Assessment of the Performance of NO and NO_2 low cost sensors over extended time periods in a real world application

Alessandro Bigi[1], Michael Mueller[2], Stuart K. Grange[3], Grazia Ghermandi[1], Christoph Hueglin[2]

[1] "Enzo Ferrari" Department of Engineering, University of Modena and Reggio Emilia, I-41125, Modena, Italy
[2] Empa, Swiss Federal Laboratories for Materials Science and Technology, CH-8600, Duebendorf, Switzerland
[3] Department of Chemistry, University of York, York, YO10 5DD, United Kingdom

presentation is based on A. Bigi et al., AMT (2018)

Motivation

- Low-cost sensors (LCS) are highly attractive
- See WMO (2018) for possibilities and limitations of LCSs
- Can LCSs be used for mapping of air pollutants (in cities) ?

Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications

valid as of May 2018

Editors: Alastair C. Lewis, Erika von Schneidemesser and Richard E. Peltier

WORLD METEOROLOGICAL ORGANIZATION

IGAC

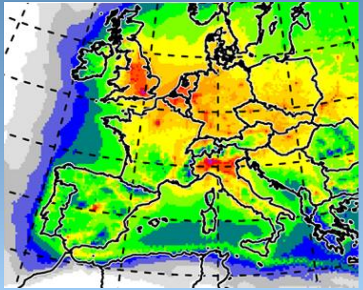
UN environment

Motivation – mapping of air pollution (in cities)

Dispersion models

- ✓ Regional scale transport
- ✓ Hourly resolution
- ✓ Simulations of few years

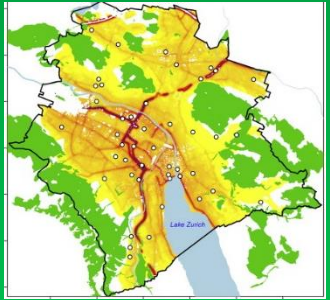
- ✗ Metrics averaged over the city
- ✗ No urban canopy



Statistical models

- ✓ Few meters of resolution
- ✓ Accurate and low computation cost

- ✗ Relying on numerous in-situ observations
- ✗ No emission sector separation

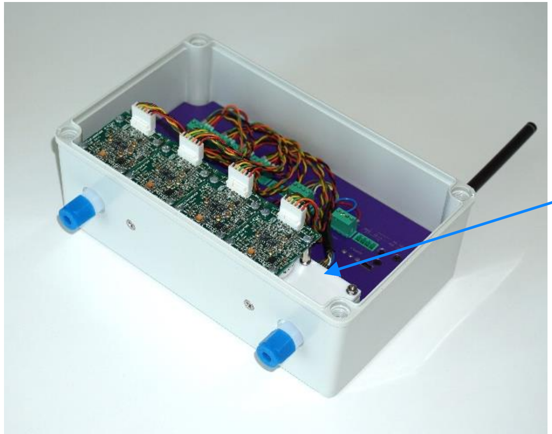


from A. Berchet, Empa

AirCube (AC) Sensor Units (DecentLab)

Air in

Air out



PTFE manifold and blower for active ventilation

- 2 NO Alphasense B4 electrochemical sensors
- 2 NO_2 Alphasense B43F electrochemical sensors
- rH/ Temp Sensirion STH21
- Battery powered
- GSM or LoRa communication

