Estimation of Land Surface Water and Energy Balance Flux Components and Closure Relation Using Conditional Sampling

Leila Farhadi

Department of Civil and Environmental Engineering
Massachusetts Institute of Technology
Cambridge, MA

Structure of the Presentation

- Motivation
- Objective
- Methodology
- Synthetic case (Test robustness)
- Field site test (Ameriflux data)
- Remote sensing application (West Africa)
- Conclusion

Motivation

- ✓ All land surface models (*LSMs*) include numerical treatment of heat and moisture diffusion in the Soil-Vegetation-Atmosphere continuum

Vertical Moisture Dynamics in the soil

- Diffusion of Moisture in the column
- Flux of Evaporation (E) at surface, Water Drainage (BC)

Vertical temperature Dynamics in the soil

- Diffusion of heat in the column
- Residual of net radiation , Sensible and Latent heat flux at the surface (BC)

- ✓ Latent heat (common boundary condition) needs to be consistent.

Motivation

- Links (and closure) for surface water and energy balance

Surface water & energy balance linked through **latent Heat** flux

The need for consistency requires a **closure function**

Over bare soil this closure often can take the form of soil moisture (s) dependent empirical functions:

1. $\beta(s) = \dfrac{\text{Actual Evaporation}}{\text{Potential Evaporation}}$
2. $h(s) = \text{Humidity at soil-atmosphere interface}$

In plant continuum, it often takes the form of :

1. Soil moisture-dependent root water extraction resistance r_g
2. Stomatal resistance r_s due to plant water stress

Motivation

Importance of the closure relationship

- All Land Surface Models – LSMs include (**explicitly or implicitly**) a form of this soil moisture dependent closure
- Land response to radiative forcing and partitioning of available energy are **critically dependent on the functional form (<u>shape</u>)** of the closure relationship.
- The function affects the **surface fluxes**, the influence reaches through the boundary layer and manifests itself in the lower atmosphere weather
- Important as these closure functions are, they still remain essentially **empirical and untested** across diverse soil and vegetation conditions.

Objective

The overarching goal of this project is to develop a <u>scale free technique</u> to better estimate the unknown parameters (e.g. the flux components) of water and energy balance equation (and the closure relation between the two) using <u>discrete observation.</u>

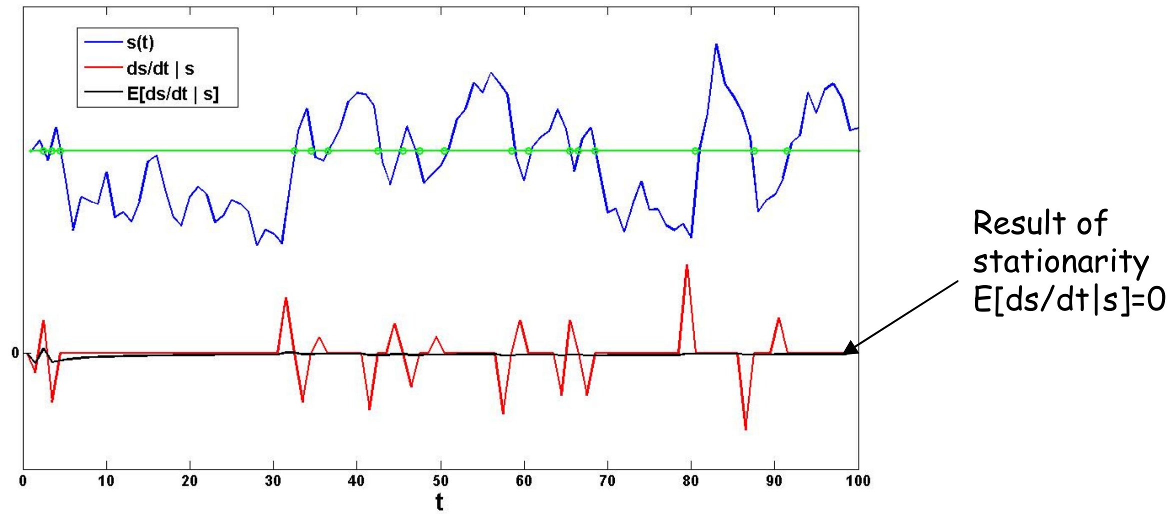
✓Estimation procedure is distinct from **"calibration"** since only forcing (P, $R_{in}^{\downarrow}$) and state (s, T_s) observations are used. No information about fluxes (e.g. flux towers) is needed.

✓The method is scale- free, i.e. it can be applied to diverse scales of states and forcing (remote sensing applications)

✓The method can be applied to diverse climates and land surface conditions using remotely sensed measurements.

Methodology

✓The overall approach is based on the conditional sampling method of Salvucci(2001) which exploits the fact that the expected value of increments of seasonally detrended soil moisture (s) conditioned on moisture is zero (E[ds/dt|s]=0) for stationary systems.



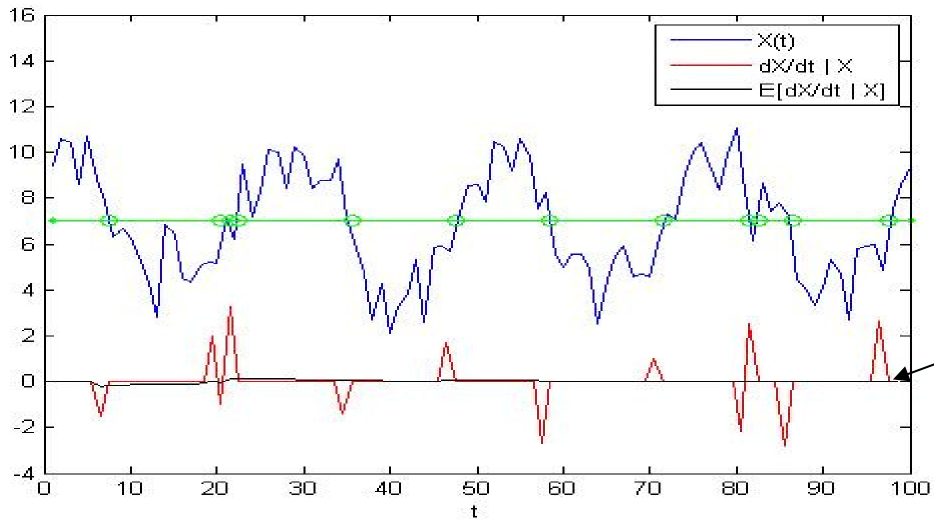
- [Mathematical proof: conditional expectation minimizes least squared loss function]

✓Model parameters (sum of evaporation and drainage) are estimated by matching the soil moisture conditional expectation of modeled fluxes to soil moisture conditional expectation of precipitation. (E[Sum of fluxes|s]=E[P|s])

- Problem in distinguishing evaporation from drainage

Methodology

✓ in my thesis I prove that for <u>seasonally (periodically) stationary</u> process (X_t), The relation $E[dX_t/dt|X_t]=0$ holds



Result of Stationarity $E[dX_t/dt|X_t]=0$

Soil moisture (S) and soil surface temperature(Ts) are seasonally stationary, Thus: $E[dS/dt|S]=0$ and $E[dTs/dt|Ts]=0$

Thus by applying to the two balance equations we can separate out drainage from evaporation(Note: both hydrologic fluxes important but not measured widely)

Methodology

Example Moisture Diffusion Eq :

$$1\frac{ds}{dt} = P - ET - D + CR$$

$E[ds/dt|s]=0$

$$E[P\,|\,s] = E[ET\,|\,s] - E[D\,|\,s] + E[CR\,|\,s]$$

Example Heat Diffusion Eq :

$$\frac{dT_s}{dt} = \left(\frac{2\sqrt{\pi\omega}}{P_i}\right)\left[\underbrace{R_n}_{R_{in}^{\downarrow} - R_{out}^{\uparrow}} - LE - H\right] - 2\pi\omega(T_s - T_D)$$

$E[dT_s/dt|T_s]=0$

$$E\left[R_{in}^{\downarrow}\,|\,T_s\right] =$$

$$E[LE\,|\,T_s] + E[H\,|\,T_s] + E[P_i.\frac{\pi\omega}{\sqrt{\pi\omega}}(T_s - T_D)\,|\,T_s] + E[\varepsilon\sigma T_s^4\,|\,T_s]$$

Process	Unknown Par's	Form
Drainage	K_s, c	$D(s)=K_s.s^c$
Capillary rise	w, n	$CR(s)=w.s^n$
Thermal Inertia	P_i	f(soil type, soil moisture)
Neutral turbulent heat coefficient (C_{HN})	α, β	$C_{HN}= \exp(\alpha LAI+\beta)$
Evaporative Fraction	a,θ_s,θ_w	$EF=1-\exp(-a(\theta/\ \theta_s - \theta_w/\ \theta_s)$

Methodology

- S and T_s are discretized to n and m ranges respectively

$$\begin{cases} E\left[\rho LP \middle| \bar{s}_i\right] = M_w(i) + \xi_1 \\ E\left[R_{in}^{\downarrow} \middle| \bar{T}_{sj}\right] = M_e(j) + \xi_2 \end{cases}$$

Units: W/m^2

Where:

Forcing uncertainty:

$$\begin{cases} \xi_1 \sim N(0, (\rho L)^2 \sigma_P^2); \\ \xi_2 \sim N(0, \sigma_{R_{in}^{\downarrow}}^2)); \end{cases} \qquad Cov(\xi_1, \xi_2) = 0$$

- **The cost function:**

$$J = \frac{1}{2}(d - m)^T.A.(d - m)$$

- d : Vector of data (n+m x1)
- M: Vector of Model Counterparts (n+m x1) ⇨ Unknown parameters: $[\theta_S, \theta_W, a, C_{HN}, P_i, K_S, c, w, n]$
- A=σ^{-2}I (n+m x n+m)

$d : \left[E\left[\rho LP \middle| s_1\right], E\left[\rho LP \middle| s_2\right], \ldots\ldots E\left[R_{in}^{\downarrow} \mid T_{s1}\right], E\left[R_{in}^{\downarrow} \mid T_{s2}\right], \ldots\right]$

- **Note:** Estimation procedure is distinct from " calibration" since only forcing data (P, $R_{in}^{\downarrow}$) and state observation (s, Ts) are used. No information on fluxes (e.g. Problematic evaporation and drainage) is needed.

