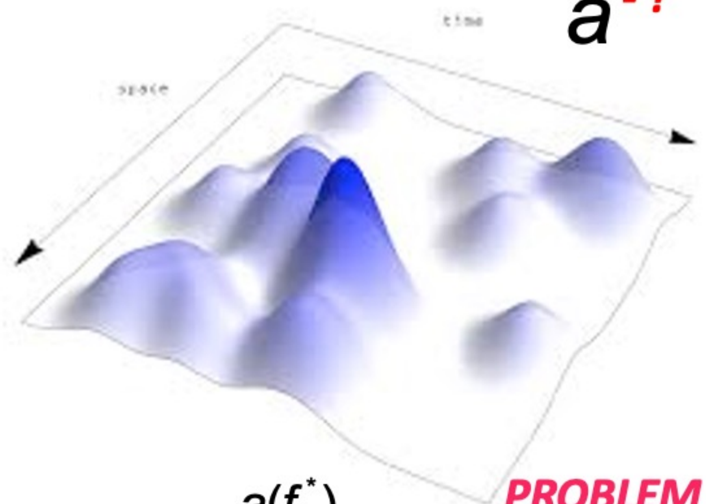


f^*

$P(\hat{a}|a(f^*))$

2 parameters

$\hat{a}$ - ?



$a(f^*)$

PROBLEM !!!

Need constraints !

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}$$

- *Diagonal matrix*

great for inversion !!!

$$\left(\mathbf{F}^T\mathbf{F}\right) \rightarrow \left(\mathbf{F}^T\mathbf{F} + \mathbf{I}\right)$$

$$\det\left(\mathbf{F}^T\mathbf{F} + \mathbf{I}\right) > 0$$

but

$$\left(\mathbf{F}^T\mathbf{F} + \mathbf{I}\right) \neq \left(\mathbf{F}^T\mathbf{F}\right)$$

???

Basic idea of constrained inversion

$$\det\left(\mathbf{F}^T\mathbf{F}\right) \to 0 \quad \Rightarrow \quad \left(\mathbf{F}^T\mathbf{F}\right)^{-1} - ???$$

$$\left(\mathbf{F}^T\mathbf{F}\right) \to \left(\mathbf{F}^T\mathbf{F} + \gamma\,\mathbf{I}\right)$$

$$\det\left(\mathbf{F}^T\mathbf{F} + \gamma\,\mathbf{I}\right) > 0$$ **and** $$\left(\mathbf{F}^T\mathbf{F} + \gamma\,\mathbf{I}\right) \approx \left(\mathbf{F}^T\mathbf{F}\right)$$

$\gamma \to 0$

Basic idea of constrained inversion

$$\det\left(\mathbf{F}^T\mathbf{F}\right) \to 0 \quad \Rightarrow \quad \left(\mathbf{F}^T\mathbf{F}\right)^{-1} - ???$$

$$\hat{\boldsymbol{a}} = \left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T\boldsymbol{f}^*$$

$$\hat{\boldsymbol{a}} = \left(\mathbf{F}^T\mathbf{F} + \gamma\,\mathbf{I}\right)^{-1}\mathbf{F}^T\boldsymbol{f}^*$$

0

Solution is unique and almost correct !!!

Base Inversion Methods

$$P(\Delta \hat{\mathbf{f}}^*) = P(\mathbf{f}(\hat{\mathbf{a}}) - \mathbf{f}^*) = P(\mathbf{f}(\hat{\mathbf{a}}) \mid \mathbf{f}^*) = max.$$

- MML

$$\hat{a} = \left(\mathbf{F}_1^T \mathbf{C}_1^{-1} \mathbf{F}_1\right)^{-1} \mathbf{F}_1^T \mathbf{C}_1^{-1} \boldsymbol{f}_1^*$$

- LSM

Methods of Constrained Inversion

$$\hat{a} = \left(\mathbf{F}^T \mathbf{F} + \gamma \mathbf{I}\right)^{-1} \mathbf{F}^T \boldsymbol{f}^*$$

- Tikhonov Regularization

$$\hat{a} = \left(\mathbf{F}^T \mathbf{C}_f^{-1} \mathbf{F} + \gamma \mathbf{S}^T \mathbf{S}\right)^{-1} \left(\mathbf{F}^T \mathbf{C}_f^{-1} \boldsymbol{f}^*\right)$$

