Estimation of NH_3 emission factors for pig production (building, storage) thanks to machine learning methods and ELFE database

THOMAS Johan – ESPAGNOL Sandrine - GUINGAND Nadine

Context, PIGAMMO Project, Aim

Livestock ammonia volatilization	Reduce environment impact Reduce health damages
Modelling	Global approach (building-storage) Comparison to measurements Simulate several scenarios
Machine Learning ("AI part who learn from data, generalize and thus perform tasks without explicit instructions")	**Explore ELFE database to predict ammonia emission factor**

Materials & Methods

ELFE database (Vigan et al. 2019)

- Data collected from international littérature

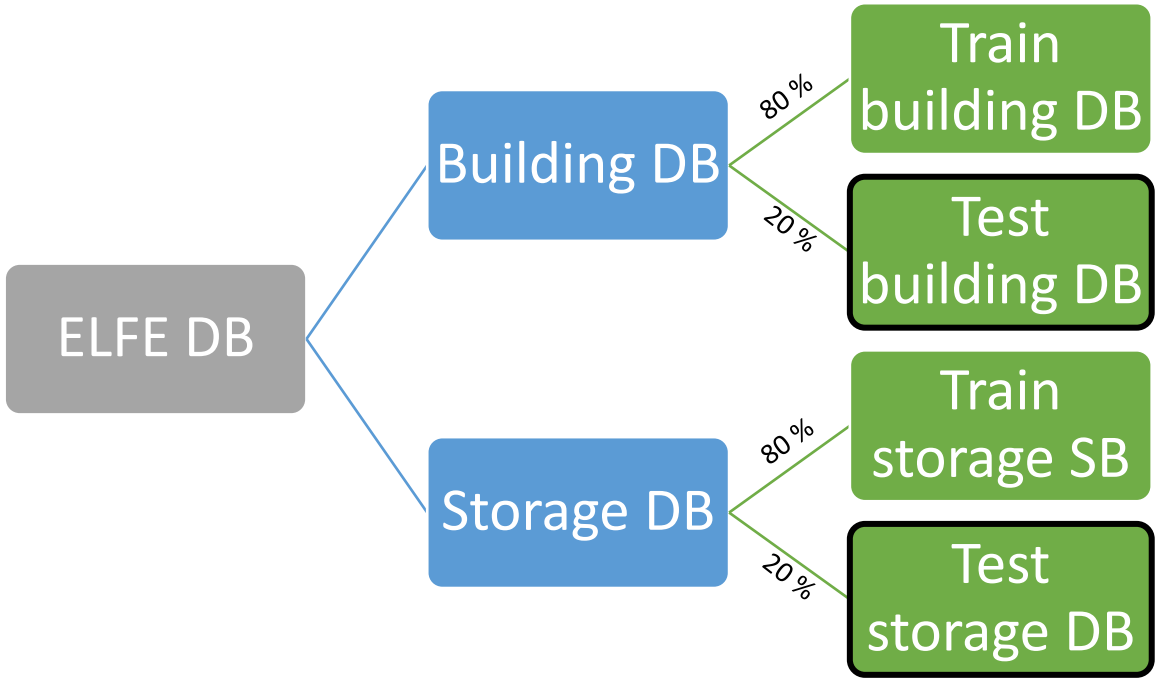
2 Subsets

- Building → Nadine Guingand as expert
- Storage → Sandrine Espagnol as expert

Several steps to improve database :

- Clean
- Complete
- Explore
- Transform
- Divide to train/test databases (80/20)

Predict ammonia emission factors thanks to Machine Learning methods



Vigan, A., Hassouna, M., Guingand, N., Brame, C., Edouard, N., Eglin, T., Espagnol, S., Eugène, M., Génermont, S., Lagadec, S., Lorinquer, E., Loyon, L., Ponchant, P., Robin, P., 2019. Development of a Database to Collect Emission Values for Livestock Systems. Journal of Environmental Quality. http://doi.org/10.2134/jeq2019.01.0007

ifip Institut du porc

Materials & Methods

Very large database …

But numerous features and missing data

- Data from various sources
- Heterogeneous units of expression

Complete …

- By expert knowledge
- Thanks to automatic methods

Machine Learning

- Supervised learning
- Several regression methods (*LinearRegressor, RandomForestRegressor, ElasticNet, Support-Vector Machine, Multi Layer Perceptron, …)*

Data used

ELFE DB

- Building
  - 1 010 observations
  - 978 variables
- Storage
  - 360 observations
  - 332 variables

