Statistical Performances measures - models comparison

L Patryl[a], D. Galeriu[a] ...

[a] Commissariat à l'Energie Atomique, DAM, DIF, F-91297 Arpajon (France)
[b] "Horia Hulubei" Institute for Physics & Nuclear Engineering (Romania)

September, 12th 2011

energie atomique • energies alternatives

1. Statistical performance measure
2. Simple statistical analysis on wheat experiments
3. Conclusions

Introduction.

In order to compare predictions from a model and observations measurements, several statistical performances measures can be used (U.S. Environmental Protection Agency).

Some of these performance measures are:

- the fractional bias (FB)
- the geometric mean bias (MG);
- the normalized mean square error (NMSE);
- the geometric variance (VG)
- the correlation coefficient (R)
- the fraction of predictions within a factor of two of observations (FAC2)

energie atomique • energies alternatives

Introduction.

In order to compare predictions from a model and observations measurements, several statistical performances measures can be used (U.S. Environmental Protection Agency).

Some of these performance measures are:

- the fractional bias (FB)
- the geometric mean bias (MG);
- the normalized mean square error (NMSE);
- the geometric variance (VG)
- the correlation coefficient (R)
- the fraction of predictions within a factor of two of observations (FAC2)

energie atomique • energies alternatives

Introduction.

In order to compare predictions from a model and observations measurements, several statistical performances measures can be used (U.S. Environmental Protection Agency).

Some of these performance measures are:

- the fractional bias (FB)
- the geometric mean bias (MG);
- the normalized mean square error (NMSE);
- the geometric variance (VG)
- the correlation coefficient (R)
- the fraction of predictions within a factor of two of observations (FAC2)

energie atomique • energies alternatives

Introduction.

In order to compare predictions from a model and observations measurements, several statistical performances measures can be used (U.S. Environmental Protection Agency).

Some of these performance measures are:

- the fractional bias (FB)
- the geometric mean bias (MG);
- the normalized mean square error (NMSE);
- the geometric variance (VG)
- the correlation coefficient (R)
- the fraction of predictions within a factor of two of observations (FAC2)

energie atomique • energies alternatives

Introduction.

In order to compare predictions from a model and observations measurements, several statistical performances measures can be used (U.S. Environmental Protection Agency).

Some of these performance measures are:

- the fractional bias (FB)
- the geometric mean bias (MG);
- the normalized mean square error (NMSE);
- the geometric variance (VG)
- the correlation coefficient (R)
- the fraction of predictions within a factor of two of observations (FAC2)

energie atomique • energies alternatives

Introduction.

In order to compare predictions from a model and observations measurements, several statistical performances measures can be used (U.S. Environmental Protection Agency).

Some of these performance measures are:

- the fractional bias (FB)
- the geometric mean bias (MG);
- the normalized mean square error (NMSE);
- the geometric variance (VG)
- the correlation coefficient (R)
- the fraction of predictions within a factor of two of observations (FAC2)

Introduction.

In order to compare predictions from a model and observations measurements, several statistical performances measures can be used (U.S. Environmental Protection Agency).

Some of these performance measures are:

- the fractional bias (FB)
- the geometric mean bias (MG);
- the normalized mean square error (NMSE);
- the geometric variance (VG)
- the correlation coefficient (R)
- the fraction of predictions within a factor of two of observations (FAC2)

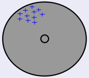
A perfect model would have

MG, VG, R, and FAC2=1.0;

FB and NMSE = 0.0.

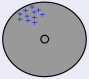
energie atomique • energies alternatives

Systematic errors.



- the systematic bias refers to the ration of Cp to Co
- FB and MG are measures of mean bias and indicate only systematic errors which lead to always underestimate or overestimate the measured values,
- FB is based on a linear scale and the systematic bias refers to the arithmetic difference between Cp and Co,
- MG is based on a logarithmic scale.

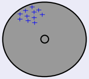
Systematic errors.



- the systematic bias refers to the ration of Cp to Co
- FB and MG are measures of mean bias and indicate only systematic errors which lead to always underestimate or overestimate the measured values,
- FB is based on a linear scale and the systematic bias refers to the arithmetic difference between Cp and Co,
- MG is based on a `logarithmic scale`.

$$FB = \frac{\sum_i (C_{oi} - C_{pi})}{0.5 \sum_i (C_{oi} + C_{pi})} = FB_{FN} - FB_{FP}$$

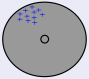
Systematic errors.



- the systematic bias refers to the ration of Cp to Co
- FB and MG are measures of mean bias and indicate only systematic errors which lead to always underestimate or overestimate the measured values,
- FB is based on a linear scale and the systematic bias refers to the arithmetic difference between Cp and Co,
- MG is based on a logarithmic scale.

$$FB = \frac{\sum_i (C_{oi} - C_{pi})}{0.5 \sum_i (C_{oi} + C_{pi})} = FB_{FN} - FB_{FP}$$

Systematic errors.

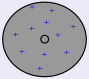


- the systematic bias refers to the ration of Cp to Co
- FB and MG are measures of mean bias and indicate only systematic errors which lead to always underestimate or overestimate the measured values,
- FB is based on a linear scale and the systematic bias refers to the arithmetic difference between Cp and Co,
- MG is based on a `logarithmic scale`.

$$MG = e^{\left(\overline{lnC_o} - \overline{lnC_p}\right)}$$

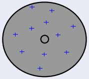
energie atomique • energies alternatives

Systematic and Random errors.



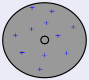
- Random error is due to unpredictable fluctuations We don't have expected value
- Values are scattered about the true value, and tend to have null arithmetic mean when measurement is repeated.
- NMSE and VG are measures of scatter and reflect both systematic and unsystematic (random) errors.

Systematic and Random errors.



- Random error is due to unpredictable fluctuations We don't have expected value
- Values are scattered about the true value, and tend to have null arithmetic mean when measurement is repeated.
- NMSE and VG are measures of scatter and reflect both systematic and unsystematic (random) errors.

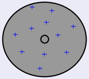
Systematic and Random errors.



