Multidimensional Global Sensitivity Analysis for Aircraft Infrared Signature Models with Dependent Inputs

S. Lefebvre ONERA - sidonie.lefebvre@onera.fr

J.P. Gauchi INRA MIA

SAMO 2013

ONERA

THE FRENCH AEROSPACE LAB

return on innovation

DOTA

Outline

- Context
- Metamodel: PLS Regression
- DOE choice
- Sensitivity indices
- Results
- Concluding remarks

Aircraft IRS

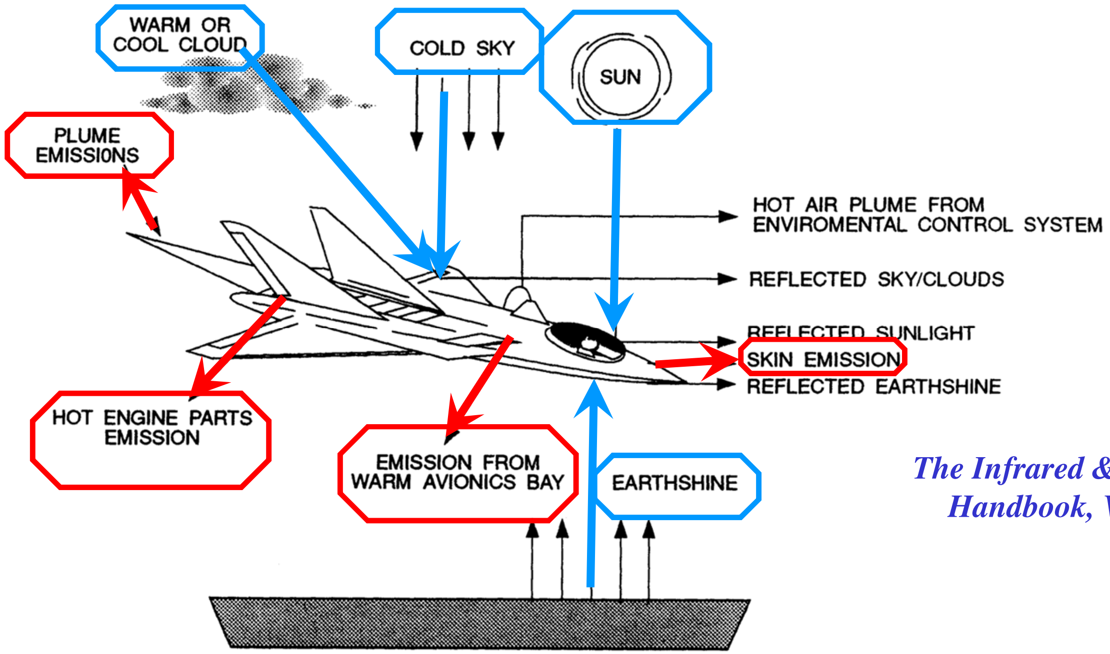
- Context: optimization of a **multispectral** optronics sensor