Cleaned PigAmmo DB

- Building
  - 633 observations
  - 25 variables
- Storage
  - 327 observations
  - 22 variables

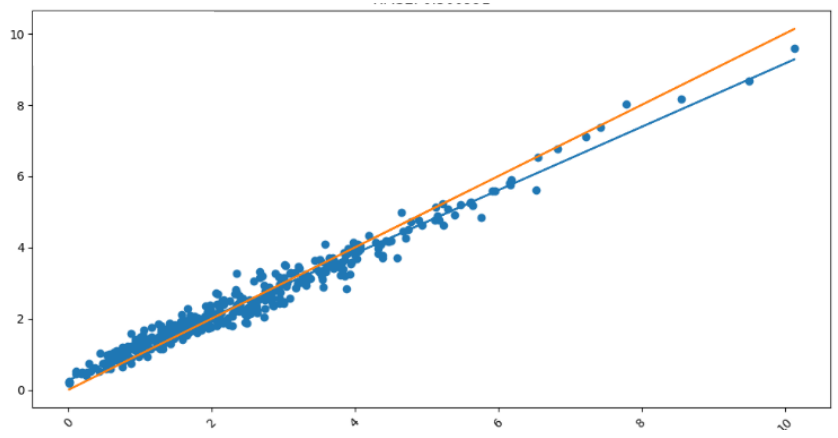
Variable examples:
- Physiological stage
- Temperature
- Floor type
- Storage volume
- Average daily gain
- Airflow
- …