Methodology

- **Minimize nonlinear Cost Function J**

- **Estimation of Uncertainty Bounds**

- Inverse of Hessian of Cost function is an approximation for the Covariance matrix.

- Covariance matrix is used to estimate the uncertainty of any model output and thus determine which aspects of the model are poorly determined by the data

- First Order Second Moment propagation of uncertainty (FOSM) analysis, or Monte Carlo method is used to define the uncertainty around different flux components.

❑ Determining the sufficiency of a particular data set to determine the model state

1- Uncertainty of each individual parameter should be reasonable in physical sense.

2-Uncertainty of the least well-determined combination of variables given by the eigenvectors of Hessian should be reasonable.

3- Correlation matrix between unknown variables should be reasonable.

-Linear dependency between variables is a sign of discrepancy between data and model

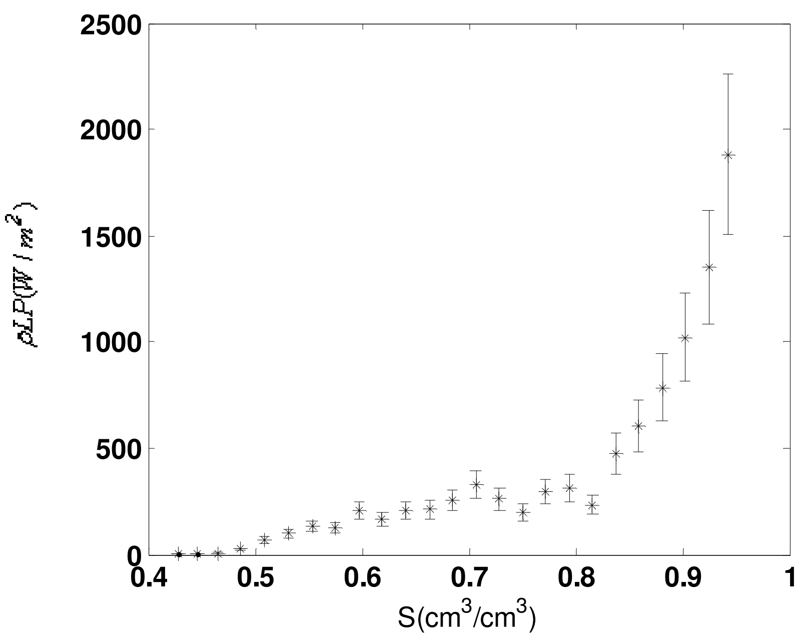
-Best scenario: The correlation between all the parameters is small,

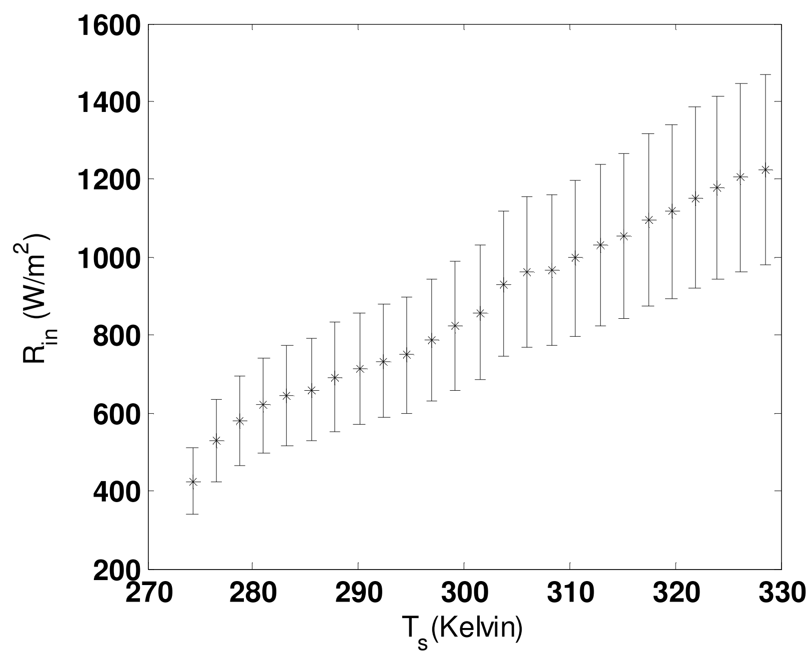
-The next best scenario: High correlation is only between parameters representing one flux type and suggests the model is robust with regard to flux components

-The worse scenario: The correlation between parameters representing different flux types is high and/ or physically not meaningful.

Synthetic data case

- 30 year of hourly meteorological data for humid climate of Charlotte North Carolina obtained from " Solar and Meteorological Surface Observational Network" (SAMSON) [National Climate Data Center];
- Simultaneous Heat and Water(SHAW) model was used to derive consistent hourly time series of state and fluxes
- Assume 20% precipitation and radiation error.
- Daily Water balance is coupled to midday energy balance→ Unknown parameters are obtained.





⇨ 50 equations

1- Optimization with 9 unknown variables

$$\alpha = [\text{Drainage flux par's, Capillary rise par's}, C_{HN} \text{ function par's}, P_i, \text{EF function Par's}]$$

$$\Rightarrow \quad \alpha = \left[K_s, c, w, n, C_{HN}(=\exp(\beta)) \; , P_i, a, S_w\left(=\frac{\theta_w}{\theta_s}\right), \theta_s \right]$$

- P_i is considered a constant effective value
- Optimization is insensitive to the value of thermal inertia (P_i) . The optimum is either the same as the initial guess for P_i value picked from the physically accepted range or a value in its close proximity. (This is consistent with findings of other studies)
- Thermal inertia is a property of soil composition, porosity and soil moisture. Murray and Verhoeff (2007) method is used.

Process	Unknown Par's	Form
Drainage	K_s, c	$D(s)=K_s.s^c$
Capillary rise	w, n	$CR(s)=w.s^n$
Neutral turbulent heat coefficient (C_{HN})	α, β	$C_{HN}= \exp(\alpha LAI+\beta)$
Evaporative Fraction	a, θ_s, θ_w	$EF=1-\exp(-a(\theta/ \theta_s- \theta_w/ \theta_s)$

2- Optimization with 8 unknown variables

$$\alpha = [K_s, c, w, n, C_{HN}, a, S_w, \theta_s]$$

Estimated model variables for the system with 8 unknown variables

Par	Dimension	Opt solution ± std	Relative error (%)
K_s	m/hr	0.0021±0.0003	14.3%
w	m/hr	0±1.257	$\rightarrow \infty$
C_{HN}	[ ]	0.0032±0.0002	6.25
a	[ ]	6.44±0.14	2.15
n	[ ]	146.11±152.1	104.09
S_w	cm^3/cm^3	0.46±0.0014	0.3
C	[ ]	9.41±0.2544	2.7
θ_s	cm^3/cm^3	0.474±0.0000	0.00

Uncertainty of combination of variables determined by eigen vectors

Eigen Values	Relative error (%) $\frac{\sigma_{e_i^T X}}{E[e_i^T X]} \times 100$
4.3225e-005	104.1
0.6147	1920.5
17.6851	2.72
54.6745	1.82
7.7479e+005	0.27
5.5736e+007	0.89
1.6690e+008	1.41
2.6540e+017	0.00

- Data is insufficient to determine the model states with acceptable accuracy- linear dependency is generated as seen in the correlation matrix

Correlation Matrix between variables of the system

	K_s	w	C_{HN}	a	n	S_w	C	θ_s
K_s	1.00							
w	-0.124	1.00						
C_{HN}	-0.78	0.13	1.00					
a	0.89	-0.15	-0.89	1.00				
n	-0.00	0	0.00	-0.00	1.00			
S_w	0.55	-0.09	-0.33	0.58	-0.00	1.00		
C	-0.11	0.02	0.097	-0.12	-0.36	-0.10	1.00	
θ_s	-0.94	0.19	0.82	-0.97	0.00	-0.62	0.12	1.00

- W ~0; its variation is high; in addition, n is large, S^n is very small (0<S<1) Thus, <u>WS^n is negligible</u>

- Due to high linearity btw " Ks ,θ_s" and " a, θs" Taking θs out of the parameter space will improve the condition number of Hessian ;

  (**<u>replace : θs ~ max(recorded θ)</u>**)

- This is not a sample correlation but derived from Hessian and related to shape of J around minimum. Used for diagnosing collinearity and has no statistical significance.