- Phillips-Tikhonov-Twomey

$$\hat{a} = \left(\mathbf{F}^T \mathbf{C}_f^{-1} \mathbf{F} + \mathbf{C}_a^{-1}\right)^{-1} \left(\mathbf{F}^T \mathbf{C}_f^{-1} \mathbf{f}^* + \mathbf{C}_a^{-1}\, a^*\right)$$

- « Optimal estimations », C. Rodgers

LSM for multiple data sets:

P (□□) - Probability Density Function (**Likelihood**)

1. One set of input data:

MML: Method of Maximum Likelihood

$$P \sim \exp\left(-\frac{1}{2}\frac{\Delta \boldsymbol{f}_1^T \Delta \boldsymbol{f}_1}{\sigma^2}\right) = max \quad \longrightarrow \quad \frac{\left(\Delta \boldsymbol{f}_i^T \Delta \boldsymbol{f}_i\right)}{\sigma_1^2} = min$$

2. Multi-Source Data:

$$P_{1,2,3} = P_1 P_2 P_3 \ldots \sim \exp\left(-\frac{1}{2\sigma_1^2}\sum_i \frac{\sigma_1^2}{\sigma_i^2}\left(\Delta \boldsymbol{f}_i^T \Delta \boldsymbol{f}_i\right)\right) = max \quad \longrightarrow \quad \sum_i \frac{\sigma_1^2}{\sigma_i^2}\left(\Delta \boldsymbol{f}_i^T \Delta \boldsymbol{f}_i\right) = min$$

where $\Delta_i = f_i^* - f_i(a)$ and f_i^* - measurements or *a priori data*

Single-sensor data

$$\mathbf{F}\,\boldsymbol{a} = \boldsymbol{f}^{*} = \boldsymbol{f} + \Delta\boldsymbol{f}$$

$$\hat{\boldsymbol{a}} = \left(\mathbf{F}^{T}\mathbf{C}^{-1}\mathbf{F}\right)^{-1}\left(\mathbf{F}^{T}\mathbf{C}^{-1}\boldsymbol{f}^{*}\right)$$

LSM

Multi-sensor data

Independent !!!

sensor 1, *sensor 2*

$$\begin{cases} \boldsymbol{f}_1^{*} = \mathbf{F}_1\,\boldsymbol{a} + \Delta_1 \\ \boldsymbol{f}_2^{*} = \mathbf{F}_2\,\boldsymbol{a} + \Delta_2 \\ \dots \end{cases}$$

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_1 & 0 \\ 0 & \mathbf{C}_2 \end{pmatrix} \qquad \boldsymbol{f} = \begin{matrix} \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \end{matrix} \qquad \mathbf{F} = \begin{matrix} \mathbf{F}_1 \\ \mathbf{F}_2 \end{matrix}$$

Multi-Term LSM

(e.g. see Dubovik 2004)

$$\hat{\boldsymbol{a}} = \left(\mathbf{F}_1^{T}\mathbf{C}_1^{-1}\mathbf{F}_1 + \mathbf{F}_2^{T}\mathbf{C}_2^{-1}\mathbf{F}_2 + \dots\right)^{-1}\left(\mathbf{F}_1^{T}\mathbf{C}_1^{-1}\boldsymbol{f}_1^{*} + \mathbf{F}_2^{T}\mathbf{C}_2^{-1}\boldsymbol{f}_2^{*} + \dots\right)$$

---

sensor, *a priori*

$$\begin{cases} \boldsymbol{f}^{*} = \mathbf{F}\,\boldsymbol{a} + \Delta_f \\ \boldsymbol{a}^{*} = \boldsymbol{a} + \Delta_a \end{cases}$$

$$\hat{\boldsymbol{a}} = \left(\mathbf{F}^{T}\mathbf{C}_f^{-1}\mathbf{F} + \mathbf{C}_a^{-1}\right)^{-1}\left(\mathbf{F}^{T}\mathbf{C}_f^{-1}\boldsymbol{f}^{*} + \mathbf{C}_a^{-1}\,a^{*}\right)$$

Optimal estimations (Rodgers 2000)

Optimal estimations - the most popular and widely used in version strategy in remote sensing

$$\hat{a} = \left(\mathbf{F}^T \mathbf{C}_f^{-1} \mathbf{F} + \mathbf{C}_a^{-1}\right)^{-1} \left(\mathbf{F}^T \mathbf{C}_f^{-1} \boldsymbol{f}^* + \mathbf{C}_a^{-1}\, a^*\right)$$

