Modélisation de la température du sol avec un bilan d'énergie, application à la prédiction de l'émergence du maïs (Zea mais)

Etienne Claverie

université PARIS-SACLAY

CentraleSupélec

NNT : 2018SACLC028

Modelling soil temperature with an energy balance model, application to prediction of maize (*Zea mais*) emergence

Thèse de doctorat de l'Université Paris-Saclay
préparée à CentraleSupélec

École doctorale n°573 Interfaces
Spécialité de doctorat : Mathématiques appliquées

Thèse présentée et soutenue à Gif-sur-Yvette, le 18 mai 2018, par

Étienne CLAVERIE

Composition du jury :

Paul-Henry COURNÈDE MICS, CentraleSupélec, Université Paris-Saclay	Directeur de thèse
Jérémie LECOEUR Seeds Research, Syngenta Crop Protection AG	Co-directeur
Isabelle COUSIN UR SOLS, INRA	Rapporteuse
Thierry FOURCAUD UMR AMAP, CIRAD	Rapporteur
Véronique LETORT MICS, CentraleSupélec, Université Paris-Saclay	Examinatrice
Alexandra JULLIEN UMR ECOSYS, AgroParisTech	Présidente du jury

Thèse de doctorat

"Be this as it may, I believe that the human imagination has never invented anything that is not true, whether in this world or in other worlds . . ."

Gérard de Nerval, *Aurélia & Other Writings*

université PARIS-SACLAY
ÉCOLE DOCTORALE
INTERFACES
Approches Interdisciplinaires : Fondements, Applications et Innovation

Title : Modelling soil temperature with an energy balance model, application to prediction of maize (*Zea mais*) emergence

Keywords : Soil model, energy balance, sensitivity analysis, emergence prediction, maize

The beginning of crop growth is influenced by soil temperature and water content near the surface. The aims of this work are three-fold: I) to develop a soil model to predict seed's local temperature and water content using easily available meteorological data, II) run global sensitivity analysis to identify the components of the model with the largest contribution to the output uncertainty and III) to apply the model to predict maize emergence.
The model implements heat and water diffusion partial differential equations with the resolution of the energy balance equation as upper boundary condition to simulate the dynamics of heat and water within a 1-D homogeneous soil profile. The system of equation was resolved using an explicit scheme. With the explicit scheme, we were able to obtain a good accuracy with a relatively coarse discretization. The soil parameters were estimated with Saxton's pedotransfer functions. For the sensitivity analysis, a semi-global (Morris) and variance-based (Sobol) method was applied to identify the effect of parameters on model output.
The parameters for net radiation were estimated using data from a net radiometer. After calibration, model relative error was lower than 10 % for temperature and water content at 30 cm depth for the whole year 2016 in Stein, Switzerland. Right below the surface, at 5 cm, the relative error for temperature was under 20 % and considered accurate enough for emergence prediction. The prediction of emergence was done using a thermal time model calibrated with field data. With the combination of the soil temperature and thermal time model, the emergence was better predicted for standard sowing conditions compared to a model using air temperature, the classical approach.
This work is an example of applying complex biophysics model to understand an agronomic problem. The results of this work will participate in optimising breeding efforts for cold-tolerant crop varieties. Future investigations should consider a better estimation of evaporation and decomposition of the two phases composing the emergence process, germination and elongation.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin,
France

université PARIS-SACLAY
ÉCOLE DOCTORALE
INTERFACES
Approches Interdisciplinaires : Fondements, Applications et Innovation

Titre: Modélisation de la température du sol avec un bilan d'énergie, application à la prédiction de l'émergence du maïs (*Zea mais*)

Mots-clés: Modèle de sol, bilan d'énergie, analyse de sensibilité, prédiction de l'émergence, maïs

La croissance en début de cycle des grandes cultures est influencée principalement par la température et la teneur en eau du sol. L'objectif de ce travail est triple: I) développer un modèle capable de prédire la température et la teneur en eau du sol autour d'une graine, à l'aide de données climatiques largement disponibles, II) analyser la sensibilité des sorties du modèle pour identifier les composants qui contribuent le plus largement à leur incertitude et III) appliquer le modèle à la prédiction de l'émergence du maïs.
Le modèle simule la dynamique de la chaleur et de l'eau dans un profil de sol homogène à une dimension en utilisant un bilan d'énergie comme condition à la borne supérieure. Les équations de diffusion sont résolues grâce à un schéma explicite. Ce schéma s'est révélé stable avec une discrétisation à larges couches. Le paramétrage du sol a été réalisé avec les fonctions de pédotransfert de Saxton. Une méthode d'analyse semi-globale (Morris) et une autre basée sur la variance (Sobol) ont été appliquées pour déterminer les paramètres les plus influents du modèle. Grâce aux données d'un bilanmètre, les paramètres du bilan radiatif ont pu être estimés. Après calibration, l'erreur du modèle est inférieure à 10% pour la température et teneur en eau à 30 cm de profondeur pour une simulation sur toute l'année 2016 à Stein en Suisse. Sous la surface à 5 cm, l'erreur pour la température est en dessous de 20% et a été considéré suffisamment précise pour être utilisée pour prédire l'émergence.
La prédiction de l'émergence s'est fait grâce à un modèle de temps thermique calibré grâce à des expériences aux champs. Avec la combinaison du modèle de sol et du modèle de temps thermique, l'émergence du maïs a été mieux prédite en utilisant notre température du lit de semences simulée qu'en utilisant la température de l'air, la variable la plus couramment utilisée pour prédire l'émergence.
Ce travail est une application d'un modèle biophysique complexe à un problème agronomique. Les résultats participeront à l'optimisation de l'effort de sélection des variétés tolérantes au froid. Deux pistes de recherche peuvent être considérées pour des futurs travaux: une meilleure modélisation de l'évaporation et une décomposition de l'émergence.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin,
France

Acknowledgements

I would like to thank Jérémie Lecoeur for giving me the chance to join the "Environmental and Crop Modelling" adventure in Syngenta and the Syngenta Talent Development program for covering the scholarship fees.

I would like to thank also the Environmental and Crop Modelling team (JiSu Bang, Thomas Galinier, Audrey Etienne, Ludovic Tambour, Hubert Vincent Varella and Pierre Anduze) for their constant support during this challenging task. In particular, thanks to Hubert for helping me with the Morris sensitivity analysis, Thomas for mining our internal databases and Audrey for the system diagram of Sophia and the listening and ideas to solve my problems.

I would like to thank Paul-Henry Cournède, my thesis director for accepting me in the Biomathematics lab in CentraleSupélec and all the team (Benoit Bayol, Gautier Viaud, Xangtuo Chen, Antonin Della Noce, Charlotte Baey, Chloé Adam and Jean-Christophe Attard) for their good mood and the wonderful Biomathematics workshops organized every year in nice French cities (Chartres, Troyes, Étretat and Marseille)

My familly deserves a warm thanks as well, my mother Anne, my father André, my brother Colin, my step-father Greg, my grand-mother Raymonde and the other relatives who were present during the defense.

My deepest thanks go to two exceptional women. First, Véronique Letort, my supervisor, whose gentleness and patience gave me the perfect atmosphere to adventure into the continent of modelling and applied mathematics. Second, Pamela Duboc, for helping me clarify my writing and wonderfully editing the manuscript. You are magic!

Foreword

This thesis project is a joint collaboration between the Environmental and Crop Modeling team of Syngenta Crop Protection AG, and the Biomathematics team in CentraleSupélec Paris. Within Syngenta R&D, the Environmental and Crop Modelling team focuses on crop simulations to predict the abiotic stresses and help the breeders, business and marketing units to promote new cereal varieties. The Biomathematics lab is a research lab focused on mathematical modelling, analysis of mathematical properties, statistical inference, machine learning and optimal control.

This thesis is a work of engineering. It shows the steps to develop a model, uses various methods to show the sensitivity of factors and shows an example of application. The model development was as straightforward as possible to be able to deploy it rapidly in an industrial context.
The thesis work was done between March 2014 and March 2018 from the offices of Syngenta in the Stein Research Centre near Basel, Switzerland and the final document was written in Basel University Library between July and September 2017.
Parts of this thesis have been published in Etienne Claverie et al. (2016). "Modeling soil temperature to predict emergence". In: *Functional-Structural Plant Growth Modeling, Simulation, Visualization and Applications (FSPMA), International Conference on*. IEEE, pp. 28–37

The thesis was written in LaTeX on the Overleaf platform, graphs were done from Rstudio with R (R Core Team, 2015) and package ggplot2 (Wickham, 2009). The figures in introduction are hand drawn by myself. The references were managed by Docear (Beel et al., 2011). The thesis was written on an HP EliteBook 840 G1 with 8 GB of RAM.

About Syngenta : Syngenta is one of the world's leading companies with more than 26 000 employees in over 90 countries dedicated to its purpose: Bringing plant potential to life. Through world-class science, global reach and commitment to customers, Syngenta helps to increase crop productivity, protect the environment and improve health and quality of life. www.syngenta.com
About CentraleSupélec : CentraleSupélec is a French institute of research and higher education in engineering and science. It was established on 1 January 2015 as a result of a strategic merger between two leading grandes écoles in France, Ecole Centrale Paris and Supélec. It is a key founding member of the University of Paris-Saclay (consortium of research universities in France) and the TIME (Top Industrial Managers for Europe) network. http://www.centralesupelec.fr/en

Contents

Abstract **v**

Résumé **vii**

Acknowledgements **ix**

Foreword **xi**

Introduction: soil conditions and crop canopy establishment **1**
1 Research context 1
2 Soil model: domain of application 5
2.1 Objectives and specifications 6
2.2 The soil system 7
2.3 Combination of soil and emergence models 10
3 Soil model: main equations 12
3.1 Near surface soil modelling 12
3.2 Soil surface conditions 14
3.3 Transport of heat within the soil 16
3.4 Transport of water within the soil 17
3.5 Model inputs 17
3.6 The modelling challenges 18
4 The scientific plan 18

1 Analysis of soil temperature in Stein, Switzerland **21**
1 Description of weather station sensors 21
1.1 Sensors 21
1.2 Data processing 23
1.3 Soil texture 26
2 Data analysis 26
2.1 Temperature profile in air and soil 26
2.2 Difference between air and soil temperatures at 5 cm is greater in winter 29
2.3 The variability of air temperature explains the variability of soil temperature at 5 cm 32
2.4 Correlations of soil temperature with other variables . 34
3 Chapter conclusion 36

2 Development of the soil model SOPHIA **39**
1 Conceptual diagram 40
1.1 Forrester diagram 40

1.2 What SOPHIA does not include . . . 40
2 Biophysics equations . . . 42
2.1 Heat movements . . . 42
2.2 Water movements . . . 44
2.3 Energy balance equation at the soil surface . . . 45
2.4 Bottom of the soil profile . . . 51
3 Numerical implementation . . . 53
3.1 Discretisation of soil into layers . . . 53
3.2 Formal expression of heat and water movements . . . 55
3.3 Formal expression of boundary conditions . . . 56
3.4 Resolution of the soil model with an explicit scheme . . . 58
3.5 Implementation . . . 61
4 Simulation: preliminary tests . . . 63
4.1 Test of the explicit scheme . . . 63
4.2 Verification with Müller's model . . . 67
4.3 Simulation with different texture classes . . . 70
4.4 Estimation of water retention parameters . . . 76
4.5 Model performance in Stein, Switzerland . . . 78
5 Discussion and summary . . . 87
5.1 Discussion and future developments . . . 87
5.2 Summary . . . 90

3 Sensitivity analysis, parameter estimation and simulation 93
1 Sensitivity analysis . . . 94
1.1 Range and distribution of parameters . . . 95
1.2 Local OAT sensitivity analysis . . . 100
1.3 Morris sensitivity indices . . . 106
1.4 Sobol sensitivity indices . . . 114
1.5 Discussion on sensitivity analysis . . . 118
2 Parameter estimation . . . 121
2.1 Estimation of Brunt's coefficients . . . 123
2.2 Cloudiness correction for atmospheric emissivity . . . 126
2.3 Estimation of soil emissivity . . . 130
2.4 New formalism for albedo . . . 132
3 Evaluation of the submodules improvements on model outputs 139
3.1 Evaluation of the complete model using the 2015 dataset, used for the calibration of some of its sub-modules . . . 139
3.2 Model performance in Stein, Switzerland for 2016 . . . 145
4 Discussion and summary . . . 153
4.1 Summary . . . 156

4 Application of soil model to emergence prediction 157
1 Context . . . 157
2 Material and method . . . 159
2.1 Thermal time models for emergence . . . 159
2.2 Thermal time models . . . 162
2.3 Soil temperature simulation . . . 163

2.4 Model selection procedure for cultivar specific thermal time model . . . 163
3 Results . . . 164
3.1 Emergence experiment . . . 164
3.2 Thermal time models for emergence . . . 168
3.3 Combination of soil simulation and adapted emergence model . . . 172
3.4 Uncertainty propagation in the thermal time models . . . 174
4 Discussion . . . 176

Conclusion and outlook **181**
1 Results . . . 181
2 Outlooks . . . 185
2.1 Necessary improvements in the short-term . . . 186
2.2 Outlooks for industrial applications . . . 187
3 General conclusion . . . 189

A Glossary of soil modelling terms **191**

B The R package 'sophia' **197**

C Resolution of the soil model with a semi-implicit scheme **207**

D Résumé de la thèse en français **211**

Bibliography **223**

List of Figures

1 Overview of maize growth stages . . . 3
2 Anatomy of maize seeding . . . 3
3 Use of soil temperature . . . 5
4 The soil under study . . . 8
5 The soil system . . . 9
6 The four energy fluxes . . . 15
7 The five meteorological inputs of the model . . . 19

1.1 Sensors on the weather station . . . 22
1.2 Profiles of temperature in May . . . 27
1.3 Gradient of temperature across seasons and horizons . . . 30
1.4 Difference between air and soil temperature over time . . . 31
1.5 Scatter plot of soil temperature and air temperature . . . 33
1.6 Matrix of bivariate scatter plots of soil temperature and other variables . . . 35

2.1 Forrester Diagram of the soil temperature model . . . 41
2.2 Discretisation of soil system . . . 54
2.3 Sketch of flow and state variables at time n and n+1 . . . 60
2.4 Three ways to discretise the soil . . . 64
2.5 Sensitivity of soil temperature to space discretisation . . . 65
2.6 Sensitivity of water content to space and time discretisation . . . 66
2.7 Soil temperature simulations . . . 71
2.8 Water content simulations . . . 72
2.9 Simulation of energy balance terms . . . 73
2.10 Simulation with six texture classes . . . 75
2.11 Measure of water retention curves . . . 77
2.12 Simulation moisture retention curve . . . 79
2.13 Comparion of temperature in Stein . . . 82
2.14 Comparion of water content in Stein . . . 83
2.15 Comparion of net radiation in Stein . . . 85

3.1 Model response for surface parameters . . . 102
3.2 Model response to changes in soil parameters . . . 104
3.3 $\mu^* - \sigma$ plot . . . 109
3.4 Morris global ranking of parameters . . . 111
3.5 Sobol sensitivity analysis on 14 parameters . . . 117
3.6 Evaluation of estimated Brunt's coefficients on clear sky data . . . 125
3.7 Parametrisation of downward long-wave simulation . . . 128
3.8 Comparison of different soil emissivities in April 2015 . . . 133

3.9 Distribution of albedo value 135
3.10 Albedo and soil water content at 5 cm 136
3.11 Albedo on net short-wave radiation 138
3.12 Evaluation of simulation of temperature at 5 cm 152

4.1 Application of soil temperature model to predict emergence: a sketch 160
4.2 Soil temperature and moisture during emergence 165
4.3 Cumulative emergence curves 167
4.4 Emergence rate vs. Temperature for the 4 cultivars 171
4.5 Error of simulation around base temperature 175

A.1 The form of water in the soil 195

B.1 Tutorial: energy balance simulation 202
B.2 Tutorial: soil temperature simulation 203
B.3 Tutorial: soil water content simulation 204

D.1 Principales équations de SOPHIA 215

List of Tables

1.1 List of sensors . . . 24
1.2 Additional variables . . . 24
1.3 Soil texture in Stein . . . 27

2.1 Müller and SOPHIA differences . . . 69
2.2 Soil properties of different texture classes . . . 77
2.3 Parameter values for Stein . . . 80
2.4 First evaluation of the model . . . 87

3.1 Ranges of parameters used for the sensitivity analysis . . . 96
3.2 Parameter ranking for three outputs . . . 113
3.3 AIC for downward long-wave radiation . . . 131
3.4 Errors with soil emissivity in 2015 . . . 133
3.5 AIC for albedo . . . 137
3.6 Model versions for 2015 evaluation . . . 141
3.7 RMSE in 2015 for Rnet and Lu . . . 143
3.8 RMSE in 2015 for soil temperature . . . 144
3.9 RMSE in 2015 for soil water content . . . 146
3.10 Model versions used for performance . . . 147
3.11 Evaluation of the soil model . . . 148
3.12 Comparison RMSE and RMAE for radiation . . . 151

4.1 Results of emergence model selection . . . 168
4.2 Emergence rate for the 4 cultivars . . . 170
4.3 Evaluation of emergence prediction . . . 173

B.1 List of parameters and their default value for the simulation in this example . . . 204

List of Abbreviations

FC	**F**ield **c**apacity
Ld	**L**ong-wave radiation **d**ownward
Lu	**L**ong-wave radiation **u**pward
PDE	**P**artial **d**ifferential **e**quation
PTF	**P**edo-**t**ransfert **f**unction
PWP	**P**ermanent **w**ilting **p**oint
Rnet	**Net r**adiation
RMSE	**R**oot **m**ean **s**quare **e**rror
RRMSE	**R**elative **r**oot **m**ean **s**quare **e**rror
USDA	**U**nited **S**tates **D**epartement of **A**griculture
SHAW	**S**imultaneous **H**eat **A**nd **W**ater model
VPD	**V**apor **p**ressure **d**eficit
VWC	**V**olumetric **w**ater **c**ontent

Physical Constants

General physical constants	
Absolute zero	$0\,\mathrm{K} = -273.15\,^{\circ}\mathrm{C}$
Acceleration due to gravity	$g = 9.81\,\mathrm{m\,s^{-2}}$
Gas constant	$R = 8.314\,\mathrm{J\,mol^{-1}\,K^{-1}}$
Latent heat of vaporisation at $0\,^{\circ}\mathrm{C}$	$A_L = 2501\,\mathrm{J\,g^{-1}}$
Slope of latent heat of vaporisation function of temperature	$B_L = -2.37\,\mathrm{J\,g^{-1}\,C^{-1}}$
Molecular weight of water	$M_{H_2O} = 18\,\mathrm{g\,mol^{-1}}$
Pi	$\pi = 3.14\,-$
Stefan-Boltzmann constant	$\sigma = 5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$
Von Karmann constant	$k = 0.4\,-$
Soil physical constants	
Mean particle density	$\rho_s = 2.6\,\mathrm{g\,cm^{-3}}$
Density of water	$\rho_{H_2O} = 1\,\mathrm{g\,cm^{-3}}$
Density of clay	$\rho_{clay} = 2.65\,\mathrm{g\,cm^{-3}}$
Density of quartz (main component of sand)	$\rho_{quartz} = 2.66\,\mathrm{g\,cm^{-3}}$
Density of air	$\rho_{air} = 0.0012\,\mathrm{g\,cm^{-3}}$
Density of organic matter	$\rho_o = 1.3\,\mathrm{g\,cm^{-3}}$
Specific heat of air	$c_{air} = 1.01\,\mathrm{J\,g^{-1}\,K^{-1}}$
Specific heat of clay	$c_{clay} = 0.9\,\mathrm{J\,g^{-1}\,K^{-1}}$
Specific heat of organic matter	$c_o = 1.92\,\mathrm{J\,g^{-1}\,K^{-1}}$
Specific heat of quartz	$c_{quartz} = 0.8\,\mathrm{J\,g^{-1}\,K^{-1}}$
Specific heat of water	$c_{H_2O} = 4.18\,\mathrm{J\,g^{-1}\,K^{-1}}$
Thermal conductivity of air	$k_{air} = 0.025\,\mathrm{W\,m^{-1}\,K^{-1}}$
Thermal conductivity of clay	$k_{clay} = 2.92\,\mathrm{W\,m^{-1}\,K^{-1}}$
Thermal conductivity of organic matter	$k_o = 0.25\,\mathrm{W\,m^{-1}\,K^{-1}}$
Thermal conductivity of quartz	$k_{quartz} = 8.8\,\mathrm{W\,m^{-1}\,K^{-1}}$
Thermal conductivity of water	$k_{H_2O} = 0.57\,\mathrm{W\,m^{-1}\,K^{-1}}$
Weighting factor for thermal conductivity in air	$w_{air} = 1.4\,-$
Weighting factor for thermal conductivity in clay	$w_{clay} = 0.4\,-$
Weighting factor for thermal conductivity in organic matter	$w_o = 0.4\,-$
Weighting factor for thermal conductivity in quartz	$w_{sand} = 0.4\,-$

List of Symbols

Roman symbols

a	Water-retention curve parameter a	$-$
b	Water-retention curve parameter b	$-$
b_{Kw}	Hydraulic conductivity parameter	$-$
B_a	Brunt parameter a for atmospheric emissivity ϵ_a	$-$
B_b	Brunt parameter b for atmospheric emissivity ϵ_a	$-$
C_h	Soil volumetric heat capacity	$\mathrm{J\,m^3\,{}^\circ C^{-1}}$
cc	Cloudiness correction factor	$-$
E	Potential evaporation	$\mathrm{m\,s^{-1}}$
E_{act}	Actual evaporation	$\mathrm{m\,s^{-1}}$
e_a	Ambient vapour pressure	kPa
e_s	Saturated vapour pressure	kPa
Ext	Extraterrestrial radiation	$\mathrm{W\,m^{-2}}$
f_{clay}	Volume fraction of clay	$\mathrm{m^3\,m^{-3}_{solids}}$
f_m	Volume fraction of minerals (sand and clay)	$\mathrm{m^3\,m^{-3}_{solids}}$
FT	thermal flow	$\mathrm{W\,m^{-2}}$
f_{sand}	Volume fraction of sand particles	$\mathrm{m^3\,m^{-3}_{solids}}$
f_o	Volume fraction of organic matter	$\mathrm{m^3\,m^{-3}_{soil}}$
FW	water flow	$\mathrm{m\,s^{-1}}$
G	Ground heat flux	$\mathrm{W\,m^{-2}}$
H	Sensible heat flux to the atmosphere	$\mathrm{W\,m^{-2}}$
I	Factor in atmospheric emissivity corrected for cloudiness	$-$
Kw	Hydraulic conductivity	$\mathrm{m\,s^{-1}}$
Kws	Hydraulic conductivity at saturation	$\mathrm{m\,s^{-1}}$
L	Latent heat of vaporisation	$\mathrm{W\,m^{-2}}$
Lu	Long-wave upward	$\mathrm{W\,m^{-2}}$
Ld	Long-wave downward	$\mathrm{W\,m^{-2}}$
P_{tot}	Total porosity	$\mathrm{m^3_{pores}\,m^3_{soil}}$
P_{air}	Volume fraction of air in the soil	$\mathrm{m^3_{air}\,m^3_{soil}}$
r_{ah}	**A**erodynamic **r**esistance to **h**eat flux	$\mathrm{s\,m^{-1}}$
Rn	Net radiation	$\mathrm{W\,m^{-2}}$
S	Soil emissivity factor	-
St	Global (direct + diffuse) short-wave radiation	$\mathrm{W\,m^{-2}}$
$T_{surface}$	Surface temperature	$\mathrm{{}^\circ C}$
T_{sky}	Sky temperature	$\mathrm{{}^\circ C}$
u	Wind speed	$\mathrm{m\,s^{-1}}$
U^*	Friction velocity	$\mathrm{m\,s^{-1}}$
z_{ref}	Reference height	m

z_0	Roughness length	m
Greek symbols		
α	Albedo, reflection coefficient	-
ϵ_a	Atmosphere emissivity	-
ϵ_{acc}	Atmosphere emissivity corrected with cloudiness factor	-
ϵ_s	Soil emissivity	-
Ψ	Hydraulic potential	$\mathrm{m_{H_2O}}$
Ψ_g	Gravitational potential	$\mathrm{m_{H_2O}}$
Ψ_m	Matric potential	$\mathrm{m_{H_2O}}$
Ψ_{min}	Minimum hydraulic potential	$\mathrm{m_{H_2O}}$
ρ_{va}	Vapour density in air	$\mathrm{g_{H_2O}\,m_{air}^{-3}}$
ρ_{vs}	Vapour density at soil surface	$\mathrm{g_{H_2O}\,m_{air}^{-3}}$
ρ_m	Density of minerals (quartz and clays)	$\mathrm{g\,m^{-3}}$
ρ_b	Dry bulk density	$\mathrm{g_{solids}\,m_{soil}^{-3}}$
θ	Volumetric water content	$\mathrm{m_{H_2O}^3\,m_{soil}^{-3}}$
θ_{sat}	Volumetric water content at saturation	$\mathrm{m_{H_2O}^3\,m_{soil}^{-3}}$
θ_e	Volumetric water content at air-entry suction	$\mathrm{m_{H_2O}^3\,m_{soil}^{-3}}$
τ	atmospheric transmission	-

To Anne and André, my parents
To the four elements
embodied by the Soil, Sun, Rain and Wind
which sustain all Life on Earth
and are the key players in my model

Introduction: soil conditions and crop canopy establishment

1 Research context

During the first month of their development, annual crops like maize (*Zea mays*) develop their canopy and root system to capture the resources they need: light, water and nutrients. A successful season starts with a fast and homogeneous development of canopy and roots, which will drive an efficient use of available resources and lead ultimately to high yield and good quality of the crop. This fast development will determine the degree of uniformity and density of the crop, and also has an impact of the degree of weed infestation (Hadas, 2004).

This period from planted seed to a functional canopy is called the establishment period. As defined by the FAO (Food and Agricultural Organization of the United Nations), it spans in maize the period from sowing to five-leaf stage (a stage called V5 in S. Ritchie, Hanway, and Benson (1992), figure 1). For maize, the establishment period will take on average 15 to 25 days (FAO, n.d.).

To maximise the length of the growing season and capture more resources, growers tend to sow as early as possible, which leads to more stresses and damage during establishment (Spaeth, 1994), since sowing happens in colder conditions. In 25 years, the maize sowing date in the US corn belt has shifted on average ten days earlier (Sacks and Kucharik, 2011). Grain yield has been shown to decrease linearly with decreasing early-season soil temperature (Bollero, D. G. Bullock, and Hollinger, 1996) in a study where the seed environment was artificially cooled until V5. Early planting can thus be considered as risky. In addition, the no-till practice, which consists in leaving residues from the previous crop on the soil surface, prevents the soil from warming up. In fact, residue-covered soils stay colder than bare soil surfaces. Hence, the no-till practice further affects emergence (Hayhoe et al., 1993). Moreover, the residue layer can be heterogeneous, with some patches of bare ground

alternating with covered soil. This creates differences in temperature that may lead to uneven establishment (Liu et al., 2004). Hence, there is a good reason for research institutes and agro-industry to find solutions to improve establishment of early-sown maize.

Maize establishment

Maize establishment happens as a series of successive processes : germination, elongation of primary root and coleoptile (figure 2), switching to the autotrophic phase around V3 (Bhosale et al., 2007) and finally shifting of the meristem from an underground to an overground position around V5. In maize, the growing point (or meristem) during stand establishment is thus located below the soil surface and is more sensitive to soil conditions than to atmospheric conditions (Vinocur and J. T. Ritchie, 2001). Emergence, which results from coleoptile elongation that allows the plant to reach the surface, is a crucial stage during establishment. It marks the end of the dramatic physiological changes that characterise the evolution from a dry seed to a functional plant (McDonald, 1994).

The environment around the seed: the seedbed

The zone where the seed is sown is called the seedbed. It is defined as the first $20\,\mathrm{cm}$ of soil below the surface. The seedbed is often prepared through tilling to maximise the chance for seeds to germinate and emerge. Seeds need to sense sufficient moisture to start germination. Then temperature further drives development of the seed (germination and elongation) all the way to emergence (Hadas, 2004).

Soil temperature and water content drive many biological, physical and chemical processes in the soil (figure 3). The germination and emergence of plants are just a couple of them. Temperature set the speed of biological processes such as seed germination, seedling elongation, root development and microbial activity. Soil temperature and moisture also have an impact on physical processes such as evaporation and solute transport. In regards to chemistry, they govern the carbon and nitrogen cycles as well as the degradation of active chemicals.

The most important environmental factors affecting stand establishment are soil temperature, soil water potential and mechanical impedance (soil strength) (Schneider and Gupta, 1985; Whalley and W. E. Finch-Savage, 2011). Soil temperature drives the timing of germination and elongation, while water

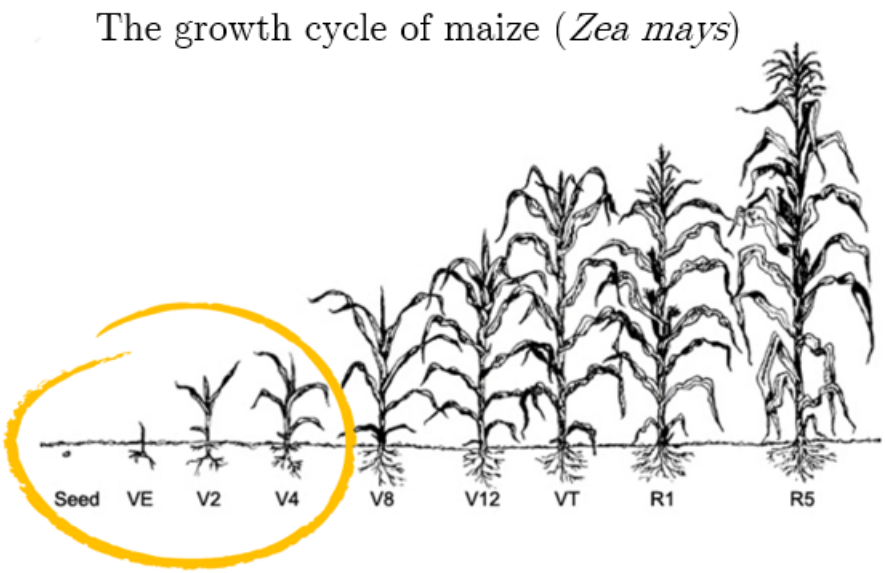

FIGURE 1: The growth cycle of maize with vegetative (V) and reproductive (R) growth stages. VE is emergence. The establishment period corresponds to the circled part : from seed to V5 (V5 stage is right after V4 but not shown on the image). Image downloaded in June 2016 from Purdue University Extension

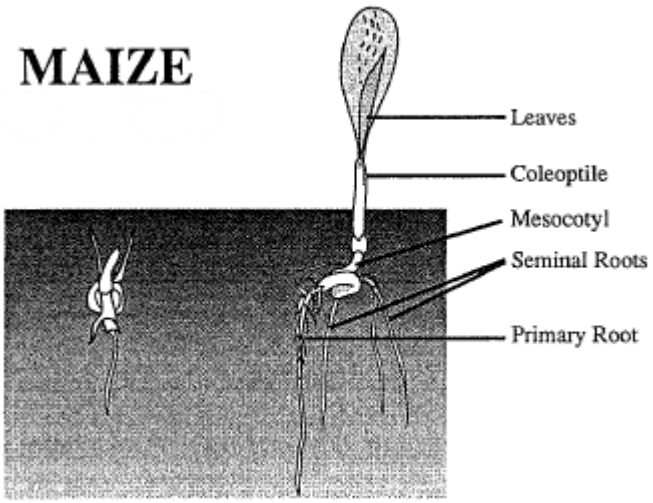

FIGURE 2: A closer look at the maize seedling just after germination (left), and after emergence (right). Germination is done when the root has protruded from the seed. Emergence sees the coleoptile appear above the ground. In favorable conditions, the first leaves unfurl one day after emergence, as represented in the image. Image taken from McDonald (1994)

potential will have an impact on germination, and mechanical impedance on elongation (W. E. Finch-Savage and Bassel, 2015). Soil temperature affects root systems, as lower temperatures cause reduction in the space explored by the roots (Nagel et al., 2009), thus hindering their access to water and nutrients, which are critical resources for establishment.

Data on soil temperature, water content or soil strength is seldom available to growers and researchers although it is valuable information for current research topics. Information on soil temperature helps to breed for chilling tolerance (Bhosale et al., 2007), to comprehend the effects of seed treatments (Bradley, 2008) or to test biological hypotheses concerning germination and emergence modifiers. Research on better seed treatments is especially relevant in the context of western Europe, where the spraying of chemicals over crops is becoming a topic of heavy public criticism. In such a context, seed treatment would offer an alternative to protect plants from diseases in the first month of their growth. Many other factors, including agricultural practices, seedling diseases, genetics, and seed quality (Stoll and Saab, 2013) do affect the establishment period but are not taken into account in this work.

Agronomy of stand establishment

Maize agronomists from Europe and North America recommend plantation to happen when average soil temperature is expected to stay above 10 °C for the next 48 hours (University of Nebraska Lincoln). These 48 hours correspond to the seed imbibition period when maize seeds are sensitive to a temperature below this threshold (Herner, 1990). If temperatures are too low, the seed coat integrity is damaged, and imbibition cannot take place. The seed needs water to allow the starch degrading enzymes to function. These cut the starch in sugars for use as energy in the growth of the embryo. As the embryo has reached a size that finally forced it to tear out of the seed coat, its shoot and primary root start to elongate. Germination is defined as the time when the radicle bursts out of the seed. Elongation, which happens at a pace that is mainly a function of temperature (Atkinson and Porter, 1996), will bring the first visible leaf above the soil surface, usually after five to ten days.

The scientific problem

From a R&D perspective, having access to information about soil conditions during establishment is crucial for the improvement of stand establishment. Knowledge of soil conditions could help breeders and biochemists develop

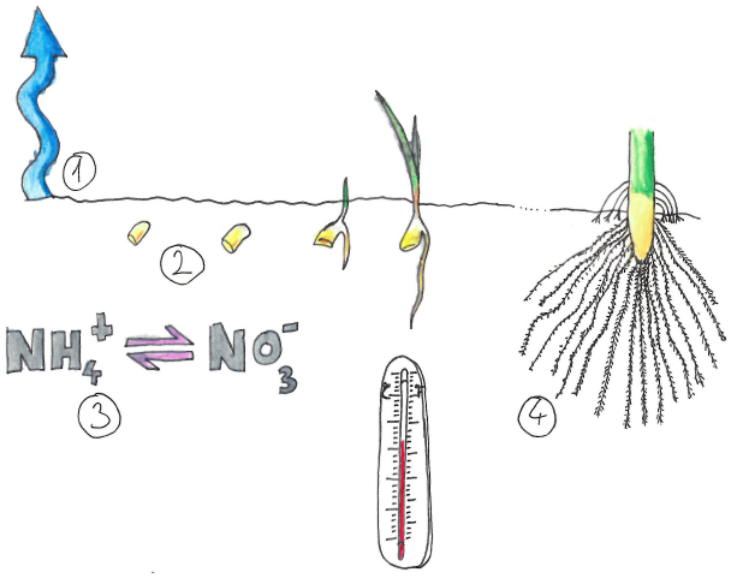

FIGURE 3: Example of the use of the soil temperature in understanding processes like ① soil evaporation, ② seed germination and elongation of the epicotyle toward the soil surface (emergence) ③ nitrogen cycle and other microbial activity, ④ root development

new varieties and new seed treatments. In order to develop these new technologies, they need to better understand the interactions between seed, plant, chemistry and soil processes from the very beginning of the crop cycle.

One way to study these interactions is to develop models that integrate knowledge from different fields. In this work, we will focus on modelling the soil temperature around the seed, since it drives many soil processes. We will show an application of this model to emergence prediction and address questions such as : What are the processes influencing soil temperature near the surface? What are the biophysical equations describing these processes? How to translate the equations into a numerical model? What are the sensitive parameters of this model? How can we estimate them? How accurate the soil model needs to be to predict accurately emergence?

2 Soil model: domain of application

When starting a modelling exercise, three steps require particular attention: *(i)* define the objective of our model, *(ii)* describe our representation of the system being modelled and *(iii)* match the complexity of that model with the

available data (L. Zhang, Walker, and Dawes, 2002). These points will be the focus of the subsequent paragraphs.

2.1 Objectives and specifications

Our long-term objective is to obtain a detailed description of the beginning of the crop cycle, a period during which soil temperature plays a vital role. In Syngenta's current crop models (unpublished), water balance is modelled in a simple way with a reservoir approach and does not include soil temperature. These models rely on the prediction of critical phenological stages of the crop development and are thus strongly dependent on the temperature input. Because the beginning of the crop cycle is driven by soil temperature, taking only air temperature leads to important errors in predicting the phenology of the entire cycle (P Cellier et al., 1993; Jame and Cutforth, 2004; Sacks and Kucharik, 2011). Hence, our objective is to develop a soil temperature and moisture model that accurately represents the topsoil conditions. The model will serve the study of the beginning of the crop cycle and the analysis of the effect of several factors (climatic, agricultural practices) on system behaviour.

More precisely, the objectives and specifications of the model can be stated as follows:

- To predict the soil temperature and moisture at any depth and time for a field with a bare soil surface
- To run accurately on the scale of a small field and for time periods from a week to several seasons
- To be suited to an agricultural soil (i.e. not for specific soil such as peat soils, which are too acidic)
- To be valid under a large range of conditions: sunny or rainy weather, day and night, extreme events and different types of soil
- To provide soil temperature and moisture information to other models (seed development, emergence, nitrogen cycle, root development, fate of chemicals in the soil)
- To require only a reduced set of necessary input variables so that it can be easily run at any location where there are a modern weather data acquisition system and soil characteristics

- To be expressed mathematically as an input-output system so that model analysis methods (sensitivity analysis, parameter estimation) can be easily tested

2.2 The soil system

In a modelling project, members of the project must precisely define the system under study. They must agree on what is included in the system, what are the models' goals and what type of analysis will be carried out. In our case, the system we want to represent is the soil. However, a more precise system description is necessary.

The soil is defined as the loose and fragmented outer layer of the earth's continental surface (Daniel Hillel, 2003). The soil layer is the result of the weathering (via physical, chemical and biological processes) of rocks and is constantly evolving through the action of wind, rain, biological and human activity.

From a physical point of view, the soil is composed of three compartments: the air, water and mineral compartments. Solid mineral particles are surrounded by thin layers of water and form a porous media dotted with air-filled pores. Mineral particles like clay have water-retention properties that modify the drying or wetting of the soil. At the surface of minerals, physical, chemical and biological processes and their interactions are influencing the germination, establishment, growth (accumulation of biomass), development (e.g. time until flowering) and yield.

The soil physical properties, mainly texture and structure, provide a basis for root development and define the water retention properties of the soil. The chemical and biological properties drive the availability of nutrients to the plant.

The interaction between plant and soil is complex given that both are composed of many different elements, and their interactions change with time and space. This level of complexity makes the prediction of the state of the soil by expertise or intuition impossible and calls for a modelling approach.

A photograph of a soil profile is presented in figure 4. The photograph captures the soil between 0 (surface) to approximately 50 cm depth. This is the soil on which we have installed the weather station used in this work (see next chapter, chapter 1). From the photograph, the composition of that soil is visible : particles of different sizes, with small particles near the surface and more clumped and aggregated soil particles towards the bottom of the

FIGURE 4: The soil profile near the Syngenta field station in Stein, Switzerland. The soil is an agglomeration of particles of various size and nature that host a wide biodiversity. The particles range from microscopic clays to macroscopic stones. Some debris of plant roots can also be identified in the picture. Some cracks are visible at the bottom and are referred to as macro-pores. Overall, the colour of the soil doesn't change, suggesting a homogeneous texture. What is visibly changing is the structure (size of cracks and pores) and the percentage of stones. The vertical grooves that we see at the bottom are traces of the mechanical shovel used to dig the hole. The photograph was taken in May 2014.

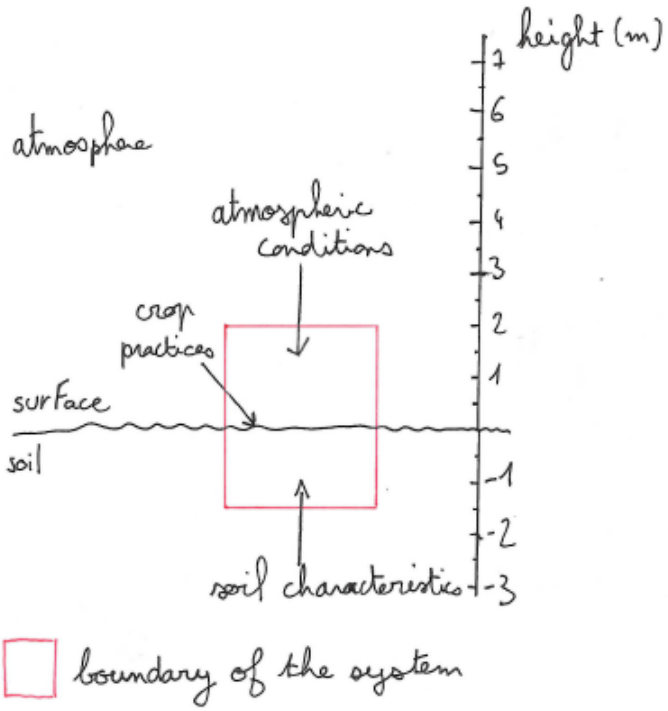

FIGURE 5: The soil system, a simplification of the soil represented in figure 4. The soil is considered homogeneous, without macro-pores. Available atmospheric data can be used to compute energy exchange between soil surface and atmosphere. Air temperature, humidity, wind speed are measured by the weather station at 2 m above ground, thus defining the upper limit of the system. Farming practices modify the system by determining which period needs to be studied, the presence of residue or mulch and the irrigation. Soil characteristics are defined as the soil texture and the thermal and hydraulic properties linked to that texture. The soil profile depth is set to 1.5 m but can be longer if needed.

soil profile. Some rocks and root debris are also visible. The colour of this soil profile is uniform, with no distinct layer of different types of soil. After analysis, the soil is homogeneous in texture (composition of various minerals) and thus easily represented as a system.

We describe our soil system (figure 5) as:

- The system is surrounded by an agricultural environment: influencing factors are weather, soil properties and farming practices (irrigation, fertilisation, soil management)
- The environment affects the system, but the system doesn't affect the environment

- The system is a one-dimensional profile of homogeneous soil, from the surface to a depth noted D.
- The surface is bare without vegetation
- There are micro-pores but no macro-pores

The assumption of homogeneous soil is realistic for our problematic of simulating the seedbed which is a shallow layer of soil, and the soil is not likely to change within this shallow soil layer. This assumption is however questionable for the simulation of water dynamics over the full length of the soil profile.

The consideration of macro-pores extends the complexity of the model to a point that is irrelevant to our focus on the seedbed and the emergence of seedlings. Models taking into account macro-pores are so-called dual porosity models (Jirka Simunek et al., 2003), which are mainly used in concern with pollutant transports within large volume of soils.

The modelling approach is essential to improve our understanding of the complex soil-plant-atmosphere system. The scientific community needs to integrate research results into whole-system models that can help improve crop management. The system/model approach is necessary to analyse and interpret results from experiments that deal with several effects within changing environments (Ma et al., 2001).

2.3 Combination of soil and emergence models

As a consequence of this diversity of approaches, simulated soil temperature is seldom used in whole plant cycle models. However, the meteorological inputs necessary to compute energy balance are nowadays readily available on locations equipped with a standard weather station which measures radiation, air temperature and humidity, wind speed and rainfalls.

Many models still use air temperature to predict emergence (Jamieson et al., 1998) because of the lack of soil temperature data. In a sunflower model, the error of prediction on the duration of the sowing-flowering period is mostly due to errors relating to the sowing-emergence phase (Casadebaig et al., 2011) because the authors used air instead of soil temperature.

Without accurate prediction of the timing of each phase, the chance of errors between simulation and observation increase. When the emergence is predicted too late, there is a shift in the prediction of the development and flowering can happen much later than what is observed, which will increase

model error (Jame and Cutforth, 2004). Given that the error tends to be largest for the sowing to emergence phase, it is natural to focus on it to improve crop models.

There are already some published models that emphasise the effects of the environment on germination or emergence (W. Finch-Savage, Rowse, and Dent, 2005; W. E. Finch-Savage and Bassel, 2015). These models use either thermal time model with air temperature, with a linear variation of seed development rates for specific ranges of temperature, while some others have adopted a non-linear temperature response curve, like the Beta curve (Jame and Cutforth, 2004; H. Wang et al., 2009). The complexity of simulating seedbed temperature has pushed most models to use air temperature instead of soil temperature (Jame and Cutforth, 2004) or to use a statistical relationship between air temperature and soil temperature (Dwyer, Hayhoe, and Culley, 1990).

The SIMPLE (SIMulation of PLant Emergence) model of C Durr et al. (2001) combines a soil physical model, a seedbed structure model and an emergence model (germination plus elongation) for sugarbeet. The physical model is able to predict soil crusting, whereas the seedbed structure help to estimate a hypocotyl path. Finally, the emergence model computes the germination and emergence time. The structure of the seedbed is rarely taken into account in emergence models and SIMPLE model stays a reference in the area. However, this model needs many inputs to run, inputs which are not easily obtainable and hence reduce its potential application to a large scale.

The prediction of soil conditions driving plant emergence has received particular interest from weed scientists to predict the emergence of weeds (Forcella et al., 2000). Indeed, knowledge of soil conditions allows prediction of the timing of their emergence, which can then be used in recommendations for herbicide applications. To predict the soil conditions for emergence, the authors either use direct measurement from soil temperature and moisture sensors, or use models to obtain the conditions at a particular depth. For temperature, some models just extrapolate from soil measurements done at another soil depth (Roman, Murphy, and Swanton, 2000) or use more or less complex models to infer the soil temperature from air temperature.

Soil temperature is estimated in some some process-based crop models such as CERES-Maize or STICS from air temperature coupled with an amplitude parameter (J. W. Jones et al., 2003; Mary et al., 2009). It is then used to accumulate growing degree days and to predict the time to emergence. Once the sum reaches a given threshold, the algorithms of biomass accumulation

and light interception start. The period after emergence can be described with various levels of complexity, from simple reservoirs (process-based models) to complex functional-structural plant models (FSPM) which focus on plant architecture and light interception (Fourcaud et al., 2008).

Because the seed is planted close to the surface, we need to precisely model the surface temperature to expect realistic seedbed conditions. To obtain surface temperature, the approach of choice is the energy balance approach (John Monteith and M. Unsworth, 2013). This approach has seldom been used in process-based plant models, probably due to its complexity, but has already been used in FSPM to predict with accuracy the temperature of the apex when it is close to the ground (P Cellier et al., 1993; Fournier and Andrieu, 1998).

The study of Weaich, Bristow, and Cass (1996) coupled a soil energy balance model, a soil strength model, and a mechanistic coleoptile elongation model to study the effect of high temperature on maize emergence. For wheat (*Triticum aestivum*), simulated soil temperature has been shown to either be a better predictor of emergence than air temperature (Bullied, P. R. Bullock, et al., 2014), or to present no added value (H. Wang et al., 2009) depending on the accuracy of the soil temperature model. The first reference (Bullied, P. R. Bullock, et al., 2014) uses a complex energy balance model named SHAW (G. N. Flerchinger, 2000) whereas the second (H. Wang et al., 2009) use a simple relationship between soil and air temperature (J. W. Jones et al., 2003).

Such examples seem to prove that, even if it is obvious that using air temperature instead of soil temperature causes errors in the prediction of the establishment phase, it is however not sufficient to use a simple model predicting soil temperature to overcome this issue. Soil temperature prediction must be realized accurately and with great care, which translates into complex biophysical modelling. Certainly one reason why it has not been much developed and used in crop models and agronomical studies so far.

3 Soil model: main equations

3.1 Near surface soil modelling

Energy balance is not the only approach to estimating soil temperature. Soil temperature follows a sinusoidal pattern over the diurnal cycle. Parton and Logan (1981) developed a sinusoidal model that predicts diurnal soil temperature variation at any depth provided that the depth at which there is no

more variation of soil temperature is known. However, this approach is more widely used as a tool to know the amplitude of temperature at a certain depth rather than its exact variation (Campbell and Norman, 1998).

The energy balance approach belongs to the domain of soil-vegetation-atmosphere transfer (SVAT) models. Amongst these models, several approaches exist that are related to the number of layers represented in the soil and the number of above-ground elements (shrubs, trees, sky). In our study, the energy balance is considered for a bare soil surface, since at the beginning of the crop cycle, the soil is bare, and the seedlings are so small that they do not yet affect the energy balance. At the beginning of the crop cycle, the seeds absorb negligible amounts of water compared to the soil's water reserve.

In agricultural research, energy balance is used with various degrees of simplification. In agronomy, the Penman-Monteith equation that computes reference evapotranspiration is a simplified version of the energy balance equation (I. A. Walter et al., 2000). In bioclimatology (a branch of micrometeorology applied to plants), its use is widespread, mostly to estimate the transpiration of canopies and stomatal conductances at various scales (Shuttleworth and Wallace, 1985; Steduto and Hsiao, 1998). It is also used to predict the temperature of organs to have a better estimation of phenology (Guilioni et al., 2000) or to estimate limiting factors of epicotyl elongation (Weaich, Bristow, and Cass, 1996).

There are published models that predict the soil temperature and moisture with an energy balance at the surface. The SHAW model (G. N. Flerchinger, 2000) or HYDRUS-1D (J Simunek et al., 2013) are among the best known. Both models have different purposes: SHAW is mainly designed to simulate soil freezing and thawing, with several layers of residue, plant cover, snow and vegetation. On the other hand, HYDRUS-1D is more focused on solute and pollutant transport within a deep soil. It has a complex water and solute transport part, but also includes an energy balance at the surface.

Models are generally briefly described in the literature. More in-depth information can only be found in biophysics books (Campbell and Norman, 1998; John Monteith and M. Unsworth, 2013; Daniel Hillel, 2003). Although one can find detailed equations of soil and atmosphere processes, their translation into a working model is never at hand. To bridge this gap between equations and model, the book of Müller (1999), *Modelling Soil-Biosphere Interactions* is a convenient guide to develop a soil model. The book is accompanied by the models in the software package Model Maker. The availability of the mode allows to play with, test the model and see how the equations are

connected to each other. We use this model as a reference in our work.

Indeed, soil modelling is a vast research area because of its many implications for human activities, which span topics as varied as agriculture, waste management, resource management, building construction or remediation (Bartolo et al., 2011). In this introduction, we will briefly show the basic equations for modelling the three parts of the soil we are interested in : the surface conditions, the movement of heat and the movement of water. We have restricted our literature review to models describing one-dimensional flows over a few meters, which is sufficient to adequately capture the variability of soil temperature and water content movement near the surface.

3.2 Soil surface conditions

Our objective is to obtain the temperature and water content near the soil surface because this is where seeds are planted. Hence, we need an accurate prediction of energy exchanges happening at the surface to have a good estimation of the temperature and evaporation (John Monteith and M. Unsworth, 2013), and this is the purpose of the energy balance equation.

$$Rn + H + G + LE = 0 \tag{1}$$

The four terms in the energy balance equation represent the four types of energy exchanges at the soil surface: (i) Rn [Wm^{-2}] stands for net radiation and represents the radiative energy exchange, (ii) H [Wm^{-2}] is the convective heat exchange between surface and atmosphere, (iii) G [Wm^{-2}] is the conductive heat exchange from surface to deeper soil and finally (iv) LE [Wm^{-2}] is the latent heat exchange, which is, for a bare soil, the evaporation of water. A representation of these fluxes during the day is shown in figure 6.

The advantage of the energy balance equation is that it relies on physics equations that are universal. Therefore, this formalism can be applied to a wide variety of environments and contexts (weather prediction, building construction, military or agriculture, (Bartolo et al., 2011)). The disadvantage is its complexity because each term is itself a model that simplifies complex physical processes. For example, the effect of wind and air turbulence on heat transport as it is used in most energy balance formalisms is a simplification of fluid dynamics, which is still a topic of discussion (Bittelli, Ventura, et al., 2008).

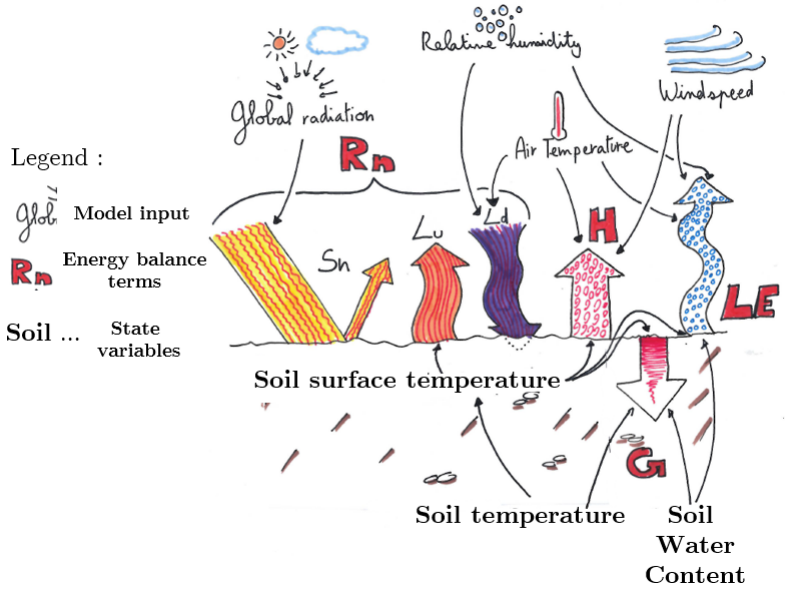

FIGURE 6: The four energy fluxes of the energy balance equation placed in the context of soil temperature prediction and the available input for the model (handwritten). The figure shows the energy balance during the day (the net short-wave radiation (Sn) is high). The energy fluxes are shown in red, and the three state variables of the model are in printed characters. The three state variables are soil surface temperature, soil temperature and water content. They will change according to the energy fluxes. The energy fluxes will be computed with the atmospheric inputs: global radiation, relative humidity, air temperature and wind speed. Rain is omitted because it does not participate directly in the computation of the energy fluxes. These are Rn, net radiation; H, sensible convective heat flux; G, sensible conductive heat flux and LE, latent heat flux. Net radiation is decomposed in three terms: Sn, net short-wave; Lu, long-wave upward; Ld, long-wave downward as $Rn = Sn + Ld - Lu$

This complexity has pushed many scientists to rewrite the energy balance equation with their symbols and notation and simplify some parts to tailor the model to their applications. In this dissertation, we are not immune to this tendency and, although we have used the most common terms present in the literature, we have still rewritten the energy balance model in our terms and adapted the model to our situation by simplifying some assumptions. This complexity also makes rewriting necessary for understanding the energy balance model. Rewriting also gives an opportunity to write the model using a clear mathematical expression, which provides an appropriate starting point for studying the model's properties in depth. One key property of the model is the numerical scheme that we will choose to solve the system of equations.

3.3 Transport of heat within the soil

The energy balance equation gives the temperature of the soil surface and its evaporation rate. The transfer of heat within the soil itself is modelled by the continuity equation for heat flow (Campbell and Norman, 1998)

$$Ch \cdot \frac{\partial T}{\partial t} = \frac{\partial FT}{\partial z} \tag{2}$$

This partial differential equation relates the change of heat ($Ch \cdot \partial T$) over time t to the change of flow of heat (∂FT, [$\mathrm{W\,m^{-2}}$]) over depth z. Heat is expressed in [$\mathrm{J\,m^{-2}}$] by multiplying temperature T [$^{\circ}\mathrm{C}$] with the volumetric heat capacity Ch [$\mathrm{J\,m^{-3}\,{}^{\circ}C^{-1}}$].

The expression of flow of heat FT known as Fourier's law, is expressed as (Campbell and Norman, 1998)

$$FT = KT \cdot \frac{\partial T}{\partial z} \tag{3}$$

The thermal conductivity KT [$\mathrm{W\,m^{-1}\,K^{-1}}$] changes according to the flow of water and vapour (Daniel Hillel, 2003). Indeed, water is a better thermal conductor than air, and a wet soil will conduct heat more rapidly than dry soil.

How heat diffuses into the soil is modelled in the same way for most soil models having a temperature component.

3.4 Transport of water within the soil

Water transport in the soil is a complex process. Water flows through the micro- and macro-pores of the soil, alternating between liquid and vapour forms. As the soil becomes drier, water becomes more tightly bound to the mineral constituents of the soil with a force known as water potential. The water potential is denoted by Ψ in reference books (Daniel Hillel, 2003). A dry soil has a highly negative water potential (i.e. well below zero) whereas a wet soil, with water moving around smoothly, has a water potential close to zero.

According to soil type, the relationship between water content and water potential changes. For a given water potential, the water content differs whether the soil is richer in clay or in sand.

In developing the soil model, the use of the equation known as the Richards equation was beneficial, as it conveniently expresses the movement of water in the same way as the flow of heat. To model the diffusion of water in the soil, the change of water content can be expressed [$\mathrm{m}^3_{\mathrm{H_2O}}\ \mathrm{m}^{-3}_{\mathrm{soil}}$] in a continuous equation that has the same form as heat diffusion (equation 2) :

$$\frac{\partial \theta}{\partial t} = \frac{\partial FW}{\partial z} \tag{4}$$

The flow of water FW [$\mathrm{m\,s^{-1}}$] is expressed according to Darcy's law. Henry Darcy was a French engineer who discovered that water flows from high water potential (close to zero) to low water potential (far from zero) :

$$FW = KW(\theta) \cdot \frac{\partial \Psi}{\partial z} \tag{5}$$

In this equation, water conductivity KW [$\mathrm{m\,s^{-1}}$] has a strong dependence on the water content θ, which makes it more mathematically complex than the equation for heat (Campbell and Norman, 1998). As a result, the Richards equation does not manage to capture all the variation of water movement. Therefore, several authors have tried to improve it, notably by taking into account the differential flow of water outside and inside soil pores in dual-porosity models (Jirka Simunek et al., 2003)

3.5 Model inputs

Our goal was to develop a model capable of running with readily available inputs. Information on temperature, wind speed, radiation, rainfall and humidity are collected by most weather stations and are publicly available

in many locations. Were these variables not available, global convection models (GCM) make a satisfactory second choice. These can interpolate weather conditions for any point on the planet (for example the ECMWF model (ECMWF, 2014)).

The complexity of the model has to match the available data. With five input variables (see an example of them in figure 7), we have to keep in mind that our model will have to approximate actual reality.

3.6 The modelling challenges

The challenge of the energy balance model development is to get the right level of complexity to solve our initial problem of temperature and water content near the surface to predict emergence. Since a large variety of different expressions of soil energy balance models exists, the choice of the formalism best adapted to the site of application was essential. In addition, several points are still the matter of debate. For example, the formulation of the effect of wind in the transfer of heat between surface and atmosphere is very complex, and no consensus exists on which formalism needs to be applied (Bittelli, Ventura, et al., 2008).

Another challenge relates to the choice of the numerical scheme to resolve the partial differential equations. In addition to environmental physics, simulation of the energy balance requires knowledge in partial differential equations and numerical analysis. The soil energy balance model has particular boundary conditions at the surface which make the application of a standard resolution algorithm impossible, and calls for a tailored solution (Ross, 2003)

Finally, even though sensitivity analysis represents an essential step in model development, it is rarely recorded in the literature on soil energy balance. In the case of complex models for which there is a wealth of interactions, the application of global sensitivity analysis method can be difficult (Cournede et al., 2011). The goal here was to find a suitable global sensitivity analysis method that helps understand the behaviour of the model, supports its calibration and identifies points to be improved.

4 The scientific plan

To support the development of the soil model, we installed a weather station near our experimental station in Stein in north-western Switzerland. The

FIGURE 7: An example of the five sensors that record data to use as input for the energy balance model. ① global radiation sensor; ② ③ combined air temperature and relative humidity sensor; ④ wind speed sensor; ⑤ rain gauge

station records the input variables needed for the model simulation, as well as other variables useful for model evaluation and calibration: soil temperature, soil water content and net radiation. The first chapter (chapter 1, page 21) will present the weather station and a concise analysis of its two-year record of variables. As such a data set is rarely found in the literature, we chose to show a descriptive analysis of soil temperature before starting the model development. This study helps underline the essential aspects of soil temperature variations across the diurnal cycle (day and night) and seasons. Chapter 2, page 39, will take us into the core of the thesis: the description of the soil model. First, we shall define the system we wanted to model. Then will come the description of the biophysics equation that translates the physical phenomena within our system. Next, we shall clarify the explicit numerical scheme that we have chosen to perform a model simulation. To assess the results of our model development, we shall compare our model version with a simplified model version of Müller (1999). Finally, we shall show the effect of different soil textures on the soil model and explain how we calibrated the model to suit the soil located directly under the weather station.

In chapter 3, page 93, a high number of model simulations are performed in order to find the parameters that most affect the two main model responses: soil temperature and soil water content at 5 cm. A first exploration of the parameters is described with a simple approach known as the One-At-A-Time approach : each parameter was allowed to vary in turn so as to gain a primary visual feel for the variation of the model response. In order to obtain a global sensitivity index for each parameter, the Morris sensitivity analysis method was applied and the parameters were ranked according to their direct effect on model output. The Morris sensitivity analysis was completed by the variance-based approach of the Sobol method. This permitted a better characterisation

of the parameters that have a strong direct effect and of the ones which have a strong effect only in interaction with other parameters.

The parameters identified as relevant were then calibrated using either measurement from the weather station or an estimation procedure to find their right value. Finally, an evaluation of the model performance was realised in several situations over the years 2015 and 2016.

Once the performance of our model was demonstrated, its output was applied in predicting the emergence of maize. The results are presented in chapter 4, page 157. During 2015 and 2016 maize seeds were planted at various times of the year to generate different climatic conditions. The data about their emergence was collected and served to infer cultivar specific thermal time models. These were used in combination with the soil temperature simulation to predict emergence.

Chapter 1

Analysis of soil temperature in Stein, Switzerland

The development of a soil temperature model based on the energy balance first and foremost requires adequate data for its calibration and evaluation. To this end, a weather station was installed near the Syngenta Field Station in Stein. To the day of writing the final copy of this thesis, it still records the soil temperature and moisture, along with radiation and air temperature, moisture and wind speed. For the purpose of our study, we included data collected every 15 minutes from June 2015 to January 2017. The wealth of information gathered constituted the source from which we extracted global tendencies and explored correlations.

In this chapter, we shall describe and analyse this data with a focus on the variability of soil temperature and its relationship with other atmospheric variables.

1 Description of weather station sensors

1.1 Sensors

A weather station was installed in June 2014 by Meteotest (Bern, Switzerland) with sensors from Adcon Telemetry GmBH (Klosterneuburg, Austria). Its geographic coordinates are latitude 47.4532 and longitude 3.5683. It is installed on a plot of bare soil, on a small hill of disturbed, homogeneous loamy soil. The soil is kept bare with regular tilling and herbicide applications from July to October, when weed growth is hard to control via tilling only. The maintenance of the soil is done by the team of the Syngenta Field Station in Stein. The installation site is close to the Rhine river, which leads to high humidity levels, especially at night.

