ERAHUMED DSS

Valerio Gherardi   Pablo Amador Crespo   Andreu Rico

Table of contents

Appendices 24

Preface

The purpose of this book is to provide a comprehensive reference for the ERAHUMED Decision Support System. Here you can find the technical descriptions of the algorithms employed by the system, as well as the user manual for the accompanying software.

The Support System and, hence, this book are currently under development on Github. In particular, the {erahumed} R package is hosted here.

For general information on the ERAHUMED project, please refer to the official website. If you want to get in touch, you can contact any of us via e-mail:

- Andreu Rico (Coordinator)
- Pablo Amador (PhD Researcher)
- Valerio Gherardi (Software Developer)

1 Introduction

This is a book created from markdown and executable code.

See Martínez-Megías et al. (2024) for additional info.

Part I

Technical description

2 The ERAHUMED model: a bird's eye view

3 ERAHUMED model components

This Chapter provides detailed descriptions of the various components of the ERAHUMED model, briefly introduced in Chapter 2.

3.1 INP: Input Data

The purpose of this model component is simply to collect the empirical data that provides the observational input to all subsequent modeling layers. This data consists of hydrological and meteorological time series data for the Albufera lake in the desired time frame.

3.1.1 Input

The only inputs to this modeling layer are the two aforementioned time-series data-sets.

Hydrological data consists of the direct measurements of the Albufera lake's daily water levels and outflows. Since the lake has three estuaries (*Gola de Pujol*, *Gola del Perellonet* and *Gola del Perelló*), this data amounts to four daily time series. The actual dataset bundled with the ERAHUMED software was obtained from the public data repository compiled by the *Confederación Hidrográfica del Júcar* (TODO: Ref.), and is described in more detail in Appendix A.1.

Meteorological data consists of the daily measurements of precipitation and evapotranspiration per unit are, and temperature (average, maximum and minimum). This amounts to five daily the series. The actual data bundled with the ERAHUMED software was obtained (TODO: where), and is described in more detail in Appendix A.2.

The time-series inputs described above are collected in the table below (the common index t refers to the time - *i.e.* day - of observation).

Table 3.1: Observational input to the ERAHUMED model (INP model component).

Input	Units	Description
L_t	m	Lake water level
O_t^{Pujol}	m^3	*Pujol* daily outflow
$O_t^{\text{Perelló}}$	m^3	*Perelló* daily outflow

Input	Units	Description
$O_t^{\text{Perellonet}}$	m^3	*Perellonet* daily outflow
P$_t$	mm	Precipitation (per unit area)
ETP$_t$	mm	Evapotranspiration (per unit area)
T_t^{max}	°C	Maximum temperature
T_t^{min}	°C	Minimum temperature
T_t^{ave}	°C	Average temperature

3.1.2 Output

This modeling layer does not involve any actual computation, and its output is simply a time-series data-set obtained as the combination of the two input data-sets, *i.e.* the collection of time series of Table 3.1.

3.2 HBA: Hydrological Balance of the Albufera lake

The purpose of this model component is to compute the total daily inflow to the Albufera lake. The relevant equation expressing the hydrological balance is:

$$\text{Volume Change} = \text{Inflow} - \text{Outflow} + \text{Precipitation} - \text{Evapotranspiration} \tag{3.1}$$

where the unknown is Inflow, whereas the remaining terms are obtained from observational data, as described in the previous Section (Section 3.1).

3.2.1 Input

The inputs specific to this layer are two numerical functions converting water heights into the water *volumes* that appear in Equation 3.1. Specifically:

- The lake's *storage curve* converts the measured lake's level into a total water volume.
- The *precipitation-evapotranspiration volume function* (or, for the sake of brevity, *P-ETP function*), that converts precipitation and evapotranspiration values *per unit area* into an overall water volume difference.