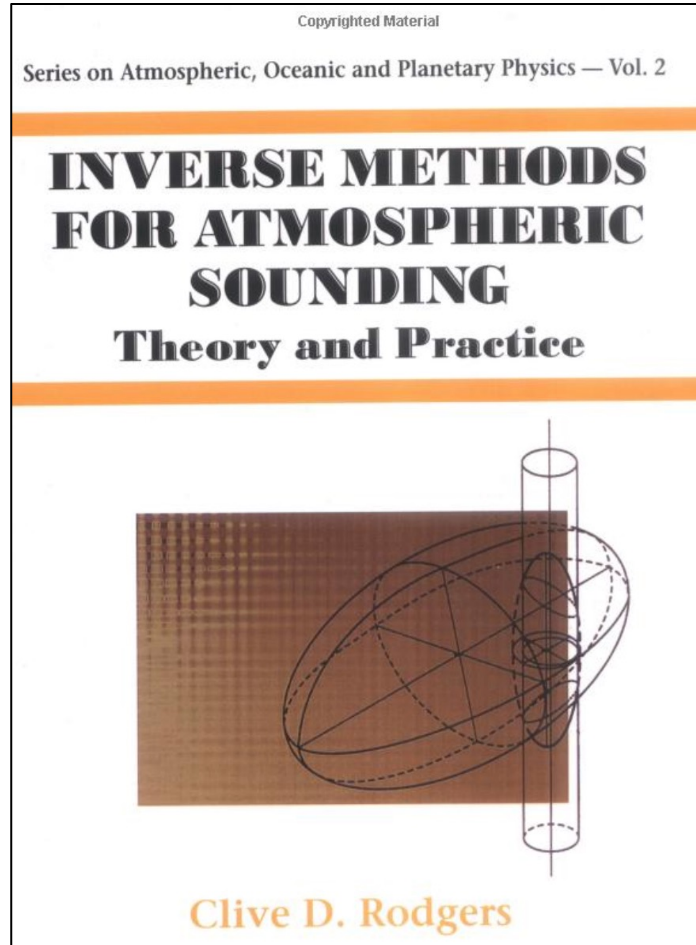
- « Optimal estimations », C. Rodgers

In remote sensing most constrained inversions use this approach for defining a priori constraints

$$\Psi(\hat{\boldsymbol{a}}) = (\boldsymbol{f}^* - \mathbf{F}\hat{\boldsymbol{a}})^T \mathbf{C}_{\boldsymbol{f}}^{-1} (\boldsymbol{f}^* - \mathbf{F}\hat{\boldsymbol{a}}) + (\boldsymbol{a}^* - \hat{\boldsymbol{a}})^T \mathbf{C}_{\boldsymbol{a}}^{-1} (\boldsymbol{a}^* - \hat{\boldsymbol{a}})$$

$$\left\|\boldsymbol{f}^* - \mathbf{F}\boldsymbol{a}\right\|^2 + \gamma \left\|h(\boldsymbol{a}^* - \boldsymbol{a})\right\|^2 = \min$$

- Byaesian Approach



Historical prospective:

Ronald Fisher was a prominent opponent of the Bayesian statistics and introduced Method of Maximum Likelihood as an alternative strategy.

(see Agresti and Hichcock, 2005)

Indeed, if no objective link of the assumed a priori information is established, the Bayesian approach technique becomes vulnerable to subjective assumptions of the developer, which contradicts somewhat the principle of scientific objectivity.

Ironically, the highly popular Optimum Estimation approach promoted by the textbook Rodgers [2000] often directly associated with MML somewhat promotes the Bayesian ideas.

Ronald Fisher

FRS

Fisher in 1913

Born	Ronald Aylmer Fisher 17 February 1890 London, England, UK
Died	29 July 1962 (aged 72) Adelaide, SA, Australia

LOA LABORATOIRE D'OPTIQUE ATMOSPHÉRIQUE

cnrs

Université de Lille

Multi-sensor data

Multi-Term LSM

(e.g. see Dubovik 2004, 2021)

sensor 1

sensor 2

Independent !