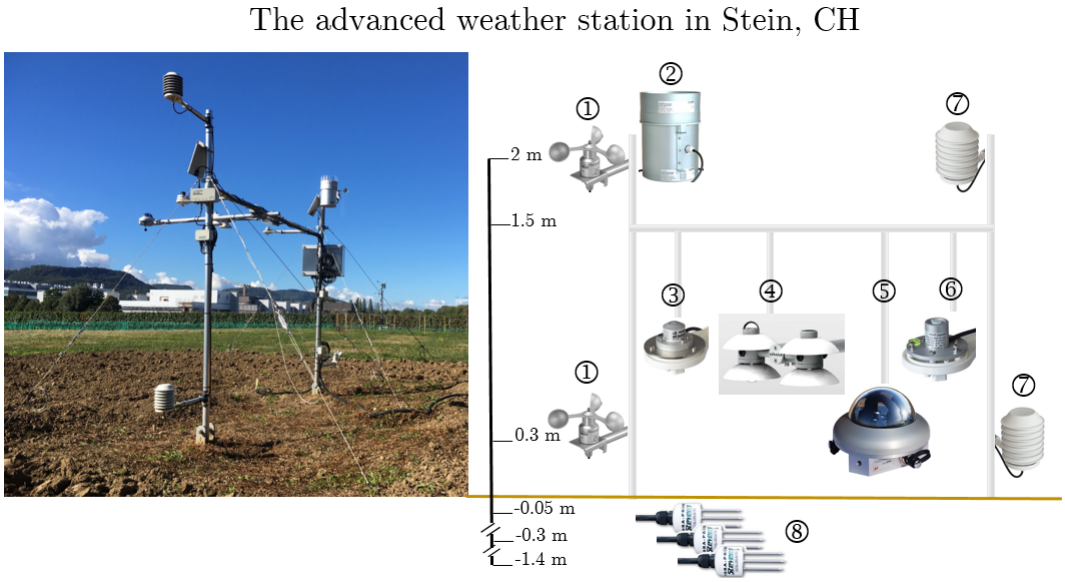

FIGURE 1.1: Left: the advanced weather station used to collect the data analysed in this chapter. It is installed on a plot of bare soil near the Syngenta Field Station in Stein, Switzerland. Photo taken on October 3rd, 2016. Right: schematic representation of the weather station and its sensors. ① Cup anemometer placed at 0.3 and 2 m, ② rain gauge with tipping bucket, ③ silicon pyranometer, a standard short-wave radiation sensor for agricultural use, ④ net radiometer, a montage of four radiation sensors that capture short- and long-wave radiation from the soil and the atmosphere, ⑤ diffuse radiation sensor, measures the part of short-wave coming from sources other than the sun, ⑥ PAR (Photosynthetically Active Radiation) sensor, ⑦ combination of air relative humidity and air temperature sensors, ⑧ soil water content and temperature sensor.

Figure 1.1 and table 1.1 present the sensors installed in the weather station. These are weather sensors commonly used for agricultural purposes (radiation, rain, temperature, humidity, wind speed) as well as more specific sensors. Indeed, the particularity of the dataset lies in the specific radiation and soil sensors installed in the weather station. Such data is not often available in classic soil temperature studies (eg. Banimahd and Zand-Parsa (2013)). It enables us to gain better understanding of the processes driving its variation.

Among the specific sensors, the net radiometer NR01 from Hukseflux (④ in figure 1.1) measured the four components of the surface radiation balance : it is a sensor composed of four radiation sensors, two sensors facing the sky and two sensors facing the soil. On each side are two radiation sensors sensitive to different wavelengths. One sensor, a pyranometer, measured short-wave radiation ($300\,\mathrm{nm}$ to $2800\,\mathrm{nm}$) while the other, a pyrgeometer, measured long-wave radiation ($4500\,\mathrm{nm}$ to $50\,000\,\mathrm{nm}$). The pyrgeometer differs from the pyranometer in that it is covered with a small glass dome that prevents short-wave radiation from reaching in. In environmental physics, radiations are separated into short and long components : short-wave radiation corresponds to solar radiation, which is an important input of energy to the soil during daytime, while long-wave radiation corresponds to infra-red radiation, which has a more prevalent role during energy exchanges at night.

Below the soil surface, three soil temperature/soil moisture sensors were placed at $5\,\mathrm{cm}$, $30\,\mathrm{cm}$ and $140\,\mathrm{cm}$. These depths are of agronomic relevance : 5 cm gives information on temperature and moisture of the seedbed (the layer of soil where the seeds are planted), 30 cm is pertinent for the nitrogen cycle and 140 cm is suited to the assessment of the temperature and moisture at the bottom of the soil profile - the boundary of our system.

1.2 Data processing

Data from the weather station was retrieved from the AdCon ADDvantage Pro 6.5 portal accessed online. The raw dataset includes measurements collected by the sensors with a 15-min interval for over 2 years, between 15 June 2014 and 2 January 2017 (928 days and 9 hours). It has 90613 rows with 46 variables in total for each row and weighs 13 MB. After download, the data was first used to compute some additional variables using the equation gathered in table 1.2. Note that tables 1.1 and 1.2 summarise all the available data collected by the weather station.

TABLE 1.1: List of sensors installed on the advanced weather station : name, technology and manufacturer

Sensor	Technology	Name	Manufacturer
Radiation	Silicon pyranometer	SP Lite2 pyranometer	Adcon Telemetry, Austria
Temperature, relative humidity	PRT*/capacitive	SEN-R Combisensor TR2	Adcon Telemetry, Austria
Wind speed	Cup anemometer	Pro 10/2	Adcon Telemetry, Austria
Precipitations	Tipping bucket	RG1	Adcon Telemetry, Austria
Net radiometer	Four thermopile	NR01	Hukseflux, The Netherlands
Diffuse radiation	Thermopiles	SPN1	Delta-T, USA
PAR*	Photodiode	PAR1	Adcon Telemetry, Austria
Soil temperature/moisture sensor	Hydra Probe II	PRT/dielectric permittivity	Stevens, USA

*PRT : platinum resistance thermometer, PAR : Photosynthetically active radiation

TABLE 1.2: Additional variables computed from the data retrieved online

Variable name	Symbol	Unit	Equation
long-wave upward	Lu	$\mathrm{W\,m^{-2}}$	$LW_{up} + \sigma \cdot (T_{pyrgeometer} + 273.15)^4$
long-wave downward	Ld	$\mathrm{W\,m^{-2}}$	$LW_{down} + \sigma \cdot (T_{pyrgeometer} + 273.15)^4$
Surface Temperature	$T_{surface}$	°C	$(Lu/\sigma)^{1/4} - 273.15$
Sky Temperature	T_{sky}	°C	$(Ld/\sigma)^{1/4} - 273.15$
Albedo	$albedo$	-	SW_{up}/SW_{down}
Saturated vapour Pressure	e_s	kPa	$0.611 \cdot \exp\left(\frac{17.27 \cdot T}{T + 240.97}\right)$
Ambient vapour Pressure	e_a	kPa	$e_s \cdot RH$
Vapour Pressure Deficit	VPD	kPa	$e_s - e_a$

Legend: σ: Stefan-Boltzmann constant; $T_{pyrgeometer}$: Temperature of pyrgeometer, the long-wave radiation sensor; SW_{up},SW_{down}: output of the pyranometer, the short-wave radiation sensor receiving upward (up) or downward (down) radiation; $T_{surface}$: the radiative temperature of the soil surface, ignoring the emissivity factor; T_{sky}: the radiative temperature of the sky, ignoring the emissivity factor; LW_{up},LW_{down}: output of the long-wave radiation sensor receiving upward or downward radiation; T: air temperature in °C; RH: air relative humidity

In order to compute the first four additional variables, we combined the Stefan-Boltzmann law with the output from the net radiometer. The Stefan-Boltzmann law (1884) expresses the radiation flux density RFD ($\mathrm{W\,m^{-2}}$) emitted by a body with a surface temperature of T (K or °C + 273.15) by :

$$RFD = \epsilon \cdot \sigma \cdot T^4 \tag{1.1}$$

In the above equation, σ is the Stefan-Boltzmann constant ($5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$), ϵ is an emissivity factor equal to 1 for a black body (or perfect emitter), and is lower than 1 for other materials. The RFD emitted by the pyrgeometer itself, $RFD_{pyrgeometer}$, can be computed using the temperature $T_{pyrgeometer}$, measured by a thermometer installed within the pyrgeometer of the net radiometer sensor. The output of the net radiometer, denoted by LW, may then be corrected by adding the $RFD_{pyrgeometer}$ to obtain a range similar to what is found in the literature (John Monteith and M. Unsworth, 2013).

By inverting the Stefan-Boltzmann law, the sky temperature (T_{sky}) and soil surface temperature ($T_{surface}$, table 1.2) can be computed from the corrected RFD, Ld and Lu. An issue arises when inverting the Stefan-Boltzmann law: the emissivity ϵ values were unknown and therefore taken at a default value of 1. As a result, the soil surface and sky temperatures obtained this way are probably under-estimated and can therefore only be considered as a mere indication of the magnitude of soil surface temperature, but cannot be used to measure an error of prediction when we will later develop the model. The measurement of soil emissivity requires emissivity plates (Ham and Senock, 1992). Such a device was not available for this study and the process necessary to obtain the value of emissivity requires a dedicated work that was impossible to carry out during the time frame of this thesis.

The last variable computed with the net radiometer was the albedo, or short wave reflection coefficient, which is defined as the proportion of downward short-wave radiation that is reflected back to the atmosphere. The albedo will be studied in chapter 3, section 2.4 where we will see that its calibration is fundamental to the improvement of the model's performance.

The next additional variables use relative humidity (RH). The relative humidity *per se* is seldom considered as an environmental variable (Campbell and Norman, 1998), but is necessary to obtain ambient vapour pressure (e_a) and vapour pressure deficit (VPD) as functions of air temperature T (see equations in table 1.2). To this end, the saturated vapour pressure e_s, which

is the maximum vapour pressure allowed by the air temperature, was computed with the Tetens formula (table 1.2). Then the ambient vapour pressure was simply obtained by multiplying e_s with relative humidity RH. Finally, the vapour pressure deficit VPD was calculated as the difference between saturated and ambient vapour pressures. This gives an indication of the evaporative demand of the atmosphere : the higher it is, the more water will be lost through either evaporation (soil) or evapotranspiration (plants).

For specific visualisation or computation purposes, those variables (those directly obtained from the measurements and those additionally computed) can be averaged out over specific time periods. In this chapter, we chose to average the 15-minute raw data into hourly data in order to obtain clearer, less clustered graphs for visualisation. Herafter, each data point thus represents the average of 4 raw measurements, except for the rainfall variable that represents the cumulated rainfall amounts for each hour.

1.3 Soil texture

In Stein, soil texture was measured by granulometric analysis from samples taken during the installation of the soil sensors at 5, 30 and 140 cm. The soil samples were analysed by the SADEF, an external soil analysis laboratory in France. For texture, two independent measurements were done on each sample and the table 1.3 presents the average of the two measurements, weighted so that clay, silt and sand account for 100 % of the solid phase.

The texture analysis shows that the samples for 5 and 30 cm were similar, belonging to the Clay Loam texture class, whereas the deep soil at 140 cm was sandier and belonged to the Loam texture class (Soil Survey Division Staff, 1993).

2 Data analysis

2.1 Temperature profile in air and soil

The data collected made it possible to first plot the temperatures across the soil and atmosphere in a given day. The graph in figure 1.2 shows temperature data for 1 May 2015. It displays the vertical profile of temperature across air and soil, obtained by using temperature sensors placed along a vertical axis. The profile is plotted at 01:00, 04:00, 07:00, 10:00, 13:00, 16:00, 19:00 and 22:00

TABLE 1.3: The texture measured by SADEF for 6 soil samples (3 for each depth) in Stein below the weather station. The mineral fraction (clay, silt and sand) are given in percentage of the solid phase. The mineral fraction accounts for 100 % of the solid phase. The organic matter (OM) is given as a percentage of the total soil volume (solid, liquid and gaseous phase)

Fractions	sample 5 cm	sample 30 cm	sample 140 cm	Texture class (USDA)
clay [$< 2\mu m$]	27 %	26.7 %	19.0 %	Clay Loam
silt [$2 - 50\mu m$]	36.7 %	36.2 %	31.4 %	Clay Loam
sand [$50\mu m - 2mm$]	36.3 %	37.1 %	49.6 %	Loam
Organic matter	3 %	2.5 %	2.2 %	-

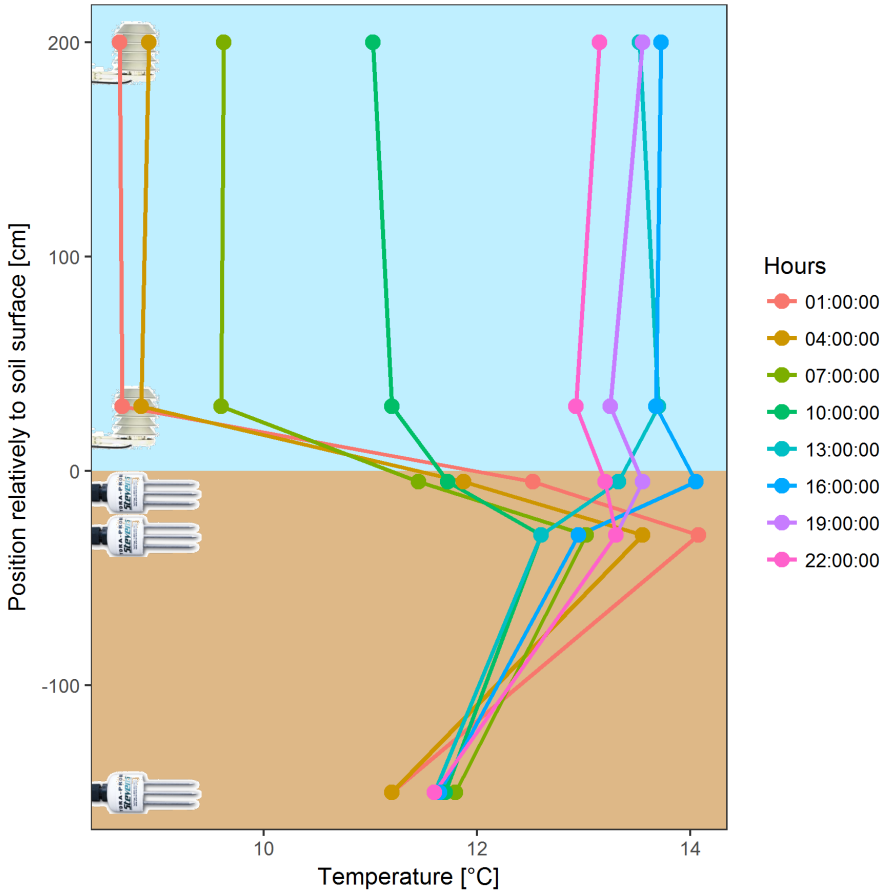

FIGURE 1.2: Representation of the vertical profile of temperature on 1 May 2015. Under the weather station, sensors were placed at 5, 30 and 140 cm below soil surface. This figure is a snapshot of temperatures across different heights and depths, at different times of the day. The plot shows that air temperature held the greatest variability, while strong temperature differences resided between air temperature and soil temperature near surface at nighttime

in order to visualize the differences between the temperature gradients at different times of the day.

Vertical profiles Figure 1.2 reveals that the temperature profiles are not vertical but are more akin to a zigzag : at no moment of the day was there a uniform temperature across soil and atmosphere. At 01:00, the temperature in the air was at its minimum, while it was at its maximum at 30 cm in the soil. At this time, the difference between these two points was greater than 5 $^{\circ}\mathrm{C}$. The temperature profile was closer to vertical at 22:00, with a difference of only 0.5 $^{\circ}\mathrm{C}$ between 2 m overground and 30 cm underground. On the one hand, large differences between soil and atmosphere temperatures occurred in the morning (01:00, 04:00 and 07:00), between +30 and -30 cm of depth. On the other hand, the overall profiles were closer to vertical in the afternoon. The substantial difference in temperature around the soil surface (between +30 and -30 cm) at night was due to the soil having accumulated some heat during the day. At night, this difference created a temperature gradient, which drove the intense exchange of heat between soil and atmosphere. This exchange phenomenon or flux of heat has a strong impact on the change in soil temperature and needs to be taken into account if we are to accurately reproduce soil temperature variations.

Diurnal variation of soil and air temperature Another interesting insight emerges from interpreting and comparing in terms of physical processess the respective variabilities of soil and air temperatures during the day. For instance, for the considered day (1 May 2015), air temperature ranged from 9 to 13.5 $^{\circ}\mathrm{C}$ whereas soil temperature at 5 cm ranged only from 11.5 to 14 $^{\circ}\mathrm{C}$. Air temperature has a greater variability than soil temperature because soil is a system that stores heat and is more resistant to the flow of heat than air. In the soil, heat is transported via conduction between successive layers where part of the heat is stored. In the air, the variability of temperature is similar for 2 m and 30 cm above soil surface. This is a consequence of the action of eddies, circular currents that transport heat through the air, resulting in a good and rapid mixing of the heat and an uniform distribution between 2 m and 30 cm. Finally, at any hour of the day, the temperature profiles almost converge at 140 cm in the soil: the soil temperature does not show any diurnal variation. This provides us with useful information about the boundary conditions of our systems, always a crucial point when modelling transport or diffusion phenomena. Here we can assume that the boundary condition at the bottom of

the system is stable within a day. In general, the temperature of soil between 1.5 to 2.0 m (depending on the type of soil) is close to the average of annual air temperature. The main challenge will lie in modelling the temperature variation at the surface.

Gradients in the air, at the surface and in the soil according to season and diurnal cycle In the previous paragraph, we have chosen to focus on a particular day for the sake of clarity but the pattern of gradients changes over the seasons (figure 1.3). To show this change, we have computed three differences of temperature for each hour in the dataset : $\Delta\mathrm{Air}$, the difference between air at 200 cm and 20 cm, $\Delta\mathrm{Surface}$, the difference between air temperature at 30 cm and soil temperature at 30 cm and $\Delta\mathrm{Soil}$, the difference between soil temperature at 30 and 140 cm. We have then compared these three differences.

For many hours in winter, the maximum Δ is for $\Delta\mathrm{Soil}$, meaning that it is in the soil that the gradient of temperature is the highest, at any time of the diurnal cycle. This is also observed in Autumn, but it is only during the day that the maximum gradient is found in the soil. The contrary happens during summer and especially during summer nights, when the maximum is $\Delta\mathrm{Surface}$. During summer nights, the heat stored in the soil during the day makes the surface gradient of temperature more important. $\Delta\mathrm{Air}$ is never the maximum of the 3 gradients, there are very low gradients between 20 cm and 2 m for any given day of the year, due to the efficient mixing of air by eddies. For summer and spring nights, the largest differences occurs at the soil surface.

In the soil-atmosphere system presented in figure 1.2, gradients in the soil are most important in winter and at the soil surface during spring and summer nights. To reproduce these gradients, the modelling of heat transport had to focus specifically on the surface and the transport within the soil. There is no difference in temperature between 20 cm and 2 m above ground.

2.2 Difference between air and soil temperatures at 5 cm is greater in winter

We have seen that the gradients of temperature were largest in the soil in winter and autumn, and at the surface during spring and summer nights. It was then important to find out how much soil temperature at 5 cm (where crop seeds are usually planted) differs from air temperature at 2 m, where temperature is usually measured. In figure 1.4, we show 2 and a half years of

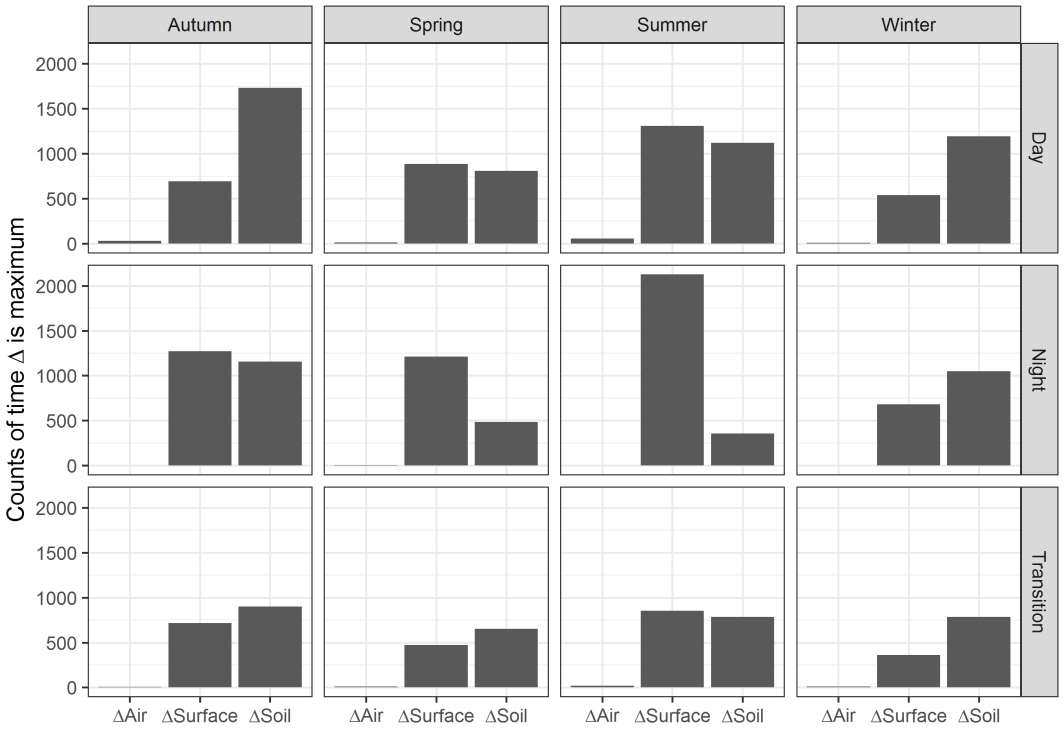

FIGURE 1.3: Number of hours when each of the three temperature gradient is the largest according to the season and diurnal period. The difference (or Δ) of hourly temperature was calculated for 3 horizons : air ($\Delta\text{Air} = |T_{air,200cm} - T_{air,30cm}|$), surface ($\Delta\text{Surface} = |T_{air,30cm} - T_{soil,30cm}|$) and soil ($\Delta\text{Soil} = |T_{soil,30cm} - T_{soil,140cm}|$). The matrix of graphs shows how many times this Δ becomes the highest of the 3 Δs. In effect, this graph shows the position of the maximum temperature gradient for each season and diurnal period. The diurnal periods are separated into: day from 9:00 to 17:00; night from 21:00 to 5:00 and transition for the hours outside these ranges.

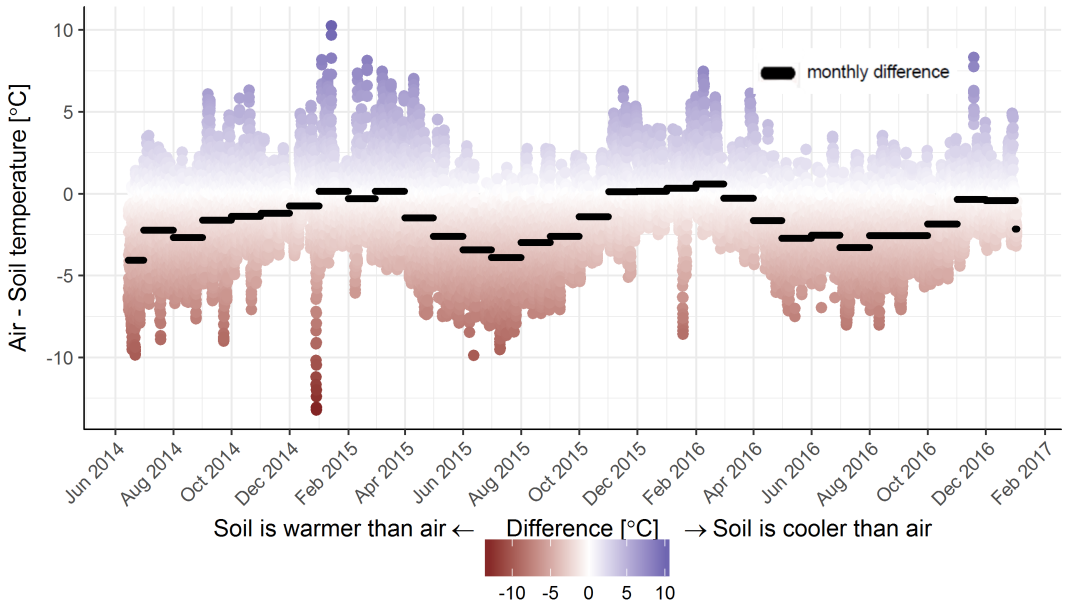

FIGURE 1.4: Difference between air and soil temperature at 5 cm computed for each hour (dots) and for monthly average (dashes). The difference over two years and a half shows a regular sinusoidal pattern, with a soil warmer than air in summer and no difference in winter. For this period and this location, soil is hotter than air in general. The difference computed hourly shows a magnitude of 10 °C of difference. Soil starts to get warmer than air in April for both years

hourly differences between soil temperatures at 5 cm and air temperatures at 2 m. On average over the whole dataset, the difference is -1.6 °C, meaning that the soil at 5 cm is generally warmer than air temperature.

A regular sinusoidal pattern can be observed within the variations of this average difference over a month: the difference is negative in summer (the soil is warmer on average) while it neighbours zero in winter (the soil is as cold as the air). For the 2 years of study, the difference in temperature in March was consistently close to 0. Only in April and May, the conventional sowing months for maize, did the soil become warmer than air. The soil was 1.5 °C warmer than air in April and 2.6 °C warmer in May. Over the summer (July, August and September), the soil was about 2.8 °C warmer.

On an hourly scale, a large variability in the difference appears (figure 1.4). The maximum recorded negative difference (soil hotter than air) was -13.2 °C on 29 December 2014 at 7:00, minutes before sunrise, when air temperature reaches its minimum. The maximum positive difference (soil colder than air) was 10.2 °C and happened on 13 January 2015 at 23:00. On that day, fresh snow had just covered the soil. This prevented the soil from warming up since fresh snow reflects back more than 80 % of solar radiation. The soil stayed cold, which explains the large difference between air and soil temperature.

To summarise, soil and air share on average the same temperature in winter, with a monthly difference close to 0. In winter, on an hourly scale, large temperature differences may also arise, with moments when the soil is much colder than air (since soil could not warm during the day, because of a fresh snow cover) and other moments when it is much hotter (air temperature is very low before sunrise). Soil started to get significantly warmer than air in April. At this point, the monthly difference between air and soil temperature was -1 °C. For summer, this difference increased to -2.8 °C.

2.3 The variability of air temperature explains the variability of soil temperature at 5 cm

The bi-variate scatter plot in figure 1.5 shows the relationship between air temperature at 2 m and soil temperature at 5 cm, 30 cm and 140 cm and the fit with a linear model. For the temperature at 5 cm, a good agreement exists between both variables, with a R^2 of 91, meaning that air temperature variability explains 91 % of soil temperature variability at 5 cm.

With depth, R^2 decreases : the air temperature is not as efficient at explaining soil temperature variation at 140cm as it is at 30cm. Some extreme values

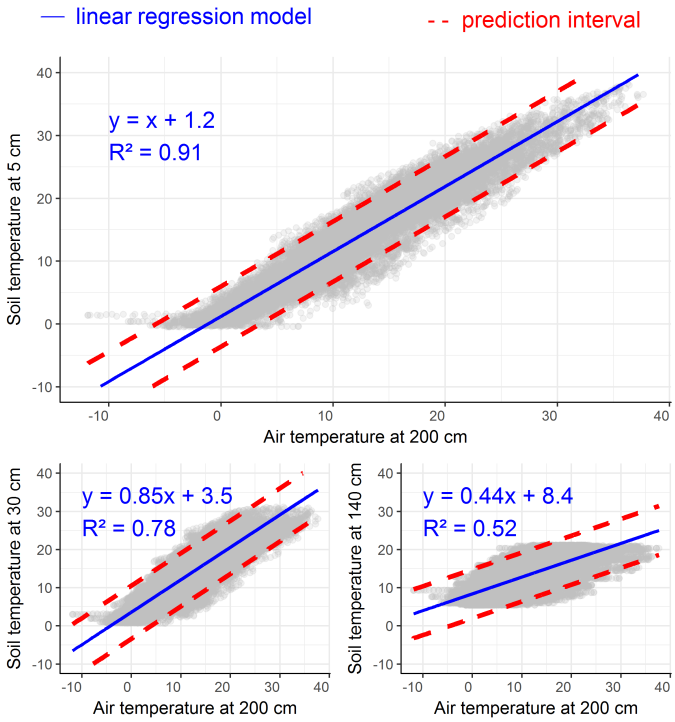

FIGURE 1.5: Scatter plot showing the relationship between hourly air and soil temperature at 5, 30 and 140 cm in Stein, Switzerland, between 15 July 2014 and 2 January 2017. We chose to represent the prediction interval at a level of 95 % to show that the linear model is unable to predict extreme values, both when the soil is hotter and colder (outside the red dashed lines). The confidence intervals are not shown here because they are too narrow due to the high amount of data. The R2 decreases with depth, showing a decreasing relationship of air temperature with soil temperature at greater depth. N=22281

are outside the prediction interval at 95 %. These value would be impossible to predict by the linear model. The statistical model is not able to reproduce these extreme values and that is why we need a model based on physics. Moreover, the statistical model is valid only for this specific location, and extrapolation to other locations may give extremely inaccurate results. In the next chapter, we develop a mechanistic model based on physics to be able to reproduce the extreme temperature differences, the very ones that impact tremendously the early development of crops.

2.4 Correlations of soil temperature with other variables

So far, we have focused on the relationship between soil temperature and air temperature. However, the weather station collects other valuable variables that may be correlated with soil temperature. Such correlations help explore relationships between soil temperature and other variables as well as take into account the most important ones into the development of our model.

In figure 1.6, we show variables that have a significant correlation. The selected variables are soil temperature, air temperature, global radiation, humidity, long-wave radiation upward (emitted by the soil) and downward (emitted by the atmosphere). The presented matrix of bivariate scatter plots conveys a lot of information that requires careful examination. In the matrix, the colour blue represents data collected during the night (21:00 to 5:00 in the morning) while red symbolises data collected during the day (9:00 to 17:00). We have avoided the transition period when sun angles are low and distort the radiation measurements (John Monteith and M. Unsworth, 2013). We have separated the dataset into day and night since, during the day, the high amount of radiation masks the effect of long-wave radiation and other types of energy exchange.

The diagonal shows the distribution of each variable for years 2015 and 2016, separated between day and night. Soil temperature at 5 cm had a different distribution in the daytime and at nighttime : temperature at night (colored blue) peaked at 5 °C with a spread between 0 and 30 °C while the distribution of day temperature in red shows two peaks, one at 6 °C and the other at 21 °C. This bimodal distribution of day soil temperature reflects the continental climate of Switzerland, characterized by hot summers (with an average day temperature of 21 °C) and cold winters (average 6 °C). The effect of the continental climate is also visible in the distribution of air temperature, with the same bimodal distribution for daily temperature. The distribution

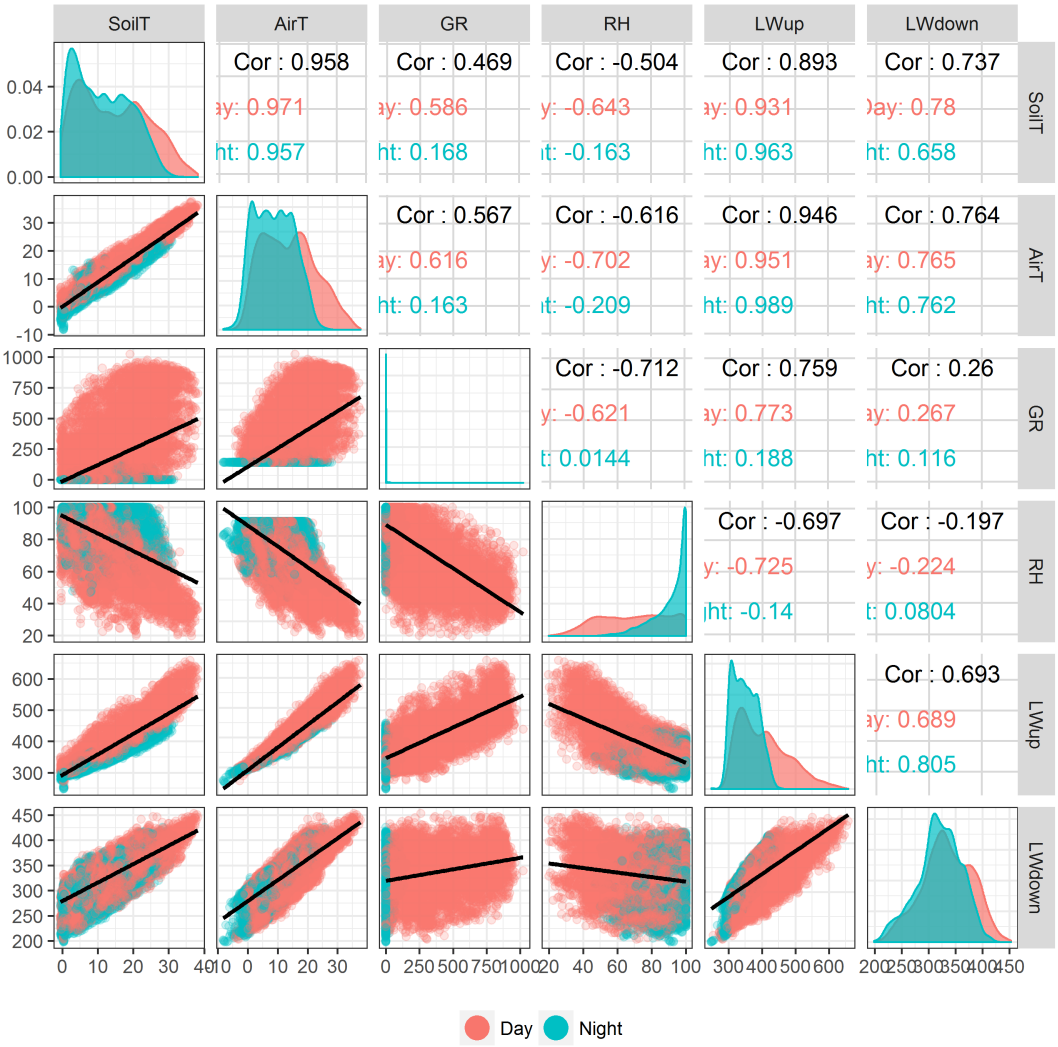

FIGURE 1.6: Matrix of bivariate scatterplot to show the relationships between variables aggregated at the hourly level for 2015 and 2016. The scatter plots show N=13127 data points, visually separated between day (red, between 09:00 and 17:00) and night (blue, between 21:00 and 05:00). On the diagonal there is the distribution of the variables, in the upper part the correlation in black over the entire dataset, and the correlation for each part of the diurnal cycle with the corresponding colour (only 'ay :' for day is readable and 'ht' for night). The lower part of the matrix shows the scatter plots with a linear regression line in black. SoilT: soil temperature at 5 cm [°C]; AirT: air temperature at 200 cm [°C]; GR: global radiation [$\mathrm{W\,m^{-2}}$]; RH: relative humidity [%]; LWup: upward long-wave radiation [$\mathrm{W\,m^{-2}}$], LWdown: downward long-wave radiation [$\mathrm{W\,m^{-2}}$]

for air temperature at night had several peaks between -3 and 10 °C. The fact that nighttime air temperature was differently distributed from nighttime soil temperature indicates that air and soil do not store heat in the same way.

The correlation between soil temperature at 5 cm and air temperature was close to 1, as shows figure 1.5, and was slightly lower at night than during the day.

The next variable is global radiation, the first source of energy (up to 1000 $\mathrm{W\,m^{-2}}$ can reach the soil surface) in the soil-atmosphere system. It had a correlation of 0.6 with soil temperature at 5 cm during the day and an almost null line at night since global radiation is nonexistent at night. Upward and downward long-wave radiation had a greater correlation with soil temperature than the global radiation (0.89 and 0.74 for long-wave, 0.47 for global radiation). Correlation with upward long-wave was high because the emission of radiation is directly proportional to the soil surface temperature. The high correlation of soil temperature at 5 cm and long-wave radiation shows the importance of long-wave radiation for the computation of radiative balance in the physical model developed in the next chapter.

Relative humidity in air showed a negative correlation with global radiation and upward long-wave radiation. High daytime humidity most likely indicates a cloudy day with less radiation coming to the surface as well as less energy emitted from the soil via long-wave radiation.

3 Chapter conclusion

We analysed the hourly data coming from the weather station installed on a bare soil in Stein, Switzerland. The analysis of two and a half years of data provided us with insight into soil temperature, our variable of interest.

Our preliminary study of local soil and air temperature data showed us that soil temperature at 5 cm under a bare surface is closely linked to air temperature at $2\,\mathrm{m}$ (R2=0.91) (chapter 1). Hence, whenever an initial dataset of measurements is available to allow the estimation of the regression coefficients, using this correlation is probably the most efficient way to predict soil temperature at 5 cm from measurements of air temperature. However, these coefficient values are specific to the year 2015, our location and our soil settings: they could not be used straight away in other situations, especially ones in which the soil has a different texture and structure, is covered by a

layer of residues or vegetation, or if there is a requirement to simulate deeper soil temperature.

For instance, R2 decreased with depth: the correlation was 0.78, down 15 $\%$, at $0.3\,\mathrm{m}$ and 0.52, down 43 $\%$ at $1.4\,\mathrm{m}$. This means that, below $0.3\,\mathrm{m}$, air temperature no longer is a good proxy for soil temperature.

On average, soil temperature at 5 cm below the soil surface was $1.6\,^{\circ}\mathrm{C}$ higher than air temperature at 2 m. This difference could go up to 10 $^{\circ}\mathrm{C}$ higher or lower, especially during cold winter nights when the soil was still warm or covered in snow and had not been warmed up by solar radiation. In summer, the ground surface was hotter than air, and the difference between atmospheric and soil temperature was highest around the surface. Other findings are presented in the following list:

- The soil started to get warmer than air in April
- The soil temperature at 1.4 m depth did not change during the day
- Air temperature at 2 m was not different from temperature at 20 cm because of the air-mixing action of eddies

These findings are limited to our location Stein in Switzerland and to bare soil, but similar data analysis of soil temperature for different regions shows little difference with these results (Campbell and Norman, 1998).
Our soil model should first and foremost be able to describe and predict the phenomena that happen at the surface, since this is where the energy exchange happens. This will be the subject of the next chapter.

Chapter 2

Development of the soil model

SOPHIA

This chapter aims at describing in detail the development of a model for predicting temperature and water movement in the soil. We named this model "SOPHIA", for soil physics in agronomy. This choice is meant to underline the model's vocation to solve problems in agronomy. The following steps describe the model's development and execution :

- Transformation of the physical phenomenon into equations
- Discretisation of the system and resolution
- Translation into computer code
- Simulation and analysis of the results

This chapter shall first show a Forrester diagram of the soil that highlights the link between the inputs (measured by a weather station) and state variables. Moving forward, it will describe the equations that drive the exchange of energy at the surface and the movement of heat and water in the soil.

In the second part, the model's resolution shall be explained. For this purpose, the equations will need to be rewritten with a clearer mathematical expression. This rewriting will empower the derivation of a numerical scheme, which will be implemented into a computer program.

Following the programming step, we shall compare the model's results to those described in Müller (1999), one of our models of reference. The numerical scheme will be tested in order to determine both the number of layers the soil should be divided into as well as the time step best suited for the simulation.

Finally, we shall close the chapter with a description of the method used to procure the soil parameters. These soil parameters are important when

testing the soil model against real observations. The soil parameters are obtained using the soil moisture retention curve and the Saxton and Rawls (2006) model.

A glossary, found in appendix A, clarifies some of the terms mentioned in the description of the model.

1 Conceptual diagram

1.1 Forrester diagram

Before starting to describe the model, it is important to represent it in a diagram that quickly identifies what it represents and what it does not. We have chosen a Forrester representation (Wallach et al., 2014) to describe the system, its components and the way that they connect to one another.

The system diagram (figure 2.1) shows our 1-D representation of the soil. The soil is represented by its surface and soil layers. The surface cannot be considered a layer since it only represents the interface between soil and air. It receives and emits the energy fluxes and holds a single state variable, the soil surface temperature. The value of this variable is obtained by resolving the energy balance equation. Contrary to the surface, the soil layers have a definite size, and their number depends on the choice of the user. For each soil layer, the temperature and water content are computed.

Between the soil surface and the first soil layer, heat is transferred via conductive heat flux. Note that it is called heat flux when the same quantity ($\mathrm{W\,m^{-2}}$) in the soil is called heat flow. The difference between flux and flow is an adopted convention: at the surface, there are energy fluxes, but in the soil, we refer to heat and water flow, not flux (Campbell and Norman, 1998).

1.2 What SOPHIA does not include

The diagram also reveals what our model does not include. Firstly, the model does not take into account the movement of vapour but solely liquid water movement. Secondly, the model does not take into account the impact of temperature onto water flow. The movement of vapour is a function of the temperature and water content in the soil. To infer vapour movement, it is required to couple together the movement of heat and water, as exemplified in the literature (De Vries, 1958). To begin with, we kept the flow of heat and

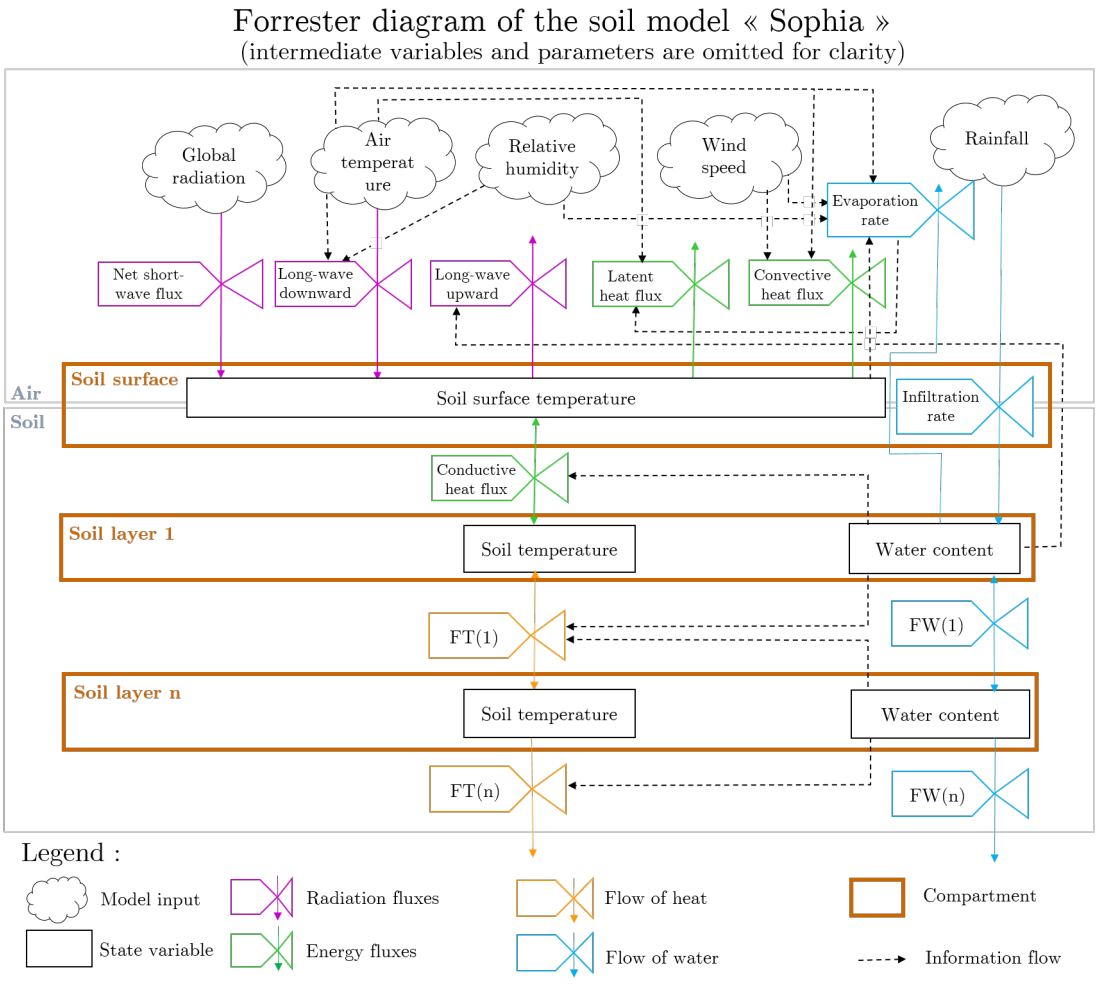

FIGURE 2.1: The Forrester or system diagram of the soil model "SOPHIA". The two large grey boxes are the air and soil components of the soil-atmosphere system depicted in the introduction in figure 5. The meteorological inputs (depicted in the cloud-like shape) are measured in the air component, while the soil component is divided into several soil layers. The soil surface is a specific layer since it constitutes the interface between the air and soil components. Unlike the other layers, it does not hold a measurable size. In the soil surface layer, energy under different forms (radiative, conductive, and convective) is received or emitted. For soil layers, there is only an exchange of conductive heat or water. This conceptual diagram omits the intermediate variables and parameters of the model because it aims at emphasising the relationships between inputs and layers. It shows that the model focuses on soil temperature rather than on water flow. The water flow is assumed to be isothermal (insensitive to temperature) and we do not take the flow of vapour into account.

water as independent as possible, in order to make the initial development easier.

Hence, for water movement, we assumed a constant temperature of 20 °C for hydraulic conductivity. This will not change depending on temperature. In reality, when the temperature rises, pressure in the liquid and gaseous phases increases as well, making water move from warm regions (with high pressure) to cold regions (with low pressure).

This phenomenon happens particularly in the winter in temperate climates, when the snow covers the soil (G. N. Flerchinger, 2000). In order for the model to run accurately in the winter, this would be a necessary addition

2 Biophysics equations

In our model development, we have described the system and showed a diagram of what the model takes into account. Moving forward, we need to write into equations the described physical phenomena. Each of the fluxes and flows mentioned in the diagram will be characterised in the following sections. In doing so, we dive into the subject area of environmental physics. All the equations follow the description of the soil model in Müller (1999). We shall first take the initial step of describing how we compute the flow of heat within the soil. Then we will go through the computation of the flow of water which introduces the concept of water potential. Finally, we will describe the more complex environmental physics equations that provide the different terms of the energy balance equation.

2.1 Heat movements

The equation for the flow of heat FT [$\mathrm{W\,m^{-2}}$] is given by Fourier's law. The flow is driven by the gradient of temperature (∂T, [°C]) in z [m], and its intensity is proportional to the thermal conductivity KT [$\mathrm{W\,m^{-1}\,K^{-1}}$] of the material. KT is dependent on the water content so we write $KT(\theta(t,z))$ to show it explicitely.

$$FT = KT(\theta(t,z)) \cdot \frac{\delta T}{\delta z} \tag{2.1}$$

The thermal conductivity KT [$\mathrm{W \cdot m^{-1} \cdot K^{-1}}$] depends on the texture, the organic matter content, soil water content, and thermal properties of the soil's constituents. The texture is the solid phase proportion attributed to clay, sand and silt. This proportion or fraction is denoted f_m, m for minerals (sand and

clays). The $\sum$ symbol in equation 2.2 signifies that we are adding up all the properties of these minerals. The organic matter fraction (denoted f_o [ratio]) is relative to the total weight of the soil and not merely the solid phase. The thermal properties are the thermal conductivity of water, minerals and organic matter (noted $k_{H_2O}, k_m, k_o, k_{air}$ $[\mathrm{W} \cdot \mathrm{m}^{-1} \cdot \mathrm{K}^{-1}]$). The equation for KT also uses the volumetric water content (θ $[\mathrm{m}^3_{\mathrm{H_2O}}\,\mathrm{m}^{-3}_{\mathrm{soil}}]$), the air porosity P_{air}, as well as factors w_o, w_m and w_{air} [-].

$$KT = \frac{\theta \cdot k_{H_2O} + \sum f_{m,o} \cdot k_{m,o} \cdot w_{m,o} + P_{air} \cdot k_{air} \cdot w_{air}}{\theta + \sum f_{m,o} \cdot w_{m,o} + P_{air} \cdot w_{air}} \tag{2.2}$$

The w factors are ratios that determine the thermal conductivities relatively to the water phase. In other words, the gradients of heat in organic matter, minerals and air are set relatively to the gradient of heat in water. P_{air} $[\mathrm{m}^3_{\mathrm{air}}\,\mathrm{m}^{-3}_{\mathrm{soil}}]$ is the volume of air per volume of soil ($P_{air} = P_{tot} - \theta$). Index m in w_m, k_m and f_m stands for minerals and refers to quartz (the main mineral in sand particles) and clay $f_m = f_{clay} + f_{sand}$. Total porosity P_{tot}, the volume of pores per volume of soil $[\mathrm{m}^3_{\mathrm{air+H_2O}}\,\mathrm{m}^{-3}_{\mathrm{soil}}]$ is necessary for the computation of air porosity P_{air} :

$$P_{tot} = 1 - \frac{\rho_b}{\rho_s} \tag{2.3}$$

P_{tot} needs the value of ρ_b $[\mathrm{g}_{\mathrm{solids}}\,\mathrm{m}^{-3}_{\mathrm{soil}}]$, the dry bulk density, and ρ_s $[\mathrm{g}_{\mathrm{solids}}\,\mathrm{m}^{-3}_{\mathrm{solids}}]$, the density of solids or mean particle density. Dry bulk density can either be measured or inferred with a statistical model (see section 4.3).

If we replace the expression of FT in the expression of the heat movement equation seen in the introduction (equation 2) we obtain the expression of the heat flux as a function of the state variable T, the temperature of the soil and θ the water content.

$$C_h \cdot \frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \cdot \left(KT(\theta(t,z)) \cdot \frac{\partial T}{\partial z} \right) \tag{2.4}$$

Finally, the last remaining term to explain is the soil volumetric heat capacity C_h $[\mathrm{J}\,\mathrm{m}^{-3}\,^{\circ}\mathrm{C}^{-1}]$. It is expressed as

$$C_h = \theta\ \rho_{H_2O}\ c_{H_2O} + \sum \rho_{m,o}\ c_{m,o}\ f_{m,o} + \rho_{air}\ c_{air}\ P_{air} \tag{2.5}$$

The specific heat is the amount of energy needed to warm up by 1 °C a volume unit of soil. It converts the temperature [°C] into a volumetric heat energy $[\mathrm{J}\,\mathrm{m}^{-3}]$. It depends on the water content θ $[\mathrm{m}^3_{\mathrm{H_2O}}\,\mathrm{m}^{-3}_{\mathrm{soil}}]$, the density ($\rho$) of water,

minerals (clay and quartz), organic matter and air ($\rho_{H_2O}, \rho_m, \rho_o, \rho_{air}$ $[\mathrm{g\,cm^{-3}}]$) and their specific heats ($c_{H_2O}, c_m, c_o, c_{air}[\mathrm{J\,g^{-1}\,{}^{\circ}C^{-1}}]$), multiplied by their volume fractions (f_m and f_o [ratio]).

2.2 Water movements

The heat equation is rather straightforwardly described. The movement of water is more complex than the movement of heat because water moves according to change in matric and gravitational potential, two of the four components of water potential. We calculate water potential as the sum of only matric and gravitational potentials, since its two other components, osmotic and pressure potentials, only play a negligible part at our scale of interest or for this part of the study. However, the presence of gravitational potential in the equation makes the derivation of the water equation more complex than for heat (Campbell and Norman, 1998).

In a porous medium like soil, water flow (FW $[\mathrm{m\,s^{-1}}]$) is driven by gradients of water potential (Ψ $[\mathrm{m}]$). Water potential represents the sum of matric potential and gravitational potential. Matric potential is a measure of the strength with which the water is linked to the soil matrix whereas gravitational potential is the energy stored in the water due to its position relatively to the soil surface. In our system, matric and gravitational potentials are negative. A wet soil will have a negative water potential value close to zero, whereas a dry soil's water potential is far below zero.

The flow of water is proportional to the hydraulic conductivity KW $[\mathrm{m \cdot s^{-1}}]$, which depends on the water content θ:

$$FW = KW(\theta(t,z)) \cdot \frac{\delta\Psi(\theta(t,z))}{\delta z} \tag{2.6}$$

This equation is more complex than the heat flux equation (equation 2.1) because the transformation of water content into water potential is based on empirical models that have a limited validity range.

We followed the method developed by Campbell (Campbell, 1974) to determine matric potential and hydraulic conductivity. Ψ_m is expressed using a power relationship with respect to water content θ as follows:

$$\Psi_m = e^a \cdot (\theta/\theta_{sat})^b \tag{2.7}$$

We derived $a = -2.3$ and $b = -5.2$ from water-retention curves measured at the site of the experiment (see sections 4.4), and $\theta_{sat} = P_{tot}$ is the soil

volumetric water content at saturation. The term e^a is interpreted as air-entry water potential (noted Ψ_e or air-entry suction). It is the threshold at which the large pores start to empty and it corresponds to a volumetric water content $\theta_e = 0.49$ [$\mathrm{m}^3_{\mathrm{H_2O}}\,\mathrm{m}^{-3}_{\mathrm{soil}}$].

Water potential (Ψ) is the sum of the matric and gravitational potentials. Gravitational potential is the force linked to the position of the water relatively to the soil surface. We represent this force by D [m], the depth at which the potential is calculated.

$$\Psi = \Psi_m - D \tag{2.8}$$

The osmotic potential, another form of potential linked to the presence of solutes in the soil, is not taken into account because its contribution to water movement is negligible. It becomes important when the transport of solutes within the soil is studied.

The hydraulic conductivity KW [$\mathrm{m\,s^{-1}}$] is expressed with another power relationship (Campbell, 1974)

$$KW = Kws \cdot (\theta/\theta_{sat})^{2 \cdot b_{Kw}+3} \tag{2.9}$$

where Kws [$\mathrm{m\,s^{-1}}$] is the saturated soil hydraulic conductivity and b_{Kw} is an empirical constant, the value of which, for different texture classes, can be found in table 2.2. When we combine equations 4, 2.6, 2.8 and 2.9, we obtain Richards equation that states that the change in water content θ is proportional to the hydraulic conductivity KW and its direction is set by the difference in water potential Ψ.

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z} \cdot \left(- KW(\theta(t,z)) \cdot \frac{\partial \Psi(\theta(t,z))}{\partial z} \right) \tag{2.10}$$

Before translating these equations into a computer program, we will describe the energy balance at the soil surface.

2.3 Energy balance equation at the soil surface

The energy balance equation can be be written

$$Rn + G + H + LE = 0 \tag{2.11}$$

with Rn [$\mathrm{Wm^{-2}}$] the net radiation, H [$\mathrm{Wm^{-2}}$] the convective heat exchange between surface and atmosphere, G [$\mathrm{Wm^{-2}}$] the conductive heat exchange from surface to deeper soil and finally LE [$\mathrm{Wm^{-2}}$] the latent heat exchange.

In the next paragraphs, we will present the biophysical equation of the four terms of the energy balance. This formalism follows the description of the energy balance model in Müller (1999), excepted the aerodynamic resistance one r_{ah}.

Net radiation Rn

The first term of the energy balance is the net radiation Rn. This form of energy is radiative, as it is emitted by the sun, the atmosphere or the soil itself. The sun, with a surface temperature of approximately 6000 °C, emits most of its radiation within the visible range, between $300\,\mathrm{nm}$ to $1000\,\mathrm{nm}$, and this is termed short-wave radiation. On the other hand, the clouds, the particles in the atmosphere and the soil exhibit a temperature between $0\,^{\circ}\mathrm{C}$ and $50\,^{\circ}\mathrm{C}$. For these temperatures, the radiation emitted is a long-wave radiation in the range of $1000\,\mathrm{nm}$ to $10\,000\,\mathrm{nm}$. The distinction is made between the two types of radiations because they are not computed in the same way.

The net radiation equation is defined in equation 2.12 as the sum of, respectively, the total energy transported by short waves (St, [$\mathrm{W\,m^{-2}}$]) attenuated by a reflection coefficient or albedo (α, [ratio]), the downward radiant flux density emitted by clouds and the atmosphere (Ld, [$\mathrm{W\,m^{-2}}$]) as well as the upward radiant flux density emitted by the soil (Lu, [$\mathrm{W\,m^{-2}}$]).

$$Rn(T,\theta) = St \cdot (1-\alpha) + Ld(T) - Lu(T,\theta) \tag{2.12}$$

The long-wave terms are computed with the Stefan-Boltzmann law in equation 2.13:

$$\begin{aligned} Lu(T,\theta) &= \epsilon_s(\theta)\sigma T^4 \\ Ld(T) &= \epsilon_a(T_{air})\sigma T^4 \end{aligned} \tag{2.13}$$

The emissivities ϵ_s and ϵ_a concern, respectively, the soil and the atmosphere. Soil emissivity depends on the volumetric water content near the soil surface θ [$\mathrm{m^3_{H_2O}\,m^{-3}_{soil}}$] (equation 2.14) and the atmospheric emissivity depends on the density (ρ) of vapour in the air, ρ_{va} (equation 2.16).

$$\epsilon_s(\theta) = S + \frac{\theta \cdot \theta_{min}}{P_{tot}} \tag{2.14}$$

where $S = 0.9$ is the soil emissivity factor and θ_{min} $[\mathrm{m}^3_{\mathrm{H_2O}}\ \mathrm{m}^{-3}_{\mathrm{soil}}]$ is the minimum volumetric water content according to the minimum water potential Ψ_{min}.

$$\theta_{min} = \frac{\Psi_{min} - d_1}{-exp(a)} \cdot P_{tot} \tag{2.15}$$

The distance d_1 is the distance between surface and the mid-layer point of layer 1 (see figure 2.2)

P_{tot} the total porosity of the soil, which is the space taken by pores in a volume of soil, is computed with previously mentioned equation 2.3.

The atmospheric emissivity ϵ_a is expressed as:

$$\epsilon_a(T_{air}) = B_a + B_b \cdot \sqrt{1.41 \cdot \rho_{va}(T_{air})} \tag{2.16}$$

The equation 2.16 is referred as Brunt's formula and uses per default $B_a = 0.605$ and $B_b = 0.039$ (Müller, 1999). The 1.41 coefficient converts the vapour density ρ_{va} from $\mathrm{g}_{\mathrm{H_2O}}\ \mathrm{m}^{-3}_{\mathrm{air}}$ into hPa.

The vapour density or absolute humidity requires the ambient vapour density, which is computed as in the equation for e_a (equation in table 1.2), the molecular weight of water (M_{H_2O} $[\mathrm{g\,mol^{-1}}]$), the gas constant (R $[\mathrm{J\,mol^{-1}\,K^{-1}}]$) as well as the air temperature expressed in $^\circ\mathrm{C}$:

$$\rho_{va}(T_{air}) = \frac{e_a \cdot M_{H_2O} \cdot 1000}{R \cdot (T_{air} + 273.16)} \tag{2.17}$$

Sensible heat flux H

The energy flux H is the flux of heat travelling from soil to atmosphere. Its mode of transfer is convective, meaning that heat is transported by a moving fluid. In our system, this moving fluid is the air above the soil surface. It is driven by the difference between air temperature (T_{air} [°C]) and the temperature of the soil surface T_0 [°C]. Above the soil surface, the turbulences in the air slow down the transfer of heat. These turbulences are causing a resistance to the heat flux that is called aerodynamic resistance to heat flux, noted r_{ah}, that will be explained later.

$$H(T) = ch_{air}\ \rho_{air} \cdot \frac{T_0 - T_{air}}{r_{ah}} \tag{2.18}$$

The proportionality factor is the volumetric specific heat of air $ch_{air} = c_{air}\ \rho_{air}$ $[\mathrm{J\,m^{-3}\,K^{-1}}]$ where c_{air} $[J.g^{-1}.K^{-1}]$ is the specific heat of air (representing the

amount of energy needed to increase 1 g of air by 1 °C) and ρ_{air} $[\mathrm{g\,m^{-3}}]$ is the air density. The aerodynamic resistance will be detailed in the next paragraph.

Aerodynamic resistance r_{ah}

On its way from soil to atmosphere, the movement of heat is slowed down by two obstacles: surface resistance and aerodynamic resistance. Surface resistance is caused by a thin layer of static air above the soil (called the surface boundary layer) and is not taken into account in the model. The aerodynamic resistance happens between the static layer of air and a reference height ($z_{ref}[\mathrm{m}]$), here of two meters. In this zone, wind mixes the air and causes turbulences that increase or decrease the transport of heat and evaporation. Strong wind will decrease the resistance and hence increase the movement of heat and vapour, while slow wind increases the resistance and reduces the transfer of heat and vapour.

The aerodynamic resistance is the term in our equation which differs most from Müller's. The original expression was simplified by assuming only neutral conditions in the atmosphere and thus avoiding the computation of the atmospheric stability factor. This stability factor is necessary to know if the atmosphere is in neutral, stable or unstable conditions. In the neutral state, we assume that the wind profile (i.e. the distribution of wind speed along height in the atmosphere) is logarithmic, meaning that the wind speed is low next to the soil surface and increases exponentially with height. In non-neutral conditions, the atmosphere is either in a state called stable or in a state called non-stable, and the shape of the wind profile changes from the logarithmic form. During non-stable conditions, it becomes linear while in stable conditions, it takes an exponential shape. The non-stable conditions are most often met in summer, when the soil is warm and the intense heat flux creates strong disturbances of air above the ground (Müller, 1999). Because our model is intended to be used in spring for the beginning of the maize cycle, non-stable conditions are unlikely to be met until July in the region, and we hence decided not to include a stability parameter into our aerodynamic resistance equation.

The simplified expression for neutral-only conditions was taken from Acs, Mihailovic, and Rajkovic (1991). Aerodynamic resistance r_{ah} is expressed in $\mathrm{s\,m^{-1}}$ and depends on wind speed u $[\mathrm{m\,s^{-1}}]$, friction velocity U^* $[\mathrm{m\,s^{-1}}]$ and Von Karmann's constant $k = 0.4$ [-]. In the equation, the parameter called roughness length z_0 is an imaginary height where the wind speed is null. The

common value for z_0 ranges between 0.02 and 0.060 m for a tilled bare soil surface (Campbell and Norman, 1998). We finally obtain:

$$r_{ah} = \frac{0.74 \ln(z_{ref}/z_0)}{kU^*} \tag{2.19}$$

with

$$U^* = \frac{uk}{\ln(z_{ref}/z_0)} \tag{2.20}$$

Ground heat flux G

Ground heat flux is a conductive heat transfer between the surface and the deeper layers of soil, following the well-known heat equation. The gradient of temperature just below the surface generates the movement of heat. We take:

$$G(T) = K_T \cdot \frac{T_0 - T_1}{d_1} \tag{2.21}$$

with d_1 the distance between the soil surface and the point under the surface, where the temperature is T_1.

Latent heat flux LE

In the latent heat flux, energy is transferred through evaporation of water. It is the multiplication of the evaporation rate $E(T, \theta)$ $[\mathrm{m\,s^{-1}}]$ by the latent heat of vaporization denoted by $L(T)$ $[\mathrm{J\,g^{-1}}]$:

$$L(T) = A - B \cdot T_0 \tag{2.22}$$

$$E(T, \theta) = \frac{\rho_{vs}(T, \theta) - \rho_{va}(T_{air})}{r_{ah} \cdot \rho_{H2O}} \tag{2.23}$$

In equation 2.22, L is an affine function of T_0 [°C], the soil surface temperature, with intercept A $[\mathrm{J\,g^{-1}}]$ corresponding to the latent heat of vaporization at 0°C and slope B $[\mathrm{J\,g^{-1}\,°C^{-1}}]$. The evaporation rate (equation 2.23) depends on the gradient of density of vapour between air ρ_{va} and soil surface ρ_{vs}, and is hampered by the same aerodynamic resistance r_{ah} as the heat flux. The density of vapour at the soil surface is expressed similarly to the density of vapour in the air (equation 2.17), but the ambient vapour pressure is replaced

by the matric potential near the soil surface $\Psi_m(\theta)$.

$$\rho_{vs}(T,\theta) = \rho_{vs,sat}(T) \cdot \exp\left(\frac{\Psi(\theta)\cdot M_{H_2O}}{R\cdot(T_0+273.16)}\cdot\frac{g}{1000}\right) \tag{2.24}$$

The newly introduced term $\Psi_m(\theta)$ is the matric potential at the upper limit near the soil surface, depending on the volumetric water content θ. M_{H_2O} and R were mentioned in equation 2.17 and g is the acceleration due to gravity. The saturation of vapour concentration is expressed as:

$$\rho_{vs,sat}(T) = \frac{e_s\cdot M_{H_2O}\cdot 1000}{R\cdot(T_0+273.16)} \tag{2.25}$$

where e_s is the saturation vapour pressure defined in table 1.2 and calculated with soil temperature.

