

Estimation de ligne de base de capteurs d'humectation : intégration et minimum locaux à différentes échelles

Jean-Yves BAUDAIS¹, Melen LECLERC² et Christophe LANGRUME²

¹IETR, UMR CNRS 6164, Beaulieu, Rennes, France

²IGEPP, UMR 1349 INRAE, Domaine de la Motte, Le Rheu, France

jean-yves.baudais@insa-rennes.fr, {melen.leclerc, christophe.langrume}@inrae.fr

Résumé – Les capteurs diélectriques d'humectation sont utilisés en agriculture pour détecter la présence d'eau déposée sur le feuillage et prédire le risque de développement de maladies. Le signal électrique mesuré présente une dérive du niveau de base qui vient biaiser les alertes. Nous proposons une méthode d'estimation de cette ligne de base exploitant la norme L1 et sélectionnant des minimums locaux à une échelle d'observation. Les performances de l'estimateur sont évaluées sur des données simulées et comparées à celles des estimateurs de la littérature.

Abstract – Dielectric humectation sensors are used in agriculture to detect the presence of water on foliage and to predict the risk of disease development. The measured electrical signal has a base level drift that skews the alerts. We propose a method for estimating this baseline using L1 and selecting local minimums at an observation scale. The performance of the estimator is evaluated on simulated data and compared to the literature estimators.

1 Introduction

L'humectation foliaire, c.-à-d. l'eau présente sur les feuilles, est une donnée importante pour prédire le développement de nombreuses maladies des plantes [1]. Elle est mesurée *in situ* avec des capteurs diélectriques. Le capteur le plus utilisé est le capteur LWS [2]. Ce capteur capacitif reproduit grossièrement la forme et les propriétés thermodynamiques, radiatives et hydrophobes d'une feuille. Il mesure la constante diélectrique qui dépend de la quantité d'eau présente sur le capteur [3]. Le signal produit correspond à une succession d'excitations indiquant la présence d'eau sur le capteur (p. ex. rosée, pluie, gel). Le traitement de ce signal consiste généralement à extraire par seuillage les périodes sèches et humides qui sont ensuite utilisées pour des recherches en épidémiologie végétale ou pour réaliser des alertes pendant les périodes propices aux maladies des cultures. Néanmoins, lorsque les capteurs sont laissés pendant de longues périodes en extérieur, une dérive lente du niveau de base du signal apparaît et vient biaiser la détection de l'humectation.

L'estimation et la correction de la ligne de base est une étape de pré-traitement essentielle pour analyser, quantifier et interpréter correctement des mesures physiques, chimiques ou biologiques [4]. De nombreuses méthodes sont proposées pour réaliser cette estimation de manière quasi-automatique une fois l'acquisition réalisée [5, 6]. Elles s'appuient sur l'estimation des pics [7], la régression quantile [8], la décomposition en ondelettes [9], l'ajustement de polynômes [10] ou des moindres carrés pénalisés ou pondérés [11]. Ces méthodes nécessitent soit l'ensemble des données, soit l'optimisation d'hyperparamètres pour estimer le niveau de base et corriger le signal. Nous proposons ici une méthode estimant la ligne de base en temps réel et nécessitant une

faible profondeur de mémoire. La méthode proposée offre la possibilité d'un traitement *in situ*, c.-à-d. dans un nœud de capteur, afin d'améliorer les outils d'aide à la décision en agriculture pour la lutte contre les maladies et le gel.

Le reste de l'article est organisé de la façon suivante. Le paragraphe 2 présente le modèle d'humectation utilisé et la méthode d'estimation de la ligne de base. Le paragraphe 3 donne quelques résultats de simulation, quantifie les performances de la méthode proposée et les compare aux performances d'algorithmes de la littérature. Le paragraphe 4 conclut l'article.

2 Modèle et méthode

La figure 1 donne un exemple de 9858 points de mesure d'humectation dans un couvert de pois, obtenus de mars à juin 2014. La préconisation du constructeur des capteurs est de fixer la ligne de base à 274 mV pour une tension d'excitation de 2,5 V, ou d'adapter empiriquement ce seuil à chaque capteur. Aucune des deux méthodes ne donnent de résultats pertinents pour ensuite positionner le seuil d'humectation. Par exemple, les données complètes obtenues avec plusieurs capteurs donnent des valeurs minimales sur les cinq premiers jours allant de 260,5 à 264,8 mV, des valeurs bien en deçà de la ligne de base préconisée.