$$\begin{cases} f_1^* = \mathbf{F}_1 \, a + \Delta_1 \\ f_2^* = \mathbf{F}_2 \, a + \Delta_2 \\ \ldots \end{cases}$$

$$\hat{a} = \left(\mathbf{F}_1^T \mathbf{C}_1^{-1} \mathbf{F}_1 + \mathbf{F}_2^T \mathbf{C}_2^{-1} \mathbf{F}_2 + \ldots\right)^{-1} \left(\mathbf{F}_1^T \mathbf{C}_1^{-1} f_1^* + \mathbf{F}_2^T \mathbf{C}_2^{-1} f_2^* + \ldots\right)$$

Single-sensor data

$$\hat{a} = \left(\mathbf{F}^T \mathbf{C}_f^{-1} \mathbf{F} + \mathbf{S}^T \mathbf{C}_{0^*}^{-1} \mathbf{S}\right)^{-1} \left(\mathbf{F}^T \mathbf{C}_f^{-1} f^*\right)$$

sensor

a priori

$$\begin{cases} f_1^{\bullet} = f^* = \mathbf{F} \, a + \Delta_f \\ f_2^{\bullet} = 0^* = \mathbf{S} a + \Delta(\Delta a) \end{cases}$$

Phillips-Tikhonov-Twomey formula

Coefficients of differences/derivatives:

e.g. for second dif. (k=2),

$\Delta^2 = \hat{a}_{i+2} - 2\,\hat{a}_{i+1} + \hat{a}_i$:

$$\mathbf{S}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & \ldots & & & \\ 0 & 1 & -2 & 1 & 0 & \ldots & & \\ 0 & 0 & 1 & -2 & 1 & 0 & \ldots & \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & 0 & 1 & -2 & 1 \end{pmatrix}$$

Multi-Term LSM

$$\hat{a} = \left(F_1^T C_1^{-1} F_1 + F_2^T C_2^{-1} F_2 + \ldots\right)^{-1} \left(F_1^T C_1^{-1} f_1^* + F_2^T C_2^{-1} f_2^* + \ldots\right)$$

Other possibilities of a priori constraining

sensor

GRASP **a priori constraints:**

$$\begin{cases} f^* = F\,a + \Delta_f \\ a^* = a + \Delta_a \\ f_2^* = 0^* = Sa + \Delta(\Delta a) \end{cases}$$

← direct a priori estimates (for $a^* = a + \Delta_a$)

← a priori smoothness restrictions (for $f_2^* = 0^* = Sa + \Delta(\Delta a)$)

$$\hat{a} = \left(F^T C_f^{-1} F + S^T C_{0^*}^{-1} S + C_a^{-1}\right)^{-1} \left(F^T C_f^{-1} f^* + C_a^{-1} a^*\right)$$

Identity ???

???

$$\hat{a} = \left(F^T C_f^{-1} F + C_a^{-1}\right)^{-1} \left(F^T C_f^{-1} f^* + C_a^{-1} a^*\right)$$

Optimal estimations (Rodgers 2000)

Single - Pixel Retrieval:

O. Dubovik
M. Herman
J.-L. Deuzé
F. Ducos
D. Tanré

f_j^* - PARASOL data:

Angular measurements (~15 angles) of

- ***Intensity*** (λ = 0.49; 0.67; 0.87; 1.02 □m)
- ***Polarization*** (λ = 0.49; 0.67; 0.87 □m)

a_j - Parameters to be retrieved:

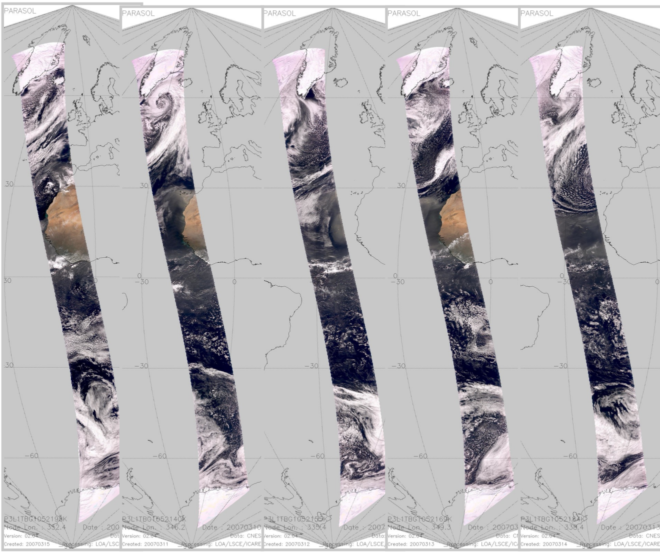
-***Aerosol*** propetries:
- size distribution; - real refractive index
- imaginary refractive index; - particle shape, - height