Water infiltration

At the surface, the infiltration of water into the soil depends on both the water content of the first layer and on the rainfall. The infiltration rate of water into the soil is expressed in $\mathrm{m\,s^{-1}}$. We first need to express the rainfall in the same unit : a rainfall rate R, the amount of rainfall per second. Measurements of rainfall are available hourly in $\mathrm{mm\,hour^{-1}}$ from the tipping bucket rain gauge installed on the weather station. We convert this rainfall to a rainfall rate by dividing by the time step of the simulation.

$$R = \frac{\mathrm{rainfall[mm\,h^{-1}]}\cdot 3600\mathrm{[s]}}{dt\mathrm{[s]}} \tag{2.26}$$

Water cannot infiltrate the soil if the water content of the soil layer just below the surface is at saturation, that is to say, all the pores are filled with water, or $\theta = P_{tot}$ [$\mathrm{m^3_{H_2O}\,m^{-3}_{soil}}$]. When the layer is at saturation and the rainfall rate is superior to the flow of water leaving the first layer, happens the phenomenon of surface water runoff expressed as:

$$\mathrm{Runoff} = \begin{cases} R - FW_2, & \text{if } \theta \geq P_{tot} \text{ and } R > FW_2 \\ 0, & \text{otherwise} \end{cases} \tag{2.27}$$

In the absence of run-off, water can infiltrate into the soil. The infiltration rate will depend on the infiltration capacity of the first layer. The infiltration rate is equal to rainfall rate R [$\mathrm{m\,s^{-1}}$] when it is inferior to the infiltration

capacity IC. When the rainfall rate is superior to the capacity, then the infiltration rate is equal to IC

$$IR = \begin{cases} R, & \text{if } R < IC \\ IC, & \text{otherwise} \end{cases} \tag{2.28}$$

IC is the infiltration capacity, it is a threshold expressed in $\mathrm{m\,s^{-1}}$. The infiltration capacity is equal to the conductivity at saturation Kws (equation 2.9) when the water content is superior to water content at air-entry suction θ_e (equation 2.7). When the water content is inferior to the air-entry suction, the soil is in an unsaturated state, and the infiltration capacity is equal to the flow of water due to hydraulic potential Ψ.

$$IC = \begin{cases} Kws, & \text{if } \theta_1 > \theta_e \\ \dfrac{Kws + KW_1}{2} \cdot \dfrac{-\Psi}{d_1}, & \text{otherwise} \end{cases} \tag{2.29}$$

E_a is the actual evaporation, which depends on the saturation of the layer of soil below the surface. To compute the first flux of water into the soil, with use actual evaporation and not directly the evaporative term E of the energy balance. This E is a potential evaporation from which we deduce actual evaporation E_{act}:

$$E_{act} = \begin{cases} -FW_2, & \text{if } \Psi < \Psi_{min} \text{ and } E > -FW_2 \\ E, & \text{if } \Psi < \Psi_{min} \text{ and } E < -FW_2 \\ E, & \text{otherwise} \end{cases} \tag{2.30}$$

The flow of water from the surface to the first layer mid-point FW_1 [$\mathrm{m\,s^{-1}}$] will depend on this infiltration rate (equation 2.28) and on the evaporation rate E (equation 2.23), obtained from the resolution of the energy balance equation (equation 2.11). We can write :

$$FW_1 = IR - E_{act} \tag{2.31}$$

2.4 Bottom of the soil profile

Water flow

For the bottom of the soil profile, there are three different cases:

- the bottom of the soil profile is a water table, i.e. the soil below our soil profile is always at saturation
- the bottom of the soil profile is an impermeable rock layer, i.e. the water does not flow outside the soil profile.
- the bottom of the soil profile is also soil, and the water runs freely into this deeper soil, i.e. this corresponds to a situation of free drainage (case depicted in the representation of the soil system in the introduction in figure 5).

If we imagine a soil with k layers, then the expression of the water flow from the mid-layer point to the bottom of the soil profile, FW_{k+1}, reflects these three different cases. When there is a water table, FW_{k+1} is represented as a standard flow of water. When there is an impermeable rock layer the water cannot flow out the system, so FW_{k+1} is null. When the water can flow freely beyond the bottom and out of system, we set FW_{k+1} to the hydraulic conductivity of the last layer KW_k. k is the index of the last layer of the soil profile (see section 3.1)

$$FW_{k+1} = \begin{cases} \frac{KW_k + KW_{k+1}}{2} \cdot \frac{\Psi_k - \Psi_{k+1}}{d_{k+1}}, & \text{water table} \\ 0, & \text{rock layer} \\ KW_k & \text{free drainage} \end{cases} \tag{2.32}$$

where d_k is the distance between the last layer's mid-point (layer k) and the bottom of the soil profile. We set the hydraulic potential below the soil profile Psi_{k+1} to the gravitational potential equal to the total depth $-D$ (potential are always negative). The hydraulic conductivity KW_{k+1} is a function of the water content θ in $k+1$.

$$\theta_{k+1} = \begin{cases} P_{tot}, & \text{water table} \\ P_{tot}, & \text{rock layer} \\ \theta_k & \text{free drainage} \end{cases} \tag{2.33}$$

Temperature

For the bottom of the soil profile, we set a null flow of temperature FT_{k+1}.

$$FT_{k+1} = 0 \tag{2.34}$$

This null flow means that the bottom of the soil profile is absorbing all the heat from the layers above.

3 Numerical implementation

In the previous section, we have defined the soil system by its boundary, its components and its state variables. Based on biophysics laws and modelling hypotheses, we have derived the set of equations describing the system's behaviour. Our next step is model simulation: to generate a set of numerical results that reproduce the behaviour of our dynamic system (Wallach et al., 2014).

The numerical implementation of the model aims at applying the system of equations in a clear structure that allows computers to resolve the system and produce a numerical result. Firstly, we shall discretise the soil system, meaning that we shall divide the soil into layers of given sizes and time into definite time steps t.

Afterwards, we will be able to rewrite the PDEs (partial differential equations) that describe the evolution of our two state variables, namely the temperature and water content in a discretised notation. The discretised notation of the PDEs will constitute the numerical scheme, which details how the equations can be resolved with a computer. The final step before starting to test the model shall be the actual implementation in computer code.

We implemented the model in the R language (R Core Team, 2015) because it is a common language used in the applied mathematics community. All the procedure to resolve the energy balance, the methods for calibration and sensitivity analysis are already available, which makes the use of the attractive for a complete modelling study.

3.1 Discretisation of soil into layers

To resolve numerically the partial differential equations 2.4 and 2.10, we use the finite difference method: we discretise the space and time to replace the partial differential with differences applied to the space-time grid made of discrete points. This method is suited for our 1-D soil system and has the advantage of allowing fast computation. We discretise our soil of total depth D with layers indexed by j, $j = 1, \ldots, k$ (figure 2.2). Each layer has a specific size z and a mid-layer node i where the state variables, temperature (T), water content (θ) and their respective conductivities are calculated. Hence, the

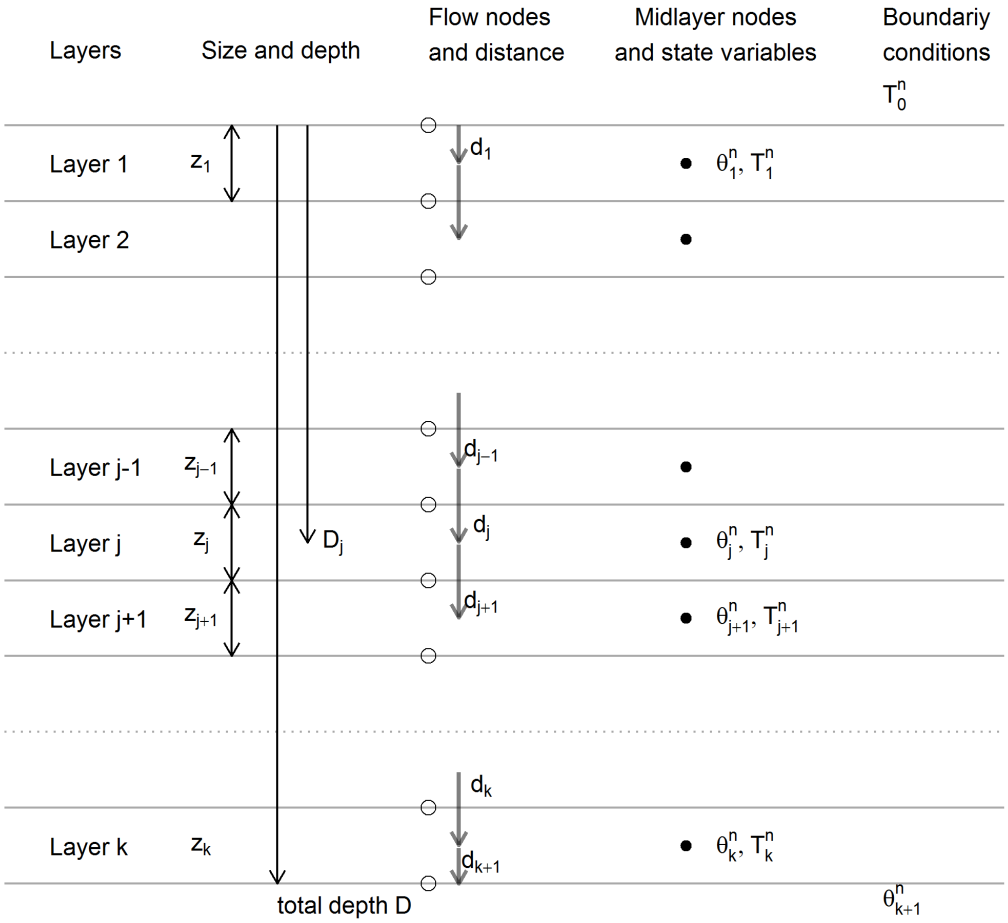

FIGURE 2.2: Discretisation of the 1-D soil profile. Each layer corresponds to a soil layer in the Forrester diagram (figure 2.1). The soil profile has a total depth denoted by D. All the variables have the n superscript meaning they are at time n. Each layer has a size z and a mid-layer node (•) where the state variables temperature T and water content θ are calculated. The flows of heat and water are calculated at the flow nodes (○). Flows are movement of water or heat along the distance $\mathrm{d_j}$. Note that there are k mid-layer nodes and k+1 flow nodes. The upper boundary condition $\mathrm{T_0}$ is the solution of the energy balance and the lower boundary condition θ_{k+1} depends on the case, see equation 2.33.

temperature of layer 1 corresponds to the temperature at the mid-layer node of layer 1. If we want to simulate the temperature at $5\,\mathrm{cm}$ without interpolation, we need to make sure that one mid-layer node falls near five cm. The state variables are updated with the computation of the flow of heat and water at the flow nodes. The flow nodes are at the interface of two adjacent layers.

The flow node represents flows over the distance d, which is the distance between two mid-layer nodes:

$$d_j = \frac{z_{j-1} + z_j}{2} \tag{2.35}$$

Layer 1 and layer k have particular distances:

$$\begin{aligned} d_1 &= \frac{z_1}{2} \\ d_{k+1} &= \frac{z_k}{2} \end{aligned} \tag{2.36}$$

The gravitational potential (equation 2.8) is equal to the depth D_j, which is the distance between soil surface to the mid-point of each layer :

$$D_j = \sum_{j=1}^{j} d_j \tag{2.37}$$

3.2 Formal expression of heat and water movements

In the search for a suitable numerical scheme, one must write the equation shown above in a clear mathematical form. In this notation, all parameter values are assumed to be positive. Everything that is constant is bulked into constant symbols and the dependencies between state variables temperature T and water content θ are made explicit. We start by denoting:
T_j^n the soil temperature of the mid-layer node of layer j at time step n
θ_j^n the volumetric water content of the mid-layer node of layer j and at time step n

We rewrite the system of equations for the movement of heat and water (equations 2.4 and 2.10) in a formal way :

$$\begin{cases} (l\theta + d\theta)\dfrac{\partial T}{\partial t} = \dfrac{\partial}{\partial z}\big(f(\theta)\dfrac{\partial T}{\partial z}\big) & \text{with } f(\theta) = \dfrac{a\theta + b\theta}{\theta + b\theta} \\ \dfrac{\partial \theta}{\partial t} = \dfrac{\partial}{\partial z}\big(g(\theta)\dfrac{\partial \Psi(\theta)}{\partial z}\big) & \text{with } g(\theta) = A\theta^{2p+3} \\ & \text{and } \Psi(\theta) = -B\theta^{-c} - D \end{cases} \tag{2.38}$$

In this system, it is visible that the movement of heat depends on water content, whereas the movement of water does not depend on temperature. The functions g and Ψ are the formal expression of the hydraulic conductivity (equation 2.9) and hydraulic potential (equation 2.8). These equations are nonlinear, especially in the case of the water movements with $g(\theta)$ and $\Psi(\theta)$ containing power relationships. The term D in $\Psi(\theta)$ makes its derivation difficult (Campbell and Norman, 1998).

3.3 Formal expression of boundary conditions

Upper boundary condition To resolve system 2.38, we also need an expression for the boundary conditions. Usually in physics, system boundaries are simple constant or sinusoidal function but in our case, we resolve the energy balance expression at the upper boundary.

In the formal expression of the energy balance, we denote the dependencies to the temperature at the surface T_0, water content just below surface θ_1 and time t. The dependency to time t shows when an observation of atmospheric conditions is needed (radiation, air temperature, air humidity, wind speed or rainfall)

For the energy balance expression, we denote by:

T_0 the soil surface temperature

θ_1 the volumetric water content of the first layer.

$$\begin{cases} Rn + H + G + LE = 0 \\ \text{with} \\ Rn(t, T_0, \theta_1) = \gamma i(t) + (\epsilon - \lambda + \nu\theta_1)\sigma T_0^4 \\ H(t, T_0) = \alpha T_0 + \beta(t) \\ G(T_0) = K\dfrac{\partial T}{\partial z} \\ L(T_0) = A_L - B_L T_0 \\ E(t, T_0, \theta_1) = \dfrac{1}{r(t)}\left(\dfrac{n}{mT_0 + l}\exp\left(\dfrac{o\Psi(\theta_1)}{uT_0 + v} - s\right)\right) \end{cases} \tag{2.39}$$

The energy balance gives implicitly the temperature at the surface T_0 and term E, which is needed for the computation of the boundary condition for the water equation:

$$\begin{cases} \dfrac{\partial \theta}{\partial z} = E(t, T, \theta) - I(\theta) \\ \text{with } I(\theta) = \begin{cases} R(t), & \text{if } R(t) < \text{treshold} \\ A, & \text{if } \theta > S \\ g(\theta)\Psi(\theta), & \text{otherwise} \end{cases} \end{cases} \tag{2.40}$$

Lower boundary condition We set the last flow of temperature FT_{k+1} as null:

$$FT_{k+1} = 0 \tag{2.41}$$

This null flow expresses that the bottom of the soil profile is absorbing all the heat coming from the rest of the soil profile. This is practical when we simulate soil profiles that are fairly shallow (< 2 m). If we simulate deeper soil profiles, another option is to set temperature of the last layer to a constant. As we are mostly interested the events near the surface, we often simulate shallow soil profile and hence prefer the null flow option.

For water, the several options presented in section 2.4 can be written:

$$\begin{cases} \dfrac{\partial \theta}{\partial z} = g(\theta)\Psi(\theta) \\ \dfrac{\partial \theta}{\partial z} = 0 \\ \dfrac{\partial \theta}{\partial z} = g(\theta) \end{cases} \tag{2.42}$$

Initial conditions

We assume that the temperature and water content at time t=0 are known. We can set the same value for the whole profile or derive temperature and water content profiles from the measurements. Here we show an example in which we set the whole profile with 20 °C and 0.3 $\mathrm{m}^3_{\mathrm{H_2O}}\,\mathrm{m}^{-3}_{\mathrm{soil}}$ at the beginning of the simulation:

$$\begin{cases} T_z^0 = 20, \quad \forall z = 1, \ldots, k \\ \theta_z^0 = 0.3, \quad \forall z = 1, \ldots, k \end{cases} \tag{2.43}$$

3.4 Resolution of the soil model with an explicit scheme

A challenging part of the model's development was to understand how to articulate the flow of heat and water (equation 2.38) with the energy fluxes at the surface (equation 2.39). Studies do not usually describe the numerical schemes used to resolve soil energy balance model and mention only a few key words to hide the technical difficulty.

Our first attempt at system resolution was to use a simple explicit scheme (forward Euler) to resolve the equations of the model. The model described in Banimahd and Zand-Parsa (2013) are also resolving the system of equations with an explicit scheme. We write the explicit scheme for the heat equation:

$$\gamma T_j^{n+1} = \gamma T_j^n + \left[f(\theta^n_{j,j+1}) \frac{T^n_{j+1} - T^n_j}{d_{j+1}} - f(\theta^n_{j-1,j}) \frac{T^n_j - T^n_{j-1}}{d_j} \right] \cdot \Delta t \tag{2.44}$$

where $\gamma = (l\theta^n_j + d\theta^n_j) \cdot z_j$, and $f()$ is the thermal conductivity associated with the flow of heat between layer $j-1$ and j. The conductivity is the average between the respective conductivity between layer $j-1$ and j

$$f(\theta^n_{j-1,j}) = \frac{z_{j-1} - z_j}{\left(\frac{z_{j-1}}{f(\theta^n_{j-1})} + \frac{z_j}{f(\theta^n_j)} \right)} \tag{2.45}$$

and the average conductivity between layer j and $j+1$

$$f(\theta^n_{j,j+1}) = \frac{z_{j+1} - z_j}{\left(\frac{z_{j+1}}{f(\theta^n_{j+1})} + \frac{z_j}{f(\theta^n_j)} \right)} \tag{2.46}$$

We write the explicit scheme for the water movement in the same form:

$$z_j \cdot \theta_j^{n+1} = z_j \cdot \theta_j^n + \left[g(\theta_{i,i+1}^n) \frac{\Psi(\theta_{j+1}^n) - \Psi(\theta_j^n)}{d_{j+1}} - g(\theta_{i-1,i}^n) \frac{\Psi(\theta_j^n) - \Psi(\theta_{j-1}^n)}{d_j} \right] \cdot \Delta t \tag{2.47}$$

The multiplication by z_j is important to keep the unit integrity of the equation. The term g is the expression of the average of conductivities of layer $j-1$ and layer j

$$g(\theta_{j-1,j}^n) = \frac{z_{j-1} - z_j}{\left(\frac{z_{j-1}}{g(\theta_{j-1}^n)} + \frac{z_j}{g(\theta_j^n)} \right)} \tag{2.48}$$

and the average of conductivities between layer j and layer $j+1$

$$g(\theta_{j,j+1}^n) = \frac{z_{j+1} - z_j}{\left(\frac{z_{j+1}}{g(\theta_{j+1}^n)} + \frac{z_j}{g(\theta_j^n)} \right)} \tag{2.49}$$

A graphical representation of this explicit scheme is depicted in figure 2.3. In the explicit scheme, the flows at time n are computed with the variables at the same time n. An implicit scheme would not use T^n for computing FT^n, but a mean temperature between T^n and T^{n+1}. When a pure explicit scheme is chosen, the heat transfer is higher than what actually occurs (Bittelli, Campbell, and Tomei, 2015). Hereafter, we describe the boundary conditions and explain why we developed an explicit numerical scheme.

Boundary conditions Suppose we know all the temperature and water content at time n for all layers, we may use this information to compute the energy balance at time n to obtain the temperature of the surface T_0. This step is required before computing θ^{n+1} and T^{n+1}.

$$T_0^n = Rn(t, T_0^n, \theta_1^n) - H(t, T_0^n) - L(T_0^n) \cdot E(t, T_0^n, \theta_1^n) - G(T_0^n) \tag{2.50}$$

With $G(T_0^n)$ is the upper boundary condition for the heat movement (FT_1) and $E(t, T_0^n, \theta_1^n) - I(\theta_1^n)$ is the upper boundary condition for water movement FW_1.

The numerical resolution of PDE is subject to questions of stability, consistency and convergence. Stability is the property that ensures that the numerical solution and the exact solution are within a certain range around the initial conditions. Consistency is the property that ensures that the numerical solution gets close to the exact solution when the discretisation steps

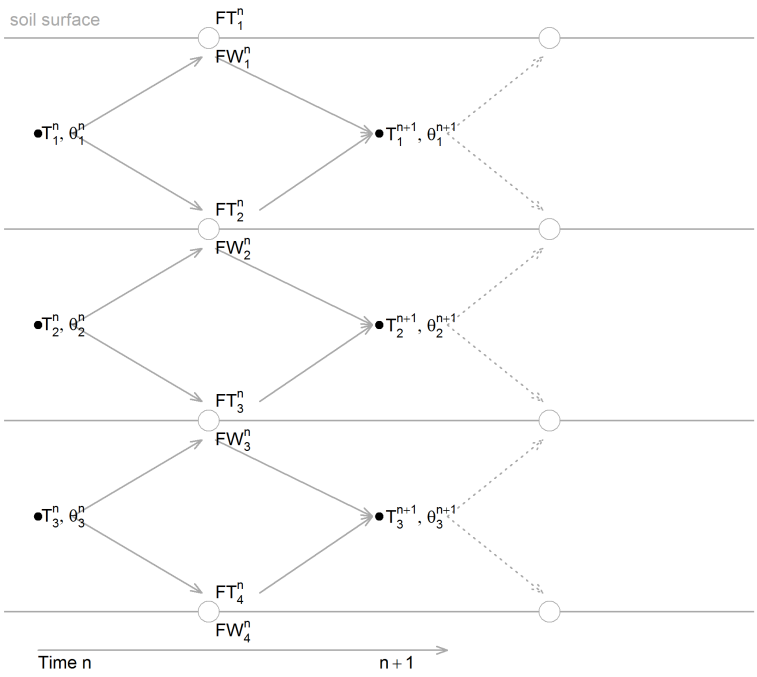

FIGURE 2.3: Flow and state variables at time n and $n+1$ in the explicit scheme to solve the heat and water equations of the SOPHIA model. For a simulation step, we know the soil water content θ and temperature T at time n and we use them to compute the flow of water FW and heat FT. These flows are then used to compute the state variables at the next time step $n+1$.

in time and space decrease to 0. Finally, convergence is a property ensuring that the numerical solution tends toward the exact solution of the continuous equation.

To converge, the time step of the model should always be proportional to the size of the layer. This condition is called a Courant-Friedrichs-Lewy (CFL) condition, and for the heat equation, it is:

$$\frac{f(\theta)}{l\theta + d\theta}\frac{\Delta t}{\Delta z^2} < 0.5 \tag{2.51}$$

As we can see, the CFL depends on $f(\theta)$ which is the thermal conductivity, which in turn depends on the water content, the texture of the soil and the porosity (equation 2.2). The CFL will hence change according to the type of soil and water content. Because the model is nonlinear, there are no general theoretical conditions that guarantee the convergence of the model. The closest general condition is the CFL, but this does not ensure that the system will converge, merely that it will be more likely to do so.

The explicit scheme has the advantage of being fast, simple to understand and easy to program, but it is ridden with stability issues (Haverkamp et al., 1977). To solve the problem, one approach consists in changing the time step dynamically during the simulation. The time step will be reduced to small steps when the water content is close to saturation and to larger steps when the water content is low. Such an adaptive time step slows down the simulation, which is why it is rarely used in advanced soil models (Haverkamp et al., 1977). In our current implementation of the model, the time step is kept constant. The time step is chosen according to the size of the smallest layer and the desired accuracy of the model.

In a first numerical approach for this study, we privileged the straightforward strategy consisting of designing an explicit scheme, with a constant and small enough time step to ensure the CFL conditions. As detailed later, no numerical instability was observed in our simulations. However, for future works it would be interesting to turn to an implicit scheme. An attempt at writing this implicit scheme is presented in appendix C

3.5 Implementation

Input data cleaning

When the model is deployed at a large scale, and especially for industrial application, the availability of the whole set of necessary input data is often

problematic. Indeed, the model requires that the weather input data has no missing values. Addressing this point is therefore a priority. A linear interpolation between available values is the proposed approach to fill the missing values.

Interpolation of input data for simulation

To run the model, we use an hourly input file containing global radiation (GR) [$\mathrm{W\,m^{-2}}$] measured at 1.5 m, air temperature T_{air} [$^{\circ}\mathrm{C}$], air humidity RH [$\%$], wind speed u [$\mathrm{km\,h^{-1}}$] and rainfalls R [mm] measured at 2 m. When the time step Δt is smaller than 1 hour we use a linear interpolation to obtain the weather for the desired time step. Rain is excluded from this process, since, instead of being interpolated, it needs to be distributed across the time step with the formula

$$\mathrm{Rainfall} = \frac{R}{i} \cdot \Delta t \tag{2.52}$$

where i is the duration interval in the input dataset and Δt is the time step set by the user.

Brief description of the SOPHIA package

In order to increase the portability of the simulator and as a way to build a rigorous programming structure, we decided to implement the model in an R package. The SOPHIA package contains all the functions needed to compute the energy balance at the soil surface and the diffusion of heat and water in the soil. It can run with hourly or sub-hourly meteorological data described above. The sequence to resolve the water content and temperature at each time step followed a given path. We started with solving the energy balance that gives the boundary conditions for the movement of heat and water. Only then were we able to compute the water content at the next time step for all layers. Finally, we updated to the next time step the temperature for all layers. After reaching this stage, we started a new cycle, by computing the energy balance with the newly found temperature and moisture of the soil profile, and continuing as described above (figure 2.3) The manual of the SOPHIA package can be found in appendix B

4 Simulation: preliminary tests

Now that the model and the numerical scheme have been described, we will present some simulation tests and model evaluations. We will first and foremost verify the robustness of the numerical scheme with different discretisations of space and time. Afterwards, we will validate our SOPHIA model SOPHIA against the training data set from Müller (1999). Before evaluating the model against experimental data, we will describe the calibration of the model for different types of soil.

4.1 Test of the explicit scheme

The soil model is composed of partial differential equations (PDE) that need to be resolved in order to figure out the changes in heat and water across time and space.

Three different discretisation grids of the soil system are considered here to test the stability of the explicit scheme: (i) a coarse one, that can be used to run fast simulations, and is composed of layers of 10 cm; (ii) a finer one with layers of 2 cm; (iii) a last one made up of a discretisation commonly used in agronomy, where the soil profile is separated into a shallow, top soil layer and a deeper sub soil layer.

Because of the explicit scheme, the value of the time step depends on the size of the layers. The simulation could not converge with a time step of 1 hour and a fine discretisation of 2 cm layers. The figure 2.5 shows situations in which the model converges. For these situations, the different space or soil discretisation do not have a big impact on the simulation of temperature or water content. The third soil configuration shows a slight shift in the temperature at 5 cm and 140 cm. The large layers of this discretisation make sense for agronomy but are likely to provide inaccurate soil temperature simulation.

For the soil water content, like for temperature, the agronomic layers discretisation also shows a bias compared to the other simulations, with an overestimation of the water content. In figures 2.6 and 2.5, the simulation with 10 cm layers and a time step of 1 hour is as accurate as the finer discretisation of 2 cm layers and 200 seconds.

With R, a simulation of SOPHIA takes under 1 minute for a typical simulation of 2 months with a time step of 1 hour and a soil discretisation of 10 layers of 10 cm.

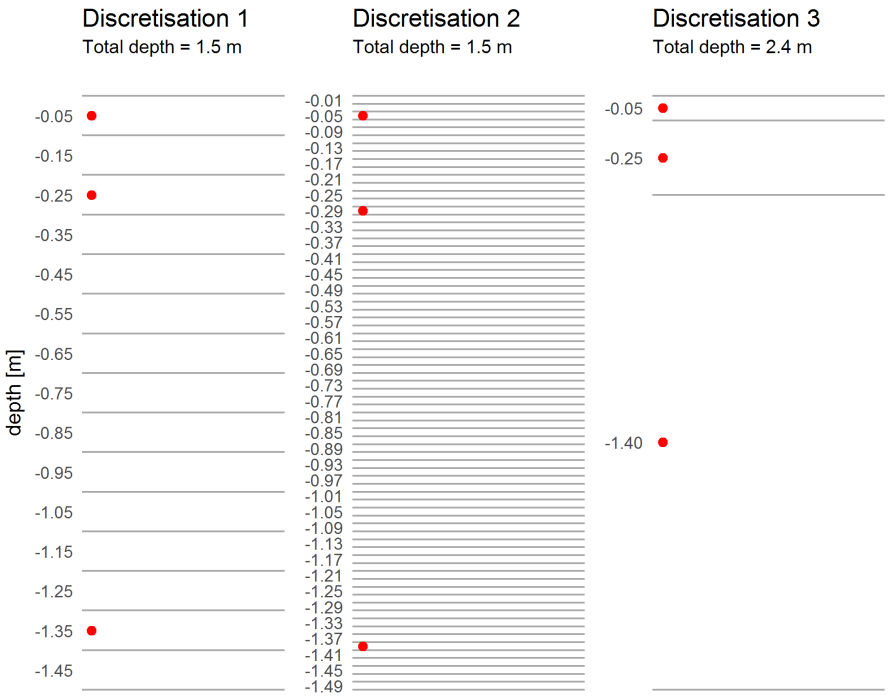

FIGURE 2.4: Different discretisation to test the sensitivity of the numerical scheme. The layers in discretisation 1 and 2 have equal size while in discretisation three the layers are increasingly larger. This discretisation is common for agronomists because they separate the soil into top soil and subsoil. The discretisation was created so as to have mid-layer points (•) close to 5, 30 and 140 cm. The total depth is 1.5 m for the profiles except for discretisation 3, where the size of the third layer had to be increased in order to have a mid-layer point at 1.4 m below ground.

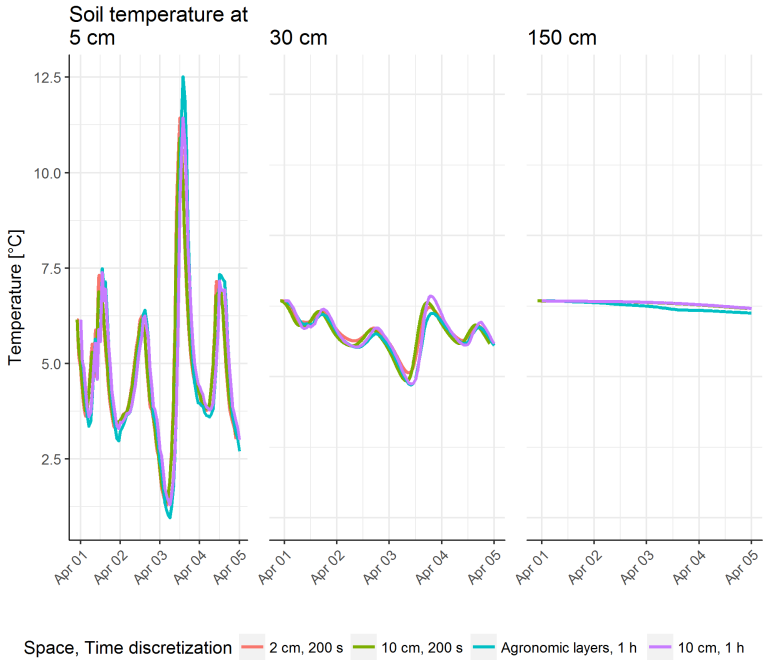

FIGURE 2.5: Comparison of different soil discretisation 2.4. The model was run with free drainage conditions at the bottom, and with two different time steps: 1 hour and 200 seconds, and the three types of discretisation. With a time step of 1 h, the simulation does not converge with the 2 cm discretisation and is not represented

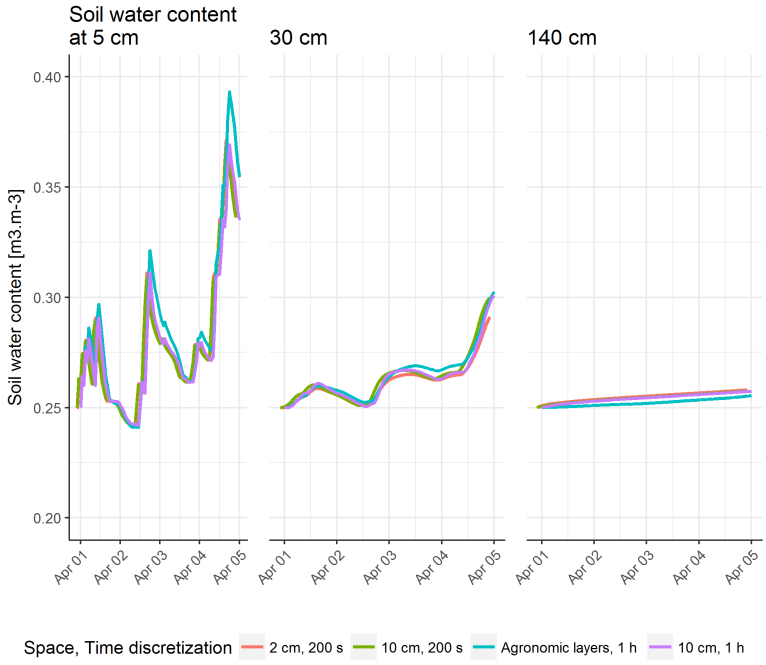

FIGURE 2.6: Comparison of different soil discretisation 2.4 on the simulation of water content. The model was run with free drainage conditions at the bottom, and with two different time steps: 1 hour and 200 seconds, and the three types of discretisation. With a time step of 1 h, the simulation does not converge with the 2 cm discretisation and is not represented

We conclude that, for the time being, this coarse discretisation leads to satisfactory results for the simulation because it is fast, stable and show good agreement with soil temperature measurement at 5 cm.

4.2 Verification with Müller's model

During the development step, it was necessary to regularly check the simulation against the results from Müller's model. One of the big advantages of the model described in Müller (1999) is that it is accompanied by the model written in the software package Model Maker. Model Maker is a practical tool to visualise the model as a box and arrows diagram and to run a simulation with the same software. In Müller (1999), each chapter is accompanied by Model Maker's '.mod' file, that lets readers view and test the model themselves. The energy balance model coupled with water and heat transport in the soil is described in chapter 5. The version of the model that we are using is the file named `Mod5-1b.mod`, that we have simplified to reflect the actual development of the SOPHIA model and to make the verification valid and accurate.

Müller's model comes accompanied with a training dataset, that is to say, a set of real observed data where the model is known to run properly. We used this dataset to compare both models. The training dataset is composed of weather data recorded in May 1997 in Linden, Germany. It contains the five inputs needed to run the model: global radiation GR, air temperature T_{air}, dew-point temperature T_{DP}, wind speed u and precipitation P at 30-minutes intervals. Müller's model needs dew-point temperature rather than relative humidity (table 2.1). The dew point temperature is more precise than relative humidity to compute air absolute humidity (Campbell and Norman, 1998) but is less routinely available than relative humidity RH.

SOPHIA model takes relative humidity as input. We hence modified the file `Mod5-1b.mod` from Müller (1999) so that it takes relative humidity as input. We wrote the equation for absolute humidity (or vapour density) $\rho_{va}(T_{air})$, equation 2.17 in Müller's code. The entry point of this equation, the relative humidity, was obtained by converting the dew-point temperature (T_{DP}) of the input file into relative humidity RH as follows.

$$RH = \frac{100 \cdot \exp\left(\frac{17.27 \cdot T_{DP}}{240.97 + T_{DP}}\right)}{\exp\left(\frac{17.27 \cdot T_{air}}{240.97 + T_{air}}\right)} \tag{2.53}$$

Another modification to the `Mod5-1b.mod` file concerned the aerodynamic resistance r_{ah} (see section 2.3 for the description of SOPHIA's aerodynamic resistance)

Finally, to run Müller's model for verification, we chose the Euler integration method for its likeness to our explicit scheme. For both model, the time discretisation was the same, with a time step of 120 seconds and 14 soil layers increasing progressively in size from surface (layer 1 is 0.01 m) to bottom (layer 14 is 0.1 m, details in table 2.1). The soil is considered homogeneous, and the parameters values were set the same in both models.

The first verification concerns the temperature profiles (figure 2.7). The two models give the same results with a little difference for deep soil temperature. The temperatures of SOPHIA at 0.75 cm are slightly higher than for Müller's model. The temperature at 1 m end above 15 $^{\circ}\mathrm{C}$ for SOPHIA while it is below for Müller's model. This difference in deep soil temperature is linked to the differences in water content simulation (figure 2.8).

The simulation of water content shows some differences for the deepest layers of soil and the increase of the water content after the rain even on day 4. In both cases, the effect of rain on the fourth and fifth days are visible: the water content first rises, then the soil progressively dries, but it seems that the model SOPHIA is less sensitive to rain, as the first peak of water content does not get above 0.4 $\mathrm{m}^3\,\mathrm{m}^{-3}$. For both models, the bottom conditions were set to "free drainage" meaning the water flows freely out of the system, but we see that SOPHIA's simulation is losing its water much faster than Müller's model. This difference might be explained by the difference in numerical implementation (explicit vs implicit scheme), or an error in the mass balance of water, where SOPHIA's simulation looses more water than there is in the system. The lower water content for the deep soil layers allows the soil to be warmer and this can explain the small difference in deep soil temperatures.

In figure 2.9, we show the comparison of soil surface temperature and energy balance terms simulations with SOPHIA and Müller's model. The comparison is shown only for three days, the second, third and fourth days of the training data set. Both models have the same surface temperature, with some very small differences linked to the difference in numerical implementation. The net radiation terms Rn, Ld and Lu are the same for both models. There are small differences between each term of the energy balance from the two models. During the day, the term LE in Müller's model is slightly different from SOPHIA's. The largest difference concerns the term G, the ground heat flux, which has a higher amplitude in SOPHIA's simulation. This difference

TABLE 2.1: Difference between the SOPHIA model, Müller's model modified for verification (figure 2.7 to 2.8) and the original Müller's model, `Mod5-1b.mod`

	Müller	SOPHIA	Müller modified
Inputs	GR, T, T_{DP}, u, P	GR, T, RH, u, P	GR, T, RH, u, P
Aerodynamic resistance (r_{ah})	$\frac{\ln(z_{ref} - d + z_h/z_h)}{k\ c_{air}\ \rho_{air}\ U^*} + s_h$	$\frac{0.74 \ln(Z_{ref}/z_0)}{kU^*}$	$\frac{0.74 \ln(z_{ref}/z_0)}{kU^*}$
Friction velocity (U^*)	$\frac{uk}{\ln(z_{ref} - d + z_m/z_m)} + s_m$	$\frac{uk}{\ln(z_{ref}/z_0)}$	$\frac{uk}{\ln(z_{ref}/z_0)}$
Stability parameter	$\frac{-k\ z_{ref}\ g\ H}{c_{air}\ \rho_{air}\ T_{air}\ U^{*3}}$	no	no
Integration method	Runge-Kutta	Euler (explicit)	Euler (implicit)
Time step [s]	600	120	120
Space discretisation [m]	14 layers increasing size [1]	same	same

Symbols : z_{ref} reference height; d height of zero plane displacement; z_h roughness length for heat; s_h stability correction for heat; z_m roughness length for momentum; s_m stability correction for heat; z_0 roughness length for bare soil in SOPHIA model, GR global radiation; T air temperature; RH relative humidity; T_{DP} dew-point temperature; u wind speed; P precipitation, for other symbols please refer to the symbol list on page xxv

[1] size of layers : 0.01,0.02,0.03,0.04,0.05,0.05,0.1,0.1,0.1,0.1,0.1,0.1,0.1 and 0.1 m

might be linked to the way the energy balance is resolved. In Müller, the term G is computed as the result of the other energy balance terms

$$G = Rn - H - LE \tag{2.54}$$

and then the temperature of the surface is computed with this G:

$$T_0 = T_1 + \frac{G \cdot d_1}{KT_1} \tag{2.55}$$

with T_0 the soil surface temperature, T_1 the temperature at depth d_1 and KT_1 the thermal conductivity of the first layer.

In SOPHIA, the term G is computed explicitly as

$$G = KT_1 \cdot \frac{T_0 - T_1}{d_1} \tag{2.56}$$

and then the temperature of the surface T_0 is found by solving

$$Rn - G - LE - H = 0 \tag{2.57}$$

With our way of solving the energy balance equation, G is allowed to adopt a larger amplitude of values than when it is the result of the difference between the other terms. This might explain the difference of amplitude between G in Müller's model and G in SOPHIA.

Conclusion The verification with the Müller model shows that the development of SOPHIA gives equal results on the training data set included in Müller (1999). There are small differences in the soil surface temperature and water content, especially for deep soil. The other noticeable difference is for the term G. These differences are small enough to consider the models as equivalent for the surface temperature and the net radiation terms.

4.3 Simulation with different texture classes

Before running the soil model for a given location, the model needs soil specific parameters in order to represent accurately its properties. Soil properties vary depending on the texture, structure, organic matter content, level of compaction (Richard et al., 2001) or rock fragments (Cousin, Nicoullaud, and Coutadeur, 2003). Some of the soil parameters, such as the texture fraction or organic matter content, are easily measured, but some others are harder

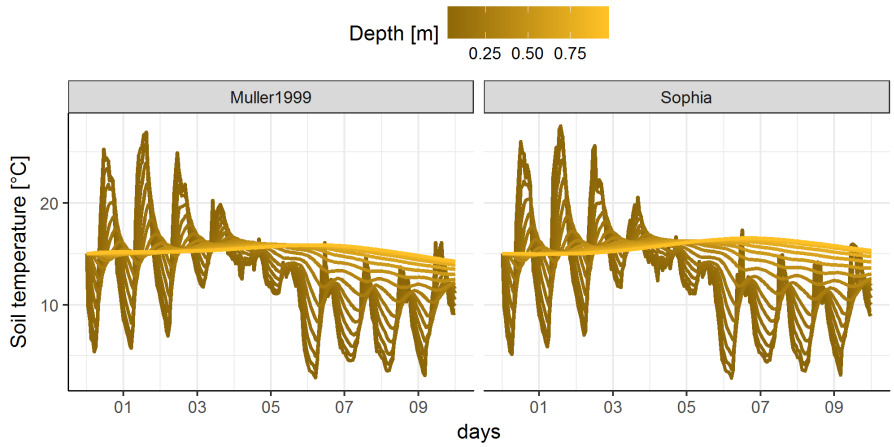

FIGURE 2.7: SOPHIA and Müller's model simulation of soil temperature on the training data set in Müller (1999), which contains weather data for May 1997 in Linden, Germany. The curves correspond to the 14 layers of soil representing a $1\,\mathrm{m}$ profile. We simulated ten days, with a Euler integration method and time steps of 120 seconds. Müller's model (left panel) was modified to take relative humidity as input, and the expression of aerodynamic resistance was simplified to match SOPHIA's (see 2.1). The SOPHIA model was run with the same soil discretisation (14 layers) and parameters.

to obtain because they need special devices or fresh samples, such as water retention properties, hydraulic conductivity at saturation or bulk density.

To correctly parametrise the soil, height parameters are needed : the fraction of sand, clay and organic matter (f_{clay}, f_{sand}, f_o), the parameters of the water retention curve a and b, the parameter for water content-hydraulic conductivity relationship (b_{Kw} and Kws, the hydraulic conductivity at saturation) and the bulk density ρ_b. These eight parameters are used in equations 2.2, 2.7, and 2.9.

From the fraction of sand, clay and organic matters, it is possible to infer water retention parameters, hydraulic conductivity parameters and bulk density with pedo-transfer functions (PTFs). PTFs are function used to compute the parameters of the curve linking water content to matric potential. Many PTFs are described in the literature (for a review, see Patil and Singh 2016) because i) there exists a variety of soil database from which pedo-transfer functions are created and, ii) no generic pedotransfer function has yet been found (Patil and Singh, 2016).

Among the PTF, we have chosen the Saxton model (Saxton and Rawls, 2006) because it uses the largest database to infer the pedo-transfer functions.

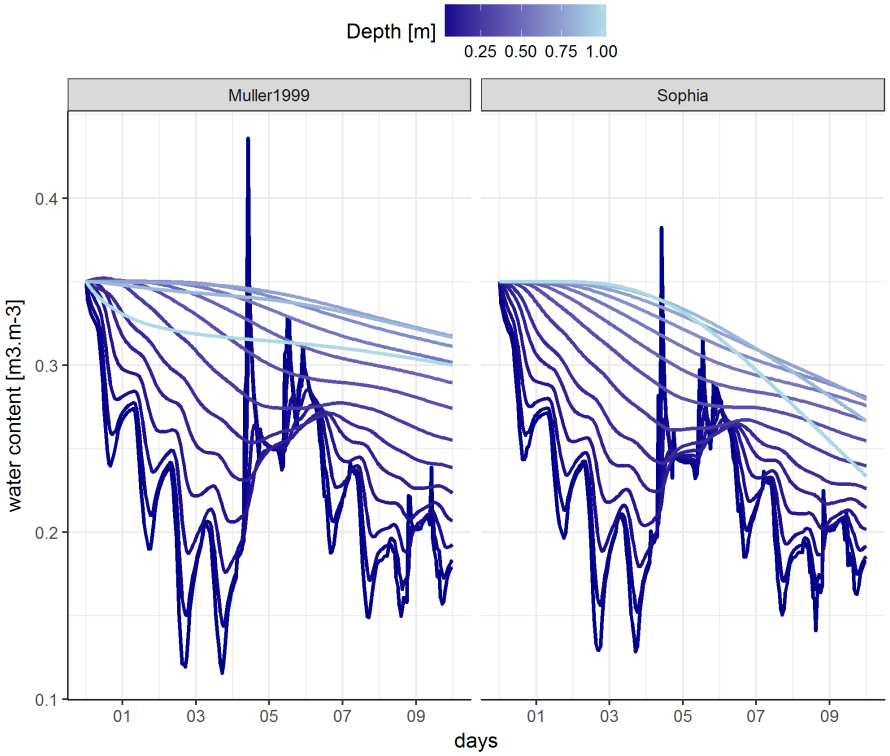

FIGURE 2.8: SOPHIA and Müller's model simulation of soil water content on the training data set in Müller (1999). The curves correspond to the 14 layers of soil representing a 1 m profile. We simulated ten days, with a Euler integration method and time steps of 120 seconds. Müller's model (left panel) was modified to take relative humidity as input, and the expression of aerodynamic resistance was simplified to match SOPHIA's (see table 2.1). The SOPHIA model was run with the same soil discretisation (14 layers) and parameters. Both models were run with "free drainage" as the bottom conditions. In the training dataset, there were rain events superior to 1 mm on day 4, 5 and

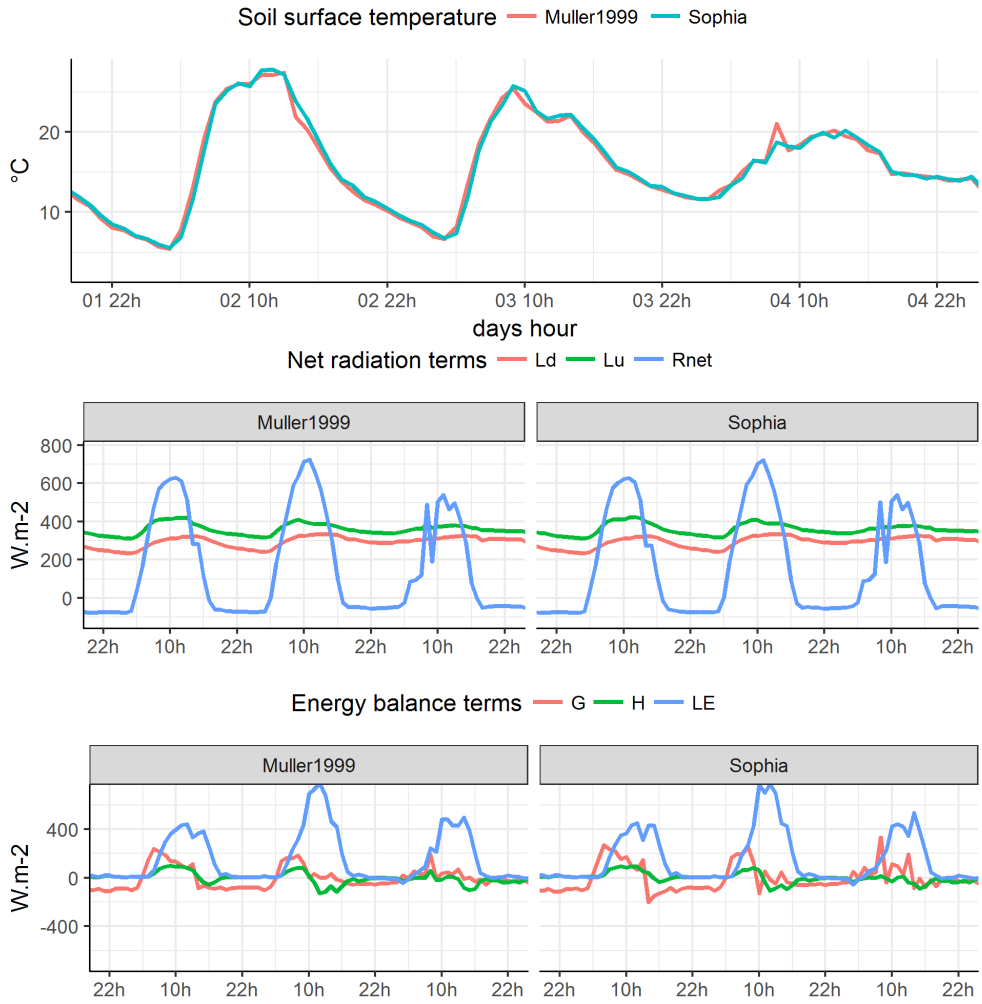

FIGURE 2.9: Simulation of the energy balance terms for 3 days, the 2nd, 3rd and 4th day in Linden, Germany by a simplified version of Müller's model (noted Muller1999) and SOPHIA (see the caption of figure 2.7 for details about the simulation). TS: surface temperature; G: ground heat flux; H: convective heat flux; LE: latent heat; Ld: downward long-wave; Lu: upward long-wave; Rnet: net radiation

The Saxton model uses empirical relationships between soil texture and water retention parameters drawn from a dataset of about 2000 soil samples of A horizons (20 first centimetres of soil) coming from the USDA soil sample database (USDA Natural Resource Conservation Service, 1994). Additionally, it is the reference model used in our research department.

The Saxton model takes as input the fraction of sand, clay and organic matter f_{clay}, f_{sand}, f_o and infer parameters a, b, Kws and ρ_b.

The soil parameters for twelve USDA texture classes (Soil Survey Division Staff, 1993) and two samples of the soil in Stein inferred by the Saxton model are presented in table 2.2. The sample of soil in Stein at 5 cm is close to a Clay Loam and the sample at 140 cm is close to a Loam texture class.

We ran the SOPHIA model with each texture class and show their effect on the temperature and moisture near the surface for a six of them in figure 2.10 (six texture classes were enough to represent the variability of all the texture classes). The change in texture class has a greater impact on the simulation of soil water content than on temperature. The simulation was made using real input data between 1st of April and 10 April 2015, when a rainy period (until 5 April) was followed by sunny weather.

The simulation of water content was the highest for the clay texture class and lowest for the sand, while the loam and silt texture classes were in between these two textures. The model is able to reproduce the typical behaviour of clay and sandy soils, with clay presenting the largest water holding capacity and sand the smallest.

The soil texture classes do not have a big impact on the beginning of the simulation of the soil temperature. At the start of the simulation, the water content is still high, and there is no clear difference between soil texture classes. When the water content gets very low with the sand and sandy loam texture classes, the simulation of temperature starts to differ from the rest of the texture classes as the soil gets warmer. As the soil gets dryer, its thermal conductivity increases, and so it absorbs more energy than the soil with larger water content.

Estimation of the soil parameter is a major step before running the model on the site of experimentation. Out of the eight parameters necessary to describe the soil, the three fractions (sand, clay and organic matter) are easily obtained from lab analysis or soil maps. The other four parameters can be inferred with the Saxton model (bulk density, a, b and Kws) and the last one (b_{Kw}) is found in Campbell and Norman (1998).

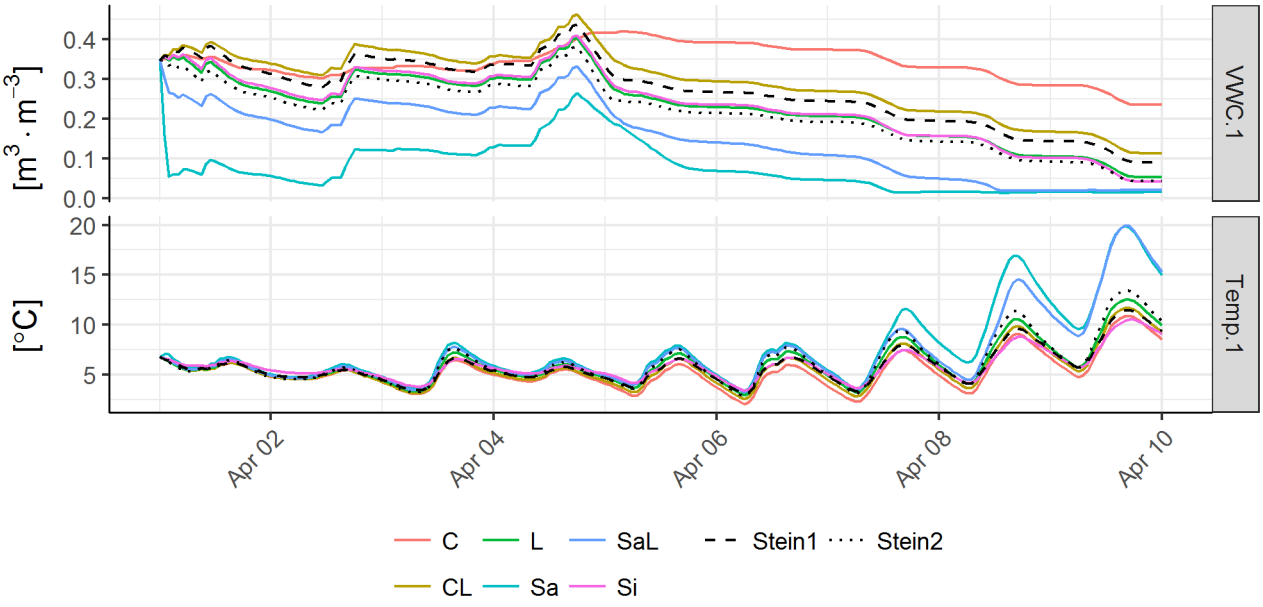

FIGURE 2.10: Simulation of the SOPHIA model between April 1st and April 10th, 2015 with 15 layers of 10 cm and a time step of 900 seconds, with free drainage at the bottom. The model was run with the soil parameters of table 2.2 for six texture classes and the two soil samples from Stein. The figure shows the water content and soil temperature at 5 cm below the soil surface (VWC.1 and Temp.1, the 1 denote the first layer of the soil discretisation). The simulation was run with observed weather data. C stands for Clay; L, Loam; SaL, Sandy loam; CL, Clay Loam; Sa, Sand and Si, Silt.

4.4 Estimation of water retention parameters

To evaluate the Saxton model for our soil, soil samples from the soil below the weather station were sent to a specialised soil laboratory (SADEF, France) to measure water retention curves. Water retention parameters a and b can be inferred from water-retention curves. The laboratory measured the water retention curves or pF curves soil samples made at three different depths (5, 30 and 140 cm). The pF curves (figure 2.11) represent the relationship between the logarithm of suction (pF, expressed in cm of water, noted cmH2O, equivalent to matric potential) and soil water content (Acs, Mihailovic, and Rajkovic, 1991). The pF notion is a logarithmic scale analogous to the pH (De Parcevaux and Huber, 2007). Suction is a positive value whereas matric potential is negative. In figure 2.11, the curves for soil sampled at 5 and 30 cm are similar, meaning that the soil at this depth holds the water in a similar way. At 140 cm, the soil texture is sandier, and by the same water content, there is a lower pF, i.e. less suction and thus a lower ability to retain water.

From figure 2.11, we can derive the parameters a and b for equation 2.7. A common method is the one of Campbell (Campbell, 1974) which defines the relationship between matric potential Ψ and water content θ with a power law, which in its simplest form is written:

$$\Psi = A(\theta)^b \tag{2.58}$$

where $A = exp(a)$ and a is the intercept and b is the slope of the log-log curve. The log-log curve is the plot of the log of matric potential (in $\mathrm{m_{H_2O}}$) versus the log of relative water content (relative because we use θ/θ_s where θ_s is water content at saturation or total porosity, here fixed at 0.5). The intercept and slope are determined by fitting a linear regression model. Because the curve is decreasing, a and b are negative. Making the log-log curve and extracting the slope and intercept yields the following parameters $a = -2.3$ and $b = -5.2$ for the sample at 5 cm and $a = -2.6$ and $b = -4.6$ for the sample at 140 cm. The result from this fitting are presented for the sample at 5 cm in figure 2.12, along with an estimation of these parameters with a specific soil properties model, the Saxton model (Saxton and Rawls, 2006) described in the next paragraph.

The obtention of water-retention curves is a long and expensive process, requiring specific soil pressure chambers. As a consequence, these curves are not often available at the location of interest. On the other hand, texture and organic matter content are often known by the agronomist or farmer because they can easily be obtained from labs or soil maps. There are models that don't

TABLE 2.2: The twelve USDA soil texture classes and soil sample from Stein with their properties inferred by the Saxton model. The texture classes are ordered in decreasing order of clay content. The data is without salinity, gravel or density adjustments. The percentage of OM is relative to the total soil volume, while for clay and sand it is related to the solid fraction. b_{Kw} are from Campbell and Norman (1998)

Texture Class	Code	Clay	Sand	OM	ρ_b g cm^{-3}	a	b	Kws m s^{-1}	b_{Kw}
Clay	C	0.50	0.25	2.5	1.33	-2.70	-11.06	3.1e-07	7.6
Silty Clay	SiC	0.45	0.10	2.5	1.26	-1.96	-9.00	1.03e-06	7.9
Sandy Clay	SaC	0.40	0.50	2.5	1.47	-3.00	-10.23	3.9e-07	6.0
Silty Clay Loam	SiCL	0.35	0.10	2.5	1.30	-1.27	-6.66	1.58e-06	6.6
Clay Loam	CL	0.35	0.30	2.5	1.39	-1.92	-7.70	1.19e-06	5.2
Stein (5 cm)	-	0.27	0.37	3.0	1.39	-3.90	-6.50	2.67e-06	5.2
Sandy Clay Loam	SaCL	0.25	0.60	2.5	1.50	-3.00	-8.00	3.14e-06	4.0
Loam	L	0.20	0.40	2.5	1.43	-1.44	-5.35	4.31e-06	4.5
Stein (140 cm)	-	0.19	0.50	2.2	1.48	-1.92	-5.73	4.92e-06	4.5
Silty Loam	SiL	0.15	0.20	2.5	1.38	-0.41	-3.74	4.47e-06	4.7
Sandy Loam	SaL	0.10	0.65	2.5	1.46	-2.05	-4.79	1.397e-05	3.1
Silt	Si	0.05	0.10	2.5	1.38	0.34	-2.29	6.11e-06	4.7
Loamy Sand	LSa	0.05	0.80	2.5	1.43	-2.52	-4.45	2.686e-05	2.1
Sand	Sa	0.05	0.88	2.5	1.43	-3.70	-5.33	3.003e-05	1.7

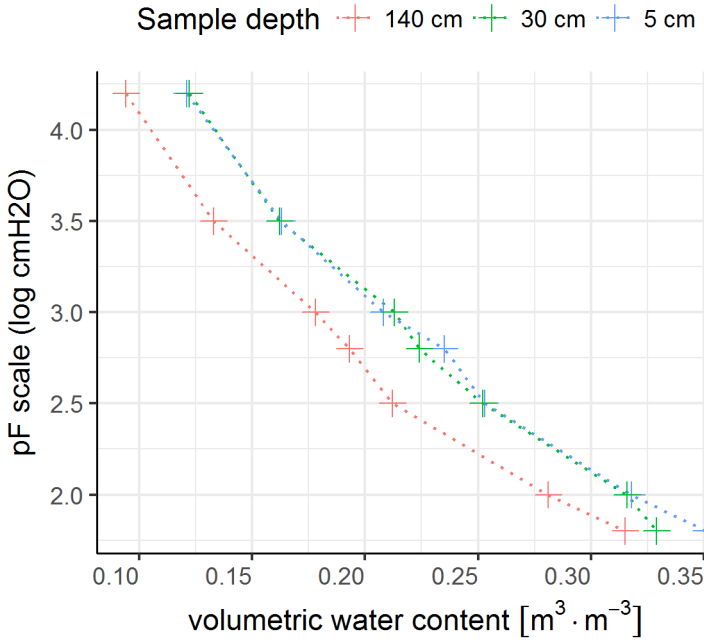

FIGURE 2.11: pF curves from soil analysis of the soil under the weather station in Stein, Switzerland. pF is defined as the base 10 logarithm of the suction expressed in cm of water. These water retention curves are for soil sampled at 5, 30 and 140 cm

even need to know the particle size distribution, but only the textural class (coarse, medium, medium fine, fine and very fine) to obtain water retention properties with a good accuracy (Al Majou et al., 2007).

The figure 2.12 shows the difference between the $\Psi - \theta$ curve from the Saxton model and the $\Psi - \theta$ curve predicted from the parameters extracted from the log-log of the pF curve. The Saxton model fits well with the observations between 0.12 and 0.2 $\mathrm{m}^3\,\mathrm{m}^{-3}$, but then the two models depart. The models hence differ in their estimation of the water content at field capacity, the water content for -10 kPa or -1.01 $\mathrm{m}_{\mathrm{H_20}}$. Field capacity is the water content of a particular soil when it has been soaked in water and then drained to remove the excess of water. Field capacity is predicted at 0.26 $\mathrm{m}^3\,\mathrm{m}^{-3}$ by the Saxton parameters and 0.32 $\mathrm{m}^3\,\mathrm{m}^{-3}$ for the regression model. The Saxton model uses three equations to model the $\Psi - \theta$ relationships, whereas here we only use the exponential part, which is only valid between permanent wilting point and field capacity.

To conclude, if water retention curves are not available, the Saxton model gives a good approximation of the a and b parameters used to convert water content into matric potential.

4.5 Model performance in Stein, Switzerland

In section 4.2, the verification of SOPHIA against a simplified version of Müller's model with Müller's Linden data set proved that the development of SOPHIA is comparable and that, except for small differences, there was no visible unexpected behaviour.

We will now run the two models with the data from our weather station installed in Stein, Switzerland. The weather station on the experimental site offers observation of soil temperature, water content and net radiation with which we will compare the performance of the two models.