3- Optimization with 5 unknown variables

$$\alpha = [K_s, C, C_{HN}, a, S_w]$$

Estimated model variables for the system with 5 unknown variables

Par	Dimension	Opt solution ± std	Relative error (%)
K_s	m/hr	0.0020±0.0003	15%
C_{HN}	[ ]	0.0028±0.0004	14.28%
a	[ ]	6.55±0.49	7.52%
S_w	cm^3/cm^3	0.416±0.011	2.6%
C	[ ]	9.05±0.3	3.31%

Uncertainty of combination of variables determined by eigen vectors

Eigen Values	Relative error (%) $\frac{\sigma_{e_i^T X}}{E[e_i^T X]} \times 100$
3.89	13.43
13.22	2.61
9888.5	1.77
1.2277e+007	19.34
2.5527e+007	8.7
3.89	13.43

Correlation Matrix between variables

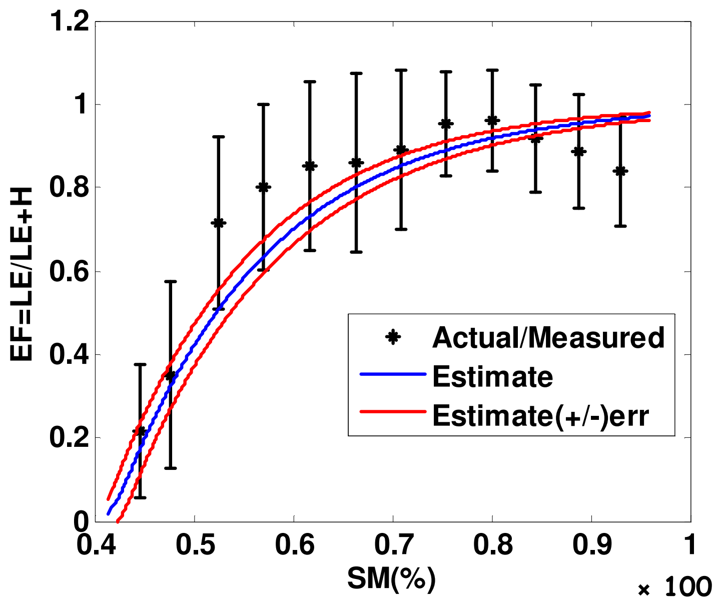
	K_s	C_{HN}	a	S_w	C
K_s	1.00				
C_{HN}	0.18	1.00			
a	-0.45	-0.64	1.00		
Sw	-0.11	0.40	-0.18	1.00	
C	0.7	0.13	-0.32	-0.23	1.00

-Parameters are estimated reasonably well

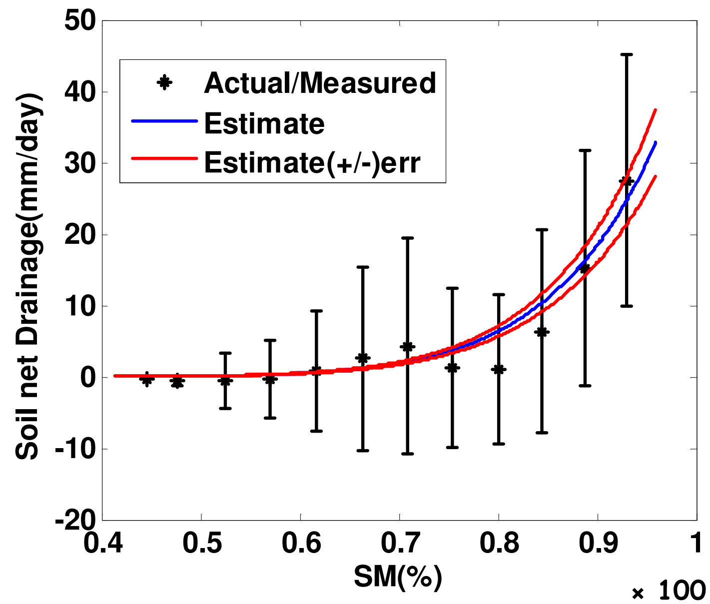
-High correlation between Ks and C is the sign of robust estimation of Drainage.
Ks increases→ $Ks.S^c$ increases
C increases→ $Ks.S^c$ Decreases

-"C_{HN} and a" parameters have negative Correlation;
Increase in parameter "C_{HN}" → Increase in estimated sensible heat flux
Decrease in parameter " a" → Decrease in estimated Latent heat flux

This result is physically meaningful, since the sum of sensible heat flux (H) and Latent heat flux (LE) represent the available energy to the system (Rn-G) and when the available energy to the system is constant, an increase in H results in a decrease in LE and vice versa.



Comparing Actual EF and model estimate of EF



Comparing Actual /measured net soil water flux and its model counterpart

- ✓ The closure function EF(s)=LE/LE+H is well estimated in this synthetic data set
- ✓ This approach is robust at point scale

Field tests

3 Field sites investigated are as follow:

- ✓ Vaira Ranch, grassland, CA, Mediteranation climate
- ✓ Audubon Research ranch, grassland, AZ, Arid/semi arid climate
- ✓ Santa Rita Mesquite , woody savannah, AZ, Arid/semi arid climate

❑ **Source of Data (estimation and validation)**

- **AMERIFLUX** : Soil water content θ ; Wind speed(u), Air temperature (T_a), Soil surface Temperature (T_s), Precipitation (P), Net radiation (R_n)

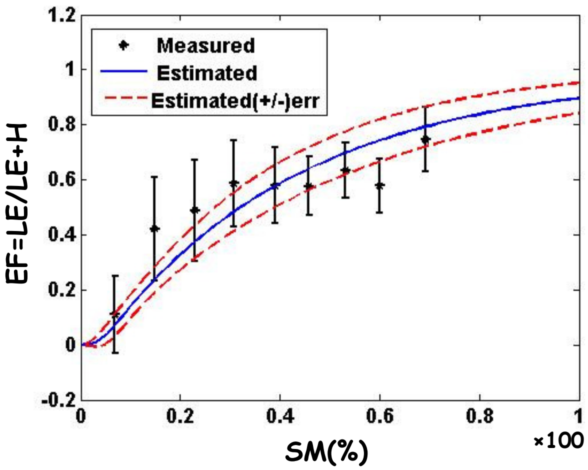
- **MODIS:** LAI

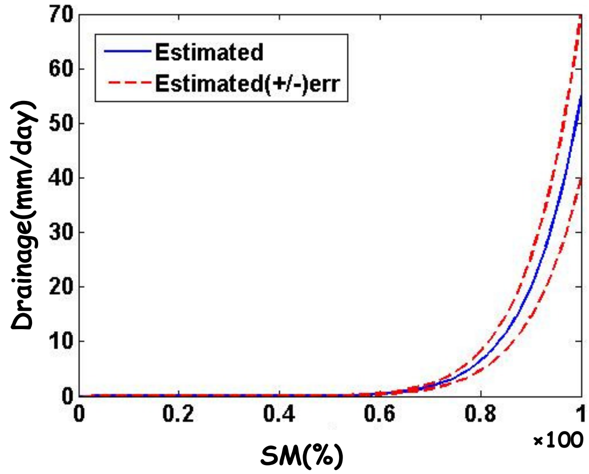
❑ Daily water balance equation is linked to midday energy balance equation

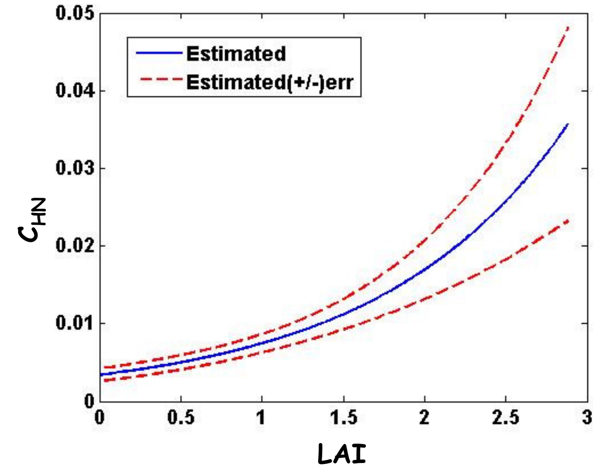
❑ Error of data

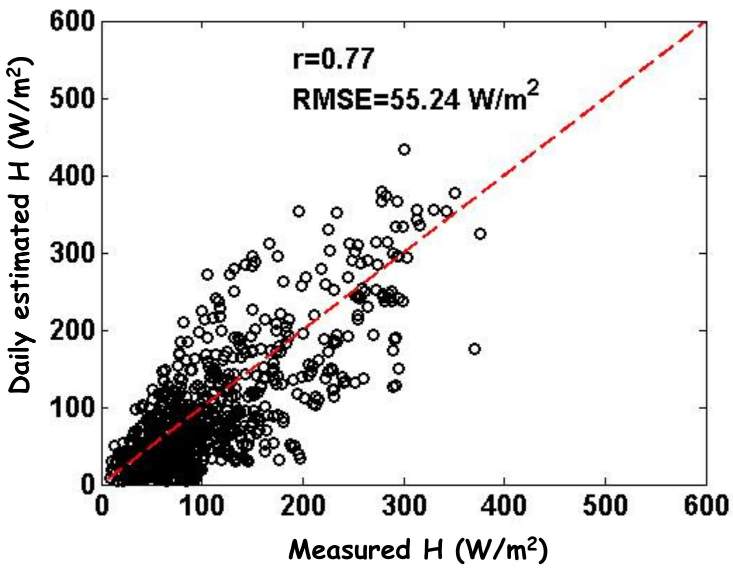
$\varepsilon_{E[\,P|s]} \sim N(0, (6\%\, E[\,P|S])^2)$; $\varepsilon_{E[\,Rin|s]} \sim N(0, (8\%\, E[\,R_{in}|T_s])^2)$;

