

Estimation de l'incertitude pour les réseaux de neurones : application pour la prédiction de mouvement d'éolienne en mer

Noura DRIDI¹ Lucas DRUMETZ² Adrien HIRVOAS³ Romain RIBAUT³

¹FEMTO-ST institut, Université de Bourgogne Franche-Comté, CNRS, ENSMM, Besançon, France

²IMT Atlantique, UMR CNRS 6285, Lab-STICC, F-29238 Brest, France

³France Energies Marines, Brest, France

Résumé – L'estimation de l'incertitude pour les Réseaux de Neurones (RN) est une étape importante pour les algorithmes d'apprentissage profond pour la régression et/ou la classification. Cette information supplémentaire sur l'incertitude attribue un niveau de confiance aux prédictions, rendant ainsi les décisions plus robustes. Le dropout a récemment été appliqué pour estimer l'incertitude des RN, en se basant sur le résultat que l'optimisation de la fonction de perte d'un RN avec dropout est équivalente à une approximation variationnelle d'un processus gaussien dans un cadre bayésien. Cette méthode a l'avantage d'un coût calculatoire faible par rapport à l'approche bayésienne classique. Nous proposons d'appliquer cette méthode pour la régression afin de prédire les mouvements de plateforme d'une éolienne flottante en mer (FOWT). Les résultats obtenus attestent de la bonne performance et de la pertinence de la mesure d'incertitude pour cette application industrielle.

Abstract – Estimating the uncertainty of Neural Networks (NN) is crucial in deep learning for both regression and classification tasks. This additional information about uncertainty assigns confidence to predictions, therefore making decisions more robust. dropout has recently been applied to estimate the NN uncertainty, based on the result that optimizing the loss function of an NN with dropout is equivalent to a Bayesian variational approximation of a Gaussian process. The main advantage of this approach compared to the classical Bayesian approach is the negligible additional computational cost. We propose to apply this method for regression in order to predict the platform motions of a Floating Offshore Wind Turbine (FOWT). The obtained results attest the good performance and relevance of uncertainty measure for this industrial application.

1 Introduction

Avec les problèmes environnementaux actuels, l'appel à des sources d'énergie renouvelable ne cesse d'augmenter. Dans ce contexte, des travaux sont menés pour développer et commercialiser la technologie d'éoliennes flottantes en mer (FOWT). Ces dernières produisent plus d'électricité par unité installée, profitant des vitesses de vent généralement plus constantes et plus élevées que sur terre. Un défi majeur est la maintenance prédictive de ces structures flottantes en mer. La surveillance des FOWT comprend le contrôle de l'état de santé des différentes composantes et la détection d'anomalies. Les deux processus sont liés à l'évaluation de l'intégrité des structures en fonctionnement grâce à une surveillance continue des mouvements du flotteur. En effet, superviser les différents mouvements du flotteur est un enjeu important qui peut prévenir les défaillances en fournissant des alertes précoces et permettre ainsi de meilleurs plans de maintenance. Dans ce travail, nous nous intéressons à la prédiction des mouvements de translation du flotteur sous différentes conditions de vitesse de vent. La surveillance en temps réel des structures FOWT peut être réalisée à l'aide d'approches basées sur les modèles ou les données. Les approches basées sur les modèles supposent une bonne connaissance du comportement et de la physique du système. Cependant, les FOWT sont définis par un système complexe à large échelle. Le fonctionnement d'une FOWT est défini par différents types de variables et de phénomènes physiques tels que l'hydrodynamique, l'aérodynamique, le servo-contrôle, etc. La non linéarité des signaux de vibration mesurés en raison des conditions opérationnelles (charges variables et des

vitesses fluctuantes) ajoute une difficulté supplémentaire. Dans [7], les auteurs illustrent la faible performance lors de l'application de méthodes de surveillance de l'état structurel (SHM) basées sur les propriétés modales à la structure FOWT en raison de plusieurs violations des hypothèses sur le système et les environnements. D'autre part, les approches basées sur les données telles que l'apprentissage automatique et l'apprentissage profond (AP) sont très efficaces dans diverses applications. Dans le contexte des éoliennes, on peut citer l'exemple de [2] qui se focalise sur la détection des dommages structurels sur les pales d'une éolienne en mer en utilisant des réseaux neuronaux récurrents tels que LSTM et Gated Recurrent Unit (GRU). Les méthodes d'AP apprennent directement le comportement du système à partir des données, contrairement à l'approche basée sur le modèle où une relation préalable est imposée entre les entrées et les sorties. Par ailleurs, pour la prédiction du mouvement du flotteur, étant donné la non linéarité des signaux, les approches type AP sont alors mieux appropriées pour apprendre la relation entrée/sortie.