- Random error is due to unpredictable fluctuations We don't have expected value
- Values are scattered about the true value, and tend to have null arithmetic mean when measurement is repeated.
- NMSE and VG are measures of scatter and reflect both systematic and unsystematic (random) errors.

$$NMSE = \frac{\overline{(C_o - C_p)^2}}{\left(\overline{C_o}\,\overline{C_p}\right)}$$

Systematic and Random errors.



- Random error is due to unpredictable fluctuations We don't have expected value
- Values are scattered about the true value, and tend to have null arithmetic mean when measurement is repeated.
- NMSE and VG are measures of scatter and reflect both systematic and unsystematic (random) errors.

$$VG = e^{\left(\overline{lnC_o} - \overline{lnC_p}\right)}$$

energie atomique • energies alternatives

Correlation coefficient R.

- Reflects the linear relationship between two variables
- It is insensitive to either an additive or a multiplicative factor
- .A perfect correlation coefficient is only a necessary, but not sufficient, condition for a perfect model.
- For exemple, scatter plot might show generally poor agreement, however, the presence of a good match for a few extreme pairs will greatly improve R.
- to avoid using

$$R = \frac{\overline{\left(C_o - \overline{C_0}\right)\left(C_p - \overline{C_p}\right)}}{\sigma_{C_o}\sigma_{C_p}}$$

Correlation coefficient R.

- Reflects the linear relationship between two variables
- It is insensitive to either an additive or a multiplicative factor
- .A perfect correlation coefficient is only a necessary, but not sufficient, condition for a perfect model.
- For exemple, scatter plot might show generally poor agreement, however, the presence of a good match for a few extreme pairs will greatly improve R.
- to avoid using

$$R = \frac{\overline{\left(C_o - \overline{C_0}\right)\left(C_p - \overline{C_p}\right)}}{\sigma_{C_o}\sigma_{C_p}}$$

Correlation coefficient R.

- Reflects the linear relationship between two variables
- It is insensitive to either an additive or a multiplicative factor
- .A perfect correlation coefficient is only a necessary, but not sufficient, condition for a perfect model.
- For exemple, scatter plot might show generally poor agreement, however, the presence of a good match for a few extreme pairs will greatly improve R.
- to avoid using

$$R = \frac{\overline{\left(C_o - \overline{C_0}\right)\left(C_p - \overline{C_p}\right)}}{\sigma_{C_o}\sigma_{C_p}}$$

Correlation coefficient R.

- Reflects the linear relationship between two variables
- It is insensitive to either an additive or a multiplicative factor
- .A perfect correlation coefficient is only a necessary, but not sufficient, condition for a perfect model.
- For exemple, scatter plot might show generally poor agreement, however, the presence of a good match for a few extreme pairs will greatly improve R.
- to avoid using

$$R = \frac{\overline{\left(C_o - \overline{C_0}\right)\left(C_p - \overline{C_p}\right)}}{\sigma_{C_o}\sigma_{C_p}}$$

energie atomique • energies alternatives

Correlation coefficient R.

- Reflects the linear relationship between two variables
- It is insensitive to either an additive or a multiplicative factor
- .A perfect correlation coefficient is only a necessary, but not sufficient, condition for a perfect model.
- For exemple, scatter plot might show generally poor agreement, however, the presence of a good match for a few extreme pairs will greatly improve R.
- **to avoid using**

$$R = \frac{\overline{\left(C_o - \overline{C_0}\right)\left(C_p - \overline{C_p}\right)}}{\sigma_{C_o}\sigma_{C_p}}$$

FAC2.

- FAC2 is the most robust measure, because it is not overly influenced by high and low outlier.

$$FAC2 = \text{fraction of data that satisfy } 0.5 \leq \frac{C_p}{C_o} \leq 2.0$$

Properties of Performance measures.

- multiple performance measures have to be considered
- Advantages of each performance measure are partly determined by the distribution of the variable
- For a log normal distribution, MG and Vg provide a more balanced treatment of extremely high and low values
- MG and VG would be more appropriate for dataset were both predicted and observed concentrations vary by many orders of magnitude.
- However, MG and VG are strongly influenced by extremely low value and are undefined for zero values $\longrightarrow$ It is necessary to impose a minimum threshold for data which can be the limit of detection (LOD). In this case, if Cp or Co are lower than the threshold, they are set to the LOD
- FB and NMSE are strongly influenced by infrequently occurring high observed and predicted concentration.
- FAC2 is the most robust measure, because it is not overly influenced by high and low outlier.

energie atomique • energies alternatives

Properties of Performance measures.

- multiple performance measures have to be considered
- Advantages of each performance measure are partly determined by the distribution of the variable
- For a log normal distribution, MG and Vg provide a more balanced treatment of extremely high and low values
- MG and VG would be more appropriate for dataset were both predicted and observed concentrations vary by many orders of magnitude.
- However, MG and VG are strongly influenced by extremely low value and are undefined for zero values $\longrightarrow$ It is necessary to impose a minimum threshold for data which can be the limit of detection (LOD). In this case, if Cp or Co are lower than the threshold, they are set to the LOD
- FB and NMSE are strongly influenced by infrequently occurring high observed and predicted concentration.
- FAC2 is the most robust measure, because it is not overly influenced by high and low outlier.

Properties of Performance measures.

- multiple performance measures have to be considered
- Advantages of each performance measure are partly determined by the distribution of the variable
- For a log normal distribution, MG and Vg provide a more balanced treatment of extremely high and low values
- MG and VG would be more appropriate for dataset were both predicted and observed concentrations vary by many orders of magnitude.
- However, MG and VG are strongly influenced by extremely low value and are undefined for zero values $\longrightarrow$ It is necessary to impose a minimum threshold for data which can be the limit of detection (LOD). In this case, if Cp or Co are lower than the threshold, they are set to the LOD
- FB and NMSE are strongly influenced by infrequently occurring high observed and predicted concentration.
- FAC2 is the most robust measure, because it is not overly influenced by high and low outlier.

energie atomique • energies alternatives

Properties of Performance measures.