Hypothèse et caractéristique du signal L'hypothèse d'additivité des composantes du signal en sortie du capteur conduit au modèle de signal suivant

$$s(t) = h(t) + b(t) + n(t), \quad (1)$$

où $s(t)$ est la réponse du capteur, $h(t)$ est le signal d'humectation à proprement parler, c.-à-d. le signal d'intérêt, $b(t)$ est

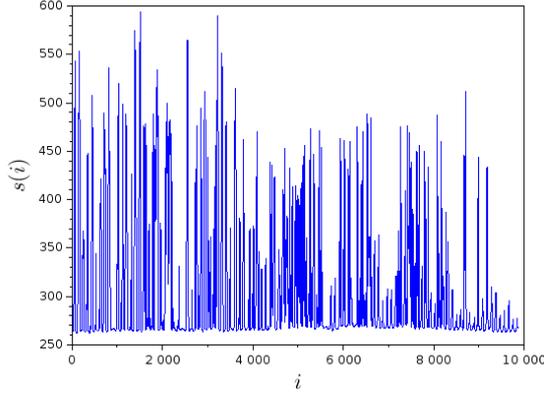


FIGURE 1 – Mesures expérimentales d’humectation en mV dans un couvert de pois, en fonction de l’indice de la mesure.

la ligne de base et $n(t)$ le bruit électronique ou électromagnétique de l’environnement. Les mesures sont échantillonnées et moyennées sur 15 minutes, d’où un bruit faible, et il y a entre 8500 et 10000 échantillons par capteur sur une campagne de mesure. Le modèle discret est alors, $i \in [1, N] \subset \mathbb{N}$

$$s(i) = h(i) + b(i) + n(i). \quad (2)$$

Il n’y pas de fonction ou fonctionnelle à priori modélisant la ligne de base et qui pourrait être exploitée par l’estimateur. Cependant, en plus de l’hypothèse d’additivité, la ligne de base a des amplitudes globalement plus faibles, des variations globalement plus lentes que les amplitudes et variations du signal d’humectation propre, sans pour autant présenter des valeurs simplement séparables. L’estimation de la ligne de base est alors faite sans hypothèse supplémentaire sur $h(i)$ ni sur $b(i)$.

Modèle Pour éprouver la méthode d’estimation et évaluer ses performances, un modèle simple est proposé où

1. $h(i)$ est un mélange de gaussiennes, avec une gaussienne par pic d’humectation, dont les distributions sont obtenues à partir de la répartition empirique de l’amplitude des pics, de leur largeur et de la distance entre les pics. La figure 2 illustre un exemple de comparaisons entre les mesures expérimentales et le modèle;
2. $b(i)$ est une marche aléatoire de pas $\pm 0,1$ lissée par moyenne glissante avec un filtre de longueur 100,
3. $n(i)$ est un bruit blanc gaussien centré.

Une meilleure modélisation du signal d’humectation serait d’utiliser des gaussiennes généralisées avec différents paramètres de forme permettant de prendre en compte l’asymétrie des lobes ou des fronts plus ou moins abruptes. Cependant, et dans la mesure où il n’y a pas d’estimation des paramètres du modèle, la forme des fonctions utilisées n’est pas un point clé de notre étude.

Mesure Le seuil d’humectation est un paramètre dépendant du phénomène biologique ou climatique à surveiller. Quelle que soit sa valeur, il doit suivre la ligne de base afin de séparer au plus juste les phases sèches des phases humides. Une

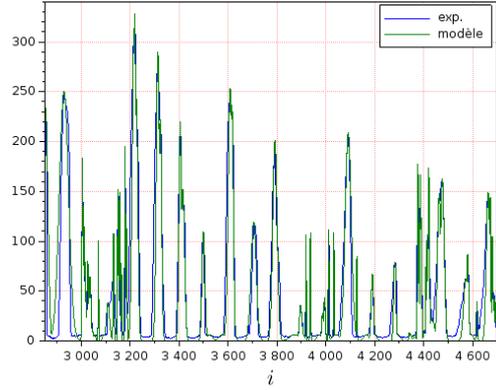


FIGURE 2 – Données expérimentales $s(i)$ normalisées et modèle d’humectation $h(i)$, amplitude en mV.