For this comparison, we are going to compare the original version of Müller's model `Mod5-1b.mod` (table 2.1) without any simplification with the SOPHIA model presented in this chapter. To run Müller's model with the weather station input, we converted the relative humidity into dew-point temperature using the ambient vapour pressure e_a (equation for e_a in table 1.2)

$$T_{DP} = \frac{240.97 \cdot \log(e_a/0.61)}{17.27 - \log(e_a/0.61)} \tag{2.59}$$

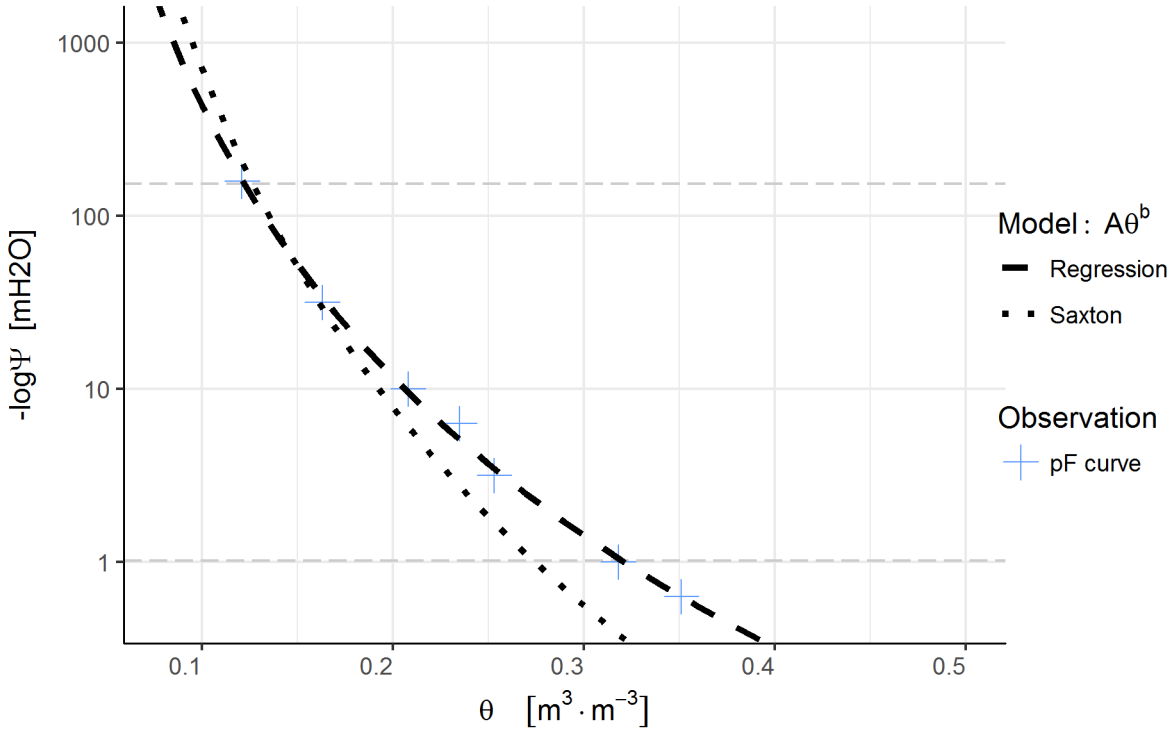

FIGURE 2.12: Relationship between the matric potential Ψ and the volumetric water content θ for the soil at 5 cm with two different parametrisations of Campbell exponential model $A\theta^b$ (Campbell, 1974). The first parametrisation comes from a log transformation of the pF curves (figure 2.11) and yields $A = exp(-2.3)$ and $b = -5.2$. The second parametrisation is obtained by the Saxton model with $A = exp(-3.9)$ and $b = -6.5$ (Saxton and Rawls, 2006) for the texture and organic matter of the Stein sample at 5 cm. We have only used the exponential part of the Saxton model. The dashed horizontal lines mark the permanent wilting point (-1500 kPa, top line) and field capacity (-10 kPa, bottom line)

For this comparison, we did not simplify Müller's model `Mod5-1b.mod`: here, the aerodynamic resistance is computed with a stability parameter, and the integration method is Runge-Kutta. The Müller and SOPHIA models were parametrized according to table 2.3, with a soil discretization of 14 layers (layer size from layer 1 to layer 14 is $0.01\,\mathrm{m}$, $0.02\,\mathrm{m}$, $0.03\,\mathrm{m}$, $0.04\,\mathrm{m}$, $0.05\,\mathrm{m}$, $0.05\,\mathrm{m}$, $0.2\,\mathrm{m}$, $0.2\,\mathrm{m}$, $0.2\,\mathrm{m}$, $0.2\,\mathrm{m}$, $0.2\,\mathrm{m}$, $0.1\,\mathrm{m}$, $0.1\,\mathrm{m}$ and $0.1\,\mathrm{m}$). The simulation with 14 layers outputs the state variables water content and temperature at 4.5, 30 and 145 cm whereas the soil sensors were positioned at 5, 30 and 140 cm. The discretisation in soil layers makes it difficult to keep the mid-point of layers at the same depth as the sensors. We assume that this point is insignificant and that measurements and simulation are comparable.

We compared the two models in a simulation between 1st of April and 10 April 2015. At the beginning of this period (first to five April), there was some rainfall and the radiation was low, whereas, for the next five days (5-10 April), there was no rain and the radiation was higher (it was sunny). We chose the month of April 2015 because air and soil temperature remained low, which diminished the likelihood of an effect of vapour movements created by high temperature. The effect of temperature on water movement (non-isothermal water flow) is not taken into account in either SOPHIA or `Mod5-1b.mod` (other version of Müller's model contain the non-isothermal water flow)

In table 2.3, we list the relevant parameters that we have set for the simulations and comparison of the two models in Stein. The soil parameters were either measured, inferred from water-retention curves, computed by the Saxton model or looked up in the reference tables. The surface parameters are the default values provided by Müller (1999).

TABLE 2.3: Parameter values for model run in Stein, Switzerland in figure 2.13 to 2.15

Symbol	Equation (#)	Value	Unit	Calibration from
Soil parameters				
a	Matric potential (2.7)	-2.3	-	fitting
b	Matric potential (2.7)	-5.2	-	fitting
b_{KW}	Hydraulic cond. (2.9)	5.2	-	Campbell and Norman (1998)
KW_{sat}	Hydraulic cond. (2.9)	2.6e-06	$\mathrm{m\,s^{-1}}$	Saxton and Rawls (2006)
f_{clay}	Thermal cond. (2.2)	0.27	-	measurements
f_o	Thermal cond. (2.2)	0.03	-	measurements
f_{sand}	Thermal cond. (2.2)	0.36	-	measurements

Symbol	Equation (#)	Value	Unit	Calibration from
ρ_b	Thermal cond. (2.3)	1.39	$\mathrm{g\,m^{-3}}$	Saxton and Rawls (2006)
Ψ_{min}	Surface infiltration (2.29)	-1000	m	Müller (1999)
θ_e	Surface infiltration (2.28)	0.49	$\mathrm{m^3\,m^{-3}}$	Müller (1999)
Surface parameters				
α	Albedo (2.12)	0.05	ratio	Müller (1999)
B_a	Atmosph. emissivity (2.16)	0.605	-	calibrated
B_b	Atmosph. emissivity (2.16)	0.039	-	calibrated
S	Soil emissivity factor (2.14)	0.9	-	Müller (1999)
z_0	Aerodyn. resistance (2.19)	0.01	m	Acs, Mihailovic, and Rajkovic (1991)
z_{ref}	Aerodyn. resistance (2.19)	2	m	measurement

The first comparison is soil temperature at three different depths in Stein, Switzerland. The RMSE is between 1.3 and 3.2 °C. The highest RMSE is found near the soil surface, where significant energy transfer occurs. The simulation of Müller predicts higher temperatures, especially during the sunny days at the end of the simulation. SOPHIA on the other hand, predicts colder temperatures at night. This difference is linked to differences in aerodynamic resistance (not shown here). Without stability parameter, the aerodynamic resistance is lower in SOPHIA and results in higher fluxes of sensible heat at night. The soil looses more energy than observed. During the day, the higher aerodynamic resistance of Müller lowers the sensible heat flux, more energy is kept in the soil and hence it is hotter than observed. The difference of hotter temperature for Müller and colder for Sophia is transferred to the deeper depth, at 30 and 140 cm.

When comparing simulations of soil water content with the observation in Stein, the importance of initial conditions is obvious, especially when a short period as this one is simulated and the effect of the initial conditions have insufficient time to disappear.

The RMSE for water is between 0.02 and 0.07 $\mathrm{m^3\,m^{-3}}$. The highest error affects the water content at 140 cm, where initial conditions are set at 0.35 $\mathrm{m^3\,m^{-3}}$ when they should have been set at 0.42 $\mathrm{m^3\,m^{-3}}$.

For this simulation, we set the initial water content as observed at 5 cm by the water sensor to have the right starting point to simulate conditions at the top, but obviously, we would need to set different starting conditions for 30 and 145 cm.

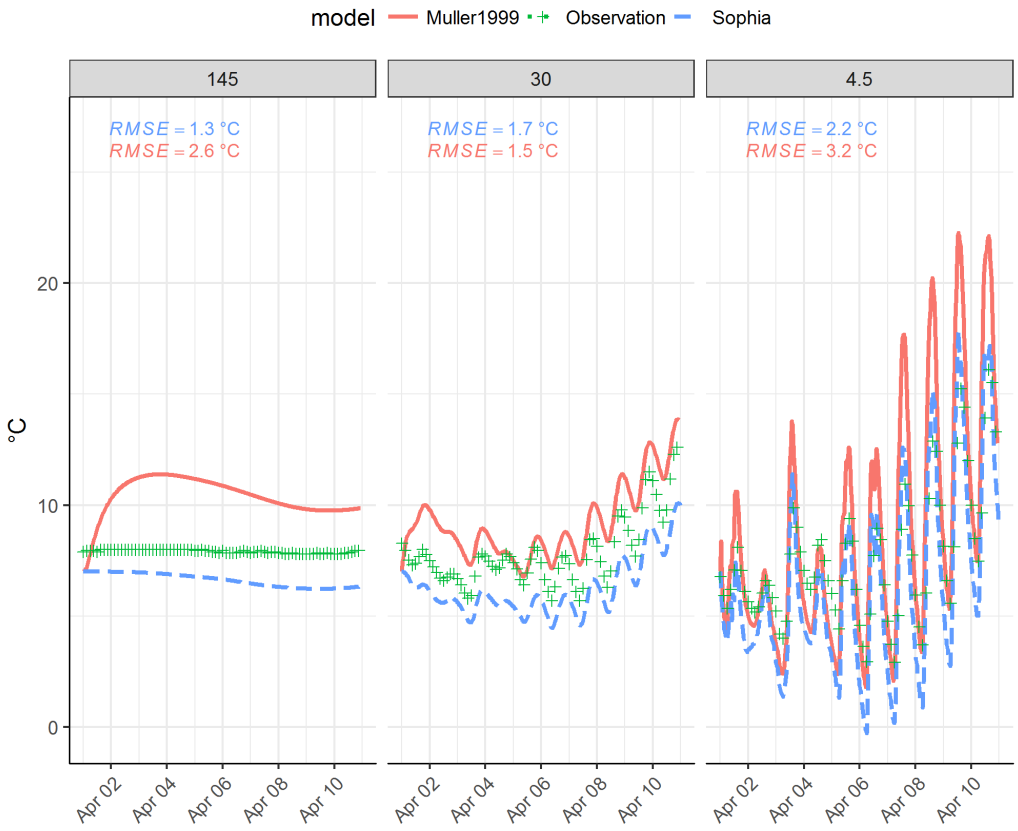

FIGURE 2.13: Simulation of temperature at three depths by the SOPHIA and Müller models compared to the observation from the automated weather station. Initial conditions were set to 7 °C for all layers for SOPHIA. The time step was 120 seconds for SOPHIA, and the soil was set to a total depth of 1.5 m divided into increasing size layers, with free drainage conditions at the bottom. The grey boxes on top of the graph are the depth of the model output in cm. It is compared with the output of Hydraprobe II temperature sensors (Stevens, USA) placed at 5, 30 and 140 cm and the RMSE for each model is indicated in color: red for Müller's model and blue for SOPHIA. For this simulation, Müller's model `Mod5-1b.mod` was not simplified, and its parameters were set to the values from table 2.3.

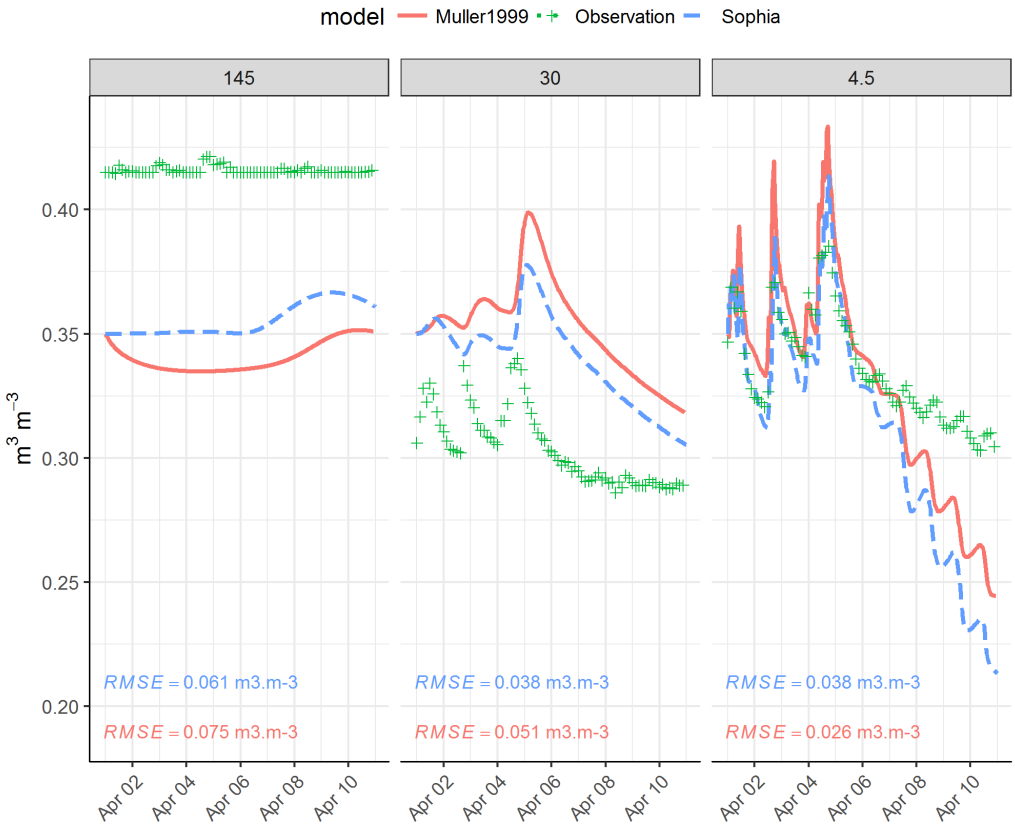

FIGURE 2.14: Simulation of soil water content at 3 depths by the SOPHIA and Müller models and observations from the automated weather station. Initial conditions were set to 0.35 $\mathrm{m}^3\,\mathrm{m}^{-3}$ for 5 cm, 0.3 $\mathrm{m}^3\,\mathrm{m}^{-3}$ for 30 cm and 0.4 $\mathrm{m}^3\,\mathrm{m}^{-3}$ for 150 cm. The time step was 120 seconds and the soil had a total depth of 1.5 m divided into increasing size layers, with free drainage condition at the bottom. The grey boxes indicate depth of the model output in cm. It is compared with the output of Hydraprobe II sensors (Stevens, USA) placed at 5, 30 and 140 cm. For this simulation, Müller was not simplified. RMSE is coloured according to the model: blue is for SOPHIA and red is for Müller's model.

The effect of rain is visible for both depths, at 4.5 and 30 cm. Both models react similarly, with the water content predicted by SOPHIA slightly lower than the one predicted by Müller. The error is smaller for SOPHIA at 30 cm but higher at 5 cm. Near the surface, the effect of rainfall is visible by the sharp peaks of water content. The peaks are higher for the simulations than for the observations.

Once the sun comes out on 5 April and the soil starts to dry, the soil is drying much faster than what is observed. In 4 days, the water contents drops from saturation to 0.20 $\mathrm{m}^3\,\mathrm{m}^{-3}$. If we carry on the simulation over the next days, the soils of the two simulation will be at their driest, 0.08 $\mathrm{m}^3\,\mathrm{m}^{-3}$. It is not possible to loose so much water, and there is obviously a problem for water content at the surface for both models.

The last opportunity to compare the SOPHIA and Müller models with measured data is with the net radiation $Rnet$ observations (figure 2.15). To recall, the Rnet equation is

$$\mathrm{Rnet} = \mathrm{GR}(1 - \alpha) + \mathrm{Ld} - \mathrm{Lu}$$

where GR is measured global radiation, α is the albedo and Ld and Lu are long-wave radiation downward (Ld) and upward (Lu). During rainy days, the simulations of long-wave downward radiation (Ld) are simulated similarly by both models but differ from observations. The algorithm used for long-wave radiation is only valid for clear-skies, and here we have a good example of how inaccurate it is when the sky is cloudy. In our algorithm, long-wave radiation depends solely on absolute humidity, but in reality, clouds are playing an important role (G. Flerchinger, Xaio, et al., 2009; John Monteith and M. Unsworth, 2013). When the weather is sunnier, and the sky is clear from clouds, the simulation and observation of long-wave downward radiation are in agreement.

For upward long-wave radiation, the simulation during rainy days fits well with the observations, but for sunny days, the observation of upward long-wave radiation is higher than what is simulated, especially by SOPHIA. The error is greater for SOPHIA than for Müller. Long-wave upward depends on the emissivity of the soil, which is difficult to estimate with certainty (Ham and Senock, 1992).

For the net radiation, Müller and SOPHIA are in agreement with each other, although Müller shows a lower error. The greatest disagreement from the measured observations happens at night during rainy days, when the net

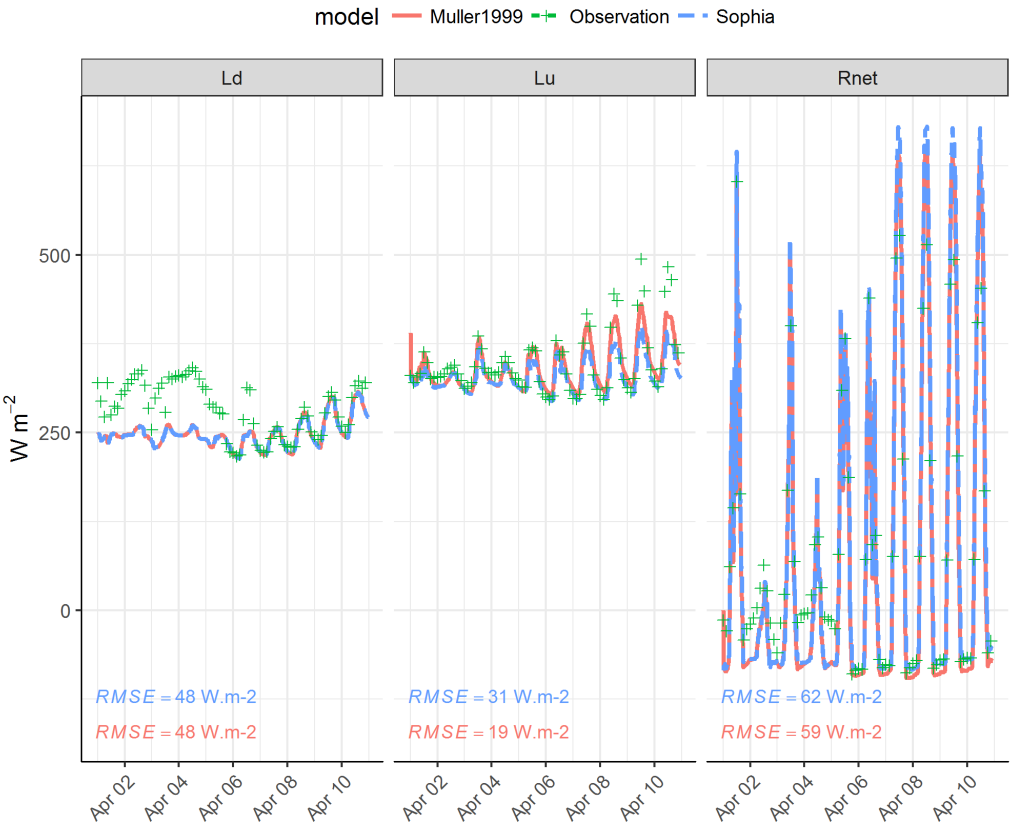

FIGURE 2.15: Simulation of net radiation (Rnet), long-wave downward (Ld) and long-wave upward (Lu) by the SOPHIA and Müller models and observations from the automated weather station. The models were run with the same discretization and parameters (see legend figure 2.14). RMSE is coloured according to the model: blue is for SOPHIA and red is for Müller's model.

radiation is higher than what is observed, around 0 $\mathrm{W\,m^{-2}}$. At night, the global radiation (input of the model) is null, and it is the difference between upward and downward long-wave radiation that drives the value of the net radiation. Hence, the observed net radiation is close to 0 because, during cloudy nights, the downward long-wave radiation is equal to the long-wave upward radiation and both terms cancel each other out. In the simulation, upward long-wave radiation is higher than downward long-wave radiation and they create a negative value of net radiation.

In table 2.4, we have compiled the RMSE displayed in figures 2.13 to 2.15 and added the relative RMSE. It gives an evaluation of the model at the end of the development phase. The main difference between the two models is the expression of the aerodynamic resistance. Sophia's error is smaller than Müller's for soil temperature at 4.5 and 145 cm and soil water content at 30 and 145 cm. The error of 2.2 $^{\circ}\mathrm{C}$ represent still an relative error of 30 $\%$ showing that the model would need additional improvement. These improvements would be efficient if targeted at the net radiation, where there is the highest relative error. The net radiation is decomposed with the upward and downward long-wave radiation. The error for upward long-wave radiation is relatively low, which bring us to the conclusion that the error would lie in the downward long-wave radiation and short-wave term of the net radiation. The algorithm used for long-wave radiation downward is not adapted for cloudy days, and a possible improvement, as we shall see in the next chapter, would be to include a cloud correction factor.

Conclusion of the comparison of the Müller and SOPHIA models on the experimental site of Stein The comparison of both models against observations shows that the models behave similarly overall, but that there are still some differences from the actual observations. The simulation for profound soil temperature and water content depends on initial conditions, and they should always be set as accurately as possible. For the soil at 30 cm, the simulations of water content and temperature had the lowest error in both models. At the surface, the error is greater, with a bigger difference between models and observations, notably for temperature, downward long-wave radiation and net radiation.

TABLE 2.4: Evaluation of the performance of `Mod5-1b.mod` from Müller (1999) and SOPHIA model on ten days in April 2015 in Stein, Switzerland. The parameters and detail of the simulation are detailed in table 2.1 and table 2.3. RMSE: root mean square error

	RMSE (relative RMSE)	
Variable	`Mod5-1b.mod`	SOPHIA
Net radiation [$\mathrm{Wm^{-2}}$]	59 (68 %)	62 (72 %)
Downward long-wave radiation[$\mathrm{Wm^{-2}}$]	48 (17 %)	48 (17 %)
Upward long-wave radiation [$\mathrm{Wm^{-2}}$]	19 (5 %)	30 (9 %)
Soil temperature at 4.5 cm [°C]	3.2 (41%)	2.2 (29%)
Soil temperature at 30 cm [°C]	1.5 (19%)	1.7 (21%)
Soil temperature at 145 cm [°C]	2.6 (33%)	1.3 (17%)
Soil water content at 4.5 cm [$\mathrm{m^3\,m^{-3}}$]	0.026 (8%)	0.038 (11%)
Soil water content at 30 cm [$\mathrm{m^3\,m^{-3}}$]	0.051 (17%)	0.038 (13%)
Soil water content at 145 cm [$\mathrm{m^3\,m^{-3}}$]	0.075 (18%)	0.061 (15%)

5 Discussion and summary

5.1 Discussion and future developments

The development of the model is now complete. We described the essential steps required to model the soil temperature and water content for a bare soil with an energy balance equation at the surface. Literature on energy balance models in agricultural context is abundant (Campbell and Norman, 1998; Daniel Hillel, 2003; John Monteith and M. Unsworth, 2013) but books detailing the implementation of the model as in Müller (1999) or the more recent Bittelli, Campbell, and Tomei (2015) are more scarce. These nevertheless represent essential sources to reliably reproduce the work presented by the authors. Reproducing their work was in fact the first logical step of our study, which initially aimed at exploring these models' properties, at comparing them and at evaluating their adequacy with our original dataset and our modelling objectives. Therefore, the choice of Müller (1999) as the main inspiration of our work was a natural one.

During the processes of model implementation, we included some adaptations from the original formulation of Müller (1999). We also applied these adaptations to Müller's model in order to have two comparable versions of the model to assist us in model development. At the end of model development, both models yielded equal soil temperatures on Müller training dataset. The model was then tested with different soil parameters from different texture classes. The model behaviour reacted as expected for the different texture

classes: fast drying in sandy soils and good water holding in clay soils.

The final model developed in this thesis, called SOPHIA, consists of 30 equations. These adaptations allowed us to keep the model as simple as possible, in order to have reasonable simulation durations, a fine control over all the aspects of the model and a modular development.

We will now use SOPHIA in this form to perform sensitivity analysis and estimate the parameters. The first version of SOPHIA gave satisfying results, both regarding stability, comparison with the reference model of Müller (1999), sensitivity to soil parameters and first evaluation.

We coded the model in the R language because it is rapid for prototyping. Besides, the statistical methods for sensitivity analysis and parameter estimation are already implemented. However, we should switch to a compiled language like C to improve the speed of the model in case we want to apply it to a large scale (for example with a grid of input data) or we need to run a large number of simulations for sensitivity analysis.

Besides the writing of the model equations themselves, the simulation aspect is also a challenging task, i.e. run the model so that it provides numerical results. Some relatively recent publications use an explicit scheme (Banimahd and Zand-Parsa, 2013) but most models use the implicit scheme (Grifoll, Gastó, and Cohen, 2005; J Simunek et al., 2013; Bittelli, Campbell, and Tomei, 2015), to avoid problems of stability of the water transport. In our experimental setting, the explicit numerical scheme is stable with the selected time and space discretisation, but we noted issues of stability at high moisture content or when heavy rain episodes occurred.

With such a nonlinear model, an implicit scheme is recommended. The implementation is much more difficult not only because it needs solid mathematical knowledge in equation derivation, but also because the implicit scheme is rarely described and rather only mentioned in publications. The numerical scheme is only described in books like Bittelli, Campbell, and Tomei (2015) or software manuals such as the one from software HYDRUS-1D (J Simunek et al., 2013). The book Bittelli, Campbell, and Tomei (2015) details the implementation of the implicit scheme in the Python language and will be a good reference for the improvement of SOPHIA.

After the first evaluation, without calibration of surface parameters, the error of the model is 2.2 °C on a period of 10 days in April 2015, better than the simulation of by Müller's model. This is a promising result showing that an energy balance model simulates soil temperature near the surface in the correct range. Other similar energy balance models, for example the SHAW

model from G. N. Flerchinger (2000) applied in Bullied, G. N. Flerchinger, et al. (2014) reports error around 2 °C as well. However, in relative terms, this is an error of 29 %, which is above our objective of 20 % of error. It is leading us to discuss aspects of the SOPHIA model that could be improved.

The highest relative error is for net radiation. There is a clear improvement that needs to be done for this term. Improvement should consider the calibration of the parameters involved in the net radiation, B_a, B_b in equation 2.16, S in equation 2.14 and albedo α in equation 2.12. The calibration of these term will be showed in the next chapter.

A weak point of SOPHIA model is that is has only been validated at one location: in Stein, Switzerland. Before its deployment, the model should be validated at other locations in different climatic areas. The limiting factor within this validation process is to obtain bare soil temperature data to validate the model. Most of the weather stations are installed on grass, which made them unfit for our model's evaluation. The best approach would be to add a layer of vegetation to the model, so it is able to simulate soil temperature under a grass cover and could be compared with this data that is available easily. The modification of the thermal conductivity of the first layer of soil could be a way to transform a bare soil model into a soil model covered with grass, as suggested by Kearney et al. (2014).

A last point to discuss is the addition of new features. SOPHIA, in its actual form, is a minimum viable product intended to be improved according to its future use. One could think of the addition of a layer of mulch or residues over the bare soil surface. In this case, recommendation from Bussière and Pierre Cellier (1994) will be helpful. The addition of this feature will make the model more suited to regions of the world where the practice of no-till is widespread, like North and South America.

A non-homogeneous soil is another important feature that could be added, i.e. a soil with layers of different properties. We could also think of dynamically changing soil properties because it is known that, near the surface, the soil parametrisation changes also with time in response to biology, climate and management (Gupta, Lowery, et al., 1991; Vereecken et al., 2016). Plant canopies and episodes of drought and torrential rains turn the soil structure into a thick crust that slows down the emergence of maize (C Durr et al., 2001).

Soil properties are also a topic of interest when it comes to improving our model. In the current version of SOPHIA, we use a stand-alone Saxton model to obtain the parameters for the soil. However, a future improvement could consist of directly integrating the equation of the Saxton model with the code

of SOPHIA. With this integration, the model would need only the particle size fraction as input and parameters a, b, ρ_b and Kw_{sat} would be computed automatically. We could also take as an input the fraction of rock fragments, which have a significant influence on soil properties (Cousin, Nicoullaud, and Coutadeur, 2003).

The last feature that could be implemented is the coupling of temperature and water transport. This addition would be necessary if the model were to be deployed in arid climates or if the model needed to run at times when the soil was covered in snow. In such cases, the gradients of temperature within the soil create gradients of water vapour that have an influence on liquid water movements (G. N. Flerchinger, 2000; Grifoll, Gastó, and Cohen, 2005; Bittelli, Campbell, and Tomei, 2015).

With these possible improvements as a guideline for future development, we have to keep focused on the simplicity of the model, which was one of the main objectives when designing SOPHIA. Such simplicity will prove essential in the global sensitivity analysis that we will run in the upcoming chapter. In fact, it will be showed that our model's relative low number of parameters supports the efficiency of simulations.

5.2 Summary

The development of SOPHIA, our soil model presented in this chapter, is the central pillar of the thesis. Its equations are derived from Müller (1999). The aerodynamic resistance was simplified and relative humidity was taken as the input instead of the dew-point temperature. SOPHIA has the minimum set of equations needed to simulate soil temperature and water content with accuracy. With more than 30 equations and 43 parameters, among which 15 may be used for calibration, the model, while kept as straightforward as possible, still is large and complex. The energy balance equation is composed of many sub-models, each representing a small part of the complex reality at the soil-atmosphere interface.

The innovative part of this chapter is the mathematical formalism simplifying the physics equation and enabling a better view of the general form of the model. This form was used to consider, together with mathematicians, the numerical scheme needed to resolve the equation. We chose an explicit scheme because it was easy to implement, and we showed that the explicit

numerical scheme is stable with selected time and space discretisation. However, to increase the stability of the model for high water content, an implicit numerical scheme should be programmed.

The validation of the SOPHIA by comparison to Müller's model showed that both models behave similarly, with a few small differences that could be explained by the numerical scheme.

With, in our hands, a model that is capable of simulating the soil temperature and moisture appropriately, we may take the next step : perform a sensitivity analysis and estimate the relevant parameters.

Chapter 3

Sensitivity analysis, parameter estimation and simulation

Although often overlooked by researchers, model analysis represents an important step in the process of model development. Deeper knowledge of the behaviour of a model may help identify its strengths and weaknesses. These in turn point towards future developments, whether new or to be made a priority, so as to define the model's range and level of validity as well as to identify its scope of applicability.

This chapter aims at exploring in more detail the behaviour of our model. We will first present both local and global sensitivity analysis of the SOPHIA model. Methods of global sensitivity analysis have been developed to explore the effect of parameters on their whole intervals of variations and including their potential interactions (Saltelli, Tarantola, et al., 2004). The sensitivity analyses that have been performed, albeit rarely, on such soil energy balance models in the past, have mostly preferred one-at-a-time approaches over global ones (Alvenäs and Jansson, 1997; Banimahd and Zand-Parsa, 2013). Global sensitivity methods are more commonly applied to land surface models used for climate prediction (Collins and Avissar, 1994; Petropoulos et al., 2009).

Once the important parameters are identified, a step of parameter estimation can be performed to refine the values of these critical parameters and improve the simulation results in comparison with observed data. The added value of parameter estimation can be evaluated computing the AIC criterion (Wallach et al., 2014).

In practice, most parameters are determined based on literature or expertise, and only a few are estimated. Thanks to the net radiometer installed on the weather station, this chapter presents a refined analysis of the net radiation term for which we propose a new expressions to improve model accuracy.

To finish, the calibrated model was evaluated over different periods in

Stein in 2015 and 2016. The prediction of the model agrees well with observed values. It will also appear that the temperature prediction is less accurate in summer than in winter, probably due to vapour movements that are not taken into account by the model.

1 Sensitivity analysis

Sensitivity analysis (SA) studies how uncertainties about the inputs of a model impact its output. These inputs may for instance be parameters, initial values or control variables and will hereafter be designated as *factors*. Additionally, we assume that the output of interest is of dimension 1; if several outputs are to be considered, they will receive independent treatment. SA aims at identifying the most important factors and, in general, at ranking them according to their relative effect on the output. Ranking may help determining which factors need calibration and where the model's weaknesses sit. This may prove instrumental in correctly interpreting the model results (J Cariboni et al., 2007).

Sensitivity analysis comprises local and global techniques. Local techniques study the output's sensitivity using "One-At-a-Time (OAT)" methods, where everything is held constant except for the one tested factor. OAT methods are easily performed on any model, whichever the number of factors. They provide a first glance at the model's sensitivity at low computational cost. However, OAT methods show their limits when one wishes to observe the effect of the variation of one factor as other factors vary as well. In other words, the OAT methods miss the effect of between-factors interactions. OAT methods are suitable for a first glance at the model but it is recommended to deepen their results using global techniques. Global techniques intend to estimate the effect of a factor on the output while all the other factors are varying as well, i.e. including the effects of its interactions with other factors and not only its individual effect. Such global methods tend to explore the whole parameter space, in contrast to local methods that provide results only at a given point (Saltelli and Annoni, 2010). Several global SA methods have been developed, and two of them are used in this chapter: a global screening method (Morris method) and a variance-based method (Sobol's method).

1.1 Range and distribution of parameters

We would like to emphasise that in sensitivity analysis, the term *factors* generally encompasses parameters as well as, for instance, the control variables of the model (in our case, the environmental variables; e.g. global radiation). Here, we have restricted our analysis of the sensitivity of the model to the variations of parameters, namely the constants used in the model's equations. Therefore, the term "parameter" shall be preferred over "factor" in this chapter.

The SOPHIA model presents a total of 42 parameters: 28 of them are physical constants, listed in the preamble, page xxiii, for which there is evidently no uncertainty to take into account. Therefore only 14 parameters are relevant for the sensitivity analysis (table 3.1).

These 14 parameters are categorized into two classes: soil and surface parameters. Soil parameters are found in equations related to water and heat movement within the soil while surface parameters are involved in the energy balance equations.

Proceeding to a sensitivity analysis first requires to define the range of uncertainty for each parameter: the subsequent paragraphs report how these ranges were chosen.

Table 3.1 shows the reference value and range of parameters selected for sensitivity analysis. For all of them, we selected uniform distributions, although it is likely that in reality some values are more probable than others. In fact, this kind of information about the parameters' shape of distribution is difficult to obtain and very uncertain: for instance, gathering datasets on soil parameters from a meta-analysis on published studies or from a large sampling program would provide biased results, because of the predominance of certain types of soil in nature while we would prefer a uniform sampling over all types of soil, without over-representations of some types. Besides, we are here interested in sensitivity indices of the model parameters over all their ranges of uncertainty without favouring certain parameter regions. Finally, from a more pragmatic point of view, the uniform distribution is conveniently defined merely by its minimal and maximal values, and it is easy to draw samples from it. For soil parameters, the reference value of the fraction of organic matter f_0 is measured. The reference value for a, b, b_{kw}, Kws and ρ_b are obtained with the Saxton model using as an input 27 $\%$ clay and 36.3 $\%$ sand, 3 $\%$ organic matter, no gravel and a density factor equal to 1. The rest of soil parameters reference values are taken from Müller (1999) or Campbell and Norman (1998).

TABLE 3.1: List of the reference value, the inferior and superior boundaries for the 14 parameters included in the OAT (One-at-A-Time) and Morris sensitivity analysis. The reference value is either taken from Müller (1999), measured or computed with the Saxton model with a density factor of 1 or no gravel. Kws is expressed in $\mathrm{m\,s^{-1}}$. This range corresponds to $0.1\,\mathrm{mm\,h^{-1}}$ and $26.4\,\mathrm{mm\,h^{-1}}$.

Symbol	Reference value	Inferior boundary	Superior boundary	Boundary reference
Surface parameters				
B_a	0.605	0.5	0.66	Iziomon, Mayer, and Matzarakis (2003)
B_b	0.039	0.037	0.066	same as B_a
S	0.9	0.8	0.91	Ham and Senock (1992)
α	0.05	0.02	0.5	Campbell and Norman (1998)
z_0	0.01	0.001	0.1	Campbell and Norman (1998)
z_{ref}	2.0	0.2	3	expertise
Soil parameters				
f_o	0.03	0	0.08	Saxton model
a	-3.9	-6.32	-3.1	Saxton model
b	-6.5	-7.9	-6.16	Saxton model
b_{Kw}	6.5	6.16	7.9	Campbell and Norman (1998)
Kws	2.67e-6	2.04e-8	7.33e-6	Saxton model
ρ_b	1.39	1.1	1.7	Saxton model
Ψ_{min}	-1000	-1500	-152	Müller (1999)
θ_e	0.49	0.4	0.5	Müller (1999)

Symbols surface parameters: B_a and B_b Brunt's coefficient; S soil emissivity factor; α albedo,z_0 the roughness length [m] and z_{ref} the reference height [m], both involved in aerodynamic resistance on convective and latent heat flux. f_o the organic matter content; **Soil parameters**: a and b soil-moisture retention curves; b_{Kw} the exponent in the relationship between hydraulic conductivity and water content; ρ_b the bulk density [$\mathrm{g\,cm^{-3}}$]; Kws the hydraulic conductivity at saturation [$\mathrm{m\,s^{-1}}$]; Ψ_{min} the minimum water potential [m] and θ_e the air-entry water content [$\mathrm{m^3\,s^{-3}}$].

Surface parameters

There are six surface parameters: B_a and B_b, the Brunt's coefficients (equation 2.16); S, the soil emissivity factor (equation 2.14); α, the albedo (equation 2.12); z_0, the roughness length (equation 2.19) and z_{ref} (equation 2.19) the reference height.

$\mathbf{B_a}$ and $\mathbf{B_b}$ The ranges for the Brunt's parameter B_a and B_b are extracted from a review by Iziomon, Mayer, and Matzarakis (2003). In this review, the authors report the parameter values of the Brunt's formulation for clear-sky radiation from seven previous studies. We borrowed the minimum and maximum values from these studies to obtain the ranges for these parameters.

$\mathbf{S}$ The soil emissivity factor S is used to compute the soil emissivity ϵ_s. The boundaries of the soil emissivity factor (S) intervals 0.8 and 0.91, which generates a variation of soil emissivity ϵ_s between 0.82 to 1 when in parallel θ changes from $0.1\,\mathrm{m^3\,m^{-3}}$ to $0.5\,\mathrm{m^3\,m^{-3}}$. Ham and Senock (1992) found an emissivity between 0.82 and 0.84 for a sandy soil and between 0.91 and 0.96 for a silt loam soil. Thus, the range of values of ϵ_s obtained from this reference is well covered by the interval generated by the variability of S within its range. It is worth noting that our interval for ϵ_s includes the value of 1 because we noticed that when $\epsilon_s = 1$, the error on net radiation is reduced (see later, section 2.3). The range for S was slightly different for Sobol as explained in the application of Sobol sensitivity analysis page 116.

α The range for albedo α is wide: from 0.02 to 0.5. Indeed, albedo changes with the soil colour, the surface roughness, the slope of the terrain and the presence of snow (Daniel Hillel, 2003). Albedo circles around 0.5 when the soil is covered with an old snow layer. For a bare soil without snow cover, the albedo varies between 0.08 for a wet dark soil to 0.18 for a light dry soil, and between 0.24 and 0.26 for grass cover (Campbell and Norman, 1998).

$\mathbf{z_0}$ The roughness length z_0 is between 0.02 and 0.06 for a bare soil surface according to John Monteith and M. Unsworth (2013); we chose a range from 0.001 to 0.1 in order to include values of roughness length corresponding to a young vegetative cover.

$\mathbf{z_{ref}}$ The standard height for measuring atmospheric variables (temperature, wind speed and humidity) in agriculture is z_{ref} =2 m (WMO, 2012). However, this is not consistently respected as sometimes sensors found installed lower or higher than this recommended height. We wanted to check the impact of reference height on model output and hence we chose to vary this factor between $0.2\,\mathrm{m}$ and $3\,\mathrm{m}$ above ground.

Soil parameters

Ten soil parameters need to be included in the sensitivity analysis as they present some uncertainty: f_{clay}, f_{sand}, f_o, the content or volume fraction of clay, sand and organic matter (equation 2.2); a and b, the water-retention curve parameters (equation 2.7); b_{Kw}, exponent for hydraulic conductivity (equation 2.9); Kws, hydraulic conductivity parameter (equation 2.9); ρ_b, the bulk density (equation 2.2); Ψ_{min}, the minimum water potential (equation 2.15) and θ_e, the water content at air-entry suction (equation 2.29).

To obtain consistent parameters ranges for the sensitivity analysis, we fixed the clay and sand content and varied the value of other soil parameters. Clay and sand content were fixed to their values measured at 5 cm below ground in the experimental field in Stein (27 % clay, 36.3 % sand). Then, we relied on the Saxton model (Saxton and Rawls, 2006) to find the inferior and superior boundaries of a, b, b_{Kw}, Kws and ρ_b by varying the content of organic matter, density factor and percentage of gravel in the soil for the soil texture in Stein. As seen in chapter 2, the Saxton model predicts the value of these five soil parameters using soil texture data.

We have chosen the ranges of parameters corresponding to only one texture class because soil parameters are not independent from each other: they depend on soil texture. Soil hydraulic conductivity is not the same in a sandy soil and in a clay soil because sand and clay hold water differently: unlike sand, clay is capable of retaining a lot of water. Hence, in order to run the model with consistent parameter values (i.e. not mixing for instance a sandy texture with the hydraulic conductivity of clay), soil parameter ranges were extracted from Saxton model results corresponding to only one texture class.

For a given soil texture class, the Saxton model has three additional parameters that generate some variability within one texture class. These three parameters are the content in organic matter f_o, a density factor and the percentage of gravel:

- Organic matter has a direct impact on the water retention curve parameters a and b because they influence soil aggregation and associated pore space distribution. Soils with large organic matter content have a greater capacity to hold and conduct water. We set values of organic matter between 0 and 8 %.

- The density factor has an impact on soil pores geometry (Richard et al., 2001) and conductivity at saturation (Kws). An unusually loose soil

(with a density factor of 0.9) usually holds more water and conducts water faster than a compacted soil (1.2 or more). We set values between 0.9 and 1.3 for the density factor.

- The percentage of gravel (i.e. soil particles larger than 2.0 mm) decreases the pore space and modifies the water retention properties (Cousin, Nicoullaud, and Coutadeur, 2003).With high gravel content, the soil is capable of holding less water. Gravel contents between 0 and 40 $\%$ of soil weight (corresponding to 0 and 26 $\%$ of soil volume) were used to create variability in soil parameters.

Thus, by varying the three parameters in the Saxton model, we were able to obtain inferior and superior boundary values for the other five soil parameters (a, b, b_{Kw}, Kws and ρ_b, see table 3.1).

a and b The density factor has the highest impact on a and b. With a low density factor and high organic matter, we find the upper boundary for a (-3.1) and b (-6.15). With the high density factor, a and b shift their lower boundary values (-6.3 and -7.9 respectively).

The range for a was modified for the Sobol sensitivity analysis (lower -2.74, upper -1.34). For Sobol, we wanted to investigate a range which contains the value of a found with measurement of moisture-retention curves, as presented in section 4.4, page 76.

$\mathbf{b_{kw}}$ In Campbell (1974), the curvature parameter for hydraulic conductivity b_{kw} is defined as the inverse of b (the water retention curve parameter). For the range of b_{kw} we could therefore take the inverse of the boundaries for b.

Kws The hydraulic conductivity at saturation is maximum with high organic matter content because the low density of organic matter allows a better conductivity for high moistures. The lower boundary value is obtained with a high density factor. When the pore size is decreased with compaction, the hydraulic conductivity decreases as well.

$\rho_\mathbf{b}$ The superior boundary of bulk density ($1.7\,\mathrm{g\,cm^{-3}}$) is obtained with a gravel content of 40 $\%$ of soil weight. Gravels have a high density and hence increase the mass of solids per volume of soil. The lower boundary for bulk density ($1.1\,\mathrm{g\,cm^{-3}}$) is obtained with a high ($8\,\%$) content of organic matter, because organic matter particles exhibit a low density.

Two last soil parameters were not determined by the Saxton model: the minimum water potential Ψ_{min} and the water content at air-entry suction θ_e. Their inferior boundary values were taken from Müller (1999). The superior

boundary of Ψ_{min} was assigned to -152 $\mathrm{m_{H_2O}}$, which corresponds to the permanent wilting point of the sample of Stein soil at 5 cm. The water content at air-entry suction θ_e conditions the rate of water infiltration. Above this threshold, the soil is at saturation; below this threshold, the largest soil pores start to empty. It means that the lower the point of saturation, the slower the soil dries after a rain event. Its boundary values were chosen to be 0.4 and 0.5 $\mathrm{m^3\,m^{-3}}$ to be close or equal to saturation.

1.2 Local OAT sensitivity analysis

Following a classical approach, we began the exploration of our model behaviour by a One-At-a-Time and local sensitivity analysis consisting in simply running the model while changing only one parameter at a time and looking graphically at the results (Wallach et al., 2014). Each parameter's uncertainty interval was discretised in 25 evenly distributed values across their respective ranges (table 3.1 page 96). During the simulations, the other parameters were fixed at the reference value.

Choice of two contrasted days for performing the sensitivity analysis

In figure 3.1, page 102, and figure 3.2, page 104, we chose to stop the model at two different times, 4 April 2015 at 17:00 and 25 April 2015 at 13:00 to observe the effect of various meteorological conditions on the response of the model. The simulation was started 24 hours before with initial conditions equal to the observed conditions at 5 cm. 4 April 2015 was a cold and rainy day, 17:00 was a time with low radiation and high water infiltration while 25 April 2015 was a sunny, bright-sky day and at 13:00, we were close to the maximum in soil temperature at 5 cm and radiation. We chose both days in April as it is a usual sowing month in the region and those two days were representative of the conditions encountered during this month.

For the simulations, we chose a time step of 900 seconds and our optimal soil discretisation of 10 layers of 10 cm in order to ensure good stability of the model with the explicit numerical scheme.

The graphical study of the model response provides a first indication on the local monotony of the model in the vicinity of the reference point. Namely, it indicates whether the output is an increasing or decreasing function of the parameter value, or whether the reference point is a local optimum. Figure 3.1, page 102 presents the effect of changing surface parameters (B_a, B_b, S, α,

z_0 and z_{ref}) on soil temperature and water content at 5 cm. Figure 3.2, page 104 does the same with soil parameters.

Amplitude of output variations in response to independent and standardised changes of the parameters

Some differences may be highlighted in the model's response to parameter variation between the two days chosen for this local sensitivity analysis (figures on page 102 and page 104). When the surface parameters were varied on 4 April, a rainy day with low radiation, the temperature reached a maximum of $6\,°C$ and a minimum of $4.7\,°C$, amounting to an amplitude of $1.3\,°C$. Taking as a reference point the temperature of $5.2\,°C$ obtained in the simulation with the reference parameter values, the variation caused by surface parameters represented $25\,\%$ of temperature variation. For 25 April, the variation caused by surface parameter variation was comparable: $21\,\%$. Regarding the water content, the variation due to surface parameters reached only $0.3\,\%$ on 4 April and $8\,\%$ on 25 April. Hence surface parameters had comparable impact on soil temperature at 5 cm for both days, and a lower impact on soil water content, especially on 4 April 2015.

The amplitude of simulated temperature is narrower with varying soil parameters (figure 3.2) than it is with surface parameters. This was expected as we are studying the model response of two variables at low depth i.e. very close to the surface, hence emphasising the importance of the associated parameters. The amplitude of model responses induced by variations in soil parameters were no higher than 0.5 °C both for 4 and 25 April, which relatively amounts to respectively 10 % and 3 % of the reference. On the other hand, the amplitude generated by changing soil parameters is higher for water content. The variation due to the change in soil parameters was $38\,\%$ on 4 April and $14\,\%$ on 25 April, which is a ten-fold and two-fold increase compared to what was observed with surface parameters. Hence, when soil parameters vary, there is a strong effect on soil water content, while when surface parameters vary, it most powerfully affects temperature fluctuation, especially on sunny days.

A closer look at the effect of surface parameters (figure 3.1, page 102)

The responses to change in surface parameters have opposite variation directions for temperature and water content. For instance, an increase in albedo α induces a temperature drop and a water content rise. This is logical, since a

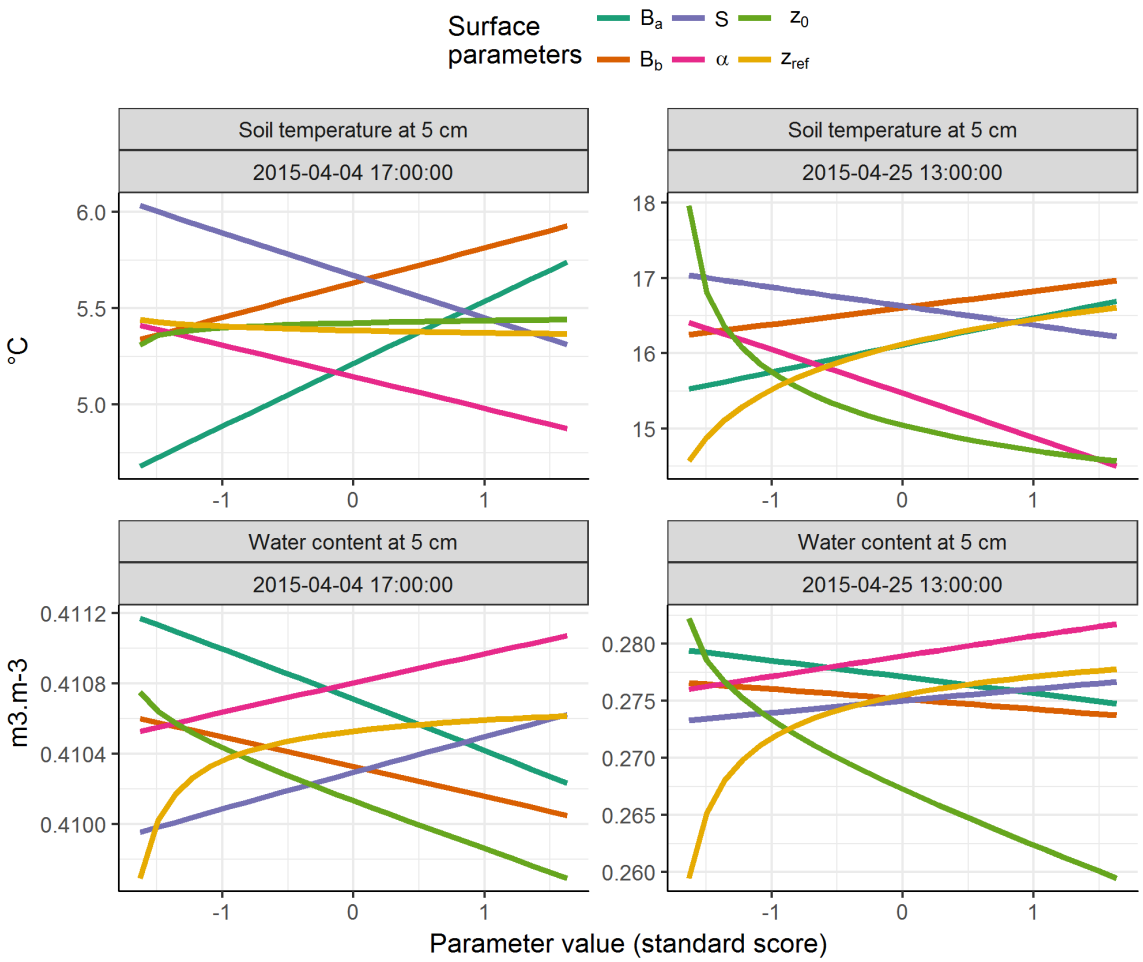

FIGURE 3.1: One-(factor-)At-a-Time and local sensitivity analysis of surface parameters for soil temperature and water content at 5 cm for a clay loam soil from Stein. The response of the model was studied at two different times: 4 April 2015 at 17:00, the end of a rainy day, and on 25 April 2015 at 13:00, in the middle of a sunny day. The parameter values were drawn from the ranges given in table 3.1, page 96. For the graphical representation, we computed the standard score of parameter values so the parameters exhibit the same mean and standard deviation. 1 on the scale represents one standard deviation. The model was run with a time step of 900 seconds, and the ten soil layers of 10 cm were initialised with the observed conditions at 5 cm.

drop in temperature will lower evaporation, therefore leaving a higher water content in the soil.

On 4 April, the parameters for aerodynamic resistance z_0 and z_{ref} had a nearly nonexistent influence on soil temperature, as their curve is approximately constant throughout the change in parameter value. For soil water content at 5 cm, the variation of parameters did provoke a change, albeit a small one. On 25 April, the change in z_0 and z_{ref} caused important variations in both soil temperature and soil water content.

These results are mainly interesting as regards their interpretation, which may shed light onto the model behaviour. To begin with, the effect of soil albedo α was expected, since it directly impacts the amount of radiative energy received by the ground surface. The higher albedo is, the more radiation from the sun is reflected back into the atmosphere. Hence, a higher albedo means a colder surface of the ground. There is a similar effect for the soil emissivity factor S. When it increases, the long-wave upward radiation builds up as well, allowing more outward radiation and thus more energy loss from the soil. A greater soil emissivity also equates to a colder soil. Hence, the soil is colder when parameters increase loss of radiative, sensible or latent energy from it.

The roughness length z_0 is the height of the static air layer above the ground. When this layer of static air increases, it means that the layer of mixed air between z_0 and z_{ref} decreases, thus causing the aerodynamic resistance to decline. This reduction of resistance increases the convective heat flux H and the evaporation E: the soil surface loses more energy and hence is colder. There is also an effect of z_0 on water content: when z_0 increases, the resistance drops and the evaporation rises, hence taking more water away from the soil.

The Brunt's coefficients related to long-wave radiation downward also affect soil temperature. The greater they are, the more energy is conveyed through long-wave radiation downward into the ground. This additional radiative energy towards the ground results in higher soil temperature.

Soil parameters in details (figure 3.2, page 104)

Among soil parameters, three parameters (b, ρ_b and Kws) show an interesting model response.

The parameter b drives the slope of the relationship between water content and matric potential (equation 2.7). The closer b is to 0 (superior boundary), the flatter the relationship, which means that the water is only loosely held

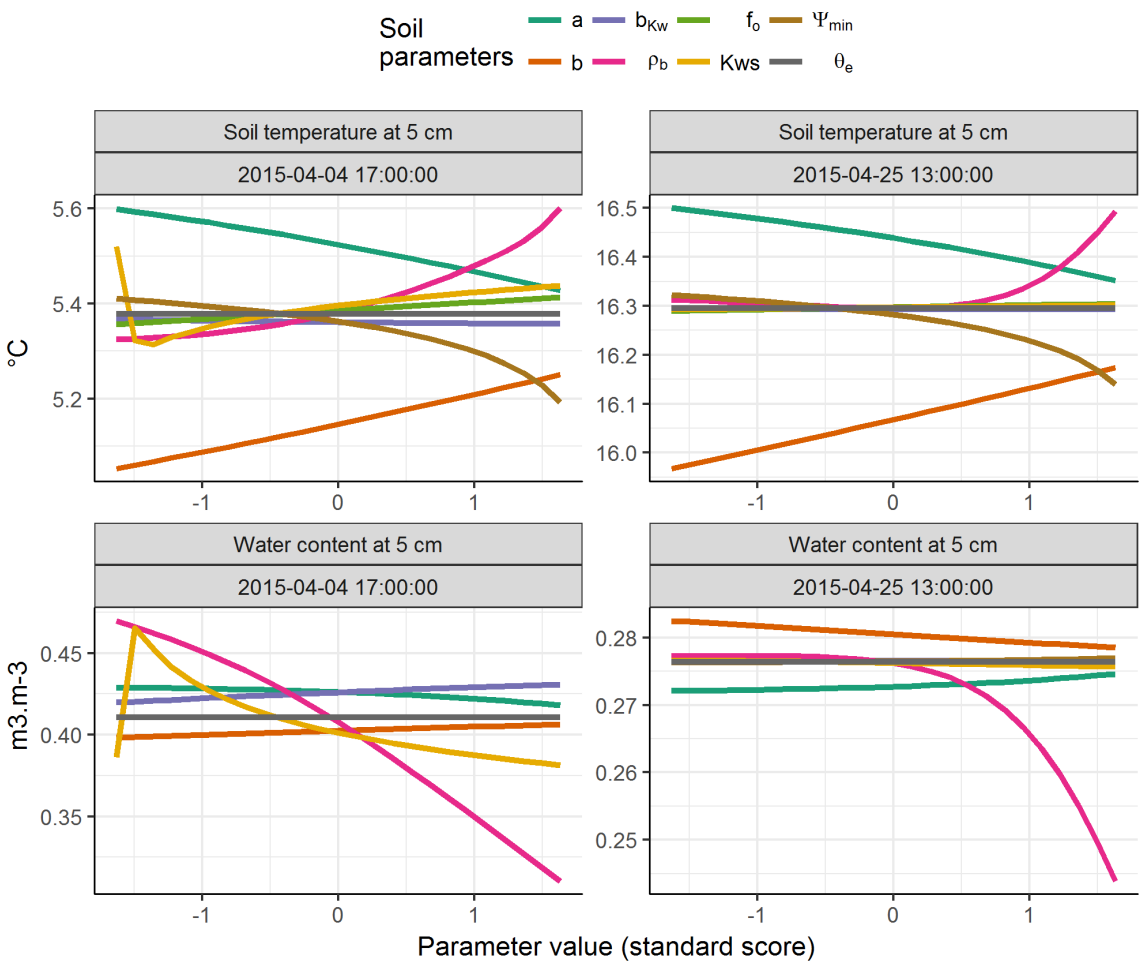

FIGURE 3.2: One-(factor-)At-a-Time and local sensitivity analysis for soil temperature and water content at 5 cm for a clay loam soil from Stein. The parameter values were taken across the ranges given in table 3.1, page 96. The response of the model was studied for the same dates in figure 3.1: on 4 April 2015 at 17:00 and on 25 April 2015 at 13:00. The model was run with a time step of 900 seconds, and the ten soil layers of 10 cm were initialised with the observed conditions at 5 cm.

in the soil. On the contrary, a negative b (lower boundary) represents a soil which has a high affinity for water. This question of affinity for water is only visible at low water contents (less than $0.2\,\mathrm{m^3\,m^{-3}}$). Indeed, the water content of these two days was too high and does not let us see a strong effect of b on soil water content. Interestingly, especially on 4 April, b had a stronger effect on soil temperature than on water content, generating 3.8 $\%$ variation of temperature (relative to the reference temperature, 5.2 $^\circ\mathrm{C}$) compared to 2 $\%$ of variation of water content (relative to 0.41 $\mathrm{m^3\,m^{-3}}$). The variation of b towards values closer to 0 produces a soil that dries faster (water loosely fixed to the soil), which could explain the effect on temperature (less water means higher temperature). However, in this OAT graphs, the change of b does not show an effect on soil water content.

The parameters Kws and ρ_b have a more significant impact on water content. The bulk density ρ_b is directly linked to porosity and defines the maximum water content within the soil. The hydraulic conductivity at saturation is crucial for water flow as it defines the speed of water movement.

The bulk density ρ_b had a significant effect on water content on 4 April and 25 April 2015: the water content changed by 0.16 $\mathrm{m^3\,m^{-3}}$ on 4 April (that is $38\,\%$ of the reference value). This effect may be explained by the fact that when bulk density increases, the pore space is reduced, hence diminishing the water content.

The parameter Kws has a non-monotonic effect on the two considered output variables (figure 3.2). The temperature response drops with very low conductivities, inferior to $3.2 \times 10^{-7}\,\mathrm{m\,s^{-1}}$ (equivalent to $1.2\,\mathrm{mm\,hours^{-1}}$). Such low conductivities are caused by compacted soil (density factor superior to 1.2). Beyond this threshold, the response of the model is moderate: it decreases or increases slowly depending on the model output (increasing for soil temperature and decreasing for water content).

This non-monotonic response shows that the model is nonlinear and complex, and encourages to perform global sensitivity analysis such as the on described in the next section.

Discussion model response to OAT

The OAT and local approach with a graphical analysis of the results is the most intuitive approach for catching a first glimpse of the parameter effects on the model response. In this method, we have chosen to focus on soil temperature and water content at 5 cm because this is where farmers plant seeds of most

crops. Moreover, the effect of parameters for deeper soil layers would only be visible after a consequent amount of simulation time, due to the time needed for diffusion in the soil. The change of state for deep soil layers may take up to 10 days and would consequently increase the time needed to perform this analysis.

For the two days selected for the sensitivity analysis study, the amplitude of model response varied greatly. For soil temperature, the amplitude generated by parameter variation ranged from 3 to 25 $\%$ of the reference temperature. This high amplitude is due to the variation in surface parameters.

For soil water content, the amplitude ranged from 0.3 to 38 $\%$ of the reference water content. The highest amplitude is caused by variation in soil parameters. We might have induced more model variability by choosing wider parameter ranges; however, we preferred to keep them consistent with the associated uncertainty that we had defined.

The next step in sensitivity analysis consists in ranking each parameter according to a sensitivity index. This sensitivity index can be the derivative of the curves in figure 3.1 and 3.2, but it would not make sense with an exponential response such as the one for roughness length z_0 or the non-monotonic response of Kws. We will hence adopt a more global approach and compute sensitivity indexes with two methods: the Morris method (Morris, 1991) and Sobol method (Saltelli, Tarantola, et al., 2004).

1.3 Morris sensitivity indices

Description of the Morris method

The Morris method belongs to the OAT (One-at-A-Time) method of sensitivity analysis: it assesses the influence of each factor independently while all the others are fixed. Nevertheless, through the computation of *elementary effects* over the whole parameter space, the method gives in addition an indicator of the interactions between factors so it is sometimes referred as a semi-global approach. Let us first define an elementary effect. We note the model as a function f, that produces a response $Y \in \mathbb{R}$ from a vector of parameters $x \in \mathbb{R}^n$.

$$Y = f(x_1, \ldots, x_i, \ldots, x_n)$$

An elementary effect d_i for the i^{th} parameter is the difference between f at an initial point x^k and $f(x^k + a_i \cdot e_i)$ for a given step $a_i \in \mathbb{R}$ in the direction e_i, a unit vector of the i^{th} dimension of x. This elementary effect is computed for

several initial points of the parameter space x^k for $k = 1, \ldots, r$ with $r \in \mathbb{N}^*$. The expression is hence written for parameter i:

$$d_i^k = \frac{f(x_1^k, \ldots, x_i^k + a_i, \ldots, x_n^k) - f(x_1^k, \ldots, x_i^k, \ldots, x_n^k)}{a_i} \tag{3.1}$$

The difference is divided by a_i to scale it. The value of a_i is a function of the number of sampling points l from the parameter distribution.

$$a_i = \frac{l}{2 \cdot (l-1)} \tag{3.2}$$

The elementary effects d_i^k are computed for r points and averaged to obtain the mean of elementary effect μ_i for parameter i:

$$\mu_i = \frac{1}{r} \sum_{k=1}^{r} d_i^k \tag{3.3}$$

In other words, k is the number of repetitions used to compute the mean elementary effect: it is the number of initial points in the parameter space. This mean elementary effect is the first sensitivity index given by the Morris method.