  sensor must detect aircraft far ahead

- Computer program CRIRA => aircraft IRS according to aircraft properties

  weather conditions

  attack profiles



The Infrared & Electro-Optical Systems Handbook, Vol 7, Countermeasure Systems

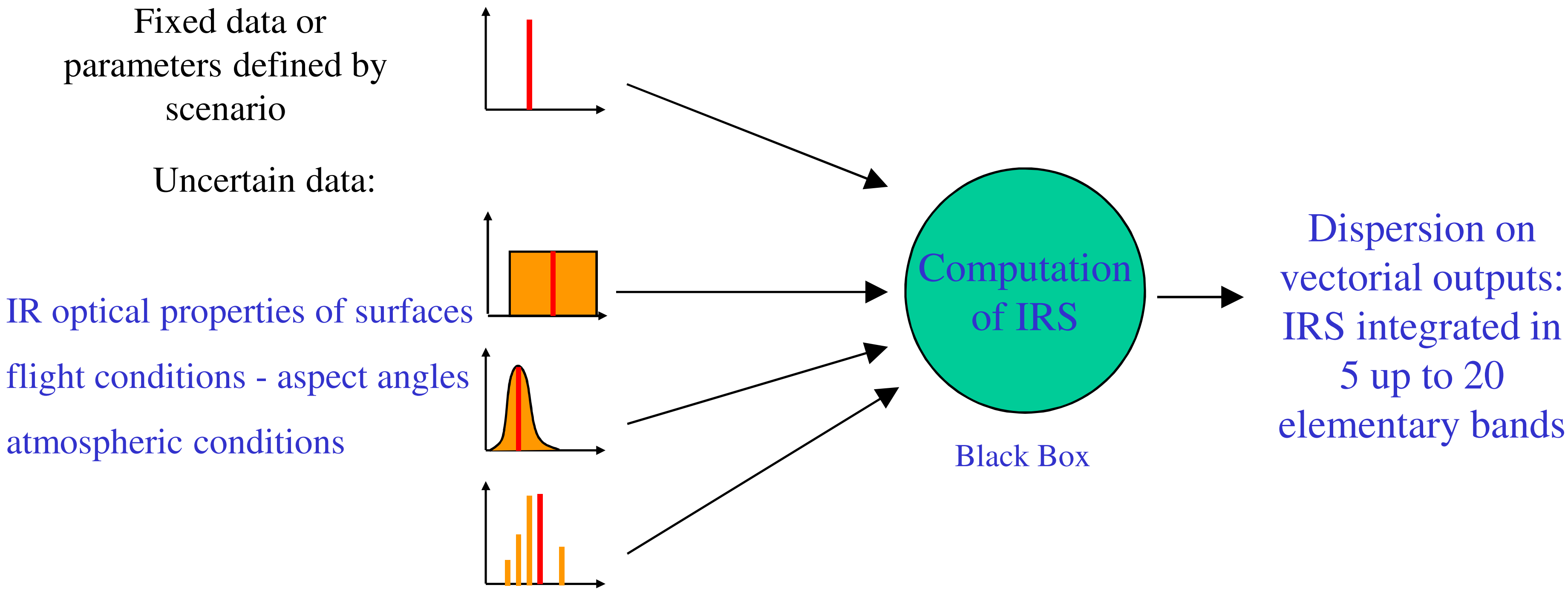
Several contributions to IRS: heat source emission

airframe reflected light from background

DOTA
ONERA

Uncertainty on input data for a given scenario

- Several data types:



Take IRS dispersion into account to estimate optronics sensor properties

➡ **Sensitivity Analysis: most important input variables**

to acquire as a priority

Constraints for Sensitivity Analysis

26 input variables: 23 continuous – 3 categorical

Four constraints: 1. Number of simulation runs must be **small** (<1000)

2. Some **qualitative inputs**: MODEL, CLOUDS, IHAZE

3. Some **correlated quantitative inputs**: RH, TA, HBASE

4. **Correlated multidimensional outputs** (5 up to 20)

⇒Use of a metamodel based on a small computer experimental design => estimation of sensitivity indices

Many input variables + correlations + multidimensional outputs
=> Choice of PLS regression for this study

- Context
- Metamodel: PLS Regression
- DOE choice
- Sensitivity indices
- Results
- Concluding remarks

Metamodel

Constraints 2-3-4 are taken into account by the use of a PLS regression metamodel

Here we assume that the outputs Y can be approximated by an incomplete polynomial model of degree 3 $\boldsymbol{M}$, with well chosen monomials:

26 variables – **381 monomials** specified by the infrared signature expert

- the 23 original continuous inputs,
- their 23 squared terms,
- their 23 cubic terms,
- all the 253 two-input interaction terms between continuous inputs
- the 3 categorical inputs coded with their (0/1)-indicator variables
- 54 two-input interactions terms between categorical and continuous inputs

$\Rightarrow$ for a simulation run i with input variables $X_i = (X_{i,1}, \ldots, X_{i,26})$: $Y_i = M(X_i) + \gamma_i$