phase est humide si l’amplitude $s(i)$ du signal est au-dessus de ce seuil S , elle est sèche sinon. En appelant $\hat{b}(i)$ la ligne de base estimée, la comparaison du signal $s(i)$ au seuil théorique $b(i) + S$ et au seuil estimé $\hat{b}(i) + S$ permet de définir les fausses alarmes (FA), sous entendu vis-à-vis de la phase humide,

$$\mathcal{I}_{\text{FA}} = \{i \mid \hat{b}(i) + S < s(i) \leq b(i) + S\} \quad (3)$$

et les détections manquées (MD)

$$\mathcal{I}_{\text{MD}} = \{i \mid b(i) + S < s(i) \leq \hat{b}(i) + S\}. \quad (4)$$

La durée d’humectation étant une donnée importante d’un point de vue biologique, la distance $d(i)$ à la phase humide est mesurée et définie par

$$\forall i \in \mathcal{I}_{\text{FA}} \cup \mathcal{I}_{\text{MD}}, d(i) = \min_{\{j \mid s(j) > b(j) \wedge s(j) > \hat{b}(j)\}} |i - j|. \quad (5)$$

Le modèle permet de donner des probabilités ou densités de probabilité de FA, MD et $d(i)$, en plus de la classique erreur quadratique moyenne (EQM). Dans le cas des données de terrain, le seuil fixe S et le seuil adaptatif $\hat{b}(i) + S$ sont utilisés à la place des seuils $\hat{b}(i) + S$ et $b(i) + S$ du modèle.

Méthode d’estimation L’estimation de la ligne de base est faite à partir de $s(t)$. Elle est d’autant plus fiable au point t que $h(t)$ atteint une valeur nulle, qui est un minimum de la fonction, avec des variations au voisinage de t suffisamment faibles. Ce voisinage correspond alors à une période sèche. Cette observation permet de sélectionner un ensemble de points favorables pour être des points de la ligne de base et ensuite de les interpoler pour reconstruire la ligne de base estimée.

Plutôt que d’utiliser la dérivée, une propriété locale de la fonction sensible au bruit, on minimise une fonction coût $C_T(t_0)$ exploitant l’intégrale à différentes échelles d’observation T

$$C_T(t_0) = \frac{1}{2T} \int_{t_0-T}^{t_0+T} \phi(s(\tau) - s(t_0)) d\tau, \quad (6)$$

où t_0 est un minimum local à l’échelle observée et vérifie

$$s(t_0) = \min_{\tau \in [-T, T]} s(t_0 + \tau). \quad (7)$$

Les méthodes itératives cherchent à minimiser l'influence des pics via la fonction ϕ , en exploitant des coûts différents des coûts quadratiques [10]. Dans notre cas, au contraire, les pics sont utilisés pour maximiser le coût et rejeter ainsi les minimums locaux en période humide, peu fiables pour estimer $b(t)$. Nous avons choisi $\phi(x) = |x|$ qui donne des résultats satisfaisants.

Il est immédiat que l'estimateur ne peut pas suivre les variations de la ligne de base si le signal $h(t)$ est nul pour tout t . Dans ce cas, seuls les minimums locaux pourront correctement être estimés. En pratique, ce cas n'existe pas car il y a les phénomènes quasi quotidiens de rosée nocturne qui conduisent nécessairement à des valeurs de $h(t)$ non nulles.