Another useful index uses the absolute value of each elementary effect, and is noted μ_i^* (Saltelli, Tarantola, et al., 2004):

$$\mu_i^* = \frac{1}{r} \sum_{k=1}^{r} |d_i^k| \tag{3.4}$$

The μ_i^* sensitivity index avoids that positive and negative elementary effects compensate each other upon computation of the mean.

The Morris' second sensitivity index is a proxy of the interaction of parameter i with all other parameters. This proxy is the standard deviation of elementary effects in the i^{th} direction. The standard deviation gives an idea of how different the elementary effects are at different points of the parameter space: large values of this standard deviation can hint at the interactions this factor may have with others. The standard deviation is noted σ_i and defined as:

$$\sigma_i = \sqrt{\frac{1}{r} \sum_{k=1}^{r} (d_i^k - \mu_i)^2} \tag{3.5}$$

The model SOPHIA is run for 24 hours and a sample of the output is taken

every 2 hours to obtain the time series $\left(Y_{t_j}\right)_{j\in\mathbb{N}}$. The analysis of the first two indices for each time step reveals that, unsurprisingly, the effect of parameters changes over time (figure 3.3). Hence, we developped a third sensitivity index, based on the previous two, that is suited to study the influence of parameters over time. This index was inspired from previous studies (Lamboni, Makowski, et al., 2009; Lamboni, Monod, and Makowski, 2011) where it had been developed for variance-based methods. By analogy, we adaptated it to propose an estimation of the global direct effect over time using the Morris method, if we consider a large number of elementary effects and sufficient number of levels for each parameter.

Hence, with the time series, we compute a global sensitivity index $\mu^*_{\cdot,i}$ for each parameter i as the pondered mean of the direct effects $\mu^*_{t_j,i}$ obtained at each time point t_j with weights corresponding to the total variance of the output $V(Y_{t_j})$ at each time point t_j:

$$\mu^*_{\cdot,i} = \frac{\sum_j V(Y_{t_j}) \cdot \mu^*_{t_j,i}}{\sum_j V(Y_{t_j})} \tag{3.6}$$

As the variance of the output changes across time, it is important to use it as a weight to sum all the parameters.

Application of the Morris method

To run the Morris method on the SOPHIA model, we discretised each parameter into 8 sampling points ($l = 8$) sampled in the range given in table 3.1. $l = 8$ gives a step $a_i = 0.57$, using equation 3.2. We computed 100 elementary effects ($r = 100$) for each parameter.

For the simulations, the model was run with a time step of 900 seconds, and the soil was discretised into ten layers of 10 cm to have an accurate and fast simulation. We studied three outputs of the model: the soil temperature and soil water content at 5 cm and the net radiation. We chose these three outputs because we hold their observed values. These observed values will allow a parameter estimation step for the identified relevant parameters.

5cm depth corresponds to the mid-point of the first layer below soil surface. We run simulations from 18:00 the previous day to midnight the following day for two contrasted days (as in the OAT analysis in the previous section), starting 3 April at 18:00 and 24 April 2015 at 18:00, respectively.

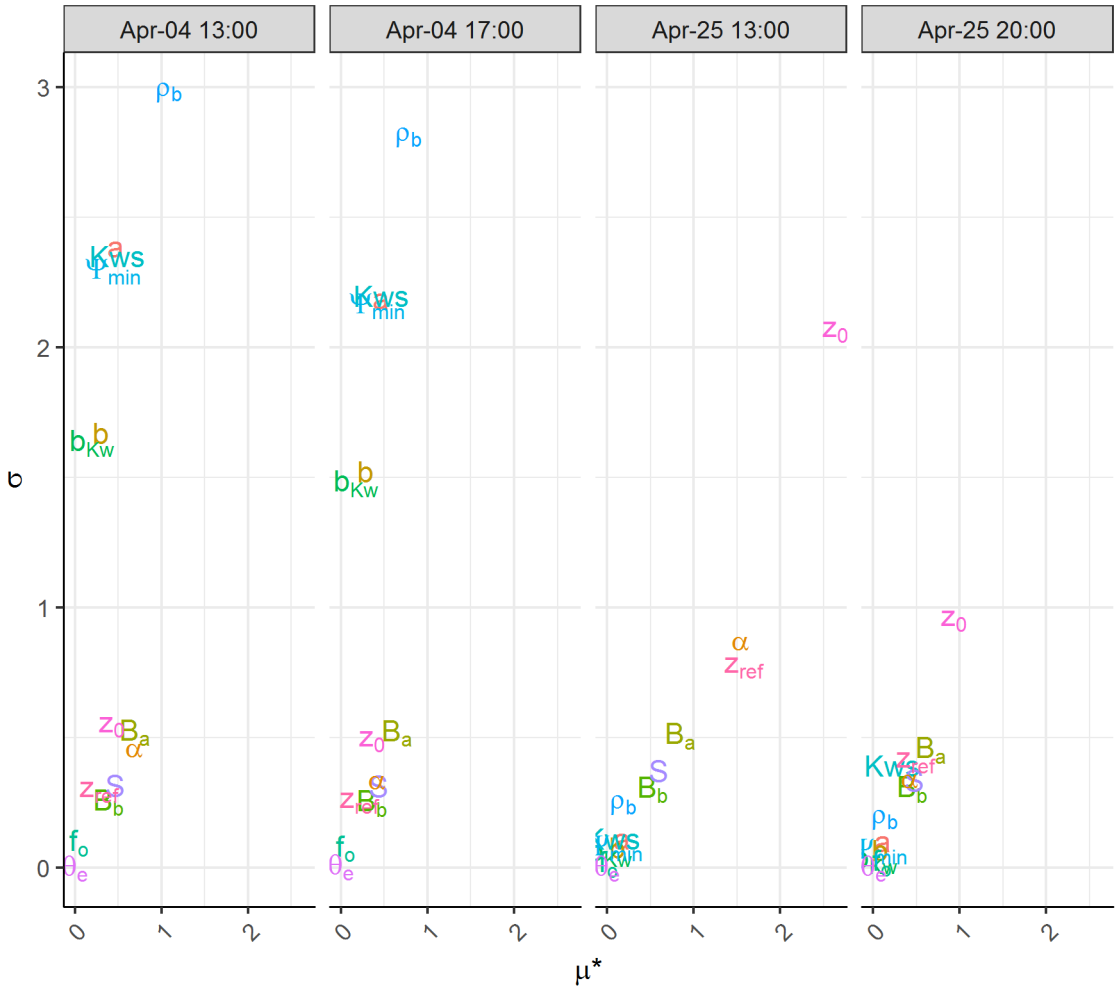

FIGURE 3.3: The 14 parameters plotted according to the two sensitivity indices given by the Morris method, $\mu*$ and σ, computed for soil temperature at 5 cm, at 13:00 and 17:00 for 4 April (a rainy day) and at 13:00 and 20:00 for 25 April 2015 (a sunny day).

Symbols: a and b soil-moisture retention curves; b_{Kw} the exponent in the relationship between hydraulic conductivity and water content; ρ_b the bulk density; f_o the organic matter content; Kws the hydraulic conductivity at saturation;Ψ_{min} the minimum water potential and θ_e the air-entry water content; B_a and B_b Brunt's coefficient; S soil emissivity factor; α albedo, z_0 the roughness length and z_{ref} the reference height

Direct effects and interaction effects across time

In figure 3.3, we plotted the mean elementary effect μ^* of each parameter *vs.* its standard deviation σ. This plot is shown at two different times of the day, 13:00 and 17:00, for 4 April 2015, and 13:00 and 20:00 for 25 April 2015. As explained in section 1.2, these two days were chosen because of their different meteorological conditions.

The parameters that stand out most differ markedly on 4 and 25 April. 4 April was a rainy, cold day with low clouds. In these circumstances, the parameters are dispersed along the σ axis (maximum of 3), but not along the $\mu*$ axis (maximum of 1). σ is an indicator of the interaction effect of the parameter while $\mu*$ is an indicator of their direct effect on soil temperature at 5 cm. The six soil parameters (ρ_b, Kws, a and Ψ_{min}, b and b_{Kw}) show a strong interaction or non-linear effect with a σ above 1.5, while the low μ^* indicates a low direct effect. Between 13:00 and 17:00, the σ of these six parameters diminishes on average by 7.4 $\%$ only, which shows that their interaction effect remains high even at dusk.

On the sunny 25 April at 13:00, other parameters stood out. These parameters z_0, z_{ref} and α are related to surface processes. Their $\mu*$ was superior to 1, which shows that, in clear-sky condition, these parameters have an important direct effect on soil temperature at 5 cm. The roughness length z_0 also has a strong curvature effect with a σ above 2. Solar radiation brings a lot of energy to the soil which increases the exchanges of energy between land and atmosphere, hence giving more importance to parameters related to surface processes.

At night (at 20:00), these three parameters showed a σ decrease of 53 $\%$ on average and exhibited indices below 1. The drop in σ was more important on 25 April than it was on 4 April. Hence, when the solar radiation vary during a clear-sky day, the interaction effect of surface parameters vary as well.

During both days, the position of the remaining five parameters not mentioned above (B_a, B_b, f_o, S and θ_e) did not change whatever the conditions or time of the day (for less than 5 $\%$).

At 20:00, the parameters were bulked around the origin, the direct and interaction effects had diminished compared to 12:00. The position of parameters also differed from 4 April at 17:00, when water content was high. On 25 April, the parameters Kws, ρ_b and a did not stand out as much as on 4 April, mainly due to a dryer soil.

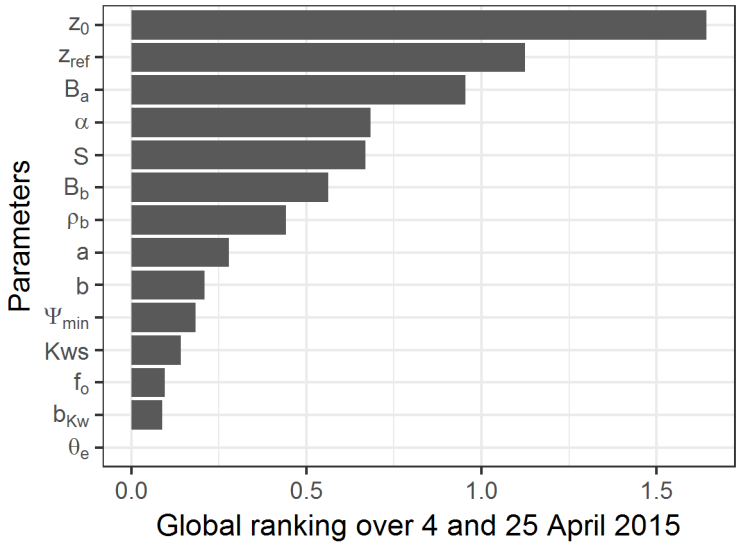

FIGURE 3.4: Global sensitivity index ranking for the 14 parameters on soil temperature at 5 cm. The sensitivity index integrates the μ^* calculated every 15 minutes on 4 and 25 April 2015 starting 18:00 the previous day and ending 00:00 that day. Symbols see figure 3.3, page 109

Global ranking of parameters

It is also possible to rank the parameters according to their global sensitivity index (equation 3.6), by integrating their sensitivity indices over the entire simulation (that is to say, integrating the μ^* calculated for the outputs of the model between 3 April, 18:00 and 4 April, 24:00 and between 24 April, 18:00 and 25 April 24:00)

The most influential parameter in figure 3.4 is the roughness length z_0: its large direct effect $\mu*$ on 25 April at 13:00 probably contributed to this score. This parameter is also reported as an important one by Hou et al. (2015) and Collins and Avissar (1994), with the latter specifying that the roughness length is the most important parameter when the model is used to reproduce a heterogeneous land area. The parameter ranking second, z_{ref}, is also involved in aerodynamic resistance. Aerodynamic resistance is the resistance term for both the sensible heat (H) and latent heat (LE) fluxes. The top position of these two parameters highlights the importance of aerodynamic resistance in the energy balance.

The following parameter is Brunt's coefficient B_a (3rd rank), which belong to the upward long-wave radiation equation. Soil albedo α appears in the fourth position, followed by soil emissivity factor S and the second Brunt's coefficient B_b. Theses parameters (from 3rd to 6th rank) are the parameters

involved in the net radiation term ($Rnet$). $Rnet$ appears, in addition to aerodynamic resistance, as an essential term for soil temperature at 5 cm. Within these parameters, the parameter involved in long-wave downward radiation has the largest influence.

Seventh in rank, the parameter ρ_b indicates the influence of soil water content on thermal conductivity and hence on soil temperature variation.

The parameters that display the least direct effect on soil temperature are the water content at air-entry suction θ_e, the curvature parameter for hydraulic conductivity b_{Kw} and the percentage of organic matter f_o

In table 3.2, we performed a sensitivity analysis on two other outputs of the model, the net radiation and the soil water content at 5 cm. We present the global ranking for these two outputs along the ranking for temperature at 5 cm already shown in figure 3.4.

As expected, the most influential parameters for net radiation are the parameters involved in net radiation: albedo α, followed by the Brunt's parameters B_a and S and Brunt's B_b. The soil albedo α ranks first because the change in albedo greatly affects the amount of solar radiation received by the system. Solar radiation is the main input of energy for the soil system. What is more interesting is the presence of roughness length z_0 at the 5th rank, following the expected net radiation parameters. It shows the importance of this parameter, even if it is not involved directly in the equation. This parameter also influences net radiation because it impacts the energy exchange from which the soil temperature is computed; besides, soil temperature is involved in net radiation.

The ranking for volumetric water content shows the top soil parameters. The most influential parameter is the bulk density ρ_b followed by hydraulic conductivity at saturation Kws and moisture-tension parameter a on the 3rd rank. Interestingly, we find again z_0 and z_{ref} on the 4th and 5th rank. These parameter, involved in the latent heat flux (evaporation) via the aerodynamic resistance, are also important for the movement of water near the surface.

At the bottom of the ranking for net radiation and water content, there are the same parameters: percentage of organic matter and water content at air-entry. The organic matter alone might not have a great influence, since in the model it is independent from bulk density. In fact, organic matter has an impact on bulk density as the density of organic matter is very low. In a model where organic matter and bulk density would be linked, certainly the fraction of organic matter would show up with more influence.

TABLE 3.2: Parameter ranking according to Morris global sensitivity index for three model outputs: net radiation, volumetric water content and soil temperature at 5 cm. Symbols see figure 3.3, page 109

Model output					
Net radiation		Vol. water content at 5 cm		Soil temperature at 5 cm	
Rank	Parameter	Rank	Parameter	Rank	Parameter
1	α	1	ρ_b	1	z_0
2	B_a	2	Kws	2	z_{ref}
3	S	3	a	3	B_a
4	B_b	4	z_0	4	α
5	z_0	5	z_{ref}	5	S
6	a	6	b_{Kw}	6	B_b
7	z_{ref}	7	b	7	ρ_b
8	b	8	Ψ_{min}	8	a
9	Ψ_{min}	9	B_a	9	b
10	ρ_b	10	S	10	Ψ_{min}
11	Kws	11	α	11	Kws
12	b_{Kw}	12	B_b	12	f_o
13	f_o	13	f_o	13	b_{Kw}
14	θ_e	14	θ_e	14	θ_e

Conclusion on Morris analysis Surface parameters powerfully affect soil temperature near the surface. The effect of these parameters change with both time and weather conditions. If we integrate the direct effect across two contrasted days, the roughness length has the largest effect, followed by the parameters for net radiation. For the sensitivity analysis of net radiation, it is the albedo that produces the largest effect. For soil water content, the bulk density ρ_b and hydraulic conductivity at saturation Kws show the most significant effect. The parameter with the lowest effect is the water content at air-entry suction.

The Morris sensitivity analysis is computationally cheap and allows a fast and efficient exploration of the parameter space, hence its denomination as a *screening* method since it may help quickly extract the most influential parameters. A more thorough analysis is then recommended to estimate more precisely their effects and interactions, since the values of direct and interaction effects may only be considered as proxys for the values obtained with variance-based method (Saltelli, Tarantola, et al., 2004). In the next chapter, we shall present an analysis of the model's behaviour with the Sobol method.

1.4 Sobol sensitivity indices

Description of the Sobol method

The Sobol method is a variance-based, global sensitivity analysis technique. Its principle is based on decomposing the uncertainty of the model response Y with respect to the relative contribution of each parameter. It relies on the assumption that the uncertainty of a given variable can be adequately characterized by its variance (other alternatives exist, such as entropy). The goal is thus to decompose the variance of Y into the respective contributions of each parameter and all their possible combinations.

The main drawback of this method is that it is computationally expensive. This explains why our strategy was to begin with running a faster and simpler approach, like Morris, as a first screening for relevant parameters, only to follow with selecting a few critical parameters to get precise sensitivity indices with the Sobol method.

Let us reintroduce the notations adopted in section 1.3. We consider the model f:

$$Y = f(X_1, \ldots, X_i, \ldots, X_n)$$

where $Y \in \mathbb{R}$ is the model response and $X_1, \ldots, X_n$ are model parameters. Parameters are supposed to be independent random variables, each with a given probability distribution.

To quantify the effect of the parameter X_i on the variance $V(Y)$, let us first examine how the variance of Y would be modified, were it possible to eliminate the uncertainty associated to this parameter, i.e. if X_i was set to a particular value, noted x_i^*. We therefore consider the difference between the variance of Y and the variance of Y when $X_i = x_i^*$:

$$V(Y) - V(Y|X_i = x_i^*)$$

The main issue is that the value of x_i^* is unknown to us, and could potentially adopt any value on the distribution of X_i. Instead of arbitrarily selecting a value for x_i^*, we take the mathematical expectation of X_i, E_{X_i} and thus rather consider the quantity:

$$V(Y) - E_{x_i}(V_{X \sim i}(Y|X_i))$$

The symbol $X \sim i$ indicates that the variance of Y is computed over all the parameters except the i^{th}, X_i. When this difference is large, it means that

fixing X_i induces a large modification of the variance of Y. Hence we may conclude that X_i has a strong effect. This effect is a sensitivity index S_i that might be appropriately rewritten, using the law of total variance:

$$S_i = \frac{V_{X_i}(E_{X\sim i}(Y|X_i))}{V(Y)} \tag{3.7}$$

S_i is called the first-order sensitivity index for parameter X_i. It is normalised by the variance of Y to obtain an index that ranges between 0 and 1. The first order sensitivity index of X_i represents its contribution in the decomposition of the variance of Y.

The sum of first-order sensitivity indices of all parameters is inferior or equal to 1. When it is strictly inferior to 1, one can raise the question of the source of the remaining part of variance. This is related to another question: what happens when two parameters X_i and X_j are changing together at the same time? An interaction effect of these parameters might exist, i.e. an effect on the variance might only be visible when these two parameters are fixed. Hence the answer: the remaining part of the variance comes from interactions between parameters. To quantify the interaction, we introduce the second-order sensitivity index, S_{ij} that indicates the effect of parameters X_i and X_j :

$$S_{ij} = \frac{V(E(Y|X_i, X_j))}{V(Y)} - S_i - S_j \tag{3.8}$$

Similarly, third-order sensitivity index can be written

$$S_{ijk} = \frac{V(E(Y|X_i, X_j, X_k))}{V(Y)} - S_i - S_j - S_k - S_{ij} - S_{jk} - S_{ik}$$

And so on until all the parameters n are taken into account. Sobol's decomposition states that, for a response Y of a model with n parameters:

$$\sum_{i=1}^{n} S_i + \sum_{i=1}^{n}\sum_{j>i} S_{ij} + \sum_{i=1}\sum_{j>i}\sum_{k>j} S_{ijk} + \ldots + S_{ijk\ldots n} = 1 \tag{3.9}$$

Computing all these sensitivity indices would be time-consuming: it requires to compute $2^n - 1$ sensibility indices. To reduce the computation time, Homma and Saltelli (1996) introduced an additional index; the total-order sensitivity index, ST_i, which is the sum of all its sensitivity indexes at different orders, that is, the sum of all contributions that include X_i (i.e. the first-order index and all its interactions). We find the ST_i by removing the contribution

of all the other parameters:

$$ST_i = 1 - \frac{V_{X\sim i}(E_{X_i}(Y|X_{\sim i}))}{V(Y)} \tag{3.10}$$

The term on the right-hand side represents the contribution of all the parameters other than i ($X \sim i$) and their interactions to the variance of Y. By removing this from 1, we get the contribution of parameter X_i and all its interaction to the variance of Y.

Once again using the law of total variance, we get the following expression of the total-order sensitivity index:

$$ST_i = \frac{E(V(Y|X_{\sim i}))}{V(Y)} \tag{3.11}$$

To make things clear on an example, if we have a model with three parameters ($n = 3$), then:

$$ST_1 = S_1 + S_{12} + S_{13} + S_{123}$$

We can see that S_1 will always be inferior or equal to ST_1, and this is a general property, that is to say

$$ST_i \geq S_i$$

To estimate the S_i and ST_i indices, we use a Monte Carlo sampling method based on two identical matrices containing N sampled value of the parameters. The first matrix is called the sampling matrix while the other is the resampling matrix. The two matrices are combined to obtain scenarios that allow computing the first and total indices.

Application of the Sobol method

To apply Sobol method to the model SOPHIA, we selected the same parameters and ranges as in the Morris method except for parameters a and S as mentioned in section 1.1 page 95. The new sensitivity analysis was an opportunity to change parameter ranges which were arguable. The range for a was modified to contain the value of a found with measurements from water-retention curves (see section 4.4) while the range for S was extended.

We used a matrix of $N = 15000$ sampled values to obtain good estimates of the sensitivity indices. We started the simulation on 24 April at 18:00 and stopped the simulation at 00:00, 05:00, 08:00 and 11:00 on 25 April to compute the sensitivity indices at these four times. The model was run with a time

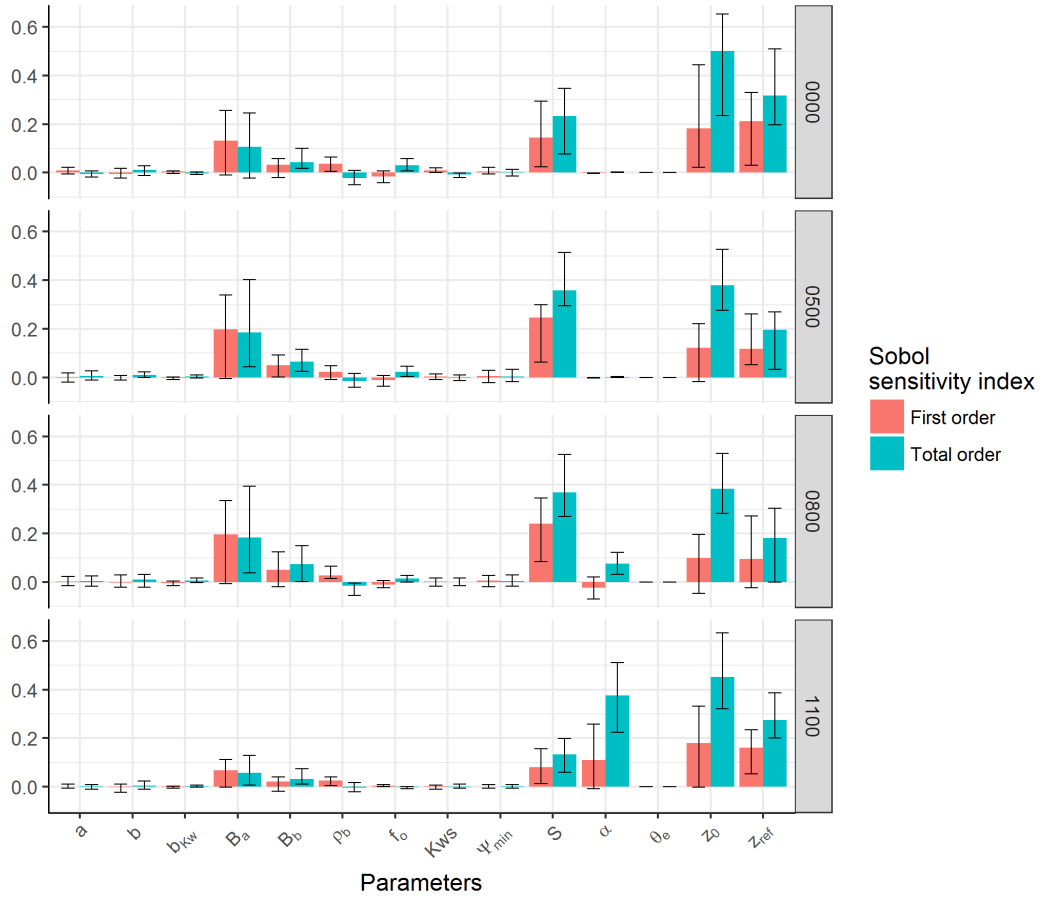

FIGURE 3.5: Sobol's first- and total-order sensitivity indices for 14 parameters and four time points for the temperature at 5 cm on 25 April 2015. The simulation was started on 24 April 2015 at 18:00, with a time step of 900 s and a soil discretization of 10 layers of 10 cm. The error bars denote 95 % confidence intervals.

step of 900 seconds and a soil discretised into 10 layers of 10 cm. The initial conditions for the whole soil profile were set to the observed soil temperature and moisture at 5 cm at the start of the simulation.

Figure 3.5 presents the results for the first and total order sensitivity indices, the error bars representing 95 % confidence interval. These errors are relatively large for most parameters, which shows the challenge of obtaining precise measures of the sensitivity indices. The negative values are artefacts due to these remaining uncertainties in the estimation. To narrow down the confidence interval, we should increase the number of the sample matrix. However, to achieve the presented results, we ran the model 75 000 times for each time of the day, which already took so long that we may consider the experiment as not reproducible with our computer equipment (25 days or 600 hours of continuous running on a HP EliteBook 840 G1 with 8 GB of RAM).

Further studies with Sobol method would require access to more computing power.

What stands out from the sensitivity analysis are the six parameters already identified with the Morris method: B_a,B_b,S,α,z_0 and z_{ref}. The effect of parameters involved in long-wave radiation (B_aB_b and S) is great at night and early morning (00:00, 05:00 and 08:00) whereas α, the albedo linked to short-wave radiation has a larger effect at 11:00. The effect of z_0 and z_{ref}, both involved in the aerodynamic resistance, are similar whatever the time.

For B_a, the first and total-order indices are nearly equal, suggesting that this parameter does not have many interactions with other parameters. In contrast, S, α, z_0 and z_{ref} have larger total-order indices compared to their first indices, which indicates that these parameters have interactions with other parameters.

1.5 Discussion on sensitivity analysis

We started the exploration of our model sensitivity with a graphical and local OAT (One-at-a-time) analysis and subsequently performed global analysis using the (semi-global) Morris and Sobol methods.

The simulations where the same for each analysis, except a slight change in range for two parameters in the Sobol analysis, for which there was uncertainty. The simulation settings were all the same. Keeping the same settings for all the simulations is important as our preliminary work (not shown here) has revealed that for some combinations of parameter values and/or meteorological conditions, our initial explicit discretisation scheme showed some limitations so we had to refine it: we chose a discretisation (10 layers of 10 cm and 900 seconds, i.e. 15 minutes, time step) that was stable within our whole parameter variation space.

Also from a technical point of view, our implementation of the Morris method allowed to compute sensitivity indices at every time step of the model and to integrate them in a global ranking, which was not possible with the implementation of the Sobol method that we used. For the Sobol sensitivity analysis, due to high computation time (25 days), we focused on one day (25 April 2015) and one output, namely the soil temperature at 5 cm. We computed the sensitivity analysis only for four time points (00:00, 05:00, 08:00 and 11:00).

The Morris method is classically used for crop models (Wallach et al., 2014), but the Sobol method is known to be more precise and to provide more complete information. In our case, its high computation time did unfortunately not make it a method of choice.

Globally, the three methods all indicate that the surface parameters have more influence on soil temperature at 5 cm than the soil parameters. For Morris and Sobol analysis, the parameters with significant direct and first-order effects are the same: the roughness length z_0, the reference height z_{ref}, the soil emissivity factor S, the albedo α and the Brunt's parameters B_a and B_b.

In the discretisation chosen, the soil temperature at 5 cm is the temperature of the mid-layer node of the first layer (see figure 2.2 page 54), i.e a soil layer that is in direct contact with the surface. As we are very close to the surface, it is expected that the most influential parameters for soil temperature be parameters linked to surface exchanges.

Another important result is the change of effect of certain significant parameters according to the time and type (sunny/rainy) of day. The parameters linked to long-wave radiation (S, B_a and B_b) display more effect at night or early in the day than the albedo α, a parameter only linked to short-wave radiation coming from the sun.

Depending on the type of day taken into account, different parameters might show significant effect. For example, for the computation of the global ranking with the Morris method, we used both 4 and 25 April 2015 altogether. As a result, the bulk density ρ_b appears as a significant parameters in the Morris analysis but is shown as insignificant in the Sobol analysis. It is therefore important to compute the indices at different times of the day and for different types of days if we are to identify the parameters that, in any weather conditions, have a significant effect on the output. It is important because the identification of significant parameters drives the selection of candidates for calibration.

The Sobol sensitivity analysis based on variance could give precise information on single and total effect terms. However, due to its high computational cost, performing the analysis over several days was not feasible in our context. A perspective would be either to use a high-performance computing platform to run the model in parallel on many cores or to recode the model in a compiled, faster program: thus we could run a Sobol analysis for different days of the year.

As is always the case, performing a model's sensitivity analysis required

making both choices and assumptions. We chose here to restrict our study to our three main variables of interest. Further analyses could involve other soil textures, other situations (summer day or winter day) and other model outputs such as the temperature and water content at greater depth.

In particular, the sensitivity analysis presented here was performed for only one texture class, namely a clay loam soil. This restriction allowed for a consistent range of soil parameters even when they were drawn independently from one another. However, their relatively narrow ranges might also explain why the soil parameter set does not have a significant effect on soil temperature at 5 cm. In order to deal with wider uncertainty ranges for these parameters without obtaining an unrealistic set of parameters (which can cause instabilities in the simulations), one should consider correlations between the different parameters, as in the approach developed in Sainte-Marie, Viaud, and Cournède (2017) and Chastaing, Gamboa, and Prieur (2015). An additional study would therefore be needed in prerequisite, in order to be able to precisely define these correlations.

It is also important to note that we have not considered the uncertainty brought by environmental inputs of the model, that is to say, the measurement of radiation, temperature and rain. A study by Banimahd and Zand-Parsa (2013) shows that air temperature, radiation and initial conditions are nevertheless the most significant inputs with respect to the error of the model. An error of 20 % on air temperature induces a variation of 20 % of the soil temperature at $5\,\mathrm{cm}$. Similarly, an error of $20\,\%$ in solar radiation induces a variation of 24.5 % of soil temperature near the surface. In those cases the variation in the output is therefore of the same magnitude order than that of the input. In spring, the 20 % change in air temperature or radiation triggers an increase of 96.8 and 41 % of the error on soil temperature, respectively. Hence, the variation of the output presents a two to four-fold higher magnitude than the input.

Our weather station was new and regularly checked for issues, therefore we assumed that the associated measurement uncertainties were reduced, although residual uncertainties inevitably still exist. We keep in mind that this could be a potential source of error, especially when the model is deployed at a larger scale. In section 3.2, we have for instance shown that the model's error for temperature at $5\,\mathrm{cm}$ changes according to the season: it is small in spring and autumn, but high in summer.

We have identified the relevant parameters that influence the three main model outputs for which we hold measurement. The next step is to explore

the effect of the calibration of these parameters on simulation accuracy.

2 Parameter estimation

Parameters to estimate To choose the parameters to estimate, we relied on the sensitivity analysis to find parameters that have a significant effect on soil temperature at 5 cm, water content or net radiation.

For soil temperature at 5 cm, six parameters were candidates for estimation (global ranking over 0.5): z_0, z_{ref}, B_a, α, S and B_b (figure 3.4, page 111).

The step of parameter estimation of our model must be considered with care. Indeed, the set of parameters to estimate should be kept as limited as possible in order to (i) avoid identifiability problems, (ii) avoid over-parameterisation problems and allow its use by other researchers or farmers at other sites without requiring heavy experimental work for its parameterisation, (iii) avoid computational problems linked to the optimisation procedure (presence of multiple local optima, prohibitive computational cost of the procedure...). The larger the number of parameters to be estimated, the higher the probability that these problems arise.

There are four options when parameters are identified as target for estimation after a ranking by sensitivity analysis. The first one consists in searching the literature for a consensus. The second approach takes a direct measurement of the parameter, or measurements that help to estimate easily the value of the parameter, as we have seen with water-retention curve in section 4.4. The third option entails estimating the parameter with an optimisation algorithm to fit simulated and observed values. Finally, the fourth option adds new expressions to the model in order to better represent the phenomenon, but such an approach is quasi invariably at the cost of adding new parameters.

The parameter with the greatest effect on soil temperature near the surface, z_0, cannot be estimated with the present dataset. Martano (2000) describe a procedure with several sonic anemometers placed at different heights to measure precisely the wind profile. In the literature, the consensual values lie between 0.01 (Acs, Mihailovic, and Rajkovic, 1991) and 0.02 for a bare soil surface (Campbell and Norman, 1998). There is a formula to compute z_0 that is mentioned in Campbell and Norman (1998) but its application with only two cup anemometers yielded a very noisy estimation and it was deemed unreliable.

The parameter z_{ref} is a fixed parameter and hence will not be estimated. It is fixed to $2\,\mathrm{m}$, the height where the wind speed and temperature are measured, on our weather station.

As z_0 and z_{ref} will not be estimated, that leaves us with the net radiation parameters B_a, α, S and B_b. The data coming from the net radiometer installed on the weather station will support their estimation. This dataset is particularly interesting as it allows to explore the variability of short- and long-wave radiation and estimate their parameters. Moreover, we can estimate these parameters independently from the others by uncoupling the equation associated with each type of radiation from the rest of the model.

Evaluation To assess the quality of different parameter estimation or the introduction of new equations in the model, we computed some classical statistical criteria: the bias, RMSE (root mean square error), RRMSE (relative root mean square error), MAE (mean absolute error), RMAE (relative mean absolute error) and AIC.

$$Bias = \bar{y}_{obs} - \bar{y}_{sim} \tag{3.12}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(y_{obs_i} - y_{sim_i}\right)^2}{N}} \tag{3.13}$$

$$RRMSE = \frac{RMSE}{\bar{y}_{obs}} \tag{3.14}$$

where, for a given observable y, $\bar{y}_{obs}$ denotes the mean of the observed values $y_{obs} \in \mathbb{R}^N$ and $y_{sim} \in \mathbb{R}^N$ denotes the corresponding simulated values. N is the number of values used to compute these indicators.

RMSE is know to give over-weighting to large differences (Brun et al., 2006). An alternative is to use the mean absolute error (MAE) instead of root mean square error (RMSE):

$$MAE = \frac{\sum_{i=1}^{N} |y_{obs_i} - y_{sim_i}|}{N} \tag{3.15}$$

and the relative mean absolute error RMAE:

$$RMAE = \frac{MAE}{|\bar{y}_{obs}|} \tag{3.16}$$

where $|\bar{y}_{obs}|$ is the mean of absolute values of y_{obs}. We use the MAE and RMAE to evaluate the fit on net radiation because large differences for value above $450\,\mathrm{W\,m^{-2}}$ artificially increased the RMSE.

AIC To quantitatively compare the different models resulting from the estimation of parameter or from new equations, we computed the AIC (Akaike's Information Criterion) for each model. The AIC allows for comparison between nested and non-nested models. Originated from the field of Information Theory, it provides a way to assess the compromise between the fitting accuracy (reduction of bias) and the increase in the number of parameters that could come along (increase of estimators variance). Within the large family of criteria for penalized model selection, AIC is known to be preferable for selecting a model intended for prediction (Yang, 2005). In this work we will use the AIC formula adapted for least squares estimation with normally distributed errors (Burnham and Anderson, 2003):

$$AIC = N \log\left(\frac{\sum_{i=1}^{N}(y_{obs_i} - y_{sim_i})^2}{N}\right) + 2p \tag{3.17}$$

where N is the number of data points used to compute the criterion and p is the number of estimated parameters. The model with the smallest AIC is preferred.

2.1 Estimation of Brunt's coefficients

The first parameters to estimate are the Brunt's coefficients. The long-wave radiation emitted downward from atmosphere to the soil is expressed with the Stefan-Boltzmann law:

$$Ld = \epsilon_a \sigma T_{air}^4$$

The atmospheric emissivity ϵ_a is computed with the Brunt's formula. The Brunt's formula is one among many long-wave radiation downward algorithms that predict long-wave radiation under clear sky (long-wave radiation is also termed clear sky irradiance). The Brunt's formula, already mentioned in the model description (equation 2.16) is a relationship between atmospheric emissivity (ϵ_a) and vapour density ρ_{va} [$\mathrm{g\,m^{-3}}$]:

$$\epsilon_a = B_a + B_b \cdot \sqrt{1.41 \cdot \rho_{va}}$$

This linear relationship is valid only under clear sky (G. Flerchinger, Xaio, et al., 2009). For this reason, we selected only clear-sky days in 2015 to estimate B_a and B_b. Clear sky days are defined as days where the maximum global radiation is superior to 800 $\mathrm{W\,m^{-2}}$. Then, we used an ordinary least square

procedure to estimate B_a and B_b. Estimation amounts to an optimisation problem and can be solved as (Wallach et al., 2014):

$$\theta_{\text{optimized}} = \underset{\theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \big(y_i - f(x_i, \theta)\big)^2 \right\} \tag{3.18}$$

where y_i is the observation for situation i, $f(x_i, \theta)$ is the model output for situation i and $\theta_{\text{optimized}}$ is the optimized value of parameter θ. In our case, situations i correspond to every hour of 2015. The numerical resolution was performed using the `nls` function in the `stats` package in R (R Core Team, 2015), that embeds the Gauss-Newton algorithm.

Figure 3.6 allows comparison of the long-wave downward radiation simulated both with parameters B_a and B_b from Müller (1999) (0.605 and 0.039, respectively) and with the optimised values using our clear-sky days datasets of 2015. The large positive bias with Müller parameters indicates a general under-prediction of long-wave radiation downward. With the estimated parameter values, the bias is close to 0, which means that there is neither under- nor over-prediction. The RMSE goes from 29.8 $\mathrm{W\,m^{-2}}$ down to 19.2 $\mathrm{W\,m^{-2}}$. The estimated values were B_a =0.58 and B_b =0.066.

The estimated values are consistent with those reported in Iziomon, Mayer, and Matzarakis (2003) ($B_a = 0.60$ and $B_b = 0.064$) in their review of different clear sky irradiance algorithms that were tested at Bremgarten, Germany, 40 km away from our current study site. The review also mentions that the performance of the Brunt's formula is similar to other algorithms under clear-sky conditions. According to the same authors, B_b is the parameter that changes the most with location. In our study case, we have also observed that the largest change between the default and the estimated value was on B_b (default 0.039, estimated 0.066) compared to B_a (default 0.605, estimated 0.58).

The simulations obtained with these new parameter values, estimated in clear sky conditions in 2015, also need to be confronted to data in cloudy conditions.

In the panel A in figure 3.7, we have plotted the simulated *vs.* observed downward long-wave radiation with all 2015 data and we have a RMSE of 37 $\mathrm{W\,m^{-2}}$. The clear-sky algorithm is thus not suited to all-sky conditions. This leads us to introduce a cloudiness correction factor to account for cloudy days.

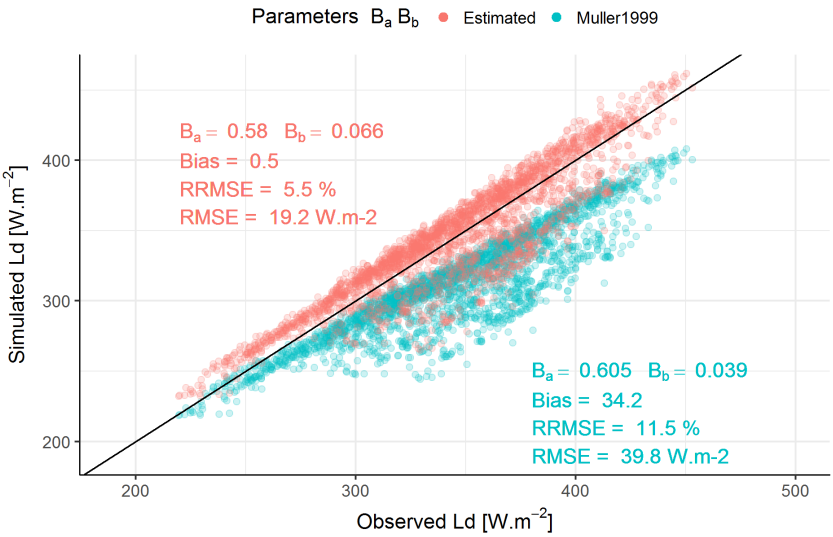

FIGURE 3.6: Simulated *vs* observed long-wave downward radiation (Ld) with the default B_a and B_b values (in blue) and with the values estimated on clear sky days datasets (in 2015, maximum global radiation superior to 800 $\mathrm{W\,m^{-2}}$) in red. In the legend, Muller1999 corresponds to the Ld computed with the default parameter values from Müller (1999). The plot shows the data of hourly long-wave data in clear skies in 2015. The black line denotes the 1:1 line. RRMSE: relative root mean square error, RMSE: root mean square error. Number of points: n=2376

2.2 Cloudiness correction for atmospheric emissivity

The literature abounds with algorithms for representing clear sky downward long-wave radiation and their cloudiness correction factor. For instance, G. Flerchinger, Xaio, et al. (2009) compares 13 expressions for clear-skies and 10 expressions to adapt the clear-sky algorithm to cloudy situations. For clear-sky, among the 13 possibilities, we decided to keep the expression presented in chapter 2, the Brunt's formula, because the authors of Iziomon, Mayer, and Matzarakis (2003) showed that Brunt's formula performs well (11 $\%$ relative error) for their site in Bremgarten, Germany, located 40 km from our site and with a similar elevation.

Brunt's formula only applies to clear skies. Therefore, to account for overcast days, we need to introduce a cloud correction factor. We describe hereafter the development of a cloud-correction factor that follows the description in M. H. Unsworth and JL Monteith (1975). This cloud correction factor is widely used, for example in a similar soil physics model described by Bittelli, Campbell, and Tomei (2015).

To implement the cloudiness correction, we consider data at the daily level. We first computed the extraterrestrial radiation for our location with the R package `sirad` (Bojanowski, 2016). Extraterrestrial radiation depends on the day of the year and the latitude. We then estimated an atmospheric transmission τ as the ratio between global radiation (total short-wave radiation St [$\mathrm{W\,m^{-2}}$], measured, and extraterrestrial radiation Ext [$\mathrm{W\,m^{-2}}$]:

$$\tau = \frac{St}{Ext} \tag{3.19}$$

Then the cloudiness factor cc is given by the empirical formula (M. H. Unsworth and JL Monteith, 1975):

$$cc = \begin{cases} 0 & \text{if } \; 2.33 - 3.33\tau \leq 0 \\ 1 & \text{if } \; 2.33 - 3.33\tau \geq 1 \\ 2.33 - 3.33\tau & \text{otherwise} \end{cases} \tag{3.20}$$

The cloudiness factor being between 0 and 1, all the values below 0 or above 1 are set to 0 and 1. The value of atmospheric emissivity ϵ_a is then corrected with the cloudiness factor cc (M. H. Unsworth and JL Monteith, 1975):

$$\epsilon_{acc} = \epsilon_a(1 - Icc) + Icc \tag{3.21}$$

where ϵ_a is the clear-sky emissivity (equation 2.16, section 2.1) and $I = 0.84$ a parameter given in M. H. Unsworth and JL Monteith (1975). The computation of cloudiness correction equation that gives ϵ_{acc} can take many different forms where parameter I varies (G. Flerchinger, Xaio, et al., 2009).

We evaluate the corrected atmospheric emissivity ϵ_{acc} in figure 3.7. We present the comparison between three parametrisations of downward long-wave radiation. The clear sky parametrisation ($B_a = 0.066$ and $B_b = 0.56$), the clear sky parametrisation with correction for cloudiness ($B_a = 0.066$ and $B_b = 0.56$ plus equation 3.21) and the all sky parametrisation.

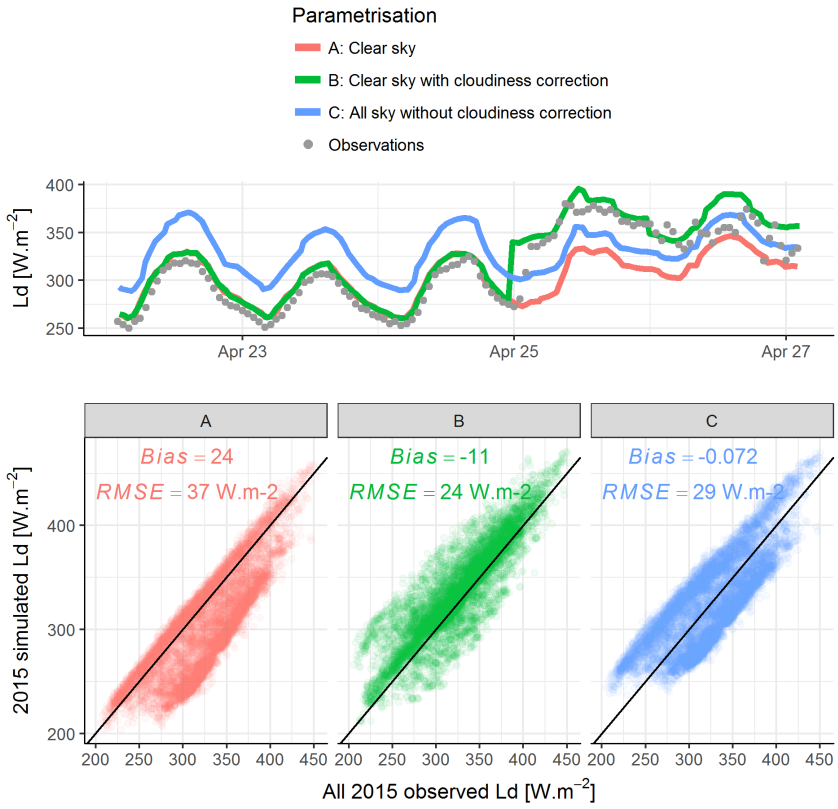

FIGURE 3.7: Comparison of different parametrisation for long-wave downward radiation simulation. A: 'clear sky' means that we used only clear sky days (maximum daily global radiation superior to 800 $\mathrm{W\,m^{-2}}$) from 2015 to estimate $B_a = 0.58$ and $B_b = 0.066$; B: same B_a and B_b values but integrating the cloudiness correction from equation 3.21; C: 'all sky' means that we used all the data in 2015 (cloudy and clear sky days) for estimating $B_a = 0.77$ and $B_b = 0.029$. TOP: simulated and observed value showed as a time-series for 22 to 26 April 2015. These days were selected for this illustration because 22, 23 and 24 April are clear sky days while 25 and 26 April are cloudy days. BOTTOM: simulated downward long-wave radiation (Ld) *vs* observed for the entire year 2015. The black line denotes the perfect agreement line. Number of data points: n=8741 (hourly observation of downward long-wave)

The third parametrisation, 'all sky' is an attempt to overcome the computation of the cloudiness factor by estimating parameters B_a and B_b directly using all the downward long-wave data in 2015, that is to say both cloudy and clear skies. This estimation results in $B_a = 0.77$ and $B_b = 0.029$. The objective of this third parametrisation was to test whether the computation of cloudiness factor could be avoided, as it requires some additional steps in the SOPHIA model and introduces an additional parameter.

We show the comparison as a time series for the period from 22 April to 26 April 2015, and as a scatter plot for all the hourly data in 2015. The period 22-26 April 2015 was chosen for this illustration because the first three days were clear-skied while the remaining two were cloudy. It made the effect of the cloudiness factor better discernible. When the downward long-wave radiations are simulated with the clear sky parametrisation, they fit well with the data on clear sky days and reproduce faithfully the diurnal variation of downward long-wave radiation. This is shown on the first scatter plot (A, figure 3.7) where an elongated, dense zone of data points is close to the perfect agreement line. This dense zone of points is the same as the points in figure 3.6 and designates the clear sky hours of 2015 (clear sky hours belongs to the days where the maximum global radiation St is superior to $800\,\mathrm{W\,m^{-2}}$).

In the same panel A, one can also identify a second zone of dense data points which are underestimated, below the 1:1 line. These points are not visible in the clear sky days data represented in figure 3.6 which leads us to conclude that these points are the cloudy days. When we introduce the cloudiness correction in panel B, the cloudy days in panel A are repositioned along the 1:1 line. In panel B, there is only one dense zone of points and is located around the 1:1 line. There is no distinction any more between the clear and cloudy days, suggesting that the correction factor helps to better simulate downward long-wave during cloudy days.

Indeed, the introduction of the cloudiness correction factor improves the bias and the RMSE: the latter decreases from 37 $\mathrm{W\,m^{-2}}$ to 24 $\mathrm{W\,m^{-2}}$. The bias decreases in absolute value (of 54 $\%$) and switches from positive (under-estimation) to negative (over-estimation). This over-estimation is visible on the time series presentation (figure 3.7, top), the curve B (green) slightly over-estimates the observed downward long-wave of 26 April. However, it does predict very well the data on 25 April.

Our attempt to parametrise the Brunt's parameters with all the downward long-wave data of 2015 resulted in parameter values $B_a = 0.77$ and $B_b = 0.029$ (panel C, figure 3.7, page 128). The bias is close to 0 and RMSE at 29 $\mathrm{W\,m^{-2}}$. This is better than than the clear sky parametrisation (panel A) but, one can see on panel C that there are still two distinctive dense zones. The parametrisation reduces the error for cloudy days but creates another for clear-sky days. Hence, the better performance of this parametrisation is obtained by making smaller errors for the two types of sky rather than only large error for cloudy skies as in panel A. This error is apparent in the times series: the C curve (in blue) over-estimates the observations during clear-sky days and under-estimates

the observations on 25 April, a cloudy day.

Moreover, the values $B_a = 0.77$ and $B_b = 0.029$, obtained by calibrating with the totality of the 2015 data sit outside the range found in the literature. This confirms that it doesn't make sense to calibrate a clear-sky algorithm such as Brunt's with downward long-wave data from clear and cloudy days. This third parametrisation is therefore not considered as reliable.

The estimation of parameters can be evaluated with the AIC. The penalization term includes two (B_a and B_b) or three parameters (with the addition of the cloud correction factor parameter I, equation 3.21). In table 3.3, the AIC is computed for the initial parameters values from Müller (1999), and parametrisation A, B and C as mentioned in figure 3.7. The parametrisation B, with estimated parameters on clear sky days and the cloud correction factor yields the lowest AIC. Despite its three estimated parameters, it is therefore the parametrisation of choice for downward long-wave radiation.

To conclude on this downward long-wave radiation term, the parametrisation B on clear sky days with an addition of a cloudiness correction factor brought about the most significant reduction of the error and the lowest AIC in the Ld term of the energy balance. Promising perspectives have been highlighted in our study, such as testing other expressions of the cloudiness factor and taking into account the suggestions in the G. Flerchinger, Xaio, et al. (2009) review.

2.3 Estimation of soil emissivity

In the previous section, we estimated the significant parameters B_a and B_b. The S parameter, listed as the fifth most influential in the global ranking from the Morris analysis, might also be estimated using the net radiometer and the inversion of the model.

However, in the literature, the reference value of 0.9 for S is, to our knowledge, not questioned. Its expression has never changed since the early works on energy balance models, such as Van Bavel and DI Hillel (1976), even until the recent HYDRUS-1D (J Simunek et al., 2013). Therefore, instead of trying to estimate S, we concentrated our effort on ϵ_s, the soil emissivity. We recall here the relationship between S the soil emissivity factor, ϵ_s the soil emissivity

TABLE 3.3: AIC (Akaike's Information Criterion) computed for different parametrisation of atmospheric emissivity (ϵ_a). The AIC was computed with hourly downward long-wave data recorded by the weather station in Stein, Switzerland in 2015

Parametrisation of ϵ_a	B_a	B_b	Estimated parameters	AIC	ΔAIC
Initial values	0.6	0.039	0	70538.65	0.0
A: calibrated with clear sky data	0.58	0.066	2	63485.5	-7053.2
B: calibrated with clear sky data + cloud cover correction	0.58	0.066	3	55868.7	-14669.9
C: calibrated with all-sky data	0.77	0.029	2	60600.2	-9938.4

and Lu, the upward long-wave radiation:

$$Lu = \epsilon_s \sigma T_{\text{surface}}^4$$
$$\epsilon_s = S + \frac{\theta_1 \cdot \theta_{min}}{P_{tot}}$$

where T_{surface} is the soil surface temperature, θ_1 is the volumetric water content of the first layer, θ_{min} the minimum volumetric water content and P_{tot} the total porosity. The main term of this equation is the T_{surface}^4, and hence the prediction of Lu is very sensitive to the prediction of the surface temperature, a result of the energy balance equation.

Without estimation ($S = 0.9$) and using the equations above, the bias for upward long-wave radiation is 32.5, showing a large under-prediction of the model, and a relative error of 11.6 % (table 3.4, page 133) for all the data available in 2015. We first tested whether this error was caused by an error in the simulation of water content and tested the same equations using the measured water content (θ_{obs}) from 2015 instead. Indeed, the bias and relative error decreased, though only by 11 % for the relative error. The Δ AIC, the difference between AIC of the reference and the other tested expression, shows its lowest value for $\epsilon_s = 1$. The change from the expression of emissivity containing water content to $\epsilon_s = 1$ adds one estimated parameter to the model.

In the case of a bare soil surface, Campbell and Norman (1998) mentions values of ϵ_s between 0.93 and 0.96. (Ham and Senock, 1992) measures an

emissivity between 0.8 and 0.84 for a sandy plot and between 0.91 and 0.96 for a silt loam. In order to further reduce the error, we fixed the emissivity to 1, its maximum possible value according to the Stefan-Boltzmann law. This makes the soil a black body, or perfect emitter. With $\epsilon_s = 1$, the bias was down to 7.7 $\%$ while relative error dropped to 6.5 $\%$, representing a decrease of 76 $\%$ relative to Lu equation with simulated water content (θ_1).

Graphically, the difference between the different parametrisation of ϵ_s is showed in figure 3.8 page 133 for 5 days in April 2015 (same days as in figure 3.7). The first three days (22, 23 and 24 April) were sunny while the remaining two (25 and 26 April) were rainy.

On the sunny days, observed upward long-wave downward radiation peaked above 500 $\mathrm{W\,m^{-2}}$ while simulated only reached around 450 $\mathrm{W\,m^{-2}}$ when ϵ_s was equal to 1. The two remaining parametrisation remained under the observations (high positive bias in table 3.4). On the rainy days (25 and 26 April), the simulated values with $\epsilon_s = 1$ agreed well with the observed values. A wet soil is darker and its properties may be closer to a perfect emitter. This could explain why the parametrisation $\epsilon_s = 1$ agrees better with the data on rainy days.

This parametrisation of Lu is only a local adaptation to increase the accuracy of the simulation for this particular place. The choice of $\epsilon_s = 1$ remains to be tested in other locations, with other types of soil.

In summary, the sensibility analysis, by estimating the most significant parameters, paved the way towards model improvement. We have now estimated B_a, B_b, introduced a cloud correction factor, and set the soil emissivity to 1. One parameter remains to be investigated, this time belonging to the net short-wave radiation, the albedo α. Once again, thanks to the net radiometer, the variability of this parameter may be explored.

2.4 New formalism for albedo

In the model SOPHIA, the albedo, or reflection coefficient, is involved in the total short-wave radiation St. Large α values imply that a larger quantity of short waves will be reflected directly back to the atmosphere while less radiative energy will be transferred to the soil. Here we recall the formula for net radiation showing the position of α

$$Rn = St(1 - \alpha) + Ld - Lu$$

TABLE 3.4: Bias, RMSE, relative RMSE (RRMSE) and Δ AIC of upward long-wave radiation with different expressions of soil emissivity ϵ_s. The indices were computed with 8741 data points, corresponding to all the hours in 2015 of upward long-wave radiation recorded by the weather station in Stein, Switzerland.

Emissivity	Bias	RMSE [$\mathrm{W\,m^{-2}}$]	RRMSE [%]	Estimated parameters	Δ AIC
$\epsilon_s = S + \frac{\theta_1 \cdot \theta_{min}}{P_{tot}}$	32.5	43.7	11.6	0	0
$\epsilon_s = S + \frac{\theta_{obs} \cdot \theta_{min}}{P_{tot}}$	28.6	39.1	10.4	0	-1932.6
$\epsilon_s = 1$	7.7	24.7	6.5	1	-9950.6

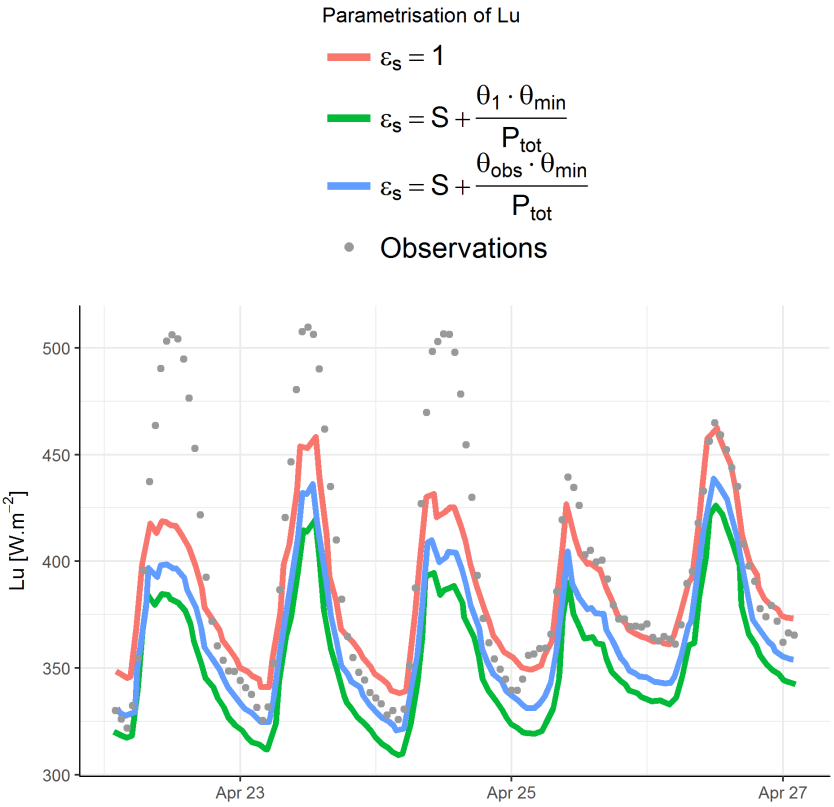

FIGURE 3.8: Comparison of different parametrisation of upward long-wave radiation (Lu) by modifying the soil emissivity ϵ_s between 22 and 27 April 2015 in Stein, Switzerland. We set $S = 0.9$, $\theta_{min} = 0.08\,\mathrm{m^3\,m^{-3}}$, $P_{tot} = 0.46\,\mathrm{m^3\,m^{-3}}$, θ_{sim} simulated volumetric water content at 5 cm with SOPHIA model, θ_{obs} the volumetric water content observed at 5 cm. The first three days were clear sky days while the last two were rainy.

The estimation of long-wave parameters was possible thanks to the measurements of the net radiometer. Similarly, the net radiometer allows us to compute the *observed* albedo α_{obs} as the ratio of St_{upward}, the total upward short-wave radiation reflected from soil to atmosphere, to St_{downward}, the total downward short-wave radiation coming from the sun and other elements in the atmosphere:

$$\alpha_{obs} = \frac{St_{\text{upward}}}{St_{\text{downward}}} \tag{3.22}$$

However this method produces a very noisy albedo (outside the expected values between 0.001 and 0.5) if straightforwardly computed without additional filtering. The albedo varies with the solar zenith angle (K. Wang et al., 2005). At low zenith angles (near sunrise and sunset), the albedo is very high: we had to first discard this data to study the albedo variation.

One approach consists in normalizing the albedo using its value at a solar angle of $60°$. This normalization didn't reduce the noise as we would have expected but nonetheless showed us to look at using the solar angle to clean the albedo data up. Hence, we computed the sun elevation using the R package `insol` (Corripio, 2003).

We chose to consider albedo at solar noon only, namely when the sun is at its highest elevation, to obtain a dataset as homogeneous as possible. We then discarded some residual outlier values that were still very high, especially in winter. We defined as outliers values superior to the 0.9 percentile, i.e. above 0.25.

The distribution of albedo at solar noon between January and December 2015 is presented in figure 3.9. A visual inspection of this histogram led us to assume the distribution of albedo being a mixture of two normal distributions whose mean and standard deviation were estimated applying the Expectations-Maximization (EM) algorithm (Bishop, 2006) with the R package `mixtools` (Benaglia et al., 2009).

This iterative algorithm returns the mixing coefficients, i.e. the mean and standard deviation of a mixture of Gaussian models by making an initial guess of parameters and iteratively refining them through 2 steps: a step of evaluation of responsibilities (posterior probabilities) and a re-estimation of parameters. Here we obtained that the two distributions of albedo were: $\mathcal{N}(0.10, 0.01)$ and $\mathcal{N}(0.15, 0.01)$.

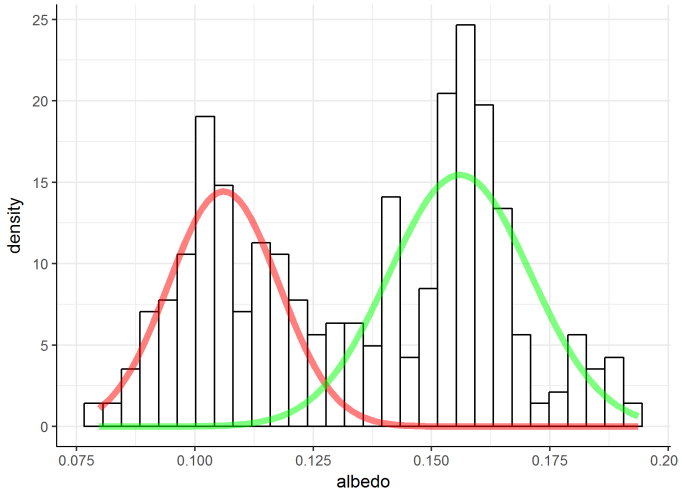

FIGURE 3.9: Density distribution of albedo at solar noon during the year 2015. Outliers with an albedo superior to 0.25 were removed. The density curves were found by running an EM algorithm for mixtures of univariate normals. Red curve: mean $\mu = 0.10$, standard deviation $\sigma = 0.01$; green curve: $\mu = 0.15$, $\sigma = 0.01$

In Müller (1999), the albedo was set to a constant value of 0.05, but the bimodal distribution that we observed suggests that albedo might change according to some factors. Campbell and Norman (1998) note that albedo varies between 0.08 and 0.18 according to the type of soil and its water content. A heavy, wet soil has an albedo around 0.08, while a light and dry soil has a higher albedo, around 0.18. The clay loam at the surface in Stein is a considered a light soil, and we could hypothesise that the two means of the two distribution could be the average values of albedo when the soil is respectively wet ($\alpha = 0.10$) and dry ($\alpha = 0.15$).

Van Bavel and DI Hillel (1976) propose an expression to compute the albedo that takes into account the water content of first layer θ_1:

$$\begin{aligned} \theta_1 > 0.25\,\mathrm{m^3\,m^{-3}} \quad & \alpha = 0.1 \\ \theta_1 < 0.1\,\mathrm{m^3\,m^{-3}} \quad & \alpha = 0.25 \\ 0.1 < \theta_1 < 0.25\,\mathrm{m^3\,m^{-3}} \quad & \alpha = 0.1 + (0.25 - \theta_1) \end{aligned} \tag{3.23}$$

This equation is used in most recent soil models, like HYDRUS 1-D (J Simunek et al., 2013). However, when evaluated against our albedo data at noon in 2015, this model revealed that it did not cover the parameter's entire variability. A closer look at the variation of albedo as a function of volumetric water content showed that the model from Van Bavel and DI Hillel (1976) is only valid for a wetting soil (figure 3.10, page 136), but not for a drying soil.