Onehot encoding of categorical features

ifip Institut du porc

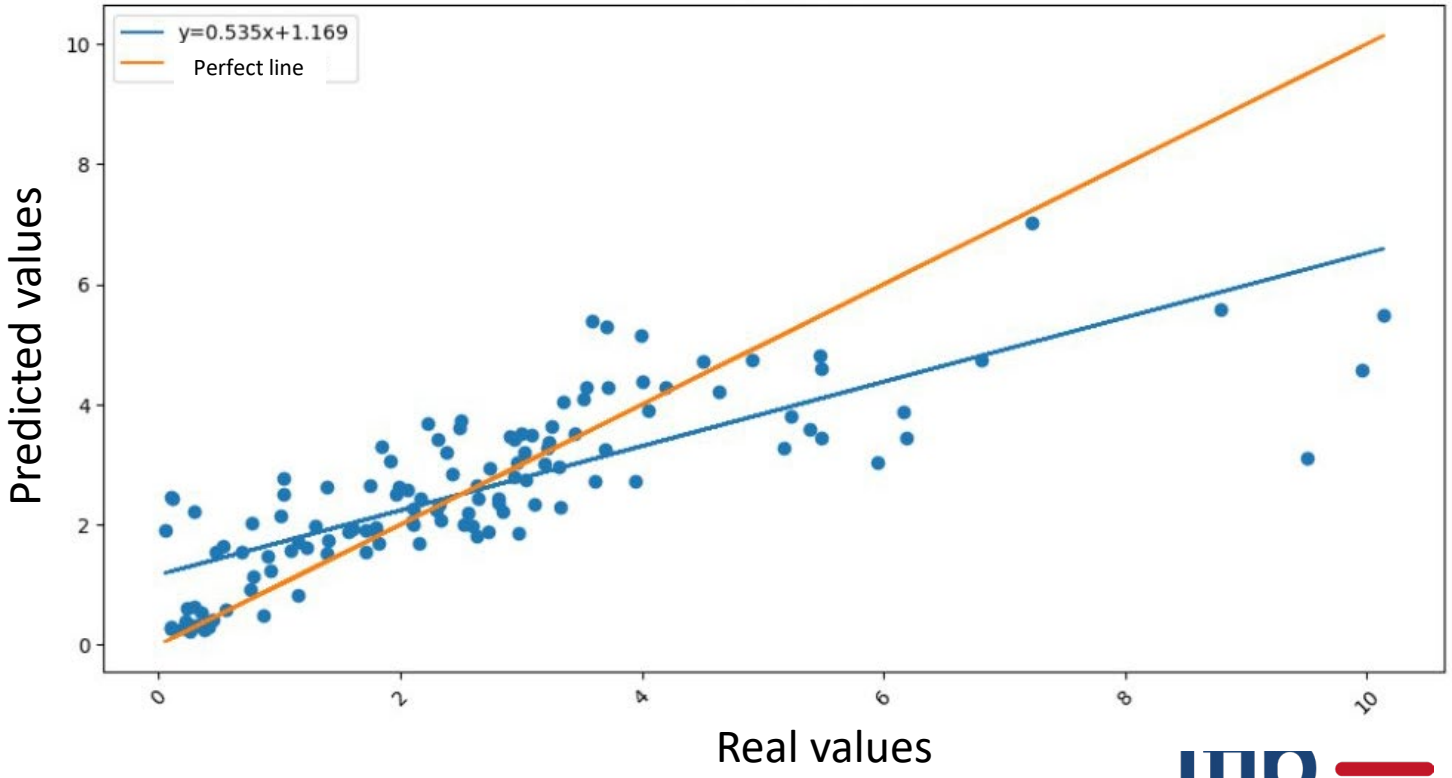
Results – Pig building

- **Building → Random Forest**
  - Train $r^2 = 0.89$
  - Test $r^2 = 0.61$
- **Overfitting ?**
- **Heterogeneous data ?**
- **Select only specific data :**
  - Fattening stage
  - Fully slatted floor
  - Dynamic ventilation