- multiple performance measures have to be considered
- Advantages of each performance measure are partly determined by the distribution of the variable
- For a log normal distribution, MG and Vg provide a more balanced treatment of extremely high and low values
- MG and VG would be more appropriate for dataset were both predicted and observed concentrations vary by many orders of magnitude.
- However, MG and VG are strongly influenced by extremely low value and are undefined for zero values $\longrightarrow$ It is necessary to impose a minimum threshold for data which can be the limit of detection (LOD). In this case, if Cp or Co are lower than the threshold, they are set to the LOD
- FB and NMSE are strongly influenced by infrequently occurring high observed and predicted concentration.
- FAC2 is the most robust measure, because it is not overly influenced by high and low outlier.

energie atomique • energies alternatives

Properties of Performance measures.

- multiple performance measures have to be considered
- Advantages of each performance measure are partly determined by the distribution of the variable
- For a log normal distribution, MG and Vg provide a more balanced treatment of extremely high and low values
- MG and VG would be more appropriate for dataset were both predicted and observed concentrations vary by many orders of magnitude.
- However, MG and VG are strongly influenced by extremely low value and are undefined for zero values $\longrightarrow$ It is necessary to impose a minimum threshold for data which can be the limit of detection (LOD). In this case, if Cp or Co are lower than the threshold, they are set to the LOD
- FB and NMSE are strongly influenced by infrequently occurring high observed and predicted concentration.
- FAC2 is the most robust measure, because it is not overly influenced by high and low outlier.

energie atomique • energies alternatives

Properties of Performance measures.

- multiple performance measures have to be considered
- Advantages of each performance measure are partly determined by the distribution of the variable
- For a log normal distribution, MG and Vg provide a more balanced treatment of extremely high and low values
- MG and VG would be more appropriate for dataset were both predicted and observed concentrations vary by many orders of magnitude.
- However, MG and VG are strongly influenced by extremely low value and are undefined for zero values $\longrightarrow$ It is necessary to impose a minimum threshold for data which can be the limit of detection (LOD). In this case, if Cp or Co are lower than the threshold, they are set to the LOD
- FB and NMSE are strongly influenced by infrequently occurring high observed and predicted concentration.
- FAC2 is the most robust measure, because it is not overly influenced by high and low outlier.

Properties of Performance measures.

- multiple performance measures have to be considered
- Advantages of each performance measure are partly determined by the distribution of the variable
- For a log normal distribution, MG and Vg provide a more balanced treatment of extremely high and low values
- MG and VG would be more appropriate for dataset were both predicted and observed concentrations vary by many orders of magnitude.
- However, MG and VG are strongly influenced by extremely low value and are undefined for zero values $\longrightarrow$ It is necessary to impose a minimum threshold for data which can be the limit of detection (LOD). In this case, if Cp or Co are lower than the threshold, they are set to the LOD
- FB and NMSE are strongly influenced by infrequently occurring high observed and predicted concentration.
- FAC2 is the most robust measure, because it is not overly influenced by high and low outlier.

energie atomique • energies alternatives

Interpretation of Performance measures.

- FB is symmetrical and bounded; values for the fractional bias range between -2.0 (extreme underprediction) to +2.0 (extreme overprediction)
- The fractional bias is a dimensionless number, which is convenient for comparing the results from studies involving different concentration levels
- Values of the FB that are equal to -0.67 are equivalent to underprediction by a factor of two
- Values of the FB that are equal to +0.67 are equivalent to overprediction by a factor of two
- Model predictions with a fractional bias of 0 (zero) are relatively free from bias
- Values of the MG that are equal to +0.5 are equivalent to underprediction by a factor of two
- values of the MG that are equal to +2 are equivalent to overprediction by a factor of two
- Value of NMSE that are equal to 0.5 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction
- Value of VG that are equal to 1.6 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction

$$\frac{\overline{C_p}}{\overline{C_o}} = \frac{1-0.5FB}{1+0.5FB}$$

energie atomique • energies alternatives

Interpretation of Performance measures.

- FB is symmetrical and bounded; values for the fractional bias range between -2.0 (extreme underprediction) to +2.0 (extreme overprediction)
- The fractional bias is a dimensionless number, which is convenient for comparing the results from studies involving different concentration levels
- Values of the FB that are equal to -0.67 are equivalent to underprediction by a factor of two
- Values of the FB that are equal to +0.67 are equivalent to overprediction by a factor of two
- Model predictions with a fractional bias of 0 (zero) are relatively free from bias
- Values of the MG that are equal to +0.5 are equivalent to underprediction by a factor of two
- values of the MG that are equal to +2 are equivalent to overprediction by a factor of two
- Value of NMSE that are equal to 0.5 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction
- Value of VG that are equal to 1.6 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction

$$\frac{\overline{C_p}}{\overline{C_o}} = \frac{1-0.5FB}{1+0.5FB}$$

Interpretation of Performance measures.

- FB is symmetrical and bounded; values for the fractional bias range between -2.0 (extreme underprediction) to +2.0 (extreme overprediction)
- The fractional bias is a dimensionless number, which is convenient for comparing the results from studies involving different concentration levels
- Values of the FB that are equal to -0.67 are equivalent to underprediction by a factor of two
- Values of the FB that are equal to +0.67 are equivalent to overprediction by a factor of two
- Model predictions with a fractional bias of 0 (zero) are relatively free from bias
- Values of the MG that are equal to +0.5 are equivalent to underprediction by a factor of two
- values of the MG that are equal to +2 are equivalent to overprediction by a factor of two
- Value of NMSE that are equal to 0.5 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction
- Value of VG that are equal to 1.6 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction

$$\frac{\overline{C_p}}{\overline{C_o}} = \frac{1-0.5FB}{1+0.5FB}$$

energie atomique • energies alternatives

Interpretation of Performance measures.

- FB is symmetrical and bounded; values for the fractional bias range between -2.0 (extreme underprediction) to +2.0 (extreme overprediction)
- The fractional bias is a dimensionless number, which is convenient for comparing the results from studies involving different concentration levels
- Values of the FB that are equal to -0.67 are equivalent to underprediction by a factor of two
- Values of the FB that are equal to +0.67 are equivalent to overprediction by a factor of two
- Model predictions with a fractional bias of 0 (zero) are relatively free from bias
- Values of the MG that are equal to +0.5 are equivalent to underprediction by a factor of two
- values of the MG that are equal to +2 are equivalent to overprediction by a factor of two
- Value of NMSE that are equal to 0.5 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction
- Value of VG that are equal to 1.6 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction

$$\frac{\overline{C_p}}{\overline{C_o}} = \frac{1-0.5FB}{1+0.5FB}$$

Interpretation of Performance measures.