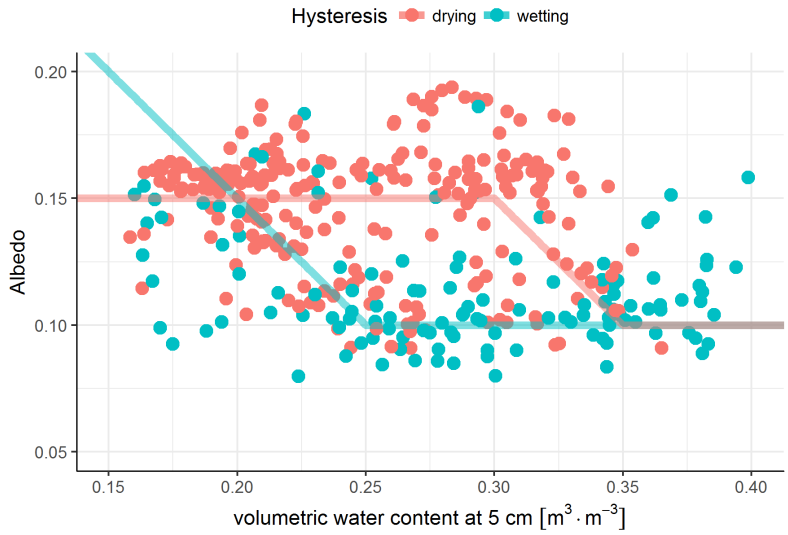

FIGURE 3.10: Relationship between albedo (at solar noon) and soil water content at 5 cm for each day of 2015. The colour of each point is linked to the hysteresis of the soil, namely whether it is drying or wetting. The set of rules for albedo (equation 3.23) from Van Bavel and DI Hillel (1976) is represented by the blue line and only accounts for situations when the soil is wetting. We developed a new set of rules to account for the albedo variation when the soil is in a drying process (red line, equation 3.24).

We determined whether the soil was in a wetting or drying state by computing a simple water balance at the surface, using the amount of rain and the actual evaporation computed by the energy balance in the SOPHIA model. When this balance is positive, there is more rain than evaporation, the soil is wetting. On the opposite, when the water balance is negative, the evaporation is more important than rain, therefore the soil is drying.

We thus decided to develop a new set of rules that better account for the albedo when the soil is drying (red dots in figure 3.10). We set the dry albedo value to 0.15, which is the estimated value of the second peak of the distribution.

$$\begin{aligned} \theta_1 &> 0.35\,\mathrm{m^3\,m^{-3}} \quad \alpha_{drying} = 0.1 \\ \theta_1 &< 0.30\,\mathrm{m^3\,m^{-3}} \quad \alpha_{drying} = 0.15 \\ 0.30 < \theta_1 &< 0.35\,\mathrm{m^3\,m^{-3}} \quad \alpha_{drying} = 0.1 + (|\theta_1 - 0.35|) \end{aligned} \tag{3.24}$$

We switch from one set of equations to the other depending on the ongoing process of hysteresis, namely the wetting or drying phases. Hence, we set the value of albedo to α_{drying} when the soil near the surface is drying (negative

TABLE 3.5: AIC (Akaike's Information Criterion) computed for different models and parametrisations of albedo (α). The AIC was computed with hourly albedo data recorded by the weather station in Stein, Switzerland in 2015. WB: surface water balance, θ_1: volumetric water content of the first layer below soil surface. Note that the ΔAIC are the differences between the AIC of the current model and the reference one (first line) i.e. the reference is not chosen as the one with the lowest AIC as more classically done.

Albedo	Nb estimated param.	AIC	ΔAIC
$\alpha = 0.05$	0	-41823.4	0.0
$\alpha = 0.10$	1	-53780.0	-11956.6
$\alpha = 0.15$	1	-59685.8	-17862.4
$\alpha = f(\mathrm{WB}, \theta_1)$	8	-64562.5	-22739.2

WB, surface water balance (rainfall-evaporation)

surface water balance). When the soil is wetting (positive surface water balance), we use the value of albedo α.

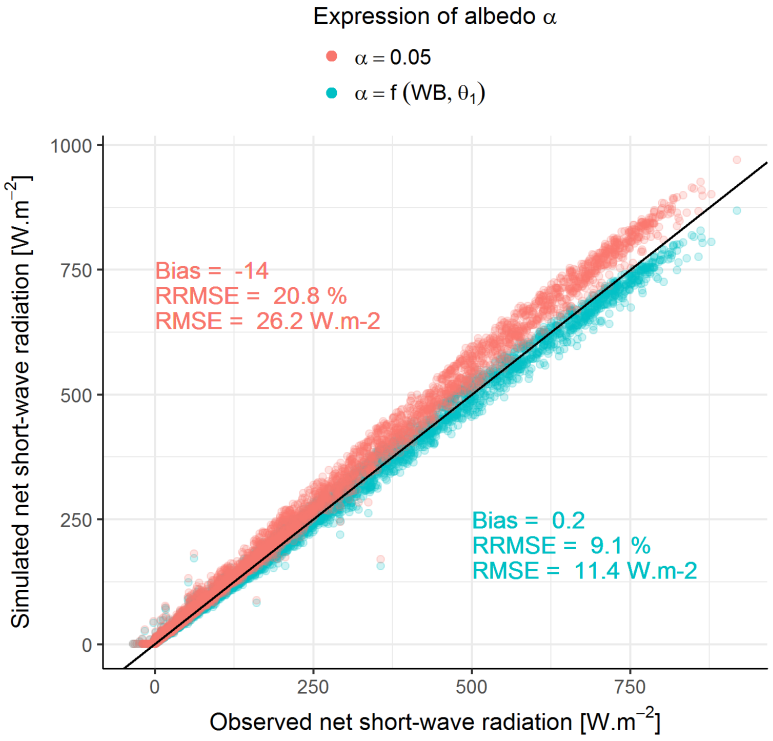

FIGURE 3.11: Simulated *vs* observed net short-wave downward radiation with the default expression of albedo $\alpha = 0.05$ in red and the new expression proposed in equation 3.23 and 3.24 in blue. The plot shows the data of hourly net short-wave data for the whole year 2015. The black line denotes the 1:1 line. RRMSE: relative root mean square error, RMSE: root mean square error. Number of points: n=8760

To compare the different expressions for albedo, we computed the AIC to determine which expression is better suited to simulate the albedo (table 3.5, page 137). The initial constant value of $\alpha = 0.05$ is the value for albedo in Müller (1999). When this parameter is estimated to 0.10 (albedo when the soil is wet) or 0.15 (albedo when the soil is dry) we obtain a decrease of AIC of 40% and 50% respectively. For the expression of albedo in function of surface water balance and volumetric water content θ_1, the AIC decreases by 56% in reference to the default value of $\alpha = 0.05$. The lowest AIC is obtained with this expression, despite the higher number of estimated parameters required (8).

The performance in terms of RMSE of the new albedo expression on net short-wave radiation is presented in figure 3.11, page 138. The figure presents all the hours for the entire year 2015. When albedo is fixed at 0.05, as in Müller (1999), there is a negative bias showing an over-prediction of net short-wave. The error is $26.2\,\mathrm{W\,m^{-2}}$ corresponding to a relative error of $21\,\%$. The relative

error is reduced by $56\,\%$ if the albedo function of surface water balance is computed. In addition, the bias is reduced to almost 0, which shows a good agreement between observed and simulated values.

To conclude, the new expression of albedo requires access to albedo data to estimate the eight parameters, but it improves the simulation net shortwave radiation by reducing the error by $56\,\%$. This expression of albedo should be tested against another reflection coefficient dataset, from another location, in order to evaluate its value.

3 Evaluation of the submodules improvements on model outputs

3.1 Evaluation of the complete model using the 2015 dataset, used for the calibration of some of its sub-modules

In the previous section, we have proposed calibration and/or new expressions for some components of the energy balance. We are going to evaluate each of them, once integrated in the full model, over the whole year 2015 (from 1 January 2015 to 31 December 2015, N=8760 hours) for simulation of net radiation, soil temperature and soil water content. These three variables were selected because net radiation is the most important energy input of the model, and soil temperature and water content are our two outputs of interests.

We will evaluate the three model versions presented in the previous chapter (calibration of Brunt's coefficient B_a B_b,introduction of the cloud correction factor, set of ϵ_s to 1 and new expression for albedo α) and the calibration of parameters a and b derived from the water-retention moisture curves. As presented in table 3.6, model versions that included only one change in an expression or in the set of estimated parameters are denoted by a single letter A to D.

- The initial version includes the default surface parameter values from Müller (1999) and the soil parameters computed by the Saxton model (see chapter 2).
- Model version A is the same model as the initial version, only with the soil emissivity ϵ_s set to 1. The new expression replaces the expression

of ϵ_s shown in chapter 2 (equation 2.14). The value $\epsilon_s = 1$ was set after studying the upward-longwave radiation over the whole year 2015.

- Model version B corresponds to the model with estimation of a and b from the water-retention curves measured in Stein (section 4.4, chapter 2).

- Model version C is a model version where the downward long-wave radiation has been improved with the calibration of B_a and B_b on clear-sky days of 2015 and where the atmospheric emissivity has been computed with the addition of a cloud correction factor cc.

- Finally, model version D includes the new expression for albedo α proposed in the previous section. The expression was set by analysing the daily albedo data from 2015 computed with measurements from the weather station. It includes eight parameters because the water content thresholds and the two albedo values when the soil is either wet (0.10) or dry (0.15) were derived from the data analysis of our dataset. It is likely that these parameters could change in other places and would have to be recalibrated.

The whole set of model versions is thus a list of combinations of the letters A to D in the 11 possible ways, resulting in a total of 15 models to evaluate for output prediction.

As our different calibration steps are performed on terms composing the net radiation, we first present the bias, MAE and relative MAE for net radiation, and RMSE and relative RMSE for the long-wave upward term in table 3.7 page 143. For net radiation, the bias was computed as $|bias| = |\bar{y}_{obs}| - |\bar{y}_{sim}|$ to avoid problem with negative net radiation values. The bias is negative and shows an important over prediction of net radiation. The lowest bias (closest to 0) is -10.6 and is found when the three versions A, C and D are combined. On their own, that is to say model version A, C and D, with only one improved net radiation terms, do not improve the bias compared to the initial version, except for the new formula of albedo (version D).

Version A also shows a $11\,\%$ increase in error for net radiation, despite that it reduces the error by 38 % for upward long-wave radiation. An important decrease of the error on net radiation and long-wave upward is clear for version ACD, with $45\,\%$ and $37\,\%$ error decrease, respectively. For net radiation, the version ACD shows a final relative error of $23.2\,\%$, or $26\,\mathrm{W\,m^{-2}}$. As the

TABLE 3.6: List of the 15 model versions resulting from the combinations of the 4 possible variants presented in section 2 of this chapter. The initial version of SOPHIA is the one presented at the end of chapter 2, including default surface parameters from Müller (1999) and soil parameters computed by the Saxton model. Model versions A to D each correspond to modification of only one module and/or the calibration of its associated parameters. The two-, three- and four-letter codes correspond to their possible combinations

Version	Changes compared to initial version	Number of estimated parameters
Initial	Default parameters	0
A	$\epsilon_s = 1$	1
B	a and b estimated from moisture-retention curves	2
C	B_a, B_b estimated and cloud correction factor cc added	3
D	$\alpha = f(\mathrm{WB}, \theta_1)$	8
Combination of 2 calibrations		
AB	a, b and $\epsilon_s = 1$	3
AC	B_a, B_b with cc and $\epsilon_s = 1$	4
BC	a,b and B_a, B_b with cc	5
AD	α and $\epsilon_s = 1$	9
BD	a,b and α	10
CD	B_a, B_b with cc and α	11
Combination of 3 and 4 calibrations		
ABC	a,b,$\epsilon_s = 1$ and B_a, B_b with cc	6
ABD	a,b,$\epsilon_s = 1$ and α	11
ACD	B_a, B_b with cc,$\epsilon_s = 1$,α	12
BCD	a,b, B_a, B_b with cc,α	13
ABCD	a,b, B_a, B_b with cc,α,$\epsilon_s = 1$	14

calibrated parameters in version B (moisture-tension parameters) do not have an effect on net radiation, the version ABCD shows a similar low error.

In the same table (table 3.7, we do not show the error for short-wave radiation and downward long-wave radiation because these terms are sensitive only to the calibration of albedo or brunt's coefficient, respectively. Model version C, alone or in combination produces always an error reduction of $57\,\%$ (as shown in figure 3.6) for downward long-wave. Similarly, version D always decrease the error on net short-wave by $56\,\%$. There is no interaction between calibrated parameters for these terms.

We showed that model ACD is the best for the simulation of net radiation, can we expect a lower error for soil temperature as well?

For models with only one module change (versions A, B, C or D), the calibration of B_a and B_b with the cloud correction factor cc (version C) brings the lowest error to soil temperature at 5 and 30 cm, $16.1\,\%$ ($2.1\,^{\circ}\mathrm{C}$) and $9.5\,\%$ ($1.3\,^{\circ}\mathrm{C}$), which represent a decrease of $32\,\%$ and $56\,\%$ compared to the initial version (table 3.8, page 144). The version ACD, performant for $Rnet$, presents a decrease of $17\,\%$ of error. The reduction of error on $Rnet$ does not have a large effect on soil temperature because the sensible and latent heat transfer at play in the energy balance are playing an important role as well.

For soil water content at 5 cm, version ABD has the lowest error, $30\,\%$ ($0.08\,\mathrm{m^3\,m^{-3}}$), which is a reduction of $14\,\%$ of error. It is worth noting that the same version ABD, good for water content, is the one with the largest error for soil temperature at 5 cm ($31.5\,\%$). As the bias for temperature is positive, there is a general under-prediction of soil temperature which could lead to a reduced evaporation and hence a better soil water content simulation. Indeed, one explanation for the antagonism in the simulation performances between temperature and water content near the surface lies in the evaporation term. One of the known weakness of SOPHIA is the over-prediction of evaporation. The evaporation is over-predicted when the aerodynamic resistance is too low. As we raised artificially the aerodynamic resistance to 3000 $\mathrm{s\,m^{-1}}$, we lowered the evaporation and obtained soil water content simulation very close to observation. The problem is that when we had a good simulation of water content at the surface, the soil temperature near the surface was too high. We did not have sufficient time to investigate further the relationship between soil temperature and moisture with aerodynamic resistance.

The calibration of moisture-retention curve parameters a and b (version

TABLE 3.7: Goodness-of-fit indicators for simulation with different versions of SOPHIA over the whole year 2015 (from 1 January 2015 to 31 December 2015, N=8760 hours). The MAE and RMSE are expressed in $\mathrm{W\,m^{-2}}$. The rMAE, rRMSE and relative to initial columns are percentages. μ is the mean and σ the standard deviation of the indicators of model versions A to ABCD (not including the initial version).

	Net radiation				Upward long-wave			
Version	|bias|	MAE	rMAE	Relative to initial	bias	RMSE	rRMSE	Relative to initial
Initial	-31.6	46.6	41.8	0	31.1	42.7	11.3	0
A ($\epsilon_s = 1$)	-34.1	51.5	46.2	11	12.2	26.6	7.0	-38
B (a and b)	-31.5	46.5	41.7	0	31.9	43.3	11.5	2
C (B_a, B_b, cc)	-27.5	50.6	45.3	8	22.2	38.1	10.1	-11
D (albedo)	-19.2	39.8	35.7	-15	32.5	44.7	11.8	4
AB	-34.2	51.6	46.2	11	12.3	26.7	7.1	-37
AC	-23.4	36.2	32.4	-23	1.4	25.3	6.7	-41
BC	-27.9	51.5	46.1	10	23.3	38.8	10.3	-9
AD	-22.1	46.7	41.9	0	14.1	28.9	7.6	-33
BD	-19.2	39.5	35.4	-15	33.2	45.3	12.0	6
CD	-14.4	38.6	34.6	-17	23.8	40.1	10.6	-6
ABC	-23.5	36.3	32.6	-22	1.6	25.4	6.7	-41
ABD	-22.3	46.7	41.9	0	14.1	29.0	7.7	-32
ACD	-10.5	25.9	23.2	-45	3.4	26.9	7.1	-37
BCD	-14.9	39.5	35.4	-15	24.9	40.9	10.8	-4
ABCD	-10.6	26.1	23.4	-44	3.6	27.0	7.2	-36
μ	-22.4	41.8	37.5	-10	17.0	33.8	8.9	-21
σ	7.8	8.5	7.6	18	11.4	7.8	2.1	18

B) improves the relative error by $20\,\%$ on soil water content at 30 cm, and only $5\,\%$ for soil water content at 5 cm. When all the calibrated parameters are combined in version ABCD, the soil water content at 30 cm presents a reduction of error of $23\,\%$ compared to the initial version.

The computation of RMSE opens up the exploration of the parameters which should be estimated in priority to enhance the model's fitting. As the lowest error is not given by the same model for each output, the answer is not straightforward. To get a good fit for the temperature at 5 cm, one might reckon that estimating Brunt's coefficient and adding the cloud correction factor provide the best improvement. However, this assumption remains to be tested for other locations.

TABLE 3.8: Goodness of fit indicators for soil temperature simulation with different versions of SOPHIA over the whole year 2015 (from 1 January 2015 to 31 December 2015, N=8760 hours). RMSE is expressed in $\mathrm{W\,m^{-2}}$. rRMSE and relative to initial column are percentages. μ is the mean and σ the standard deviation of the indicators of model versions A to ABCD (not including the initial version).

	Temperature							
	at 5 cm				at 30 cm			
Version	bias	RMSE	rRMSE	Relative to initial	bias	RMSE	rRMSE	Relative to initial
Initial	2.5	3.1	23.8	0	2.7	2.8	21.5	0
A ($\epsilon_s = 1$)	3.3	3.8	29.2	23	3.5	3.6	27.4	27
B (a and b)	2.5	3.1	23.7	0	2.7	2.8	21.4	-1
C (B_a, B_b, cc)	0.3	2.1	16.1	-32	0.5	1.3	9.5	-56
D (albedo)	2.8	3.4	25.8	8	3.1	3.2	24.2	13
AB	3.3	3.8	29.3	23	3.5	3.6	27.5	28
AC	1.2	2.4	18.3	-23	1.4	1.8	13.7	-36
BC	0.3	2.1	16.1	-32	0.5	1.3	9.6	-55
AD	3.6	4.1	31.4	32	3.8	4.0	30.0	40
BD	2.8	3.3	25.7	8	3.0	3.2	24.1	12
CD	0.7	2.2	16.6	-30	0.9	1.5	11.3	-47
ABC	1.2	2.4	18.4	-23	1.4	1.9	14.0	-35
ABD	3.6	4.1	31.5	32	3.8	4.0	30.1	40
ACD	1.5	2.6	19.8	-17	1.7	2.1	16.2	-25
BCD	0.7	2.2	16.6	-30	0.9	1.5	11.3	-47
ABCD	1.6	2.6	19.9	-16	1.8	2.2	16.5	-23
μ	2.0	2.9	22.6	-5	2.2	2.5	19.1	-11
σ	1.2	0.7	5.8	24	1.2	1.0	7.6	35

In the next section, we will evaluate the performance of model version C and ABCD over the whole year 2016 (2016 dataset was not used for calibration). Version C proposes the best improvement for soil temperature with only 3 estimated parameters, while version ABCD proposes a good error reduction for all six outputs analysed in this section.

3.2 Model performance in Stein, Switzerland for 2016

Evaluation over the whole year 2016, a dataset not used for the estimation step

The three selected model versions (initial, C and ABCD) were evaluated over the whole year 2016 dataset, which was not used for calibration.

For the three simulations, we used a time step of 15 minutes and a soil discretisation of 15 layers of 10 cm, representing a total soil depth of 150 cm. Table 3.10 synthesizes the main features that distinguish the three models from to the initial version of SOPHIA.

The initial conditions for all layers was set to the conditions measured on first January at 5 cm ($T^0_{1...15} = 4.2°\mathrm{C}$ and $\theta^0_{1...15} = 0.3\mathrm{m}^3\,\mathrm{m}^{-3}$). All the layers were assigned a clay content of 27 %, sand content of 36.7 % and organic matter of 3 %.

TABLE 3.9: Goodness of fit indicators for soil temperature simulation with different versions of SOPHIA over the whole year 2015 (from 1 January 2015 to 31 December 2015, N=8760 hours). RMSE is expressed in $\mathrm{W\,m^{-2}}$. rRMSE and relative to initial columns are percentages. μ is the mean and σ the standard deviation of the indicators of model versions A to ABCD (not including the initial version).

	Volumetric Water Content							
	at 5 cm				at 30 cm			
Version	bias	RMSE	rRMSE	Relative to initial	bias	RMSE	rRMSE	Relative to initial
Initial	0.053	0.090	34.7	0	0.015	0.042	15.4	0
A ($\epsilon_s = 1$)	0.041	0.085	33.0	-5	0.009	0.039	14.3	-7
B (a and b)	0.042	0.087	33.7	-3	0.000	0.034	12.3	-20
C (B_a, B_b, cc)	0.076	0.104	40.2	16	0.027	0.049	17.9	16
D (albedo)	0.045	0.085	32.9	-5	0.011	0.039	14.4	-7
AB	0.030	0.083	31.8	-8	-0.006	0.033	12.2	-21
AC	0.070	0.099	38.3	10	0.024	0.046	17.0	10
BC	0.074	0.105	40.5	17	0.013	0.037	13.6	-12
AD	0.033	0.081	31.3	-10	0.004	0.037	13.5	-12
BD	0.033	0.082	31.6	-9	-0.005	0.033	11.9	-23
CD	0.072	0.101	38.9	12	0.025	0.047	17.3	12
ABC	0.064	0.097	37.4	8	0.009	0.035	12.7	-18
ABD	0.023	0.078	29.9	-14	-0.010	0.033	12.1	-21
ACD	0.065	0.095	36.8	6	0.021	0.044	16.2	5
BCD	0.069	0.100	38.8	12	0.011	0.035	13.0	-16
ABCD	0.055	0.089	34.5	-1	0.005	0.033	11.9	-23
μ	0.053	0.091	35.3	2	0.009	0.038	14.0	-9
σ	0.018	0.009	3.6	10	0.012	0.006	2.1	14

TABLE 3.10: The three model versions and their differences in terms of parameter values or expressions. The parameters not mentioned in this table are the same as in table 2.3, page 80

	Sophia model		
Parameters	Initial version	Version C	Version ABCD
a	-3.9	-3.9	-2.3
b	-6.5	-6.5	-5.2
B_a	0.605	0.58	0.58
B_b	0.039	0.066	0.066
ϵ_a	ϵ_a	$\epsilon_a(1-0.84cc)+0.84cc$	$\epsilon_a(1-0.84cc)+0.84cc$
ϵ_s	$S+\frac{\theta_1 \cdot \theta_{min}}{P_{tot}}$	$S+\frac{\theta_1 \cdot \theta_{min}}{P_{tot}}$	1
α	0.05	0.05	$f(\mathrm{WB},\theta_1)$ see section 2.4

WB, surface water balance (rainfall-evaporation). Other symbols are explained page xxv

TABLE 3.11: Bias, RMSE and relative RMSE (RRMSE) for ten outputs of the SOPHIA model for the whole year 2016 in Stein. A negative bias means general over-prediction. The RMSE is in the unit of the variable, while RRMSE is the RMSE relative to the average of the observation in percent. N=8784 (366 days of hourly data).

	Initial SOPHIA version			SOPHIA version C			SOPHIA version ABCD		
Model output	Bias	RMSE	RRMSE	Bias	RMSE	RRMSE	Bias	RMSE	RRMSE
Soil temperature at 5 cm [°C]	1.59	2.56	21 %	-1.06	2.46	20 %	0.32	1.82	15 %
Soil temperature at 30 cm [°C]	1.72	2.19	17 %	-0.90	1.73	14 %	0.48	1.15	9 %
Soil temperature at 140 cm [°C]	1.75	3.95	31 %	-0.79	3.67	29%	0.58	3.02	24 %
Net radiation [$\mathrm{Wm^{-2}}$]	15.65	55.59	78 %	-36.07	61.16	86 %	-7.65	32.69	46 %
Net short-wave radiation [$\mathrm{Wm^{-2}}$]	-16.04	30.80	27 %	-16.04	30.80	27 %	-3.22	13.37	12 %
Upward long-wave radiation [$\mathrm{Wm^{-2}}$]	20.89	28.75	8 %	9.25	24.10	6 %	-6.35	19.36	5 %
Downward long-wave radiation [$\mathrm{Wm^{-2}}$]	52.58	58.92	18 %	-10.78	24.23	7 %	-10.78	24.23	7 %
Soil water content at 5 cm [$\mathrm{m^3\,m^{-3}}$]	0.03	0.07	25 %	0.05	0.08	29 %	0.03	0.07	26 %
Soil water content at 30 cm [$\mathrm{m^3\,m^{-3}}$]	0.00	0.03	12 %	0.02	0.04	14 %	0.00	0.03	10 %
Soil water content at 140 cm [$\mathrm{m^3\,m^{-3}}$]	0.13	0.14	35 %	0.14	0.15	37 %	0.13	0.13	33 %

The evaluation of the three model simulations for 2016 is presented in table 3.11, page 148. We computed for the whole year 2016 (that is, 8784 hours) the bias, RMSE and relative RMSE (RRMSE). The best simulation, that is, the lowest error for all evaluated variables (except water content at 5 cm) is obtained with the full model version ABCD. With this version, the average error on soil temperature at 5 cm, our main variable of interest, is 1.82 °C which represents a relative error of 15 %. The bias of 0.32 is very low and shows that this model version does not over or under-predicts the soil temperature at 5 cm. The bias increases to 0.48 and 0.58 for soil temperature at 30 and 140 cm, respectively. The average error is lower at 30 cm (1.15 °C) and higher for 140 cm (3 °C). This error at high depth is probably linked to the different texture at this depth. It is a sandier soil (clay 19 %, sand 49 %) and the gravel content is higher. This error could be corrected by setting different soil parameters for each soil layer.

The model version C represents only a small improvement compared to the initial version (6 % on average) and is less accurate in all the variables than version ABCD. The fact that version C is of course included in version ABCD was a reason to expect this trend but not a sufficient one since the variables used for this evaluation were not all used for its calibration. It is likely that the model version ABCD is more accurate because the estimation of albedo is important for summer months, when the radiation is high.

Analyzing the performance of SOPHIA model version ABCD in more details reveals that it is very accurate (less or equal to 10 %) for soil temperature and water content at 30 cm, as well as down- and upward long-wave radiation. The relative error of 15 % for soil temperature at 5 cm shows us that our model version ABCD can reproduce well the seedbed temperature over the whole year.

Compared to the initial version of the model, version ABCD shows a decrease in the relative error of 30 % on average. The most significant reduction in error appertains to both the net short-wave radiation (due to the improvement of albedo, 57 % decrease) and long-wave downward (due to the calibration of Brunt's coefficient and the cloud correction algorithm, 59 % decrease). There was no improvement in the water content at 5 cm.

The largest relative RMSE concerns net radiation (46 % or 32.7 $\mathrm{W\,m^{-2}}$) even if the error of its components (Sn, Lu and Ld) is low: net short-wave radiation (12 %), upward long-wave radiation (5 %) and downward long-wave (7 %).

The computation of MAE and RMAE is shown for net radiation and net short-wave radiation in table 3.12, page 151. The mean absolute error

(MAE) is lower because MAE does not over-weight large difference between observation and simulation. The RMAE for model version ABCD is 22 $\%$. The difference between 22 $\%$ and 46 $\%$ of RRMSE suggests that there exist large differences that weight strongly in the computation of RMSE.

However, all the radiation terms show a negative bias (table 2.4): they are over-predicted compared to the observations. The highest bias applies to downward long-wave radiation, meaning that testing for other formulations of the cloud cover correction for atmospheric emissivity would be valuable.

To conclude the evaluation in 2016; the variables that are the most accurately simulated are soil temperature and soil water content at 30 cm below soil surface. We improved the simulation of the temperature at 5 cm by 30 $\%$ by estimating parameters and introducing new expressions for soil emissivity and albedo. The simulation of water content at 5 cm did not improve with the calibration of surface parameters. Our ongoing research has suggested that the error on water content could be linked to our over-estimation of evaporation, but this problem remains to be investigated in more details.

Evaluation of soil temperature from ABCD model in different situations

As a last performance check, our model was tested in "real-use conditions". This was the last step before applying the model to simulate temperature for emergence. The model ran for 50 days, in situations for which no prior information about the soil temperature or water content was available. Weather conditions, however, were considered as available since they may be readily recorded. Therefore, we assigned the air temperature measured at 2 m as the initial value for the entire soil profile. Concerning the soil water content, we gave every layer as a starting value 0.35 $\mathrm{m}^3\,\mathrm{m}^{-3}$ of water, which is slightly above the soil field capacity (0.32 $\mathrm{m}^3\,\mathrm{m}^{-3}$).

We ran the model version ABCD for 11 virtual sowing windows that have been dispatched through 2015 and 2016 to obtain different conditions in Stein, in order to explore how the model would behave in these contrasted situations. Although some of these 11 periods are of course not traditional sowing periods for maize, they are actually the ones that have been used in our experiment on maize emergence, as we shall present in the next chapter.

To reproduce, as well as possible, "real-use conditions", we added a 15 days burn-in period, meaning that we started the simulation 15 days before the sowing date to let the soil profile reach a stable regime and avoid the

TABLE 3.12: Comparison of RMSE and RMAE for net radiation and net short-waver radiation on whole year 2016. The RMAE (equation 3.16) has the advantage of not over-weighting large differences between observation and simulations. Rnet, net radiation; Sn, net short-wave radiation

Version		Rnet	Sn
Initial Version	RMSE	55.6	30.8
	MAE	45.8	16.1
	RRMSE	78 %	27 %
	RMAE	44 %	14 %
C	RMSE	61.2	30.8
	MAE	41.0	16.1
	RRMSE	86 %	27 %
	RMAE	39 %	14 %
ABCD	RMSE	32.7	13.4
	MAE	22.6	6.5
	RRMSE	46 %	12 %
	RMAE	22 %	6 %

influence of the initial conditions. To capture the beginning of the crop cycle, we continued the simulation for 35 days after the sowing date. The results of the simulation of soil temperature at 5 cm are presented in figure 3.12. The simulated soil temperature of the model was compared to hourly observations from the weather station in Stein, Switzerland.

For all sowing windows, the average error on soil temperature at 5 cm is 2.9 °C with a coefficient of variation of 36 %: the average error strongly varies depending on the sowing window.

The best performance on soil temperature at 5 cm was obtained for the simulations starting 27 September, 23 March, and 7 September 2015. During these situations, the error was below 2 °C and relative RMSE between 11 and 15 %. The error was between 2 and 3 °C for simulations starting 9 April, 27 April and 26 August 2015 and was above 4 °C for simulations starting 26 May, 17 June, 30 July 2015. Globally, there is a trend to larger errors during hot summer months (June, July and August).

For summer months, the simulation starting on 17 June displayed the highest RMSE (4.7 °C) and an under-prediction of the soil temperature. Our investigations suggest that this under-prediction, also found for simulations started on 30 July and 26 August 2015, might be linked to the soil losing more energy at night in simulations than what actually occurs.

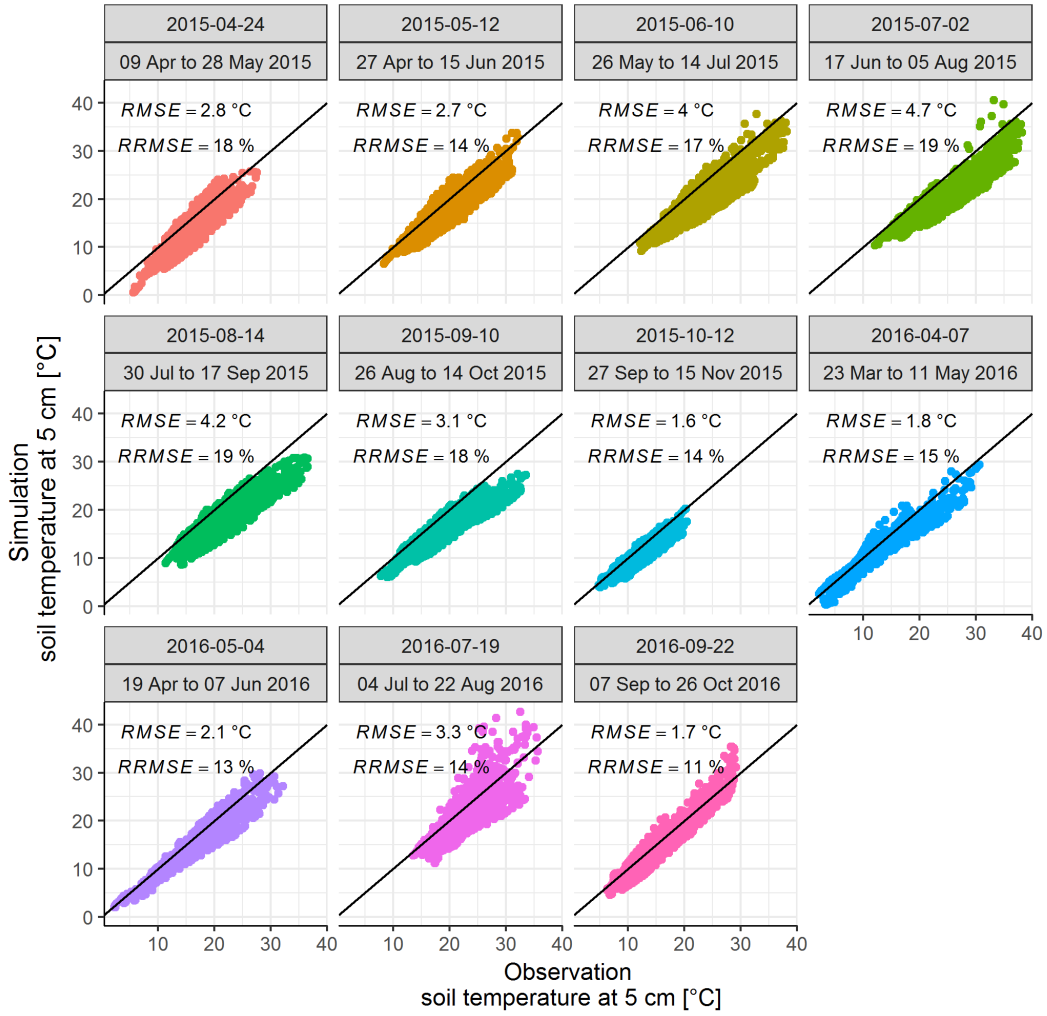

FIGURE 3.12: Simulated vs. observed soil temperatures at 5 cm over several periods centred around our experiments' sowing dates in 2015 and 2016 in Stein, Switzerland. The model version used is ABCD. The date above each graph is the sowing date and the period of simulation is shown below. The model was run with a time step of 900 seconds and a soil discretisation of 15 layers of 10 cm. The bottom condition for the soil profile was chosen as a water table at 1.5 m, reflecting the reality in Stein. The initial conditions for the whole profile for water content was 0.35 $\mathrm{m}^3\,\mathrm{m}^{-3}$ and, for temperature, set equal to air temperature at the beginning of the simulation.

On the contrary, in July 2016, the model produces an overprediction of temperature, probably caused by an unusually dry period, where simulated water content was lower than observed and would cause the simulation to display a higher temperature than observed.

The simulation for normal sowing situations (starting in April) obtains RMSE below 20 %. The average error for this situation is higher than the overall model evaluation for the whole year 2016. Testing an extended burn-in period (starting on January 1st) slightly increased the accuracy of the model, but not significantly.

4 Discussion and summary

In this section, we have presented a mathematical analysis of our SOPHIA model, including sensitivity analysis, estimation of parameters, model selection and evaluation. Although still far from being complete, especially given the complexity of the model and the large number of possible ways to improve each of its submodules, this first analysis nevertheless provides us some interesting interpretations and perspectives.

Sensitivity analysis. After model development and tests with texture classes, we performed, a sensitivity analysis of three model outputs (soil temperature at 5 cm, water content at 5 cm and net radiation) to rank the parameters according to their respective contributions to the output variations. The ranges of the surface parameters were drawn from the literature. For soil parameters, we fixed the specific soil texture of the soil in Stein, and obtained the parameter ranges by changing the compaction, organic matter content and content in gravels. The result from the Morris sensitivity analysis showed that the parameters have different interactions with each other depending on weather conditions. The interaction term for parameters like bulk density and hydraulic conductivity at saturation is higher when it is raining than when the sky is clear. The detailed investigation of these interactions with a Sobol analysis appears as one of this work's perspectives. Indeed, high computing cost using the R `sensitivity` package impeded its convenient or repeated use. We should consider to recode the model in a faster program and use high-performance computing to obtain precise sensitivity indices and information on the interaction between parameters. For soil temperature at 5 cm, the two most influential parameters at first order were roughness length and reference

height. Roughness length parameter is also reported as an important one by Hou et al. (2015) and Collins and Avissar (1994), with the latter specifying that the roughness length is the most important parameter when the model is used to reproduce a heterogeneous land area. Parameter B_a and albedo were also ranking high, suggesting the critical role of downward long-wave radiation and net short-wave on soil temperature. For water content at 5 cm, the most influential parameters were bulk density, hydraulic conductivity at saturation and moisture retention curve parameter a. Bulk density influences the porosity of the soil and has a large effect on water content.

Parameter estimation We took advantage of the net radiometer data to estimate parameters related to the net radiation. We performed their estimation and evaluated their effect using our hourly data of the year 2015: it comprises 8785 points of observation, thus providing us a reliable way to test our procedure. The estimation of parameters B_a and B_b was done on clear-sky days of 2015. Clear-sky days were selected as days with maximal global radiation superior to 800 $\mathrm{W\,m^{-2}}$. After checking with observed data, we introduced a cloud correction factor as detailed in M. H. Unsworth and JL Monteith (1975). The most influential parameter for net radiation is albedo. To improve this term, we borrowed the expression of albedo in Van Bavel and DI Hillel (1976), still in use in the recent HYDRUS-1D model (J Simunek et al., 2013). Our data analysis showed that this expression fits well with the albedo when the soil is wetting but does not account for the albedo when the soil is drying. We introduced another function to compute the albedo when the soil is drying. We devised whether the soil was drying or wetting by performing an instantaneous water balance at the surface (Rain - Evaporation). The water balance was an efficient way to discriminate between wetting-drying cycle. The new expression for albedo reduced the average error on net short-wave radiation by 56 $\%$ (from 26.2 $\mathrm{W\,m^{-2}}$ to 11.5 $\mathrm{W\,m^{-2}}$) when applied to the entire year 2015.

Selection of model with lowest error As we estimated different parameters and added new expressions to the model, it was natural to evaluate which model was fitting well with the available observed data. We evaluated the 15 possible versions of the resulting fully integrated model, combining the different calibration and new expressions. We evaluated the models on the whole year 2015 to cover a wide range of weather conditions.

The results showed that we obtained the lowest error for soil temperature at 5 cm by estimating parameters B_a, B_b and adding the cloud cover correction to the emissivity (version C). However, the error was not improved for the

other variables, namely net radiation and soil water content. The best compromise that improved the error of the three variables was obtained with model version ABCD, combining all the calibration and new expressions presented in this chapter.

Model evaluation We evaluated the relative performance of the initial version of the model, model C and model ABCD. We computed RMSE indicators for 2016 (whole-year simulation) and also applied the model in "real-use" conditions to predict the emergence of maize, namely for a shorter period with unknown initial conditions. The data in 2016 were not used for model development and calibration. Although not strictly independent from the calibration data since it was measured on the same location and same soil, the 2016 dataset may nevertheless be considered as an interesting validation dataset, since the meteorological conditions changed. In 2016, the best model for every variable was the ABCD version. The lowest error was obtained with soil temperature and water content at 30 cm with less than 10 $\%$ of error. Near the surface, the error increased slightly but was still under 15 $\%$ for soil temperature and 20 $\%$ for water content.

When evaluating the model with unknown initial conditions on different situations, the performance of the model version ABCD depended on the considered period, with good performance (RMSE below or around 2 $^{\circ}\mathrm{C}$) for spring or autumn months, and less (above 3 $^{\circ}\mathrm{C}$) for hotter, summer months. For comparison, we reviewed the error levels in similar published models. Bittelli, Ventura, et al. (2008) reports an error of 1.8 $^{\circ}\mathrm{C}$ at 2 cm and 1.6 $^{\circ}\mathrm{C}$ at 5.5 cm depth for the month of October. Pertaining to water content, they found an error of 0.04 $\mathrm{m}^3\,\mathrm{m}^{-3}$ and 0.02 $\mathrm{m}^3\,\mathrm{m}^{-3}$ for 2 and 5.5 cm depth respectively, which is much smaller than the error obtained with SOPHIA simulations. In Banimahd and Zand-Parsa (2013), the error on 1 year of soil temperature data is 2.2 $^{\circ}\mathrm{C}$ and it is 0.017 $\mathrm{m}^3\,\mathrm{m}^{-3}$ for the water content at 5 cm. Lastly, with the SHAW model (G. N. Flerchinger, 2000), error for soil temperature at various depths also circles around 2 $^{\circ}\mathrm{C}$ in their study of soil temperature simulation to predict wheat emergence (Bullied, G. N. Flerchinger, et al., 2014). The error of the SOPHIA model for temperature during cold months stays within the range of other published model. The model also needs to be tested in further locations to strengthen the validation. The difficulty here is to obtain quality data and to ensure that the soil is maintained bare without vegetation cover. The simulation of water content still needs some improvement. Our preliminary study of the aerodynamic resistance has shown us that it is responsible for too-high evaporation that leads to the soil

losing much more water than it should. Overestimation happened with our simplified expression but also the expression with stability parameter as well, that we implemented following the code presented in Bittelli, Campbell, and Tomei (2015). Our next steps will be to investigate further the aerodynamic resistance to improve water content simulation. To conclude, for our period of interest, i.e. the beginning of the crop cycle, usually in April and May for summer crops like maize, the model error was around 2 °C which is within the acceptable range of error for energy balance models.

4.1 Summary

This chapter was divided into four parts: the sensitivity analysis, the estimation of parameters, the evaluation of the model in 2015 (data used for calibration) and final evaluation in 2016 in Stein, Switzerland.

The sensitivity analysis was restricted to one soil type, a clay loam, to avoid model instability and obtain a consistent, sensitive parameter range. The Morris method, a semi-global sensitivity analysis, gave rapid results that were confirmed by the more computationally costly global method of Sobol.

The parameters with the most significant effect on soil temperature were shown to be the roughness length and parameters involved in the net radiation balance: soil emissivity factor, atmospheric emissivity parameters and albedo.

These parameters were estimated using the measurements from a net radiometer installed on the weather station. The net radiation data allowed us to implement a cloud correction factor to atmospheric emissivity and to propose a new formalism to simulate albedo when the soil is drying. The best fit for upward long-wave radiation was obtained with an emissivity equal to 1.

Finally, these two model versions were evaluated for the whole year 2016. The average error was the lowest for temperature and water content at 30 cm (relative error of 9 and 10 %) and upward and downward long-wave radiation (relative error of 5 and 7 %). For soil temperature at 5 cm, the relative error was 1.8 °C, which is a relative error of 15 %. We consider the accuracy sufficient to apply our model to emergence prediction.

There is still room for improvement of the water content simulations near the surface. Preliminary research has shown us that the error on water content is linked with an overestimation of evaporation due to a low aerodynamic resistance. One potential perspective would entail taking a closer look at this equation.

Chapter 4

Application of soil model to emergence prediction

This chapter is based on an article in preparation entitled "Combining a soil temperature model with selection of a thermal time model to improve the prediction of emergence in the field for maize cultivars."

1 Context

The soil model Sophia has been developed in order to provide information for other models to better understand complex interactions within the soil-plant-atmosphere system. One of the applications that we propose here is the prediction of emergence in the field. As we have seen in the introduction, emergence is a major step in the establishment of the crop canopy and hence an important step to achieving high yields. For crops, the timing and homogeneity of emergence are of the utmost importance. In this chapter, we will focus on the timing of emergence, as it is the starting point of biomass algorithm timelines for crop models. There are several ways to predict the timing to emergence.

Modelling timing to emergence Emergence is a phenological stage that is classically denoted as 'VE', standing for Vegetative Emergence (S. Ritchie, Hanway, and Benson, 1992). The time to VE is the period needed from planting to when 50 % of the plants have their coleoptile above the surface.

Usually, the timing of certain particular stages of development is predicted using the concept of thermal time. In this theory, the passage from one stage to the next happens on a given thermal time threshold, which is a specific amount of degree days or degree hours. Weed scientists use thermal time to predict the emergence of weeds (Forcella et al., 2000; H. Wang et al., 2009; Bullied, P. R. Bullock, et al., 2014), as this information gives a better timing for herbicide

application. The thermal time models use either the air or soil temperature, and the soil temperature models used in these approaches are often based on simple and empirical relationships (Roman, Murphy, and Swanton, 2000) rather than on the energy balance approach. The thermal time approach requires knowing the three cardinal temperatures of the organism: (i) the base temperature, a theoretical threshold below which no development is possible, (ii) the optimal temperature at which the development rate is maximum and (iii) the maximum temperature which defines the upper limit for development. These thresholds can change according to the period of interest. For example, it is known that the thresholds are not the same depending on whether the plant is in the heterotrophic phase or the autotrophic phase (Bhosale et al., 2007): this difference is related to the fact that the enzymes involved in starch remobilisation and germination are different from the enzymes involved in photosynthesis.

A few more mechanistic approaches have been also been developed to predict the time to emergence of crops, e.g. by simulating the coleoptile's elongation towards the surface. The crop model STICS (Brisson et al., 1998) uses a model of stem elongation towards the soil surface, and emergence is determined to occur when the stem is longer than the planting depth. Here, the required parameter is the threshold temperature for germination and then the sharpness coefficient of coleoptile elongation curve. This kind of model requires data on maize seedling elongation that may only be obtained in the lab and thus may not be representative of the actual conditions in the soil. This approach has received less attention from weed scientists since data is harder to come by and genetic variability is higher than for crops.

Weaich, Bristow, and Cass (1996) developed another seedling elongation model, where two different exponential models are used to represent the respective growth of internode and coleoptile. They coupled the resulting elongation model with a soil temperature and moisture model. They used the temperature to predict seedling elongation and soil moisture to predict soil strength. With their model, they studied the effect of different kinds of evaporative demands (due to wind or to radiations) on seedling elongation. They found that when the radiation term dominates the evaporative demand, maize emergence might fail because of this high temperature linked to high radiation, whereas when the evaporation is mostly determined by wind, the emergence might fail because of high soil strength resulting from the fast drying due to the wind.

Be they based on thermal time or coleoptile elongation approaches, these

models need soil temperature as input: this variable is a key driver of the seed and seedling development (Schneider and Gupta, 1985). However, because this information is seldom available, models rely on air temperature to predict the timing to emergence. This may be a source of errors since soil temperature differs significantly from air temperature: in April 2015 in Stein, Switzerland, the maximum difference was up to 6 °C (see chapter 1). In a study on wheat, H. Wang et al. (2009) found that air temperature was a better predictor of emergence timing than soil temperature, because the estimation of soil temperature were not accurate enough. The model of soil temperature that they use was not based on the energy balance approach. The soil temperature was predicted with the soil module of the DSSAT model (J. W. Jones et al., 2003), which uses air temperature and a deep soil temperature computed from average annual air temperature and amplitude of monthly mean temperatures. On the contrary, for Bullied, P. R. Bullock, et al. (2014), the use of an energy balance soil model improved their predictions of wheat emergence planted at different depths.

In our case study, we chose to stick to a thermal time approach. Indeed, no details on the physiology of maize emergence were neither available nor of interest for our aim that was to predict the date of emergence of maize in various sowing conditions at a given location where soil temperature measurements are available. The objective is to evaluate different thermal time models and in particular to test whether model forms other than the standard base temperature one could be more suited to describe it in a range of climatic conditions obtained by sowing at different times during the year.

2 Material and method

2.1 Thermal time models for emergence

Numerous thermal time models have been developed to predict plant development. The classical response of organism development follows an arrhenius curve (Bonhomme, 2000). For a certain range of temperatures, it is assumed that development stays proportional with temperature. It is possible to compute the accumulation of thermal time to predict attainment of a certain stage of development with the following formula:

$$TT_{stage} = \sum_{i=1}^{n} (T_i - T_b) \tag{4.1}$$

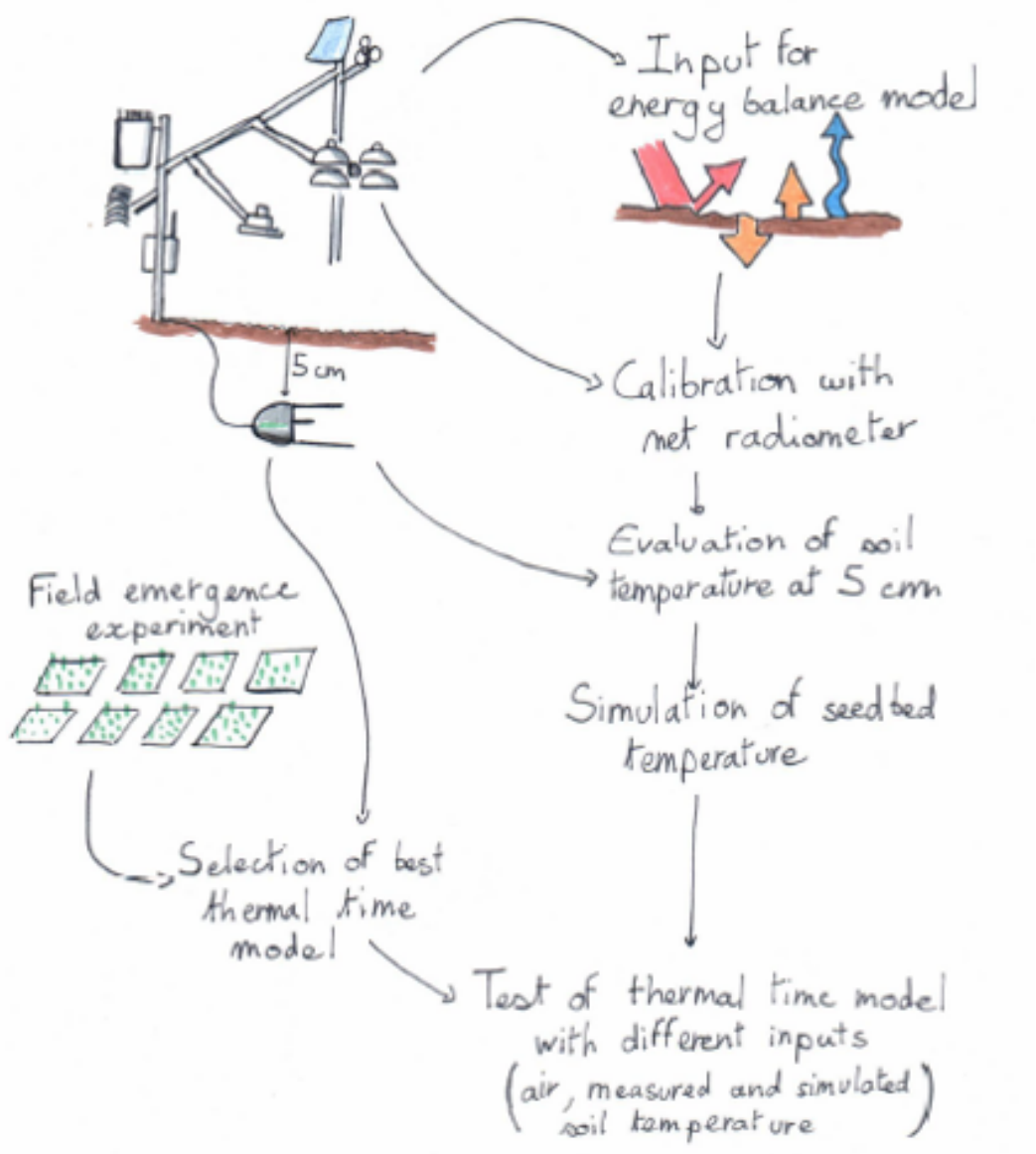

FIGURE 4.1: This sketch summarises the material and method of chapter 4. The weather station provides the input for the soil model. The soil temperature sensor at 5 cm is used to select the best emergence model and evaluate the soil temperature simulation. Then we combine simulation of soil temperature and thermal time model to predict emergence for four sowing dates in 2016.

where T_i is the temperature of the organism and T_b is the base temperature, an hypothetical temperature that extrapolates the linear part of the development model to null development (Bonhomme, 2000).

Our assumption, justifying the work presented in this part, is that the underground growth period of a plant do not obey the same rules and dynamics as the aerial one. In fact, a germinating and growing seedling does not have the same limiting temperature than a photosynthetically active young plant (Brandolini et al., 2000). Therefore it is assumed that not only the model parameters (threshold values) but also the model structure is to be designed and evaluated. To this end, one needs the widest possible range of different temperature conditions and accurate measurements of the environmental situation of the organism (Bonhomme, 2000). Hence, maize seeds were planted at various time of the year to generate different temperature conditions. Maize was planted close to the weather station described in chapter 1 (figure 1.1, page 22).

Environmental conditions The environmental conditions were recorded with an automated weather station (Adcon Telemetry, Austria) equipped with standard meteorological sensors (temperature, moisture, wind speed and rain gauge at two meters, global radiation at 1.5 meters) and soil temperature and humidity sensors (Hydraprobe II from Stevens, USA) located at 5, 30 and 140 cm below the ground surface. The weather station additionally included a net radiometer (NR01 from Hukseflux, The Netherlands) in order to measure the radiative term of the energy balance equation. The data generated by the sensors were aggregated hourly.

The hourly soil temperature at 5 cm was used as the temperature of the seed bed to select the best thermal time model for the prediction of crop emergence (figure 4.1)

Emergence The planting work and count of emerged seedling was carried out entirely by the author of this dissertation. Five meters from the weather station, four commercial cultivars of maize (*Zea mays*) 'NK Falkone', 'NK Cobalt', 'SY Multitop' (Syngenta France S.A.S) and 'NK Famoso' (Syngenta Italia S.p.A) were planted in bare soil at different sowing dates to obtain contrasted conditions for emergence. These cultivars are early varieties used for grain harvest. The recommended geographic distribution zone of 'NK Famoso' is southern Europe while 'NK Falkone' is a northern Europe cultivar. 'NK Cobalt' and 'SY Multitop' are not latitude-specific. The soil was prepared before each sowing dates with a mechanical soil tilling by the field station team and flattened with a rake by the author. The soil is a clay loam, homogeneous

soil with texture sand: 36.3 %, silt: 36.7 %, clay: 27 % and 3 % of organic matter at 5 cm.
For each cultivar and each sowing dates, 121 seeds were sown within $50 \times 50 \times 4$ cm (length $\times$ width $\times$ depth) metal frames. The metal frames were assembled at the on-site workshop by external partners. Then the seed were covered with the same soil, sieved with a 0-10 mm sieve. Four cm is the usual sowing depth in the region. Each cultivar was replicated 4 times at each sowing date. Three litres of water were poured over the planted seeds to give all the replications the same start for germination. These three litres correspond approximately to a 1.2 mm rain, which is negligible for the soil water balance. Newly emerged plant were marked with a coloured marker every day to obtain the cumulative emergence curves. The sowing dates were, in 2015: 24 April, 12 May, 10 June, 2 July, 14 August, 10 September, 12 October and, in 2016: 7 April, 4 May, 19 July and 22 September. On 19 July 2016, only 'NK Falkone' was planted because, instead of testing several cultivars, we tested seed treatments (results will not be shown). The results of emergence for 19 July taken in this study is the emergence of 'NK Falkone' without treatment. On 22 September, we tested several sowing depths with 'NK Famoso' (results will not be shown) and, in our present analysis, we consider the emergence of 'NK Famoso' planted at 4 cm.

2.2 Thermal time models

Although VE is formally defined as a phenological stage at individual scale, we are generally interested by its distribution within a given population. In practice, the definition of VE corresponds to the time when 50% of the final number of plants is reached. Its practical characterisation therefore relies on measurements done at a population level. The cumulated number of emerged plants is usually modelled with a transition function such as the logistic one. This function is a relevant choice in our context since the underlying parameters follow a normal distribution, which is what is assumed when working with commercial maize cultivars where there is neither genetic variability nor seed-to-seed variability. The equation of the logistic giving the number of emerged plants E w.r.t time t is:

$$E(t) = \frac{a}{1 + e^{\left(\frac{-4b \cdot (x - VE)}{a}\right)}} \tag{4.2}$$

where a is the maximum cumulative emergence proportion, b the maximum slope and VE the time when 50% of the emergence is completed. The parameters b and VE were estimated by fitting against the measurements of emergence, expressed in hours after sowing; a was directly fixed as the maximum counted emergence because not 100 % of the plants emerged all the time.

Prediction of VE Our objective is to predict the value of VE, the time to emergence, based on the thermal time approach. We have chosen to test three thermal time models with respectively one, two or three parameters corresponding to the base temperature T_{base}, the optimal temperature T_{opt} at which the rate of emergence is maximum and the maximum temperature T_{max}, after which the emergence rate is null. The input variable of the three thermal time models is T_i, the hourly soil temperature at 5 cm at time i. We write the accumulation of thermal time, TT_i, with a recurrence relation $TT_i = TT_{i-1} + \Delta TT_i$ with

$$\Delta TT_i = \begin{cases} \left.\begin{array}{l} T_i - T_{base} \end{array}\right\} & \text{Model } TT_1 \\ \left.\begin{array}{l} T_i - T_{base} \text{ if } T_i \leq T_{opt} \\ T_{opt} - T_{base} \text{ if } T_i > T_{opt} \end{array}\right\} & \text{Model } TT_2 \\ \left.\begin{array}{l} T_i - T_{base} \text{ if } T_i \leq T_{opt} \\ T_{opt} - T_{base} \cdot (\dfrac{T_{max} - T_i}{T_{max} - T_{opt}}) \text{ if } T_i > T_{opt} \end{array}\right\} & \text{Model } TT_3 \end{cases} \tag{4.3}$$

2.3 Soil temperature simulation

The method and results of the simulation of the soil temperature for the different sowing dates are presented in the section 3.2 in chapter 3.

2.4 Model selection procedure for cultivar specific thermal time model

When expressed in calendar time, the emergence curves are visibly different for different sowing date, even for the same genotype, because the associated recorded temperature sequences are different. Our methodology relies on the idea that there exists a function of the recorded temperature sequences that entitles reduction of this variability of emergence trajectories of a given genotype sown in various conditions, once expressed with this new thermal time unit. It can be expected that if such a function is identified, it might be considered as a genotype-specific attribute.

We chose to characterise the variability of emergence by the VE values (parameter VE of the logistic curve), that is, the time to 50 $\%$ of emergence (hours). The procedure was therefore to select the model that minimises the coefficient of variation of VE computed over the seven emergence curves, corresponding to the seven sowing dates of the year 2015. Since the problem did not fit within the standard frame of model parametric estimation with regards to a set of experimental data but corresponded to the minimisation of a particular objective function (VE variability), the usual AIC/BIC penalised criteria could not be applied for this model comparison. Therefore the adopted selection procedure was a leave-one-out cross-validation scheme, for each cultivar independently. More precisely, it consisted in iteratively optimising the parameters of the thermal time model (T_{base}, T_{opt} and T_{max}, finding a common value for VE on six sowing dates and computing an absolute error with the remaining seventh sowing date. Each planting date was chosen in turn to be left aside.

For the optimisation, the parameters were given a lower boundary of 0 °C and upper boundary of 50 °C for the cardinal temperatures. At the end of the iterations, averaging over the 7 iterations provides a mean absolute error:

$$MAE = \sum_{1}^{7} \frac{|VE_{sim} - VE_{obs}|}{7} \tag{4.4}$$

The model with the smallest MAE was declared the best model for the given cultivar. It was then evaluated against the data of 2016, that was not used in this selection procedure.

3 Results

3.1 Emergence experiment

Soil conditions during emergence In 2015 and 2016, we planted maize seeds at different dates to get different conditions between sowing and emergence within a short time span. The 11 planting dates spanned temperatures from below 10 °C up to 35 °C, whereas the volumetric water content in the soil ranged from 0.15 $\mathrm{m}^3\,\mathrm{m}^{-3}$ to almost 0.35 $\mathrm{m}^3\,\mathrm{m}^{-3}$. In Figure 4.2, the interquartile ranges (IQR) of soil temperature at 5 cm and soil water content at 5 cm were plotted for each period from sowing date to observed VE. The intersection of the segments is the median for both variables. The 11 planting dates cover

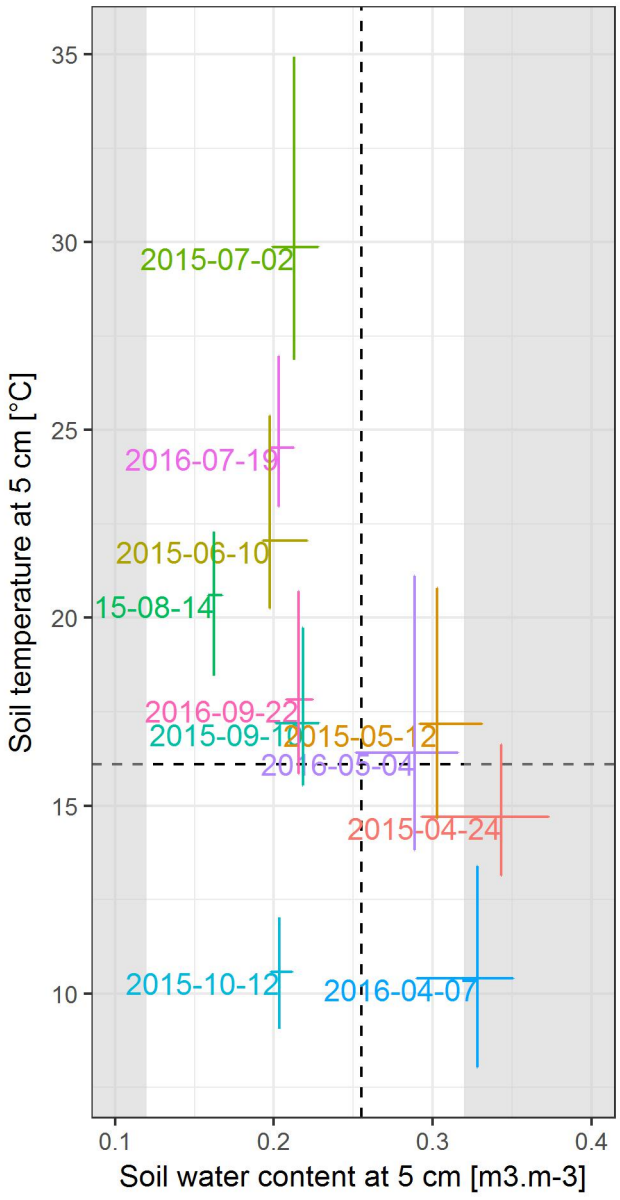

FIGURE 4.2: Soil temperature and soil water content between sowing and emergence (VE) for 11 planting dates in 2015 and 2016. The experiment took place in the same location in north-western Switzerland, with sowing at a different time of the year to generate variability in the conditions. The bars represent the respective interquartile ranges for soil temperature (vertical) and soil water content at 5 cm (horizontal), and their intersection is positioned at their medians. The dashed lines symbolise the averages for soil temperature and water content among all the sowing dates. For a better interpretation of the moisture content, we greyed out zones below the permanent wilting point (0.12 $m^3\,m^{-3}$) and above field capacity (0.32$m^3\,m^{-3}$) measured at this location.

a broad range of temperature, with sowing dates in April 2016 and October 2015 exhibiting temperatures even below the usual 6°C threshold (the base temperature for maize growth in Europe, based on air temperature (Durand *et al.* 1982)). The warmest temperatures, experimented during emergence in July 2015, were above 30°C, which nevertheless stays below the maximal recorded temperature for maize growth : 38 °C (Hatfield and Prueger 2015).