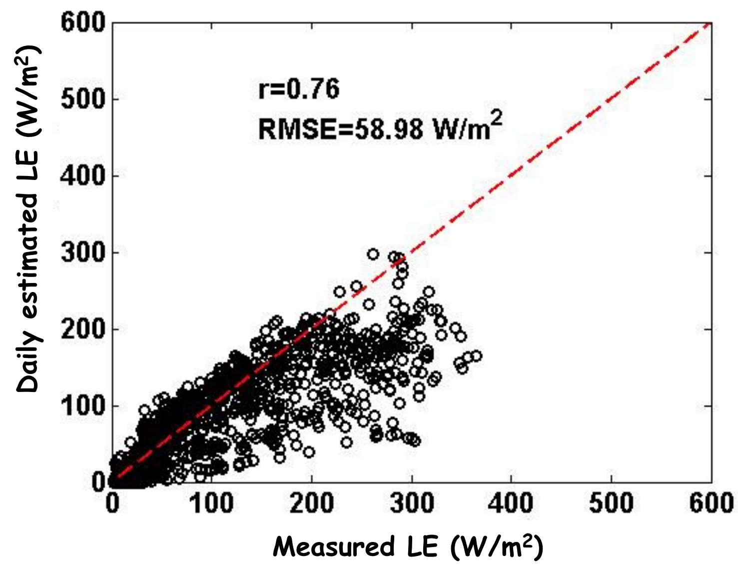
✓ Vaira Ranch, grassland, CA



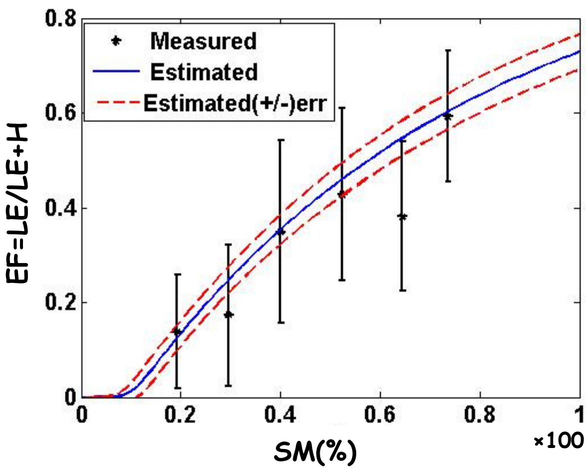


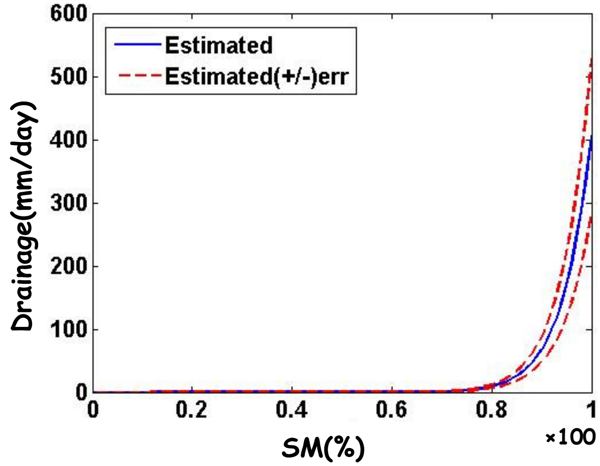


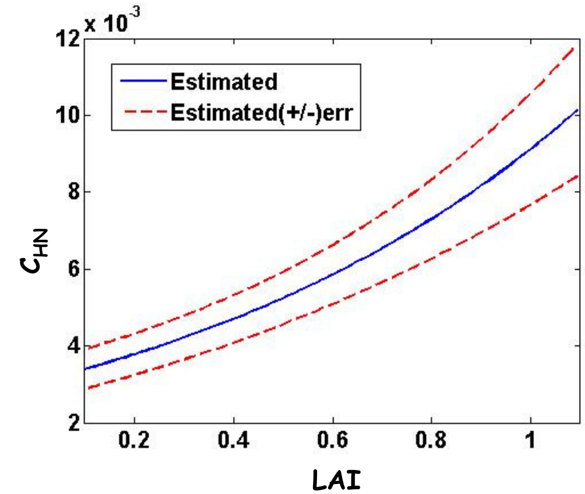


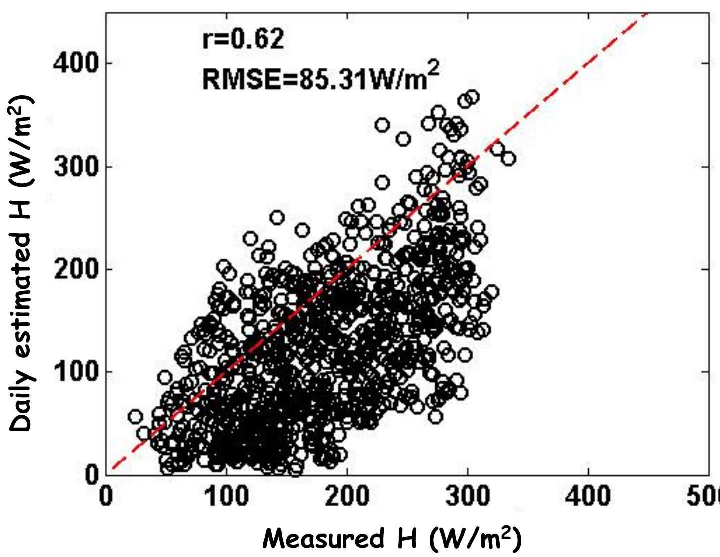


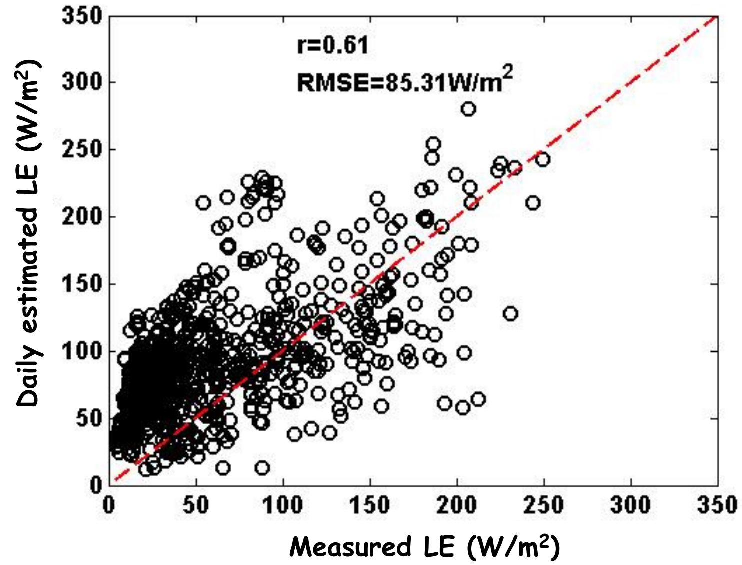
✓Audubon Research Ranch, grassland, AZ



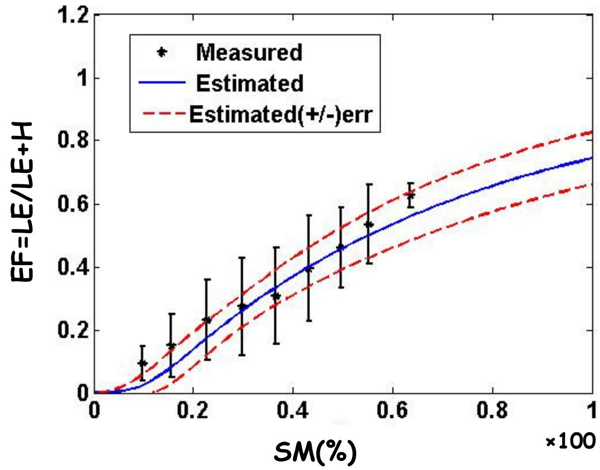


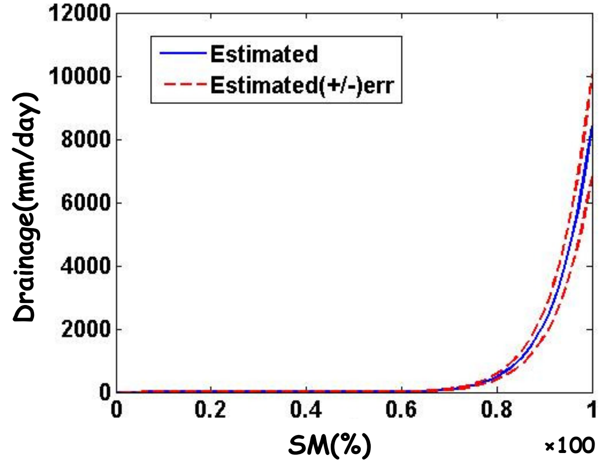


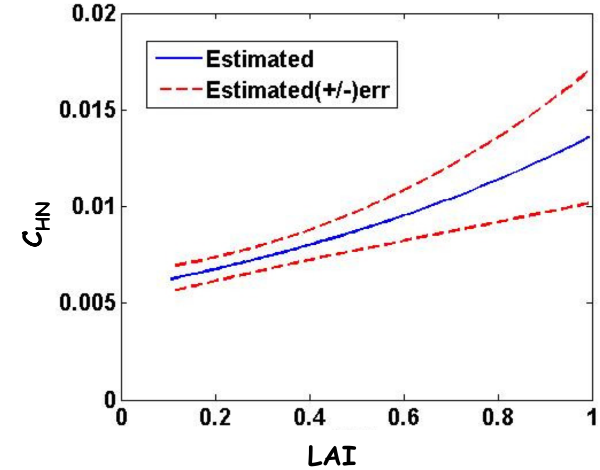


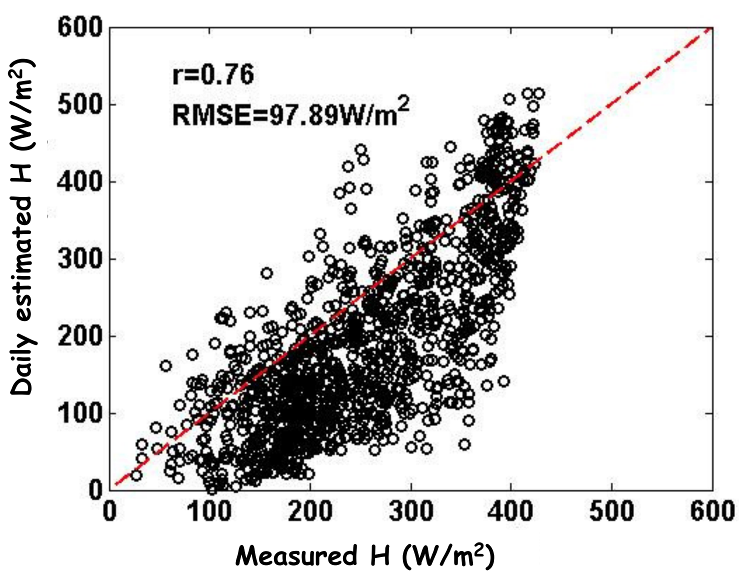


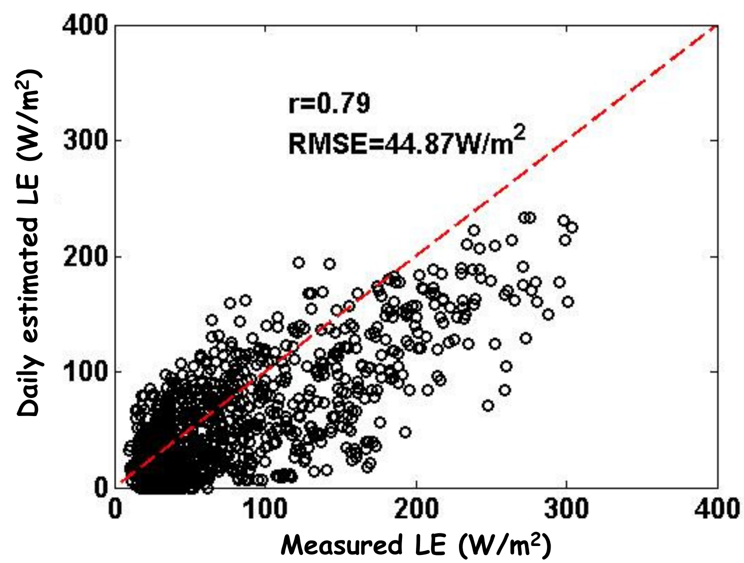
✓ Santa Rita Mesquite, woody savannah, AZ