Dans cet article, nous proposons une méthode AP pour prédire les mouvements de la plateforme FOWT à partir des vitesses du vent. En particulier la méthode fournit une mesure de l'incertitude sur les prédictions. En effet, en raison de la complexité du système, les prédictions intègrent des incertitudes inhérentes au système, à l'environnement et aux conditions opérationnelles. Par conséquent, mesurer l'incertitude associée à une estimation ponctuelle est crucial pour assister la décision. Par exemple, lorsqu'une anomalie est détectée avec un faible niveau de confiance, une vérification supplémentaire à l'aide

d'une autre source d'information est ajoutée (image, intervention humaine...) avant de planifier des actions de réparation. Plus généralement, une quantification fiable de l'incertitude de la prédiction permet de faire confiance ou non aux prévisions pour des entrées interpolant/extrapolant les situations vues dans l'ensemble d'entraînement. Dans ce travail, nous proposons d'évaluer l'incertitude de la prédiction en utilisant une méthode simple : le dropout. Nous proposons de suivre l'approche proposée dans [3] pour notre application. Nous proposons de prédire le mouvement de la plateforme à partir des données de vitesse du vent. L'architecture se base sur les réseaux de neurones convolutionnels (CNN) 1D qui sont plus adaptés aux données spectrales que nous avons, par rapport à un réseau neuronal récurrent ou un perceptron multicouche (MLP), car les fréquences sont fortement corrélées sans relation causale, contrairement au domaine temporel.

2 État de l'art

Il existe, dans la littérature, deux types d'incertitude : l'incertitude aléatoire inhérente aux données, et l'incertitude épistémique liée au modèle, comme pour un réseau de neurones entraîné NN. Dans ce qui suit, nous nous concentrons sur le deuxième type. L'incertitude du modèle peut être quantifiée à l'aide d'approches basiques telles que la régression linéaire ou les processus gaussiens. En apprentissage profond, les réseaux de neurones bayésiens fournissent un outil mathématique pour mesurer l'incertitude du modèle. L'idée est de remplacer les valeurs déterministes des paramètres du RN (poids et biais) par des distributions de probabilité avec un a priori gaussien, pour ensuite apprendre les paramètres de ces distributions. Une fois entraîné, les sorties du RN peuvent être évaluées plusieurs fois pour obtenir une distribution empirique des prédictions. Plusieurs approches ont été proposées pour les réseaux de neurones bayésiens, basées, par exemple, sur l'approximation de Laplace [5] ou des méthodes de Monte Carlo Hamiltonien [6]. Toutefois, ces deux approches considèrent généralement des architectures de réseaux de neurones simples pour éviter une complexité prohibitive. Pour les éoliennes terrestres, un modèle d'approximation avec quantification de l'incertitude est présenté dans [4]. En utilisant une approche basée modèle, les auteurs proposent de quantifier l'impact de l'incertitude liée aux paramètres du modèle sur le comportement de fatigue de l'éolienne. En outre une formulation bayésienne est proposée pour calibrer ces paramètres à l'aide de mesures in situ.

3 Présentation de la méthode

3.1 Description du système

Le système des éoliennes en mer est composé par les principaux éléments : les pales, la nacelle, la tour et le système de fondation (lignes d'ancrages, la plateforme flottante). Il existe différents types d'éoliennes en fonction de la capacité, de la structure de la plate-forme, des propriétés du système, etc. Dans cette étude, les données sont simulées à l'aide d'un modèle physique du système. Ils représentent l'éolienne FOWT OC3-Hywind¹ avec l'éolienne en mer de 5 MW de référence du National Renewable Energy Laboratory (NREL). L'éolienne est de type semi-submersible à trois pales, avec une capacité de 5 MW, maintenue dans l'eau par une bouée-spar appelé "Hywind", développé par la société Equinor.

¹OC3 désigne le projet Offshore Code Comparaison Collaboration.