Calibration

Apr 12 – Jul 6, 2017

Haerkingen, rural/traffic

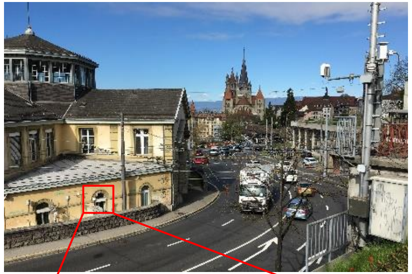
Calibration (training) dataset

Deployment

Jul 30 – Dec 5, 2017

Zurich, urban

Lausanne, urban/traffic



Deployment (test) dataset

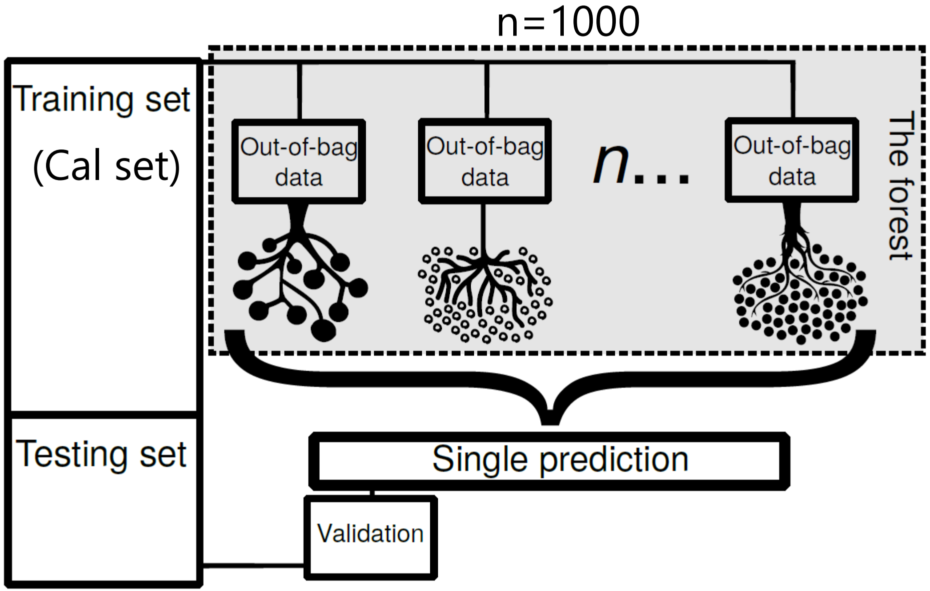
Calibration Models

Method		
Linear Models (LM)	LM	
Robust Linear Models (RLM)	RLM	
Random Forest Regression (RF)	RF	
Support Vector Regression (SVR)	SVR	(see Bigi et al. AMT 2018)

Target variable	Signal reference instrument
Explanatory variables	$V_{Sensor\ A}$, $V_{Sensor\ B}$, $(V_{Sensor\ A} + V_{Sensor\ B})/2$, T_{AC}, RH_{AC}, $DRH_{t\ AC}$, $V_{Sensor\ Co\text{-}Pollutant}$
Interactions	✓
Time resolution	10 min
Number observations	8'315 – 11'454 (calibration/training) 16'535 – 17'083 (deployment/test)

AC:	AirCube Sensor Unit
$DRH_{t\ AC}$:	see Mueller et al., AMT (2017)

Random Forest Model



Conceptual diagram of a random forest model

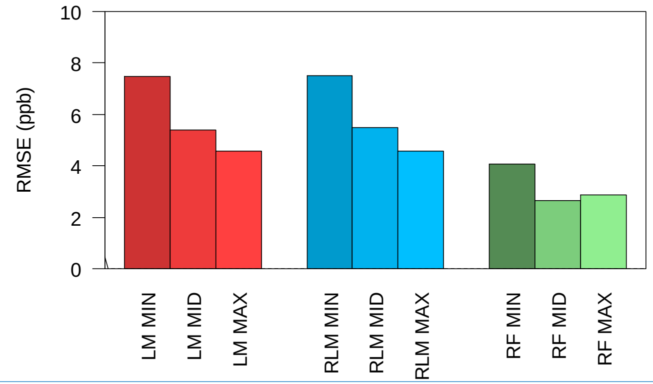
- Tuning of RF's model parameters was conducted

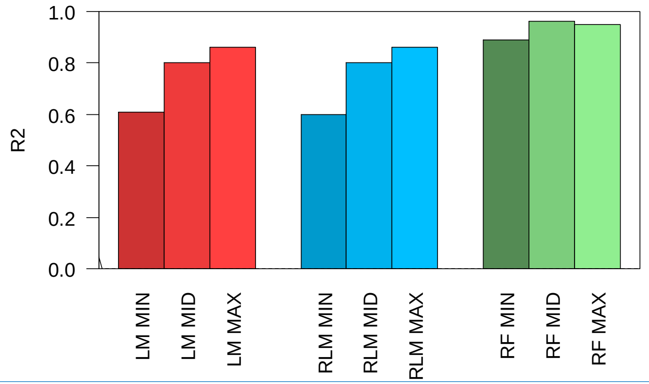
- Measurement at the Basel St. Johann reference site from March 6 – Aug 13, 2018
- Split of measurements in training (67%) and testing (33%) data sets, 3-fold cross-validation