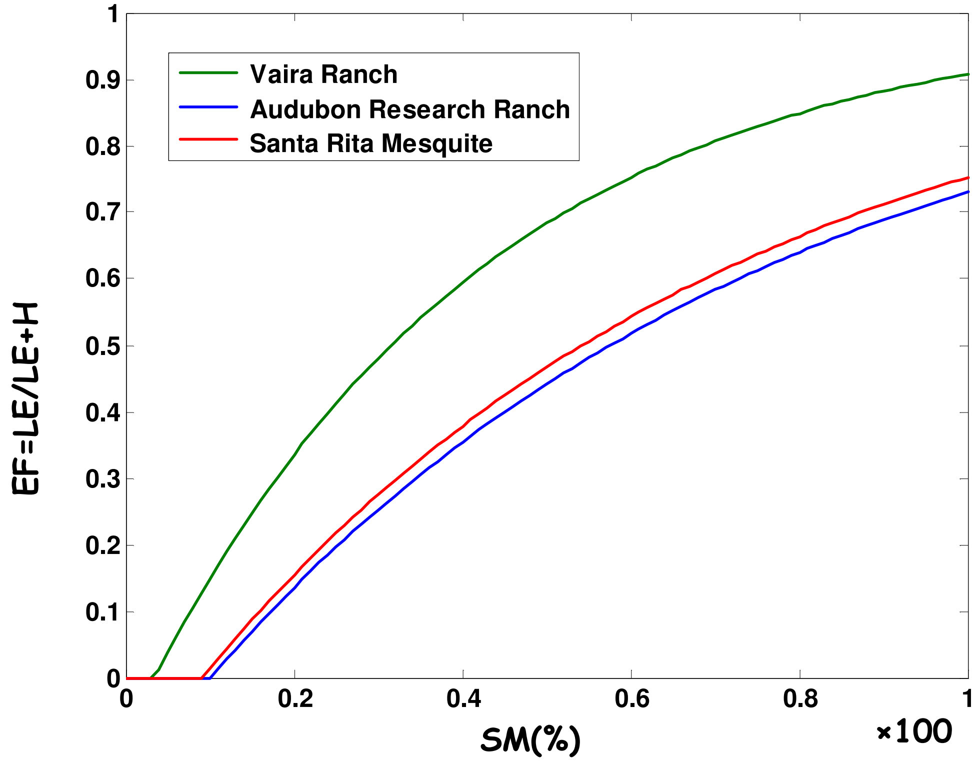








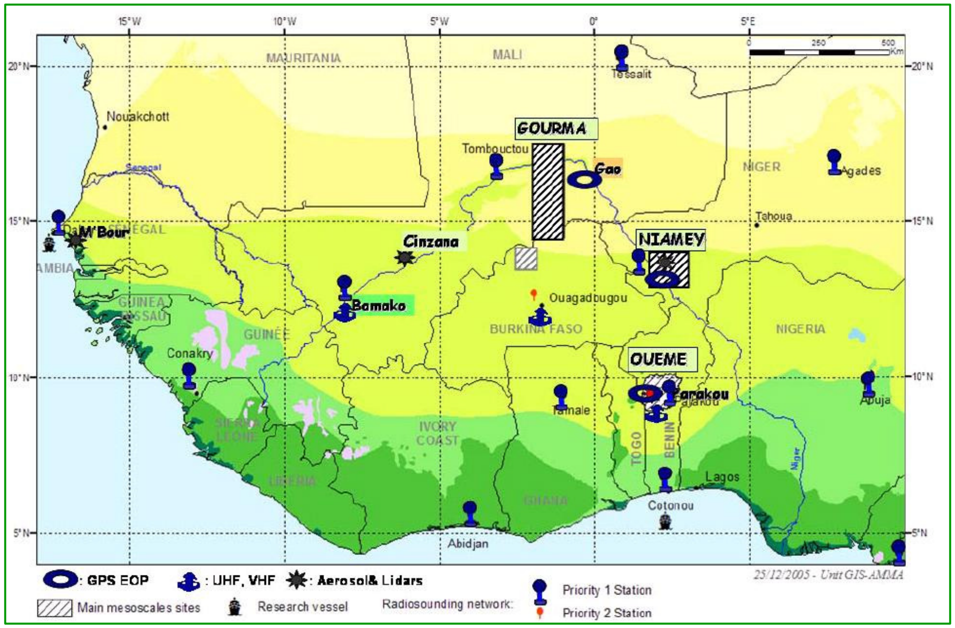
- Closure function (EF) at different field sites



❑ EF is distinct for each site ; hence it needs to be mapped using remote sensing.

Remote Sensing

- The Gourma meso scale site in Mali of West Africa is an area located in the Gourma region. This region stretches from the loop of Niger River southward down the border region with Burkina- Faso. Location of the Gourma meso- scale site is (14.5-17.5 ^{O}N, 1-2 ^{O}W). Thus it is a 40,000 km^{2} area.