3.2 Formulation du problème

L'objectif est de prédire les mouvements de la plateforme sous différentes conditions environnementales, cela permet de détecter d'éventuelles anomalies. En effet, si des mouvements dépassant le niveau de sécurité sont détectés, l'utilisateur est informé et des actions de maintenance sont planifiées. Les mouvements incluent 6 degrés de liberté (3 rotations et 3 translations). Dans ce genre de situations, il est crucial d'obtenir une approximation de l'incertitude du modèle liée à la prédiction donnée. Par exemple, pour les prévisions avec un faible niveau de confiance, une vérification additionnelle est ajoutée avant le plan d'intervention. Alternativement, il est possible d'utiliser une architecture simple et rapide du RN pour la prédiction, et une autre plus élaborée mais plus lente uniquement sur les entrées pour lesquelles l'architecture simple est incertaine.

3.3 Dropout comme approximation bayésienne

Le dropout consiste à supprimer aléatoirement un nœud d'un réseau avec toutes ses connexions entrantes et sortantes. Cette procédure peut être effectuée pour chaque neurone en entrée et chaque couche sauf la dernière. Une unité est supprimée avec une probabilité notée p_{drop} , fixe pour le passage forward et backward. Généralement, le dropout est utilisé pour éviter le surapprentissage. Ici, nous proposons de l'appliquer afin d'obtenir une mesure de l'incertitude de prédiction. Le travail est basé sur l'approche proposée dans [3], où les auteurs montrent que l'utilisation du dropout dans les réseaux de neurones peut être interprétée comme une approximation bayésienne d'un processus gaussien. Soit f la fonction inconnue reliant l'entrée X à la sortie Y , à apprendre par le réseau de neurones. $F|X$ et $Y|F$ sont modélisées par deux processus Gaussien :

$$F|X \sim \mathcal{N}(0, K(X, X)) \quad \text{and} \quad Y|F \sim \mathcal{N}(F, \tau^{-1}I_N)$$

Selon l'approche proposée, la minimisation de la fonction de perte d'un RN avec dropout équivaut à une approximation dans le cadre d'une approche bayésienne avec processus Gaussien. Pour la régression, la fonction de perte est définie par :

$$L = \frac{1}{N} \sum_{n=1}^N \|y_n - \hat{y}_n\|_2^2 + \lambda \sum_{i=1}^M (\|W_i\|^2 + \|b_i\|^2) \quad (1)$$

avec N le nombre d'observations, M le nombre de couches, W_i et b_i correspondent respectivement aux matrices de poids et les biais dans le RN. λ est un hyperparamètre du terme de régularisation utilisé pour éviter le surapprentissage. D'autre part, la distribution a posteriori $P(\omega|X, Y)$ (ω les paramètres du RN) est complexe, elle est alors optimisée par rapport à ω en utilisant l'approche variationnelle classique. La distribution cible est remplacée par une autre plus simple notée $q(\omega)$, et la divergence de Kullback-Leibler est minimisée entre $q(\omega)$ et $P(\omega|X, Y)$ ce qui est équivaut à la maximisation de la borne inférieure de l'évidence (ELBO). En outre, une distribution séparable est choisie pour $q(\omega)$, où les poids W et les biais b suivent la loi normale multivariée [3]. La distribution prédictive est enfin approchée par :

$$q(y^*|x^*) = \int P(y^*|x^*, \omega)q(\omega)d\omega \quad (2)$$

En suivant l'approche proposée dans [3], les moments d'ordre 1 et 2 de y^* sont estimées par la méthode de Monte Carlo,

en utilisant l'approximation de la distribution prédictive Eq.2.

Ainsi, la moyenne empirique égale : $E(y^*) \approx \frac{1}{T} \sum_{t=1}^T \hat{y}^*(x^*)$,

avec x^* et y^* désignant respectivement une entrée et une sortie, $\hat{y}^*(x^*)$ est la prédiction de x^* par le RN. T est le nombre de forward pass utilisé pour la prédiction. L'estimation de la variance est donnée par :

$$V(y^*) \approx \frac{I_D}{\tau} + \frac{1}{T} \sum_{t=1}^T \hat{y}^*(x^*) \hat{y}^*(x^*)^T - E(y^*) E(y^*)^T \quad (3)$$