γ_i is a metamodel error

PLS Regression

PLS regression = bilinear method for relating *d* inputs to *N* outputs (Tenenhaus, Gauchi, Menardo 1995)

$$Y_{I\times N} = X_{I\times d}\beta_{d\times N} + \varepsilon_{I\times N}$$

Multivariate output matrix — Input matrix — Coefficients matrix — Error terms matrix

<u>Principle</u>: carry out a PCA of the set of inputs X_j, $j=1..d$ subject to the constraint that the orthogonal principal components t_h are as explanatory as possible of the set of outputs variables Y_k, $k=1..N$

=>NIPALS iterative algorithm (Tenenhaus 1998)

The significant number *H* of principal components is obtained thanks to a specific cross-validation test (Lazraq, Cléroux, Gauchi 2003)

⇒ We obtain $\hat{Y}_i = X_{i\times d}\hat{\beta}_{d\times N}$

PLS estimation of β

And thus $Y_i = \widehat{M}(X_i) + \varepsilon_i + \gamma_i$

- Context
- Metamodel: PLS Regression
- DOE choice
- Sensitivity indices
- Results
- Concluding remarks

DOE Choice

Constraint 1 is taken into account thanks to the construction of a small D-optimal computer experimental design *P*

Step 1: We consider a large ~ 20000 network of candidate simulations (space filling desing):

- LHS for all quantitative variables except for meteorological ones
- non parametric reconstruction for the distribution of VIS and CLOUDS
- discrete distribution for the qualitative variables MODEL and IHAZE

- RH, TA and HBASE correlated => distribution estimation from 3200 measured data

Non parametric kernel reconstruction of marginal distributions + dependence modeling based on a Normal copula *C* (Nelsen 2006) using **OpenTURNS**

Cumulative distribution fonction $C\left(F_1\left(X_{RH}\right), F_2\left(X_{TA}\right), F_3\left(X_{HBASE}\right)\right)$

F_1, F_2, F_3 marginal distributions estimated by Kernel Smoothing

$$C\left(u_1, u_2, u_3\right) = \Phi_{\Sigma}^{3}\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \Phi^{-1}(u_3)\right)$$

u_i uniform law on [0,1]

Φ normal CDF – dim 1

Φ_{Σ}^{3} multivariate normal CDF

DOE Choice

Step 2: We build a sequence of D-optimal designs with size N_P increasing from 381 (monomials nb) up to 1000

D-optimal design: max $\det(X^T X)/{N_P}^{381}$ - obtained thanks to Fedorov exchange algorithm

The size of the final design *P* is chosen as a trade off between a moderate size, and a good value of the normalized determinant of the information matrix X^TX associated to *M*

⇒ Final size = 400

We perform the 400 simulations, collect the multivariate outputs

Q_2 ~0.7 for the different outputs – ok for Sensitivity Analysis

to improve for estimation of IRS dispersion

=> Estimation of sensitivity indices

- Context
- Metamodel: PLS Regression
- DOE choice
- Sensitivity indices
- Results
- Concluding remarks

SIVIP sensitivity indices

VIP (Variable Importance Projection) statistics (Tenenhaus 1998):

$$Rd(Y, t_h) = \frac{1}{N}\sum_{k=1}^{N} cor^2(Y_k, t_h)$$

$$Rd(Y, t_1, \dots, t_H) = \sum_{h=1}^{H} \frac{1}{N} \sum_{k=1}^{N} cor^2(Y_k, t_h)$$

For H crossvalidated components a VIP is defined for each monomial X_j

It represents the monomial contribution to Y variance

$$VIP_{Hj} = \left[\frac{381}{Rd(Y, t_1, \dots, t_H)} \sum_{h=1}^{H} Rd(Y, t_h) {w_{hj}}^2\right]^{1/2}$$