Even though the R interface to the ERAHUMED software allows arbitrary definitions, the graphical user interface assumes a linear approximation for both these functions. Therefore, the storage curve is assumed to take the form:

$$V_t = m \cdot L_t + q, \tag{3.2}$$

where V_t is the (daily) water volume of the lake, L_t the corresponding water level (*cf.* Table 3.1), and m and q are numerical coefficients[1]. Correspondingly, the P-ETP function reads:

$$\Delta V_t^{\text{P-ETP}} = \alpha \cdot \text{P}_t - \beta \cdot \text{ETP}_t, \tag{3.3}$$

where the left-hand side represents the relevant water volume difference, whereas the P and ETP terms in the right-hand side are the measured precipitation and evapotranspiration levels per unit area (*cf.* Table 3.1). The actual default values of m, q, α and β are documented in Appendix A.4.

3.2.2 Output

The output of this model consists of the time-series collected in the table below.

Table 3.2: HBA model component outputs

Variable	Units	Description
V_t	m^3	Lake water volume
I_t	m^3	Lake total inflow
O_t	m^3	Lake total outflow

3.2.3 Details

Using the notation introduced in the previous Subsections, we can rewrite Equation 3.1 as follows:

$$V_{t+1} - V_t = I_t - O_t + \Delta V_t^{\text{ETP}}, \tag{3.4}$$

where V_t and ΔV_t^{ETP} are computed through the storage and P-ETP functions, as in Eqs. 3.2 and 3.3. The only terms that requires further clarification in Equation 3.4 is O_t, the total outflow from the lake.

Strictly speaking, O_t is not merely the sum of O_t^{Pujol}, $O_t^{\text{Perelló}}$ and $O_t^{\text{Perellonet}}$, but is rather calculated as follows:

$$O_t = \max \left[O_t^{\text{Pujol}} + O_t^{\text{Perelló}} + O_t^{\text{Perellonet}}, \, \Delta V_t^{\text{ETP}} + V_t - V_{t+1} \right]. \tag{3.5}$$

---

[1]In fact, for the purpose of computing volume *changes* entering the hydrological balance Equation 3.1, only the slope m is required.

The rationale is that the simple sum of estuaries outflows omits potentially important contributions from *water recirculation*, that is to say, water being pumped out from the lake for rice-field irrigation, by the so-called *tancats*. Such amount of recirculated water is hard to estimate and, in the lack of a better model, we simply assume this to be negligible, *except* when a positive amount is required by Equation 3.4 itself, due the physical constraint that $I_t \geq 0$.

Once O_t is calculated through Equation 3.5, I_t can be immediately obtained from Equation 3.4. Notice that whenever the aforementioned compensating outflow term due to water recirculation is included (which happens when the maximum in Eq. 3.5 is given by the second term), the total inflow is always estimated to be zero.

3.3 HBP: Hydrological Balance of rice Paddy clusters

This modeling layer simulates the local hydrology (*i.e.* water levels, inflows and outflows) of rice paddy clusters in the Albufera Natural Park. The simulation, that literally involves the generation of a synthetic data-set, is based on:

- The hydrology of the Albufera lake, which allows to constrain the total outflow of all rice paddies, assumed to be equal to the lake's total inflow (the main output of the HBA layer, see Section 3.2).
- An estimate of the fraction of the total lake's inflow that comes from each of the 26 ditches that flow into the Albufera.
- A subdivision of the rice fields surface into clusters that are assumed to be drain water into a single ditch - meaning that the sum of outflows of clusters from the same group must equal the total flow through the corresponding ditch.
- An ideal yearly management plan for the irrigation and draining of rice paddies.

3.3.1 Input

As already mentioned, the simulation is based on a spatial clustering, which is described in greater detail in Appendix A.5. In particular each cluster is labeled according to its rice variety, and to its field type (*tancat* or regular field), which we will indicate in the following through the categorical variables $v = $ Clearfield, Sendra or Bomba and $\theta = $ Tancat, Regular respectively.

The yearly management plan for these clusters, is described in Appendix A.3, and provides, for each day of the year d, rice variety v and field type θ the following:

- $\text{Ir}_{d,v,\theta}$: boolean expressing whether the cluster is supposed to be irrigated (if true) on day of year d.

- $\text{Dr}_{d,v,\theta}$: boolean expressing whether the cluster is supposed to be drained (if true) on day of year d.
- $\mathcal{H}_{d,v,\theta}$: ideal water level (in cm) of the cluster on day of year d.