avec D est la dimension de la sortie. Le second terme correspond à la variance empirique de l'estimateur et dans le premier τ égal à : $\tau = \frac{l^2(1-p_{drop})}{(2N\lambda)}$ avec λ le coefficient associé au terme de régularisation et l un facteur d'échelle liée à la distribution a priori des poids, $w \sim \mathcal{N}(0, l^{-2}I)$. L'éq. 3 est obtenue en utilisant la proposition D de l'annexe de [3]. τ représente la précision du modèle, sa valeur est faible en présence de données bruitées, ainsi λ est élevé pour régulariser par rapport aux données et vice-versa. Dans ce qui précède nous avons présenté le lien entre le dropout et l'approche variationnelle pour justifier la base théorique du dropout comme mesure de l'incertitude. Plus précisément, nous avons détaillé lien entre la minimisation de la fonction perte d'un RN avec dropout et l'approche bayésienne où l'incertitude est mesurée par le moment d'ordre deux de la distribution prédictive. Finalement, il est important de signaler que l'approche est simple à implémenter, indépendante de l'architecture du RN et n'ajoute aucun coût calculatoire.

4 Résultats de simulation

4.1 Présentation de données

Dans ce travail, le mouvement est limité à la translation selon l'axe x , appelé surge, en fonction de la vitesse du vent. Les données sont générées par deux logiciels :

-Turbsim : simule un champ de vent turbulent complet. Il génère des champs de vent tridimensionnels réalistes, décrivant le vent longitudinal, transversal et vertical.

-OpenFast : est un simulateur basé sur la physique des éoliennes. C'est un outil multi-physique, multi-niveau de fidélité pour simuler la réponse dynamique couplée des éoliennes. Les deux logiciels sont développés par le National Renewable Energy Laboratory (NREL) ² et sont gratuits et libres. Les données sont obtenues en trois étapes. Tout d'abord, la vitesse moyenne du vent est fixée au niveau du hub de l'éolienne et le champ de vent est généré par Turbsim. Ce dernier est modélisé par un processus Gaussien de moyenne et écart-type fixe. Ensuite, le champ de vent est introduit comme entrée dans OpenFast pour obtenir différentes séries temporelles liées au comportement de l'éolienne. Le déplacement de translation (surge) est l'unique mouvement pertinent en l'absence de variations de vagues et de courants. Troisièmement, les logarithmes des densités spectrales de puissance (DSP) des séries temporelles sont calculés, pour obtenir les données d'entrée/sortie du RN. L'ensemble des données est composé de $N = 200$ spectres des champs de vent, pour des vitesses allant de 3 à 22 [m/s], avec 10 réalisations pour chaque vitesse. Un spectre est formé d'un signal à trois composantes le long des axes x , y

²<https://openfast.readthedocs.io/download/en/main/pdf/>

et z , de longueur 600 correspondant aux fréquences mesurées en hertz [Hz], ainsi la dimension des données d'entrée est de (200, 600, 3). Les données de sortie représentent le surge de la plateforme, le signal est de dimension (200, 600, 1).