- FB is symmetrical and bounded; values for the fractional bias range between -2.0 (extreme underprediction) to +2.0 (extreme overprediction)
- The fractional bias is a dimensionless number, which is convenient for comparing the results from studies involving different concentration levels
- Values of the FB that are equal to -0.67 are equivalent to underprediction by a factor of two
- Values of the FB that are equal to +0.67 are equivalent to overprediction by a factor of two
- Model predictions with a fractional bias of 0 (zero) are relatively free from bias
- Values of the MG that are equal to +0.5 are equivalent to underprediction by a factor of two
- values of the MG that are equal to +2 are equivalent to overprediction by a factor of two
- Value of NMSE that are equal to 0.5 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction
- Value of VG that are equal to 1.6 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction

$$\frac{\overline{C_p}}{\overline{C_o}} = \frac{1-0.5FB}{1+0.5FB}$$

Interpretation of Performance measures.

- FB is symmetrical and bounded; values for the fractional bias range between -2.0 (extreme underprediction) to +2.0 (extreme overprediction)
- The fractional bias is a dimensionless number, which is convenient for comparing the results from studies involving different concentration levels
- Values of the FB that are equal to -0.67 are equivalent to underprediction by a factor of two
- Values of the FB that are equal to +0.67 are equivalent to overprediction by a factor of two
- Model predictions with a fractional bias of 0 (zero) are relatively free from bias
- Values of the MG that are equal to +0.5 are equivalent to underprediction by a factor of two
- values of the MG that are equal to +2 are equivalent to overprediction by a factor of two
- Value of NMSE that are equal to 0.5 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction
- Value of VG that are equal to 1.6 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction

$$\frac{\langle C_p \rangle}{\langle C_o \rangle} = \frac{1}{MG}$$

Interpretation of Performance measures.

- FB is symmetrical and bounded; values for the fractional bias range between -2.0 (extreme underprediction) to +2.0 (extreme overprediction)
- The fractional bias is a dimensionless number, which is convenient for comparing the results from studies involving different concentration levels
- Values of the FB that are equal to -0.67 are equivalent to underprediction by a factor of two
- Values of the FB that are equal to +0.67 are equivalent to overprediction by a factor of two
- Model predictions with a fractional bias of 0 (zero) are relatively free from bias
- Values of the MG that are equal to +0.5 are equivalent to underprediction by a factor of two
- values of the MG that are equal to +2 are equivalent to overprediction by a factor of two
- Value of NMSE that are equal to 0.5 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction
- Value of VG that are equal to 1.6 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction

$$\frac{\langle C_p \rangle}{\langle C_o \rangle} = \frac{1}{MG}$$

energie atomique • energies alternatives

Interpretation of Performance measures.

- FB is symmetrical and bounded; values for the fractional bias range between -2.0 (extreme underprediction) to +2.0 (extreme overprediction)
- The fractional bias is a dimensionless number, which is convenient for comparing the results from studies involving different concentration levels
- Values of the FB that are equal to -0.67 are equivalent to underprediction by a factor of two
- Values of the FB that are equal to +0.67 are equivalent to overprediction by a factor of two
- Model predictions with a fractional bias of 0 (zero) are relatively free from bias
- Values of the MG that are equal to +0.5 are equivalent to underprediction by a factor of two
- values of the MG that are equal to +2 are equivalent to overprediction by a factor of two
- Value of NMSE that are equal to 0.5 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction
- Value of VG that are equal to 1.6 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction

$$\frac{\overline{C_p}}{\overline{C_o}} = \frac{2+NMSE \pm \sqrt{(2+NMSE)^2 - 4}}{2}$$

energie atomique • energies alternatives

Interpretation of Performance measures.

- FB is symmetrical and bounded; values for the fractional bias range between -2.0 (extreme underprediction) to +2.0 (extreme overprediction)
- The fractional bias is a dimensionless number, which is convenient for comparing the results from studies involving different concentration levels
- Values of the FB that are equal to -0.67 are equivalent to underprediction by a factor of two
- Values of the FB that are equal to +0.67 are equivalent to overprediction by a factor of two
- Model predictions with a fractional bias of 0 (zero) are relatively free from bias
- Values of the MG that are equal to +0.5 are equivalent to underprediction by a factor of two
- values of the MG that are equal to +2 are equivalent to overprediction by a factor of two
- Value of NMSE that are equal to 0.5 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction
- Value of VG that are equal to 1.6 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction

$$\frac{\overline{C_p}}{\overline{C_o}} = \frac{2+NMSE \pm \sqrt{(2+NMSE)^2 - 4}}{2}$$

Interpretation of Performance measures.

- FB is symmetrical and bounded; values for the fractional bias range between -2.0 (extreme underprediction) to +2.0 (extreme overprediction)
- The fractional bias is a dimensionless number, which is convenient for comparing the results from studies involving different concentration levels
- Values of the FB that are equal to -0.67 are equivalent to underprediction by a factor of two
- Values of the FB that are equal to +0.67 are equivalent to overprediction by a factor of two
- Model predictions with a fractional bias of 0 (zero) are relatively free from bias
- Values of the MG that are equal to +0.5 are equivalent to underprediction by a factor of two
- values of the MG that are equal to +2 are equivalent to overprediction by a factor of two
- Value of NMSE that are equal to 0.5 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction
- Value of VG that are equal to 1.6 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction

$$\frac{\langle C_p \rangle}{\langle C_o \rangle} = exp[\pm\sqrt{lnVG}]$$

Interpretation of Performance measures.