We stress that the values of these yearly series are specific to each rice variety and field type, and the corresponding values for a cluster are obtained by matching with the corresponding cluster's attributes.

In addition to these rather complex input, the model component requires the following numerical parameters:

- k_{flow} (*Ideal flow rate*). Rate at which water flows through rice paddies when these are being simultaneously irrigated and drained, with the overall level being kept constant. Expressed in $\text{cm} \cdot \text{day}^{-1}$.

- h_{thres} (*Height threshold*). Maximum allowed water level for a cluster to be considered emptied, used in the calculation of draining/irrigation plan delays. Expressed in cm.

Finally, it should be noted that the simulation involves random sampling (see Subsection 3.3.3), so that the seed of random number generation is an additional pseudo-parameter.

3.3.2 Output

The output of this layer provides a simulation for the local hydrology of each cluster, summarized by the variables in the table below, in which the indices c and t refer to cluster and time, respectively.

Table 3.3: HBP model component outputs.

Variable	Units	Description
$V_{c,t}$	m³	Cluster's water volume
$I_{c,t}$	m³	Cluster's inflow
$O_{c,t}$	m³	Cluster's outflow

3.3.3 Detail

In the notation introduced above, the purpose of this model layer is to simulate for each cluster c the hydrological time series of Table 3.3, satisfying the hydrological balance:

$$V_{c,t+1} - V_{c,t} = I_{c,t} - O_{c,t} + (\text{P}_t - \text{ETP}_t) \times A_c \tag{3.6}$$

where P_t and ETP_t stand for precipitation and evapotranspiration per unit area, as in table Table 3.1, whereas A_c denotes the cluster's area. As mentioned in the introduction to this Section (see 3.3), this simulation rests on various simplifications, which we detail in what follows.

To begin with, we assume that each cluster only drains water into a single one out of the twenty-six main ditches that cross the Albufera Natural Park. In what follows, we will focus on the set of all clusters connected with a given ditch, and we will enumerate clusters through the index $c = 1, 2, ..., N_C$. On the other hand, clusters are assumed to be irrigated from the external of the Albufera system, *i.e.* not through any of the ditches which eventually flow into the lake.

These two assumptions imply that the sum of cluster outflows is constrained by:

$$\sum_{c=1}^{N_C} O_{c,t} = Q_t \tag{3.7}$$

where Q_t denotes the daily flow through the ditch. This quantity is, in turn, not measured, and we estimate it through the following simple model:

$$Q_t = \frac{\text{Area of clusters draining into ditch}}{\text{Area of all clusters}} \times \text{Lake's Inflow}_t \tag{3.8}$$

where the lake's inflow is the one computed by the HBA model component (see Section 3.2).

Qualitatively speaking, the simulation works by finding daily values of $I_{c,t}$ and $O_{c,t}$ compatible with Eqs. 3.6 and 3.7, and such that the resulting hydrology is as close as possible with the "ideal" one prescribed by the specified management plan. This is accomplished in three steps:

1. Computing ideal inflows and outflows, given current water levels and management plans.
2. Computing actual inflows and outflows, and corresponding actual water level changes.

3. Computing required management plan delays, if some clusters were scheduled to be drained, but could not due to insufficient flow through the common ditch (see below for details).

These steps are iterated on a daily basis, starting from some initial time (say $t = 0$) at which all cluster water levels coincide with the ideal ones, and no plan delays are present.

Concerning the last step, a few words may serve to clarify the algorithm described below. The purpose of plan delays is to ensure that during each year's sowing season, all cluster's are emptied at some point, as required for ground
applications of chemicals (modeled in subsequent layers of ERAHUMED). This is achieved by delaying by one day the management plan, whenever an emptying condition is not met. When

far outside of the sowing window, the delays are reset then to zero, in order to ensure that delays are not carried over indefinitely, which would be quite unrealistic.

In what follows, the conversion between cluster water volumes and depths is provided by $V_{c,t} = A_c \cdot h_{c,t}$, and we denote by $\delta_{c,t}$ the time-series of plan delays for cluster c, which is initialized by $\delta_{c,0} = 0$.