Reference [AMMA Documentation]

- **Why this region**

1- vast spatial and temporal coverage, remote sensing data which give access to surface variables in this area;

2- Gourma region is located in Sahara & Sahelian-Sahara climate; Evaporation is generally water limited (EF=EF(S)) ;

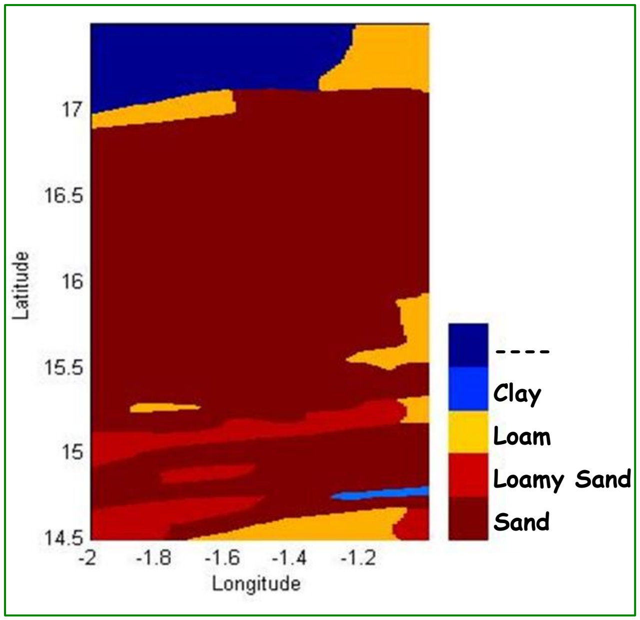
3- Runoff can be considered negligible in most areas;

Gourma site- Source of data

Variable	Definition	Source of Data	Spatial Resolution	Temporal Resolution
u	Wind speed	AMMA-ECMWF	50km	6hr
T_a	Air Temp	AMMA-ECMWF	50km	6hr
Rs	Down Welling short wave	SEVIRI	3km	15 min
α	albedo	SEVIRI	3km	Daily
LAI	Leaf Area Index	SEVIRI	3km	Daily
P	Precipitation	PERSIANN	4km	hrly , daily
s	Soil Moisture	AMSR-E	25km	1:30 pm;1:30 am
T_s	Surface Temperature	SEVIRI	3km	15 min
T_D	Soil Deep Temperature	Filtering Ts	3km	15 min

- 2008 data sets were selected.
- Data were aggregated to present daily time step (temporal resolution).
- Data are interpolated on a 3km*3km grid (spatial resolution)

- In order to reduce dimensionality, categorical soil maps are used to find common soil hydraulic parameters in similar regions (alternative dimensionality reduction approaches can be applied)



Sand
Loamy Sand
Clay
Loam

▪Water Bodies and a 9km area around them are neglected
(removes shallow water table locations where capillary rise may be significant)

-Also runoff can be considered negligible in most areas

Sand region....

Sand pixels ~ 81% of the pixels corresponding to the 4 different soil categories)

$$\boxed{\alpha = [K_s, \alpha, \beta, a, S_w, C]}$$

Estimated model variables

Par	Dimension	Opt solution $\pm\sigma$	Relative error (%)
K_s	$\frac{m}{hr}$	5.61±13.59	242%
α	[ ]	-5.56±0.06	1.4%
a	[ ]	7.59±1.09	14.4%
S_w	$\frac{cm^3}{cm^3}$	0.14±0.014	9.25%
C	[ ]	10.2±2.47	23.8%
β	[ ]	0.98±0.23	23.2%

Uncertainty of combination of variables determined by eigen vectors

Eigen values	Relative error (%) $\frac{\sigma_{e_i^T X}}{E[e_i^T X]} \times 100$
0.0053	186%
0.768	11.2%
14.4	5.7%
40.65	4.7%
257.51	1.1%
91524	0.5%

Correlation Matrix between variables of the system

	K_s	α	a	S_w	C	β
K_s	1.00					
α	-0.10	1.00				
a	0.22	0.085	1.00			
S_w	0.05	0.48	0.83	1.00		
C	0.98	-0.08	0.38	0.18	1.00	
β	0.12	-0.12	-0.07	-0.02	0.15	1.00

- Collinearity exists between parameter K_s and C

Sand region….

Vector of parameters is reduced to:

$$\boxed{\alpha = [K_s, \alpha, \beta, a, S_w)]}$$

✓ Expectation Maximization (EM) method is used to reduce the parameter space

✓ In EM method **__the aim is__** to maximize the probability (likelihood) of seeing the observed values (find the parameter values which would maximize the likelihood of our observations)

- Select a typical C value based on soil type ; C=2b+3
- Solve the coupled system with 5 Unknowns
- K_s should be within the appropriate range for the soil type

✓ Iterate untill Ks value is consistent with __soil type__

Example of look up table used by LSM Community(Based on Clapp and Hornberger)

Soil Texture	Φ , cm^3/cm^3	K_s, cm/s	$\lvert\Psi_{ae}\rvert$, cm	b
Sand	0.395 (0.056)	1.76×10^{-2}	12.1 (14.3)	4.05(1.78)
Loamy sand	0.410 (0.068)	1.56×10^{-2}	9.0 (12.4)	4.38(1.47)
Sandy loam	0.435 (0.086)	3.47×10^{-3}	21.8(31.0)	4.9(1.75)
Silt loam	0.485 (0.059)	7.2×10^{-4}	78.6(51.2)	5.30(1.96)
Loam	0.451 (0.078)	6.95×10^{-4}	47.8(51.2)	5.39(1.87)
Sandy clay loam	0.420 (0.059)	6.30×10^{-4}	29.9(37.8)	7.12(2.43)
Silty clay loam	0.477 (0.057)	1.70×10^{-4}	35.6(37.8)	7.75(2.77)
Clay loam	0.476 (0.053)	2.45×10^{-4}	63.0(51.0)	8.52(3.44)
Sandy clay	0.426 (0.057)	2.17×10^{-4}	15.3(17.3)	10.4(1.64)
Silty clay	0.492 (0.064)	1.03×10^{-4}	49.0(62.1)	10.4(4.45)
Clay	0.482 (0.050)	1.28×10^{-4}	40.5(39.7)	11.4(3.70)

Sand region….

$$\alpha = [K_s, \alpha, \beta, a, S_w)]$$

Estimated model variables for the system

Par	Dimension	Opt solution± σ	Relative error (%)
K_s	m/hr	0.755±0.0999	13.3%
α	[ ]	-5.69±0.076	1.4%
a	[ ]	4.029±0.452	11.2%
S_w	cm^3/cm^3	0.07±0.02	29.4%
β	[ ]	1.165±0.397	34%

Uncertainty of combination of variables determined by eigen vectors

Eigen values	Relative error (%) $\frac{\sigma_{e_i^T X}}{E[e_i^T X]} \times 100$
4.5338	12.8%
6.6974	13.4%
206.57	4%
358.98	1.1%
3732	14.4%

Correlation Matrix

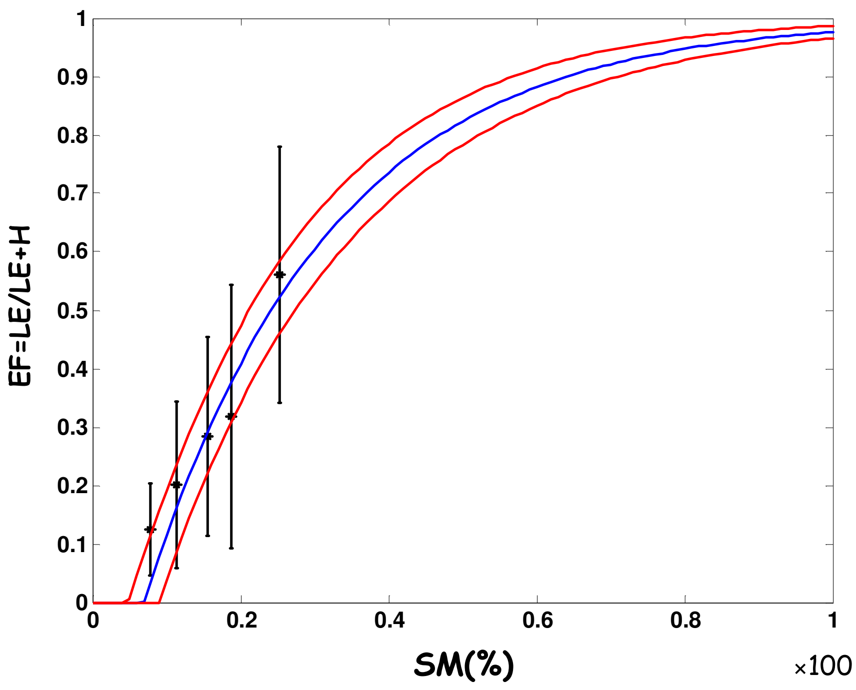
	K_s	α	a	S_w	β
K_s	1.00				
α	0.49	1.00			
a	-0.61	-0.66	1.00		
S_w	-0.03	0.025	0.23	1.00	
β	-0.26	-0.16	-0.15	0.58	1.00