-***Surface*** properties (***over land***):
- BRF parameters; - BPRF parameters

!!!

$$\begin{cases} f_j^* \\ 0_j^* \end{cases} = \begin{pmatrix} \mathbf{F}_j \\ \mathbf{S}_j \end{pmatrix} \mathbf{a} + \begin{matrix} \Delta_j^m \\ \Delta_j^a \end{matrix}$$

PARASOL



A Priori Constraints limiting derivatives (e.g. *Dubovik 2004*) of

- for aerosols (e.g. in AERONET, *Dubovik and King 2000)* ***:***
- aerosol size distribution variability over size range;
- spectral variability of complex refractive index;

- for surface (e.g. in AERONET/satellite retrievals, *Sinuyk et al. 2007)* ***:***
- spectral variability of BRF/ PBRF parameters.

Multi-term LSM statistically optimized ***Solution*** (*Dubovik and King 2000, Dubovik 2004*) **:**

$$\mathbf{a}_j = \left(\mathbf{F}_j^T \mathbf{W}_j^{-1} \mathbf{F}_j + \gamma_j \Omega_j\right)^{-1} \left(\mathbf{F}_j^T \mathbf{W}_j^{-1} \mathbf{f}_j^*\right)$$

,where $\Omega_j = \mathbf{S}_j^T \mathbf{S}_j$; $\mathbf{W}_i = \frac{1}{\varepsilon_f^2} \mathbf{C}_f$; $\gamma_j = \frac{\varepsilon_f^2}{\varepsilon_a^2}$

Need to be inverted:

Satellite data processing POLDER

$$\boldsymbol{f}^* = \boldsymbol{f}(\boldsymbol{a}) + \Delta_f$$

- Observed atmospheric radiation (non-linear function !!!)

$$\boldsymbol{a} = \begin{pmatrix} \boldsymbol{a}_v & \boldsymbol{a}_n & \boldsymbol{a}_k & a_{sph} & a_{Vc} & a_h & \mathbf{a}_{brdf,1} & \mathbf{a}_{brdf,2} & \mathbf{a}_{brdf,3} & \mathbf{a}_{bpdf} \end{pmatrix}^T$$

- *size distribution;*
- *$n(\lambda)$*
- *$k(\lambda)$*
- *fraction of sphericity;*
- *aerosol height*
- *parameters of BRDF and BPRDF*

Array matrices:

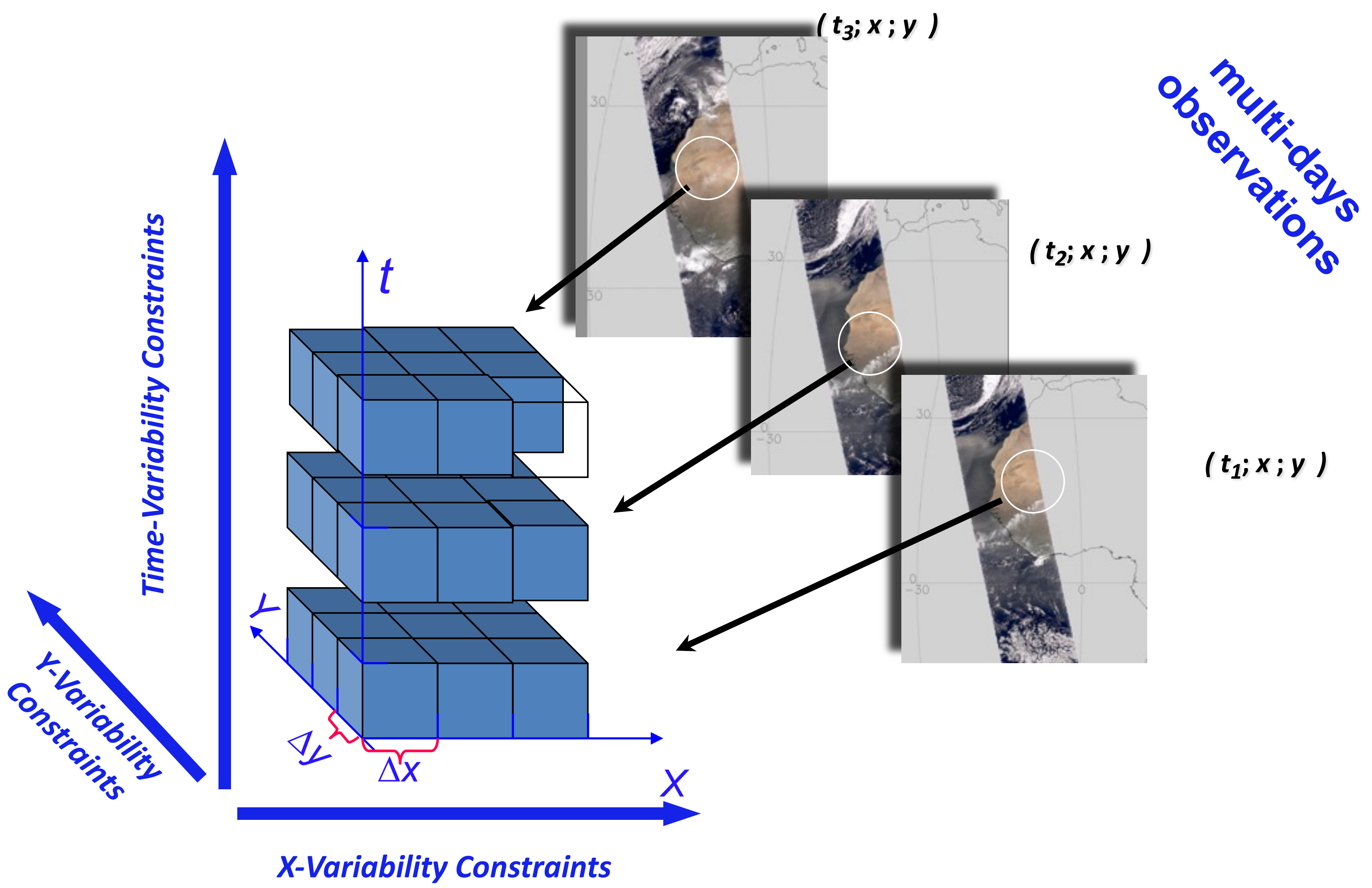
$$\begin{cases} \mathbf{f}_j^* \\ 0_j^* \end{cases} = \begin{pmatrix} \mathbf{F}_j \\ \mathbf{S}_j \end{pmatrix} \mathbf{a}_j + \begin{matrix} \Delta_j^m \\ \Delta_j^a \end{matrix}$$

$$\mathbf{Sa} = \begin{pmatrix} \mathbf{S}_v & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{S}_n & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{S}_k & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{S}_{brdf,1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{S}_{brdf,2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{S}_{brdf,3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{S}_{bpdf} \end{pmatrix} \begin{pmatrix} \mathbf{a}_v \\ \mathbf{a}_n \\ \mathbf{a}_n \\ a_h \\ a_{sph} \\ a_{Vc} \\ \mathbf{a}_{brdf,1} \\ \mathbf{a}_{brdf,2} \\ \mathbf{a}_{brdf,3} \\ \mathbf{a}_{bpdf} \end{pmatrix}$$

$$\mathbf{a}_j = \left(\mathbf{F}_j^T \mathbf{W}_j^{-1} \mathbf{F}_j + \gamma_j \Omega_j\right)^{-1} \left(\mathbf{F}_j^T \mathbf{W}_j^{-1} \mathbf{f}_j^*\right)$$

$$\gamma_\Delta \Omega = \begin{pmatrix} \gamma_{\Delta,1}\Omega_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma_{\Delta,2}\Omega_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_{\Delta,3}\Omega_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \gamma_{\Delta,4}\Omega_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma_{\Delta,5}\Omega_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma_{\Delta,6}\Omega_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma_{\Delta,7}\Omega_7 \end{pmatrix}$$

The concept of multi-pixel retrieval



Multi - Pixel Retrieval:

$$
\begin{cases}
\begin{pmatrix} f_1^* \\ O_1^* \end{pmatrix} \\
\begin{pmatrix} f_2^* \\ O_2^* \end{pmatrix} \\
\begin{pmatrix} f_3^* \\ O_3^* \end{pmatrix} \\
\dots \\
O_t^* \\
O_x^* \\
O_y^*
\end{cases}
=
\begin{pmatrix}
\begin{pmatrix} \mathbf{F}_1 \\ \mathbf{S}_1 \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\
\begin{matrix} 0 \\ 0 \end{matrix} & \begin{pmatrix} \mathbf{F}_2 \\ \mathbf{S}_2 \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\
\begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} & \begin{pmatrix} \mathbf{F}_3 \\ \mathbf{S}_3 \end{pmatrix} \\
\dots & \dots & \dots \\
\mathbf{S}_{t,1} & \mathbf{S}_{t,2} & \mathbf{S}_{t,2} \\
\mathbf{S}_{x,1} & \mathbf{S}_{x,2} & \mathbf{S}_{x,3} \\
\mathbf{S}_{y,1} & \mathbf{S}_{y,2} & \mathbf{S}_{y,3}
\end{pmatrix}
\begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{pmatrix}
+
\begin{pmatrix}
\begin{pmatrix} \Delta_1^m \\ \Delta_1^a \end{pmatrix} \\
\begin{pmatrix} \Delta_2^m \\ \Delta_2^a \end{pmatrix} \\
\begin{pmatrix} \Delta_3^m \\ \Delta_3^a \end{pmatrix} \\
\Delta_t^a \\
\Delta_x^a \\
\Delta_y^a
\end{pmatrix}
$$

Single-Pixel Data (PARASOL measurements and physical a priori constraints) ***are used by the same way as in Single-Pixel retrieval.***

Multi-Pixel a priori constraints (*e.g.Dubovik et al. 2008*):

- limited ***spatial*** variability of each aerosol /surface parameter
- limited ***temporal*** variability of each aerosol /surface parameter

NOTE: *degree of variability constraints (smoothnes) can be different and adequately chosen for each parameter*

Multi-term LSM ***Multi-Pixel Solution:***

$$
\begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{pmatrix}
=
\left[
\begin{pmatrix}
\mathbf{F}_1^T \mathbf{W}_1^{-1} \mathbf{F}_1 & 0 & 0 \\
0 & \mathbf{F}_2^T \mathbf{W}_2^{-1} \mathbf{F}_2 & 0 \\
0 & 0 & \mathbf{F}_3^T \mathbf{W}_3^{-1} \mathbf{F}_3
\end{pmatrix}
+
\begin{pmatrix}
\gamma_1 \Omega_1 & 0 & 0 \\
0 & \gamma_2 \Omega_2 & 0 \\
0 & 0 & \gamma_3 \Omega_3
\end{pmatrix}
+ \gamma_x \Omega_x + \gamma_y \Omega_y + \gamma_t \Omega_t
\right]^{-1}
\begin{pmatrix}
\mathbf{F}_1^T \mathbf{W}_1^{-1} \Delta \mathbf{f}_1^p \\
\mathbf{F}_2^T \mathbf{W}_2^{-1} \Delta \mathbf{f}_2^p \\
\mathbf{F}_3^T \mathbf{W}_3^{-1} \Delta \mathbf{f}_3^p
\end{pmatrix}
$$

,where $\Omega_x = \mathbf{S}_x^T \mathbf{S}_x;\ \Omega_y = \mathbf{S}_y^T \mathbf{S}_y;\ \Omega_t = \mathbf{S}_t^T \mathbf{S}_t;\ \gamma_x = \frac{\varepsilon_f^2}{\varepsilon_x^2};\ \gamma_y = \frac{\varepsilon_f^2}{\varepsilon_y^2};\ \gamma_t = \frac{\varepsilon_f^2}{\varepsilon_t^2}$

Multi-pixel constraints array matrices:

$$\Omega_x = \mathbf{s}_x^T \mathbf{s}_x; \quad \Omega_y = \mathbf{s}_y^T \mathbf{s}_y; \quad \Omega_t = \mathbf{s}_t^T \mathbf{s}_t;$$

$$\mathbf{s}_x^T \mathbf{s}_x = \begin{pmatrix} \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D} \end{pmatrix} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \\ \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} & \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D} \end{pmatrix} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \\ \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} & \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D} \end{pmatrix} \end{pmatrix};$$