- FB is symmetrical and bounded; values for the fractional bias range between -2.0 (extreme underprediction) to +2.0 (extreme overprediction)
- The fractional bias is a dimensionless number, which is convenient for comparing the results from studies involving different concentration levels
- Values of the FB that are equal to -0.67 are equivalent to underprediction by a factor of two
- Values of the FB that are equal to +0.67 are equivalent to overprediction by a factor of two
- Model predictions with a fractional bias of 0 (zero) are relatively free from bias
- Values of the MG that are equal to +0.5 are equivalent to underprediction by a factor of two
- values of the MG that are equal to +2 are equivalent to overprediction by a factor of two
- Value of NMSE that are equal to 0.5 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction
- Value of VG that are equal to 1.6 corresponds to an equivalent factor of two mean bias
- It doesn't differentiate whether the factor of two mean bias is underprediction or overprediction

$$\frac{\langle C_p \rangle}{\langle C_o \rangle} = exp[\pm\sqrt{lnVG}]$$

How good is good enough ?

- Fraction of prediction within a factor 2 of observation is about 50% or greanter ($FAC2 > 0.5$)
- The mean bias is within $\pm 30\%$ of the mean ($|FB| < 0.3$ *or* $0.7 < MG < 1.3$)
- Random scatter is about a factor of two to three of the mean ($NMSE < 1.5$ *or* $VG < 4$)

How good is good enough ?

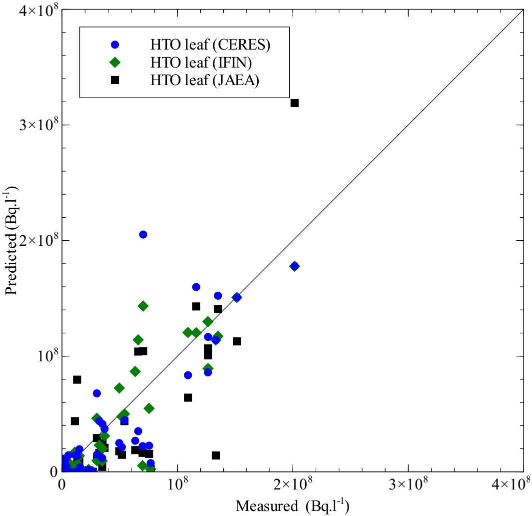
- Fraction of prediction within a factor 2 of observation is about 50% or greanter ($FAC2 > 0.5$)
- The mean bias is within $\pm 30\%$ of the mean ($|FB| < 0.3$ *or* $0.7 < MG < 1.3$)
- Random scatter is about a factor of two to three of the mean ($NMSE < 1.5$ *or* $VG < 4$)

How good is good enough ?

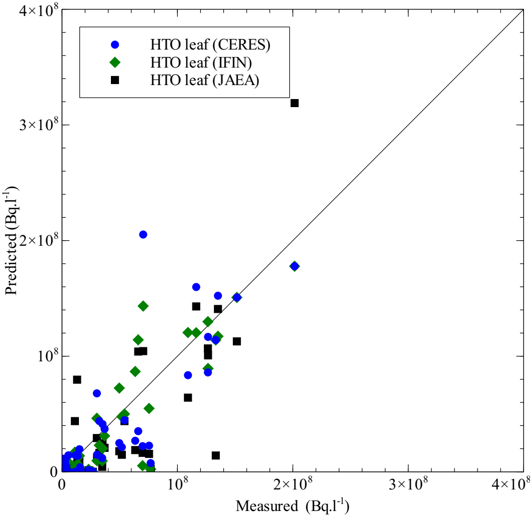
- Fraction of prediction within a factor 2 of observation is about 50% or greanter ($FAC2 > 0.5$)
- The mean bias is within $\pm 30\%$ of the mean ($|FB| < 0.3$ *or* $0.7 < MG < 1.3$)
- Random scatter is about a factor of two to three of the mean ($NMSE < 1.5$ *or* $VG < 4$)

energie atomique • energies alternatives

- Difficult to say which model is better
- Difficult to say if models make overprediction ou underprediction



- Difficult to say which model is better
- Difficult to say if models make overprediction ou underprediction



energie atomique • energies alternatives

- 61 experiments
- 3 models (CEA, JAEA, IFIN)
- Some of values equal 0 →without detection threshold or other informations we use only arithmetic scale (FB and NMSE)
- More than a factor 2 for CEA and JAEA (radom and systematic errors)
- Only about 30% value are within a factor of 2 of observations

Model/Performance (factor 2)	NMSE (0.5)	FB (±2/3)	FAC2	R
CEA	0.7	0.16	0.31	0.858
JAEA	1.13	0.26	0.30	0.818
IFIN	0.42	0.15	0.36	0.912

- 61 experiments
- 3 models (CEA, JAEA, IFIN)
- Some of values equal 0 →without detection threshold or other informations we use only arithmetic scale (FB and NMSE)
- More than a factor 2 for CEA and JAEA (radom and systematic errors)
- Only about 30% value are within a factor of 2 of observations

Model/Performance (factor 2)	NMSE (0.5)	FB (±2/3)	FAC2	R
CEA	0.7	0.16	0.31	0.858
JAEA	1.13	0.26	0.30	0.818
IFIN	0.42	0.15	0.36	0.912

- 61 experiments
- 3 models (CEA, JAEA, IFIN)
- Some of values equal 0 $\rightarrow$without detection threshold or other informations we use only arithmetic scale (FB and NMSE)
- More than a factor 2 for CEA and JAEA (radom and systematic errors)
- Only about 30% value are within a factor of 2 of observations

Model/Performance (factor 2)	NMSE (0.5)	FB ($\pm 2/3$)	FAC2	R
CEA	0.7	0.16	0.31	0.858
JAEA	1.13	0.26	0.30	0.818
IFIN	0.42	0.15	0.36	0.912

energie atomique • energies alternatives

- 61 experiments
- 3 models (CEA, JAEA, IFIN)
- Some of values equal 0 →without detection threshold or other informations we use only arithmetic scale (FB and NMSE)
- More than a factor 2 for CEA and JAEA (radom and systematic errors)
- Only about 30% value are within a factor of 2 of observations