Grange et al ACP (2018)

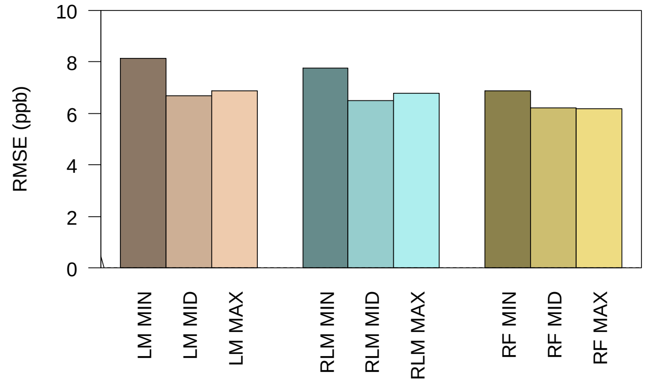
Goodness-of-Fit Metrics for NO_2 (AC009 – AC012)

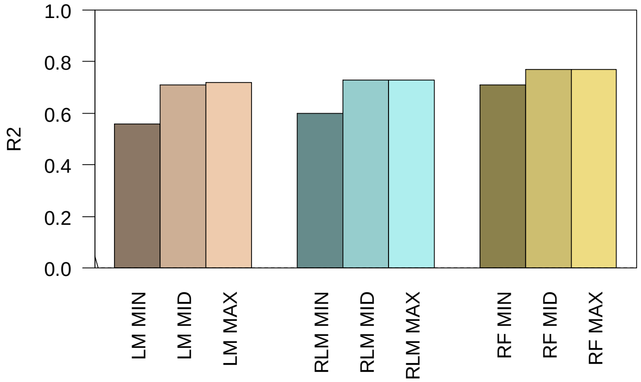
Mean of models using only one NO_2 sensor in AC (n=8) – **Calibration data set**





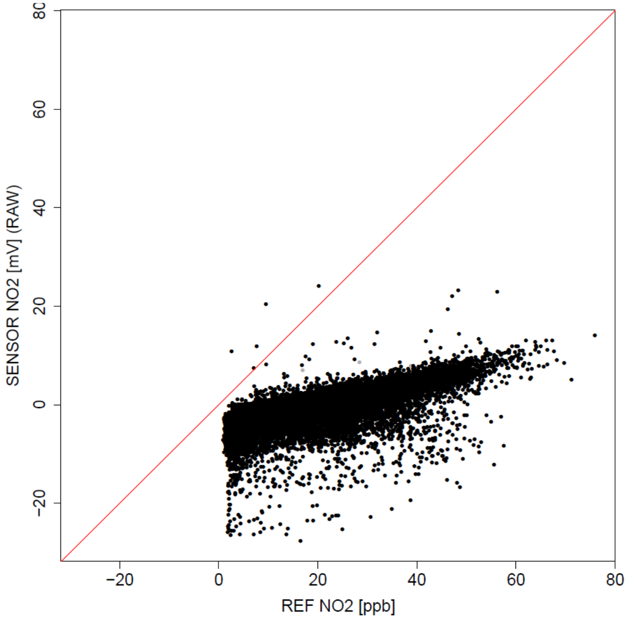
– **Deployment data set**





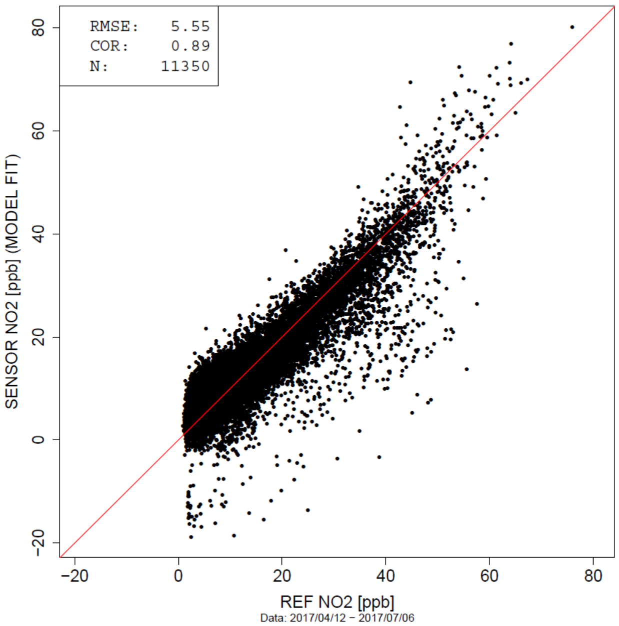
Calibration data for NO_2 sensor A in AC 012