4.2 Performance de prédiction

Une architecture de type CNN 1D est choisie pour le RN, avec trois couches précédées chacune par une couche dropout et deux couches denses. Les hyperparamètres sont fixés comme suit : taux d'apprentissage $lr = 10^{-4}$, poids de régularisation $\lambda = 5.10^{-5}$, probabilité de dropout $p_{drop} = 0,5$ (nous avons testé pour des valeurs plus faibles mais les performances sont les mêmes), taille du batch = 8, nombre d'époques = 10^4 . Étant donné que nous avons des données à haute fréquence, le paramètre l défini dans le facteur τ de l'Eq.3 est fixé à une valeur faible $l = 0,1$, ceci permet de réduire la régularisation sur les poids du RN et augmenter ainsi leur ordre de grandeur afin d'identifier les caractéristiques de ces données haute fréquence [3]. 80% des données sont utilisées pour entraîner le modèle et 20% pour le test. Notons ici que la répartition est faite de telle sorte que nous n'avons pas les mêmes vitesses du vent dans la partie entraînement et test pour montrer la capacité de généralisation de l'algorithme. L'algorithme proposé est appelé CNN-Dropout. Afin de valider la pertinence du choix de l'architecture CNN, l'algorithme est comparé avec le MLP (avec deux couches cachées) et un RN basé sur la régression linéaire appelé RN-LR. Ce dernier inclut une seule couche cachée, à l'entrée une vectorisation du champ de vent est considérée, ce qui explique le nombre élevé des paramètres, voir tableau 1. Pour une comparaison correcte, les mêmes hyperparamètres sont fixés pour les trois algorithmes avec un choix de l'architecture du MLP donnant un nombre comparable de paramètres à ceux du CNN. Une fois le réseau entraîné, une prédiction du surge est calculée en moyennant sur $T = 1000$ échantillons. En outre, une comparaison avec une méthode non basée sur les RN est présentée, il s'agit de la Régression Bayésienne Généralisée (RBG). Chaque fréquence est considérée comme une fonction des trois composantes de la vitesse du vent le long de trois axes. Le modèle reliant l'entrée/sortie est donné par $y = \phi(x) + \varepsilon$ où $\phi(x) = (1, x, x^2, x^3)$ est la projection de données dans un espace de dimension plus élevée pour obtenir une régression polynomiale. La distribution prédictive est calculée et utilisée pour générer plusieurs prédictions et calculer l'intervalle de confiance [1]. Dans un premier temps, les performances de prédiction sont évaluées. La racine de l'erreur quadratique moyenne (REQM) est calculée entre la vraie DSP et celle prédite. La moyenne et l'écart type sont calculés sur les ensembles de données d'entraînement/test. Le

Algorithme	Nb paramètres	REQM moyen	Ecart-type-REQM	Temps(s)
RN-RL	1.080.600	0.32/0.69	0.13/0.48	202
MLP	69416	0.36/0.55	0.10/0.35	221
CNN-Dropout	68072	0.26/0.39	0.04/0.21	298
RBG	-	0.49/0.85	0.20/1.13	-

TABLE 1 : Performance de prédiction.

tableau 1 présente le nombre de paramètres entraînés par le RN, la moyenne et l'écart-type de la REQM et le temps de calcul en

secondes. La RBG est moins performante que les algorithmes de RN, car elle est limitée par une relation a priori entre l'entrée et la sortie. Pour les ensembles d'entraînement et de test, le CNN-Dropout offre des meilleures performances par rapport aux algorithmes RN-RL, MLP et RBG en termes de moyenne et d'écart-type de la REQM, et avec moins de paramètres. Cela atteste de l'efficacité de l'architecture CNN pour apprendre automatiquement les principales caractéristiques, en effet la corrélation entre les bandes de fréquences est prise en compte, permettant un apprentissage efficace, contrairement au RN-RL où on suppose qu'il y a une relation linéaire entrée/sortie. Pour les données de test, le MLP généralise mieux que le RN-RL, mais il est moins performant que le CNN-Dropout car il n'est pas adapté à la structure des données considérées. De plus, pour l'application actuelle, le surapprentissage est expliqué par la faible vitesse du vent. En effet, pour des telles vitesses, le mouvement est limité et il est plus difficile de le prédire. La figure 1 illustre la moyenne et l'écart-type de la REQM en fonction de la vitesse du vent, comme observé pour une vitesse de vent $> 5[m/s]$, l'erreur augmente légèrement entre l'entraînement et le test. L'algorithme fournit

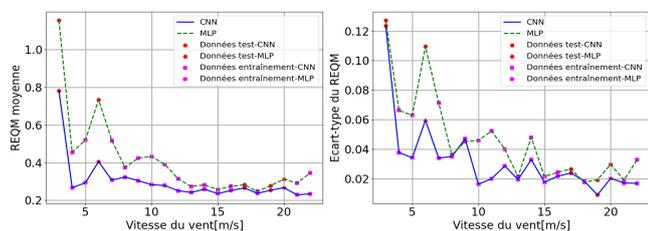


FIGURE 1 : Moyenne et écart-type du REQM en fonction de la vitesse du vent.