R2 score: 0.613
RMSE: 1.2761

Expressed in kg NH3-N.animal place^{-1}.yr^{-1}

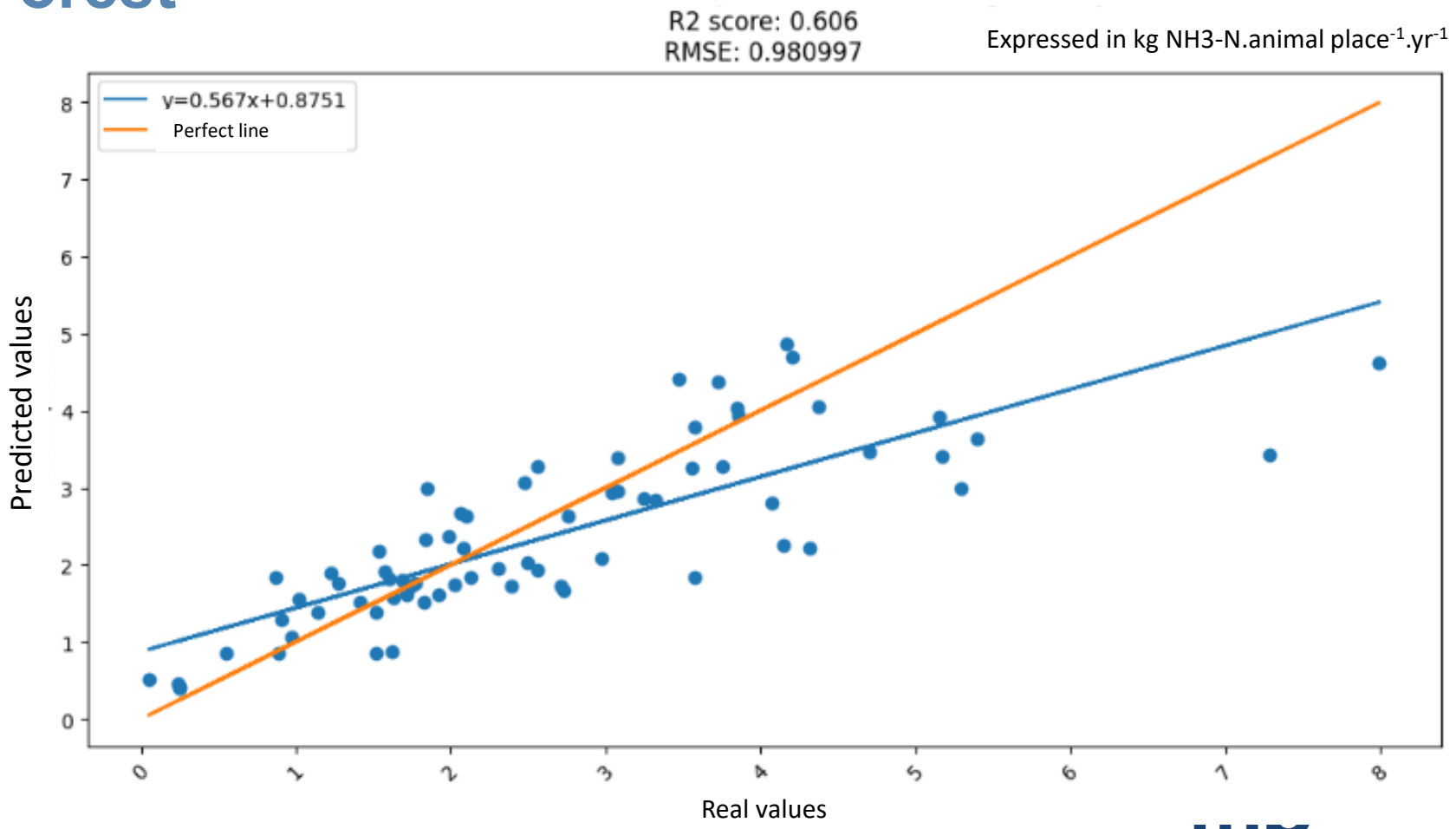


Results – Fattening pig building

Building → Random Forest

- Train r^2 = 0.96
- Test r^2 = 0.61

No improvement



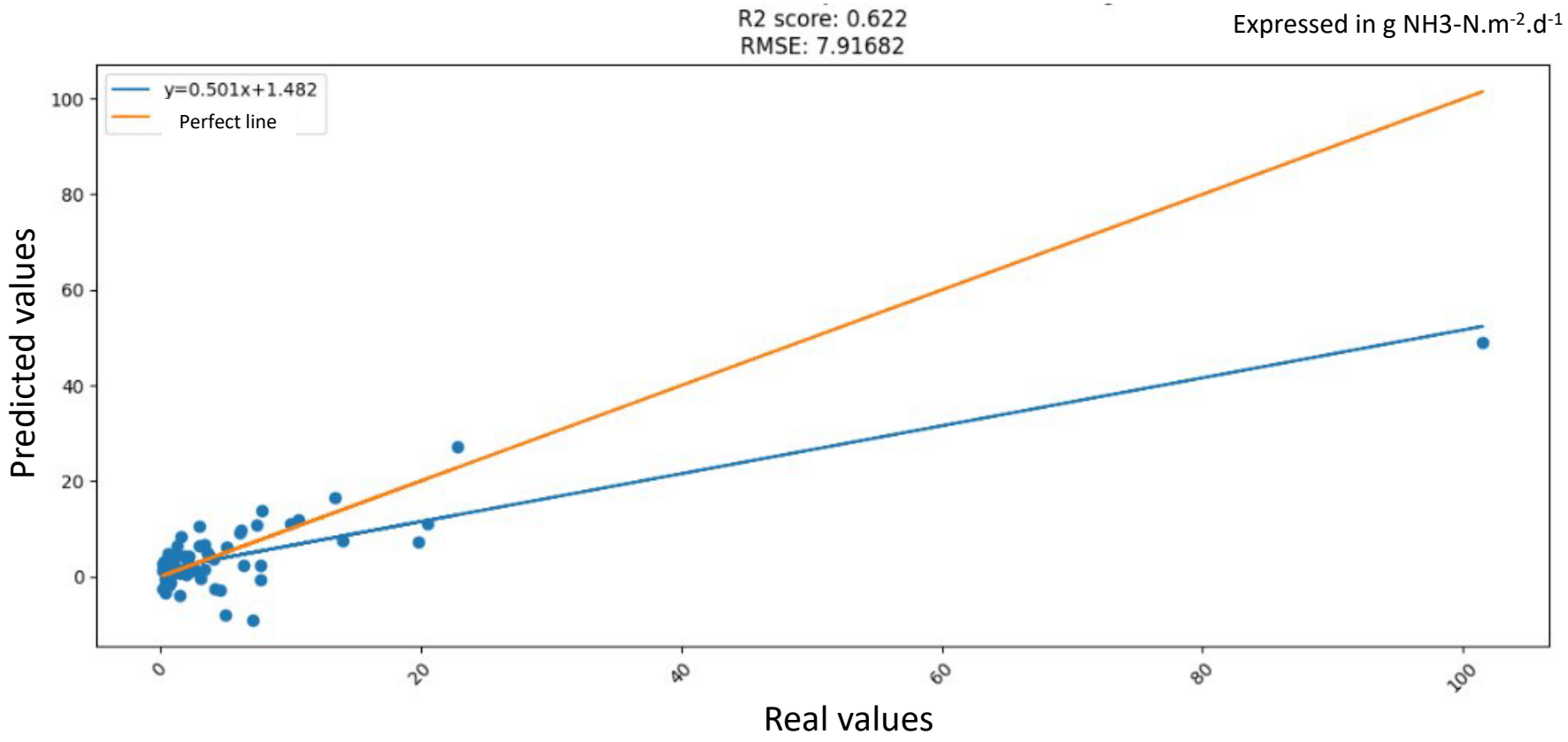
Results - Storage

Storage → Ridge

- Train r^2 = 0.87
- Test r^2 = 0.62

Overfitting ?

Outliers ?