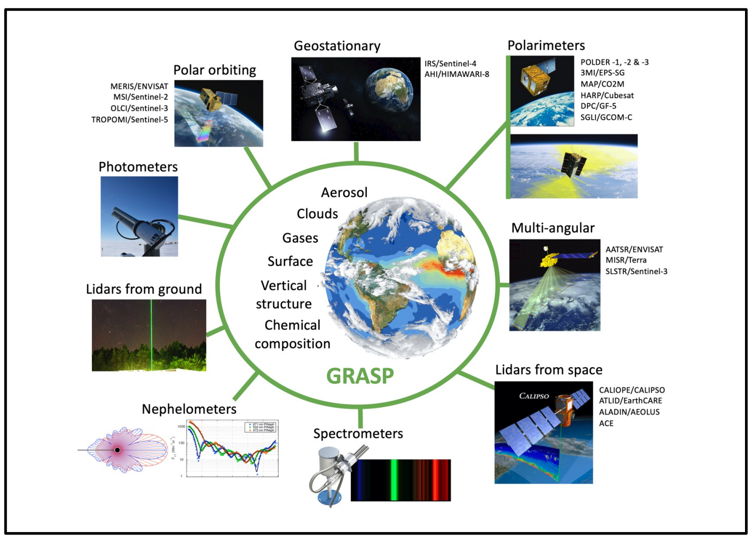
$$\mathbf{s}_y^T \mathbf{s}_y = \begin{pmatrix} \begin{pmatrix} \mathbf{I}_{d_{11}} & \mathbf{I}_{d_{12}} & \mathbf{I}_{d_{13}} \\ \mathbf{I}_{d_{21}} & \mathbf{I}_{d_{22}} & \mathbf{I}_{d_{23}} \\ \mathbf{I}_{d_{31}} & \mathbf{I}_{d_{32}} & \mathbf{I}_{d_{33}} \end{pmatrix} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \\ \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} & \begin{pmatrix} \mathbf{I}_{d_{11}} & \mathbf{I}_{d_{12}} & \mathbf{I}_{d_{13}} \\ \mathbf{I}_{d_{21}} & \mathbf{I}_{d_{22}} & \mathbf{I}_{d_{23}} \\ \mathbf{I}_{d_{31}} & \mathbf{I}_{d_{32}} & \mathbf{I}_{d_{33}} \end{pmatrix} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \\ \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} & \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} & \begin{pmatrix} \mathbf{I}_{d_{11}} & \mathbf{I}_{d_{12}} & \mathbf{I}_{d_{13}} \\ \mathbf{I}_{d_{21}} & \mathbf{I}_{d_{22}} & \mathbf{I}_{d_{23}} \\ \mathbf{I}_{d_{31}} & \mathbf{I}_{d_{32}} & \mathbf{I}_{d_{33}} \end{pmatrix} \end{pmatrix};$$

$$\mathbf{s}_t^T \mathbf{s}_t = \begin{pmatrix} \begin{pmatrix} \mathbf{I}_{d_{11}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_{11}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{d_{11}} \end{pmatrix} & \begin{pmatrix} \mathbf{I}_{d_{12}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_{12}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{d_{12}} \end{pmatrix} & \begin{pmatrix} \mathbf{I}_{d_{13}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_{13}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{d_{13}} \end{pmatrix} \\ \begin{pmatrix} \mathbf{I}_{d_{21}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_{21}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{d_{21}} \end{pmatrix} & \begin{pmatrix} \mathbf{I}_{d_{22}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_{22}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{d_{22}} \end{pmatrix} & \begin{pmatrix} \mathbf{I}_{d_{23}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_{23}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{d_{23}} \end{pmatrix} \\ \begin{pmatrix} \mathbf{I}_{d_{31}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_{31}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{d_{31}} \end{pmatrix} & \begin{pmatrix} \mathbf{I}_{d_{32}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_{32}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{d_{32}} \end{pmatrix} & \begin{pmatrix} \mathbf{I}_{d_{33}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_{33}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{d_{33}} \end{pmatrix} \end{pmatrix}.$$

Key messages:

- **Statistical optimization** is a fundamental basis for overall theory of inverse problems
- **Optimal Estimation** (Rodgers, 2000) – very popular concept in remote sensing inversion, while it has methodological limitations
- **Multi-Term LSM** – efficient concept for constraining inversion
  – proposes additional potential for designing inversion compare to **Optimal Estimation**

GRASP: Generalized Retrieval of Atmosphere and Surface Properties



Dubovik et al. "A Comprehensive Description of Multi-Term LSM for Applying Multiple a Priori Constraints in Problems of Atmospheric Remote Sensing: GRASP Algorithm, Concept, and Applications", Front. Remote Sens., 2021

GRASP is advanced algorithm for retrieval of aerosol, gas and surface properties from diverse remote sensing observations and any combination of them based on:

Forward Model for rigorous simulation of atm. radiation.

+

Inversion with applying **multiple *a priori* constraints**

Thank you!