également une quantification de l'incertitude des prédictions. Pour attester la pertinence de l'approche, nous considérons un scénario avec 50% des données pour l'entraînement (vitesse du vent de 13 à 22[m/s]) et 50% pour le test (vitesse du vent de 3 à 12[m/s]). Ce choix est justifié par le fait que plus la vitesse est faible moins le flotteur bouge et donc plus c'est difficile pour l'algorithme de capter le mouvement. Dans ce cas, l'écart de performance entre entraînement et test est plus remarquable. Deux exemples sont présentés choisis de l'ensemble d'entraînement fig.2-(b) et test fig.2-(a). La vraie DSP et la prédite DSP sont présentées avec l'intervalle de largeur deux fois l'écart-type. Dans la figure 2-(c) l'incertitude est présentée par la moyenne de la variance des prédictions sur l'ensemble des fréquences pour la partie apprentissage et test. Comme illustré sur les figures 2-(a) et 2-(b), la DSP est prédite

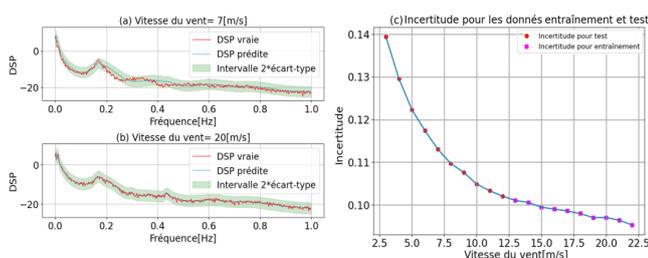


FIGURE 2 : DSP et représentation de l'incertitude.

dans l'intervalle de confiance. En outre, le CNN-Dropout est capable de capter les principales caractéristiques du signal :

les prédictions ont le même pic que la vraie DSP. Ceci est important, car les pics de la DSP se réfèrent au mouvement de quelques composantes principales de l'éolienne (plateforme, rotor, tour, etc.)³. Enfin, comme observé sur la figure 2-(c), l'incertitude est plus élevée pour la partie test et diminue en s'approchant de l'ensemble d'entraînement. Cette information pertinente alerte l'utilisateur sur une prédiction avec un faible niveau de confiance.

5 Conclusion

Dans cet article, nous avons proposé une méthode basée sur l'AP pour modéliser et prédire le mouvement de la plateforme flottante d'une éolienne en mer. L'architecture CNN est utilisée pour prédire le mouvement de translation surge en fonction de la vitesse du vent. Par ailleurs, une mesure de l'incertitude et du niveau de confiance des prédictions est fournie. En effet, nous avons utilisé une interprétation probabiliste de la technique de dropout dans le RN, qui relie l'approche bayésienne avec processus gaussien à la fonction de perte du RN avec dropout. Les résultats montrent une amélioration par rapport aux algorithmes de référence. Parmi les perspectives, nous allons étudier avec plus de détails la relation entre l'incertitude et la vitesse du vent. Pour le RN, nous explorerons des architectures plus élaborées et d'autres méthodes pour quantifier l'incertitude afin d'améliorer la capacité de généralisation et de fournir des incertitudes plus précises. Enfin, nous prévoyons d'appliquer la méthode sur des données réelles collectées à partir de l'éolienne en mer installée au large de la Norvège.

Références

- [1] C.M. BISHOP : Pattern recognition and machine learning (information science and statistics). *Springer*, 2007.
- [2] D. E. CHOE, H.C. KIM et M. H. KIM : Sequence-based modeling of deep learning with LSTM and GRU networks for structural damage detection of floating offshore wind turbine blades. *Elsevier journal on renewable Energy*, 174:218–235, 2021.
- [3] Y. GAL et Z. GHAHRAMANI : Dropout as a bayesian approximation : Representing model uncertainty in deep learning. *ICML*, 48:1050–1059, 20–22 Jun 2016.
- [4] A. HIRVOAS, C. PRIEUR, E. ARNAUD, F. CALEYRON et M. M. ZUNIGA : Quantification and reduction of uncertainties in a wind turbine numerical model based on a global sensitivity analysis and a recursive bayesian inference approach. *International Journal for Numerical Methods in Engineering*, 122:2528–2544, 2021.
- [5] D.J.C. MACKAY : A practical bayesian framework for back-propagation networks.neural computation. *Neural computation*, 4(3), 1992.
- [6] R.M. NEAL : Bayesian learning for neural networks. *PhD thesis, University of Toronto*, 1995.
- [7] D. TCHERNIAK, S. CHAUHAN et M. H. HANSEN : Applicability limits of operational modal analysis to operational wind turbines. *Proceedings of the Society for Experimental Mechanics Series*, 1:317–327, 2011.

³www.nrel.gov/publications/2013