Raw signal



RLM MID

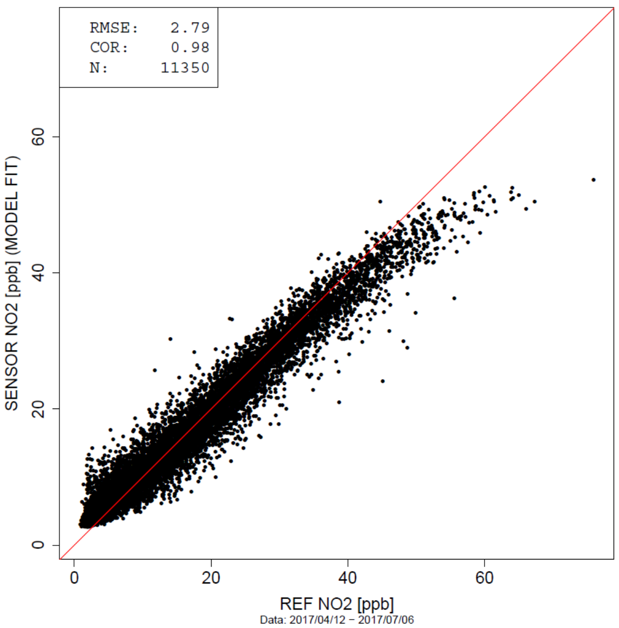
$NO_2 = NO_{2A}+T+RH+DRH60 + NO_{2A}*T$



RF MID

$NO_2 = f(NO_{2A}, T, RH, DRH60)$

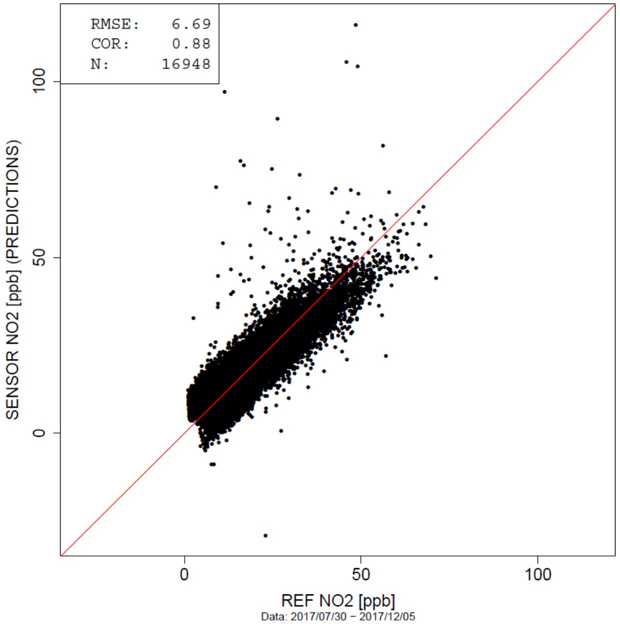
nodesize = 30



Scatterplot for NO_2 sensor A in AC 012 during Deployment

RLM MID

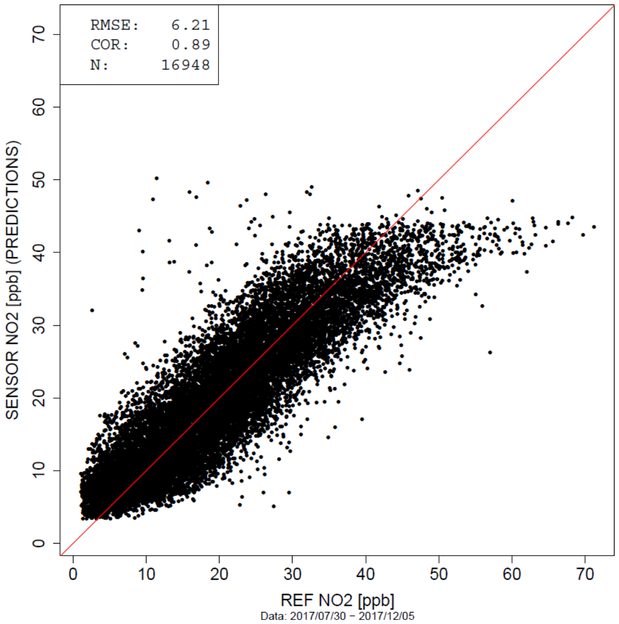
$NO_2 = NO_{2A} + T + RH + DRH60 + NO_{2A} * T$