3.3.3.1 Step 1: ideal balance

Ideal balance quantities for each cluster c are obtained from the management plan data-set, whose relevant row is identified by the cluster's rice variety v and field type θ, and the delayed day:

$$d_{c,t+1} = d_{t+1} - \delta_{c,t}, \tag{3.9}$$

where d_{t+1} denotes the day of year corresponding to time $t+1$, and $\delta_{c,t}$ the accumulated plan delay.

Let $h^{id}_{c,t+1}$ denote the ideal depth for cluster c at time $t+1$ retrieved in this way, and $\mathrm{Ir}_{c,t}$, $\mathrm{Dr}_{c,t}$ the corresponding ideal irrigation and draining plans. Furthermore, denote by $V_{c,t+1}^{id} = A\_c\ h\_{c,t+1}^{id}$ the corresponding ideal water volume.

In order to compute ideal inflow and outflow, we require (*cf.* Equation 3.6):

$$V^{id}_{c,t+1} = \max\{V^{id}_{c,t} + (\mathrm{P}_t - \mathrm{ETP}_t) \times A_c, 0\} + I^{id}_{c,t} - O^{id}_{c,t} \tag{3.10}$$

Clearly, Equation 3.10 alone does not individually specify $I^{id}_{c,t}$ and $O^{id}_{c,t}$, but only their difference $\Delta^{id}_{c,t} = I^{id}_{c,t} - O^{id}_{c,t}$. In order to fix both these quantities:

$$(I^{id}_{c,t})^{(0)} = \begin{cases} k_{\mathrm{flow}} & \text{if } \mathrm{Ir}_{c,t} = \mathrm{Dr}_{c,t} = 1 \\ 0 & \text{otherwise} \end{cases},$$
$$(O^{id}_{c,t})^{(0)} = (I^{id}_{c,t})^{(0)} - \Delta^{id}_{c,t}. \tag{3.11}$$

and, in order to ensure that flows are positive, we finally set:

$$O^{id}_{c,t} = \max\{(O^{id}_{c,t})^{(0)}, 0\}$$
$$I^{id}_{c,t} = O^{id}_{c,t} + \Delta^{id}_{c,t}, \tag{3.12}$$

which satisfy Equation 3.10 and give rise to positive $\hat{}\ O^{id}_{c,t}$ and $I^{id}_{c,t}$.

3.3.3.2 Step 2: real balance

At each time-step t, the cluster's index set is randomly permuted [2], and the real flows are calculated as:

$$O_{c,t} = \min\{O_{c,t}^{\mathrm{id}}, Q_t - \sum_{c'<c} O_{c',t}\} + \frac{\max\{0, Q(t) - \sum_{c'} O_{c,t}^{\mathrm{id}}\}}{N_C},$$

$$I_{c,t} = \max\{I_{c,t}^{\mathrm{id}} - O_{c,t}^{\mathrm{id}} + O_{c,t}, 0\}$$

$$(3.13)$$

In words, clusters are emptied in a random order within the allowed capacity of the corresponding ditch; if the sum of ideal outflows is less than capacity, the remaining outflow volume is shared uniformly among clusters. Using Equation 3.13, we finally determine the real water level achieved as:

$$V_{c,t+1} = \max\{V_{c,t} + (\mathrm{P}_t - \mathrm{ETP}_t) \times A_c, 0\} + I_{c,t} - O_{c,t} \qquad (3.14)$$

to be compared with Equation 3.10.

3.3.3.3 Step 3: updating the plan delay

The updated value $\delta_{c,t+1}$ is obtained as follows. If d_{t+1} (the *actual* day of year) is outside of the window $W = $ [20th of April, 15th of October], then $\delta_{c,t+1} = 0$. Otherwise, if $h_{c,t}^{\mathrm{id}} > 0$ or $h_{c,t} < h_{\mathrm{thres}}$, the plan delay is unchanged: $\delta_{c,t+1} = \delta_{c,t}$. Finally, if $h_{c,t}^{\mathrm{id}} = 0$ but $h_{c,t} > h_{\mathrm{thres}}$, we add one day of delay: $\delta_{c,t+1} = \delta_{c,t} + 1$.

3.4 CA: Chemical Applications

TBD.