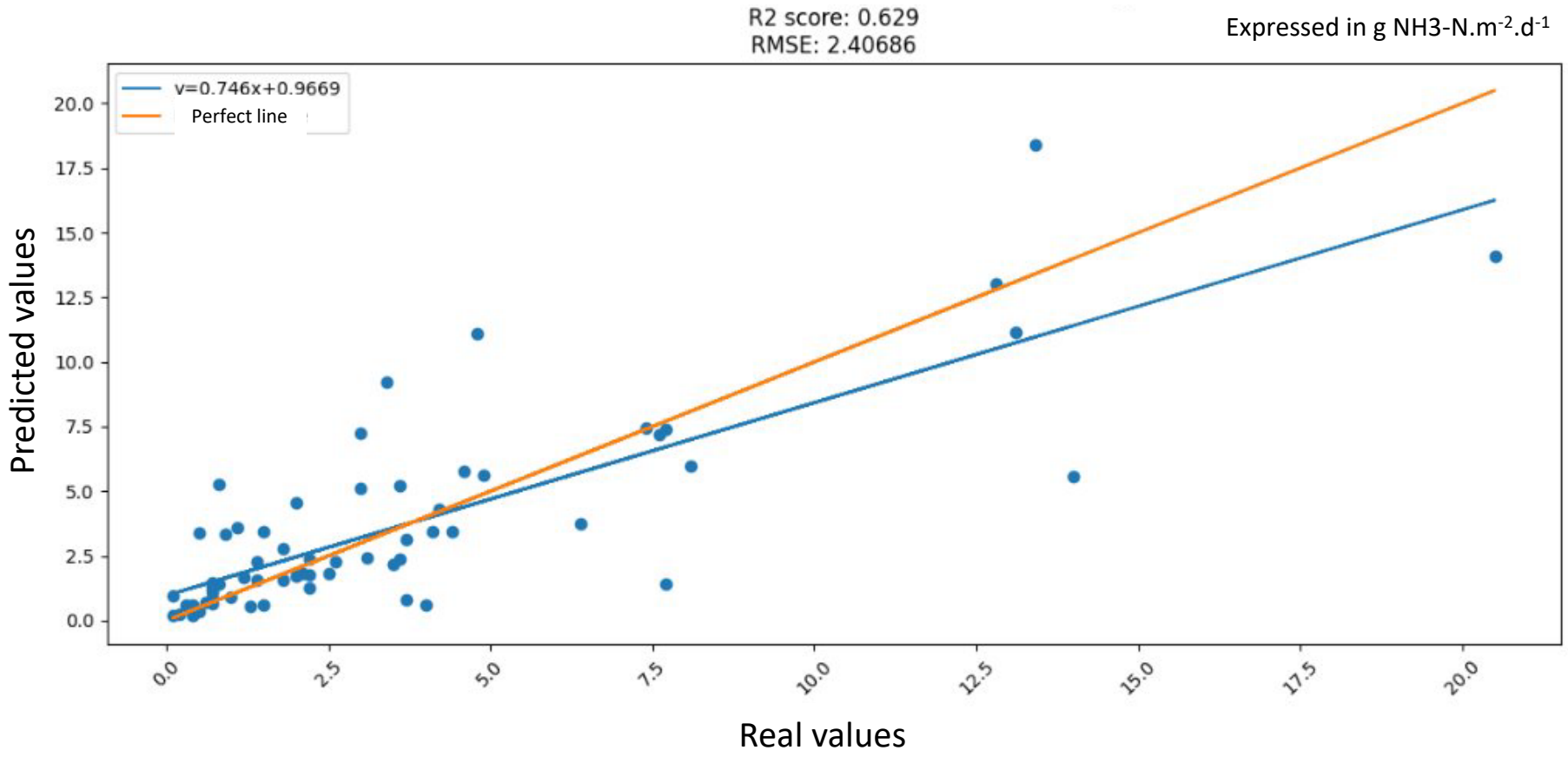


Results – Storage without outliers

Storage → MLP

- Train r^2 = 0.90
- Test r^2 = 0.63

No improvement



Conclusions

Insufficient results for reliable prediction of ammonia emission factors, and unsatisfaction in relation to project aims

Explanations :

- **Lack of data**
  - Due to: incomplete variables, data obtained in specific context, database split,…
- **Heterogenous and imbalanced data**
- **Completing database increases the reliability**

Improvement ways extend ELFE database with standardization of publication writing

ifip Institut du porc

ifip — Institut du porc

ADEME — AGENCE DE LA TRANSITION ÉCOLOGIQUE

CORTEA N° : 2003C0031

Thanks for your attention

THOMAS Johan

johan.thomas@ifip.asso.fr