Proposition 1. *Si $C_T(t_0) < \epsilon$ alors $s(t_0)$ est un bon candidat pour l'estimation de $b(t_0)$.*

Démonstration. La formulation de la proposition est probabiliste mais il n'y a pas de preuve formelle de l'implication dans la mesure où il est toujours possible de construire des fonctions et des points qui ne vérifient pas la proposition. Cependant, $\forall t_0$ vérifiant (7)

$$\begin{aligned} C_T(t_0) &\leq \frac{1}{2T} \int |h(t) - h(t_0)| dt + \frac{1}{2T} \int |b(t) - b(t_0)| dt \\ &\quad + \frac{1}{2T} \int |n(t) - n(t_0)| dt \\ &\leq a_h(t_0) + a_b(t_0) + a_n(t_0). \end{aligned} \quad (8)$$

Ainsi,

- Sur une période sèche, $a_h(t_0) < \epsilon_h(t_0)$ car $h(t) \approx h(t_0) \approx 0$. À l'opposé et autour d'un minimum local $s(t_0)$, les variations de $h(t)$ sur une période humide sont plus importantes;
- La ligne de base ayant des variations faibles, $b(t) \approx b(t_0)$ sur une fenêtre $2T$ qui reste à préciser mais qui ne doit pas être pas « trop grande », d'où $a_b(t_0) < \epsilon_b$;
- Enfin, l'intégration permet de réduire l'influence du bruit et $a_n(t_0) < \epsilon_n$ pour des valeurs de T qui restent à quantifier.

Dans ces conditions, $C_T(t_0) < \epsilon_h(t_0) + \epsilon_b + \epsilon_n = \epsilon$ et $s(t_0) \approx b(t_0)$. \square

La proposition 1 donne une méthode simple de sélection des points de $s(t)$ favorables à l'estimation de la ligne de base. L'estimation de la ligne de base se termine par une interpolation des points sélectionnés pour construire $\hat{b}(t)$.

3 Simulation et résultats

Performances de l'estimateur Le modèle présenté au paragraphe 2 est utilisé pour quantifier les performances de l'estimateur. Nous commençons par valider la proposition 1 avec une EQM $\sigma_T^2(C_0)$ calculée uniquement sur les points t_0 (7) sélectionnés à l'échelle T et pour un seuil C_0

$$\sigma_T^2(C_0) = \frac{1}{|\mathcal{T}|} \sum_{t_0 \in \mathcal{T}} |\hat{b}(t_0) - b(t_0)|^2, \quad (9)$$

avec

$$\mathcal{T} = \{t_0 \mid C_T(t_0) < C_0\} \quad (10)$$

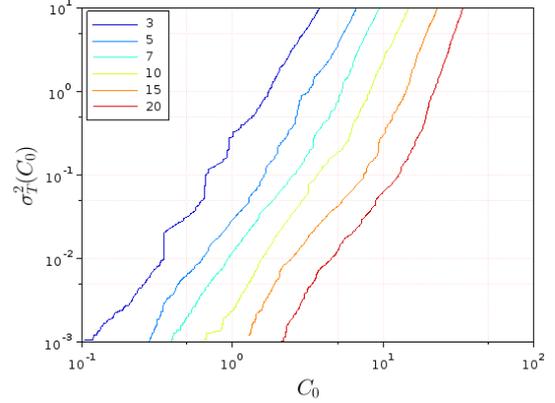


FIGURE 3 – EQM $\sigma_T^2(C_0)$ en fonction du seuil C_0 et de l'échelle d'observation $T \in \{3, 5, 7, 10, 15, 20\}$.

$T \setminus C_0$	$ \mathcal{T} $		EQM		$P_{FA} (10^{-3})$		$P_{MD} (10^{-3})$	
	0,1	1	0,1	1	0,1	1	0,1	1
3	308	352	0,025	0,46	1	0,6	0,6	3,4
5	163	200	0,037	0,071	1,5	0,9	0,7	1,3
7	96	129	0,068	0,042	1,9	1,3	0,9	0,9
10	47	75	0,176	0,061	3	1,8	1,6	0,8
15	17	37	0,645	0,168	6,1	2,9	3	1,4
20	6	19	2,02	0,414	15,4	4,7	4,2	2,3

TABLE 1 – Performances de l'estimateur.

et $\hat{b}(t_0) = s(t_0)$. Ainsi, les points t_0 sont d'autant plus favorables que l'EQM $\sigma_T^2(C_0)$ est faible. La proposition 1 et (10) nous disent que C_0 doit également être faible.