j component of the w_h eigenvector

Interesting property: $\sum_{j=1}^{381} {VIP_{Hj}}^2 = 381$

SIVIP sensitivity indices

1. ISIVIP (Individual Sensitivity Indices VIP) for each monomial X_j

$$ISIVIP_j = \frac{{VIP_{Hj}}^2}{381} \qquad \sum_{j=1}^{381} ISIVIP_j = 1$$

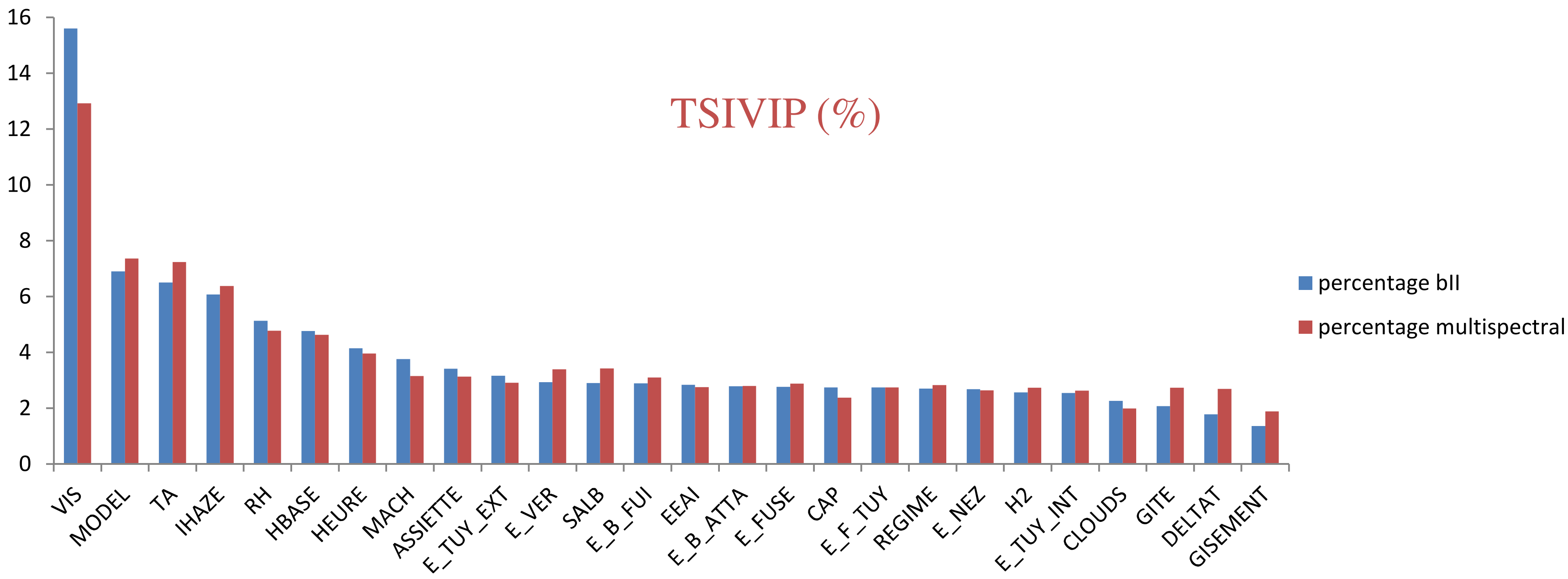
2. TSIVIP Total Sensitivity Indices for each original input variable:

$$TSIVIP_i = \sum_{u=1}^{J} ISIVIP_{\Omega_{iu}}$$

u^{th} index set where the i index is present (total number of such set = J)

- Context
- Metamodel: PLS Regression
- DOE choice
- Sensitivity indices
- Results
- Concluding remarks