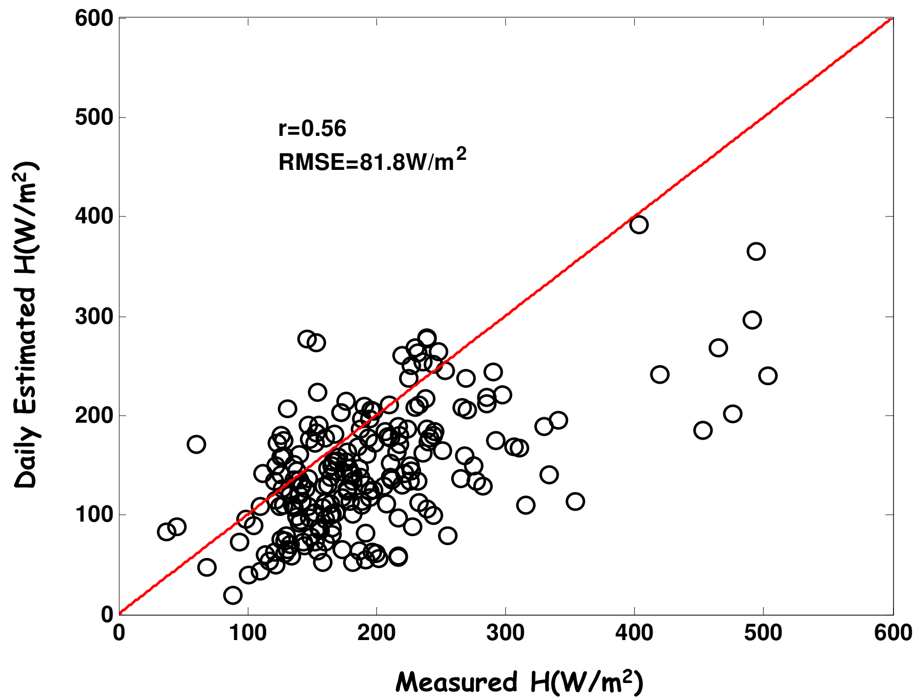
- ✓ Parameters are estimated robustly
- ✓ Correlation btwn different parameters is reasonable & physically meaningful

Validating The results ….

Agoufa flux tower site

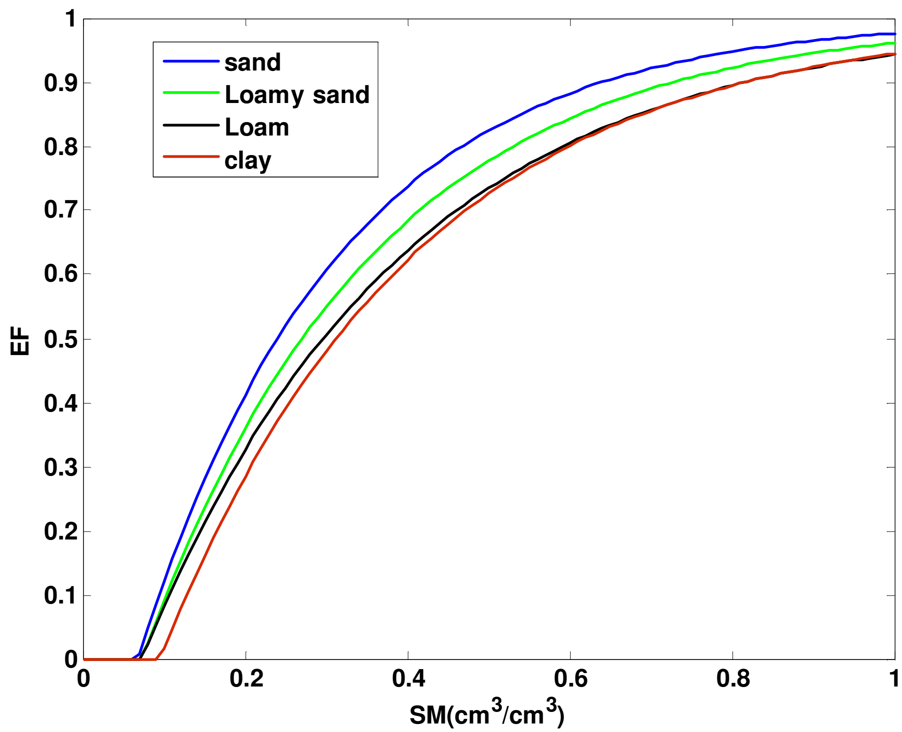
- hourly H, LE, LE/(LE+H)
- Soil type: Sand
- Vegetation type: Grassland
- Soil water content: AMSR-E data interpolation





Validating the results …

❑ **EF-SM relationship for different soils**



- Soil water potential increases between coarser to finer soils.

- Higher water potential is a barrier to water extraction, thus the rate of Evaporation from soils with coarser texture is higher than from soils with finer texture.

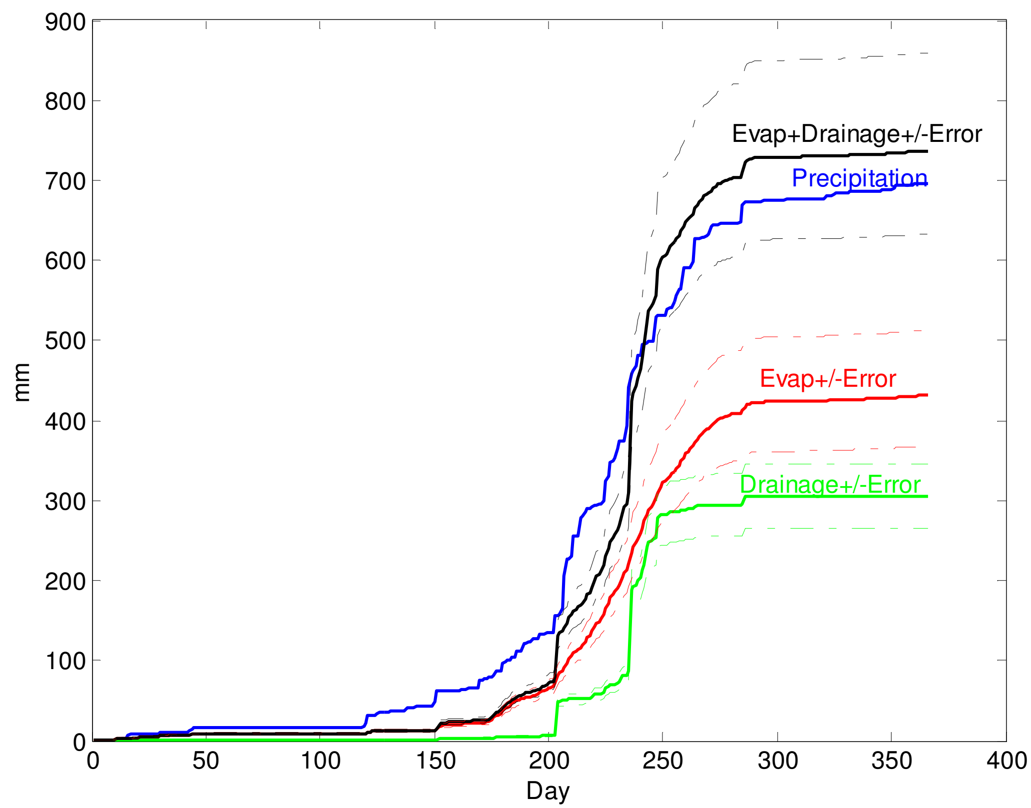
Validating the results …

❑ Endorheic property

Gourma is globally **endorheic system**, meaning it contributes little water to, nor receives water from, the Niger river.

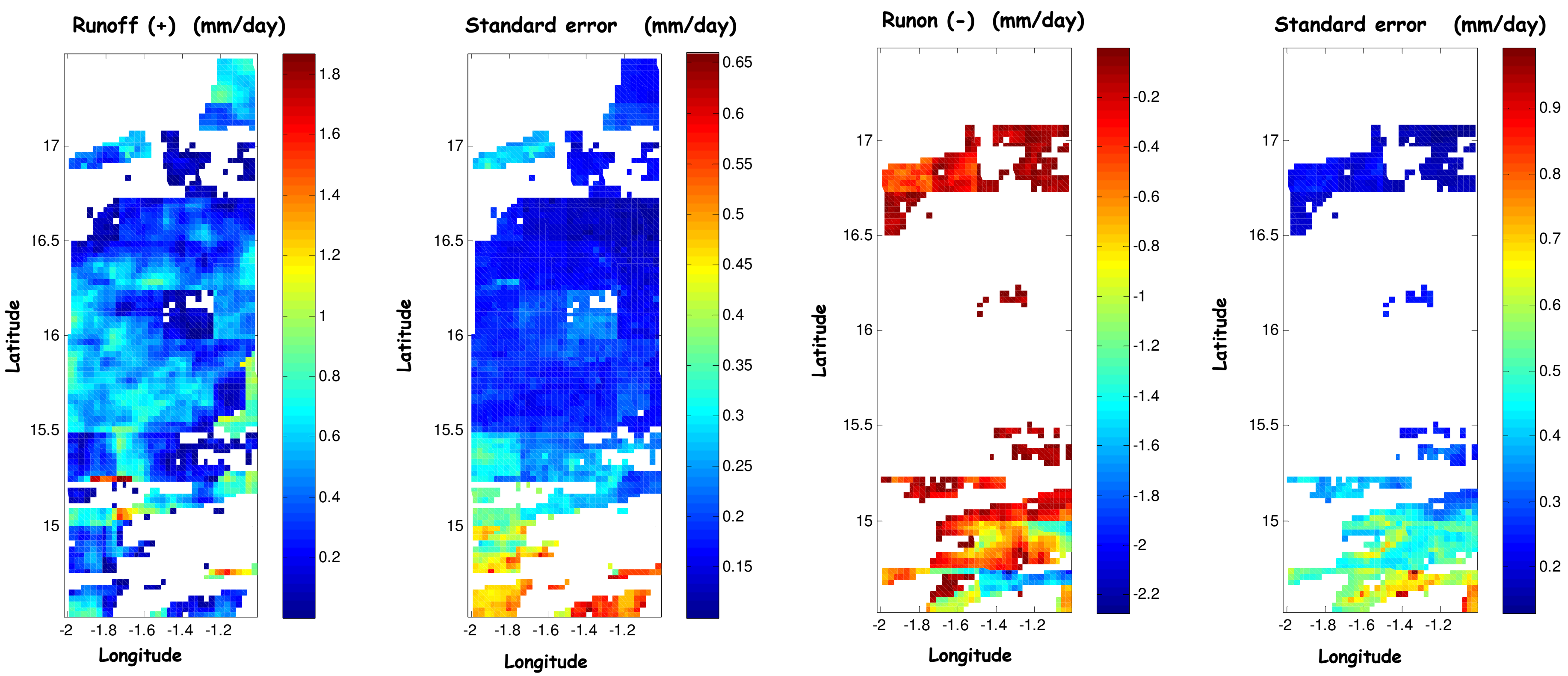
✓Cumulative distribution of the estimated drainage and evapotranspiration averaged over the entire pixels within the Gourma region is close to cumulative distribution of precipitation