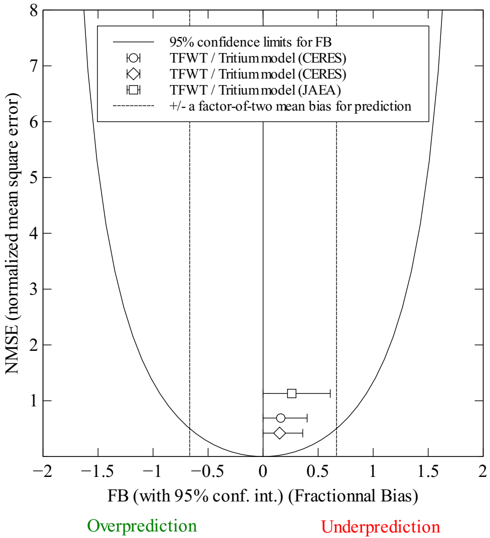
Model/Performance (factor 2)	NMSE (0.5)	FB (±2/3)	FAC2	R
CEA	0.7	0.16	0.31	0.858
JAEA	1.13	0.26	0.30	0.818
IFIN	0.42	0.15	0.36	0.912

- 61 experiments
- 3 models (CEA, JAEA, IFIN)
- Some of values equal 0 $\rightarrow$without detection threshold or other informations we use only arithmetic scale (FB and NMSE)
- More than a factor 2 for CEA and JAEA (radom and systematic errors)
- Only about 30% value are within a factor of 2 of observations

Model/Performance (factor 2)	NMSE (0.5)	FB ($\pm 2/3$)	FAC2	R
CEA	0.7	0.16	0.31	0.858
JAEA	1.13	0.26	0.30	0.818
IFIN	0.42	0.15	0.36	0.912

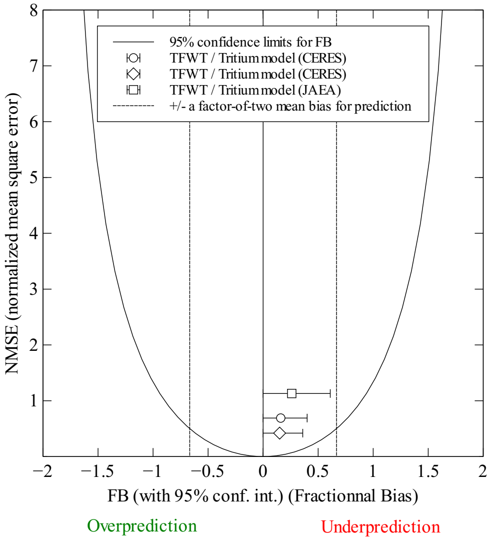
HTO IN WHEAT LEAF (3/3)

- All models tend to underestimate activity in leaf (less than a factor of 2)
- Surely due to very low values

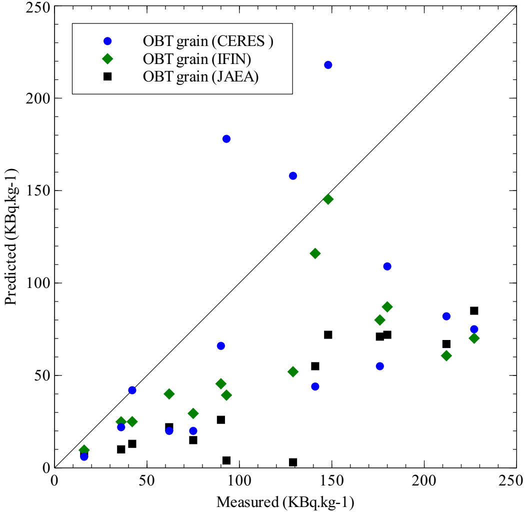


HTO IN WHEAT LEAF (3/3)

- All models tend to underestimate activity in leaf (less than a factor of 2)
- Surely due to very low values

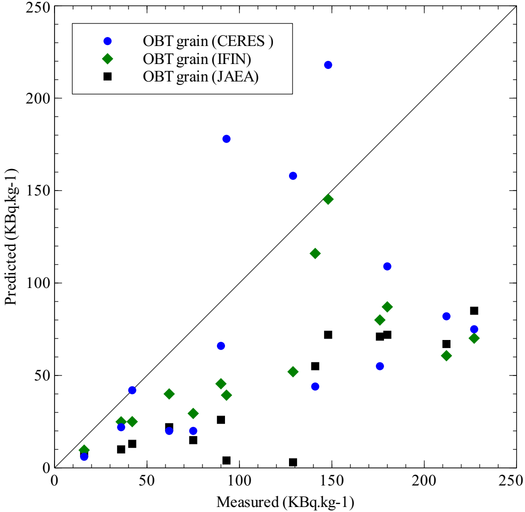


- IFIN and JAEA seems make underprediction OBT at the end of harvest but how much ?
- Difficult to say which model is better



OBT IN GRAIN WHEAT (1/4)

- IFIN and JAEA seems make underprediction OBT at the end of harvest but how much ?
- Difficult to say which model is better



energie atomique • energies alternatives

- 14 experiments at the end or harvest
- 3 models (CEA, JAEA, IFIN)
- Use arithmetic and logarithmic scale →gives about the same results)
- More than a factor 2 for all models (radom and systematic errors)
- All model made underprediction (more than a factor of 2 for JAEA)

Model/Performance (factor 2)	NMSE (0.5)	FB (±2/3)	FAC2	R
CEA	0.7	0.4	0.5	0.41
JAEA	1.8	1.0	0.07	0.86
IFIN	0.8	0.7	0.5	0.66

Model/Performance (factor 2)	VG (1.6)	MG (2.0 or 0.5)	FAC2	R
CEA	2.1	1.8	0.5	0.76
JAEA	15.2	4.0	0.07	0.61
IFIN	1.8	1.9	0.5	0.89

energie atomique • energies alternatives

- 14 experiments at the end or harvest
- 3 models (CEA, JAEA, IFIN)
- Use arithmetic and logarithmic scale $\rightarrow$gives about the same results)
- More than a factor 2 for all models (radom and systematic errors)
- All model made underprediction (more than a factor of 2 for JAEA)