3.5 CT: Chemical Transport

This model component describes the evolution of chemical masses in the three compartments of foliage, water and sediment. This is obtained by solving the system of differential equations that describes the dynamics of chemicals, through a semi-analytic approach with suitable approximations and simplifications.

---

[2]With some abuse of notation, we assume the indexes c and c' in Equation 3.13 to be sorted according to this random permutation.

3.5.1 Input

TBD.

3.5.2 Output

TBD.

3.5.3 Details

The evolution of masses in the foliage, water and sediment compartments is described by the following system of differential equations:

$$
\begin{aligned}
\frac{\mathrm{d}m_f}{\mathrm{d}t} &= -(k_f + w)m_f + a_f \\
\frac{\mathrm{d}m_w}{\mathrm{d}t} &= -(k_w + d_w + s + \frac{O}{V})m_w + d_s m_s + w m_f + a_w - \sigma(\frac{m_w}{\rho V}) \\
\frac{\mathrm{d}m_s}{\mathrm{d}t} &= (d_w + s)m_w - (k_s + d_s)m_s + a_s + \sigma(\frac{m_w}{\rho V})
\end{aligned}
\tag{3.15}
$$

where:

- $a_{f,w,s}$ are the mass application rates of the chemical in the three compartments,
- $k_{f,w,s}$ are the degradation rates of the chemical in the three compartments,
- d_w and d_s are the water-sediment diffusion rates,
- s is the settling rate,
- w is the washout rate,
- O is the outflow rate of the rice field,
- V is the volume of water in the rice field,
- ρ is the chemical solubility in water,
- $\sigma(x)$ is a function (not further specified, see below) that grows quickly for $x > 1$, and vanishes for $x \leq 1$. This function accounts for solubility. (TODO: motivate the inclusion of exactly these processes)

Strictly speaking, all these terms have instantaneous time dependence[3]. Apart from making the system 3.15 hard to attack by analytic means, such a dependence is troubling because we don't have access to the exact time dependence of the majority of these terms (*e.g.* outflow, volume or chemical applications), our input consisting of simple daily average/cumulative values. On the other hand, if all the terms were constant, and if we could neglect σ, the solution

---

[3]This observation also includes pure physico-chemical "constants" such as degradation rates, where the time dependence would stem from *temperature* dependence.

of Equation 3.15 would be immediate, as the corresponding system becomes a linear ODE with constant coefficients, in addition whose eigenvalues and eigenvectors can be computed explicitly.

What we use in practice is an intermediate semi-analytic approach that allows us to compute daily values of $m_{f,w,s}$, which we can briefly summarize as follows. The time evolution of $m_{f,w,s}$ in a daily time-step (from t to $t+1$, say, assuming t is measured in days) is obtained through the following four consecutive stages:

1. We compute the exact evolution of $m_{f,w,s}$ according to the linear ODE obtained by disregarding the processes of outflow, chemical application and solubility, and using daily constant values for all the remaining constants involved (described in more detail below).
2. We compute the m_w losses due to outflow as if they happened instantaneously after the processes computed in the previous step took place, with the water volume of the rice paddy varying from $V(t) + O(t)$ to $V(t)$[4]

3. We compute the mass applications again as instantaneous, after the losses in the water compartment due to outflows took place.
4. We compare the resulting m_w with the maximum amount allowed by the solubility in water, that is $m_w^{\max}(t) = \rho V(t)$ and transfer any excess to the sediment compartment (m_s).

In the following, we describe in full detail the computations involved in these four steps.

3.5.3.1 Step 1: linear ODE evolution (physico-chemical processes)

Disregarding outflow, mass application and solubility, we get a linear ODE of the form:

$$
\begin{aligned}
\frac{\mathrm{d}m_f}{\mathrm{d}t} &= \gamma m_f, \\
\frac{\mathrm{d}m_w}{\mathrm{d}t} &= a_{ww}m_w + a_{ws}m_s + wm_f, \\
\frac{\mathrm{d}m_s}{\mathrm{d}t} &= a_{sw}m_w + a_{ss}m_s,
\end{aligned}
\tag{3.16}
$$

with:

---

[4]Recall that, in the notation of TODO, $V(t)$ denotes the final volume on day t.

$$\gamma = -(k_f + w),$$
$$a_{ww} = -(k_w + d_w + s)$$
$$a_{ws} = d_s \tag{3.17}$$
$$a_{sw} = d_w + s$$
$$a_{ss} = -(k_s + d_s)$$

where the various physico-chemical parameters are assumed to be constant during a daily time-step (see Sec. TODO for the actual numerical values).