✓Expected values of water balance residual (potentially Runon/ Runoff) ~0.11 mm/day (small)



Validating the results …

- Map of water balance residual (runoff/runon) over the Gourma region

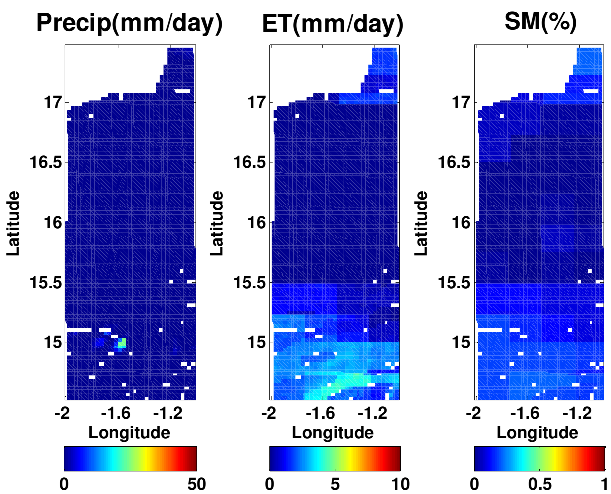


- Yearly average water balance equation over all the pixels results in the map of runoff/ runon (+/-)
- The errors in this estimation methodology manifests itself in the form of runoff/ run residuals
- The map of runoff/ runon corresponds well with the characteristics of Gourma region

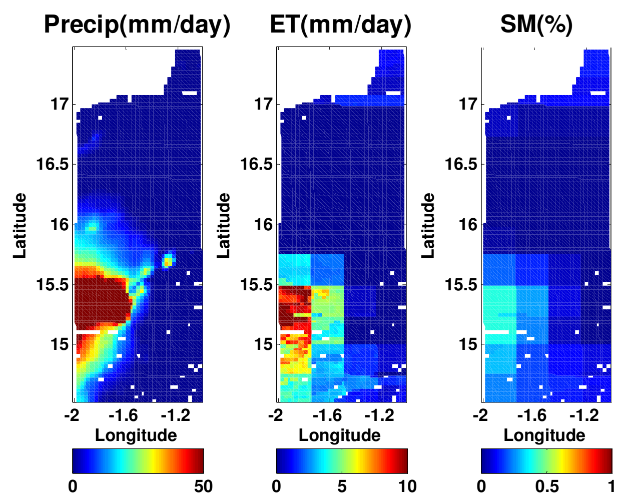
Evaluating the results …

❑ Precipitation- Evaporation patterns

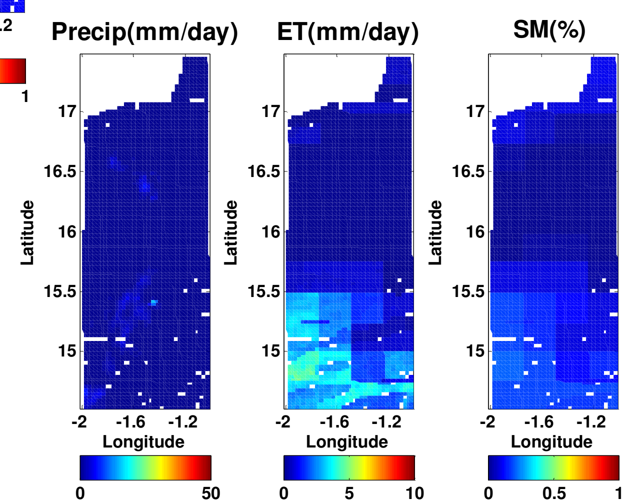
Note: No water balance or soil moisture accounting used



Day 190



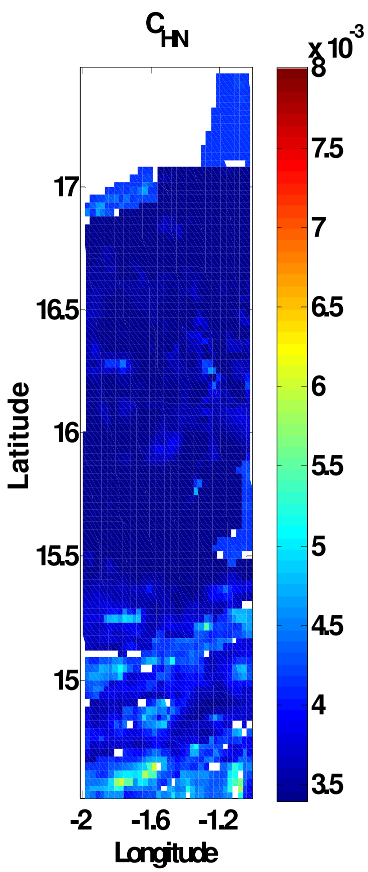
Day 191



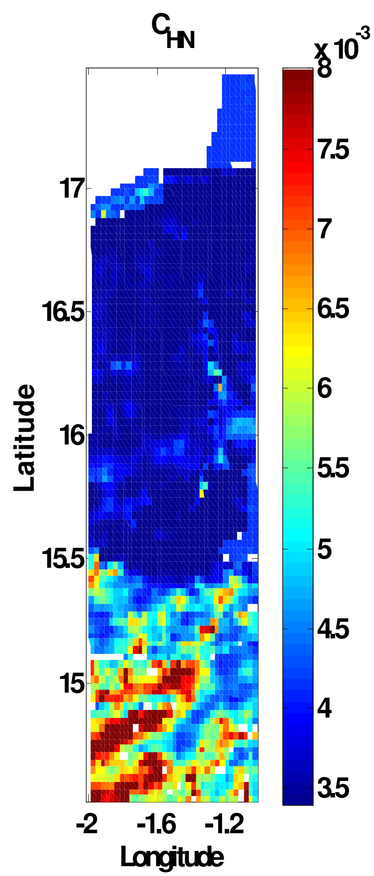
Day 192

Evaluating the results …

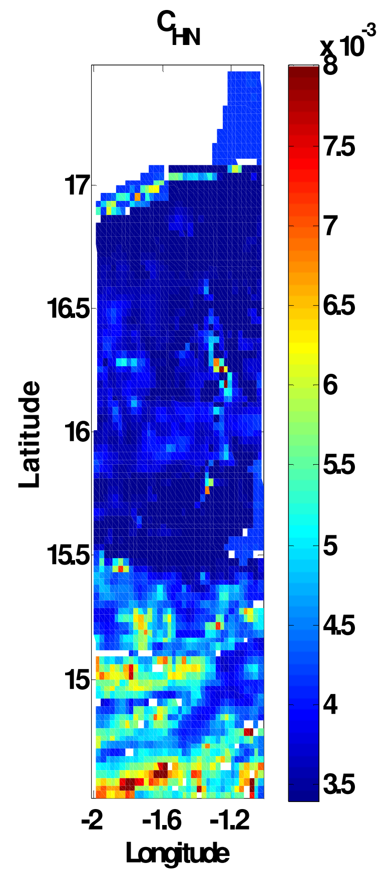
- C_{HN} patterns



June(Beginning of Monsoon)



August(Peak of Monsoon)



October(end of Monsoon)

- C_{HN} effective value over pixel area
- Mapped values
- Dynamics

Conclusions

- Methodology developed to use both water and energy balance to constrain parameter estimation of surface energy and moisture balance

- Method is distinct from traditional calibration because it does not need flux information (eg. problomatic drainage and evaporation data) to estimate parameters

- Only forcing (P, $R_{in}^{\downarrow}$) and states (s,T_s) used; hence scalable for remote sensing and mapping applications

- Feasibility demonstrated at point-scale with synthetic data (true parameters known for evaluation) and Ameriflux field site data

- Application over West Africa using remote sensing shows feasibility of using satellite data to estimate effective values of important land surface model parameters

Questions ?

Leila Farhadi
farhadi@mit.edu