RF MID

$NO_2 = f(NO_{2A}, T, RH, DRH60)$

nodesize = 30



Hourly residuals ($NO_{2\ reference} - NO_{2\ sensor}$), e.g. AC 012

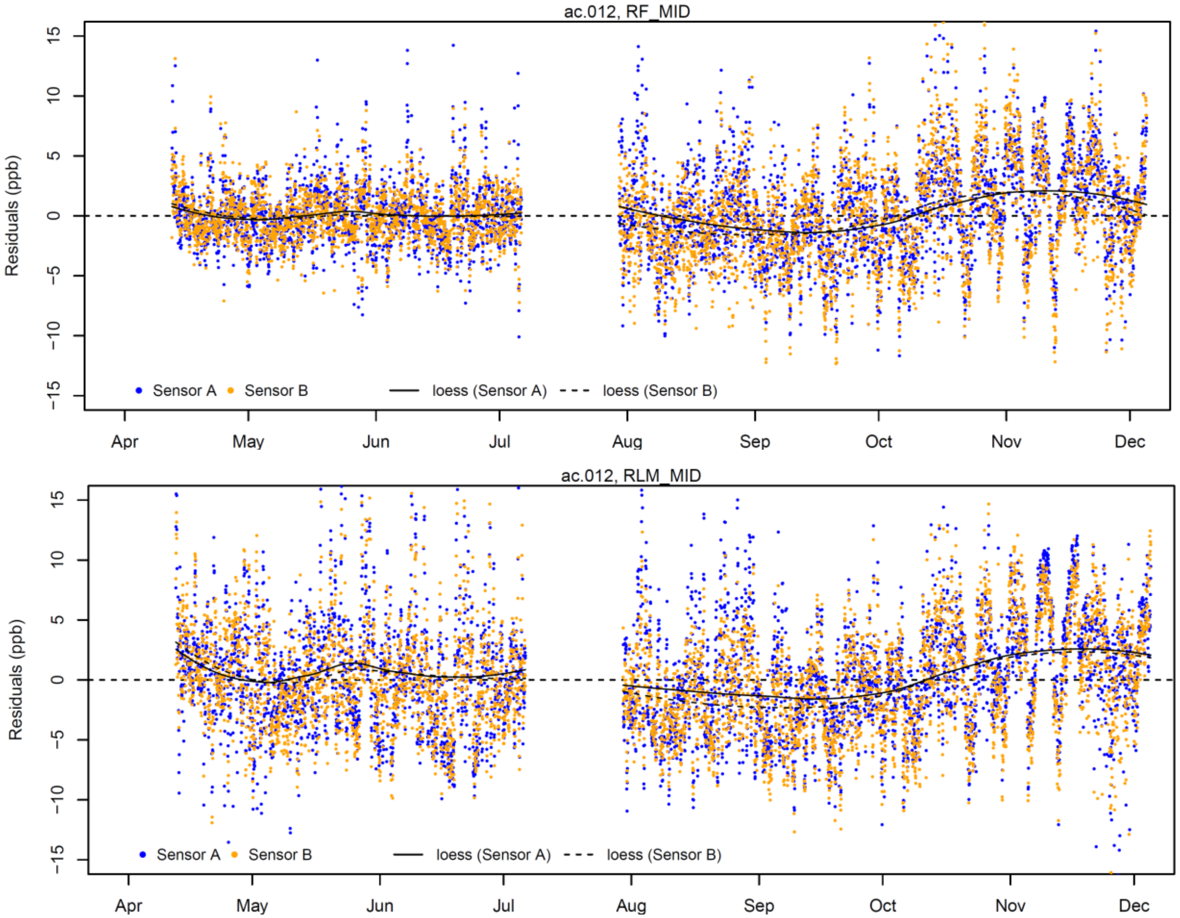
RF MID

$NO_2 = f(NO_{2A}, T, RH, DRH60)$

nodesize=30

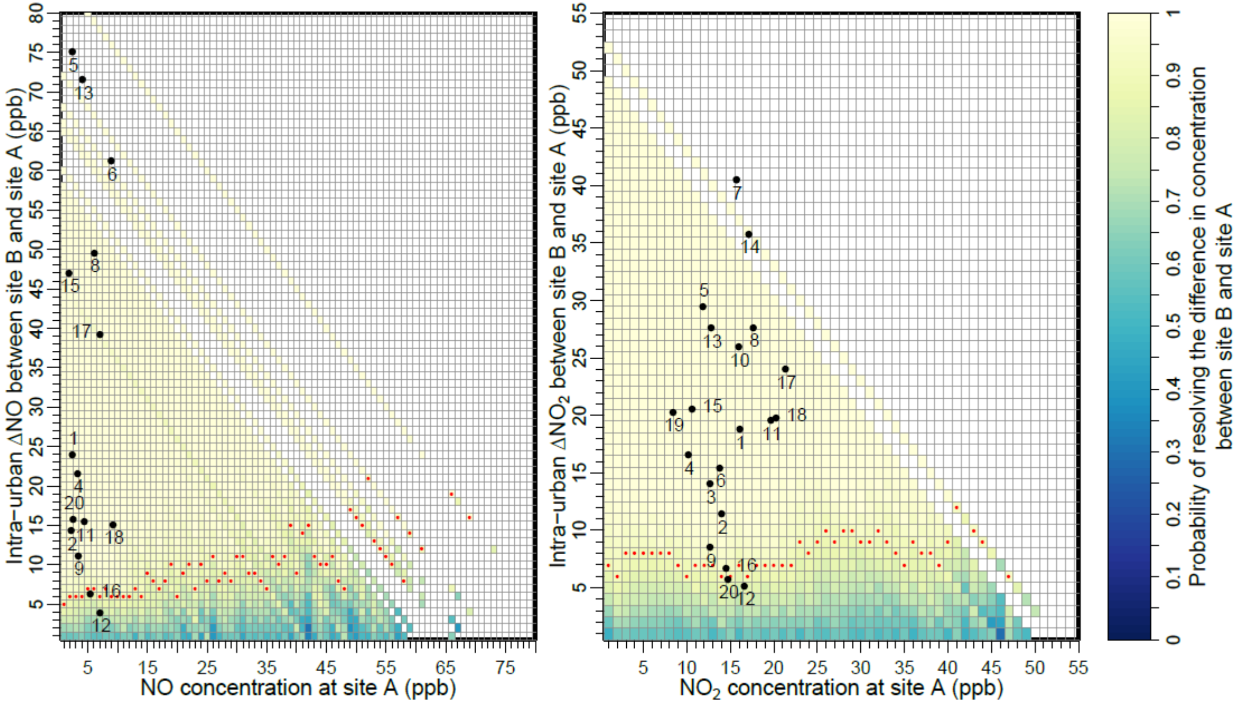
RLM MID

$NO_2 = NO_{2A}+T+RH+DRH60 + NO_{2A}*T$



Probability of resolving intra-urban difference in NO and NO_2

site A: urban background, site B: urban hotspot (e.g. traffic)



1. Barcelona
2. Bologna
3. Brussels
4. Budapest
5. Hamburg
6. Krakow
7. London
8. Madrid
9. Manchester
10. Marseille
11. Milan
12. Modena
13. Munich
14. Paris
15. Prague
16. Reggio Emilia
17. Rome
18. Turin
19. Vienna
20. Zürich

Bigi et al AMT (2018)

Conclusions

- Air sensors need individual calibration (e.g. co-location to reference instruments)
- Different calibration algorithms have been tested. Random Forest (RF) regression showed the best performance
- Data quality for RF models has during deployment been lower than during calibration (Overfitting?)

  ⇒ Model parameters need to be carefully chosen and calibration models should be validated (cross-validation, use at different reference site, …)
- Sensor drift was not the limiting factor for data quality during deployment
- Measurement error of tested electrochemical sensors as determined in a real world application is around 6 ppb for both NO and NO_2 (Bigi et al. 2018)

Thank you !

Many thanks to

- Swiss National Science Foundation (International Short Visit Grant IZK0Z2-174969)
- Anthony Wild for provision of the Wild Fund Scholarship
- Markus Camenzind for data from Basel St. Johann