Model/Performance (factor 2)	NMSE (0.5)	FB ($\pm 2/3$)	FAC2	R
CEA	0.7	0.4	0.5	0.41
JAEA	1.8	1.0	0.07	0.86
IFIN	0.8	0.7	0.5	0.66

Model/Performance (factor 2)	VG (1.6)	MG (2.0 or 0.5)	FAC2	R
CEA	2.1	1.8	0.5	0.76
JAEA	15.2	4.0	0.07	0.61
IFIN	1.8	1.9	0.5	0.89

energie atomique • energies alternatives

- 14 experiments at the end or harvest
- 3 models (CEA, JAEA, IFIN)
- Use arithmetic and logarithmic scale →gives about the same results)
- More than a factor 2 for all models (radom and systematic errors)
- All model made underprediction (more than a factor of 2 for JAEA)

Model/Performance (factor 2)	NMSE (0.5)	FB (±2/3)	FAC2	R
CEA	0.7	0.4	0.5	0.41
JAEA	1.8	1.0	0.07	0.86
IFIN	0.8	0.7	0.5	0.66

Model/Performance (factor 2)	VG (1.6)	MG (2.0 or 0.5)	FAC2	R
CEA	2.1	1.8	0.5	0.76
JAEA	15.2	4.0	0.07	0.61
IFIN	1.8	1.9	0.5	0.89

energie atomique • energies alternatives

- 14 experiments at the end or harvest
- 3 models (CEA, JAEA, IFIN)
- Use arithmetic and logarithmic scale $\rightarrow$gives about the same results)
- More than a factor 2 for all models (radom and systematic errors)
- All model made underprediction (more than a factor of 2 for JAEA)

Model/Performance (factor 2)	NMSE (0.5)	FB ($\pm 2/3$)	FAC2	R
CEA	0.7	0.4	0.5	0.41
JAEA	1.8	1.0	0.07	0.86
IFIN	0.8	0.7	0.5	0.66

Model/Performance (factor 2)	VG (1.6)	MG (2.0 or 0.5)	FAC2	R
CEA	2.1	1.8	0.5	0.76
JAEA	15.2	4.0	0.07	0.61
IFIN	1.8	1.9	0.5	0.89

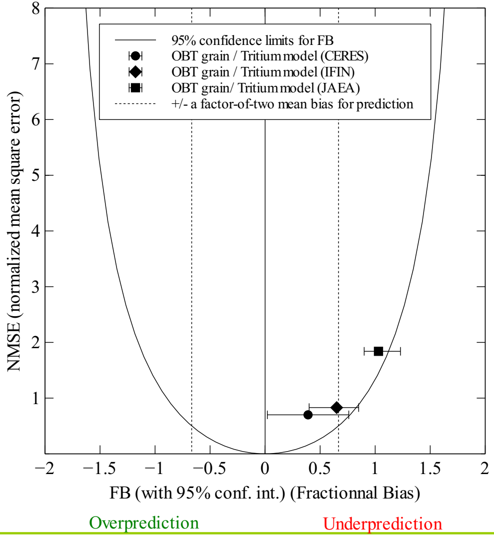
- 14 experiments at the end or harvest
- 3 models (CEA, JAEA, IFIN)
- Use arithmetic and logarithmic scale →gives about the same results)
- More than a factor 2 for all models (radom and systematic errors)
- All model made underprediction (more than a factor of 2 for JAEA)

Model/Performance (factor 2)	NMSE (0.5)	FB (±2/3)	FAC2	R
CEA	0.7	0.4	0.5	0.41
JAEA	1.8	1.0	0.07	0.86
IFIN	0.8	0.7	0.5	0.66

Model/Performance (factor 2)	VG (1.6)	MG (2.0 or 0.5)	FAC2	R
CEA	2.1	1.8	0.5	0.76
JAEA	15.2	4.0	0.07	0.61
IFIN	1.8	1.9	0.5	0.89

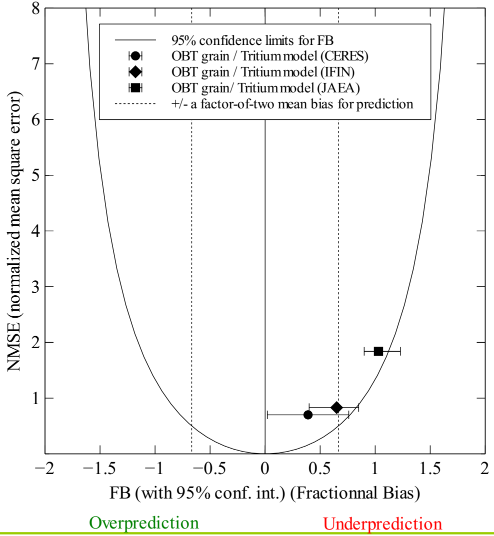
OBT IN GRAIN WHEAT (3/4)

- CEA and IFIN models tend to underestimate activity in leaf (less than a factor of 2), JAEA underestimates about a factor of 3
- Surely due to very low values



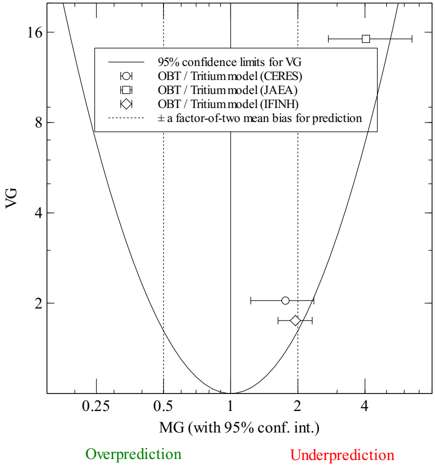
OBT IN GRAIN WHEAT (3/4)

- CEA and IFIN models tend to underestimate activity in leaf (less than a factor of 2), JAEA underestimates about a factor of 3
- Surely due to very low values

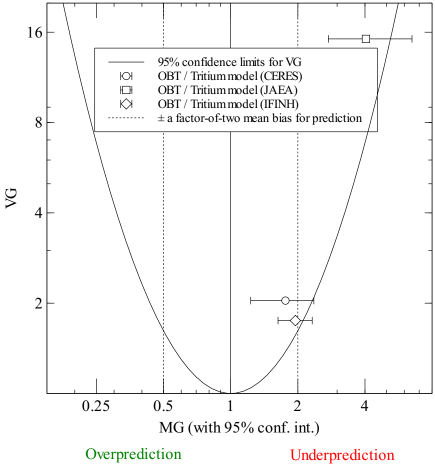


HTO IN WHEAT LEAF (4/4)