The evolution of m_f is easily obtained:

$$m_f(t) = e^{\gamma t} m_f(0). \tag{3.18}$$

Plugging this into the three-dimensional system 3.16 we obtain a reduced two-dimensional system for the water-sediment compartments:

$$\frac{\mathrm{d}m_w}{\mathrm{d}t} = a_{ww}m_w + a_{ws}m_s + we^{\gamma t}m_f(0),$$
$$\tag{3.19}$$
$$\frac{\mathrm{d}m_s}{\mathrm{d}t} = a_{sw}m_w + a_{ss}m_s,$$

which is solved explicitly with standard methods[5].

---

[5]Letting $x = (m_w, \, m_s)^T$ and $b = w(m_f(0), 0)^T$, the system can be rewritten in the form:

$$\dot{x} = Ax + e^{\gamma t}b$$

where

$$A = \begin{pmatrix} a_{ww} & a_{ws} \\ a_{sw} & a_{ss} \end{pmatrix}$$

The general solution reads:

$$x(t) = e^{At}x(0) + (e^{At} - e^{\gamma t}I)(A - \gamma I)^{-1}b$$

where I is the identity matrix. Notice that the exponential can be computed explicitly from the eigenvalues of A:

$$e^{At} = e^{\lambda_+ t}P_+ + e^{\lambda_- t}P_-,$$

with:

$$\lambda_\pm = \frac{\mathrm{Tr}A}{2} \sqrt{\left(\frac{\mathrm{Tr}A}{2}\right)^2 - \det A}$$

and

$$P_\pm = \pm\frac{1}{\lambda_+ - \lambda_-}(A - \lambda_\mp)$$

3.5.3.2 Step 2: mass losses due to outflow

Mass losses due to outflow are computed as if the outflow process happened instantaneously after the processes discussed in the previous step took place. The amount of mass in the water compartment is affected as follows:

$$m_w \to \frac{V_f}{V_i} m_w, \tag{3.20}$$

where V_i and V_f denote the initial and final volume for the outflow process. We assume that V_f coincides with the value $V(t)$ simulated by the HBP model component (*cf.* Sec TODO), so that $V_i = V(t) + O(t)$, where $O(t)$ denotes the total simulated outflow for day t.

3.5.3.3 Step 3: chemical applications

Applications are computed as instantaneous and subsequent to outflow. Masses in the various compartments ($i = f$, w, s) are modified as follows:

$$m_i \to m_i + a_i \tag{3.21}$$

3.5.3.4 Step 4: solubility

The chemical density in the water compartment after Steps 1, 2, and 3 is eventually compared with the chemical's solubility ρ, and any mass excess is instantaneously transferred to the sediment compartment. This implies the following modifications:

$$m_s \to m_s + \max(0, m_w - \rho V), \quad m_w \to \max(m_w, \rho V), \tag{3.22}$$

where, for the volume V we again use the value $V(t)$ simulated by the HBP model component.

Part II

User Manual

4 The ERAHUMED DSS User Interface

5 The {erahumed} R package

References

Martínez-Megías, Claudia, Alba Arenas-Sánchez, Diana Manjarrés-López, Sandra Pérez, Yolanda Soriano, Yolanda Picó, and Andreu Rico. 2024. "Pharmaceutical and Pesticide Mixtures in a Mediterranean Coastal Wetland: Comparison of Sampling Methods, Ecological Risks, and Removal by a Constructed Wetland." *Environmental Science and Pollution Research* 31 (10): 14593–609.

A Input Data

A.1 Hydrological data

A.2 Meteorological data

A.3 Albufera Rice Paddies Management

A.4 Storage curve and P-ETP function

A.5 Definition of rice clusters