Concerning water content, there were two sowing dates with high moisture, in April 2015 and 2016. At these two dates, more than 25 % of the measurements were above the field capacity, and the soil could not absorb any more water. Figure 4.2 shows that for most of the sowing dates, high temperatures are associated with low moisture and reciprocally. Only exception to this trend : the cold and dry conditions recorded when maize was planted on 12 October. Nor warm and wet conditions, neither very dry ones (below 0.15 $\mathrm{m}^3\,\mathrm{m}^{-3}$), were encountered. This last observation lets us assume that water was not a limiting factor for emergence in our experiments.

Variability of emergence The variability of environmental conditions generated different cumulative emergence curves and a different value of VE that we will use to find the parameters of the thermal time models. In figure 4.3 we show, for each sowing date, the emergence curves (in percent of total sown seeds), the average time to VE in days and the standard deviation σ in hours. VE was obtained by fitting the logistic curve to the data points of each cultivar and each sowing date. In average, VE varies from 4.2 days in July 2016 to 26 days in October 2016. The maximum emergence (100 %) was not part of the fitting since it was not achieved every time, as in July 2015, October 2015 and September 2016.

On certain sowing dates, the difference between cultivar is more visible than in others. For local standard planting dates such as April 2015, May 2015 and May 2016, the cumulative emergence curves are close to each other, with 'NK Cobalt' demonstrating faster emergence in May, although less pronounced in 2016. The sowing dates in August and September 2015 generated steep development curves with a tiny standard deviation among cultivars. The conditions for these planting dates seem optimal for maize emergence, with a median temperature between 17 and 20 °C.

Emergence curves from sowing dates with high temperatures (June 2015, July 2015 and 2016), show a more highly contrasted behaviour in cultivars. In June 2015, 'NK Famoso' displayed a fast emergence but a low final emergence count. On the other hand, in July 2015, 'NK Famoso' and 'NK Falkone' exhibited a slow emergence and a lower final count than 'SY Multitop' and

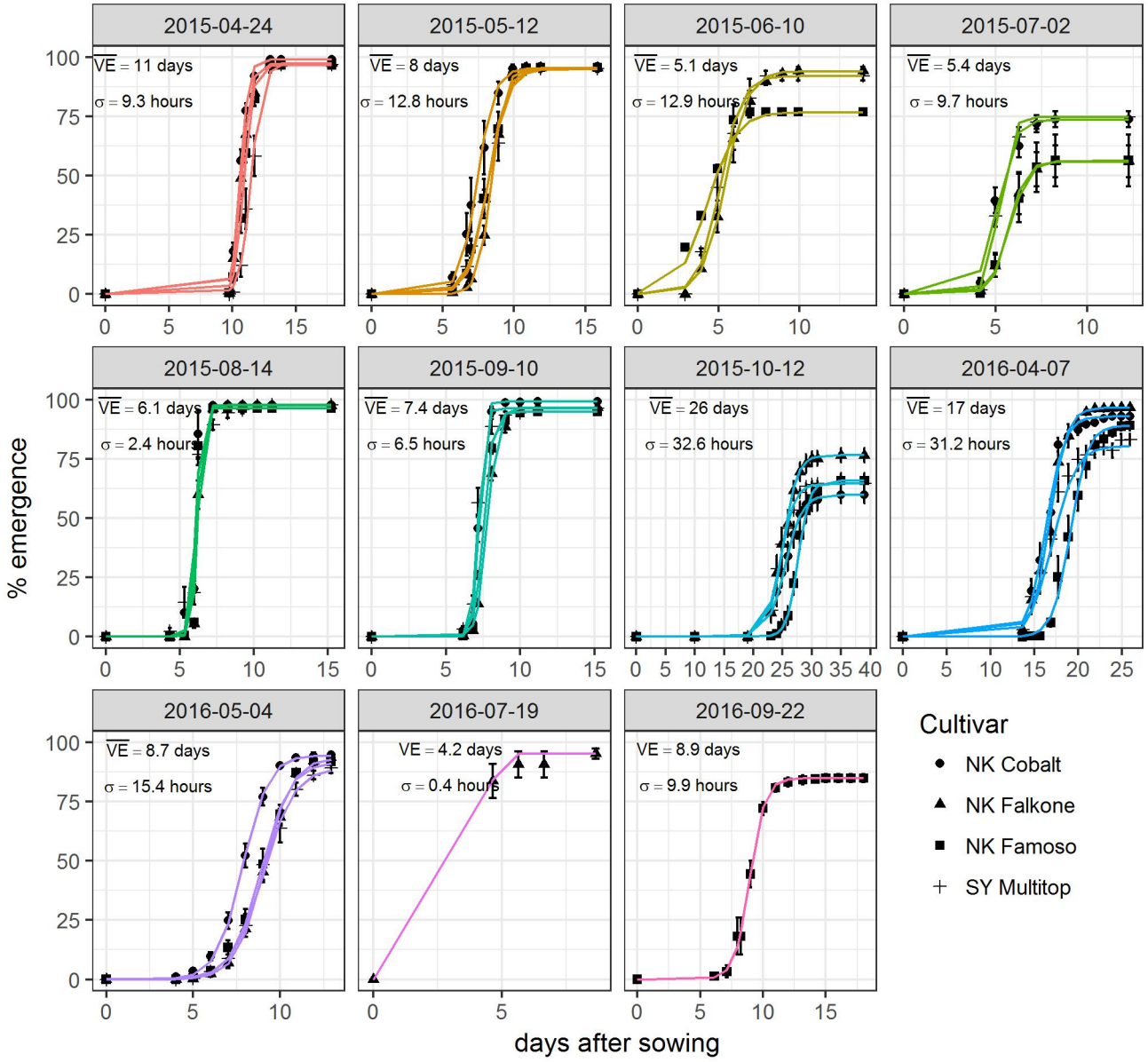

FIGURE 4.3: Cumulative emergence percentage for the 11 sowing dates and the four cultivars (dots). For each planting date, the average VE in days (for all varieties planted at this time) and its standard deviation σ in hours are indicated. The error bars on the dots correspond to the standard error of the four replications for each cultivar. A logistic function was fitted to the emergence data of each cultivar. Only 'NK Falkone' emergence was recorded in July 2016 and 'NK Famoso' in September 2016.

'NK Cobalt'. It appears that they suffered more deeply from the heat than other cultivars did. On cold sowing dates (in April 2016 and October 2015), 'NK Famoso' exhibited the longest time until emergence. It is interesting to see that, even with similar temperatures for both cold sowing dates, the time to emergence was higher and the maximal count was lower in October 2015. The quantity of available water might explain this difference, as we noticed that wet conditions in April 2017 stimulated emergence. The experimental variability in emergence dynamics obtained by our protocol seems satisfactory for our modelling objectives.

3.2 Thermal time models for emergence

The next step is to find the best thermal time model that, when applied to the conversion of VE values from calendar to thermal scales, would reduce as much as possible its coefficient of variation. To choose the best model among the three proposed thermal time models TT_1, TT_2 and TT_3, we ran a leave-one-out cross-validation scheme and obtained the prediction error for each leave-one-out iteration. From this, the mean absolute error (MAE) was finally computed for each model (table 4.1).

TABLE 4.1: Mean absolute error (MAE $\pm$ standard error) of model prediction during the iterative leave-one-out cross validation scheme. For each cultivar, the best adapted thermal time model is the one with the lowest MAE

Cultivar	MAE TT_1 [h]	MAE TT_2 [h]	MAE TT_3 [h]
NK Cobalt	$70(\pm 40.7)$	$5.6(\pm 1.8)$	$15.7(\pm 6.6)$
NK Falkone	$60.2(\pm 22.8)$	$29.9(\pm 16.4)$	$12.6(\pm 4.8)$
NK Famoso	$75.4(\pm 41.7)$	$22.8(\pm 8.5)$	$14.9(\pm 5.2)$
SY Multitop	$48.5(\pm 26.3)$	$20.6(\pm 3.6)$	$12.9(\pm 3.2)$
Average	63.5	19.7	14

The increasing complexity of the thermal time models reflects the degree of linearity of cultivar emergence response to soil temperature: a high linearity for the simplest model, TT_1 and a peak response to temperature for the model TT_3. TT_1 is the classical and most used model applied to thermal time computation, known to be valid only for a certain range of temperature (Sharpe and DeMichele, 1977).

Since in our experiments, temperate-climate maize cultivars were sown

in summer, we faced situations outside the range of temperatures where the simple thermal time model for maize is valid. And indeed, the model showing the lowest MAE is found to be model TT_3 in average: it is the most suitable model for three out of four cultivars, making it the model of choice if we would have to predict the emergence of an unknown variety. It has to be noted however that the standard deviation values indicate a considerable variability among iterations, especially for model TT_1.

For one cultivar only, 'NK Cobalt', the selected model is not TT_3 but TT_2, with an MAE of 5.6 hours. This difference with the other cultivars may be an artefact linked with a missing observation of Cobalt emergence on June 10th, 2015, so that the optimisation was done on six sowing dates only instead of seven. Nevertheless, the error of model TT_3 for 'NK Cobalt' (15.7 hours) is in the same range as the MAE of model TT_3 for the other cultivars (respectively 12.6, 14.9 and 12.9 °C for 'NK Falkone', 'NK Famoso' and 'SY Multitop'). This observation leads us to the conclusion that model TT_3 better describes the emergence in our conditions than do the other models, especially since it takes into account the effects of warm sowing conditions.

In terms of parameter values, the three proposed thermal time models gave consistent ranges of values for T_{base}, T_{opt} and T_{max}, whichever the genotype: T_{base} ranged from 4.9 to 8.4 °C, T_{opt} from 22.8 to 31.3 °C and T_{max} from 35.8 to 44.5 °C (table 4.2). In model TT_3, 'NK Cobalt' has the highest T_max of 44.5 °C, suggesting that it is the most tolerant cultivar to high temperature for emergence. This is consistent with model TT_2 having the lowest MAE for 'NK Cobalt', since this model does not take into consideration any limitation of the emergence rate at elevated temperatures (above 27.9 °C) and therefore also reveals a high heat tolerance. The opposite behaviour can be found with 'NK Famoso', which displays the narrowest linear response and the sharpest peak response with the highest T_{base} (8.2°C) and the steepest decrease from T_{opt} to T_{max} with, respectively 31 and 35.8°C (figure 4.4). The slope of the reduction in emergence for high temperatures is more than twice that of the two other cultivars with the same model. Concerning the required sum of growing degree hours to reach VE (GDH to VE, table 4.2), 'NK Famoso' is the fastest whereas 'NK Falkone' is the slowest, needing on average 200 °C h more to reach VE, whichever the model. The thermal time model TT_3 captures well the various behaviours at high temperatures but fails to highlight differences at base temperature, which sits between 7 and 8°C.

This comparison of parameter values highlights the genetic variability of emergence sensitivity to soil temperature. 'NK Cobalt' appears as a robust

TABLE 4.2: For each cultivar, values of the parameters for each thermal time model ± standard error. The value of the parameter is the average of the leave-one-out cross validation process (average over seven iterations). GDH to VE are the sum of the growing degree hours required to reach the VE stage (continued below)

	Model TT_1	
Cultivar	T_{base} [°C]	GDH to VE [°C h]
NK Cobalt	$5.4(\pm1)$	$2479.3(\pm178.7)$
NK Falkone	$4.9(\pm0.8)$	$2696.6(\pm157.9)$
NK Famoso	$5.7(\pm1)$	$2477.3(\pm178)$
SY Multitop	$5.6(\pm0.5)$	$2476(\pm104.7)$

	Model TT_2		
Cultivar	T_{base} [°C]	T_{opt} [°C]	GDH to VE [°C h]
NK Cobalt	$7.5(\pm0)$	$22.8(\pm0.1)$	$1853(\pm9)$
NK Falkone	$7.4(\pm0.2)$	$23.2(\pm1.3)$	$1923.9(\pm48.9)$
NK Famoso	$8.4(\pm0.2)$	$22.9(\pm1.4)$	$1677.4(\pm49.8)$
SY Multitop	$7.5(\pm0.4)$	$24.7(\pm1.4)$	$1913.1(\pm83.4)$

	Model TT_3			
Cultivar	T_{base} [°C]	T_{opt} [°C]	T_{max} [°C]	GDH to VE [°C h]
NK Cobalt	$7.3(\pm0.1)$	$27.9(\pm0.9)$	$44.5(\pm1.6)$	$1911.7(\pm17.7)$
NK Falkone	$7(\pm0.1)$	$30.2(\pm0.5)$	$39.1(\pm0.7)$	$2086.2(\pm17.8)$
NK Famoso	$8.2(\pm0.1)$	$31.3(\pm0.4)$	$35.8(\pm0.9)$	$1775.3(\pm30.5)$
SY Multitop	$7.3(\pm0.2)$	$30.6(\pm0.3)$	$39.6(\pm0.2)$	$1995.1(\pm39.8)$

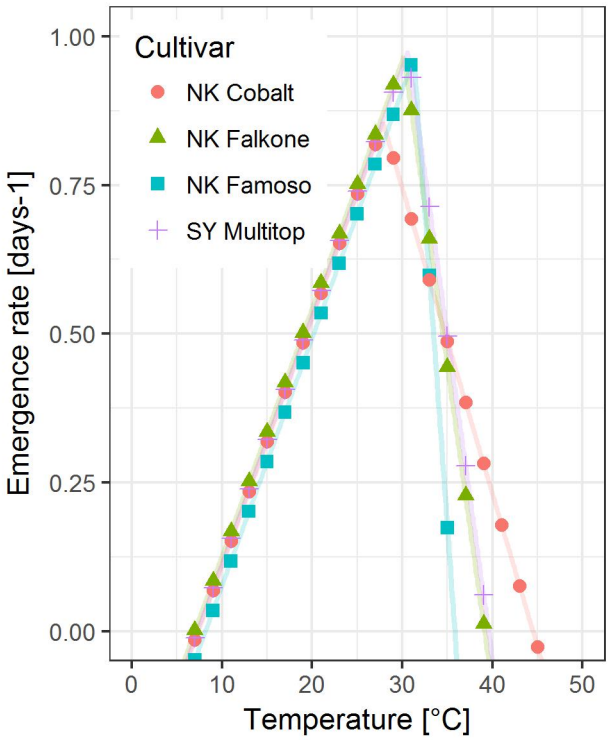

FIGURE 4.4: Response of emergence [days^{-1}] to temperature with model TT_3 for the four cultivars after parameter estimation with the leave-one-out cross validation scheme. The cultivars have a similar response with small differences at base temperature but larger one at maximum temperature.

cultivar for a brand range of soil temperatures, especially at high temperatures. 'NK Famoso' would be a sensitive variety both for low and high soil temperatures regarding its highest T_{base}, whichever the emergence model and the strongest decrease in emergence rates, for temperatures over 31°C.

3.3 Combination of soil simulation and adapted emergence model

To further test the performance of the model TT_2 for 'NK Cobalt' and model TT_3 for the other cultivars, we used the parameters in table 4.2 to predict the date of emergence for sowing dates in 2016 that were not used for parameter estimation and model selection. The four sowing times in 2016 spanned both low and high-temperature conditions (figure 4.2), hence providing reliable evaluation tests. Table 4.3 presents the prediction of VE obtained respectively with the recorded soil temperature (considered as a reference), with the soil temperature simulated by our soil model (calibrated version "ABCD") and with the recorded air temperature. Air temperature, as a widely available measurement, is in fact classically used as a proxy for soil temperature to predict phenological stages in many models. Using air temperature is sometimes the only possibility when neither soil temperature nor a soil temperature model is available. It can be thus considered as a baseline for evaluating the benefits of our modelling approach compared to this classical approach.

Ninety % of the time, simulated soil temperatures improve the prediction of the time to emergence compared to air temperatures. On one day only, on 22 September, did the use of air temperature as an input to the thermal time model yield a better prediction than soil temperature. Overall, with simulated soil temperatures, the relative error is negative. We over-predict the date of emergence for all sowing dates and cultivars, excepted in September.

In 40 % of the cases, the simulated soil temperature also gives better results than the recorded soil temperature at 5 cm. For a standard sowing date, 4 May 2016, simulated soil temperatures generated relative error lower than 10 % compared to observations, NK Cobalt excepted - it showed error over 20 %. For this sowing date, NK Cobalt was the fastest to emerge and the thermal time model did not manage to properly capture this behaviour.

For all the sowing dates in 2016, the error on soil temperature prediction sits between 11 and 15 % (figure 3.12, page 152). The error on soil temperature is largest for 7 April (15 %) and it is also the temperature for which the error on emergence prediction is the greatest. For 7 April, the date of emergence

TABLE 4.3: Comparison of observed time to emergence (VE_{obs}) in days and the time to emergence predicted VE_{sim} by the thermal time model for 4 sowing dates in 2016. As input of the thermal time model, we used either recorded or simulated soil temperature at 5 cm or air temperature. The Δ is the relative difference between VE_{obs} and VE_{sim} ($\Delta_{obs} = \frac{VE_{obs} - VE_{sim}}{VE_{obs}} \cdot 100$)

			Soil temperature at 5 cm				Air temp. at 2 m	
			Recorded		Simulated		Recorded	
Sowing date 2016	Cultivar	VE_{obs} days	VE_{sim} days	Δ_{obs} %	VE_{sim} days	Δ_{obs} %	VE_{sim} days	Δ_{obs} %
7 April	NK Cobalt	16.3	17	-4.3	25.3	-55.2	28.1	-72.4
	NK Falkone	16.7	17.2	-3	25.1	-50.3	28.1	-68.3
	NK Famoso	19.1	22	-15.2	27	-41.4	29.2	-52.9
	SY Multitop	17.1	19.9	-16.4	25.6	-49.7	28.6	-67.3
4 May	NK Cobalt	7.8	7.3	6.4	9.5	-21.8	11.9	-52.6
	NK Falkone	9.1	7.2	20.9	9.8	-7.7	12.9	-41.8
	NK Famoso	8.9	6.8	23.6	9.6	-7.9	13	-46.1
	SY Multitop	9	7	22.2	9.7	-7.8	12.9	-43.3
19 July	NK Falkone	4.2	5	-19	5.5	-31	5.8	-38.1
22 September	NK Famoso	9.1	7.3	19.8	8	12.1	9.5	-4.4

is over-predicted by an average of 50 % of the observed emergence date. To conclude, using our model to simulate soil temperature is advantageous for emergence prediction, compared to the use of air temperature.

The next paragraph further investigates which level of accuracy in the simulated soil temperature is necessary to get acceptable prediction performance for VE.

3.4 Uncertainty propagation in the thermal time models

To better understand the effects of soil temperature uncertainties on the prediction of VE, we generated different sequences of simulated soil temperatures from reference ones that were created using 24 hours-periodic sinusoidal variations around a constant mean value that ranged from 8°C to 40°C. For each reference sequence, the different biased sequences were simulated by systematically adding respectively 0.1, 1, 2, 3 and 4 °C to all the terms of the sequence. Figure 4.5 presents the relative errors obtained on VE (error on VE divided by the reference VE) with model $TT_3(T_{base} = 7.4, T_{opt} = 30, T_{max} = 40)$ for these different sequences, for each value of average temperatures. We considered as acceptable a relative error of less than 10%, representing an error of about two days for the prediction of emergence for typical VE value of 20 days. Interestingly, the curves display a 'U' shape, showing that soil temperatures close to the base or maximal temperatures mechanistically imply large error levels, even with small input errors on these simulated temperatures. As can be seen in Figure 4.5, a simulated error on temperature of 0.1 °C gives a relative error for emergence prediction that stays below the 10% threshold. However, a simulation error of 1 °C, which would still appear a reasonable modelling performance, results in a threshold that is only respected between 16 and 33 °C and an error increased up to 20 % outside this interval. Finally, when the error of the model exceeds 2 °C, the threshold is never met. This simulation helps us to visualise that even with a very accurate soil temperature prediction, the error on VE will consistently be significant in extreme temperatures.

The error comes in fact from the linearity of the thermal time model itself and can be mathematically justified. Let us simply take into account the first linear part of the model, between T_{base} and T_{opt}. We set a sequence of constant temperature $T(t) \geq T_{base}$ and a constant bias of ΔT. We set d_1 as the day of emergence with sequence $(T(t))_{t=1...\infty}$ and d_2 the day of emergence with $(T(t) + \Delta T)_{t=1...\infty}$. Note that $d_2 \leq d_1$ since the bias ΔT is positive and

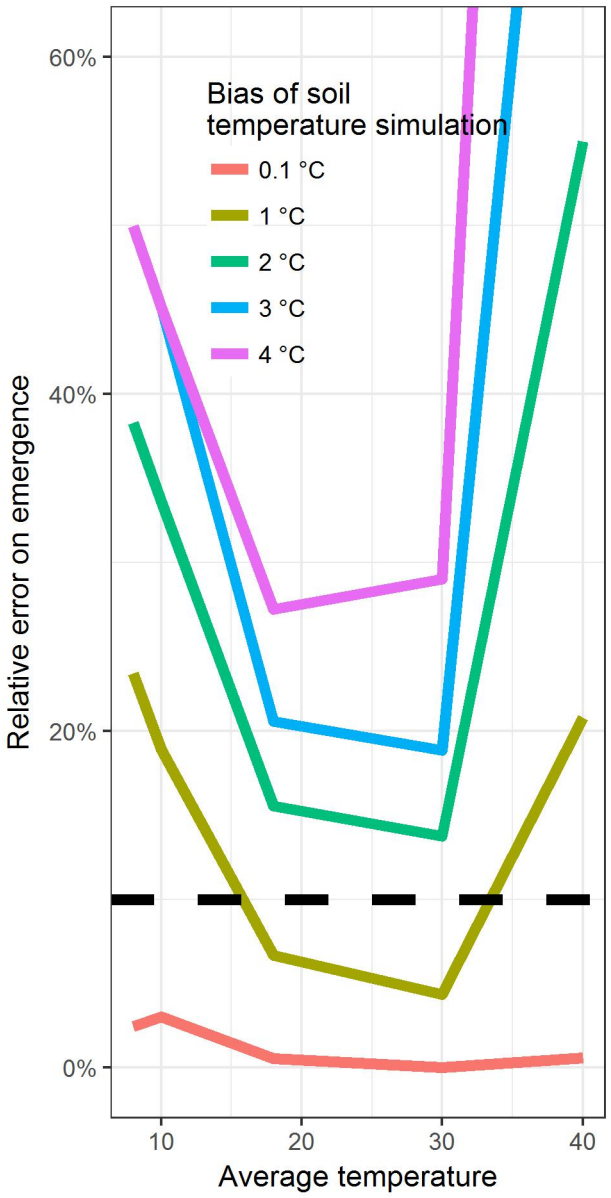

FIGURE 4.5: Impact of soil temperature simulation error on emergence prediction. An error of 0.1, 1, 2, 3 and 4 °C was successively applied to artificial reference sequences of temperatures with diurnal variations around 8, 10, 18, 30 and 40 °C averages. The closer the average temperature is from cardinal temperature T_{base} (here fixed at 7.4 °C) and T_max (here fixed at 40 °C), the larger the impact of the error on the prediction of VE. The error is a percentage of the total duration from sowing to emergence. An acceptable error for prediction would be approximatively 10 %, as represented by the dashed line

emergence stage is reached faster in hotter conditions. Expliciting the equality of sum of degree days required to reach the same emergence stage with the two temperature conditions gives, using the integral definition of thermal time:

$$\int_0^{d_1} \max(0, T(t) - T_{base})dt = \int_0^{d_2} \max(0, T(t) + \Delta T - T_{base})dt$$

With our assumption that $T(t)$ is constant and above T_{base}, it implies:

$$d_1(T - T_{base}) = d_2(T + \Delta T - Tbase)$$

Introducing the notation $\Delta d = d_1 - d_2$, we get:

$$\Delta d = \frac{d_1 \Delta T}{T + \Delta T - T_{base}}$$

which obviously shows that Δd tends to infinity when T tends towards $T_{base} - \Delta T$ This proves that, as soon as soil temperature approaches from the base temperature, the model becomes very sensitive to the bias of the soil temperature model. Even with an extremely accurate soil temperature model, a little bias mechanistically causes a large error on prediction if the environmental conditions are within these extreme zones, as was graphically illustrated in figure 4.5.

4 Discussion

In this chapter, we combined a soil temperature model with a simple emergence rate model to improve the prediction of emergence in the field for maize cultivars. The energy balance approach gives satisfactory levels of accuracy of the near soil surface temperature, which is the zone of interest with regards to seeds and seedlings. We have used hourly recorded soil temperatures to account for diurnal variation of soil temperature. However, we showed that, no matter how accurate the soil temperature model or emergence rate model, the error on predicting emergence will automatically be significant when we get closer to the base or maximal temperatures.

Emergence rate model A considerable variability in emergence timing was generated by sowing at different times, even largely outside the standard planting periods in the area (which spans from 15 April to 15 May in north-western Switzerland). Thus, maize seeds and seedlings were placed in stress

conditions, so that we might obtain the different cardinal temperatures that are unique to the sowing to emergence period. By sowing every month between April and October, we generated a range of temperatures spanning the range of development temperature of maize. In contrast, this objective was not achieved regarding the soil water content, with no sowing dates exhibiting soil moisture content close to the soil's permanent wilting point. Hence, water was not a limiting factor for these emergence times, and this is a reason why we chose to use a thermal time model and not a hydrothermal time model. Moreover, for spring wheat, Bullied, P. R. Bullock, et al. (2014) showed that the use of hydrothermal time model improved the prediction of emergence time only right at the soil surface when it is dry. As our seeds were planted at 4 cm depth, using a hydrothermal time model would only have a slight impact on our predictions, according to these authors.

Regarding the set of thermal time models considered in the selection procedure, only simple linear or piecewise linear models were chosen. The literature abounds with alternatives to this simple approach. For example, the relationship between emergence rate and temperature can also be modelled with a beta function (Jame and Cutforth, 2004; Bullied, P. R. Bullock, et al., 2014; Edalat and Kazemeini, 2014). These model proposals could straightforwardly be included into our selection procedure in further studies. Simple models present the advantage of rapid parameter computation. Since we are working with commercial cultivars, which typically exhibit a high germination rate and a normal distribution of seed to seed variability for weight and quality, the simple thermal time model grants comparison between cultivars without taking sub-population thresholds into account (S. Hardegree and Van Vactor, 1999).

Our entire study is based on the central assumption that the cardinal temperatures for the period from sowing to emergence are different from that of the aerial, heterotrophic growth. Previous works support this hypothesis (Jame and Cutforth, 2004; Bhosale et al., 2007) and highlight the physiological differences between a seed, a growing seedling using its endosperm reserve, and an autotrophic plant relying on photosynthesis. The enzymes involved in germination and seedling elongation are different from the enzymes used in autotrophic growth, with different activation energy or sensibility to heat (S. Hardegree and Van Vactor, 1999). To exploit the precision of the dataset, we chose to use an hourly integration of soil temperature while many other studies use daily average temperatures to compute the degree day unit. We acknowledge that the use of hourly temperatures can create different results

from daily averages (Bonhomme, 2000).

With the emergence recorded in 2015 and the recorded soil temperatures at 5 cm, we optimised three thermal time models to find the best thermal time model describing the period from sowing to the emergence and the value of parameters. We used the recorded soil temperatures to optimise the models in order to obtain only one set of parameters for the evaluation. An optimisation using air temperatures would not have made sense since we required cardinal temperatures linked to soil temperatures.

The optimisation of the thermal time model was done by minimising the coefficient of variation between observed VE and simulated VE. The method took advantage of the restricted number of situations (seven situations).

The simple logistic emergence model can capture the observed genotypic diversity and appears to be an interesting alternative methodology for phenotyping the cultivar sensitivity to chilling and heat during emergence. It relies on the powerful hypothesis that, for a given variety, there exists an intrinsic, probably genotype-specific, emergence dynamic curve that should be the same in any environmental conditions, once expressed in the appropriate thermal time unit. Here this methodology was focused on VE only (the objective was thus to minimise its coefficient of variation) but could further be extended to the whole emergence curve, with the difficulty that the plateau value (the maximum emerged seedling) would have to be modelled independently, as it depends on the environmental conditions (C Durr et al., 2001). Nevertheless, different patterns in the sensitivity of the maize emergence have been identified among the four cultivars, which have different maturity levels. 'NK Famoso' is a hybrid recommended in Mediterranean countries which was known to have a narrower allowable range of temperature for emergence than the other cultivars. The earliest cultivar is 'NK Falkone', which is recommended for Northern Europe; 'SY Multitop' and 'NK Cobalt' are favoured all throughout Europe. Even though the parameter values were purely statistical and did not correspond to real, physiological values, they might help us distinguish cultivars, provide interesting information on their emergence-related characteristics and guide the recommendation strategies.

Combination of thermal time model and soil conditions simulation For the last part of this analysis, the thermal time model found for each cultivar was combined with the simulation of soil temperature for an evaluation of the predictive capacity of the resulting model on 2016 datasets, not used in the calibration. The VE predicted using simulated soil temperatures is more accurate than when using air temperatures, last sowing day of 'NK Famoso'

excepted, on 22 September 2016, when air temperatures fared better than soil temperatures (both simulated and recorded). This is probably an artefact due to some other factors that might have influenced the time to VE on this sowing date.

The relationship between error in VE prediction is not linearly dependent on simulation of soil temperature error. We showed that, with (at least locally) linear forms of the thermal time model, a small error in soil temperature simulation has an enormous impact on the prediction of VE when temperatures are close to T_{base} (or to T_{max} for high temperatures). This effect was also reported with the use of a beta model (Jame and Cutforth, 2004). Similarly, in (S. P. Hardegree, 2006), the large prediction errors exhibited by their germination model near T_{base} led the authors to discard linear models for prediction of the emergence of weeds in the field. Therefore, a new research avenue might develop better-adapted models for emergence prediction for maize and other crops in cold spring temperatures, and for better understanding the adaptation of maize to cold conditions.

The use of energy balance models mainly stays restricted to biophysicists and bioclimatologists who have, for instance, developed such models to better predict the development of growing seedlings (Weaich *et al.* 1996) or of young maize apex (Guilioni *et al.* 2000). This chapter continued those studies by proposing a simplified energy balance model combined with a piecewise linear thermal time model. Its assets make it a practical tool for predicting emergence, which is known to be a critical stage for agronomists.

Conclusion and outlook

Soils are complex systems. Any given soil combines its inherently porous nature with physical and chemical interactions between solid, liquid and gaseous phases. Such complexity affects crop yields in ways that are still poorly understood. It is therefore crucial, for both public and industrial research, to develop a deeper knowledge of the relationship between soil processes and crop productivity. Such insight will drive innovation and new technology development for a productive and sustainable agriculture.

In this thesis, we focused on modelling soil temperature, as it plays a central role in regulating biological (microbial respiration), chemical (nutrient ad- and absorption) and physical processes (biogeochemistry) which in turn all influence crop growth and development. As a case study, we demonstrated that soil temperature simulator may constitute an efficient tool to better predict the emergence of maize.

1 Results

Soil temperature data analysis As a preliminary step before building the model, we introduced in the chapter 1 of this thesis a brief analysis of soil temperature variability with the help of soil temperature sensors installed at 5, 30 and 140 cm in the soil in Stein, Switzerland. The main results show that, during year 2015, soil temperatures at 5 cm under a bare surface were closely linked to air temperatures at $2\,\mathrm{m}$ (R2=0.91), with a regression coefficient of 1.6. However, this coefficient value is specific to 2015, our location and soil settings. In most places, data is not available to estimate the regression coefficient, which gives great value to a model capable of predicting soil temperature using only classical above-ground meteorological measurements.

In order to build a versatile model, we picked the energy balance approach, which uses the laws of physics to describe energy exchanges at the surface. Such a model is expected to adequately simulate soil temperatures over a wide array of environments and soils.

Model development Our model's development was inspired by the model described in Müller (1999)'s book. This particular book proved conducive to our project, since it also supplied the model's source code. Our model was christened SOPHIA for SOil PHysics In Agronomy, and the procedures undertaken to build it are described in chapter 2. This includes the equations, numerical implementation and preliminary simulations. The partial differential equations for heat and water were resolved with an explicit Euler scheme that was stable for a soil discretised in layers of 10 cm and a time step of 1 hour.

In addition, we have simplified the expression of the aerodynamic resistance in SOPHIA to test from the outset whether that complex part of the energy balance was necessary for obtaining satisfactory accuracy.

The soil parameters were gathered using the Saxton model (Saxton and Rawls, 2006), which infers water-retention and hydraulic conductivity at saturation parameters merely from soil texture and organic matter content. The Saxton model gave the different parameters of 12 texture classes, to which we tested the model's response. SOPHIA was able to consistently differentiate clay from sandy soils.

The model's development was concluded with a first evaluation against the values of soil temperature measured in Stein, Switzerland. For 10 days in April 2015, the average error was 2.2 $^{\circ}\mathrm{C}$ (29 $\%$) near the surface and 1.3 $^{\circ}\mathrm{C}$ (17 $\%$) at 145 cm. For soil water content, the error was 0.04 $\mathrm{m}^3\,\mathrm{m}^{-3}$ (11 $\%$) at 5 cm and 0.06 $\mathrm{m}^3\,\mathrm{m}^{-3}$ (15 $\%$) at 145 cm. SOPHIA had a lower error than Müller's model for soil temperature at 5 cm, despite the simplification of the aerodynamic resistance.

Despite these encouraging results, the error for temperature and water increased when longer period of time where evaluated. We suspected that error in net radiation caused error in the energy balance, resulting in lower soil temperature at night and excessive evaporation during the day.

Our version, not including simultaneous heat and water transport as well as a simplified version of the aerodynamic resistance, is also well suited to sensitivity and other analyses, as well as to communication purposes. This represents an important aspect in the context of our company, since this means it may be readily presented to, understood and hopefully adopted by its end-users.

Model improvement In order to improve the model, we first performed a sensitivity analysis to identify candidate parameters for calibration. The semi-global Morris sensitivity analysis, supported by the variance-based Sobol analysis, revealed that the parameters of aerodynamic resistance, roughness length and reference height held the most influence over soil temperature at 5 cm. The parameters appearing next in ranking were the parameters involved in the net radiation. All of those high-ranking parameters are involved in the energy balance at the surface and influence the most the temperature at 5 cm. The other, low-ranking parameters, are classified as soil parameters and do not impact the near surface temperature as much.

We further investigated the error on net radiation by examining closely both the reflection coefficient and the upward and downward long-wave radiation. We used all the hourly net radiation data of 2015 to calibrate and evaluate the improvement of net radiation. This high number of data points strengthened our calibration.

Concerning albedo, we borrowed an equation that makes this parameter into a function of the soil's water content from the most recent version of energy balance model HYDRUS-1D (J Simunek et al., 2013). However, using our own physical measurements, we demonstrated that said equation was valid only to describe the albedo when the soil was wetting. We developed another expression for the drying soil. As a result the error on albedo was reduced by 56 %.

In regard to long-wave radiation, setting the soil emissivity to 1, its maximum value, reduced the error for upward long-wave radiation by 44 %. Pertaining to downward long-wave radiation, the estimation of parameters and the introduction of a cloud correction factor reduced the error by 36 %.

These numerous model improvements needed to be evaluated on our variable of interests: soil temperature and moisture at 5 cm. As a result, the model combining all the improvements (SOPHIA version 'ABCD') was the one chosen for the final evaluation for a whole year simulation in 2016 and for sowing windows simulation in 2015 and 2016.

The evaluation of the model version 'ABCD' on whole 2016 yielded a relative error of less than 10 % for soil temperature and soil water content at 30 cm depth. Near the surface at 5 cm depth, the relative error on soil temperature was 15 % and 26 % for soil water content. The improvements made to the initial version of the model represented a relative error reduction of 28 % for soil temperature at 5 cm and 47 % for soil temperature at 30 cm. The focus of the improvement was the net radiation terms. For them, the reduction

of relative error compared to the initial version was 41 % for net radiation, 55 % for net short-wave radiation and 61 % for downward long-wave radiation.

Prediction of emergence In chapter 4, we have combined the soil temperature model with a cultivar-specific emergence model to describe this all-important stage in canopy establishment. Accurate predictions of emergence dates are crucial to the truthfulness of all plant growth models. Indeed, these models start their biomass accumulation algorithms on the day of emergence. Since this date turns out to be unavailable in most cases, it has to be predicted using air temperatures. Our soil temperature model produces more trustworthy emergence dates, which in turn could significantly enhance the accuracy of plant models' simulations. Besides, running the emergence model for past seasons will shed light onto the events that took place during the early development of a crop, especially the temperatures experienced by and the quantities of water available to the seedling.

We drew different ranges of temperatures by sowing at different times of the year and by comparing three model structures. The three models enclosed incremental numbers of parameters to describe the response of emergence to temperature. The most adapted thermal time model, as exhibited by our model selection procedure, turned out to be the one exhibiting three thresholds : a base, optimal and maximal temperature. This selection may be interpreted in consideration with the hot weather that best fits maize sowing but calls for a penalty in high heat conditions.

The model parameters, i.e. the thermal time thresholds, were assumed to be cultivar specific. When tested to predict emergence on our validation datasets, the model predicted emergence with less than 1 day of error in standard sowing conditions. However, the model error was 6 days for a sowing date at the beginning of April, when temperatures were still cold for maize, as some days were close to the base temperature.

We proved, by performing an analysis of uncertainty propagation (section 3.4, chapter 4), that this increased error finds its cause in the linear assumption behind the thermal time model concept. Near the base and maximum temperature, the relationship between developmental rate and temperature is not linear. As a consequence, the prediction of emergence close to base temperature will invariably be prone to error, even if the temperature is known with accuracy. An error of only 1 °C of prediction for a temperature near the base temperature (6 °C) will result in an error of 20 % in emergence prediction, whereas our analysis shows that the same error will cause only an error of

10 % in prediction for temperature above 16 °C. This seems an important consideration to highlight since it also concerns the thermal time models used in plant growth models (Bonhomme, 2000) or germination models (S. P. Hardegree, 2006).

The trend to sow earlier, in colder conditions, will affect emergence prediction based on thermal time. It will be uncertain, especially if simulated temperatures are the only available ones. This warrants the development of a new emergence model adapted to prediction of emergence in conditions near the base temperature. As we submit this manuscript, no such model has, to our knowledge, yet been proposed for maize.

2 Outlooks

Regarding systems as complex as the soil-plant system, which encompasses such multifarious processes over numerous scales, opportunities for further enhancements abound, especially when it comes to including additional functions to the models. However, one has to bear in mind that the aim of a model is not to build a complete representation of reality but to answer specific research questions and fulfill specific industrial needs.

In the context of our company's interests, the research question will stay focused on a better simulation of soil water content and temperature in crop models until emergence. Therefore, the associated model specifications are: keeping the adequate level of complexity, that is keeping it as simple as possible to still have good enough prediction performance but without including unnecessary processes except if they are needed and can be calibrated. A second specification is to have a model that uses as few inputs as possible, in particular using only standard meteorological measurements and simple soil description. The diverse analyses that we performed helped us in the process of iteratively designing the current version of our model: a satisfactory one although maybe not yet the final one since several ways of improvements still deserve to be explored.

We chose to categorise the potential model improvements into two categories: those that are necessary regardless of the future use of the model, and those that are needed for particular projects, be it hypothesis testing or a research question. Then a third part opens up the outlooks of the use of soil temperature and emergence modelling in global challenges.

2.1 Necessary improvements in the short-term

Semi-implicit scheme

We developed an explicit numerical scheme to resolve numerically our coupled system of non-linear partial differential equations. With a soil discretised into layers of 10 cm and a time step of 15 minutes, the simulations were stable across the weather conditions recorded in Stein, Switzerland. However, writing of an implicit scheme is necessary to achieve a reasonably fast simulation with smaller layers of soil. The objective would be to go down to roughly one minute for simulating periods of 60 days on a standard laptop (currently an Intel Core i5 @ 1.9 GHz, 8 GB ram). Such an improvement is also important in view of the application of the model to other locations and climatic conditions since it could happen that the Euler scheme diverge in some cases (the CFL condition is only a necessary but not sufficient condition due to the system non-linearity and our tests could not be exhaustive since the possible combinations of environmental conditions is infinite). Some promising steps have been undertaken in that direction, with developing and testing a semi-implicit scheme. Although this scheme is in a stage still too preliminary to be fully presented in this manuscript, it may be found in appendix C.

Variable texture throughout soil profile

The majority of agricultural soils have different layers in their profile due both to the pedogenesis process and to human activity. At the location of our experiment, the soil was homogeneous down to 60 cm. Therefore, we assumed a homogeneous soil. Improving the adaptability of our model to a wider range of soil profiles may involve that we endow it with the ability to describe the properties of each soil layer. All the soil models mentioned in this thesis (Müller, 1999; G. N. Flerchinger, 2000; J Simunek et al., 2013) exhibit this feature.

Integration of pedo-transfer functions

Evaluating our soil temperature and moisture model for other soil profiles in other locations is worthwhile. However, we will most likely not gain water-retention characteristic curves for these locations nor for all the layers of the profiles. A way to circumvent the problem is to implement directly pedo-transfer functions (PTF) into the main program. With this new feature, texture data will be the only ones needed to recover soil parameters. Pedo-transfer

functions are a set of simple operations and should not increase significantly the execution time of the model.

We have used the Saxton model to access the soil parameters. In the future, we might want to evaluate other pedo-transfer functions, such as the popular Van Genuchten-Mualem PTF (Patil and Singh, 2016), by checking their effect on soil water content simulation.

These three improvements are required to ensure that the model is both more robust and more flexible before it is applied in an industrial context. The outlooks for such industrial applications will be detailed below.

2.2 Outlooks for industrial applications

Breeding for early vigour

In a near future, our model, once completed and evaluated, could become a milestone in the breeding process, especially taking part in two aspects: (i) bridging the gap from the experimental laboratory to the field and (ii) refining the characterisation of the environment.

Concerning the first aspect (i), an important application of the soil temperature model would be to provide information about the patterns and gradients of soil temperature at locations of interest. Provided that we have access to the hourly atmospheric data for some locations, we could answer the following questions: what are the soil temperatures for these areas ? What are the maximum and minimum soil temperatures? How many days present a minimum temperature below a certain threshold?

One of the goals of the breeding industry is to select for cold-tolerant varieties. The kind of information mentioned above is critical when designing cold resistance screenings of new cultivars and new chemical compounds. When this information is missing, low temperatures that would never occur in the field or that the crop experiences only once every ten years might be tested in the lab. Soil temperature knowledge will support the design of laboratory experiments that reproduce realistic field conditions in greenhouses and climate chambers.

Our model will also assist the transposition of laboratory results to the field. For example, let's imagine a new maize hybrid that emerges two days earlier. Would this give an advantage to meet optimal condition later in the cycle? Quite the contrary, faster emergence could be prejudicial, since, in some regions, it might lead the plant to surface before a damaging frost. In

this regard, the model helps test virtually whether a new product might give an advantage or not in the field. The model might even help defining new research directions. For instance, the objective in breeding for cold-tolerance might be to select for a stable emergence of 10 days, which would ensure that the minimum temperature and moisture requirements are met when planting within a certain sowing window.
With reference to the second aspect (ii), the characterisation of the environment is important to breeders as it helps them relate the performance of a given variety (eg. time to emergence, yield) with environmental variables. Running backwards, with recorded weather data and known soil texture, our model reveals the soil conditions and temperatures that were experienced by the seedling. The simulation will be used to define the level of stress felt by the seedling. The level of stress would then be used to classify the trials. This classification will improve interpretation of the trial results. For example, the use of simulated soil temperature information could help trialists to understand the reason why the efficacy of a seed treatments vary by knowing the exact temperature of the soil around the active roots during the trials. The field trialist's toolbox for product selection (chemical compound or new variety) that is usually centered on crop performance, phenology and classic weather data, will be greatly enriched by soil temperature dynamics data.

The characterisation of soil temperature will also help reduce the number of trials needed to check the performance of the cultivars' stand establishment. This is a valuable prospect for our industry since trials are time-consuming and expensive. Identifying regions with similar soil temperature scenarios might diminish the number of trials necessary. Moreover, we could use the scenarios to endorse sales of the crop varieties suited to regions that display similar scenarios (Galinier, 2018).

As a tool to characterise the environment and design experiments that are faithful to actual field conditions, our model may be used straightaway in projects that improve the process of cold-tolerant varieties production.

Seed coating chemical availability for the plant

Another important application for foreseeable future would be to integrate our soil model into a larger model, that will describe the degradation of seed coating chemicals in the soil. A quickly-degrading chemical is safer for the environment but offers a less potent protection to the plant. Syngenta R&D

focuses on discovering chemicals that are as environmentally safe as they are efficient.

The key objective is to develop chemicals that are rapidly available to the plants but are degraded quickly in the soil. Chemicals degradation depends on soil microbial respiration, which in turn depends on soil temperature (W. J. Jones and Ananyeva, 2001). Our model is ideal to provide soil temperature information as an input to a soil carbon cycle model. The maximum number of days necessary for plant uptake may also be defined by adding a plant growth model. This will result in different scenarios of degradation speed and plant uptake. Endowed with this information, chemists may then improve the structure of the chemicals so they meet the desired degradation/uptake objectives.

3 General conclusion

The soil is a complex system that is impossible to overlook when studying the beginning of the crop cycle. The first steps of the crop cycle depend on soil temperature and water content near the surface, which are determined by the exchange of energy between soil and atmosphere.

This thesis presented the entire process that led to the development of a one-dimensional soil model that includes an algorithm of surface energy exchange. Through the major stages of the development process, we defined the system, equations and numerical scheme, we ran a sensitivity analysis to identify the important parameters, we estimated some of these parameters and made a final evaluation over several periods in 2015 and 2016.

As an example of application, we tested the model with a cultivar-specific thermal time model for emergence and proved that using a soil model increases the accuracy of emergence prediction in normal sowing conditions.

We listed some improvements that might be brought shortly to the model to enhance its reliability and versatility, as well as some longer-term developments that could make it suitable for new applications. The model needs now to be deployed at a larger scale in order to provide useful information to the industry and to contribute to more general open research subjects such as those regarding the consequences of climate change on soil temperature. It is worth mentioning that Y. Zhang et al. (2005) and Mellander, Löfvenius, and Laudon (2007) have studied the impact of changes in air temperature on soil temperature in northern ecosystems. They both identified that soil

temperature has a complex response to air temperature increase and that the change in snow cover affects the variability of soil temperatures in these area. These study could be extended to other ecosystems to gain a broader view on this issue.

The use of such models in agriculture is still in its infancy, mainly because of the complexity of the system under study. However, I believe that the interdisciplinary approach we adopted when building our model will be key in leading, step by step, to innovations that will benefit farmers, food production, environmental preservation and adaptation to climate change.

Appendix A

Glossary of soil modelling terms

A glossary is necessary to give clear definition of the numerous term used in chapter 2. For example, to express the quantity of water in the soil, we have different terms: soil water content, soil moisture, soil moisture content, water content, soil volumetric water content, volume wetness, volume fraction of soil water, water content mass basis.... and all of them characterize either a mass of water per mass of soil or a volume of water per volume of soil!

Bulk density (dry): The mass of a given soil volume after oven drying. It gives an indication of the porosity of the soil. The lower it is, the more pores are in the soil. A higher porosity means the that the water is conducted faster. Bulk density of a standard soil is around $1.3\,\mathrm{g\,cm^{-3}}$. Soil with high organic matter content have a lower bulk density (around $1\,\mathrm{g\,cm^{-3}}$) because organic material has a low density.

Darcy's law : named after a French engineer that first stated that water is flowing from low (close to 0) to high water potential (far from zero)

Fourier's law : the law that states that the heat flows from high temperatures to cold temperatures regions

Heat equation : the partial differential equation stating that the change of temperature in a solid over time is proportional to the change of heat flow over space.

Heat flow : a conductive transfer of heat between two adjacent layers of soil. The heat flows from hotter to colder regions according to Fourier's law. The heat flow is proportional to heat conductivity.

Heat flux : The transfer of heat in the energy balance equation, from soil to atmosphere. Heat is transferred under two important form, sensible heat (that you can feel) and latent heat (through water evaporation).

Latent heat flux : the flux of heat linked the change of state of a substance. Here we are speaking of the change of water from liquid to vapour, or evaporation

Layer : a 1-D layer of soil that is defined by its size z, a mid-layer point and the distance from the mid-layer point from the soil surface. In the current model development, all layers have the same texture.

Matric potential : The energy associated with water that is held in the soil matrix. In earlier soil water works, the water was assigned different discrete energy levels (adsorbed or capillary water, see figure A.1) but the more recent notion of matric potential is a continuous value best suited for water flow equations.

Moisture tension : equivalent to water potential, the term is used when water potential is measured with a tensiometer in the field

Precipitations : snowfall, hail, sleet, drizzle or rainfall, it is any product of the condensation of vapour in the atmosphere that falls with gravity. It's measurement requires specific equipment compare to rainfall.

Rainfall : an amount of liquid precipitations, measured by a rain gauge with a tipping bucket. The unit is $\mathrm{mm\,hour^{-1}}$ or more often $\mathrm{mm\,day^{-1}}$. In the Sophia model, we use rainfall and not precipitation to run the model.

Richards equation : the equivalent of the heat equation for water flow. Its the application of the mass balance to Darcy's law (Ross, 2003). It is more complicated than the heat equation because the hydraulic conductivity depends on water and there is an additional term linked to gravitational potential.

Sensible heat flux : the flux of heat that you can feel with your finger tips, for example the heat emitted by a radiator.

Soil matrix : the arrangement of the soil minerals and organic components and the pores within this arrangements, see figure A.1

Soil moisture : another term for soil water content, used as much as soil water content in the literature. It is mostly used in the expression "soil moisture characteristics", indicating the relation between water content and matric potential for the soil. Digression : soil moisture rhymes with soil temperature.

Soil water content : the content of water in the soil. It should be precise if it is the mass of water per unit mass of soil or the volume of water per volume of soil. In this dissertation, we will always refer to soil water content as the volumetric water content [$\mathrm{m}^3_{\mathrm{H_2O}}\,\mathrm{m}^{-3}_{\mathrm{soil}}$], but we will not always write "volumetric"

Soil wetness : the preferred term for soil water content used in Daniel Hillel (2003) because for the authors, soil water content contains the idea of an absolute amount of water in the soil whereas it is always a relative amount of water to a given volume, and hence wetness convey this is idea better. As it has not been widely adopted by the scientific community so far, we will keep soil water content in the dissertation.

Soil-moisture characteristic curve : its the graphical representation of the relationship between soil water content and water potential. It is used to compute the water potential from the water content.

Surface temperature : the temperature at the surface of the soil. The soil surface has no dimension. One can imagine it thin layer like a skin. Surface temperature is sometimes called skin temperature in biophysics reference book (John Monteith and M. Unsworth, 2013)

Thermal conductivity : the ability of soil to conduct heat. Conductivity increase with water content. It is calculated with equation 2.2

Volumetric water content : the volume of water per volume of soil, noted θ, this is the one we adopt in this dissertation, using volumetric water content or soil water content indistinctly

Water content, volume fraction : another name for volumetric water content, the volume of water per volume of soil as mentioned in Marshall, Holmes, and Rose (1996). It is opposed to water content, mass basis, which is the mass of water per unit mass of soil.

Water flow : the movement of water that follows the gradient of water potential and is proportional to the hydraulic conductivity of the soil

Water potential : it's called potential because its a form of energy that is in the water when it is in a porous media. It's the energy state of the water

Soil water in the soil matrix

In a porous media like soil, water flow (F_w $[\mathrm{m\,s^{-1}}]$) is driven by gradients of water potential (Ψ $[\mathrm{m}]$), a notion that translates how strong the water is held by the soil matrix. In a soil that is not saturated with water, that is to say an unsaturated soil that has 3 phases (air, water and soil), the water between soil particles is held with different forces, mainly the strong adsorptive and capillary forces (figure A.1 in appendix). Water potential is the sum of matric potential and gravitational potential, which is the energy stored in the water due to its position relative to the soil surface. In reality, the two adsorptive and capillary forces are difficult to distinguish and are bundled in the term matric potential. Matric potential is a measure of the strength of how much the water is linked to the soil matrix.

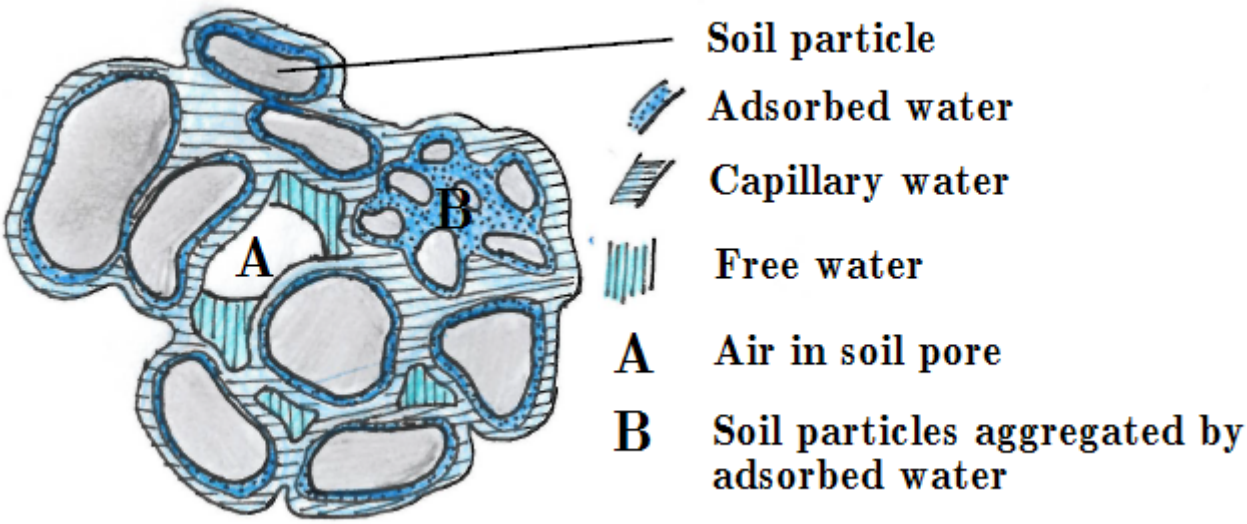

FIGURE A.1: Figure derived from Gaucher (1968) that shows the different way water is linked in a silty, unsaturated soil matrix. The capillary and adsorptive forces have been replaced by the concept of matric potential, that is stronger (more negative) when the soil gets dryer. Water potential is an indication on how difficult it is to extract the water from the soil. When the soil is getting dryer and dryer, the water is more and more difficult to extract due to the strong capillary and adsorptive forces, translated by the highly negative water potential

Appendix B

The R package 'sophia'

Introduction

The package is called sophia, and it stands for **SO**il **PH**ysics **I**n **A**griculture. The package is a collection of function that uses environmental and soil physics relationships to predict heat, water and energy movement for an agricultural soil with a bare surface (without vegetation). When I wrote this package, I had always a companion by my side. This companion is a book written by Christoph Müller called *Modelling soil-biosphere interactions* (Muller 1999). The book introduces the complex concepts of environmental physics by providing examples in a software called ModelMaker. Without this book and the provided examples, I would not have been able to write this package.

Objective of this package

The objective of this package was to write the model contained in Müller's book in order to perform sensitivity and uncertainty analysis on the model. R is a now widely use language and offer many possibilities for model analysis.

Objective of this vignette

The objective on this vignette is to give you a quick introduction on how to use the package's main function `soilflo`

Installation of the sophia package

To start, you need to have R install. I recommend to install along RStudio, a practical IDE (Integrated development environment) for R. The `sophia`

package comes in a zip `sophia_0.1.zip`. You can ask for a copy to Etienne Claverie (`etienne.claverie@syngenta.com`) or Jérémie Lecoeur (`jeremie.lecoeur@syngenta.com`).

Once RStudio is launched, you can start a new script and write in the top left window.

```
 # Change the path to where you have saved the zip
install.packages("C:/sophia_build/sophia_0.1.zip", repos=NULL)
```

Then press `ctrl + R` with the cursor placed anywhere on the same line as the `install.package` to execute the line and install the package in your R package library.

Once the package is install, you have to call it, like saying "Hey Sophia, I need to use you!", by executing

```
library(sophia)
# you can access this tutorial from within R by executing
# vignette("sophia-tutorial")
```

`soilflo` function

The main function of the `sophia` package is `soilflo`. Behind `soilflo` is a mechanistic model that simulates movement of heat and water within the soil. The objective of soilflo is to provide information on soil temperature and soil water content. It is working at the hourly or sub-hourly time scale. The soil profile is simplified as a 1-D layered system of a given depth (1 meter for example). Soilflo is not yet concerned with nitrogen, nutrient or carbon fluxes within the soil.

Actually the `soilflo` function is maybe the only one you will use as an external user. To get to know how to deal with the function, start with its `help`

```
help("soilflo")
```

Following is a more documented version of the examples in the `help("soilflo")` page.

Input

To execute the `soilflo` function, you need input data. Input of this model is the atmospheric conditions over a bare soil surface. There is an example dataset in the package `sophia`, called `weather_stein`. Here is how to load it into the workspace.

```
# Load the example of weather dataset
data(weather_stein_hourly)
# More information about this dataset and its content
# help("weather_stein_hourly")
```

Currently, the model only works with the `weather_stein_hourly` dataset. The dataset contains many variable for evaluating and fitting the model. However, the only needed variables to run `soilflo` are `DateTime`, `ADV_AirTemperature_200cm_C`, `ADV_GlobalRadiation_150cm_Wm2`, `ADV_Precipitation_150cm_m`, `ADV_RelativeHumidity_200cm_RH` and `ADV_WindSpeed_200cm_kmh`. If you want to make the model work with another dataset, the column names must be the same as this one (beware of the units)

Time discretization

The models contains partial differential equations. Their numerical solutions requires the time to be discretised. The example `weather_stein_hourly` contains data every 15 minutes from June 15th, 00:00 to June 21st, 00:00. We are discretising the time in the same manner. In this way, the simulation will run for each of these steps.

```
# Define a start and stop of the simulation
START <- "2014-06-15 00:00"
STOP <- "2014-06-21 00:00"
# Define the time discretization
dt <- 3600 # seconds #= 1 hour
# More info
# help("time_discretization")
```

Space discretization

The soil profile is divided into layers. The size of the layers can be chosen by the user. As a default, I chose 10 layers of 10 cm to represent the soil profile. The function `soil_discretization` creates a list with N, DEPTH and DIST

```
# Divide the soil into 10 layers of 0.1 m
z<-rep(0.1,10)
soil <- soil_discretization(z)
# More info
# help("soil_discretization")
```

Initial conditions

For the moment, we provide constant initial conditions for the temperature and volumetric water content. As the `weather_stein` dataset is in June 2014, an initial condition of 18 degrees for the soil and 0.2

```
VWCini <- rep(0.2,soil$N)
Tini <- rep(18,soil$N)
```

Boundary conditions

The model includes an energy balance equation

$$Rn + H + G + LE = 0$$

where :
Rn $[\mathrm{W\,m^{-2}}]$ = Net radiation, the radiative transfer of energy
H $[\mathrm{W\,m^{-2}}]$ = Sensible (ie you sense it with your fingertips) heat flux, the convective transfer of energy from soil surface to atmosphere
G $[\mathrm{W\,m^{-2}}]$ = Sensible heat flux downward, the conductive transfer of energy from soil surface to lower layers
LE $[\mathrm{W\,m^{-2}}]$ = Latent heat flux, a latent energy transfer from soil surface to atmosphere in the form of water evaporation (latent is a synonym of potential, this type of energy cannot be felt by our senses). This equation computes exchange of energy at the soil surface. The energy balance give the boundary condition of the heat and water flow equations.

Running the model

Executing the `soilflo` function can be done as follow. The result of the model are in the object `res`. The model run for less than a minute with this set-up.

```
# Load the sophia library
library(sophia)

# Load the weather data
data(weather_stein_hourly)

#' # Define a start and stop of the simulation
START <- "2014-06-15 00:00"
STOP <- "2014-06-21 00:00"

# Put the weather dataset into a list structure, limited to START and S
weatherlist <- weatherdataintolist(weather_stein_hourly,START, STOP)

# Divide the soil into 10 layers of 0.1 m
z<-rep(0.1,10)
N <- length(z)

# Define the time discretization
dt <- 3600 # seconds

# Execute the model
res <- soilflo(par=sophia.define.param(),
        weather=weatherlist,
        z = z,
        dt = dt,
        VWCini = 0.2,
        Tini = 18,
        lower_cond = 3)
```

Outputs of `soilflo`

See figure B.1, figure B.2 and figure B.3.