- CEA and IFIN models tend to underestimate activity in leaf (less than a factor of 2), JAEA underestimates about a factor of 4
- Random scatter is less than a factor of 3 (CEA, IFIN) and 5 (JAEA)



- CEA and IFIN models tend to underestimate activity in leaf (less than a factor of 2), JAEA underestimates about a factor of 4
- Random scatter is less than a factor of 3 (CEA, IFIN) and 5 (JAEA)



1. Statistical performance measure
2. Simple statistical analysis on wheat experiments
3. Conclusions

energie atomique • energies alternatives

- Statistical analysis can seriously help the models comparison
- Performance measures have to be used to compare predictions to observations
- In case of wheat all models have systematic errors
- HTO modelling in wheat leaf seems good for the 3 models
- Systematic errors : $\left(\frac{\overline{C_p}}{\overline{C_o}} = 0.76(JAEA)\ 0.86(IFIN\&CEA\right)$
- OBT modelling in wheat grain seems make underprediction for all model
- Systematic errors : $\left(\frac{\overline{C_p}}{\overline{C_o}} = 0.3(JAEA)\ 0.48(IFIN)\ 0.7(CEA)\right)$

energie atomique • energies alternatives

- Statistical analysis can seriously help the models comparison
- Performance measures have to be used to compare predictions to observations
- In case of wheat all models have systematic errors
- HTO modelling in wheat leaf seems good for the 3 models
- Systematic errors : $\left(\frac{\overline{C_p}}{C_o} = 0.76(JAEA)\ 0.86(IFIN\&CEA)\right)$
- OBT modelling in wheat grain seems make underprediction for all model
- Systematic errors : $\left(\frac{\overline{C_p}}{C_o} = 0.3(JAEA)\ 0.48(IFIN)\ 0.7(CEA)\right)$

- Statistical analysis can seriously help the models comparison
- Performance measures have to be used to compare predictions to observations
- In case of wheat all models have systematic errors
- HTO modelling in wheat leaf seems good for the 3 models
- Systematic errors : $\left(\frac{\overline{C_p}}{\overline{C_o}} = 0.76(JAEA)\ 0.86(IFIN\&CEA\right)$
- OBT modelling in wheat grain seems make underprediction for all model
- Systematic errors : $\left(\frac{\overline{C_p}}{\overline{C_o}} = 0.3(JAEA)\ 0.48(IFIN)\ 0.7(CEA)\right)$

energie atomique • energies alternatives

- Statistical analysis can seriously help the models comparison
- Performance measures have to be used to compare predictions to observations
- In case of wheat all models have systematic errors
- HTO modelling in wheat leaf seems good for the 3 models
- Systematic errors : $\left(\frac{\overline{C_p}}{C_o} = 0.76(JAEA)\ 0.86(IFIN\&CEA\right)$
- OBT modelling in wheat grain seems make underprediction for all model
- Systematic errors : $\left(\frac{\overline{C_p}}{C_o} = 0.3(JAEA)\ 0.48(IFIN)\ 0.7(CEA)\right)$

- Statistical analysis can seriously help the models comparison
- Performance measures have to be used to compare predictions to observations
- In case of wheat all models have systematic errors
- HTO modelling in wheat leaf seems good for the 3 models
- Systematic errors : $\left(\frac{\overline{C_p}}{\overline{C_o}} = 0.76(JAEA)\ 0.86(IFIN\&CEA\right)$
- OBT modelling in wheat grain seems make underprediction for all model
- Systematic errors : $\left(\frac{\overline{C_p}}{\overline{C_o}} = 0.3(JAEA)\ 0.48(IFIN)\ 0.7(CEA)\right)$

- Statistical analysis can seriously help the models comparison
- Performance measures have to be used to compare predictions to observations
- In case of wheat all models have systematic errors
- HTO modelling in wheat leaf seems good for the 3 models
- Systematic errors : $\left(\frac{\overline{C_p}}{\overline{C_o}} = 0.76(JAEA)\ 0.86(IFIN\&CEA\right)$
- OBT modelling in wheat grain seems make underprediction for all model
- Systematic errors : $\left(\frac{\overline{C_p}}{\overline{C_o}} = 0.3(JAEA)\ 0.48(IFIN)\ 0.7(CEA)\right)$

- Statistical analysis can seriously help the models comparison
- Performance measures have to be used to compare predictions to observations
- In case of wheat all models have systematic errors
- HTO modelling in wheat leaf seems good for the 3 models
- Systematic errors : $\left(\frac{\overline{C_p}}{\overline{C_o}} = 0.76(JAEA)\ 0.86(IFIN\&CEA \right)$
- OBT modelling in wheat grain seems make underprediction for all model
- Systematic errors : $\left(\frac{\overline{C_p}}{\overline{C_o}} = 0.3(JAEA)\ 0.48(IFIN)\ 0.7(CEA) \right)$

energie atomique • energies alternatives

ARE MODELS IN ACCEPTANCE CRITERIA

HTO Leaf			
Test/models	CEA	IFIN	JAEA
FAC2 > 0.5	no	no	no
Mean bias within $\pm 30\%$ of the mean ($\lvert FB \rvert < 0.3$ *or* $0.7 < MG < 1.3$))	ok	ok	ok
Random scatter ($NMSE < 1.5$ *or* $VG < 4$)	ok	ok	ok
Acceptance	ok ?	ok ?	ok ?

OBT Grain			
Test/models	CEA	IFIN	JAEA
FAC2 > 0.5	ok	ok	no
Mean bias within $\pm 30\%$ of the mean ($\lvert FB \rvert < 0.3$ *or* $0.7 < MG < 1.3$))	no	no	no
Random scatter ($NMSE < 1.5$ *or* $VG < 4$)	ok	ok	no
Acceptance	ok ?	ok ?	no