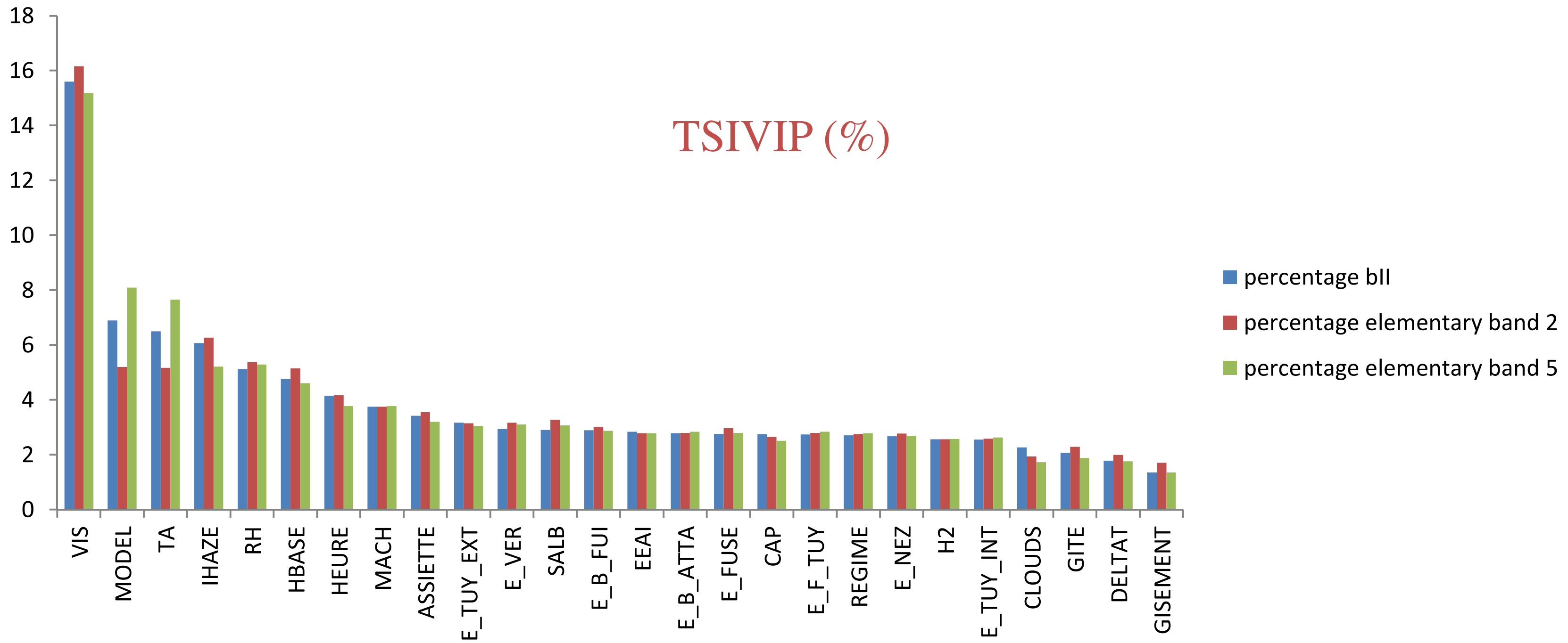
Results: integrated and multispectral IRS



Integrated IRS: About **10 really important variables – 7 first associated to meteorological conditions**
The 4 most important are meteorological ones
Multispectral IRS: Same 7 most important variables
some differences / integrated IRS for the 2 over important ones (SALB – E_ver)

=> If we want to reduce IRS uncertainty, we can combine the optics sensor with some detectors that can measure these atmospheric data

Results: 2 elementary bands



Same 7 most important variables in all configurations, but the rankings are different for the 2 elementary bands

☺ **Very easy to consider different selections and merging of bands**

- Context
- Metamodel: PLS Regression
- DOE choice
- Sensitivity indices
- Results
- Concluding remarks

Concluding remarks

- Simple and effective approach well-adapted to our 4 constraints, when the computer simulation can be approximated by a polynomial of degree ≤ 3 (no large nonlinearity)
  R package to appear (person to contact: JP Gauchi)

- Enable to select **10 variables among 26** that have an important impact on IRS variability

- Sensitivity Analysis simultaneously in 10 spectral bands
  Very easy to consider different selections and merging of bands

$\Rightarrow$ Carry through the specification of a multispectral sensor:
metamodel with 10 input variables - uncertainty propagation => IRS dispersion

Work in progress:

- Adaptive construction of the design of experiments
- Adaptive selection of the monomials

Questions ?

sidonie.lefebvre@onera.fr

ONERA

THE FRENCH AEROSPACE LAB

return on innovation

DOTA