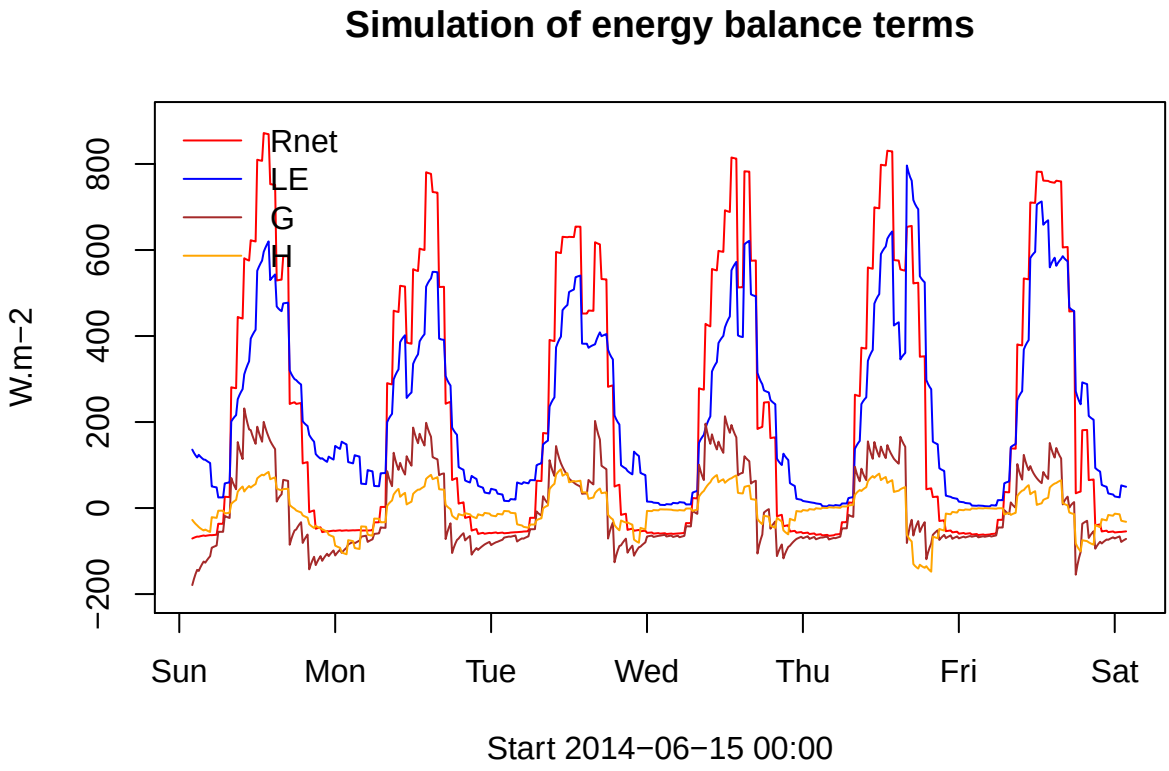

FIGURE B.1: Simulation of energy balance terms at the bare soil surface with meteorological conditions of Stein, CH, from 15 June to 21 June 2014. Rnet: net radiation, LE: latent heat, H: upward convective sensible heat transfer, G: conductive heat flux in the ground

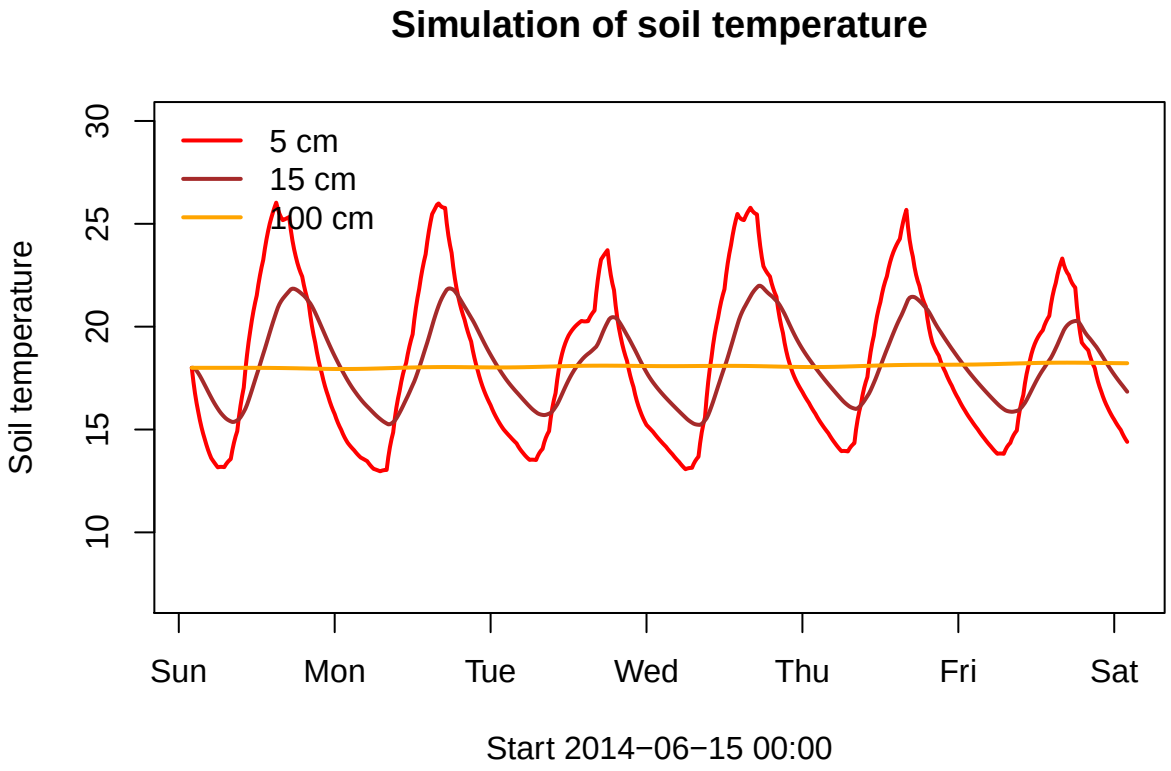

FIGURE B.2: Simulated soil temperature of layer 1 (5 cm), 2 (15 cm) and 10 (1 m) with meteorological conditions of Stein, CH, from 15 June to 21 June 2014.

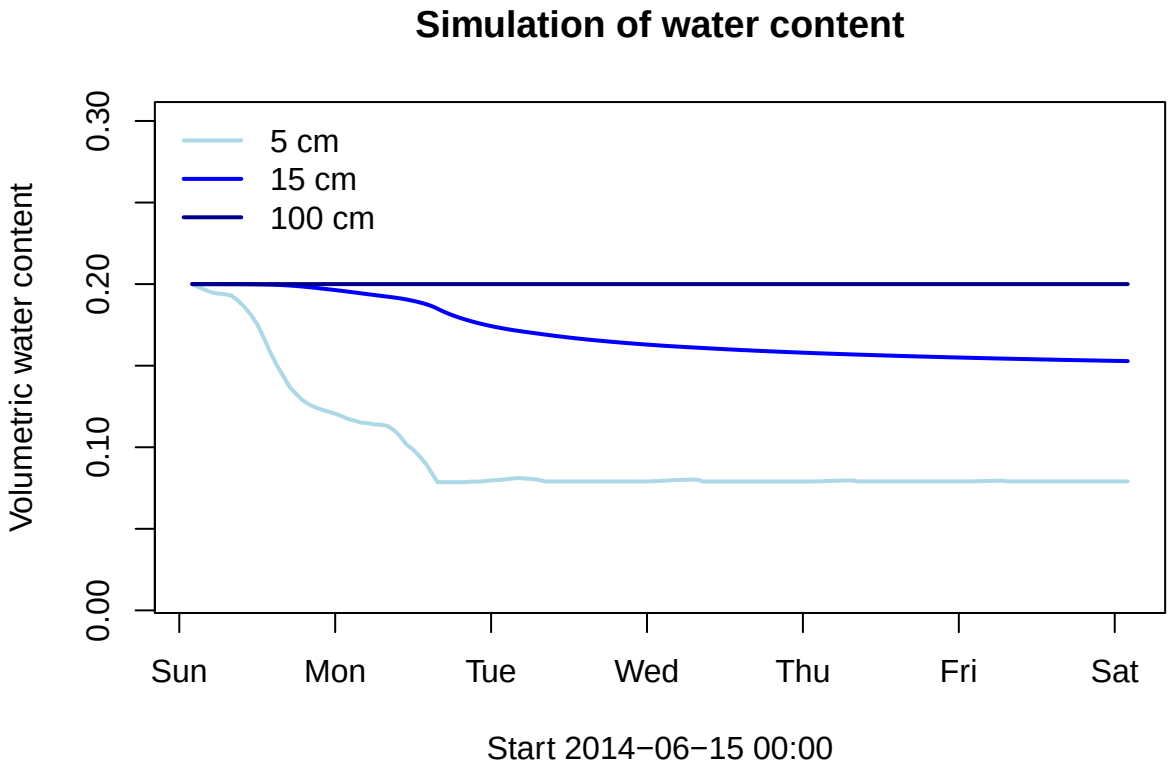

FIGURE B.3: Simulated water content of layer 1 (5 cm), 2 (15 cm) and 10 (1 m) with meteorological conditions of Stein, CH, from 15 June to 21 June 2014.

Default parameters

TABLE B.1: List of parameters and their default value for the simulation in this example

name	description	value
a	Ψ-θ curve	-2.20
A_L	latent heat of vaporisation at °C	2501.00
b	Ψ-θ curve	-5.3
b_Kw_vwc	Kw-θ curve	1.5
B_L	slope of latent heat of vaporaisation and temperature	-2.37
boltzmanConstant	longwaves equation	5.67e-8
Brunt_a	air emissivity	0.60
Brunt_b	air emissivity	0.04
C_air	specific heat	1.01

name	description	value
C_clay	specific heat	0.90
C_om	specific heat	1.92
C_quartz	specific heat (sand and silt)	0.80
C_water	specific heat	4.18
Dens_air	density	0.00
Dens_bulk	mass of solids/total soil volume	1.30
Dens_clay	density	2.65
Dens_om	density	1.30
Dens_quartz	density	2.66
Dens_solid	mass of a volume of solid	2.60
Dens_water	density	1.00
f_clay	volume fraction of clay	0.27
f_om	volume fraction of organic matter	0.03
f_sand	volume fraction of sand	0.36
gasconstant	gas constant	8.31
gravity	gravity constant	9.81
Ka_air	ratio cond. air/water	1.40
Ks_clay	ratio cond. clay/water	0.40
Ks_om	ratio cond. om/water	0.40
Ks_quartz	ratio cond. quartz/water	0.40
KT_air	thermal cond. of air	0.02
KT_clay	thermal cond. of clay	2.92
KT_om	thermal conductivity of organic matter	0.25
KT_quartz	thermal conductivity of quartz	8.80
KT_water	thermal conductivity of water	0.57
Kws	hydraulic cond. at saturat^{ion}	5e-7
MinPot	hydraulic potential parameter	-1000
Molwt_water	mol. weight of H2O	18.00
pi	pi	3.14
soil_emissivity_factor	longwaves upward	0.90
soilAlbedo	reflection coefficient	0.05
tetensConstant_a	Magnus Tetens formula	0.61
tetensConstant_b	Magnus Tetens formula	17.27
tetensConstant_c_degC	Magnus Tetens formula	240.97
tetensConstant_c_degK	Magnus Tetens formula	36.00
vonKarmanConstant	Constant for aerodynamic resistance	0.30

name	description	value
VWC_e	air-entry suction treshold	0.49
z0	roughness parameter	0.01
zref	reference height	2.00

References

Campbell, Gaylon S, and John M Norman. 1998. *An Introduction to Environmental Biophysics*. Springer.

Hillel, Daniel. 2003. *Introduction to Environmental Soil Physics*. Academic press.

Muller, Christoph. 1999. *Modelling Soil-Biosphere Interactions.* Cabi Publishing.

Zhang, Lu, Glen R Walker, and Warrick R Dawes. 2002. "Water Balance Modelling: Concepts and Applications." *ACIAR Monograph Series* 84. Australian Centre for International Agricultural Research: 31–47.

Appendix C

Resolution of the soil model with a semi-implicit scheme

With such a nonlinear model, an implicit scheme is recommended. The implementation is much more difficult because it needs knowledge in matrix inversions. Often in publication, the implicit scheme is just mentioned but not detailed. For example, Bittelli, Ventura, et al. (2008) mention a Cranck-Nicholson scheme resolved with a Newton-Raphson algorithm. The Cranck-Nicholson scheme is an implicit scheme but the detail is not given. An exception where the numerical scheme is well described is in HYDRUS-1D (J Simunek et al., 2013) documentation. The authors mention a mass-lumped linear finite elements scheme to resolve Richards' equation. They use a Picard iterative solution at each time step to find the right value of Ψ.

If we discretise the equation for water movement (equation 2.38) in the differential form we have

$$\frac{\theta_j^{n+1} - \theta_j^n}{\Delta t} = \frac{g\left(\frac{\theta_{j+1}^n + \theta_j^n}{2}\right)\Psi'\left(\frac{\theta_{j+1}^n + \theta_j^n}{2}\right)(\theta_{j+1}^{n+1} - \theta_j^{n+1})}{\Delta z^2} - \frac{g\left(\frac{\theta_j^n + \theta_{j-1}^n}{2}\right)\Psi'\left(\frac{\theta_j^n + \theta_{j-1}^n}{2}\right)(\theta_j^{n+1} - \theta_{j-1}^{n+1})}{\Delta z^2} \tag{C.1}$$

The terms divided by 2 is the average of the water content between two adjacent layers. It is made to increase the accuracy of the solution. The term Ψ' is the derivative of the function $\Psi(\theta)$ (equation 2.38). The derivative form is needed to write the differential form of the partial differential equations.

We rewrite equation C.1 to single out θ_j^n :

$$\theta_j^n = \theta_j^{n+1} + \frac{\Delta t}{\Delta z^2}\left[- g\left(\frac{\theta_{j+1}^n + \theta_j^n}{2}\right)\Psi'\left(\frac{\theta_{j+1}^n + \theta_j^n}{2}\right)(\theta_{j+1}^{n+1} - \theta_j^{n+1}) + g\left(\frac{\theta_j^n + \theta_{j-1}^n}{2}\right)\Psi'\left(\frac{\theta_j^n + \theta_{j-1}^n}{2}\right)(\theta_j^{n+1} - \theta_{j-1}^{n+1})\right] \tag{C.2}$$

We can switch to a matrix notation that we will detail hereafter

$$\theta^n + b(\theta^n) = (I + \frac{\Delta t}{\Delta z^2}M(\theta^n))\theta^{n+1} \tag{C.3}$$

In this equation θ^n is a vector as :

$$\theta^n = \begin{bmatrix} \theta_1^n \\ \vdots \\ \theta_j^n \\ \vdots \\ \theta_k^n \end{bmatrix}$$

I is the identity matrix

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $M(\theta)$ a tridiagonal matrix that is multiplied by θ^{n+1}

$$\begin{bmatrix} d_2 & r_2 & 0 & 0 & 0 \\ l_{j-1} & d_{j-1} & r_{j-1} & 0 & 0 \\ \vdots & l_j & d_j & r_j & \vdots \\ 0 & 0 & l_{j+1} & d_{j+1} & r_{j+1} \\ 0 & 0 & 0 & l_k & d_k \end{bmatrix} \begin{bmatrix} \theta_1^{n+1} \\ \vdots \\ \theta_j^{n+1} \\ \vdots \\ \theta_k^{n+1} \end{bmatrix} \tag{C.4}$$

with :

$$d_2 = g\left(\frac{\theta_2^n + \theta_1^n}{2}\right)\Psi'\left(\frac{\theta_2^n + \theta_1^n}{2}\right)$$

$$d_j = g\left(\frac{\theta_{j+1}^n + \theta_j^n}{2}\right)\Psi'\left(\frac{\theta_{j+1}^n + \theta_j^n}{2}\right) + g\left(\frac{\theta_j^n + \theta_{j-1}^n}{2}\right)\Psi'\left(\frac{\theta_j^n + \theta_{j-1}^n}{2}\right)$$

$$d_k = g\left(\frac{\theta_k^n + \theta_k^n - 1}{2}\right)\Psi'\left(\frac{\theta_k^n + \theta_k^n - 1}{2}\right)$$

and

$$r_j = -g\left(\frac{\theta_{j+1}^n + \theta_j^n}{2}\right)\Psi'\left(\frac{\theta_{j+1}^n + \theta_j^n}{2}\right)$$

and

$$l_j = g\left(\frac{\theta_j^n + \theta_{j-1}^n}{2}\right)\Psi'\left(\frac{\theta_j^n + \theta_{j-1}^n}{2}\right)$$

Finally, $b(\theta^n)$ is a vector having the top boundary conditions and 0 otherwise

$$b(\theta^n) = \frac{\Delta t}{\Delta z^2}\begin{bmatrix} g\left(\frac{\theta_1^{n_1} + \theta_0^n}{2}\right)\Psi'\left(\frac{\theta_1^{n_1} + \theta_0^n}{2}\right) \cdot \frac{\Delta z}{2}\left(E(T_0^n\theta_1^n) - I(\theta_1^n)\right) \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \tag{C.5}$$

Appendix D

Résumé de la thèse en français

Avant-propos

Ce travail de thèse est le fruit d'une collaboration entre l'entreprise suisse Syngenta Crop Protection AG et l'école d'ingénieur française CentraleSupélec. Plus précisément, la collaboration s'est effectuée entre deux équipes : d'une part au sein du département Recherche et Développement de Syngenta, où l'équipe de modélisation des cultures et de l'environnement (ECM) a pour mission principale d'identifier les stress abiotiques perçus par la plante et d'aider sélectionneurs et commerciaux à promouvoir de nouvelles variétés de céréales; d'autre part à CentraleSupélec, avec l'équipe de biomathématiques, qui se consacre à l'étude et à l'analyse des propriétés mathématiques des modèles biologiques, l'inférence statistique, l'apprentissage automatique et la commande optimale.

Introduction

Les cultures annuelles telles que le maïs développent leur canopée et leur système racinaire durant le premier mois de leur croissance. L'établissement rapide d'une canopée homogène est un facteur clé pour pouvoir capturer les ressources nécessaires à la croissance du couvert végétal, à savoir la lumière, l'eau et les nutriments. Un établissement rapide du couvert végétal permet de réduire la présence de mauvaises herbes en les privant de lumière (Lipiec, Nosalewicz, and Pietrusiewicz, 2011). L'interception des ressources détermine la quantité et la qualité du produit récolté (Atkinson and Porter, 1996). Pour le maïs, cette période d'établissement de la canopée se situe entre la date de semis et le stade cinq feuilles et dure entre 15 et 25 jours (FAO, n.d.).

Pour maximiser la durée de cycle et le rendement, les producteurs de maïs sèment de plus en plus tôt. Entre 1981 et 2005, la date de semis s'est

avancée en moyenne de 10 jours aux États-Unis (Sacks and Kucharik, 2011). Cette avancée augmente la durée de cycle et le rendement potentiel, mais augmente également le risque de rencontrer des épisodes froids après le semis, avec pour conséquence un établissement du couvert moins homogène et donc un rendement moindre (Lipiec, Nosalewicz, and Pietrusiewicz, 2011). Les pratiques agricoles de non-labour affectent également la température du sol, son contenu en eau, sa structure, et donc modifient les chances d'un établissement réussi (G. Flerchinger, Sauer, and Aiken, 2003).

La vigueur de cet établissement est très dépendante du processus d'émergence, c'est-à-dire la percée de la surface du sol par le coléoptile (W. E. Finch-Savage and Bassel, 2015). Pendant l'émergence, la graine, et donc la zone de croissance, se trouve sous la surface du sol. Le processus de germination et d'élongation dépend de l'environement physique de la graine, caractérisé principalement par la température, la teneur en eau et la structure du sol (Gupta, Swan, and Schneider, 1988). Des informations fiables sur les conditions environmentales autour de la graine pourrait apporter aux chercheurs, agriculteurs et agronomes une meilleure compréhension des facteurs limitants de l'émergence et donc favoriser une meilleure gestion des risques qui y sont liés (Bartolo et al., 2011). Or, même si la température, la teneur en eau et la structure du sol sont des informations importantes, elle sont rarement disponibles en pratique. Il serait avantageux de pouvoir modéliser ces informations grâce à un modèle mathématique, à partir de mesures des conditions atmosphèriques qui, pour leur part, sont mesurées plus systématiquement.

Le sol est un système complexe, composé de phases solides, liquides et gazeuses qui interagissent entres elles à différents pas de temps et d'espace. Le modèle doit prendre en compte cette complexité si l'on veut être en mesure de comprendre et d'agir sur un tel système. Le coeur de notre travail a consisté à développer et analyser un modèle principalement dédié à la prédiction de la température du sol. Dans ce contexte, nous avons notamment abordé les questions scientifiques sous-jacentes suivantes : par quoi sont influencées les variations de température sous la surface du sol? Quelles sont les équations biophysiques qui décrivent les processus influents? Comment transformer les équations en un modèle numérique? Quels sont les paramètres influents de ce modèle? Comment peut-on les estimer? Le modèle a fait l'objet d'une application à la prédiction du temps d'émergence pour le maïs. Cette application soulève des questions sur le modèle d'émergence le plus adapté ainsi que sur le niveau de précision du modèle nécessaire pour bien prédire l'émergence.

L'objectif de cette thèse est ainsi triple : I) développer un modèle capable

de prédire la température et la teneur en eau du sol autour d'une graine, à l'aide de données climatiques largement disponibles, II) analyser la sensibilité des sorties du modèle pour identifier les composants qui contribuent le plus largement à leur incertitude et III) appliquer le modèle à la prédiction de l'émergence du maïs.

L'objet d'étude : analyse descriptive de la température du sol à Stein (chapitre 1)

Le premier chapitre commence par la présentation de la station météorologique installée spécifiquement pour l'étude de la température du sol et des échanges d'énergie radiatifs. La station est équipée de sondes mesurant la température et l'humidité relative de l'air, la vitesse du vent, la pluviométrie et les radiations (courtes et grandes longueurs d'onde, voir paragraphe suivant).

Pour les échanges radiatifs, la station météo est équipée d'un bilanmètre. Un bilanmètre est une association de quatre thermopiles (sondes mesurant la densité de flux des radiations) qui mesurent séparément la radiation globale, la réflection solaire du sol, les radiations de grandes longueurs d'onde issues du sol (montantes) et de l'atmosphère (descendantes). Le bilanmètre a donc deux sondes tournées vers l'atmosphère et deux sondes tournées vers le sol.

La station météo est également équipée de sondes pour la température et la teneur en eau du sol (Hydraprobe II, Stevens, USA) placées 5, 30 et 140 cm sous la surface. Ces profondeurs ont été choisies pour leur pertinence agronomique : 5 cm correspond au lit de semence, 30 cm correspond à la profondeur nécessaire pour prendre en compte le cycle de l'azote et 140 cm est une profondeur suffisante pour ne plus détecter les variations journalières de la température du sol et connaître la teneur en eau en profondeur.

Le jeu de données provenant de la station météo s'étale sur une période de 2 ans et demi, du 15 juin 2014 au 1 janvier 2017, et est agrégé au niveau horaire (moyenne des 4 mesures enregistrées tous les quart d'heure).

Notre courte analyse descriptive montre que la température du sol proche de la surface du sol est proche de la température de l'air, mais que les variations de cette dernière ne suffisent pas pour expliquer les variations extrêmes de température du sol ni de la température en profondeur. Pour obtenir ces températures, il est alors nécessaire de développer un modèle mécaniste que nous présentons dans le chapitre suivant.

Développement du modèle (chapitre 2)

Puisqu'il est pratique de nommer un modèle pour s'y référer, nous l'avons nommé SOPHIA, qui est une contraction de l'anglais *Soil Physics in Agronomy* et qui souligne l'utilisation visée du modèle en agronomie.

Le développment du modèle commence par sa conception sous forme de diagramme. Ce diagramme de Forrester (Brun et al., 2006) présenté en figure 2.1 page 41 est intéressant pour donner une vision globale des compartiments, des entrées, des flux et des variables d'états du modèle.

Le modèle prend en entrée cinq variables météorologiques : la température de l'air, l'humidité relative, la radiation globale, la vitesse du vent et la pluviométrie. Il nécessite également en entrée un jeu de valeurs de paramètres décrivant les propriétés physiques du sol (non représenté sur le diagramme).

Le modèle peut être résumé par le système d'équations algébro-différentielles décrit dans la figure D.1 où l'on note $T = T(z,t)$ la température du sol et $\theta = \theta(z,t)$ la teneur en eau à une profondeur z et un temps t. Ce système contient deux équations différentielles partielles et les équations aux bornes : la résolution du bilan d'énergie pour la borne supérieure et les diverses conditions pour le bas du profil de sol.

Pour résoudre ce système d'équation, nous utilisons un schéma aux différences finies avec une écriture explicite (d'Euler), avec des couches de sol d'épaisseur variable et en prenant en compte le couplage entre les deux variables pour définir l'ordre de résolution.

Concernant l'implémentation, nous avons choisi de développer un package (paquet de fonctions) dans le langage de programmation R (R Core Team, 2015) et avons également rédigé une fiche d'introduction du paquet B. Lancer une simulation nécessite un fichier d'entrées avec de la donnée météorologique au niveau horaire sans données manquantes.

Une première évaluation du modèle sur les 10 premiers jours d'avril 2015 montre une erreur relative de 29 % pour notre modèle SOPHIA pour la température à 5 cm sous le sol, la variable la plus importante pour prédire l'émergence des cultures. Cette valeur est à comparer avec l'erreur relative de 41 % que l'on obtient avec le modèle de Müller (Müller, 1999), la référence dont nous nous sommes inspirés pour l'architecture générale du modèle et un certain nombre de ses modules. Une analyse plus fine de nos résultats montre que l'erreur la plus importante concerne les radiations nettes, avec 78 % d'erreur relative. Pour abaisser cette erreur, il est important d'analyser le modèle pour découvrir les paramètres importants.

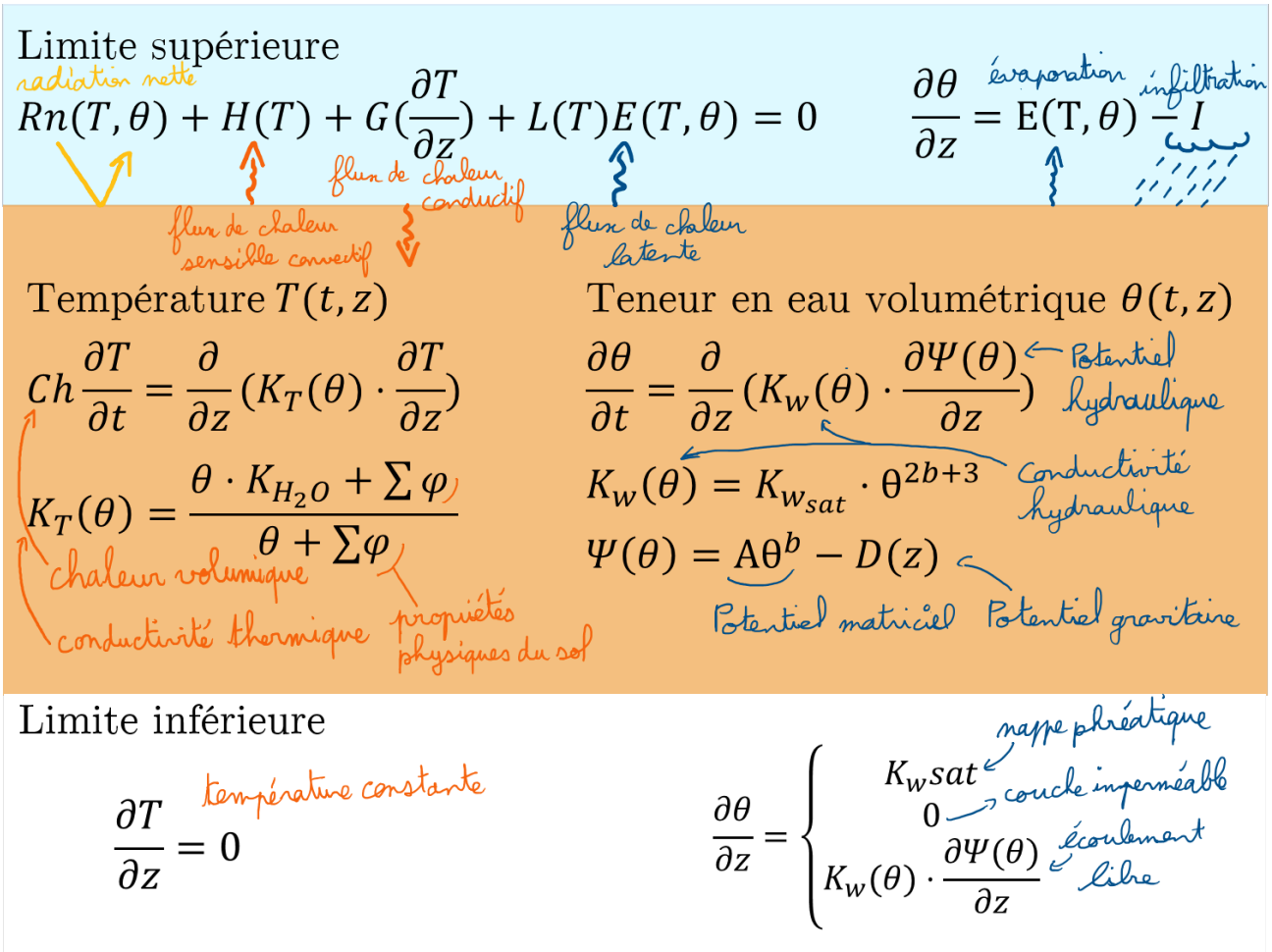

FIGURE D.1: Principales équations du modèle SOPHIA. Les équations différentielles partielles pour la température et la teneur en eau volumétrique avec les conditions aux limites inférieures et supérieures. L'équation du bilan d'énergie à la surface contient de nombreux sous-modèles.

Analyse de sensibilité, calibration et évaluation du modèle (chapitre 3)

Analyse de sensibilité

Nous présentons trois méthodes d'analyse de sensibilité et leur application au modèle SOPHIA. Une première analyse locale et graphique, la méthode semi-globale de Morris (Campolongo, Jessica Cariboni, and Saltelli, 2007) et la méthode de Sobol basée sur la décomposition de la variance (Saltelli, Tarantola, et al., 2004).

Nous nous intéressons à 14 paramètres du modèle SOPHIA parmi lesquels 7 sont des paramètres prenant part aux équations du bilan d'énergie et 7 sont des paramètres de sol, impliqués dans les équations de diffusion de la chaleur et de l'eau dans le sol. Nous avons défini la gamme de variation de chaque paramètre grâce à la littérature pour les paramètres de surface et grâce à un modèle de pédotransfert, le modèle de Saxton (Saxton and Rawls, 2006), pour les paramètres du sol.

Les trois types d'analyse de sensibilité (graphique, Morris et Sobol) que nous avons réalisés sur notre modèle SOPHIA révèlent l'importance prépondérante des paramètres de surface pour la température proche de la surface, par rapport aux paramètres du sol. Il s'agit en particulier des paramètres de la résistance aérodynamique de surface et des paramètres de la radiation nette.

Estimation des paramètres de la radiation nette

Les paramètres de la radiation nette peuvent être calibrés indépendamment des autres paramètres du modèle grâce au bilanmètre installé sur la station météo de Stein. Grâce aux mesures des grandes longueurs d'onde descendantes, nous avons pu les calibrer et cette étude nous a amenés à introduire un nouveau facteur de correction de l'ennuagement qui permet la réduction de l'erreur de plus de 50 %.

Pour l'albedo, également mesuré avec le bilanmètre, nous avons également pu réduire l'erreur de 50 % sur ce terme en en introduisant une nouvelle expression. Celle-ci prend en compte un bilan hydrique à la surface pour révéler si le sol à la surface est en phase de séchage ou d'hydratation, et mesurer la teneur en eau de la première couche de sol. Cette nouvelle expression de l'albedo est un apport original de la thèse.

Le dernier paramètre concernant la radiation nette est le facteur d'émissivité du sol. L'analyse des données mesurées par le bilanmètre nous a incités à

fixer l'émissivité du sol à 1 afin de se rapprocher des données observées, ce qui représente une simplification intéressante du modèle.

Les paramètres de la résistance aérodynamique, identifiés comme importants par l'analyse de sensibilité, n'ont pu être calibrés faute d'instruments précis (anémomètre sonique). Cela reste une piste importante d'amélioration du modèle que nous n'avons pas eu le temps d'explorer dans le cadre de cette thèse.

Évaluation du modèle après calibration

En combinant les différentes possibilités de valeurs de paramètres calibrées et de nouvelles expressions développées pour certains modules, nous avons obtenu 15 versions du modèle SOPHIA que nous avons testées et comparées. Nous nous sommes servi pour cela, comme jeu de test de prédiction, des données collectées durant l'année 2016 qui n'ont pas été utilisées pour la calibration. La version combinant toutes les améliorations a une erreur plus faible pour toutes les variables par rapport à la version de référence, c'est-à-dire avant la calibration. Elle présente une erreur de 15 % sur la température à 5 cm, 9 % à 30 cm, 46 % pour la radiation net, 26 % pour la teneur en eau à 5 cm et 10 % pour la teneur en eau à 30 cm.

Une analyse plus fine de la performance du modèle en fonction de la saison nous montre que l'erreur du modèle est plus grande pendant les mois d'été. Le modèle sous-estime alors la température du sol. Cette sous-estimation de la température est liée surtout à une sous-estimation de la température pendant la nuit : le modèle prédit des températures plus froides que ce qui est réellement observé pendant la nuit.

Pour les mois de semis, avril et mai, la performance du modèle est comparable à celle d'autres modèles publiés et nous validons donc son usage pour notre exemple d'application : la prédiction de l'émergence du maïs.

Application à la prédiction de l'émergence (chapitre 4)

L'objectif de ce chapitre est de montrer la plus-value d'un modèle de température de sol combiné à un modèle de temps thermique spécifique à l'émergence pour améliorer la prédiction du temps entre semis et émergence pour le maïs. Nous définissons l'émergence VE (Vegetative Emergence), en tant que stade

phénologique d'une population, comme la date à laquelle 50 % des coléoptiles de la population sont visibles au-dessus de la surface du sol.

La prédiction de stade phénologique se base sur l'utilisation du concept de temps thermique : un stade phénologique est atteint lorsqu'un certain nombre de degrés-jours est dépassé. Pour calculer les degrés jours, on utilise classiquement la formule (Bonhomme, 2000) :

$$TT^{n}_{stade} = \sum_{i=1}^{n}(T_i - T_b) \tag{D.1}$$

Où TT_{stade} est le nombre de degrés-jours, T_i la température de l'organisme (ou de son proche environnement) au jour i et T_b la température de base de l'organisme. La température de base est une température théorique en deçà de laquelle le développement de l'organisme s'arrête. Cette température est obtenue en extrapolant la partie linéaire du développement jusqu'à 0.

L'organisme sur lequel nous nous concentrons dans ce travail sur le temps thermique est le maïs entre la phase graine et la phase émergence. Pour obtenir la température de base de cette phase, nous avons mis en place une expérimentation. Des graines de quatre cultivars de maïs ont été plantées à 11 dates différentes au cours des années 2015 et 2016. Les variétés ont été choisies car elles montrent une importante variabilité sur le temps de germination en laboratoire. Les 11 dates de semis ont permis d'obtenir une variabilité de température du lit de semence pendant l'émergence. Pour chaque date de semis, les quatres cultivars ont été plantés avec quatre répétitions. Chaque jour après le semis, le nombre de plantes émergées, c'est-à-dire dont le coléoptile dépassait d'un cm au-dessus de la surface, était relevé. Les dynamiques d'émergence (nombre de plantes émérgées au cours du temps) générées ont été ajustées avec une courbe logistique selon l'équation :

$$E(t) = \frac{a}{1 + e^{\left(\frac{-4b \cdot (x - VE)}{a}\right)}} \tag{D.2}$$

avec a la proportion maximale de plantules qui ont émergé, b le taux maximal d'émergence de plantes par heure et VE l'heure après émergence à laquelle 50 % des plantes ont déjà émergé. Pour ajuster la courbe, nous avons utilisé un algorithme d'optimisation sur b et VE. Le nombre maximal de plantes ayant émergé a a été évalué séparément, car il était parfois inférieur à 100 %. En 2015, b varie de 0.34 à 6 émergences de plante par heure et VE varie de 5 à 26 jours après semis.

	Model TT_3			
Cultivar	T_{base} [°C]	T_{opt} [°C]	T_{max} [°C]	GDH to VE [°C h]
NK Cobalt	$7.3(\pm 0.1)$	$27.9(\pm 0.9)$	$44.5(\pm 1.6)$	$1911.7(\pm 17.7)$
NK Falkone	$7(\pm 0.1)$	$30.2(\pm 0.5)$	$39.1(\pm 0.7)$	$2086.2(\pm 17.8)$
NK Famoso	$8.2(\pm 0.1)$	$31.3(\pm 0.4)$	$35.8(\pm 0.9)$	$1775.3(\pm 30.5)$
SY Multitop	$7.3(\pm 0.2)$	$30.6(\pm 0.3)$	$39.6(\pm 0.2)$	$1995.1(\pm 39.8)$

Avec un modèle de temps thermique, il est possible de dire que la plante émerge pour une certaine somme de degrés-heure de température du lit de semence, qui sera la même pour chaque date de semis. C'est la puissance du concept de temps thermique. Nous avons donc voulu trouver un modèle de temps thermique qui puisse prédire le temps de l'émergence, quelle que soit la date de semis, à l'aide de la température du lit de semence. Nous avons ainsi testé trois formes de modèles de temps thermiques : un modèle simple, avec un seul paramètre T_{base}, qui représente une relation linéaire, un modèle avec 2 paramètres, qui possède une partie linéaire et un plateau qui commence à T_{opt} et un modèle à 3 paramètres qui forment un pic. La partie linéaire croit jusqu'à T_{opt} puis décline jusqu'à T_{max}, la température maximale de développement de l'organisme.

Pour choisir le modèle de temps thermique qui prédit au mieux le temps de VE, nous avons sélectionné le modèle qui, lorsqu'appliqué pour transformer le temps calendaire en temps thermique, réduit le plus possible le coefficient de variation du paramètre VE de la courbe logistique. Ainsi, c'est une fonction de minimisation de coefficient de variation que nous avons appliqué. Pour obtenir le meilleur modèle, nous avons effectué une procédure originale qui nous garantit de sélectionner le meilleur modèle de temps thermique. Cette procédure estime, avec la fonction de minimisation, les paramètres avec six dates de semis, puis calcule une erreur sur une septième date de semis. Ce processus est réitéré avec les 6 autres dates de semis. Enfin l'erreur est moyennée sur les 7 itérations. Le modèle pour lequel l'erreur est la plus faible est le modèle à 3 paramètres, c'est-à-dire le modèle dans lequel le développement est pénalisé par les fortes températures. Les valeurs de la température de base, optimale et maximale pour chaque cultivar (moyenne des 7 itérations de la sélection du modèle) sont résumées dans le tableau ci-dessous.

Pour montrer le bénéfice de notre modèle SOPHIA, nous avons utilisé, en entrée du modèle de temps thermique à 3 paramètres, la température de sol

simulée par SOPHIA. L'évaluation s'est faite sur 4 dates de semis en 2016 : le 7 avril, le 4 mai, le 19 juillet (seulement NK Falkone) et le 22 septembre (seulement NK Famoso). Pour référence, nous utilisons le modèle de temps thermique simulé avec la température de l'air en entrée, ce qui est la procédure la plus couramment utilisée pour prédire les stades phénologiques dans la littérature. Dans 90 % des cas, utiliser la température du sol simulée améliore la prédiction de l'émergence par rapport à l'utilisation de la température de l'air.

Pour clore le chapitre, nous nous posons alors la question de savoir quel niveau d'erreur du modèle SOPHIA est tolérable pour avoir une bonne prédiction de l'émergence. La réponse dépend de la température du sol. Plus celle-ci est proche de la température de base, plus le modèle de temps thermique peine à prédire l'émergence. Dans ce cas-là, même si l'erreur du modèle de température du sol est égale à 1 %, il y aura une erreur importante sur la prédiction de l'émergence. Nous concluons que, pour des dates de semis très précoces, le modèle de temps thermique en 2 segments n'est pas adapté, même si l'on a un très bon modèle de température. Une des perspectives de recherche de ce travail est donc de trouver un modèle d'émergence plus adapté lorsque les dates de semis sont très proches de la température de base.

Conclusion

La température du sol contrôle des processus importants du début de cycle des cultures, tels que l'émergence, la disponibilité en nutriment et le devenir des molécules actives dans le sol. Cette thèse présente les étapes de développement d'un modèle, SOPHIA, qui prédit la température du sol et qui s'appuie sur le concept de bilan d'énergie à la surface. L'apport original de cette thèse est l'application des techniques d'analyse de sensibilité sur ce type de modèle et la calibration du modèle à l'aide des données du bilan-mètre. À l'issue de cette calibration, le modèle affiche une erreur de moins de 10 % pour la température du sol à 30 cm et de 20 % pour la température à 5 cm du sol sur une simulation de toute l'année 2016 à Stein en Suisse. Comme exemple d'application, nous montrons que la température du sol peut servir comme entrée d'un modèle de temps thermique et ainsi mieux prédire l'émergence par rapport à l'utilisation de la température de l'air.

Cet exercice de modélisation ouvre de nombreuses pistes de recherche.

Le bilan d'énergie est composé de plusieurs sous-modèles. Chaque sous-modèle possède plusieurs formalismes dans la littérature. Il serait intéressant d'élaborer une stratégie formelle pour tester et évaluer ces différents formalismes. Les sous-modèles que nous avons identifiés comme les plus intéressants à explorer seraient les fonctions de pédo-transferts, la formulation de la résistance aérodynamique et le couplage des mouvements d'eau et de chaleur.

Dans un contexte industriel, le modèle SOPHIA peut servir à l'élaboration de scénarios de température du sol, c'est-à-dire caractériser les variations de la température du sol pour certaines périodes, certaines profondeurs et certains lieux. Ces scénarios peuvent être transmis au laboratoire qui reproduira ces variations de température dans une chambre de culture et sera ainsi plus proche des conditions extérieures. Le modèle SOPHIA peut aussi servir à l'identification de stress liés à la température du sol pour la période semis-émergence sur un réseau d'essais agronomiques. Par exemple, en faisant des simulations avec les données météo récoltées sur la parcelle, nous pourrons savoir quelles ont été les températures de sol et donc si la graine a subi des épisodes trop chauds ou trop froids. Enfin, le modèle SOPHIA pourra servir de brique de base pour l'élaboration d'un modèle multi-échelle plus complexe qui comprendrait également des modèles mécanistes de développement de la graine, des modèles de croissance racinaire et des modèles de devenir de molécules d'intérêt dans le sol.

Bibliography

Acs, F, DT Mihailovic, and B Rajkovic (1991). "A coupled soil moisture and surface temperature prediction model". In: *Journal of Applied Meteorology* 30.6, pp. 812–822.

Al Majou, Hassan et al. (2007). "Variation of the water-retention properties of soils: Validity of class-pedotransfer functions". In: *Comptes Rendus Geoscience* 339.9, pp. 632–639.

Alvenäs, Gunnel and Per-Erik Jansson (1997). "Model for evaporation, moisture and temperature of bare soil: calibration and sensitivity analysis". In: *Agricultural and Forest Meteorology* 88.1, pp. 47–56.

Atkinson, David and John R Porter (1996). "Temperature, plant development and crop yields". In: *Trends in Plant Science* 1.4, pp. 119–124.

Banimahd, SA and Sh Zand-Parsa (2013). "Simulation of evaporation, coupled liquid water, water vapor and heat transport through the soil medium". In: *Agricultural Water Management* 130, pp. 168–177.

Bartolo, S De et al. (2011). "Preface "Modeling soil system: complexity under your feet"". In: *Biogeosciences* 8.11, pp. 3139–3142.

Beel, Joeran et al. (2011). "Docear: An academic literature suite for searching, organizing and creating academic literature". In: *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*. ACM, pp. 465–466.

Benaglia, Tatiana et al. (2009). "mixtools: An R Package for Analyzing Finite Mixture Models". In: *Journal of Statistical Software* 32.6, pp. 1–29. URL: http://www.jstatsoft.org/v32/i06/.

Bhosale, SU et al. (2007). "Chilling tolerance of central European maize lines and their factorial crosses". In: *Annals of botany* 100.6, pp. 1315–1321.

Bishop, Christopher M (2006). *Pattern recognition and machine learning*. springer.

Bittelli, Marco, Gaylon S Campbell, and Fausto Tomei (2015). *Soil Physics with Python: Transport in the Soil-Plant-Atmosphere System*. OUP Oxford.

Bittelli, Marco, Francesca Ventura, et al. (2008). "Coupling of heat, water vapor, and liquid water fluxes to compute evaporation in bare soils". In: *Journal of Hydrology* 362.3, pp. 191–205.

Bojanowski, Jedrzej S. (2016). *sirad: Functions for Calculating Daily Solar Radiation and Evapotranspiration*. R package version 2.3-2/r80. URL: http://R-Forge.R-project.org/projects/sirad/.

Bollero, German A, Donald G Bullock, and Steven E Hollinger (1996). "Soil temperature and planting date effects on corn yield, leaf area, and plant development". In: *Agronomy Journal* 88.3, pp. 385–390.

Bonhomme, Raymond (2000). "Bases and limits to using degree.day units". In: *European journal of agronomy* 13.1, pp. 1–10.

Bradley, CA (2008). "Effect of fungicide seed treatments on stand establishment, seedling disease, and yield of soybean in North Dakota". In: *Plant Disease* 92.1, pp. 120–125.

Brandolini, A et al. (2000). "Variation among Andean races of maize for cold tolerance during heterotrophic and early autotrophic growth". In: *Euphytica* 111.1, pp. 33–41.

Brisson, Nadine et al. (1998). "STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn". In: *Agronomie* 18.5-6, pp. 311–346.

Brun, Francois et al. (2006). *Working with dynamic crop models: evaluation, analysis, parameterization, and applications*. Elsevier.

Bullied, W John, Paul R Bullock, et al. (2014). "Process-based modeling of temperature and water profiles in the seedling recruitment zone: Part II. Seedling emergence timing". In: *Agricultural and Forest Meteorology* 188, pp. 104–120.

Bullied, W John, Gerald N Flerchinger, et al. (2014). "Process-based modeling of temperature and water profiles in the seedling recruitment zone: Part I. Model validation". In: *Agricultural and Forest Meteorology* 188, pp. 89–103.

Burnham, Kenneth P and David R Anderson (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Bussière, François and Pierre Cellier (1994). "Modification of the soil temperature and water content regimes by a crop residue mulch: experiment and modelling". In: *Agricultural and Forest Meteorology* 68.1-2, pp. 1–28.

Campbell, Gaylon S (1974). "A simple method for determining unsaturated conductivity from moisture retention data." In: *Soil science* 117.6, pp. 311–314.

Campbell, Gaylon S and John M Norman (1998). *An introduction to environmental biophysics*. Springer.

Campolongo, Francesca, Jessica Cariboni, and Andrea Saltelli (2007). "An effective screening design for sensitivity analysis of large models". In: *Environmental modelling & software* 22.10, pp. 1509–1518.

Cariboni, J et al. (2007). "The role of sensitivity analysis in ecological modelling". In: *Ecological modelling* 203.1, pp. 167–182.

Casadebaig, Pierre et al. (2011). "SUNFLO, a model to simulate genotype-specific performance of the sunflower crop in contrasting environments". In: *Agricultural and forest meteorology* 151.2, pp. 163–178.

Cellier, P et al. (1993). "Estimating the temperature of a maize apex during early growth stages". In: *Agricultural and Forest Meteorology* 63.1-2, pp. 35–54.

Chastaing, Gaelle, Fabrice Gamboa, and Clémentine Prieur (2015). "Generalized Sobol sensitivity indices for dependent variables: numerical methods". In: *Journal of Statistical Computation and Simulation* 85.7, pp. 1306–1333.

Claverie, Etienne et al. (2016). "Modeling soil temperature to predict emergence". In: *Functional-Structural Plant Growth Modeling, Simulation, Visualization and Applications (FSPMA), International Conference on*. IEEE, pp. 28–37.

Collins, Dan C and Roni Avissar (1994). "An evaluation with the Fourier amplitude sensitivity test (FAST) of which land-surface parameters are of greatest importance in atmospheric modeling". In: *Journal of Climate* 7.5, pp. 681–703.

Corripio, Javier G (2003). "Vectorial algebra algorithms for calculating terrain parameters from DEMs and solar radiation modelling in mountainous terrain". In: *International Journal of Geographical Information Science* 17.1, pp. 1–23.

Cournede, P-H et al. (2011). "Some parameter estimation issues in functional-structural plant modelling". In: *Mathematical Modelling of Natural Phenomena* 6.2, pp. 133–159.

Cousin, Isabelle, Bernard Nicoullaud, and Caroline Coutadeur (2003). "Influence of rock fragments on the water retention and water percolation in a calcareous soil". In: *Catena* 53.2, pp. 97–114.

De Parcevaux, Sane and Laurent Huber (2007). *Bioclimatologie: Concepts et applications*. Quae.

De Vries, DA (1958). "Simultaneous transfer of heat and moisture in porous media". In: *Eos, Transactions American Geophysical Union* 39.5, pp. 909–916.

Durr, C et al. (2001). "Simple". In: *Soil Science Society of America Journal* 65.2, pp. 414–423.

Dwyer, LM, HN Hayhoe, and JLB Culley (1990). "Prediction of soil temperature from air temperature for estimating corn emergence". In: *Canadian Journal of Plant Science* 70.3, pp. 619–628.

ECMWF, ECMWF (2014). *IFS documentation CY40r1*. Tech. rep. consulted August 3rd, 2015.

Edalat, Mohsen and Seyed Abdolreza Kazemeini (2014). "Estimation of cardinal temperatures for seedling emergence in corn". In: *Australian Journal of Crop Science* 8.7, p. 1072.

FAO. *FAO Crop Water Information*. http://www.fao.org/land-water/databases-and-software/crop-information/maize/en/. Accessed: 2017-06-07.

Finch-Savage, WE, HR Rowse, and KC Dent (2005). "Development of combined imbibition and hydrothermal threshold models to simulate maize (Zea mays) and chickpea (Cicer arietinum) seed germination in variable environments". In: *New Phytologist* 165.3, pp. 825–838.

Finch-Savage, William E and George W Bassel (2015). "Seed vigour and crop establishment: extending performance beyond adaptation". In: *Journal of Experimental Botany*, erv490.

Flerchinger, Gerald N (2000). "The simultaneous heat and water (SHAW) model: Technical documentation". In: *Northwest Watershed Research Center USDA Agricultural Research Service, Boise, Idaho*.

Flerchinger, GN, TJ Sauer, and RA Aiken (2003). "Effects of crop residue cover and architecture on heat and water transfer at the soil surface". In: *Geoderma* 116.1, pp. 217–233.

Flerchinger, GN, Wei Xaio, et al. (2009). "Comparison of algorithms for incoming atmospheric long-wave radiation". In: *Water Resources Research* 45.3.

Forcella, Frank et al. (2000). "Modeling seedling emergence". In: *Field Crops Research* 67.2, pp. 123–139.

Fourcaud, Thierry et al. (2008). "Plant growth modelling and applications: the increasing importance of plant architecture in growth models". In: *Annals of Botany* 101.8, pp. 1053–1063.

Fournier, Christian and Bruno Andrieu (1998). "A 3D architectural and process-based model of maize development". In: *Annals of botany* 81.2, pp. 233–250.

Galinier, Thomas (2018). "Multifactorial analysis of crop performance - Method and automation of agronomical, environmental and socio-economic data integration - Example of non-irrigated corn for grain in North America." PhD thesis. Université Paris-Saclay.

Gaucher, Gilbert (1968). *Traite de pedologie agricole, le sol et ses caracteristiques agronomiques*. Dunod.

Grifoll, Jordi, Josep Ma Gastó, and Yoram Cohen (2005). "Non-isothermal soil water transport and evaporation". In: *Advances in Water Resources* 28.11, pp. 1254–1266.

Guilioni, Lydie et al. (2000). "A model to estimate the temperature of a maize apex from meteorological data". In: *Agricultural and forest meteorology* 100.2, pp. 213–230.

Gupta, SC, Birl Lowery, et al. (1991). "Modeling tillage effects on soil physical properties". In: *Soil and Tillage Research* 20.2-4, pp. 293–318.

Gupta, SC, JB Swan, and EC Schneider (1988). "Planting depth and tillage interactions on corn emergence". In: *Soil science society of America journal* 52.4, pp. 1122–1127.

Hadas, Amos (2004). "Seedbed preparation: The soil physical environment of germinating seeds". In: *Handbook of seed physiology: Applications to agriculture*, p. 480.

Ham, Jay M and RS Senock (1992). "On the measurement of soil surface temperature". In: *Soil Science Society of America Journal* 56.2, pp. 370–377.

Hardegree, SP and SS Van Vactor (1999). "Predicting germination response of four cool-season range grasses to field-variable temperature regimes". In: *Environmental and Experimental Botany* 41.3, pp. 209–217.

Hardegree, Stuart P (2006). "Predicting germination response to temperature. I. Cardinal-temperature models and subpopulation-specific regression". In: *Annals of Botany* 97.6, pp. 1115–1125.

Haverkamp, Roland et al. (1977). "A comparison of numerical simulation models for one-dimensional infiltration". In: *Soil Science Society of America Journal* 41.2, pp. 285–294.

Hayhoe, HN et al. (1993). "Tillage effects on corn emergence rates". In: *Soil and Tillage Research* 26.1, pp. 45–53.

Herner, Robert C (1990). "The effects of chilling temperatures during seed germination and early seedling growth". In: *Chilling injury of horticultural crops*, pp. 51–69.

Hillel, Daniel (2003). *Introduction to environmental soil physics*. Academic press.

Homma, Toshimitsu and Andrea Saltelli (1996). "Importance measures in global sensitivity analysis of nonlinear models". In: *Reliability Engineering & System Safety* 52.1, pp. 1–17.

Hou, Ting et al. (2015). "Parameter sensitivity analysis and optimization of Noah land surface model with field measurements from Huaihe River Basin, China". In: *Stochastic environmental research and risk assessment* 29.5, pp. 1383–1401.

Iziomon, MOSES G, HELMUT Mayer, and ANDREAS Matzarakis (2003). "Downward atmospheric longwave irradiance under clear and cloudy skies: Measurement and parameterization". In: *Journal of Atmospheric and Solar-Terrestrial Physics* 65.10, pp. 1107–1116.

Jame, YW and HW Cutforth (2004). "Simulating the effects of temperature and seeding depth on germination and emergence of spring wheat". In: *Agricultural and Forest Meteorology* 124.3, pp. 207–218.

Jamieson, PD et al. (1998). "Sirius: a mechanistic model of wheat response to environmental variation". In: *European Journal of Agronomy* 8.3, pp. 161–179.

Jones, James W et al. (2003). "The DSSAT cropping system model". In: *European journal of agronomy* 18.3, pp. 235–265.

Jones, William J and Nadezhda D Ananyeva (2001). "Correlations between pesticide transformation rate and microbial respiration activity in soil of different ecosystems". In: *Biology and Fertility of Soils* 33.6, pp. 477–483.

Kearney, Michael R et al. (2014). "Microclimate modelling at macro scales: a test of a general microclimate model integrated with gridded continental-scale soil and weather data". In: *Methods in Ecology and Evolution* 5.3, pp. 273–286.

Lamboni, Matieyendou, David Makowski, et al. (2009). "Multivariate global sensitivity analysis for dynamic crop models". In: *Field Crops Research* 113.3, pp. 312–320.

Lamboni, Matieyendou, Hervé Monod, and David Makowski (2011). "Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models". In: *Reliability Engineering & System Safety* 96.4, pp. 450–459.

Lipiec, Jerzy, Artur Nosalewicz, and Jacek Pietrusiewicz (2011). "Crop Responses to soil physical conditions". In: *Encyclopedia of Agrophysics*. Springer, pp. 167–176.

Liu, Weidong et al. (2004). "Response of corn grain yield to spatial and temporal variability in emergence". In: *Crop Science* 44.3, pp. 847–854.

Ma, L et al. (2001). "Integrating system modeling with field research in agriculture: Applications of the Root Zone Water Quality Model (RZWQM)". In: *Advances in Agronomy* 71, pp. 233–292.

Marshall, Theo John, John Winspere Holmes, and Calvin W Rose (1996). *Soil physics*. Cambridge University Press.

Martano, Paolo (2000). "Estimation of surface roughness length and displacement height from single-level sonic anemometer data". In: *Journal of Applied Meteorology* 39.5, pp. 708–715.

Mary, Bruno et al. (2009). *Conceptual basis, formalisations and parameterization of the STICS crop model*. Quae.

McDonald, Miller B (1994). "Seed germination and seedling establishment". In: *Physiology and determination of crop yield* physiologyandde, pp. 37–60.

Mellander, Per-Erik, Mikaell Ottosson Löfvenius, and Hjalmar Laudon (2007). "Climate change impact on snow and soil temperature in boreal Scots pine stands". In: *Climatic Change* 85.1, pp. 179–193.

Monteith, John and Mike Unsworth (2013). *Principles of Environmental Physics: Plants, Animals, and the Atmosphere*. Academic Press.

Morris, Max D (1991). "Factorial sampling plans for preliminary computational experiments". In: *Technometrics* 33.2, pp. 161–174.

Müller, Christoph (1999). *Modelling soil-biosphere interactions.* Cabi Publishing.

Nagel, Kerstin A et al. (2009). "Temperature responses of roots: impact on growth, root system architecture and implications for phenotyping". In: *Functional Plant Biology* 36.11, pp. 947–959.

Parton, William J and Jesse A Logan (1981). "A model for diurnal variation in soil and air temperature". In: *Agricultural Meteorology* 23, pp. 205–216.

Patil, Nitin Gorakh and Surendra Kumar Singh (2016). "Pedotransfer functions for estimating soil hydraulic properties: A review". In: *Pedosphere* 26.4, pp. 417–430.

Petropoulos, G et al. (2009). "A global Bayesian sensitivity analysis of the 1d SimSphere soil–vegetation–atmospheric transfer (SVAT) model using Gaussian model emulation". In: *Ecological Modelling* 220.19, pp. 2427–2440.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Richard, Guy et al. (2001). "Effect of compaction on the porosity of a silty soil: influence on unsaturated hydraulic properties". In: *European Journal of Soil Science* 52.1, pp. 49–58.

Ritchie, SW, JJ Hanway, and GO Benson (1992). "How a corn plant grows". In: *Iowa State Univ Sci Tech Coop Ext Serv Rep* 48.

Roman, Erivelton S, Stephen D Murphy, and Clarence J Swanton (2000). "Simulation of Chenopodium album seedling emergence". In: *Weed Science* 48, pp. 217–224.

Ross, PJ (2003). "Modeling soil water and solute transport—fast, simplified numerical solutions". In: *Agronomy journal* 95.6, pp. 1352–1361.

Sacks, William J and Christopher J Kucharik (2011). "Crop management and phenology trends in the US Corn Belt: impacts on yields, evapotranspiration and energy balance". In: *Agricultural and Forest Meteorology* 151.7, pp. 882–894.

Sainte-Marie, Julien, Gautier Viaud, and Paul-Henry Cournède (2017). "Indices de Sobol généralisés aux variables dépendantes: tests de performance de l'algorithme HOGS couplé à plusieurs estimateurs paramétriques". In: *Journal de la Société Française de Statistique* 158.1, pp. 68–89.

Saltelli, Andrea and Paola Annoni (2010). "How to avoid a perfunctory sensitivity analysis". In: *Environmental Modelling & Software* 25.12, pp. 1508–1517.

Saltelli, Andrea, Stefano Tarantola, et al. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models*. John Wiley & Sons.

Saxton, KE and WJ Rawls (2006). "Soil water characteristic estimates by texture and organic matter for hydrologic solutions". In: *Soil Science Society of America Journal* 70.5, pp. 1569–1578.

Schneider, EC and SC Gupta (1985). "Corn emergence as influenced by soil temperature, matric potential, and aggregate size distribution". In: *Soil Science Society of America Journal* 49.2, pp. 415–422.

Sharpe, Peter JH and Don W DeMichele (1977). "Reaction kinetics of poikilotherm development". In: *Journal of Theoretical Biology* 64.4, pp. 649–670.

Shuttleworth, W James and JS Wallace (1985). "Evaporation from sparse crops-an energy combination theory". In: *Quarterly Journal of the Royal Meteorological Society* 111.469, pp. 839–855.

Simunek, Jirka et al. (2003). "Review and comparison of models for describing non-equilibrium and preferential flow and transport in the vadose zone". In: *Journal of Hydrology* 272.1, pp. 14–35.

Simunek, J et al. (2013). "The HYDRUS-1D software package for simulating the movement of water, heat, and multiple solutes in variably saturated media, version 4.17, HYDRUS software series 3". In: *Department of Environmental Sciences, University of California Riverside, Riverside, California, USA*, p. 342.

Soil Survey Division Staff (1993). "Soil survey manual". In: *United States Department of Agriculture*.

Spaeth, Stephen C (1994). "Germination and Seedling Establishment: Discussion". In: *Physiology and Determination of Crop Yield* physiologyandde, pp. 61–63.

Steduto, Pasquale and Theodore C Hsiao (1998). "Maize canopies under two soil water regimes.: I. Diurnal patterns of energy balance, carbon dioxide flux, and canopy conductance". In: *Agricultural and Forest Meteorology* 89.3, pp. 169–184.

Stoll, Marian and Imad Saab (2013). "Soil Temperature and Corn Emergence". In: *Crop insights* 23.1, pp. 1–4.

Unsworth, Michael H and JL Monteith (1975). "Long-wave radiation at the ground I. Angular distribution of incoming radiation". In: *Quarterly Journal of the Royal Meteorological Society* 101.427, pp. 13–24.

Van Bavel, CHM and DI Hillel (1976). "Calculating potential and actual evaporation from a bare soil surface by simulation of concurrent flow of water and heat". In: *Agricultural Meteorology* 17.6, pp. 453–476.

Vereecken, Harry et al. (2016). "Modeling soil processes: Review, key challenges, and new perspectives". In: *Vadose zone journal* 15.5.

Vinocur, Marta G and Joe T Ritchie (2001). "Maize leaf development biases caused by air–apex temperature differences". In: *Agronomy Journal* 93.4, pp. 767–772.

Wallach, Daniel et al. (2014). *Working with dynamic crop models second edition: methods, tools and examples for agriculture and environment*. Elsevier.

Walter, Ivan A et al. (2000). "ASCE's standardized reference evapotranspiration equation". In: *Watershed management and operations management 2000*, pp. 1–11.

Wang, H et al. (2009). "Predicting the time to 50% seedling emergence in wheat using a Beta model". In: *NJAS-Wageningen Journal of Life Sciences* 57.1, pp. 65–71.

Wang, Kaicun et al. (2005). "Variation of surface albedo and soil thermal parameters with soil moisture content at a semi-desert site on the western Tibetan Plateau". In: *Boundary-Layer Meteorology* 116.1, pp. 117–129.

Weaich, Karl, Keith L Bristow, and Alfred Cass (1996). "Simulating maize emergence using soil and climate data". In: *Agronomy Journal* 88.4, pp. 667–674.

Whalley, WR and William E Finch-Savage (2011). "Crop emergence, the impact of mechanical impedance". In: *Encyclopedia of Agrophysics*. Springer, pp. 163–167.

Wickham, Hadley (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-0-387-98140-6. URL: http://ggplot2.org.

WMO (2012). *Guide to Agricultural Meteorological Practices*. World Meteorological Organization, Geneva, Swtizerland.

Yang, Yuhong (2005). "Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation". In: *Biometrika* 92.4, pp. 937–950.

Zhang, Lu, Glen R Walker, and Warrick R Dawes (2002). "Water balance modelling: concepts and applications". In: *ACIAR Monograph Series* 84, pp. 31–47.

Zhang, Yu et al. (2005). "Soil temperature in Canada during the twentieth century: Complex responses to atmospheric climate change". In: *Journal of Geophysical Research: Atmospheres* 110.D3.

université PARIS-SACLAY

ÉCOLE DOCTORALE
INTERFACES
Approches Interdisciplinaires : Fondements, Applications et Innovation

Titre: Modélisation de la température du sol avec un bilan d'énergie, application à la prédiction de l'émergence du maïs (*Zea mais*)

Mots-clés: Modèle de sol, bilan d'énergie, analyse de sensibilité, prédiction de l'émergence, maïs

La croissance en début de cycle des grandes cultures est principalement influencée par la température et la teneur en eau du sol. Nous avons développé un modèle capable de prédire ces variables grâce à l'utilisation de données climatiques largement disponibles. Des analyses de la sensibilité du modèle nous ont permis d'identifier les composants qui contribuent à son incertitude. Après calibration, une erreur moyenne relative de moins de 10 % est constatée pour la température et la teneur en eau à 30 cm de profondeur. Dans des conditions de semis standard en Suisse, l'émergence du maïs a été mieux prédite en utilisant notre température de sol simulée plutôt que la température de l'air, plus couramment utilisée. Ce travail est une application d'un modèle biophysique complexe à un problème agronomique. Les résultats participeront à l'optimisation de l'effort de sélection des variétés tolérantes au froid. Deux pistes de recherche peuvent être considérées pour des futurs travaux: une meilleure modélisation de l'évaporation et une décomposition de l'émergence.

Title : Modelling soil temperature with an energy balance model, application to prediction of maize (*Zea mais*) emergence

Keywords : Soil model, energy balance, sensitivity analysis, emergence prediction, maize

The beginning of crop growth is influenced by soil temperature and water content near the surface. We have developed a model that predicts the local temperature and water content surrounding the seed using easily available meteorological data. Our global sensitivity analysis helped us identify the components of the model with the largest contribution to the output uncertainty. After calibration, the model showed less than 10 % relative error for temperature and water content at 30 cm. In standard sowing conditions in north-western Switzerland, the emergence was better predicted when using our simulated seed bed temperature than air temperature, the classical proxy variable. Combining the emergence model with soil temperature simulation, an accurate prediction of emergence was achieved. This work is an example of applying complex biophysics model for understanding an agronomic problem. The results of this work will participate in optimising breeding efforts for cold-tolerant crop varieties. Future investigations should consider a finer modelling of processes for evaporation and emergence.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France