La figure 3 donne l'évolution de cette EQM en fonction du seuil C_0 pour plusieurs valeurs de l'échelle d'observation T . On observe bien une EQM qui décroît avec C_0 . Ces résultats ne permettent cependant pas de sélectionner un seuil car

1. Les EQM, si elles sont faibles, sont des valeurs moyennes qui présentent de fortes dispersions. La proposition 1 n'est alors vérifiée qu'en moyenne;
2. Le nombre de points favorables diminue quand C_0 diminue et conduit à une impossibilité à suivre correctement la ligne de base si ce nombre est trop faible ou si la répartition des points n'est pas satisfaisante.

À noter que, du fait de l'échantillonnage irrégulier généré par la sélection des points t_0 , une interpolation spline cubique conduit à des sur-oscillations limitant les performances de l'estimateur lorsque l'EQM est calculée sur l'ensemble des points de la ligne de base. L'interpolation linéaire est alors privilégiée.

Le tableau 1 donne quelques valeurs moyennes du nombre de points sélectionnés, de l'EQM calculée sur tous les points de la ligne de base, des taux FA et MD pour plusieurs échelles d'observation et valeurs C_0 . Contrairement aux résultats de la figure 3, il n'y a plus une simple décroissance de l'EQM, ou des P_{FA} et P_{MD} , avec l'augmentation de l'échelle d'observation pour une valeur de C_0 donnée car la mesure de l'erreur n'est plus seulement réalisée sur les points t_0 . Des valeurs de T entre 5 et 10, voire 15, pour C_0 égale à 0,1 ou 1 donnent des EQM et des taux FA et MD comparables.

Les performances de la méthode proposée sont ensuite

BEADS	airPLS	régression	proposition
0,080	0,97	1,8	0,071

TABLE 2 – EQM des estimateurs de ligne de base.

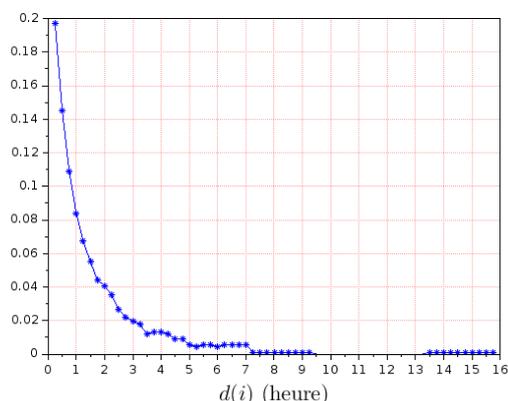


FIGURE 4 – Distribution de la distance à la phase humide $d(i)$.

comparées à celles obtenues avec les méthodes BEADS ($d = 1$, $f_c = 0,03037$, $r = 9,62$) [7], airPLS ($\lambda = 125577$) [11], où les paramètres optimaux ont été choisis, et de régression quantile [8], développées pour estimer les lignes de bases des spectres de spectroscopie RMN ou de chromatographie. Les valeurs données dans le tableau 2 sont les EQM calculées sur 55 réponses $s(i)$ de 9858 échantillons, $i \in [1, 9858]$. La puissance du bruit est $E[n(i)^2] = 0,01$, $C_0 = 1$ et $T = 5$. La méthode proposée paraît bien adaptée pour traiter le problème de dérive de la ligne de base des capteurs d'humectation. Sur ces données simulées, elle est la plus performante devant BEADS, airPLS puis la régression polynomiale sur quantile, avec des écarts entre les EQM allant de 0,009 à 1,73, cf. tableau 2.

Mesures expérimentales En pratique, une donnée de mesure est la moyenne de l'humectation sur 15 minutes. Si la réduction du bruit est alors suffisante, il reste des irrégularités locales à lisser. Une simple moyenne glissante est préférée au classique algorithme de Savitzky-Golay qui, en étalant les fronts abruptes, introduit des irrégularités supplémentaires en relation avec les coefficients négatifs du filtre et les fortes variations d'amplitude. Un filtre de longueur 10 est utilisé. L'ensemble de la procédure d'estimation est alors la suivante

1. Filtrage par moyenne glissante des données;
2. Sélection des échantillons $\mathcal{T}(10)$;
3. Reconstruction de la ligne de base par interpolation linéaire des échantillons sélectionnés.

En prenant les configurations du tableau 1 les plus favorables, on observe peu de différence sur les taux de FA, MD et sur la répartition $d(i)$, (5). La figure 4 donne un exemple de taux de répartition de $d(i)$, pour $S = 5$ sur un relevé de 8833 points de mesures, $C_0 = 1$ et $T = 10$. L'utilisation de la méthode proposée, comparée à un seuil fixe, conduit à environ 10 % de points de mesure humide en moins, soit 238 heures d'humectation en moins, ce qui indique en moyenne une période moins favorables aux maladies que ne le suggère le traitement par défaut.

4 Conclusion

La ligne de base des capteurs d'humectation foliaires varie en environnement opérationnel et biaise la mesure des périodes sèches et humides. Dans ce travail, nous avons proposé une méthode d'estimation de cette ligne de base qui sélectionne, sur la base de la norme L1, des minimums locaux et une échelle d'observation. L'estimateur est validé sur un modèle de ligne de base et ses performances sont comparées à celle des algorithmes de la littérature. La mise en pratique de l'estimateur nécessite de fixer deux paramètres : un seuil et une échelle d'observation. L'analyse menée montre une faible sensibilité de l'estimateur au choix de ces paramètres. La très faible complexité de l'algorithme et la faible profondeur de mémoire nécessaire permettent d'envisager des traitements *in situ* en temps réel dans les microcontrôleurs intégrés dans les nœuds des capteurs déployés pour la surveillance des cultures. Toutefois, la dépendance des performances de l'estimateur vis à vis du modèle n'est pas totalement exclue. Il est alors nécessaire de construire des bases de données expérimentales pour éprouver la robustesse de la proposition.

Références

- [1] R. MAGAREY, T. SUTTON et C. THAYER, « A simple generic infection model for foliar fungal plant pathogens », *Phytopathology*, vol. 95, no. 1, p. 92–100, 2005.
- [2] DECAGON DEVICES INC., *Dielectric Leaf Wetness Sensor*. Operator's Manual, 2016.
- [3] P. THIEBEAU et G. ALAVOINE, « Estimation de la quantité d'eau apportée par une rosée à l'aide de capteurs diélectriques de durée d'humectation », *Cahier des Techniques de l'INRA*, vol. 94, p. 1–16, 2018.
- [4] D. BERTRAND et É. DUFOUR, *La spectroscopie infrarouge et ses applications analytiques*. Sciences et techniques agroalimentaires, Lavoisier, 2e éd., 2006.
- [5] G. PEARSON, « A general baseline-recognition and baseline-flattening algorithm », *Journal of Magnetic Resonance (1969)*, vol. 27, no. 2, p. 265–272, 1977.
- [6] W. DIETRICH, C. RÜDEL et M. NEUMANN, « Fast and precise automatic baseline correction of one and two-dimensional nmr spectra », *Journal of Magnetic Resonance (1969)*, vol. 91, no. 1, p. 1–11, 1991.
- [7] L. DUVAL, A. PIRAYRE, X. NING et I. SELESNICK, « Suppression de ligne de base et débruitage de chromatogrammes par pénalisation asymétrique de positivité et dérivées parcimonieuses », in *Colloque GRETSI*, (Lyon, France), p. 1–4, sept. 2015.
- [8] L. KOMSTA, « Comparison of several methods of chromatographic baseline removal with a new approach based on quantile regression », *Chromatographia*, vol. 73, p. 721–731, avril 2011.
- [9] F. QIAN, Y. WU et P. HAO, « A fully automated algorithm of baseline correction based on wavelet feature points and segment interpolation », *Optics & Laser Technology*, vol. 96, p. 202–207, 2017.
- [10] V. MAZET, C. CARTERET, D. BRIE, J. IDIER et B. HUMBERT, « Background removal from spectra by designing and minimising a non-quadratic cost function », *Chemometrics and Intelligent Laboratory Systems*, vol. 76, no. 2, p. 121–133, 2005.
- [11] Z. ZHANG, S. CHEN et Y. LIANG, « Baseline correction using adaptive iteratively reweighted penalized least squares », *Analyst*, vol. 135, p. 1138–1146, mai 2010.